IL NUOVO CIMENTO **47 C** (2024) 78
DOI 10.1393/ncc/i2024-24078-y

COLLOQUIA: IFAE 2023

Studies of Higgs boson pair production at the LHC(*)

F. MONTEREALI(1)(2) on behalf of the ATLAS and CMS COLLABORATIONS
(1) *INFN, Sezione di Roma Tre - Roma, Italy*
(2) *Dipartimento di Matematica e Fisica, Università Roma Tre - Roma, Italy*

received 13 February 2024

Summary. — The most recent results of the ATLAS and CMS searches for Higgs boson pair production at the LHC using the full Run-2 data collected in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV with an integrated luminosity of about 140 fb^{-1} are presented. The focus is on the analyses of most sensitive di-Higgs channels, $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau\tau$, and their combination in the non-resonant context, and a brief overview on the resonant processes and projections for the future data taking LHC runs are discussed.

1. – Introduction

The observation of the Higgs boson (H) in 2012 at the CERN Large Hadron Collider (LHC) [1] by both ATLAS [2] and CMS [3] experiments provided the last building brick for the Standard Model (SM) of particle physics. Since its discovery, most of its properties such as mass, spin, production cross-section and its coupling to fermions and bosons have been measured. However, the couplings of the Higgs boson to itself are very far from being experimentally constrained.The Higgs boson self-coupling provides information about the structure of the Higgs potential and consequently of the spontaneous electroweak symmetry breaking (EWSB) realized by the Brout-Englert-Higgs mechanism, which is responsible for the mass generation of elementary particles.

The Higgs potential reads

$$V(H) = \frac{1}{2}m_H^2 H^2 + \lambda_3 v H^3 + \frac{\lambda_4}{4}H^4 \tag{1}$$

where H is the Higgs scalar field, m_H is the Higgs boson mass, λ_3, λ_4 are its trilinear and quartic self-couplings, and $v = (\sqrt{2}G_F)^{-\frac{1}{2}} \simeq 246$ GeV [4] is the vacuum expectation value. In the SM the λ_3 is predicted as $\lambda_3^{SM} = \frac{m_H^2}{2v^2} \sim 0.13$.

While λ_4 is likely to remain inaccessible for a long time, the trilinear self-coupling can be directly probed by studying Higgs boson pair (HH) production, that is an extremely rare process (not yet observed). At the LHC, HH is mainly produced via gluon-gluon Fusion (ggF) and Vector-Boson Fusion (VBF) modes, that are affected by the trilinear

(*) IFAE 2023 - "Energy Frontier" session

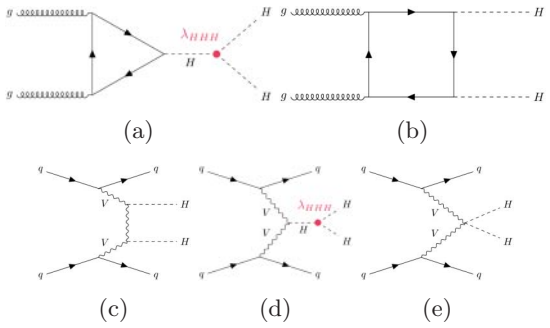


Fig. 1.: Feynman diagrams for ggF (top) and VBF (bottom) HH production.

self-interaction vertex. A deviation from the SM predicted self-coupling λ_3^{SM}, corresponding to a deviation of its modifier parameter $\kappa_\lambda = \lambda_3/\lambda_3^{SM}$ from unity, may point to physics Beyond the Standard Model (BSM).

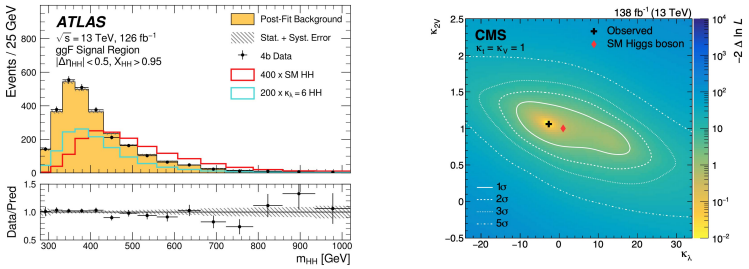
2. – Di-Higgs production at the LHC

In the Standard Model, the leading contributions to ggF HH production are the 'triangle diagram' (fig. 1(a)), which includes a Higgs boson self-coupling vertex, and the heavy-quark 'box diagram' (fig. 1(b)), that has two fermion-fermion-Higgs vertices. These diagrams interfere destructively resulting in a SM non-resonant HH cross section of three orders of magnitude smaller than that of single Higgs boson production: $\sigma_{ggF}^{HH} = 31$ fb at 13 TeV for $m_H = 125$ GeV [5]. The VBF process provides a sub-leading source of HH production and has a cross section of $\sigma_{VBF}^{HH} = 1.73$ fb [6]. This production mode depends on the strength of the interaction of pairs of vector bosons V $(= W, Z)$ with a single (VVH) and a pair $(VVHH)$ of Higgs bosons (figs. 1(c), 1(d), 1(e)), whose values compared to the SM prediction are parametrized by the modifiers κ_V and κ_{2V}, respectively. The results of the HH searches can be used to infer the value not only of the coupling modifier κ_λ but also of κ_{2V} to which the VBF production mode has a unique sensitivity.

Due to the large Higgs boson Branching Ratio (BR) to bottom quarks (59%), most searches require at least one $H \to b\bar{b}$ decay while different decay modes of the second Higgs boson are considered.

This proceeding reviews the HH searches with full Run 2 data by the ATLAS and CMS experiments in the three most sensitive channels: $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau\tau$, and their combination. In all three ATLAS and CMS analyses, the jets are reconstructed using the anti-κ_t jet clustering algorithm with a radius parameter of 0.4 and those originating from b quarks are identified through specific multivariate classification techniques, the CMS *DeepJet* [7] and ATLAS *DL1r* [8] algorithms, that are Deep-Learning Neural Networks (DNN). For the VBF production mode, events must contain two additional not b-tagged and energetic jets having large invariant mass, that are tagged as "VBF jets".

Non-resonant $HH \to b\bar{b}b\bar{b}$ process. This fully hadronic final state is characterized by the largest BR of 34% and the challenging multi-jet background. Both in ATLAS [9] and in CMS [10] the event selection requires exactly four reconstructed and b-tagged jets, that are then paired, according to the increasing distance between highest-p_T jets, to form the two Higgs boson candidates H_1, H_2 ($> 90\%$ correct $H \to b\bar{b}$ matching). A data-driven

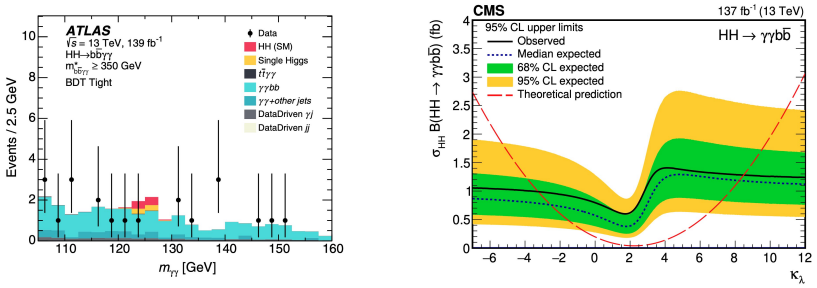


(a) (b)

Fig. 2.: ATLAS m_{HH} distribution in a $|\Delta\eta|$, X_{HH} ggF signal region category [9] (left) and the CMS profile likelihood test statistic scan on data as function of κ_λ, κ_{2V} [11] (right) in the $HH \to b\bar{b}b\bar{b}$ channel.

technique is employed by both Collaborations for the background estimation by using Machine Learning (ML) techniques to perform a reweighting of data from an alternative phase space to model the background in the region of interest. In ATLAS, the ggF and VBF signal regions are cathegorized by the pseudorapidity separation $|\Delta\eta|$ and the discriminant $X_{HH} = \sqrt{(\frac{m_{H_1}-124GeV}{0.1m_{H_1}})^2 + (\frac{m_{H_2}-117GeV}{0.1m_{H_2}})^2}$ (fig. 2(a)), in order to improve the HH analysis sensitivity. The invariant mass of the Higgs boson candidate pair, m_{HH}, is used as the signal-background discriminating variable. In CMS, the signal is extracted via a multivariate BDT score and the m_{HH} in the ggF and VBF subcategories, respectively. There is no deviation from a background-only hypothesis and the observed (expected) upper limits at 95% confidence level (CL) are set on HH production cross section compared to the SM expectation σ/σ_{SM} to 5.4(8.1) and on the coupling modifiers $-3.9 < \kappa_\lambda < 11.1(-4.6 < \kappa_\lambda < 10.8)$ and $-0.03 < \kappa_{2V} < 2.11(-0.05 < \kappa_{2V} < 2.12)$ for ATLAS, while to $\sigma/\sigma_{SM} < 3.9(7.8)$, $-2.3 < \kappa_\lambda < 9.4(-5.0 < \kappa_\lambda < 12.0)$ and $-0.1 < \kappa_{2V} < 2.2(-0.4 < \kappa_{2V} < 2.5)$ for CMS. The most stringent limits on κ_{2V} are obtained from the CMS boosted regime analysis, [11]: $0.62 < \kappa_{2V} < 1.41(0.66 < \kappa_{2V} < 1.37)$ at 95% CL (fig. 2(b)), which excludes $\kappa_{2V} = 0$ at 6.3σ. In this analysis, only events with both Higgs bosons in the highly Lorentz-boosted regime are selected and reconstructed via a Graph Neural Network (GNN) algorithm, *ParticleNet* [12], increasing signal purity by improving the jet mass estimation.

Non-resonant $HH \to b\bar{b}\tau\tau$ process. This decay channel has a moderate branching fraction (7.3%) and a good trade-off between BR and the final state reconstruction, but the electroweak and top backgrounds can mimic signal. The events are selected in three separate signal categories characterized by the presence of exactly two b-tagged jets and two oppositely charged τ leptons that can both decay hadronically, $\tau_{had}\tau_{had}$, or one decays hadronically and the other leptonically, $\tau_{lep}\tau_{had}$. In ATLAS [13], the main backgrounds including the production of top-quark pairs ($t\bar{t}$), single top quarks, W/Z+jets and fake-τ, are estimated by both Monte Carlo (MC) simulation-based and data-driven methods. Multivariate discriminants, BDT in the hadronic category and NN in leptonic one, are used to extract the signal. Upper limits are set at the 95% CL only on σ/σ_{SM} corresponding to 4.7(3.9). In CMS [14], hadronic τ leptons are identified using a convolutional neural network, the *DeepTau* algorithm [15]. The main background sources are $t\bar{t}$ production, $Z/\gamma^* \to ll$ production and QCD multijet events: the first two contribu-



(a) (b)

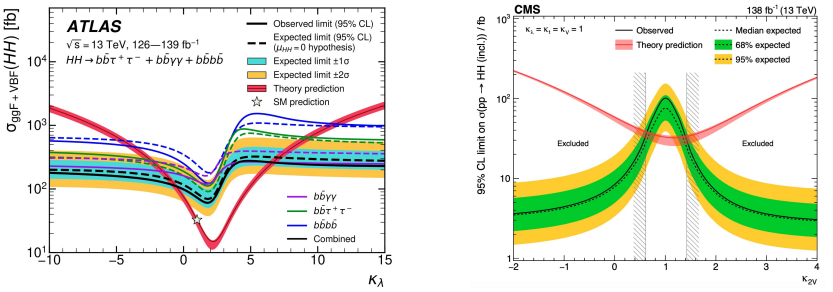
Fig. 3.: Left: ATLAS $m_{\gamma\gamma}$ distribution in high-mass BDT tight category, that is the most sensitive to SM signal [17]. Right: the CMS expected and observed 95% CL upper limits on $\sigma^{HH} \times \mathcal{B}(HH \to \gamma\gamma b\bar{b})$ obtained for different values of κ_λ [16].

tions are modeled using simulation methods, the third using data-driven approach. A binned maximum likelihood fit of a DNN discriminant is performed in all the ggF and VBF categories. Upper limits at 95% CL are set both on σ/σ_{SM} to 3.3(5.2) and on $-1.7 < \kappa_\lambda < 8.7(-2.9 < \kappa_\lambda < 9.8)$ and $-0.4 < \kappa_{2V} < 2.6(-0.6 < \kappa_{2V} < 2.8)$.

Non-resonant $HH \to b\bar{b}\gamma\gamma$ process. This final state is characterized by a small branching ratio (0.26%), but also by a clear experimental signature and a smooth background, mainly di-photon continuum ($\gamma\gamma$+jets) and single Higgs boson, that is estimated using data-driven methods in CMS [16] and in combination with MC simulation in ATLAS [17].The invariant mass of the diphoton plus b-tagged jets system, $m^*_{b\bar{b}\gamma\gamma} = m_{b\bar{b}\gamma\gamma} - m_{b\bar{b}} - m_{\gamma\gamma} + 250$ GeV is used in ATLAS to divide the selected events into two regions, targeting the SM signal, $m^*_{b\bar{b}\gamma\gamma} > 350$ GeV, and the BSM one, $m^*_{b\bar{b}\gamma\gamma} < 350$ GeV. In each region, a dedicated BDT is trained to isolate HH signals from backgrounds (fig. 3(a)). The diphoton invariant mass is used as the final discriminant variable to obtain the statistical results via a maximum likelihood fit to its distribution in the range $[105, 160]$ GeV, performed simultaneously over all the categories. No significant excess of events is observed in the data with respect to the expected background and observed (expected) upper limits at 95% CL are set on the HH cross-section and trilinear self-coupling modifier, $\sigma/\sigma_{SM} < 4.2(5.7)$ and $-1.5 < \kappa_\lambda < 6.7(-2.4 < \kappa_\lambda < 7.7)$.

In the CMS analysis, a particular attention is reserved to the ttH background and in order to reduce it, a DNN, *ttHScore*, is used. A 2D fit in the ($m_{\gamma\gamma}$, m_{jj}) plane, in the mass range $100 < m_{\gamma\gamma} < 180$ GeV and $70 < m_{jj} < 190$ GeV, is performed for the signal extraction simultaneously in all the ggF and VBF categories. The observed (expected) 95% CL limits on σ/σ_{SM} is 7.7(5.2) and on coupling modifiers are $-1.7 < \kappa_\lambda < 8.7(-2.9 < \kappa_\lambda < 9.8)$ (fig. 3(b)), and $-1.3 < \kappa_{2V} < 3.5(-0.9 < \kappa_{2V} < 3.1)$.

Analyses combinations. The $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau\tau$ analyses are combined to improve the HH sensitivity. In CMS the combination includes also the $b\bar{b}ZZ$, $b\bar{b}WW$ and $WW\gamma\gamma$ and Multilepton decay modes. In both Collaborations, the assumption in performing a statistical combination, as the product of the individual analysis likelihoods in ATLAS and Poisson probability functions in CMS [18], is that the analyses are statistically independent. The ATLAS combination [19] sets observed (expected) limits at 95% CL of $\sigma/\sigma_{SM} < 2.4(2.9)$, while CMS [20] of $\sigma/\sigma_{SM} < 3.4(2.5)$, improving their previous



(a) (b)

Fig. 4.: The observed (expected) 95% CL exclusion limits on the $\sigma^{HH}_{ggF+VBF}$ as a function of κ_λ for the three HH search channels and their combination for ATLAS [19] (left), and on the HH production cross section for different values of κ_{2V} for CMS [20] (right).

results at 36.1 fb^{-1} [21] and 35.9 fb^{-1} [22] by a factor of ~ 3 and > 5, respectively. The ATLAS upper limits on modifier couplings are $-0.6 < \kappa_\lambda < 6.6 (-2.1 < \kappa_\lambda < 7.8)$ (fig. 4(a)) and $0.1 < \kappa_{2V} < 2.0 (0.0 < \kappa_{2V} < 2.1)$, a 8x improvement on previous analysis κ_λ, while the corresponding CMS limits are $-1.24 < \kappa_\lambda < 6.49 (-2.28 < \kappa_\lambda < 7.94)$ and $0.67 < \kappa_{2V} < 1.38 (0.61 < \kappa_{2V} < 1.42)$ (fig. 4(b)) excluding $\kappa_{2V} = 0$ with a significance of 6.6σ, that establishes existence of the quartic coupling $VVHH$. The improvements compared to the analyses with the partial Run-2 data are mainly due to the optimization of the tagging algorithms and the extensive use of ML techniques.

3. – Resonant processes

Searches for resonant HH production are motivated by many beyond the Standard Model theories predicting new heavy scalar particles that can decay into pairs of Higgs bosons.

The analysis strategies adopted for the HH resonant processes are similar to non-resonant ones in both the Collaborations. Focusing on spin-0 resonance, the complementarity of the three analyses $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau\tau$, permits to explore the existence of a scalar particle in the mass range $250 < m_X < 3000$ GeV (fig. 5). The $b\bar{b}\gamma\gamma$ channel is sensitive at low mass $m_X \in [250, 1000]$ GeV, and CMS results [23], via a 2D fit on $(m_{\gamma\gamma}, m_{b\bar{b}})$, indicate the largest deviation from the background only hypothesis at 650 GeV with a local (global) significativity of $3.8\sigma(2.8\sigma)$. In ATLAS the largest deviation is observed in the $b\bar{b}\tau\tau$ search [13], that covers the mass range $260 - 1600$ GeV, at 1.1 TeV with a local (global) significativity of $3.2\sigma(2.1\sigma)$, after a fit to Parametrized Neural Network (applied on m_X) score. The $b\bar{b}b\bar{b}$ decay channel is sensitive at high m_X, up to 3 TeV, and both the Collaborations' analyses ([24], [25]) focus on the boosted regime.

4. – HL-LHC projections

The High Luminosity LHC (HL-LHC) project (2029-2040) is expected to deliver a total integrate luminosity of 3000 fb^{-1} at $\sqrt{s} = 14$ TeV, resulting in a statistics increased by a factor of 10 compared to combining Run-2+3 but also with a higher number of pile-up events, about 140-200.

The ATLAS projections [26] rely on an extrapolation of the different HH analyses based on the Run-2 results and scaled by multiplicative factors to take into account the

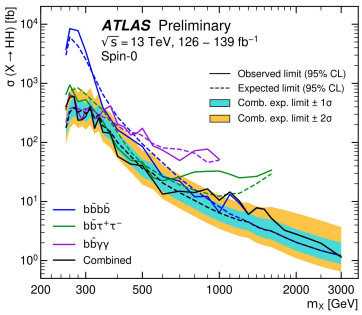


Fig. 5.: ATLAS observed(expected) 95% CL upper limits on $\sigma(X \to HH)$ for a spin-0 resonance as a function of its mass m_X in the $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau\tau$ searches [24].

increase in integrated luminosity and $\sqrt{s}$ from Run-2 to the HL-LHC. The CMS prospects [20] are based on parametric simulation performed with the DELPHES software of the detector to model its response in the HL-LHC conditions. Both the Collaborations expect a limit $\sigma/\sigma_{SM} < 1.0$ after the combination of $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau\tau$ channels, indicating that the sensitivity is sufficient to confirm the existence of the SM HH production.

REFERENCES

[1] Evans L. and Bryant P., *JINST*, **3** (2008) S08001.
[2] ATLAS Collaboration, *JINST*, **3** (2008) S08003.
[3] CMS Collaboration, *JINST*, **3** (2008) S08004.
[4] Particle Data Group, *PTEP*, **2022** (2022) 083C01.
[5] Di Micco B., Gouzevitch M., Mazzitelli J., Vernieri C., *et al.*, *Rev. Phys.*, **5** (2020) 100045.
[6] Ling L. S., Zhang R. Y., Ma W. G., Guo L., Li W. H. and Li X. Z., *Phys. Rev. D*, **89** (2014) 073001.
[7] Bols E., Kieseler J., Verzetti M., Stoye M. and Stakia A., *JINST*, **15** (2020) P12012.
[8] ATLAS Collaboration, *Eur. Phys. J. C*, **79** (2019) 970.
[9] ATLAS Collaboration, *Phys. Rev. D*, **108** (2023) 052003.
[10] CMS Collaboration, *Phys. Rev. Lett.*, **129** (2022) 081802.
[11] CMS Collaboration, *Phys. Rev. Lett.*, **131** (2023) 041803.
[12] Qu H. and Gouskos L., *Phys. Rev. D*, **101** (2020) 056019.
[13] ATLAS Collaboration, *JHEP*, **07** (2023) 040.
[14] CMS Collaboration, *Phys. Lett. B*, **842** (2023) 137531.
[15] CMS Collaboration, *JINST*, **17** (2022) P07023.
[16] CMS Collaboration, *JHEP*, **03** (2021) 257.
[17] ATLAS Collaboration, *Phys. Rev. D*, **106** (2022) 052001.
[18] ATLAS, CMS Collaborations and LHC Higgs Combination Group, CMS-NOTE-2011-005.
[19] ATLAS Collaboration, *Phys. Lett. B*, **843** (2023) 137745.
[20] CMS Collaboration, *Nature*, **607** (2022) 60.
[21] ATLAS Collaboration, *Phys. Lett. B*, **800** (2020) 135103.
[22] CMS Collaboration, *Phys. Rev. Lett.*, **122** (2019) 121803.
[23] CMS Collaboration, CMS-PAS-HIG-21-011.
[24] ATLAS Collaboration, ATLAS-CONF-2021-052.
[25] CMS Collaboration, *Phys. Lett. B*, **842** (2023) 137392.
[26] ATLAS Collaboration, ATL-PHYS-PUB-2022-053, `https://cds.cern.ch/record/2841244`.