PROCEEDINGS OF SCIENCE

Di-Higgs Searches at CMS

Bruno Alves[a,*] **for the CMS Collaboration**
[a] *École Polytechnique, Route de Saclay, 91128 Palaiseau Cedex, France*
E-mail: bruno.alves@cern.ch

The production of Higgs boson pairs (HH) at the LHC is the main handle for the measurement of the Higgs boson interaction with itself and is thus a fundamental test of the Standard Model theory as well as for determining the Higgs potential. The most recent results from the CMS Collaboration on measurements of nonresonant HH production using different final states, and their combination using the data set collected by the CMS experiment at a centre of mass energy of 13 TeV, will be presented.

31st International Workshop on Deep Inelastic Scattering (DIS2024)
8–12 April 2024
Grenoble, France

*Speaker

PoS(DIS2024)153

1. Introduction

The production of Higgs boson [1–3] pairs (HH), despite its rarity and experimental challenges, has become a central piece of the scientific programme of the CMS experiment [4] at the Large Hadron Collider (LHC). Measurements of the Higgs self-coupling κ_λ provide a unique Standard Model (SM) consistency test, probe the stability of the vacuum of our Universe, and might help establishing, for instance, new theoretical explanations for baryogenesis [5]. We report the most recent CMS HH results.

2. HH $\rightarrow \gamma\gamma\tau\tau$

CMS has recently explored the extremely rare $\gamma\gamma\tau\tau$ decay channel (0.028% branching ratio), for the first time, using the Run 2 (2016-2018) dataset [6]. The analysis covers both resonant and nonresonant production. We focus on the latter. The analysis exploits di-γ triggers only, as triggering on the τ's is found to have negligible impact. Standard photon and lepton selections are applied. The dominant backgrounds are peaking single-H and nonresonant γ+jets, $\gamma\gamma$+jets, $t\bar{t}+\gamma$, $t\bar{t}+\gamma\gamma$, $W\gamma$ and $Z\gamma$. Minor backgrounds are taken from simulation. The signal is fitted independently for different categories and data-taking years with a double crystal ball function. The background also includes a H $\rightarrow \gamma\gamma$ contribution which is modeled just like the signal. The background continuum, instead, uses the discrete profiling method [7], which considers multiple analytical functions and penalizes those with many parameters. The tau lepton pairs are reconstructed in their 6 possible decay modes (ee, $\mu\mu$, $e\mu$, $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$), and additionally the single τ_h and τ_h + track channels are considered when no electron or muon passes the selection. Events compatible with Z $\rightarrow$ ll or Z $\rightarrow$ llγ close to the mass of the Z boson are rejected using a selection cut on the mass. A (di-photon) mass-independent Boosted Decision Tree (BDT) is used, considering input features related to the kinematic properties of the event. Sequential boundaries are applied to the BDT's output to create two categories of different signal purities. The splitting maximizes signal sensitivity. Events not belonging to one of those categories are discarded. The final results are obtained performing a simultaneous maximum likelihood fit to $m_{\gamma\gamma}$ in the two signal-enriched categories. As predicted by the SM, the analysis observes very few $\gamma\gamma\tau\tau$ candidates. It delivers observed (expected) κ_λ 95% CL upper limits, -13 (-11) $<$ k_λ $<$ 18 (16), and on the HH proton-proton cross-section, $\sigma_{\text{HH}} < 930$ (740) fb or $\sigma_{\text{HH}} < 33$ (26) $\sigma_{\text{HH}}^{\text{SM}}$. The limits are around one order of magnitude worse than the ones obtained with the most sensitive channels.

3. QCD multijet modeling with hemisphere mixing in ZZ/ZH $\rightarrow$ bbbb

Currently available multijet simulations lack the required precision and the available number of events for a robust background estimate, particularly in distribution tails. Data-driven methods are therefore usually employed to model the multijet background. The methods usually take an "ABCD-like" approach. The idea is to find fully uncorrelated variables on which the signal region (SR) selection depends, and invert the cuts to obtain control regions (CRs). The latter can be used to estimate both the shape and normalization of the QCD multijet background in the SR, without using the SR directly. The background is derived in CRs, and thus requires an extrapolation to the

SR. In order to validate the extrapolation, a validation region (VR) is usually employed. However, the definition of an additional region will necessarily deplete the signal region. Additionally, the extrapolation cannot be tested directly, since the VR differs from the signal region inasmuch as it is not signal-enriched. Additionally, both CRs and VRs often suffer from a small number of events, and become a dominant source of systematic uncertainties. There is therefore a need to develop new methods to estimate and validate the QCD multijet background that are not sensitive to small number of events. It would also be beneficial to directly test the ABCD extrapolation in the SR.

The hemisphere mixing technique [8] first creates a library of "hemispheres", which arise from the splitting of events along the plane orthogonal to the transverse thrust axis. The latter is in turn defined as the axis where the sum of the absolute values of the p_{T} projections of all the jets in the event is maximal. For each hemisphere a set of four variables is calculated: mass, longitudinal momentum, and transverse momentum perpendicular and parallel to the thrust axis. A second pass on data mixes pairs of hemispheres by minimizing their distances with respect to the normalized sum of the summary variables. The two hemispheres must belong to different events.

ZZ and ZH decays represent standard candles to validate HH analyses, given their larger cross-sections. The ZZ/ZH → bbbb processes (31 and 8 times the HH → bbbb cross-section, respectively) are therefore expected to be observed before HH → bbbb. The ZZ/ZH → bbbb analysis [9] contributes with two improvements to the original hemisphere mixing technique, where the splitting is done using a sample of events with four b-tagged jets, thus pure in signal events. Firstly, the mixing step is performed with 3-tagged data in order to increase the sample's size and make (4-tagged) signal contamination negligible. Statistics are also increased by lowering the b-tag working point used on the three jets. Secondly, the non-negligible presence of $t\bar{t}$ events is mitigated by removing such events from the mixing stage. This is done event-by-event via a classification neural network, which calculates the probability P for each event to be multi-jet, where a random number X is generated between 0 and 1. If $\mathrm{X} > P$, the event is rejected. For the validation of the background model, we have to ensure the size of the synthetic dataset is comparable to the one used for the model. The hemisphere dataset is thus sub-sampled, and 15 separate mixed models are formed, given the available statistics. Systematic uncertainties of the multijet modeling are determined using the synthetic dataset in three different ways: *i*) differences between mixed models, arising from limited statistics, are quantified using their average, *ii*) the background model is compared with the mixed models in the signal region, and *iii*) an unconstrained signal template is added to the signal + background fit to verify if a spurious signal can be mimicked by the background model. Despite not yet being used in the most recent HH → bbbb results, a principled and precise way of measuring the most important systematics directly in the SR is thus available. We note that, given appropriate modifications, a similar method could be extended to the HH → bb$\tau\tau$ analysis.

In the ZZ/ZH → bbbb analysis, as mentioned, the synthetic dataset is used for the validation of the multijet background, which is in turn built on a control region, defined by requiring three b-tagged jets instead of four. The model is weighted by two sets of weights, to account for additional jet activity and subsisting kinematic mismatches, in order to match the SR. The weights are derived in a di-jet mass sideband. The observed (expected) 95% CL upper limits on the production cross sections correspond to 3.8 (3.8) and 5.0 (2.9) times the SM prediction, for the ZZ → bbbb and ZH → bbbb processes, respectively. The analysis indicates that ZH → bbbb will likely be observed before ZZ → bbbb.

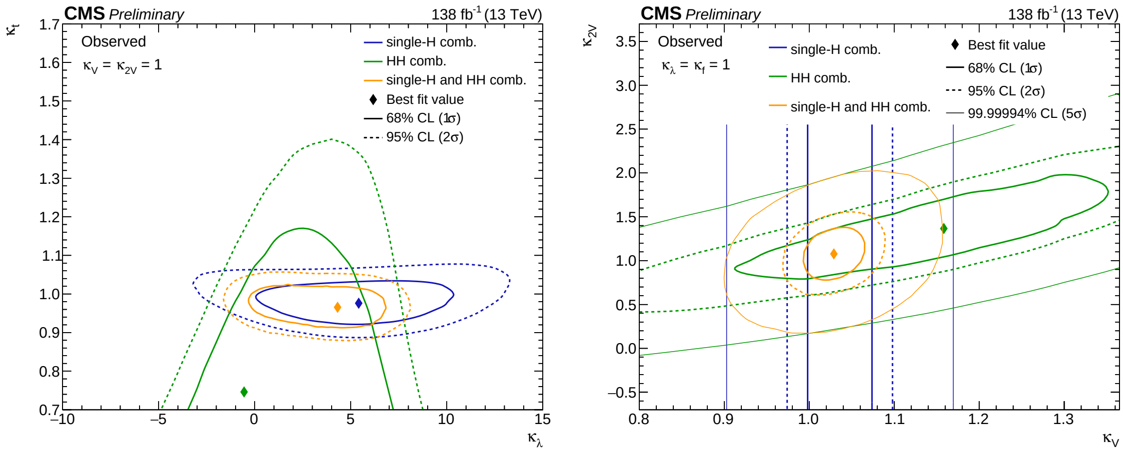

Figure 1: Observed two-dimensional likelihood scans of (κ_λ, κ_t) (left) and (κ_V, κ_{2V}) (right). The strong complementarity between the single and double Higgs processes is well illustrated. The remaining coupling modifiers are set to their SM value. Taken from [10].

4. Single and double Higgs combination

On top of HH searches, κ_λ can be constrained by exploiting electroweak corrections to single Higgs processes. Single Higgs cross-sections can indeed be quite sensitive to κ_λ variations, especially in the VH and ttH production modes, where up to 10% differences are expected. There is also a strong complementarity between direct and indirect searches. HH searches generally provide weaker constraints on the Higgs boson coupling to fermions and vector bosons, due to the lower production cross-section with respect to single Higgs. On the other hand, double Higgs processes provide stronger constraints on κ_λ. The main challenge of the combination [10] recently performed by CMS consists in estimating and efficiently removing overlaps between signal regions of different analyses. Whenever overlaps exist, one of two approaches is taken: either additional selections are applied, or the least sensitive category or analysis is removed. The modeling of systematics in HH processes is generally simpler when compared to single H, due to the limited statistics of HH. CMS observed exclusion intervals at 95% CL of $-2.0 < \kappa_\lambda < 7.7$, assuming other Higgs couplings to follow the SM, or $-1.2 < k_\lambda < 7.5$ otherwise. For comparison, ATLAS [11] observed $-0.4 < \kappa_\lambda < 6.3$, assuming SM couplings, or $-1.4 < k_\lambda < 6.1$ otherwise [12]. We show (κ_λ, κ_t) and (κ_V, κ_{2V}) scans in Fig. 1, where the complementarity between the two types of processes is clearly highlighted. CMS was also able to once again [5] exclude $\kappa_{2V} = 0$ at more than 5σ, this time without fixing $\kappa_V = 1$.

5. Run 3 and beyond

Upcoming nonresonant HH searches will bring further constraints on κ_λ as well as on other couplings not discussed in this report. Yet unexplored HH production modes and decay channels are being studied. On top of the recent $\kappa_{2V} = 0$ exclusion, and assuming $\kappa_\lambda = 1$, CMS might measure nonresonant HH via a multi-channel combination by the end of the High Luminosity LHC (HL-LHC) [5], possibly in combination with ATLAS. Uncertainties are still dominated by the limited size of the data sample, but gluon-gluon fusion theory uncertainties might soon become important. For the moment, Run 3 is an opportunity to bring improvements before the start of the HL-LHC. New techniques, including better estimates of QCD background and new machine learning methods, will make existing results quickly obsolete. The usage of Particle Net (PNet) [13] for τ-initiated jets and the application of transformer technology to jet tagging [14] might have a strong impact. Additionally, an improved trigger strategy has been implemented, considering both data scouting and parking [15], and including PNet b-tagging and τ-tagging at trigger level. We also expect that some HH analysis might benefit from the inclusion of synthetic datasets. The first CMS Run 3 HH results will soon become available.

References

[1] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, September 2012.

[2] The CMS Collaboration. 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, September 2012.

[3] The CMS Collaboration. Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV. *Journal of High Energy Physics*, 2013(6), June 2013.

[4] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, Aug 2008.

[5] The CMS Collaboration. A portrait of the Higgs boson by the CMS experiment ten years after the discovery. *Nature*, 607(7917):60–68, July 2022.

[6] The CMS Collaboration. Search for the nonresonant and resonant production of a Higgs boson in association with an additional scalar boson in the $\gamma\gamma\tau\tau$ final state. Technical report, CERN, Geneva, 2024.

[7] 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, apr 2015.

[8] Pablo de Castro Manzano, Martino Dall'Osso, Tommaso Dorigo, Livio Finos, Grzegorz Kotkowski, Giovanna Menardi, and Bruno Scarpa. Hemisphere Mixing: a fully data-driven model of QCD multijet backgrounds for LHC searches. *PoS*, EPS-HEP2017:370, 2017.

[9] The CMS Collaboration. Search for ZZ and ZH production in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV. *preprint arXiv:2403.20241*, 2024.

[10] The CMS Collaboration. Constraints on the Higgs boson self-coupling with combination of single and double Higgs boson production. Technical report, CERN, Geneva, 2023.

[11] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003, Aug 2008.

PoS(DIS2024)153

[12] The ATLAS Collaboration. Constraints on the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV. *Physics Letters B*, 843:137745, 2023.

[13] Huilin Qu and Loukas Gouskos. Jet tagging via particle clouds. *Phys. Rev. D*, 101:056019, Mar 2020.

[14] Huilin Qu and Congqiao Li and Sitian Qian. Particle Transformer for Jet Tagging. *preprint arXiv:2202.03772*, 2024.

[15] The CMS Collaboration. Novel strategy targeting HH and HHH production at High Level Trigger in Run 3. *CERN Document Server: 2868787*, 2023.