Università degli Studi di Milano Bicocca

Facoltà di Scienze Matematiche Fisiche e Naturali

Scuola di Dottorato di Scienze
Corso di Dottorato di Ricerca in Fisica ed Astronomia
Settore Scientifico Disciplinare FIS/04

Study of $\gamma\gamma b\bar{b}$ final state topologies at LHC and search for high mass resonances decaying into two Higgs bosons with the CMS detector.

Coordinatore della Scuola di Dottorato: prof. **Giuseppe Chirico**
Relatore: prof. **Tommaso Tabarelli de Fatis**

Tesi di dottorato di
Badder Marzocchi
matricola n° 709790

Ciclo XXVIII Anno Accademico 2015/2016

Abstract

The discovery of the Higgs boson at the LHC completes the standard model (SM) of particle interactions. Albeit very successful, the SM does not provide answers to critical questions, such as the nature of dark matter or the hierarchy problem. For this reason, theories that predict the existence of new phenomena beyond the standard model (BSM theories) have been proposed. Many BSM theories predict the existence of new particles coupled to the Higgs boson. Therefore, the consistency of the SM and the viability of these BSM theories can be probed indirectly through precision measurements of the Higgs boson properties or directly through the search for an anomalous production of Higgs bosons. These complementary approaches share to a large extent the same set of experimental procedures to reconstruct the Higgs boson from its final state products.

In this thesis, we report on the search for production of Higgs boson pairs at the LHC with the CMS detector, where the two Higgs bosons decay into a photon pair and into two b quarks. This final state combines the advantage of high signal efficiency and high signal to background ratio. It also leverages in the excellent performance of the electromagnetic calorimeter (ECAL) and of the silicon tracker of CMS. An anomalous resonant production of Higgs boson pairs would indicate of the existence of new high-mass boson. Among BSM theories, models with warped extra dimensions (WED), which address specifically the hierarchy problem of the SM, provide boson candidates such as the radion (spin 0) or the graviton (spin 2) with masses that can be probed at the LHC. Among the possible di-Higgs resonant searches (e.g. HH$\rightarrow$4b or HH$\rightarrow \tau\tau$bb), the $\gamma\gamma$bb decay channel is the most sensitive for resonant candidates with low mass, $m_X < 400$ GeV. The resonant di-Higgs decays allow a very good rejection of spurious events due to non-resonant production of hadronic jets and photons. Upon optimization of the diphoton invariant mass resolution, of the efficiency to photons ($\approx$90%), and of the identification of jets coming from b-quarks ($\approx$60%) the sensitivity achieved with 19.7 fb^{-1} of data collected in pp collisions at a centre-of-mass energy of 8 TeV permitted to exclude a relevant fraction of the parameter space of models that predict the existence of a radion. No significant excess of events is observed above the expected background in the search mass window of 260 and 1100 GeV, and an upper limit on the cross-section from 3.30 to 0.31 fb is placed. The sensitivity of this search is expected to improve at the centre-of-mass energy of 13 TeV of the LHC Run II.

The data analysis has required a detailed preparatory work to understand and optimize the photon reconstruction and selection. An account of this experimental work is presented in the report, with detailed discussions of the calibration method of the ECAL, and of the optimization of the simulation of the ECAL response. In particular, the optimization of a method exploiting the azimuthal symmetry of the energy flow in pp collisions, used in the equalization of the ECAL response across the different readout channels, is presented. The photon energy calibration and identification rely to a large extent on the Monte Carlo simulation of ECAL response, which in reality is subject to variation in time due to radiation-induced damage to the crystals and photodetectors. To account for this, response variations measured from data have been modelled and appropriately propagated to the MC simulation. The adoption of this time-dependent simulation has contributed substantially also to improve the CMS sensitivity of the analysis that enabled the standalone observation of the Higgs bosons through the H$\rightarrow \gamma\gamma$ decay, and the measurement of the Higgs boson mass with a systematic uncertainty of of 0.12%. This precision makes the uncertainty of the Higgs boson mass subleading in precision tests of the SM.

CONTENTS

CHAPTER 1

THE STANDARD MODEL AND BEYOND

In this Chapter the basic concepts of the model describing the elementary particles and their interactions, the standard model (SM), are introduced. The electroweak spontaneous symmetry breaking and the Higgs mechanism are then described as the way to explain the origin of particles masses. An overview of the Higgs boson production, decay mechanisms and precision measurement of its properties at the LHC is given. However, as it is, the SM is incomplete since observed physical phenomena such as the gravity or the dark matter masses cannot be described by it. Therefore, a description of possible models beyond the SM that can solve some of these issues is given in view of searches for new massive resonant particles. In particular, in the context of hadron colliders, the search for resonant di-Higgs production is a very sensitive channel to probe the existence of massive particles beyond the SM ones.

1.1 Theoretical introduction

In order to describe the observed phenomena, four forces are believed to be enough and fundamental, i.e. directly associated with matter properties. The forces are: the electromagnetic, strong, weak and gravitational force. The SM explains how the particles interact, through the fundamental forces. Unfortunately, by now we are not able to include in the SM the gravitational force.

The background of the SM is the relativistic quantum field theory. Basing on their spin, all particles are split into two groups: fermions and bosons. Fermions have half-integer spin, follow the Fermi-Dirac statistics and are the constituents of matter. On the other hand, bosons have integer spin, follow the Bose-Einstein statistics and can be seen as the intermediates of the interaction forces. Tables 1.1 and 1.2 show the fermions and bosons described by the SM and their main properties.

The matter is composed by two types of spin-1/2 constituents: leptons and quarks. Besides the masses and the electromagnetic charges, quarks and leptons differ for the couplings, i.e. quarks have additional charges, the colours: r, b, g, $\bar{r}$, $\bar{b}$ and $\bar{g}$. Because of the colour charges, quarks

Table 1.1: Spin-1/2 fermions [1] masses and charges

	1^{st} **family**	2^{nd} **family**	3^{rd} **family**	**Q**	**Colour Charge**
Leptons	ν_e (~ 0)	ν_μ (~ 0)	ν_τ (~ 0)	0	0
	e (511 keV/c^2)	μ (105.7 MeV/c^2)	τ (1.777 GeV/c^2)	-1	0
Quarks	**u** (1.7-3.1 MeV/c^2)	**c** ($1.29^{+0.05}_{-0.11}$GeV/c^2)	**t** ($172.9^{+1.1}_{-1.1}$GeV/c^2)	1/3	$r,\ b,\ g$
	d (4.1-5.7 MeV/c^2)	**s** (100^{+30}_{-20}MeV/c^2)	**b** ($4.19^{+0.18}_{-0.01}$GeV/c^2)	-2/3	$r,\ b,\ g$

Table 1.2: Spin-1 bosons [1] masses and charges

	Mass (GeV/c^2)	**Q**	**Colour Charge**
Photon (γ)	0	0	0
Gluon (g)	0	0	$r,\ b,\ g$
W	80.385 ± 0.015	±1	0
Z^0	91.188 ± 0.002	0	0

cannot be detected alone as affected by the asymptotic freedom, i.e. the coupling is asymptotically weaker as energy increases and distance decreases. Hence, only their bound states can be detected: mesons ($q\bar{q}$) and baryons (qqq).

A quantum field is associated to each particle, based on the group of symmetries $SU_C(3) \times SU_L(2) \times U_Y(1)$. The three local symmetry groups dictate the three interactions between the particles. $SU_C(3)$ is related to the colour charges, while $SU_L(2)$ is related to the chirality. Each quark can be seen as a triplet under $SU_C(3)$ symmetry due to the colour charges, whereas each lepton is a colourless singlet. The chirality is associated to the properties of the bispinors, i.e. representation of the fermions as solution of the Dirac equation:

$$\mathcal{L} = \bar{\psi}(i\partial\!\!\!/ - m)\psi \longrightarrow (-i\partial\!\!\!/ + m)\psi = 0 \tag{1.1}$$

Indeed, a bispinor can be separated as: $\psi = \psi_R + \psi_L$, where ψ_R (right component) and ψ_L (left component) are the eigenstate of chirality, γ_5. In the Dirac representation:

$$\psi_L = \frac{1-\gamma_5}{2}\psi$$

$$\psi_R = \frac{1+\gamma_5}{2}\psi$$

Right bispinors and left bispinors interact in different ways. In fact, given the $SU(2)$ symmetry it is possible to identify the pairs of the left-handed fermions as isospin doublets, while right-handed fermions are isospin singlets. The last symmetry, $U_Y(1)$ is related to the weak hypercharge.

The isospin and the hypercharge are quantum numbers associated with the electroweak interactions between the fermions and the bosons. Three bosons are associated with the electroweak force: W$^\pm$ (charged), Z^0 and photon (A$^\mu$ field).

The coupling between the fermions and the charged bosons (charged currents) is associated to the vectorial components of the isospin. On the other hand, the neutral currents are generated by the interactions between the fermions and a mixing of photon and Z fields [2], where

the couplings are given by the third component of the isospin and the hypercharge. The mixing between photons and Z bosons is performed by means of the Weinberg angle ($\sin^2\theta_W^{M_Z} = 0.23102\pm0.00005$ [1]). The relation between the electric charge Q, the third component of the isospin (T_3) and the hypercharge (Y) is:

$$Q = T_3 + \frac{Y}{2}$$

In order to find out the Lagrangian terms that describe the currents, the gauge invariance principle is used. Consider a Dirac field ψ and the local gauge transformation:

$$\psi \rightarrow e^{ig\alpha(x)}\psi$$

The Lagrangian of Equation 1.1 is invariant under this gauge transformation only with the introduction of a gauge boson field (A^μ) which allows the following gauge transformation:

- $A_\mu \rightarrow A_\mu$ - $\partial_\mu\alpha(x)$
- $\partial_\mu\psi \rightarrow \partial_\mu\psi$ + $igA_\mu\psi$

Hence, using a representation of the $SU_L(2)$ symmetry and the gauge invariance, after the definition of:

$$L^\ell \equiv \begin{pmatrix}\psi^\nu\\ \psi^\ell\end{pmatrix}_L \ , \ R^\ell \equiv \begin{pmatrix}\psi^\nu\\ \psi^\ell\end{pmatrix}_R$$

$$\tau^\pm = \frac{\sigma_1 \pm i\sigma_2}{2}$$

where σ_i are the Pauli matrices, generators of SU(2) symmetry, the Lagrangian terms for the charged currents becomes:

$$\mathcal{L}_{CC} = \frac{g}{\sqrt{2}}\left[\bar{L}^\ell \not{W}^+ \tau^+ L^\ell + \bar{L}^\ell \not{W}^- \tau^- L^\ell\right]$$

Concerning the terms for the neutral currents, after the definition of:

$$\Psi^\ell = \begin{pmatrix}\psi^\nu_L\\ \psi^\ell_L\\ \psi^\nu_R\\ \psi^\ell_R\end{pmatrix} \ , \ T_3 = \begin{pmatrix}1/2 & & & \\ & -1/2 & & \\ & & 0 & \\ & & & 0\end{pmatrix} \ , \ Y = \begin{pmatrix}Y_{\nu_L} & & & \\ & Y_{\ell_L} & & \\ & & Y_{\nu_R} & \\ & & & Y_{\ell_R}\end{pmatrix}$$

they can be decoupled between electromagnetic and weak interaction:

$$\mathcal{L}_A = eQ\bar{\Psi}^\ell \not{A}\Psi^\ell$$

$$\mathcal{L}_Z = \frac{e}{2\sin\theta_W\cos\theta_W}\bar{\Psi}^\ell \not{Z}(g_V - g_A\gamma^5)\Psi^\ell$$

where:

- $Q \equiv T_3 + \frac{Y}{2}$
- $g_V \equiv T_3 - 2Q\sin^2\theta_W$
- $g_A \equiv T_3$

Table 1.3: Left: leptons quantum numbers; Right: quarks quantum numbers

	T	T_3	Y	Q		T	T_3	Y	Q
ν_L^ℓ	$\frac{1}{2}$	$\frac{1}{2}$	-1	0	q_L^{up}	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$
ℓ_L	$\frac{1}{2}$	$-\frac{1}{2}$	-1	-1	q_L^{down}	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{3}$	$-\frac{1}{3}$
ν_R^ℓ	0	0	0	0	q_R^{up}	0	0	$\frac{4}{3}$	$\frac{2}{3}$
ℓ_R	0	0	-2	-1	q_R^{down}	0	0	$-\frac{2}{3}$	$-\frac{1}{3}$

It is possible to generalize to all the leptons and quarks families, using the values of isospin, hypercharge and electromagnetic charges of Table 1.3. Since ν_R has all the quantum numbers equal to zero it doesn't interact.

The mass terms:

$$m_W^2 W^{+\mu} W_\mu^- + \frac{1}{2} m_Z^2 Z^\mu Z_\mu + \sum_f m_f(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R)$$

cannot be put directly in the Lagrangian because they are not gauge invariant. Maintaining the gauge invariance, a way to give mass to these particles is the Higgs mechanism [3], which introduces the last particle in this theory: the Higgs boson, which is a massive, chargeless, spin-0 particle. The Higgs boson is an elementary particle which was proposed at the beginning of the 1960s in order to explain in a very simple way how all the massive particles gain the mass.

1.2 The Higgs mechanism

Consider a complex scalar field ϕ, a vector field A_μ and the Lagrangian:

$$\mathcal{L} = D^\mu \phi^* D_\mu \phi - \mu^2 |\phi|^2 - \lambda |\phi|^4 - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \tag{1.2}$$

with $D_\mu \phi = [\partial_\mu + iqA_\mu]\phi$ and $V(x) \equiv \mu^2 |\phi(x)|^2 + \lambda |\phi(x)|^4$.

Because ϕ is a complex field it can be decomposed in two fields ϕ_1 and ϕ_2. Indeed, the potential $V(x)$ has very different physical meanings depending on μ^2 and λ signs. In order to have a below limit the potential must have λ positive. If $\mu^2 > 0$ the potential has only one minimum in zero, it becomes the Klein-Gordon potential with an extra auto-interaction term and μ^2 can be interpreted as the square mass of the particle. On the other hand, if $\mu^2 < 0$, $V(x)$ has a maximum in zero and minima in $\phi = \sqrt{\frac{-\mu^2}{2\lambda}} e^{i\theta}$; thus, it is not possible to use the perturbation theory in $\phi = 0$. Figure 1.1 shows the potential for $\mu^2 < 0$.

It is possible to expand the field on the minima-circle, choosing one specific minimum and breaking the θ-symmetry of the potential (Spontaneous Symmetry Breaking). Fixing one minimum, v, expanding and inserting this expansion in the Lagrangian 1.2, it becomes:

$$\phi(x) = \frac{1}{\sqrt{2}}[v + h(x) + i\eta(x)]$$

$$\mathcal{L} = \frac{1}{2}\partial^\mu h \partial_\mu h(x) + \frac{1}{2}\partial^\mu \eta(x) \partial_\mu \eta(x) - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \frac{1}{2}(qv)^2 A_\mu(x) A^\mu(x) - \frac{1}{2}(2\lambda v^2) h^2(x) +$$

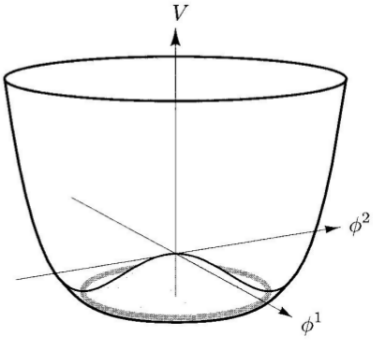


Figure 1.1: Higgs boson potential if $\mu^2 < 0$

$$+vqA^{\mu}(x)\partial_{\mu}\eta(x) \; + \quad \text{interaction terms}$$

This Lagrangian is inconsistent since we begin with ϕ and A^{μ}, thus with 4 degrees of freedom, and after the spontaneous symmetry breaking, we arrive to 5 degrees of freedom: a massless real scalar field $\eta(x)$ (Goldstone boson), a massive real scalar field $h(x)$ and a massive vector fields $A^{\mu}(x)$. By means of the invariance under $U(1)$, it is possible to perform a gauge transformation which eliminates the non physical field, $\eta(x) = 0$:

$$\mathcal{L} = \frac{1}{2}\partial^{\mu}h(x)\partial_{\mu}h(x) - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{2}(2\lambda v^2)h^2(x) + \frac{1}{2}(qv)^2 A_{\mu}(x)A^{\mu}(x) \; + \quad \text{interaction terms}$$

Now there are one massive vector field $A^{\mu}(x)$ with mass qv and a massive scalar field $h(x)$ with mass $\sqrt{2\lambda v^2}$, therefore 4 degrees of freedom.

In order to apply this mechanism to SM, it is necessary to introduce an Higgs boson doublet $\Phi(x)$, because of the SU(2) symmetry. With the Lagrangian of Equation 1.2 it is possible to give mass only to the bosons. In order to give mass to the fermions the additional Yukawa interaction term has to be added:

$$\mathcal{L}_{YK} = -g_{\ell}[\bar{L}^{\ell}\psi_R^{\ell}\Phi + \Phi^{\dagger}\bar{\psi}_R^{\ell}L^{\ell}]$$

Hence, the total Lagrangian to which apply the spontaneous symmetry breaking mechanism is:

$$\mathcal{L} = [D^{\mu}\Phi]^{\dagger}[D_{\mu}\Phi] - \mu^2\Phi^{\dagger}\Phi - \lambda[\Phi^{\dagger}\Phi]^2 -$$

$$-g_{\ell}[\bar{L}^{\ell}\psi_R^{\ell}\Phi + \Phi^{\dagger}\bar{\psi}_R^{l}L^{\ell}] +$$

$$+i[\bar{L}^{\ell}\not{D}L^{\ell} + \bar{\psi}_R^{\ell}\not{D}\psi_R^{\ell} + \bar{\psi}_R^{\nu}\not{D}\psi_R^{\nu}] \tag{1.3}$$

Now, if $\lambda > 0$ and $\mu^2 < 0$, $< 0|\Phi|0 > \neq 0$, it is possible to expand Φ near the minimum v as:

$$\Phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} \eta_1(x) + i\eta_2(x) \\ v + h(x) + i\eta_3(x) \end{pmatrix}$$

where $\eta_1(x)$, $\eta_2(x)$ and $\eta_3(x)$ are Goldstone bosons. After the expansion, all the $SU(2) \times U(1)$ symmetry but $U_{EM}(1)$ is broken, since the photon is massless and cannot interact with the Higgs boson. As previously done, it is possible to cancel all the Goldstone bosons performing gauge transformation, taking to:

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

Inserting this in the Lagrangian of Equation 1.3, interactions terms between Higgs Boson, fermions and gauge bosons appear, as well as the desired mass terms, where:

- $m_\gamma^2 = 0$
- $m_\ell^2 = \dfrac{g_\ell v}{\sqrt{2}}$
- $m_W^2 = \dfrac{g^2 v^2}{4}$
- $m_Z^2 = \dfrac{v^2 e^2}{\sin^2\theta_W \cos^2\theta_W}$
- $m_H^2 = 2\lambda v^2$

1.3 Higgs boson production and decay modes at the LHC

The SM Higgs boson production cross-section at a p-p hadron collider of center-of-mass energy of $\sqrt{s} = 8$ TeV [4] is shown in Figure 1.2. For a Higgs boson mass of 125 GeV, the total production cross section is expected to be $\sigma = 17.5$ pb at $\sqrt{s} = 7$ TeV and $\sigma = 22.3$ pb at 8 TeV. In Figure 1.3 the corresponding leading-order production Feynman diagrams are shown.

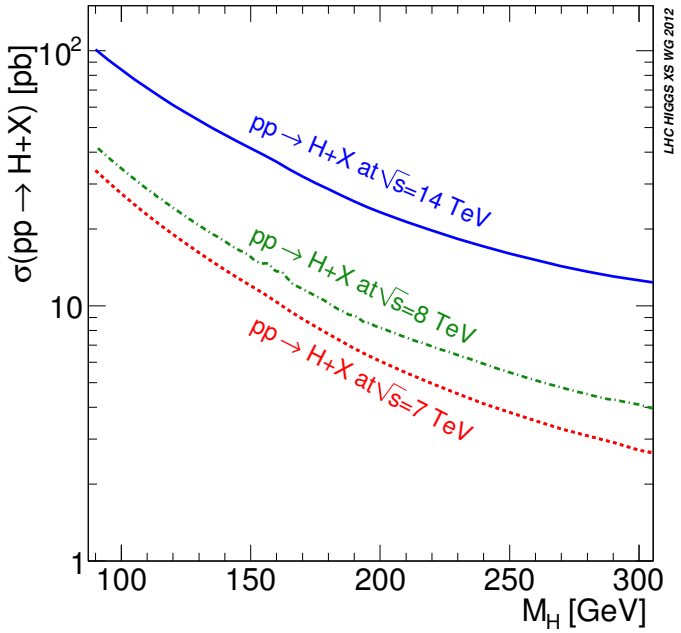


Figure 1.2: Higgs production cross-sections at $\sqrt{s} = 7$, 8 and 14 TeV as a function of the Higgs boson mass. NNLO QCD corrections as well as NLO EKW corrections are taken into account [4].

The **gluon fusion** ($gg \to H$) is the dominant Higgs boson production process over the entire mass range accessible at the LHC. It proceeds with a heavy quark triangle loop, as shown in Figure 1.3. Because of the Higgs boson couplings to fermions, the t-quark loop is the most important.

The second largest production mechanism of the Higgs boson is by means of **vector boson**

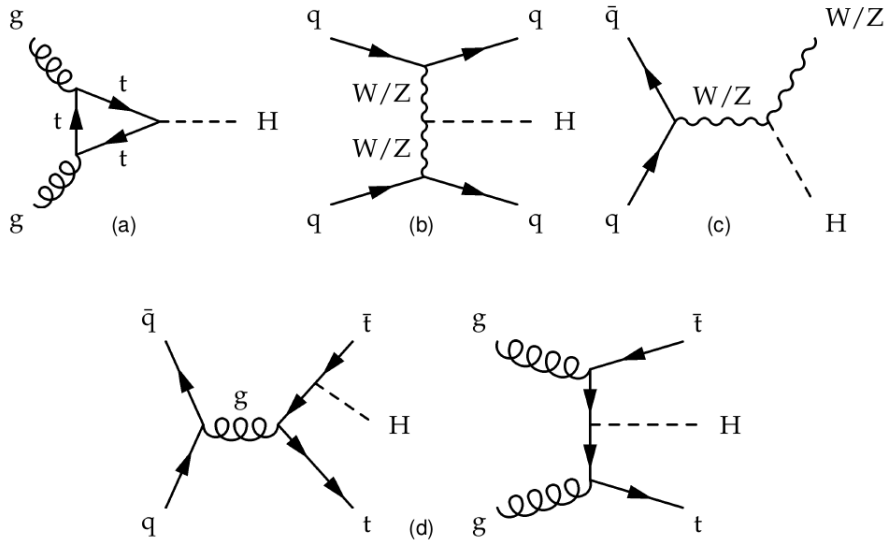


Figure 1.3: Feynman diagrams for the most important LO production processes of the SM Higgs boson: (a) gluon fusion, (b) vector boson fusion, (c) Higgs-strahlung, (d) $t\bar{t}$ associated production.

fusion (VBF,qq $\rightarrow$ qqH). In this process the Higgs boson is originated from the fusion of two weak bosons radiated off the incoming quarks. Their hadronization produces two forward jets of high invariant mass which can be used to tag the event and differentiate it from backgrounds.

In the **Higgs-strahlung** ($q\bar{q}' \rightarrow$ WH, $q\bar{q} \rightarrow$ ZH) and $t\bar{t}$ **associated production** (gg, $q\bar{q} \rightarrow t\bar{t}$H) processes the Higgs boson is produced in association with a W/Z boson or a pair of t quarks. In both cases, their decay products can be used to tag the event.

Depending on the Higgs boson mass, different decay channels can be exploited to detect the particle. The Higgs boson total decay width and its different decay branching ratios depend on the Higgs couplings to the vector bosons and to the fermions in the SM Lagrangian of Equation 1.3. Due to the dependence of Higgs couplings on the particle masses, the Higgs boson tends to decay into the heaviest particles which are kinematically allowed. Figure 1.4a shows the Higgs boson decay branching ratios including also NLO QCD and EWK corrections. Light-fermion decay modes contribute only in the low mass region (up to $\sim$ 150 GeV/c^2). Once the decay into a pair of weak bosons is possible, it quickly dominates. The Higgs boson does not couple to photons and gluons at tree level, but such couplings can arise via fermion loops and they give a contribution in the low mass region. The total width, given by the sum over all the possible decay channels, is shown in Figure 1.4b. It quickly increases with the Higgs boson mass due to the opening of new channels.

1.4 Higgs boson properties

After the discovery of the Higgs boson, its properties have been measured by CMS [5] and ATLAS [6] experiments. These precise measurements are necessary to test the SM and find possible discrepancies which would be a clear sign of physics beyond it.

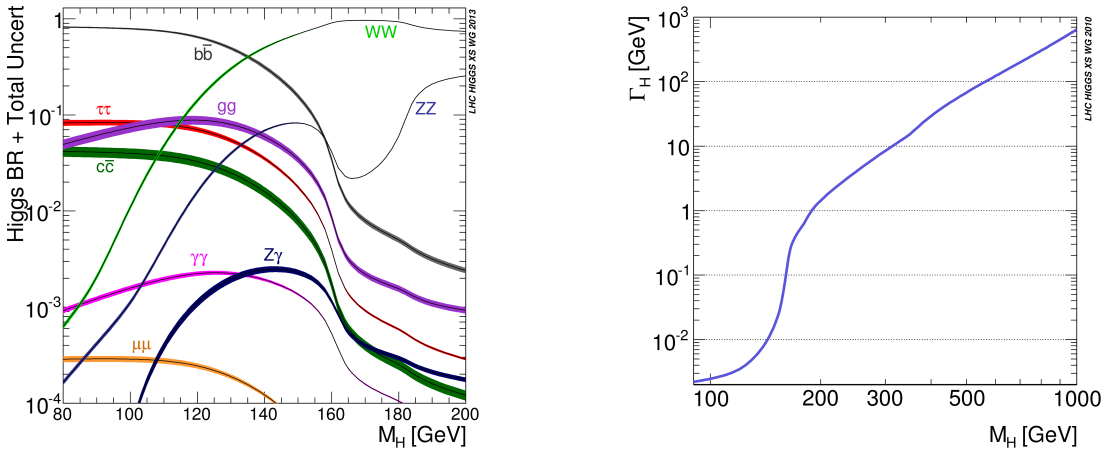


Figure 1.4: Left: Decay branching ratios of the SM Higgs boson in the different channels versus its mass.Right: Total decay width (in GeV) of the SM Higgs boson with respect to its mass [4].

CMS has used events coming from a total integrated luminosity of 5.1 fb^{-1} at $\sqrt{s}$ = 7 TeV (2011 data) and 19.7 fb^{-1} at 8 TeV (2012 data) [7]. Measurements of H $\to \gamma\gamma$, H $\to ZZ^* \to 4\ell$, H $\to WW^*$, H $\to b\bar{b}$ (via the VH production), H $\to \tau^+\tau^-$, H $\to \mu^+\mu^-$ and H to invisible decays, in addition with the ttH procution mode, have been combined to extract the final results, taking into account correlations among several sources of systematic uncertainty.

The different analyses have different sensitivities to the presence of the SM Higgs boson. The H $\to \gamma\gamma$, H $\to ZZ^* \to 4\ell$ (where $\ell = e, \mu$) channels play a special role because of their high sensitivity and excellent mass resolution. The H $\to WW^* \to \ell\nu\ell\nu$ measurement has a high sensitivity due to large expected yields but relatively poor mass resolution because of the presence of neutrinos in the final state. Finally the b$\bar{\text{b}}$ and $\tau^+\tau^-$ decay modes are beset with large background contributions and have relatively poor mass resolution, resulting in lower sensitivity. Table 1.4 shows for each decay channel the significance, i.e. the probability in terms of standard deviation of rejecting the null hypothesis given that it is true.

In the following sections an overview of the CMS results about the Higgs boson properties are presented. Similar results have been published by the ATLAS collaboration [9], therefore also the CMS and ATLAS combined results [8] of the Higgs boson properties are presented. Among them the most precise measurement of Higgs Boson mass is made combining CMS and ATLAS results, exploiting H $\to \gamma\gamma$, H $\to ZZ^* \to 4\ell$ decay channels [10].

Compatibility with the SM

The best-fit value for the common signal strength modifier $\mu = \sigma/\sigma_{SM}$ obtained from the combined analysis of all channels, provides the simplest compatibility test. As shown in Figure 1.5 (left), the observed μ, assuming m_H = 125.0 GeV, is $1.00^{+0.14}_{-0.13}$, consistent with unity and the expectation for the SM Higgs boson. On the right side the measured signal strength for each production mode is shown for the CMS and ATLAS combined results. As it can be seen, some possible discrepancies with respect to the SM are present in the ttH exclusive production. This

Table 1.4: The observed and median expected significances of the excesses for each decay mode, assuming $m_H = 125.0$ GeV

channel	observed (σ)	expected (σ)
$H \to ZZ$	6.5	6.3
$H \to \gamma\gamma$	5.6	5.3
$H \to WW$	4.7	5.4
$H \to \tau\tau$	3.9	3.9
$H \to b\bar{b}$	2.1	2.5
$H \to \mu\mu$	<0.1	0.4

can be a hint of new physics beyond the SM and it needs further investigation with additional data and precise measurements.

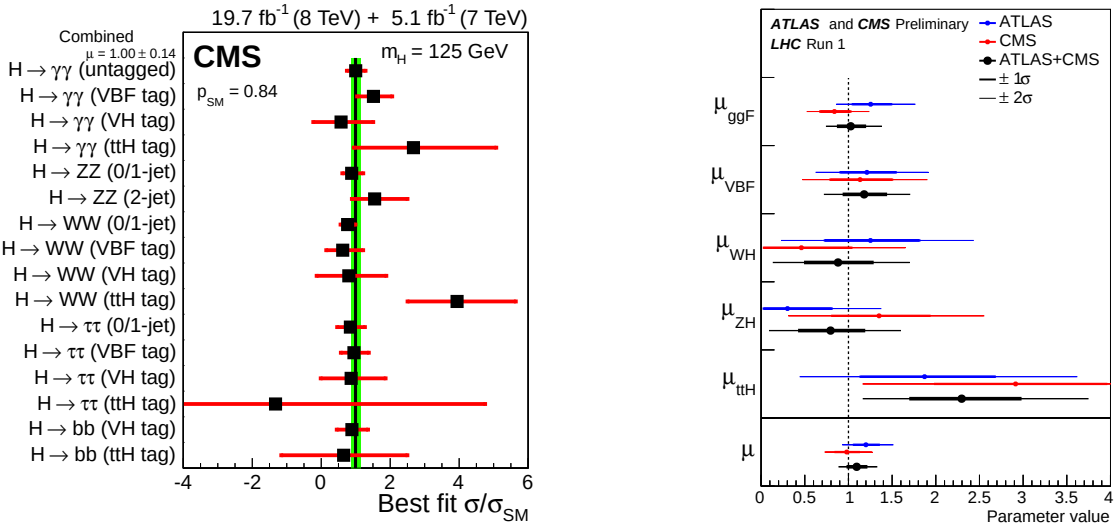


Figure 1.5: Left: Values of the best-fit σ/σ_{SM} for the overall combined analysis (solid vertical line) and separate combinations grouped by predominant decay mode and additional tags targeting a particular production mechanism. The σ/σ_{SM} ratio denotes the production cross section times the relevant branching fractions, relative to the SM expectation. The vertical band shows the overall σ/σ_{SM} uncertainty. The horizontal bars indicate the ± 1 standard deviation uncertainties in the best-fit σ/σ_{SM} values for the individual combinations; these bars include both statistical and systematic uncertainties. Right: Best-fit results for the production signal strengths for the combination of ATLAS and CMS. Also shown for completeness are the results for each experiment. The error bars indicate the 1σ (thick lines) and 2σ (thin lines) intervals. The measurements of the global signal strength μ are also shown.

Mass

The final ATLAS and CMS combined mass measurement is:

$$m_H = 125.09 \pm 0.21(stat) \pm 0.11(syst)\ GeV$$

A great effort has been spent for computing the most precise systematic uncertainty, so that only mass parameter left in the SM to be measured with great precision is the W boson mass. The H $\rightarrow \gamma\gamma$, H $\rightarrow ZZ^* \rightarrow 4\ell$ are fundamental for the precise measurement of the Higgs boson mass. In particular, concerning the H $\rightarrow \gamma\gamma$ a specific optimization of the simulation has been performed in order to reduce the systematic uncertainties in the mass measurement. These optimizations are described in details in Chapter 4.

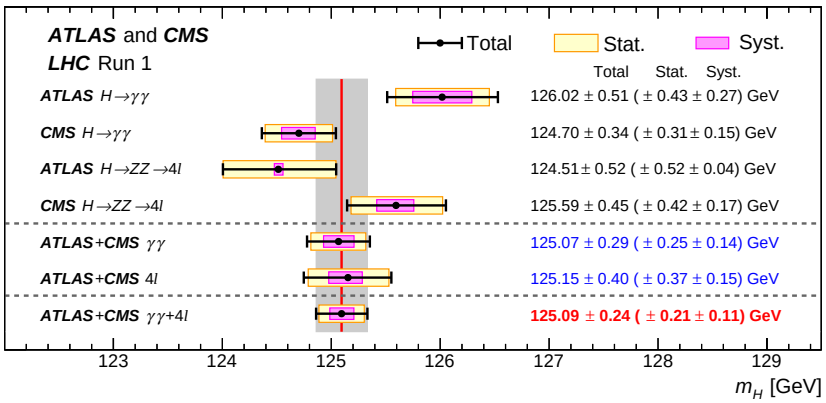


Figure 1.6: Summary of Higgs boson mass measurements from the individual analyses of ATLAS and CMS and from the combined analysis presented here. The systematic (narrower, magenta-shaded bands), statistical (wider, yellow-shaded bands), and total (black error bars) uncertainties are indicated. The (red) vertical line and corresponding (gray) shaded column indicate the central value and the total uncertainty of the combined measurement, respectively.

Width

For $m_H \sim 125$ GeV the SM Higgs boson width is predicted to be $\Gamma_H \sim 4$ MeV. From the study of off-shell Higgs boson production, CMS has set an indirect limit on the total width, $\Gamma_{tot}/\Gamma_{SM} < 5.4(8.0)$ observed (expected) at the 95% CL [11]. It relies on assumptions on the underlying theory, such as the absence of contributions to Higgs boson off-shell production from particles beyond the standard model. In contrast, a direct limit is only limited by the experimental resolution. The direct limit measurement relies on the combination of results from H $\rightarrow \gamma\gamma$, H $\rightarrow ZZ^* \rightarrow 4\ell$ decay channels. The final combined observed (expected) upper limit at the 95% CL is 1.7 (2.3) GeV.

Couplings

Under the assumption that there are no interactions of the Higgs boson other than to the massive SM particles, the data allow a fit for deviations in κ_W, κ_Z, κ_b, κ_τ, κ_t. In this fit, the loop-induced processes (σ_{ggH}, Γ_{gg} and $\Gamma_{\gamma\gamma}$) are expressed in terms of the above tree-level κ parameters and are scaled according to their SM loop structure. Figure 1.8 shows no significant deviations from the SM expectation.

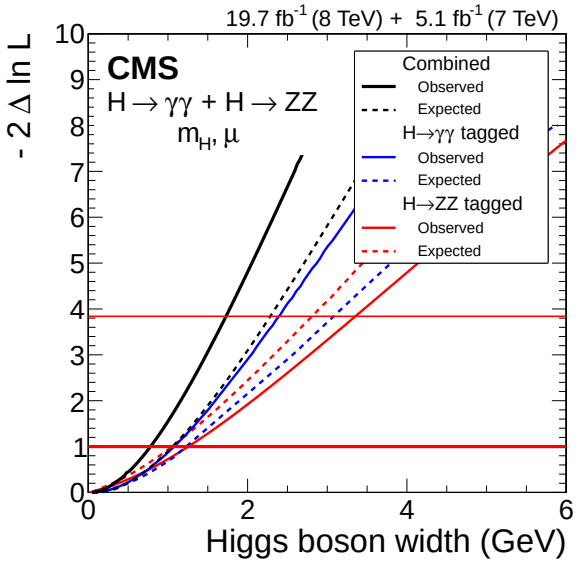


Figure 1.7: Likelihood scan as a function of the width of the boson. The continuous (dashed) lines show the observed (expected) results for the H $\to \gamma\gamma$ analysis, the H $\to ZZ^* \to 4\ell$ analysis, and their combination. The data are consistent with $\Gamma_{SM} \sim$ 4 MeV and for the combination of the two channels the observed (expected) upper limit on the width at the 95% CL is 1.7 (2.3) GeV.

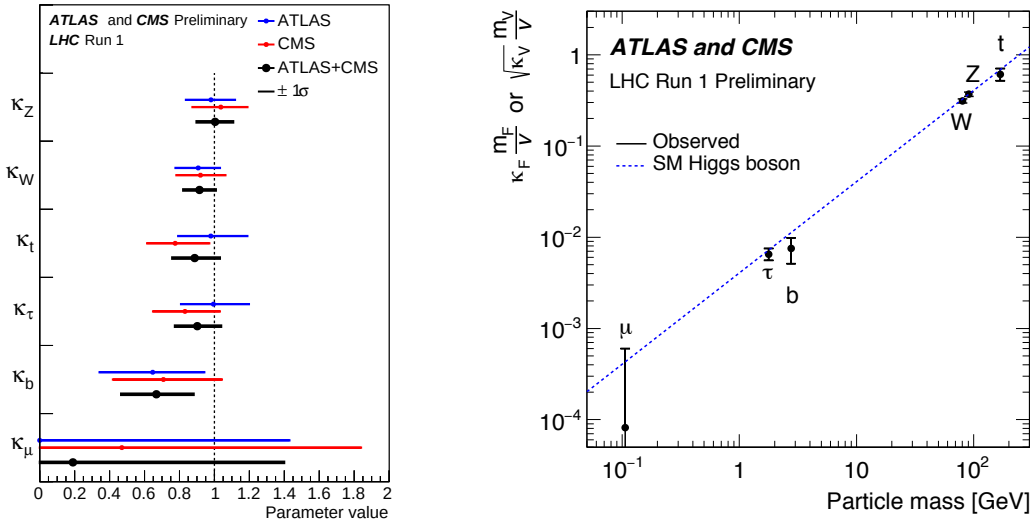


Figure 1.8: Left: Best-fit values of parameters for the combination of ATLAS and CMS and separately for each experiment, for the parametrization assuming the absence of BSM particles in the loops. Right: Fit results for the combination of ATLAS and CMS in the case of the parametrization with reduced coupling modifiers for the weak vector bosons, $y_{V,i} = \sqrt{\kappa_{V,i}\frac{g_{V,i}}{2v}} = \sqrt{\kappa_{V,i}}\frac{m_{V,i}}{v}$ and $y_{F,i} = \kappa_{F,i}\frac{g_{F,i}}{\sqrt{2}} = \kappa_{V,i}\frac{m_{F,i}}{v}$ for the fermions, as a function of the particle mass. The dashed line indicates the predicted dependence on the particle mass for the SM Higgs boson.

Spin and parity

Extensive studies have been performed comparing $J^P = 0^+$ hypothesis with many other possibilities [12]. Kinematic and especially angular variables of $H \to VV$ ($VV = \gamma\gamma$, ZZ, WW or $Z\gamma$) decays are exploited in order to discriminate the different spin-parity hypotheses. All observations are consistent with the expectations for a scalar SM-like Higgs boson and, as shown is Figure 1.9, $J^{PC} = 0^{++}$ is favoured among all the other possibilities.

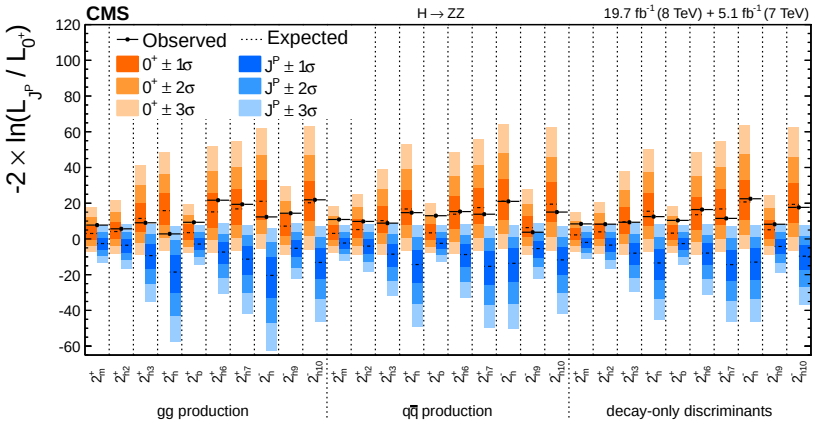


Figure 1.9: Distributions of the test statistic $q = 2ln(L_{J^P}/L_{0^+})$ for the spin-two J^P models tested against the SM Higgs boson hypothesis in the $X \to ZZ$ analyses. The expected median and the 68.3%, 95.4%, and 99.7% CL regions for the SM Higgs boson (orange, the left for each model) and for the alternative JP hypotheses (blue, right) are shown. The observed q values are indicated by the black dots.

1.5 Beyond the Standard Model

New phenomena beyond the SM are expected to be observed at some energy scale, because of the dissatisfaction in the SM description of some observed physical phenomena. In general, the most compelling unsolved issues related to the SM are:

- The electroweak vacuum is not stable in the SM [13], if one takes the measurements of m_t and m_H and extrapolates naively to high scales without introducing new physics
- The SM has no candidate for cosmological dark matter [14]
- The SM does not explain the origin of matter, nor the matter and anti-matter asymmetry
- The SM does not include the mixing and masses of the neutrino
- The SM does not explain the origin and naturalness of the hierarchy of mass scales in physics
- The SM does not accommodate cosmological inflation [15]
- The SM does not include a quantum theory of gravity

Among them the hierarchy problem is one of the issues that directly originates from electro-weak unification and the Higgs mechanism. The main problem is why the weak force is 10^{32} times stronger than gravity and, in a stricter way, why the Higgs boson and the electroweak scale are so much lower than the Planck mass, $M_P \equiv \sqrt{\frac{\hbar c}{G}} \approx 1.2209 \times 10^{19} GeV/c^2$.

One would expect that the large quantum contributions to the square of the Higgs boson mass would make the mass huge, comparable to the energy at which new physics appears, unless there is an incredible fine-tuning cancellation between the quadratic radiative corrections and the bare mass. The problem cannot even be formulated in the strict context of the SM, in fact the Higgs boson mass cannot be calculated because of the its quadratic divergences on the quantum corrections. Many solutions to this problem have been proposed, such as super-symmetry (SUSY) [16], which predict a lot of possible models such as MSSM [17], and models with warped extra dimensions in addition to the 4D space-time, such as [18].

SUSY models predict additional super-symmetric partners to the SM particles, which affect the quantum corrections of Higgs boson mass, due to the cancellation between the couplings with fermions and bosons.

Other models such as the warped extra dimension (WED) models predict extra dimensions, compactified on a line segment, beyond the canonical four, in which the gravity is strong, whereas it is weak in the canonical ones because of a loss of flux into the other dimensions. WED models can be a good solution to the problem since they do not need any fine tuning of the model parameters in order to pass from quantum gravity energy scale to SM energy scale. Moreover, these models are phenomenologically interesting since they predict the existence of massive particles which can be directly probed at LHC.

1.5.1 Warped Extra Dimension models

Warped dimensional theories approach electroweak and cosmological (Planck) scales allowing the gravity to propagate into a compact extra dimension with a warp factor. Those are generic families of Einstein equation solutions [19]:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

The most simple solution for the warp function is known in literature as Randall-Sundrum (RS). The solution of Einstein's equation for this configuration assumes a negative bulk cosmological constant, which means that the warped geometry is anti-deSitter (AdS) space [20].

In the RS model the background metric for the case of single extra dimension is given by

$$ds^2 = e^{-2ky} \eta_{\mu\nu} dx^\mu dx^\nu - dy^2,$$

$$g_{ij} = \begin{pmatrix} e^{-2ky} & & & & \\ & -e^{-2ky} & & & \\ & & -e^{-2ky} & & \\ & & & -e^{-2ky} & \\ & & & & -1 \end{pmatrix} \tag{1.4}$$

where y refers to the coordinate in the 5th dimension, k is related to its curvature, and g_{ij} is the 5 dimension metric. The so-called ultraviolet (UV) and infrared (IR) branes are introduced at $y = 0$ and $y = \pi R$ respectively, as shown in Figure 1.10.

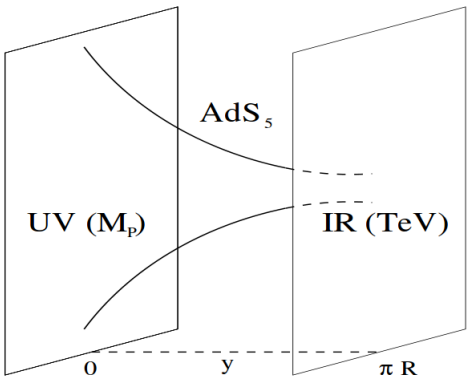


Figure 1.10: A slice of AdS5: the Randall-Sundrum scenario

In particular let H be a complex scalar field with the action that at $y = \pi R$ (i.e. in the usual 4 dimension world) becomes:

$$S_H = -\int dx^5\sqrt{-g}[g^{\mu\nu}\partial_\mu H^\dagger\partial_\nu H - M_5^2|H|^2 + \lambda|H|^4]\delta(y-\pi R) =$$
$$= -\int dx^4[e^{-2k\pi R}\eta^{\mu\nu}\partial_\mu H^\dagger\partial_\nu H - M_5^2 e^{-4k\pi R}|H|^2 + \lambda e^{-4k\pi R}|H|^4] =$$
$$= -\int dx^4[\eta^{\mu\nu}\partial_\mu H^\dagger\partial_\nu H - (M_5 e^{-k\pi R})^2|H|^2 + \lambda|H|^4]$$

where $dx^5\sqrt{-g}$ is the 5-dimensional invariant volume element under general coordinate transformations, g is the metric determinant and M_5 represents a value near the 5D cutoff scale, $M_5 \approx M_P$. The second line is obtained by using the metric definition and performing the y integration, while the third line is obtained rescaling H $\rightarrow e^{k\pi R}$H. As shown in the third line, the original mass parameter M_5 is scaled down or redshifted by an amount $e^{-k\pi R}$ at $y = \pi R$, so $m_{Higgs} \approx e^{-k\pi R}M_P$. In particular, the location in the warped dimension determines the local physical scales. Thus, it does not appear implausible that quantum-gravity effects can start revealing themselves at energies much lower than M_P, possibly as low as the weak scale.

Given the metric g_{ij}, it is possible to identify the massless gravitational fluctuations about vacuum solution, i.e. expanding the 4-dimensional $g_{\mu\nu}$ as:

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x)$$

where $\eta_{\mu\nu}$ is the classical Minkowski metric. With the same criterion the fifth component can be expanded as a small fluctuation about the constant value, $T(x)$. Therefore the metric becomes:

$$ds^2 = e^{-2T(x)y}[\eta_{\mu\nu} + h_{\mu\nu}(x)]dx^\mu dx^\nu - T(x)dy^2$$

With this expansion, the compactification radius R becomes the vacuum expectation value of the $T(x)$. It can be demonstrated that $T(x)$ is stabilized at its vacuum expectation value R, with a mass of about 10^{-4} eV. In general, the stabilization problem can lead to inconvenient fine tuning of the parameters, therefore a more realistic descriptions of WED picture requires a

suitable additional mechanism to stabilize the distance between the UV and IR branes.

A simple way to stabilize the extra dimension size in RS scenarios is to introduce a bulk scalar with specific potential [21], "bulk" scenarios. This bulk scalar with a y-dependent vacuum expectation value can generate a potential to stabilize R without having to fine tune the parameters of the model.

Moreover, in a simple model as the minimal RS, any change of coordinates, such as $z \equiv e^{ky}/k$ eliminates the warping factor $e^{-k\pi R}$ of the 4-dimensional physical quantities. On the other hand, stabilization by means of a potential allows for a minimum independent of coordinates, i.e. warping factor independent of the coordinate systems. Another important feature of these kind of models is that the excitations of the 5-dimensional metric give rise to two gravity mediators which in terms of 4-dimensional indices can be spin 2 (graviton) and spin 0 (radion).

When the gauge symmetry is extended to the bulk of an extra dimension, it is necessary to consider gauge fields on all the directions where the symmetry needs to be restored. If for example a five dimensional abelian local symmetry is assumed, the gauge field introduced to restore local symmetry has five components: $A_M(x,y) = (A_\mu(x,y), A_5(x,y))$. The physics of bulk gauge bosons is fully determined by geometry and boundary conditions. The effect of brane terms to the physics of the bulk fields have been investigated for example at [22].

An additional ingredient in the theory is the propagation of the matter fields (fermions and gauge fields) in the extra dimension, which modifies their interactions with the gravity mediators and makes the theory phenomenologically rich: the matter fields are projected on our 4-dimensional world, producing for each field a tower of excited particles, called Kaluza-Klein (KK) states. This way a consistent structure for the couplings and masses of the new bosons [23] can be found.

The couplings between the gravity mediators, which are now called KK radion and graviton, and the matter fields occur by means of the energy momentum tensor. The Kaluza-Klein modes are different if spin 0 or spin 2 particles are considered. In fact, for spin 0 particles the 0-mode is massive, while for the spin 2 particles only modes greater than 0 are massive. For this reason, we will refer explicitly to "KK" graviton to indicate the 1-mode, while for the radion the 0-mode is always assumed.

Moreover, if the propagation in the extra dimensions is considered, the light fermions become UV localized, heavy quarks (top) become IR localized, and the couplings to the light quarks are very suppressed. On the other hand, the KK graviton profile is IR localized. This makes the overlap between these profiles to be less than unity and the coupling between gluons and KK gravitons to be suppressed. Finally, profiles of non-massive gauge bosons (gluon and photon) are flat. This is shown in detail in Figures 1.11 and 1.12. In particular, Figure 1.12 gives a schematic view of the profiles of the zero KK modes matter fields along the extra dimension y in contrast to the KK graviton profile.

The exact Lagrangian of the interaction with the matter fields is shown in equation 1.5, where $G^{\mu\nu(1)}$ stands for first KK graviton and ϕ for the radion field [24, 25].

$$\mathcal{L} = -\frac{c_i}{\Lambda_G} G^{\mu\nu(1)} T_{\mu\nu} - \frac{d_i}{\Lambda_R} \phi Tr T_{\mu\nu} \tag{1.5}$$

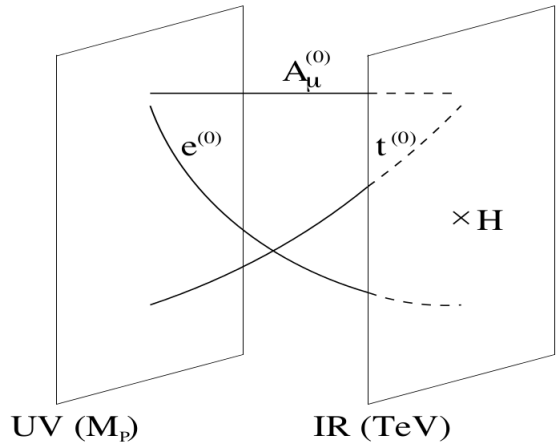


Figure 1.11: The Standard Model in the warped five-dimensional bulk

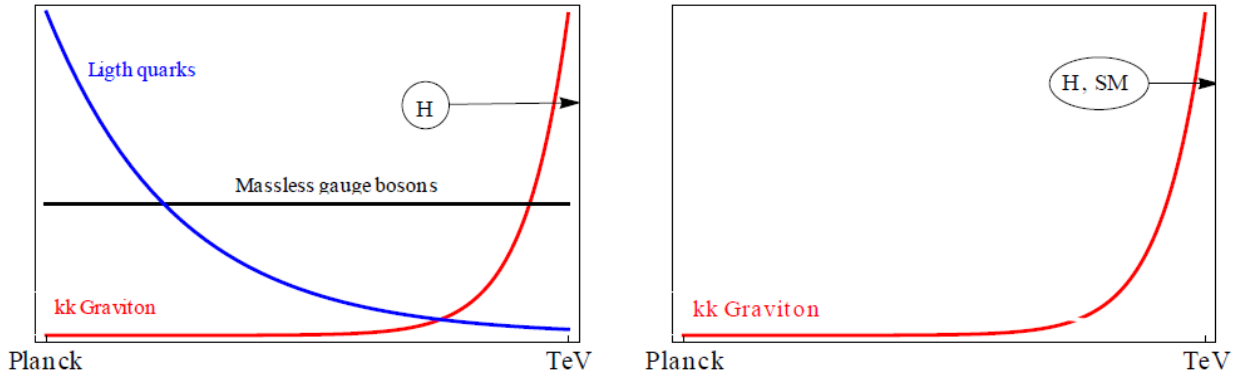


Figure 1.12: Scheme of matter localization on the different WED scenarios. Left: Bulk scenario. Right: RS1 scenario, i.e. the SM fields do not propagate in the extra dimension. The combination of Exponential and Bessel functions makes the KK graviton profile to be very TeV localized.

In this phenomenological presentation, all the relevant parameters are treated as independent. In equation 1.5, c_i and d_i are constants which depend on the behaviour of the matter fields on the bulk, while $T_{\mu\nu}$ is the energy momentum tensor (EMT) for the matter fields and the the scales Λ_ϕ and Λ_G are respectively the radion vacuum expectation value and the ultraviolet mass scale of the theory. Both the scales can be interpreted as cut off of the theory and are expected to be in the 1 TeV energy range or beyond.

In addition, in bulk scenarios the ultraviolet mass scale of the theory Λ_G is also directly related to the first KK graviton mass by $m_G = \frac{k\sqrt{8\pi}}{M_P} x_1 \Lambda_G$, where $x_1 = 3.83$ is the first zero of the Bessel function J_1.

1.6 WED searches at hadron colliders

In this section a description of the phenomenology at the colliders is given [26], [27], [28]. The production cross section of KK graviton on both RS1, i.e. a scenario where the SM fields do not propagate in the extra dimension, and bulk scenarios is shown in Figure 1.13 on the inclusive, associated weak boson and VBF productions for the parameter $k/M_P = 0.2$. The coupling of the KK graviton to bulk light fermions is negligible, therefore only the gluon fusion as inclusive production of bulk KK graviton is considered.

Figure 1.14 shows the KK graviton branching fractions to SM particles. The left part of this figure shows its branching ratios on bulk scenario. The right part shows KK graviton branching ratios on RS1 scenario. The couplings of RS1 KK graviton with SM particles are democratic with respect to the type of the particle, while the couplings of bulk KK graviton with SM particles are predominantly to the fields localized on TeV brane (massive bosons and Higgs boson field). Finally, the graviton width is shown in Figure 1.15 for different scenarios with $k/M_P = 0.2$.

The cross section, branching fraction and width values are given in a specific model with specific values of k/M_P. However, it is simple to extrapolate to other models since the cross section and the width scale with $(k/M_P)^2$.

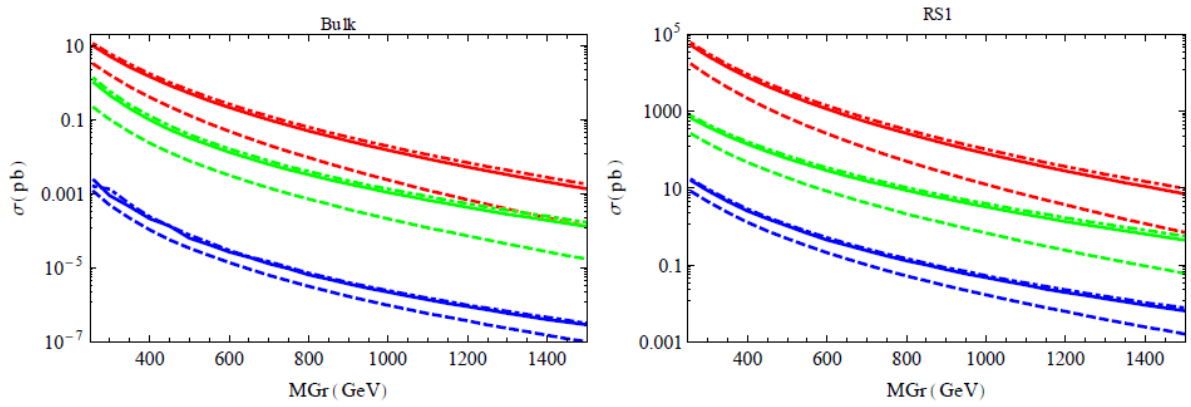


Figure 1.13: Production cross sections with $k/M_P = 0.2$ in a proton-proton collider, in pb. The red curves correspond to the inclusive production, the green ones to VBF and the blue ones to the hadronic weak boson associated production. The dot-dashed, continuous and dashed curves are respectively for the center of mass energy of 14 TeV, 13 TeV and 8 TeV. Right: Bulk scenario. Left: RS1 scenario. The cross section values scale with $(k/M_P)^2$ on all production mechanisms.

The radion field couples to the trace of energy momenta, so it couples to the mass. Its couplings to massless gauge bosons, in particular to the gluon, happen by means of fermion loops, trace anomaly and bulk effects. In particular, trace anomaly is dominant with respect to the fermion loops, and this causes the radion production to be gluon fusion dominated in the full range of masses at LHC.

The decay modes are predominant to massive states: a pair of electroweak vector bosons, a pair of Higgs bosons or a pair of tops. Radion production and decay are fixed by the SM field masses and are practically independent on allowing or not the matter fields to propagate on the

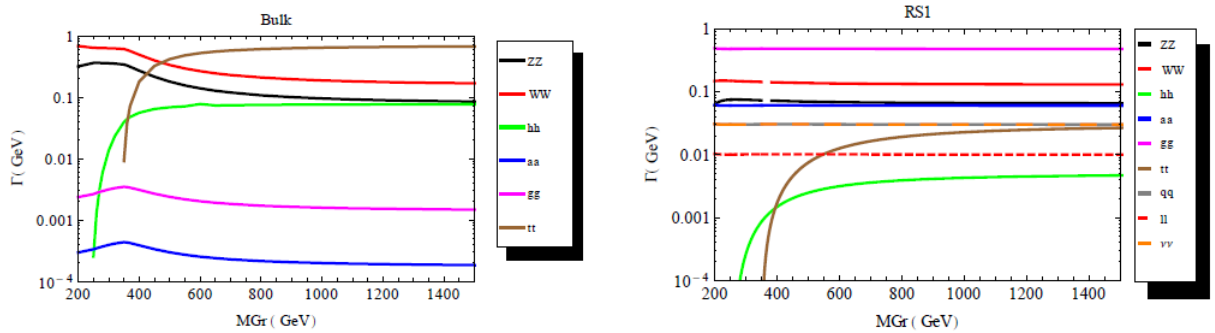


Figure 1.14: KK graviton branching fractions. Left: Bulk scenario. Right RS1 scenario. The dashed line represents the individual KK graviton decay rates to a pair of light fermions $f\bar{f}$. The branching ratios are independent of k/M_P parameter values

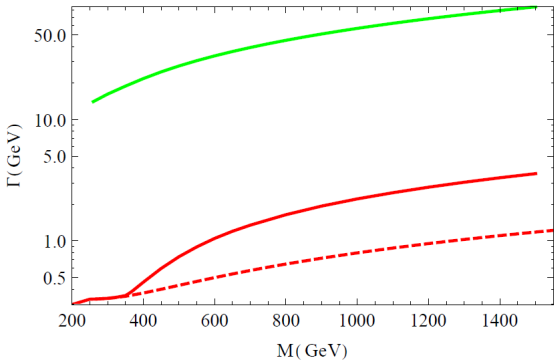


Figure 1.15: Total width for $k/M_P = 0.2$. The green curve represents the RS1 scenario and the red curves represents the bulk scenario. Two different hypothesis of treating KK graviton couplings to bulk third family of quarks are shown: the continuous curve considers a fully composite top quark [29] while dashed curve ignores KK graviton coupling with top quark. Total width values scale with $(k/M_P)^2$ parameter.

bulk [30, 31].

Similarly to the graviton, Figure 1.16 shows on the left the production cross sections of a RS1 Radion as a function of its mass, with $\Lambda_R = 3$ TeV and $kR = 35$ and centre-of-mass energy of 8 TeV, for gluon fusion, VBF, W/Z and $t\bar{t}$ associated production mechanisms. Furthermore, the branching ratio of different decay modes as a function of the mass is shown, assuming no mixing between the Higgs boson and the radion. As it is shown, the gluon pair final state dominates the branching fractions in the mass range less than $2m_W$.

The experimental resolution on the width of the resonance depends on the investigated channels and also of the mass hypothesis probed. The experimental search presented in this thesis is designed using signal models with negligible width with respect to experimental resolution.

Regarding the searches for these models at the LHC, many searches for radion and graviton candidates at the hadronic colliders have been performed. Simple RS1 and bulk models have been extensively tested, searching for candidates decaying either directly into photons, jets or

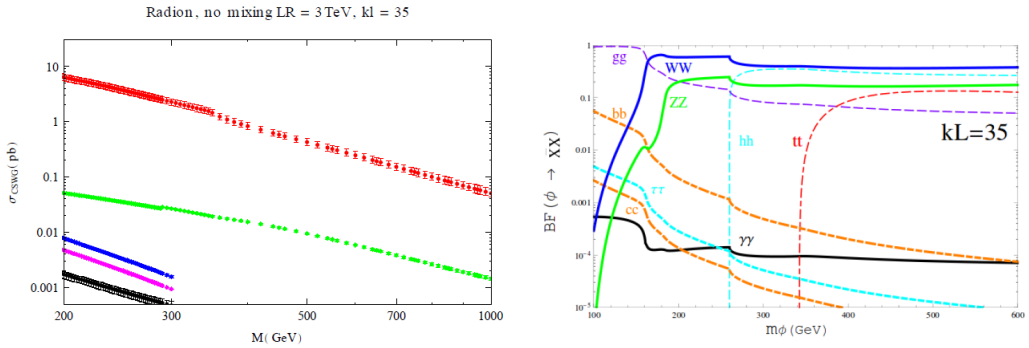


Figure 1.16: Left: Production cross sections of a Radion with $\Lambda_R = 3$ TeV and $k = 35$ on a proton-proton collider with 8 TeV centre-of-mass energy. No Higgs-radion mixing is assumed. Red for gluon fusion production, green for VBF production, blue for W associated production, magenta for Z associated production and black for $t\bar{t}$ associated production. The radion production cross section scales with Λ^2 for all the mass points in the RS1 scenario. Right: Radion decay rates without Higgs-radion mixing for a Higgs boson of 130 GeV [32]

.

leptons pairs, either decaying into W/Z or H bosons. Both for the graviton and the radion, many different final states are tested, since, given the interplay between the branching fraction, the cross section and the various analyses sensitivities, no specific channel could cover all the parameter space (mass, k and Λ).

In the direct decays, graviton candidates have been tested by the CMS and ATLAS experiments at LHC scanning the dijet mass spectrum [33], the diphoton spectrum [34] and the dilepton spectrum [35, 36]. Also the decays into vector bosons have been testes, such as searches for massive resonances in the dijet system containing jets tagged as W or Z boson [37, 38] or for pairs of W/Z bosons decaying semileptonically [39, 40, 41].

Even if no new particles compatible with the graviton are observed, these analyses can exclude many theoretical models with different parameters (such as $k/M_P = 1.0, 0.1$ or 0.01), different intrinsic widths (such as 10 MeV or up to 30% of the particle mass) along a wide mass range of masses, up to 5 TeV, depending on the analysis sensitivity. Figures 1.17, 1.18 and 1.19 show the exclusion limits at 95% CL on the production as a function of the graviton mass.

Concerning the radion searches, while the diphoton final state is tested also for this candidate [34] (exclusion limit scanning on the candidate mass and width is shown in Figure 1.20), the main tested channel of decay is the production of two Higgs bosons in the final state. This signature is used also to test graviton candidates. The tagging of the two Higgs bosons is performed by means of 4 b-quarks (H $\rightarrow$ b$\bar{\text{b}}$ for both Higgs bosons) [42, 43], b$\bar{\text{b}}\gamma\gamma$ [44], b$\bar{\text{b}}\tau\tau$ or b$\bar{\text{b}}WW$.

Also for the radion candidates searches, no excess over the expectation from the SM processes had been found. The final state in b$\bar{\text{b}}$b$\bar{\text{b}}$, which exclusion limits is shown in Figure 1.21, is the most sensitive channel for high mass resonances, $m_X \gtrsim 500$ GeV, because of the experiments capability of identifying jets coming from b-quarks till high masses.

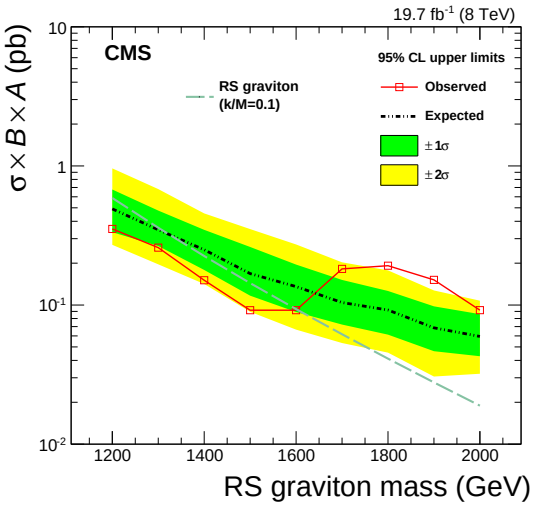


Figure 1.17: The observed 95% CL upper limits on cross section ×BR×Acceptance (σBA) for RS graviton resonances scanning dijet spectrum [33]. The limits are shown as points and solid lines. Also shown are the expected limits (dot-dashed dark lines) and their variation at the 1σ and 2σ uncertainties levels (shaded bands). Predicted cross sections calculated at LO for various narrow resonances are also shown.

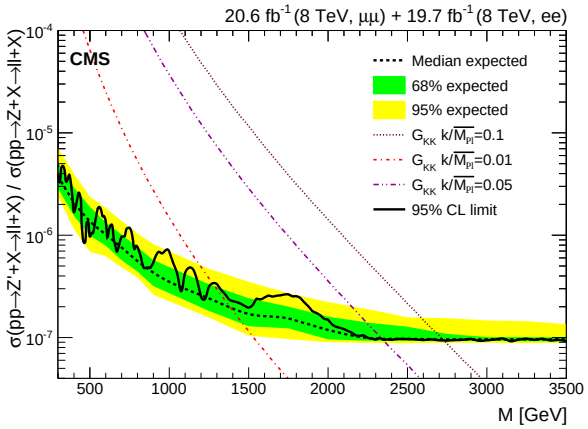


Figure 1.18: Upper limits as a function of the resonance mass M on the ratio of the product of cross section and branching fraction into lepton pairs relative to that of Z bosons [35]. The shaded bands correspond to the 68% and 95% quantiles for the expected limits. Theoretical predictions for spin 2 RS gravitons are shown for comparison.

1.6.1 HH→ $\gamma\gamma$bb search

In this thesis the search for the di-Higgs decay of a radion or graviton candidate, where one Higgs boson decays into $\gamma\gamma$ and the other into a pair of b-quarks, is presented. Figure 1.22 shows the topology used in the simulation of the resonant signals.

As radion and graviton couple to the Higgs boson with non-negligible branching ratio, Figures 1.14 and 1.16, this signature is really relevant for these searches. Furthermore, only the gluon-gluon production is considered in the simulation as it is the dominant production mode

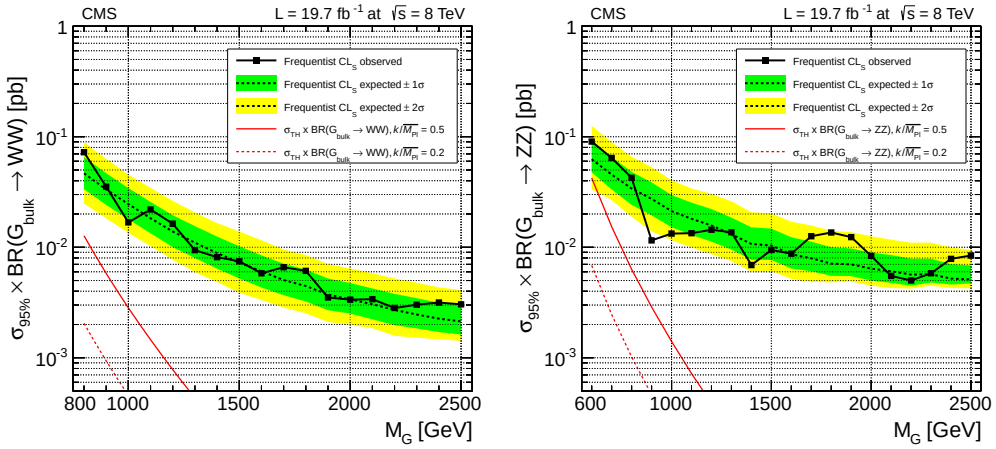


Figure 1.19: Observed (solid) and expected (dashed) 95% CL upper limits on the product of the graviton production cross section and the branching fraction of a graviton $G_{bulk} \to WW$ (left) and $G_{bulk} \to ZZ$ (right) [39]. The cross section for the production of a bulk graviton multiplied by its branching fraction for the relevant process is shown as a red solid (dashed) curve for $k/\overline{M_{Pl}}$=0.5 (0.2), respectively.

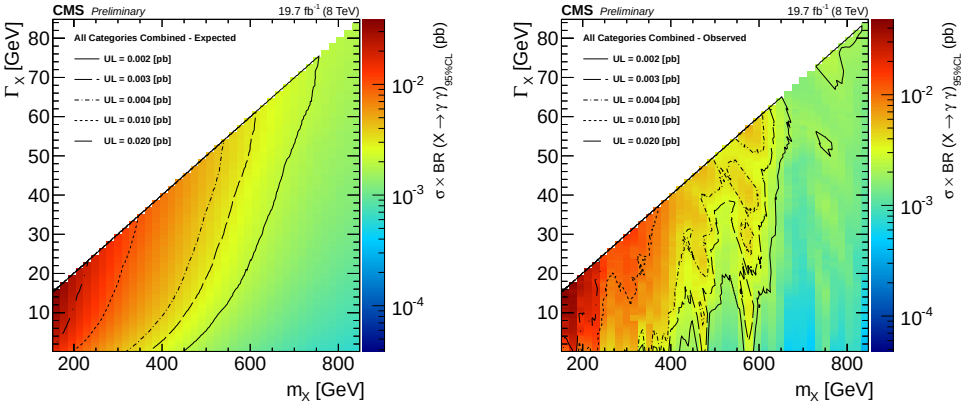


Figure 1.20: Exclusion limit at 95% CL on the cross section times the branching ratio of a spin 0 gluon fusion produced resonance decaying into two photons as a function of the resonance mass hypothesis and width combining the four classes of events [34]. Expected limits are on the left and observed limits are on the right.

for both radion and graviton signals.

Table 1.5 shows the branching ratios of the di-Higgs decay channels, assuming the SM Higgs decay rates (for a 125 GeV Higgs). Indeed, $\gamma\gamma b\bar{b}$ is a clear and fully reconstructible channel to probe the double SM Higgs production. Although the branching fraction is small, 0.26% (Table 1.5), in the context of the CMS experiment, because of the the excellent identification and reconstruction of diphotons coming from H $\to \gamma\gamma$ decay and in particular the high efficiency of the diphoton trigger, $\gamma\gamma b\bar{b}$ signature is the most sensitive for low mass candidates, $m_X < 400$ GeV. In fact, at low mass its sensitivity is better than the $b\bar{b}b\bar{b}$ analysis, which is affected by the trigger inefficiencies.

In addition, the excellent resolution of H $\to \gamma\gamma$ invariant mass, the high efficiency of photons

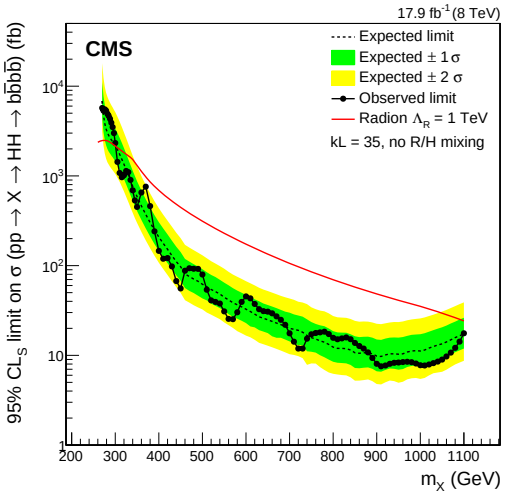


Figure 1.21: The observed and expected upper limits on the cross section for $pp \to X \to$ H(b$\bar{\text{b}}$)H(b$\bar{\text{b}}$) at a 95% confidence level, where the resonance X has spin 0 [42]. The theoretical cross section for the RS1 radion, with $\Lambda_R = 1$ TeV , $k = 35$, and no radion-Higgs boson mixing, decaying to four b jets via Higgs bosons is overlaid on the left plot.

reconstruction (≈90%) and identification of jets coming from b-quarks (≈60%) have permitted to reach enough sensitivity to exclude a relevant fraction of the parameter space of theoretical models.

To achieve these results, it is necessary to have a detector for the photon energy reconstruction perfectly calibrated, described in Chapter 3, and to have a really precise simulation of it, which described in detail in Chapter 4. Furthermore, the complete description of the $\gamma\gamma$b$\bar{\text{b}}$ search is given in Chapter 5.

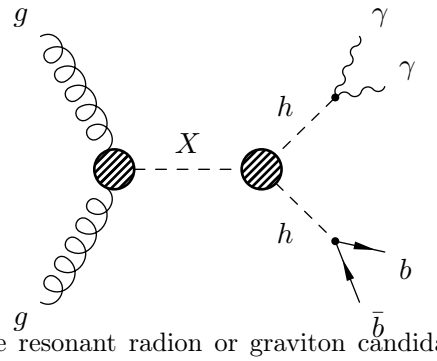


Figure 1.22: Topology of the resonant radion or graviton candidate production, decaying into di-Higgs.

Table 1.5: Branching ratios of the di-Higgs decay channels, assuming the SM Higgs boson decay rates (for a 125 GeV Higgs boson).

Channel	Branching ratio(%)
$h(b\bar{b}, c\bar{c}, gg)h(b\bar{b}, c\bar{c}, gg)$	47.86
$h(b\bar{b})h(b\bar{b})$	33.30
$h(b\bar{b}, c\bar{c}, gg)h(VV^*)$	33.40
$h(b\bar{b})h(\tau^+\tau^-)$	7.29
$h(VV^*)h(VV^*)$	5.83
$h(l^+l^-)h(VV^*)$	3.06
$h(l^+l^-)h(l^+l^-)$	0.40
$h(b\bar{b}, c\bar{c}, gg)h(\gamma\gamma)$	0.32
$h(b\bar{b})h(\gamma\gamma)$	0.26
$h(b\bar{b}, c\bar{c}, gg)h(\mu^+\mu^-)$	0.03
$h(l^+l^-)h(\gamma\gamma)$	0.03

CHAPTER 2

LHC AND CMS

In this Chapter a brief description of the Large Hadron Collider and of the CMS detector are presented in order to contextualize the physics analyses that are described in the following chapters. In particular, the CMS subdectors are described, since they are fundamental for the reconstruction of particles, such as photons and products of partons hadronization.

2.1 The Large Hadron Collider LHC

The Large Hadron Collider (LHC) is a 27 km long proton collider installed 100 m underneath the ground level in the tunnel previously built for the LEP e^+e^- accelerator [45]. The prime motivation of the LHC accelerator was the investigation of the electroweak symmetry breaking for which the Higgs mechanism was presumed to be responsible [46]. The search for a possible Higgs boson was not the only motivation, indeed the accelerator allows the study of the consistency of the Standard Model (SM) at the scale $\Lambda \simeq 1$ TeV. In addition, precision studies of QCD, electroweak and flavour physics are possible which might also open a window onto new physics phenomena.

Furthermore, heavy ion collisions physics have experienced a giant leap: LHC can accelerate the ions till an energy of $\sqrt{s} = 5.5$ TeV, to be compared with the 200 GeV attainable by RHIC [47].

Given the machine design centre of mass energy and luminosity, $7 + 7$ TeV and $\mathscr{L} = 10^{34}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for p-p collisions, the LHC represents a unique challenge from the point of view of technologies and human resources involved.

2.1.1 LHC properties

The high beam intensities necessary for reaching a luminosity of $\mathscr{L} = 10^{34}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ makes the use of two separate proton beams necessary. The collision of two beams of equally charged particles requires opposite magnet dipole fields in both beams. The LHC is therefore designed as a

proton-proton collider with separate magnet fields and vacuum chambers in the main arcs, with common sections only at the insertion regions where the experiments are located.

The choice to reach at regime centre-of-mass energies of 14 TeV has forced to have a magnetic field of $\sim$ 8.3 T, 9300 liquid Helium cooled superconducting magnets made of a Niobium-Titanium compound at a temperature of 1.9 K, by means of super-fluid Helium. Figure 2.1 shows all the acceleration steps the particles have to perform in order to reach 14 TeV energies. To reach the nominal luminosity, up to 2808 bunches per beam, with about 1.1×10^{10} protons each, are collided every 25 ns.

On the LHC ring four main experiments are located: CMS [5], ATLAS [6], LHCb [48] and ALICE [49]. CMS and ATLAS are general purpose experiments, with complementary features and detector choices. CMS will be described in detail in the next sections. The LHCb collaboration aim to perform precision measurements on CP violation and rare decays in order to reveal possible indications for new physics phenomena. ALICE is dedicated to heavy ions physics and the goal of the experiment is the investigation of the behaviour of the strongly interacting hadronic matter resulting from high energy Pb nuclei collisions. In those extreme energy densities the formation of a new phase of matter, the quark gluon plasma, is expected.

2.1.2 Phenomenology of proton-proton collisions

Even though the LHC is a proton collider, the actual particles that interact are the partons (quarks and gluons). The events at a hadron collider can be divided into two main categories: long range and short range collisions.

The long range ones occur when there is a small momentum transfer between the incoming partons (soft collisions), which reduces the particles scattering at large angle. The particles produced in the final state of such interactions have large longitudinal momentum, but a small transverse momentum relative to the beam line and most of the collision energy escapes in regions not accessible by the detectors. The products of these soft interactions represent by far the majority of the p-p collisions, even if they are not interesting for most of physics research purposes.

On the other hand, in case head-on collisions occur between two partons of the incoming protons, interacting at small distances, a large momentum can be transferred (hard scattering). In these conditions, final state particles can be produced at large angles with respect to the beam line with creation of massive particles. However, these are rare events compared to the soft interactions. The total proton-proton cross section at 7 TeV is approximately $\sigma = 110$ mb, while, for example, the production of a W boson through the annihilation of a quark-antiquark pair has a cross-section of about $\sigma = 90$ pb. A comparison of the cross sections of the typical processes at the LHC is shown in Fig. 2.2.

2.1.3 Hadron collider kinematics

In a typical proton-proton collision, the fractions x_a and x_b of the momentum carried by the interacting partons are in general different, and the rest frame of the hard collision is boosted

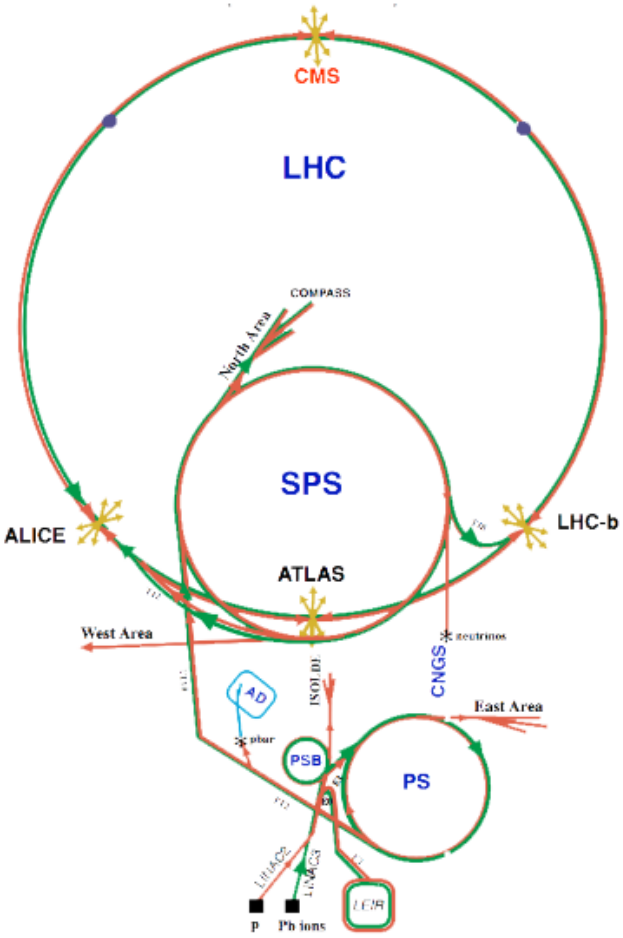


Figure 2.1: The LHC collider ring and the various caverns dedicated to the four experiments: CMS, ATLAS, ALICE, LHCb. At the bottom of the picture is indicated the SPS synchrotron which injects the beams into the LHC machine

along the beam line with respect to the laboratory frame. A convenient set of kinematic variables for particles produced in hadronic collisions, derived from the usual particle quadrimomentum:

$$p^\mu = (E, p_x, p_y, p_z)$$

and, considering the collision axis as the z axis, the transverse momentum p_{T}, the rapidity Y and the azimuthal angle ϕ are defined as

$$p_T = \sqrt{p_x^2 + p_y^2}, \quad Y = \frac{1}{2}\log\frac{E + p_z}{E - p_z}, \quad p_x = p_T\cos\phi, \quad p_y = p_T\sin\phi.$$

These variables have simple transformation properties under boosts along the beam line direction, p_T and ϕ being invariant, and

$$Y \to Y + \frac{1}{2}\log\frac{1+\beta}{1-\beta}$$

where β is the boost velocity along the z direction. Thus $\Delta Y = Y_1 - Y_2$, is invariant under boosts along z. Usually, the rapidity is replaced by the pseudorapidity variable, defined as

$$\eta = -\log\left(\tan\frac{\theta}{2}\right)$$

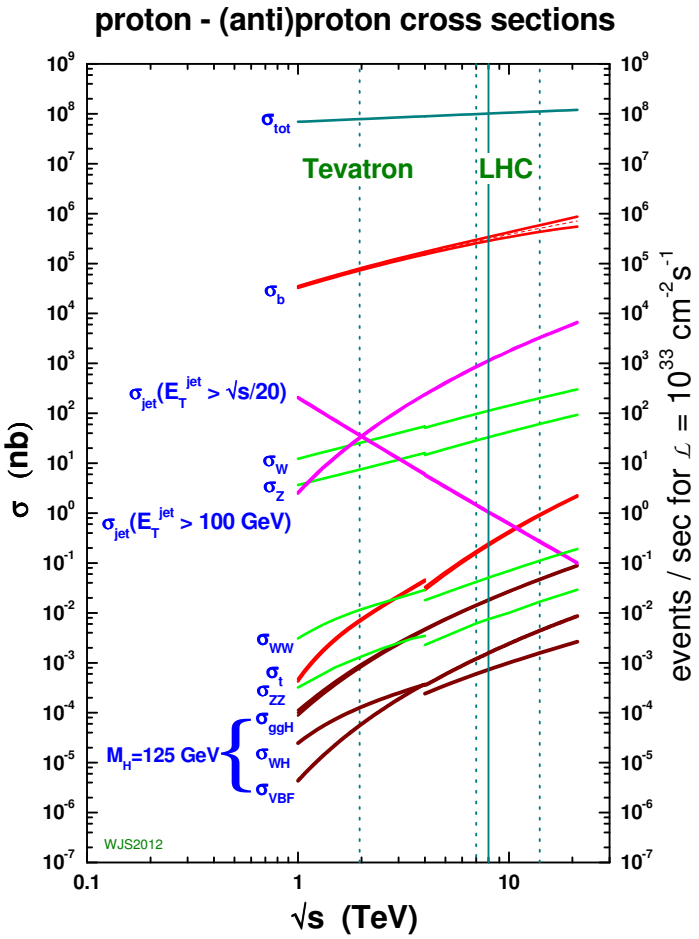


Figure 2.2: Cross sections at LHC for different center of mass energies [4].

where θ is the polar angle between $\vec{p}$ and the z-axis. The pseudo-rapidity is equal to the rapidity for massless particles, therefore the difference in η, namely $\Delta\eta$, is in first approximation invariant under longitudinal boosts.

2.2 CMS Detector

The Compact Muon Solenoid (CMS) is a general purpose detector that is installed at the interaction point number 5 along the LHC tunnel [5]. It is 22 m long, its diameter is 15 m and its weight is about 12500 t. A magnetic field of 3.8 T, enough to bend charged particles to achieve the goal momentum precision, is produced through a superconducting solenoid (see Fig. 2.3), 13 m long and with a inner diameter of 5.9 m. The tracker and the calorimetry are placed in the solenoid, while the 1.5 m of iron of the holding structure is long enough so that the return magnetic field saturates. This struct contains the 4 layers of muon detectors for the outer muon tracking. Figures 2.3 and 2.4 show pictorial views of the CMS experiment subdetectors.

A description of the CMS subdetectors is given in the following sections. In particular, in the context of precision measurements of the Higgs boson properties through the $H \to \gamma\gamma$ decay and the searches for di-Higgs resonances coming from the $\gamma\gamma b\bar{b}$ signature, the tracker is fundamental for the reconstruction of particles coming from the hadronization of partons (jets), especially for momentum and interaction point measurements. Indeed, the interaction point measurement is necessary for the identification of jets coming specifically from b-quarks. In addition, the hadronic calorimeter is necessary for the energy reconstructions of jets, while the electromagnetic calorimeter is the main ingredient for the photon identification and reconstruction.

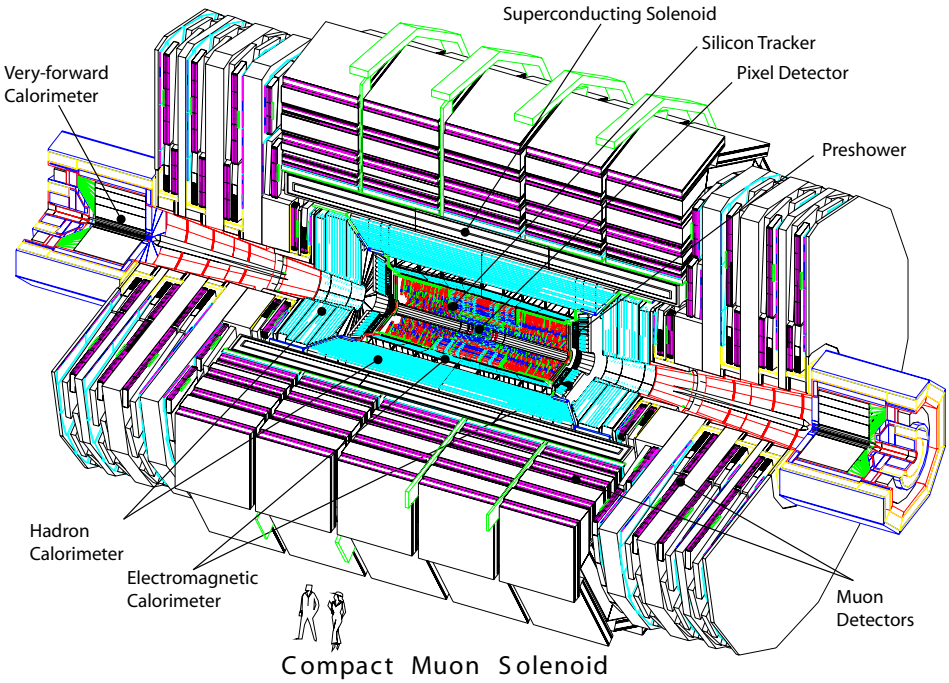


Figure 2.3: Pictorial view of the CMS detector.

2.2.1 Tracker System

The inner CMS tracking system [50] is designed to provide a precise and efficient measurement of the trajectories and the momentum of charged particles, as well as a precise reconstruction of secondary vertices. It surrounds the interaction point and has a length of 5.4 m, a diameter of 2.4 m and a volume of 24.4 m^3, with a running temperature of $-10°$ C.

Given the high luminosity and the large number of particles created in each collision, a detector technology featuring high granularity and fast response is required. Since intense particle flux will also cause severe radiation damage, the main challenge in the design of the tracking system was then to develop detector components able to operate in this harsh environment for an expected lifetime of ten years. All of these requirements led to design a detector technology entirely based on silicon.

Surrounding the beam line, the Silicon Pixel Detector (SPD), organised in three layers, is meant

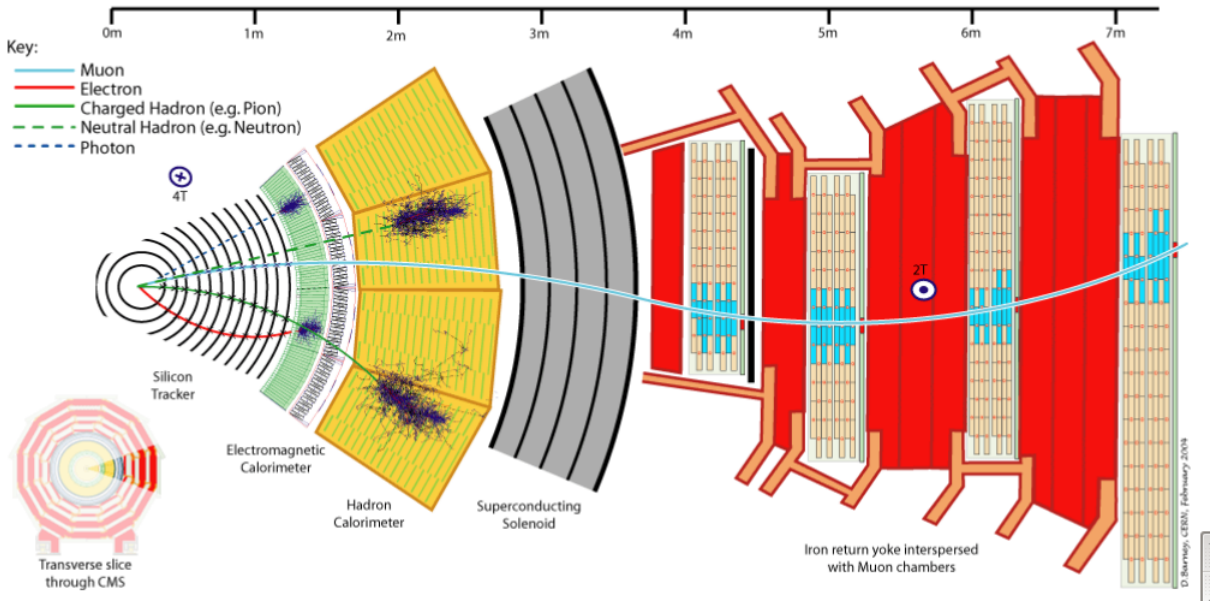


Figure 2.4: A slice of CMS: the picture shows the sub-detector sequence. Paths of different particles are also drawn.

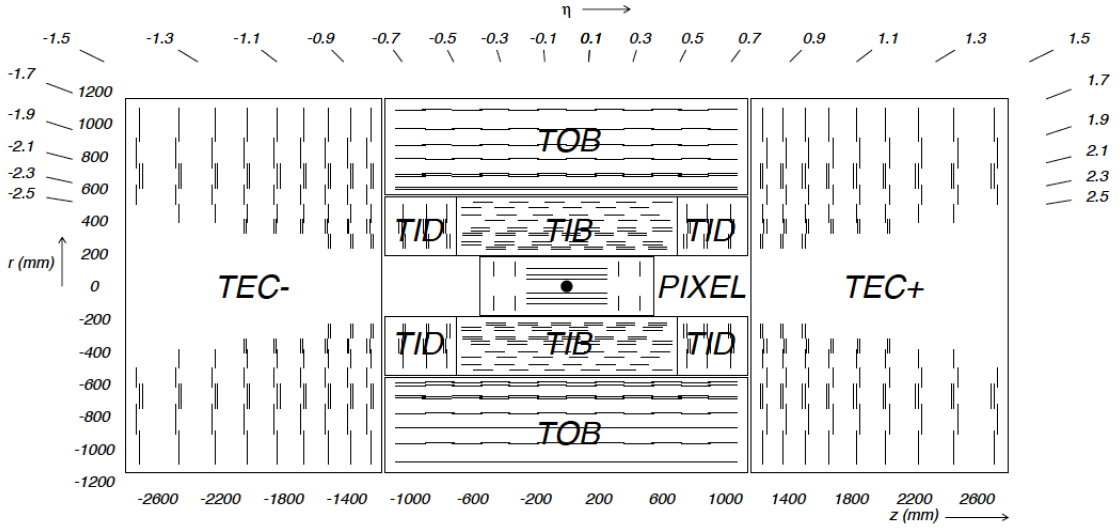


Figure 2.5: The CMS tracker detector: an overview. This device is composed of Silicon Pixel Detectors and Silicon Strip Detectors. The Silicon Strip Tracker is composed of three sub-detectors: Tracker Inner Barrel (TIB), Tracker Inner Disk (TID), Tracker Outer Barrel (TOB) and Tracker End Cap (TEC).

to reconstruct the secondary vertices of the interaction, which are fundamental for the identification of jets coming from b-quarks. It is made of silicon pixels with 150×150 μm^2 surface and 250 μm thickness, divided into modular units of 6.4×1.6 cm^2. It covers the region $|\eta| < 2.6$. A total of 1440 pixels modules are mounted in three barrel layers at radii between 4.4 cm and 10.2 cm and two endcap disks on each side of the barrel.

A second sub-detector, the Silicon Strip Detector (SSD), surrounds the SPD. It is composed

of silicon microstrips arranged in 10 cylindrical layers and 9 disks per endcap. It covers the region $|\eta| < 2.5$. A total of 15148 silicon strip modules are arranged in 10 barrel detection layers extending outward to a radius of 1.1 m and 9 disks on each side of the barrel. The active silicon area is about 200 m^2 active, making the CMS tracker the largest silicon tracker ever built.

One of the major constraint in the design of a tracking system is to reduce as much as possible the amount of material distribution in front of the subsequent calorimeters. The material budget of the tracker system, shown in Fig. 2.6, constitutes the main source of uncertainty in accurate calorimetric measurements of electrons and photons, because it causes the conversion of photons into e^+e^- pairs and the bremsstrahlung of the electrons.

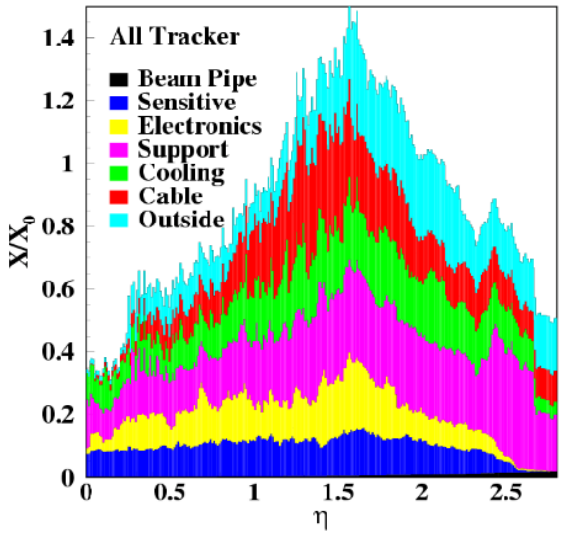


Figure 2.6: CMS Tracker budget material in units of radiation length X_0 as a function of η. The maximum is reached in the region of transition between the barrel and the endcaps.

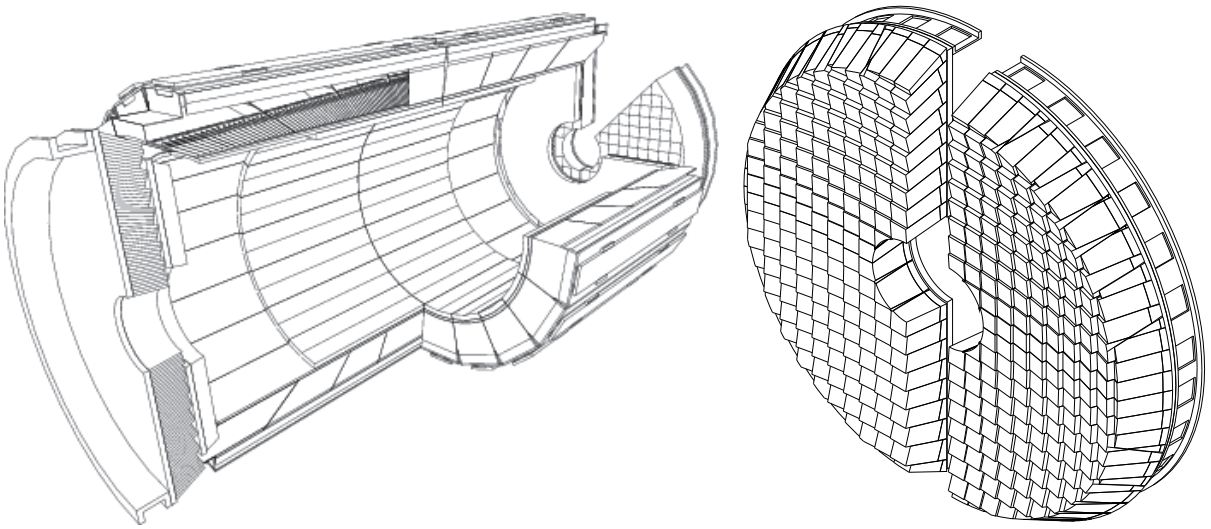
2.2.2 The electromagnetic calorimeter: ECAL

During first period of LHC operation (Run 1), the CMS electromagnetic calorimeter (ECAL) [51] played an essential role in the search for the Higgs boson by measuring with high precision the energy of photons arising from $H \to \gamma\gamma$ decays and of electrons from $H \to ZZ$ decays.

To achieve this, a homogeneous and hermetic calorimeter, made of scintillating lead tungstate crystals, has been designed. This material is suitable for operation at LHC due to its fast emission (80% of the scintillation light is emitted within 25 ns) and its resilience to irradiation. Moreover, thanks to crystals' short radiation length ($X_0 = 0.89$ cm) and small Molière radius ($r_M = 21.9$ mm), most of an electron or photon's energy can be collected within a small matrix of crystals. In addition these scintillating crystals are really fast, reaching a 25 ns scintillating time.

As shown in Figure 2.7, ECAL crystals are divided into the two main parts of the calorimeter:

- Barrel (EB): it is made of 61200 crystals and it is shaped as a cylinder, with an inner radius



(a) ECAL (b) ECAL EE

Figure 2.7: CMS ECAL geometry schema. The ECAL barrel (EB) is made of 36 Super-Modules: 18 in EB+ ($z > 0$) and 18 in EB- ($z < 0$) as depicted in (a). The ECAL endcap (EE) is divided in 4 Dees: 2 Dees in EE+ ($z > 0$) and 2 Dees in EE+ ($z > 0$) as depicted in (b).

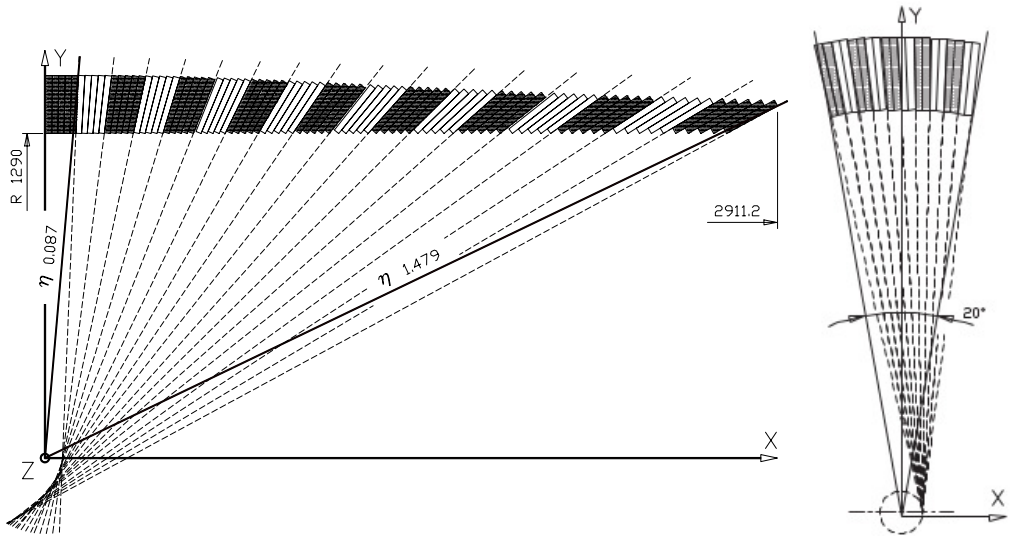
of 1.290 m. It covers a pseudorapidity range of $0 < |\eta| < 1.479$ and its granularity is 360 crystals in ϕ and 2×85 crystals in η. Crystals in EB are mounted in a quasi-projective geometry to avoid cracks aligned with particle trajectories, so that the crystal axes are tilted at an angle of 3° in both the ϕ and η projections, as shown in Figure 2.8.

- Two endcaps (EE): they cover the pseudorapidity range $1.479 < |\eta| < 3.0$ and consist of identically shaped crystals, grouped into carbon-fiber structures of 5×5 elements, called supercrystals. Each endcap is divided into 2 halves, or Dees, holding 3662 crystals each.

The relatively low light yield of $\simeq 30$ $\gamma/$ MeV makes it necessary to use intrinsic high-gain photodetectors, capable of operating in high magnetic fields. Avalanche PhotoDiodes (APDs) are used for barrel crystals and Vacuum PhotoTriodes (VPTs) are used for endcaps crystals. APDs have a nominal gain of 50 and a variation of $\Delta G/\Delta T$ = -2.4%/°C and $\Delta G/\Delta V$ = 3.1%/V. On the other hand VPTs are more radiation resilient but have a gain variation of 25% over all the endcaps.

Moreover, a preshower (ES) is placed in front of EE crystals with the aim of identifying neutral pions in the endcaps and improving the position determination of electrons and photons. The preshower consists of two layers made of passive lead radiators ($2X_0$ and $1X_0$), which initiate electromagnetic showers, and active silicon strip sensors placed after each radiator (4288 strips, thickness of 310 μm and area of 1.9mm×61mm), which measure the deposited energy and the transverse shower profiles.

ECAL operates with a temperature of 18°C which is maintained by a dedicated cooling system [52]. The temperature dependence of the crystal light yield (−2%°C) and of the APD gain (−2%°C) demand a precise temperature stabilization of better than 0.05°C in the EB. In the endcaps, the dependence of the VPT response on the temperature is negligible, and a stabiliza-



(a) η tilt (b) φ tilt

Figure 2.8: The disposal of CMS ECAL crystals. The crystal tilt in a transverse view (a) and construction of the crystal φ tilt (b). To produce a non-pointing geometry in η, crystal longitudinal axes are all inclined by 3°with respect to the line joining the crystal front face centre to the interaction point.

tion of better than 0.1°C for the crystals is sufficient. These specifications limit the contribution to the constant term of the energy resolution to be less than 0.2%.

2.2.3 The hadronic calorimeter: HCAL

The goal of the hadronic calorimeter (HCAL) [53] is to measure the energy of hadronic jets. HCAL is a sampling calorimeter and HCAL barrel is radially restricted between the outer extent of the electromagnetic calorimeter (r = 1.77 m) and the inner extent of the magnet coil (r = 2.95 m). Because of this constraint on radiation absorbing material, an outer hadronic calorimeter is placed outside the solenoid. Furthermore, Beyond $|\eta| = 3$, the forward hadron calorimeters are placed at 11.2 m from the interaction point and extend the pseudorapidity coverage down to $|\eta| = 5.2$. A schematic view of HCAL location in the CMS detector is shown in Figure 2.9.

The central calorimeter is divided into a barrel part (HB, $0 < |\eta| < 1.3$) and two endcaps (HE,$1.3 < |\eta| < 3$), with a transverse granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087°$ and $\Delta\eta \times \Delta\phi \simeq 0.17 \times 0.17°$ respectively. The 3.7 mm thick active layers of plastic scintillators are interleaved with 5 to 8 cm thick brass absorbers. Wavelength-shifter are used to bring out the scintillation light and read the signal. The HB effective thickness increases with polar angle θ as $1/\sin\theta$, resulting in a total absorber thickness spanning from 5.82 interaction length (X) at 90° to 10.6 X at the end of the barrel. In the endcaps, the total length of the calorimeter is about 10 interaction lengths.

As in the central pseudorapidity region the combined stopping power of EB plus HB does not provide sufficient containment for hadron showers, hadron calorimeter is extended outside the

solenoid with an additional absorber equal to $1.4/\sin\theta$ interaction lengths (HO).

The forward calorimeter (HF) experiences unprecedented particle fluxes, thus the design of the HF calorimeter was guided by the necessity to survive in this environment, preferably for at least a decade (10 MGy of absorbed dose are expected at $|\eta| = 5$ after ten years of LHC operation). The calorimeter consists of a steel absorber structure composed of 5 mm thick grooved plates. Quartz fibers are inserted in these grooves and constitute the calorimeter active medium, detecting energy emitted by particles via Cherenkov radiation.

Regarding the HCAL performance, Figure 2.10 reports the jet energy resolution expected from the simulation and measured in data within the tracker acceptance as a function of transverse momentum p_T [54].

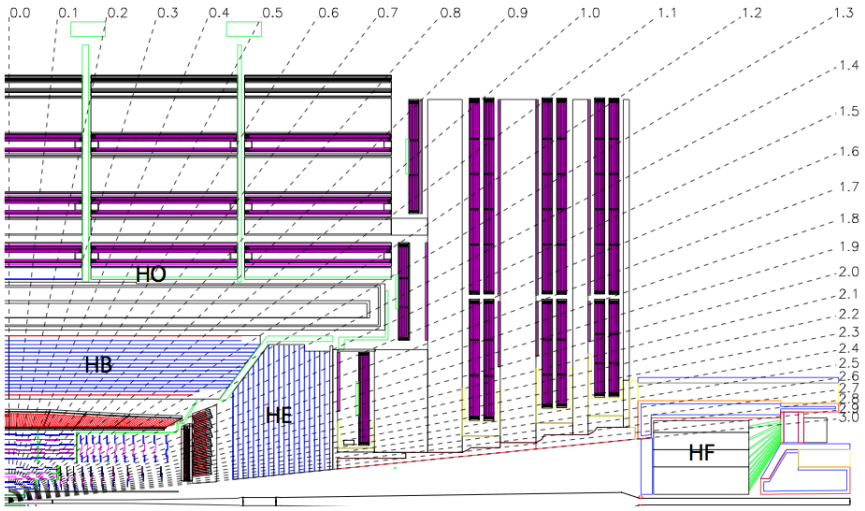


Figure 2.9: Schematic HCAL composition

2.2.4 Muon system

The CMS muon system [55] provides full geometric coverage for muon measurement up to $|\eta| = 2.4$. The detectors are embedded in the magnet return yoke, so that muon momentum and charge measurements can also exploit the strong magnetic return field. This is particularly important for muons with transverse momentum in the ~ 1 TeV range, for which the complementary tracker measurements degrade.

CMS muon spectrometer is composed of a barrel part ($|\eta| < 1.2$) and a forward region ($0.9 < |\eta| < 2.4$). The barrel consists of five wheels, in which drift-tube (DT) detectors and resistive-plate chambers (RPC) are placed in concentric muon stations around the beam line. The total radius is between about 4 and 7 m. In the forward region cathode-strip chambers (CSC) and RPC are mounted perpendicular to the beam line in overlapping rings on the endcaps.

In Figure 2.11 an (r,z) view is given of the different parts of a quarter of the CMS muon system. The CMS muon system uses three types of gaseous detectors. The choice of detector technologies has been driven by the need for fast triggers, excellent resolution, coverage of a very

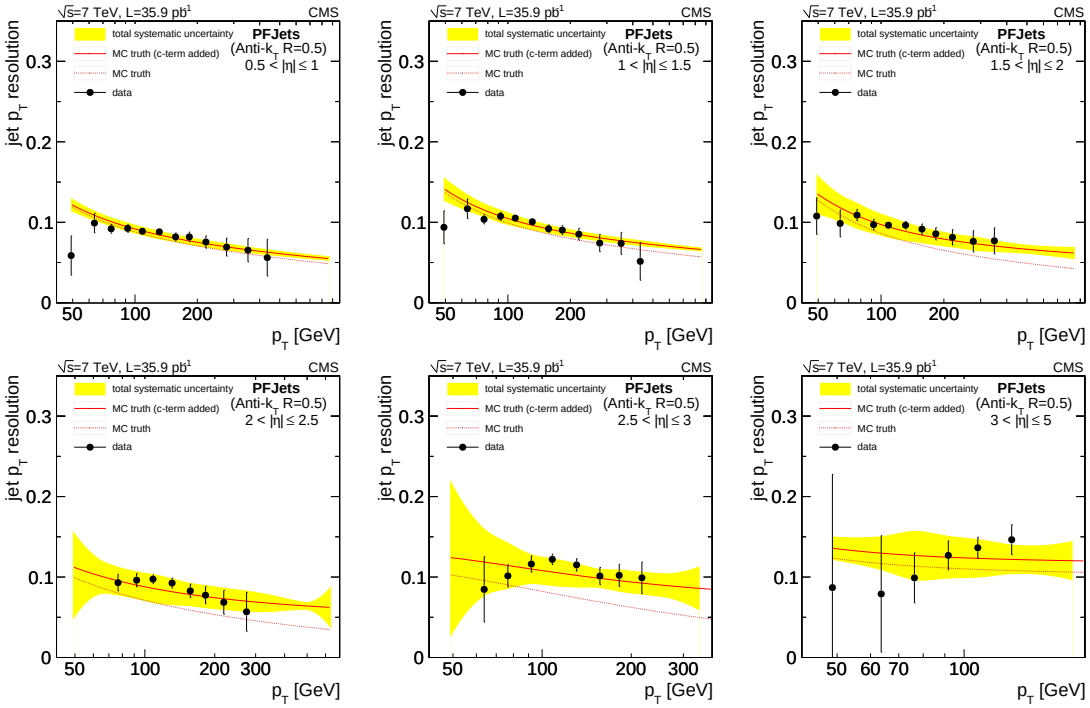


Figure 2.10: Bias-corrected data measurements, compared to the generator-level MC (denoted as MC-truth) resolution, before (red-dashed line) and after correction for the measured discrepancy between data and simulation (red-solid line), for jets in different η ranges [54]

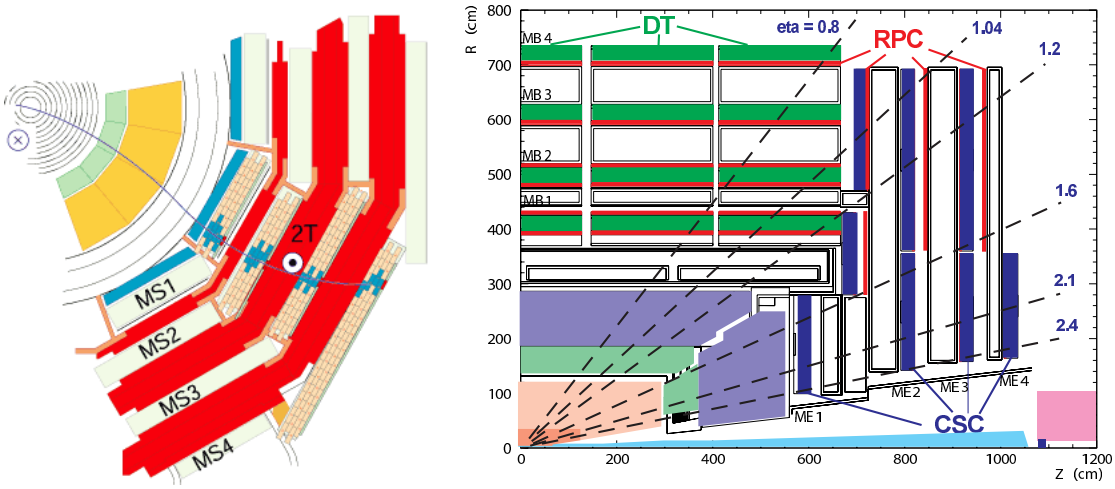
large total surface and operation in dense radiation environments.

- The drift tubes (DT): they are located in the barrel part, where the muon rate is expected to be low. Each station is designed to measure muon positions with about 1 mrad resolution in ϕ;
- The cathode strip chambers (CSC): they are used in the endcap regions, where the magnetic field is very intense (up to several Tesla) and very inhomogeneous.
- The resistive plate chambers (RPC): the position resolution from the RPCs is poorer than for the DTs and CSCs, but the collection of charges on the strips is very fast. Therefore these chambers are used mainly for trigger purposes and for an unambiguous identification of the bunch crossing.

2.2.5 The trigger system

As the production rate ($40{\cdot}10^6 s^{-1}$) is high compared to the affordable acquisition rate ($\sim 600 s^{-1}$), a powerful trigger system has been implemented, based on two levels. The first one is hardware implemented on each subdetector, but the tracker (Level-1 trigger), the second is the High Level Trigger (HLT), which runs on a dedicated farm of commercial PCs [56].

The Level-1 triggers involve the calorimetry and muon systems, as well as some correlation of information between these systems. The Level-1 decision is based on the presence of trigger



(a) Lateral view (b) Longitudinal view

Figure 2.11: The CMS muon stations integrated in the iron return yoke. The trajectory of a typical muon is displayed in the lateral view.

primitive objects such as photons, electrons, muons, and jets above set transverse energy thresholds. It also employs global sum of $E_T \equiv \sqrt{m^2 + p_T^2}$ and E_T^{miss} [57]. Reduced-granularity and reduced-resolution data are used to reconstruct trigger candidates.

The HLT reduces the output rate down to few 100 Hz. The idea of the HLT software is the regional reconstruction on demand, so that only the candidates in the useful regions are reconstructed and the uninteresting events are rejected as soon as possible. This leads to the development of two virtual trigger levels: at the first level only the full information of the muon system and of the calorimeters is used, in the second level the data from the tracker pixels are added.

In addition to triggers where all events that satisfies the requirements are saved (*unprescaled triggers*), a set of utility triggers, whose rate would be too high due to band saturation, have been developed, where a good event is saved only once every N times (*prescaled triggers*, with a prescale parameter N). The latter ones are used for detector studies and to study kinematical regions of object reconstruction, such as low p_{T} leptons.

CHAPTER 3

ECAL DETECTOR CALIBRATION

In this chapter a full description of the ECAL detector calibration procedure is presented. Indeed, photon identification and energy resolution have a direct impact on the sensitivity of searches using diphoton resonances, such as $X \to HH \to \gamma\gamma b\bar{b}$. Data analysis, described in detail in Chapters 4 and 5, rests on a detailed instrumental work to reconstruct and calibrate the energy of the electromagnetic showers in the ECAL detector, and to understand and model the detector response in the simulation.

3.1 Energy Reconstruction

When a photon or an electron strikes the detector it showers through the crystals. The energy of these electromagnetic showers is deposited in crystal matrices. On average the electrons/photons leave 94% of their total energy in a 3×3 crystal matrix and 97% of their total energy in a 5×5 crystal matrix. Electrons are reconstructed combining ECAL and tracker measurements [58], while the photon reconstruction is based on ECAL [59].

The energy reconstruction is complicated by the fact that electrons and photons start showering in the tracker, through bremsstrahlung and photon conversion processes. Furthermore, the energy radiate in the tracker is spread along ϕ due to the intense magnetic field of CMS.

Clustering algorithms are used to sum together energy deposits in adjacent crystals belonging to the same electromagnetic shower. The clustering algorithms begin first with the formation of "basic clusters", corresponding to local maxima of energy deposits. The basic clusters are then merged together to form a "supercluster", which is extended in ϕ, to recover the radiated energy. The energy in a supercluster can be expressed as:

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

the sum runs over all the crystals composing the supercluster and the different terms are:

- E_{ES}: the fraction of energy measured in the preshower.

- A_i: the ADC amplitude.
- C_i: the intercalibration term, which takes into account crystal-to-crystal variations in the response by equalizing the different crystal responses.
- $S_i(t)$: the time dependent transparency correction.
- G: the ADC to GeV conversion factor, it is determined separately for the barrel and the endcaps from $Z \rightarrow e^+e^-$ decays.
- $F_{e,\gamma}$: energy corrections, applied to the superclusters to take into account the η and ϕ dependent geometry and material effects as well as the fact that electrons and photons shower slightly differently.

Because of the differences between the geometric arrangement of the crystals in the barrel and endcap regions, a different clustering algorithm is used in each region. The algorithms do not make any hypothesis as to whether the particle originating from the interaction point is a photon or an electron, consequently electrons from $Z \rightarrow e^+e^-$ events can provide excellent measurements of the photon reconstruction and identification efficiencies, and of the photon energy scale and resolution. The clustering algorithms achieve a rather complete (≈95%) collection of the energy of photons and electrons, even those that undergo conversion and bremsstrahlung in the material in front of the ECAL.

About half of the photons convert in the material upstream of the ECAL. The R_9 variable, defined as the energy sum of the 3×3 crystals centred on the most energetic crystal in the supercluster divided by the energy of the supercluster, is useful for discriminating photons and electrons coming from the tracker interaction and prompt photons and electrons coming from the hard event. The showers of photons that convert before reaching the calorimeter have wider transverse profiles and lower values of R_9 than those of unconverted photons, as shown in Figure 3.1.

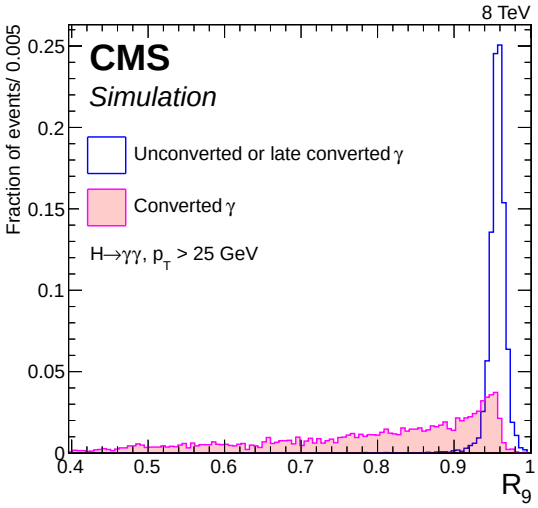


Figure 3.1: Distributions of the R_9 variable for photons in the ECAL barrel that convert in the material of the tracker before a radius of 85 cm (solid filled histogram), and those that convert later, or do not convert at all before reaching the ECAL (outlined histogram)

3.1.1 Amplitude reconstruction

The front-end electronics of the EB, EE, and ES use 12-bit analogue-to-digital converters (ADC) to sample the analogue signals from APDs, VPTs, and silicon sensors at 40 MHz. In EB and EE ten consecutive samples are stored for each trigger received, while in the ES only three samples are stored. Reading out a larger number of samples allows the identification of the out-of-time pileup, which is caused by early interactions with respect to in-time events, an event-by-event subtraction of the pedestal and some reduction of the noise contribution.

The electronic noise is driven by the photodetector electronics. Figure 3.2 shows the noise in ADC counts for different η regions during 2011 and 2012 data taking. Both the APDs and VPTs noises have an η dependence, while only the APDs noise grew during the 2011 and 2012. In the barrel part, the noise is mildly recoverable, as it can be seen in the data taking stop between 2011 and 2012, whereas for the endcaps the equivalent noise in energy strongly depends on the response time variations, which are described in the following sections.

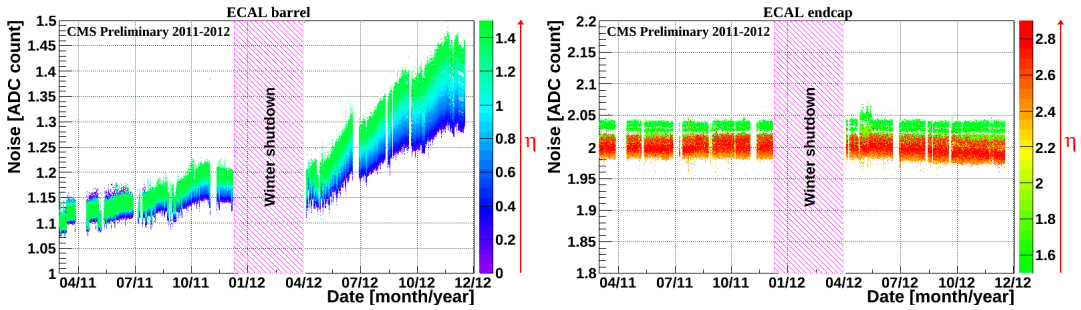


Figure 3.2: Single channel noise measured on the pre-samples of the laser events taken during standard monitoring sequences in 2011 and 2012 for barrel and endcap

The readout phase of ECAL is adjusted such that the signal pulse starts on the forth sample and reaches its maximum on the sixth sample, while the baseline pedestal can be estimated from the first three samples. Figure 3.3 shows the profile of the signal pulse from a crystal of a super-module using an electron beam of 120 GeV. In the ES the pedestal is in the first sample and the signal is in the two following samples. In both barrel and endcaps the amplitude of the signal is reconstructed in the same way using a linear combination of the samples:

$$A = \sum_j w_j \cdot s_j \tag{3.2}$$

where s_j is the sample value in ADC counts, w_j is a weight and j runs over the samples. The w_j weight is computed using a "3+5 pedestal-subtracting weights" method, optimized for noise reduction using the average pulse shapes, measured in beam tests in the respective detectors [60]. The optimization of the weights using the noise covariance matrix results in giving more weight to the sample on the peak.

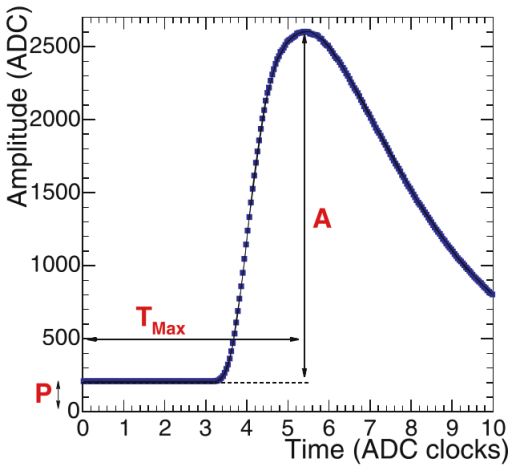


Figure 3.3: Profile of the signal pulse from a crystal of a super-module using an electron beam of 120 GeV. The peaking time T_{Max} , the pedestal P and the amplitude of the signal A are shown

3.2 Monitoring System

The optical transmission within crystals at the scintillation wavelengths is affected by the production of colour centres under ionizing radiation. This transparency loss process is not permanent, in fact spontaneous annealing of the colour centres occurs also at room temperature and leads to a transmission recovery, which is evident when the crystals are not irradiated, such as during machine-fill gaps. The net time effect is an equilibrium between the rates of colour centres production and their annealing.

Crystals produced for ECAL are optimized to reduce the relative variations in light transmission during an LHC collision running period (luminosity of $10^{34}\ cm^{-2}s^{-1}$) to less than 6% for barrel crystals (dose rates of 0.15 Gy/h) and less than 20% for the endcpas at $|\eta| = 2.5$ (dose rates of 1.9 Gy/h) [61].

Uncertainties in the relative measurement of the optical transmission, from crystal to crystal, contribute directly to the energy resolution. *In situ* light transmission measurements are performed through a laser monitoring system [61] which is based on injecting light into the crystal during collisions gaps.

The laser light pulses are directed to individual crystals via a multi-level optical-fibre distribution system. The basic operations for barrel geometry are the following: laser pulses transported via an optical fibre are injected at a fixed position at the crystal's front face, the injected light is collected, with the pair of APDs glued to the crystal's rear face, as for scintillation light from an electromagnetic shower. Although the optical light path is different from that taken by shower scintillation photons, this design guarantees that the light transmission is measured in the relevant region. The underlying principle is similar for ECAL endcaps; however, laser light is injected at a corner of each endcap crystal's rear face, and the light is collected (as for scintillation) via a VPT glued on the crystal's rear face.

The energy correction factor extracted by means of the laser monitoring system depends on

the light collection mechanisms of both electromagnetic showers and injected laser. It is out of scope to explain it in detail the light collection mechanism. It is possible to find a relation between the shower signal amplitude (S) and the injected laser signal amplitude (R) [61]. The demonstration begins considering the average light optical path (Λ) and the average light attenuation coefficient (λ), which is directly related to the light transmission. If we consider a shower, with initial amplitude S_0, which goes through the crystal, the measured amplitude S is:

$$S = S_0 \; e^{-\frac{\Lambda_S}{\lambda_S}}$$

with the same idea, if we consider the injected light, with initial amplitude R_0:

$$R = R_0 \; e^{-\frac{\Lambda_R}{\lambda_R}}$$

this gives:

$$\frac{S}{S_0} = \left(\frac{R}{R_0}\right)^{\frac{\Lambda_S \lambda_R}{\lambda_S \Lambda_R}} = \left(\frac{R}{R_0}\right)^{\alpha} \tag{3.3}$$

Thus the laser correction to apply to final amplitude S is:

$$\mathrm{LC} = \left(\frac{R}{R_0}\right)^{-\alpha}$$

where α is an empirical parameter.

Since Equation 3.3 is an approximation, in cases where there are big transparency losses (as in the endcaps) some modifications are applied when these features are measured, like dividing samples in η-zones so that the overall R/R_0 range does not vary too much.

Another important feature is that in the endcaps the absorbed radiation dose is much higher than in the barrel. Therefore, during the LHC running VPT response can deteriorate, leading to an inaccurate R/R_0 measurement. ECAL cannot monitor these losses, thus the net effect is a reduction on the effective α value.

Actually, crystals were grown in two different facilities, one in China (SIC crystals) and the other in Russia (BTCP crystals). The stability of the final laser monitoring system and a measure of the α was performed with *insitu* measurements during 2011 and 2012 data taking. It was verified with excellent agreement the relation between S and R and it was measured that the α of BTCP barrel crystals is 1.52, the effective α value of the encaps BTCP crystals is 1.16, whereas the measured α of the endcaps SIC crystals is 1.00.

Figure 3.4 shows the *in situ* measurement of the transparency loss during 2011, 2012 and the beginning of 2015 data taking for different η regions. The recovery of the crystals during the LHC shut-downs is evident. However the response is not fully recovered, in fact there is also a permanent damage of crystals and photodetectors, especially in the endcaps.

3.3 Energy Calibration for H $\rightarrow \gamma\gamma$ precision measurements

The H $\rightarrow \gamma\gamma$ decay channel provides a clean final state topology with a mass peak that can be reconstructed with high precision. ECAL is designed to provide excellent energy resolution in

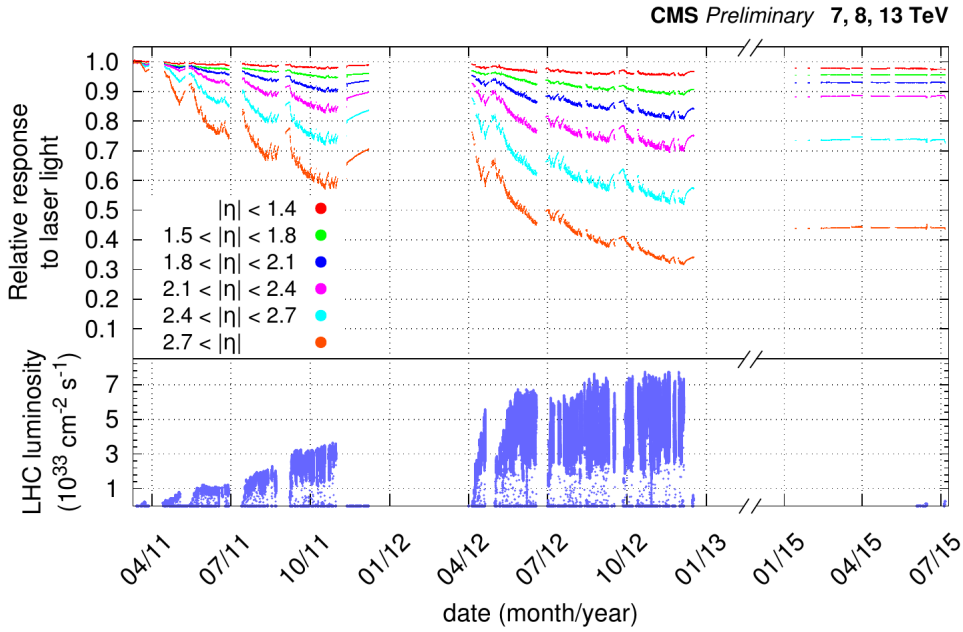


Figure 3.4: Relative response to laser light (440 nm in 2011 and 447 nm from 2012 onwards) injected in the ECAL crystals, measured by the ECAL laser monitoring system, averaged over all crystals in bins of pseudorapidity, for the 2011, 2012 and early 2015 data taking periods. The response change observed in the ECAL channels is up to 6% in the barrel and it reaches up to 30% at $\eta \sim 2.5$, the limit of the tracker acceptance. The response change is up to 70% in the region closest to the beam pipe. The recovery of the crystal response during the Long-Shutdown-1 period (from the end of the 2012 to the beginning of the 2015) is visible. The response is not fully recovered, particularly in the region closest to the beam pipe. These measurements are used to correct the physics data. The bottom plot shows the instantaneous LHC luminosity delivered during this time period.

order to maintain the advantage of such a narrow width. Because of this a precise calibration of the detector is needed in order to perform Higgs bosons precision measurements, such as the Higgs boson mass, in the $H \to \gamma\gamma$ decay.

3.3.1 Validation of the response corrections using collision data

The response corrections were tuned and validated using the energy of electrons from W boson decays, the reconstructed mass from $\pi^0/\eta \to \gamma\gamma$, and the energy resolution measured with $Z \to e^+e^-$ events. The tuning involves the optimization of the value of α, for BTCP and SIC crystals in EB and EE separately, to obtain the best *in situ* measurement of the resolution of the invariant mass of the Z boson [62].

Isolated electrons from W boson decays are used to provide an energy scale to validate response corrections over periods of days to weeks. The ratio E_{SC} /p_{tk}, which is the ratio of the electron energy, E, measured in the ECAL, to the electron momentum, p, measured in the tracker, has been used to monitor the stability and uniformity of the ECAL response. This is computed in each event and a reference E_{SC} /p_{tk} distribution is obtained at a given time, position or laser

correction. This distribution is then scaled to fit subsets of data, properly partitioned in order to measure the response relative to the reference.

To avoid biases related to the imperfect description of the data by the simulation, the reference distribution has been in general derived from data themselves. The scale factors provide a measure of the relative response and are shown in Figure 3.5 for 2011 data taking, as a function of time. For example, a stable response to electromagnetic showers is achieved throughout 2012 with an RMS of 0.09% in EB and 0.28% in EE.

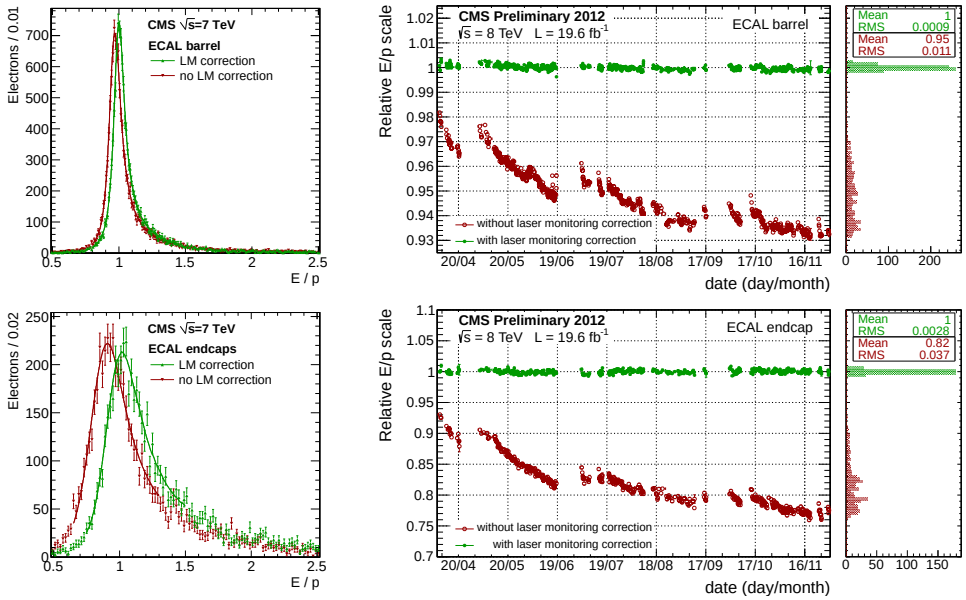


Figure 3.5: Relative energy response variation for EB (top) and EE (bottom) determined from the E/p analysis of electrons in W boson decays. Left: examples of fits to the E/p distributions before (red) and after (green) LM corrections. Middle: Response stability during the 2012 *p-p* data-taking period before (red open circles) and after (green points) response corrections; the blue line shows the inverse of the average LM corrections. Right: Distribution of the projected relative energy scales.

The response corrections for EE are calculated using an effective α value of 1.16 for all BTCP crystals. This value of α is extracted from *in situ* measurement and it is shown to give the most stable and optimal mass resolution as a function of time.

Figure 3.6 shows how the α value can be measured directly from the data, using the E/p distributions and binning them in laser corrections values. The slope of the logarithm of Equation 3.3 (red points of Figure 3.6) gives the independent α measurement. Indeed, an α value of 1.16 is measured. This is attributed to interplay between larger crystal transparency losses in EE and the VPT noise. This α value for the EE crystals makes the the corrected E/p distributions stable and improves the dispersion.

Finally, the validation of the response corrections is also carried out by monitoring the ECAL energy resolution using events with a Z boson decaying into two electrons. The mass resolution is dominated by the energy resolution of the electron reconstruction. The Z peak is fitted with the convolution between a Breit-Wigner and a Crystal-Ball (Gaussian with an exponential tail).

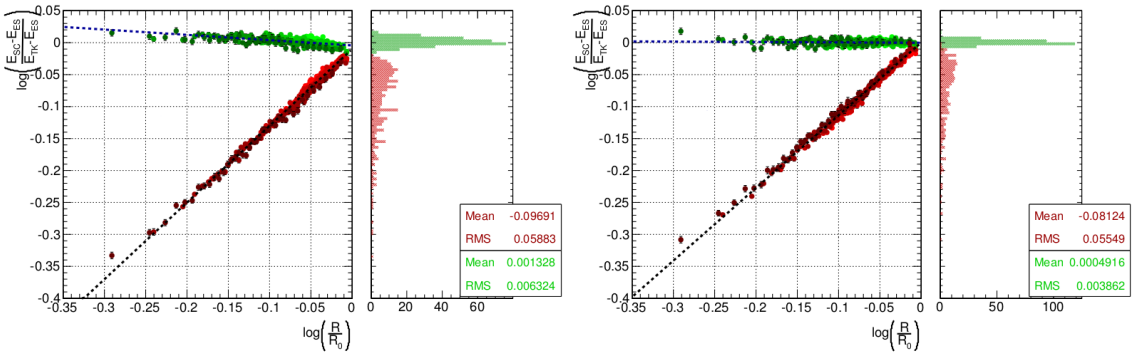


Figure 3.6: Stability plots with the E/p methods as a function of the response $\log(\frac{R}{R_0})$ for the BTCP endcap crystals. The slope of the uncorrected points (red) gives a direct measure of $\alpha = 1.16$. In the left plot the α value of 1.52 is used to correct the E/p distributions (green points), while in the right plot the value of 1.16 is used.

Figure 3.7 shows the contribution to the instrumental mass resolution for the Z boson peak, σ_{CB}/m_Z , as a function of time for events with both electrons in EB (left) or both in EE (right). The results are shown for 2011 data, however similar procedure is used also on the 2012 data. The mass resolution, after the application of the response corrections, is stable within an RMS spread of 0.1% and 0.2% for events with both electrons in EB or EE, respectively.

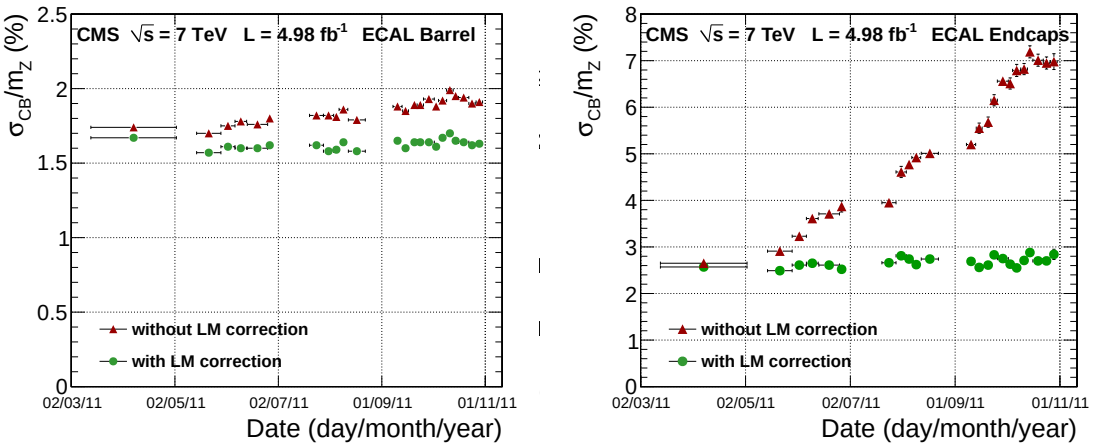


Figure 3.7: Mass resolution for the reconstructed Z boson peak, from $Z \to e^+e^-$ decays, as a function of time for EB (left) and EE (right) before (red dots) and after (green dots) LM corrections are applied.

3.3.2 Single-channel intercalibration

The ECAL channels are calibrated by using relative and absolute calibrations. Relative calibrations, C_i , between one channel and another, are referred to as intercalibrations. Intercalibration with collision data involves many methods: π^0/η, isolated electrons from W and Z decays and ϕ-symmetry. All these methods are used to intercalibrate channels at the same pseudorapidity. The precision of the intercalibrations has been studied for each method with the aid of

Monte Carlo simulations and by a cross comparison of the calibration coefficients derived by each method.

π^0/η-calibration

In order to take advantage of the high rate of π^0 decays, a specialized data taking stream has been developed. In addition, a separate calibration stream has been implemented to select $\eta \rightarrow \gamma\gamma$ decays. An iterative procedure is used to determine the intercalibration constants. The π^0/η invariant mass distribution is fitted with a Gaussian function, for the signal, and a fourth-order polynomial for the background. Then the intercalibration constants are updated iteratively to correct the fitted mass value in each channel.

Isolated electrons from W and Z decays

Once electron candidates are selected with the same selections used for the monitoring of the stability, Section 3.3.1, the L3 recursive algorithm [62] is used to calculate the intercalibration coefficients. For each l^{th} crystal, which is the seed crystal of N good electrons, the intercalibration coefficient at the n^{th} iteration of the procedure is linked to the coefficient at the $(n-1)^{th}$ step by the E/p template distribution, which depends on η. Each single crystal is assigned a weight w corresponding to the fraction of the total supercluster energy deposited in that particular channel. The intercalibration coefficients are evaluated for each single crystal and normalized to the η-ring average intercalibration. In contrast to the other methods, this intercalibration method is more limited by the statistical precision in 2011 and 2012.

ϕ-symmetry

The ϕ-symmetry method is based on the expectation that for a large sample of soft interaction events the total deposited transverse energy (E_T) should be the same for all the crystals in a ring, at fixed η. Intercalibration in ϕ is performed by comparing the total transverse energy ($\sum \mathrm{E}_T$) deposited in one crystal with the mean of the total $\sum \mathrm{E}_T$ collected by crystals at the same absolute value of η. Hence, for each ring in ϕ the average of 360 intercalibration constants c_i is equal to the unity by construction. A measurement of the precision as a function of time was performed as shown in Figure 3.8. The figure shows the standard deviation of the IC ratio with respect to the adjacent in time IC values (red) and start values (blue). The red dots are flat with time while the blue ones have a dependency mainly due to the spread variation of the laser corrections.

Combination of the intercalibration constants

The combination was obtained from a mean of the intercalibration constants in fixed ϕ rings from the π^0/η, the E/p, and the ϕ-symmetry methods, weighted on the respective precisions. The precision of each intercalibration set used in the combination has been derived by means of Monte Carlo simulation studies. In addition, the precision was estimated from the cross-comparison of

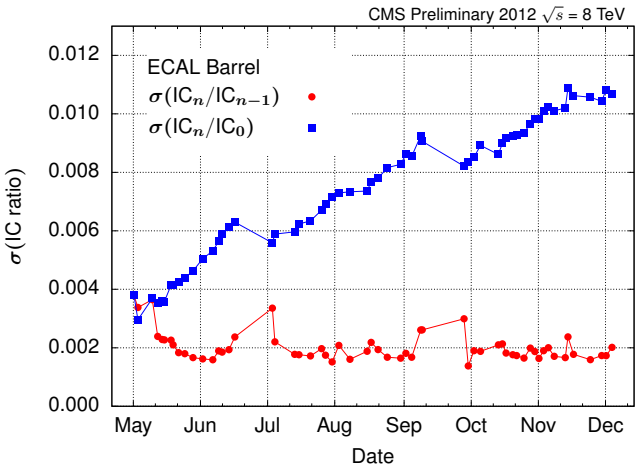


Figure 3.8: Time evolution of the standard deviation of the ratio distribution between two sets of inter-calibration constants (IC). The intercalibration is performed with the ϕ-symmetry method for individual ECAL barrel crystals using 2012 data. The standard deviation is determined by a Gaussian fit to the ratio distribution.

the results of the different intercalibration techniques.

In each ϕ-ring, the variance of the difference between the intercalibration constants for every pair of intercalibration sets was derived. This variance is assumed to be the sum in quadrature of the uncertainty of the constants in each set. Consequently, the precision of each intercalibration set was extracted by solving three simultaneous equations for the three variances. The values obtained with this method were found to be consistent with the expected precisions based on the simulation studies. Figure 3.9 shows the separate intercalibration procedures precision and the final combination as a function of η for the barrel and the endcaps. The precision of the ϕ-symmetry and photon calibrations is at the level of the systematic errors, while the precision of the electron calibration is still dominated by the statistical errors for $\eta > 1$.

3.4 ϕ-symmetry intercalibration optimization

As it shown in Figure 3.9, the ϕ-symmetry is the intercalibration method with the biggest uncertainties. However, it is the most suited method for short term monitoring to cross check stability corrections derived from laser data. In addition, the intercalibration constants coming from this method have the advantage of being constant in time. Therefore, this method can be used for the transport of the intercalibration constants from the 2012 to the beginning of the 2015 Run II data taking.

In view of the new data taking, an optimization of the intercalibration procedure has been performed. In the following sections a detailed description of the this calibration method is given, the main sources of uncertainties are described and the optimization with respect to the Run I calibration procedure is shown.

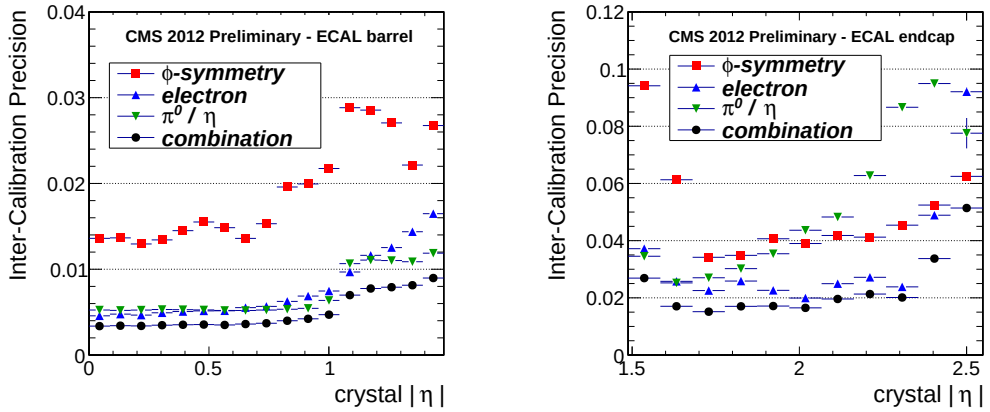


Figure 3.9: Precision of channel inter-calibration, using energy deposits, as a function of pseudo-rapidity in the ECAL barrel (left) and endcap (right) detectors. The precision for measuring the inter-calibration constants from ϕ-symmetry, from $\pi^0 \to \gamma\gamma$ and $\eta \to \gamma\gamma$ decays, and from W and Z decay electrons, is shown as a function of η in EB and EE, using 2012 data. The black points represent the precision of the combination of the three methods (weighted average).

3.4.1 Event Selection

The physics events useful for the calibration are the soft interactions. They represent by far the majority of the p-p collisions, even if they are not interesting for most of physics research purposes. Due to the huge amount of this kind of events, the events are selected online by a dedicated calibration trigger and recorded with reduced event content. This kind of output is able to record all the triggered events with a reduced event size, in order to make a negligible impact on the CMS bandwidth. During Run I for each event only the ECAL information of the crystals with an energy greater than 150 MeV in EB and 750 MeV in EE was saved in the output.

This choice of thresholds is driven by the necessity to have a small output size and to require enough energy to be much above the equivalent noise. Because of the large response variation, it is difficult to choose a unique energy selection in EE. In fact, even if the mean of the noise distribution is flat in η, the response loss and therefore the laser correction strongly depends with η, giving also an η dependence to the final energy associated to the equivalent noise.

Figure 3.10 shows the effect of the laser correction on the conversion from ADC noise to the equivalent energy in the EE. For these reasons, during Run II it has been decided to select on the crystal ADC amplitude, in order to directly cut the electronic noise. The electronic noise during the 2015 data taking is foreseen to span from values similar to the values of September 2012 data to values similar to the end of that Run I, shown in Figure 3.2. For these reasons, a selection of 8 ADC (12 ADC) in EB (EE) is chosen for the Run II, which is almost six times the mean of the noise distributions, therefore well above it.

3.4.2 Offline Selections

The ϕ-symmetry intercalibration is performed within a constant η ring of crystals by comparing the total transverse energy, $\sum E_T$, deposited in a crystal with the mean $\sum E_T$ of the crystals

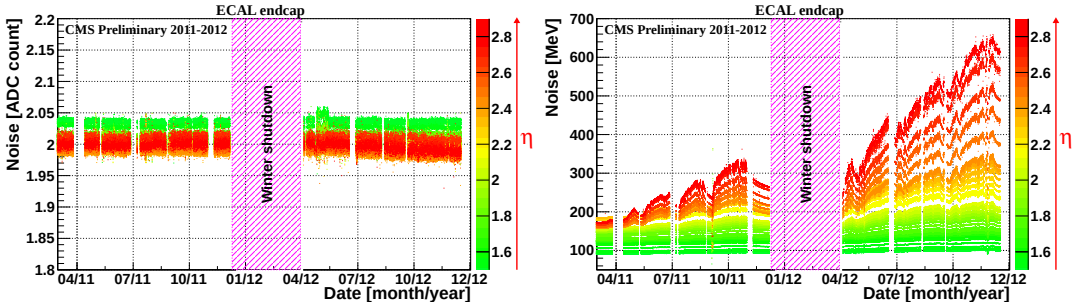


Figure 3.10: Single channel noise measured on the pre-samples of the laser events taken during standard monitoring sequences in 2011 and 2012. Left: ADC noise plots as a function of η and time, Right: left plot after the application of the ADC2GeV factor, intercalibration and Laser Correction.

in the ring. In practice, the transverse energy sums are truncated, summing the values of E_T of all deposits with E_T between fixed upper and lower thresholds. The truncation is necessary for removing the noise component in the low part of the spectrum, whereas the upper threshold is applied to minimize the fluctuations induced by rare deposits of very high E_T in the short time scale of the ϕ-symmetry intercalibration procedure.

Figure 3.11 shows a simulation of the energy distribution for the ring of crystals at $\eta = 0$ for $60 < \mathrm{E} < 200$ MeV. There is a distinct change in the slope of the distribution at about 120 MeV, below which the noise is dominant. A Gaussian is superimposed on this plot, showing that the part of the distribution below about 120 MeV is dominated by the noise.

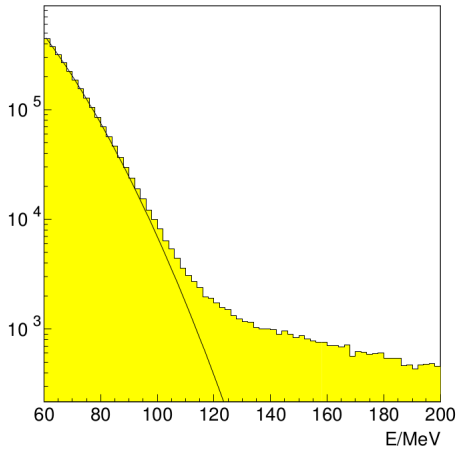


Figure 3.11: Simulation of energy distribution for $60 < \mathrm{E} < 200$ MeV for 500000 minimum bias events, for a ring of 360 crystals at $\eta = 0$. A Gaussian MeV is superimposed, normalized to the total number of events, showing the effect of the noise.

As the case of the selections for the trigger output, the lower threshold in EB may be constant since the equivalent noise is approximately constant, whereas in EE an η dependence is necessary. The choice of the threshold is driven by the necessity to reach a statistical precision of order of 0.1% within an LHC fill, i.e. $\sim$10 hours. To reach a similar precision, about 1000 hits per crystal are necessary.

Extensive occupancy studies have been performed using the simulation with different LHC Run II conditions: 20 or 40 interactions per bunch-crossing (PU) and 50 or 25 ns bunch-spacing (bx). The values of the thresholds are chosen multiplying the ADC threshold for the energy conversion, IC×LC×ADCtoGeV, Section 4.1, which is taken from ECAL conditions similar to those foreseen for the 2015. The average energy conversion value in the barrel is 40 MeV/ADC, while in the endcaps it has a parabolic η dependence.

In order to perform the intercalibration as accurately as possible, an upper threshold on the energy must be applied in addition to the lower one. Indeed, a few very high E_T hits can have a significant effect on the value of the summed transverse energy. The variation of the intercalibration precision versus the upper threshold was studied for a range of values of η in both the barrel and the endcaps. It was found that the attainable precision is not very sensitive to the threshold value which are more than a few 100 MeV above the lower threshold. A value for the upper energy threshold of 1GeV $\times \cosh \eta$ above the lower threshold is found to be always close to optimal and is applied for all the η rings.

Occupancy studies for the barrel and the endcaps, shown in Figure 3.12 and Figure 3.13, have proven that an ADC lower threshold of 13 ADC counts in EB or 20 ADC counts in EE, which is then multiplied to the energy conversion, is high enough to make the noise contribution negligible while reaching a 0.1% statistical precision. The optimized lower and upper energy thresholds are shown in Figure 3.14. The thresholds for Run II are higher than those for Run I, because of this optimization of the energy thresholds in order to reduce the impact of the electronic noise.

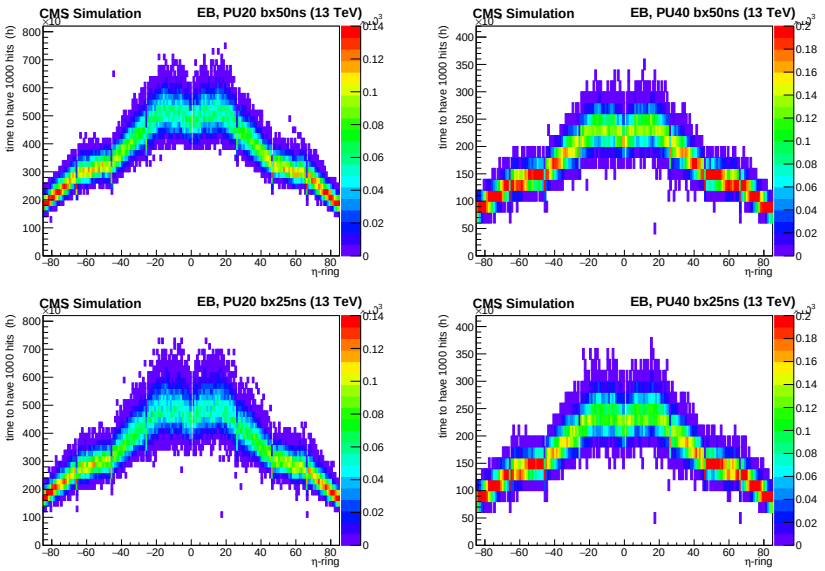


Figure 3.12: Occupancy studies for different LHC simulated conditions of the barrel. For all the crystals a precision of about 0.1% can be reached within 10 h.

In conclusion Figure 3.15 and Figure 3.16 show some of the truncated E_T and $\sum E_T$ distributions in EB after the offline selections obtained with 1 million events coming from the ϕ-symmetry trigger, for individual rings at three values of $|\eta|$. Figure 3.17 and Figure 3.18 show similar distributions for the endcaps.

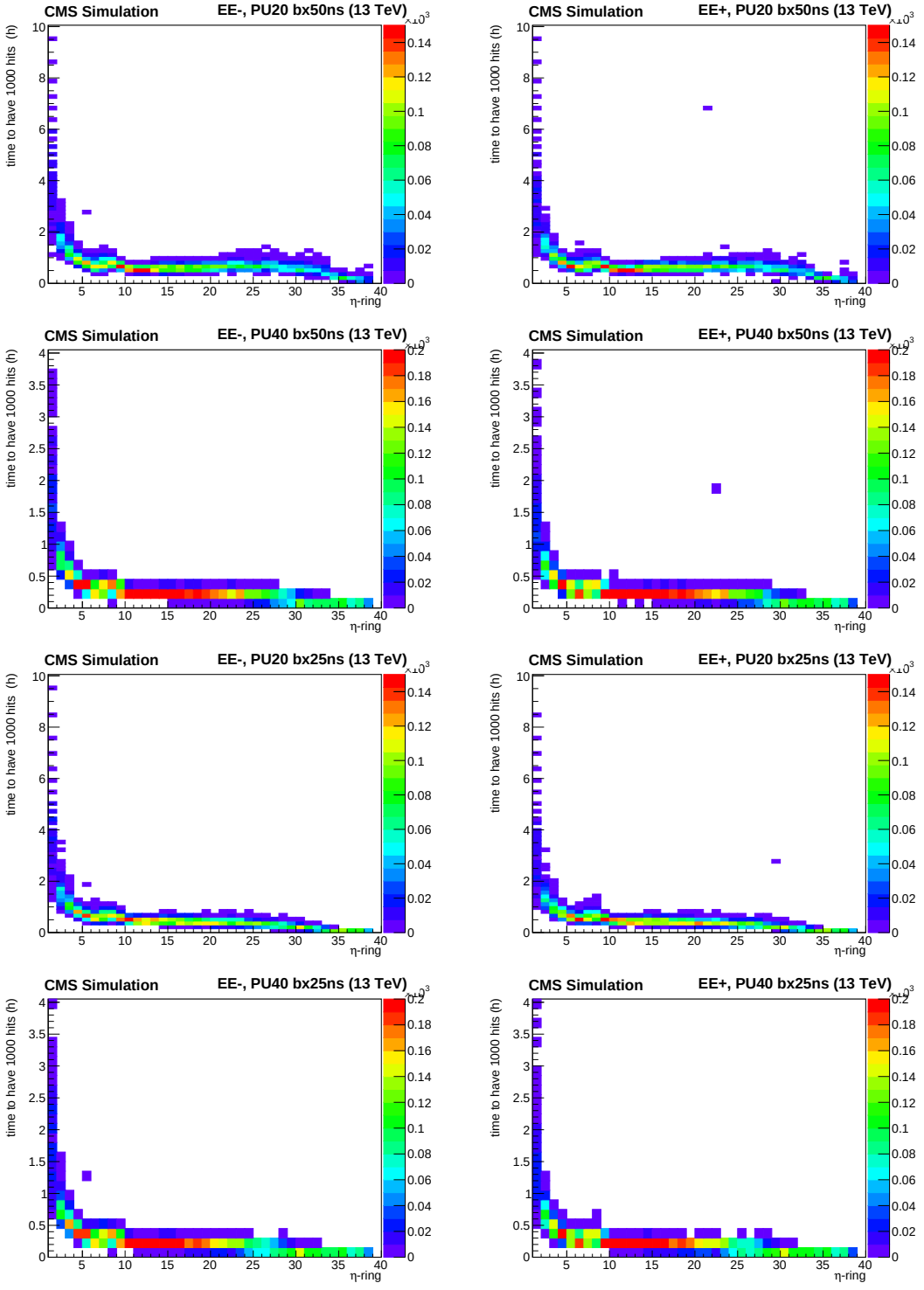


Figure 3.13: Occupancy studies for different LHC simulated conditions of the endcaps. For most of the crystals a precision of about 0.1% can be reached within 10 h.

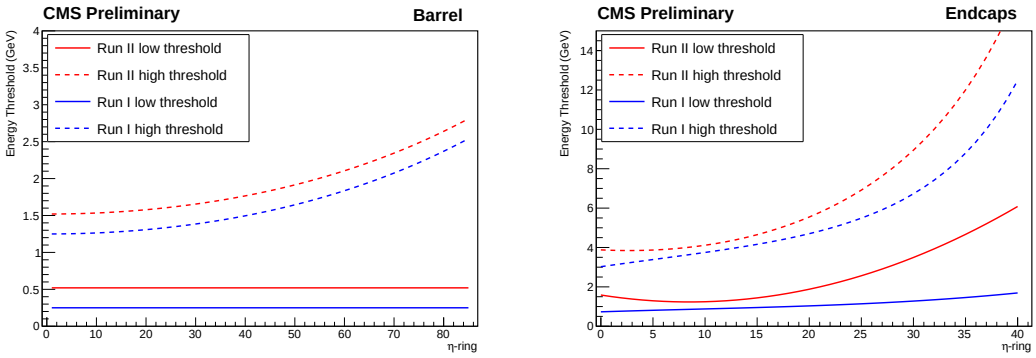


Figure 3.14: Offline energy thresholds used for Run II compared with the ones for the Run I for EB (left) and for EE (right).

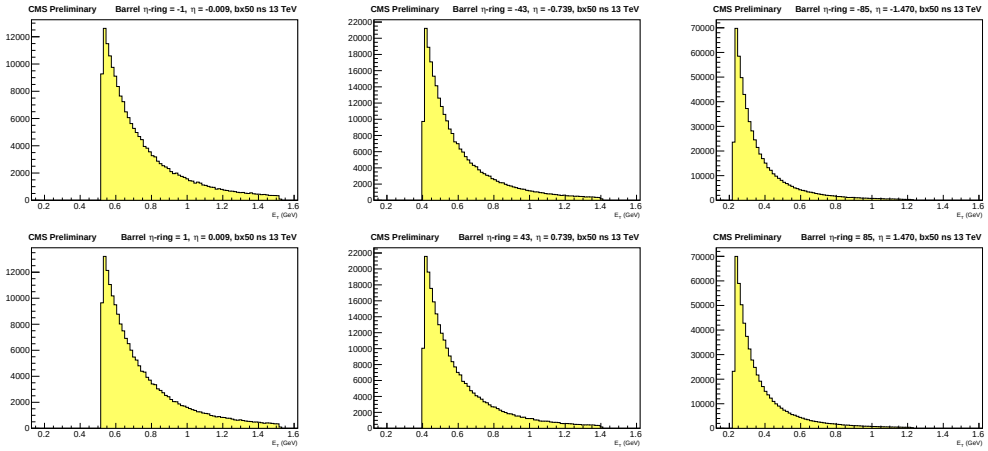


Figure 3.15: Examples of truncated E_T distributions obtained with 1 million of events coming from the ϕ-symmetry trigger at 13 TeV energy, for individual EB rings at six values of $\eta = \pm 0.009$, ± 0.739 and ± 1.470.

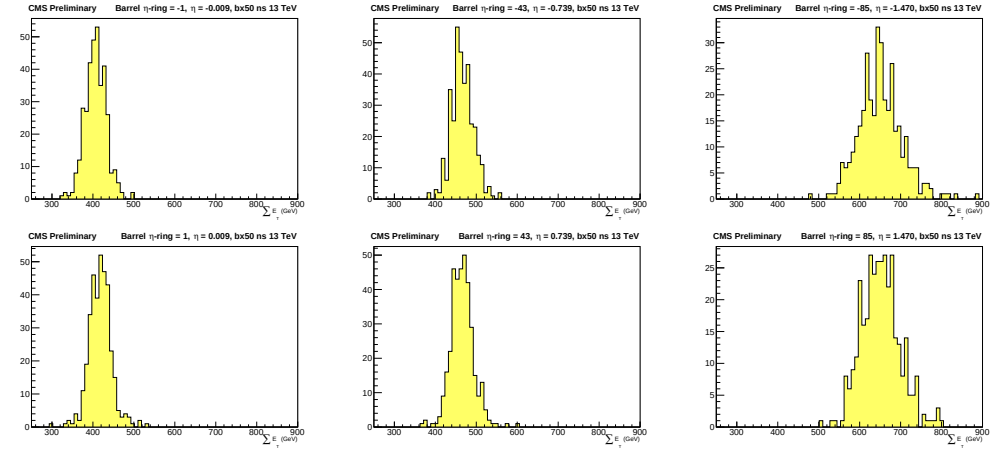


Figure 3.16: Examples of truncated $\sum E_T$ distributions obtained with 1 million of events coming from the ϕ-symmetry trigger at 13 TeV energy, for individual EB rings at six values of $\eta = \pm 0.009$, ± 0.739 and ± 1.470.

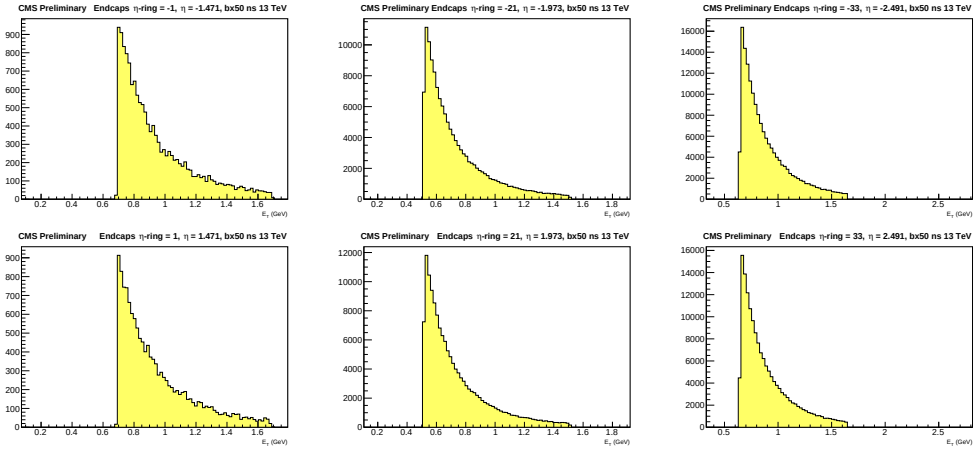


Figure 3.17: Examples of truncated E_T distributions obtained with 1 million of events coming from the ϕ-symmetry trigger at 13 TeV energy, for individual EE rings at six values of $\eta = \pm 1.471$, ± 1.973 and ± 2.491.

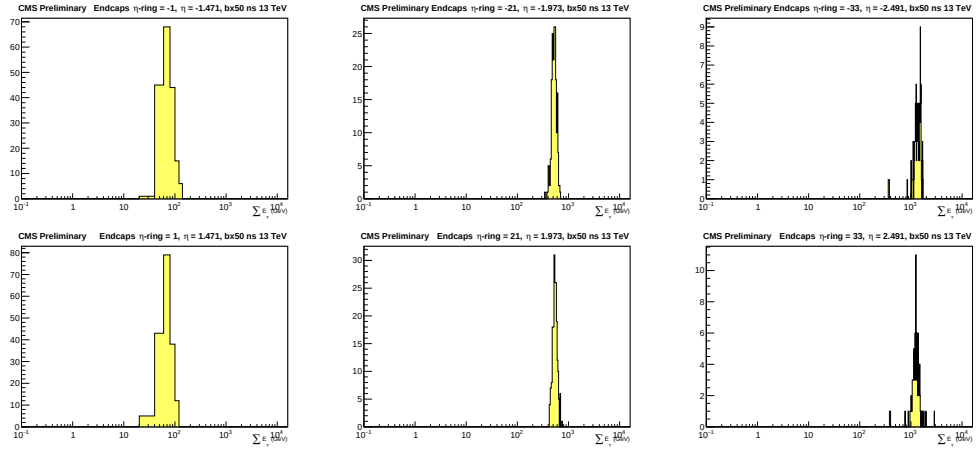


Figure 3.18: Examples of truncated $\sum E_T$ distributions obtained with 1 million of events coming from the ϕ-symmetry trigger at 13 TeV energy, for individual EE rings at six values of $\eta = \pm 1.471$, ± 1.973 and ± 2.491.

3.4.3 Intercalibration Computation

Due to the fact that the E_T sum is obtained from a truncated E_T distribution, a small change in the calibration constant of a crystal will result in a proportionate change to the measured $\sum E_T$ in that crystal, with constant of proportionality, k-factor, different from unity. These k-factors are determined empirically for each ring of crystals. Figure 3.19 shows the fractional change ϵ_T in the measured $\sum E_T$, when crystals are given a miscalibration ϵ_M, i.e. when the energy is multiplied by $1 + \epsilon_M$. For the range of miscalibrations shown, the relationship between ϵ_T and ϵ_M, the k-factor, is given by a straight line with a slope of 1.6.

A new set of intercalibration constant is always derived on top of a previous set. Therefore, the output of the ϕ-symmetry method is a relative correction factor to be applied to the intercalibrations injected in the reconstruction of the crystal energy. Once the k-factor is computed

the final ϕ-symmetry correction factor to be applied to a crystal is given by:

$$\mathrm{Corr} = \left[\left(\frac{\sum E_T}{< \sum E_T >} - 1 \right) \cdot \frac{1}{k} + 1 \right]^{-1}$$

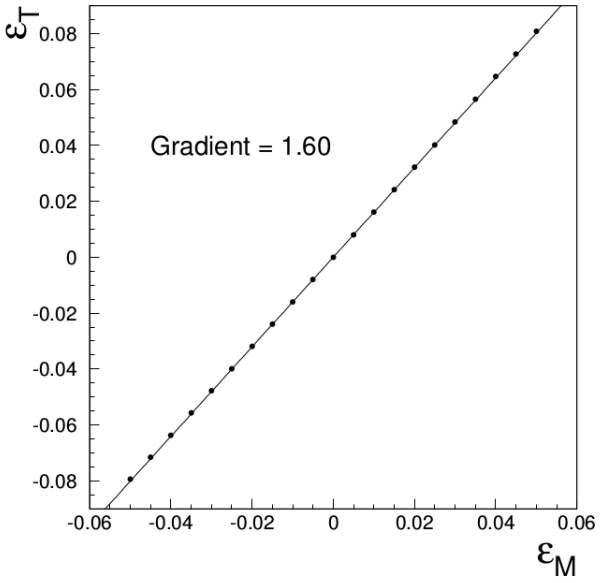


Figure 3.19: Fractional change ϵ_T in the total energy measured as a function of the introduced miscalibration ϵ_M , for crystals in the two rings at $|\eta| = 0.23$.

3.4.4 Intercalibration Precision

Assuming no systematic effect and no truncation in the E_T distributions, the precision would scale as $1/\sqrt{N}$, where N is the number of hits in the crystal. Since the systematic effects and the truncation are not negligible, the final precision as a function of the number of hits is:

$$\sqrt{c^2 + \frac{s^2}{N^2}}$$

where s takes into account the dilution in sensitivity related to the k-factors and c parametrizes all the systematic uncertainties (noise, geometric asymmetries and material in front of ECAL). With enough amount of data (i.e. ~ 1000 events per crystal) the precision is dominated by the systematic uncertainties.

Regarding the various components of systematic uncertainties, as a consequence of the 3° off-pointing arrangement of crystals, crystals on one side of an inter-module gap receive a greater number of hits than other crystals, as illustrated in Figure 3.20. This causes a systematic effect that has to be taken into account in order to compute correct intercalibration constants. In order to mitigate this effect a module-by-module correction factor is extracted averaging the average intercalibration differences.

An other systematic effect is given by the variation with azimuth of the amount of material

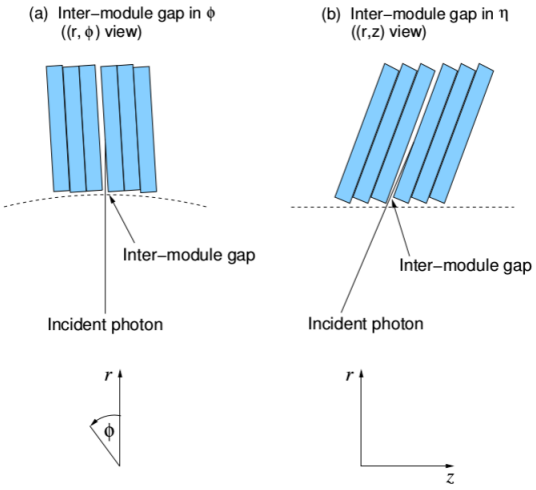


Figure 3.20: Illustration of the effect of the presence of inter-module gaps. The incident photon travels in a straight line from the interaction point. Due to the 3° off-pointing arrangement, such a particle can enter crystals on one side of an inter-module gap from a side face as well as from the front face. Such crystals therefore receive more hits than other crystals.

between the calorimeter and the interaction point. The material causes the conversion of photons coming from the interaction point. Therefore, any asymmetry in the material distribution directly affects the intercalibration, as crystals in the same η-ring do not result in similar energy spectra.

This systematic effect can be reduced evaluating a correction factor coming from the difference of material effects between data taken with and without the magnetic field. In fact, charged particles which are affected by the magnetic field spread in ϕ, passing through more material in front of ECAL.

Finally the effect of the noise on the intercalibration calculation is mitigated choosing properly the offline thresholds, as described in Section 3.4.2.

A precise and independent intercalibration of the crystals, for example using the E/p of electrons from W decay, provides a measure of these systematic variations. During the 2012 data taking, using independent calibration algorithms, a total systematic uncertainty associated to the ϕ-symmetry of about 2% was estimated. Similar procedures will be used to estimate the systematics uncertainties also with Run II data.

The ϕ-symmetry intercalibration procedure started as an early calibration method, due to its short time needed for reaching an uncertainty dominated only by the systematics. Then, during the 2011 and 2012 data taking, it was also used as a method to monitor the ECAL stability, as the ϕ-symmetry is stable with time. Furthermore, because of this, the procedure is fundamental to transport the ECAL detector conditions from the 2012 to the Run II data taking. Hence, even if the ϕ-symmetry is the intercalibration method with the highest systematic uncertainties, it is fundamental and extensively used also during the 2015 Run II data taking.

CHAPTER 4

TUNING OF MONTE-CARLO SIMULATION OF ECAL AND IMPACT ON THE H→ $\gamma\gamma$ SEARCH

The baseline software framework for the full ECAL simulation was set up before the start of the LHC operation and at first validated during test-beam campaigns. With the start of the LHC collisions and the evolution of the data taking conditions, efforts have been made to continuously improve the simulation description. Of particular importance are a realistic description of the ECAL noise and the channels response evolution, representative of the conditions measured in data. Such elements have indeed high impact on cluster shape variables and the energy resolution, whose accurate description in the Monte Carlo simulation is fundamental for physics studies. Among them the most sensitive target is the H $\rightarrow \gamma\gamma$ analysis [63], which makes a great use of the cluster shape variables for photon identification and event classification by means of multivariate analyses and energy corrections, in particular. At the end of Run I data, an optimize version of the Monte Carlo simulation was developed to improve the sensitivity and the precision of the H $\rightarrow \gamma\gamma$ analysis. In this chapter the Monte Carlo simulation of the ECAL detector and its impact on the H $\rightarrow \gamma\gamma$ analysis is described. In particular, the tuning of the MC simulation conditions has improved the analysis sensitivity of about 10% with respect to the previous results [64]. Other refinements of the analysis, and in particular a detailed study of the photon energy scale and response linearity, have permitted the mass measurement with the precision of 0.12%. These studies were not conducted within this thesis work and will not be described in detail. In addition, any improvement in the H $\rightarrow \gamma\gamma$ analysis affects also the search of $\gamma\gamma b\bar{b}$ final states, Chapter 5.

4.1 General description of the ECAL simulation

The ECAL simulation within the official CMS software (CMSSW) framework is organized in different blocks. The Geant4 [65] software is used to track particles in the magnetic field, through the detector itself. Particles in each event are generated according to the information stored in the physics list integrated in Geant4. The initial particles are then propagated through the detector relying on their specific properties and the following behaviour of the sensitive components of the detector is simulated. In this first stage, the interaction of particles with the

detector is considered. Relying on models for the electromagnetic shower development, the energy deposited by particles in each ECAL crystal is estimated and stored in the CaloHit object. In a second stage, the effect of the detector electronics is also considered, updating the estimate of the CaloHit energy [66].

Within the ECAL simulation, a simplified description of the physics processes involved is used, to deal with the emission of the scintillation light and its propagation through the detector, as well as the explicit behaviour of APD and VPT.

Upon the Geant4 event generation, several energy deposits are simulated along each crystal, belonging to the development of the electromagnetic shower and summed together:

$$E_{CaloHit} = \sum_i E_i \times f_i$$

where f_i takes into account the ECAL channel response variation in the different crystal position, with values which range from 1 to 1.2.

Afterwards, the estimated energy is converted into an average number of photo-electrons (GeVToNpe), expected in each channel, according to laboratory measurements. The conversion factors used are 2.25 for EB and 1.8 photo-electrons/MeV for EE. The aim of this conversion is to correctly simulate the Poisson photo-statistic fluctuations that take part in the conversion process and that are the main contribution to the stochastic term in the energy resolution. These conversion factors already account for the additional excess noise factor due to the fluctuations in the avalanche process in the sensitive devices. The fluctuation is expressed as:

$$E_{SimHit} = \text{Poissonian}\left(\frac{E_{CaloHit}}{\text{LC} \times \text{GeVToNpe}}\right) \times \text{NpeToGeV}$$

Where LC is the laser correction. It is injected at this step because it is important to simulate the response variation as it directly affects the signal/noise ratio.

The simulated energy measurements are finally passed through the modelling of the detector electronics response, to provide the same digitisation output that is coming from the real detector. The electronic shaping of the ECAL readout is simulated using proper digitisation shapes for both barrel and endcap crystals, which are derived from test beam measurements. The shapes are shown in Figure 4.1.

In order to deal with a simulation output coherent with the one measured during the LHC data taking, the simulation of the digitisation step is performed this way:

$$ADC_i = \text{Digitisation}\left(\frac{E_{SimHit}}{\text{ADCToGeV} \times \text{IC}}\right) \oplus \text{Noise}$$

where ICs are the intercalibration constants and ADCToGeV is the conversion factor from ADC counts to energy. This digitization is performed on the deconvolution of the energy from the intercalibration constants, in order to take into account for the effect of the noise in the digitization step. Once the digitization is simulated, the energy is reconstructed in two steps, as it is done for the data: amplitude reconstruction and energy calibration, Equations 3.2 and 3.1

$$E_{Uncalibrated} = \sum_{i=0}^{10 \text{ digis}} \text{ADC}_i \times w_i$$

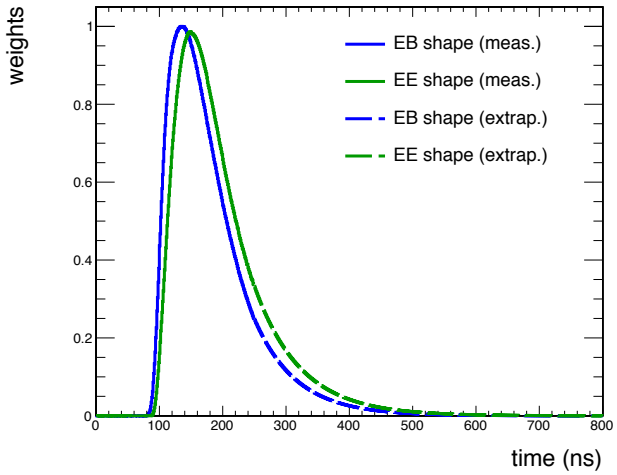


Figure 4.1: MGPA signal shape implemented in the simulation of the electronic readout. The pulse shapes measured at the test beam (solid line) are shown for both EB (blue) and EE (green) with the analytical extrapolation (dashed line) superimposed.

$$E_{RecHit} = E_{Uncalibrated} \times \text{ADCToGeV} \times \text{IC} \times \text{LC}$$

As for the data energy reconstruction, ICs and LCs are fundamental for the correct simulation of the equivalent energy of the signal and the noise. Over the years, the re-implementation of the noise model in CMSSW has been refined with great care for the description of all the details of the electronics.

4.2 Update of conditions

The update campaign for the ECAL simulation of 2013 aimed at providing simulated samples with a realistic description of the detector response in 2012. The main updates are:

- **Simulation time-window**: the time-window for the pileup simulation is extended, from [-50 ns, 50 ns] to [-300 ns, 50 ns], in order to better take into account for the out-of-time pileup (OOT PU). In fact OOT PU signals, early with respect to in-time events, contribute with a negative weight to the amplitude reconstruction. The extension of the digitisation window, from the previous [-50ns,+50ns], brings to an average decrease of the energy measured in the ECAL. This new window essentially covers the whole range used for the integration of the ECAL signal pulse shape.
- **Noise:** The conditions to describe the 2012 data are extracted from specific pedestal runs representative of the mean noise during three different 2012 periods and uploaded in the database, to be used in inputs to the noise simulation. Figure 4.2 shows the three different noise scenarios as a function of η for EB and EE (to be compared with Figure 3.2).
- **Response loss:** As for the noise conditions, three different scenarios have been chosen for the response loss to be put into the simulation. Figure 4.3 shows the three different laser correction scenarios as a function of η for EB and EE (to be compared with Figure 3.4).

where the three different conditions correspond to 2012 Run A+B, 2012 Run C and 2012 Run D data taking periods. The MC simulated events are generated with a probability for having one of the three conditions given by the ratio of the accumulated luminosities, $p(RunA+B) = 5.3/19.7 = 26.9\%$, $p(RunC) = 7.1/19.7 = 36.0\%$ and $p(RunD) = 7.1/19.7 = 37.1\%$.

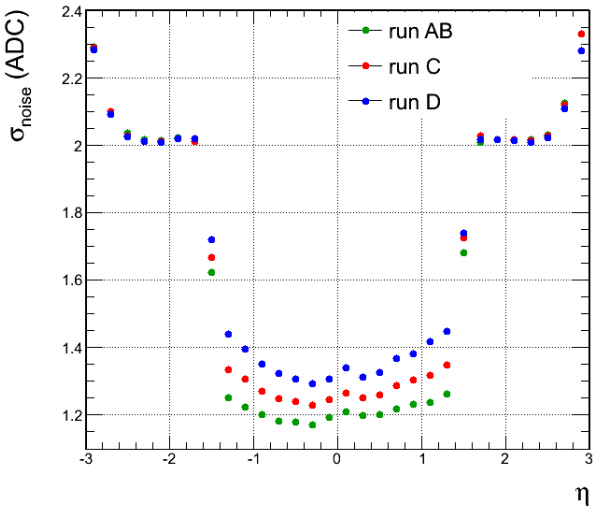


Figure 4.2: The conditions representative of the mean noise during run 2012 A+B, C and D used as inputs to the simulation are shown in ADC counts, as average values versus η.

4.3 Validation

A validation was performed upon re-digitisation of Drell-Yan events to include the updated inputs for noise and response loss description, which allowed to test the proper description of the variables useful for physics. Such production, made with the pileup scenario representative of the data taking allowed the careful validation of the MC simulation against data. A full comparison between the standard MC simulation (in red), the new one upon re-digitisation with Run2012D conditions (in blue) and data corresponding to the same Run2012D (in black dots) is shown in Figure 4.4, for the ECAL isolation variables, both in the barrel and the endcap part, and for the average calorimetric energy density ρ.

The effect of the OOT PU is visible, in fact the improved description of the OOT PU reflects in the good agreement between the ρ variable distributions for data and the re-digitised MC simulation. Finally, in Figure 4.5 the impact of the updated input to the simulation is shown in terms of energy resolution for the ECAL, which is computed for the standard MC simulation, the new MC simulation and data. The $Z \to e^+e^-$ events are used as a reference. The comparison clearly shows how the resolution is degraded in the new MC simulation, especially in the endcaps and in the central EB region, with a progressive degradation with time (blue vs orange and yellow), approaching the performance for data. Some residual discrepancy between the new MC simulation and data is still present and is possibly due to the mismodelling of the simulated material budget upstream ECAL.

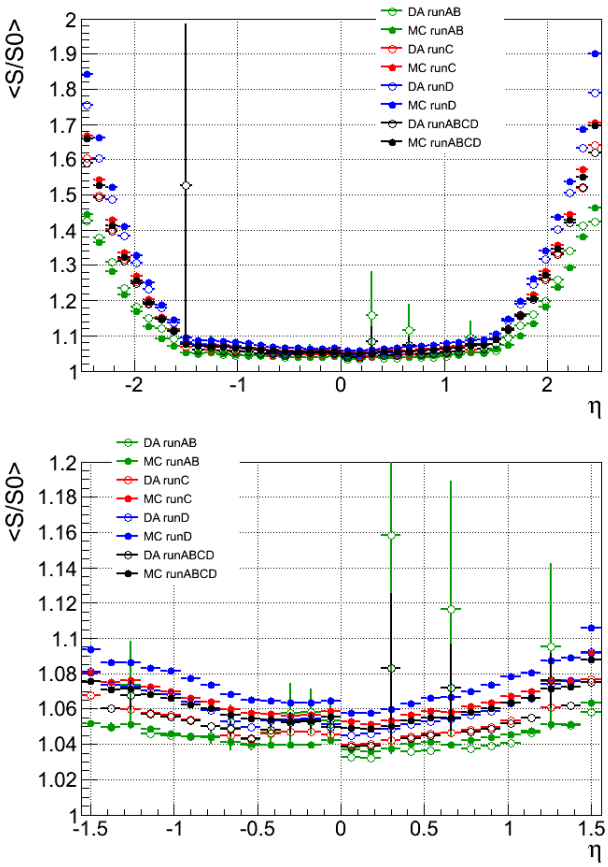


Figure 4.3: Conditions representative of the mean response loss during run 2012 A+B, C and D used as inputs to the simulation are shown as average values versus η: over the whole EB-EE range on the top and with focus on the EB only region on the bottom.

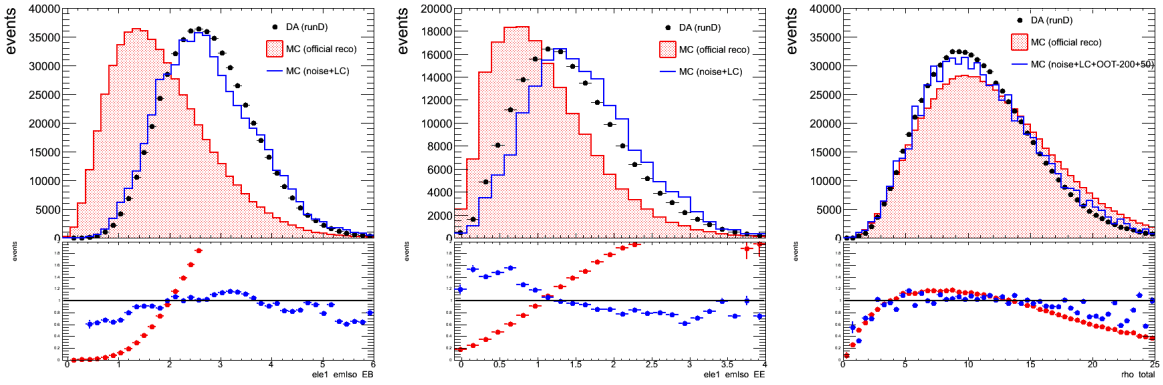


Figure 4.4: Ecal isolation for single electrons in the EB (left), in the EE (centre) and energy density of the event ρ (right). The distributions are shown for the standard MC simulation (red), the one upon re-digitisation with updated conditions (blue) and data (black dots). Conditions representative of the Run 2012D are chosen for the comparison, as those where the maximum change is expected.

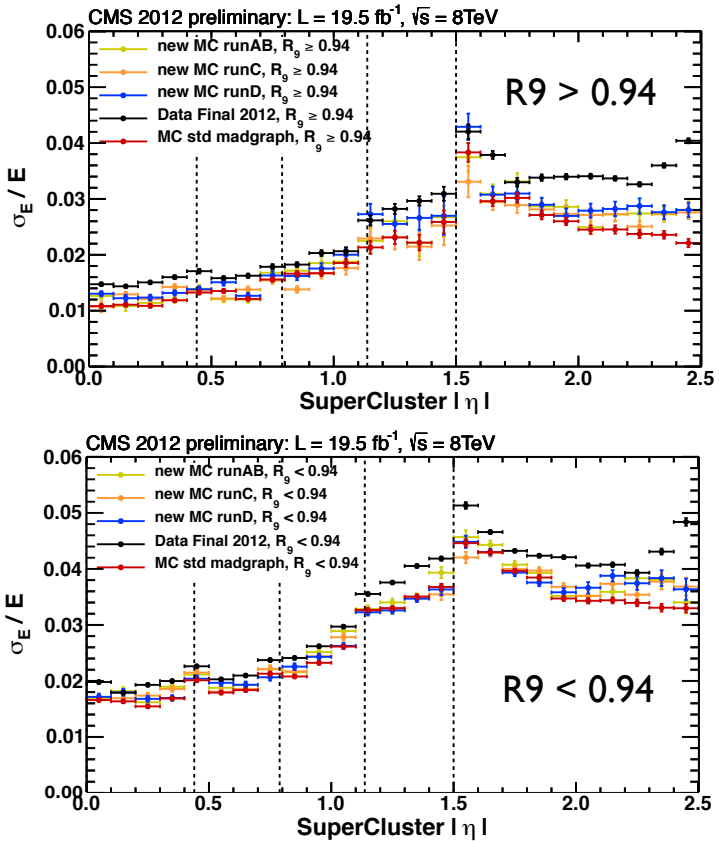


Figure 4.5: ECAL energy resolution as a function of η for $R_9 > 0.94$ electron selection (top) and $R_9 < 0.94$ (bottom). The comparison is shown between data (black), standard MC simulation (red), and the simulation accounting for the noise and response loss conditions representative of 2012 period A+B (yellow), period C (orange) and period D (blue).

4.4 Photon reconstruction performances

The implementation on these ECAL conditions in the simulation affects directly the photon reconstruction performances. This work is described in detail in [59] and a direct investigation is not performed in this thesis. However, for sake of completeness a description of the impact on the photon performances and on the H → γγ analysis is given.

A correction function, $F_{e,\gamma}$, derived from MC simulation, is applied to the supercluster energy to account for energy containment effects, including both shower containment in the calorimeter, and energy containment in the supercluster for particles that shower in the material in front of ECAL. The energy corrections have been tuned for electrons and photons separately to account for the differences in the way they interact with the material in front of the ECAL. The primary validation tool for this corrections is to compare data and MC simulation performance for electrons in Z and W boson decays.

Corrections for photons have been optimized using a multivariate regression technique based on a boosted decision tree (BDT) implementation. The regression has been trained on prompt photons (from γ+jets MC simulated samples) using the ratio of generator level photon energy to the supercluster energy (including the preshower energy) as the target variable.

The input variables are the η and ϕ coordinates of the supercluster, a collection of shower shape variables and a set of local cluster coordinates to measure the distance of the clusters from crystal boundaries within ECAL. The number of primary vertices is also included as input to the BDT in order to correct for the dependence of the shower energy on spurious energy deposits due to pileup events. The local coordinates provide information on the amount of energy which is likely to be lost in crystal, module gaps and cracks, and drive the level of local containment corrections predicted by the regression. The other variables provide information on the likelihood and location of a photon conversion, and the degree of showering in the material and they are correlated with η and ϕ.

The very good agreement between data and MC simulation on a shower shape variable is shown in Figure 4.6 using single photons coming from $Z \to \mu^+\mu^-\gamma$ decays. In addition, the energy resolution is optimized using simulated electrons or photons. As it was verified with $Z \to e^+e^-$ events that the MC simulation provides a good description of data, Figures 4.7 and 4.8, the simulation of $H \to \gamma\gamma$ events can be used to estimate the photon energy resolution. This is shown in Figure 4.8 as a function of pseudorapidity for low and high R_9 photons.

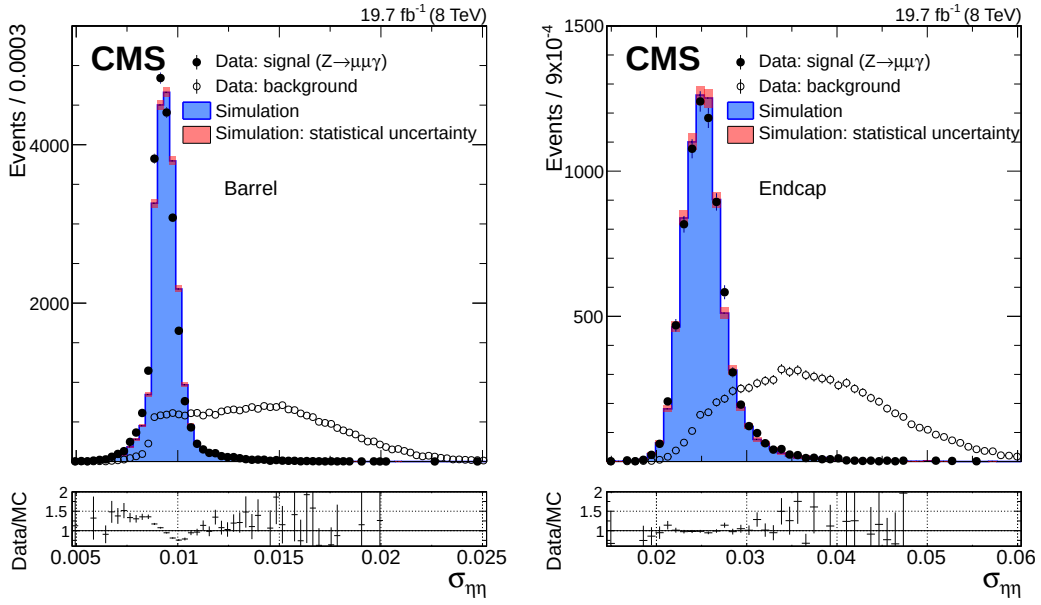


Figure 4.6: Distribution of the shower-shape variable, $\sigma_{i\eta i\eta}$, for FSR photons in $Z \to \mu^+\mu^-\gamma$ events in data (solid circles) and simulation (histogram), and for background-dominated photon candidates in dimuon triggered events (open circles). The barrel and endcaps are shown separately. The simulated signal and background distributions are normalized to the number of signal photons in the data. The ratios between the photon signal distributions in data and simulation are shown in the bottom panels.

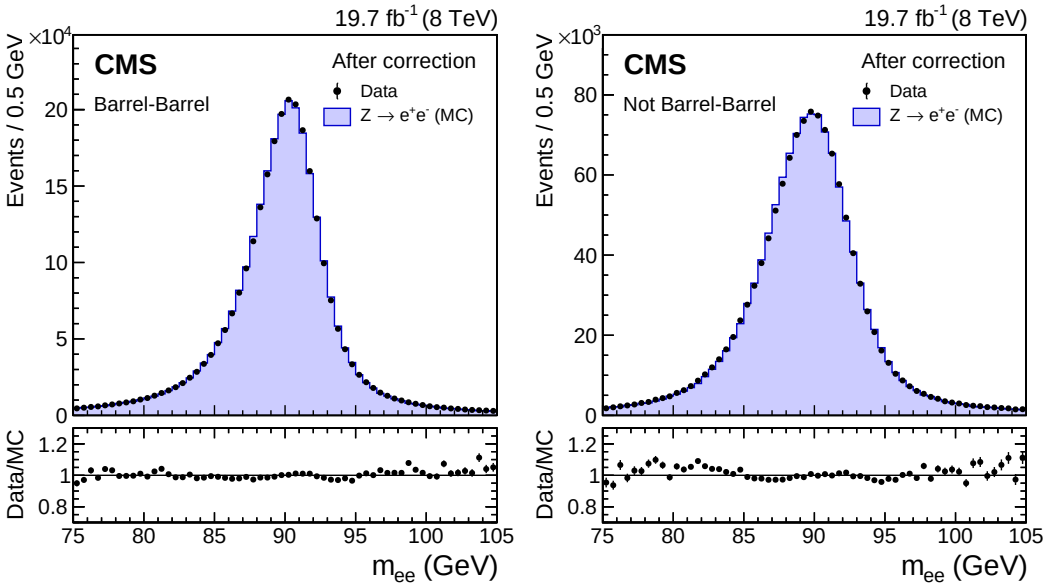


Figure 4.7: Invariant mass of e^+e^- pairs in $Z \to e^+e^-$ events in the 8 TeV data (points), and in simulated events (histogram), in which the electron showers are reconstructed as photons, and the full set of photon corrections and smearings are applied. The comparison is shown for (left) events with both showers in the barrel, and (right) the remaining events. For each bin, the ratio of the number of events in data to the number of simulated events is shown in the lower main plot.

4.5 Impact on the $H \to \gamma\gamma$ precision measurements

In this section, a description of the $H \to \gamma\gamma$ analysis, aiming to the measurement of the Higgs boson properties, is presented. The photon candidates are reconstructed from energy deposits in the ECAL using algorithms described in 3.1. Since the algorithms do not make any hypothesis as to whether the particle is a photon or an electron, electrons from $Z \to e^+e^-$ events provide measurements of the photon trigger, reconstruction and identification efficiencies, and are proxy for the photon energy scale and resolution. Differences in response between electrons and photons contribute to the systematic uncertainty of the method.

After the reconstruction the photon candidates are required to be within the fiducial region $|\eta| < 2.5$, excluding the barrel-endcap transition region $1.44 < |\eta| < 1.57$, where the photon reconstruction is suboptimal. In order to obtain the best energy resolution, the calorimeter signals are calibrated and corrected for several detector effects using methods described in 3.3.

4.5.1 Photon identification

A boosted decision tree (BDT) is trained to separate prompt photons from photon candidates resulting from misidentification of jet fragments passing the preselection requirements. The following variables are used as inputs to the photon identification BDT:

- Lateral shower shape variables. No significant differences are observed between these variables in data and in MC simulated $Z \to e^+e^-$, $Z \to \mu^+\mu^-\gamma$ events.

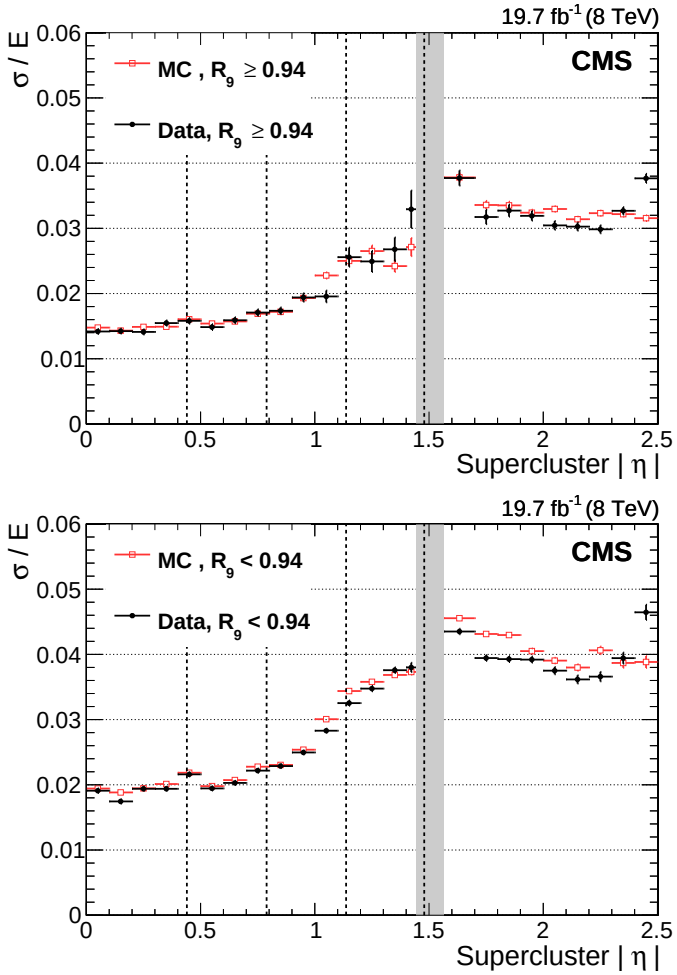


Figure 4.8: Relative photon energy resolution measured in small bins of absolute supercluster pseudorapidity in $Z \to e^+e^-$ events, for data (solid black circles) and simulated events (open squares), where the electrons are reconstructed as photons. The resolution is shown for (upper plot) showers with $R_9 > 0.94$ and (lower plot) $R_9 < 0.94$. The vertical dashed lines mark the module boundaries in the barrel, and the vertical grey band indicates the range of $|\eta|$, around the barrel/endcap transition, removed from the fiducial region.

- Isolation variables, based on the Particle-Flow algorithm [67], summing the p_T of photons and of charged hadrons, within regions of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$ around the candidate.
- The energy median density per unit area in the event, ρ, introduced to take into account the pileup dependence of the isolation variables.
- The pseudorapidity and energy of the supercluster corresponding to the reconstructed photon.

Figure 4.10 shows the photon identification BDT score for events passing the preselection in the 8 TeV dataset, for simulated background events and for signal events. The agreement of their BDT score distribution with that in data is good. The bump at 0.1 value corresponds to events where both photons are prompt and, therefore, signal-like.

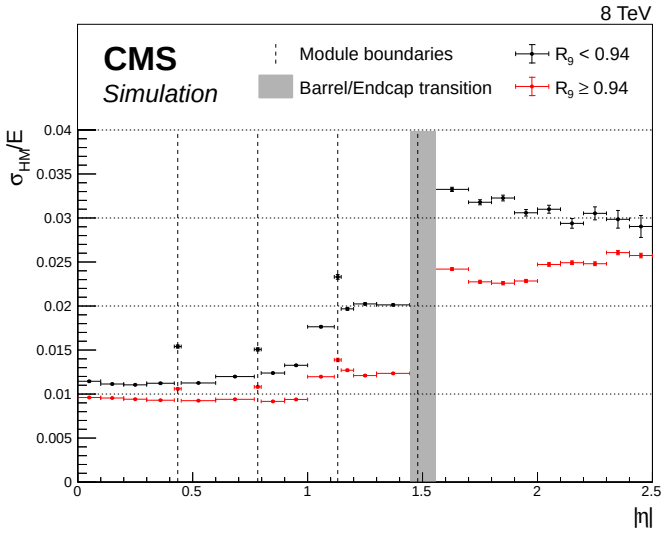


Figure 4.9: Relative energy resolution, σ_{eff}/E, as a function of $|\eta|$, in simulated H $\rightarrow \gamma\gamma$ events, for photons with $R_9 > 0.94$ (solid circles) and photons with $R_9 < 0.94$ (open squares). The vertical dashed lines mark the module boundaries in the barrel, and the vertical grey band indicates the range of $|\eta|$, around the barrel/endcap transition, removed from the fiducial region.

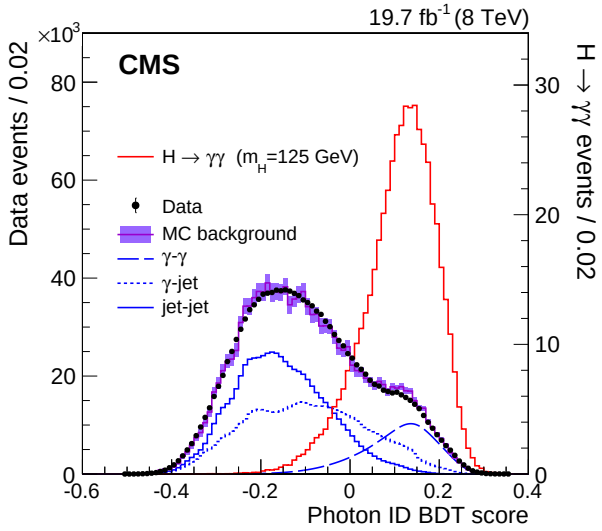


Figure 4.10: Photon identification BDT score of the lower-scoring photon of diphoton pairs with an invariant mass in the range $100 < m_{\gamma\gamma} < 180$ GeV, for events passing the preselection in the 8 TeV dataset (points), and for simulated background events (histogram with shaded error bands showing the statistical uncertainty). Histograms are also shown for different components of the simulated background, in which there are either two, one, or zero prompt signal-like photons. The tall histogram on the right (righthand vertical axis) corresponds to simulated Higgs boson signal events.

4.5.2 Diphoton vertex identification

The diphoton mass resolution has contributions from the resolution of the measurement of the photon energies and the measurement of the angle between the two photons. If the vertex from

which the photons originate is known to within about 10 mm, then the experimental resolution on the angle between them makes a negligible contribution to the mass resolution. The mean number of p-p interactions per bunch crossing is 9 in the 7 TeV dataset and 21 in the 8 TeV dataset. The longitudinal direction of the interaction vertices has a distribution with an RMS spread of about 6 (5) cm in the 7 (8) TeV dataset.

No charged particle tracks result from photons that do not convert, so the diphoton vertex is identified indirectly, using the kinematic properties of the diphoton system and its correlations with the kinematic properties of the recoiling proton debris against the photon pair. If either of the photons converts, the direction of the resulting tracks can provide additional information. Four discriminating variables are calculated for each reconstructed primary vertex:

- $\sum \vec{p}_T^{\,2}$
- $-\sum(\vec{p}_T \cdot \frac{\vec{p}_T^{\,\gamma\gamma}}{|\vec{p}_T^{\,\gamma\gamma}|})$
- $(|\sum \vec{p}_T| - |\vec{p}_T^{\,\gamma\gamma}|)/(|\sum(\vec{p}_T| + |\vec{p}_T^{\,\gamma\gamma}|)$
- If there is at least one converted photon: $g_{conv} = |z_e - z_{vtx}|/\sigma$

where the sums are over the transverse momentum vectors of the charged tracks, $\vec{p}_T$, $\vec{p}_T^{\,\gamma\gamma}$ is the transverse momentum vector of the diphoton system, z_e is the estimate of the primary vertex longitudinal position with associated uncertainty σ and z_{vtx} is the longitudinal position of the reconstructed vertex. The variables are used as the inputs to a multivariate system based on a BDT to choose the reconstructed vertex to be associated with the diphoton system.

The vertex finding efficiency, defined as the efficiency that the chosen vertex is within 10 mm of the true vertex location, has been measured using $Z \to \mu^+\mu^-$ events, where the same reconstruction method is applied to all the tracks of the same vertex, excluding the two muons. The ratio of the efficiency measured in data to that measured in MC simulation is within 1% of unity when viewed as a function of the number of vertices in the event and increases to a few percent in the region where $p_T^Z < 15$ GeV.

Figure 4.11 shows the efficiency with which a diphoton system is assigned to a vertex reconstructed within 10 mm of the true diphoton vertex in simulated Higgs boson events in the 8 TeV dataset, as a function of the transverse momentum of the diphoton system.

4.5.3 Diphoton BDT classifier

The analysis uses events with two photon candidates satisfying the preselection requirements, with an invariant mass in the range $100 < m_{\gamma\gamma} < 180$ GeV, $p_T^{\gamma\text{-leading}} > m_{\gamma\gamma}/3$ and $p_T^{\gamma\text{-subleading}} > m_{\gamma\gamma}/4$. The use of p_T thresholds scaled by $m_{\gamma\gamma}$ prevents the distortion of the low end of the $m_{\gamma\gamma}$ spectrum. In the case of multiple diphoton candidates, the one with the highest $p_T^{\gamma 1} + p_T^{\gamma 2}$ is selected. In addition, the photon identification BDT scores are required for both photons to be greater than -0.2. This requirement retains more than 99% of simulated signal events, while removing about 24% of events in data.

After these selections, a multivariate event classifier, the diphoton BDT, is constructed so that it assigns a high score to events that have good diphoton mass resolution and/or high probabil-

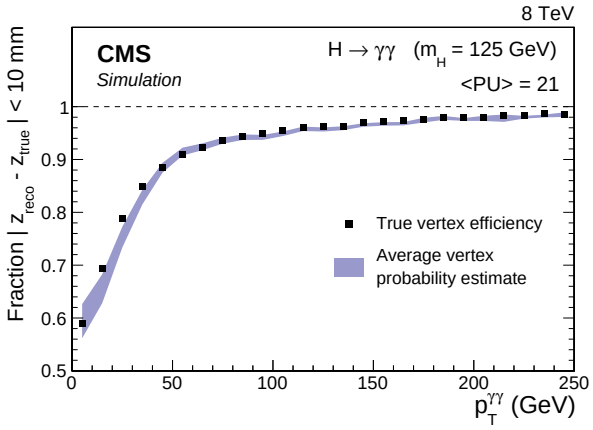


Figure 4.11: Fraction of diphoton vertices (solid points) assigned, by the vertex assignment BDT, to a reconstructed vertex within 10 mm of their true location in simulated Higgs boson events, $m_H = 125$ GeV, $\sqrt{s} = 8$ TeV, as a function of $p_T^{\gamma\gamma}$. Also shown is a band, the centre of which is the mean prediction, from the vertex probability BDT of correctly locating the vertex. The mean is calculated in $p_T^{\gamma\gamma}$ bins, and the width of the band represents the event-to-event uncertainty in the estimates.

ity of being signal rather than background. The Higgs boson signal-to-background ratio, S/B, varies with the kinematic properties of the diphoton system mainly through the η of the photons (highest S/B when both are in the barrel), and $p_T^{\gamma\gamma}$ (highest S/B for large $p_T^{\gamma\gamma}$). Moreover, the classifier should not select events according to the mass of the diphoton system relative to the particular mass of the Higgs boson signal used for training. The complete list of variables used in the BDT is:

- The scaled photon transverse momenta, $p_T^{\gamma 1}/m_{\gamma\gamma}$ and $p_T^{\gamma 2}/m_{\gamma\gamma}$, which are scaled to avoid any dependence on m_H.
- The pseudorapidities of both photons.
- The photon identification BDT classifier values for both photons.
- The cosine of the angle between the two photons in the transverse plane.
- The expected relative diphoton mass resolutions under the hypotheses of selecting the correct/a wrong interaction vertex.
- The probability of selecting the correct vertex.

The relative abundance of events from different production processes in the sample is set according to the expectations for a SM Higgs boson with that mass. In addition, it has been verified that selecting simulated background events with high diphoton BDT score does not result in any peak in the diphoton invariant mass.

4.5.4 Event classification

To achieve the best analysis performance, the events are separated into classes based on both their mass resolution and their relative probability to be due to signal rather than background.

First, the events tagged by the presence of additional particles (e.g. muons [68], electrons [58], b-jets [69] and missing transverse energy, E_T^{miss} [57]) in the final state in addition with the photon pair, are extracted. These additional particles give the event a signature characteristic of one of the production processes. The Higgs bosons produced by VBF are associated with a pair of jets separated by a large rapidity gap. Those resulting from the VH production mechanism may be accompanied by one or more charged leptons, large E_T^{miss}, or jets from the decay of the W or Z boson. Photons resulting from ttH production are, as a result of the decay of the top quarks, accompanied by b-quarks, and may be accompanied by charged leptons or additional jets.

All of these exclusive tags significantly increase the overall sensitivity of the analysis and precision on the measured signal strength, and increases the sensitivity to deviations of the Higgs boson couplings from their expected values. The remaining untagged events, which constitute the majority ($\approx 99\%$) of the events used in the analysis, are classified according to the variable constructed using multivariate techniques.

In total there are 14 event classes for the analysis of the 8 TeV dataset and 11 for the analysis of the 7 TeV dataset. To ensure that the classes are mutually exclusive, events are tested in a fixed order: first the production-signature tagged classes ranked by expected signal-to-background ratio, then the untagged classes. Table 4.1 lists the classes together with their key selection requirements.

4.5.5 Final results and mass measurement

Figure 4.12 (left) shows the $m_{\gamma\gamma}$ distribution of the combined data in the 7 and 8 TeV samples, together with the sum of the signal-plus-background fits to the 25 event classes which results in a best-fit mass $m_H = 124.7$ GeV. The local p-value, which quantifies the probability for the background to produce a fluctuation as large, or larger, than the apparent signal observed, is shown in Figure 4.12 (right) in the mass range $110 < m_H < 150$ GeV, calculated separately for the 7 and 8 TeV datasets as well as their combination.

The signal strength is quantified by $\mu = \sigma/\sigma_{SM}$, where σ/σ_{SM} denotes the production cross section times the relevant branching fractions, relative to the SM expectation. Table 4.2 shows the best-fit signal strength, $\hat{\mu}$, when m_H is treated as an unconstrained parameter, for the 7 TeV, 8 TeV, and combined datasets. The best-fit signal strength when the value of m_H is treated as an unconstrained parameter in the fit, is $\hat{\mu} = 1.14^{+0.26}_{-0.23}$, with the corresponding best-fit mass being $\hat{m} = 124.70 \pm 0.31(\text{stat}) \pm 0.15(\text{syst})$ GeV. Figure 4.13 shows the single categories (left) and the single production mechanisms (right) contribution to the final combined $\hat{\mu}$ value.

In particular, concerning the mass measurement systematic uncertainties, the three main sources that contribute to the uncertainty in the measured mass are shown in Table 4.3. The largest contributions are due to the possible imperfect simulation of differences in detector response to electrons and photons and the energy scale non-linearity in the extrapolation from the Z boson mass to the Higgs boson mass. A further contribution comes from the uncertainties in the energy scale measurement procedure and methodology, which uses measurement of the invariant mass in Z $\rightarrow e^+e^-$ events in which the electron showers are reconstructed as photons.

Other additional Higgs boson properties have been measured, such as the couplings and the

Table 4.1: Event classes for the 7 and 8 TeV datasets and some of their main selection requirements. Events are tested against the selection requirements of the classes in the order they are listed here.

Label	No. of classes 7 TeV	8 TeV	Main requirements
ttH lepton tag	$\star$	1	$p_T^{\gamma^1} > m_{\gamma\gamma}/2$ 1 b-tagged jet + 1 electron or muon
VH tight ℓ tag	1	1	$p_T^{\gamma^1} > 3m_{\gamma\gamma}/8$ [e or μ, $p_T > 20$ GeV, and $E_T^{miss} > 45$ GeV] or [$2e$ or 2μ, $p_T^{\ell} > 10$ GeV; $70 < m_{\ell\ell} < 110$ GeV]
VH loose ℓ tag	1	1	$p_T^{\gamma^1} > 3m_{\gamma\gamma}/8$ e or μ, $p_T > 20$GeV
VBF dijet tag 0-2	2	3	$p_T^{\gamma^1} > m_{\gamma\gamma}/2$ 2 jets; classified using combined diphoton-dijet BDT
VH E_T^{miss} tag	1	1	$p_T^{\gamma^1} > 3m_{\gamma\gamma}/8$ $E_T^{miss} > 70$ GeV
ttH multijet tag	$\star$	1	$E_T^{miss} > m_{\gamma\gamma}/2$ 1 b-tagged jet + 4 more jets
VH dijet tag	1	1	$E_T^{miss} > m_{\gamma\gamma}/2$ jet pair, $p_T^j > 40$ GeV and $60 < m_{jj} < 120$ GeV
Untagged 0-4	4	5	The remaining events, classified using diphoton BDT

$\star$ For the 7 TeV dataset, events in the ttH lepton tag and multijet tag classes are selected first, and combined to form a single event class.

spin. The results have been already discussed in Section 1.4 and compatibility with the SM Higgs boson have been found.

In general, the improvement with respect to discovery results [64] in the H $\rightarrow \gamma\gamma$ search and its properties measurement is due to a complete detector calibration with procedures described in Chapter 3.1, the tuning of the detector conditions of the MC simulation, which directly affect superclusters isolation variables, and an improved method for the background and signal fit. The implementation of these new features affect the final results with an improvement of about 10% each.

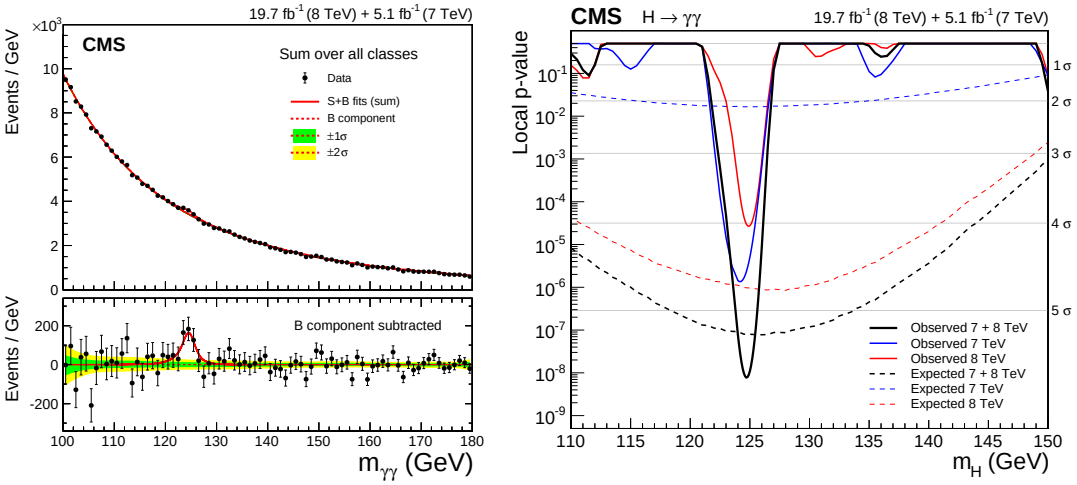


Figure 4.12: Left: Sum of the 25 signal-plus-background model fits to the event classes in both the 7 and 8 TeV datasets, together with the data binned as a function of $m_{\gamma\gamma}$. The 1σ and 2σ uncertainty bands shown for the background component of the fit are computed from the fit uncertainty in the background yield in bins corresponding to those used to display the data. These bands do not contain the Poisson uncertainty that must be included when the full uncertainty in the number of background events in any given mass range is estimated. The lower plot shows the residual data after subtracting the fitted background component. Right: Local p-values as a function of m H for the 7 TeV, 8 TeV, and the combined dataset. The values of the expected significance, calculated using the background expectation obtained from the signal-plus-background fit, are shown as dashed lines.

Table 4.2: Values of the best-fit signal strength, $\hat{\mu}$, when m_H is treated as a nuisance parameter, for the 7 TeV, 8 TeV, and combined datasets. The corresponding best-fit value of m_H, $\hat{\mu}$, is also given.

	$\hat{\mu}$	$\hat{m}_H$ (GeV)
7 TeV	$2.22^{+0.62}_{-0.55}$	124.2
8 TeV	$0.90^{+0.26}_{-0.23}$	124.9
Combined	$1.14^{+0.26}_{-0.23}$	124.7

Table 4.3: Magnitude of the uncertainty in the best fit mass induced by the systematic uncertainties in the signal model. These numbers have been obtained by quadratic subtraction of the statistical uncertainty. The statistical uncertainty includes all uncertainties in the background modelling.

Source of uncertainty	**Uncertainty in $\widehat{m}_{\mathbf{H}}$ (GeV)**
Imperfect simulation of electron-photon differences	0.10
Linearity of the energy scale	0.10
Energy scale calibration and resolution	0.05
Other	0.04
All systematic uncertainties in the signal model	0.15
Statistical	0.31
Total	**0.35**

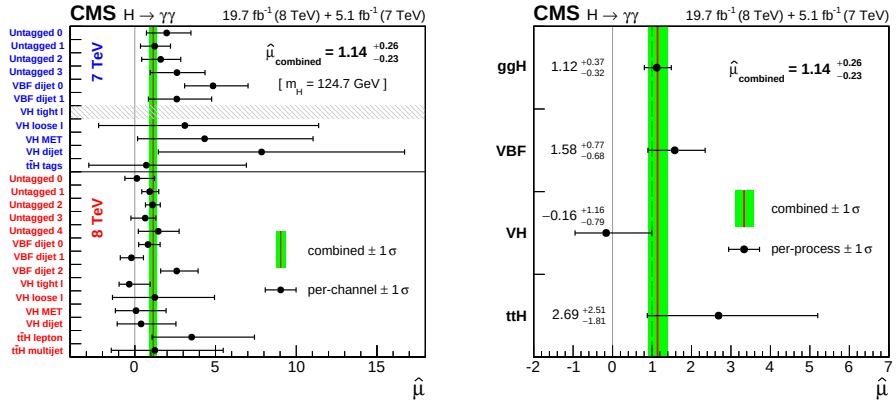


Figure 4.13: Left: Values of $\hat{\mu}$ measured individually for all event classes in the 7 and 8 TeV datasets, fixing $m_H = 124.7$ GeV. The horizontal bars indicate $\pm\sigma$ uncertainties in the values, and the vertical line and band indicate the best-fit signal strength in the combined fit to the data and its uncertainty. Right: Best-fit signal strength, $\hat{\mu}$, measured for each of the production processes in a combined fit where the signal strengths of all four processes have been allowed to vary independently in the fit. The signal mass, common to all four processes, is treated as an unconstrained parameter in the fit. The horizontal bars indicate $\pm 1\sigma$ uncertainties in the values for the individual processes. The band corresponds to $\pm 1\sigma$ uncertainties in the value obtained from the combined fit with a single signal strength.

CHAPTER 5

SEARCH FOR DI-HIGGS PRODUCTION DECAYING INTO DIPHOTON AND B-JETS PAIR

As pointed out in Section 1.5 the Higgs mechanism is theoretically unsatisfactory. It contains naturalness problems which manifest themselves as quadratic divergences on the quantum corrections of the Higgs mass. Hence, there is wide consensus that the SM of elementary particles is an effective theory, and that new physics models beyond it have to be probed. As described in Section 1.5.1, Warped Extra Dimension models can be a solution to the naturalness problem. The WED models predict at least two new massive particles, the radion (spin 0) and the graviton (spin 2). In the context of WED or in general new physics models beyond the SM, the enhanced double Higgs production is predicted at the LHC and it is a powerful tool to probe the existence of new particles with mass greater than twice the Hiss boson mass. In addition to the search of explicit resonant particles, the modifications of SM couplings of the di-Higgs production, due to higher order operators, can be probed with this signature. Therefore, also the non-resonant SM double Higgs production and possible deviation from the SM predictions can be tested. In this Chapter the search for a heavy resonance of spin 0 or 2 decaying into two 125 GeV Higgs bosons is described, using data collected during Run I of the LHC. In addition, the SM non-resonant double Higgs production and couplings are probed and deviation from the SM values are tested.

5.1 Data samples

This analysis is performed on the full dataset collected in 2012, reconstructed with the final detector calibration values in p-p collisions at the LHC, with an integrated luminosity of 19.7 fb^{-1} at 8 TeV. The 5.1 fb^{-1} data at 7 TeV collected in 2011 are not included in the analysis as they provide a marginal improvement to the sensitivity.

The events used in the analysis were selected by diphoton triggers with asymmetric transverse energy thresholds and complementary photon selections. One selection requires a loose calorimetric identification based on the shape of the electromagnetic shower and loose isolation requirements on the photon candidates, while the other requires only that the photon candidate has a high value of the R_9 shower shape variable. High trigger efficiency is maintained by allowing both photons to satisfy either selection. An efficient trigger at low p_T is one of the advantages of this

decay mode, compared to the other modes listed in Table 1.5.

5.2 Monte Carlo simulations

The Monte Carlo full simulations of the CMS detector at 8 TeV are used, where ECAL detector is simulated as described in Chapter 4, with detector conditions for 8 TeV events described in 4.2. Simulated events include simulation of the multiple p-p interactions taking place in each bunch crossing. Thus, in each recorded event they simulate the effects of the pileup, i.e. the presence of signals from multiple p-p interactions in multiple bunch crossings. All Monte Carlo samples are re-weighted to match the pileup distribution in the data by applying an event weight based on the number of simulated pileup events and the instantaneous luminosity per bunch-crossing averaged over the data-taking period considered in this analysis.

5.2.1 Resonant signal simulations

The bulk radion is used as a signal benchmark model. The events simulate a gluon fusion production of an on shell narrow-width radion, decaying into $H \rightarrow \gamma\gamma$ and $H \rightarrow b\bar{b}$ using MadGraph5 [70]. The interference between the Higgs and the radion is neglected. The events are hadronized under Pythia6 [71] and processed under the full CMS simulation. Also the graviton (spin 2) and the MSSM CP-even heavy neutral higgs samples are produced with the similar simulation procedures, but are marginally used. The analysis is performed testing signals with masses that span from 260 GeV to 1100 GeV. The upper bound for the search range is defined by the approximate transition between the two distinct jets regime and the one jet regime, in which the boosted decay products of the hadronic Higgs are reconstructed into a single fat jet. Table 5.1 shows the cross-sections for the simulated radion samples used in the analysis, which are used for the final interpretation of the results.

5.2.2 Non-resonant signal simulations

The $HH \rightarrow \gamma\gamma b\bar{b}$ signature enables also the non-resonant double Higgs production to be probed. In the SM two diagrams are mainly responsible for the non-resonant production of two Higgs bosons in the final state via gluon fusion, as can be seen on the two left diagrams of Figure 5.1. One process involves the Higgs self-coupling λ, linked to the Higgs boson mass in the SM: $m_H = 2\lambda/G_F$. This parameter is fundamental to probe if the Higgs potential at the origin of the electroweak symmetry breaking is as predicted by the SM. The other parameter is the Yukawa coupling with the top quark, y_t . In the SM, these two diagrams interfere destructively almost perfectly, which is the main reason for the low cross-section of the process, $\sigma \approx 10$ fb at 8 TeV and $\sigma \approx 40$ fb at 14 TeV. A comparison of the four-body invariant mass spectrum between the SM non-resonant production and the spin 0 resonant production can be seen on Figure 5.2.

By means of an effective parametrisation of the Lagrangian with dimension 6 operators, it is possible to vary the prediction of these two couplings λ and y_t. In this context, a new production diagram can be added, as represented in Figure 5.1 right, which opens the possibility for a

Table 5.1: The radion gluon fusion production cross section is estimated up to NLL QCD + NLO EW precision. Note: The cross section calculations holds up to 1 TeV. The radion branching ratio to HH is assumed to be 25% in all spectra, $BR(H \to \gamma\gamma) = 0.00228$, $BR(H \to b\bar{b}) = 0.577$.

	Radion $\Lambda_R = 1$ TeV & kl = 35 & no mixing	
mass (GeV)	σ(pb)	$\sigma \cdot BR(\text{HH})(2 \cdot BR(\gamma\gamma) \cdot BR(b\bar{b}))$(fb)
270	27.7	18.22
300	20.6	13.55
350	13	8.55
400	7.5	4.93
450	5.2	3.42
500	3.8	2.50
550	2.9	1.91
600	2.2	1.45
650	1.7	1.12
700	1.4	0.92
900	0.62	0.41
1000	0.45	0.30

quartic coupling ttHH denoted by the parameter c_2.

Given the low SM cross section for the HH production, as well as the fact that λ and c_2 are not constrained directly experimentally, and that there is quite some freedom in the current constraints on y_t, there is a strong motivation into looking for anomalous non-resonant di-Higgs production. In particular, even if the sensitivity to the standard model cross section for di-Higgs boson production is out of reach for the LHC Run II, a preliminary search for this particular signature provides an insight on how to optimize this specific analysis.

The ATLAS collaboration has published results for this search, where non-resonant production of Higgs di-bosons with cross section larger than 2.2 pb is excluded at 95% CL. The observed (expected) sensitivity of the analysis is $2.2(1.0^{+0.5}_{-0.2})$ pb [44].

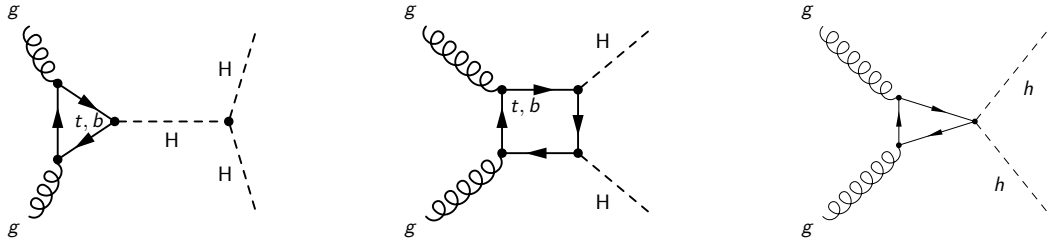


Figure 5.1: Feynman diagrams of the processes responsible for non-resonant gluon fusion production of two Higgs bosons in the final state. These diagrams involve the Higgs boson self-coupling (left), the top quark Yukawa coupling (middle), predicted in the case of the Standard Model, as well as the anomalous coupling ttHH (right), which is open in case of an effective parametrization of the Lagrangian by dimension 6 operators (beyond SM).

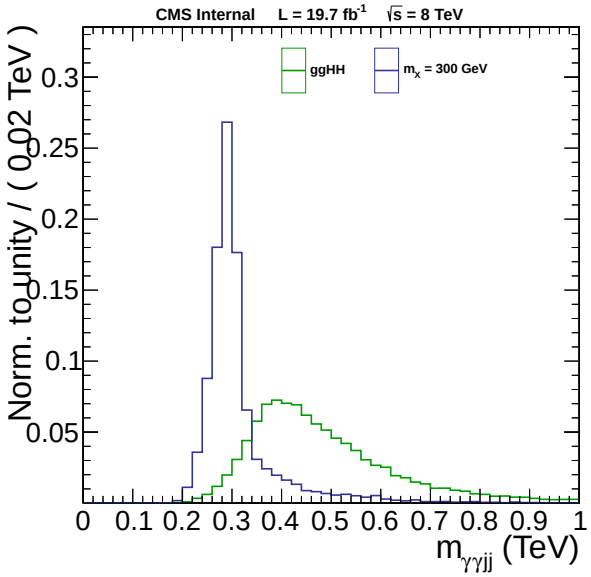


Figure 5.2: Shape comparison of a radion with $m_X = 300$ GeV and SM di-Higgs non-resonant production.

Regarding the Monte Carlo generation, three coupling parameters are considered in this study to scan beyond the standard model: $\kappa_\lambda \equiv \lambda/\lambda^{SM}$ (variation with respect to the SM value of λ parameter), $\kappa_t \equiv y_t/y_t^{SM}$ (variation with respect to the SM value of y_t parameter) and c_2. Anomalous couplings are introduced when these parameters are different from the SM values, $\kappa_\lambda = 1$, $\kappa_t = 1$ and $c_2 = 0$. Signal samples are generated at leading order, using MadGraph5 models showered with Pythia6. In total, 124 signals have been generated varying $\kappa_\lambda = -20$, -15, -10, (-2), 0,1, (2), (3), (5), 10, 15, 20, $\kappa_t = 0.75$, 1.00, 1.25 and $c_2 = -3$, -2, 0, 2, 3, where $\kappa_\lambda = -2$, 2, 3, 5, 10 scenarios are simulated only with $\kappa_t = 1$ and $c_2 = 0$. In this work, even when anomalous couplings are involved, the 125 GeV narrow width final state SM branching ratios are assumed valid at first approximation.

5.2.3 Background topologies and simulations

The main contribution to the diphoton background is the prompt-prompt photon contribution. This contribution is given by two photons coming from QED interaction which are reconstructed as coming from the hard interaction, mimicking the signal and making this background irreducible. Figure 5.3 shows the main diagrams contributing to the prompt-prompt background.

The other important contributions are given by the jets misidentified as fake photons. Typically these photon candidates come from one or more neutral mesons that take a substantial fraction of the total jet p_T and are thus relatively isolated from hadronic activity in the detector. In the transverse momentum range of interest, the photons from neutral pion decays are rather collimated and are reconstructed as a single photon. These contributions can be reduced with a suitable photon identification, so that the misidentification efficiency becomes as low as possible.

At 8 TeV the diphoton continuum processes are simulated using SHERPA [72]. The SHERPA samples give a noticeably improved description of diphoton continuum events accompanied by

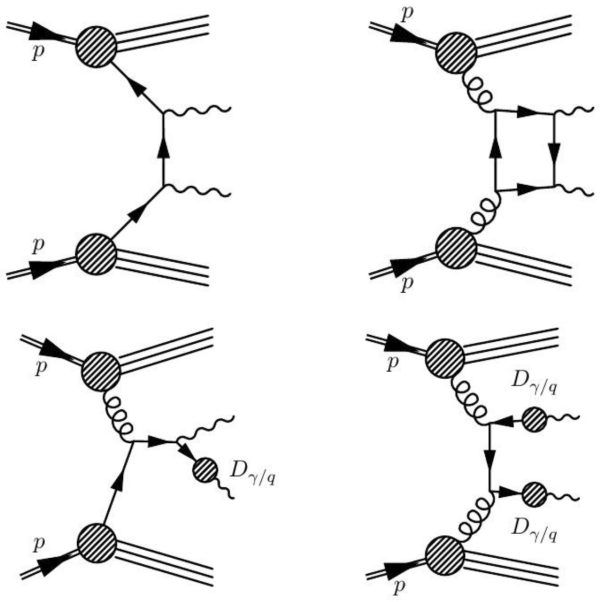


Figure 5.3: Main contributing diagrams to the prompt-prompt background: born (top-left), box (top-right) and fragmentations (bottom)

one or two jets. This is the dominant background and represents typically 3/4 of the total.

The remaining processes, prompt+fake and fake-fake, where one or two photon candidates arises from misidentified jet fragments are simulated using PYTHIA alone, with events enriched in jets with electromagnetic decays. This non-resonant $\gamma\gamma$ background represents typically 1/4 of the total. The contribution with two jets misidentified as photons appears to be mostly negligible.

The other non-resonant backgrounds contribute less than 1% each. They are composed mainly from associated photon production with electroweak boson and top quarks that behave as a source of b-quarks or with neutral Drell-Yan events where the electrons are misidentified as photons. The resonant backgrounds corresponding to the different processes of SM Higgs boson signals (ggH, VBF, VH and ttH) and double Higgs bosons production are also considered.

The background samples are not used for extracted directly the final results, in fact the fit of the background is made to the data. However, these samples are used for the selections optimization, as described in the following sections.

5.2.4 bbH associated production

With similar diagrams to those involved in the ttH associated production (Figure 1.3d), it is possible to produce a Higgs boson in association with two b-quarks. Given the prompt nature of the two b-quarks for this process, this production mode is an irreducible background for the analyses presented in this chapter when considering the diphoton decay mode of the Higgs boson. The samples used here have been produced at LO with Madgraph5 and showered with Pythia6.

Due to the lower mass of the b-quark with respect to the mass of the top-quark, the kine-

matics of the bbH dijet pair is much softer than the ttH one, as it can be seen in Figure 5.4. This leads to a lower selection efficiency compared to ttH, and despite the higher cross-section, the final bbH selection yield is smaller than the ttH one. For these reasons, a low sensitivity for this exclusive channel is expected, thus a specific analysis aiming the bbH $\rightarrow \gamma\gamma$ search is not performed using the 8 TeV data.

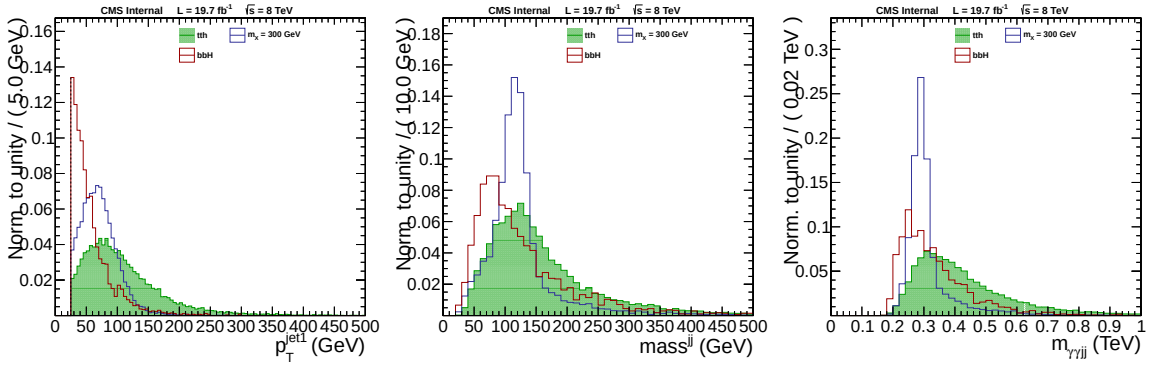


Figure 5.4: Shape comparison of ttH, bbH and radion (m_X = 300 GeV) samples at reconstructed level, for the leading jet (left), the dijet invariant mass (middle) and the four-body invariant mass (right).

5.3 Photon reconstruction and identification

Photon candidates are reconstructed using the procedure described in Section 4.5. As in H $\rightarrow \gamma\gamma$ analysis, in order to have similar conditions in MC simulation and data, resolution smearings are applied to the MC simulation and a scale shift is applied to the data [63].

5.3.1 Photon preselections

The photons entering the analysis are required to satisfy preselections similar, but slightly more stringent, than the trigger requirements:

- $p_T^{\gamma\text{-leading}} > 33$ GeV and $p_T^{\gamma\text{-subleading}} > 25$ GeV
- a selection on the hadronic energy leakage of the shower, H/E
- a loose selection based on isolation and the shape of the shower
- an electron veto, which removes the photon candidate if its supercluster is matched to an electron track with no missing hits in the innermost tracker layers, thus excluding almost all Z $\rightarrow e^+e^-$ events.

The selection requirements are applied with different stringency in four categories defined as: EB and $R_9 > 0.9$, EB and $R_9 < 0.9$, EE and $R_9 > 0.9$ and EE. and $R_9 < 0.9$. The efficiency of the photon preselection is measured in data using a "tag-and-probe" technique [73], by means of Z $\rightarrow e^+e^-$ events. The ratio of the photon efficiency measured in data to that found in simulated Z $\rightarrow e^+e^-$ events, $\epsilon_{data}/\epsilon_{MC}$, is consistent with unity in all categories. It is used to correct the

simulated signal sample and the associated uncertainties are taken into account as systematic uncertainties in the signal extraction procedure.

5.3.2 Photon identification

Differently from the SM H $\rightarrow \gamma\gamma$ analysis, a cut-based photon identification is used in this analysis. This choice is driven by boosted topology of the Higgs decaying into diphoton. In case of a low mass resonance (300 GeV) decaying to HH, the kinematics of the H $\rightarrow \gamma\gamma$ system is very similar to the one from the SM Higgs, as shown in Figure 5.5. On the other hand, for higher radion masses the photons are more boosted, consequently easier of identify. This property could be used to improve the photon identification efficiency in the high mass region, where anyway the background is much smaller.

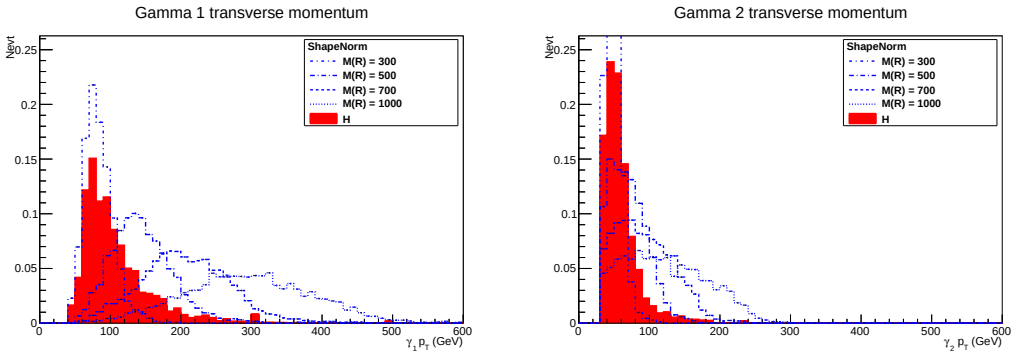


Figure 5.5: Transverse momentum of the two selected photons for the radion signal at different masses compared with the SM Higgs production.

The cut-based photon identification was studied and optimized as an alternative to the photon MVA identification of the SM H $\rightarrow \gamma\gamma$ analysis, resulting in a less performing analysis [63]. The selections have been optimized separately in four categories defined in terms of η and R_9. These categories have significantly differing levels of background and mass resolution. The use of categories rather than a unique sample provides an increased sensitivity.

The selections were set to get the highest efficiency of accepted photons for a chosen purity (S/B). This means that the selections are tighter in the low R_9 category than in the high R_9 category and tighter in the endcap than in the barrel. The selections have been optimized using subleading photons from H $\rightarrow \gamma\gamma$ events as signal, and subleading photons from γ+jet events as background.

Isolation and topology variables are exploited and optimized, and the final selections are listed in Table 5.2, which are the "CiC Super Tight" working point. The chosen photon identification selections are applied to both legs of the diphoton pair. As for the pre-selections, Section 5.3.1, efficiencies have been computed, and the ratio between the efficiencies in data and in simulation has been used to correct the signal MC sample (Table 5.3). Simulation efficiencies as a function of photon η and p_T are shown in Figure 5.6.

Even if in the current analysis the photon identification is directly taken from H $\rightarrow \gamma\gamma$ analysis,

Table 5.2: Photon ID selection values. The selections are applied to both the leading and subleading photons.

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

Table 5.3: Photon identification efficiencies measured in the 4 photon categories using tag and probe with $Z \to e^+e^-$ events (for all selections except electron rejection).

	DATA			MC		R	
	Eff.	Stat. Err.	Syst. Err.	Eff.	Stat. Err.	Eff.	Err.
Barrel; R_9 >0.94	0.9086	0.0007	0.0025	0.9064	0.0003	1.002	0.003
Barrel; 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
Endcap; R_9 <0.94	0.4554	0.0015	0.0115	0.4435	0.0006	1.027	0.026

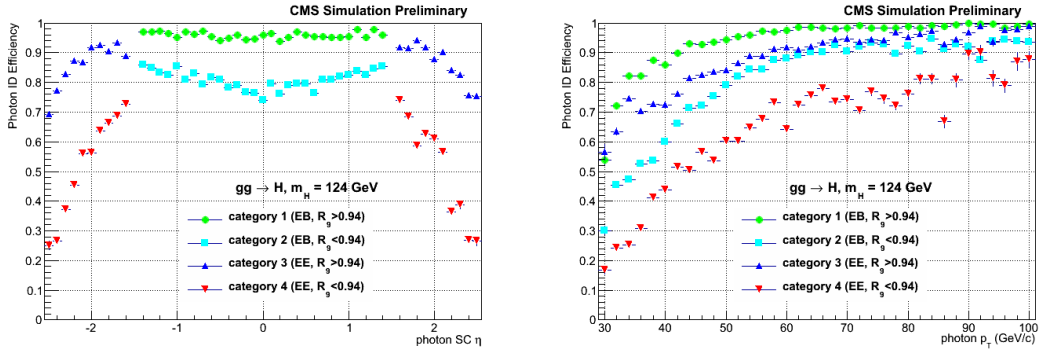


Figure 5.6: Photon Identification efficiency for signal photons (simulated Higgs bosons with $m_H = 124$ GeV, as a function of pseudorapidity (a) and p_T (b), for each of the four photon categories.

additional studies show that up to 17% signal acceptance can be recovered by loosening the isolation criteria on the subleading photon. While this relaxing increases the amount of background evaluated from data by 45%, this turns also into an increase of the final sensitivity of about 13%, using Monte Carlo simulation samples as background.

In addition, also the photon MVA identification has been tested. It is verified that the use of the standard $H \to \gamma\gamma$ working points, > 0.02 for EB and > 0.1, Section 4.5.1, does not improve the resonant and the SM non-resonant analyses sensitivity and that a new working point with a looser selection on the subleading photon should be used to reach an improvement.

In addition photons are required to have:

- $p_T^{\gamma 1} > m_{\gamma\gamma}/3$ and $p_T^{\gamma 2} > m_{\gamma\gamma}/4$
- $|\eta| < 2.5$
- $100 < m_{\gamma\gamma} < 180$ GeV

5.3.3 Vertex identification

Usually the direction of a photon cannot be directly reconstructed with the CMS detector. To reconstruct the photon four-momentum, and the diphoton invariant mass, a proper identification of the interaction vertex is necessary (except for photons with early conversions in the tracker). The solution exploited in the $H \to \gamma\gamma$ analysis for the identification of the primary vertex, described in detail in Section 4.5.2, is to rely on the kinematic properties of the relics of the proton-proton collision and their correlation with the Higgs boson production. In fact, in $H \to \gamma\gamma$ decays the photon pair is the only direct product of the hard scattering gluon-gluon Higgs process, thus a BDT technique is used to enhance the vertex identification efficiency.

In the $HH \to \gamma\gamma b\bar{b}$ analysis, the two photons come from the same vertex as the two b-jets. The presence of charged particles associated to the hard interaction process is an advantage that can be used for the identification of the primary vertex. The most natural solution is to select the vertex with the largest value of $\sum p_T^2$ of the event, where the sum runs over the tracks associated to the vertex, since that vertex is the most probable associated to the production of the two b-jets.

The performance of the two methods in the choice of the primary vertex, H$\gamma\gamma$-BDT and $\sum p_T^2$, are compared. For this study the signal events are identified by the presence of at least two jets with $p_T > 25$ GeV and one reconstructed photon pair, selected according to Section 5.3.2, matching the generator level photon pair within a distance $\Delta R < 0.05$.

The vertex identification efficiency is defined as the probability of the $HH \to \gamma\gamma b\bar{b}$ generator level vertex to be within a given distance from the reconstructed vertex $|z_{gen} - z_{ID}| < \Delta z$. It is known from the $H \to \gamma\gamma$ analysis that for $\Delta z < 1$ cm the resolution on the diphoton invariant mass is not limited by the angle resolution. For the diphoton mass reconstruction it is therefore sufficient to localize the vertex within about 1 cm.

For a 300 GeV mass hypothesis the efficiency is above 98% from a distance $\Delta z \leq 1$ cm for both methods (see Figure 5.7), and it increases as a function of Δz. The bigger the Δz tolerance, the bigger is the random coincidence of a wrongly identified vertex within this tolerance from the generated vertex.

The right part of Figure 5.7 shows that the improvement that we have in choosing the $\sum p_T^2$ vertex instead of the H$\gamma\gamma$-BDT one is more modest for higher masses of the resonance. This can be explained in this way: at a resonance mass of 300 GeV the two Higgs (125 GeV) bosons are produced with little transverse momentum with respect to the beam, so the dijet system of $H \to b\bar{b}$ has a small angular correlation with the diphoton system. At higher resonance masses the correlation is restored and the vertex identification from the BDT algorithm has the same efficiency as the discriminator based on $\sum p_T^2$ alone.

For resonance masses higher than 500 GeV the efficiency reaches more than 99%, and matches the target. This is explained with the fact that one of the inputs of the BDT is based on the

recoil properties of the interaction debris against the photon pair. Therefore, $\sum p_T^2$ is chosen as a vertex-identification method for all the mass hypotheses.

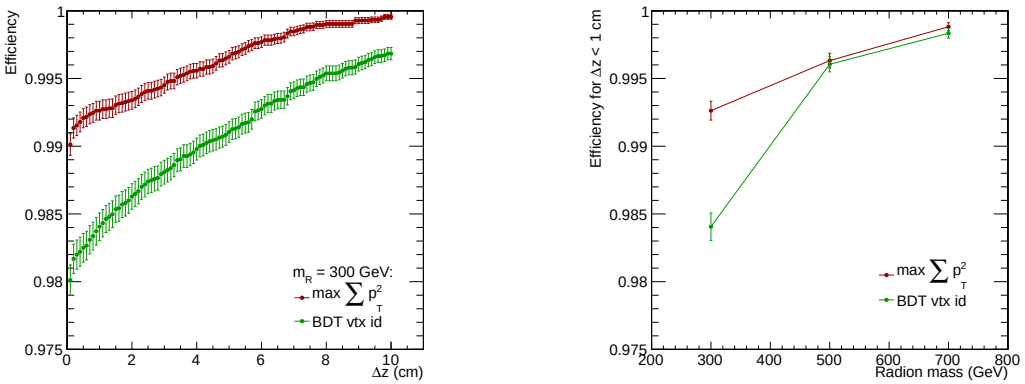


Figure 5.7: Vertex identification efficiency as function of $|z_{gen} - z_{ID}| < \Delta z$ for the $\sum p_T^2$ (red dots) and the BDT (green dots) algorithms for a resonance signal of mass 300 GeV. The errors are statistical only, and depend on the size of the simulated sample. Right: Efficiency for $\Delta z < 1$ cm as function of the resonance mass, for the $\sum p_T^2$ (red dots) and the BDT (green dots) algorithms. The lines are to guide the eye.

5.4 Jets reconstruction and identification

A high-energy quark or gluon emitted in a hard proton-proton collision does not appear in the detector: as it reaches large distances from the rest of the partons, the strong force potential favours the radiation of softer and collinear gluons and quarks, until a point where a non-perturbative transition causes the partons to combine into colourless hadrons. The result is a spray of collimated particles, referred to as a jet, which, due to energy conservation, reflects the energy and the flight direction of the initial parton. Jets are detected as a cluster of tracks and energy deposits in a defined region of the detector. Due to the intrinsic compositeness of such objects, a jet cannot be defined until an algorithmic procedure of clustering is chosen.

The jets are reconstructed using the Particle Flow algorithm [67], not considering the charged-hadron subtraction and the anti-k_T jets with R = 0.5 are used. In addition, at pre-selection level the jet candidates must be separated from the two selected photons by at least $\Delta R = 0.3$. They are required also to have $p_T > 20$ GeV, $|\eta| < 2.5$ and $100 < m_{jj} < 180$ GeV, where the dijet pair choice is described in Section 5.4.3. If the regression corrections (Section 5.4.5) are applied, these selections are performed after the regression.

In CMS, the adopted jet clustering algorithm, is the so-called anti-k_T [74]. This algorithm proceeds via the definition of two distances for each particle i in the list of particles, namely

$$\begin{aligned} d_{ij} &= \min\left(\frac{1}{p_{\mathrm{T}i}^2}, \frac{1}{p_{\mathrm{T}j}^2}\right)\frac{\Delta R_{ij}^2}{R^2} \\ d_{iB} &= \frac{1}{p_{\mathrm{T}i}^2} \end{aligned} \tag{5.1}$$

In the above equation, d_{ij} can be interpreted as the distance between the particle i and an other particle j among those still to be clustered, while d_{iB} represents the distance between the particle i and the beam line. ΔR_{ij} is the distance between the two particles in the $\eta \times \phi$ plane, while R is the algorithm radius parameter. For each particle i, the algorithm looks if there is another particle j such that d_{ij} is smaller than d_{iB}. If this happens, then particles i and j are recombined by adding together their four-momenta, otherwise the i particle is promoted to jet. The whole procedure is iterated and the algorithm stops when only jets are left.

It can be easily seen that particles at a distance greater than R from the jet axis are not clustered together, thus leading to the construction of cone-shaped jets. The standard radius parameter R adopted in CMS in the $\eta \times \phi$ plane during the Run I is 0.5. The jet momentum is determined as the vectorial sum of all particle momenta in it.

The Particle Flow (PF) (also known as Global Event Description, GED) is a whole-event reconstruction technique whose purpose is to reconstruct and identify each single particle produced in proton-proton collision with an optimized combination of all sub-detectors information. In this process, the identification of the particle type (photon, electron, muon, charged hadron, neutral hadron) plays a crucial role in the determination of the particle direction and energy.

While no substantial changes are expected for the reconstruction of high-energy electrons and muons (e.g. from W boson decay), the PF allows to significantly improve the resolution of jets and E_T^{miss} [57] with respect to the standard, pure calorimetric reconstruction. Since on average only about the 15% of a jet energy is carried by neutral, long-lived hadrons (neutrons, Λ baryons, etc.), and for the remaining 85% carried by charged particles, the coarse HCAL information is combined with the more precise tracker momentum measurements, thus allowing for a largely better jet reconstruction.

The missing transverse energy, E_T^{miss}, is the only physics quantity definable at hadron colliders used as a signature of invisible particles like neutrinos. E_T^{miss} is defined as the negative vector sum of the transverse momenta of all final-state particles in the event.

In the hypothesis that all detectable particles are properly reconstructed, E_T^{miss} coincides with the sum of the four-momenta of all undetectable particles (i.e. neutrinos, or BSM particles such as neutralinos in more exotic scenarios), since the initial p-p collision occurs between two particles of negligible transverse momentum ($\lesssim$ 1GeV). In practice, this is not possible: since a fraction of the total event energy is unavoidably lost in the beam pipe or only coarsely reconstructed in the forward calorimeters, additional contributors besides undetectable particles affect E_T^{miss}.

5.4.1 Pileup jet identification

During jets reconstruction, the particles produced in the pileup interactions can be sometimes collected by the jet clustering algorithm into objects with large enough p_T. The resulting pileup jets are removed using selection criteria based on the jets width and on the compatibility of the jet tracks with the primary vertex which is selected for the event. Two variables are discriminant enough to be used for the cut-based pileup rejection: the fraction of charged particle-flow candidates and the jet width.

The fraction of charged particle-flow candidates, attached to the primary vertex is quantified by the β^* variable. It is defined as the sum of the p_T of all the PF objects attached to a given vertex, balanced by the total jet p_T, see Equation 5.2. A selection based on β^* is efficient only within the tracker acceptance, i.e. $|\eta_{jet}| < 2.5$. On the other hand, the jet width is quantified by the jet RMS. This variable is defined by the median of the distance of each two constituents (ΔR), balanced by p_T, see Equation 5.3.

The selections were optimized for the (VBF) H $\rightarrow \gamma\gamma$ analysis, Section 4.5. Different selections are applied in different regions of the detector and are summarized in Table 5.4.

$$\beta^* = \frac{\sum_{vertex} p_T^{tk}}{\sum_{jet} p_T^{tk}} \tag{5.2}$$

$$RMS = \frac{\sum_{\text{constituents}} p_T^2 \cdot \Delta R}{\sum_{\text{constituents}} p_T^2} \tag{5.3}$$

Jet-ID	β^*	**RMS**
$\eta < 2.5$	$< 0.2 \log(\text{Nvtx} - 0.64)$	< 0.06
$2.5 < \eta < 2.75$	$< 0.3 \log(\text{Nvtx} - 0.64)$	< 0.05

Table 5.4: pileup jet identification working point.

A study of the efficacy of this identification on this analysis has been performed, which has proven that this selection applied to the jets used for this analysis gives no significant improvement to the final sensitivity. In fact, Figure 5.8 shows that the signal jets are correctly selected even if the pileup jet ID is removed. In fact, the distance ΔR between the b-quark direction at the MC generator level and the selected jet after the full reconstruction is essentially the same with or without these selections. A signal jet is defined as "matched" if ΔR(jet,b-quark)$<$ 0.4. The impact on yield of the signal, the background estimated from the MC simulation, and the data are presented in Table 5.5. The same behaviour is observed for all the considered mass hypotheses (270, 300, 350, 400 GeV) and for leading and subleading jets.

Thefore, this study has proven that, within statistical errors, the efficiencies of picking pileup jets are the same applying or not these selections on the simulated signal hypotheses. Without these selections, the backgrounds increase by about 40%. The impact of this selection removal on the analysis was quantified examinating the sensitivity using the diphoton MC sample as background.

The removal of the pileup jet ID selection allowed to increase the signal acceptance by about 18%, without impacting the sensitivity of the analysis. The statistics increase of about 40% for background and 18% for signal allowed also for a finer analysis optimization described in the following sections.

5.4.2 Identification of b-jets

The hard fragmentation, long lifetimes (i.e. secondary production vertex), high masses of B hadrons, and the relatively high fraction of semileptonic decays distinguish jets arising from bottom-quark hadronization (b jets) from those originating from gluons, light quarks and - to a

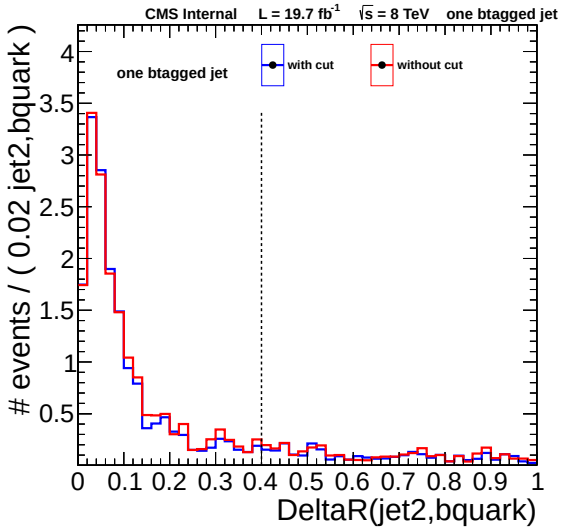


Figure 5.8: Distance in ΔR between the generated b-quark and the reconstructed jet. This plot is for radion with $m_X = 270$ GeV for the subleading jet of the event, without asking any b-tag requirement. Similar behaviour is observed for the leading jets, and the mass hypotheses 300, 350, 400 GeV.

Table 5.5: Number of events after baseline selections for the different samples with or without jet PU ID selection. 2-btag and 1-btag categories refer to the categorization on the number of b-jets described in the next section, Section 5.4.2.

	# events 2btag		# events 1 btag	
Sample	with jet PU ID	without jet PU ID	with jet PU ID	without jet PU ID
Data	143.0	170.0 (+ 18.9 %)	1476.0	1847.0 (+ 25.1 %)
diphojet sherpa	79.4	104.7 (+ 31.9 %)	1422.1	1831.1 (+ 28.8 %)
Radion m270	1684.5	1825.3 (+ 8.4 %)	2945.9	3218.7 (+ 9.3 %)
Radion m300	3625.5	3956.2 (+ 9.1 %)	6299.2	6821.2 (+ 8.3 %)
Radion m350	2149.0	2318.9 (+ 7.9 %)	3532.8	3811.1 (+ 7.9 %)
Radion m400	2491.6	2702.6 (+ 8.5 %)	4015.6	4230.6 (+ 5.4 %)

lesser extent - from c quarks.

The presence of a secondary vertex and the kinematic variables associated with this vertex can be used to discriminate between b and non-b jets. Two of these useful variables for discrimination are the flight distance and direction, using the vector between primary and secondary vertices. The other variables are related to various properties of the system of associated secondary tracks such as the multiplicity, the mass (assuming the pion mass for all secondary tracks), or the energy [69].

The *Simple Secondary Vertex* (SSV) algorithms use the significance of the flight distance (the ratio of the flight distance to its estimated uncertainty) as the discriminating variable. The algorithm's efficiency is limited by the secondary vertex reconstruction efficiency to about 65%.

A more complex approach involves the use of secondary vertices, together with track-based lifetime information. By using these additional variables, the *Combined Secondary Vertex* (CSV) algorithm provides discrimination also in cases when no secondary vertices are found, increasing the efficiency with respect to the SSV algorithms. In many cases, tracks can be combined in a

"pseudo vertex", allowing for the computation of a subset of secondary-vertex-based quantities even without an actual vertex fit.

Two likelihood ratios are built from the combination of track-based and secondary vertex variables. They are used to discriminate between b and c jets and between b and light-parton jets. The distributions of the vertex multiplicity and of the CSV discriminator for multijet samples are presented in Figure 5.9 using 8 TeV data.

Given the topology of the $\gamma\gamma b\bar{b}$ search, two jets produced by b-quarks are expected. The most natural approach consists in requiring that at least two jets in the event are b-tagged. Nevertheless, according to the choice of the algorithm and working point, one of the jets could fail the b-tagging selection, as the average b-tagging efficiency of the signal hypotheses is at maximum $\approx 80\%$. Hence, the events are divided into two categories: events with only 1 b-tagged jet ("1-btag", medium purity category) and events with at least 2 b-tagged jets ("2-btag", high purity category). As for the $ttH \rightarrow \gamma\gamma$ channel, the medium working point of the *Combined Secondary Vertex* algorithm, CSV > 0.679, is chosen for the b-jet identification.

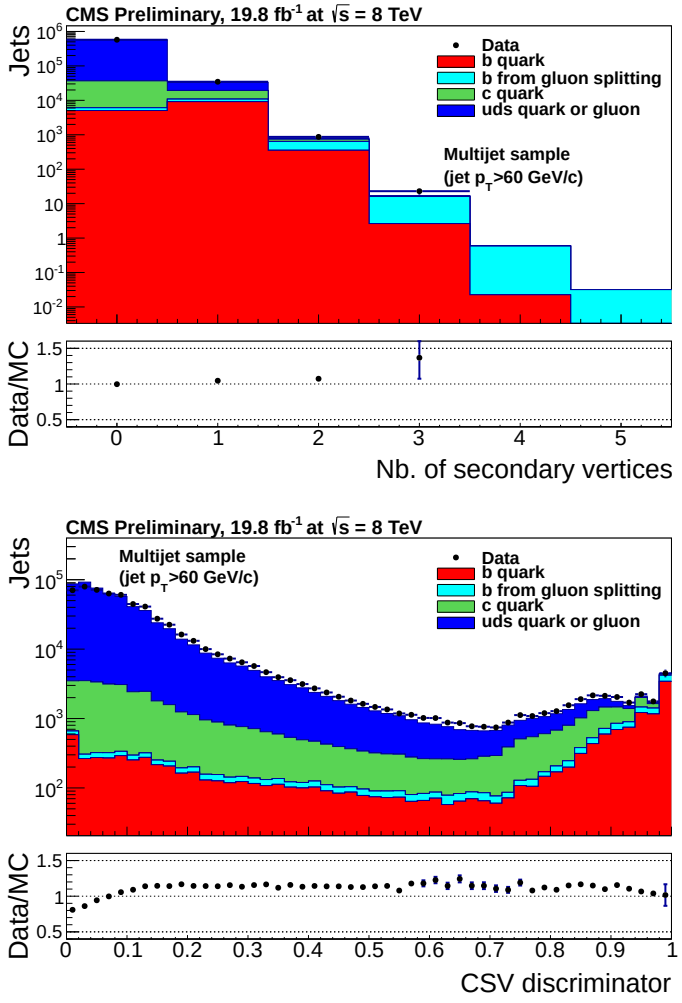


Figure 5.9: Distributions of (top) the secondary vertex multiplicity and (bottom) the CSV discriminator for 8 TeV data of multijet samples [69].

5.4.3 Jets pair choice

After the jets selections described above, there are three possible type of events: events with only one b-tagged jet and one or more non-b-tagged jets, events with two or more b-tagged jets and events with no b-tagged jet. The events without any b-tagged jet are excluded from the analysis. Once the events are divided into the two b-tag categories ("1-btag" and "2-btag"), in the case of two or more combinatoric jet pairs, the choice of the jet pair to used for the signal extraction can be driven by many criteria. The following criteria are tested:

1. The individually highest p_T jets.
2. The combination of jets with highest $p_T(jj)$
3. The combination of jets with highest $p_T(jj)/m_{jj}$

In all these choices, in case the b-tagging information is used, a priority to the b-tagged jets is given, basing on the category.

The possible choices are compared in signal events, measuring the probability of matching the pair of jets at the MC generator level with the pair coming from the tested choice. Table 5.6 shows the result of this study, where results are shown also for the case where the b-tagging information of the jets is not used, i.e. no priority to the b-tagged jets is given. The usage of the b-tagging information improves significantly the probability that the dijet choice is correct and makes all the proposed methods equivalent.

Moreover, the effect of the different choices on the background shape is verified. It was proven that the choice of jets with the highest $p_T(jj)$ is the one which modify the least the shape of the background, as shown in Figure 5.10. Therefore, the choice of jets with highest $p_T(jj)$ is the one that is used for this analysis.

Strategy	not using b-tagging	using b-tagging
1	0.67%	0.87%
2	0.64%	0.87%
3	0.66%	0.87%

Table 5.6: Percentage of signal events (signal mass of 300 GeV) in which the reconstructed pair of jets closest to the MC truth is selected for different dijet combinatoric choices

5.4.4 b-tagging efficiency scale factors

The b-tagging efficiency for jets is different between data and MC simulations. To account for this, scaling factors (SF) to correct the MC simulation efficiencies are provided. These factors are computed from the samples, such as QCD events and events where top-quarks are produced [69]. These are computed upon measurements of the b-tagging efficiencies (ϵ_{DATA} and ϵ_{MC}) as:

$$SF = \frac{\epsilon_{\text{DATA}}}{\epsilon_{\text{MC}}}$$

The scale factors corrections are applied as event weights only to signal events, because data are

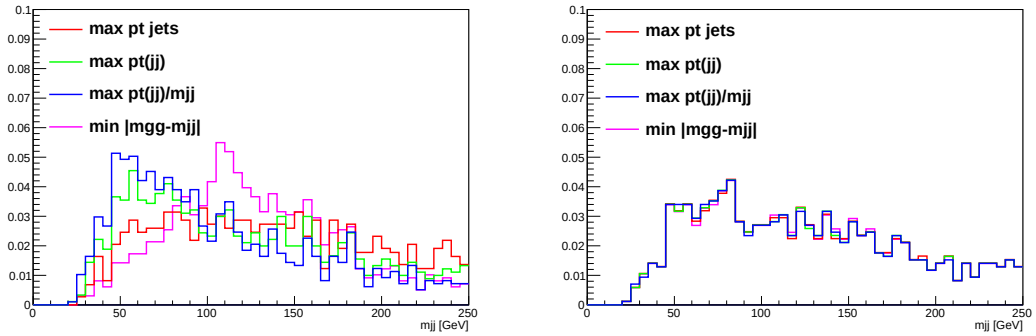


Figure 5.10: Dijet invariant mass for the 2-btag category for different jet choices, not using (left) or using (right) the b-tagging information of the jets. The events come from the diphoton background after the above photons and jets selections. The distributions are normalized to the unity.

used as background. These weights are applied per category, and do not produce any migration-effect between categories. In the case of events with the same b-tagging working point (e.g. medium CSV), the weights are defined as follows:

$$weight = \frac{\prod_{i=tagged} SF_i \epsilon_i \prod_{j=not\ tagged}(1 - SF_j \epsilon_j)}{\prod_{i=tagged} \epsilon_i \prod_{j=not\ tagged}(1 - \epsilon_j)}$$

where ϵ_i is the b-tagging efficiency, measured on jets that pass the analysis selection, and both SF_i and ϵ_i depend on η, p_T and the flavour of the jet.

The b-tagging SFs values for 8 TeV data are shown in Figure 5.11, while the b-tagging efficiencies are shown in Figure 5.12 for jets coming from the 300 GeV signal sample.

The systematic uncertainty for each weight is computed through a complete error propagation of the SF and ϵ uncertainties. The scale factors for light quarks are uncorrelated with SF_c and SF_b, while SF_c is conservatively taken as fully correlated to SF_b. Because of this, an anti-correlation in the final weights uncertainty has been found between the two categories of the analysis, "1-btag" and "2-btag". The systematic uncertainties associated to the procedure are evaluated on the $m_{\gamma\gamma jj}$ invariant mass distribution, by moving $\pm 1\sigma$ the weight for each event.

5.4.5 Jet transverse momentum regression

As the jet population used in this analysis is composed by b-jets, even if not necessarily tagged, it is possible to improve the correction to the jet energy by exploiting the properties of these jets. This in turn improves the resolution of the dijet mass, with a few percent shift on the value of the mass peak.

One of the main motivations for doing our specific correction is to recover the fraction of energy of the b-jets coming from the missing energy. Indeed, because of semileptonic b-hadron decays, jets from b quarks contain, on average, more leptons and a larger fraction of missing energy than jets from light quarks or gluons. This kinematic difference between the b-jets and the other

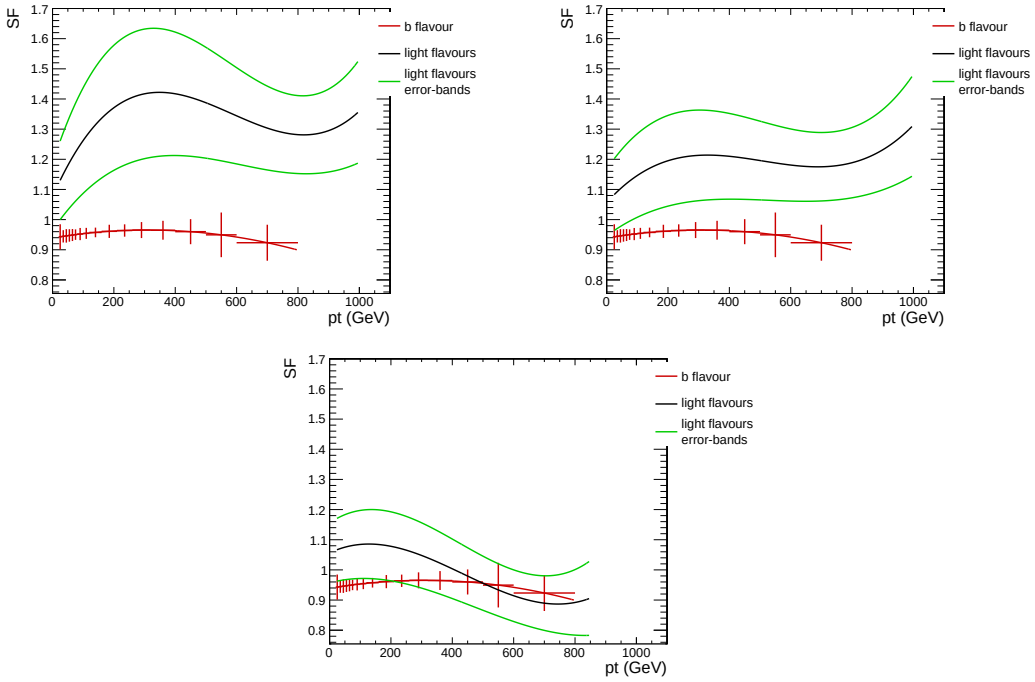


Figure 5.11: B-tagging efficiencies SFs [69], [75], for different η-bins. From the left to the right: $0.0 < \eta < 0.8$, $0.8 < \eta < 1.6$ and $1.6 < \eta < 2.4$. In red SFs for the b or c flavoured jets are shown, while in black for the light flavoured quarks. SFs for c-flavoured jets are the same as the b-flavoured ones but have doubled uncertainties.

flavours of jets is important especially for low mass signals, as for high mass hypotheses the kinematic is sufficiently different.

A jet transverse momentum regression with TMVA Boosted Decision Tree algorithm using the gradient boost is applied. It acts as a multidimensional calibration tuned to the specific jet properties of signal. High mass signal hypotheses ($m_{\gamma\gamma b\bar{b}} \gtrsim 400$ GeV) are excluded because of the strongly different jet kinematics. Moreover, there are very few events that drive the analysis sensitivity of high mass regime, therefore the expected gain from this technique is very low.

This kind of regression has been inspired by the two SM $H \to b\bar{b}$ analyses in the VBF [76] and VH [77] production modes. In fact, their implementation is mirrored up to the training sample. This means that the inputs, target, and structure of the BDT are the same, although di-Higgs samples train the BDT instead of SM Higgs samples.

The main reason why a retraining is performed is the difference in kinematic of the jets coming from the di-Higgs signals and SM Higgs boson signals. In fact, in this analysis the Higgs boson has a wider range of boost than the SM $H \to b\bar{b}$ decay. Therefore, suboptimal performance is expected using a regression trained for SM Higgs boson production, such as a ZH sample with $Z \to \ell\ell$ and $H \to b\bar{b}$.

One unique regression is trained for all the low-mass signal hypotheses. As the distributions of the p_T of the jet coming from the generator level are very different, as shown in Figure 5.13,

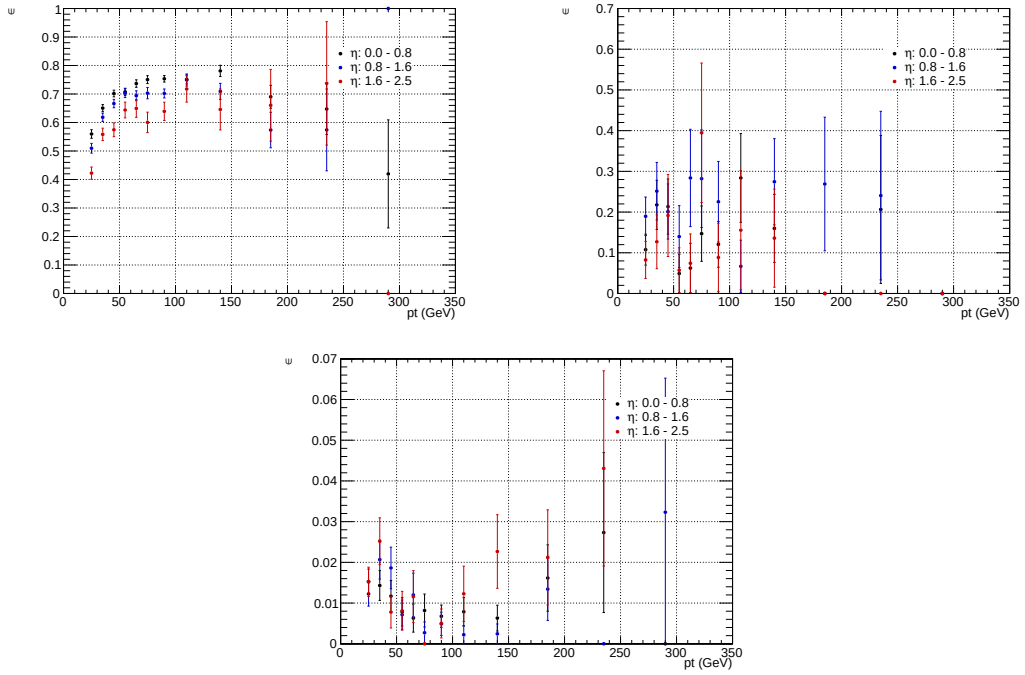


Figure 5.12: B-tagging efficiencies as a function of the jet p_T for different η-bins, [0.0,0.8] (black), [0.8,1.6] (blue) and [1.6,2.5] (red), and flavours (from the left to the right, for b, c and light flavours). The plots refer to jets from the 300 GeV signal sample that pass the jet selections described in previous sections.

the ratio of the p_T of the jet coming from the generator level (gen-p_T) to the reconstructed jet p_T is chosen as a target for the regression. The signal jets passing these preselections: $pt_T > 20$ GeV, $|\eta| < 2.5$ and $\Delta R(genjet, jet) < 0.4$, are used to train the BDT.

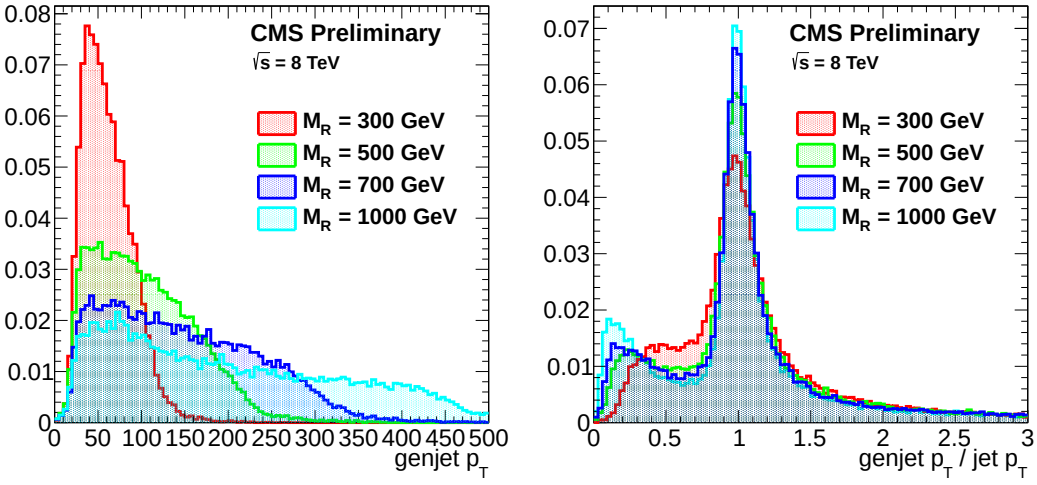


Figure 5.13: Comparisons for the distributions of generated jet p_T (left) and ratio of generated jet p_T to jet p_T (right) for the four signal samples. The ratio is chosen as the target for the regression over the generated jet p_T.

The input variables to the regression are:

- Jet transverse momentum, p_T
- Jet transverse mass, $M_T \equiv \sqrt{M^2 + p_x^2 + p_y^2}$
- Jet pseudorapidity, η
- Jet PF photon energy fraction
- Neutral hadron energy fraction
- Number of neutral and charged PF jet constituents
- Lead track p_T associated to the jet
- Jet secondary vertex flight distance (if there is a secondary vertex)
- Jet secondary vertex mass (if there is a seconardy vertex)
- Soft lepton p_T (if there is a soft lepton in the jet)
- Soft lepton relative p_T in direction of the jet (if there is a soft lepton in the jet)
- PF MET with Type 1 correction [57] + dedicated H $\to \gamma\gamma$ azimuthal corrections [63]
- $\Delta\phi(jet, \text{MET})$
- Median jet energy per jet area, ρ

Possible overtraining has been studied and considered negligible. The figure of merit to quantify the improvement given by the regression is the resolution as measured in the selected dijet mass spectrum of the signal sample, where the choice of the dijet pair is described in detail in the previous Section 5.4.3.

The resolution improvement in the m_{jj} and $m_{\gamma\gamma jj}$ spectra is quoted in two ways. One is through a fit to each spectrum using the sum of a Crystal Ball (Gaussian with an exponential tail) and third-order polynomial for each of the events categories ("1-btag" and "2-btag"). The parameters of the Crystal Ball give estimates of the spread and central value of the distribution, and their ratio gives the resolution. A second way is through the effective σ used in the H $\to \gamma\gamma$ search. The σ_{eff} is defined as the smallest interval containing 68% of the distribution, and the central value in this case is taken to be the mean value and the resolution is the ratio of these two values.

The measured resolutions are summarized in Table 5.7 and Figure 5.14 shows the m_{jj} and $m_{\gamma\gamma jj}$ fits before and after the regression. The improvement for m_{jj} is between 13% and 23% in the "2-btag" category, the one that drives the sensitivity in the analysis. For $m_{\gamma\gamma jj}$ the improvement is between 13% and 20%. A big motivation for using this technique is also to improve the separation between signal and background. It has proven that the regression shifts the background slight to higher values of the invariant mass, giving an increase in background yield of about 10% but not changing its shape.

A separate regression was trained for the SM non-resonant signal, but it brought no improvement to the analysis when propagated to the statistical analysis for signal extraction. Moreover, since the many BSM non-resonant signals have different jet kinematics, it is difficult to have a unique regression with good performance for all the signals. Thus, the regression is not applied to the non-resonant search.

Table 5.7: Improvement from regression on $m_{\gamma\gamma jj}$ and m_{jj} spectra. All numbers are in units of percentage unless noted.

	$m_{\gamma\gamma jj}$, spectrum				m_{jj}, spectrum			
	cat0		cat1		cat0		cat1	
M_X (GeV)	Imp. CB	Imp. σ_{eff}	Imp. CB	Imp. σ_{eff}	Imp. CB	Imp. σ_{eff}	Imp. CB	Imp. σ_{eff}
270	19.72	15.16	3.10	5.64	15.08	14.72	12.24	-2.07
300	16.64	13.02	8.70	8.69	16.05	12.99	14.19	21.11
350	19.76	17.53	13.62	9.63	23.07	14.18	18.95	4.25
400	19.82	20.20	21.23	13.37	17.03	18.43	14.74	4.60

5.4.6 Kinematic Fit

Once the dijet selections are applied, it is possible to improve the resolution in the four-body spectrum exploiting our knowledge that the dijet mass is expected to peak around the Higgs boson mass (125 GeV) in the case of signal. In the case of the background, this assumption introduces a constraint which is not correlated to the underlying physical process, and therefore has the effect of shuffling randomly the events in the final four-body invariant mass spectrum.

The constraint of the jets to be compatible with the Higgs boson decay may be done in different ways. In order to optimally scale the dijet quadrimomentum to the Higgs boson mass, a kinematic fit is used to the two jets which exploits the parametrizations of the expected jet transverse momentum and position resolutions (energy resolution of about 15 GeV) as a functions of the jet p_T and η. It constrains the mass of the dijet system to the value of the Higgs mass by modifying for each jet the values of E_T, η and ϕ in accordance to their expected resolutions, which are estimated from the Monte Carlo simulation. Such kinematic constraint technique is widely used in CMS for measurements where the decay proceeds through quarks, for example the semi-leptonic H $\to VV$ searches at high masses [78].

In Figure 5.15 the impact of the kinematic fit on the resonant samples is shown, where the improvement in resolution is quite large. Such an improvement is most significant for lower resonance masses because the relative resolution improves with the jet p_T, and the higher the p_T the smaller the resolution gain by the kinematic fit. Concerning the effect on the background, the kinematical constrain is not exploiting a physical process and the four-body mass is not peaking even when the fit is applied.

The final selections on photons and jets, which are required before the signal and background extraction, are summarized in Table 5.8.

5.5 Control sample

In order to choose the best background fit function, Section 5.7, a control sample built from data is used. The events are selected where one of the two photons is requested to pass the same

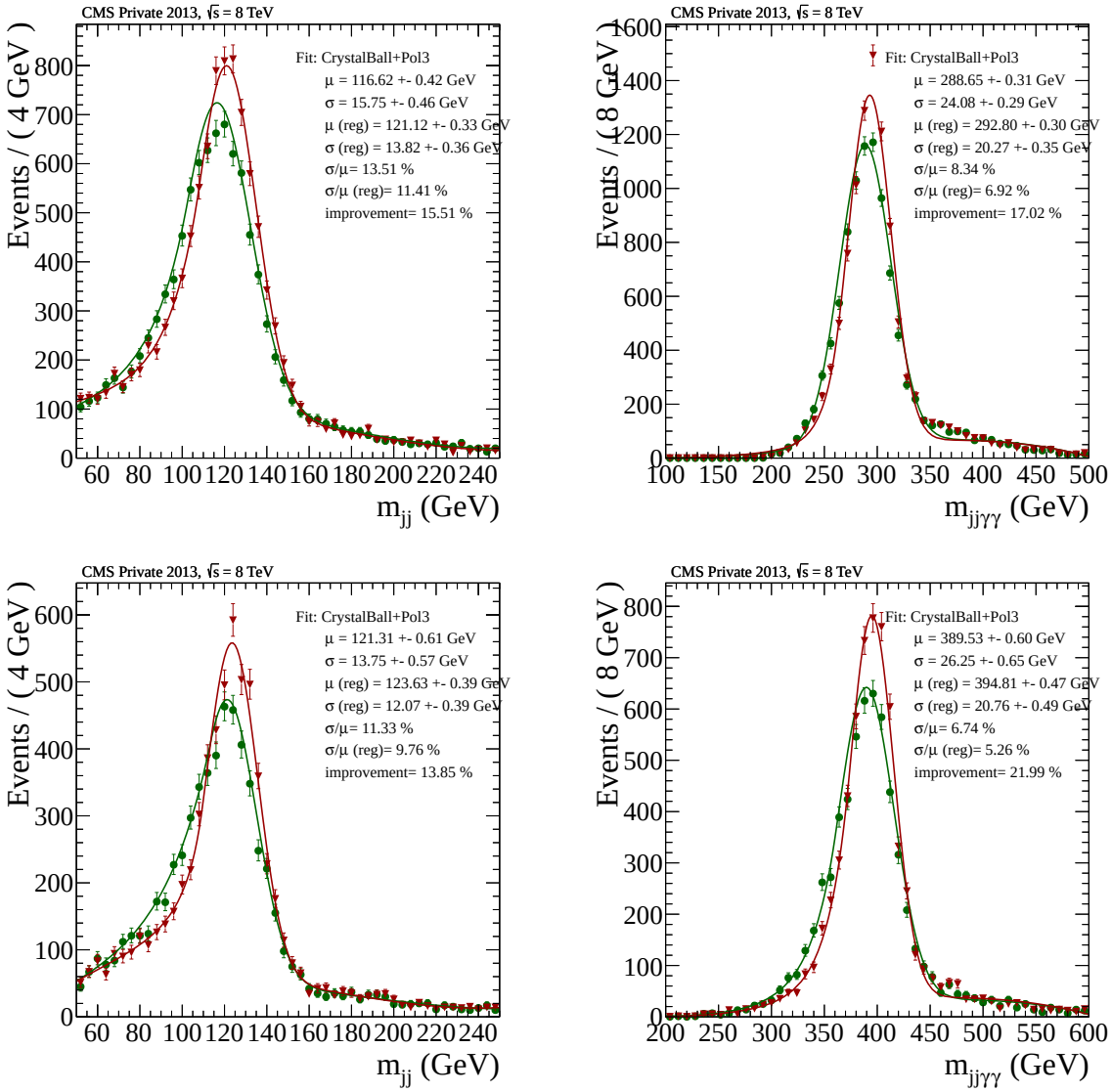


Figure 5.14: Resolution improvements for the m_{jj} (left) and $m_{\gamma\gamma jj}$ spectra at mass points $m_X = 300, 400$ GeV mass points for both categories together. The green distribution is before applying the regression, and the red one is after applying the regression. The fit model is the sum of a Crystal Ball and third-order polynomial.

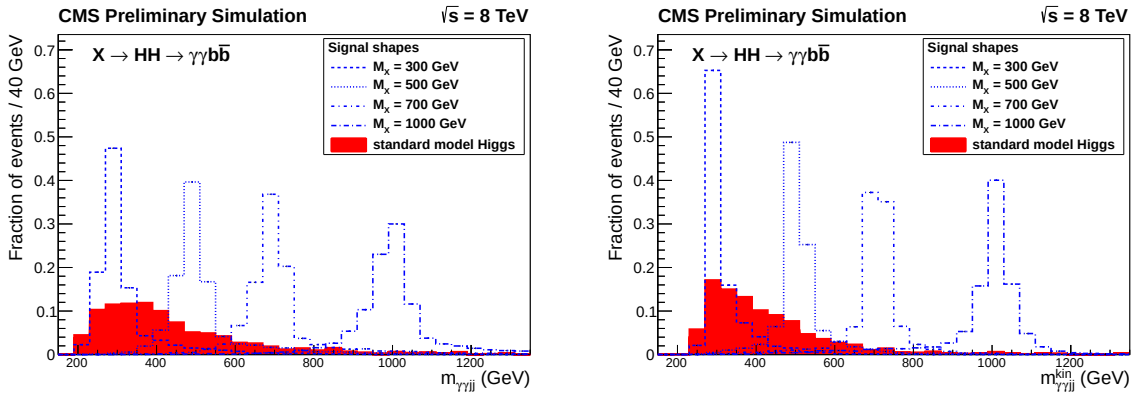


Figure 5.15: The effect of the kinematic fit on the $m_{\gamma\gamma jj}$ spectra. $m_{\gamma\gamma jj}$ spectrum for resonant and H $\to \gamma\gamma$ signals before (left) and after (right) the kinematic fit.

Table 5.8: Summary of the analysis preselections. In case the search is performed using the jet energy regression, the selections are applied directly on those jets.

Photons	Jets	Events classification
tight photon identification	loose jet identification	
$p_{T\gamma_1}/m_{\gamma\gamma} > 1/3$		
$p_{T\gamma_2}/m_{\gamma\gamma} > 1/4$	$p_T^j(p_T^{j,reg}) > 25$GeV	
$\lvert\eta_\gamma\rvert < 2.5$	$\lvert\eta_j\rvert < 2.5$	
$100 < m_{\gamma\gamma} < 180$GeV	$60 < m_{jj}(m_{jj}^{reg}) < 180$GeV	
		High purity: ≥ 2 b–tagged jets
	at least 1 b–tagged jet	Medium purity: $= 1$ b–tagged jet

tight cut-based identification and isolation selection used in the analysis, and the other one to be anti-selected, i.e. isolation and identification looser than the CiC Super Tight working point. Therefore, this sample is dominated by the contribution of 1γ+3jets instead of 2γ+2jets, as the signal region.

Due to this, some of the main kinematic variables, in particular those related to the photons, have a different shape with respect to the signal region. Hence, in order to build a viable background model a reweighting of the control sample events is needed to make the kinematic distributions in the control region match those in the signal region. Different reweighing strategies have been tried to match appropriately the data spectra, exploiting both the photons and the jets p_T and η distributions. The one that was found to be more performing consists in building a weight-map in 2 dimensions in the plane identified by the transverse momenta of the photons, $p_T^{\gamma 1} \times p_T^{\gamma 2}$. To better match the number of events in data, the normalization of the control sample is fixed to the that in the data. The $m_{\gamma\gamma}$ spectrum for data (dots) and control sample (filled histograms) is shown in Figure 5.16 after the reweighting.

5.6 Strategy for the signal extraction

The notable property of the resonant signal is the presence of three invariant mass spectra with three distinctive peaks: dijet and diphoton invariant mass peaks at the Higgs boson mass (125

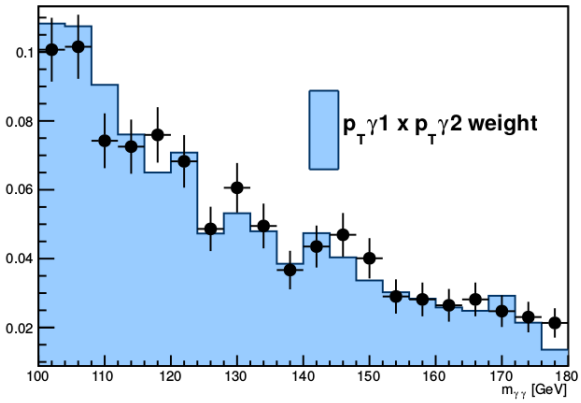


Figure 5.16: Diphoton invariant mass normalized to unity in data (dots) and in the data control sample (filled histograms) after the $p_T^{\gamma 1} \times p_T^{\gamma 2}$-reweighting.

GeV), and the resonance peak in the four-body invariant mass distribution, Figures 5.17 and 5.18.

5.6.1 Resonant analysis

In the invariant mass spectra, Figures 5.17 and 5.18, after the photons and jets selection described above, there are two different kinematic regions:

- Low mass region, $250 < m_X < 400$ GeV: the four-body spectrum exhibits a kinematic peak around 200-300 GeV, related to the kinematic constraints applied to the jet and photon transverse momenta. A resonant signal in this region would appear as a small peak on top of this kinematic peak. Due to this, it is hard to fit a resonant shape in the four-body invariant mass without a detailed prediction of the background shape. For these reasons, a simultaneous fit (2D-fit) to the $m_{\gamma\gamma} \times m_{jj}$ spectra is performed.
- High mass region, $400 < m_X < 1100$ GeV: the four-body spectrum above 400 GeV is smoothly falling, while the signal has a resonant shape, giving a large discrimination power to the signal. In addition, the observation of a peak in the four-body invariant mass would be a direct indication of a presence of a resonant particle. Therefore, the signal is extracted from a fit of signal and background components to the four-body invariant mass spectrum of the events surviving the selections.

At values of resonant mass $m_X \gtrsim 1000$ GeV, the two b-quarks produced by the Higgs boson start to merge into a single jet for the anti-$k_T$algorithm with size parameter R = 0.5. This is visible in the distribution of the leading jet mass shown in Figure 5.19: at high resonance masses the leading jet mass can be the same as the Higgs boson mass. The study of this region therefore requires considering the possibility of a leading Higgs boson jet with a substructure rather than two separated b-jets.

As selections on the resonant mass spectra are the best handle of the kinematic properties of the double Higgs bosons production, the final fits are performed after selections on these variables. In particular, a selection on the $m_{\gamma\gamma jj}^{KinFit,reg}$ spectrum is applied before performing the 2D fit to the $m_{\gamma\gamma} \times m_{jj}^{reg}$ of the low mass resonant search, where the regression is applied to the dijet and the four-body mass spectra . On the other hand, for the high mass resonant search selections on

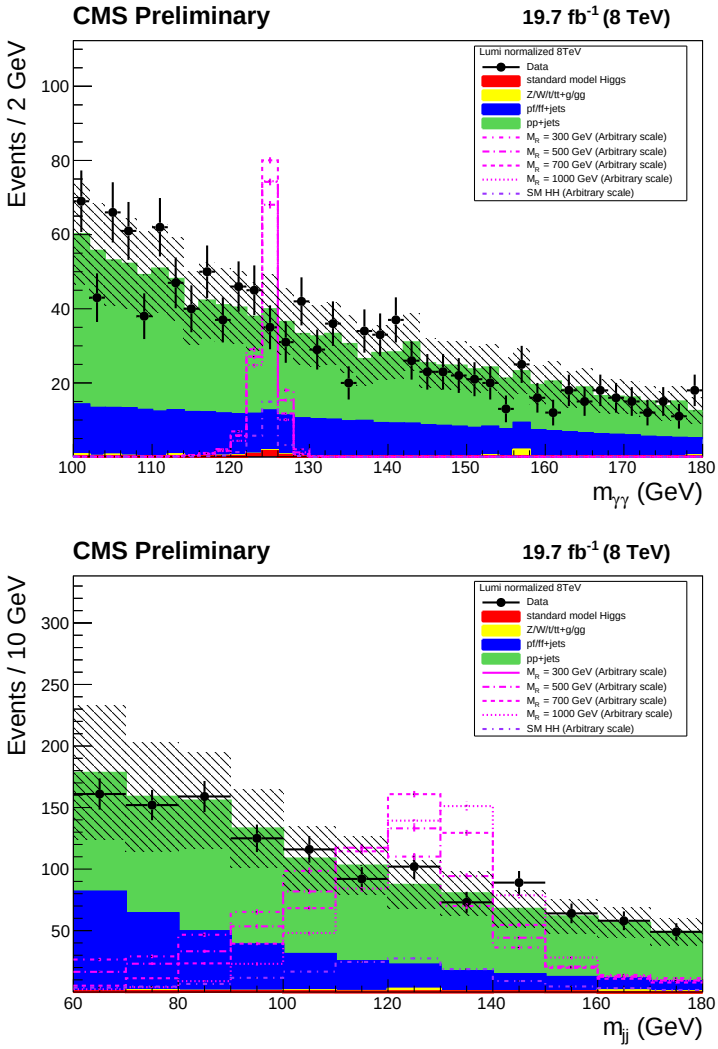


Figure 5.17: Diphoton and dijet distributions of data, MC background simulations and MC signal simulations, after basic selections on photons and jets of Table 5.8 and requesting at least one b-tagged jet. Signals resonant (magenta) and non-resonant (violet) samples are normalized to arbitrary scale, while the MC backgrounds are normalized to the data. A systematic uncertainty of 20% is applied to the pp+jets sample, due to the missing order in the generation, and a total statistical uncertainty of 60% is applied to the pf/ff MC samples.

the $m_{\gamma\gamma}$ and m_{jj} mass spectra are applied before fitting the $m_{\gamma\gamma jj}^{KinFit}$ spectrum.

The optimization of the mass windows has been performed running the procedure for signal extraction, using the Monte Carlo simulation samples as a background, where the mass window which gives the best sensitivity is chosen. Figure 5.20 shows an example of how the window optimization of the m_{jj} spectrum of the "2-btag" category for the 500 GeV mass hypothesis is performed. Even though the mass windows were optimized for each category and each mass point, little difference in the sensitivity was seen if the windows were enlarged of about 20%. Hence, one unique window for each mass hypothesis and category, containing the largest window extremes, was chosen, shown in Table 5.10.

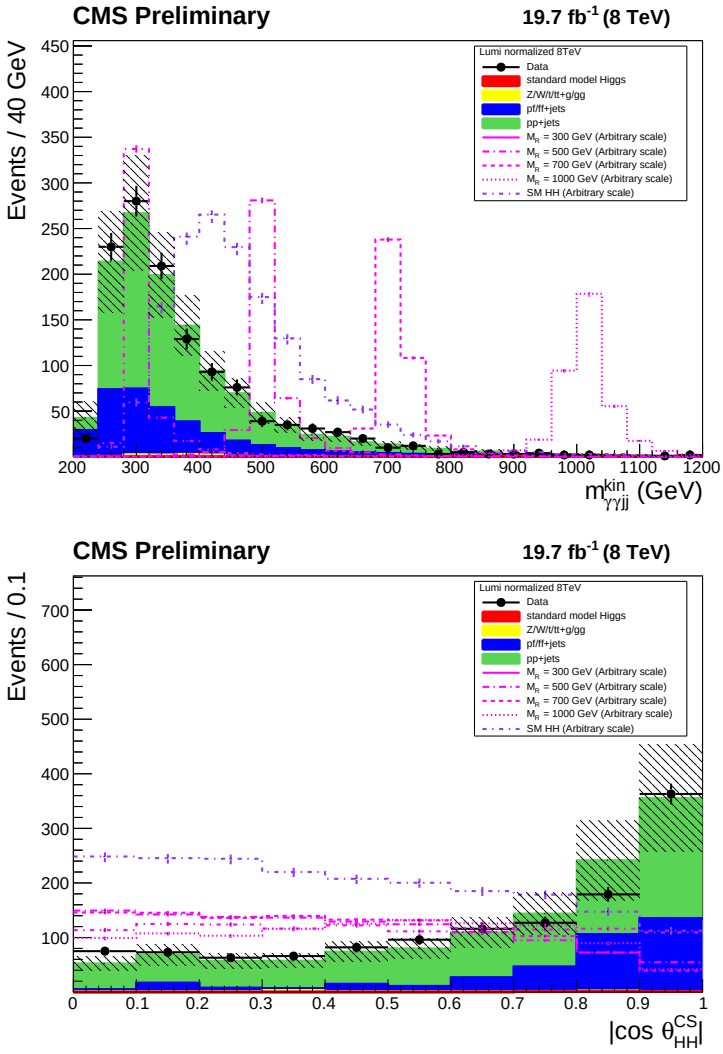


Figure 5.18: Four-body and $|\cos\theta_H^{CS}H|$ distributions of data, MC background simulations and MC signal simulations, after basic selections on photons and jets of Table 5.8 and requesting at least one b-tagged jet. Signals resonant (magenta) and non-resonant (violet) samples are normalized to arbitrary scale, while the MC backgrounds are normalized to the data. A systematic uncertainty of 20% is applied to the pp+jets sample, due to the missing order in the generation, and a total statistical uncertainty of 60% is applied to the pf/ff MC samples.

Regarding the low-mass analysis, similar procedure was used for the four-body mass windows, whose final results are shown in Table 5.9. Even if the optimization was performed separately for the two categories, a common selection was chosen since almost no improvement comes from the separation.

Table 5.9: Additional selections applied for low mass search hypotheses.

m_X (GeV)	260	270	300	350
$m_{\gamma\gamma jj}^{KinFit,reg}$ selection (GeV)	$[250, 270]$	$[260, 280]$	$[290, 310]$	$[330, 375]$

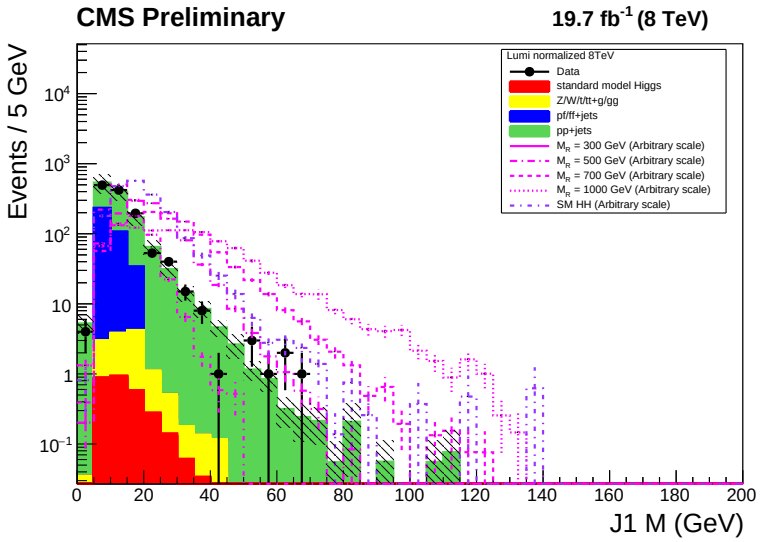


Figure 5.19: Invariant mass distribution of the leading distribution of data, MC background simulations and MC signal simulations, after basic selections on photons and jets of Table 5.8 and requesting at least one b-tagged jet. Signals resonant (magenta) and non-resonant (violet) samples are normalized to arbitrary scale, while the MC backgrounds are normalized to the data. A systematic uncertainty of 20% is applied to the pp+jets sample, due to the missing order in the generation, and a total statistical uncertainty of 60% is applied to the pf/ff MC samples.

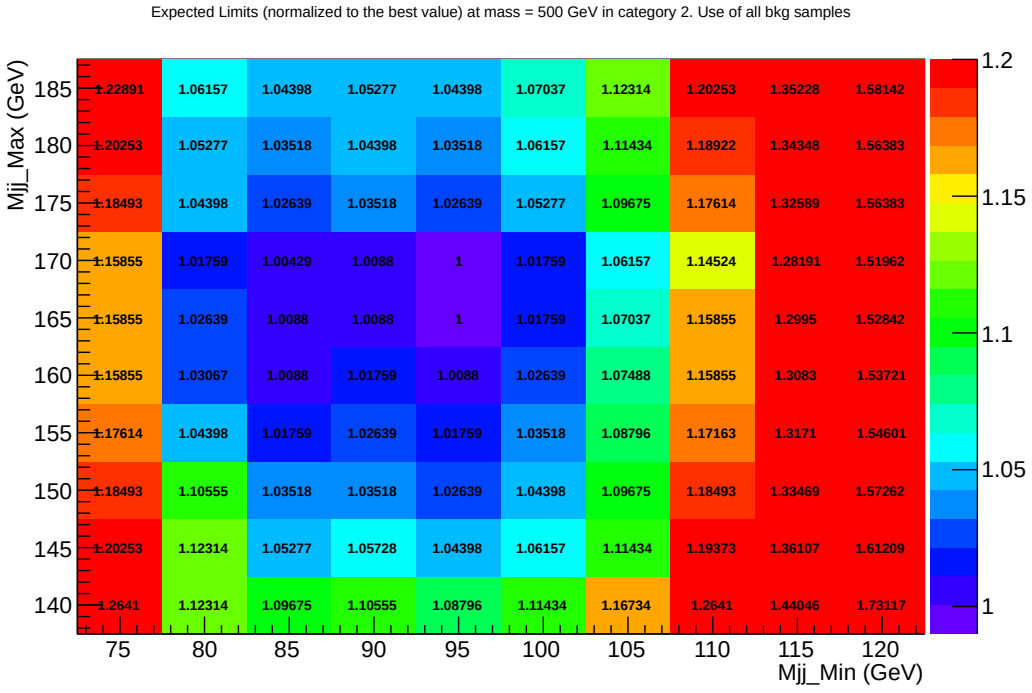


Figure 5.20: Expected 95% CL exclusion limits, normalized to the best value, as a function of the mass window of m_{jj} spectrum for category 0 (2-btag) at the 500 GeV mass hypothesis.

Table 5.10: Additional selections applied for high mass search hypotheses.

Selection (GeV)	Medium purity	High purity
$m_{\gamma\gamma}$	$[122, 128]$	$[120, 130]$
m_{jj}	$[85, 170]$	

5.6.2 Non-resonant analysis

The analysis strategy used as baseline results for non-resonant analysis is similar to the low-mass strategy described above. The method used to extract signal in the low-mass search is to select a signal region in the four-body mass spectrum while fitting the diphoton and dijet mass spectra (2D-fit). In the non-resonant case, the resonant structure in the four-body mass spectra is no longer present, however the difference between the signal and the background four-body spectrum makes this variable a good discriminant for enhancing the sensitivity.

Since the $m_{\gamma\gamma jj}^{KinFit}$ is a model dependent variable, as shown in Figure 5.21, its selections are optimized on the SM non-resonant signal, using procedure similar to the optimization of the resonant analysis, i.e. using the MC simulations as a background and running all the signal extraction procedure, aiming to the best sensitivity. The discriminant value is found to be $m_{\gamma\gamma jj}^{KinFit} = 350$ GeV. In order to exploit the difference in $m_{\gamma\gamma jj}^{KinFit}$ of the BSM non-resonant hypotheses, an additional categorization in $m_{\gamma\gamma jj}^{kinFit}$ is performed, giving rise to a total of four categories for the non-resonant analysis.

Moreover, also $|\cos\theta_{CS}^{*}|$ is a good variable for discriminating signal and background, as shown in Figure 5.18. The angle θ_{HH}^{CS} is defined, in Colin-Soper HH rest frame, as the angle between the Higgs decaying into two photons and the line that bisects the acute angle between the colliding protons. In the rest frame the two Higgs candidates are collinear thus the choice of the one decaying to photons as reference one is arbitrary.

An optimization of $\cos\theta_{CS}^{*}$ selections has been performed on the SM non-resonant signal, as for the $m_{\gamma\gamma jj}^{kinFit}$ spectrum selection. The optimized selections are: $|\cos\theta_{CS}^{*}| < 0.65$ for 1-btag categories and $|\cos\theta_{CS}^{*}| < 0.90$ for 2-btag categories.

The final selections for the non-resonant search are shown in Table 5.11, while Table 5.12 summarizes the final strategies for the signal extraction of resonant and non-resonant searches. Moreover, Tables 5.13, 5.14 and 5.15 show the yields of the MC signal, MC backgrounds and data for the resonant low mass search at 300 GeV, the resonant high mass search and the non-resonant search of the SM di-Higgs production, after the selections described in the previous tables.

Table 5.11: Additional selections applied for non-resonant search.

b-tag category	High purity		Medium purity	
Maximum $\cos\theta_{CS}^{*}$	0.9		0.65	
$m_{\gamma\gamma jj}^{KinFit}$ categorization (GeV)	< 350	> 350	< 350	> 350

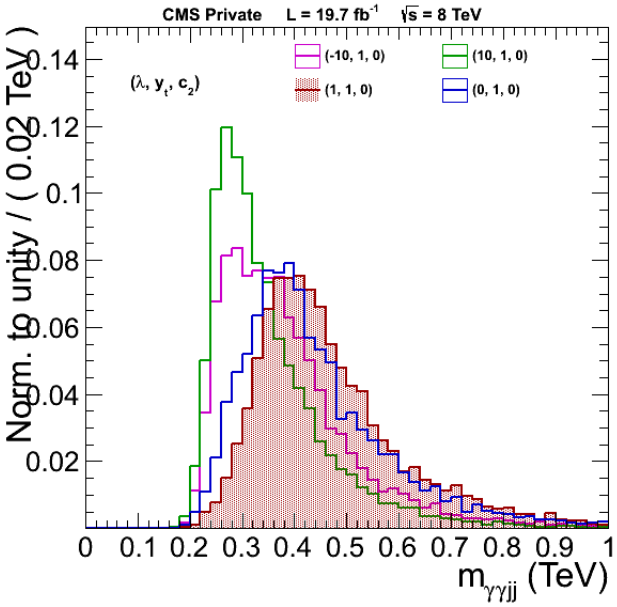


Figure 5.21: Non-resonant $m_{\gamma\gamma jj}$ spectra for different BSM hypotheses.

Table 5.12: Summary of the search analysis methods.

signal hypothesis	fit	select	categorization
Resonant $m_X \leq 400$GeV	$m_{\gamma\gamma}$, m_{jj}^{reg}	$m_{\gamma\gamma jj}^{KinFit,reg}$	2 categories: b tagging
Resonant $m_X \geq 400$GeV	$m_{\gamma\gamma jj}^{KinFit}$	$m_{\gamma\gamma}$, m_{jj}	2 categories: b tagging
Non-resonant	$m_{\gamma\gamma}$, m_{jj}	$\cos\theta^*_{CS}$	4 categories: b tagging and $m_{\gamma\gamma jj}^{KinFit}$

Table 5.13: Yields of the MC signal, MC backgrounds and data for the resonant low mass search at 300 GeV.

Sample	High Purity	Medium Purity
Radion (300 GeV, σ=1 pb)	1.45	1.42
$ggH(\rightarrow\gamma\gamma)$	0.02	0.16
VBF $(H\rightarrow\gamma\gamma)$	-	0.03
$WH(\rightarrow\gamma\gamma)$	-	0.03
$ZH(\rightarrow\gamma\gamma)$	0.01	0.02
$t\bar{t}H\rightarrow\gamma\gamma$	0.04	0.06
$b\bar{b}H\rightarrow\gamma\gamma$	-	0.02
$\gamma\gamma+$ jets	6.2	131
$\gamma+$ jets	-	39
QCD	-	-
DY+Zg+Wgg	-	3.9
tgg+ttgg+ttgj	0.60	0.87
Data	17	161

5.7 Signal and Background models

The signal model is built for each mass hypothesis by fitting the $m_{\gamma\gamma}$, m_{jj} or four-body peak in the Monte Carlo signal, separately for each category. For the $m_{\gamma\gamma}$-fit a sum of a Crystal Ball and

Table 5.14: Yields of the MC signal, MC backgrounds and data for the resonant high mass search.

Sample	High Purity	Medium Purity
Radion (500 GeV, σ=1 pb)	2.68	2.80
Radion (700 GeV, σ=1 pb)	3.43	3.50
Radion (1000 GeV, σ=1 pb)	3.12	3.60
$\gamma\gamma +$ jets	4.4	58
$\gamma +$ jets	-	27
QCD	-	-
DY+Zg+Wgg	-	0.07
tgg+ttgg+ttgj	0.44	0.76
Data	8	29

Table 5.15: Yields of the MC signal, MC backgrounds and data for the non-resonant search of the SM di-Higgs production.

	High Purity		Medium Purity	
Sample	high $m_{\gamma\gamma jj}^{KinFit}$	low $m_{\gamma\gamma jj}^{KinFit}$	high $m_{\gamma\gamma jj}^{KinFit}$	low $m_{\gamma\gamma jj}^{KinFit}$
SM: $\kappa_\lambda = 1$, $\kappa_t = 1$, $c_2 = 0$	2.03	0.28	1.99	0.20
$\kappa_\lambda = 20$, $\kappa_t = 1$, $c_2 = 0$	78.7	102	86.5	96.5
$\kappa_\lambda = 1$, $\kappa_t = 1$, $c_2 = -2$	103	16.2	101	16.5
$ggH(\rightarrow \gamma\gamma)$	0.05	0.04	0.29	0.32
VBF $(H \rightarrow \gamma\gamma)$	0.01	0.01	0.05	0.05
$WH(\rightarrow \gamma\gamma)$	-	-	0.12	0.09
$ZH(\rightarrow \gamma\gamma)$	0.04	0.02	0.07	0.05
$t\bar{t}H \rightarrow \gamma\gamma$	0.16	0.17	0.30	0.17
$b\bar{b}H \rightarrow \gamma\gamma$	-	0.01	0.01	0.04
$\gamma\gamma +$ jets	13	21	151	268
$\gamma +$ jets	-	4.3	28	53
QCD	-	-	-	-
DY+Zg+Wgg	-	0.01	2.3	0.18
tgg+ttgg+ttgj	1.3	2.2	3.3	3.4
Data	41	37	136	319

a Gaussian is used to account for the different classes of photons contributing to the core of the distribution, and it is tuned to describe the MC simulation distributions well. Also for the m_{jj} and four-body fit the sum of a Crystal Ball and a Gaussian is used to account for the different classes of jets contributing to the core of the distribution.

In the low-mass search, the peak position and resolution exhibit limited dependence on the resonance mass and on the category, while for the high mass the peak position follows the m_X evolution and the peak resolution improves for higher masses, corresponding to higher momenta of the jets.

Figure 5.22, 5.23, and 5.24 show the signal fits of $m_{\gamma\gamma}$ and m_{jj} distributions for the radion 300 GeV hypothesis, for the non-resonant analysis and the signal $m_{\gamma\gamma jj}^{KinFit}$ distribution fit for the high mass radion 700 GeV hypothesis, respectively.

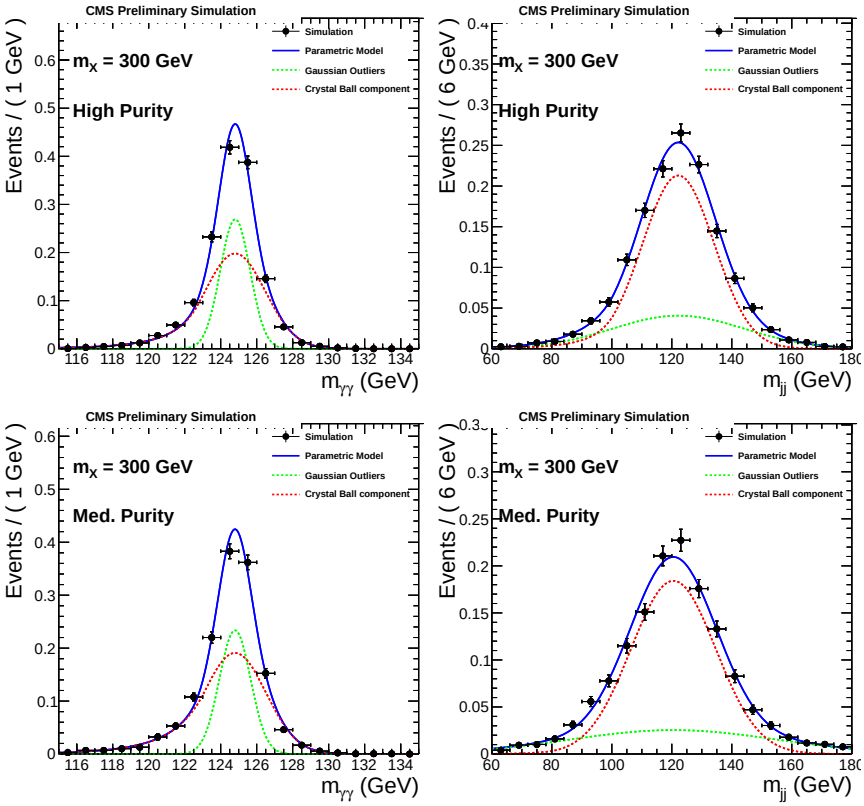


Figure 5.22: Signal modelling for the resonant search at 300 GeV. Top: the fits for the "2-btag" category. Bottom: fits for the "1-btag" category

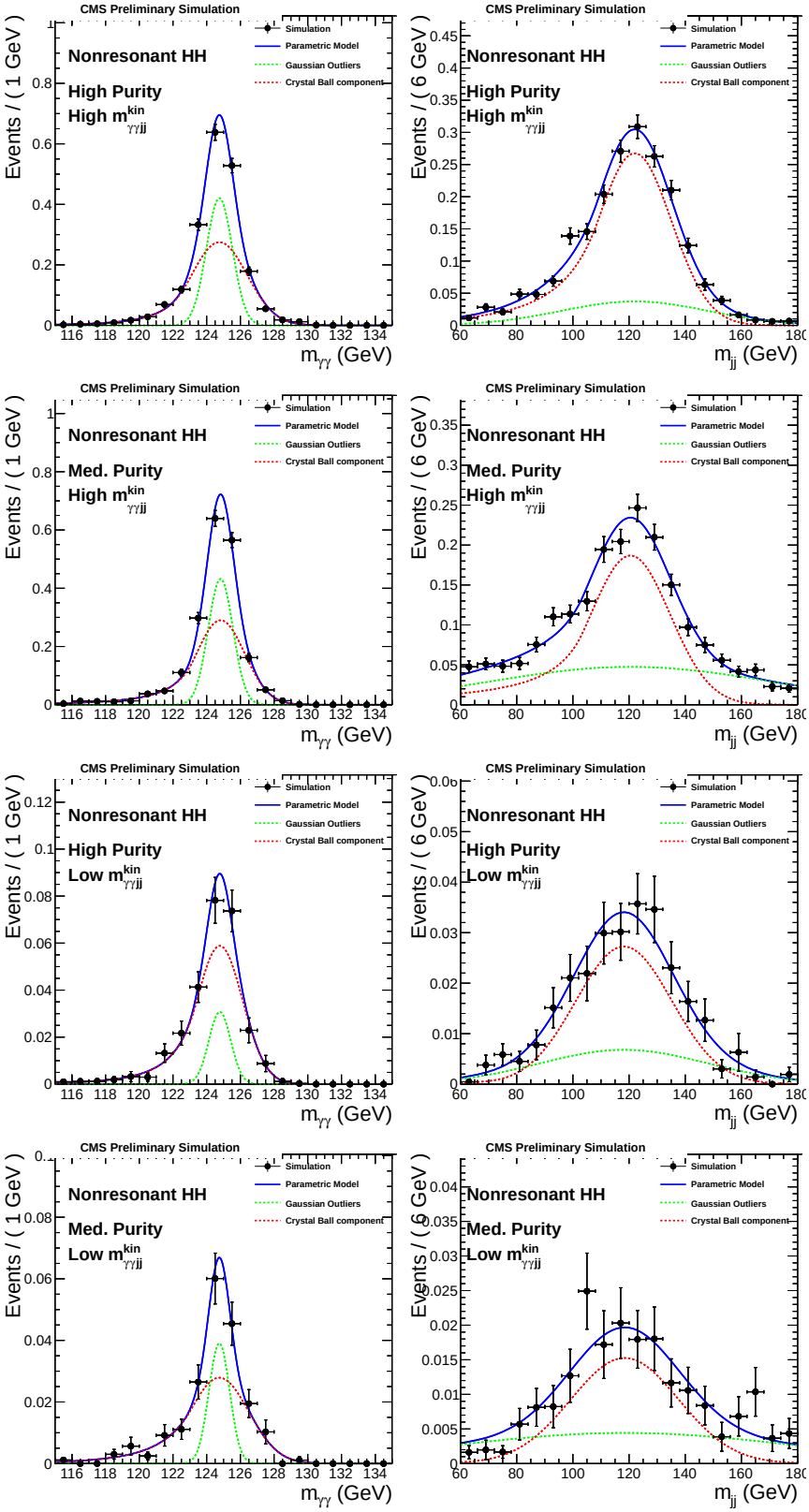


Figure 5.23: Signal modelling for the non-resonant search. Going from the top to the bottom, the $m_{\gamma\gamma}$ (left) and m_{jj} (right) fits are shown for the 4 categories.

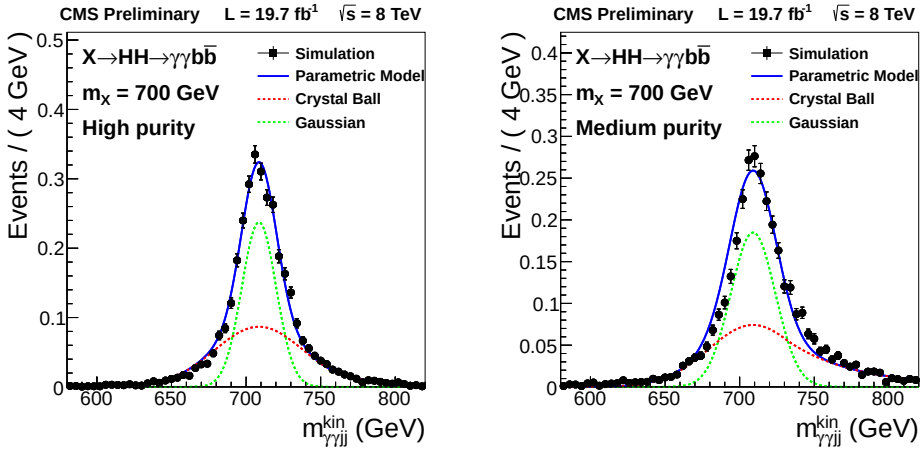


Figure 5.24: Signal modelling of $m_{\gamma\gamma jj}^{KinFit}$ spectrum for the resonant search at 700 GeV.

While the optimization of selections have been performed on simulated background, the final background fit is performed on data. As shown in Figures 5.17 and 5.18, all the mass spectra are smooth falling functions. The choice of the background model is justified by the bias studies.

The background model desired is the one that gives an accurate normalization in the signal region with as few model parameters as possible. This has a direct effect on our final sensitivity: a model that underestimate (overestimate) the number of events in the signal region will worsen (improve) the final sensitivity. The bias studies have been performed using the data control samples as background, Section 5.5

The first step of the bias studies is to construct several truth models fit to the data control sample. The purpose of the truth models is to have several templates from which toy MC experiments can be generated and fitted by candidate fit functions. Three smoothly falling truth models are considered for each dimension (2D-fit case):

- $Ber_N(x) = \sum_{i=0}^{N} \beta_i b_{i,N}(x)$, where $b_{i,N}$ are the Bernstein basis polynomials
- $Exp_N(x) = \sum_{i=0}^{N} \beta_i e^{-\alpha_i x}$
- $Pow_N(x) = \sum_{i=0}^{N} \beta_i x^{-\alpha_i}$

The Bernstein basis polynomials of degree N form a basis for the vector space of polynomials of degree at most N. $Ber_N(x)$, $Exp_N(x)$ and $Pow_N(x)$ functional forms have been chosen as they can easily model a smooth falling background.

The order N to which each of these functions are expanded is determined using a likelihood-ratio test from fits to the control sample, treating each dimension separately. The expansion is continued until the p-value of the test is greater than 0.05.

The truth models are used as templates to generate a large number of toy datasets. Candidate background models and the signal component are fitted to each toy dataset. For the 2D-fit, using the notation $m_{\gamma\gamma} \times m_{jj}$, the background candidates for the 2D-fit are: Exp1×Exp1, Pow1×Pow1, Ber1×Ber1, Ber1×Ber2, Ber2×Ber1 and Exp2×Exp2, Pow2×Pow2 in case of second order truth functions. The bias of a given candidate background model with respect to a truth model is

defined as:

$$\text{Bias} = \text{Median}\left(\frac{N_{sig,gen} - N_{sig,fit}}{\sigma_{N_{sig,fit}}}\right) \tag{5.4}$$

where $N_{sig,gen}$ ($N_{sig,fit}$) is the number of signal events generated in the truth model (fitted by the candidate model).

The candidate background function is considered unbiased if, for every truth model, the equation: Bias < 0.14 holds, i.e. the deviation in the signal strength between the truth model and candidate background function does not exceed the uncertainty of the candidate prediction.

For most of these studies, there was no signal injected into the truth model, so $N_{sig,gen} = 0$. However, for the high-purity category of the low-mass resonant search, because of the small number of events, one signal event was injected into the generated toys in order to better stabilize the fits.

Figure 5.25 shows an example of the full procedure of the bias studies for a truth model and a fit function of the non-resonant analysis, whereas all the non-resonant and resonant 2D-fit results are shown in Table 5.16.

Regarding the 2D-fit, the chosen background model for the 2-btag categories of the non-resonant analysis is a first-order exponential for both dimensions, while for the low-mass resonant analysis a first order Bernstein for both dimensions is chosen. A first order polynomial for both dimensions is chosen for the 1-btag categories of low-mass and non-resonant search.

In addition, the correlation between $m_{\gamma\gamma}$ and m_{jj} spectra was tested injecting it in the toys, by means of a conditional PDF with a linear correlation, and a negligible additional bias was observed.

In the high mass search, a similar procedure is used. However, instead of 2D functions, a 1D generated or fitted function is used to model the four-body spectrum. The truth model functions that are tested are exponential, Landau, Laurent polynomials and power law, as they are the ones that perform the best in shaping the spectrum. The results for the chosen functions are shown in Table 5.17. The functions that prove to be the choice with the least bias are:

$$\text{1-btag: } 1/(x^2 + b)^a$$

$$\text{2-btag: } 1/x^a$$

The fits of the background spectra are shown in Figure 5.26 for the radion 300 GeV hypothesis, in Figure 5.27 for the non-resonant analysis, and in Figure 5.28 for the high mass radion 700 GeV hypothesis.

5.8 Systematic uncertainties

The expected number of signal events is estimated using simulation, Tables 5.13, 5.14 and 5.15. Possible discrepancies in the photon or jet reconstruction and identification, as well as differences in the b-tagging efficiency between data and Monte Carlo samples are corrected through data-to-simulation scale factors. The experimental uncertainties are applied to the reconstructed objects in simulated events by scaling and smearing the relevant observables.

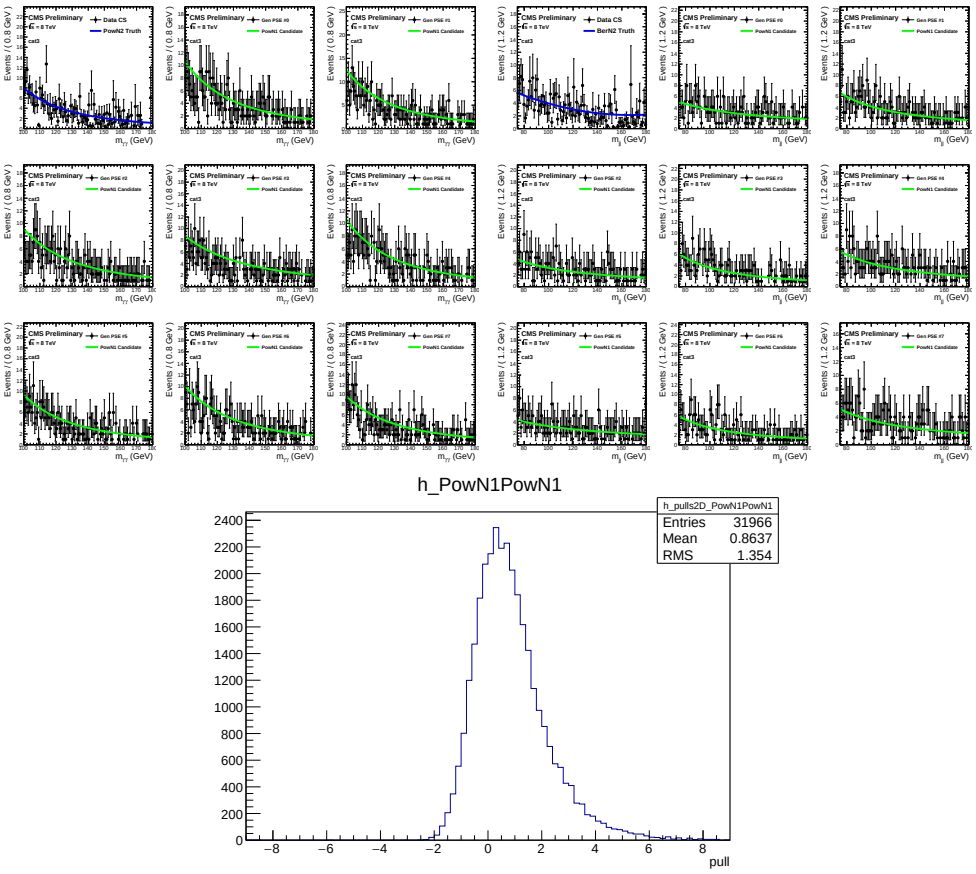


Figure 5.25: 9 plots on the left (right): top left plot shows the fit, in blue colour, with the Pow2 (Ber2) truth model of the $m_{\gamma\gamma}$ (m_{jj}) spectrum of data control sample for the "2-btag $m_{\gamma\gamma jj}^{KinFit} < 350$ GeV" category of the non-resonant analysis, while the other 8 plots show some generated toys, in dots, fitted with the Pow1 (Pow1) background function, in green colour. Bottom: pull distributions of the toys described in the top left and top right plots.

Table 5.16: Bias study results for non-resonant and resonant searches. For each subtable the rows correspond to different truth models and the columns correspond to the candidate background functions.

nonres cat0: 2 b-tag, $m^{\text{kin}}_{\gamma\gamma\text{jj}} > 350$ GeV, $N_{\text{bkg}} = 34.2$

Truth Model	Exp1,Exp1	Pow1,Pow1	Ber1,Ber1	Ber1,Ber2	Ber2,Ber1
BerN0,BerN0	-0.090	-0.096	-0.147	-0.152	-0.155
BerN0,ExpN1	-0.105	-0.147	-0.074	-0.136	-0.061
BerN0,PowN1	-0.048	-0.103	-0.031	-0.110	-0.030
ExpN1,BerN0	-0.092	-0.107	-0.157	-0.147	-0.147
ExpN1,ExpN1	-0.107	-0.159	-0.073	-0.128	-0.063
ExpN1,PowN1	-0.053	-0.112	-0.035	-0.109	-0.018
PowN1,BerN0	-0.018	-0.103	-0.141	-0.152	-0.149
PowN1,ExpN1	-0.105	-0.145	-0.070	-0.137	-0.067
PowN1,PowN1	-0.066	-0.096	-0.036	-0.112	-0.038

nonres cat1: 1 b-tag, $m^{\text{kin}}_{\gamma\gamma\text{jj}} > 350$ GeV, $N_{\text{bkg}} = 139.4$

Truth Model	Exp1,Exp1	Pow1,Pow1	Ber1,Ber1	Ber1,Ber2	Ber2,Ber1
BerN1,BerN1	-0.007	-0.095	0.108	0.103	0.117
BerN1,ExpN1	0.071	-0.005	0.195	0.120	0.190
BerN1,PowN1	0.133	0.058	0.236	0.132	0.235
ExpN1,BerN1	-0.005	-0.075	0.123	0.120	0.112
ExpN1,ExpN1	0.094	0.005	0.205	0.137	0.189
ExpN1,PowN1	0.154	0.069	0.243	0.147	0.247
PowN1,BerN1	0.024	-0.054	0.151	0.146	0.123
PowN1,ExpN1	0.121	0.039	0.242	0.163	0.207
PowN1,PowN1	0.181	0.094	0.288	0.178	0.249

nonres cat2: 2 b-tag, $m^{\text{kin}}_{\gamma\gamma\text{jj}} < 350$ GeV, $N_{\text{bkg}} = 50.8$

Truth Model	Exp1,Exp1	Pow1,Pow1	Ber1,Ber1	Ber1,Ber2	Ber2,Ber1
BerN0,BerN1	-0.060	-0.058	-0.042	-0.039	-0.029
BerN0,ExpN1	-0.003	-0.046	-0.017	-0.045	-0.015
BerN0,PowN1	0.025	-0.010	0.002	-0.038	0.012
ExpN1,BerN1	-0.061	-0.175	0.065	0.075	0.054
ExpN1,ExpN1	-0.049	-0.164	0.089	0.077	0.068
ExpN1,PowN1	-0.025	-0.130	0.123	0.078	0.098
PowN1,BerN1	-0.005	-0.105	0.124	0.132	0.094
PowN1,ExpN1	-0.002	-0.095	0.145	0.123	0.107
PowN1,PowN1	0.028	-0.076	0.180	0.141	0.155

nonres cat3: 1 b-tag, $m^{\text{kin}}_{\gamma\gamma\text{jj}} < 350$ GeV, $N_{\text{bkg}} = 317.4$

Truth Model	Exp1,Exp1	Pow1,Pow1	Ber1,Ber1	Ber1,Ber2	Ber2,Ber1	Exp2,Exp2	Pow2,Pow2
BerN0,BerN2	0.217	0.171	0.261	0.068	0.064	0.203	0.158
BerN0,ExpN1	0.130	0.056	0.180	0.069	0.184	0.116	0.050
BerN0,PowN1	0.193	0.124	0.256	0.084	0.251	0.174	0.112
ExpN1,BerN2	0.169	-0.050	0.482	0.275	0.168	0.120	-0.109
ExpN1,ExpN1	0.049	-0.182	0.412	0.279	0.298	0.008	-0.219
ExpN1,PowN1	0.132	-0.094	0.499	0.272	0.373	0.086	-0.135
PowN2,BerN2	0.341	0.106	0.655	0.429	0.305	0.275	0.010
PowN2,ExpN1	0.224	0.022	0.562	0.437	0.419	0.165	-0.098
PowN2,PowN1	0.287	0.056	0.644	0.435	0.489	0.228	-0.016

res 270 GeV cat0: 2 b-tag, $m^{\text{r, kin}}_{\gamma\gamma\text{jj}} \in [260,280]$ GeV, $N_{\text{bkg}} = 12.6$

Truth Model	Exp1,Exp1	Pow1,Pow1	Ber1,Ber1	Ber1,Ber2	Ber2,Ber1	Exp2,Exp2	Pow2,Pow2
BerN0,BerN0	-0.051	0.080	-0.009	-0.034	-0.004	-0.073	-0.022
BerN0,ExpN2	-0.064	-0.054	-0.018	-0.028	-0.005	0.000	-0.042
BerN0,PowN1	-0.049	-0.089	0.005	-0.015	0.023	0.008	-0.038
ExpN2,BerN0	-0.120	-0.139	0.015	0.009	-0.039	-0.114	-0.121
ExpN2,ExpN2	-0.098	-0.176	0.018	0.023	-0.022	-0.073	-0.108
ExpN2,PowN1	-0.140	-0.143	0.010	0.015	-0.055	-0.103	-0.114
PowN1,BerN0	-0.033	-0.169	0.055	0.033	-0.047	-0.099	-0.070
PowN1,ExpN2	-0.114	-0.135	0.066	0.049	-0.026	-0.109	-0.086
PowN1,PowN1	-0.086	-0.146	0.070	0.013	-0.037	-0.081	-0.116

res 270 GeV cat1: 1 b-tag, $m^{\text{r, kin}}_{\gamma\gamma\text{jj}} \in [260,280]$ GeV, $N_{\text{bkg}} = 143.2$

Truth Model	Exp1,Exp1	Pow1,Pow1	Ber1,Ber1	Ber1,Ber2	Ber2,Ber1
BerN0,BerN2	0.121	0.033	0.110	0.098	0.082
BerN0,ExpN1	0.073	-0.008	0.013	0.100	0.233
BerN0,PowN1	0.170	0.072	0.332	0.109	0.340
ExpN1,BerN2	0.116	-0.055	0.283	0.271	0.072
ExpN1,ExpN1	0.062	-0.098	0.415	0.272	0.224
ExpN1,PowN1	0.171	0.005	0.527	0.291	0.350
PowN1,BerN2	0.198	0.026	0.359	0.351	0.113
PowN1,ExpN1	0.156	-0.020	0.499	0.349	0.272
PowN1,PowN1	0.258	0.080	0.603	0.358	0.385

res 300 GeV cat0: 2 b-tag, $m^{\text{r, kin}}_{\gamma\gamma\text{jj}} \in [290,310]$ GeV, $N_{\text{bkg}} = 17.6$

Truth Model	Exp1,Exp1	Pow1,Pow1	Ber1,Ber1	Ber1,Ber2	Ber2,Ber1	Exp2,Exp2	Pow2,Pow2
BerN0,BerN1	-0.150	-0.125	-0.097	-0.059	-0.055	-0.120	-0.018
BerN0,ExpN2	-0.167	-0.086	-0.042	-0.052	-0.011	-0.076	-0.017
BerN0,PowN1	-0.074	-0.088	-0.003	-0.055	0.005	-0.062	-0.041
ExpN2,BerN1	-0.140	-0.140	-0.064	-0.075	-0.037	-0.108	-0.055
ExpN2,ExpN2	-0.039	-0.106	-0.005	-0.109	0.023	-0.088	-0.057
ExpN2,PowN1	-0.037	-0.080	-0.006	-0.024	0.007	-0.061	-0.039
PowN1,BerN1	-0.140	-0.131	-0.073	-0.084	-0.053	-0.136	-0.043
PowN1,ExpN2	-0.153	-0.081	-0.020	-0.067	-0.009	-0.105	-0.042
PowN1,PowN1	-0.071	-0.068	0.024	-0.053	0.012	-0.068	-0.030

res 300 GeV cat1: 1 b-tag, $m^{\text{r, kin}}_{\gamma\gamma\text{jj}} \in [290,310]$ GeV, $N_{\text{bkg}} = 195.5$

Truth Model	Exp1,Exp1	Pow1,Pow1	Ber1,Ber1	Ber1,Ber2	Ber2,Ber1
BerN0,BerN3	-0.070	-0.142	-0.461	0.099	0.086
BerN0,ExpN1	0.121	0.042	0.226	0.102	0.228
BerN0,PowN1	0.213	0.146	0.309	0.101	0.326
ExpN1,BerN3	-0.155	-0.139	-0.463	0.132	0.091
ExpN1,ExpN1	0.088	-0.051	0.277	0.135	0.234
ExpN1,PowN1	0.185	0.039	0.369	0.145	0.321
PowN1,BerN3	-0.106	-0.141	-0.421	0.181	0.103
PowN1,ExpN1	0.136	-0.002	0.314	0.187	0.251
PowN1,PowN1	0.242	0.090	0.399	0.177	0.339

Table 5.17: Results of the bias study for the $m_{\gamma\gamma jj}^{KinFit}$ spectrum. For each subtable the rows correspond to different truth models, and the columns correspond to the candidate functions at three different points on the four-body spectrum.

1 btag	Candidate at 500 GeV	Candidate at 700 GeV	Candidate at 1000 GeV
ExpN1	-0.0655	0.0813	0.0135
LanN1	0.0244	-0.1257	0.1398
LauN1	-0.0483	0.0105	0.0701
PowN1	-0.0129	0.0030	0.0621
2 btag	Candidate at 500 GeV	Candidate at 700 GeV	Candidate at 1000 GeV
ExpN1	0.0942	0.0169	-0.0431
LanN1	0.0604	-0.0055	-0.0062
LauN1	0.0059	0.0053	0.0036
PowN1	0.0105	0.0024	0.0022

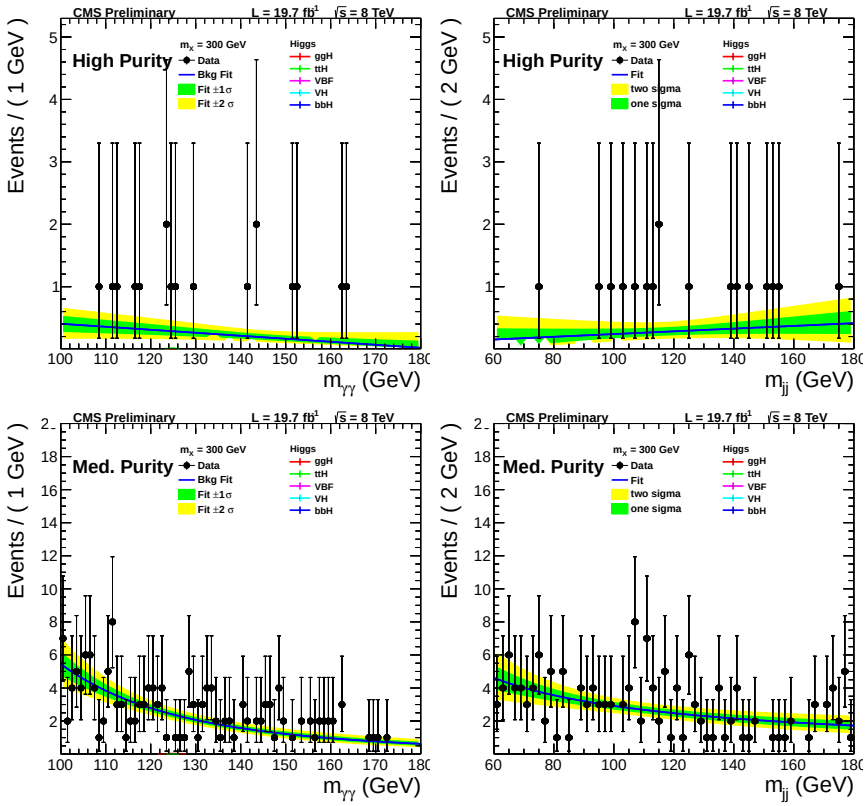


Figure 5.26: Background modelling for the resonant search at 300 GeV. Top: fits for the "2-btag" category. Bottom: fits for the "1-btag" category.

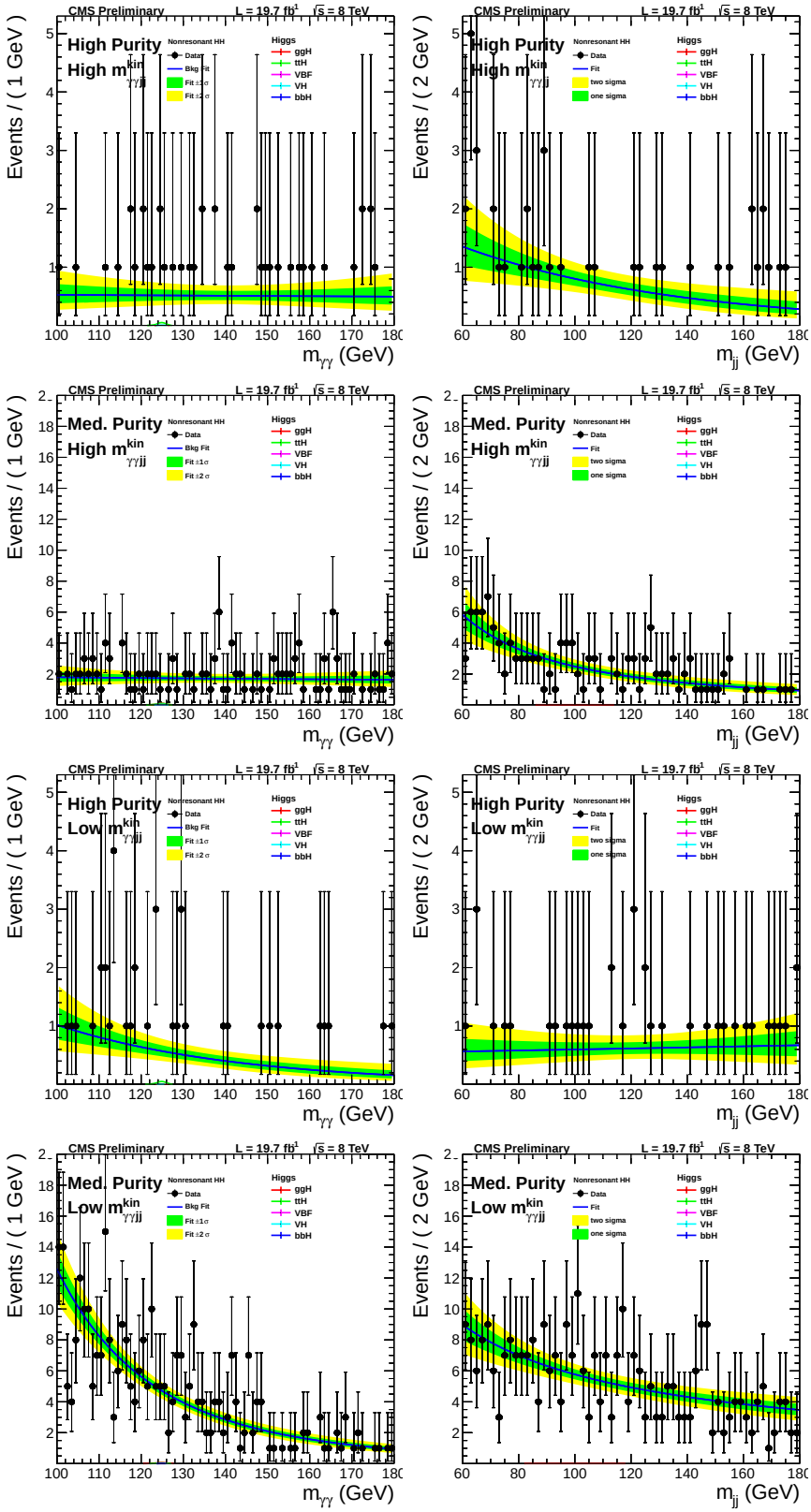


Figure 5.27: Background modelling for the non-resonant search. Going from the top to the bottom, the $m_{\gamma\gamma}$ (left) and m_{jj} (right) fits are shown for the 4 categories.

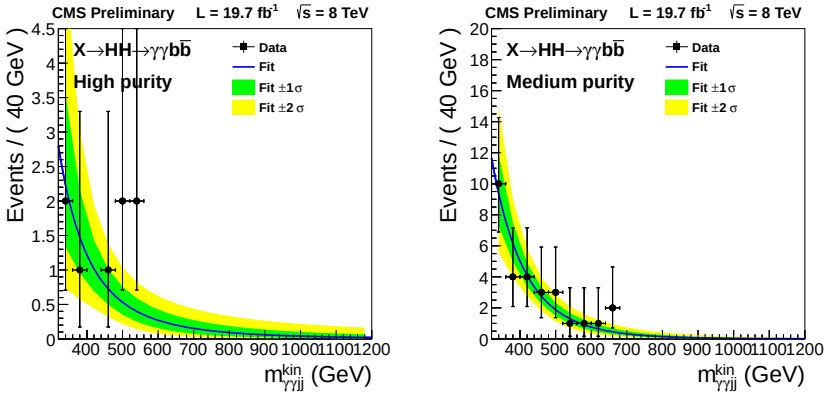


Figure 5.28: Background modelling of $m_{\gamma\gamma jj}^{KinFit}$ spectrum for the resonant search at 700 GeV.

The total normalisation uncertainty related to the uncertainty in the estimation of the collected luminosity is taken to be 2.6% [79]. The other main contributions to the systematic uncertainties can be separated in two categories: the photon-related and jet-related

- **Photons:** the photon-related uncertainties are taken from [63]. An uncertainty between 0.23 and 0.93% is considered on the energy resolution (PER) and between 0.12 and 0.88% on the energy scale (PES), depending on η_γ and the electromagnetic shower shape. When $p_T^\gamma > 100$ GeV the uncertainty of the energy scale is conservatively increased up to 1%. A 1% normalization uncertainty is assumed on the offline photon selection efficiency and on the trigger efficiency. An additional conservative normalization uncertainty of 5% is assumed for the high mass region to account for the differences in the p_T spectrum of the signal photons and of the electrons from $Z \to e^+e^-$ used to estimate the quoted uncertainties.

- **Jets:** The jet energy scale uncertainty (JES) is accounted for by varying the jet response by 1-2%, depending on the kinematics, while the jet energy resolution uncertainty (JER) by varying the jet resolution by 5% [80]. The $m_{\gamma\gamma jj}^{KinFit}$ acceptance uncertainties of the "$m_{\gamma\gamma jj}^{KinFit} > 350$ GeV" and "$m_{\gamma\gamma jj}^{KinFit} < 350$ GeV" categories have been found to be anti-correlated. An additional 1% uncertainty on the four-body mass response accounts for effects in the high-mass region related to the partial overlap between the two b-jets coming from the Higgs boson decay. The uncertainty on the b-tagging efficiency is estimated by varying the b-tagging scale factor by one standard deviation in each category, as described in Section 5.4.4, and the related systematics were shown to be anti-correlated between the two categories.

The additional theory systematic uncertainties are considered for the SM single contribution including scale dependence to account for the missing order effects and the dependency on proton parton density functions [4]. On the other hand, no theory systematics are assumed on the other signals. Finally, an additional systematic uncertainty of 0.24 GeV is assigned to the experimental knowledge of the Higgs mass [10].

The impact of the quoted systematic uncertainties on the results is summarized in the Table 5.18. As the analysis is limited by the statistical precision, the systematic uncertainties worsen

the sensitivity by 1.5% (3.8%) at most in the resonant (non-resonant) searches.

Table 5.18: Systematic uncertainties by analysis.

Common normalization uncertainties	
Luminosity	2.6%
Diphoton trigger acceptance	1.0%
Resonant low mass and non-resonant analyses: 2D fit to $m_{\gamma\gamma}$ and m_{jj}^{reg} (m_{jj})	
Normalization uncertainties	
Photons selection acceptance	1.0%
$p_{T,j}$ selection acceptance (JES & JER)	1%
b tagging efficiency uncertainty 2 b tag category	5.0%
b tagging efficiency uncertainty 1 b tag category	
Low mass resonant and non-resonant $m_{\gamma\gamma jj}^{KinFit} < 350$ GeV	2.1%
Non-resonant $m_{\gamma\gamma jj}^{KinFit} > 350$ GeV	2.8%
$m_{\gamma\gamma jj}^{KinFit}$ acceptance (PES $\oplus$ JES & PER $\oplus$ JER)	
Low mass resonant	1.5%
Non-resonant $m_{\gamma\gamma jj}^{KinFit} < 350$GeV categories	1.5%
Non-resonant $m_{\gamma\gamma jj}^{KinFit} > 350$GeV categories	0.5%
Shape uncertainties	
$m_{jj}^{regression}$ parametric resolution shift (JER)	$\frac{\Delta\sigma_{\rm jj}}{\sigma_{\rm jj}} = 10\%$
$m_{jj}^{regression}$ parametric scale shift (JES uncertainty)	$\frac{\Delta m_{jj}}{m_{jj}} = 2.6\%$
$m_{\gamma\gamma}$ parametric resolution shift (PER)	$\frac{\Delta\sigma_{\gamma\gamma}}{\sigma_{\gamma\gamma}} = 5\%$ or $\frac{\Delta\sigma_{\gamma\gamma}}{m_{\gamma\gamma}} = 0.05\%$
$m_{\gamma\gamma}$ parametric scale shift (PES$\oplus$M(H) uncertainty)	
Low mass resonant	$\frac{\Delta m_{\gamma\gamma}}{m_{\gamma\gamma}} = 0.4 \oplus 0.2\%$
Non-resonant	$\frac{\Delta m_{\gamma\gamma}}{m_{\gamma\gamma}} = 0.5 \oplus 0.2\%$
High mass resonant analysis: 1D fit to $m_{\gamma\gamma jj}^{KinFit}$	
Normalization uncertainties	
Photons selection acceptance	1.0%
b tagging efficiency uncertainty 2 b tag category	5.%
b tagging efficiency uncertainty 1 b tag category	2.8%
m_{jj} and $p_{T,j}$ selection acceptance (JES & JER)	1.5%
$m_{\gamma\gamma}$ selection acceptance (PES & PER)	0.5%
Extra high $p_{T,\gamma}$ normalization uncertainty	5.0%
Shape uncertainties	
Parametric absolute shift (PES $\oplus$ JES)	$\frac{\Delta m_{\gamma\gamma\rm jj}}{m_{\gamma\gamma\rm jj}} = 0.45 \oplus (0.8 \oplus 1.0) = 1.4\%$
Parametric shift (PER $\oplus$ JER)	$\frac{\Delta\sigma}{\sigma_{\gamma\gamma\rm jj}} = 10\%$

5.9 Results

5.9.1 Resonant search results

The evolution of the signal efficiency for the selections described in section 5.6 is shown in Figure 5.29. In particular, the left plot shows the various step efficiencies associated to the final selection Tables 5.8, 5.9 and 5.10:

- Red: at least 2 photons are required in the event, following the usual photon η acceptance requirements, $p_{T\gamma,\rm lead} > 40$ GeV, $p_{T\gamma,\rm sublead} > 30$ GeV, at least 2 jets are required in the

event, following $p_T > 10$ GeV, $\Delta R(jet, \gamma) > 0.3$, $|\eta| < 2.5$.

- Pink: the H $\to \gamma\gamma$ preselections are required on the photons, $p_{T\gamma}/m_{\gamma\gamma} > 1/3(1/4)$ for the leading (trailing) photon, and a $m_{\gamma\gamma}$ mass window 100-180 is required.
- Blue: the H $\to \gamma\gamma$ cut-based photon ID is required (a.k.a. CiC SuperTight working point) on the two photons.
- Cyan: the jets are require to pass the loose jet ID, $p_T > 25$ GeV, $\Delta R(jet, \gamma) > 0.5$.
- Green: at least one jet is required to be b-tagged by the CSV Medium working point
- Dark yellow: the analysis mass selections are applied: $m^{KinFit,reg}_{\gamma\gamma jj}$, for the low mass analysis (circles) ; m_{jj}, and $m_{\gamma\gamma}$, selections for the high mass analysis (squares).

The SM single Higgs samples are shown for comparison, with the mass window selections corresponding to the $m_X = 300$ GeV hypothesis. The final efficiency (after mass selections) does not contain the data/MC correction factors (in particular the btag SF).

The signal efficiency increases from $m_X = 260$ GeV to 900 GeV because of better photon and jet reconstruction efficiencies with the mass increase. The efficiency starts to drop after 900 GeV due to the merging of b-quarks into a single jet. The contribution of both categories to the total efficiency is roughly equally.

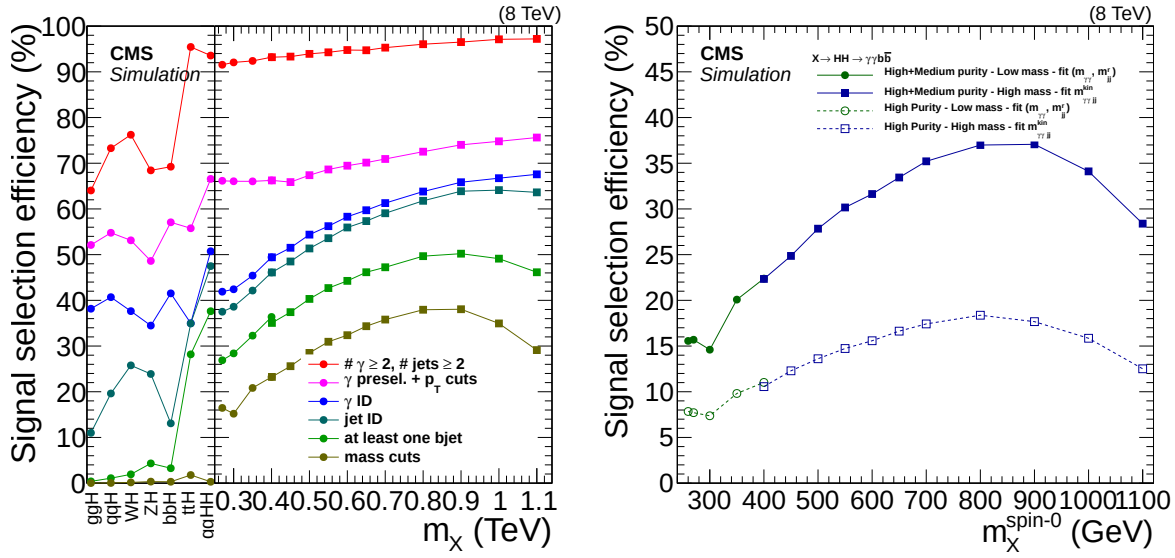


Figure 5.29: Cut flow efficiency (left) and signal efficiency (right) divided in categories, associated to the final selection Tables 5.8, 5.9 and 5.10.

After applying the selections and the signal extraction strategies from the previous sections, no significant excess is observed over the background in the low and high mass searches, therefore upper limits are computed. The modified frequentist approach CL_s, taking the profile likelihood as a test statistics [81, 82], is used with an asymptotic approximation.

The 95% CL expected and observed median upper limits are shown in Figure 5.30. The green and yellow bands represent the 1σ and 2σ confidence intervals around the expected limit. A zoom in the low mass region for the high-purity category only is also shown. The vertical dashed

line indicates the transition region between the low mass and the high mass analysis. The limit for the mass hypothesis $m_X = 400$ GeV is shown for both the methods.

Finally, Table 5.19 provides the values of the observed and expected limits. The analysis excludes the presence of a radion with masses below 975 GeV for the radion scale $\Lambda_R = 1$ TeV. The results are consistent with the expected sensitivity for the hypothesis with no signal.

Interpretation of the WED models are given, by means of theory lines overlaid with the limits, where "bulk" ("RS1") refers to models where the SM fields do (do not) propagate in the extra dimension. On the other hand, "radion" refers only to radion bulk models with specific values of Λ.

The analysis selections were designed to minimize the sensitivity to the spin hypothesis. Thus, the results shown for the spin 0 model, such as radion, can be applied also to a spin 2 model such as KK graviton. In fact, the difference in efficiency between the radion and the KK graviton does not exceed 3%. The RS1 KK graviton with masses between 320 and 400 GeV with $k/M_P = 0.2$ is excluded.

Table 5.19: Observed and median expected 95% CL limit from the low mass ($m_X \leq 400$GeV) and high mass regions ($m_X \geq 400$GeV).

m_X	Observed limit (fb)	Expected limit (fb)	Observed limit (fb)	Expected limit (fb)
			High-purity category only	
260	1.55	1.65	1.75	1.99
270	2.34	1.68	1.70	1.96
300	3.30	1.98	5.35	2.38
350	0.99	1.62	1.45	1.98
400	1.19	1.18	1.71	1.42
400	1.79	1.52		
450	1.48	1.18		
500	1.86	0.96		
550	1.38	0.83		
600	1.11	0.68		
650	0.86	0.60		
700	0.36	0.53		
800	0.31	0.44		
900	0.31	0.40		
1000	0.32	0.39		
1100	0.39	0.45		

5.9.2 Non-resonant search results

The same as the resonant searches statistical approach is used to estimate the significance of a potential excess over the background and to set the 95% CL upper limits.

For the SM-like search no significant excess is observed over the background. The observed (expected) upper limit on the SM-like HH $\rightarrow \gamma\gamma b\bar{b}$ production is 1.85 (1.56) fb for HH produc-

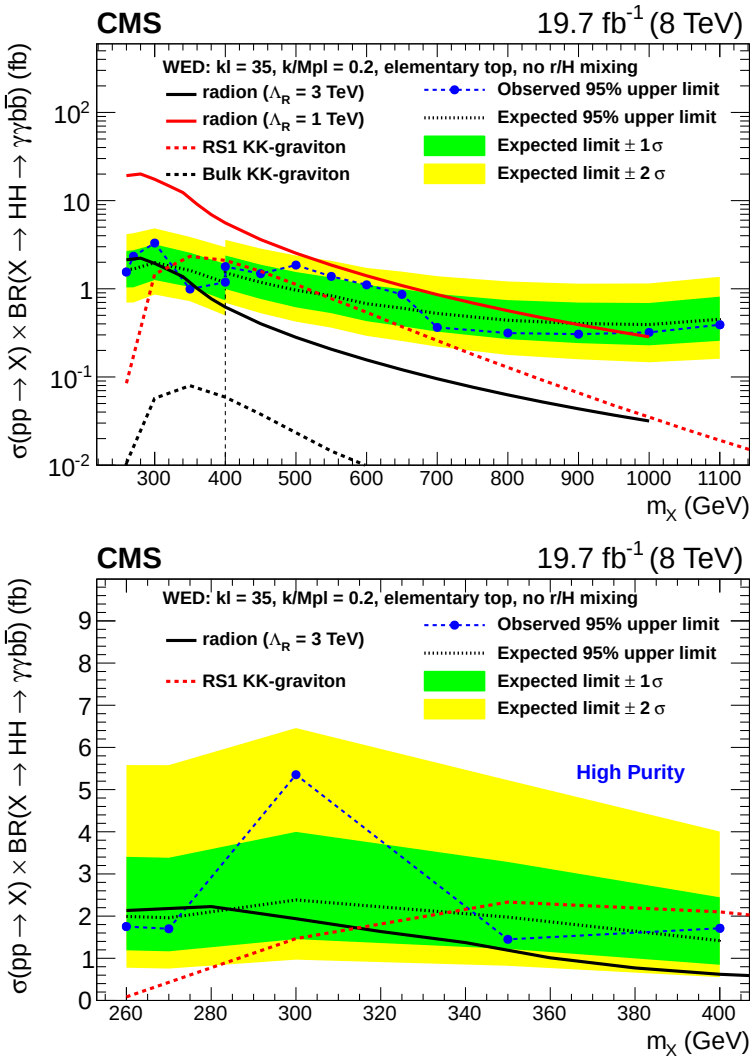


Figure 5.30: Expected 95% CL upper limits on the cross section times branching ratios $\sigma(pp \to X) \times \mathrm{BR}(X \to HH \to \gamma\gamma \mathrm{b\bar{b}})$ obtained by a combination of two categories (top) and a zoom at low mass with high purity category only (bottom). Theory lines corresponding to WED models with radion, RS1 KK graviton and bulk KK graviton are also shown. The results are obtained using the asymptotic CL_s approach. The vertical dashed line shows the separation between the low mass analysis and high mass analysis. The limits for $m_X = 400$ GeV are shown with both methods.

tion and, assuming the SM Higgs branching fraction, it becomes 0.71 (0.60) pb. Using the SM signal strength modificator an observed (expected) limit of $\mu_{\mathrm{HH}} = 74\ (62^{+37}_{-22})$ is obtained, where the theoretical uncertainty of the SM cross section prediction are applied.

Regarding the BSM non-resonant search, Figure 5.31 provides how the signal efficiency evolves, after the selections of the non-resonant analysis strategy, as function of κ_t, c_2 and κ_λ in bins of $m^{KinFit}_{\gamma\gamma jj}$. This figure gives a further insight on the kinematics of the various BSM signals.

In Figure 5.32 the cross section 95% CL limit as a function of the κ_λ parameter is shown, where the limits are compared to the predictions assuming the SM Higgs branching fractions [83]. The κ_λ values are excluded below -17.5 and above 23. Those results are obtained by extrapolating the limits above the highest simulated value of κ_λ. This approach is relying on the fact that the signal shapes are very similar for large values of κ_λ [84].

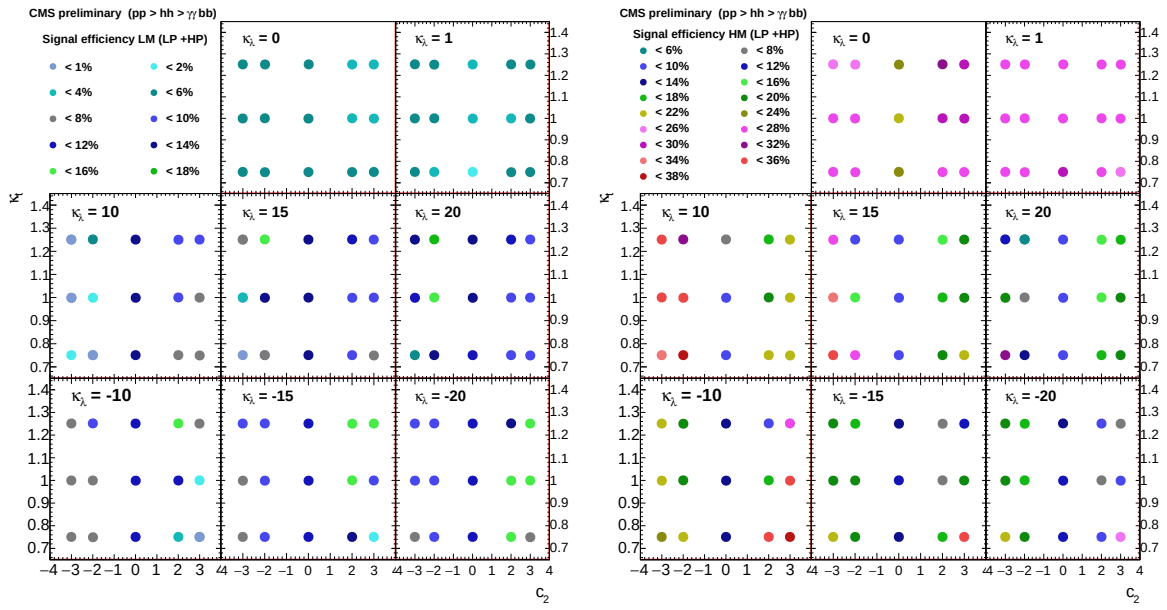


Figure 5.31: Signal efficiencies in terms c_2, κ_t and κ_λ for the low m_{HH} categories (left) and high m_{HH} categories (right). In both figures we sum the middle and high b-tag purity categories.

Finally, the upper Figure 5.33 shows the observed and expected limits as a function of κ_λ, κ_t and c_2 parameters. Whereas the bottom Figure 5.33 shows the comparison of the observed limits with the signal theoretical calculation contours [83]. Large values of c_2, κ_λ or κ_t can be excluded depending on the interference pattern specific for each combination of the parameters.

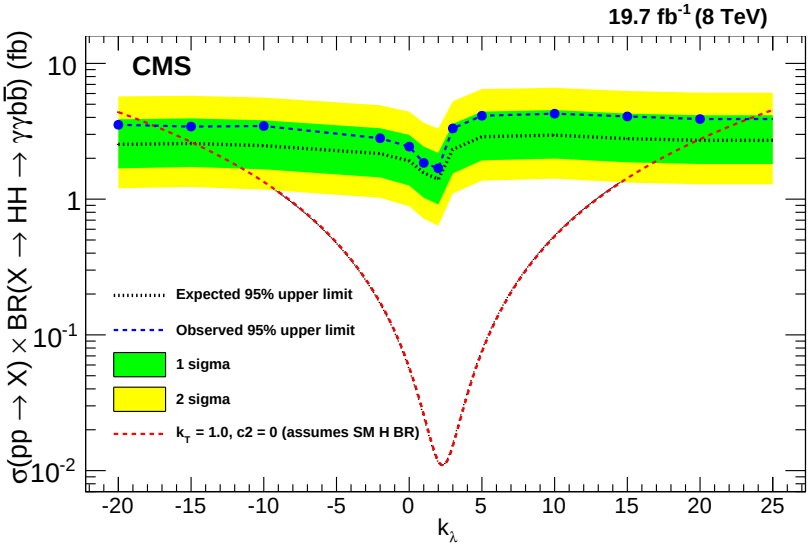


Figure 5.32: Expected 95% CL upper limits on the cross section times branching ratios $\sigma(pp \to X) \times \mathrm{BR}(X \to HH \to \gamma\gamma b\bar{b})$ for the BSM-like analysis varying only κ_λ, while all the other parmeters are fixed to the SM predictions, $kappa_t = 1$ and $c_2 = 0$.

5.10 Future prospects

In this Chapter results of the resonant and non-resonant di-Higgs production in the $\gamma\gamma\mathrm{b\bar{b}}$ final channel decay are presented, using data collected during LHC Run I by the CMS experiment at centre-of-mass energy of 8 TeV.

The sensitivity to these searches can be strongly enhanced at the higher centre-of-mass energy of the LHC Run II, 14 TeV. Despite the high photon resolution, the sensitivity of the search for high mass candidates is limited by jet merging. On the other hand, the sensitivity of low mass resonant searches can be ameliorated with a specific optimization of the photon identification.

Concerning the high mass searches, the more the Higgs is boosted the more the two b-jets are merged, worsening the performance of default algorithms for jets reconstruction. The usage of specific jets reconstruction algorithms that exploit merged topology can improve the sensitivity. These algorithms allow the identification and reconstruction of sub-jets within the unique jet coming from the two merged jets. These type of jet reconstruction and identification methods have been successfully used in many high mass searches such as resonant HH →4b [85]. After the implementation of the merged jet reconstruction, boosted $\gamma\gamma\mathrm{b\bar{b}}$ topologies require an additional optimization of the merged jet selections in order to reach an optimal background rejection.

In the standard H $\to \gamma\gamma$ analysis a multivariate photon identification is used in order to optimally exploit the final state kinematic in the signal events. A similar identification method for photons can be used in order to improve the reconstruction of low mass di-Higgs searches. However, because of the different kinematics a specific multivariate photon identification is needed, in fact in the di-Higgs search both Higgs boson systems are boosted. This provides an additional handle, different from H $\to \gamma\gamma$ decay search, to maximize the signal to background ratio for

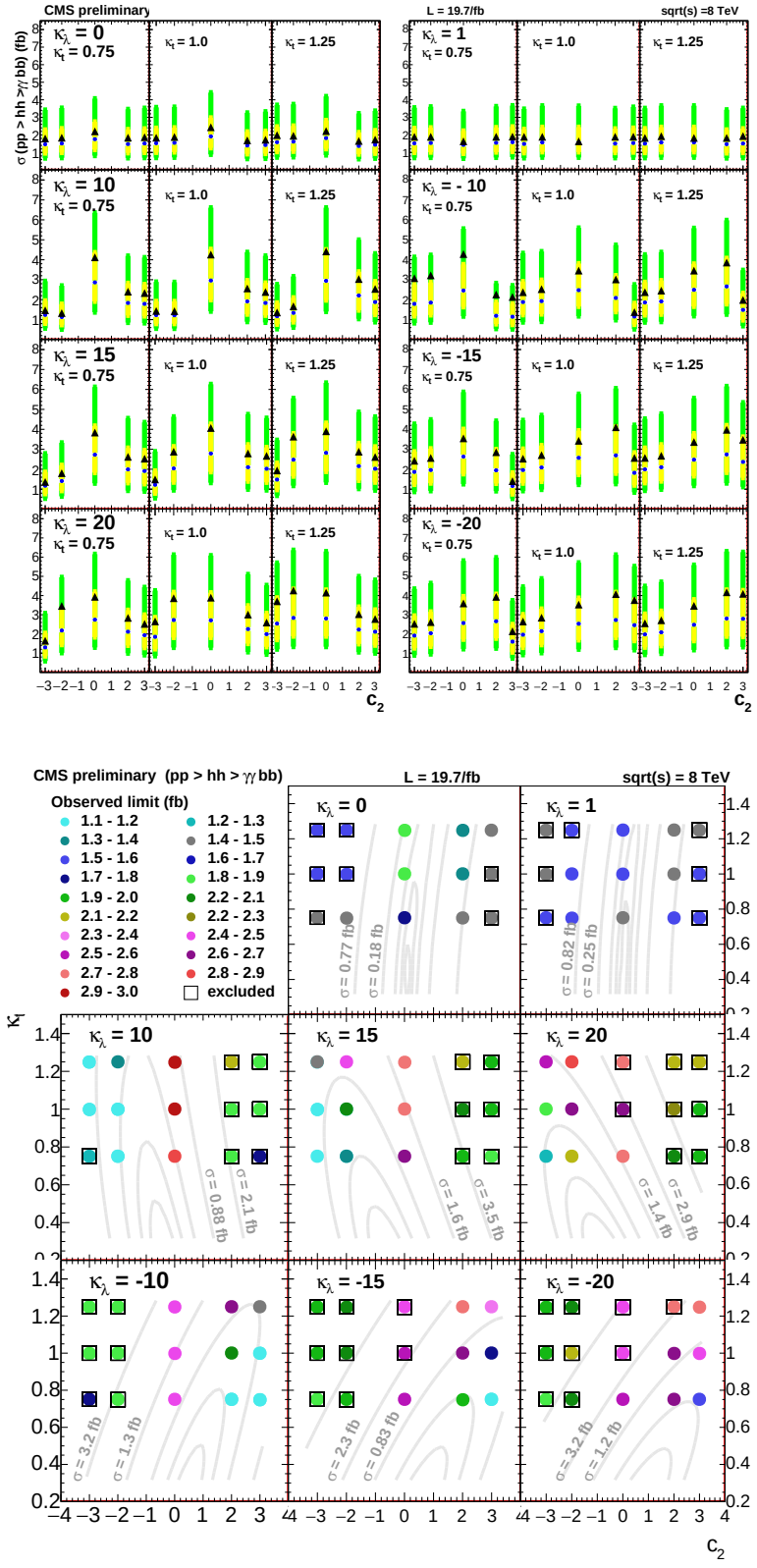


Figure 5.33: Top: Expected and observed limits for non-resonant double Higgs production as function of c_2 for different values of κ_t and κ_λ. Bottom: Observed limits for non-resonant double Higgs production in the c_2 and κ_t planes for different values of κ_λ. The results are compared to theory prediction. The gray lines represent the contour of cross section, as calculated in [83]. The exclusions are derived assuming SM Higgs boson branching fractions. The black boxes symbolise the excluded combination of parameters.

photons in this topology.

An alternative method to probe any deviation with respect to the SM is the precise measurements of the Higgs boson couplings, especially to massive particles which give the largest quantum corrects. In particular, the study of the Higgs-top coupling is fundamental because a change in the Yukawa sign and/or absolute value with respect to its SM value ($\approx$1) would be a signal of an origin of the fermion masses different from the SM one and could spoil the unitarity and renormalizability of the theory. Additionally, any deviation from the SM of the Higgs-top coupling would also affect the loop quantum corrections to the Higgs mass.

Within the Standard Model, b-jets in the final state can be originated in processes with associated production of the Higgs boson with b or t-pairs. The two b-jets would be non-resonant and the analysis of these events would give access to the study of the ttH or bbH coupling strengths. Because of the different decay products, separate specific analyses have to be prepared regarding ttH and bbH production, in fact top-quark decays into a b-quark and an additional W boson.

Although the current measurement of the Higgs-top coupling using final states such as $H \to \gamma\gamma$, $H \to WW$ and $H \to b\bar{b}$ [7] is consistent with the prediction, the measurement of the exclusive ttH production has a cross-section 2.3 times bigger than the SM, thus further investigation is mandatory.

Given the increase in the production cross-sections from 8TeV to 14 TeV, i.e. about 4 times bigger for the ttH production and 2.5 times for the bbH production, with a less accentuated background cross-section growth, a great improvement in the sensitivity of these searches is foreseen from a specific optimization of the analysis. Because of all these reasons, Run II analyses with $\gamma\gamma b\bar{b}$ topology are absolutely necessary to better shed light on the couplings.

Furthermore, any deviation from the SM of the Higgs boson coupling to the top-quark affects also the Higgs self-coupling, due to interference with other di-Higgs production modes. Thus these two measurements are intimately related and any sensitivity improvement will permit to further probe deviations in the non-resonant di-Higgs production modes due to anomalies in the Higgs self-coupling and/or in the Higgs-top coupling.

CHAPTER 6

CONCLUSIONS

The most relevant achievements of my three-years Ph.D. work in the CMS experiment have been described. This research activity has been based on testing the standard model (SM) of particle interactions. Indeed, both a direct and indirect search of phenomena beyond the SM (BSM) are presented. Concerning the indirect search, the contribution to the improvement of the sensitivity of precision measurements of the Higgs boson properties in the $H \to \gamma\gamma$ channel is described. The improvement of the sensitivity has been achieved by means of detector calibration and implementation of the detector conditions into the Monte Carlo simulation. On the other hand, the direct search has been performed looking for the di-Higgs production decaying into $\gamma\gamma b\bar{b}$ final state.

On the detector side, during the Ph.D. I have concentrated my work on many aspects of the ECAL detector calibration and validation of its conditions. For example, even though not specifically addressed in this thesis, I was involved in the validation of the ECAL conditions for the data collected during the 2011 and 2012. This permitted a full comprehension of the detector evolution, and gave the knowledge and the tools to implement and validate the ECAL conditions of the noise and the response variation injected in the Monte Carlo simulation. The improvement in the simulation directly affected the $H \to \gamma\gamma$ analysis, in particular the photons' shower shapes and isolation, giving an improvement of about 10% to the final sensitivity.

In addition, in the context of the commissioning of ECAL calibration for Run II, during the months before the starting of Run II data taking, I have concentrated also upon the optimization of the selections of soft interaction events, which are used by the ϕ-symmetry intercalibration method. The aim was to have the ϕ-symmetry procedure on Run II data fully commissioned. Indeed, the ϕ-symmetry procedure is fundamental for a short term monitoring to cross check the stability corrections derived from data.

Regarding the search for di-Higgs production, the resonant and non-resonant analyses are presented. A detailed description is given starting from the background main components, going through the particles reconstruction, identification and selection. In particular, the application of kinematic corrections, such as the jet p_T-regression or the kinematic fit, and the optimization of the selections have permitted to reach a good sensitivity. This allowed the exclusion of a relevant fraction of the parameter space of theoretical models, using 19.7 fb^{-1} of data with a

centre-of-mass energy of 8 TeV. The resonant analysis excludes the presence of a radion with masses below 975 GeV for the radion scale $\Lambda_R = 1$ TeV and the RS1 KK graviton with masses between 320 and 400 GeV with $k/M_P = 0.2$. The results are consistent with the expected sensitivity for the hypothesis with no signal.

Moreover, the analysis of the SM and BSM non-resonant double Higgs production is presented and the final results for different signal hypotheses are given. For the SM-like search no significant excess is observed over the background. The observed (expected) upper limit on the SM-like HH $\rightarrow \gamma\gamma b\bar{b}$ production is 1.85 (1.56) fb and it becomes 0.71 (0.60) pb, assuming the SM Higgs branching fraction. Using the SM signal strength modifier, an observed (expected) limit of $\mu_{\mathrm{HH}} = 74$ (62^{+37}_{-22}) is obtained. With this specific analysis, a better sensitivity is achieved with respect to the ATLAS result, which is an observed (expected) limit of 2.2 (1.0) pb, i.e. about 100 times the SM cross section. About the BSM hypotheses, some scenarios with anomalous Higgs boson self-coupling (λ), top-Yukawa coupling (y_t) and/or quartic coupling with the top (c_2) are excluded. In particular, the scenarios with $\kappa_\lambda \equiv \lambda/\lambda^{SM}$ values below -17.5 and above 23 are excluded.

The sensitivity to $\gamma\gamma b\bar{b}$ searches can be strongly enhanced at the higher centre-of-mass energy of the LHC Run II, 14 TeV. Despite the high photon resolution, the sensitivity of the search for high mass candidates is limited by the jet merging. The use of specific jets reconstruction algorithms that exploit merged topology can improve the sensitivity. These algorithms allow the identification and reconstruction of sub-jets within the unique jet coming from the two merged jets. On the other hand, the sensitivity of low mass resonant searches can be ameliorated with an optimization of the photon identification. A specific MVA identification method for photons, exploiting the boosted kinematics, can be used in order to improve the reconstruction for low mass di-Higgs searches.

Regarding other $\gamma\gamma b\bar{b}$ topologies, the bbH and ttH channels are fundamental for the understanding of the couplings within the SM. As their production modes will experience an increase in the cross sections from 8 TeV to 14 TeV, i.e. about 4 times bigger for the ttH production and 2.5 times for the bbH production, it is necessary to concentrate also upon these topologies. In particular, bbH and ttH together with the non-resonant di-Higgs production, are fundamental to better shed light on the couplings and to find possible deviations from the SM, which are a clear hint of physics beyond it.

BIBLIOGRAPHY

[1] K.A. Olive, *et al.*, *"Particle Data Group"*, Chin. Phys. C, 38, 090001 (2014)

[2] Franz Mandl, Graham Shaw, *"Quantum Field Theory"*, Second edition,2010, Wiley

[3] Peter W. Higgs, *"Broken Symmetries and the Masses of Gauge Bosons"*, Phys. Rev. Letters. 13 508 (1964)

[4] LHC Higgs Cross Section Working Group Collaboration, *"Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables"*, CERN-2011-002, http://arxiv.org/abs/1101.0593, doi:10.5170/CERN-2011-002

[5] CMS Collaboration, *"CMS, the Compact Muon Solenoid: technical proposal"*, CERN-LHCC-94-38, https://cds.cern.ch/record/290969

[6] ATLAS Collaboration, *"ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN"*, CERN-LHCC-94-43, https://cds.cern.ch/record/290968

[7] CMS Collaboration, *"Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV"*, Eur. Phys. J. C 75 (2015) 212, http://arxiv.org/abs/1412.8662, doi:10.1140/epjc/s10052-015-3351-7

[8] ATLAS Collaboration and the CMS Collaboration, *"Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV"*, ATLAS-CONF-2015-044, CMS-PAS-HIG-15-002, https://cds.cern.ch/record/2053103

[9] ATLAS Collaboration, *"Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment"*, CERN-PH-EP-2015-125, http://arxiv.org/abs/1507.04548

[10] ATLAS Collaboration and the CMS Collaboration, *"Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments"*, Phys. Rev. Lett. 114, 191803 (2015), http://arxiv.org/abs/1503.07589, doi:10.1103/PhysRevLett.114.191803

[11] CMS Collaboration, *"Constraints on the Higgs boson width from off-shell production and*

decay to Z-boson pairs", Phys. Lett. B 736 (2014) 64, http://arxiv.org/abs/1405.3455, doi:10.1016/j.physletb.2014.06.077

[12] CMS Collaboration, *"Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV"*, Phys. Rev. D 92, 012004 (2015), http://arxiv.org/abs/1411.3441, doi:10.1103/PhysRevD.92.012004

[13] S. Alekhin, A. Djouadi, S. Moch, *"The top quark and Higgs boson masses and the stability of the electroweak vacuum"*, Phys. Lett. B716 (2012) 214, http://arxiv.org/abs/1207.0980, doi:10.1016/j.physletb.2012.08.024

[14] Matts Roos, *"Dark Matter: The evidence from astronomy, astrophysics and cosmology"*, http://arxiv.org/abs/1001.0316

[15] Daniel Baumann, Hiranya V. Peiris, *"Cosmological Inflation: Theory and Observations"*, Adv. Sci. Lett. 2:105-120,2009, http://arxiv.org/abs/0810.3022, doi:10.1166/asl.2009.1019

[16] Stephen P. Martin, *"A Supersymmetry Primer"*, http://arxiv.org/abs/hep-ph/9709356

[17] A. Djouadi *et. al.*, *"Fully covering the MSSM Higgs sector at the LHC"*, KCL-PH-TH 15-08, LPT-Orsay 15-15, http://arxiv.org/abs/1502.05653

[18] Lisa Randall, Raman Sundrum, *"A Large mass hierarchy from a small extra dimension"*, Phys. Rev. Lett. 83:3370-3373,1999, http://arxiv.org/abs/hep-ph/9905221, doi:10.1103/PhysRevLett.83.3370

[19] Albert Einstein, *"Die Feldgleichungen der Gravitation" (The Field Equations of Gravitation)*, Preussische Akademie der Wissenschaften, Sitzungsberichte, 1915 (part 2), 844-847. https://en.wikisource.org/wiki/Translation:The_Field_Equations_of_Gravitation

[20] Ingemar Bengtsson, *"Anti-deSitter space"*, http://www.fysik.su.se/~ingemar/Kurs.pdf

[21] Walter D. Goldberger, Mark B. Wise, *"Modulus stabilization with bulk fields"*, Phys. Rev. Lett. 83 (1999) 4922-4925, http://arxiv.org/abs/hep-ph/9907447, doi:10.1103/PhysRevLett.83.4922

[22] F. del Aguila, M. Perez-Victoria, Jose Santiago, *"Bulk fields with general brane kinetic terms"*, JHEP 0302 (2003) 051, http://arxiv.org/abs/hep-th/0302023, doi:10.1088/1126-6708/2003/02/051

[23] Tony Gherghetta, *"TASI Lectures on a Holographic View of Beyond the Standard Model Physics"*, http://arxiv.org/abs/1008.2570

[24] Gian F. Giudice, Riccardo Rattazzi, James D. Wells, *"Quantum gravity and extra dimensions at high-energy colliders"*, Nucl. Phys. B544 (1999) 3-38, http://arxiv.org/abs/hep-ph/9811291, doi:10.1016/S0550-3213(99)00044-9

[25] Csaba Csaki, Michael L. Graesser, Graham D. Kribs, *"Radion dynamics and electroweak physics"*, Phys. Rev. D63:065002, 2001, http://arxiv.org/abs/hep-th/0008151, doi:10.1103/PhysRevD.63.065002

[26] Csaba Csaki, Jay Hubisz, Seung J. Lee, *"Radion Phenomenology in Realistic Warped Space Models"*, Phys.Rev.D76:125015,2007, http://arxiv.org/abs/0705.3844, doi:10.1103/PhysRevD.76.125015

[27] Alexandra Oliveira, *"New Physics from Warped compact Extra Dimensions: from model building to colliders signals"*, Ph.D. thesis, IFT-T.004/14

[28] Alexandra Oliveira, *"Gravity particles from Warped Extra Dimensions, a review. Part I - KK Graviton"*, High Energy Physics - Phenomenology, http://arxiv.org/abs/1404.0102

[29] Tuomas Hapola and Oleg Antipin, http://cp3-origins.dk/research/units/ed-tools

[30] D. Dominici, B. Grzadkowski, J.F. Gunion, M. Toharia, *"The Scalar sector of the Randall-Sundrum model"*, Nucl.Phys. B671 (2003) 243-292, http://arxiv.org/abs/hep-ph/0206192, doi:10.1016/j.nuclphysb.2003.08.020

[31] Gian F. Giudice, Riccardo Rattazzi, James D. Wells, *"Graviscalars from higher dimensional metrics and curvature Higgs mixing"*, Nucl. Phys. B595:250-276, 2001, http://arxiv.org/abs/hep-ph/0002178, doi:10.1016/S0550-3213(00)00686-6

[32] Vernon Barger, Muneyuki Ishida, *"Randall-Sundrum Reality at the LHC"*, http://arxiv.org/abs/1110.6452, doi:10.1016/j.physletb.2012.01.073

[33] CMS Collaboration, *"Search for resonances and quantum black holes using dijet mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV"*, Phys. Rev. D 91, 052009 (2015), http://arxiv.org/abs/1501.04198, doi:10.1103/PhysRevD.91.052009

[34] CMS Collaboration, *"Search for diphoton resonances in the mass range from 150 to 850 GeV in pp collisions at $\sqrt{s} = 8$ TeV"*, CMS-HIG-14-006, CERN-PH-EP-2015-113, http://arxiv.org/abs/1506.02301 doi:10.1016/j.physletb.2015.09.062

[35] CMS Collaboration, *"Search for physics beyond the standard model in dilepton mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV"*, JHEP 04 (2015) 025, http://arxiv.org/abs/1412.6302, doi:10.1007/JHEP04(2015)025

[36] ATLAS Collaboration, *"Search for high-mass dilepton resonances in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector"*, Phys. Rev. D. 90, 052005 (2014), http://arxiv.org/abs/1405.4123, doi:10.1103/PhysRevD.90.052005

[37] CMS Collaboration, *"Search for massive resonances in dijet systems containing jets tagged as W or Z boson decays in pp collisions at $\sqrt{s} = 8$ TeV"*, JHEP 08 (2014) 173, http://arxiv.org/abs/1405.1994, doi:10.1007/JHEP08(2014)173

[38] ATLAS Collaboration, *"Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector"*, CERN-PH-EP-2015-115, http://arxiv.org/abs/1506.00962

[39] CMS Collaboration, *"Search for massive resonances decaying into pairs of boosted bosons in semi-leptonic final states at $\sqrt{s} = 8$ TeV"*, JHEP 08 (2014) 174, http://arxiv.org/abs/1405.3447, doi:10.1007/JHEP08(2014)174

[40] ATLAS Collaboration, *"Search for resonant diboson production in the $\ell\ell q\bar{q}$ final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector"*, Eur. Phys. J. C (2015) 75:69, http://arxiv.org/abs/1409.6190, doi:10.1140/epjc/s10052-015-3261-8

[41] ATLAS Collaboration, *"Search for production of WW/WZ resonances decaying to a lepton, neutrino and jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector"*, Eur. Phys. J. C (2015) 75:209, http://arxiv.org/abs/1503.04677, doi:10.1140/epjc/s10052-015-3593-4

[42] CMS Collaboration, *"Search for resonant pair production of Higgs bosons decaying to two bottom quark-antiquark pairs in proton-proton collisions at 8 TeV"*, Phys. Lett. B 749 (2015) 560, http://arxiv.org/abs/1503.04114, doi:10.1016/j.physletb.2015.08.047

[43] ATLAS Collaboration, *"Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector"*, Eur. Phys. J. C (2015) 75:412, http://arxiv.org/abs/1506.00285, doi:10.1140/epjc/s10052-015-3628-x

[44] ATLAS Collaboration, *"Search For Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State using pp Collision Data at $\sqrt{s} = 8$ TeV from the ATLAS Detector"*, Phys. Rev. Lett. 114, 081802 (2015), http://arxiv.org/abs/1406.5053, doi:10.1103/PhysRevLett.114.081802

[45] CERN, *"LEP design report"*, CERN-LEP-84-01, https://cds.cern.ch/record/102083

[46] CMS Collaboration, *"CMS physics: Technical design report"*, CERN-LHCC-2006-001, https://cds.cern.ch/record/922757

[47] RHIC, http://www.bnl.gov/rhic/

[48] LHCb Collaboration, *"LHCb : Technical Proposal"*, CERN-LHCC-98-004, https://cds.cern.ch/record/622031

[49] ALICE Collaboration, *"ALICE: Technical proposal for a large ion collider experiment at the CERN LHC"*, CERN-LHCC-95-71, https://cds.cern.ch/record/293391

[50] CMS Collaboration, *"The CMS tracker system project: Technical Design Report"*, CERN-LHCC-98-006, https://cds.cern.ch/record/368412

[51] CMS Collaboration, *"The CMS electromagnetic calorimeter project: Technical Design Report"*, CERN-LHCC-97-033, https://cds.cern.ch/record/349375

[52] P. Baillon, *et al.*, *"Performance of the cooling system of the CMS ECAL"*, Proceedings of the 8th conference on Astroparticle, Particle and Space Physics, Detectors and Medical Applications, doi:10.1142/9789812702708_0031

[53] CMS Collaboration, *"The CMS hadron calorimeter project: Technical Design Report"*, CERN-LHCC-97-031, https://cds.cern.ch/record/357153

[54] CMS Collaboration, *"Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS"*, JINST 6 (2011) 11002, http://arxiv.org/abs/1107.4277, doi:10.1088/1748-0221/6/11/P11002

[55] CMS Collaboration, *"The CMS muon project: Technical Design Report"*, CERN-LHCC-97-032, https://cds.cern.ch/record/343814

[56] CMS Collaboration, *"CMS TriDAS project : Technical Design Report, Volume 1: The Trigger Systems"*, CERN-LHCC-2000-038, http://cds.cern.ch/record/706847

[57] CMS Collaboration, *"Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV"*, JINST 10 (2015) P02006 http://arxiv.org/abs/1411.0511, doi:10.1088/1748-0221/10/02/P02006

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

[59] CMS Collaboration, *"Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV"*, JINST 10 (2015) P08010, http://arxiv.org/abs/1502.02702, doi:10.1088/1748-0221/10/08/P08010

[60] P. Adzic, *et al.*, *"Reconstruction of the signal amplitude of the CMS electromagnetic calorime-*

ter", CERN-CMS-NOTE-2006-037, doi:10.1140/epjcd/s2006-02-002-x

[61] Anfrevilleb M., *et al.*, *"Laser monitoring system for the CMS lead tungstate crystal calorimeter"*, CMS-NOTE-2007-028, https://cds.cern.ch/record/1073694

[62] CMS Collaboration, *"Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at sqrts = 7 TeV"*, JINST 8 (2013) P09009, http://arxiv.org/abs/1306.2016, doi:10.1088/1748-0221/8/09/P09009

[63] CMS Collaboration, *"Observation of the diphoton decay of the Higgs boson and measurement of its properties"*, Eur. Phys. J. C 74 (2014) 3076, http://arxiv.org/abs/1407.0558, doi:10.1140/epjc/s10052-014-3076-z

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

[65] S. Agostinelli, *et al.*, *"GEANT4: A Simulation toolkit"*, Nucl.Instrum.Meth. A506 (2003) 250-303, doi:10.1016/S0168-9002(03)01368-8

[66] Fabio Cossutti, *"The CMS electromagnetic calorimeter simulation"*, 2006 IEEE Nuclear Science Symposium Conference Record, doi:10.1109/NSSMIC.2006.354176

[67] CMS Collaboration, *"Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector"*, CMS-PAS-PFT-10-001, http://cds.cern.ch/record/1247373

[68] CMS Collaboration, *"Performance of CMS muon reconstruction in pp collision events at sqrts) = 7 TeV"*, JINST 7 (2012) P10002, http://arxiv.org/abs/1206.4071, doi:10.1088/1748-0221/7/10/P10002

[69] CMS Collaboration, *"Performance of b tagging at sqrts = 8 TeV in multijet, ttbar and boosted topology events"*, CMS-PAS-BTV-13-001, http://cds.cern.ch/record/1581306

[70] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, Tim Stelzer, *"MadGraph 5: Going Beyond"*, JHEP 06 (2011) 128, http://arxiv.org/abs/1106.0522, doi:10.1007/JHEP06(2011)128

[71] Torbjorn Sjostrand, Stephen Mrenna, Peter Skands, *"PYTHIA 6.4 Physics and Manual"*, JHEP 0605:026,2006, http://arxiv.org/abs/hep-ph/0603175, doi:10.1088/1126-6708/2006/05/026

[72] T. Gleisberg, S. Hoeche, F. Krauss, M. Schoenherr, S. Schumann, F. Siegert, J. Winter, *"Event generation with SHERPA 1.1"*, JHEP 0902:007,2009, http://arxiv.org/abs/0811.4622, doi:10.1088/1126-6708/2009/02/007

[73] CMS Collaboration, *"Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s}$ = 7 TeV with the CMS experiment"*, JHEP 10 (2011) 132, http://arxiv.org/abs/1107.4789, doi:10.1007/JHEP10(2011)132

[74] Matteo Cacciari, Gavin P. Salam, Gregory Soyez, *"The anti-k_t jet clustering algorithm"*, JHEP 0804:063,2008, http://arxiv.org/abs/0802.1189, doi:10.1088/1126-6708/2008/04/063

[75] CMS BtagPOG, https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation53XReReco

[76] CMS Collaboration, *"Search for the standard model Higgs boson produced through vector boson*

fusion and decaying to $b\bar{b}$", Phys. Rev. D 92, 032008 (2015), http://arxiv.org/abs/1506.01010, doi:10.1103/PhysRevD.92.032008

[77] CMS Collaboration, *"Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks"*, Phys. Rev. D 89, 012003 (2014), http://arxiv.org/abs/1310.3687, doi:10.1103/PhysRevD.89.012003

[78] CMS Collaboration, *"Search for a standard-model-like Higgs boson with a mass in the range 145 to 1000 GeV at the LHC"*, Eur. Phys. J. C 73 (2013) 2469, http://arxiv.org/abs/1304.0213, doi:10.1140/epjc/s10052-013-2469-8

[79] CMS Collaboration, *"CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update"*, Technical Report CMS-PAS-LUM-13-001, https://cds.cern.ch/record/1598864

[80] CMS Collaboration, *"Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS"*, JINST 6 (2011) 11002, http://arxiv.org/abs/1107.4277, doi:10.1088/1748-0221/6/11/P11002

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

[82] T. Junk, *"Confidence level computation for combining searches with small statistics"*, Nucl.Instrum.Meth. A434 (1999) 435, doi:10.1016/S0168-9002(99)00498-2

[83] Roberto Contino, Margherita Ghezzi, Mauro Moretti, Giuliano Panico, Fulvio Piccinini, Andrea Wulzer, *"Anomalous Couplings in Double Higgs Production"*, http://arxiv.org/abs/1205.5444, doi:10.1007/JHEP08(2012)154

[84] Martino Dall'Osso, Tommaso Dorigo, Carlo A. Gottardo, Alexandra Oliveira, Mia Tosi, Florian Goertz, *"Higgs Pair Production: Choosing Benchmarks With Cluster Analysis"*, http://arxiv.org/abs/1507.02245

[85] CMS Collaboration, *"Search for heavy resonances in the H tagged dijet mass spectrum in pp collisions at 8 TeV"*, CMS-PAS-EXO-12-053, https://cds.cern.ch/record/2043788