SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Dipartimento di / Department of

Fisica G. Occhialini

Dottorato di Ricerca in / PhD program Physics Ciclo / Cycle XXXIII

Measurement of the Higgs trilinear self-coupling from the HH → bbγγ and ttH(γγ) channels with the CMS detector

Cognome / Surname Monti Nome / Name Fabio

Matricola / Registration number 763220

Tutore / Tutor: Prof. Tommaso Tabarelli de Fatis

Cotutore / Co-tutor:
(se presente / if there is one)

Supervisor:
(se presente / if there is one)

Coordinatore / Coordinator: Prof. Marta Calvi

ANNO ACCADEMICO / ACADEMIC YEAR 2019/2020

Summary

This thesis presents a search for the non-resonant production of Higgs boson pairs in the final state with one Higgs boson decaying into a pair of $b\bar{b}$ quarks and the other Higgs boson decaying into a pair of photons $HH \to b\bar{b}\gamma\gamma$. The data analyzed were collected by the CMS detector in proton-proton collisions with a center-of-mass energy of 13 TeV, for a total integrated luminosity of 137 fb^{-1}. In the interpretation of the results, the double Higgs production processes considered are via gluon fusion (ggHH), which is the main production process, and via vector boson fusion (qqHH). The observed(expected) upper limit to the inclusive $HH \to b\bar{b}\gamma\gamma$ cross section is 0.67(0.45) fb corresponding to 7.7(5.2) times the cross section predicted by the Standard Model. Constraints on anomalous values of the Higgs coupling parameters are measured. The parameters considered are the Higgs boson trilinear self-coupling constant λ_{HHH}, the coupling constant of two Higgs bosons with two vector bosons c_{2V}, and the Yukawa coupling of the Higgs boson with the top quark y_t. The constraint on the λ_{HHH} and on the y_t parameter are significantly improved through the simultaneous fit of the cross sections of the HH processes and of the associated production of the Higgs boson with a $t\bar{t}$ quark pair ($t\bar{t}H$), with the Higgs boson decaying in two photons. Such processes are intrinsically correlated because they both depend from the λ_{HHH} and y_t constants.

After the discovery of the Higgs boson, one of the main purposes of the physics program at the LHC has been a thorough characterization of its properties. A discrepancy between the measurements and the corresponding Standard Model (SM) expectations would be a clear sign of new physics. Three among the most sensitive parameters to new physics, accessible at LHC, are the Higgs boson coupling to the top quark y_t, the Higgs boson trilinear self-coupling λ_{HHH}, and the coupling between two vector bosons and two Higgs bosons c_{2V}. The parameter y_t defines the strongest coupling of the Higgs boson with the fermions and it is fundamental, for example, to determine the SM vacuum (meta)stability. One of the processes directly sensitive to y_t is the $t\bar{t}H$ production. The parameter λ_{HHH} arises in the Standard Model from the expansion of the Higgs boson field potential around its vacuum expectation value. Therefore, its direct measurement provides a fundamental test of the SM predictions for the Higgs boson field potential shape. The most suitable process for a direct measurement of λ_{HHH} is the di-Higgs production. Also the processes of production of a single Higgs boson are sensitive to λ_{HHH} due to electroweak corrections at next-to-leading order. The coupling c_{2V} is expected in

the SM as a consequence of the spontaneous symmetry breaking mechanism. The qqHH production is the most sensitive process to the c_{2V} parameter accessible at LHC.

The $b\bar{b}\gamma\gamma$ final state is one of the most sensitive channels to the HH signal thanks to a relatively clean signature of two heavy-flavor high-energy jets (b-jets) and two high-energy photons. Furthermore, the qqHH topology is marked by two boosted jets with high invariant mass. The applied selections target the events with those features. Thanks to the fully reconstructed final state, the HH signal can be identified as a peak in the diphoton invariant mass distribution, and a peak in the invariant mass of the two b-jets, at the value of the Higgs boson mass. The background is dominated by the non-resonant production of two photons associated with jets which can be heavy-flavored or misidentified as such. Another important background source consists in the single Higgs boson production processes, with the Higgs boson decaying into a photon pair. It is dominated by the $t\bar{t}H(\gamma\gamma)$ process that has a final state similar to the HH signal.

The resolution on the photon energy is a key element to achieve the best sensitivity, as it is the largest contribution to the width of the signal peak in the diphoton mass distribution. For this reason, a significant fraction of this work focused on the performance optimization of the CMS electromagnetic calorimeter (ECAL), which provides the energy measurements for photons (and electrons) with $\mathcal{O}(1\%)$ resolution. In particular, I contributed to inter-calibration of the ECAL channels using electrons from the W and Z bosons decays. The calibration work had two purposes: firstly, to correct for the response variations in time and across the calorimeter due to the crystal radiation damage in order to maintain the design performance, and secondly, to improve the reconstruction and calibration in presence of an increasing rate of pileup collisions in LHC.

At the analysis level, exclusive categories targeting the specific HH production modes are defined to maximize the sensitivity to the HH signal. A qqHH-enriched category is populated by the events passing requirements on the two additional boosted jets. A boosted decision tree (BDT) classifier is defined to isolate the qqHH signal from the ggHH signal and from the continuum background. For the ggHH-enriched categories, a BDT classifier is used to reject the continuum background. The BDT-score as well as the four-body mass of the two photons and the two b-jets are used to classify the events in twelve ggHH-enriched categories. The four-body mass is used for the categorization because its distribution significantly varies for anomalous values of the λ_{HHH} coupling, enhancing the sensitivity to this hypothesis. The number and the boundaries of the categories are optimized to maximize the expected significance to the SM HH signal. Furthermore, a deep neural network is trained to separate the HH from the $t\bar{t}H(\gamma\gamma)$ process, because the $t\bar{t}H(\gamma\gamma)$ was found to be the dominant background in the most sensitive ggHH categories. In order to fully exploit the $t\bar{t}H$ process sensitivity to the λ_{HHH} and y_t parameters, additional categories are defined to select the $t\bar{t}H$ events. The events rejected from the HH categories are tested for the $t\bar{t}H$-enriched categories. Those

categories were developed and optimized for the measurement of the $t\bar{t}H(\gamma\gamma)$ process at CMS with a dataset equivalent to an integrated luminosity of 137 fb^{-1}. Four leptonic-$t\bar{t}H$ categories are defined to target the processes where at least one W boson, coming from the (anti)top quark decay, decays leptonically. In addition, four hadronic-$t\bar{t}H$ categories are defined to target the processes where all the W bosons, from the (anti)top quark decay, decay hadronically.
Notwithstanding these optimizations, the dataset currently available can constrain anomalous production of HH, but cannot access the HH process should it be consistent with the SM prediction. The observation of the SM HH process is among the main goals of the physics program of the high-luminosity phase of LHC (HL-LHC). This phase is planned to start in 2027, and to collect data equivalent to an integrated luminosity of at least 3000 fb^{-1} in ten years of operation. The CMS detector subsystems will be upgraded to cope with the increase in the number of simultaneous interactions at each bunch crossing (pileup), and to withstand the expected damage due to the absorbed radiation dose. For the same reason, new subsystems, such as the minimum-ionizing particles timing detector (MTD), will be integrated in the CMS detector. Providing a timing measurement with a resolution of 30-40 ps for each charged particle within acceptance, the MTD detector will allow a four dimensional reconstruction of the tracks and of the associated vertices. The effective pileup will be reduced to a level well-handled by the reconstruction algorithms. Prospects for the searches of primary interest in the HL-LHC physics program are a fundamental input for the design, the development, and the optimization of the CMS upgrade. For this reason, I worked on the prospects for the HH search in the exclusive decay channel $HH \to b\bar{b}\gamma\gamma$ at HL-LHC, within a more extensive work regarding all the main physics searches of interest, included in the CERN Yellow Report. The prospects are based on Monte Carlo simulations of proton-proton collisions for a center-of-mass energy of 14 TeV and optimized for a total integrated luminosity of 3000 fb^{-1}. The estimated impact of the CMS upgrades in terms of reconstruction and identification of the final state topologies is taken into account. The analysis focuses on the main production mode through ggHH, and the work-flow is analogous to the one previously described. The expected significance for the $HH \to b\bar{b}\gamma\gamma$ channel is 1.9σ and the $\kappa_\lambda = \lambda_{HHH}/\lambda^{SM}_{HHH}$ parameter is expected to be constrained in the interval $-0.5 < \kappa_\lambda < 5.9$ at 95% confidence level. A combination of the most sensitive HH decay channels results in an expected significance of 2.7σ and the κ_λ parameter constrained in the interval $-0.18 < \kappa_\lambda < 3.6$ at 95% confidence level. The impact of the MTD detector on the prospects for the HH searches was studied in details for the MTD technical design report. In particular, for the $HH \to b\bar{b}\gamma\gamma$ channel, the MTD detector provides an increase of the photon identification efficiency of about 3-4% as well as an increase of the b-jet identification efficiency of about 4-6%. Thanks to those improvements, the expected sensitivity to the $HH \to b\bar{b}\gamma\gamma$ channel improves of about 12%, and the expected combined sensitivity to HH improves of about 13%, which would require an additional 26% luminosity without MTD.

Contents

Chapter 1

Importance of the HH and $t\bar{t}H$ physics channels

With the increasingly large dataset of proton-proton collisions events delivered at the LHC, rare physics processes become experimentally accessible. Among those processes are the Higgs associated production with a $t\bar{t}$ quark pair ($t\bar{t}H$), and the Higgs boson pair production HH. All the measurements performed so far at the LHC, including tests of the Higgs boson sector, have confirmed the Standard Model (SM) of particle interactions [1] with great precision. An illustration of the results achieved at the LHC is given in Fig. 1.1 and 1.2. This chapter will motivate the interest in the measurement of the HH and the $t\bar{t}H$ processes to characterize the Higgs boson properties and to search for effects Beyond the Standard Model (BSM). The $t\bar{t}H$ and the HH processes are intrinsically correlated because they depend on the same set of Higgs boson couplings, especially the Higgs boson trilinear self-coupling λ_{HHH}, and the Yukawa coupling of the Higgs boson to the top quark y_t. Therefore, a simultaneous fit of the HH and $t\bar{t}H$ production cross sections is required to measure the couplings controlling, at the same time, all the possible interplays. For completeness, a summary of the SM framework is provided in Appendix A.
In this chapter, after an introduction on the Higgs boson phenomenology (Section 1.1), the SM Higgs boson couplings controlling the HH and $t\bar{t}H$ processes will be described in detail in Section 1.2, while Section 1.3 will describe the couplings beyond the Standard Model (BSM) that can provide sizable modifications to the HH and $t\bar{t}H$ cross sections. The topology of the exclusive HH and $t\bar{t}H$ decay channels considered for this work, i.e. $t\bar{t}H(\gamma\gamma)$ and $HH \rightarrow b\bar{b}\gamma\gamma$, will be described in detail in Sec. 1.4. Finally, the status of the art for the $t\bar{t}H(\gamma\gamma)$ measurements and for the HH searches will be given in Section 1.5.

1.1 The Higgs boson phenomenology

A new boson consistent with the properties of the Higgs boson was experimentally observed for the first time by the CMS [2] and the ATLAS [3] experiments in 2012

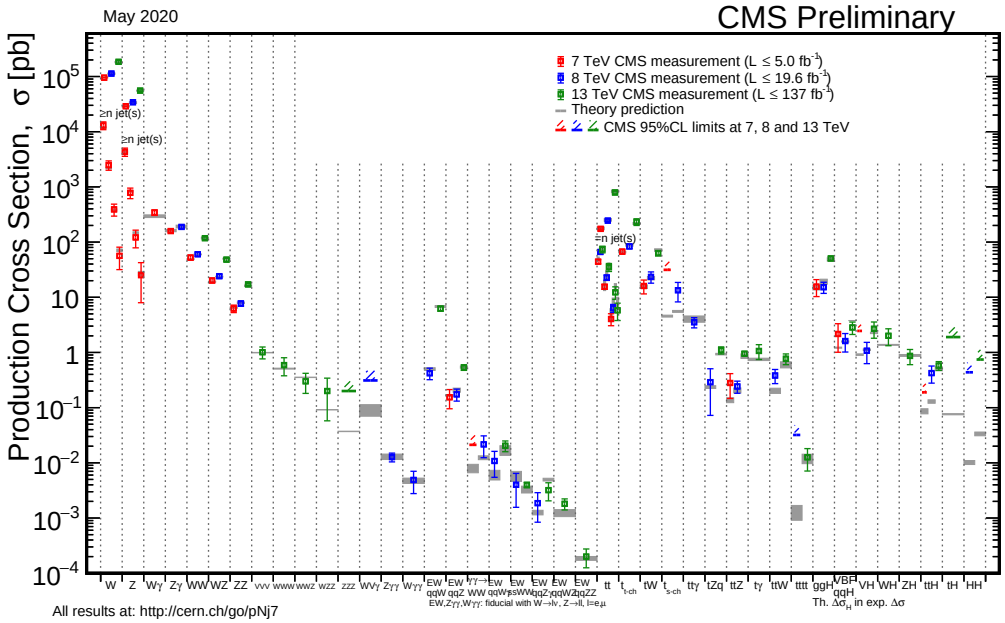


Figure 1.1: Overview of all the measurements and upper boundaries set by the CMS experiment and comparison with the corresponding SM predictions.

with a mass m_{H} of about 125 GeV. After this discovery, one of the main purposes of the CMS and the ATLAS physics program consists in the characterization of this particle. This is realized through precise measurements of its properties such as the mass, the coupling constants with the other particles and the self-coupling constant.

1.1.1 Couplings

The Higgs boson can couple to all the massive particles. The Yukawa coupling with the fermions is proportional to the mass of the fermions, while the coupling with the W and the Z bosons is proportional to the squared mass of the vector boson. Therefore, the Higgs couplings with the heaviest particles are stronger, i.e. the quarks and the lepton of the third generation, and the vector bosons. In addition, the Higgs boson can couple to the gluons and the photons through loop-induced processes dominated by the top quark contribution and for photons also by the W boson contribution. Possible anomalies in the Higgs boson couplings can be experimentally investigated by the mean of the κ-framework, presented in Section 1.3.2. The most up-to-date results achieved by the ATLAS and by the CMS experiments are presented in Fig. 1.3. All the Higgs boson coupling modifiers, described by the κ_i parameters, are found compatible with one, which is the SM value. The CMS result, on the left panel, can be compared to the black points of the ATLAS result in the right one. The only difference in the assumptions is that the CMS analysis considers $\kappa_{\mathrm{t}} > 0$ while the ATLAS analysis considers $\kappa_{\mathrm{W}}, \kappa_{\mathrm{Z}} > 0$.

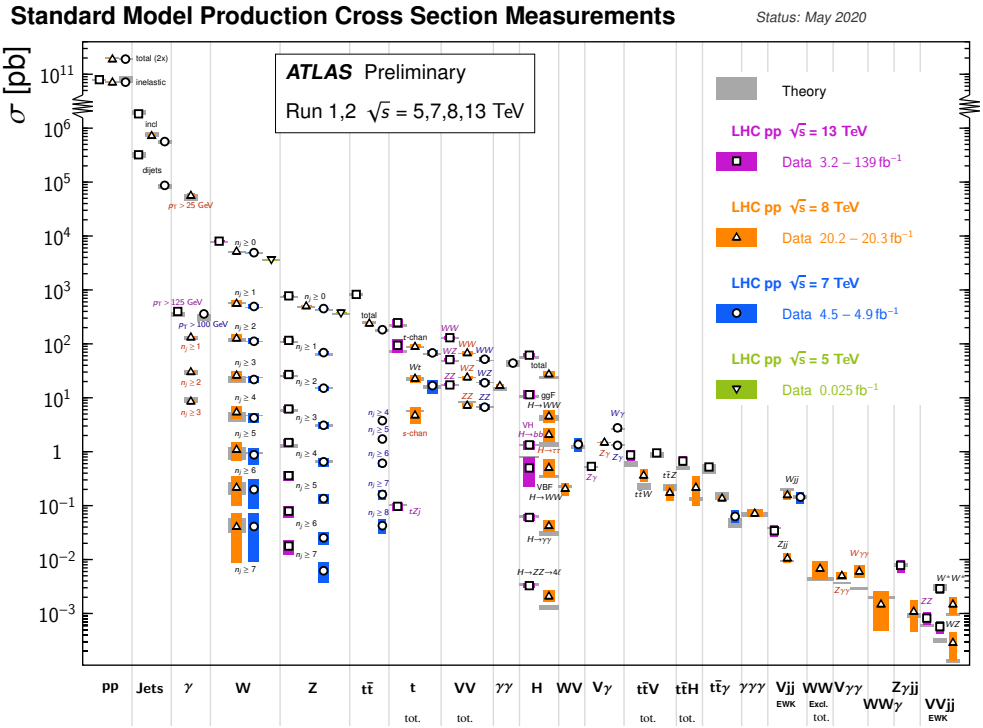


Figure 1.2: Overview of all the measurements and upper boundaries set by the ATLAS experiment and comparison with the corresponding SM predictions.

The two assumptions are equivalent since the measured processes are only sensitive to the relative sign of κ_t and κ_W,κ_Z.

1.1.2 Higgs boson production at the LHC

The Large Hadron Collider machine has delivered proton-proton collisions at a center of mass energy of 13 TeV. With the future upgrades the LHC machines will be able to deliver proton-proton collisions at a center of mass energy of 14 TeV. At these energies, there are several Higgs boson production mechanisms. The Feynamn diagrams of the main production mechanisms are shown in Fig. 1.4, while the cross sections are presented in Table 1.1. In addition, the most recent Higgs production cross section measurements of the CMS and the ATLAS are shown in Fig. 1.5. The main production mechanism is the gluon-gluon fusion (ggH) with a cross section of 48.6 pb at a center-of-mass energy of 13 TeV. The second most probable production mechanism is the vector boson fusion (VBF or qqH). In the qqH mechanism the Higgs boson is produced in association with a pair of jets. The two additional jets have a high di-jet invariant mass as well as a large difference in pseudorapidity. Therefore, they provide a clear signature of this specific production mechanism. The production of a Higgs boson in association with a vector boson (VH) is the

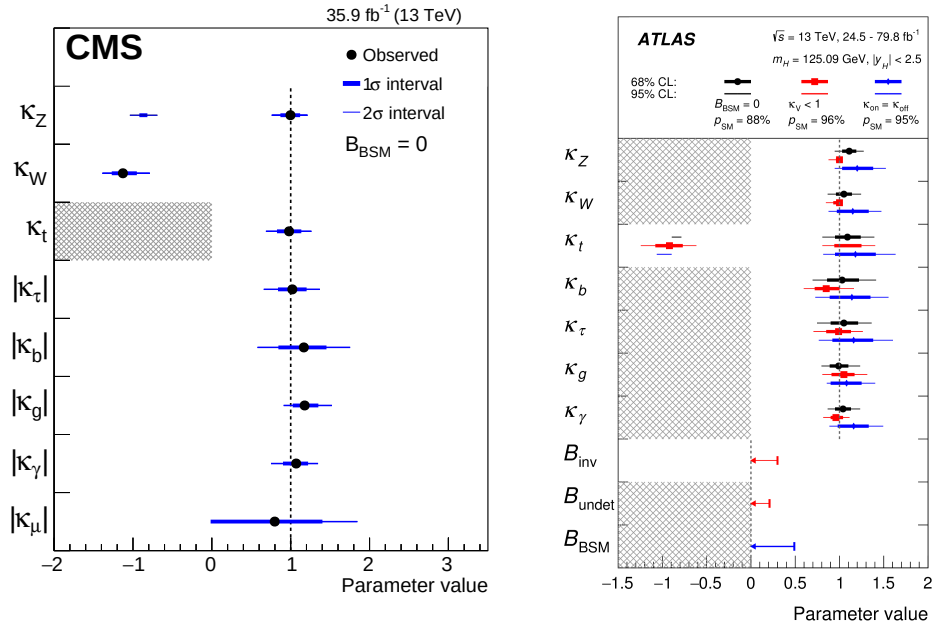


Figure 1.3: *Left:* Higgs boson coupling modifiers as measured by the CMS experiment with 35.9 fb^{-1} of integrated luminosity [4]. *Right:* Higgs boson coupling modifiers as measured by the ATLAS experiment using 24.5 fb^{-1} or 79.8 fb^{-1} of integrated luminosity depending on the channel [5].

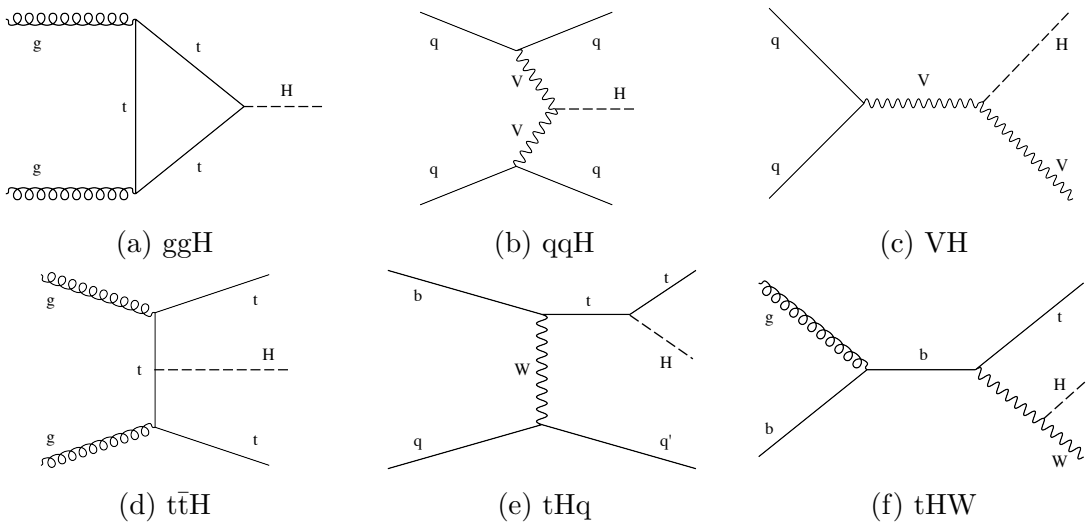


Figure 1.4: Some example of the leading order Feynman diagrams of the main Higgs production processes at the LHC.

third most probable production mechanism. The VH mechanism can be further classified in ZH and WH in case of the associated production with a Z or a W bosons, respectively. The ZH events with the Z boson decaying to e^+ e^- or to μ^+ μ^- can be successfully isolated requiring the additional lepton pair in the final state. Also the WH events with the W boson decaying to e ν_e or to μ ν_μ can be successfully isolated requiring one additional lepton and missing transverse momentum ($\overrightarrow{\not{p}}_T$) in the final state. The Higgs boson associated production with a $t\bar{t}$ quark pair ($t\bar{t}H$) has a cross section of only 0.5 pb. Whereas the $t\bar{t}H$ final state is complex and manifold, these events can be successfully selected with high purity as discussed in Section 1.4.1. The Higgs boson associated production with a $b\bar{b}$ quark pair ($b\bar{b}H$) has leading order diagrams analogous to the $t\bar{t}H$ ones and also the cross section is similar. However, the $b\bar{b}H$ events are difficult to isolate because the two b-jets in the final state lie in a phase-space region where the b-jet identification algorithms have a low discrimination power. An additional production mechanism to be taken into account is the Higgs boson associated production with one top quark and one jet (tHq). Whereas the tHq cross section is only of about $7 \cdot 10^{-2}$ pb, this mechanism should be considered because the experimental overlap between the $t\bar{t}H$ and the tHq events is quite large due to the very similar final state. Furthermore, in case of an anomalous top Yukawa coupling, the tHq contribution can be significantly enhanced. Instead, the Higgs boson associated production with a W boson and one jet (tHW) can be safely neglected due to the very small cross section of only $1.5 \cdot 10^{-2}$ pb.

The Higgs boson pair production is another rare Higgs production mechanism. This process is the object of study of this work, and is extensively discussed in Section 1.2 and 1.3.

Table 1.1: Cross sections of the main single and double Higgs boson production mechanisms from proton-proton collisions at a center-of-mass of 13 TeV and 14 TeV. The single-Higgs, ggHH, and qqHH cross sections correspond to the latest recommendations from Ref. [6], [7], and [8], respectively.

Production mechanism	Cross section (pb) $\sqrt{s} = 13$ TeV	$\sqrt{s} = 14$ TeV
ggH	$4.86^{+5\%}_{-5\%} \cdot 10^{1}$	$5.47^{+5\%}_{-5\%} \cdot 10^{1}$
qqH	$3.78^{+2\%}_{-2\%} \cdot 10^{0}$	$4.28^{+2\%}_{-2\%} \cdot 10^{0}$
WH	$1.37^{+2\%}_{-2\%} \cdot 10^{0}$	$1.51^{+2\%}_{-2\%} \cdot 10^{0}$
ZH	$8.8^{+5\%}_{-5\%} \cdot 10^{-1}$	$9.9^{+5\%}_{-5\%} \cdot 10^{-1}$
$t\bar{t}H$	$5.1^{+9\%}_{-13\%} \cdot 10^{-1}$	$6.0^{+9\%}_{-13\%} \cdot 10^{-1}$
$b\bar{b}H$	$4.9^{+20\%}_{-23\%} \cdot 10^{-1}$	$5.5^{+20\%}_{-24\%} \cdot 10^{-1}$
tHq	$7.4^{+7\%}_{-15\%} \cdot 10^{-2}$	$9.0^{+6\%}_{-15\%} \cdot 10^{-2}$
ggHH	$3.1^{+15\%}_{-14\%} \cdot 10^{-2}$	$3.7^{+15\%}_{-13\%} \cdot 10^{-2}$
qqHH	$1.7^{+2\%}_{-2\%} \cdot 10^{-3}$	$2.0^{+2\%}_{-2\%} \cdot 10^{-3}$

1.1.3 Decay channels of the Higgs boson

Table 1.2: Branching ratio of the main decay channels of the Higgs boson [6].

Decay channel	Branching ratio(%)
$H \to b\bar{b}$	$58.4^{+3.2\%}_{-3.3\%}$
$H \to W^+W^-$	$21.4^{+4.3\%}_{-4.2\%}$
$H \to \tau^+\tau^-$	$6.27^{+5.7\%}_{-5.7\%}$
$H \to ZZ$	$2.62^{+4.3\%}_{-4.1\%}$
$H \to \gamma\gamma$	$0.23^{+5.0\%}_{-4.9\%}$
$H \to Z\gamma$	$0.15^{+9.0\%}_{-8.9\%}$
$H \to \mu^+\mu^-$	$0.02^{+6.0\%}_{-5.9\%}$

The Higgs boson decay channels with their corresponding branching ratio are summarized in Table 1.2, while Fig. 1.6 shows the corresponding Feynman diagrams at the leading order. In addition, Fig. 1.7 summarizes the state of the art of the Higgs decay measurements at the CMS and the ATLAS experiment. The Higgs boson can decay to a $f\bar{f}$ fermion pair. The fermions of the third generation, excluded the ν_τ, are preferred as the Yukawa coupling is proportional to the fermion mass. However, a decay to a $t\bar{t}$ quark pair is kinematically unaccessible, even with off-shell top quarks due to the very large mass difference. The most probable decay channel is the $H \to b\bar{b}$ channel. The decay products consist in two highly energetic b-jets with a mass consistent with m_H. The $H \to \tau^+\tau^-$ channel has also a discrete branching ratio and an identifiable final state thanks to the τ lepton decay products, i.e. one lepton and $\not{p}_T$ in case of a leptonic decay or a very collimated jet in case of a tau

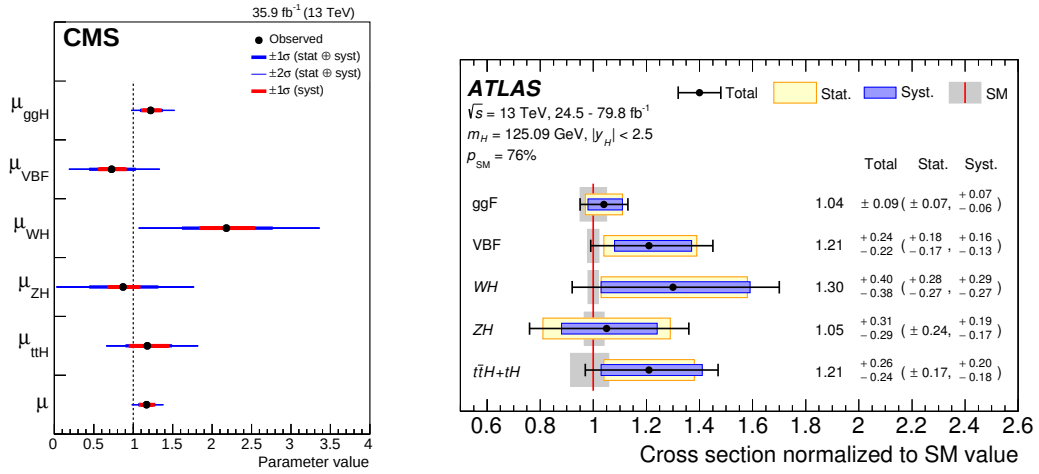


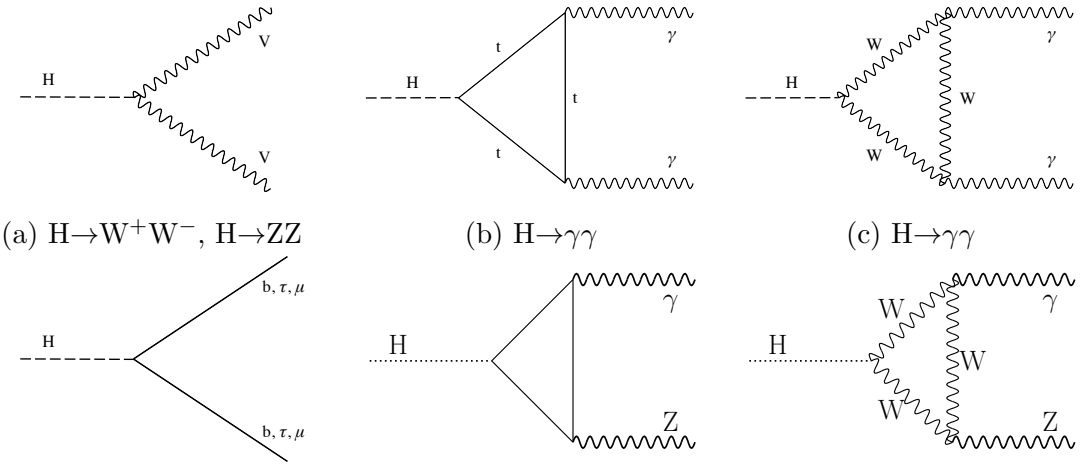
Figure 1.5: *Left:* Higgs boson production cross sections normalized to the SM prediction as measured by the CMS experiment with 35.9 fb^{-1} of integrated luminosity [4]. *Right:* Higgs boson production cross sections normalized to the SM prediction as measured by the ATLAS experiment using 24.5 fb^{-1} or 79.8 fb^{-1} of integrated luminosity (depending on the channel) [5].

hadronic decay. The Higgs boson decays to a fermion pair of the second generation (H→c$\bar{\text{c}}$ and H→$\mu^+\mu^-$) are rare decays which have not been observed yet. The Higgs boson can also decay to a W$^+$ W$^-$ or a Z Z vector boson pair with one off-shell vector boson. The decay products of the vector bosons are used to select the events compatible with this decay channel. In particular, high sensitivity is provided by the H→ZZ(4ℓ) channel thanks to the clean signature of the four leptons and a fully reconstructed final state. The Higgs boson can also decay to massless particles, i.e. a photon or a gluon pair, through loop-induced processes. In particular, the H→$\gamma\gamma$ decay channel is one of the golden channels for the Higgs boson measurements thanks to the clean signature of two highly energetic photons with an invariant mass consistent with m_{H} and the fully reconstructed final state. However, due to the higher perturbative order of the loop processes, the H→$\gamma\gamma$ decay has a relatively low branching ratio. An additional rare decay channel, produced through loop processes, is the H→Zγ channel.

1.2 SM Higgs boson couplings and the HH and tt̄H processes

1.2.1 Higgs boson trilinear self-coupling

The parameter λ_{HHH} is the Higgs boson trilinear self-coupling. It arises in the SM from the expansion of the Higgs field potential around its vacuum expectation



(a) $H\to W^+W^-$, $H\to ZZ$ (b) $H\to\gamma\gamma$ (c) $H\to\gamma\gamma$

(d) $H\to f\bar{f}$ (e) $H\to Z\gamma$ (f) $H\to Z\gamma$

Figure 1.6: Some example of the leading order Feynman diagrams of the main Higgs boson decay channels at the LHC.

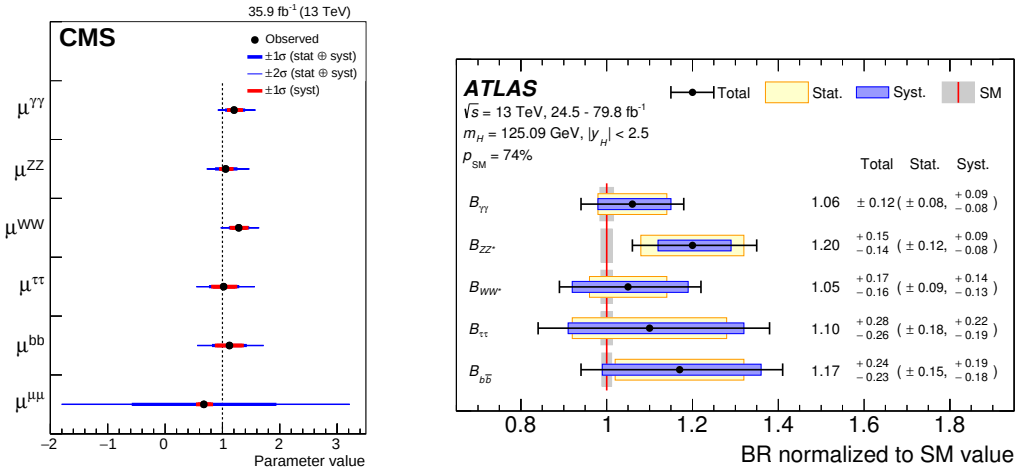


Figure 1.7: *Left:* Higgs boson branching ratios normalized to the SM prediction as measured by the CMS experiment with 35.9 fb^{-1} of integrated luminosity [4]. *Right:* Higgs boson branching ratios normalized to the SM prediction as measured by the ATLAS experiment using 24.5 fb^{-1} or 79.8 fb^{-1} of integrated luminosity (depending on the channel) and fixing the $H\to\mu^+\mu^-$ decay contribution to its SM prediction [5].

value $v/\sqrt{2}$. In the SM theory, the predicted λ_{HHH} value at the leading order is

$$\lambda_{HHH}^{SM} = \frac{m_H^2}{2v^2} = (1.291 \pm 0.003) \cdot 10^{-1} \tag{1.1}$$

where m_H is the Higgs boson mass with a measured value $m_H = 125.10 \pm 0.14$ GeV [9]. The value of v predicted by the SM is 246.22 GeV derived with an extremely good precision from the Fermi coupling constant. The direct measurement of the λ_{HHH} value provides a consistency test of the spontaneous symmetry breaking mechanism. On the other hand, several BSM theories predicts anomalous values of the λ_{HHH} value, such as the composite Higgs models [6, 10] and in general the Higgs-portal models [11]. It is practical to define $\kappa_\lambda = \lambda_{HHH}/\lambda_{HHH}^{SM}$ as a modifier of the SM value.

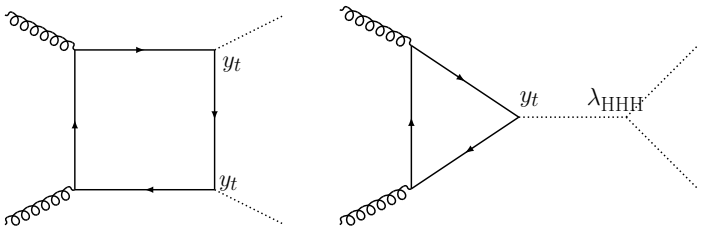


Figure 1.8: Feynman diagrams for the HH production via gluon fusion at the leading order.

The most suitable process for a direct measurement of the λ_{HHH} parameter is the Higgs boson pair production HH. The main HH production mechanism at LHC is the gluon fusion (ggHH) whose Feynman diagrams at the leading order are shown in Fig. 1.8. The tri-Higgs boson vertex appears in the diagrams so that the corresponding cross section depends at the leading order (LO) on the λ_{HHH} parameter. The ggHH cross section σ_{ggHH} predicted by the SM is about 31 fb. The second most frequent HH production mode is the vector boson fusion (qqHH), extensively discussed in Section 1.2.3, which also depends at the LO on the λ_{HHH} parameter. However, the cross section of only about 1.7 fb makes the qqHH process less sensitive to λ_{HHH} than the ggHH one. For both the ggHH and qqHH production mechanisms the SM cross section is too small for an observation with the current datasets. Therefore, the data collected by the LHC experiment so far, corresponding to about 137 fb^{-1} of integrated luminosity, are expected to provide only an upper limit to it. However, anomalous λ_{HHH} values could significantly increase the cross section, as visible in Fig. 1.9, and make the ggHH process observable. As visible in the same figure, the cross sections of other HH production modes, i.e. the HH associated production with a $t\bar{t}$ quark pair $t\bar{t}HH$, and the HH associated production with a vector boson VHH, are also sensitive to the λ_{HHH} parameter. However, the cross section of these processes is much smaller than σ_{ggHH}, hence also the corresponding sensitivity to λ_{HHH} is reduced. In this work, only the qqHH mechanism is considered along with the ggHH mechanism, while the other HH production modes are neglected. In case

of anomalous λ_{HHH} values, the HH invariant mass (m_{HH}) distribution could also be strongly modified, as shown in Fig. 1.9. This feature is properly exploited in the analysis to increase the sensitivity to the λ_{HHH} parameter.

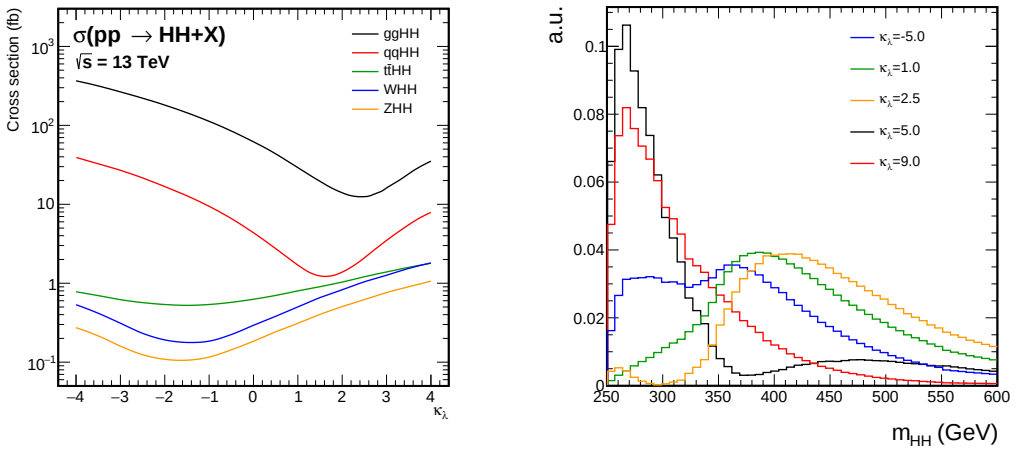


Figure 1.9: *Left:* cross sections of the main HH production modes as a function of κ_λ. *Right:* m_{HH} distributions for the ggHH process for different κ_λ hypotheses. The distributions are all normalized to unity.

The λ_{HHH} parameter can also be probed through the single Higgs boson production processes which are sensitive to the λ_{HHH} parameter via next-to-leading order (NLO) electroweak corrections. The Feynman diagrams inducing such corrections are represented in Fig. 1.10. As shown in the left panel of Fig. 1.11, an anomalous λ_{HHH} value induces a modification of the inclusive cross section. The largest effect appears on the $t\bar{t}H$ process, where the cross section modification can be as large as 30%. The effect is also large for the VH and the tHq production modes. In the $t\bar{t}H$, VH, and tHq mechanisms the Higgs boson can interact with another final-state particle via a Higgs propagator, thus a Sommerfeld enhancement effect happens in the non-relativistic regime [12]. For the same reason, in these mechanisms the λ_{HHH} value has a strong impact on the differential cross sections, e.g. on the Higgs transverse momentum p_T^H distribution [13].
Figure 1.11 compares the single- and double-Higgs production cross sections dependence on the λ_{HHH} parameter. The HH cross sections variations are much larger than the single-Higgs ones because the HH dependence on the λ_{HHH} parameter arises at the leading order, while the effect on the single-Higgs cross sections arises only at NLO electroweak. However, the single-Higgs cross sections are typically much higher than the HH ones. Therefore, they can be sensitive to smaller cross section variations.

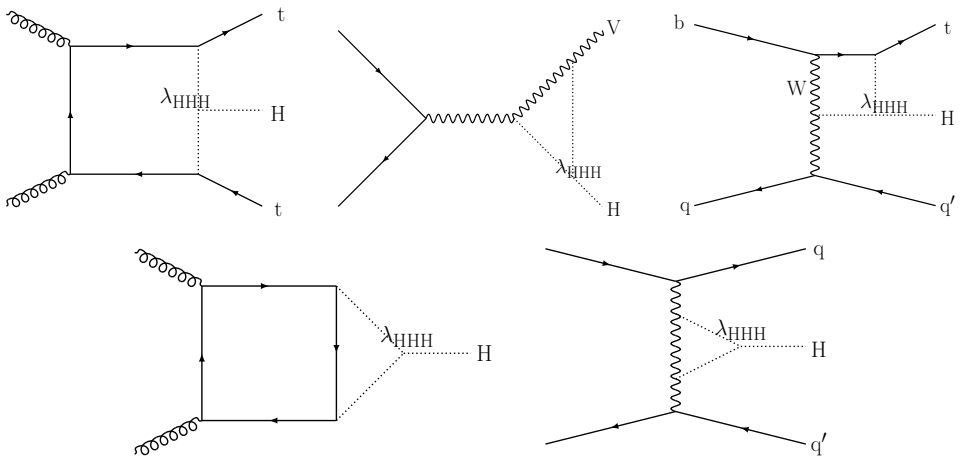


Figure 1.10: Some Feynman diagrams representing the λ_{HHH}-dependent corrections at the electroweak NLO for the main single Higgs boson production processes.

1.2.2 Yukawa coupling of the Higgs boson to the top quark

The Yukawa coupling of the Higgs boson to the top quark is predicted by the Standard Model with a strength:

$$y_{\text{t}}^{\text{SM}} = \frac{\sqrt{2}m_{\text{t}}}{v} = 0.992 \pm 0.002 \tag{1.2}$$

where m_{t} is the top quark mass with a value $m_{\text{t}} = 172.76 \pm 0.30$ [9]. As the Yukawa coupling of the Higgs boson with a fermion is proportional to the fermions mass, y_{t} is the strongest coupling of the Higgs boson with a fermion. The fact that its predicted value is close to the unity suggests that the interaction of the Higgs boson with the top quark might have some special role not disclosed in the SM. The Higgs boson - top quark interaction induces also very large corrections to the SM Higgs boson potential which could produce additional minima in the Higgs field potential. The minima values are y_{t}-dependent and determine the SM vacuum (meta)stability [14]. On the other hand, some BSM theories predict deviations of y_{t} from the SM prediction up to $20 - 30\%$ [15]. The t$\bar{\text{t}}$H process is particularly interesting for the measurement of the y_{t} parameter because its cross section depends at leading order on y_{t}. As visible in the Feynman diagrams of Fig. 1.12, the Higgs boson coupling to the top quark enters directly in the diagrams, without loop-induced processes. On the contrary, the measurement of the Higgs boson production cross section via gluon fusion and of the H$\rightarrow\gamma\gamma$ decay width provide access to the y_{t} parameter only through loop-induced processes. In such processes other BSM phenomena, e.g. new particles in the loop, could modify the final observable, enhancing or compensating the effect of an anomalous y_{t} value. As visible in Fig. 1.8, also the ggHH production depends at leading order on the y_{t} parameter. The variation of the HH and the t$\bar{\text{t}}$H cross sections for anomalous values of y_{t} is shown in Fig. 1.13 where for simplicity

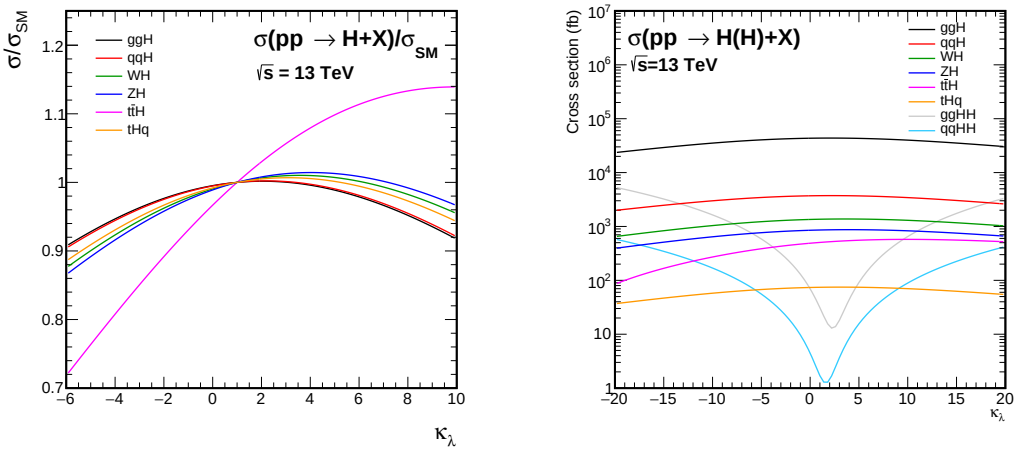


Figure 1.11: *Left:* Cross sections of the main single Higgs boson production modes as a function of κ_λ, normalized to their corresponding SM values. *Right:* Variations of the cross sections of the main single and double Higgs boson production modes as a function of κ_λ.

the κ_t parameter is defined as

$$\kappa_t = y_t / y_t^{SM} \tag{1.3}$$

The ggHH cross section σ_{ggHH} is very sensitive to $|\kappa_t| \gg 1$, because in that case the Feynman diagram with a box loop of top quarks is the dominating matrix element, thus σ_{ggHH} scales as κ_t^4. Instead, $\sigma_{t\bar{t}H}$ scales exactly as κ_t^2 and it is more sensitive to $|\kappa_t| \simeq 1$ because of its larger SM cross section.

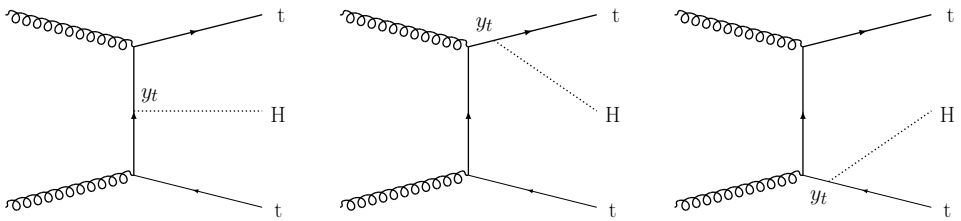


Figure 1.12: Feynman diagrams for the tt̄H production process at the leading order.

1.2.3 Higgs boson couplings to the vector bosons

The spontaneous symmetry breaking mechanism induces the couplings c_Z (c_W) between one Higgs boson and two Z (W) bosons, as well as the couplings c_{2Z} (c_{2W}) between two Higgs bosons and two Z (W) bosons. In the assumption that the BSM phenomena affect in the same way the c_Z and c_W couplings, it is practical to define c_V and c_{2V} as:

$$c_V = c_Z / c_Z^{SM} = c_W / c_W^{SM} \tag{1.4}$$

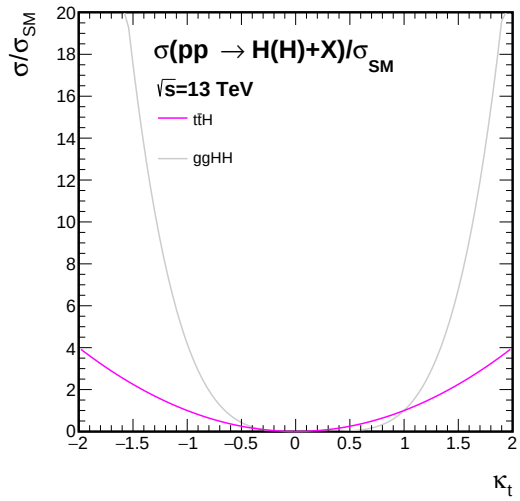


Figure 1.13: Variations of the t$\bar{t}$H and the ggHH cross sections as a function of κ_t.

$$c_{2V} = c_{2Z}/c_{2Z}^{SM} = c_{2W}/c_{2W}^{SM} \tag{1.5}$$

The c_V coupling is experimentally accessible through the measurement of the cross sections of the qqH and VH production modes, and of the decay width of the Higgs boson to WW, ZZ, and $\gamma\gamma$. The measured value of c_V is consistent with the SM prediction with an uncertainty of about 10% [16]. In the SM, considerations of perturbative unitarity of the qqHH cross section [17] require that the c_{2V} and the c_V values are related through the relation:

$$c_{2V} = c_V^2 \tag{1.6}$$

It is fundamental to test the relation in Eq. 1.6 because its violation, hence a violation of the perturbative unitarity, would be a clear signature of a BSM dynamics. In fact a BSM dynamics would be required at a higher energy scale to re-establish the unitarity. Therefore, a direct observation of the c_{2V} coupling, and a measurement of its strength, is a very important step toward the full characterization of the Higgs boson properties. The observation of the c_{2V} coupling is challenging because the most sensitive physics process, accessible at the LHC, is the qqHH whose SM cross section σ_{qqHH}^{SM} is expected to be only 1.73 fb. The Feynman diagrams for the qqHH production at the leading order are represented in Fig. 1.14, where the dependence on c_{2V} and also on λ_{HHH} is visible. Due to the small value of σ_{qqHH}^{SM} predicted by the SM, the collected data are expected to provide only an upper limit to it. However, anomalous values of c_{2V} could significantly increase the qqHH cross section, as shown in Fig. 1.15, making the process observable.

1.3 BSM Higgs boson couplings and the HH and $t\bar{t}H$ processes

1.3.1 Effective field theory approach

Theoretical considerations [18] indicate that the scale of the new physics Λ, e.g. the mass of new particles not predicted by the SM, could be at the TeV scale. From the experimental point of view, the direct searches performed so far using the LHC data exclude the presence of BSM resonances typically up to around 1 TeV. The possibility to probe masses beyond 1 TeV with direct searches is limited at LHC by the available center-of-mass energy and collected data. However, the indirect probe of high energy BSM phenomena at a smaller and accessible energy scale is possible thanks to radiative or perturbative effects. The SM effective field theory (SMEFT [6]) approach allows a quasi model-independent description of a phenomenon at an energy scale $E \ll \Lambda$. The only remnants of the high-energy dynamics are in the low-energy couplings and in the symmetries of the EFT. With the SMEFT approach new operators are added to the SM lagrangian. Such operators are built using the SM particle fields and ensuring the SM Gauge and Lorentz invariance. As a consequence, new effective couplings between the SM particles and modifications of the SM coupling constants could arise.

1.3.2 The κ-framework

The κ-framework introduced in [19], can be considered as a special case of the SMEFT to consistently parametrize the Higgs boson production cross section and decay width in presence of anomalous Higgs coupling values. In the κ-framework, only the EFT operators whose effect is the modification of the SM couplings are considered, while the other EFT operators are assumed to be negligible. In addition, the new operators are assumed to impact only on the coupling strength, and not on the tensor structure of the coupling. As a consequence, in the case of anomalous couplings only the total cross section is modified, while the kinematic distributions are assumed to be unchanged. In addition, the zero-width approximation is adopted

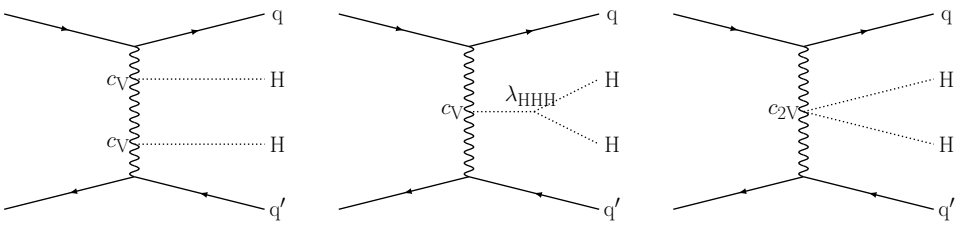


Figure 1.14: Feynman diagrams for the qqHH production process at the leading order.

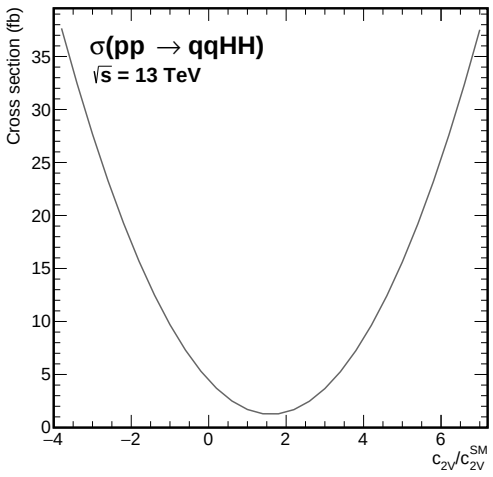


Figure 1.15: Variations of the qqHH cross section as a function of c_{2V}.

for the Higgs boson such that the signal cross section can be written as:

$$(\sigma \cdot BR)(ii \to H \to ff) = \frac{\sigma_{ii} \cdot \Gamma_{ff}}{\Gamma_H} \tag{1.7}$$

Under this assumption, κ parameters are defined to parametrize the modification of the Higgs couplings strengths in such a way that $\kappa_i^2 = \sigma_i/\sigma_i^{\mathrm{SM}}$, or $\kappa_i^2 = \Gamma_i/\Gamma_i^{\mathrm{SM}}$. The κ_{b}, κ_{t}, κ_τ, κ_μ, κ_{W} and κ_{Z} define the coupling modifiers of the Higgs boson to the bottom quark, top quark, τ lepton, μ lepton, W boson, and Z boson, respectively. In addition the H→$\gamma\gamma$ and the ggH vertices can be considered as effective vertices with coupling modifiers κ_γ and κ_g, respectively, or they can be expressed in term of the particles contributing inside the loops. Such loops are dominated by the top quark contribution, and for the H→$\gamma\gamma$ also by the W boson contribution. The κ_γ and κ_g parameters are typically used to probe whether BSM particles contribute to the effective H→$\gamma\gamma$ and ggH vertices.

1.3.3 Sensitivity of HH to the EFT couplings

As visible in Fig. 1.16, the HH production via gluon fusion, is sensitive to five Higgs EFT coupling constants κ_λ, κ_{t}, κ_g, $c_{2\mathrm{g}}$, and c_2 controlling the strength of the corresponding EFT operators. The $c_{2\mathrm{g}}$, and c_2 couplings are effective couplings that can be induced by loops dominated by new heavy BSM particles. In particular, the $c_{2\mathrm{g}}$ defines the coupling of two Higgs boson with two gluons, while the c_2 parameter defines the coupling of two Higgs bosons and two top quarks. The impact of the EFT couplings on the HH observables is double:

- They induce a variation, typically an increase, of the inclusive HH production cross section.

- They significantly modify the differential HH cross section. In particular, the distribution of the di-Higgs invariant mass m_{HH} can dramatically change in case of anomalous couplings.

Therefore, the measured HH cross section together with the m_{HH} differential information can be used to constrain the EFT parameters.

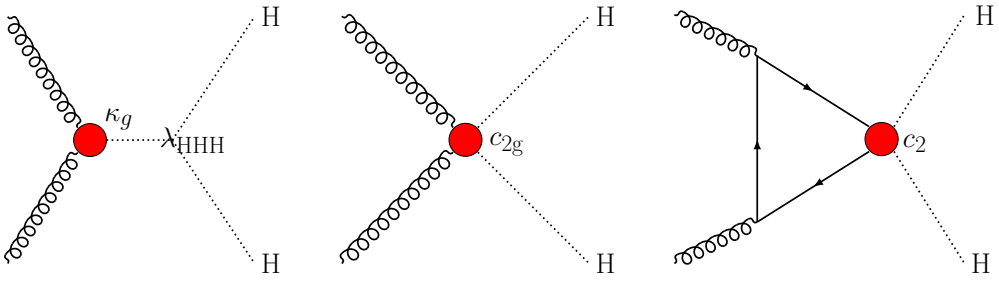


Figure 1.16: Feynman diagrams representing the Higgs boson pair production at the leading order. The black dots represents the SM vertices while the red dots represent contact interactions from contracted loops.

BSM benchmarks to probe the sensitivity to BSM

Because of the small HH cross sections values, the HH processes are not sufficient to simultaneously constrain the κ_λ, κ_{t}, κ_g, $c_{2\mathrm{g}}$, and c_2 couplings. For this reason, twelve points in the five parameters space are selected to be representative of the HH kinematics for all the possible anomalous couplings scenarios [20, 21]. Such points are called BSM benchmarks. Typically in the HH searches, the data compatibility with each specific BSM benchmark is tested. If no BSM evidences are found, upper limits on the benchmark cross sections are extracted. The coupling values for each benchmark is reported in Table 1.3, while the corresponding m_{HH} distributions are visible in Fig. 1.17. A more complete approach discussed in Ref. [11] would consist in including the main single Higgs boson production and decay channels, to reach the sensitivity needed to simultaneously constrain all the five parameters.

1.3.4 Combination of single and double Higgs to measure the EFT couplings

The single and the double Higgs boson processes are intrinsically correlated because they depend on the same coupling parameters. In particular, the ggHH cross section depends on the κ_λ and κ_{t} parameters, while the qqHH cross section depend on the κ_λ, c_{V}, and $c_{2\mathrm{V}}$ parameters. At the same time, the cross section of each single-Higgs production mechanisms depends on the Higgs boson couplings with fermions and/or bosons [19] and also on κ_λ as a result of the NLO electroweak corrections described in Section 1.2.1. In addition, the decay widths of the considered exclusive Higgs

Table 1.3: Coupling values for the twelve defined BSM benchmarks.

Benchmark	κ_λ	κ_t	c_2	κ_g	c_{2g}
0	7.5	1.0	-1.0	0.0	0.0
1	1.0	1.0	0.5	-0.8	0.6
2	1.0	1.0	-1.5	0.0	-0.8
3	-3.5	1.5	-3.0	0.0	0.0
4	1.0	1.0	0.0	0.8	-1.0
5	2.4	1.0	0.0	0.2	-0.2
6	5.0	1.0	0.0	0.2	-0.2
7	15.0	1.0	0.0	-1.0	1.0
8	1.0	1.0	1.0	-0.6	0.6
9	10.0	1.5	-1.0	0.0	0.0
10	2.4	1.0	0.0	1.0	-1.0
11	15.0	1.0	1.0	0.0	0.0
SM	1.0	1.0	0.0	0.0	0.0

decay channels can introduce additional dependences on the couplings. Therefore, a simultaneous fit of the single and double Higgs production cross sections can be performed to measure the Higgs couplings. Three levels of the combination can be identified, in terms of measured couplings:

- Measurement of the κ_λ and the κ_t parameters fixing all the other parameters to their SM expectation.
- Measurement of all the Higgs coupling modifiers including the κ_λ parameter. In this case only the SM Higgs coupling modifiers, as defined in the κ-framework, are considered while the genuine EFT couplings are fixed to zero.
- Measurement of all the EFT Higgs couplings including the κ_λ parameter.

In all the three cases, in order to reach the best sensitivity, specific strategies have to be implemented to select and isolate each single and double Higgs production and decay mechanism. The work presented in this thesis focuses on the simultaneous fit of the κ_λ and the κ_t parameters in prospect of a future extension of the combination to the most general case. For an optimal measurement of the κ_λ and κ_t parameters, it is important to distinguish the $t\bar{t}H$ and HH events from the background, and the $t\bar{t}H$ from the HH events. In fact the HH process, and especially the ggHH process, is the most sensitive process to the κ_λ parameter. On the other hand, $\sigma_{t\bar{t}H}$ depends directly, at the leading order on κ_t. In addition the $t\bar{t}H$ is the single Higgs mechanism receiving the highest κ_λ-dependent correction to its cross section.

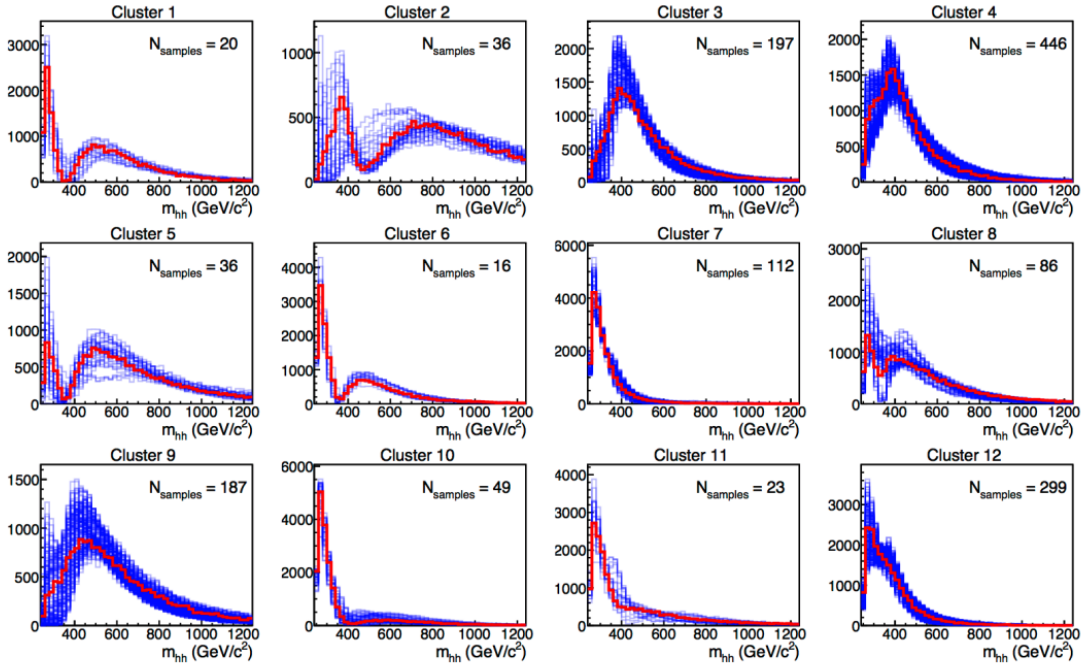


Figure 1.17: Generator-level distributions of di-Higgs boson mass for the clustered benchmarks from [20] are shown. The red distributions correspond to the chosen benchmark sample in each cluster, while the blue ones describe the other members of each cluster.

1.4 The exclusive channels $t\bar{t}H(\gamma\gamma)$ and $HH \rightarrow b\bar{b}\gamma\gamma$

The work presented in this thesis focuses on the combination of the exclusive decay channels $t\bar{t}H(\gamma\gamma)$ and $HH \rightarrow b\bar{b}\gamma\gamma$ to measure the κ_λ and the κ_t parameters. This novel approach set the base for a future desirable combination of all the sensitive decay channels to reach the best sensitivity. The $HH \rightarrow b\bar{b}\gamma\gamma$ channel is chosen for this study because it is the most sensitive HH channel for CMS [22]. The $t\bar{t}H(\gamma\gamma)$ channel is selected because of the experimental overlap with the $HH \rightarrow b\bar{b}\gamma\gamma$. In fact the final state of these two channels is very similar. Therefore, the definition of $t\bar{t}H$-enriched categories along with HH-enriched categories helps to disentangle the contribution of these two signals in the categories.

1.4.1 The $t\bar{t}H(\gamma\gamma)$ channel

The $t\bar{t}H(\gamma\gamma)$ channel has a relatively complex final state consisting of two high energy photons from the Higgs boson decay and the decay products of the $t\bar{t}$ quark pair. The $t\bar{t}$ quark pair decays to a W^+ W^- boson pair and a $b\bar{b}$ quark pair:

$$t\bar{t} \rightarrow W^+W^-b\bar{b} \tag{1.8}$$

The b quarks hadronize and are reconstructed as two highly energetic heavy-flavored jets (b-jets). The W boson can in turn decay hadronically to jets or leptonically to $\ell\ \nu$ (about 1/3 for each lepton flavor) with a branching ratio (BR) of 67% and 33%, respectively. In the W boson hadronic decay, the number of reconstructed jets depends on the specific event topology and on the jet definition. In the $W \to e\nu$ or $W \to \mu\nu$ cases, the final state contains the specific lepton and an unbalanced event transverse momentum ($\not{\vec{p}}_T$) due to the neutrino. The case of a $W \to \tau\nu$ decay can result in a final state with a τ-jet (BR$\simeq$ 67%) identifiable as a collimated jet with few tracks, or in a final state with an electron or a muon (BR$\simeq$ 33%). The possible $t\bar{t}H(\gamma\gamma)$ final states are summarized in Table1.4. The two photons and the two b-jets in the final state provide a good signature of the event. In addition, the fully reconstructed Higgs boson final state allows one to identify the signal as a peak in the diphoton invariant mass distribution over a smoothly falling background. For this reason, the $H\to\gamma\gamma$ channel provides an excellent sensitivity despite the relatively small BR of only about 0.2%. The background for the $t\bar{t}H$ signal is dominated by the non-resonant associated production of photons and jets ($\gamma\gamma$+jets), the non-resonant associated production of a single photon and jets (γ+jets) with one jet mis-identified as a photon, and the associated production of a $t\bar{t}$ quark pair and photons ($t\bar{t}\gamma\gamma$). The sensitivity on the $t\bar{t}H(\gamma\gamma)$ signal is mainly determined by the reconstruction and identification efficiency of photons and b-jets and on the photon energy resolution. The electrons and muons reconstruction and identification efficiency as well as the $|\not{\vec{p}}_T|$ resolution also affects the sensitivity at a subdominant level.

Table 1.4: Branching ratio of all the possible $t\bar{t}$ final states from [9].

Final State	Branching ratio
$t\bar{t} \to b\bar{b}ee\nu\nu$	1.1%
$t\bar{t} \to b\bar{b}\mu\mu\nu\nu$	1.1%
$t\bar{t} \to b\bar{b}\tau\tau\nu\nu$	1.2%
$t\bar{t} \to b\bar{b}e\mu\nu\nu$	2.2%
$t\bar{t} \to b\bar{b}e\tau\nu\nu$	2.4%
$t\bar{t} \to b\bar{b}\mu\tau\nu\nu$	2.4%
$t\bar{t} \to b\bar{b}qqe\nu$	14.4%
$t\bar{t} \to b\bar{b}qq\mu\nu$	14.4%
$t\bar{t} \to b\bar{b}qq\tau_h\nu$	15.2%
$t\bar{t} \to b\bar{b}qqqq$	45.6%

1.4.2 The $HH \to b\bar{b}\gamma\gamma$ channel

The $HH \to b\bar{b}\gamma\gamma$ channel has a fully reconstructed final state consisting in two high energy photons and two high energy b-jets, as illustrated by means of a Monte Carlo (MC) simulated event in Fig. 1.18. The signal can be identified as a peak

in the di-photon invariant mass distribution, and a peak in the di-jet invariant mass distribution, both at the value of the Higgs boson mass m_H. Despite the relatively small HH $\rightarrow b\bar{b}\gamma\gamma$ branching ratio, which is only of about 0.26%, its fully reconstructed final state provides a clean signature over the background. The background is dominated by the non-resonant associated production of photons and jets ($\gamma\gamma$+jets) and the non-resonant associated production of a single photon and jets (γ+jets) with one jet mis-identified as a photon. Other important sources of background for this channel are the $t\bar{t}$ quark pair production, the associated production of a $t\bar{t}$ quark pair and photons ($t\bar{t}\gamma\gamma$), and the associated production of a $t\bar{t}$ quark pair, one photon and jets with one jet misidentified as a photon ($t\bar{t}\gamma$). The events of single Higgs production with the Higgs boson decaying to a pair of photons represent an additional important source of background. In particular, the $t\bar{t}H(\gamma\gamma)$ events have a final state similar to the HH $\rightarrow b\bar{b}\gamma\gamma$ process, with two photons and two b-jets. Nevertheless, the presence of additional objects in the $t\bar{t}H(\gamma\gamma)$ events coming from the W bosons decays provides discrimination power that has been fully exploited in the analysis. The efficiency of the reconstruction and identification of the photons and the b-jets, as well as the photon and jet energy resolution are the parameters dominating the HH $\rightarrow b\bar{b}\gamma\gamma$ sensitivity.

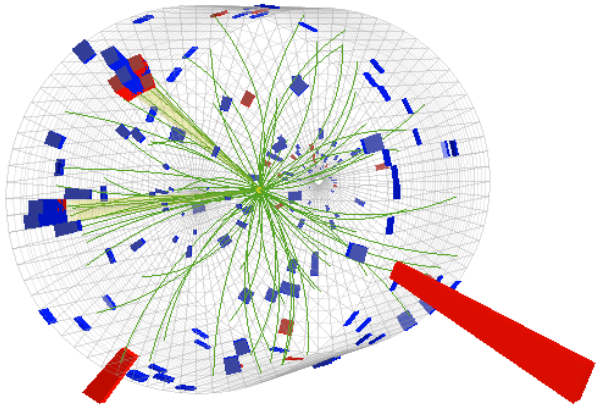


Figure 1.18: Simulated HH $\rightarrow b\bar{b}\gamma\gamma$ event reconstructed at the CMS detector. The red peaks represent the two reconstructed photons, while the yellow semi-transparent cones in top-left direction represent the two reconstructed jets.

1.5 State of the art

1.5.1 $t\bar{t}H(\gamma\gamma)$ observation

The most recent $t\bar{t}H$ measurement results obtained by the CMS and the ATLAS experiments are [23] and [24], respectively. They are obtained using the full Run 2

dataset corresponding to an integrated luminosity of about 137 fb^{-1} (CMS) and 139 fb^{-1} (ATLAS). The $t\bar{t}H$ process is observed in the exclusive $H\to\gamma\gamma$ decay channel with a significance of 6.6(CMS) and 5.2(ATLAS) standard deviations. The measured $t\bar{t}H$ cross sections are:

$$\sigma_{t\bar{t}H} \cdot BR(H\to\gamma\gamma) = 1.56^{+0.34}_{-0.32} \text{ fb by the CMS}$$
$$\sigma_{t\bar{t}H} \cdot BR(H\to\gamma\gamma) = 1.64^{+0.41}_{-0.39} \text{ fb by the ATLAS}$$

which are compatible with the SM prediction of $\sigma_{t\bar{t}H} \cdot BR(H\to\gamma\gamma) = 1.13^{+0.08}_{-0.11}$ fb. In addition the CP structure of the Higgs boson coupling to the top quark is measured, resulting in an exclusion of the pure CP-odd structure at 3.2 and 3.9 standard deviations, by the CMS and ATLAS, respectively, compatible with the SM expectation of a pure CP-even coupling structure. The analysis strategy adopted by the CMS experiment is presented in detail in Section 3.10 since the exact same selections are used in this work to define a set of $t\bar{t}H$-enriched categories.

1.5.2 HH search in the $b\bar{b}\gamma\gamma$ decay channel

In Ref. [22] the HH search in the $b\bar{b}\gamma\gamma$ exclusive decay channel using the data collected by the CMS detector for 35.9 fb^{-1} of integrated luminosity is presented. For this search, only the ggHH process was considered in the interpretation of data, while the other HH production processes were neglected because of their smaller cross sections. The analysis made use of multivariate analysis techniques for the separation of signal-like events from background-like events. The results were found in agreement with the SM prediction. The observed (expected) upper limit at 95% confidence level on the $\sigma(HH \to b\bar{b}\gamma\gamma)$ was found to be 24 (19) times the SM value. In addition, a constraint to the κ_λ parameter was set assuming all the other EFT parameters fixed to their SM value. The observed κ_λ confidence interval, in the assumption of no HH signal, was found to be $-11 < \kappa_\lambda < 17$ at 95% confidence level.

1.5.3 Combination of HH searches

In Ref. [16] the most sensitive HH decay channels considered by the CMS Collaboration, including the $HH \to b\bar{b}\gamma\gamma$ channel, were combined to maximize the sensitivity to the HH signal. An analogous combination of HH searches was performed by the ATLAS Collaboration [25]. The data used for the combinations correspond to an integrated luminosity of about 36 fb^{-1}. For these searches, only the ggHH process was considered in the interpretation of data, while the other HH production processes were neglected. No significant deviations from the SM predictions were observed in both the CMS and ATLAS results. The observed(expected) upper limits at 95% confidence level on the HH production cross section were found to be:

22.2(12.8) times the standard model value by the CMS experiment
6.9(10) times the standard model value by the ATLAS experiment

The large upper limit set by the CMS experiment is due to small excesses of signal, still compatible with the SM predictions, observed in the HH $\rightarrow b\bar{b}b\bar{b}$, HH $\rightarrow b\bar{b}\tau^+\tau^-$, and HH $\rightarrow b\bar{b}\gamma\gamma$ channels, as shown in Fig. 1.19. As visible in the figure, the most sensitive channel to the HH process in the CMS result is the HH $\rightarrow b\bar{b}\gamma\gamma$ channel, while in the ATLAS result the HH $\rightarrow b\bar{b}\tau^+\tau^-$ channel retains the largest sensitivity. In addition, a constraint to the κ_λ parameter was set assuming all the other EFT parameters fixed to their SM value. The observed (expected) κ_λ intervals at 95% confidence level, in the assumption of no HH signal, were found to be:

$$-11.8 < \kappa_\lambda < 18.8 \ (-7.1 < \kappa_\lambda < 13.6) \text{ by the CMS experiment}$$
$$-5 < \kappa_\lambda < 12 \ (-5.8 < \kappa_\lambda < 12) \text{ by the ATLAS experiment}$$

The κ_λ constraint obtained with the combination by the CMS experiment is slightly softer than the one obtained considering only the HH $\rightarrow b\bar{b}\gamma\gamma$ channel because of the small excesses of signal observed in the HH $\rightarrow b\bar{b}b\bar{b}$ and HH $\rightarrow b\bar{b}\tau^+\tau^-$ channels.
To date, these combinations provide the best upper limit to the HH cross section and the most stringent constraint to anomalous κ_λ values from the HH process. The work presented in the next chapters for the HH $\rightarrow b\bar{b}\gamma\gamma$ channel, developed in the context of this thesis, aims to probe even smaller HH cross sections and to provide more stringent constraints to the κ_λ parameter. This is achieved through the analysis of a larger dataset collected by the CMS experiment during the Run 2, equivalent to about 137 fb^{-1} of integrated luminosity, as well as the development of more advanced MVA techniques to improve the signal-over-background ratio. The data are also interpreted in terms of anomalous κ_λ and κ_t couplings values through a simultaneous fit of the single- and double-Higgs production cross sections. Categories targeting the $t\bar{t}H(\gamma\gamma)$ process are defined along with HH-enriched categories to fully exploit the sensitivity of the $t\bar{t}H$ process to the κ_λ and κ_t parameters. Furthermore, the definition of qqHH-enriched categories allows the interpretation of the data in terms of anomalous c_{2V} coupling values.

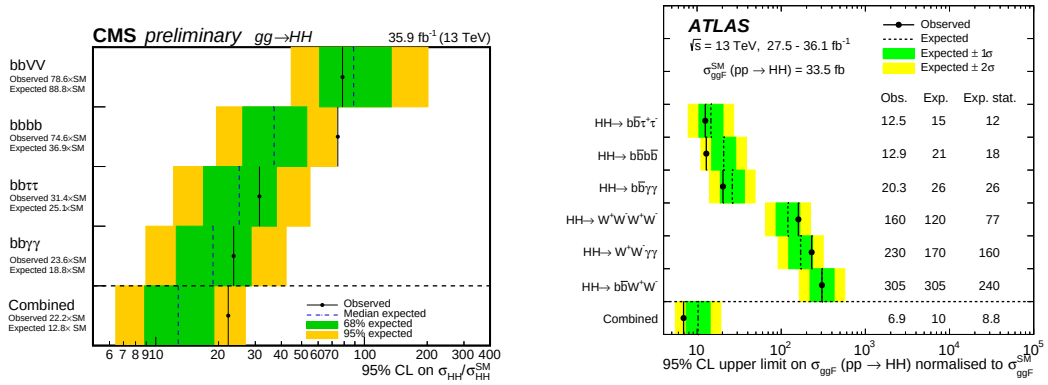


Figure 1.19: The 95% C.L. upper limits on the HH production cross section normalized to the SM value set by the CMS and ATLAS experiments on the left and the right, respectively. The green and yellow bands represent the one and two standard deviation extensions beyond the expected limit, respectively. Figures from Ref. [16] and Ref. [25].

Chapter 2

Experimental Apparatus

This chapter presents the Large Hadron Collider (LHC) machine, which provides the proton-proton collisions to study a wide range of high energy physics phenomena, including the Higgs boson properties. One of the detectors designed to reconstruct the LHC collision events is the Compact Muon Solenoid (CMS). Data from the CMS detector were used for the analysis presented in this work. A description of the CMS detector components and of the reconstruction of the final state observables, often referred to as physics objects, is given. The description will focus especially on the more relevant physics objects for the analysis i.e., photons and b-jets, and on the CMS components used for their reconstruction and identification i.e., the electromagnetic calorimeter (ECAL) and the tracking system. Furthermore, a thorough work, described in this chapter, was carried out to optimize the ECAL performance in order to achieve the best energy resolution for photons (and electrons). This was fundamental to reach the best experimental sensitivity on the $HH \to b\bar{b}\gamma\gamma$ and the $t\bar{t}H(\gamma\gamma)$ searches.

2.1 The Large Hadron Collider

The LHC [26] is the world's largest and most powerful particle accelerator. This machine can provide proton-proton collisions with a center-of-mass energy ($\sqrt{s}$) of up to 14 TeV and heavy ion collisions with a center-of-mass energy of up to 2.76 TeV per nucleon. Among the main goals of the LHC physics program was the search for the Higgs boson that was observed for the first time in 2012 [2, 3]. The LHC is a circular accelerator with a circumference of about 27 Km. It is located close to Geneva across the French and the Swiss border in a tunnel at a depth between 50 and 175 m underground. The tunnel was originally built to host the Large Electron Positron (LEP) collider operating until the year 2000. The LHC started to provide proton-proton collisions at $\sqrt{s} = 7$ TeV in 2010 and had continued its operations until 2018, with multiple planned technical stops. Currently, a long technical stop of the LHC is allowing the planned upgrades of the machine and the detectors. The collisions are planned to re-start in 2021.

2.1.1 The LHC design

The LHC consists in two adjacent pipes hosting the two counter-rotating proton beams. The LHC "ring" is made by eight 2.9 km long arcs and eight 528 m long straight sections. In the arc sections, two opposite magnetic dipole fields with an intensity of about 8 T are applied to bend the proton beams in counter-rotating directions. The dipole magnetic field is provided by electric currents circulating in superconducting coils placed on each side of the pipes. The coils are made of niobium-titanium strands with a copper cladding. In order to work in the superconducting state, a cryogenic Helium system is used to cool down the coils to a temperature of 1.9 K. In addition, a system of magnetic quadrupoles employing the same magnet technology is used to focus the beams to the center of the beam pipes. Along the arcs there is a total of 1232 dipole units with a length of about 15 m, and a total of 858 quadrupole units with lengths ranging from 2.4 to 4.8 m. Higher order multi-pole magnets are also in place to improve the beam focusing and to counteract other interactions that each beam suffers, e.g., proton bunch-bunch interactions and pipe-beam interactions.

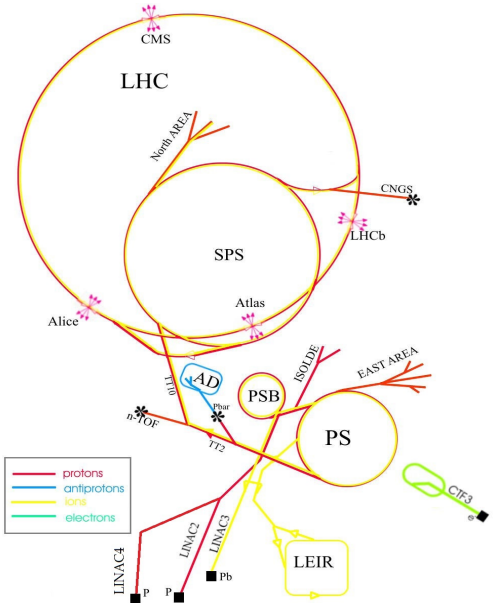


Figure 2.1: Cern acceleration system for different particles (colors). This includes the accelerators system to prepare the protons to be injected in the LHC.

The beams are made to collide in four of the straight sections, corresponding to the positions of four particle detectors: ATLAS, CMS, ALICE, and LHCb. The ATLAS and CMS are general purpose experiments especially designed for the search for the Higgs boson. The ALICE is an experiment designed for the study of the heavy-ions collisions while the LHCb is an experiment designed for the study of the b quark physics with particular interest in the CP violation in the quark interactions. In the proximity of the collision points, additional superconductive quadrupole magnets are used to focus the beams to the interaction points. In one of the straight sections, a 400 MHz superconducting cavity system accelerates the protons up to an energy of 7 TeV. The acceleration system requires the protons to be grouped together in bunches. Each bunch contains about $1.15 \cdot 10^{11}$ protons, and there can be up to 2800 bunches simultaneously circulating along each beam pipe with a nominal spacing of 25 ns. Before the injection to the LHC, the proton bunches are accelerated to 450 GeV through a sequence of accelerating machines depicted in Fig. 2.1.

2.1.2 The LHC performance

One among the main parameters characterizing the performance of a colliding machine is the luminosity $\mathcal{L}$, which is defined by the relation:

$$\mathcal{L}\sigma = \frac{dN}{dt} \tag{2.1}$$

Where σ is the cross section of a scattering process and dN/dt is the observed rate of events of that process. The integrated luminosity L in a defined time interval $[t_0, t_1]$ is defined as:

$$\mathrm{L} = \int_{t_0}^{t_1} \mathcal{L}\ dt \tag{2.2}$$

The integrated luminosity is the reciprocal of a cross section and provides a process-independent measurement of the amount of delivered events. In Fig. 2.2, the delivered luminosity at the CMS experiment is shown across the years of the LHC operations since 2010. The data-taking campaign performed until the 2013-2015 long shutdown, is called Run 1, while the following campaign terminated in 2018 is called Run 2. However, in this thesis the 2015 data-taking is not included in the Run 2 definition. The 2015 dataset amounts to about 4 fb^{-1} and is much smaller than the ones of the following years (137 fb^{-1} in total).
Another important parameter to characterize the LHC collisions, is the number of simultaneous interactions (pileup) at each collision between two proton bunches (bunch crossing). The number of vertices per bunch crossing reconstructed at the CMS detector throughout the LHC operation years is shown in Fig. 2.3. The mean number of the concurrent collisions per beam crossing is about 20-40 in Run 2. The spread of the collision region from where the interaction vertices originate is about 10 cm long along the beam axis, and has a radial extension of about 10 μm on the transverse plane.

2.2 The Compact Muon Solenoid detector

The CMS detector [27] is located in one of the interaction points along the LHC ring. The current CMS physics program targets the Higgs boson characterization as well as precision measurements of the SM parameters, especially of the electroweak sector. At the same time, the searches for BSM phenomena up to the TeV energy scale are also fundamental in the CMS physics program. The CMS detector features a 13 m long superconducting solenoid with an inner radius of 5.9 m. The solenoid is able to provide a uniform magnetic field of 3.8 T within its cylinder. This strong magnetic field provides high precision measurements of the charged particles momentum. As visible in Fig. 2.4, the components of the CMS detector are organized in concentric cylinders (barrel) closed at the two sides by disks (endcaps). The components of the CMS detector from the interaction point are:

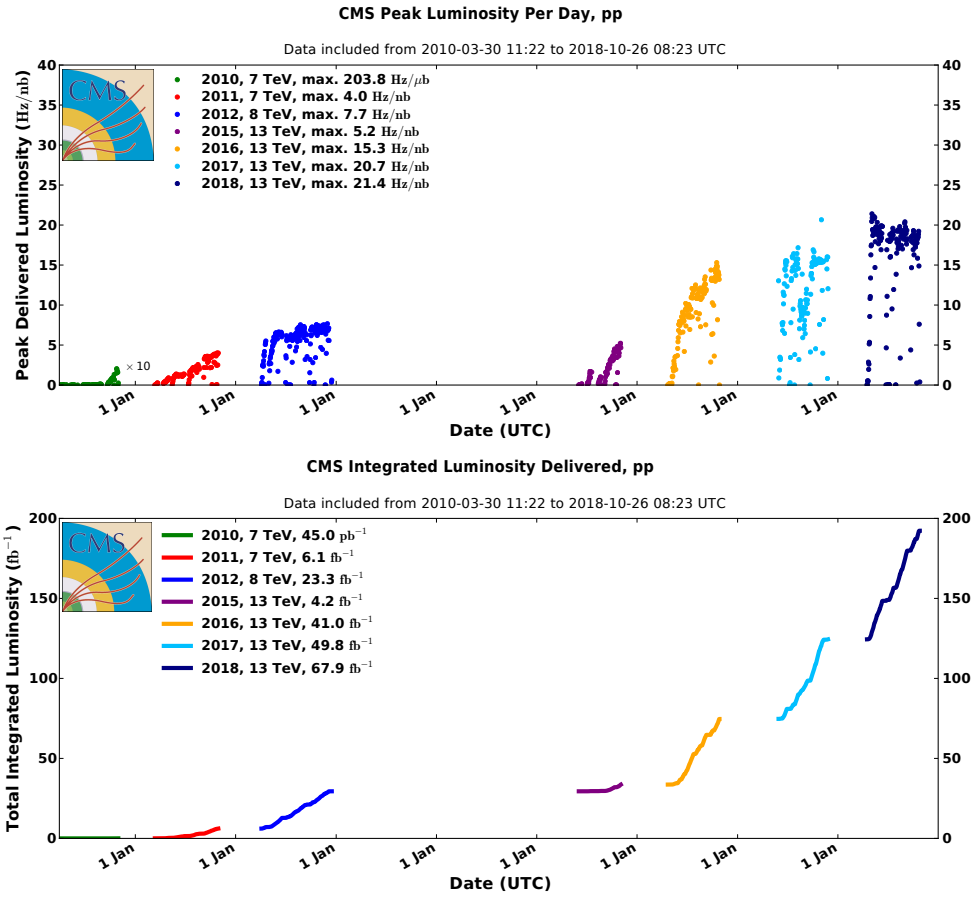


Figure 2.2: *Top:* The peak luminosity delivered by the LHC in the CMS experiment across the years since 2010. *Bottom:* The integrated luminosity delivered by the LHC in the CMS experiment across the years since 2010.

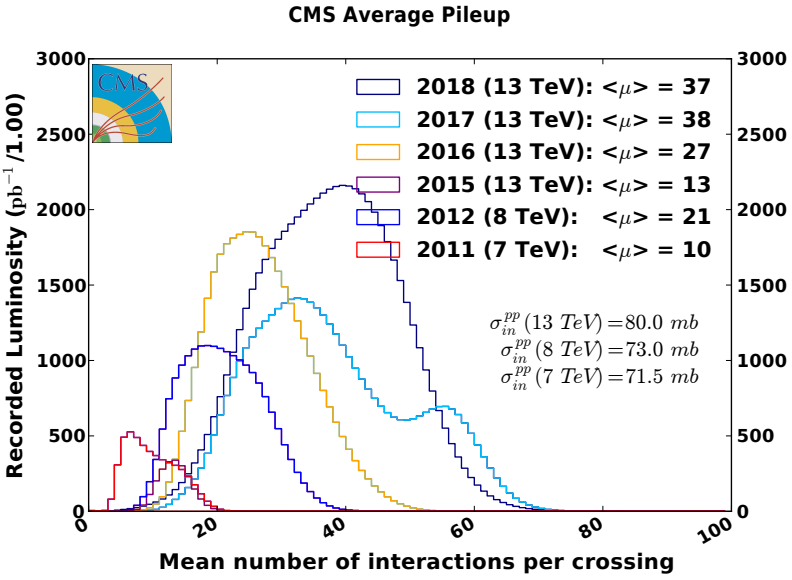


Figure 2.3: Distribution of the average number of interactions per crossing (pileup) for pp collisions delivered to the CMS detector in the different data-taking periods represented with the different colors.

- Inner tracking system: a highly segmented detector made of silicon pixels is used to reconstruct the primary vertices of the interactions and possibly secondary vertices originated by short-lived particle decays. Outside the pixel detector is a silicon strip detector to track the charged particles trajectory within the magnetic field and measure their transverse and longitudinal momentum.

- Electromagnetic calorimeter (ECAL): a homogeneous calorimeter made of scintillating crystals to measure the energy of the electromagnetic showers initiated by photons and electrons.

- Hadronic calorimeter (HCAL): a sampling calorimeter with brass radiator layers alternated with plastic scintillator layers. The purpose of the HCAL is the energy measurement of the hadrons.

- Muon system: system of three different detector technologies to provide an efficient muon reconstruction as well as a precise measurement of their momentum, especially for the high p_T muons. The barrel muon system is outside the solenoid, wedged in the iron yoke for the magnetic flux return.

The next sections focus on the description of the detectors relevant for the HH $\rightarrow$ $b\bar{b}\gamma\gamma$ search which are the inner tracking system (pixel and tracker) and the ECAL. For the remaining CMS components, the reader can find more details on the specific publications [28–31].

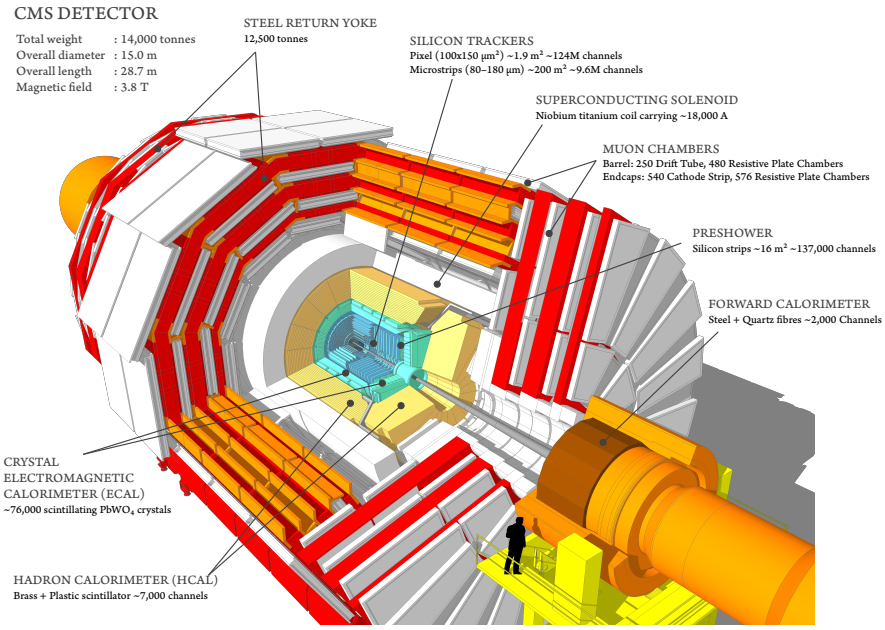


Figure 2.4: Cutaway diagram of CMS detector [32].

2.2.1 The CMS reference system

The CMS detector can be described through a right-handed reference system with the origin in correspondence to the center of the detector. The x-axis is directed toward the center of the LHC ring, the y-axis is directed toward the upward direction, and the z-axis is directed along the beam line. Due to the particle kinematics, it is useful to define a spherical coordinate system:

- The direction along the yz-plane with respect to the y-axis is defined by the polar angle θ. Typically, the pseudorapidity variable η, defined as $\eta = -\log(\tan(\theta/2))$, is used instead of the θ variable. The reason is that the scattering events produced by QCD interactions, which are the most frequent events in a hadron collider, are almost uniformly distributed in η. Furthermore, differences in pseudorapidity for (approximately) massless particles are Lorentz invariant for boosts along the z-axis.

- The direction along the xy-plane with respect to the x-axis is defined by the azimuthal angle ϕ. The xy-plane is called transverse plane. The transverse momentum vector $\overrightarrow{p_T}$ is the projection of the particle momentum on the transverse plane, while the corresponding magnitude is the transverse momentum p_T or transverse energy E_T in case of massless particles. For the momentum conservation, the transverse momentum sum of all the particles originating from a scattering event is zero (up to the transverse momentum of the partons in the initial state).

With this coordinate system, an angular distance between two particles can be defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\Phi^2} \tag{2.3}$$

The ΔR variable for (approximately) massless particles is Lorentz invariant with respect to boosts along the z direction.

2.2.2 Tracker

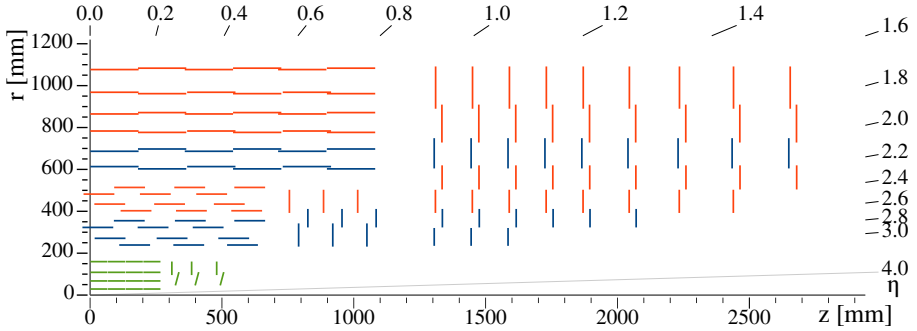


Figure 2.5: Sketch of one quarter of the Phase-1 CMS tracking system in $r-z$ view. The pixel detector is shown in green, the single-sided strip modules in red, while the double-sided strip modules in blue.

At full regime, the LHC has delivered collisions between proton bunches (bunch crossing) at intervals of 25 ns. An average number as high as 50 simultaneous interactions (pileup) is produced at each bunch crossing towards the end of Run 2. Therefore, a highly segmented tracking system is crucial to resolve the high multiplicity of tracks and vertices. The silicon technology was selected for the tracking systems because, beside the economic convenience, it ensures a precise and efficient track reconstruction and the required radiation hardness. The CMS tracking system is made of a pixel detector surrounded by a silicon strip detector as shown in Fig. 2.5. The pixel detector is the closest detector to the interaction point. After the 2016 winter shutdown the pixel detector underwent an upgrade (Phase-1 upgrade) that significantly modified its structure and improved its performance [33]. For the 2016 data-taking the pixel detector consisted in three cylindrical layers of pixels in the barrel at the radius of 4.4 cm, 7.3 cm and 10.2 cm, and two disk layers in each endcap at a distance of 34.5 cm and 46.5 cm from the interaction point. The pixel size is 100×150 μm in the $r-\phi$ and z direction, respectively, for both the legacy and the upgraded pixel detector. In total, the legacy pixel detector had $6.6 \cdot 10^7$ channels. The modifications of the Phase-1 upgrade are shown in Fig. 2.6. The number of pixel layers was increased from three to four in the barrel, reducing also the inner radius to 3.9 cm. Consequently, the total number of pixel channels has increased to $8 \cdot 10^7$ channels. The upgrade has brought a general improvement for the tracking performance and especially in the identification and measurement

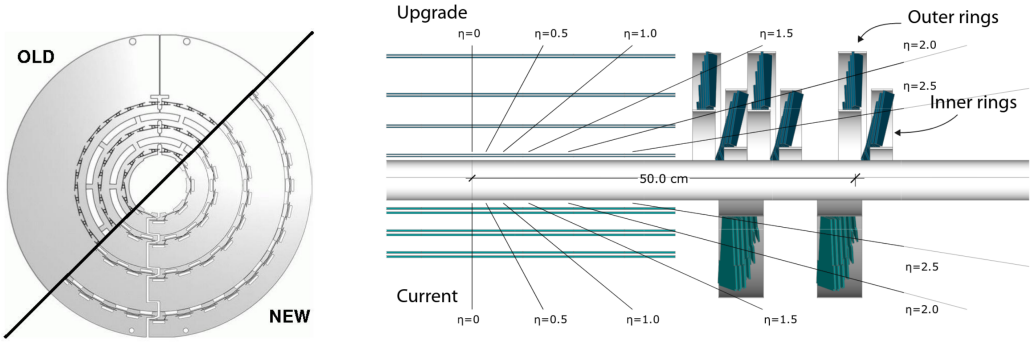


Figure 2.6: *Left:* Comparison of the barrel pixel detector before and after the Phase-I upgrade on the transverse plane. *Right:* Comparison of the barrel pixel detector on the $r-z$ plane before (on the bottom) and after the Phase-I (on the top) upgrade.

of the secondary vertices impacting significantly on the τ lepton and the b-jet identification performance. In addition, the total material in front of the calorimeters (material budget) was significantly reduced, as shown in Fig. 2.7.
Outside the pixed detector the track density is smaller, therefore silicon strip detectors with a segmentation smaller than the pixels are used. In the barrel, the inner four strip layers form the Tracker Inner Barrel (TIB), while the remaining six layers form Tracker Outer Barrel (TOB). The TIB and TOB differ mainly for the strip dimension. The strip thickness and length are smaller in the inner layers to keep the occupancy at about $2-3\%$ per bunch crossing. Analogously, the are two sets of disk layers in the two endcaps forming the tracker inner disk (TID) and the tracker end cap (TEC). Some of the layers, highlighted in Fig. 2.5, have double-sided strip modules that allow a 3D reconstruction of the hits. In total, there are $9.3 \cdot 10^6$ strips providing an angular coverage up to $|\eta| = 2.5$.

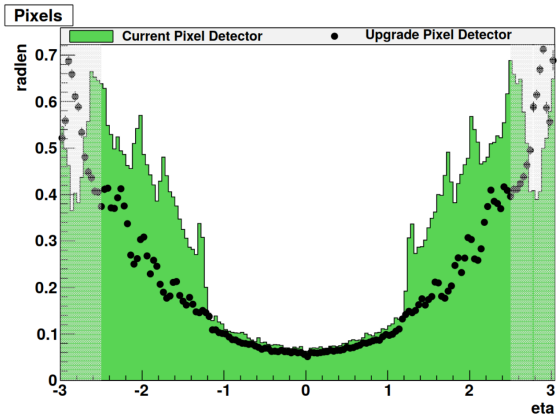


Figure 2.7: Material budget of the pixel detector in units of radiation length. The green area and the black dots correspond to the legacy and the upgraded pixel detector, respectively. The shaded region at $|\eta|> 2.5$ is outside the region for track reconstruction. Figure from Ref. [33].

2.2.3 Electromagnetic calorimeter

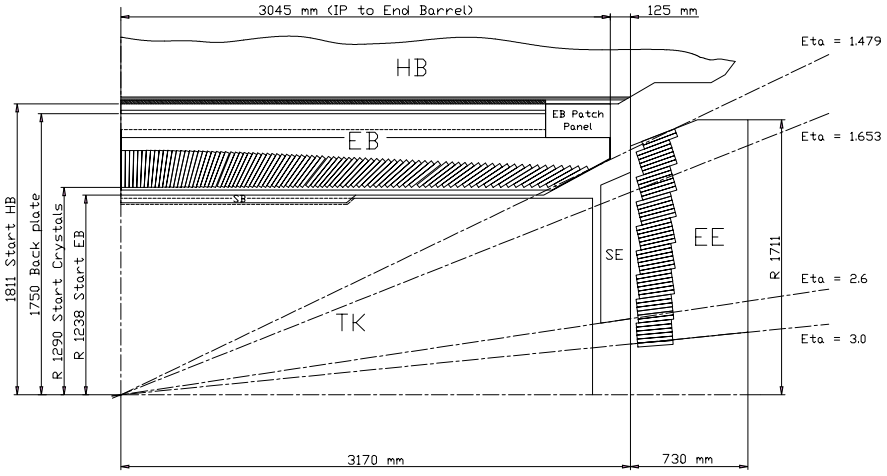


Figure 2.8: Sketch of one quarter of the CMS ECAL in $r - z$ view.

The ECAL is a hermetic homogeneous calorimeter made of 75848 scintillating crystals of $PbWO_4$. The ECAL is designed to provide photons and electrons energy measurements with a resolution on the di-photon and di-electron invariant mass of about 1% for $E \sim 100$ GeV. There are 61200 crystals in the ECAL barrel (EB) and 7324 crystals in each of the two endcaps (EE) in a quasi-projective geometry, as visible in Fig. 2.8. The tilt of the crystal direction of about 3° with respect to the nominal center of the detector prevents photons from the interaction region from leaking through the dead regions due to gaps between the crystals. The $PbWO_4$

crystals feature a scintillating decay time of 15 ns with a fast component of 5 ns. Consequently, about 80% of the scintillation light is emitted within 25 ns which is the LHC bunch crossing period. This crystal has also a short radiation length X_0 of 0.89 cm allowing the containment of the electromagnetic showers in a compact calorimeter. In addition, its Molière radius of 2.19 cm allows the separation of electromagnetic showers with a very small angular difference.
The ECAL barrel covers the pseudorapidity region up to $|\eta|= 1.479$. The front face of the EB crystals has a distance of 1.3 m from the interaction point. The crystal are arranged in modules, each containing 400 or 500 crystals depending on the η position. Groups of four modules are assembled together in a supermodule containing a total of 1700 crystals. The EB crystals are 23 cm long, corresponding to 25.8 X_0, with a cross section of 2.2×2.2 cm^2, corresponding to an angular coverage of 0.0174×0.0174 in $\eta \times \phi$. Along the ϕ direction this corresponds to an angular coverage of 1° per crystal for a total of 360 crystals. The ECAL endcaps covers the pseudorapidity region $1.479 < |\eta| < 3$. The distance along the z axis of the endcap front faces from the interaction point is 3.15 m. The endcap crystals are arranged in a x-y grid, in mechanical units of 5×5 crystals called supercrystals. For each endcap, the supercrystals are grouped in two halves forming the full disk. The endcap crystals are 22.0 cm long (24.7 X_0) and have a front face cross section of 2.9×2.9 cm^2.
The crystals provide a light yield of only about 30 photons/MeV [34], thus the light readout technology was selected to ensure a high gain and quantum efficiency. Two different technologies were selected for EB and EE according to the required radiation hardness and to the magnetic field conditions. In particular, each EB crystal is readout from the rear face by a pair of avalanche photodiodes (APD) [35]. Each APD has 5×5 mm^2 active surface and ensures a gain of about 50× as well as a quantum efficiency of 75%. The APD gain has a significant dependence on the bias voltage, i.e. $\Delta G/\Delta V = 3.1\%/\text{V}$, thus a bias voltage control system is in place to ensure the stability at the level of 10^{-1} V. The EE crystals are readout by vacuum phototriodes (VPT) [36]. Each VPT sensor has an active surface of about 220 mm^2 and provides a gain of about 8× and a quantum efficiency of 22%. The smaller gain and quantum efficiency of the VPT with respect to the APD, is partially compensated by the larger section of the EE crystals.
The temperature stability is an important requirement to achieve the best ECAL performance. In fact, at the operating temperature of 18 °C, the crystal light output varies of -2.1% °C^{-1}. Also the APD gain varies with temperature of -2.4% °C^{-1}, while the VPT gain variation is negligible. Therefore, a cooling system is in place to ensure a temperature stability at the level of 0.05 °C in the barrel and at the level of 0.1 °C in the endcaps. The $PbWO_4$ crystals show a variation of their transparency consequent from the radiation damage. The transparency loss is larger at high η, and worsen with the integrated luminosity, according to the absorbed dose. However, during the LHC shutdown periods, an expected partial recovery of the transparency happens. A laser monitoring system is in place to measure the trans-

parency variation and correct for that. More details about the monitoring system are given in Section 2.4.5.
A preshower detector is placed in front of the ECAL endcaps, covering the pseudorapidity regions $1.653 < |\eta| < 2.6$. The main purpose of the preshower detector is to improve the separation between a photon from the hard scattering and two collimated photons from a $\pi^0 \to \gamma\gamma$ decay. At each endcap side, the preshower detector is made of two alternate layers of lead radiators, to initiate the electromagnetic shower, and silicon strip detectors, to track with high granularity the shower shape. The first layer of lead radiator is 2 X_0 thick, while the second layer is 1 X_0 thick. The strips of the two layers are oriented orthogonally to each other to allow a two dimensional measurement of the electromagnetic shower.

2.2.4 The CMS trigger system

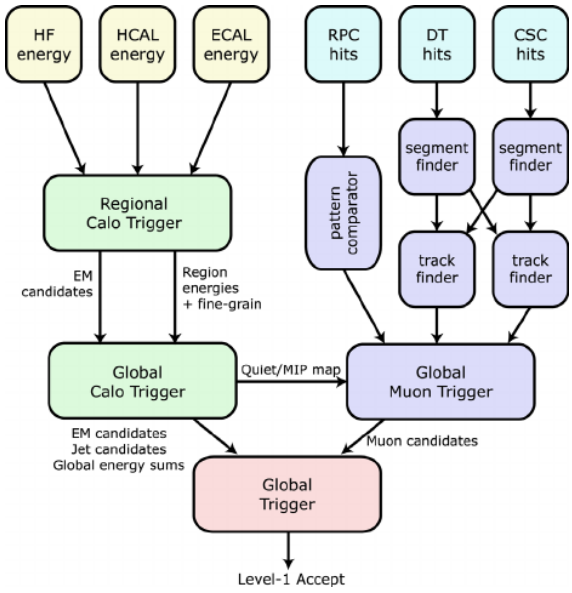


Figure 2.9: Architecture of the CMS L1 trigger.

The LHC delivers proton-proton collisions at a rate of 40 MHz, and within each collision there are 20-50 simultaneous interactions (events). However, the data storage and processing limit the output event rate to an average of 1 kHz. For this reason, a trigger system is in place to select only the interesting events to be saved and discard all the others. The CMS trigger is made of two stages: the first stage, called Level 1 (L1) trigger, reduces the event rate to 100 kHz, while the second stage, called High Level Trigger (HLT), further reduces the event rate to 1 kHz.
The scheme of the L1 trigger system is visible in Fig. 2.9. The input comes from the ECAL, the HCAL, and the muon system. The L1 decision must be made within 3.2 μs because it is limited by the buffer size in the front-end electronics

where the event is temporary stored. The tracking information is not used by the L1 trigger because it requires a longer processing time. For the same reason, the ECAL output is readout in units of 5×5 crystals, called "trigger towers" matching the HCAL granularity. In the first place, local trigger objects called trigger primitives are generated. The trigger primitives can be energy deposits in the ECAL or HCAL, or muon tracks in the muon system. The calorimeter trigger reconstructs the electron/photon, jet, tau candidates of the event, as well as global quantities such as the scalar energy sum of all jets or the $\not{\vec{p}}_T$ of the event from the combination of the ECAL and HCAL trigger primitives. The muon trigger reconstructs the muon candidates of the event combining the track candidates of the three muon detectors. The output of the calorimeter and the muon L1 trigger is sent to a global trigger that takes the final decision. The decision is based on the number of candidates passing specific p_T thresholds and on several other variables, such as invariant masses of pair of objects. In case the L1 trigger rate for a specific seed is too high, only a fixed fraction of the events is accepted (prescaling).
The events passing the L1 trigger are sent to the HLT trigger. The HLT consist in a computer farm that exploits the full readout of the CMS detector. Sequences of reconstruction and filter modules called "paths" are defined to select the interesting events. Each path is built in steps of increasing complexity and precision. This allows a prompt rejection of the events optimizing the CPU usage. Furthermore, the paths are run in parallel but common modules are shared among different paths. In case of a HLT path with an output rate exceeding the limit, the HLT trigger can be prescaled as for the L1 trigger.

2.3 Physics objects reconstruction

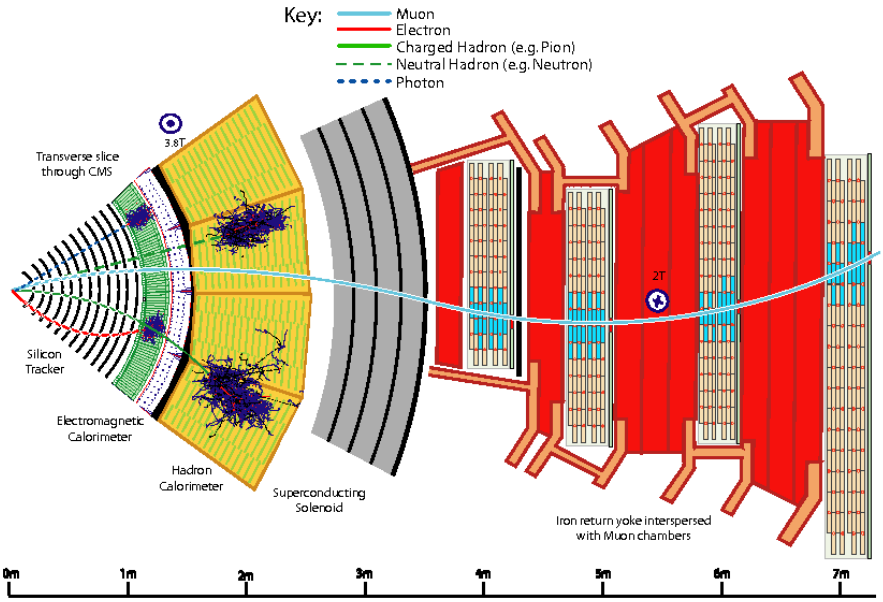


Figure 2.10: Transverse section of the CMS barrel. The detector components are shown as well as the typical interactions of the different particles.

The interactions of the different particles in the CMS components is represented in Fig. 2.10. A muon produces hits in the tracker and in the muon system. An electron produces hits in the tracker and an electromagnetic shower in the ECAL, while a photon produce only an electromagnetic shower. A charged hadron produces hits in the tracker and an energy deposit in the HCAL, while a neutral hadron can produce only an energy deposit in the HCAL. The object reconstruction at CMS relies on the Particle Flow (PF) algorithm [37]. This algorithm exploits the full information from all the CMS subsystems to identify and reconstruct all the particles produced in a collision. The reconstructed particles are then used to build, with high precision, high level objects such as the jets, the τ leptons, and the $\overrightarrow{\not{p}}_T$.

2.3.1 Global event reconstruction

The global event reconstruction is performed by the PF algorithm through the following steps:

1. Tracks and vertices reconstruction from hits in the inner tracker. An iterative tracking algorithm based on the Kalman-filter(KF) is used. At each iteration, new charged particle tracks are reconstructed from compatible sets of hits. The matched hits are then not considered for the further algorithm iterations.

Such a tracking algorithm provides a high track reconstruction efficiency with a small number of hits mis-identified as a track (fake tracks).

2. Track reconstruction in the muon system. The hits in the muon detectors are included in the tracker hits collection and the same iterative algorithm is run. This improve the efficiency of the muon reconstruction.

3. Calorimeter clustering. For each calorimeter and for the preshower, the energy deposits in neighbors cells are gathered in clusters to measure the energy and the position of electrons, photons, charged and neutral hadrons. The clustering algorithm starts from a "seed" which is the cell in the region with the highest energy deposit above a fixed threshold. Then, iteratively all the neighbors energy deposits above a fixed threshold are included in the cluster. Then an iterative algorithm is used to merge clusters compatible with substructures from the same particle. e.g. two close ECAL clusters compatible with two photons from a $\pi^0 \to \gamma\gamma$ decay.

4. Linking of tracks and calorimeter clusters. A track and a calorimeter cluster are linked if their angular distance ΔR is below a given threshold. With the same criterion, an ECAL cluster can be linked to an HCAL and/or a preshower clusters. The output is a collection of PF "blocks".

5. Processing of the PF blocks to build the final state particles. Firstly, muons are reconstructed, then isolated photons and electrons followed by the jet components, i.e., hadrons and non-isolated photons, which correspond to the neutral electromagnetic component of the jet. Finally, the $\not\!\!\vec{p}_T$ object is built. After a PF particle is reconstructed, the corresponding PF blocks are removed from the collection.

2.3.2 Isolated electron and photon reconstruction

Photons and electrons are reconstructed from clusters in the ECAL. An electron is required to have a link to a track while an isolated photon is required to have no link to tracks. Electrons can lose part of their energy for bremsstrahlung effect in the material in front of the ECAL. Bremsstrahlung photons can in turn produce secondary electromagnetic showers in the ECAL. At the same time, a photon can convert to an e^+ e^- pair in the material upstream the ECAL, initiating the electromagnetic shower before entering the ECAL. While electrons are bent in ϕ by the magnetic field, the photons follow a straight path. For this reason an algorithm is used to gather the ECAL clusters within a large window in ϕ and a small window in η, in an object called "supercluster" [38]. The supercluster window at a certain pseudorapidity takes into account the bending of the electrons in the magnetic field, and the ECAL upstream material. The tracker material budget variation with η is shown in Fig. 2.11. Because of the bremsstrahlung energy loss along its path, the

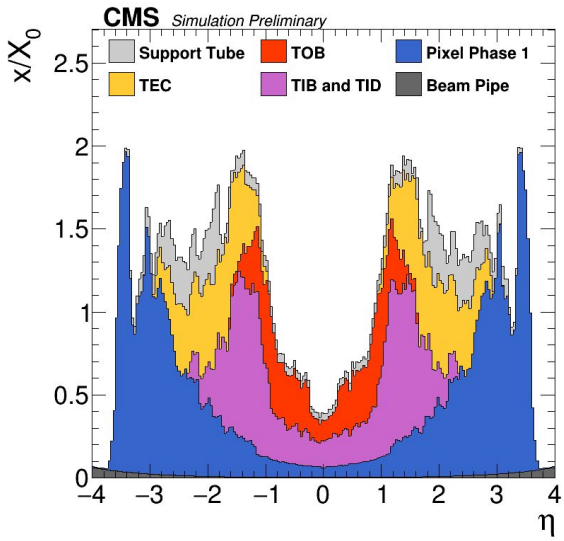


Figure 2.11: Material budget of the CMS tracker considering the upgraded pixel detector in units of radiation length X_0.

electron trajectory is not described through the KF tracking. A specific electron tracking algorithm, based on the Gaussian Sum Filter [39] (GSF) is used to properly describe the electron trajectories. The KF algorithm is used by default. For electrons candidates with low quality KF fit, the GSF algorithm is tried. An electron reconstruction can then be seeded by a GSF track or by an ECAL supercluster. In both cases, an angular matching between an ECAL supercluster and a GSF track is required. This procedure exploits the excellent performance of the tracker with low p_T eletrons down to 2 GeV, and, at the same time, the excellent performance of the ECAL with high p_T electrons. The energy of the electron is obtained from a combination of the GSF track momentum and the ECAL supercluster.
For the isolated photon reconstruction, the ECAL supercluster is required to have no link with tracks. The photon energy is obtained from the ECAL supercluster. A boosted decision tree (BDT) based regression, discussed in Section 2.4.2, is applied to the photon and electron energy measurement to correct for energy containment effects such as the energy leakages out of the supercluster, in the ECAL dead regions, and in the HCAL downstream the ECAL, as well as the additional energy from pileup interactions.
In order to distinguish isolated photons and electrons from the electromagnetic component (still photons and electrons) of a jet, the ratio between the ECAL supercluster energy and the energy deposits in the corresponding HCAL cells is required lower than 0.1. Further identification selections are applied to photons and electrons for the analyses, as discussed in Section 2.4.4 and 3.5.2, respectively.

2.3.3 Jet reconstruction

As discussed in Appendix A.3, the jet is a collimated shower of particles produced by the hadronization of quarks or gluons produced in the hard scattering. On the average, 65% of the jet energy is carried by charged hadrons, 10% by neutral hadrons, and the remaining 25% by non-isolated photons coming from hadrons decays, e.g. $\pi^0 \to \gamma\gamma$. The PF algorithm allows the compensation for the low HCAL energy resolution (about $100\%/\sqrt{E}$) with the excellent resolution of the tracker and the ECAL. The tracker (for $|\eta| < 2.5$) and the ECAL are in fact used to measure the energy fraction from charged hadrons and non-isolated photons, respectively, leaving to the HCAL only the measurement of the energy fraction from neutral hadrons.
Within the pseudorapidity region covered by the tracker, the PF-blocks consisting in ECAL or HCAL clusters linked with tracks are classified as charged hadrons, the ECAL and HCAL clusters without a link to a track are classified as non-isolated photons and neutral hadrons, respectively. Outside the tracker coverage $2.5 < |\eta| < 3$, the PF blocks are classified only as non-isolated photons or neutral hadrons. These PF particles are clustered together using the *anti-kT* algorithm [40] with a distance parameter $R = 0.4$. This algorithm is selected because it provides the required robustness against the collinear and soft particle emission. The momentum of the jet is then computed as the vectorial sum of all the PF particles forming the jet. A jet energy correction (JEC) is applied to match the measured energy scale with the one from the MC simulation. In particular, factorizable corrections for the pileup (PU), and for the mis-modeling of the detector response are derived using MC simulated events. Then, a residual factor is derived comparing real and simulated data to correct for differences among them.
Additional corrections for pileup effects, such as the Charged Hadron Subtraction (CHS), and the PileUp Per Particle Identification (PUPPI) can be used in specific analysis to improve the jet energy resolution. The details of these corrections can be found in Ref. [41].

2.3.4 Missing transverse momentum reconstruction

The missing transverse momentum ($\not\!\!\vec{p}_T$) is built from the negative vector sum of all the PF particles momenta reconstructed in the previous steps. The $\not\!\!\vec{p}_T$ variable should describe the sum of all the particles not interacting in the CMS detector, i.e. neutrinos and possibly weak interacting BSM particles. However, the $\not\!\!\vec{p}_T$ measurement is sensitive to several effects such as the applied clustering thresholds, the noise of the detector components, beside the non-hermeticity of the detector. In order to improve the $\not\!\!\vec{p}_T$ resolution, the jet energy corrections are propagated also to the $\not\!\!\vec{p}_T$.

2.4 Optimization of the ECAL energy resolution

One of the key ingredients to achieve the best sensitivity in many analyses is the photon and electron energy resolution. For the analyses with one Higgs boson decaying to a pair of photons, for example, the peak width in the $m_{\gamma\gamma}$ distribution $\sigma_{m_{\gamma\gamma}}$ is dominated by the energy resolution on the two photons, with a subdominant term dependent on the resolution on the opening angle between the two photons. Since the two photons from a H$\to\gamma\gamma$ event have about the same energy E, the $m_{\gamma\gamma}$ peak width can be estimated in first approximation as:

$$\left(\frac{\sigma_{m_{\gamma\gamma}}}{m_{\gamma\gamma}}\right)^2 \sim \frac{1}{\sqrt{2}}\frac{\sigma_E}{E} \tag{2.4}$$

where σ_E is the photon energy resolution. Therefore, it is crucial to optimize the ECAL energy resolution.
For a calorimeter in general, the energy resolution σ_E/E can be written as the quadratic sum of three contributions:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E[\mathrm{GeV}]}}\right)^2 + \left(\frac{N}{E[\mathrm{GeV}]}\right)^2 + C^2 \tag{2.5}$$

In the ECAL, S is the stochastic term, and includes effects such as the fluctuation of the produced photoelectrons, and shower containment variations. The N parameter describes the electronic noise contribution, including the APD or VPT noise induced by the leakage current. The C parameter corresponds to the constant term and accounts for the effects which are not energy dependent. The main contributions to the constant term consist of instabilities or non-uniformities of the channels responses, and non-uniformities of the light collection. For the ECAL, the three term were measured prior to the installation in CMS in a test beam [42] to be $S = 2.8\%$, $N = 12\%$, and $C = 0.3\%$. The resolution in-situ has additional contributions from incomplete clustering and energy losses in the tracker, as well as incomplete corrections for detector non-uniformities in time and across the channels. For high energy photons, the constant term is expected to dominate the energy resolution. In order to retain the design performance, corresponding to an energy resolution of about 1% for the photons from a H$\to\gamma\gamma$ decay, the constant term C must be kept below 0.5%. Therefore, it is crucial to correct for any channel response instability and non-uniformity across the detector to reduce the constant term contribution.

2.4.1 Energy reconstruction

In a reconstructed supercluster, as described in Section 2.3.2, the energy of the photon or the electron object is estimated as:

$$E_{\mathrm{e}/\gamma} = F_{\mathrm{e}/\gamma}\Big[G\sum_i (H(\eta_i)\cdot C_i \cdot S_i(t)\cdot A_i) + E_{ES})\Big] \tag{2.6}$$

Where the sum runs over all the crystals belonging to the supercluster. The factors contributing to the formula are described as follows:

- A_i is the amplitude (in ADC counts) of the signal measured in the i-th channel. The ECAL uses an amplitude reconstruction algorithm which ensures the required robustness against pileup [43].
- $S_i(t)$ is the time-dependent correction for the crystal transparency change. This factor correlates the response variation of the crystals to electrons with the response to a light source from the laser monitoring system. This relation is approximately described by a universal law presented in Section 2.4.5.
- C_i is the intercalibration constant which equalizes the response of the different ECAL channels at constant η.
- $H(\eta_i)$ equalizes the response of all the crystals at the same pseudorapidity with respect to the MC simulation. The used energy reference is the dielectron invariant mass peak in Z $\rightarrow e^+e^-$ events.
- G provides the scale conversion from ADC counts to GeV. Two different values are used for EB and for EE. It is derived from the comparison of the energy scale in data and in MC simulation using as reference the dielectron invariant mass peak in Z $\rightarrow e^+e^-$ events.
- E_{ES} is the energy measured in the preshower cluster(s), if any associated with the ECAL supercluster.
- $F_{e/\gamma}$ is an energy correction for shower containment and pileup effects. More details can be found in Section 2.4.2.

Within a wide effort to optimize all the factors entering in the ECAL reconstruction, my work focused on the optimization of the ECAL channels response stability and uniformity controlled by the S_i and C_i factors, respectively. The introduction of new corrections was required in Run 2 to address instabilities observed in the laser monitoring system.
An intercalibration procedure is performed to uniform the response of the ECAL channels across the detector. It needs to be repeated periodically because the universal law used to correlate the channels response to electrons with the response to the laser light works well on the short time-scale, but can have long-term deviations (on a timescale of one year). Small response differences between the channels can also arise from long-term variations of the temperature or bias voltage, or subsequent from the on/off CMS cycles. In addition, drifts of the laser monitoring system were observed during Run 2 that have required an additional time-dependent correction prior to the intercalibration since 2016. The intercalibration procedure is presented in Section 2.4.3, while the time dependent-correction is presented in Section 2.4.6.

2.4.2 Supercluster energy correction

The measured supercluster energy is affected by several energy containment effects. The dominant effects are the energy leakages out of the supercluster, in the ECAL dead regions (gaps between crystals, modules and supermodules), and in the material upstream and downstream the ECAL, as well as the additional energy in the supercluster from pileup interactions. In order to properly take into account the correlation between all these effects, a multivariate regression approach based on a boosted decision decision tree (BDT) is used [38, 44]. The BDT regression is trained separately for photons and electrons because they interact differently in the CMS detector. The training is performed using simulated electrons(photons) with a uniform p_T spectrum from 1(5) GeV to 300 GeV, and targets the $E_{\text{true}}/E_{\text{raw}}$ ratio, where E_{true} is the electron(photon) energy from the MC truth, and E_{raw} is the uncorrected supercluster energy. Beside E_{raw}, the variables used for the training are:

- the (η, ϕ) positions of the supercluster and of its seed to provide information especially on the energy leakages in the ECAL upstream materials and in the ECAL dead regions,
- the H/E ratio to provide information on the energy leakages in the HCAL downstream the ECAL,
- the supercluster shape observables as well as the energies, positions, and shapes observables of the clusters that form it to provide information especially on the energy leakages out of the supercluster,
- the number of reconstructed vertices as well as the average energy density of the event ρ to infer the pileup effect.

The output of the energy regression is double: firstly the correction factor for the photon or electron supercluster energy $F_{\text{e}/\gamma}$, and secondly, an estimate of the resulting energy resolution. Figure 2.12 shows, as an illustration, the impact of the energy regression correction on $\text{Z} \rightarrow \text{e}^+\text{e}^-$ MC events. The supercluster correction further improves the excellent energy resolution provided by the clustering algorithm.

2.4.3 ECAL intercalibration

The ECAL crystal intercalibration reduces the channel-to-channel response spread, improving the energy resolution. The intercalibration constants (IC) are computed typically at the end of each year of data taking profiting of the full amount of data. A subsequent reconstruction of the data including the intercalibration constants is performed. Four methods have been developed to intercalibrate the ECAL channels. These methods exploit energy scale references derived from the physics events from collisions:

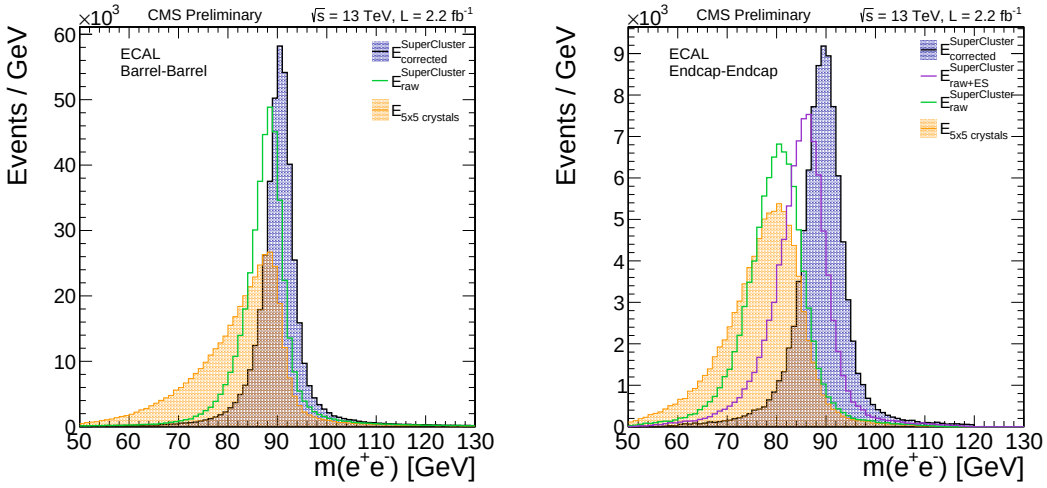


Figure 2.12: *Left:* Dielectron invariant mass distribution for Z $\rightarrow e^+e^-$ events with both electrons in the EB (on the left) or in EE (on the right). The electron energy is estimated with different levels of refinement corresponding to more sophisticated clustering and cluster correction algorithms: summing of the energy of the 5×5 crystals surrounding the crystal seed (in orange), using the energy of the reconstructed supercluster (green line) that for the EE electrons can also include the preshower energy (purple line), and finally applying the MVA-based supercluster energy correction (in blue).

- E/p ratio of electrons and positrons from the W $\rightarrow e\nu$ and Z $\rightarrow e^+e^-$ decays, where E is the ECAL supercluster energy and p is the momentum measured by the tracker which is taken as a reference after a p_T calibration of the tracker. Using a high-purity sample of energetic electrons, the E/p method provide the highest precision in the EB along with the Z $\rightarrow e^+e^-$ method. The precision of this method is limited by the statistical uncertainty, the tracker momentum resolution correlated with the material budget upstream the ECAL. For this reason, the precision of the E/p method is significantly worse in the endcaps than in the barrel.

- Invariant mass peak in Z $\rightarrow e^+e^-$ events. This intercalibration method has been developed for the Run 2 data-taking. In the EE, the Z $\rightarrow e^+e^-$ method is the most precise one and, not using the tracker as a reference, it can be extended up to $|\eta| = 3$. Since the Z boson production cross section is smaller than the W boson production cross section, the impact of the statistical uncertainty on the Z $\rightarrow e^+e^-$ method is higher than on the E/p method, and dominates the intercalibration precision.

- Invariant mass peak in $\pi^0 \rightarrow \gamma\gamma$ events. The peak position is extracted through a fit of the diphoton invariant mass distribution with a proper function to describe the signal and the background. The π^0 method makes use of a very

large dataset, thus the statistical uncertainty is negligible. However, due to the low event energies, the π^0 precision is dominated by the electronic noise and pileup effects as well as a non-negligible background contribution. These limitations make the precision of the π^0 method dramatically worse in the EE, and unusable in the region $|\eta|> 2.5$ not covered by the tracker.

- ϕ-symmetry of the average deposited energy in all the crystals at a fixed pseudorapidity in a large sample of soft interactions. Geometrically this corresponds to a ring of crystals called η-ring. The deviation of the average deposited energy in a crystal with respect to the energy deposited in the full η-ring is used to derive the intercalibration constant. The precision provided by this method is worse than the other three methods because it suffers for the non-uniformity along ϕ of the material upstream the ECAL which can not be corrected. For this reason the ϕ-symmetry method is typically used only for validation, or for checking relative time drifts.

The work for this thesis focused on the E/p method which will be described in detail in Section 2.4.4. The IC obtained with the E/p, Z $\rightarrow e^+e^-$, and π^0 methods are combined together through an average weighted on the relative IC precisions. The results are presented in Section 2.4.8.

2.4.4 Intercalibration with electron E/p

The E/p intercalibration method exploits the tracker momentum measurement that is taken as the energy scale reference. The tracker is calibrated with high precision exploiting several resonances decaying to $\pi^\pm$ or $\mu^\pm$, such as Z $\rightarrow\mu^+\mu^-$. However, the electron momentum is measured with the dedicated GSF algorithm (Section 2.3.2) that requires an additional correction because of the significant bremsstrahlung energy loss of the electrons in the tracker material. In particular, a bias of the order of 0.5% and 2% in the barrel and endcaps, respectively, is observed along the azimuthal direction both in data and in MC. For this reason, a pre-calibration of the tracker is performed to provide an unbiased momentum measurement. The procedure is described in detail in the next Sections.
As visible in Fig. 2.13, the E/p distribution for high energy electrons from W $\rightarrow e\nu$ and Z $\rightarrow e^+e^-$ decays consists in a peak centered around one. The width of the peak is dominated by the tracker momentum measurement which is more precise and accurate at small η values. The right tail of the distribution is due to energy losses for bremsstrahlung which are not fully corrected in the tracker momentum measurement through the GSF fit. As a consequence the precision of this calibration method worsen as increasing η.
The events selected by the HLT path requiring one or two electron candidates are used for the calibration. Further selections on the electron quality and isolation, and on the event kinematics, tighter than the trigger requirements, are applied to select a pure sample of electrons from the W and Z boson decays.

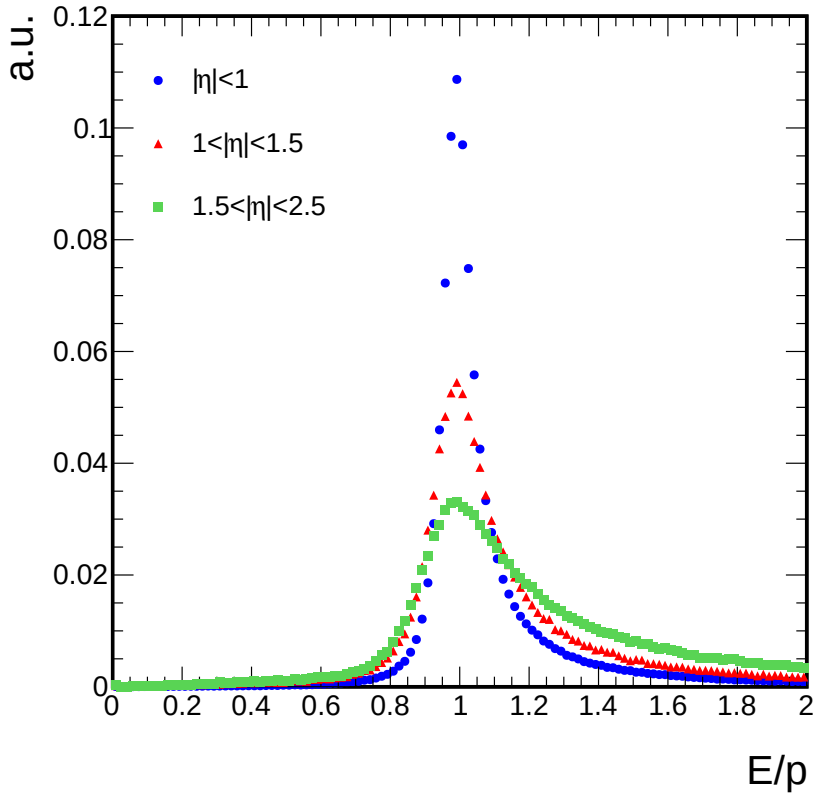


Figure 2.13: E/p distributions of the selected electrons in different intervals of pseudorapidity.

Electron identification requirements

The purpose of the identification requirements is to isolate a pure sample of electrons originating from the hard scattering from electrons originated by photon conversions in the tracker or electrons within a jet, originating for example from a heavy quark decay. Furthermore, a jet can be misidentified as an electron because of its electromagnetic component. On the other hand, a converted photon will have missing hits in the inner tracker layers and a wider electromagnetic shower shape while a jet will have a high multiplicity of tracks around the electron as well as a significant hadronic component measured in the HCAL. For these reasons, the electron identification criteria shown in Table 2.1 are defined providing tight and loose identification tags with an efficiency on signal of 70 and 90-95%, and a background contamination $\leq 1\%$ and $\leq 6\%$, significantly different between EB and EE, respectively. The isolation Iso is the total transverse energy of the charged and neutral hadrons and photons in a cone with $\Delta R = 0.3$ around the electron direction,

relative to the electron transverse momentum:

$$Iso = \frac{1}{p_{T\,ele}} \left[\sum_{\text{Charg. hadr.}} p_T + \max\Big(0, \sum_{\text{Neutr. hadr.}} p_T + \sum_{\text{Phot.}} p_T - \rho \cdot A_{eff}\Big) \right] \quad (2.7)$$

The $\rho \cdot A_{eff}$ factor mitigate the pileup contribution to the isolation terms, and it is called "ρ-correction". In fact, ρ is the average energy density of the event, estimated as the median of the transverse energy density per unit area (in this case from neutral particles) in the event, and A_{eff} is the area of the isolation region corrected for the pileup energy density dependence on the pseudorapidity. The pileup contribution to the charged hadrons isolation term of Eq. 2.7 is negligible because only the tracks that are consistent with originating from the same vertex as the electron are considered. A selection is also applied on the shape of the electromagnetic shower. In particular, the $\sigma_{i\eta i\eta}$ variable is defined as the energy-weighted covariance of the supercluster along the η direction.

Table 2.1: Electron loose and tight identification requirements. The selections are different for superclusters in EB or EE. The H/E variable is the ratio between the hadronic energy H measured in the HCAL and the electromagnetic energy E measured in the ECAL.

Variable	EB		EE	
	Loose ID	Tight ID	Loose ID	Tight ID
$\sigma_{i\eta i\eta}$	< 0.0112	< 0.0104	< 0.0425	< 0.0353
$\Delta\eta$(SC, trk)	< 0.00377	< 0.00255	< 0.00674	< 0.00501
$\Delta\phi$(SC, trk)	< 0.0884	< 0.022	< 0.169	< 0.0236
$H/E < A + B/E_{SC} + C\rho/E_{SC}$	$A = 0.05$ $B = 1.16$ $C = 0.0324$	$A = 0.05$ $B = 1.15$ $C = 0.0324$	$A = 0.0441$ $B = 2.54$ $C = 0.183$	$A = 0.0188$ $B = 2.06$ $C = 0.183$
$Iso < A + B/p_T$	$A = 0.112$ $B = 0.506$	$A = 0.0287$ $B = 0.506$	$A = 0.108$ $B = 0.963$	$A = 0.0445$ $B = 0.963$
$\lvert 1/E - 1/p \rvert$	< 0.193	< 0.159	< 0.111	< 0.0197
Missing inner hits	< 1	< 1	< 1	< 1
Conversion veto	True	True	True	True
Eff. on Drell-Yan electrons	90%	70%	95%	70%
Bkg contamination ($t\bar{t}$ sample)	1.5%	0.4%	6%	1%

Event selection

Further selections are applied to isolate the W $\to e\nu$ and Z $\to e^+e^-$ event candidates. In particular, the W $\to e\nu$ event selection requires exactly one electron with a tight ID and no other electrons with a loose ID. The electron is required to have $p_T >$

30 GeV, $\not\!\!\vec{p}_T > 25$ GeV to account for the neutrino, and a transverse mass $M_T > 50$ GeV. The transverse mass variable M_T is computed as:

$$M_T = \sqrt{(2\not\!\!\vec{p}_T E_T^{ele} \cdot (1 - \cos\Delta\phi(\not\!\!\vec{p}_T, \text{ele}))} \tag{2.8}$$

Where $\Delta\phi(\not\!\!\vec{p}_T, \text{ele})$ is the angle on the transverse plane between the electron and the $\not\!\!\vec{p}_T$ direction. Instead, the Z $\to$e$^+$e$^-$ selections requires two electrons with at least a loose ID in the event. This removes any overlap between the Z $\to$e$^+$e$^-$ enriched sample and the W $\to e\nu$ one. In the case that more than two electrons are present in the event, the pair with the highest p_T is selected. In addition, the dielectron invariant mass m_{ee} is required to be higher than 55 GeV.

Momentum calibration

A calibration of the tracker is performed beforehand to correct for the bias of the momentum measurement along the azimuthal direction ascribed to the tracker material. The selected Z $\to$e$^+$e$^-$ events are used to compute the dielectron invariant mass as:

$$m_{ee}^2(\phi_1, \phi_2) = 4p_1^{\text{trk}} E_2^{\text{SC}} [\sin\Delta\theta_{ee}]^2 \tag{2.9}$$

Where E_i^{SC}, p_i^{trk}, and ϕ_i are the supercluster energy, the tracker momentum, and the opening angle between the two electrons, respectively. Then, the m_{ee}^2 value is averaged with respect to ϕ_2:

$$\langle m_{ee}^2 \rangle(\phi_1) = \alpha_p(\phi_1) \langle 4\hat{p}_1^{\text{trk}} \alpha_E(\phi_2) \hat{E}_2^{\text{SC}} [\sin\Delta\theta_{ee}]^2 \rangle \tag{2.10}$$

Where the "hat" quantities correspond to ideal unbiased measurements. The $\alpha_p(\phi)$ and $\alpha_E(\phi)$ parameters describe the ϕ-modulation of the momentum and of the energy measurements, respectively. In the context of a MC simulation:

$$\begin{aligned} \alpha_p(\phi) &= p/p_{\text{true}}(\phi) \\ \alpha_E(\phi) &= E/E_{\text{true}}(\phi) \end{aligned} \tag{2.11}$$

The $\alpha_p(\phi_1)$ variable can be isolated because it does not depend on ϕ_2. For $\alpha_E(\phi_2)$ the situation is more complex because in the Z $\to$e$^+$e$^-$ events, ϕ_1 and ϕ_2 are anti-correlated. In fact, the ϕ_2 distribution is peaked at $-\phi_1$ with a spread of about 20°. In order to prevent for any ϕ-modulation induced by an uncalibrated ECAL detector, the momentum calibration is performed injecting the intercalibration constants derived with the Z $\to$e$^+$e$^-$ calibration method. Under this condition, the $\alpha_E(\phi_2)$ contribution can be neglected. Therefore, Eq. 2.10 can be simplified as follows:

$$\langle m_{ee}^2 \rangle(\phi_1) = \alpha_p(\phi_1) \langle 4\hat{p}_1^{\text{trk}} \hat{E}_2^{\text{SC}} [\sin\Delta\theta_{ee}]^2 \rangle = \alpha_p(\phi_1) \hat{m}_{ee}^2 \tag{2.12}$$

Since $\hat{m}_{ee} = m_{\text{Z}}$ is a well known constant, the measurement of $\langle m_{ee}^2 \rangle(\phi_1)$ provides the access to the $\alpha_p(\phi_1)$ parameter. The correction is computed in 360 ϕ_1 intervals,

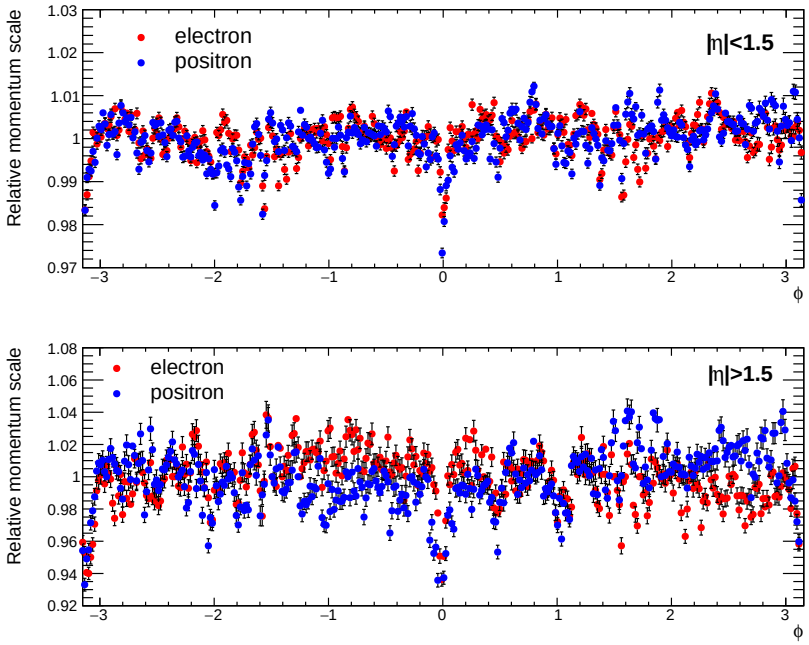


Figure 2.14: Modulation of the momentum measurement of electrons (red) and positrons (blue) as a function of the particle direction along the transverse plane in the barrel and in the endcaps on top and on the bottom panels, respectively, in the 2017 dataset.

separately for EB and EE because of the different upstream material, and separately for positrons and electrons because of the opposite bending in the magnetic field. In order to achieve the best precision, $\langle m^2_{ee} \rangle$ is estimated in each ϕ_1 interval through a template fit of the m^2_{ee} distribution. The template function is built using all the collected data (integrating along ϕ_1). The momentum modulation measured using the 2017 dataset is shown in Fig. 2.14. The main structures visible in the figure, e.g. at $\phi = 0, \pm\pi$, correlate with the tracker structures. The small offset between electrons and positrons, especially visible in the endcaps, is ascribed to a residual tracker misalignment that has an opposite effect on electrons and positrons.

Intercalibration algorithm

The E/p intercalibration algorithm consists in an iterative procedure. At the n-th iteration of the algorithm, the intercalibration constant of the i-th crystal is computed as:

$$c_i^n = c_i^{n-1} \cdot \frac{\sum_{j=1}^{N_e} w_{ij} \cdot f(E/p)_j \cdot (p/E)_j}{\sum_{j=1}^{N_e} w_{ij} \cdot f(E/p)_j} \qquad \text{with } c_i^0 = 1 \tag{2.13}$$

Where N_e is the total number of selected electrons, w_{ij} is the fraction of the j-th electron supercluster energy in the i-th crystal, and $f(E/p)_j$ is a weight correspond-

ing to the probability of finding an electron with a E/p value in a given barrel or endcap ring. The $f(E/p)$ probability is build using the E/p distribution observed in data within a given η-ring. At each algorithm iteration the $f(E/p)$ is updated accordingly to the derived c_i values. In addition, for the EB superclusters only the electrons with $0.85 < E/p < 1.15$ are selected to exclude electrons with an inaccurate momentum measurement. With the applied selections the ECAL resolution is expected to be close to the percent level, thus a higher deviation of the E/p ratio is caused by an inaccurate tracker momentum measurement. The selection on the E/p value was optimized to minimize the IC statistical uncertainty in the widest part of the EB. The IC statistical uncertainty in a given η-ring is estimated as the standard deviation of:

$$\sigma^i_{stat} = \frac{c_{ieven} - c_{iodd}}{c_{ieven} + c_{iodd}} \quad \text{with } i \in \eta\text{-ring} \tag{2.14}$$

Where c_{ieven} and c_{iodd} are the intercalibration constants derived using two orthogonal subsets of the selected events. At the same time, no bias induced by the E/p selection on the c_i values cut was observed. The map of the intercalibration constants derived with the E/p method for the 2017 is shown in Fig. 2.15. The small and big white squares in the maps correspond to non-responding ("dead") crystals and dead trigger towers, respectively. Structures matching the granularity of the ECAL modules are visible in the barrel intercalibration map (left panel). These structures originate from drifts of the laser monitoring system, as presented in Section 2.4.6. In the endcap intercalibration map (right panel), a modulation of the intercalibration constants along the ϕ direction is visible. The modulation is mostly matched with the pixel and tracker detector structures. Such a modulation is not present in the intercalibration constants derived using the $Z \rightarrow e^+e^-$ method. Therefore, the effect is ascribed to a residual miscalibration of the tracker momentum.

As presented in Section 2.4.8, the IC precision achieved by the E/p method in the barrel is on average 0.5%, while in the endcaps is worse than 1.5%. The worsening of the precision in the endcaps is expected as a consequence of the tracker momentum resolution worsening and the larger material budget.

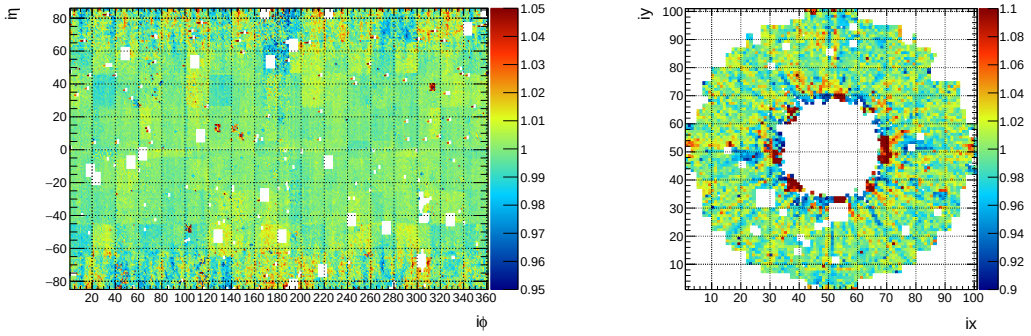


Figure 2.15: *Left:* Map of the EB intercalibration constants derived with the E/p method for the 2017 data-taking. *Right:* Map of the intercalibration constants of one of the endcaps derived with the E/p method for the 2017 data-taking.

2.4.5 Laser Monitoring system

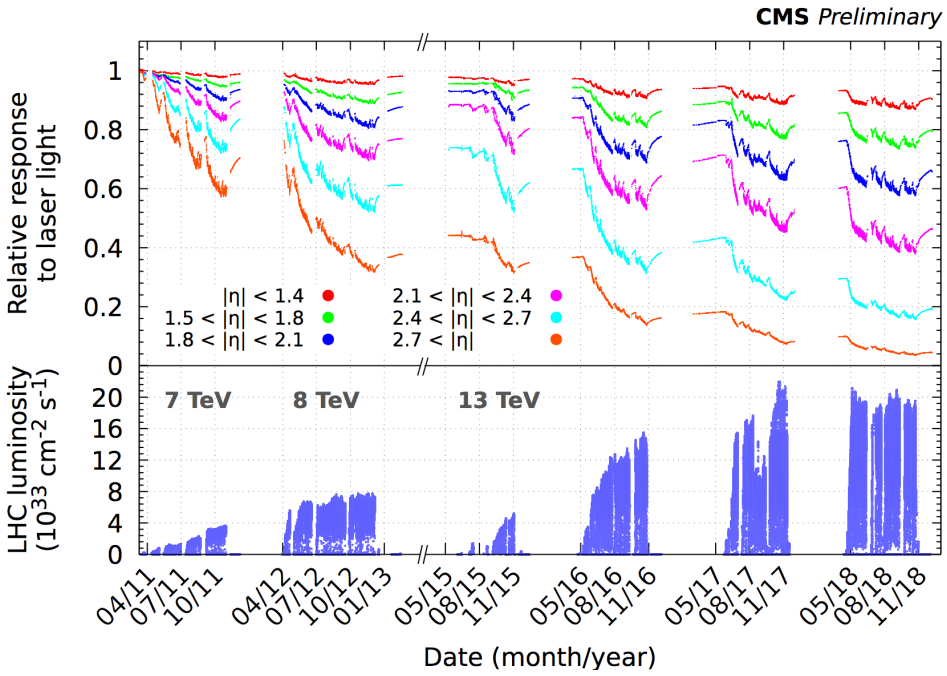


Figure 2.16: Evolution of the ECAL response to the laser light for different intervals of pseudorapidity, represented with the different colors. The corresponding LHC delivered luminosity is also shown.

The laser monitoring system is a key element for the ECAL channels stability. The crystals transparency varies over time as a consequence of the radiation damage and the spontaneous recovery [45] during the LHC shutdown periods. The loss of transparency during the LHC operations is larger in the forward regions of the

detector because of the higher absorbed dose of radiation. A laser monitoring system [46] is used to measure and correct for the crystal transparency changes. In particular, the light from two laser sources with wavelengths of 440 nm and 495 nm is distributed to each crystal through a multiplexed system of optical fibers. The variation of the channel response to the laser light, shown in Fig. 2.16, is used to correct for the crystal transparency variations. The response of the channels to the laser light L is correlated to the response to electrons (and photons) S through the relation:

$$\frac{S}{S_0} = \left(\frac{L}{L_0}\right)^{\alpha} \tag{2.15}$$

Where L_0 and S_0 are the responses to the laser light and to energy deposits, respectively, at a given reference time. α is a parameter dependent on the crystal manufacturer and on the fiber read-out configuration, which differs between EB and EE. The values of α for EB and EE were measured in test beams prior to the installation in CMS, and found to be on average 1.5 and 1.1, respectively. However, for large transparency losses, as for the EE crystals during the Run 2, an effective reduction of the α parameter is expected. For this reason, a fine tuning of α is performed in the EE for each year of the Run 2 in three separate pseudorapidity intervals. The α value selected in each interval is the one providing the most stable energy scale estimated using the $Z \rightarrow e^+e^-$ peak position in the m_{ee} distribution. The optimal α values range from 1.1 to 0.7 for the low and high pseudorapidity regions, respectively.
The light is injected in the crystals sequentially during the beam abort gaps (3 μs interval every 90 μs during the LHC collisions). This allows the scan of all the crystals with the two light sources in about 40 min. In order to correct for possible variations in time or non-uniformities across the detector of the injected light, the laser pulse is also injected in PN diodes which are used as a reference. In particular, the crystals are grouped into regions called harnesses, made up of 100 or 200 crystals in EB and 36 crystals in EE, whose light pulse is monitored by the same PN diode. Adjacent pairs of harnesses share the respective PN diodes through cross-connections to ensure the redundancy of the reference. The harness division of a supermodule is represented in Fig. 2.17, the harness pairs 2-3, 4-5, 6-7, 8-9 shares the same PN pairs, while the harness 1 share the PN with the corresponding harness in the adjacent supermodule. The PN diodes are installed on-detector upstream the ECAL crystals at a pseudorapidity corresponding to approximately the center of the monitored harness.

2.4.6 Time-dependent energy scale correction with electrons

As visible in Fig. 2.18, the laser correction improves the energy scale stability but a residual drift of up to 1% every 10 fb^{-1} of integrated luminosity is observed in the barrel. The effect of the residual drift was observed for the first time in the 2016 intercalibration constant map with a granularity matching the EB harness regions.

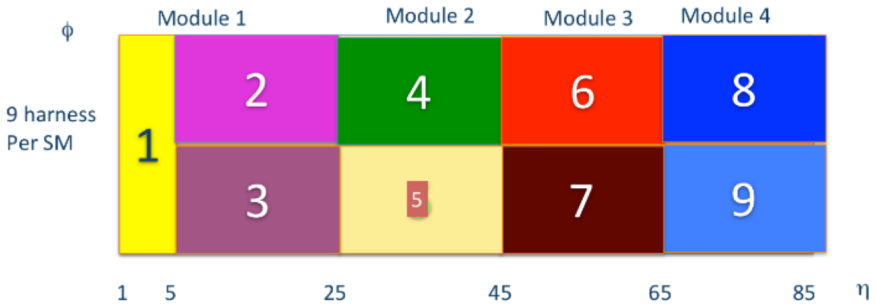


Figure 2.17: The division of a EB supermodule in harnesses.

For this reason this effect, still under study, is ascribed to the PN diodes system. The E/p ratio of electrons from W $\rightarrow e\nu$ and Z $\rightarrow e^+e^-$ events is used to measure and correct for this residual drift. The same dataset and the same selection criteria of the E/p intercalibration are applied. The selected electrons are divided in harness regions (depending on the supercluster seed position) and in integrated luminosity intervals. For each interval, a template function is fitted to the E/p distribution to extract the relative energy scale parameter, as visible in Fig. 2.19. For each harness region, the template function is built using all the electrons assigned to that harness using the data equivalent to about the first 30 fb^{-1} of integrated luminosity of that data-taking year. This ensures a negligible statistical uncertainty on the template function. As shown in Fig. 2.19, the scale drift in short time intervals is well described by a linear function. A linear fit is performed in each harness to extract a time-dependent energy scale correction.

Properties of the residual drift

The properties of the residual drift throughout the Run 2 are studied using the E/p method. The EB drift map in Fig. 2.20 shows that typically the two harness regions within the same module have a similar drift. This is expected because they share the same PN diodes pair. Although the harness drifts have changed throughout the Run 2, they have been always found coherent for the harness regions at the same $|\eta|$, as visible on the left panel of Fig. 2.21. This suggests that this effect is related to the radiation damage which is approximately uniform along ϕ. An additional evidence is visible in the example on the right panel of Fig. 2.21: when the harness drift is expressed as a function of the estimated dose absorbed by the corresponding PN sensors during the Run 2, the behavior of the harness regions at different η is more similar.

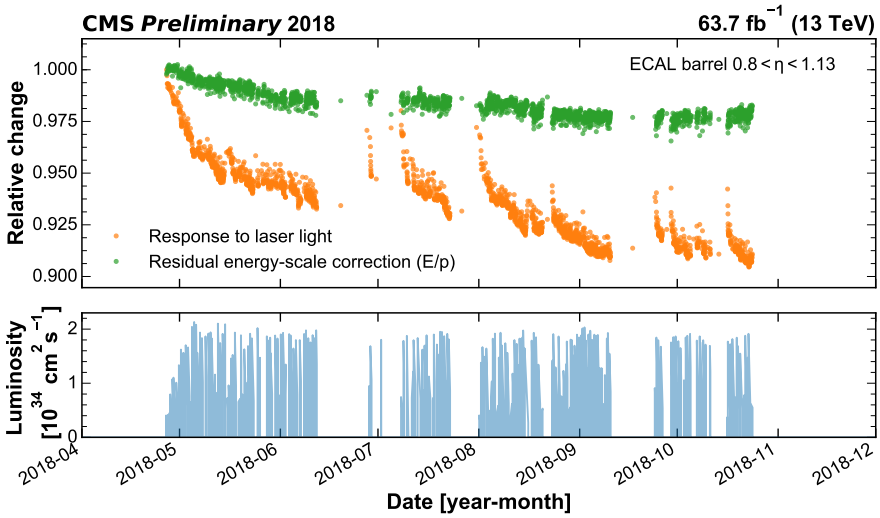


Figure 2.18: The orange dots represent the relative response to the laser light for one barrel module covering the pseudorapidity range $0.8 < \eta < 1.13$, while the green dots represent the residual energy-scale evolution determined through the E/p method as a function of the date in 2018. The values of each curve are normalized to the first corresponding measurement. In the bottom panel the LHC delivered luminosity is shown.

2.4.7 Validation of the energy scale stability

In order to validate the energy scale stability after having applied the corrections for the residual drift, the position of the $Z \rightarrow e^+e^-$ peak in the dielectron invariant mass distribution is monitored across the Run 2. The energy scale stability of barrel and endcaps are reported in Fig. 2.22. Thanks to the time-dependent E/p correction the barrel energy scale is stable within 0.1 GeV RMS. In the endcaps the stability is achieved within 0.4 GeV RMS (0.5 GeV RMS considering the 2018 data-taking only). In the endcaps, the time-dependent E/p correction can not be used due to its limited precision related to the tracker resolution. However, after the event reconstruction, residual drifts in the energy scale with time are corrected for in approximately 18-hour (corresponding to at most one LHC fill) intervals using the $Z \rightarrow e^+e^-$ peak position. The ECAL scale stability is also cross validated using the $\pi^0 \rightarrow \gamma\gamma$ peak position in the diphoton invariant mass finding an analogous result.

2.4.8 Results

A refined calibration of the ECAL was performed for each year of the Run 2 for an optimized event reconstruction. This calibration further improves the results achieved with the preliminary calibrations performed at the end of each year of the data taking, which have been used for most of the analyses, including the one

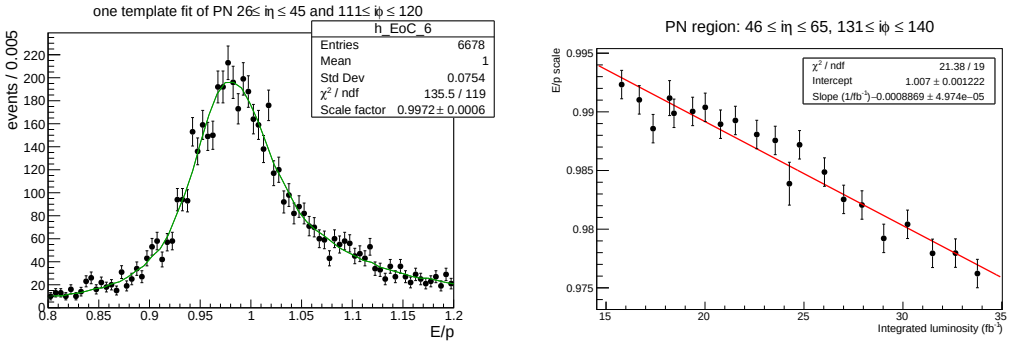


Figure 2.19: *Left:* A template fit of the E/p distribution to extract the energy scale in a certain integrated luminosity interval. *Right:* Fit of the E/p energy scale as a function of the integrated luminosity for a certain period of the 2018 data-taking.

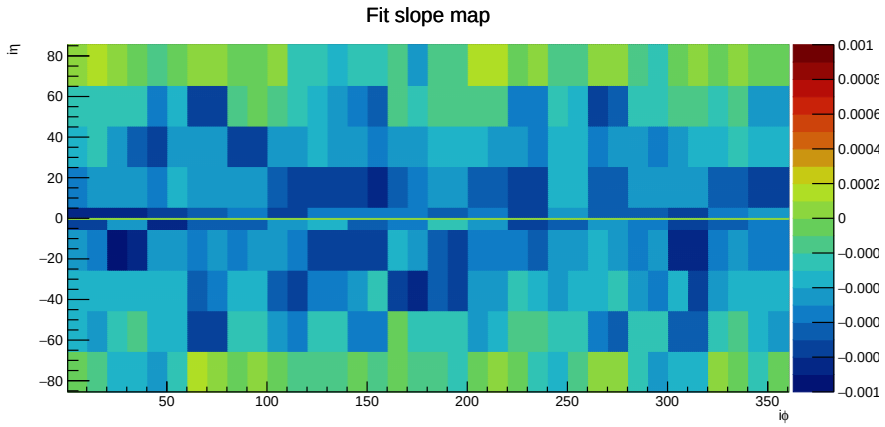


Figure 2.20: The EB map of the harness drift (in $1/\text{fb}^{-1}$) for the middle period of the 2018 data-taking.

presented in this work. The optimized events reconstruction is especially useful for the precision measurements of the SM parameters such as the Higgs boson mass measurement in the H→$\gamma\gamma$ and the H→ZZ(4ℓ) channels with the Run 2 dataset.
The intercalibration precision of the three considered methods and of their combination for the refined calibration of the 2018 data-taking is shown in the left panel of Fig. 2.23. The Z →e^+e^- and E/p methods are the most precise in the barrel while a slightly worse precision is observed for the π^0 method. In the endcaps the intercalibration is fully dominated by the Z→e^+e^- method. In particular, for $1.5 < |\eta| < 2.5$ the precisions of the E/p and π^0 methods are partially and totally above the vertical range of the figure, respectively. In the $|\eta| > 2.5$ region, out of the tracker coverage, only the Z →e^+e^- method is considered. The combined intercalibration precision is better than 0.5% and 1% in the barrel and endcaps, respectively.
The improvement of the electron energy resolution for the 2018 dataset as a func-

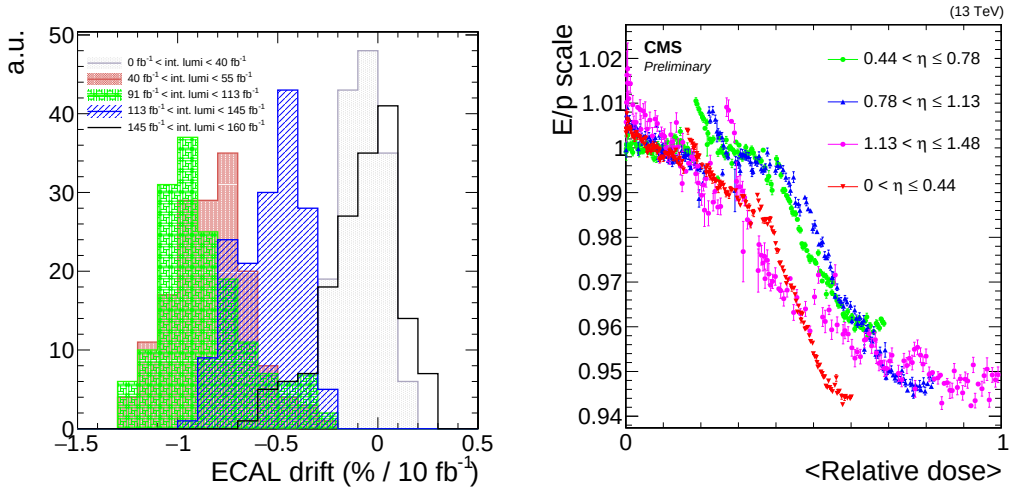


Figure 2.21: *Left:* Distributions of the energy scale drift of the harness regions at the pseudorapidity $|\eta|< 0.46$ for different periods of the Run 2. The same coherent behavior is observed in the other η regions. *Right:* E/p energy scale evolution as a function of the relative average absorbed dose of the corresponding PN sensor pair for the harness regions at $150 < \phi \leq 160$ during the Run 2, normalized to the maximum dose absorbed by the PN at highest pseudorapidity. An analogous behavior is observed in the other harness regions.

tion of the pseudorapidity is shown on the right panel of Fig. 2.23. The resolution is measured through an unbinned likelihood fit to the $Z \rightarrow e^+e^-$ peak in the dielectron invariant mass, using a Breit-Wigner convoluted with a Gaussian function as the signal model. Local structures in the figure in correspondence of $|\eta|= 0$, 0.45, 0.8, and 1.15 correlate with inter-module boundaries in the barrel.
As shown in Fig. 2.24, the electron energy resolution is comparable across the three years of the Run 2 data-taking. A small resolution worsening in time is observed, especially in the endcaps, consistent with the loss of crystal transparency and the increase of noise in the ECAL channels. The figure shows also the energy resolution achieved with the 2012 dataset in Run 1. The Run 1 resolution is 10-20% better than in Run 2. The difference between the Run 1 and the Run 2 is kept relatively small, despite the larger pileup and absorbed dose, thanks to the strategies optimized for the electron (and photon) energy reconstruction in Run 2, described throughout this chapter.
In the context of the $H\rightarrow\gamma\gamma$ analyses, the expected impact of the refined calibration is under evaluation through MC simulations. An improvement of 20 and 10% on the $m_{\gamma\gamma}$ peak resolution is expected for the 2017 and 2018 datasets, respectively, with respect to the preliminary calibrations. The resolution improvement for the 2016 dataset is smaller because an accurate preliminary calibration was derived for the Higgs boson mass measurement with the 2016 dataset [47]. Combining the three years, a 6% gain in sensitivity to the $H\rightarrow\gamma\gamma$ signal is expected. This is equivalent

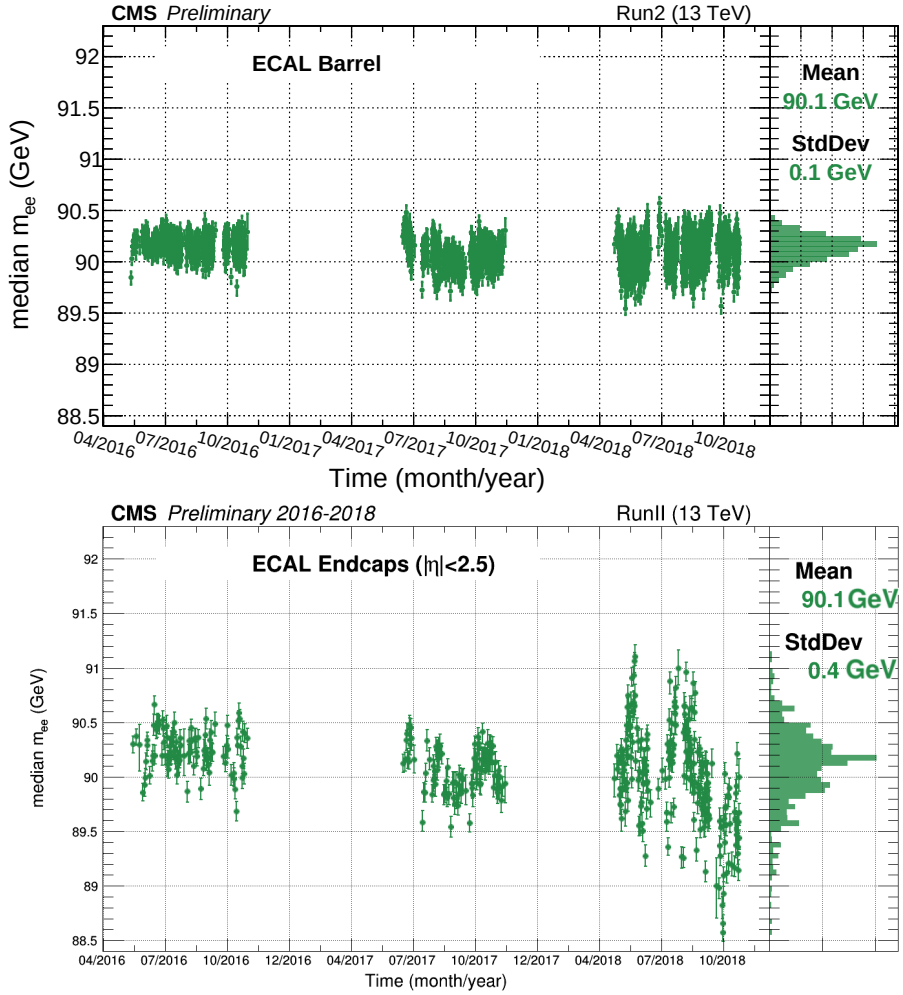


Figure 2.22: *Top:* Time stability of the dielectron invariant mass distribution for the full Run2 data-taking period using Z $\rightarrow e^+e^-$ events. Both electrons are required to be in EB. *Bottom:* Time stability of the dielectron invariant mass distribution for the full Run2 data-taking period using Z $\rightarrow e^+e^-$ events. Both electrons are required to be in EE.

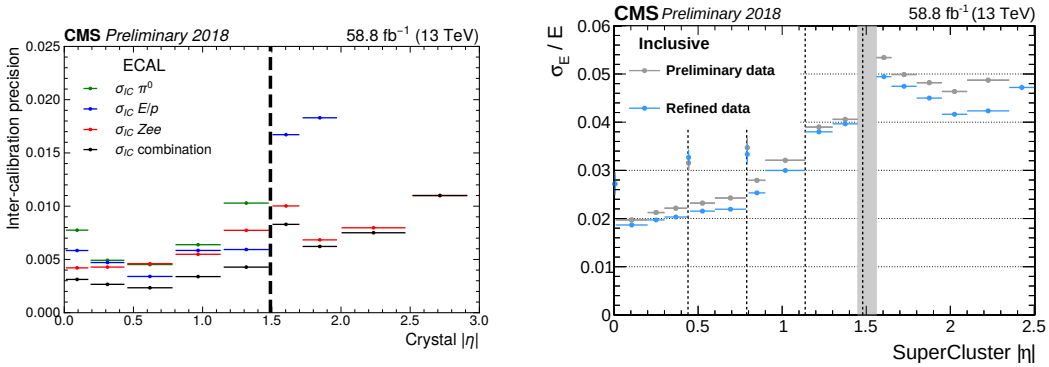


Figure 2.23: *Left:* Residual mis-calibration of the ECAL channel inter-calibration, as a function of the pseudo-rapidity with the 2018 dataset. The red, blue, and green points represent the residual mis-calibration corresponding to the three considered intercalibration methods, while the black points represent the residual mis-calibration of their combination. *Right:* ECAL energy resolution on electrons as a function of the pseudorapidity for the 2018 dataset with the preliminary and refined calibrations.

to a gain in integrated luminosity of about 17 fb^{-1}.

Data/MC comparison

The energy resolution predicted by the MC simulation on electrons and photons is typically 1-3% better than the resolution observed in data. At the same time, an energy scale difference of about 1-2% is observed. The reason is attributed to detector effects that are not included in the simulation. Therefore, additional energy smearing and scale factors are derived using $Z \rightarrow e^+e^-$ events to improve the data-simulation agreement. In a first step the energy scale observed in data is corrected to match to the one predicted by the simulation. In a second step the energy resolution observed in the simulation is corrected to match to the one observed with data. The correction factors are derived in intervals of pseudorapidity and R_9 which is defined as the ratio between the energy deposited in the 3×3 crystal matrix centered on the supercluster seed and the total supercluster energy. The R_9 variable is used to take into account the dependence of the correction from the shower shape. As shown in Fig. 2.25, an excellent data-simulation agreement is finally achieved. This is fundamental for the $H \rightarrow \gamma\gamma$ analysis because an accurate MC simulation is required to describe the photons from $H \rightarrow \gamma\gamma$ events.

2.4.9 Conclusion

In summary, in this work we have refined the calibration tools. The precision of the monitoring and calibration process has allowed us to identify new sources of instabilities that we ascribe to radiation damage effects on the monitoring system.

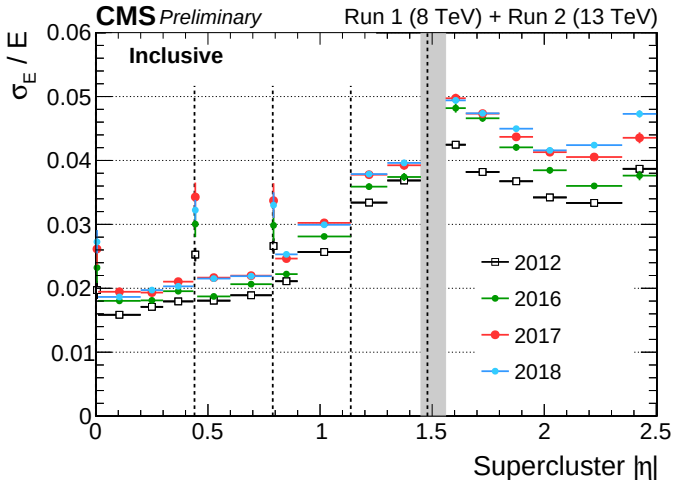


Figure 2.24: ECAL energy resolution on electrons with the refined calibration as a function of the pseudorapidity comparing the 2016, 2017, and 2018 data-taking periods, as well as the 2012 data-taking (Run 1).

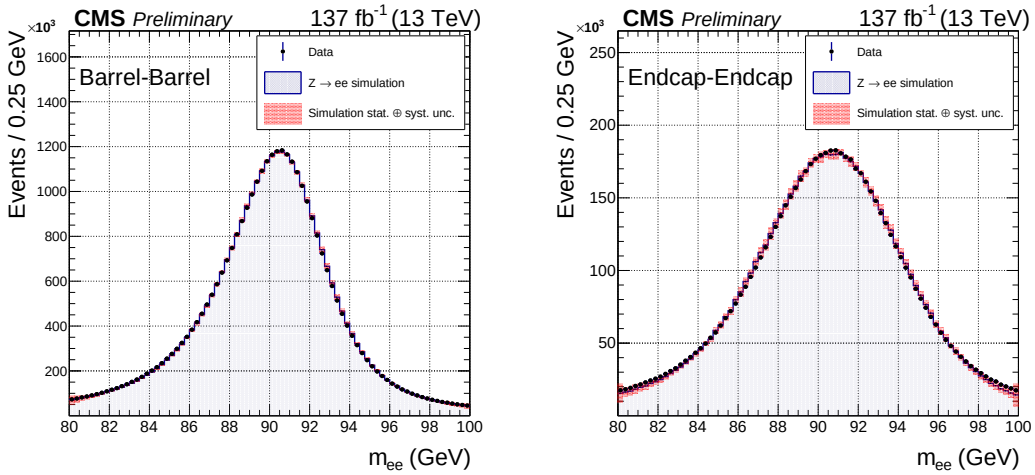


Figure 2.25: *Left:* m_{ee} distribution for data and MC simulation in Z $\rightarrow e^+e^-$ events with both electrons in the EB with the energy scale and smearing corrections applied. *Right:* m_{ee} distribution for data and MC simulation in Z $\rightarrow e^+e^-$ events with both electrons in the EE with the energy scale and smearing corrections applied. Figures from [48].

A correction procedure has been developed that enables us to fully compensate the effect and achieve in Run 2 an energy resolution very close to the Run 1.

Chapter 3

The search for the $\mathrm{HH} \to \mathrm{b\bar{b}}\gamma\gamma$ process

This chapter presents the search for the non-resonant production of Higgs boson pairs in the $\mathrm{b\bar{b}}\gamma\gamma$ final state. The work exploits the data collected by the CMS detector in proton-proton collisions with a center-of-mass energy of 13 TeV, for a total integrated luminosity of 137 fb^{-1}. The HH production is found to be consistent with the SM. Constraints on anomalous values of the Higgs coupling parameters are set. The parameters considered are the Higgs boson trilinear self-coupling constant λ_{HHH}, the coupling constant of two Higgs bosons with two vector bosons $c_{2\mathrm{V}}$, and the Yukawa coupling of the Higgs boson with the top quark y_{t}. A simultaneous fit of the HH and the $\mathrm{t\bar{t}H}$ cross sections is also performed to improve the sensitivity on the λ_{HHH} and the y_{t} parameters, and to simultaneously measure the two parameters. The $\mathrm{t\bar{t}H}$ and HH processes are intrinsically correlated because they both depend on the λ_{HHH} and y_{t} constants.

3.1 Signal topologies and backgrounds

As discussed in Section 1.4.2, the $\mathrm{HH} \to \mathrm{b\bar{b}}\gamma\gamma$ final state consists in two highly energetic photons and two highly energetic b-jets. For the qqHH topology, two additional jets with a large pseudorapidity difference are produced in the final state. The main sources of background are the $\gamma\gamma$+jets process and the γ+jets process with one jet misidentified as a photon. Other important sources of background are the $\mathrm{t\bar{t}}\gamma\gamma$ and the $\mathrm{t\bar{t}}\gamma$ processes. The events of single Higgs production with the Higgs boson decaying to a pair of photons represent an additional important source of background for the HH search. The contamination of $\mathrm{t\bar{t}H}(\gamma\gamma)$ events, which feature two photons and two b-jets in the final state, in the HH signal regions is especially pronounced.
The $\mathrm{t\bar{t}H}(\gamma\gamma)$ final state consists in two photons and the decay products of the $\mathrm{t\bar{t}}$ quark pair, as summarized in Section 1.4.1. The additional objects in the final state coming from the two W bosons decays, following the $\mathrm{t\bar{t}}$ decays, are useful to

separate the tt̄H from the HH events and to select them. The background for the tt̄H signal is dominated by the γγ+jets and γ+jets processes as well as the tt̄γγ process.
For both HH and tt̄H processes, the photon identification performance is fundamental to reduce the background sources with one or two jets misidentified as photons. At the same time, the b-jet identification performance (b-tagging) is fundamental to reduce the background sources without heavy flavor jets.

3.2 Analysis strategy

The candidates are primarily selected requiring two high energy reconstructed photons. In order to maximize the sensitivity to the HH and the tt̄H signals, the events are classified in exclusive categories each targeting a specific single(double) Higgs production mechanism, i.e. ggHH, qqHH, and tt̄H. Each category requires additional objects in the final state and a specific event topology. In addition, multivariate analysis (MVA) classifiers were developed to isolate each signal from its background in the target category. For clarity, the analysis workflow is summarized in Fig. 3.1. A sub-classification of the events based on the MVA classifiers scores is performed in each category. In the ggHH and qqHH categories, the events are further classified using the four-body mass of the photons and jets pairs forming the H→γγ and H→bb̄ candidates to improve the sensitivity to several BSM scenarios. For the tt̄H(γγ) process, the signal is identified as a peak in the diphoton invariant mass distribution at the m_{H} value. Instead, for the HH → bb̄γγ processes, the signal is identified as a peak in the $m_{\gamma\gamma}$ distribution, and a peak in the m_{jj} distribution, both at the m_{H} value. For both processes, the background is modeled from data. In particular a parametric background model is defined. In order to measure the parameters of interest, a likelihood fit of the signal plus background model to the $m_{\gamma\gamma}$ distribution is performed in each of the tt̄H enriched categories, while a simultaneous fit to the $m_{\gamma\gamma}$ and the m_{jj} distributions, assumed uncorrelated, is performed in each of the HH enriched categories.

3.3 Data samples

The data used for this analysis were collected by the CMS detector from 2016 to 2018 (Run 2) for an integrated luminosity of 137 fb^{-1}.

3.3.1 Trigger requirements

At the L1 trigger level, the events are selected requiring one or two electromagnetic (e/γ) candidates. In case of a single e/γ candidate, the minimum E_T is required to be 40 GeV (30-32 GeV in case of an isolated candidate) to maintain the trigger rate at a sustainable level. For the events with a double e/γ candidate, the E_T

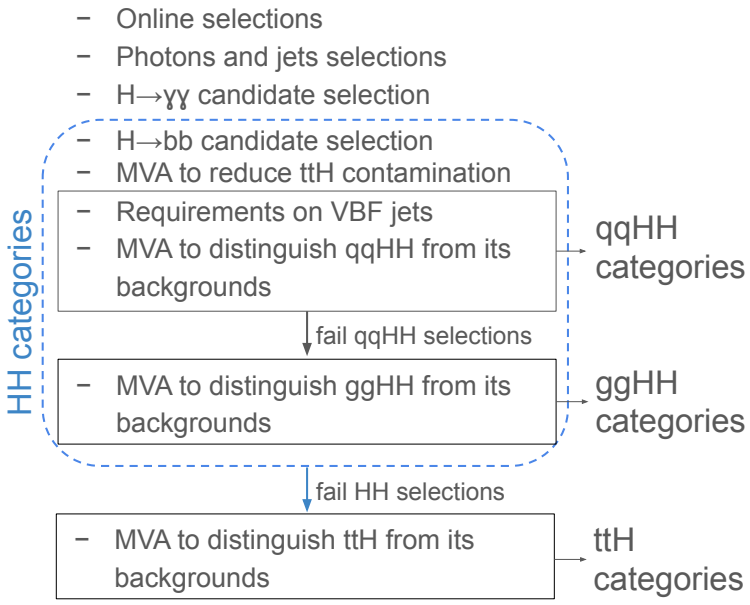


Figure 3.1: Scheme of the analysis workflow.

thresholds for the leading(subleading) candidates are set to 23(10) GeV in the 2016 data-taking period. In the following years, featuring a higher luminosity, the E_T thresholds are raised to 25(14) GeV to limit the trigger rate.
Relying on a refined reconstruction of the events, the HLT trigger requirements are tighter than the L1 trigger. Furthermore, the inclusion of the tracker information allows the separation of diphoton from dielectron candidates in dedicated trigger paths. In the HLT diphoton trigger, used for this analysis, the E_T variable is required to be higher than 30(18) GeV for the leading(subleading) photon in 2016, and than 30(22) GeV in 2017 and 2018. In order to reduce the contamination from misidentified jet, additional selections are applied on the H/E ratio, on the photons isolation, and on the photons shower shapes. The efficiency of the trigger selections is estimated from the data and then used to scale the MC simulations which do not include the trigger effects. In particular, a Tag and Probe ($T\&P$) method with Z →e^+e^- events is used [49]. The "probe" electron is treated as a photon candidate, i.e., ignoring the track information. The efficiency is estimated in intervals of p_T, η and R_9. The different kinematics of the Z →e^+e^- and H→$\gamma\gamma$ events, and the different interaction of electrons and photons with the material upstream the ECAL, resulting in a different shower shape, are properly taken into account. A rescaling of the "probe" electrons weights in intervals of η and R_9, denser than the ones considered to compute the efficiencies, is performed to match the corresponding distributions of the H→$\gamma\gamma$ photons. The scale factors are derived from MC simulations and are shown in Fig. 3.2. The main effect of the scale factors consists in a shift of the R_9 distribution to values close to unity. The shifting is

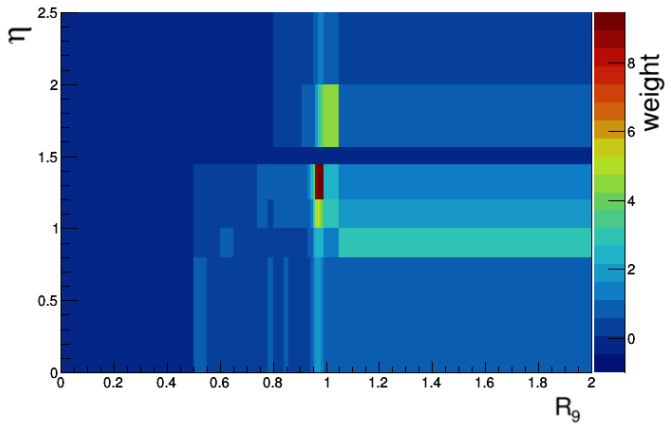


Figure 3.2: Scale factors to correct the differences in R_9 and η between the Z →e^+e^- events, used to compute the trigger efficiencies, and the H→$\gamma\gamma$ events for the 2018 dataset. Similar scale factors are obtained for the 2016 and 2017 datasets.

more pronounced in the region $1 < |\eta| < 1.5$ where the material upstream the ECAL is maximum.

3.4 Simulated samples

The sample are simulated using the MADGRAPH5 generator for LO accuracy in QCD, and MADGRAPH5_AMC@NLO [50] or POWHEG BOX2 [51] generators for up to NLO accuracy in QCD. The generators are interfaced with PYTHYA8 [52] that performs the parton showering and the hadronization. The PYTHIA tuning CUETP8M1 and CP5 [53, 54] is used for the underlying events modeling of 2016 and 2017-2018 samples, respectively. The PDFs are taken from the NNPDF3.0 [55] (2016) and NNPDF3.1 [56] (2017 and 2018) sets. The particles interaction with the detector and the subsequent readout is simulated using the Geant4 package [57].

3.4.1 Higgs boson production samples

The list of the simulated SM Higgs boson (pair) production samples is presented in Table 3.1. The Higgs boson mass is assumed to be 125 GeV.
The ggHH samples are generated using POWHEG BOX2 at the NLO accuracy in QCD and including the full top quark mass dependence [58]. The qqHH samples are generated using MADGRAPH5_AMC@NLO at the LO accuracy. The SM ggHH and qqHH samples are generated along with BSM samples with anomalous $(\kappa_\lambda, \kappa_t, c_V, c_{2V})$ coupling values, as shown in Table 3.2. A proper combination of the ggHH samples and, separately, of the qqHH samples, through the methods described in Appendix B, allows a full description of the ggHH and qqHH processes in any point of the $(\kappa_\lambda, \kappa_t, c_V, c_{2V})$ parameter-space.

In addition, ggHH samples are generated at LO for the BSM benchmarks described in Section 1.3.3 using MADGRAPH5_AMC@NLO. The twelve BSM benchmarks are added together to increase the statistical precision, and then reweighed to any coupling configuration $(\kappa_\lambda, \kappa_t, c_2, \kappa_g, c_{2g})$ using the generator-level information.
All the main single Higgs production mechanisms with the Higgs boson decaying to a pair of photons are also included. The ggH and qqH processes are generated using POWHEG BOX2 with NLO accuracy in QCD. The VH and tt̄H processes are simulated using the MADGRAPH5_AMC@NLO generator with NLO accuracy in QCD, while the tHq process is simulated at LO using the MADGRAPH5 generator. The cross sections used for the single Higgs production mechanisms correspond to the latest recommendations from [6]. For the ggHH and the qqHH cross sections, global k-factors are applied to match the next-to-next-to-leading order (NNLO) prediction [7] and the next-to-next-to-next-to-leading order prediction [8], respectively.

Table 3.1: List of the simulated SM Higgs boson (pair) production samples.

Mechanism	Generator	Cross section (fb) (including k-factor)
ggH	POWHEG BOX2	$4.41 \cdot 10^4$
qqH	POWHEG BOX2	$3.78 \cdot 10^3$
VH	MADGRAPH5_AMC@NLO	$2.25 \cdot 10^3$
tHq	MADGRAPH5	$7.4 \cdot 10^2$
bb̄H	MADGRAPH5_AMC@NLO	$5.3 \cdot 10^2$
tt̄H	MADGRAPH5_AMC@NLO	$5.1 \cdot 10^2$
ggHH (SM)	POWHEG BOX2	$3.105 \cdot 10^1$
qqHH (SM)	MADGRAPH5_AMC@NLO	$1.73 \cdot 10^0$

Table 3.2: List of the simulated ggHH and qqHH BSM samples for anomalous coupling values. The same setup of the corresponding SM sample is used for the simulation.

Mechanism	Couplings values	Cross section (fb) (including k-factor)
ggHH	$\kappa_\lambda = 0$, $\kappa_t = 1$	$6.973 \cdot 10^1$
ggHH	$\kappa_\lambda = 2.45$, $\kappa_t = 1$	$1.312 \cdot 10^1$
qqHH	$\kappa_\lambda = 1$, $c_V = 1$, $c_{2V} = 2$	$1.42 \cdot 10^1$
qqHH	$\kappa_\lambda = 2$, $c_V = 1$, $c_{2V} = 1$	$1.42 \cdot 10^0$
qqHH	$\kappa_\lambda = 0$, $c_V = 1$, $c_{2V} = 1$	$4.61 \cdot 10^0$
qqHH	$\kappa_\lambda = 1$, $c_V = 1.5$, $c_{2V} = 1$	$6.60 \cdot 10^1$
qqHH	$\kappa_\lambda = 1$, $c_V = 1$, $c_{2V} = 0$	$2.71 \cdot 10^1$

3.4.2 Background samples

The background samples are used for the training of the MVA classifiers and for the optimization of the categories selections. The list of the samples used is shown in Table 3.3. Further selections are applied at the generator level to enrich the sample with signal-like events. In particular, for the γ+jets at least one jet is required to have a high fraction of electromagnetic energy to mimic a e/γ object.

Table 3.3: List of the simulated background samples.

Mechanism	Generator	Cross section (pb)
$\gamma\gamma$+jets ($m_{\gamma\gamma} > 80$ GeV)	Sherpa	88.36
$\gamma\gamma$+jets ($m_{\gamma\gamma} > 80$ GeV, 1 b-jet)	Sherpa	0.8185
$\gamma\gamma$+jets ($m_{\gamma\gamma} > 80$ GeV, 2 b-jets)	Sherpa	0.4874
γ+jets ($m_{\gamma\gamma} > 80$ GeV, $p_T^{\gamma} > 40$ GeV, em-enriched)	Sherpa	874.2
t$\bar{\text{t}}\gamma\gamma$	MADGRAPH5_AMC@NLO	0.017
t$\bar{\text{t}}\gamma$	MADGRAPH5_AMC@NLO	4.078
t$\bar{\text{t}} \to \ell\ell\nu\nu$	MADGRAPH5_AMC@NLO	687.1

3.5 Identification requirements

3.5.1 Vertex identification

The choice of the correct vertex among an average of 30-40 vertices, corresponding to the simultaneous interactions at each bunch crossing, is fundamental to achieve the best sensitivity. The diphoton invariant mass is computed as:

$$m_{\gamma\gamma} = 2E_1E_2\sqrt{1-\cos\theta_{\gamma\gamma}} \tag{3.1}$$

where E_1 and E_2 are the energies of the two selected photons, and $\theta_{\gamma\gamma}$ is the angle between the two photons. The vertex position has a significant impact on the $\theta_{\gamma\gamma}$ value, hence on the $m_{\gamma\gamma}$ resolution. In particular, using a MC simulation it was estimated that, if the interaction point is known with a precision better than about 1 cm, the contribution of the $\theta_{\gamma\gamma}$ resolution to the $m_{\gamma\gamma}$ resolution is negligible with respect to the photon energy resolution [59]. Since the ECAL can not reconstruct the shower direction, the vertex selection for a H$\to\gamma\gamma$ event relies on the possible additional tracks from the primary vertex corresponding to charged particles recoiling against the two photons. Furthermore, in case of photons converted in the tracker, the corresponding tracks provide an additional pointing to the primary vertex. In order to optimize the vertex selection, a MVA based on a boosted decision tree (BDT) classifier is trained using MC simulations to distinguish between a

signal, defined as the closest reconstructed vertex to the true vertex at most within 1 cm, and a background made of the other reconstructed vertices distant more than 1 cm from the true vertex. The ggH, qqH, VH, and t$\bar{\text{t}}$H simulated samples are used for the training, each weighed by its cross section. In case of unconverted photons, the most three discriminating variables are:

$$sumpt2 = \sum_i |\overrightarrow{p_T}^i|^2$$

$$ptbal = -\sum_i p_T^i \cdot \frac{\overrightarrow{p_T}^{\gamma\gamma}}{p_T^{\gamma\gamma}}$$

$$ptasymm = \frac{|\sum_i \overrightarrow{p_T}^i| - p_T^{\gamma\gamma}}{|\sum_i \overrightarrow{p_T}^i| + p_T^{\gamma\gamma}}$$

where the i index runs over all the tracks associated with a given vertex. The H$\rightarrow\gamma\gamma$ vertex is expected to be the hardest vertex of the event, hence maximizing the *sumpt*2 variable. Furthermore, the diphoton system recoils against the other particles coming from the same vertex, thus the *ptbal* variable is maximized by a correct vertex choice while it is distributed around 0 otherwise. For the same reason, the *ptasymm* variable is maximized by a correct vertex choice while it is close to the minimum (-1) otherwise. In the cases of a HH $\rightarrow$ b$\bar{\text{b}}\gamma\gamma$ event or a t$\bar{\text{t}}$H event, the requirement of additional jets in the event ensures a correct choice of the vertex in more than 99% of the events.

3.5.2 Photon identification

An important background source consists in γ+jets events with one jet misidentified as a photon. In particular, a neutral hadron produced from the jet fragmentation can decay to a pair of collimated photons and can be reconstructed as a single electromagnetic object. Several properties of the reconstructed object can be used to discriminate between a photon from the primary interaction vertex, defined "prompt", and the remaining photons, defined "non-prompt". In particular, the photon isolation and the H/E variables can give a clear evidence for the presence of a jet, hence a non-prompt photon. In addition, the shower shape of two photons from a neutral hadron decay, although collimated, is typically wider than the shower shape of a prompt-photon. The photon identification requirement consists in a set of loose preselections followed by a MVA BDT-based selection. Prior to any selection, a correction to the shower shape and isolation variables is applied to the simulated events to improve the agreement in their description with data.

Shower shapes and isolation corrections

The agreement in the description of the shower shapes and isolation variables between data and MC simulation is fundamental to minimize the systematic uncertainties of the analysis. A data-simulation discrepancy on those variables would induce

a discrepancy in the photon identification BDT output, thus in the event selection efficiency and migration across the analysis categories. Although scale factors could be derived to correct the simulation for that, the corresponding systematic uncertainty would be among the dominating ones [60]. Therefore, a correction of the BDT input variables is derived. The origin of the data-simulation discrepancies are ascribed to the mis-modeling of detector effects, such as the ECAL channels pedestal, in the simulation. Such effects are expected to have a similar impact on electromagnetic showers from electrons and photons, thus, a correction is derived using the electrons from Z $\rightarrow$e$^+$e$^-$ events through a $T\&P$ method, with the electrons treated as photons. The shower shape variables to be corrected are:

- R_9;
- $\sigma_{i\eta i\eta}$ which is the the energy weighted extension of the shower in the η direction within the 5×5 crystal matrix centered on the seed crystal, in terms of crystals cells;
- σ_η and σ_ϕ which are the standard deviations of the energy weighted extensions in the η and ϕ directions, respectively, in terms of crystals cells;
- $cov_{i\eta i\phi}$ which is the η - ϕ covariance in terms of crystals cells of the 5×5 crystals centered on the supercluster seed;
- $E_{2\times2}/E_{5\times5}$ which is the ratio of the energy in the 2×2 crystals containing the crystal with the maximum energy and the energy in the 5×5 crystals centered on the supercluster seed crystal;

while the isolation variables to be corrected are:

- Iso_{ph} which is the E_T sum of all the photons (as defined by the particle-flow algorithm) falling inside a cone of size $R = 0.3$ around the photon candidate with the ρ-correction for pileup (see Section 2.4.4) applied;
- Iso_{ch} which is the transverse momentum sum of all the tracks identified as charged hadrons falling in a cone of size $R = 0.3$ around the photon candidate direction with the ρ-correction applied;
- Iso_{ch}(Worst vertex) which is the Iso_{ch} variable, as defined above, computed with respect to the vertex of the event giving the largest value (worst vertex). This is useful because a prompt photon is generally isolated with respect to any vertex of the event, unlike a non-prompt photon.

In order to properly account for the correlations among those variables, a chained quantile regression (CQR) method is used to derive the corrections [48]. The CQR method consists in an ordered sequence of BDTs, each predicting the cumulative distribution function (CDF) of a specific input variable. The CDFs extracted from

data and MC simulation are then used to derive a per-electron(photon) correction. The inputs to a BDT consist in the predictions of the BDTs prior in the sequence, and a common set of variables which are the electron (p_T, η, ϕ), and the global event energy density ρ.
Since the correlation among the isolation from charged hadrons, the isolation from photons, and the shower shape variables is negligible, independent CQRs are developed for the three sets of variables. The training is performed separately for each year of the data-taking because of the different detector conditions. The effectiveness of the corrections is validated using a set of Z $\rightarrow e^+e^-$ events from a testing sample not included in the training, and also through the $T\&P$ method with a Z $\rightarrow\mu^+\mu^-\gamma$ sample, as shown in the examples in Fig. 3.3.

Photon preselections

A set of preselections is applied both to the simulated and to the data events. Such preselections are slightly tighter than the trigger requirements to limit the effects related to the trigger reconstruction precision (turn-on effects) ensuring at the same time a high signal efficiency. Beside the shower shape and isolation variables already introduced, the preselections shown in Table 3.4, exploit also:

- Iso_{trk} which is the transverse momentum sum of all the tracks falling in a cone of size $R = 0.3$ around the photon candidate direction. In order to avoid inefficiencies for prompt photons converting the tracker, the tracks falling in an inner cone of size $R = 0.04$ are not included in the p_T sum;

- conversion-safe electron veto which consists in requiring exactly zero charged particle tracks, with a hit in the innermost layer of the pixel detector not matched to a reconstructed conversion vertex, pointing to the ECAL supercluster.

Different selections are applied to the photons in the ECAL barrel (EB) and endcaps (EE) because of the different detector geometries, and the to low and high R_9 photons roughly corresponding to converted and unconverted photons, respectively. In order to validate the preselection efficiency agreement of the MC simulation with the data, the electrons from Z $\rightarrow e^+e^-$ events, treated as photons, are used through a $T\&P$ method. Since the electron veto efficiency can not be measured with a Z $\rightarrow e^+e^-$ sample, a Z $\rightarrow\mu^+\mu^-\gamma$ sample is used instead with the $T\&P$ method. Scale factors are extracted to correct the simulation for efficiency discrepancies with data. In the photon categories with the highest purity of prompt-photons, i.e. EB categories and EE-high R_9 category, the scale factors deviate from unity by at most 3%, while in the EE-low R_9 category the scale factor deviates from unity of about 8%.

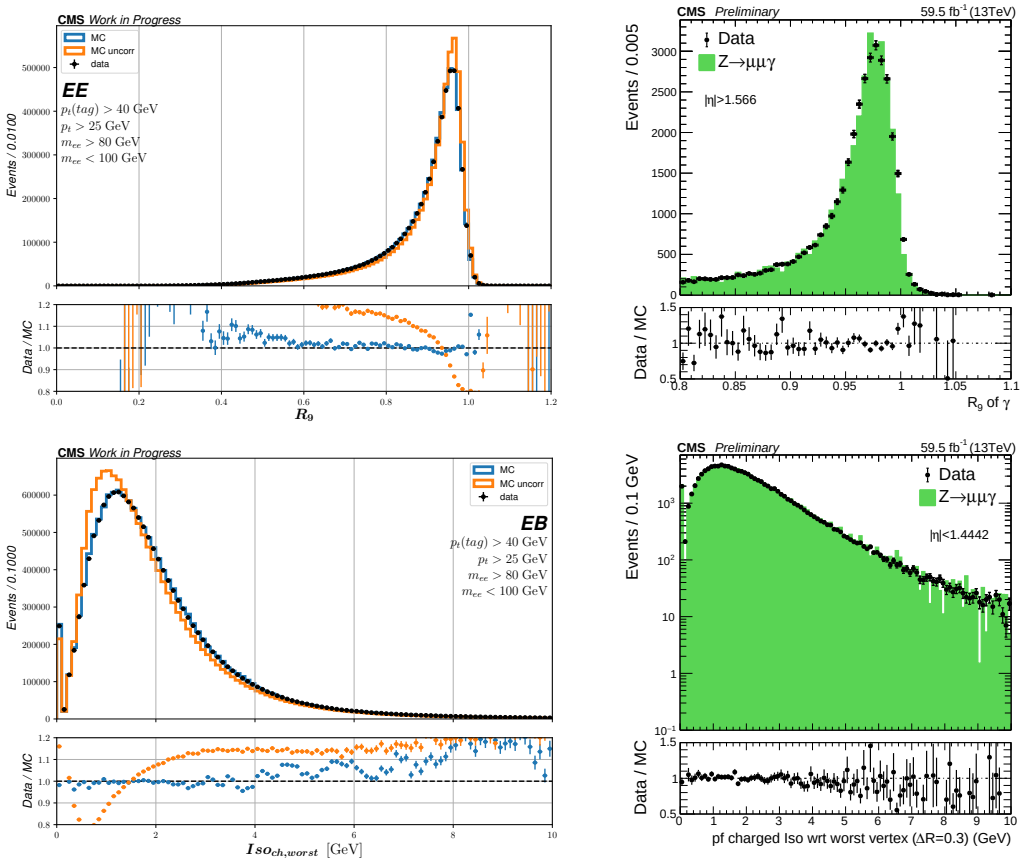


Figure 3.3: *Top:* Data-simulation comparison of the R_9 distribution in the EE using a $T\&P$ method with Z →e$^+$e$^-$ events (*left*) or Z →$\mu^+\mu^-\gamma$ events (*right*), with the 2018 dataset. *Bottom:* Data-simulation comparison of the Iso_{ch}(Worst vertex) distribution in the EB using a $T\&P$ method with Z →e$^+$e$^-$ events (*left*) or Z →$\mu^+\mu^-\gamma$ events (*right*), with the 2018 dataset. For the Z →e$^+$e$^-$ events both the uncorrected and the corrected MC simulations for shower shape and isolation variables are shown in orange and blue, respectively, while for Z →$\mu^+\mu^-\gamma$ only the corrected MC is shown.

Table 3.4: Preselections applied to the photons.

Photon category	EB $R_9 > 0.85$	EB, $R_9 \leq 0.85$	EE $R_9 > 0.9$	EE $R_9 \leq 0.9$
$\sigma_{i\eta i\eta}$	-	< 0.015	-	< 0.015
R_9	-	> 0.5	-	> 0.8
Iso_{ph}	-	< 4 GeV	-	< 4 GeV
Iso_{trk}	-	< 6 GeV	-	< 6 GeV
H/E	< 0.08			
Electr. veto	Yes			
η	$\lvert\eta\rvert < 1.44$ or $1.55 < \lvert\eta\rvert < 2.5$			
lead(sublead) p_T^γ	$> 35(25)$ GeV			
Other select.	$R_9 > 0.8$ or $Iso_{ch} < 20$ GeV or $Iso_{ch}/p_T < 0.3$			

Photon identification BDT

A MVA BDT classifier is used to discriminate between prompt and non-prompt photons. The BDT is trained using a γ+jets MC sample with photon preselections applied. The prompt photons of the events are used as signal while the non-prompt photons as background. In order to make the BDT training as independent as possible from the photon kinematics, the p_T and η distributions of the signal are weighed to match the ones of the background. Beside the variables already used at the preselection level, the $E_{2\times2}/E_{5\times5}$, $cov_{i\eta i\phi}$, σ_η, σ_ϕ, Iso_{trk}(Worst vertex), and ρ variables as well as the η and the raw energy of the supercluster are included in the training to provide additional information.
As visible in Fig. 3.4 (left), the BDT output (*photon ID*) distribution is validated using Z $\rightarrow e^+e^-$ events through a $T\&P$ method, with the electron treated as a photon. The small discrepancies in the *photon ID* distribution between data and MC are ascribed to the residual data-simulations discrepancies in the shower shape and isolation variables. The origin of those discrepancies is assumed to come from the limited size of the training samples for the chained quantile regressions. In order to estimate the related uncertainty, two sets of shower shape and isolation corrections are derived using two separate training samples of the same size. The standard deviation of the events number in intervals of *photon ID* is taken as the systematic shift for the *photon ID* variable. As visible in Fig. 3.4 (right), a good data-simulation compatibility is achieved also for a sample dominated by non-prompt photons. The analysis preselections include the *photon ID* > -0.9 requirement on both the photons, corresponding to an efficiency of 99% on prompt-photons with 49% rejection of non-prompt photons.

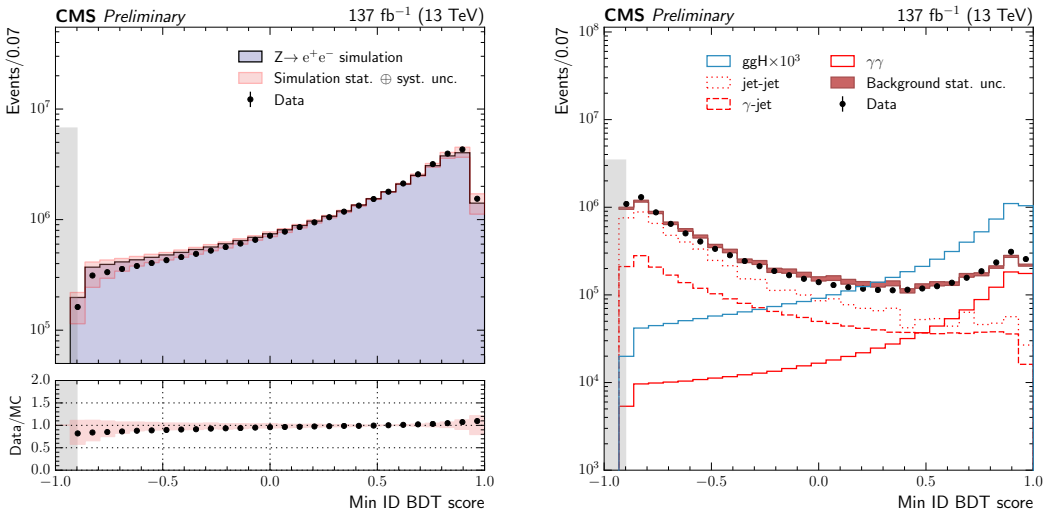


Figure 3.4: *Left: photon ID* distribution for the lowest scoring electron from the preselected Z →e^+e^- events, with the electrons treated as photons. *Right: photon ID* distribution for the lowest scoring photon for all the preselected diphoton events with an invariant mass 100 GeV< $m_{\gamma\gamma}$ < 180 GeV. The photons(electrons) with *photon ID* ≤ -0.9 (represented by the left gray bands) are excluded by the preselection requirements. Figures from [48] considering the full Run 2 dataset.

3.5.3 Jet identification

As discussed in Section 2.3.3, the considered jets are reconstructed with the *anti-kT* algorithm with a distance parameter $R = 0.4$. The charged hadron subtraction correction is also applied. The jets are required to have $p_T > 25$ GeV and $|\eta| < 4.7$. A jet identification criterion is applied to reject reconstructed jets originating from noise fluctuations in the calorimeters channels. The identification criterion exploits the jet energy composition in terms of charged/neutral hadrons, and electromagnetic fraction. The chosen working point ensures an efficiency of more than 98-99% on prompt jets from the interaction vertex and more than 98% rejection of jets from noise. Another source of contamination consists in close-by particles coming from different interaction vertices (pileup) that can be clustered and reconstructed as a jet. A "pileup jet" identification algorithm, based on a BDT discriminator, is trained to discriminate the physical jets from the pileup jets [41]. The BDT exploits observables related to the spatial profile of the jet that is typically wider for pileup jets. The contamination from pileup jets is expected to be significant only for jets in the forward region with $p_T < 50$ GeV. Therefore, the contamination from pileup jets can be significant for the selected VBF jets. Instead, for the jets originating from top quark or H boson decays, the contamination from pileup jets is negligible. Therefore, a tight PU jet ID requirement is applied only to the VBF jet candidates with $p_T < 50$ GeV corresponding to an efficiency of about 80% on prompt jets and a pileup jet rejection of about 98% and 60-80%(strongly p_T dependent) for $|\eta| < 2.5$

and $|\eta|>$ 2.5, respectively.

3.5.4 Heavy-flavored jet identification

The identification of the b-jets (b-tag) is fundamental for the sensitivity of this analysis because it is the signature of the H$\to$b$\bar{\text{b}}$ decay and of the t$\bar{\text{t}}$H production mode. Within CMS, a deep neural network (DNN) classifier (DeepJet [61]) is used. The algorithm exploits the distance between the production and the decay position of the b-hadrons, of the order of 1 mm, resulting in a secondary vertex resolvable from the primary interaction vertex. The algorithm exploits also the fact that the b hadron decay can produce leptons (and undetected neutrinos) with high p_T. Finally, because of the high bottom quark mass, the b-jet have a wider $\eta-\phi$ extension than a light flavor jet. The DeepJet score is used for the b-jets candidates selection and as input for the MVA classifiers.

b-jet energy correction

A good energy resolution on b-jets is necessary for the best sensitivity to the HH $\to$ b$\bar{\text{b}}\gamma\gamma$ channel, because the signal extraction includes a fit to the dijet invariant mass. However, the frequent presence of neutrinos causes an underestimate of the jet energy. At the same time, the b-jet energy deposition often extends over the clustering region causing an energy leakage. Two MVA-based energy regressions are applied in sequence to correct for those effects. The two regressions exploit different properties of the event. The first regression [62] provides a correction for the p_T of the b-jet. It exploits the jet composition and kinematic, the leptons (if any) kinematic and the secondary vertex information as well as the pileup information. It has a DNN architecture and it is trained on a large set of simulated b-jets. As an additional output, the regression provides also an estimate of the b-jet energy resolution.
The second b-jet energy regression provides a correction for the two-jet invariant mass m_{jj}. It exploits the fact that there is not genuine missing transverse energy in a HH $\to$ b$\bar{\text{b}}\gamma\gamma$ event. Therefore, the training with simulated HH $\to$ b$\bar{\text{b}}\gamma\gamma$ events include global event variables such as the $\not{\vec{p}}_T$ as well as the kinematic variables of the reconstructed objects. In addition, the output of the p_T regression is used to feed the m_{jj} regression. It is a BDT regression trained using the XGBOOST package [63]. The impact of the b-jet energy corrections is shown in Fig. 3.5. As expected the first energy correction improves the resolution and the energy scale, while the second correction, exploiting additional information, further improves the energy resolution. Overall the b-jet energy corrections improve the m_{jj} resolution of about 25%.

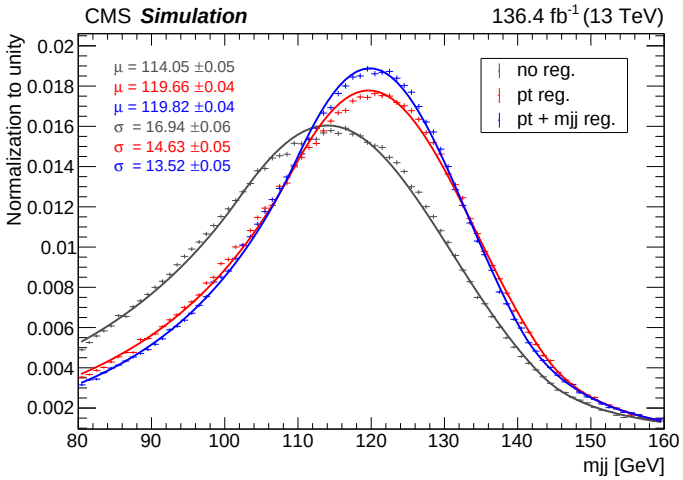


Figure 3.5: m_{jj} distribution for the HH → bb̄γγ simulation without b-jet energy correction (black), with the first energy correction (red), and with both b-jet energy corrections (blue).

3.6 Event selection

3.6.1 Selection of the H→γγ candidate

The H→γγ candidate is built using the photons passing the identification criteria described in Section 3.5.2. If more than two photons are present, the two photons with the highest p_T are selected. The leading and subleading photons are required to have $p_T^\gamma/m_{\gamma\gamma} > 1/3$ and $p_T^\gamma/m_{\gamma\gamma} > 1/4$, respectively. In addition, the diphoton invariant mass is required to be in the window 100 GeV$< m_{\gamma\gamma} <$ 180 GeV. The selections on the photon $p_T/m_{\gamma\gamma}$ ratio was proven to prevents distortions of the $m_{\gamma\gamma}$ spectrum on the low mass side with a negligible loss of efficiency on the H→γγ signal.

3.6.2 Selection of the H→bb̄ candidate

The jets passing the identification selections are also required to have $|\eta|<$ 2.4 and $|\eta|<$ 2.5 for 2016 and 2017-2018 datasets, respectively, and an angular distance from the two selected photons of at least 0.4. The two b-jets with the highest b-tag score are selected to build the H→bb̄ candidate in the HH categories. Finally the m_{jj} value is required to be in the window 70 GeV$< m_{jj} <$ 190 GeV.

3.6.3 Requirements for the qqHH topology

The events with a H→γγ and a H→bb̄ candidate (HH candidates) are required to have at least two additional jets passing the tight PU ID selection. The jets

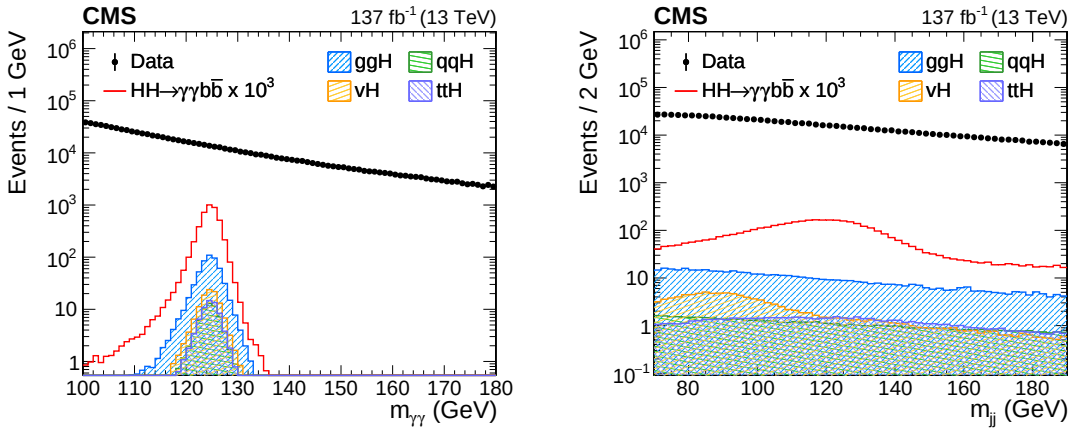


Figure 3.6: Distributions of $m_{\gamma\gamma}$ (left) and m_{jj} (right) for the selected HH candidates for data and for the simulated single and double Higgs processes.

are required to have $p_T > 30$ GeV, $|\eta| < 4.7$, and an angular distance $\Delta R > 0.4$ from each of the two selected photons and b-jets. The two jets with the highest dijet invariant mass are selected as VBF jets candidates. Finally, the leading jet is required to have $p_T > 40$ GeV.

3.7 Background rejection from the HH signal regions

The distribution of the $m_{\gamma\gamma}$ and m_{jj} distributions after requiring the presence of a H→γγ and a H→bb̄ candidate in the events are visible in Fig. 3.6. The signal appears as a peak at 125 GeV in the two distributions smeared by the experimental resolution. A continuum background mainly from γγ+jets and γ+jets events dominates the HH signal region. The single Higgs production processes represent another important source of background because it is resonant in the $m_{\gamma\gamma}$ distribution as the HH signal.
In order to maximize the separation of the signals from the background contaminations, specific MVA-based strategies were developed. MVA classifiers based on different architectures are trained using the simulated events. The classifier outputs are used to define the signal regions and also to classify the events in exclusive categories. The MC simulation events are divided in two subsets. One subset of events is used for the MVA trainings, the other subset is used for the MVA outputs validation, the category optimization, and the signal modeling.

3.7.1 Rejection of t$\bar{\text{t}}$H($\gamma\gamma$) events from the HH signal region

With two photons resonant on m_{H} and two b-jets in the final state, the t$\bar{\text{t}}$H($\gamma\gamma$) process was found to be one of the main backgrounds in the most sensitive HH categories. Therefore, a specific strategy was developed to discriminate the t$\bar{\text{t}}$H($\gamma\gamma$) from the HH $\to$ b$\bar{\text{b}}\gamma\gamma$ mechanisms. In particular, a DNN classifier is trained using the simulated t$\bar{\text{t}}$H events as background and the combination of all the twelve simulated ggHH benchmarks samples (including the SM) as signal. The twelve benchmarks are combined with the same weight. It was verified that this strategy improves the BSM ggHH selection efficiency with a negligible impact on the SM ggHH efficiency. This strategy makes also the classifier suitable to separate the t$\bar{\text{t}}$H from the qqHH events, not included in the training because the corresponding MC simulation became available only at an advanced stage of the analysis. The classifier exploits the angular variables related to the different topologies of the two processes as well as the presence of a W boson, decaying hadronically or leptonically, originated by the top quark decay. The variables used are:

- The minimum angular distance between one of the two selected photons and one of the two selected jets $\Delta R_{\text{min}}(\gamma, \text{b-jet})$;
- The cosine of the angle in the b$\bar{\text{b}}\gamma\gamma$ rest-frame between the diphoton object and the beam axis $|\cos\theta^{\text{CS}}_{\text{HH}}|$;
- The cosine of the angle in the dijet rest-frame between the leading jet and the beam axis $|\cos\theta_{\text{jj}}|$;
- The $\overrightarrow{\not{p}}_T$ absolute value and its azimuthal angles with the selected b-jets $\Delta\phi(\overrightarrow{\not{p}}_T, \text{b-jet})$;
- The p_T of the leading and subleading electrons and muons of the event, if any;
- A top quark sensitive variable χ^2_{top} defined for the events with at least two additional jets as:

$$\chi^2_{top} = \min_{j1,j2,jb}\left[\left(\frac{m_{\text{W}} - m_{j1j2}}{0.1\cdot m_{\text{W}}}\right)^2 + \left(\frac{m_{\text{t}} - m_{jbj1j2}}{0.1\cdot m_{\text{t}}}\right)^2\right] \tag{3.2}$$

  Where $j1$ and $j2$ are two among the additional jets and jb is one of the two selected b-jets. Among all the possible jet combinations, the one minimizing the quantity is chosen. A similar invariant mass resolution, roughly estimated of 10%, is expected for both the W and t masses. In case of four or more additional jets, an additional χ^2_{top} variable is built in the same way with the remaining jets.

A selection on the output score of this DNN (ttH-score) is applied to reject the t$\bar{\text{t}}$H events from the HH enriched categories. The selection on the ttH-score is

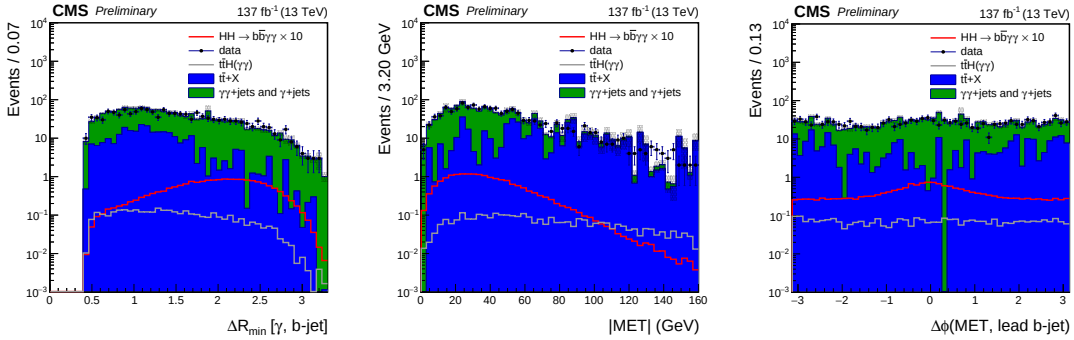


Figure 3.7: Distribution of $\Delta R_{\min}(\gamma, \text{b-jet})$ (*left*), $|\overrightarrow{\not{p}}_T|$ (*center*), and $\Delta\phi(\overrightarrow{\not{p}}_T, \text{b-jet})$ (*right*) of the selected ggHH candidates for the data as well as the simulated $t\bar{t}$ +X, $\gamma\gamma$+jets, and γ+jets events excluding the signal region $120 < m_{\gamma\gamma} < 130$ GeV. The distributions of the simulated HH → bb̄γγ and $t\bar{t}H(\gamma\gamma)$ events are also shown with red and gray lines, respectively.

optimized together with the ggHH category boundaries definition, and separately with the qqHH one, to provide the best sensitivity. The chosen working point ensures a $t\bar{t}H$ rejection of about 80% with an efficiency of about 95% on the ggHH signal. A good performance is also achieved with respect to the qqHH signal. In particular, after the qqHH additional requirements presented in Section 3.6.3, the chosen working point provides a $t\bar{t}H$ rejection of about 85% with 90% efficiency on the qqHH signal.

Validation of the method

In order to validate the training variables for the $t\bar{t}H$ discriminant, their distributions in data and MC simulation were compared. The distribution of three among the most sensitive variables are shown in Fig. 3.7. The comparison included the preselected events with a H→$\gamma\gamma$ and a H→b$\bar{\text{b}}$ candidate (HH candidates) outside the diphoton invariant mass region 115 GeV$< m_{\gamma\gamma} <$ 135 GeV containing the expected signal. The distributions from simulation were found compatible with the distributions observed in data. The distribution of the ttH-score for the different background sources was also studied by means of MC simulations considering for simplicity the 2016 and 2017 datasets, as shown in Fig. 3.8 (left). The peak at ttH-score = 0, corresponding to $t\bar{t}H$-like events, is populated by all the $t\bar{t}$ + (X) processes, while the peak at ttH-score = +1 is populated by the HH-signal and all the other backgrounds, hence dominated by the $\gamma\gamma$+jets events. Considering the full Run 2 dataset, the ttH-score distributions for the selected data events, and for the ggHH and single Higgs simulated events, are found compatible as shown in Figure 3.8.

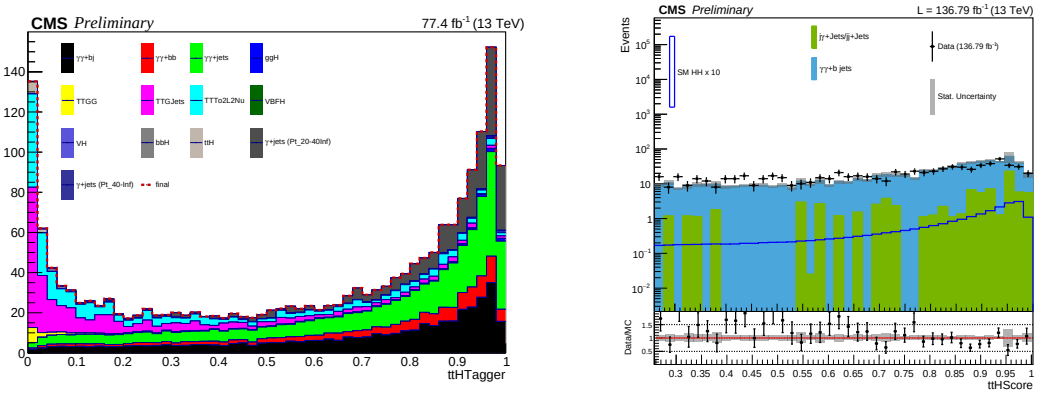


Figure 3.8: *Left:* ttH-score distribution for all the main background components from the MC simulations of the 2016 and 2017 events. *Right:* Comparison between data and MC simulation of the ttH-score distribution for the HH candidates with ttH-score > 0.26.

3.7.2 Discrimination of the continuum background from the ggHH-enriched categories

The dominating background for the ggHH signal region consists in the $\gamma(\gamma)$+jets processes. The optimal background rejection is achieved through an MVA approach using a BDT classifier trained using MC simulated events. The variables optimized for this training exploit the HH system kinematic, the identification variables to reject background processes with jets misidentified as photons or b-jets, and the resolution variables to account for the resonant nature of the signal. In particular, the variables chosen are:

- The H→$\gamma\gamma$ and H→b$\bar{\text{b}}$ candidates kinematics described by $p_T{}^{\gamma}/m_{\gamma\gamma}$ for each of the two selected photons, and $p_T{}^{j}/m_{\text{jj}}$ for each of the two selected jets;
- The transverse balance of the HH system consisting in $p_T^{\gamma\gamma}/m_{\gamma\gamma\text{jj}}$ and $p_T^{\text{jj}}/m_{\gamma\gamma\text{jj}}$ where $p_T^{\gamma\gamma}$ and p_T^{jj} are the diphoton and dijet transverse momentum respectively, and $m_{\gamma\gamma\text{jj}}$ is the four-objects invariant mass;
- The $|\cos\theta^{\text{CS}}_{\text{HH}}|$ and $|\cos\theta_{\text{jj}}|$ helicity angles as defined for the t$\bar{\text{t}}$H discriminant, and in addition, the cosine of the angle in the diphoton rest-frame between the leading photon and the beam axis $|\cos\theta_{\gamma\gamma}|$;
- The angular separation of the objects, i.e., $\Delta R_{\text{min}}(\gamma, \text{b-jet})$ as defined for t$\bar{\text{t}}$H discriminant, and additionally, the ΔR between the other selected photon and jet;
- The b-tag score provided by the DeepJet algorithm for the two selected jets;

- The photon ID output for the two selected photons;
- The energy resolution for the two selected photons, and the diphoton energy resolution estimated by the photon energy regression algorithm;
- The energy resolution for the two selected jets, and the dijet energy resolution estimated by the b-jet energy regression algorithms;
- The global event energy density ρ to account for the different pileup conditions.

The distributions of some of the most discriminating variables for data and simulated events are shown in Fig. 3.9. As for the $t\bar{t}H$ discriminant, this BDT is trained using a combination of the twelve ggHH BSM benchmarks samples as signal, and the $\gamma(\gamma)$+jets as background. Each event used for the training is weighed by the inverse of the estimated diphoton and dijet energy resolutions because the events with good resolutions are expected to provide the highest sensitivity to the HH signal. In order to prevent for any sculpting of the $m_{\gamma\gamma}$ and m_{jj} distributions, the ${p_T}^{\gamma}$ and ${p_T}^{j}$ variables are provided as input for the BDT training scaled by $m_{\gamma\gamma}$ and m_{jj}, respectively. This is fundamental for a correct estimation of the signal and background. The training is performed separately for the three years of data taking because the different detector conditions have modified the variables distributions and correlations. This strategy offsets the differences across the years, providing very similar BDTs output distributions for the three years, both for signal and backgrounds. Therefore, it is possible to merge the three distributions and then uniformly optimize the BDT selections. It is useful for the optimization of the selections on the BDT output score to have a signal BDT output uniformly distributed. To do that, a bijective transformation of the BDT output is performed using the signal CDF as mapping function. The distribution of the BDT output (ggHH-BDT score) for the selected data events, ggHH and single Higgs simulated events is shown in Fig. 3.10.

3.7.3 Background rejection from the qqHH signal region

Analogously to the ggHH process, the background for the qqHH is dominated by $\gamma(\gamma)$+jets processes. In addition, the ggHH (+jets) process represents a contamination in the qqHH signal region limiting the sensitivity to the vector boson fusion production mode. Therefore, an MVA approach is used to optimize the separation of the qqHH events from the continuum background and from the ggHH events. In particular, a BDT multi-classifier is trained using the simulated events to discriminate between three classes of processes: qqHH, ggHH, and $\gamma(\gamma) + jets$. Beside the variables already optimized for the background rejection from the ggHH signal region, additional variables exploiting the qqHH kinematic are included in the training. Such variables improve the separation of the two HH production modes and

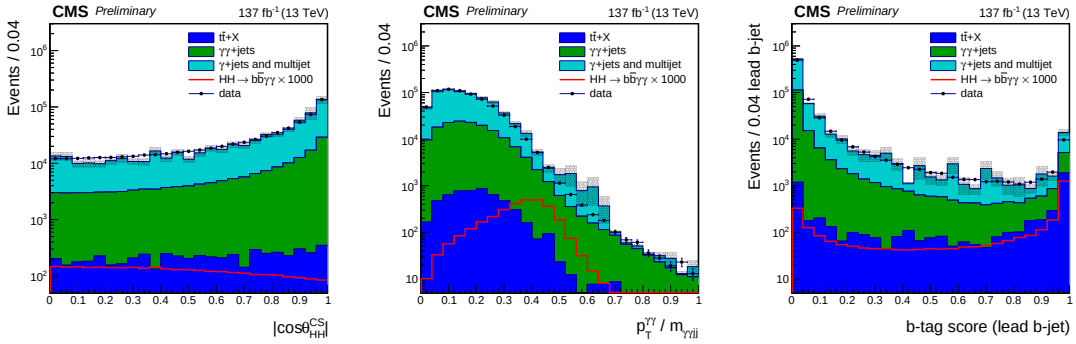


Figure 3.9: Distribution of $|\cos\theta^{\mathrm{CS}}_{\mathrm{HH}}|$ (*left*), $p_T^{\gamma\gamma}/m_{\gamma\gamma\mathrm{jj}}$ (*center*), and the b-tag score for the leading jet (*right*) of the selected HH candidates for the data as well as the simulated $\mathrm{t\bar{t}}$ +X, $\gamma\gamma$+jets, γ+jets, and multijet events excluding the signal region $120 < m_{\gamma\gamma} < 130$ GeV. The gray band represents the statistical uncertainty on the simulated events. The distribution of the simulated HH → bb̄γγ events is also shown with a red line.

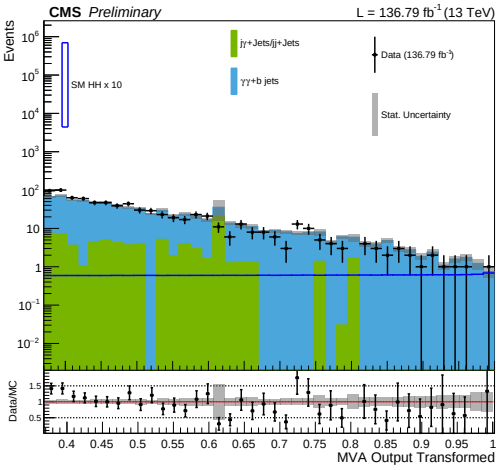


Figure 3.10: The distribution of the BDT output to discriminate the ggHH signal from the continuum background for the selected data and simulated events in the ggHH signal region (including the selection BDT score> 0.37).

also the continuum background rejection. In particular, the VBF jets are produced in opposite directions at large pseudorapidities recoiling against the diHiggs object. The VBF jets feature also a p_T typically larger than the background jets as well as a large dijet invariant mass. The additional variables are:

- The dijet invariant mass ${m_{\mathrm{jj}}}^{VBF}$
- The VBF jets kinematic described by their ${p_T/m_{\mathrm{jj}}}^{VBF}$ and pseudorapidities η_1^{VBF} and η_2^{VBF};
- Product and difference of pseudorapidity of the two VBF jets;
- Quark-gluon likelihood [64] of the two VBF jets to discriminate between jets originating from quarks and from gluons;
- Minimum angular distance between the VBF jets and the selected photons $\Delta R_{\mathrm{min}}(j^{VBF}, \gamma)$, or the selected b-jets $\Delta R_{\mathrm{min}}(j^{VBF}, \text{b-jet})$;
- The diHiggs kinematics described by the HH transverse momentum ${p_T}^{\mathrm{HH}}$ and the m_{X} variable.
- Centrality variable for the H→γγ and H→bb̄ candidates defined as:

$$C_{xx} = \exp\left[-\frac{4}{(\eta_1^{VBF} - \eta_2^{VBF})^2}\left(\eta_{xx} - \frac{\eta_1^{VBF} + \eta_2^{VBF}}{2}\right)^2\right] \quad \text{with } xx = \gamma\gamma,\ bb$$

  because the two Higgs boson candidates are typically produced centrally with respect to the VBF jets.

The distributions of some of the most discriminating additional variables for data and simulated events are shown in Fig. 3.11. The training is performed separately in two four-body mass categories because of the different qqHH kinematics (see Section 3.8). The four-body mass variable m_{X} is defined as:

$$m_{\mathrm{X}} = m_{\gamma\gamma jj} - m_{\gamma\gamma} - m_{\mathrm{jj}} + 250 \text{ GeV} \tag{3.3}$$

This definition reduces the impact of the jet and photon energy resolutions on the reconstructed four-body mass. Two m_{X} categories are defined by the selections $m_{\mathrm{X}} < 500$ GeV and $m_{\mathrm{X}} > 500$ GeV. In both categories the training is performed using as signal a combination of the SM qqHH sample and the BSM qqHH sample with $c_{2\mathrm{V}} = 0$. This strategy ensures a good sensitivity both to the SM and to the BSM hypotheses, especially to the $c_{2\mathrm{V}} = 0$ case. This BSM hypothesis is particularly relevant, as discussed in more detail in Section 3.8. The distribution of the BDT multiclassifier output relative to the qqHH class for the data events (qqHH-BDT score) as well as for the qqHH and single Higgs simulated events is shown in Fig. 3.12 for the two m_{X} categories.

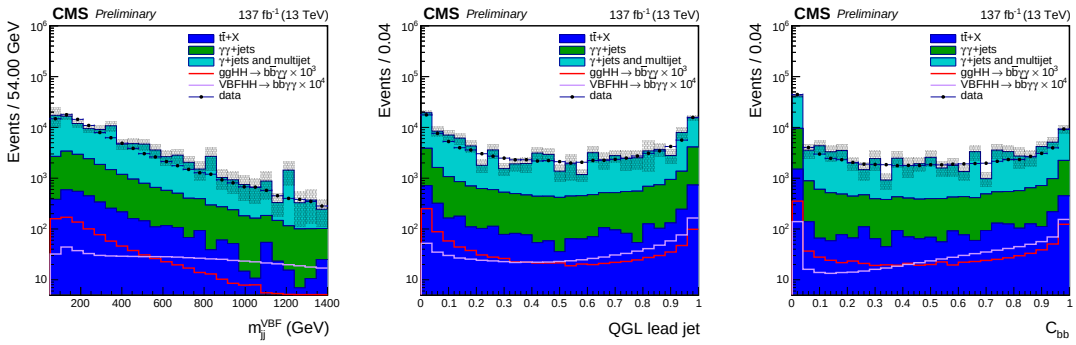


Figure 3.11: Distribution of m_{jj}^{VBF} (*left*), the quark-gluon likelihood (*center*), and the C_{bb} centrality variable for the selected qqHH candidates for the data as well as the simulated $\mathrm{t\bar{t}}$ +X, $\gamma\gamma$+jets, γ+jets, and multijet events, excluding the signal region $120 < m_{\gamma\gamma} < 130$ GeV. The gray band represents the statistical uncertainty of the simulated events. The distributions of the simulated ggHH and qqHH events are also shown with red and purple lines, respectively.

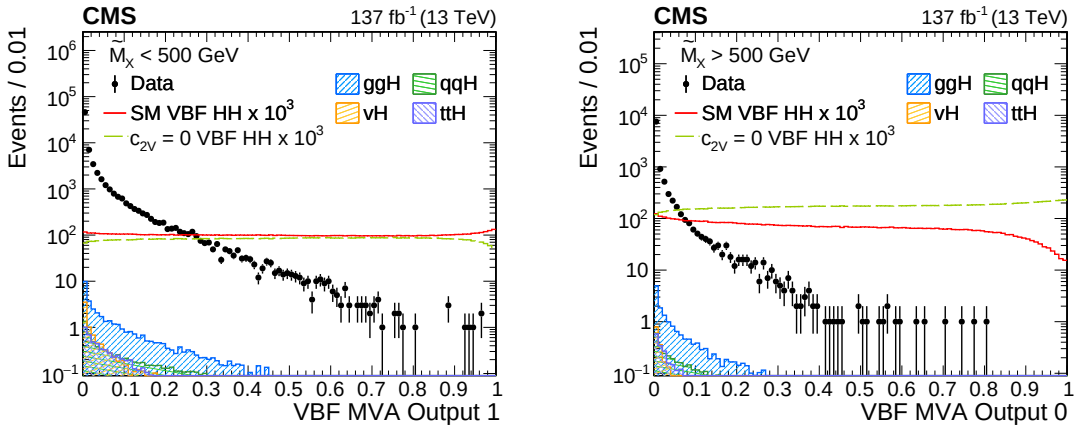


Figure 3.12: BDT multiclassifier output relative to the qqHH class for the data events as well as for the qqHH and single Higgs simulated events, for the $m_{\mathrm{X}} < 500$ GeV and $m_{\mathrm{X}} > 500$ GeV on the left and right, respectively.

3.8 qqHH-enriched categories

In the first place, the HH candidates are tested for the qqHH-enriched categories. The qqHH categories are given highest priority because the qqHH production has the smallest cross section among the signals considered, about 15 times smaller than the ggHH cross section. Therefore, a high selection efficiency on this signal is fundamental. At the same time, a selection on the qqHH-BDT score mitigates the ggHH (+jets) events migration to the qqHH-enriched categories.
The HH candidates passing the additional qqHH requirements are classified in two m_{X} categories to improve the sensitivity both to the SM signal and to the anomalous $c_{2\mathrm{V}}$ hyphothesis. The qqHH analysis is optimized to maximize the sensitivity to the SM qqHH signal and, at the same time, to the qqHH signal for anomalous $c_{2\mathrm{V}}$ values. In particular, the result of the ATLAS experiment for the HH → bb̄bb̄ channel [65] indicates that the experimental sensitivity is close to the exclusion of $c_{2\mathrm{V}} = 0$ at 95% confidence level. Therefore, the analysis is optimized to achieve the best sensitivity both for $c_{2\mathrm{V}} = 0$ and $c_{2\mathrm{V}} = 1$ (SM). For this reason, the value chosen as boundary of the two categories is $m_{\mathrm{X}} = 500$ GeV. As visible in Fig. 3.13, the $m_{\mathrm{X}} < 500$ GeV category is especially sensitive to the SM signal, while the $m_{\mathrm{X}} > 500$ GeV category is more populated by events produced with $c_{2\mathrm{V}} = 0$.
The selections on the ttH-score and on qqHH-BDT score are simultaneously optimized to maximize the expected significance on the qqHH signal. The expected significance is estimated as the sum in quadrature over the two categories of $S/\sqrt{B}$, where S and B are the expected qqHH signal and background yields in each category, respectively. Both S and B are estimated using MC events in the $122 < m_{\gamma\gamma} < 128$ GeV region. The signal considered for the optimization consist in the same mixture of the SM qqHH sample and the $c_{2\mathrm{V}} = 0$ qqHH sample used for the BDT training. The signal times the branching ratio is normalized to the expected excludable cross section estimated of 0.5 fb. The number of expected events in the sidebands (outside $115 < m_{\gamma\gamma} < 135$ GeV) of each category is required to be higher than 6. This number is found to be the minimum for a data-driven background modeling with a sufficient accuracy (see Section 3.11.3). This constraint is found to be the factor mostly controlling the optimization of the qqHH-BDT score boundaries. The significance is around its maximum for ttH-score values in the range $[0.2 - 0.3]$. Therefore, for simplicity the ttH-score > 0.26 selection is chosen, identically to the selection applied for the ggHH categories. The qqHH-BDT score boundaries for the categories definition are summarized in Table 3.5.

3.9 ggHH-enriched categories

The HH candidate events that do not pass the qqHH category selections are tested for the ggHH-enriched categories. The ggHH-BDT score is used to reject the background-like events and to classify the remaining events in three exclusive categories. The boundaries of the categories, along with the ttH-score selection, are

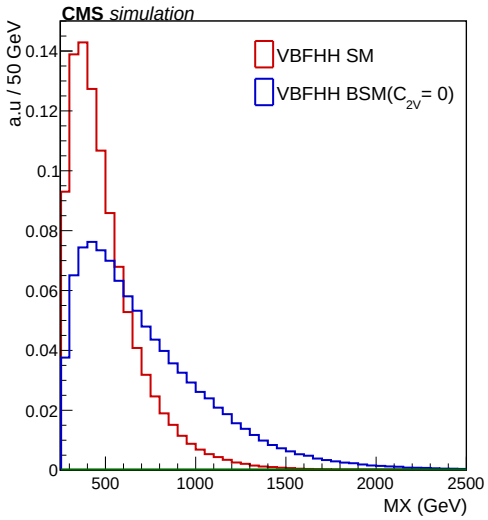


Figure 3.13: m_X distribution for the qqHH simulated events for $c_{2V} = 1$ and $c_{2V} = 0$ in red and blue, respectively.

simultaneously optimized with the same procedure used for the qqHH categories (Section 3.8). It was verified that the same ttH-score selection for all the three categories makes the optimization more robust without a significant worsening of the expected significance.
Within each of the three defined BDT score categories, four m_X exclusive categories are defined to improve the sensitivity to several BSM scenarios. As discussed in Section 1.2.1, the four-body mass is highly sensitive to the BSM benchmarks as well as to the anomalous κ_λ hypothesis. The distribution for the SM ggHH and the main backgrounds MC events is shown in Fig. 3.14. The optimization of the 3×4 m_X boundaries is performed simultaneously with the same procedure adopted for the MVA boundaries optimization.
Unlike the qqHH categories optimization, this optimization is performed with respect to the SM (ggHH) signal. It was verified that this does not penalize the BSM sensitivity thanks to the dense m_X categorization. The number of categories was also optimized repeating the procedure with a different number of BDT-score and m_X categories between one and four. The BDT-score and m_X selections for the categories is summarized in Table 3.5. As for the qqHH categories, the optimization of the boundaries for the high-purity categories is controlled by the constraint on the minimum number of expected background events in the sidebands.

3.10 tt̄H-enriched categories

The events failing both the qqHH and the ggHH selections are tested for the tt̄H categories. The tt̄H candidate selection and classification are the same developed

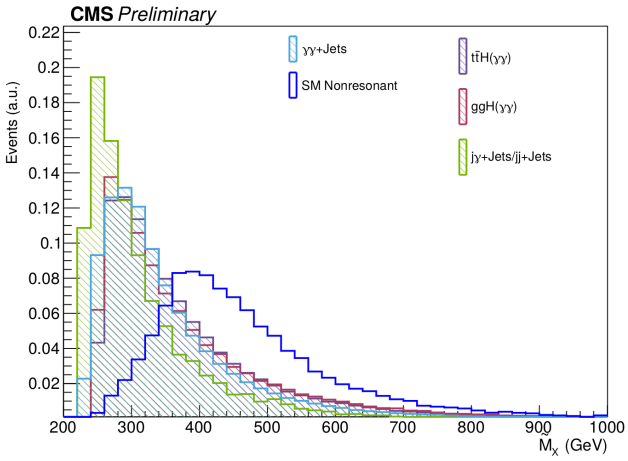


Figure 3.14: m_X distribution for the SM ggHH and the main background MC events. All the distributions are normalized to one.

Table 3.5: Optimized BDT-score and m_X selections for the HH categories. In all the categories the selection ttH-score > 0.26 is also applied.

Category	MVA	m_X (GeV)
qqHH CAT 0	0.52-1.00	>500
qqHH CAT 1	0.86-1.00	250-500
ggHH CAT 0	0.78-1.00	>600
ggHH CAT 1		510-600
ggHH CAT 2		385-510
ggHH CAT 3		250-385
ggHH CAT 4	0.62-0.78	>540
ggHH CAT 5		360-540
ggHH CAT 6		330-360
ggHH CAT 7		250-315
ggHH CAT 8	0.37-0.62	>585
ggHH CAT 9		375-585
ggHH CAT 10		330-375
ggHH CAT 11		250-330

and optimized for the $t\bar{t}H(\gamma\gamma)$ analysis with the Run 2 dataset [23]. A summary of the $t\bar{t}H$ categories definition will be given in the next sections. Should the reader be interested, more details can be found in Ref. [23].
The events are required to have a $H\rightarrow\gamma\gamma$ candidate with two photons with $photon\ ID > -0.7$. Two sets of mutually exclusive categories are defined to target two $t\bar{t}H$ signals: $t\bar{t}H$ leptonic and $t\bar{t}H$ hadronic. The $t\bar{t}H$ leptonic signal consists in $t\bar{t}H$ events where at least one W boson, coming from the top or antitop quark decays, decays leptonically. Instead, the $t\bar{t}H$ hadronic signal consists in $t\bar{t}H$ events where both the W bosons decay hadronically. A set of orthogonal selections, presented in the next Sections, is applied to the events to find either a $t\bar{t}H$ leptonic candidate, or a $t\bar{t}H$ hadronic candidate. MVA classifiers are used to maximize the separation of the $t\bar{t}H$ events from the backgrounds. The MVA outputs are used to reject the background-like events and to classify the remaining events in exclusive categories. Altough no differences in the definition of the $t\bar{t}H$ categories are introduced with respect to Ref. [23], small differences in the event reconstruction and in the analysis workflow induce variations in the expected yields:

- Vertex choice criterion: in this analysis, the standard $H\rightarrow\gamma\gamma$ vertex choice (Section 3.5.1) is applied while in Ref. [23] the vertex choice is based on the standard CMS criterion. In such a criterion, the vertex with the highest *sumpt*2 is selected. The usage of the $H\rightarrow\gamma\gamma$ vertex causes a loss of efficiency on the $t\bar{t}H$ signal selection of about 3% in the hadronic categories and 5% in the leptonic categories. Although it was verified through MC simulations that the selected vertex is within 1 cm from the true vertex in more than 99% of the cases, the impact of a different vertex choice within 1 cm slightly affects the kinematics of the reconstructed objects. In turn, this effect slightly reduces the performance of the BDT classifiers.

- Event migration to the HH categories: the loss of efficiency on the $t\bar{t}H$ signal amounts to about 12% and 4% in the $t\bar{t}H$ hadronic and leptonic categories, respectively.

Therefore, the overall loss of efficiency on the $t\bar{t}H$ events selection is of about 15% and 9% in the $t\bar{t}H$ hadronic and leptonic categories, respectively, with respect to Ref. [23]. A re-optimization of the $t\bar{t}H$ categories was not performed because a variation of the expected number of background events in the categories is also observed such that the $S/\sqrt{B}$ variation is at most 4%.

3.10.1 $t\bar{t}H$ leptonic categories

The $t\bar{t}H$ leptonic categories requires the event to have at least one isolated electron (or muon) with $|\eta|< 2.4$ and $p_T > 10(5)$ GeV, and at least one jet with $p_T > 25$ GeV. A BDT classifier is trained using the simulated samples to discriminate the $t\bar{t}H$ leptonic events from the main backgrounds including $\gamma(\gamma)$+jets, $t\bar{t} + (\gamma)(\gamma)$, $Z + \gamma$,

and $W + \gamma$. The variables used for the training include kinematic properties of the reconstructed objects (photons, leptons, and jets), object identification variables (photonID values of the two selected photons and b-tagging scores of jets), and global variables of the event such as the jet and lepton multiplicity. In the signal region, the background is dominated by $t\bar{t}\gamma\gamma$ events. For this reason, an additional DNN classifier is trained to separate this specific background from the $t\bar{t}H$ signal. The DNN output score is used as an additional input to the BDT classifier. The output score of the BDTs is used to reject the background-like events, and to classify the remaining events in four categories. The boundaries of the $t\bar{t}H$ leptonic categories were optimized along with the boundaries of the $t\bar{t}H$ hadronic categories to maximize the total expected significance of the $t\bar{t}H$ signal. The BDT output distribution as well as the optimized boundaries are shown in Fig. 3.15 (left).

3.10.2 $t\bar{t}H$ hadronic categories

The $t\bar{t}H$ hadronic categories require the event to have exactly zero isolated electron and muons (defined as for the $t\bar{t}H$ leptonic categories), at least 3 jets, and at least one b-tagged jet. These selections ensure the orthogonality with respect to the $t\bar{t}H$ leptonic categories. The background rejection strategy is the same as the one developed for the $t\bar{t}H$ leptonic signal. A BDT is trained to separate the $t\bar{t}H$ hadronic signal from the main backgrounds, which are the same of the $t\bar{t}H$ leptonic categories, with different relative proportions. In particular, the background for the $t\bar{t}H$ hadronic is dominated by the γ+jets process. In order to improve the rejection of such a background, the BDT training is performed including the data events not passing the photon ID< -0.7 selection instead of the γ+jets MC simulation. It was verified through MC simulation that the events failing the photon ID selection are mostly γ+jets with a smaller fraction of multi-jet events. The BDT performance improvement comes from the larger size of the training dataset and from the better background modeling. An additional DNN classifier is also trained to separate the $t\bar{t}\gamma\gamma$ and $\gamma\gamma$+jets processes, dominating the background in the signal-like region, from the $t\bar{t}H$ signal. The DNN classifier score is used as an additional input for the BDT.
As for the $t\bar{t}H$ leptonic categories, the BDT output score is used to reject the background-like events, and to classify the remaining events in four categories. The category boundaries were optimized along with the $t\bar{t}H$ leptonic categories to maximize the total expected significance of the $t\bar{t}H$ signal. The BDT output distribution as well as the optimized boundaries are shown in Fig. 3.15 (right).

3.10.3 Sorting of the HH and $t\bar{t}H$ categories

A detailed study was performed to optimize the order of the categories selections to maximize for each category the selection efficiency of the target signal events. The following order of mutually exclusive categories was found to be optimal (in order

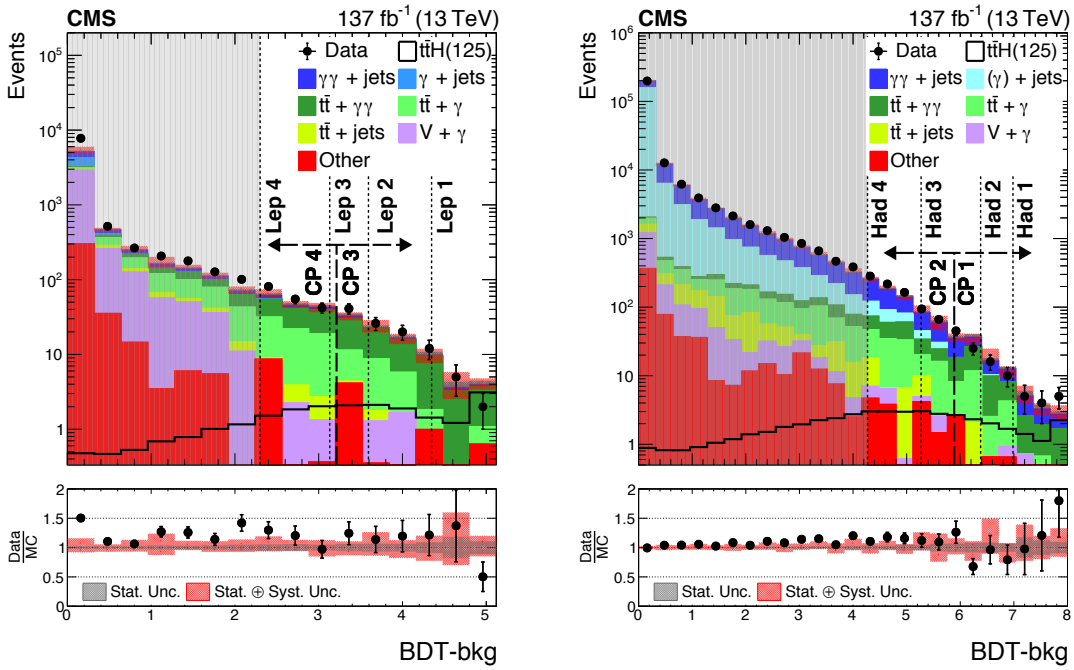


Figure 3.15: Distribution of the BDT output for selected data and simulation events as well as category boundaries for the tt̄H leptonic (*left*) and tt̄H hadronic (*right*) signals. Figures from [23].

of priority from the highest to the smallest):

1. qqHH-enriched categories
2. ggHH-enriched categories
3. tt̄H leptonic categories
4. tt̄H hadronic categories

The purpose of the ordering optimization is to:

- Maximize the selection efficiency of the HH signal in the dedicated HH categories, as this process is expected to be the most sensitive to the κ_λ parameter. As shown in Table 3.6, the events should be tested for the HH categories before the tt̄H categories otherwise about 34% of the HH events would migrate to the tt̄H hadronic categories.
- Minimize the impact on the other categories (in terms of selection efficiency of the targeted processes). As shown in Table 3.6, once the HH category has the highest priority, the migration of tt̄H events from tt̄H hadronic and tt̄H leptonic categories to HH categories is of about 12% and 4%, respectively.

The expected composition of the categories, estimated through simulation is shown in Fig. 3.16, while the corresponding numbers are presented in Table 3.7. The figure shows also the expected $S/(S+B)$ in $\pm 1\sigma_{\text{eff}}$ which is defined as the smallest interval containing 68% of the signal. The signal yield S is referred to the target process of a given category, e.g., qqHH signal yield for the qqHH-enriched categories. The background B is the sum of the expected Higgs boson background processes and of the expected continuum background estimated through the data-driven technique presented in Section 3.11.3. A $S/(S+B)$ value up to about 15% is achieved in the ggHH-enriched categories. The tt̄H contamination in the ggHH-enriched categories is maintained under control thanks to the selection on the ttH-score. In the tt̄H enriched categories a very good $S/(S+B)$ is achieved, which is found consistent with Ref. [23].

Table 3.6: Efficiency of the selection of each category for its specific targeted process for different priority orders. For the first line, corresponding to HH, the relative yields are quoted with respect to the "High HH priority" configuration, while for the other lines the relative yields are quoted with respect to the "No HH cat." configuration. The study was performed on the 2017 MC simulation.

Process	**Category**	**Relative yields**			
		No HH cat. 1. tt̄H lept. 2. tt̄H hadr.	**Low HH priority** 1. tt̄H lept. 2. tt̄H hadr. 3. HH	**Medium HH priority** 1. tt̄H lept. 2. HH 3. tt̄H hadr.	**High HH priority** 1. HH 2. tt̄H lept. 3. tt̄H hadr.
HH → bb̄γγ	HH	-	0.66	0.99	1
tt̄H(γγ)	tt̄H lep.	1	1	1	0.96
tt̄H(γγ)	tt̄H hadr.	1	1	0.88	0.88

3.11 Measurement of the HH production and interpretation

The data analysis aims to determine the compatibility of the experimental observation with the "signal + background" hypothesis against the "background only" hypothesis, or viceversa. In case a signal is observed, the data are used to measure the corresponding signal strength μ defined as:

$$\mu = \frac{\sigma_{obs}}{\sigma_{SM}} \tag{3.4}$$

A μ value compatible, within the uncertainties, with one proves the compatibility of the measured cross section with the SM prediction (σ_{SM}). In case the signal is

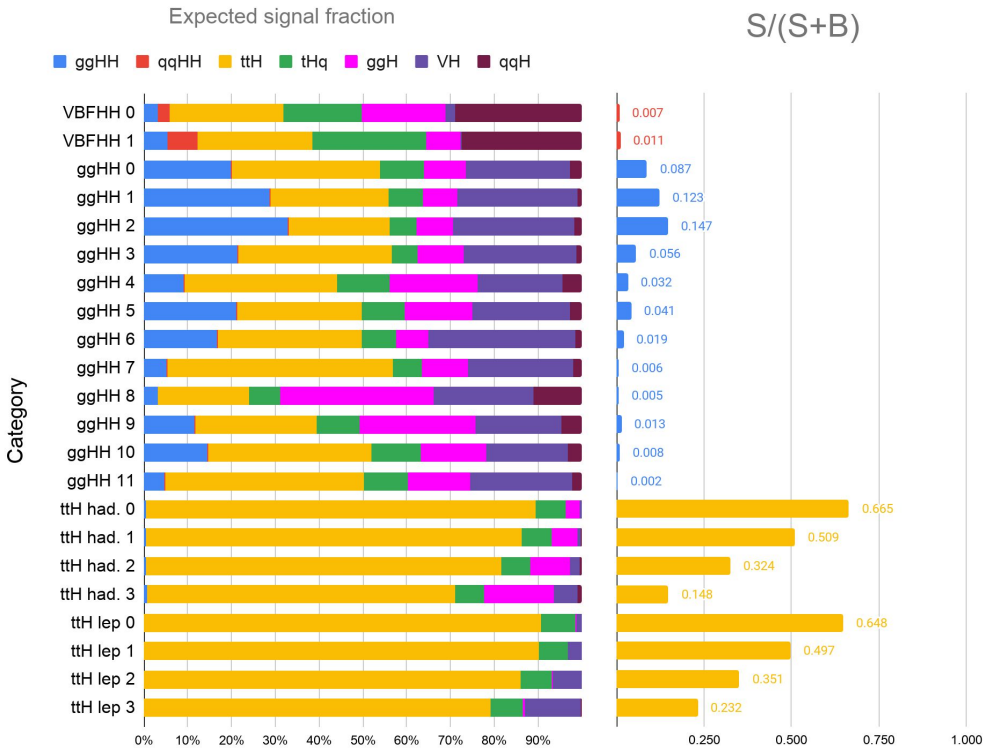


Figure 3.16: *Left:* Expected categories composition in terms Higgs boson processes. *Right:* Expected $S/(S+B)$ in $\pm 1\sigma_{\text{eff}}$ for each category. S is referred to the Higgs boson process target of each category, and B is the sum of the expected Higgs boson background processes and of the expected continuum background.

Table 3.7: Expected number of signal events per category and breakdown of contribution by production mode as well as the expected number of background in events per GeV at 125 GeV.

Category	Total	ggHH	qqHH	$t\bar{t}H$	tHq	ggH	VH	qqH	bkg(GeV^{-1})
qqHH cat. 0	0.44	3.3%	2.5%	26.0%	18.1%	18.9%	2.3%	28.9%	0.18
qqHH cat. 1	0.10	5.4%	6.9%	26.3%	25.8%	7.9%	0.2%	27.5%	0.093
ggHH cat. 0	1.23	19.9%	0.2%	33.7%	10.1%	9.5%	23.8%	2.7%	0.32
ggHH cat. 1	0.67	28.7%	0.2%	27.0%	7.7%	7.8%	27.4%	1.1%	0.18
ggHH cat. 2	1.05	32.8%	0.2%	23.0%	6.3%	8.1%	27.8%	1.7%	0.26
ggHH cat. 3	0.21	21.3%	0.3%	35.0%	5.9%	10.5%	25.9%	1.2%	0.12
ggHH cat. 4	1.54	9.0%	0.2%	34.9%	11.9%	20.1%	19.4%	4.4%	0.56
ggHH cat. 5	1.78	21.1%	0.2%	28.5%	9.7%	15.4%	22.3%	2.7%	1.48
ggHH cat. 6	0.21	16.6%	0.3%	33.0%	7.8%	7.4%	33.4%	1.6%	0.32
ggHH cat. 7	0.21	5.2%	0.3%	51.3%	6.7%	10.7%	23.9%	2.0%	0.33
ggHH cat. 8	2.91	3.2%	0.1%	20.7%	7.0%	35.2%	22.6%	11.1%	3
ggHH cat. 9	4.70	11.6%	0.2%	27.6%	9.9%	26.6%	19.4%	4.8%	7.4
ggHH cat. 10	0.94	14.5%	0.3%	37.1%	11.3%	15.1%	18.5%	3.2%	3.2
ggHH cat. 11	0.97	4.6%	0.4%	45.4%	9.9%	14.2%	23.2%	2.3%	4.4
$t\bar{t}H$ had. cat. 0	4.83	0.4%	$< 0.1\%$	89.1%	6.8%	3.1%	0.4%	0.1%	0.31
$t\bar{t}H$ had. cat. 1	3.48	0.4%	$< 0.1\%$	85.7%	6.9%	6.0%	0.8%	0.1%	0.45
$t\bar{t}H$ had. cat. 2	9.55	0.5%	$< 0.1\%$	81.0%	6.8%	9.0%	2.2%	0.4%	2.7
$t\bar{t}H$ had. cat. 3	10.69	0.7%	$< 0.1\%$	70.5%	6.5%	15.9%	5.4%	1.1%	7.6
$t\bar{t}H$ lep. cat. 0	5.17	$< 0.1\%$	$< 0.1\%$	93.0%	6.0%	0.1%	0.9%	$< 0.1\%$	0.29
$t\bar{t}H$ lep. cat. 1	4.51	$< 0.1\%$	$< 0.1\%$	90.8%	6.1%	$< 0.1\%$	2.9%	$< 0.1\%$	0.65
$t\bar{t}H$ lep. cat. 2	3.17	$< 0.1\%$	$< 0.1\%$	86.1%	6.9%	0.2%	6.6%	0.1%	0.87
$t\bar{t}H$ lep. cat. 3	5.26	0.1%	$< 0.1\%$	79.1%	7.4%	0.5%	12.8%	0.1%	2.4

not observed, the data are used to set an upper limit on its cross section. Multiple signals are tested: the ggHH signal (μ_{ggHH}), the qqHH signal (μ_{qqHH}), and the inclusive HH production (ggHH + qqHH) signal ($\mu_{\text{ggHH}} = \mu_{\text{qqHH}} = \mu_{\text{HH}}$). Instead, a measurement of the t$\bar{\text{t}}$H signal strength $\mu_{\text{t}\bar{\text{t}}\text{H}}$ is out of the scope of this work because no improvement on the t$\bar{\text{t}}$H sensitivity are expected with respect to the measurement reported in Ref. [23] using the same dataset.
Alternatively, the data can be interpreted in terms of Higgs boson coupling modifiers (κ_λ, κ_t, c_V, $c_{2\text{V}}$). Due to the limited number of considered Higgs production and decay channels, at most two coupling constants are measured simultaneously, fixing the other couplings to the SM prediction. The measurement can be performed under the assumption of a SM HH signal, or no HH signals. The statistical analysis adopts a maximum likelihood method described in Section 3.11.1 while the signal and background modeling are described in Section 3.11.2 and Section 3.11.3, respectively.

3.11.1 Likelihood definition and nuisance parameters

A likelihood function is used as test statistics. The likelihood is split into three terms:

$$\mathcal{L} = \mathcal{L}_{\text{HH}} \cdot \mathcal{L}_{\text{t}\bar{\text{t}}\text{H}} \cdot p(\theta|\tilde{\theta}) \tag{3.5}$$

where $\mathcal{L}_{\text{HH}}$ and $\mathcal{L}_{\text{t}\bar{\text{t}}\text{H}}$ are the likelihood functions corresponding to the HH and t$\bar{\text{t}}$H enriched categories, respectively. A different approach is used to define the likelihood in the two sets of categories to achieve the best sensitivity. Defined θ as the vector of all the nuisance parameters, $p(\theta|\tilde{\theta})$ is the distribution of θ given the true values (Bayesian interpretation) $\tilde{\theta}$. For the nuisance description, the frequentist approach common to the CMS and ATLAS experiment is used [66]. In particular, the nuisance pdfs $p(\theta|\tilde{\theta})$ are reinterpreted using the Bayes theorem as posteriors arising from some auxiliary measurement:

$$p(\theta|\tilde{\theta}) = p(\tilde{\theta}|\theta)p(\theta) \tag{3.6}$$

The a-priori $p(\theta)$ distribution is assumed to be uniform, thus negligible in the likelihood maximization. All the nuisances considered for this work are assumed to be distributed as a log-normal.
The $\mathcal{L}_{\text{HH}}$ factor is built exploiting the resonant nature of the HH signal in the m_{jj} and $m_{\gamma\gamma}$ distributions. Given the low statistic regime, the $\mathcal{L}_{\text{HH}}$ function is built as an unbinned likelihood. Therefore, $\mathcal{L}_{\text{HH}}$ is defined for each analysis category as:

$$\begin{aligned}\mathcal{L}_{\text{HH}} = & k^{-1} \prod_{i \in \text{events}} \Big[\sum_{\substack{j=\text{ggHH,}\\ \text{qqHH}}} \mu_j S_j f_j(m^i_{\gamma\gamma}, m^i_{\text{jj}}) + \sum_{\substack{j=\text{t}\bar{\text{t}}\text{H, tHq,}\\ \text{ggH, VH,}\\ \text{qqH}}} S_j f_j(m^i_{\gamma\gamma}, m^i_{\text{jj}}) + \\ & + B f_B(m^i_{\gamma\gamma}, m^i_{\text{jj}}|\theta) \Big] \cdot \exp\Big(\sum_{\substack{j=\text{ggHH,}\\ \text{qqHH}}} \mu_j S_j + \sum_{\substack{j=\text{t}\bar{\text{t}}\text{H, tHq,}\\ \text{ggH, VH,}\\ \text{qqH}}} S_j + B \Big)\end{aligned} \tag{3.7}$$

where k is the total number of observed events, S_j is the number of events predicted by the SM for the j-th single or double Higgs process, and f_j is the corresponding two dimensional $(m_{\gamma\gamma}, m_{\mathrm{jj}})$ parametric pdf. B is the expected number of continuum background events and f_B is the corresponding parametric pdf. For simplicity, the dependence of the S_j, f_j, and f_B quantities from the nuisance parameters θ is omitted from Eq. 3.7. The correlation between the $m_{\gamma\gamma}$ and m_{jj} variables is found to be negligible both for the signals and the backgrounds (see Sections 3.11.2 and 3.11.3). Therefore, the two dimensional models can be factorized as:

$$f_j(m_{\gamma\gamma}, m_{\mathrm{jj}}) = f_j^{\gamma\gamma}(m_{\gamma\gamma}) f_j^{\mathrm{jj}}(m_{\mathrm{jj}}) \tag{3.8}$$

where $f_j^{\gamma\gamma}$ and f_j^{jj} are the one dimensional $m_{\gamma\gamma}$ and m_{jj} models for the j-th single or double Higgs process.
For the measurement of the coupling modifiers, the signal strengths are fixed to one while the number of expected events for the single and double Higgs processes S_j is expressed as a function of the (κ_λ, κ_t, c_V, $c_{2\mathrm{V}}$) parameters. The procedure is described in detail in Appendix B. In particular, the SM and BSM ggHH and qqHH samples are properly combined to provide a per-category description of the signals rate variations. For the single Higgs processes a parametric description of the total cross section variation as a function of the coupling modifiers is used for all the categories. Although modifications of the p_T spectrum of the single Higgs processes, especially for tt̄H, are expected in case of anomalous κ_λ values, an explicit p_T categorization is not performed. Since the p_T distribution is found to be similar in all the categories, the inclusive cross section variation provides sufficient accuracy for the description of the anomalous coupling effects. A future extension of this work can be a p_T classification of the single Higgs events to improve the sensitivity to anomalous couplings. The H→γγ and H→bb̄ branching ratios variations for anomalous couplings are also considered.
The $\mathcal{L}_{\mathrm{t\bar{t}H}}$ term of Eq. 3.5 is included in the likelihood only for the couplings measurements because the impact of the tt̄H-enriched categories is found to be negligible for the HH searches. In the tt̄H-enriched categories a one dimensional binned analysis of the $m_{\gamma\gamma}$ distribution is performed. Therefore, the $\mathcal{L}_{\mathrm{t\bar{t}H}}$ factor is built as:

$$\mathcal{L}_{\mathrm{t\bar{t}H}} = \prod_{i \in \mathrm{bin}} \frac{(s_i + b_i)^{n_i}}{n_i!} e^{-(s_i + b_i)} \tag{3.9}$$

where the i index runs over the $m_{\gamma\gamma}$ bins, s_i and b are the expected number of events in the i-th bin for the single+double Higgs production and background processes, respectively. Those quantities are computed as:

$$\begin{aligned} s_i &= \sum_{\substack{j=\mathrm{ggHH,\ qqHH,}\\ \mathrm{t\bar{t}H,\ tHq,}\\ \mathrm{ggH,\ VH,}\\ \mathrm{qqH}}} S_j \int_{\mathrm{bin}\ i} f_j^{\gamma\gamma}(m_{\gamma\gamma})\ \mathrm{d}m_{\gamma\gamma} \\ b_i &= B \int_{\mathrm{bin}\ i} f_B^{\gamma\gamma}(m_{\gamma\gamma})\ \mathrm{d}m_{\gamma\gamma} \end{aligned} \tag{3.10}$$

For simplicity, the dependence of the S_j, f_j, and f_B quantities from the nuisance parameters θ is also omitted from Eq. 3.9. As for the HH categories, the numbers of expected events S_j are expressed as a function of the coupling parameters for the Higgs boson couplings measurements.

Estimation of the parameters of interest

Let a generic μ be the parameter of interest. The likelihood estimators of μ and θ, namely $\hat{\mu}$ and $\hat{\theta}$, are the values simultaneously maximizing the likelihood function, or equivalently the values minimizing the negative logarithm of the likelihood function (log-likelihood L). The latter is much easier to compute from the algorithmic point of view. In order to estimate the uncertainty on $\hat{\mu}$, a "profile likelihood" is defined as:

$$\mathcal{L}_{prof}(\mu) = -2 \log \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \tag{3.11}$$

Where $\mathcal{L}$ is the likelihood defined in Eq. 3.5, $\hat{\hat{\theta}}$ represents the set of nuisance values maximizing $\mathcal{L}$ for a given μ and the denominator is maximized over the full parameter space. The 68% and 95% confidence intervals on μ correspond to the μ values satisfying the condition $\mathcal{L}_{prof}(\mu) < 1$ and $\mathcal{L}_{prof}(\mu) < 3.84$, respectively.

Hypothesis testing to set an upper limit

In case no evidence of the double Higgs production is found with data, an upper limit can be set on the ggHH, qqHH, and HH cross sections. To do that, the following test statistic is defined:

$$q_\mu = -\log \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad \text{with } 0 \leq \hat{\mu} \leq \mu \tag{3.12}$$

where $\mathcal{L}$ is the likelihood defined in Eq. 3.5, $\hat{\mu}$ and $\hat{\theta}$ are the signal strength and the nuisances values maximizing $\mathcal{L}$, while $\hat{\hat{\theta}}$ is the set of nuisance values maximizing $\mathcal{L}$ for a given value of μ. Depending on the cross section of interest, the μ variable is referred either to the ggHH, or qqHH, or the inclusive HH signal. The constraint $\hat{\mu} \geq 0$ is to avoid the unphysical situation of negative signals, while the constraint $\mu \geq \hat{\mu}$ is required to avoid to use upward fluctuations of the data against the signal hypothesis. The modified frequentist criterion [67] is adopted for the limit extraction. The level of disagreement of the observed data with a given hypothesis is quantified through a "p-value" which is the probability to obtain results worse than or equal to the one observed under the given hypothesis. In particular, given the observed value of the test statistic q_μ^{obs}, two p-values p_μ and p_b can be derived

for the signal plus background and background only hypotheses respectively:

$$\begin{aligned} p_\mu &= Prob(q_\mu > q_\mu^{obs} | \text{signal+background}) \\ 1 - p_b &= Prob(q_\mu > q_\mu^{obs} | \text{background only}) \end{aligned} \tag{3.13}$$

Such p-values are computed using the asymptotic properties of the test statistics [68]. This avoids the computationally expensive procedure of the MC toy generation to explicitly derive the q_μ distributions. The p_μ value is not used directly for the limit extraction because it is not sufficiently robust against background under-fluctuations. Such under-fluctuations could lead to exclude small values of μ even if the sensitivity to the signal would not be sufficient. In order to prevent for that effect, the $CL_s(\mu)$ quantity is defined as:

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} \tag{3.14}$$

A signal strength μ is said to be excluded at a confidence level (CL) α if $CL_s(\mu) < 1 - \alpha$. The value commonly chosen for α is 95%. The limit on μ is in fact a limit on the cross section normalized to the corresponding SM prediction.

Hypothesis testing to quantify an excess

The following test statistics is defined to quantify the excess of events:

$$q_0 = -\log \frac{\mathcal{L}(\mu, \hat{\hat{\theta}})\Big|_{\mu=0}}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad \text{with } \hat{\mu} > 0 \tag{3.15}$$

This is in fact the likelihood computed for the background-only hypothesis normalized by the likelihood value for the best fit point. Given the observed value of the test statistic q_0^{obs}, the p-value is then defined as:

$$p_0 = Prob(q_0 \geq q_0^{obs} | \text{background only}) \tag{3.16}$$

This is the probability that a background fluctuation gives an excess larger than or equal to the observed one. For a more direct interpretation, the p_0 probability is converted to a Z significance by expressing it as a one-sided Gaussian integral:

$$p_0 = \int_Z^{+\inf} \frac{1}{2\pi} e^{-x^2/2} \, dx \tag{3.17}$$

It is customary to consider a significance larger than 3σ as an evidence, and a significance larger than 5σ as an observation of a signal above the background.

3.11.2 Modeling of Higgs processes

For each analysis category and each single/double Higgs production mechanism, the MC simulations are used to derive the expected number of events (S_j) and to model the $m_{\gamma\gamma}$ distribution ($f_j^{\gamma\gamma}$), and, for the HH categories, the m_{jj} distribution (f_j^{jj}).
The $m_{\gamma\gamma}$ peak is modeled as a sum of up to five gaussian functions. Examples of the $m_{\gamma\gamma}$ models for the t$\bar{\text{t}}$H process in the t$\bar{\text{t}}$H categories, and of the HH processes in the HH categories are visible in Fig. 3.17 and 3.18, respectively.
For the HH categories, the m_{jj} pdf parametrization depends on the specific production mechanism:

- The m_{jj} distribution for the qqHH and ggHH processes is modeled with a double-sided Crystal Ball (CB) function which is a CB function [69] with two independent exponential tails instead of one, as shown in Fig. 3.19. This function is found to provide an adequate description of the m_{jj} peak with its left and right tails related to the jet energy resolution.

- The m_{jj} distribution for the VH process, consisting in a peak in correspondence of the vector bosons masses, is modeled through a standard CB function. This function, with a lower number of free parameter than a double-sided CB, provides the required robustness and accuracy for the VH modeling. The simulated VH events are in fact affected by higher statistical uncertainties because of the limited number of selected events.

- The m_{jj} distributions for the ggH and qqH processes have a smooth falling shape, thus they are parametrized by Bernstein polynomials.

- The t$\bar{\text{t}}$H and tHq events kinematics feature a m_{jj} distribution peaking at about 110-120 GeV. Thus, a gaussian function is used to model their m_{jj} distributions.

The functions parameters are determined for each category through a fit to the selected simulated events. The final signal model is built as the product of the obtained $m_{\gamma\gamma}$ and m_{jj} distributions.
The number of simulated events for the single-Higgs production mechanisms in the HH signal regions is limited, especially for the ggH, VH, and qqH processes. Therefore, the simulated events of the three data-taking years within the same BDT-score category are merged together, and a common m_{jj} model is extracted to improve the model accuracy. It was verified that the m_{jj} models across different m_{X} categories and different years are compatible within the uncertainties. Examples of the m_{jj} models for the high BDT score category are visible in Fig. 3.20. In order to improve the accuracy of the $m_{\gamma\gamma}$ modeling of the ggH, VH, and qqH processes in the HH signal regions, in case the number of MC entries is less than 500, the used $m_{\gamma\gamma}$ model is the same one used for the t$\bar{\text{t}}$H process in that category. Then

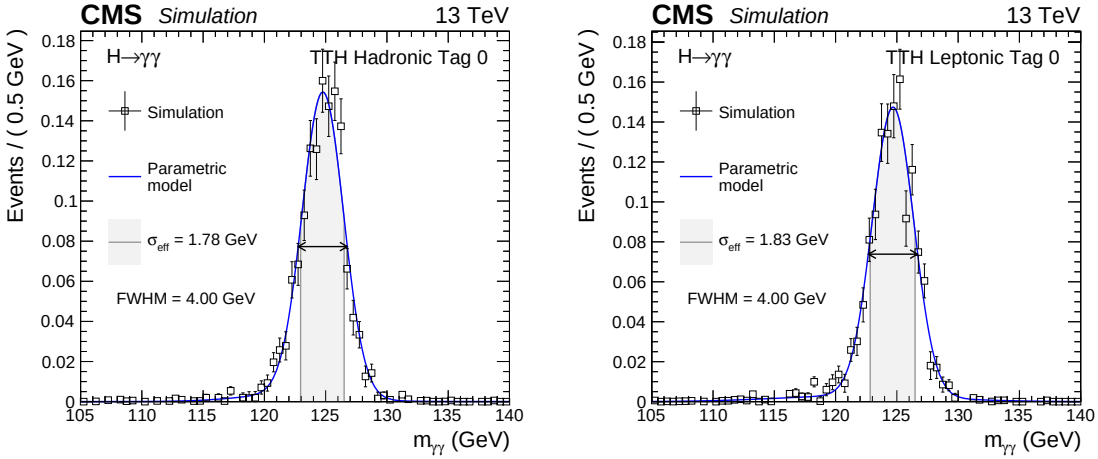


Figure 3.17: $m_{\gamma\gamma}$ signal modeling for the $t\bar{t}H$ process in the 2018 dataset for the high-BDT score hadronic (left) and leptonic (right) category.

the normalization is set according to the expected yield for that specific process. The assumption of no-correlation between the $m_{\gamma\gamma}$ and m_{jj} variables hypothesis is verified comparing the two dimensional ($m_{\gamma\gamma}$,m_{jj}) distribution of the simulated events, visible in Fig. 3.21 for the simulated ggHH events from the 2018 dataset, to the product of the derived m_{jj} and $m_{\gamma\gamma}$ models. The difference of the two distributions is found to be negligible within the statistical uncertainties related to the expected number of signal events.

3.11.3 Modeling of the continuum background

The continuum background is modeled in each category through a fit of the $m_{\gamma\gamma}$ (and m_{jj}) distribution of the selected data events. Since the parametrization of the underlying model is not known, a specific method called "envelope method" [70] is used to choose the background parametrization estimating also the related uncertainty. The envelope method considers the choice of the background functional form as an additional (discrete) nuisance parameter to be included in the likelihood definition. In particular, an integer number called "envelope index" is used to select a specific functional form among the set of given parametrizations. Therefore, the negative log-likelihood is minimized with respect to all the nuisance parameters including the envelope index. The impact of this additional nuisance parameter is a broadening of the profile likelihood of Eq. 3.11 corresponding to an increase of the uncertainty in the parameter estimation. This is expected because a nuisance parameter corresponds to a loss of information in the measurement.
For a correct usage of the envelope method, a proper choice of the set of functions is crucial. The set of functions has to provide a negligible bias in the estimation of the parameter of interest and a consistent estimation of the corresponding uncertainty (coverage). The studies presented in Ref. [70] show that for an exponentially falling

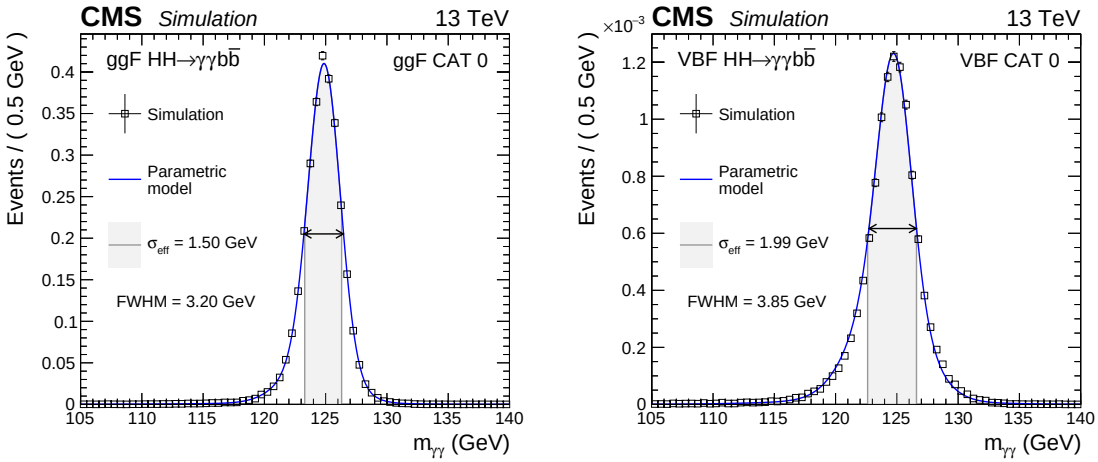


Figure 3.18: $m_{\gamma\gamma}$ modeling for the ggHH and qqHH process on the left and right, respectively, for the high m_X and high BDT score category in the 2018 dataset.

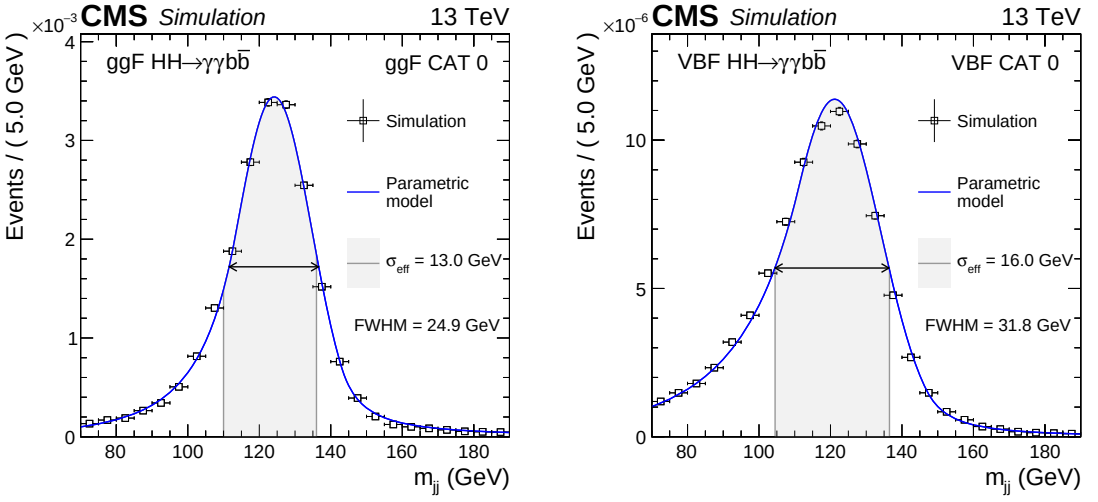


Figure 3.19: m_{jj} modeling for the ggHH and qqHH process on the left and right, respectively, for the high m_X and high BDT score category in the 2018 dataset.

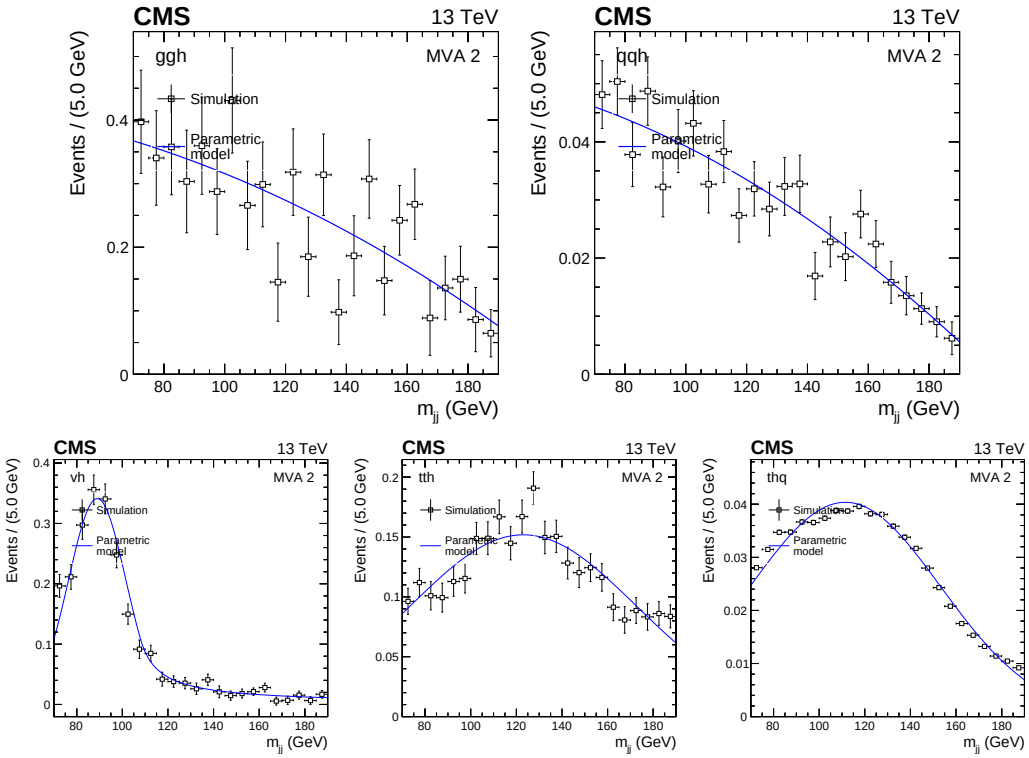


Figure 3.20: m_{jj} modeling for the ggH (top-left), qqH (top-right), VH (bottom-left), tt̄H (bottom-center), and tHq (bottom-right) processes, in the ggHH-enriched category with the highest BDT score value.

background such as the $m_{\gamma\gamma}$ background distribution in the H→γγ analysis, a good coverage is provided by the following set of function families:

$$\text{Power law sum:} f(x) = p_0 x^{p_1} + p_2 x^{p_3} + p_4 x^{p_5} + ...$$
$$\text{Exponential sum:} f(x) = p_0 e^{p_1 x} + p_2 e^{p_3 x} + p_4 e^{p_5 x} + ...$$
$$\text{Laurent series:} f(x) = \sum_{i=0}^{N} p_i / x^i$$
$$\text{Polynomial:} f(x) = \sum_{i=0}^{N} p_i x^i$$

A second important aspect is how to compare functions with different degrees of freedom in terms of data agreement. By construction, a function with a higher number of free parameters within the same family is able to better describe the data, thus the likelihood minimization will select it. However, it is more sensitive to the data fluctuations. Therefore, a correction to the likelihood penalizing the higher order functions is defined to make the method robust against the background

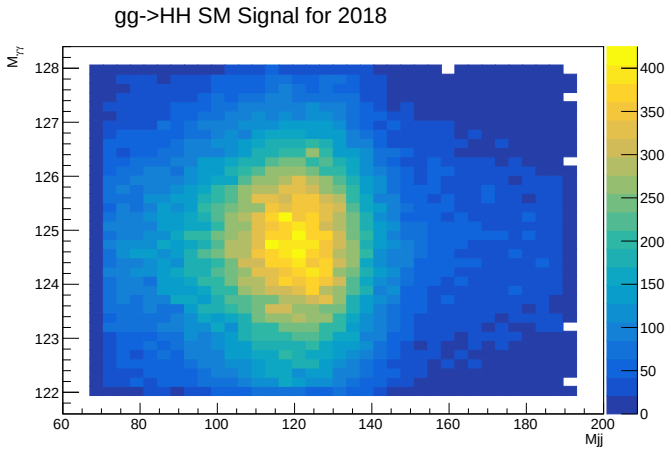
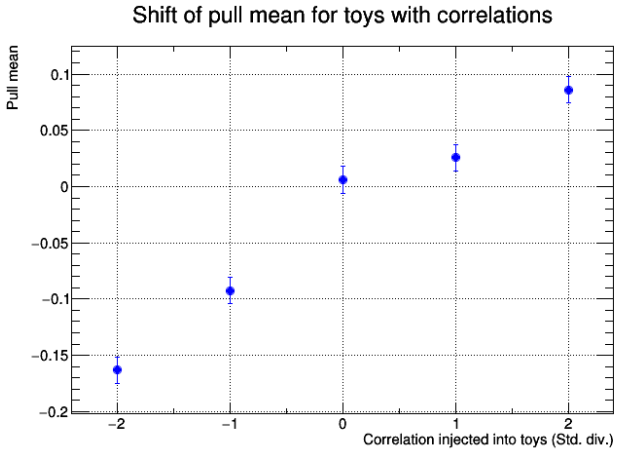


Figure 3.21: Two dimensional (m_{jj}, $m_{\gamma\gamma}$) distribution for the selected HH candidates of the simulated ggHH events using the 2018 dataset.

fluctuations. In Ref. [70], a good correction for the H→$\gamma\gamma$ case is found to be:

$$L_{corr} = L + N_{par} \tag{3.18}$$

where L is the negative log-likelihood and N_{par} is the number of free parameters of the function considered.
For the HH categories the envelope method is used for the description both of the $m_{\gamma\gamma}$ and the m_{jj} backgrounds. In the assumption of no-correlations between the $m_{\gamma\gamma}$ and m_{jj} variables, the generalization of the envelope method for the 2D fit is straightforward. The same function families defined for the $m_{\gamma\gamma}$ description can be used also for the m_{jj} variable because the background shape is analogous. The correction defined in Eq. 3.18 accounts for the sum of free parameters of the m_{jj} and $m_{\gamma\gamma}$ distributions.
From the practical point of view, the numerical minimization with discrete parameters is not reliable. Therefore, for each 2D combination of ($m_{\gamma\gamma}, m_{\mathrm{jj}}$) envelope indexes, a minimization is performed and the envelope is built afterwards. In order to reduce the computing time required for the minimization, a preliminary procedure determines the optimal order of each function family in each category. All the functions with up to six free parameters are considered. The functions with order higher than the optimal one are then not considered for the envelope construction. The range for the $m_{\gamma\gamma}$ and m_{jj} fits are 100 GeV$< m_{\gamma\gamma} <$ 180 GeV, and 70 GeV$< m_{\gamma\gamma} <$ 190 GeV. However, for the two ggHH-enriched categories with the lowest m_{X} at low BDT score, the fit region is reduced to 90 GeV$< m_{\gamma\gamma} <$ 190 GeV because the background shape is not well described. It was verified that with the reduced m_{jj} range the background modeling remains robust and that the bias induced on the expected signal strength is negligible. The changes of the expected upper limits on the HH cross section and of the constraints on anomalous couplings are found to be below 1%.



:

Figure 3.22: Average pull (over the generated MC toys) on the signal strength for different levels of $(m_{\gamma\gamma}, m_{\mathrm{jj}})$ correlation on the background distribution.

Checks of background uncorrelation

No correlation is observed between any of the families considered within uncertainties. As a test, for each category a set of toy MC samples are generated using the nominal background model with no-correlation. Then the correlation coefficient ρ measured in data is compared to the spread of the ρ distribution observed with the toy MC samples and found to be compatible with zero within one standard deviation.
Because of a very low number of selected events, the high BDT score categories can have a significant $(m_{\gamma\gamma}, m_{\mathrm{jj}})$ correlation in the background hidden within the statistical uncertainty. Therefore, it was verified with toy MC samples that the bias in the estimated μ is negligible even in case of $(m_{\gamma\gamma}, m_{\mathrm{jj}})$ correlations in the background. To this end, toy MC samples were generated with an hypothetical HH signal equivalent to 10 fb cross section, and correlated ($m_{\gamma\gamma}$,m_{jj}) background distributions. The correlation is introduced by modifying the background function with the following functions:

$$P(x, y) = x + y - 2xy \quad \text{for a negative correlation}$$
$$P(x, y) = 1 - x - y + 2xy \quad \text{for a positive correlation}$$

where $x = (m_{\gamma\gamma} - 100\ \mathrm{GeV})/80\ \mathrm{GeV}$, and $y = (m_{\mathrm{jj}} - 70\ \mathrm{GeV})/120\ \mathrm{GeV}$. In each category, a correlation at the level of 0, 1, and 2 standard deviations with respect to the ρ distribution observed with the toy MC samples is tested. A ± 2 standard deviation correlation is an unrealistic situation which is taken as a conservative upper boundary for the study. As visible in Fig. 3.22, the bias on the signal strength is found to be at most 0.15 which is negligible compared to the uncertainty on μ as we will see.

3.11.4 Systematic uncertainties

The analysis is statistically limited, therefore the impact of the systematics uncertainties on the result is small. In particular, as we will see, the upper limit on μ_{HH} including the systematics uncertainties is only 2% worse than the limit computed considering only the statistical uncertainties. The impact of the systematics uncertainties consists in a modification of the single and double Higgs yields in the categories, due to the event loss or migration across categories. The only systematic uncertainty associated with the continuum background, estimated from data, is the choice of its parametric modeling within the envelope method. The impact of the systematics uncertainties, quoted for simplicity on μ_{HH} and $\mu_{t\bar{t}H}$, is shown in Table 3.8, while their description is listed below:

- QCD scale: it accounts for the uncertainty on the renormalization and factorization scale. They are set accordingly to Ref. [6], [7], and [8] for the single Higgs, ggHH, and qqHH processes, respectively. In particular, the uncertainties associated with the ggHH and tt̄H QCD scales, whose 1σ variations change the total cross sections of up to 5 and 9%, respectively, represent the dominant uncertainties for this analysis.

- Parton distribution functions (PDF) modeling: it is computed according to the PDF4LHC15 prescriptions [71]. The PDF modeling affects the total number of events and also the event categorization since it modifies the number of jets produced in association with the Higgs boson signals. The event migration is computed using the NNPDF30 set with the MC2hessian method [72].

- α_s value: the uncertainty on the QCD coupling constant is computed along with the PDF modeling uncertainty using the PDF4LHC15 prescriptions. The α_s value affects in fact also the PDF modeling, thus the impact of the two uncertainties is estimated together.

- Parton shower modeling: this uncertainty is considered for the qqHH process because different showering schemes can significantly change some of the VBF observables [73], hence the total number of events and their classification. This uncertainty is conservatively estimated as the full symmetrized difference in yields in each category obtained from qqHH MC samples generated with different parton shower ISR and FSR configurations.

- Uncertainty in the H→γγ and H→bb̄ branching fractions which amounts to about 3 and 0.5%, respectively, according to Ref. [6].

The dominant experimental systematic uncertainties are:

- Photon identification BDT score: it accounts for the residual data-simulation discrepancy of the photon ID BDT score distribution. Such a discrepancy is

ascribed to the limited accuracy of the regression used to correct the BDT input variables to cover the residual discrepancies between data and simulation. It is estimated through the procedure described in Section 3.5.2.

- Photon energy scale and resolution: it accounts for the residual discrepancy between the data and MC simulation after the corrections (Section 2.4.8). It accounts for effects such as non-linearities of the light collection, the different shower-shape of electrons, used to derive the correction, and photons, as well as different trainings in the energy regression and variation of the binning used to derive the correction.

- Per-photon energy resolution estimate: this variable is computed by the photon energy regression. Its impact on the event selection and classification is estimated by varying the resolution of $\pm 5\%$ around its nominal value.

- Jet energy scale and resolution corrections: the energy scale and resolution of jets is measured using the p_T balance in Z(ee)+jets, Z($\mu\mu$)+jets, γ+jets, and multijet events [64]. The uncertainty on the calibration is a few percent and depends on p_T and η. The impact of jet energy scale uncertainties in event yields is estimated by varying the jet energy corrections within their uncertainties, ranging between 1 and 3% in central barrel, and propagating the effect to the final result. Correlations between years ranging between 0 and 100% are introduced for the different jet energy scale uncertainty sources.

- Jet b-tagging: such uncertainties are computed comparing the distribution of the b tagging efficiency between data and simulation. The efficiency on light flavour jets is measured using a inclusive multijet sample, while the efficiency on heavy-flavour jets is measured using muon-enriched jet samples, and $t\bar{t}$ plus one or two leptons samples [61]. The uncertainties include the statistical component on the estimate of the fraction of heavy and light flavour jets in data and simulation.

- Trigger efficiency: as discussed in Section 3.3.1, the efficiency of the trigger selection and the corresponding uncertainty is measured with Z $\rightarrow e^+e^-$ events using a $T\&P$ technique. An additional uncertainty is introduced to account for a gradual shift in the timing of the inputs of the ECAL L1 trigger in the region $|\eta| > 2$, causing a specific trigger inefficiency during the 2016 and 2017 data taking periods. Photons and also jets are affected by this inefficiency.

- Photon preselection: as discussed in Section 3.5.2, the photon preselection efficiency (including the electron veto efficiency) is estimated using Z $\rightarrow e^+e^-$ and Z $\rightarrow \mu^+\mu^-\gamma$ events with a $T\&P$ method. The uncertainty on the scale factors derived to match the efficiency of the simulation to the one measured with data is propagated throughout the analysis.

- Integrated luminosity: the related uncertainties are determined through auxiliary measurements by the CMS luminosity monitoring [74–76]. The uncertainties across the different years of data-taking are partially correlated to account for common sources of uncertainty in the luminosity measurement schemes.

- Pileup jet ID output score: it accounts for the differences between data and simulation in the distribution of the pileup jet ID variable. Only the VBF jets are affected. This uncertainty is estimated by comparing the score of jets in Z+jets events in data and simulation in intervals of p_T and η.

Other systematics uncertainties impact the signal strength by less than 1% and are thus negligible with regard to the ones described above. They include uncertainties on lepton identification and isolation efficiencies, on the correct vertex assignment efficiency, and on the missing transverse momentum.

Table 3.8: Impact of the systematic uncertainties on the signal strengths in percentage.

Systematic uncertainty	$\Delta\mu_{\mathrm{HH}}$ (%)	$\Delta\mu_{\mathrm{t\bar{t}H}}$ (%)
QCD scale	$+7/-2$	8
PDF modeling + α_s	3	6
branching ratio	3	2
parton shower modeling	< 1	< 1
Luminosity	3	2
Photon preselection	2	1
Per-photon σ_E/E	1	< 1
Photon ID	< 1	1
Trigger	< 1	1
Photon energy & res.	< 1	1
Jet energy & res.	< 1	2
b-tag efficiency	< 1	4
Pileup jet ID	< 1	< 1

3.12 Results

This work led to two results: the search for the double Higgs production presented in Section 3.12.1, and the measurement of the Higgs boson couplings presented in Section 3.12.3. All the results were found compatible with the SM predictions. In particular, no significant excesses over the background of double Higgs production events were found, thus upper limits on the HH cross sections were extracted.

3.12.1 Search for the HH process

Only the HH categories (twelve ggHH and two qqHH categories) are included in the extraction of the HH production yield. The t$\bar{\text{t}}$H categories are not considered because the HH yields in the t$\bar{\text{t}}$H categories are negligible. A likelihood fit is performed using the likelihood defined in Eq. 3.5 to measure the μ_{HH} signal modifier. The fit to the $m_{\gamma\gamma}$ and m_{jj} distributions of the fourteen categories is shown in Fig. 3.23, 3.24, 3.25, and Fig. 3.26, 3.27, 3.28, respectively. Analogous fits are performed to measure either the μ_{ggHH} or the μ_{qqHH} parameter fixing the other parameter to one. The observed and expected signal strengths are reported below:

Parameter	Expected	Observed
μ_{HH}	$1.0^{+2.7}_{-1.9}$	$2.7^{+2.6}_{-2.0}$
μ_{ggHH} fixing μ_{qqHH} =1	$1.0^{+2.7}_{-1.9}$	$2.8^{+2.7}_{-2.0}$
μ_{qqHH} fixing μ_{ggHH} =1	1^{+91}_{-65}	10^{+97}_{-62}

Due to a small excess observed in data, the observed signal strengths are larger than one, but still compatible with the SM within the uncertainties. The inclusive μ_{HH} signal strength is dominated by the ggHH process because of the larger cross section, as visible by the large uncertainty on μ_{qqHH} compared to the uncertainty on μ_{HH}. Alternatively, μ_{ggHH} and μ_{qqHH} are measured simultaneously, as shown in Fig. 3.29.
Since no evidences of double Higgs production were found, upper limits on the corresponding cross sections are extracted using the procedure described in Section 3.11.1. The observed and expected upper limits are reported as follows:

	Upper limit at 95% C.L.		
	Expected	Observed	Best published result (observed)
σ_{HH}	5.2×SM	7.7×SM	-
σ_{ggHH} fixing μ_{qqHH} =1	5.3×SM	7.8×SM	6.9×SM (HH comb. with 36 fb^{-1} [25])
σ_{VBFHH} fixing μ_{qqHH} =1	208×SM	225×SM	840×SM (HH $\to$ b$\bar{\text{b}}$b$\bar{\text{b}}$ with 126 fb^{-1} [65])

The upper limit on σ_{ggHH} of 7.8×SM is the best result from the CMS experiment to date, and it is comparable to the constraint set by the ATLAS experiment combining the most sensitive HH decay channels with the 2016 dataset. This is also the first upper limit on σ_{VBFHH} set by the CMS experiment, which improves the best constraint set by the ATLAS experiment.

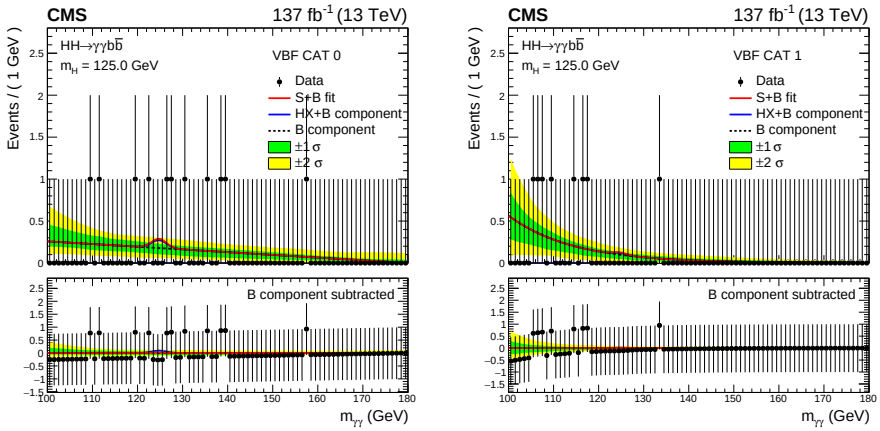


Figure 3.23: The $m_{\gamma\gamma}$ distribution for the selected events (black points) is shown for the qqHH categories with the curves corresponding to the signal + background fit (solid red) and the background only (dashed red), with bands covering the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties in the fitted background.

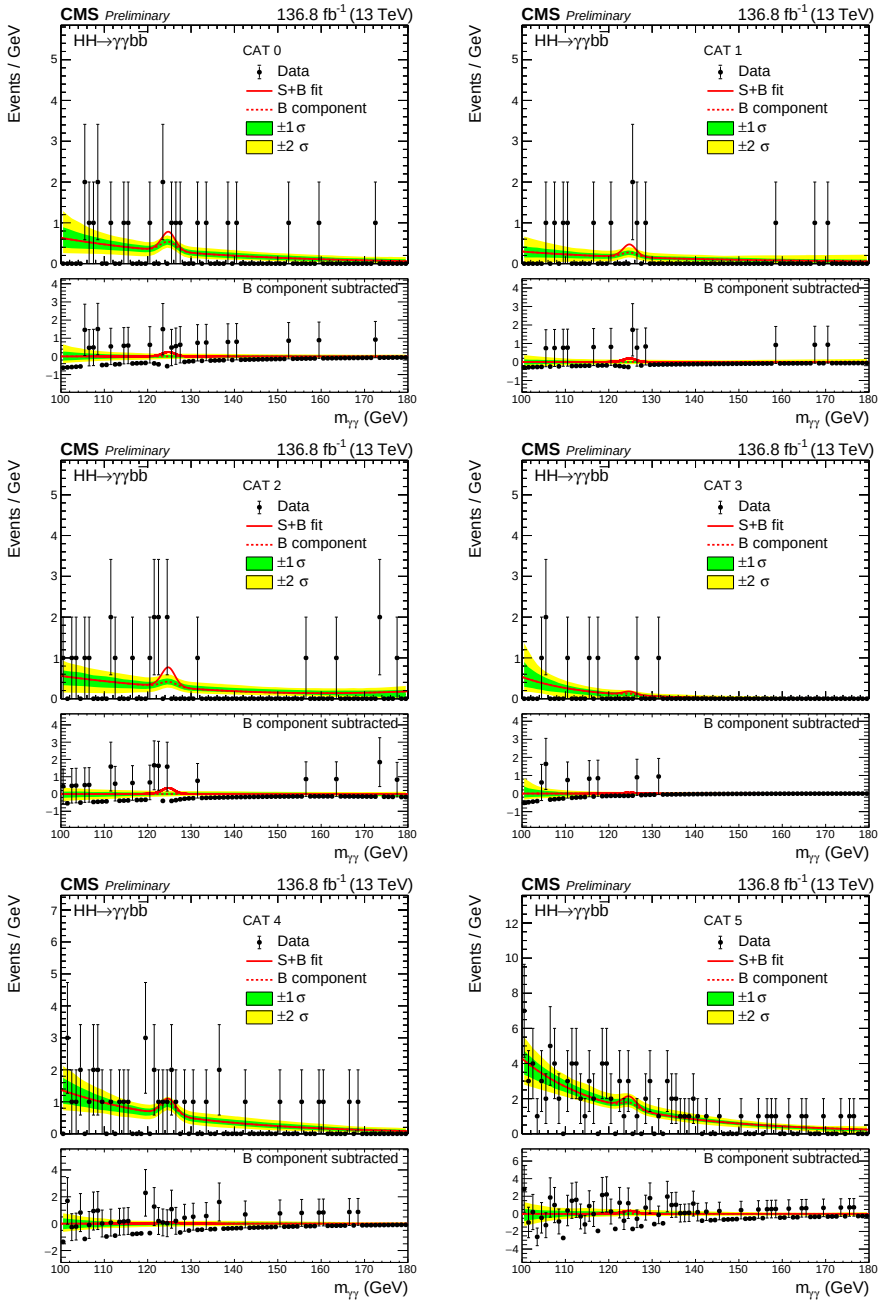


Figure 3.24: The $m_{\gamma\gamma}$ distribution for the selected events (black points) is shown for the first six ggHH categories with the curves corresponding to the signal + background fit (solid red) and the background only (dashed red), with bands covering the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties in the fitted background.

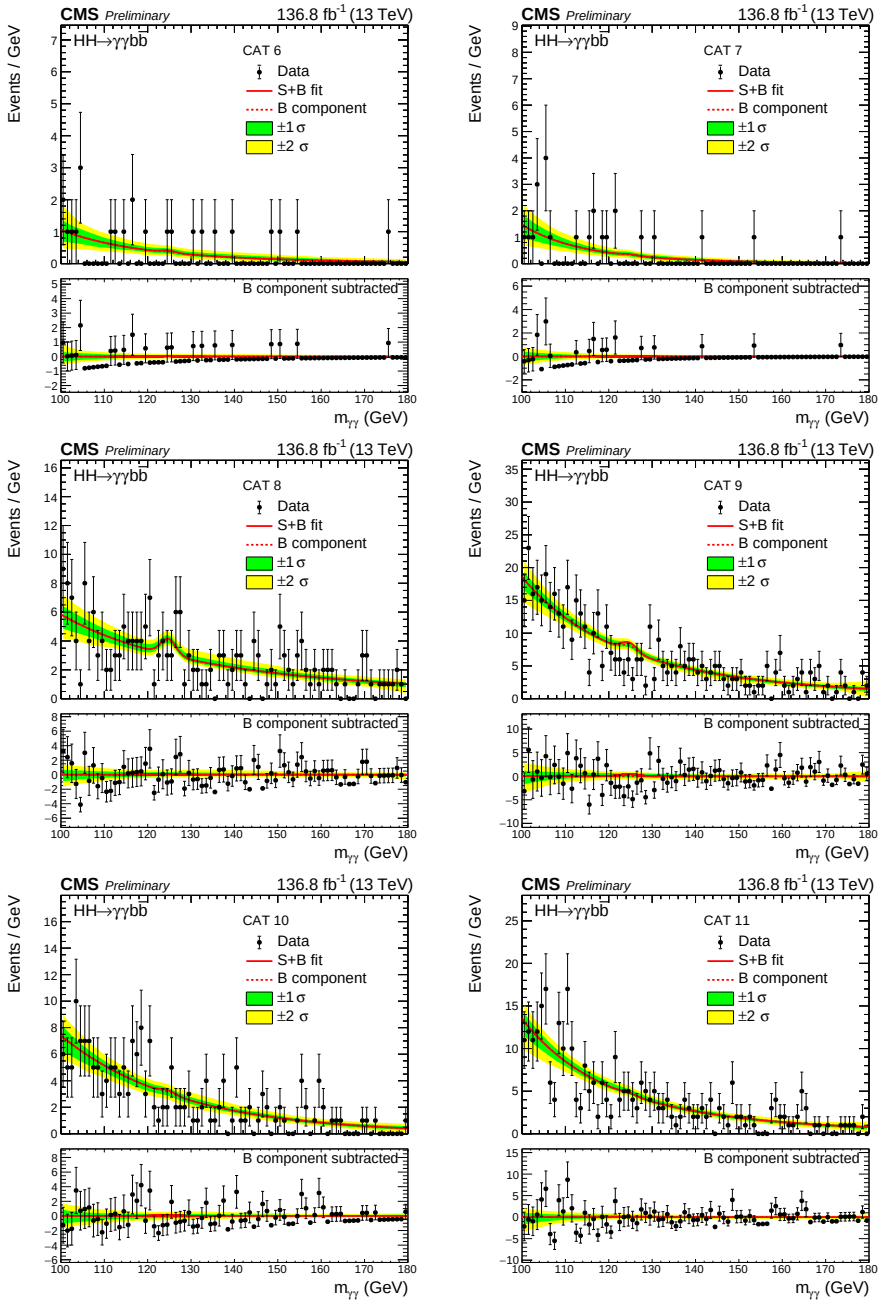


Figure 3.25: The $m_{\gamma\gamma}$ distribution for the selected events (black points) is shown for the second six ggHH categories with the curves corresponding to the signal + background fit (solid red) and the background only (dashed red), with bands covering the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties in the fitted background.

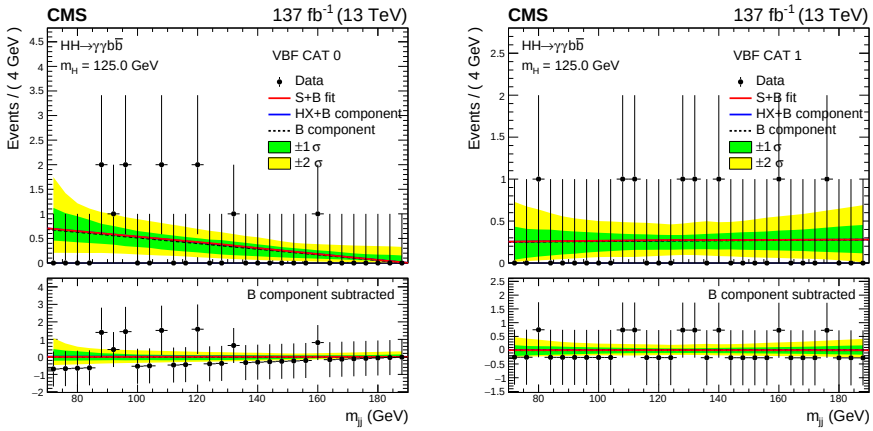


Figure 3.26: The m_{jj} distribution for the selected events (black points) is shown for the qqHH categories with the curves corresponding to the signal + background fit (solid red) and the background only (dashed red), with bands covering the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties in the fitted background.

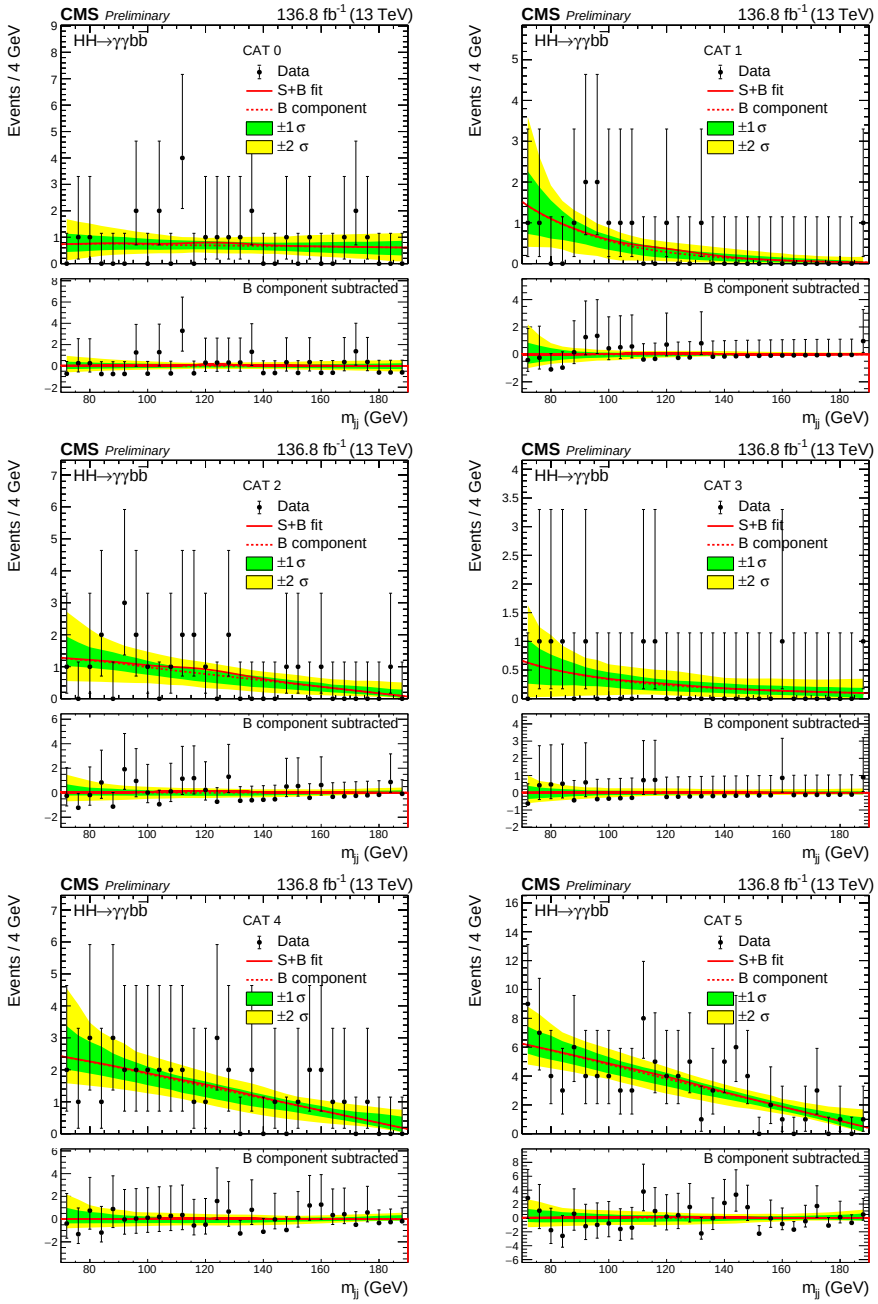


Figure 3.27: The m_{jj} distribution for the selected events (black points) is shown for the first six ggHH-enriched categories with the curves corresponding to the signal + background fit (solid red) and the background only (dashed red), with bands covering the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties in the fitted background.

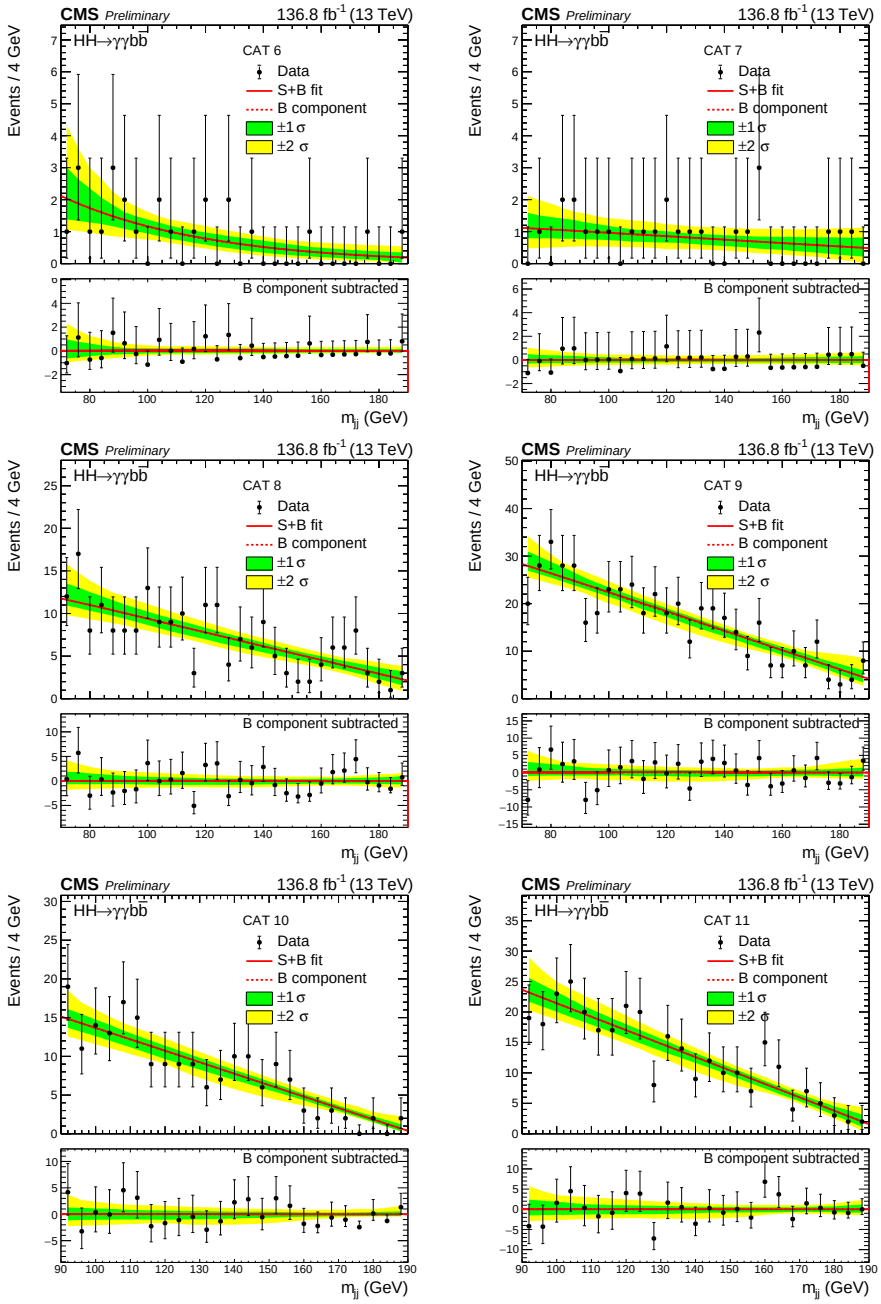


Figure 3.28: The m_{jj} distribution for the selected events (black points) is shown for the second six ggHH-enriched categories with the curves corresponding to the signal + background fit (solid red) and the background only (dashed red), with bands covering the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties in the fitted background.

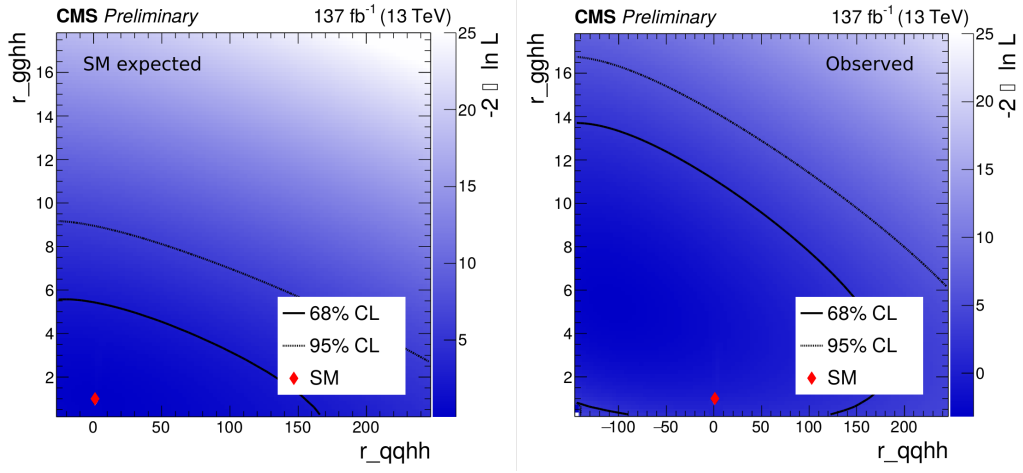


Figure 3.29: Expected and observed two dimensional likelihood scan of the μ_{ggHH} (y-axis) and μ_{qqHH} (x-axis) parameters, on the left and on the right side, respectively. The exclusion contours are also shown for the 68% (solid black line) and 95% (dashed black line) confidence levels.

3.12.2 Constraints on the BSM benchmark hypotheses

Upper limits are also set on the twelve BSM benchmark hypotheses. The ggHH events simulated at the LO for each benchmark hypothesis are used to derive the expected number of events as well as the signal model. Only the ggHH categories are considered, and the qqHH process is neglected in the statistical interpretation of the data. The observed and expected upper limits at 95% C.L. on the BSM benchmark hypotheses are shown in Fig. 3.30. The observed limits are slightly larger than the expected limits because of the small excess of events found in the ggHH-enriched categories. The expected limits span from about 0.1 to 1 fb. The different sensitivity is due to the kinematics of each specific BSM benchmark hypothesis, which change the population of the expected signal events in the four-body mass categories.

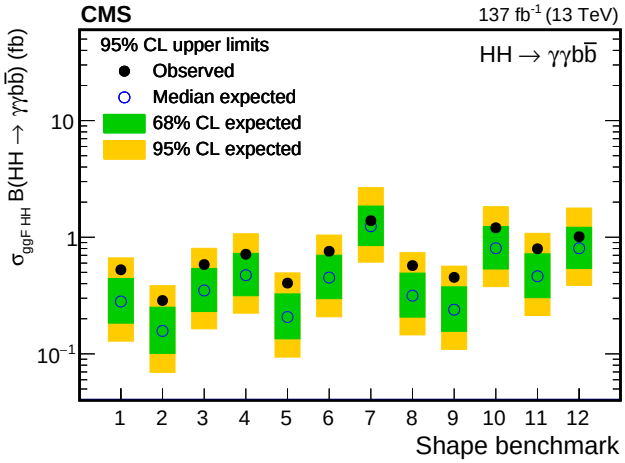


Figure 3.30: Expected and observed 95% CL upper limits on the $\sigma_{\mathrm{ggHH}} \times \mathrm{BR}(\mathrm{HH} \to \mathrm{b\bar{b}}\gamma\gamma)$ for the BSM benchmark models, shown with transparent and solid circles, respectively. The green and yellow bands represent the one and two standard deviation extensions beyond the expected limit, respectively.

3.12.3 Constraints on the Higgs boson couplings

Since no evidence of the HH signal are found, the 95% C.L. upper limits on $\sigma_{\mathrm{HH}} \times \mathrm{BR}(\mathrm{HH} \to \mathrm{b\bar{b}}\gamma\gamma)$ are derived as a function of the κ_λ parameter, as shown in the left panel of Fig. 3.31. The upper limit dependence on κ_λ is determined by the variation of the m_{X} distribution of the ggHH signal that modifies the categories population, thus the sensitivity to that signal. In particular, the high-m_{X} categories provides a higher sensitivity than the low-m_{X} categories because they have a smaller continuum background contamination. For κ_λ values in the [0,6] interval, the destructive interference between the ggHH production diagram with a box loop of top quarks and the one with the triHiggs vertex is maximum. This causes a strong variation of the m_{X} distribution which migrates from the highest energy spectrum at about κ_λ =2 to the softest spectrum at about κ_λ =5. Comparing with the theoretical prediction, the resulting constraint on the κ_λ parameter is found to be:

$$\begin{aligned} &\text{Observed:} \quad -3.26 < \kappa_\lambda < 8.48 \text{ at } 95\% \text{ C.L.} \\ &\text{Expected:} \quad -2.61 < \kappa_\lambda < 8.28 \text{ at } 95\% \text{ C.L.} \end{aligned} \tag{3.19}$$

This result improves the existing most stringent constraint using the HH signal ($-5 < \kappa_\lambda < 12$ at 95% CL), from the ATLAS HH combination with the 2016 dataset [25] (36 fb^{-1}).
The same procedure is performed to extract a constraint on the $c_{2\mathrm{V}}$ parameter, as visible in the right panel of Fig. 3.31. In this case the upper limit is derived on $\sigma_{\mathrm{VBFHH}} \times \mathrm{BR}(\mathrm{HH} \to \mathrm{b\bar{b}}\gamma\gamma)$ because the $c_{2\mathrm{V}}$ sensitivity is totally retained by the

qqHH process. As for the κ_λ parameter, the variation of the upper limit as a function of the c_{2V} value is determined by the corresponding m_X distribution variation. In this case, the interference between the three qqHH production diagrams makes the m_X distribution spectrum migrate to high energies as soon as c_{2V} deviates from its SM prediction, enhancing the sensitivity to the qqHH signal. The resulting constraint on the c_{2V} parameter is found to be:

$$\begin{aligned} &\text{Observed:} \quad -1.31 < c_{2V} < 3.45 \text{ at } 95\% \text{ C.L.} \\ &\text{Expected:} \quad -0.96 < c_{2V} < 3.07 \text{ at } 95\% \text{ C.L.} \end{aligned} \tag{3.20}$$

This is the first upper limit on the c_{2V} parameter set by the CMS experiment. The HH → bb̄γγ channel provides a good sensitivity also to the c_{2V} parameter. However, the HH → bb̄bb̄ and HH → bb̄$\tau^+\tau^-$ are expected to be more sensitive because for anomalous c_{2V} values, the Higgs boson p_T spectrum shifts to higher values. In that phase-space region, the signal-over-background ratio of the HH → bb̄bb̄ and HH → bb̄$\tau^+\tau^-$ channels, penalizing them in the SM case, is expected to significantly increase. Therefore, the larger branching ratio of those processes becomes the factor dominating the sensitivity.

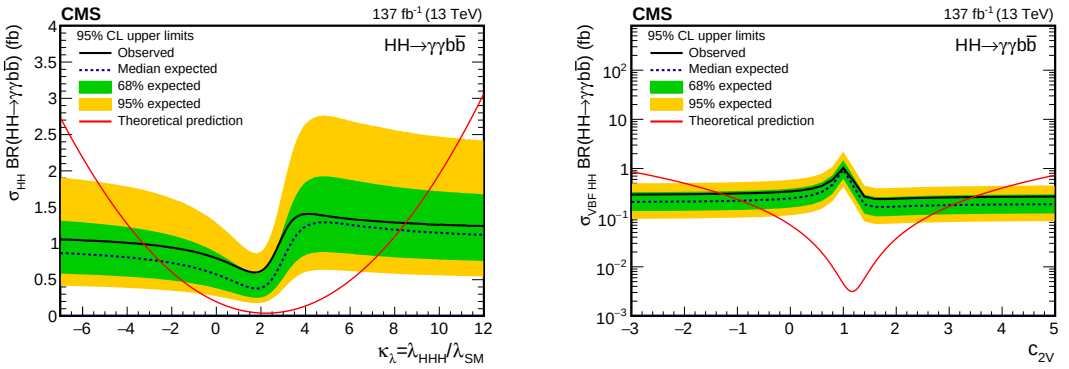


Figure 3.31: Expected and observed 95% CL upper limits on the SM-like qqHH production cross section times BR(HH → bb̄γγ) obtained for different values of κ_λ and c_{2V} on the left and right side, respectively. The green and yellow bands represent the one and two standard deviation extensions beyond the expected limit, respectively. The red lines show the theoretical predictions.

Likelihood scan for the coupling measurements

In order to measure the (κ_λ, κ_t, c_V, c_{2V}) parameters a profile likelihood is defined as:

$$\mathcal{L}_{prof}(\kappa_\lambda, \kappa_t, c_V, c_{2V}) = -2 \log \frac{\mathcal{L}(\kappa_\lambda, \kappa_t, c_V, c_{2V}, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\kappa_\lambda}, \hat{\kappa_t}, \hat{c_V}, \hat{c_{2V}}, \hat{\theta})} \tag{3.21}$$

Where $\mathcal{L}$ is the likelihood defined in Eq. 3.5 expressed as a function of the Higgs couplings. The $\hat{\kappa_\lambda}$, $\hat{\kappa_t}$, $\hat{c_V}$, $\hat{c_{2V}}$, and $\hat{\theta}$ are the parameters values maximizing $\mathcal{L}$, i.e.

their best estimate. Instead, $\hat{\hat{\theta}}$ is the set of nuisance values maximizing $\mathcal{L}$ for a given set of (κ_λ, κ_t, c_V, c_{2V}) values.
Since the Higgs boson production and decay channels considered are insufficient to constrain the full set of couplings, only one or two parameters are measured at a time while the other ones are fixed to the SM prediction. The profile likelihood is used to extract also the exclusion regions at 68% and 95% confidence level.
The $\mathcal{L}_{prof}$ value as a function of the parameter of interest (likelihood scan) is reported in Fig. 3.32 and 3.33 for the κ_λ and κ_t parameters, respectively. Each figure compares the likelihood scan with and without the inclusion of the tt̄H categories.

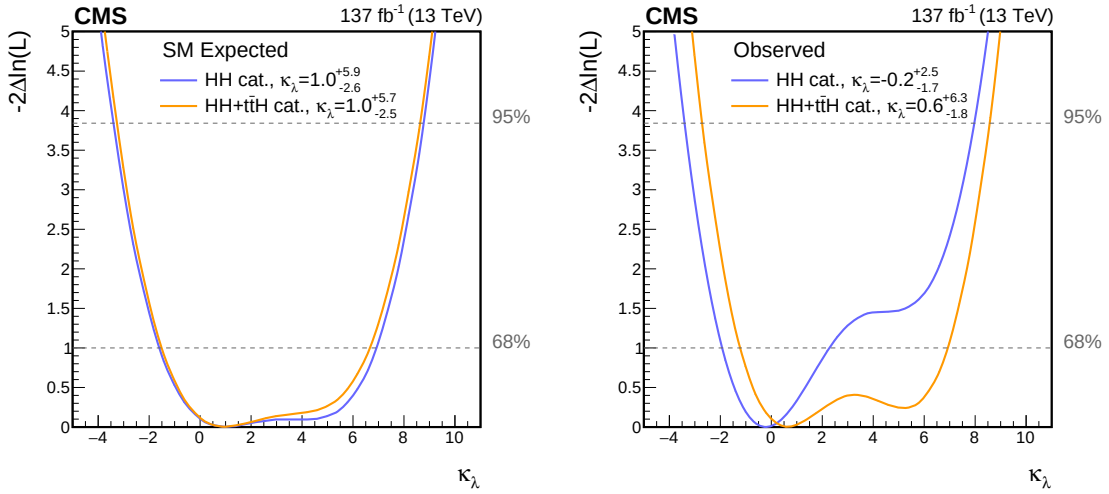


Figure 3.32: Expected and observed likelihood-scan of the κ_λ parameter, on the left and right sides, respectively. The likelihood scan is shown including the tt̄H categories (orange line) or using only the HH categories (blue line). The likelihood values corresponding to the 68% and 95% confidence levels are represented by the lower and upper horizontal dashed grey lines, respectively.

For the κ_λ-scan, when considering only the HH categories a difference is observed between the expected and the observed result because the small excess of events, observed especially in the ggHH categories at medium-m_X, tends to exclude κ_λ values moderately larger than one. The inclusion of the tt̄H categories brings the observed likelihood close to the expectation of the SM. In fact, the tt̄H categories help constrain the tt̄H contamination in the HH categories. In particular, an excess of events is observed also in the tt̄H categories according to Ref. [23] that measured a signal strength of $\mu_{t\bar{t}H} = 1.4$ with an uncertainty of about ±30%. As a consequence, the small excess of events observed in the HH categories is partially attributed to the excess of tt̄H signal that tends instead to favour a κ_λ value larger than one.

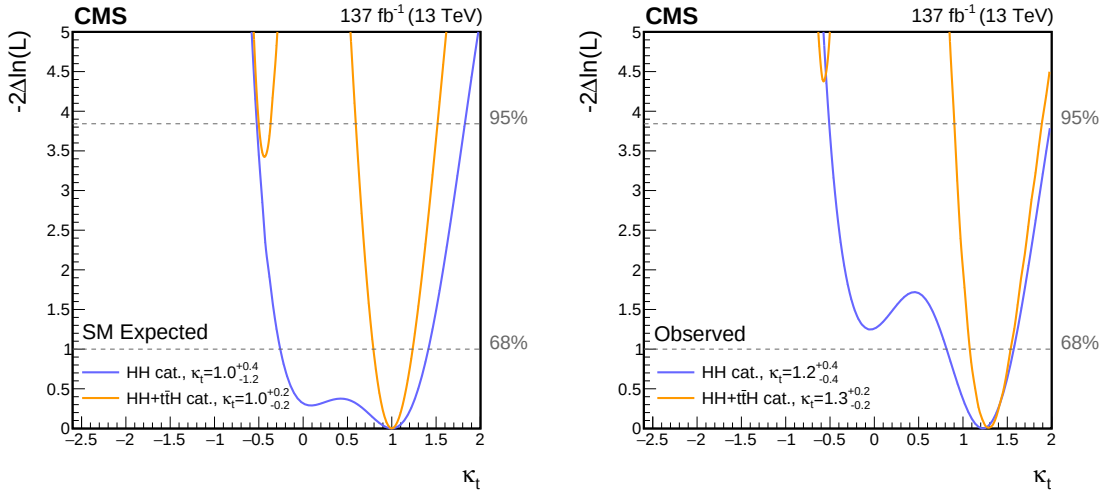


Figure 3.33: Expected and observed likelihood-scan of the κ_t parameter, on the left and right sides, respectively. The likelihood scan is shown including the $t\bar{t}H$ categories (orange line) or using only the HH categories (blue line). The likelihood values corresponding to the 68% and 95% confidence levels are represented by the lower and upper horizontal dashed grey lines, respectively.

For the κ_t-scan, the inclusion of the $t\bar{t}H$ categories significantly improve the sensitivity to the coupling, thanks to the sensitivity to κ_t both of the $t\bar{t}H$ cross section and of the H→$\gamma\gamma$ branching ratio. In particular, the $t\bar{t}H$ cross section is symmetric for $\kappa_t = \pm 1$, and approximately also the HH production cross section. Instead, the H→$\gamma\gamma$ branching ratio dependence on κ_t is strongly asymmetric and improves the sensitivity to a negative κ_t value. This allows the exclusion of a negative value of κ_t at 95% confidence level.
The c_{2V} scan is shown in Fig. 3.34. For this parameter the sensitivity comes entirely from the qqHH cross section, thus the contribution from the $t\bar{t}H$ categories is completely negligible. A c_V likelihood scan is out of the scope for this analysis because there are not experimental categories targeting the VH and qqH events that dominates the c_V sensitivity.

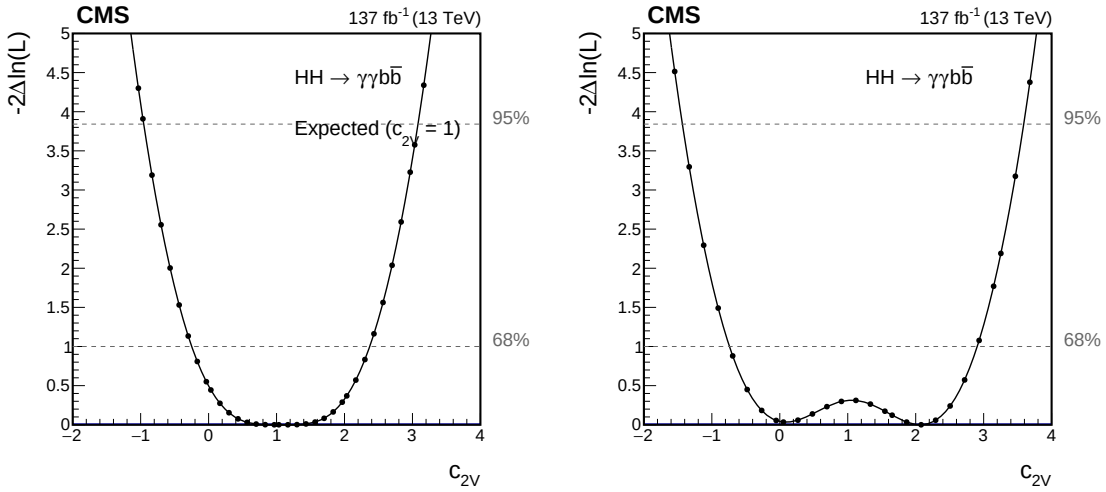


Figure 3.34: Expected and observed likelihood-scan of the c_{2V} parameter, on the left and right sides, respectively. The likelihood values corresponding to the 68% and 95% confidence levels are represented by the lower and upper horizontal dashed grey lines, respectively.

The two dimensional likelihood scan of the $(\kappa_\lambda, \kappa_t)$ parameters is reported in Fig. 3.35. In the figure, the regions where the parametrization of $\sigma_{t\bar{t}H}$ $(\kappa_\lambda,\kappa_t)$ is not reliable are shown with a gray band (see Appendix B for details). As expected, the improvement from the $t\bar{t}H$ categories is large. The result is found compatible with the SM.

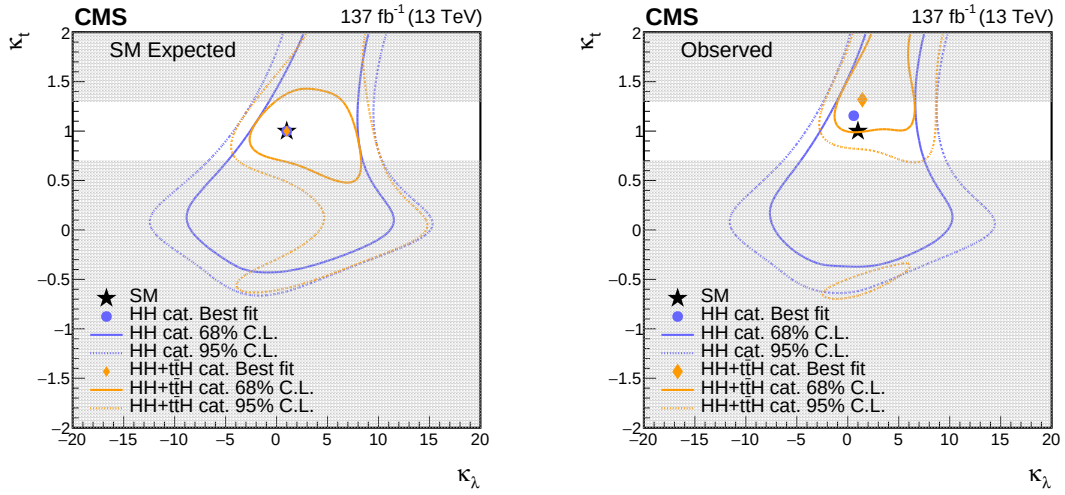


Figure 3.35: Expected (left) and observed (right) two dimensional likelihood-scans of the (κ_λ,κ_t) parameters including the t$\bar{t}$H categories (red) and using only the HH categories (black). The exclusion regions at 68% and 95% confidence level are represented with the solid and the dashed lines, respectively. The region where the $\sigma_{t\bar{t}H}$ (κ_λ,κ_t) parametrization is not reliable is highlighted in gray.

The (κ_λ, c_{2V}) likelihood scan is shown in Fig. 3.36. The sensitivity on those two parameters is dominated by the two considered HH production mechanisms. The coupling measurements are summarized in Table 3.9.

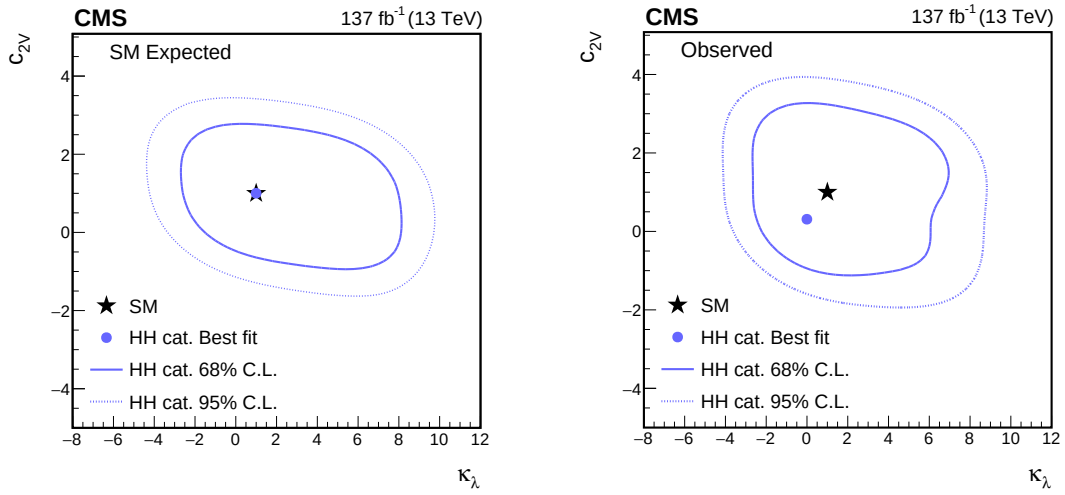


Figure 3.36: Expected (left) and observed (right) two dimensional likelihood-scans of the (κ_λ,c_{2V}) parameters. The exclusion regions at 68% and 95% confidence level are represented with the solid and the dashed lines, respectively.

Table 3.9: Best fit values for the 1D and 2D likelihood scans of the Higgs coupling parameters. For the 1D scans the 1σ uncertainties and the 95% confidence intervals are also quoted.

Parameters		Best fit $\pm 1\sigma$	Interval at 95% C.L.
κ_λ	obs	$0.6^{+6.3}_{-1.8}$	[-2.7, 8.6]
	exp	$1.0^{+5.7}_{-2.5}$	[-3.3, 8.6]
κ_t	obs	$1.3^{+0.2}_{-0.2}$	[0.90, 1.90]
	exp	$1.0^{+0.2}_{-0.2}$	[-0.50, -0.37] $\cup$ [0.59, 1.52]
$c_{2\mathrm{V}}$	obs	$2.1^{+0.8}_{-2.8}$	[-1.4,3.6]
	exp	$1.0^{+1.2}_{-1.2}$	[-2.0,3.1]
(κ_λ,κ_t)	obs	$(1.4, 1.3)$	-
	exp	$(1, 1)$	-
(κ_λ,$c_{2\mathrm{V}}$)	obs	$(0.0, 0.3)$	-
	exp	(1,1)	-

3.13 Conclusions and perspective

In summary, all the results were found compatible with the SM predictions. No evidence of the HH process is found, thus, an upper limit was set to its cross section. The observed upper limit on the inclusive HH production cross section is 7.7×SM and corresponds to the most stringent result achieved by the CMS experiment to date. In the assumption of no HH signals, constraints on anomalous κ_λ values, and, for the first time with the data collected by the CMS experiment, on $c_{2\mathrm{V}}$ values are set. The observed constraints at 95% C.L. are $-3.26 < \kappa_\lambda < 8.48$, which is the most stringent constraint among the published results [25, 77], and $-1.31 < c_{2\mathrm{V}} < 3.45$. In the assumption of a HH signal, measurements of the κ_λ and κ_t parameters are performed through a combination of HH and t$\bar{\text{t}}$H enriched categories. The measured values with one standard deviation uncertainty are $\kappa_\lambda = 0.6^{+6.3}_{-1.8}$ and $\kappa_\mathrm{t} = 1.3 \pm 0.2$. A simultaneous measurement of the κ_λ and κ_t parameters is performed and the result is shown in Fig. 3.35.
The combination of the HH and the t$\bar{\text{t}}$H enriched categories has been developed to provide a framework for a future combination including all the sensitive single and double Higgs production and decay channels with the dataset collected by the CMS experiment during the Run 2. As shown by a preliminary result of ATLAS [78], such a combination provides the sensitivity required for a simultaneous measurement of all the Higgs boson couplings, including κ_λ. This combination provides the most precise measurement of the κ_λ parameter, fixing all the other couplings to the corresponding SM prediction, corresponding to $\kappa_\lambda = 4.6^{+3.2}_{-3.8}$ (one standard deviation uncertainties).
Further data will be collected by the CMS experiment during the Run 3, equivalent to an integrated luminosity of about 300 fb^{-1}. This dataset will improve the sensitivity to the HH signal and to the coupling parameters. Preliminary stud-

ies provide a projected limit (expected SM) on the inclusive ggHH cross section of about 3.6×SM for the end of Run 3. The projected constraints at 95% C.L. on the κ_λ and $c_{2\text{V}}$ parameters, assuming no HH signal, are $-1.6 < \kappa_\lambda < 7.2$ and $-0.58 < c_{2\text{V}} < 2.7$ for the end of Run 3. Therefore, the Run 3 will not provide the required amount of data for an evidence of the HH process, should it be consistent with the SM predictions. An evidence of a (SM) HH process is expected during the high-luminosity phase of the LHC, as discussed in the next Chapter.

Chapter 4

Prospects for the HH searches at High Luminosity LHC

The measurement of the HH process is among the main goals of the physics program of the high-luminosity phase of LHC (HL-LHC [79]). This phase is planned to start in 2027. An upgraded LHC machine will provide proton-proton collisions at a center-of-mass energy of 14 TeV with a luminosity about 3-4 times larger than the Run 2 level. This will allow the collection of data amounting to 3000 fb^{-1} of integrated luminosity in about ten years of operation. The CMS detector will be upgraded to ensure the required resistance against the expected radiation levels and to handle an unprecedent level of pileup.
The prospects for the HH measurements are included in an extensive study reported in Ref. [80] covering all the physics channels of main interest at the HL-LHC. All the most sensitive HH decay channels are included in the study. The work carried out in the context of this thesis focuses on the prospects for the HH $\rightarrow$ $\text{b}\bar{\text{b}}\gamma\gamma$ channel. A combination of the projected results from ATLAS and CMS provides an expected significance for the SM HH signal of 4.0σ, and is expected to constrain anomalous κ_λ values in the interval $0.52 < \kappa_\lambda < 1.5$ at 68% CL.
The HH search is among the benchmark channels for the study of the physics impact of the new detector for timing measurements of minimum-ionizing-particles (MTD) that will be installed in the upgraded CMS detector. The HH searches feature final states with multiple objects that can individually benefit of the MTD detector. Therefore, the combined gain on the event is substantial. The work for the thesis focuses also on the estimate of the physics impact of the MTD detector on the HH $\rightarrow$ $\text{b}\bar{\text{b}}\gamma\gamma$ channel, and on the HH combination. This study is part of the MTD physics impact studies included in Ref. [81]. After an introduction in Section 4.1 on the HL-LHC phase and the CMS detector upgrade, the strategy and assumptions for the projections will be described in Section 4.2. A summary of the analysis strategy for the HH $\rightarrow$ $\text{b}\bar{\text{b}}\gamma\gamma$ channel will be given in Sec. 4.3, and finally the projection results will be discussed in Section 4.4.

4.1 HL-LHC

Figure 4.1 shows the long term LHC schedule and the luminosity forecast. The current shutdown will end in 2022 for a new data-taking campaign of 3 years duration (Phase I). The performance of the machine will be kept at the level of the last years of the Run 2 data-taking. A subsequent long shutdown of 2-3 years will allow the installation and commissioning of the LHC machine as well as to complete the HL-LHC upgrade. The HL-LHC phase (Phase II) is planned to start in 2027. Operating for about 10 years with a luminosity of $5 \cdot 10^{34}$, HL-LHC will provide data equivalent to a total of 3000 fb^{-1} of integrated luminosity. Furthermore, the center-of-mass energy for the proton-proton collisions will be increased to 14 TeV. In an ultimate performance scenario the luminosity would reach $7.5 \cdot 10^{34}$, allowing the collection of about 4000 fb^{-1} of integrated luminosity.

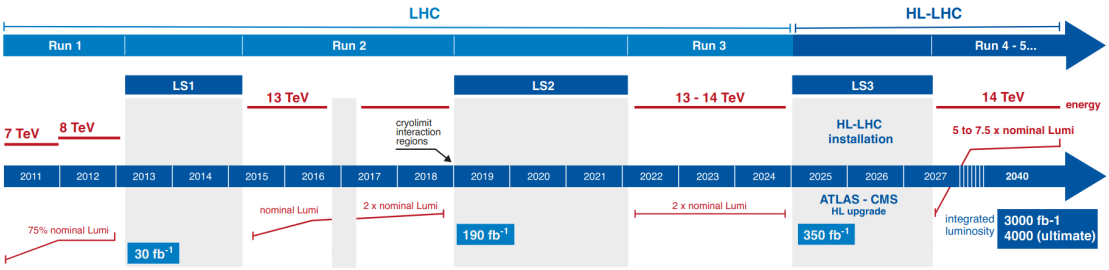


Figure 4.1: Long term schedule (nominal scenario) and luminosity forecast of the LHC.

4.1.1 CMS upgrade for HL-LHC

The CMS detector will be upgraded to be resistant against the radiation levels expected at HL-LHC, and to handle the number of simultaneous interactions at each proton-proton collision increasing from the current level of up to 50 to a 140(200) in the nominal(ultimate) performance scenario. A detailed description of the CMS upgrade is out of the scope of this work. An overview of the upgrades relevant for the HH $\rightarrow$ b$\bar{\text{b}}\gamma\gamma$ search will be given below. More details can be found in the technical design reports (TDR) referenced in the text.
To benefit of the higher luminosity, the CMS L1 and HLT triggers will be upgraded [82, 83] to sustain rates of 750 kHz and 7.5 kHz, respectively. Furthermore, the L1 trigger latency will be raised to 12.5 μs. In order to provide an efficient event selection, the L1 trigger will be able to access the inner tracking system information and to readout the ECAL barrel with crystal granularity.
The inner tracking system upgrade [84] consists in a full replacement of the current sensors with new pixel and strips silicon sensors with higher granularity to cope with the expected increase in track multiplicity. The front-end electronics will be

also upgraded to be interfaced with the L1 trigger. Furthermore, the tracker coverage will be extended up to $|\eta|= 4$.
The barrel calorimeters will also undergo an upgrade [85]. The ECAL barrel crystal and APDs will not be replaced. Instead, the front end electronics will be upgraded to meet the trigger requirements and to mitigate the APD noise consequent to the radiation damage. For the same reason, the supermodule operating temperature will be reduced to 9 ^{o}C. Similarly, the HCAL barrel scintillators and fibers will not be replaced. Instead, the hybrid photodiodes currently used to read-out the scintillation light will be replaced by Silicon Photomultipliers (SiPM) ensuring a higher radiation resistance. In addition the HCAL front-end electronics will be replaced to meet the L1 trigger requirements.
In the endcap regions, a new high-granularity calorimeter (HGCal [86]) will replace the ECAL and HCAL endcaps, which were found unable to stand the HL-LHC radiation dose. The HGCal consists in a combined electromagnetic and hadronic sampling calorimeter based primarily on silicon pad sensors. In the HGCal hadronic region far from the beam line, plastic scintillators read-out by SiPMs will be installed. Providing an excellent spatial resolution both in the transverse and in the longitudinal direction, the HGCal will provide an excellent pileup rejection and object identification efficiency.
Finally, a new subsystem for the time-of-flight measurements of minimum-ionizing-particles (MIP) will be installed in the CMS detector. The MIP timing detector (MTD [81]) will provide measurements with a time resolution of 30-40 ps RMS. This will allow a four-dimension reconstruction of the tracks, hence the vertices, reducing the effective pileup to a level comparable to the one observed during the Run 2. The MTD detector consists in a barrel and two endcaps subsystems covering the pseudorapidity regions $|\eta|< 1.5$ and $1.6 < |\eta|< 3$, respectively.

4.2 Strategy for the projections

The projections are fully based on MC simulations. The CMS detector response in the HL-LHC conditions is modeled through a parametric simulation based on the DELPHES software [87]. The pileup effect is simulated by overlaying an average of 200 simulated minimum-bias events per bunch crossing. The DELPHES reconstruction and identification performance on the physics objects is adjusted to match the results obtained using the standard CMS simulation based on the Geant4 software [57] (full CMS simulation) including the upgraded CMS detector. In case of not-finalized algorithms for the object reconstruction and/or identification, the simulation is tuned on the results of preliminary studies, or directly from the technical design reports for the Phase II upgrade of the CMS components.
For this projection study of HH searches only the gluon fusion production mechanism is considered. The most sensitive HH decay channels, reported in Table 4.1 were taken into account. The work described in this chapter, which was conducted

in the context of this thesis, focuses on the HH $\to b\bar{b}\gamma\gamma$ channel. A combination of the channels was performed to determine the overall sensitivity to the HH process. The expected impact of the MTD detector on the final state objects is summarized in Section 4.2.2. Different MTD performance scenarios, are defined to estimate the impact of the MTD on the HH sensitivity. An MTD resolution of 35 ps is assumed in the nominal scenario, which correspond to the average MTD resolution, weighted on the integrated luminosity, throughout the Phase II. A scenario without the MTD detector is also considered. Finally, a conservative scenario with a MTD resolution of 50 ps is defined.

Table 4.1: HH decay channels included in the projections for HL-LHC with the corresponding branching ratios, and the number of expected (SM) signal events prior to any selection.

HH decay channel	$b\bar{b}b\bar{b}$	$b\bar{b}\tau^+\tau^-$	$b\bar{b}VV(\ell\ell\nu\nu)$	$b\bar{b}\gamma\gamma$	$b\bar{b}ZZ(4\ell)$
Branching ratio(%)	33.9	7.3	1.7	0.26	0.015
Number of events	37000	8000	1830	290	17

4.2.1 Performance scenarios

The performance on the physics objects assumed for the projections is summarized in Ref. [88]. The algorithms implemented in the full CMS simulation used to tune the DELPHES performance are mainly the same used for the TDR studies. The assumed performance on photons and b-jets, the relevant objects for the HH $\to b\bar{b}\gamma\gamma$ search, are reported in the following sections.

Photon reconstruction and identification efficiency

The reconstruction of a photon object in the barrel region is analogous to the one performed in Run 2 presented in Section 2.3.2. The ECAL aging consequent to the radiation damage affects the clustering. However, a preliminary retuning of the clustering parameters [85] shows that the efficiency is within 1% from the value achieved in Run 2, as visible in Fig. 4.2 (left). For the photon identification, the same approach developed for Run 2 described in Section 3.5.2 is used. After preselections based on isolation and shower shape variables, a BDT is trained to discriminate the prompt from the non-prompt photons. The BDT performance is shown in Fig. 4.2 (right). The strategy for the photon reconstruction and identification in the endcaps [86] is analogous. The HGCal reconstruction and identification algorithms exploit both the longitudinal and the transversal shower shape profile to improve the performance. The reconstruction and identification efficiencies for the endcap photons are shown in Fig. 4.3.
In the DELPHES simulation a photon is reconstructed from a neutral energy excess in a simplified version of the electromagnetic calorimeter. The reconstruction and

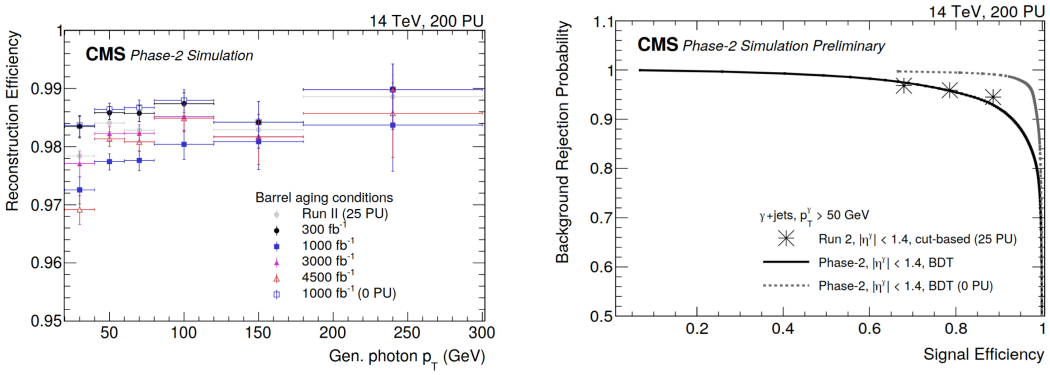


Figure 4.2: *Left:* photon reconstruction efficiency as a function of the photon p_T for different aging scenarios for barrel. *Right:* background rejection as a function of the photon identification efficiency for an average aging scenario for barrel photons. Figures taken from [85].

identification efficiencies on prompt and non-prompt photons are adjusted to match the full CMS simulation. Two working points inclusive of the reconstruction and identification efficiencies are defined. The loose working point provides an efficiency of about 90% and 6% on prompt and non-prompt photons, respectively, in the barrel region with $p_T \sim 60$ GeV which is representative of the H→$\gamma\gamma$ phase-space.

Photon energy resolution

The photon energy resolution is expected to worsen as a consequence of the ECAL aging and the pileup [85]. The corresponding worsening in the $m_{\gamma\gamma}$ resolution for the H→$\gamma\gamma$ peak is of about 54% throughout the HL-LHC, as reported in Table 4.2. In DELPHES, a $m_{\gamma\gamma}$ peak resolution of about 2.9 GeV is assumed, corresponding to the average $m_{\gamma\gamma}$ resolution weighted on the integrated luminosity throughout the HL-LHC phase. A similar $m_{\gamma\gamma}$ resolution is assumed for endcaps photons, consistently with the HGCal TDR [86]. Such a resolution estimate does not take into account the improvements from training the photon energy regression for the Phase II conditions. Assuming in quadrature the same resolution improvement observed in Run 2, the $m_{\gamma\gamma}$ resolution is expected to be of about 2.4 GeV. Therefore, the $m_{\gamma\gamma}$ value is tuned after the simulation to provide a resolution of 2.4 GeV.

Table 4.2: Resolution on the H→$\gamma\gamma$ peak in the $m_{\gamma\gamma}$ distribution at HL-LHC with 200 pileup events for different detector aging scenarios. Table from Ref. [85].

Integrated luminosity	300 fb^{-1}	1000 fb^{-1}	3000 fb^{-1}
H→$\gamma\gamma$ peak resolution	2.8 GeV	2.9 GeV	4.3 GeV

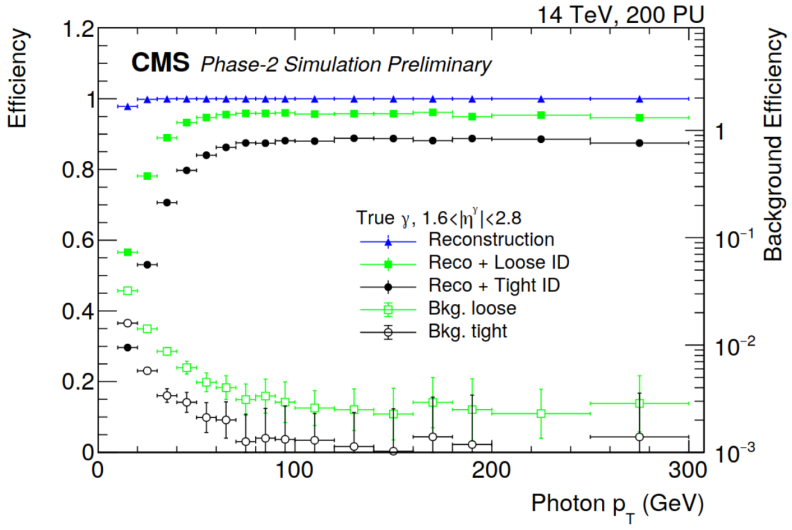


Figure 4.3: Reconstruction and identification efficiency for prompt and non-prompt endcap photons as a function of the photon p_T for different aging scenarios. Figure from [86].

Jet reconstruction and b-jet identification performance

The two b-jets from the HH $\rightarrow b\bar{b}\gamma\gamma$ process are mostly produced at low pseudo-rapidity, thus in the barrel region. Thanks to the upgraded tracker detector, the jet reconstruction efficiency is expected to be stable throughout the Phase II and compatible with the Run 2 performance. The larger pileup will increase the contamination from pileup jets. However, for jets from a H$\rightarrow b\bar{b}$ decay, the b-tag and the high p_T requirements reduce the pileup jet contamination to a negligible level. The jet energy resolution for high p_T jets is expected to be almost insensitive to pileup and to the aging effects, thanks to the tracker and HCAL radiation tolerance. In DELPHES, a loose, medium, and tight working points corresponding to a misidentification probability of light flavor jets of 10%, 1%, and 0.1% are defined.

4.2.2 Impact of the MTD detector

The performance of the final state objects reconstruction can be improved through an optimized usage of the timing information provided by the MTD detector [81]. The electron, muon, and photon reconstruction efficiencies improve of about 3-4% for constant background through the requirement of the time compatibility between the tracks and the primary vertex. This requirement allows the rejection of a large number of pileup tracks improving the isolation for prompt leptons and photons, resulting in an efficiency improvement. An example is visible in the left panel of Fig. 4.4 for a signal made of prompt muons from Z $\rightarrow \mu^+\mu^-$ and a background made of non-prompt muons from $t\bar{t}$ events. The muons with $|\eta| < 1.5$ are considered for

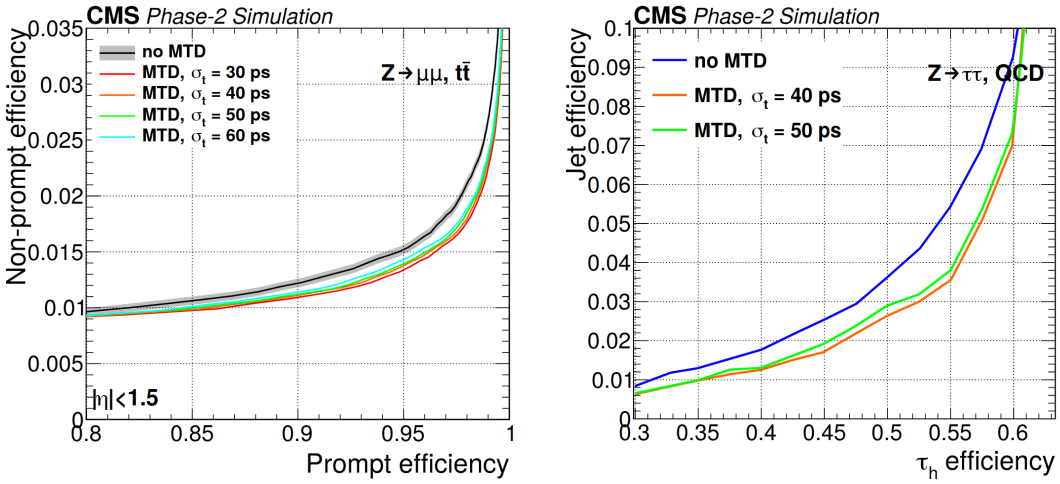


Figure 4.4: *Left:* Selection efficiency on a signal made of prompt muons from Z $\rightarrow\mu^+\mu^-$ as a function of the selection efficiency on a background made of non-prompt muons from $t\bar{t}$ events. Muons with $|\eta|<$ 1.5 are considered for different MTD performance scenarios, including the no MTD case. *Right:* Selection efficiency on a signal made of prompt hadronic tau leptons from Z $\rightarrow\tau^+\tau^-$ as a function of the selection efficiency on a background made of jets from QCD events. The curves are shown for different MTD performance scenarios, including the no MTD case.

different MTD performance scenarios corresponding to begin and end of life.
For the hadronic tau leptons, more affected by the background from jets misidentified as tau leptons, the improvement amounts to 10-15% thanks to the better rejection of pileup tracks in the isolation computation, as shown in the right panel of Fig. 4.4.
Using the same time compatibility criterion for the $\overrightarrow{\not{p}}_T$ calculation, the $|\overrightarrow{\not{p}}_T|$ resolution is expected to improve of about 10% thanks to a better rejection of pileup tracks. The HH $\rightarrow$ $b\bar{b}\tau^+\tau^-$ channel will especially benefit of that because the $|\overrightarrow{\not{p}}_T|$ variable is used for the calculation of the invariant mass of the τ^+ τ^- lepton pair, and in fact dominates the invariant mass resolution, as visible in the left panel of Fig. 4.5. Also the b-tag identification performance is expected to improve of 4-6% through the inclusion of the timing information in the secondary vertexing algorithms that allows a significant rejection of pileup tracks. As shown in the right panel of Fig. 4.5, the gain from MTD is larger for the tight b-tag working points.
In order to estimate the impact of the MTD on the HH sensitivity, the efficiencies and fake rates variations resulting from the improved lepton and photon isolation, and b-tagging, computed in (p_T,η) intervals, are applied to the events for each of the defined MTD scenarios. Although the individual gain on leptons, photons, and b-tag efficiencies is relatively small, the combined gain for the considered HH channels is sizable because the final states consist of two or four b-jets, and possibly leptons or photons.

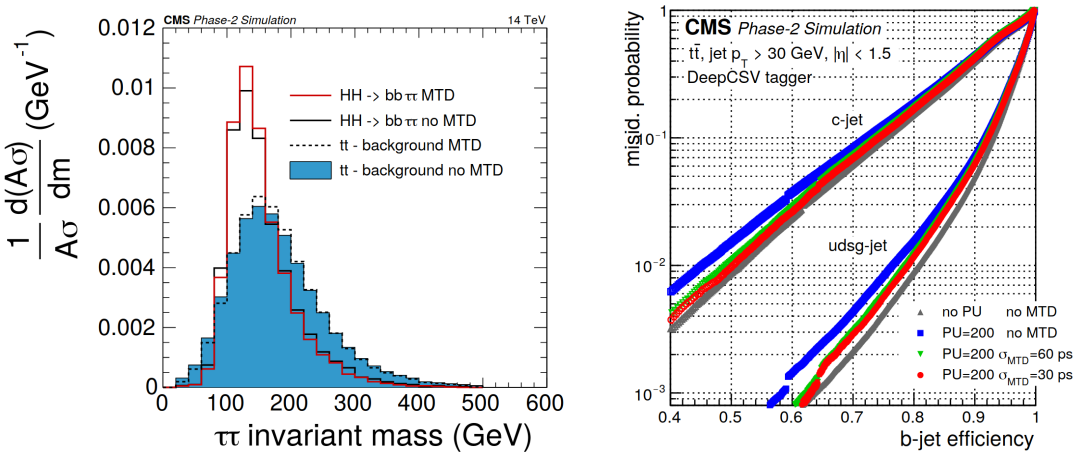


Figure 4.5: *Left:* Invariant mass of the tau leptons pair in HH $\rightarrow$ b$\bar{\text{b}}\tau^+\tau^-$ events with and without the MTD detector. *Right:* b-tag mis-identification probability for c-jets and light flavor jets as a function of the the prompt b-jet efficiency for jets in the barrel and endcap regions on the left and right, respectively, for different MTD performance scenarios.

4.2.3 Systematics uncertainties

A detailed discussion on the systematics uncertainties is give in Ref. [88]. No uncertainties are assumed on the limited size of the MC samples used for the studies. The theoretical uncertainties are expected to improve thanks to the progresses in the theoretical studies exploiting also the large HL-LHC dataset. Therefore, the factorization and the renormalization scale uncertainties are assumed to be reduced by a factor two with respect to the Run 2 values according to Ref. [89].
The systematic uncertainties related to the photon reconstruction and identification efficiency are assumed at the same level of the Run 2 corresponding to about 1% [22]. The large available dataset and the upgraded detector may reduce the uncertainties related to the data-driven background modeling, initial-state radiation, and photon efficiency with respect to the Run 2. Furthermore, the development of the MC simulation algorithms will improve the signal and background modeling. On the other hand, the high pileup level and the detector aging will affect the photon isolation, compensating for the improvements described above. For the same reasons, also the uncertainties related to the photon energy scale and resolution are assumed at the Run 2 level corresponding to 0.5% and 5%, respectively. However, this is a conservative estimate because it does not take into account the pileup mitigation effect of the MTD detector, especially for the photon isolation. The diphoton trigger is assumed to be 100% efficient, and the corresponding uncertainty is included in the photon reconstruction and identification uncertainty.
The uncertainties related to the b-tag efficiency is also considered. The upgraded detector and the large available dataset are assumed to improve the precision of the

b-tag efficiency measurements. A better modeling of the $t\bar{t}$ process would further reduce the uncertainty because it is one of the samples used for the b-tag efficiency measurement. For genuine b-jets with $p_T < 300$ GeV the projected uncertainty amounts to 1%, which is about 30% better than Run 2. Analogous considerations can be made for the jet energy scale and resolution uncertainties. Therefore, the jet energy scale and resolution uncertainties is estimated to be of about 1% and 3%, respectively, corresponding to half of their values in Run 2.
An uncertainty on the total integrated luminosity of 1% is also considered. Such an uncertainty is considered fully correlated across the considered HH decay channels.

4.2.4 Simulated events

The events are simulated using MC generators interfaced with PYTHYA8 for the hadronization and the fragmentation. The full list of used samples for the HH $\rightarrow$ $b\bar{b}\gamma\gamma$ analysis is presented in Table 4.3. For the HH signal, the MADGRAPH5_AMC@NLO generator is used at LO accuracy. The HH signals corresponding to anomalous κ_λ values are obtained through a weighting of the LO events exploiting the generator-level information [21]. The single Higgs production samples with the Higgs boson decaying to a pair of photons are generated using the MADGRAPH5_AMC@NLO generator at NLO QCD accuracy. The $t\bar{t}$ +X and $\gamma(\gamma)$+jets samples are simulated at LO accuracy using the MADGRAPH5_AMC@NLO and Sherpa generators, respectively.
The cross section assumed for the ggHH process matches the NNLO QCD prediction including the finite top quark mass corrections [7]. The cross sections for the single Higgs production mechanisms are set according to Ref. [6]. The $t\bar{t}$H cross section is computed at NLO accuracy in QCD and EW. The qqH and VH cross sections are computed at NNLO accuracy in QCD and NLO accuracy in EW. The ggH cross section is computed at N^3LO accuracy in QCD and NLO accuracy in EW. The assumed branching ratio for the H$\rightarrow\gamma\gamma$ decay is 2.28×10^{-3}.
The cross sections used for the continuum background samples are computed at LO accuracy. A data-simulation agreement for the dominating backgrounds from $\gamma(\gamma)$+jets is fundamental for a precise background modeling. Since data at 14 TeV are not available for a direct comparison, the inclusive scale factors derived for the $t\bar{t}$H analysis with the 2016 and 2017 datasets [90] to adjust the MC agreement with data, reported in Table 4.4, are applied. The $t\bar{t}$H-hadronic signal region is in fact very similar to the HH $\rightarrow$ $b\bar{b}\gamma\gamma$ one, and no big differences are expected increasing the center-of-mass energy from 13 TeV to 14 TeV.

4.3 Analysis strategy for the $HH \rightarrow b\bar{b}\gamma\gamma$ channel

The signal and the background contributions are the same as for the HH $\rightarrow$ $b\bar{b}\gamma\gamma$ analysis with the Run 2 dataset. Therefore, the same strategy described in Chap-

Table 4.3: List of the simulated samples used for the HH $\rightarrow b\bar{b}\gamma\gamma$ projections.

Process	Production cross section (fb)
ggHH	3.669×10^1
$\gamma\gamma$+jets	9.46×10^4
γ+jets ($m_{\gamma\gamma} > 80$ GeV, $p_T^{\gamma} > 40$ GeV, em-enriched)	1.04×10^6
multijets ($m_{\gamma\gamma} > 80$ GeV, $p_T^{\gamma} > 40$ GeV, em-enriched)	1.41×10^8
$t\bar{t}\gamma\gamma$	1.86×10^1
$t\bar{t}\gamma$ (hadronic)	7.92×10^2
$t\bar{t}\gamma$ (semileptonic)	7.70×10^2
$t\bar{t}\gamma$ (leptonic)	6.23×10^2
$t\bar{t}$	8.64×10^5
ggH	5.04×10^4
qqH	4.17×10^3
VH	2.39×10^3
$t\bar{t}$H	6.11×10^2

Table 4.4: k-factors applied to the simulated samples derived from [90].

Process	k-factor
$\gamma\gamma$+jets	1.25
γ+jets ($m_{\gamma\gamma} > 80$ GeV, $p_T^{\gamma} > 40$ GeV, em-enriched)	1.67
multijets ($m_{\gamma\gamma} > 80$ GeV, $p_T^{\gamma} > 40$ GeV, em-enriched)	2.45

ter 3 is used.
The photons and jets selections to define the H→$\gamma\gamma$ and H→$b\bar{b}$ candidates are summarized in Table 4.5 and Table 4.6, respectively. Although the upgraded tracker offers an extended angular coverage, the considered pseudorapidity region is limited to $|\eta|= 2.5$ to increase the signal-over-background ratio. The H→$\gamma\gamma$ candidate is built with the two selected photons with the highest p_T, while the H→$b\bar{b}$ candidate is built using the two jets with the highest b-tag discriminant. The $m_{\gamma\gamma}$ distribution for the selected HH candidate events is shown in Fig. 4.6.
A BDT classifier is trained to separate the HH → $b\bar{b}\gamma\gamma$ events from the $t\bar{t}H(\gamma\gamma)$ events. The most discriminating variables used for the training are the same presented in Section 3.7.1. However, the $|\overrightarrow{\not{p}}_T|$ variable is not used for the training because significant discrepancies are observed in the description of such a high-level variable between the DELPHES and the full CMS simulation. The optimal working point allows a $t\bar{t}H$ rejection of about 25% with an efficiency on signal of about 90%.
A second BDT classifier is trained to separate the HH signal from the continuum background and the single Higgs processes excluding $t\bar{t}H$. The most sensitive training variables are the same described in Section 3.7.2.
The events are classified in three m_X categories whose boundaries are 350 GeV and 480 GeV to improve the sensitivity, especially to the anomalous κ_λ signals. In each m_X category, the BDT scores are used to reject the background-like events and to classify the remaining events in two categories. The boundaries of the two categories are simultaneously optimized to maximize the expected naive significance.
The signal is extracted by means of a maximum likelihood fit of the $m_{\gamma\gamma}$ and m_{jj} distributions which are assumed to be uncorrelated. The signal modeling is analogous to the one described in Section 3.11.2. The m_{jj} and $m_{\gamma\gamma}$ distributions for the continuum background are both modeled with exponential functions. The bias induced on the signal strength by the background modeling is found to be negligible according to the HH → $b\bar{b}\gamma\gamma$ analysis with the 2016 dataset [22]. An example of the expected $m_{\gamma\gamma}$ and m_{jj} distributions for the signal and backgrounds in the high-m_X and high-purity category is shown in Fig. 4.7.

Table 4.5: Photon selections.

	Leading photon	Subleading photon
p_T	> 30 GeV	> 20 GeV
$p_T/m_{\gamma\gamma}$	$> 1/3$	$> 1/4$
Identification	loose ID	
$\|\eta\|$	$\|\eta\|< 1.44$ or $1.57 < \|\eta\|< 2.5$	
$m_{\gamma\gamma}$	100 GeV$< m_{\gamma\gamma} <$ 180 GeV	

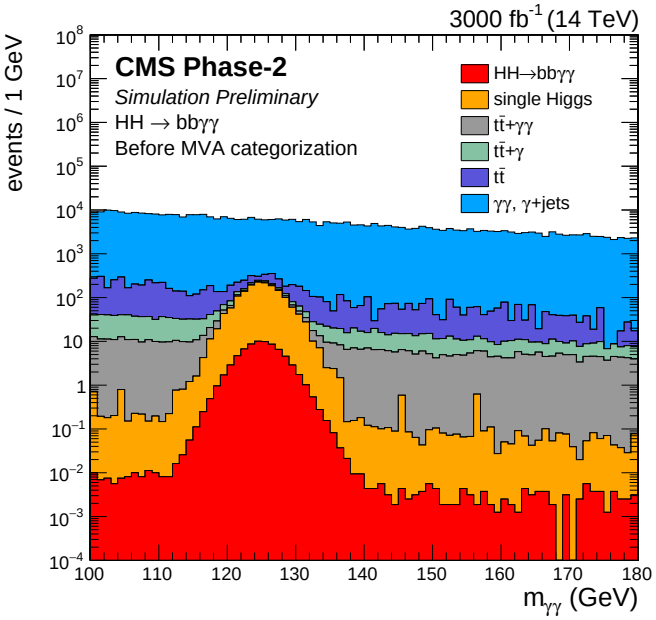


Figure 4.6: $m_{\gamma\gamma}$ distribution for the selected HH candidates for the simulated signal and background events.

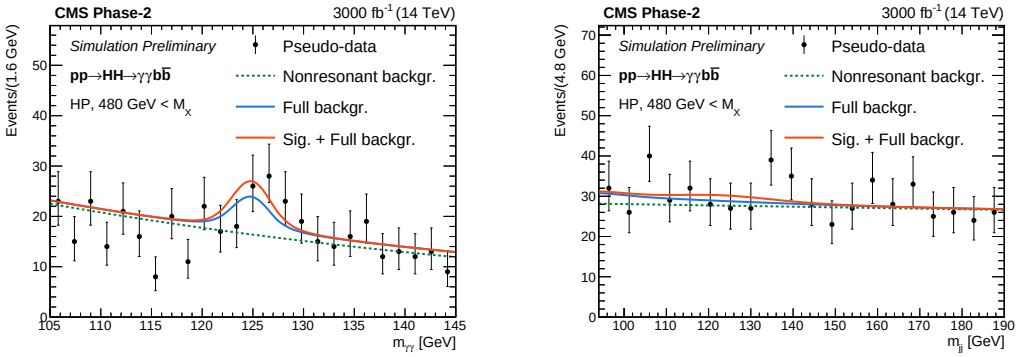


Figure 4.7: Expected $m_{\gamma\gamma}$ and m_{jj} distributions in the high-mass and high-purity category.

Table 4.6: Jet selections.

p_T	> 25 GeV
$\|\eta\|$	< 2.5
$\Delta R(\mathrm{j}, \gamma)$	> 0.4
m_{jj}	80 GeV$< m_{\mathrm{jj}} <$ 200 GeV
b-tag	At least 2 loose b-tagged selected jet

4.4 Prospects for 3000 fb^{-1} of integrated luminosity

The five considered HH decay channels are statistically combined to measure a common HH signal strength. The overlaps of the HH channels across the different signal regions is negligible thanks to the identification requirements and the MVA-based selections. The systematics uncertainties related to the same objects are correlated across the channels while the others are left uncorrelated. The Higgs branching ratios are assumed equal to their SM values.
The expected significance on the SM signal for the considered HH channels and for their combination is shown in Table 4.7 for the three considered MTD scenarios. The expected improvement is large for the channels whose background is dominated by reducible contributions. This is the case of the HH $\to$ b$\bar{\mathrm{b}}\tau^+\tau^-$ channel whose background is dominated by t$\bar{\mathrm{t}}$ events with two misidentified τ leptons. For the HH channels whose background is dominated by prompt-objects, the expected improvement is smaller. This happens for example in the HH $\to$ b$\bar{\mathrm{b}}\gamma\gamma$ channel with a background dominated by $\gamma\gamma$+(b-)jets events. The overall improvement in sensitivity on the HH signal from the MTD detector is estimated of 13%, which would require an additional 26% luminosity without MTD. In the nominal MTD scenario (35 ps resolution), the combined significance is 2.7σ. The most sensitive channels is the HH $\to$ b$\bar{\mathrm{b}}\gamma\gamma$ with an expected significance of 1.9σ.

4.4.1 Prospects for a combination of ATLAS and CMS HH searches

Projection studies for the HH search at HL-LHC have been also performed by the ATLAS experiment. The expected significances for the channels considered as well as the combined significances resulting from the CMS and ATLAS studies are shown in Table 4.8. The significances quoted for the CMS results correspond to the values from Ref. [80] that includes only the impact of the MTD detector on the b-tag efficiency. The studies of the MTD impact on photons and leptons were not completed in time for being included in Ref. [80]. This explains the difference with the results quoted in Table 4.7.

Table 4.7: Expected SM significances (in standard deviations) for the HH channels considered and their combination for different MTD scenarios. The gain in terms of integrated luminosity with respect to the scenario without MTD is also reported.

	NO MTD	MTD with 35 ps resolution	MTD with 50 ps resolution
$HH \to b\bar{b}b\bar{b}$	0.88	0.95	0.94
$HH \to b\bar{b}\tau^+\tau^-$	1.3	1.6	1.48
$HH \to b\bar{b}\gamma\gamma$	1.7	1.9	1.83
$HH \to b\bar{b}VV(\ell\ell\nu\nu)$	0.53	0.58	0.58
$HH \to b\bar{b}ZZ(4\ell)$	0.38	0.42	0.42
Combined	2.4	2.7	2.63
Luminosity gain	-	+26%	+20%

The combined expected significance for the SM signal in ATLAS is 3.0σ, and is dominated by the $HH \to b\bar{b}\gamma\gamma$ and $HH \to b\bar{b}\tau^+\tau^-$ channels. The ATLAS result for the $HH \to b\bar{b}\gamma\gamma$ channel is better than the CMS one because the ATLAS analysis was mainly optimized for the SM signal while the CMS analysis found a compromise between the SM sensitivity and the sensitivity to anomalous κ_λ values as we will see in Section 4.4.2.
A statistical combination of the CMS and ATLAS projected results is also performed in Ref. [80]. The correlation between the common systematic uncertainties, e.g. the luminosity and the theory uncertainties, is assumed to be negligible because such uncertainties have a small impact on the individual results. Since the $HH \to b\bar{b}ZZ(4\ell)$ and the $HH \to b\bar{b}VV(\ell\ell\nu\nu)$ are included only in the CMS study, they are rescaled by 6000 fb^{-1} to estimate in the first approximation also the sensitivity of the corresponding ATLAS channels. The combined expected significance is 4.0σ. Considering only the statistical uncertainty, the expected significance is 4.5σ. Therefore, the expected result is close to the 5σ threshold to claim a discovery. Further improvements not included in this prospects can really push the sensitivity over the discovery threshold. In particular, more efficient analysis approaches, especially in terms of MVA techniques, are expected throughout the next years. Additional decay channels with less but still significant sensitivity, e.g. the $HH \to b\bar{b}WW(jj\ell\nu)$ and the $HH \to WW\gamma\gamma$, will be included in the combination. Finally, analysis categories targeting the qqHH production mode will be included in the analysis.

4.4.2 Sensitivity to anomalous κ_λ values

In the assumption of a SM-like HH signal, a likelihood scan is performed on the κ_λ parameter, as visible in the left panel of Fig. 4.8. The expected κ_λ confidence intervals at 68 and 95% CL derived from the likelihood scan are (0.35, 1.9) and (-0.18, 3.6), respectively. The sensitivity to κ_λ is dominated by the $HH \to b\bar{b}\gamma\gamma$

Table 4.8: Expected significance for the SM HH signal in the exclusive decay channels studied for the HLLHC prospects by ATLAS and CMS and their combination. From [80].

	Stat.-only		Stat.+ Syst	
	ATLAS	CMS	ATLAS	CMS
$HH \to b\bar{b}b\bar{b}$	1.4	1.2	0.61	0.95
$HH \to b\bar{b}\tau^+\tau^-$	2.5	1.6	2.1	1.4
$HH \to b\bar{b}\gamma\gamma$	2.1	1.8	2.0	1.8
$HH \to b\bar{b}VV(\ell\ell\nu\nu)$	-	0.59	-	0.56
$HH \to b\bar{b}ZZ(4\ell)$	-	0.37	-	0.37
Combined	3.5	2.8	3.0	2.6
	Combined		Combined	
	4.5		4.0	

channels mainly because of the four-body mass categorization, as shown in the right panel of Fig. 4.8.

A combination between the ATLAS and the CMS results is performed under the same assumptions listed for the SM HH signal. The combined κ_λ likelihood scan is shown in Fig. 4.9 (left), while the channel-by-channel comparison is visible in Fig. 4.9 (right). The κ_λ parameter is expected to be constrained in $0.52 < \kappa_\lambda < 1.5$ at 68% CL. The second minimum is excluded at 99.4% CL. A large difference between the ATLAS and the CMS sensitivity is visible for $\kappa_\lambda > 5$. This is due to the different optimization for the $HH \to b\bar{b}\gamma\gamma$ channel. In particular, the ATLAS selections, optimized for the SM signal, rejects most of the events with a low m_X value that is the phase-space enhanced in case of a large κ_λ. Instead the CMS analysis defines specific low m_X categories to collect such events.

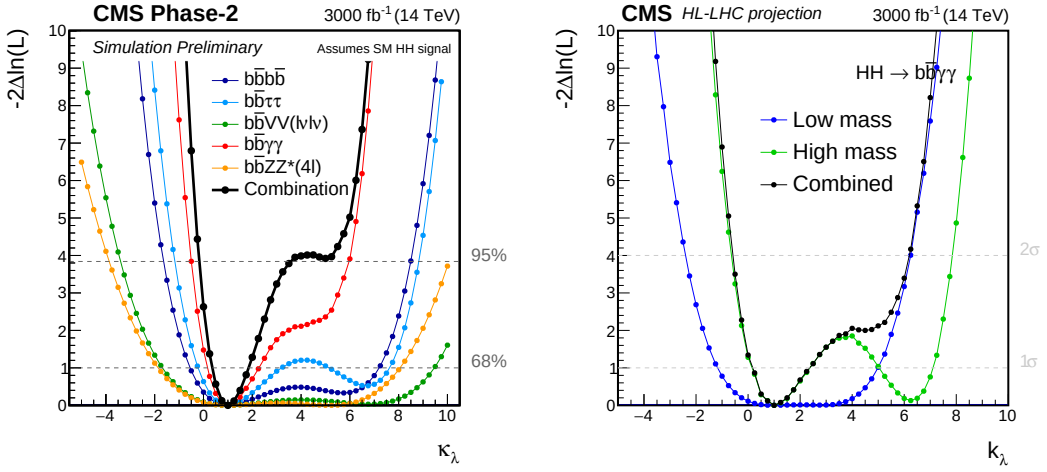


Figure 4.8: *Left:*κ_λ likelihood scan (expected SM) for the considered HH exclusive decay channels, and for their combination. *Right:* κ_λ likelihood scan (expected SM) for the two low-m_X ($m_X < 350$ GeV) and four high-m_X ($m_X > 480$ GeV) categories in blue and green, respectively. The combined likelihood is also shown in black.

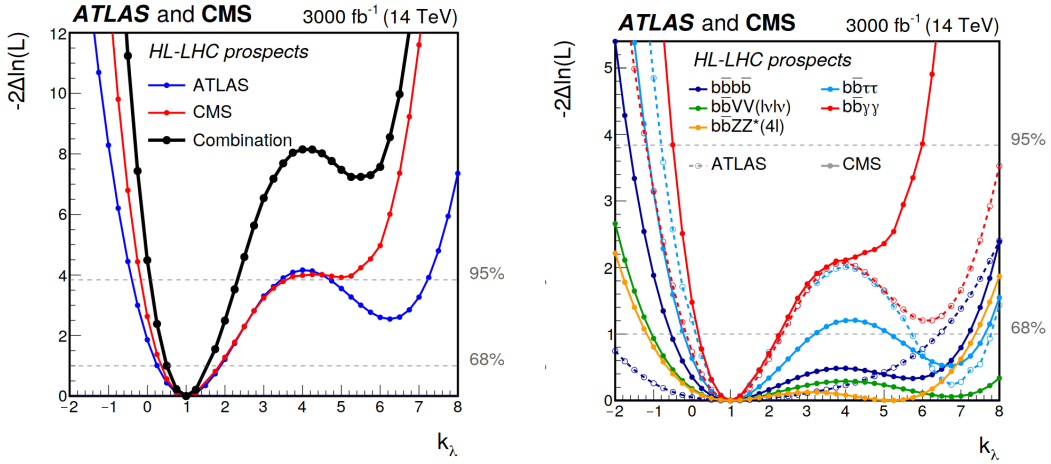


Figure 4.9: *Left:* Expected κ_λ likelihood scan from the CMS and ATLAS HH prospects, and their combination. *Right:* Comparison of the κ_λ likelihood scans between the CMS and ATLAS HH prospects for each considered decay channel. Figures from [80].

Chapter 5

Conclusions

The search for the non-resonant production of Higgs boson pairs in the HH $\rightarrow b\bar{b}\gamma\gamma$ final state is presented. The data analyzed were collected by the CMS detector in proton-proton collisions with a center-of-mass energy of 13 TeV, for a total integrated luminosity of 137 fb^{-1}. The considered double-Higgs production processes are the ggHH, which is the main production process, and the qqHH.
The resolution on the photon energy is a key element to achieve the best sensitivity because it determines the width of the H$\rightarrow\gamma\gamma$ peak in the diphoton mass distribution, which is used for the signal measurement. For this reason, a work is carried out to optimize the performance of the ECAL, which provides the energy measurements for photons (and electrons). In particular, the electrons from the W and Z bosons decays are used to calibrate the relative response of the ECAL channels and to correct for the response variations in time. A refined calibration for the optimized reconstruction of the Run 2 data provides an improvement of about 20 and 10% on the $m_{\gamma\gamma}$ peak resolution for the 2017 and 2018 datasets, respectively, with respect to the preliminary calibrations. For the analysis of the entire Run 2 dataset, this improvement in resolution translates into a gain in the sensitivity to the H$\rightarrow\gamma\gamma$ signal of about 6%, equivalent to a gain in integrated luminosity of about 17 fb^{-1}. This can be fully exploited by the analyses aiming to precision measurements of the Higgs boson parameters, such as the Higgs boson mass measurement in the H$\rightarrow\gamma\gamma$ (and H$\rightarrow$ZZ(4ℓ)) decay channel. The same benefits are predicted for the HH $\rightarrow b\bar{b}\gamma\gamma$ analysis discussed in this thesis. However, and regrettably, the time latency of the data reprocessing in CMS did not made the samples with refined calibration available in time. Still, the analysis conducted with a dataset using preliminary calibrations has marked a significant progress in the HH search. The data analysis exploits a new set of categories, both for HH and $t\bar{t}$H candidate events. A combined analysis of the two set of categories provides a significant improvement in the measurements of the κ_λ and κ_t couplings.
No significant deviations from the SM predictions are found for the HH $\rightarrow b\bar{b}\gamma\gamma$ process in the two production mechanisms considered. In particular, no evidence of the HH signal is found, thus an upper limit on its cross section is set. The observed

(expected) upper limit on the inclusive HH $\rightarrow b\bar{b}\gamma\gamma$ cross section is 7.7 (5.2) times the Standard Model. This result is the most stringent upper limit on the HH cross section set by the CMS experiment to date.
Constraints on anomalous values of the Higgs coupling parameters κ_λ, κ_t, and c_{2V} are also extracted. In the hypothesis of no HH signal and all the other Higgs couplings equal to their SM value, the κ_λ parameter is constrained in the range $[-3.3, 8.5]$ at 95% confidence level. This is the most stringent constraint to the κ_λ parameter from a HH search, to date. Under the same assumptions, the c_{2V} parameter is constrained in the range $[-1.3, 3.5]$ at 95% confidence level. The inclusion of categories targeting specifically the $t\bar{t}H(\gamma\gamma)$ events improves the sensitivity to the κ_λ and κ_t parameters. A two dimensional likelihood scan of the κ_λ and κ_t parameters exploiting the constraints from the HH and the single Higgs boson, especially $t\bar{t}H$, processes is performed. The best-fit values of the (κ_λ,κ_t) parameters are consistent with the SM within the uncertainties. The HH-$t\bar{t}H$ combination presented in this work is intended to be the basis for a future combination of all the sensitive double- and single-Higgs sensitive channels to achieve the maximum sensitivity.
The HH process is one of the key measurement for the physics program of the HL-LHC. The prospects for the HH search are studied for center-of-mass energy of 14 TeV at the HL-LHC. The most sensitive HH decay channels are included in the study. The analysis is developed using a parametric simulation of the Phase-II upgraded CMS detector and optimized for an integrated luminosity of 3000 fb^{-1}. In particular, the analysis for the HH $\rightarrow b\bar{b}\gamma\gamma$ channel is presented. The projected statistical sensitivity of this decay channel for the SM HH signal is of about 1.9σ, the highest among the considered channels. The statistical combination of the HH results from the CMS and the ATLAS projection studies provides an expected significance for the SM HH signal of 4.0σ. This value is below the customary threshold to claim a discovery. However, further optimizations of the analysis strategies, and of the reconstruction and identification algorithms can still improve the sensitivity to the HH signal. Furthermore, production modes (and decay channels) not included in the projection studies, such as the qqHH process can provide additional sensitivity. Therefore, the result of the HH prospect should be interpreted as a motivation to invest more effort to optimize and extend the HH searches in the next years, as well as in the HL-LHC phase.

Appendices

Appendix A

The Standard Model of particles physics

The Standard Model of particle physics (SM) is the theory describing the fundamental interactions between the particles, excluding the gravitational interaction. Today, it is the accepted theory by the physics community since it is able to correctly predict all the experimental observations. However, in the astrophysics and cosmology context there are still phenomena not explained by the Standard Model such as the matter-antimatter asymmetry of the universe, the composition of the dark matter, and the gravitational interaction as described by the general relativity. For this reason, the SM is considered as an effective model valid in a limited energy window of a more general underlying theory. The purpose of the current work at the LHC is to perform precision experimental measurements to compare to the SM prediction, any discrepancy would be a clear sign of the new physics.

A.1 Symmetries of the Standard Model

The Standard Model is built using the paradigms of the quantum field theory [1]. Therefore, the SM lagrangian is invariant with respect to the Lorentz transformation group. In addition, the gauge invariance of the SM lagrangian with respect to specific symmetry groups is assumed to exist. This assumption allows a consistent description of the particles interactions along with their propagation [91]. The specific gauge symmetry groups determine the properties of the interactions as well as the conserved charges of the particles. The SM lagrangian can be written as:

$$\mathcal{L} = \mathcal{L}_{kin} + \mathcal{L}_{QCD} + \mathcal{L}_{EW} \tag{A.1}$$

Thw $\mathcal{L}_{kin}$ term describes the propagation of the particles in the vacuum. The $\mathcal{L}_{QCD}$ term describes the strong interaction among particles, namely the quantum chromodynamics (QCD), presented in detail in Section A.3 while the $\mathcal{L}_{EW}$ describe the electroweak interaction (EW) among particles, presented in detail in Section A.4.

A.2 Elementary particles

An elementary particle is a particle without a substructure. The elementary particles of the SM consist in twelve fermions with spin 1/2, and their corresponding antiparticles, five vector bosons which mediate the interaction between the particles, and one scalar boson named "Higgs boson". Six of the fermions have an additional degree of freedom, called "color", exploited by the strong interaction (see Section A.3). In the same way, the gluon (g) which is the mediator of the strong interaction, has an additional degree of freedom. In particular, eight gluons are distiguishable with respect to the strong interaction. The fermions with a color charge are called quarks, while the other fermions are called leptons if electrically charged, or neutrinos if electrically neutral. The fermions can be classified in three "generations", differing mainly for the mass of the fermions. An overview of the SM particles is given in Table A.1 and Table A.2 for fermions and bosons, respectively. To date, only an upper limit to the neutrino masses have been set, although several running experiments aim to the measurement [92].

Table A.1: Fermions of the SM model.

Particle type (color charge)	Particle name	Mass (GeV)	Electric charge
Quark	d	$4.7 \cdot 10^{-3}$	$-1/3$
	u	$2.2 \cdot 10^{-3}$	$+2/3$
	s	$9.5 \cdot 10^{-1}$	$-1/3$
	c	$1.28 \cdot 10^{0}$	$+2/3$
	b	$4.18 \cdot 10^{0}$	$-1/3$
	t	$1.73 \cdot 10^{2}$	$+2/3$
Lepton or neutrino	ν_e	$< 2 \cdot 10^{-9}$	0
	e	$5.11 \cdot 10^{-4}$	-1
	ν_μ	$< 2 \cdot 10^{-9}$	0
	μ	$1.06 \cdot 10^{-1}$	-1
	ν_τ	$< 2 \cdot 10^{-9}$	0
	τ	$1.78 \cdot 10^{0}$	-1

A.3 Strong interaction

The model of the strong interaction assumes that some of the SM fermions, i.e. the quarks, have an additional degree of freedom which is called "color". In particular, each quark field consist in a triplet of fermionic fields:

$$\Psi_q = \begin{pmatrix} \Psi_R \\ \Psi_G \\ \Psi_B \end{pmatrix} \tag{A.2}$$

Table A.2: Vector and scalar bosons of the SM model.

Particle name	Mass (GeV)	Electric charge	Spin
γ	0	0	1
g	0	0	1
Z	91.2	0	1
W^+	80.4	+1	1
W^-	80.4	−1	1
H	125	0	0

The three components are identified by convention by the red(R), green(G), and blue(B) colors. The lagrangian is assumed to be invariant under any $SU(3)$ gauge transformation acting on the color space. The leptons and the neutrino are instead assumed to be singlets, thus invariant with respect to the $SU(3)$ transformations. The symmetry of the lagrangian introduces a new conserved charge in the interactions, namely the color charge, which can take three different values (R,G,B). For this reason the model of the strong interaction is also known as Quantum Chromo-Dynamics(QCD). The QCD coupling constant is called g_s. The $SU(3)$ symmetry group has eight generators t_a, hence eight gauge bosons fields A^a_μ. The corresponding particles, called gluons, mediate the strong interaction. The SM lagrangian of the strong interactions can be written as:

$$\mathcal{L}_{QCD} = \mathcal{L}_{qg} + \mathcal{L}_{king} \tag{A.3}$$

The $\mathcal{L}_{qg}$ term describes the quark-gluon interaction while the $\mathcal{L}_{king}$ term describes the gluon kinematics. The $\mathcal{L}_{qg}$ term can be written as:

$$\mathcal{L}_{qg} = g_s \bar{\Psi}_q \gamma^\mu t_a \Psi_q A^a_\mu \tag{A.4}$$

This term introduces an interaction between a gluon and a $q\bar{q}$ quark pair with a coupling constant g_s. Each gluon couples only with specific combinations of quark colors, determined by the corresponding generator t_a. The $\mathcal{L}_{king}$ term can be written as:

$$\mathcal{L}_{king} = -\frac{1}{4} F^{\mu\nu}_a F^a_{\mu\nu} \quad \text{with } F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f^{abc} A^b_\mu A^c_\nu \tag{A.5}$$

This introduce the description of the gluon propagation in the vacuum and also gluons self-interaction terms. In particular, tri-gluons and four-gluons interactions arise in the lagrangian with a coupling proportional to g_s and g_s^2, respectively.

A.3.1 Running of the g_s coupling and asymptotic freedom

The QCD model is able to explain the fact that at the energy scale of the regular matter (below 1 GeV), a free quark has never been experimentally observed. In

fact, the quarks exist only in bounded states called hadrons. The hadrons[1] can be barions, such as neutrons and protons, and mesons, such as pions. This can be explained by the running of the g_s coupling constant at different energy scales, consequent to the renormalization of the theory [1]. In particular, the α_s coupling constant, defined as $\alpha_s = g_s^2/4\pi$, diverges for small energy scales, as shown in Fig. A.1. Therefore, at energy scales below 1 GeV the QCD interaction is so intense that it is not possible to break the hadron structure and probe the single quarks. Instead, at high energy scales, the α_s value reduces making possible possible to probe the internal structure of the hadrons. This phenomenon is known as asymptotic freedom. It allows the calculation of precise QCD prediction through a perturbative order expansion. Even in this case, it is not possible to observe a quark in the final state because, after its production, it is immediately[2] subjected to the strong interaction until it forms a set of QCD singlets, i.e. the hadrons. This process takes the name of "hadronization". A precise prediction of the cross section is still possible through a proper definition of the "jet" object to gather the final products of the hadronization.

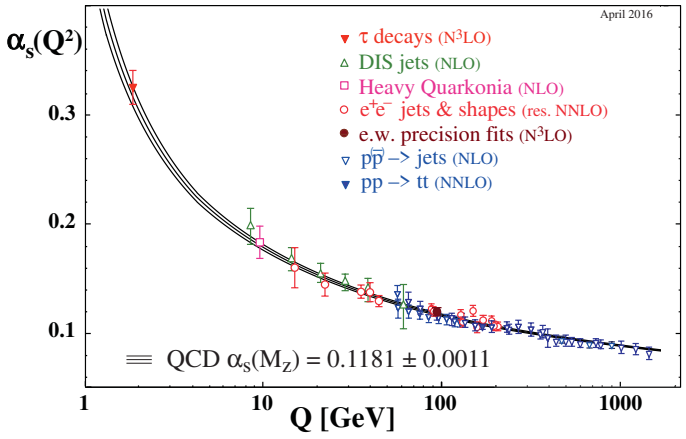


Figure A.1: The dots represent the measured values of the α_s constant at different energy scales, through different physics processes. The black lines represent the theoretical prediction with its uncertainty. Figure from [9].

[1]Other exotic hadron states predicted by the SM, such as the pentaquark, were recently obserbed at the LHCb experiment [93].

[2]With a timescale of the order of 10^{-15} s, thus not experimentally accessible.

A.4 Electroweak interaction

The electroweak interaction is associated to the gauge invariance with respect to the $SU(2)_L \times U(1)_Y$ group. The model assumes that the left-chirality components[3] of the fermions can be grouped in the following doublets:

$$\begin{pmatrix} \mathrm{u}_L^i \\ \mathrm{d}_L^i \end{pmatrix} \text{ with } \mathrm{u}^i = \{\mathrm{u}, \mathrm{c}, \mathrm{t}\} \text{ and } \mathrm{d}^i = \{\mathrm{d}, \mathrm{s}, \mathrm{b}\} \tag{A.6}$$

$$\begin{pmatrix} \ell_L^i \\ \nu_L^i \end{pmatrix} \text{ with } \ell^i = \{\mathrm{e}, \mu, \tau\} \text{ and } \nu^i = \{\nu_e, \nu_\mu, \nu_\tau\} \tag{A.7}$$

The $SU(2)_L$ transformation acts on those doublets with a coupling constant g. Due to this symmetry of the lagrangian, a weak isopin is conserved by the electroweak interactions. The third component of the weak isopin T_3, which is also conserved, characterizes the two components of the doublets. The T_3 value is assumed by convention to be $-1/2$ for the u_L^i quarks and the ℓ_L^i leptons, and $+1/2$ for the d_L^i quarks and the ν_L^i neutrinos. The right-chirality component of each fermion behaves instead as a singlet with respect to the $SU(2)_L$ transformation, hence $T_3 = 0$ for all the right-chirality components.
The $U(1)_Y$ transformation instead acts separately on each left and right chirality component of fermions with a coupling constant called g'. The invariance with respect to the $U(1)_Y$ transformation implies the existence of a conserved charge called weak ipercharge Y. The two components of each chirality-left doublet are assumed to have the same weak ipercharge, while the corresponding two chirality-right singlets can have a different hypercharge. The $SU(2)$ group has three generators, hence three gauge boson fields $W^{1,2,3}$ corresponding to the three Pauli matrices. The $U(1)$ group has one generator, hence one gauge boson field B. The electroweak interaction assumes the leptonic universality, i.e. the same T_3 and Y values for the three generations of leptons. In the quark sector, the universality of the electroweak interaction is delicate because it is assumed on the flavor eigenstates, under consideration for now, which differ from the mass eigenstates (see Section A.6). The interaction terms can be written for simplicity considering only one generic lepton generation:

$$\mathcal{L}_{EW} \supset \mathcal{L}_{CC} + \mathcal{L}_{NC} \tag{A.8}$$

with:

$$\mathcal{L}_{CC} = \frac{g}{\sqrt{2}} \Big[W_\mu^+ \bar{\nu}_L \gamma^\mu e_L + W_\mu^- \bar{e}_L \gamma^\mu \nu_L \Big]; \tag{A.9}$$

$$\mathcal{L}_{NC} = \sum_{\Psi = e_L, \nu_L, e_R, \nu_R} \Big[g T_3 \bar{\Psi} \gamma^\mu \Psi W_\mu^3 + g' \frac{Y}{2} \bar{\Psi} \gamma^\mu \Psi B_\mu \Big] \tag{A.10}$$

The $\mathcal{L}_{CC}$ term defined in Eq. A.9 represents a charged current interaction. In $\mathcal{L}_{CC}$, the W^1 and W^2 fields are re-combined to define the physical boson fields $W_\mu^\pm =$

[3] Given a fermion field Ψ, the left and right chirality components are defined as $\Psi_L = (1-\gamma_5)/2$ and $\Psi_R = (1+\gamma_5)/2$, respectively.

$(W^1_\mu \mp iW^2_\mu)/\sqrt{}(2)$. Only the left-chirality component of the fermions are selected by the charged current interaction. Instead, the $\mathcal{L}_{NC}$ term defined in Eq. A.10 represents a neutral current interaction. Purpose of the theory of the electroweak interactions was also to include the quantum electrodynamics (QED) interaction. However, neither the W^3 nor the B fields features the photon field because they both couple with neutrinos which are electrically neutral. The Weinberg theory [94], predicts that the physical gauge bosons correspond to an orthogonal rotation of the W^3 and B fields:

$$\begin{aligned} B_\mu &= A_\mu \cos(\theta_w) - Z_\mu \sin(\theta_w) \\ W^3_\mu &= A_\mu \sin(\theta_w) + Z_\mu \cos(\theta_w) \end{aligned} \tag{A.11}$$

where θ_w is the rotation angle. Under this assumption the neutral currents of Eq. A.10 can be re-written as:

$$\begin{aligned} \mathcal{L}_{NC} = \sum_{\Psi=e_L,\nu_L,e_R,\nu_R} & \bar{\Psi}\gamma^\mu \Big[g \sin(\theta_w) T_3 + g' \cos(\theta_w) \frac{Y}{2} \Big] \Psi A_\mu + \\ & \bar{\Psi}\gamma^\mu \Big[g \cos(\theta_w) T_3 - g' \sin(\theta_w) \frac{Y}{2} \Big] \Psi Z_\mu \end{aligned} \tag{A.12}$$

The A_μ field is required to be the photon field, thus the corresponding coupling must be proportional to the electric charge of leptons Q and to the QED coupling constant e. Under this constraints, there is still the freedom to set the weak ipercharge of the leptonic doublet by convention to $Y(L) = -1$. With this convention, a useful set of relations among the constants can be derived:

$$g \sin(\theta_w) = g' \cos(\theta_w) = e \tag{A.13}$$

$$Q = T_3 + Y/2 \tag{A.14}$$

Using Eq. A.14 for the ν_R field, it can be noticed that its weak ipercharge is zero. Having all the charges equal to zero, the ν_R field does not couple with any other particle of the SM, thus it is completely irrelevant for the EW (and for the QCD) physics, thus it is assumed not to exist.
The electroweak interaction predicts also interactions between the vector bosons. In fact, the lagrangian kinetic term for the gauge bosons is:

$$\mathcal{L}^{EW}_{kin} = -\frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W^i_{\mu\nu} W^{i\mu\nu} \tag{A.15}$$

with

$$\begin{aligned} B^{\mu\nu} &= \partial^\mu B^\nu - \partial^\nu B^m u \\ W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g \epsilon^{ijk} W^j_\mu W^k_\nu \end{aligned} \tag{A.16}$$

Once $\mathcal{L}^{EW}_{kin}$ is expressed in terms of the physical bosons fields $W^\pm$, A, and Z, tri-boson and four-boson interaction terms arise in the lagrangian.

A.5 Brout-Englert-Higgs mechanism

The lagrangian mass terms for the fermions, for the $W^{\pm}$, and for the Z bosons should be:

$$\mathcal{L}_{mf} = m_f(\bar{\Psi}_L\Psi_R + \bar{\Psi}_R\Psi_L) \text{ for the fermions} \tag{A.17}$$

$$\mathcal{L}_{mV} = m_{\mathrm{W}}W^{\mu+}W^{-}_{\mu} + \frac{m_{\mathrm{Z}}}{2}Z^{\mu}Z_{\mu} \text{ for the gauge bosons} \tag{A.18}$$

However, the $\mathcal{L}_{mf}$ term violates the $SU(2)_L$ gauge symmetry while the $\mathcal{L}_{mV}$ term violates both the $SU(2)_L$ and the $U(1)_Y$ gauge symmetries that were assumed to build the electroweak model. The Brout-Englert-Higgs mechanism allows to naturally introduce the mass terms in the SM lagrangian within the initial assumption of the gauge symmetries. In particular, a new $SU(2)_L$ doublet Φ is defined as:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} \tag{A.19}$$

The Φ^+ and Φ^0 are complex scalar fields. Their superscripts correspond to their electric charge as it will be proven later on. The Φ field is introduced in the lagrangian through a kinetic term:

$$\mathcal{L}_{\Phi kin} = D_{\mu}\Phi D^{\mu}\Phi \tag{A.20}$$

and a "potential" term:

$$V(\Phi) = -\mu^2\Phi^{\dagger}\Phi + \lambda(\Phi^{\dagger}\Phi)^2 \text{ with } \mu^2 > 0 \text{ and } \lambda > 0 \tag{A.21}$$

The $V(\Phi)$ potential has a set of degenerate minima defined by the condition:

$$|\Phi|^2 = \frac{\mu^2}{2\lambda} = \frac{v^2}{2} \text{ with } v = \mu/\sqrt{\lambda} \tag{A.22}$$

Therefore, $v/\sqrt{2}$ is the vacuum expectation value of the Φ field. The generic Φ vacuum state can be written as:

$$\langle\Phi\rangle = \frac{1}{\sqrt{2}}e^{\frac{i\sigma_j\alpha_j(x)}{v}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{A.23}$$

Where $\sigma_{1,2,3}$ are the three Pauli matrices while $\alpha_{1,2,3}$ and H are real scalar fields. The $SU(2)_L$ symmetry of the lagrangian ensures that the specific direction chosen by the vacuum state in the $SU(2)_L$ space is completely irrelevant for the physics. Therefore, the following vacuum state can be selected for simplicity:

$$\langle\Phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{A.24}$$

The kinetic term of Eq. A.20 can be expresses with respect to the Φ vacuum state of Eq. A.24 obtaining[4] :

$$\mathcal{L}_{\Phi kin} = \frac{1}{2}\partial^{\mu}H\partial_{\mu}H + \left(1+\frac{H}{v}\right)^{2}\left[\frac{g^{2}v^{2}}{4}W^{+\mu}W^{-}_{\mu} + \frac{1}{2}\frac{(g^{2}+g'^{)}2v^{2}}{4}Z^{\mu}Z^{\mu}\right] \tag{A.25}$$

This result has two important consequences:

- The non-zero vacuum expectation value of the Φ field introduces in the SM lagrangian the mass terms for the $W^{\pm}$ and Z bosons. The $W^{\pm}$ and Z bosons masses can be related to the v constant as:

$$m_{\mathrm{W}} = \frac{gv}{2} \quad \text{and} \quad m_{\mathrm{Z}} = \frac{\sqrt{g^{2}+g'^{2}}v}{2} = \frac{m_{\mathrm{W}}}{\cos(\theta_{w})} \tag{A.26}$$

  In addition, the v value can be related to the Fermi constant G_F, experimentally measured with a very good precision:

$$v = \sqrt{\frac{1}{\sqrt{2}G_{F}}} = 246 \text{ GeV} \tag{A.27}$$

- In the lagrangian, new interaction terms between one or two H fields and a vector boson pair arise. The corresponding vertices are proportional to the squared mass of the vector bosons.

In the same way, the potential term of Eq. A.21 can be expresses with respect to the Φ vacuum state of Eq. A.24 obtaining:

$$V(\langle\Phi\rangle) = \frac{1}{2}(2\lambda v^{2})H^{2} + \lambda v H^{3} + \frac{\lambda}{4}H^{4} - \frac{\lambda}{4}v^{4} \tag{A.28}$$

This result has important consequences:

- A mass term for the H field arise in the lagrangian. Therefore, the H scalar field describes a boson with spin 0, without electric charge, and with a mass $m_{\mathrm{H}} = \sqrt{2\lambda}v$. This particle takes the name of Higgs boson.
- Tri-linear and quadri-linear H self-coupling terms arise in the lagrangian with a coupling constant proportional to λv and $\lambda/4$, respectively.

It is worth to notice that the functional form for the Higgs boson potential of Eq. A.21 was chosen arbitrarily as the lowest order polynomial ensuring the $SU(2)_{L}\times U(1)_{Y}$ gauge symmetries and providing a vacuum expectation value different from zero. The actual functional form can differ. Therefore, in order to test the accuracy of the model predictions, it is fundamental a precision measurement of the m_{H} value and of the trilinear and the quadrilinear Higgs boson self-coupling constants.

[4] An omitted step consists in requiring no mass term for the photon field. This demands that the weak ipercharge of the Φ field is $Y(\Phi) = -1$. Consequently, the Φ field electric charges are $Q(\Phi^{0}) = 0$ and $Q(\Phi^{+}) = +1$.

A.6 Yukawa interaction of the Higgs boson with fermions

The Brout-Englert-Higgs mechanism introduces in the lagrangian also the mass terms for the fermions. The fundamental assumption is that the fermion flavor eigenstates defined as the states diagonalizing the electroweak interaction are different from the mass eigenstates defined as the states diagonalizing the mass term[5]. In particular, Yukawa-type interaction terms [95] can be added to the SM lagrangian:

$$\begin{aligned}\mathcal{L}_{Yuk} = &- \Gamma_d^{ij} \bar{Q'}^i_L \Phi d'^j_R - \Gamma_u^{ij} \bar{Q'}^i_L \Phi_c u'^j_R + \text{h.c.} + \\ &- \Gamma_l^{ij} \bar{L'}^i_L \Phi l'^j_R + \text{h.c.}\end{aligned} \tag{A.29}$$

Where Q'^i_L is the flavor eigenstate of the i-th left-chirality quark doublet, u'^j_R and d'^j_R are the flavor eigenstates of the j-th right-chirality quark singlets of up-type and down-type, respectively. L^i_L is the flavor eigenstate of the i-th left-chirality lepton doublet, while l'^j_R is the flavor eigenstate of the j-th right-chirality lepton. Finally, Γ_u, Γ_d, and Γ_l are generic 3×3 complex matrices. The term including the ν_R fields is not added because the SM assumes that those particles do not exists.
The Eq. A.29 can be expressed with respect to the Φ vacuum state obtaining:

$$\mathcal{L}_{Yuk} = -\left(1 + \frac{H}{v}\right)\left(M_d^{ij} \bar{d'}^i_L d'^j_R + M_u^{ij} \bar{u'}^i_L u'^j_R + M_l^{ij} \bar{l'}^i_L l'^j_R + \text{h.c.}\right) \tag{A.30}$$

with $M_f = \Gamma_f v/\sqrt{2}$ for each fermion type f. The M_f matrices can be diagonalized to give the mass terms through unitary transformations of the flavor eigenstates f' to give the corresponding mass eigenstates f:

$$f'_{Li} = (U_L^f)_{ij} f_{Lj} \text{ and } f'_{Ri} = (U_R^f)_{ij} f_{Rj} \tag{A.31}$$

Resulting in:

$$\mathcal{L}_{Yuk} = \sum_f m_f (\bar{f}_L f_R + \bar{f}_R f_L)\left(1 + \frac{H}{v}\right) \tag{A.32}$$

This result has important consequences:

- The spontaneous symmetry breaking mechanism induces also the mass terms for the fermions. However, this mechanism does not provide any prediction of the mass values that have to be experimentally measured.
- Interaction terms between the Higgs boson and the fermions are predicted by the theory with a coupling constant proportional to m_f/v.

[5] As shown later, the strong interaction is diagonal both on the flavor and on the mass eigenstates.

The SM lagrangian can be expressed with respect to the fermion mass eigenstates. The unitary transformations leave invariant the kinetic terms, the QCD terms, and the EW neutral current terms. Instead, the EW charged current terms are modified. In particular for the quarks:

$$\mathcal{L}_{CC} = \frac{g}{\sqrt{2}} \bar{u}_L^i V_{CKM}^{ij} \gamma_\mu d_L^j W^\mu + \text{h.c.} \tag{A.33}$$

Where V_{CKM} is a unitary 3×3 matrix. Therefore, the charged current interaction mixes quarks of different generations with a coupling proportional to the corresponding entrance of the V_{CKM} matrix. For the leptonic sector, in the assumption of no ν_R field, an additional degree of freedom can be used to cancel the mixing of leptons in $\mathcal{L}_{CC}$. However, under that assumption the Yukawa interaction does not allow the inclusion of the mass terms for the neutrinos. Instead, in case the ν_R field is assumed to exist, the Yukawa interaction is able to provide the mass terms also to the neutrinos and a leptonic mixing in the charge current term arise in the lagrangian analogously to the quark sector. The corresponding mixing matrix is called V_{PMNS}.

Appendix B

Cross sections parametrizations for anomalous couplings

B.1 HH cross section up to QCD NLO

It is possible to obtain a differential description of the ggHH and qqHH cross sections variations for anomalous couplings values at the NLO through the combination of a limited number of ggHH and qqHH samples simulated with different coupling values.
As visible in Fig. 1.8, two Feynman diagrams contribute to σ_{ggHH} at the LO. Let T be the amplitude of the diagram with tri-Higgs vertex, and B the amplitude of the diagram with a box loop of top quarks. The σ_{ggHH} dependence from the κ_λ and κ_t parameters can be parametrized as:

$$\sigma_{\text{ggHH}} = \kappa_\lambda^2 \kappa_\text{t}^2 t + \kappa_\text{t}^4 b + \kappa_\text{t}^3 \kappa_\lambda i = \kappa_\text{t}^4 \Big[\Big(\kappa_\lambda / \kappa_\text{t} \Big)^2 t + b + \Big(\kappa_\lambda / \kappa_\text{t} \Big) i \Big] \tag{B.1}$$

with $t = |T|^2$, $b = |B|^2$, and $i = (T^* B + B^* T)$. The first two terms correspond to the squared modules of the two amplitudes while the third term correspond to their interference. Whereas additional diagrams have to be included to compute σ_{ggHH} at the NLO in QCD, the dependence on the κ_λ and κ_t parameters can still be factorized as in Eq. B.1. Furthermore, the relation is also valid in selecting a specif regions of the kinematic phase-space, thus differentially. Therefore, it is possible to differentially describe $\sigma_{\text{ggHH}}(\kappa_\lambda, \kappa_\text{t})$ combining only the three MC samples corresponding to the t, b, and i terms.
More generally, a full description of $\sigma_{\text{ggHH}}(\kappa_\lambda, \kappa_\text{t})$ can be achieved using a generic set of three ggHH MC samples $\vec{s} = (s_1, s_2, s_3)$ generated with independent $(\kappa_\lambda, \kappa_\text{t})$ couplings, i.e. different $\kappa_\lambda / \kappa_\text{t}$ ratios. In particular, s_i can be written as:

$$s_i(\kappa_{\lambda i}, \kappa_{\text{t}i}) = \vec{c}_i^{\,\text{T}} \vec{v} \text{ with } c_i = ({\kappa_\lambda}_i^2 {\kappa_\text{t}}_i^2 \ , \ {\kappa_\text{t}}_i^4 \ , \ {\kappa_\text{t}}_i^3 \kappa_{\lambda i}) \text{ and } v = (t \ , \ b \ , \ i) \tag{B.2}$$

thus defining the 3×3 matrix $C = (\vec{c}_1, \vec{c}_2, \vec{c}_3)$:

$$\vec{s} = C \vec{v} \tag{B.3}$$

Choosing three s_i samples with different κ_λ/κ_t ratios ensures that the C matrix is invertible, thus:

$$\vec{v} = C^{-1}\vec{s} \tag{B.4}$$

Analogously to Eq. B.2, a generic sample $s'(\kappa_\lambda, \kappa_t)$ can be described as:

$$s'(\kappa_\lambda, \kappa_t) = \vec{c}^{\,T} C^{-1}\vec{s} \tag{B.5}$$

This is particularly useful to overcome issues related to the limited size of the MC samples producing large uncertainties in the final result. For example, a large cancellation among the three terms of Eq. B.1 is required to describe $\sigma_{ggHH}(\kappa_\lambda, \kappa_t)$ in correspondence of its minimum at $\kappa_\lambda \sim 2.45$ (κ_t =1).
The procedure to describe the qqHH cross section as a function of $(\kappa_\lambda, c_V, c_{2V})$ is the same. Since three diagrams contribute to σ_{VBFHH} at the LO (see Fig. 1.14), the dependence on $(\kappa_\lambda, c_V, c_{2V})$ can be parametrized by the sum of six terms:

$$\sigma_{VBFHH} = c_V^2\kappa_\lambda^2 a + c_V^4 b + c_{2V}^2 c + c_V^3\kappa_\lambda i_{ab} + c_V c_{2V}\kappa_\lambda i_{ac} + c_V^2 c_{2V} i_{bc} \tag{B.6}$$

Where a, b, and c represent the square module of the three amplitudes while i_{ab}, i_{ac}, and i_{bc} represent the three interference terms. Therefore, the combination of six independent MC samples are used to differentially describe σ_{VBFHH} (κ_λ,c_V,c_{2V}). The sample combination procedure is implemented within the CMS data analysis framework. Therefore, provided the set of MC samples as input, the statistic interpretation of the data in terms of anomalous couplings is handled automatically by the framework. The samples used for the ggHH and qqHH description are reported in Table B.1. In order to validate this method, additional MC samples were generated, one for the ggHH process and one for the qqHH process. It was verified that the kinematics distributions of the additional sample were correctly described by the samples combination. Also the expected upper limits on σ_{VBFHH} and σ_{ggHH} derived with the additional samples were found to be compatible to the corresponding limits derived with the samples combination.

B.2 Single Higgs production cross section

The dependence of the single Higgs production cross sections on the κ_λ and c_V parameters is parametrized accordingly to the κ-framework, summarized in Section 1.3.2. The ggH vertex is fully resolved. Therefore, no BSM particles are assumed to contribute in the loop to the gluon-fusion mechanism, so that the vertex dependence on κ_t is fully expanded. The c_{2V} parameter is not considered because the impact on the single Higgs cross sections is negligible. Instead, the κ_λ dependence is parametrized following the prescriptions of [13]. In particular, the cross section for a given Higgs production mechanism is parametrized as:

$$\begin{aligned} &\frac{\sigma}{\sigma^{SM}} = \Big[K(\kappa_t, c_V) + (\kappa_\lambda - 1)C_1/K_{EW}\Big] Z_H^{BSM} \\ &\text{with } Z_H^{BSM} = \frac{1}{1 - ({\kappa_\lambda}^2 - 1)\delta Z_H} \end{aligned} \tag{B.7}$$

Table B.1: List of the simulated ggHH and qqHH samples to describe the anomalous couplings. All the samples are generated at the QCD NLO using POWHEG BOX2.

Mechanism	Couplings values	Cross section (fb) (including k-factor)
ggHH	$\kappa_\lambda = 1$, $\kappa_t = 1$	$3.105 \cdot 10^1$
ggHH	$\kappa_\lambda = 0$, $\kappa_t = 1$	$6.973 \cdot 10^1$
ggHH	$\kappa_\lambda = 2.45$, $\kappa_t = 1$	$1.312 \cdot 10^1$
ggHH (for validation)	$\kappa_\lambda = 5$, $\kappa_t = 1$	$9.117 \cdot 10^1$
qqHH	$\kappa_\lambda = 1$, $c_V = 1$, $c_{2V} = 1$	$1.75 \cdot 10^0$
qqHH	$\kappa_\lambda = 1$, $c_V = 1$, $c_{2V} = 2$	$1.42 \cdot 10^1$
qqHH	$\kappa_\lambda = 2$, $c_V = 1$, $c_{2V} = 1$	$1.42 \cdot 10^0$
qqHH	$\kappa_\lambda = 0$, $c_V = 1$, $c_{2V} = 1$	$4.61 \cdot 10^0$
qqHH	$\kappa_\lambda = 1$, $c_V = 1.5$, $c_{2V} = 1$	$6.60 \cdot 10^1$
qqHH	$\kappa_\lambda = 1$, $c_V = 1$, $c_{2V} = 0$	$2.71 \cdot 10^1$
qqHH (for validation)	$\kappa_\lambda = 1$, $c_V = 0.5$, $c_{2V} = 1$	$1.08 \cdot 10^1$

where $K(\kappa_t, c_V)$ is the cross section scaling derived by the κ-framework. The term proportional to $(\kappa_\lambda - 1)$ represents the interference term between the LO amplitude and the NLO-electroweak amplitude(s) dependent on the κ_λ parameter. C_1 and K_{EW} are coefficients derived for each specific process describing the intensity of the interference term. Their values is reported in Table B.2. Finally, Z_H^{BSM} represents a universal κ_λ-dependent factor corresponding to the wavefunction renormalization in case of an anomalous κ_λ value. The computed value of δZ_H is $-1.536 \cdot 10^{-3}$ and it is universal for all the processes.
The validity of the formula is limited to small deviations of κ_t (and c_V) from the unity. The perturbative order expansion of the formula is in fact not fully consistent since the κ-framework parametrization is at the LO in κ_t (and c_V) while the κ_λ effects enter at the NLO-electroweak. The κ_t (and c_V) contributions at the NLO-electroweak, whose value is not available to date, are neglected in the assumptions of small deviations of κ_t (and c_V) from one. In particular, the formula is valid for κ_t or c_V deviations from the unity within 30%. This threshold is inferred from the ranges of the figures presented in [13]. As visible in Fig. B.1 for the $t\bar{t}H$, the Eq. B.7 can give unphysical scalings (negative) outside the defined renge of validity. It must be stressed that the limited validity of the formula does not affect any one-dimensional parameter scan because the perturbative order expansion in each single parameter is fully consistent.

B.3 Higgs branching ratios scaling

The $H\to\gamma\gamma$ and $H\to b\bar{b}$ branching ratio variations for anomalous couplings values are taken into account to provide a consistent description of the phenomenon. The

Table B.2: Values of the C_1 and K_{EW} coefficients to scale the single Higgs cross sections for anomalous κ_λ values from [13].

Channels	ggH	qqH	ZH	WH	$t\bar{t}H$	tHq
$C_1(\%)$	0.66	0.63	1.19	1.03	3.52	0.91
K_{EW}	1.049	0.932	0.947	0.93	1.014	0.95

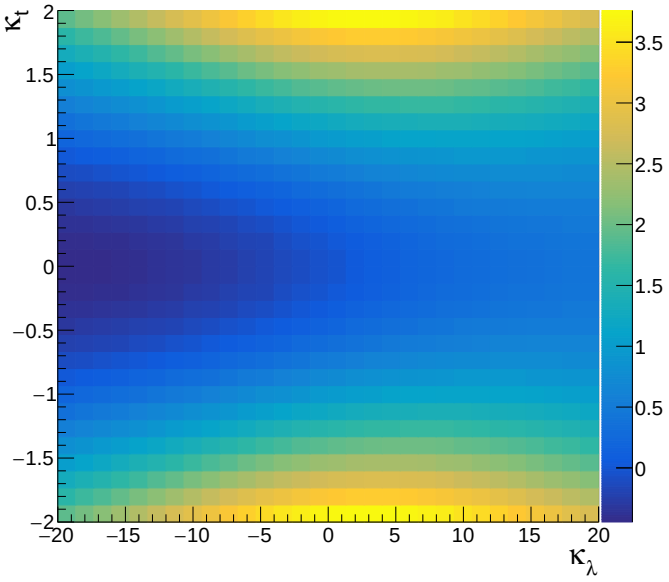


Figure B.1: The value of σ/σ^{SM} for the $t\bar{t}H$ process as a function of the κ_λ and κ_t couplings computed using Eq. B.7.

scalings of the Higgs boson decay widths for anomalous κ_t and c_V are parametrized accordingly to the κ-framework (Section 1.3.2). The impact of an anomalous κ_λ is also included following the prescriptions of [13]. In particular, the decay width to a given final state f, Γ_f is scaled as:

$$\frac{\Gamma_f}{\Gamma_f^{SM}} = Z_H\Big[K(\kappa_t, c_V) + \kappa_\lambda C_1^f\Big] \tag{B.8}$$

and the total Higgs decay width is scaled accordingly:

$$\frac{\Gamma_{tot}}{\Gamma_{tot}^{SM}} = \frac{\sum_f \Gamma_f}{\sum_f \Gamma_f^{SM}} \tag{B.9}$$

The Z_H factor represent the wavefunction resummation contribution as in Eq. B.7. The Z_H factor is irrelevant for the branching ratio calculation because it cancels out in the ratio Γ_f/Γ_{tot}. The $K(\kappa_t, c_V)$ factor represents the branching ratio scaling from the κ-framework. For the $H\to\gamma\gamma$ decay width (and also $H \to gg$) no BSM

particles are assumed to contribute in the loop. Therefore, the $K(\kappa_t, c_V)$ factor for the H→γγ decay width is fully expanded as a function of the κ_t and c_V parameters. The term of Eq. B.8 proportional to κ_λ represents the interference term between the LO amplitude and the NLO amplitudes dependent on κ_λ, if any. The coefficient C_1^f are reported in Table B.3. Due to the small value of the coefficients, the effects of κ_λ on the decay width is at most of few percent.
Figure B.2 shows that the H→γγ branching ratio is strongly dependent on κ_t and c_V while the effect of κ_λ is almost negligible. The H→b$\bar{\text{b}}$ decay width, at the considered perturbative order, does not depend on κ_t and κ_λ. However, Γ_{tot} depends on κ_t and very weakly also on κ_λ. Therefore, also the H→b$\bar{\text{b}}$ branching ratio depends on κ_t, and very weakly on κ_λ.

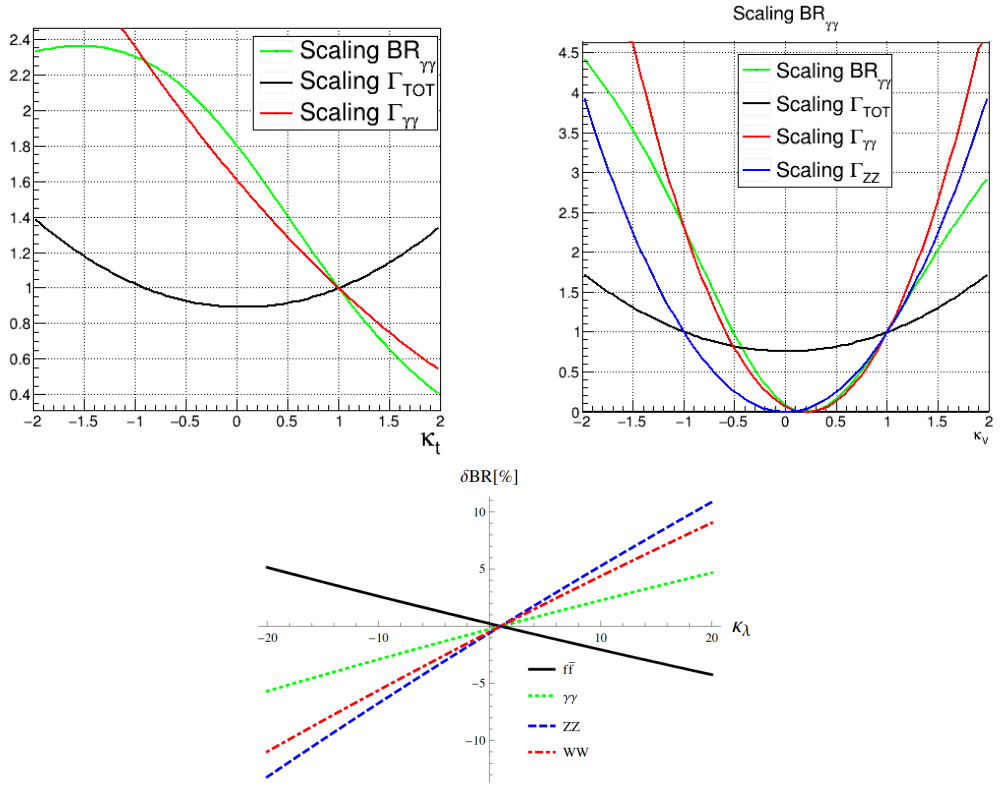


Figure B.2: Higgs boson decay width scalings as a function of κ_t, c_V, and κ_λ on top-left, top-right, and bottom, respectively. Bottom figure taken from [96].

Table B.3: Values of the C_1^f coefficients to scale the (on-shell) Higgs boson partial decay widths for anomalous κ_λ values. The values are taken from [96].

Channels	$\gamma\ \gamma$	Z Z	W W	f $\bar{\text{f}}$	$g\ g$
C_1^f	0.49	0.83	0.73	0	0.66

Bibliography

[1] Michael E Peskin and Daniel V Schroeder. *An introduction to quantum field theory*. Westview, Boulder, CO, 1995.

[2] S. Chatrchyan et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30 – 61, 2012.

[3] G. Aad et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1 – 29, 2012.

[4] Albert M Sirunyan et al. Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(arXiv:1809.10733. CMS-HIG-17-031-005):421. 67 p, Sep 2018.

[5] G. Aad et al. Combined measurements of higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the atlas experiment. *Phys. Rev. D*, 101:012002, Jan 2020.

[6] D. de Florian, C. Grojean, F. Maltoni, C. Mariotti, et al. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. Oct 2016.

[7] Massimiliano Grazzini, Gudrun Heinrich, Stephen Jones, Stefan Kallweit, Matthias Kerner, Jonas M. Lindert, and Javier Mazzitelli. Higgs boson pair production at NNLO with top quark mass effects. *JHEP*, 05(CERN-TH-2018-044):059. 19 p, Mar 2018.

[8] Frédéric A. Dreyer and Alexander Karlberg. Vector-boson fusion higgs pair production at n^3LO. *Phys. Rev. D*, 98:114016, Dec 2018.

[9] P A Zyla et al. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. 083C01.

[10] Johannes Braathen and Shinya Kanemura. Leading two-loop corrections to the higgs boson self-couplings in models with extended scalar sectors. *The European Physical Journal C*, 80, 03 2020.

[11] Stefano Vita, Christophe Grojean, Giuliano Panico, Marc Riembau, and Thibaud Vantalon. A global view on the Higgs self-coupling. *JHEP*, 9, 04 2017.

[12] J. H. Kuhn, A. Scharf, and P. Uwer. Weak interactions in top-quark pair production at hadron colliders: An update. *Phys. Rev. D*, 91:014020, Jan 2015.

[13] Fabio Maltoni, Davide Pagani, Ambresh Shivaji, and Xiaoran Zhao. Trilinear Higgs coupling determination via single-Higgs differential measurements at the LHC. *Eur. Phys. J.*, C77(12):887, 2017.

[14] Fedor Bezrukov and Mikhail Shaposhnikov. Why should we care about the top quark Yukawa coupling? *J. Exp. Theor. Phys.*, 120:335–343, 2015.

[15] Da Liu, Ian Low, and Carlos E. M. Wagner. Modification of higgs couplings in minimal composite models. *Phys. Rev. D*, 96:035013, Aug 2017.

[16] Albert M Sirunyan et al. Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C*, 79(5):421, 2019.

[17] Fady Bishara, Roberto Contino, and Juan Rojo. Higgs pair production in vector-boson fusion at the LHC and beyond. *Eur. Phys. J. C*, 77(arXiv:1611.03860. 7):481. 42 p, Nov 2016.

[18] T.L. Barklow, S. Dawson, H.E. Haber, and J.L. Siegrist, editors. *Electroweak symmetry breaking and new physics at the TeV scale*, volume 16. 1996.

[19] S Heinemeyer et al. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. CERN Yellow Reports: Monographs. Jul 2013.

[20] Alexandra Carvalho et al. Higgs Pair Production: Choosing Benchmarks With Cluster Analysis. *JHEP*, 04(arXiv:1507.02245):126. 26, Jul 2015.

[21] Alexandra Carvalho et al. Analytical parametrization and shape classification of anomalous HH production in the EFT approach. Technical report, 2016. LHCHXSWG-2016-001, arXiv:1608.06578.

[22] Albert M Sirunyan et al. Search for Higgs boson pair production in the $\gamma\gamma\mathrm{b\bar{b}}$ final state in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 788:7–36. 30 p, Jun 2018.

[23] Albert M Sirunyan et al. Measurements of $\mathrm{t\bar{t}H}$ production and the CP structure of the Yukawa interaction between the Higgs boson and top quark in the diphoton decay channel. *Phys. Rev. Lett.*, 125(6):061801. 18 p, Mar 2020.

[24] G. Aad et al. *cp* properties of higgs boson interactions with top quarks in the $t\bar{t}h$ and th processes using $h \to \gamma\gamma$ with the atlas detector. *Phys. Rev. Lett.*, 125:061802, Aug 2020.

[25] Georges Aad et al. Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector. *Phys. Lett. B*, 800(arXiv:1906.02025):135103. 23 p, Jun 2019.

[26] Thomas Sven Pettersson and P Lefèvre. The Large Hadron Collider: conceptual design. Technical Report CERN-AC-95-05-LHC, Oct 1995.

[27] S Chatrchyan et al. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004–S08004, aug 2008.

[28] V Karimäki et al. *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[29] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

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

[31] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[32] Tai Sakuma. Cutaway diagrams of CMS detector. May 2019.

[33] A Dominguez et al. CMS Technical Design Report for the Pixel Detector Upgrade. Technical Report CERN-LHCC-2012-016. CMS-TDR-11, Sep 2012.

[34] P. Lecoq et al. Lead tungstate (pbwo4) scintillators for lhc em calorimetry. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 365(2):291 – 298, 1995.

[35] D. Bailleux et al. Hamamatsu apd for cms ecal: quality insurance. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 518(1):622 – 625, 2004. Frontier Detectors for Frontier Physics: Proceedin.

[36] K. W. Bell et al. Vacuum phototriodes for the cms electromagnetic calorimeter endcap. *IEEE Transactions on Nuclear Science*, 51(5):2284–2287, 2004.

[37] A.M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(CMS-PRF-14-001. CMS-PRF-14-001-004. 10):P10003. 82 p, Jun 2017.

[38] Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s}$= 8 TeV. *Journal of Instrumentation*, 10(06):P06005–P06005, jun 2015.

[39] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45(5):910–927, 2000.

[40] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008.

[41] Pileup mitigation at CMS in 13 TeV data. Technical Report CMS-PAS-JME-18-001, CERN, Geneva, 2019.

[42] Q Ingram. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *Journal of Instrumentation*, 2(04):P04004–P04004, apr 2007.

[43] Albert M Sirunyan et al. Reconstruction of signal amplitudes in the cms electromagnetic calorimeter in the presence of overlapping proton-proton interactions. *JINST*, 15(arXiv:2006.14359. CMS-EGM-18-001-003. 10):P10002. 45 p, Jun 2020.

[44] Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s}$= 8 TeV. *Journal of Instrumentation*, 10(08):P08010–P08010, aug 2015.

[45] T. Adams et al. Beam test evaluation of electromagnetic calorimeter modules made from proton-damaged PbWO4crystals. *Journal of Instrumentation*, 11(04):P04012–P04012, apr 2016.

[46] M. Anfreville et al. Laser monitoring system for the cms lead tungstate crystal calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 594(2):292 – 320, 2008.

[47] Albert M Sirunyan et al. A measurement of the Higgs boson mass in the diphoton decay channel. *Phys. Lett. B*, 805(arXiv:2002.06398. CMS-HIG-19-004-003):135425. 37 p, Feb 2020.

[48] Measurements of Higgs boson properties in the diphoton decay channel at $\sqrt{s}$ = 13 TeV. Technical Report CMS-PAS-HIG-19-015, CERN, Geneva, 2020.

[49] Serguei Chatrchyan et al. Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s}$ = 7 TeV with the CMS experiment. *JHEP*, 10(arXiv:1107.4789. CMS-EWK-10-005. CERN-PH-EP-2011-107):132. 68 p, Jul 2011.

[50] J. Alwall et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07(CERN-PH-TH-2014-064):079. 157 p, May 2014.

[51] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[52] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159 – 177, 2015.

[53] Albert M Sirunyan et al. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *Eur. Phys. J. C*, 80(arXiv:1903.12179. CMS-GEN-17-001-003. 1):4. 61 p, Mar 2019.

[54] CMS Collaboration. Event generator tunes obtained from underlying event and multiparton scattering measurements. *European Physical Journal C - Particles and Fields*, 76, 12 2015.

[55] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[56] Richard D. Ball et al. Parton distributions from high-precision collider data. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77(CAVENDISH-HEP-17-06. 10):663. 95 p, Jun 2017.

[57] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003.

[58] G. Buchalla, M. Capozi, A. Celis, G. Heinrich, and L. Scyboz. Higgs boson pair production in non-linear Effective Field Theory with full m_t-dependence at NLO QCD. *JHEP*, 09:057, 2018.

[59] S. Chatrchyan et al. Search for the standard model higgs boson decaying into two photons in pp collisions at s=7 tev. *Physics Letters B*, 710(3):403 – 425, 2012.

[60] A.M. Sirunyan et al. Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11(CMS-HIG-16-040. CMS-HIG-16-040-003):185. 56 p, Apr 2018.

[61] A.M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *Journal of Instrumentation*, 13(05):P05011–P05011, may 2018.

[62] A deep neural network for simultaneous estimation of b quark energy and resolution. Technical Report CMS-PAS-HIG-18-027, CERN, Geneva, 2019.

[63] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery.

[64] Jet algorithms performance in 13 TeV data. Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[65] Georges Aad et al. Search for the $HH \to b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 07:108, 2020.

[66] Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011.

[67] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 434(2):435 – 443, 1999.

[68] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 73:2501, 2013.

[69] M. Oreglia. A Study of the Reactions $\psi' \to \gamma\gamma\psi$. Master thesis, 12 1980.

[70] P.D. Dauncey, M. Kenzie, N. Wardle, and G.J. Davies. Handling uncertainties in background shapes: the discrete profiling method. *Journal of Instrumentation*, 10(04):P04015–P04015, apr 2015.

[71] Jon Butterworth et al. PDF4lhc recommendations for LHC run II. *Journal of Physics G: Nuclear and Particle Physics*, 43(2):023001, jan 2016.

[72] Stefano Carrazza, Stefano Forte, Zahari Kassabov, Jose Ignacio Latorre, and Juan Rojo. An Unbiased Hessian Representation for Monte Carlo PDFs. *Eur. Phys. J. C*, 75(8):369, 2015.

[73] Barbara Jäger, Alexander Karlberg, Simon Plätzer, Johannes Scheller, and Marco Zaro. Parton-shower effects in Higgs production via Vector-Boson Fusion. *Eur. Phys. J. C*, 80(8):756, 2020.

[74] CMS Luminosity Measurements for the 2016 Data Taking Period. Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017.

[75] CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-17-004, CERN, Geneva, 2018.

[76] CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-18-002, CERN, Geneva, 2019.

[77] Albert M Sirunyan et al. Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 122(12):121803, 2019.

[78] Constraints on the Higgs boson self-coupling from the combination of single-Higgs and double-Higgs production analyses performed with the ATLAS experiment. Technical Report ATLAS-CONF-2019-049, CERN, Geneva, Oct 2019.

[79] Apollinari G., Béjar Alonso I., Brüning O., Fessia P., Lamont M., Rossi L., and Tavian L. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1.* CERN Yellow Reports: Monographs. CERN, Geneva, 2017.

[80] M. Cepeda et al. Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC. *CERN Yellow Rep. Monogr.*, 7(arXiv:1902.00134):221–584. 364 p, Dec 2018.

[81] Collaboration CMS. A MIP Timing Detector for the CMS Phase-2 Upgrade. Technical Report CERN-LHCC-2019-003. CMS-TDR-020, CERN, Geneva, Mar 2019.

[82] The Phase-2 Upgrade of the CMS L1 Trigger Interim Technical Design Report. Technical Report CERN-LHCC-2017-013. CMS-TDR-017, CERN, Geneva, Sep 2017.

[83] The Phase-2 Upgrade of the CMS DAQ Interim Technical Design Report. Technical Report CERN-LHCC-2017-014. CMS-TDR-018, CERN, Geneva, Sep 2017.

[84] The Phase-2 Upgrade of the CMS Tracker. Technical Report CERN-LHCC-2017-009. CMS-TDR-014, CERN, Geneva, Jun 2017.

[85] The Phase-2 Upgrade of the CMS Barrel Calorimeters. Technical Report CERN-LHCC-2017-011. CMS-TDR-015, CERN, Geneva, Sep 2017.

[86] The Phase-2 Upgrade of the CMS Endcap Calorimeter. Technical Report CERN-LHCC-2017-023. CMS-TDR-019, CERN, Geneva, Nov 2017.

[87] J. Favereau, C. Delaere, Pavel Demin, A. Giammanco, Vincent Lemaitre, Alexandre Mertens, and Michele Selvaggi. Delphes 3, a modular framework for fast simulation of a generic collider experiment. *Journal of High Energy Physics*, 2014, 07 2013.

[88] The CMS Collaboration. Expected performance of the physics objects with the upgraded CMS detector at the HL-LHC. Technical Report CMS-NOTE-2018-006. CERN-CMS-NOTE-2018-006, CERN, Geneva, Dec 2018.

[89] Projected performance of Higgs analyses at the HL-LHC for ECFA 2016. Technical Report CMS-PAS-FTR-16-002, CERN, Geneva, 2017.

[90] Albert M Sirunyan et al. Observation of $t\bar{t}H$ production. *Phys. Rev. Lett.*, 120(23):231801, 2018.

[91] Walter Greiner and Berndt Müller. *Gauge theory of weak interactions; 3rd rev. ed.* Springer, Berlin, 2000.

[92] Susanne Mertens. Direct neutrino mass experiments. *Journal of Physics: Conference Series*, 718:022013, may 2016.

[93] R. Aaij et al. Observation of $j/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to j/\psi K^- p$ decays. *Phys. Rev. Lett.*, 115:072001, Aug 2015.

[94] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[95] Hideki Yukawa. On the Interaction of Elementary Particles. I. *Progress of Theoretical Physics Supplement*, 1:1–10, 01 1955.

[96] Giuseppe Degrassi, Pier Paolo Giardino, Fabio Maltoni, and Davide Pagani. Probing the Higgs self coupling via single Higgs production at the LHC. *JHEP*, 12:080, 2016.