PROCEEDINGS OF SCIENCE

Search for resonant and non-resonant di-Higgs boson production at CMS using jet substructure techniques

Valeria D'Amante[a,b,*] **on behalf of the CMS Collaboration**

[a] *Università di Siena,*
Via Roma, 56, Siena, Italia
[b] *INFN Sezione di Pisa,*
Largo Bruno Pontecorvo, 3/Edificio C, Pisa, Italia

E-mail: v.damante@cern.ch

The measurement of pair-production of Higgs bosons is one of the key goals of the LHC. Also, beyond the standard model theories involving extra spatial dimensions predict resonances with large branching fractions in a pair of Higgs bosons with negligible branching fractions to light fermions. We present an overview of searches for resonant and nonresonant Higgs boson pair production at high transverse momentum, using proton-proton collision data collected with the CMS detector at the CERN LHC. These results use novel analysis techniques to identify and reconstruct highly boosted final states that are created in these topologies.

41st International Conference on High Energy physics - ICHEP2022
6-13 July, 2022
Bologna, Italy

*Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

https://pos.sissa.it/

PoS(ICHEP2022)504

1. Double Higgs production at LHC

A pair of Higgs bosons can be produced either through a high mass resonance decay or in a non resonant way. The main di-Higgs production channels at LHC are the gluon-gluon Fusion (ggF) and the Vector Boson Fusion (VBF). The non-resonant production is sensitive to the Higgs self-coupling λ_{HHH}, the couplings with the top quark, y_t, and the vector bosons, c_V and c_{2V} respectively, as shown in the Leading Order (LO) Feynman diagrams in figures 1 and 2; this provides a direct gateway to access the scalar sector properties, to perform independent Standard Model (SM) tests, and also to probe different Beyond Standard Model (BSM) theories with anomalous couplings using strength modifiers $\kappa_\lambda = \lambda_{HHH}/\lambda_{HHH}^{SM}$, $\kappa_t = y_t/y_t^{SM}$, $\kappa_V = c_V/c_V^{SM}$, and $\kappa_{2V} = c_{2V}/c_{2V}^{SM}$. Many BSM models foresee the presence of a new heavy particle in a mass range up to TeV scale decaying into a pair of Higgs bosons. The LO Feynman diagrams describing the double Higgs resonant production are reported in figure 3. The experimental signature would be an enhancement of the di-Higgs production cross section in correspondence of the mass of the resonance. Considering all couplings to their SM values, ggF and VBF cross sections expected at the LHC with a centre mass energy of $\sqrt{s} = 13$ TeV are, respectively, $\sigma_{HH}^{ggF} \simeq 31$ fb and $\sigma_{HH}^{VBF} \simeq 1.7$ fb. Although the smallness of VBF production cross section makes its measurement very challenging, the VBF process has a very clear signature, given by two forward jets with large invariant mass and highly separated in η.

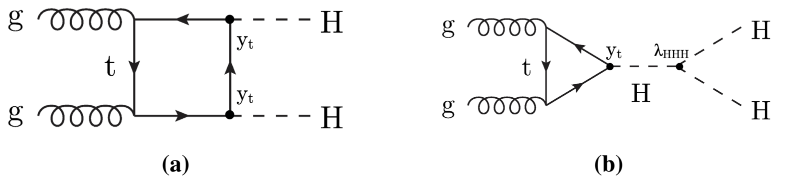


Figure 1: Feynman diagrams of the processes contributing to the non-resonant, Standard Model, double Higgs production at LO in the ggF channel.

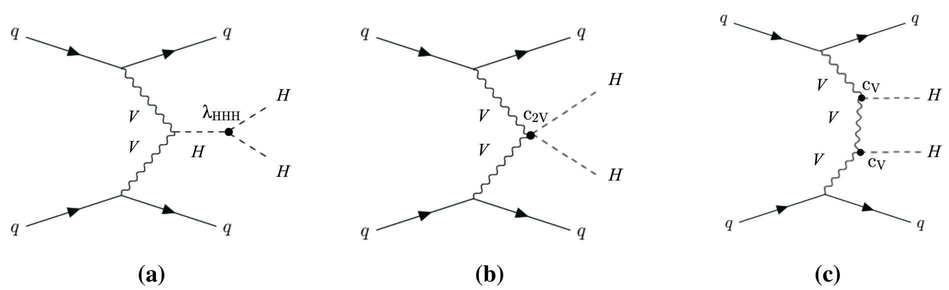
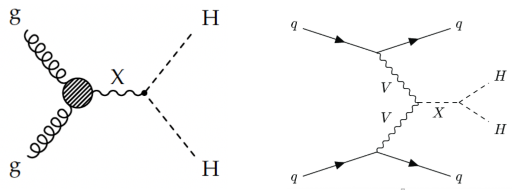


Figure 2: Feynman diagrams for the non-resonant SM double Higgs production in the VBF channel at LO.

Depending on the transverse momentum of the Higgs decay products, it is possible to distinguish the *resolved* and *boosted* regimes. In the first case, the final state particles are reconstructed as well separated, small jet cones. In the boosted regime the Higgs bosons are highly energetic so their decay products are Lorentz boosted in the detector frame and consequently detected as a single large jet. A large effort has been made to identify the substructure of these large jets and therefore to reconstruct the properties of the particles that generated them. The ggF Feynman diagrams in

PoS(ICHEP2022)504



(a) **(b)**

Figure 3: Feynman diagrams for the resonant double Higgs production in the ggF (3a) and VBF (3b) channels at LO.

figures 1a and 1b undergo to a destructive interference. This means that any small deviation from SM prediction cause large changes in the distribution of m_{HH}, thus affecting also the kinematics. This effect is particularly evident in the high invariant mass region, where Higgs bosons are more energetic. This makes boosted topologies particularly suitable to perform studies on the Higgs couplings.

2. Jet definition and reconstruction at CMS

The offline event reconstruction at the CMS experiment is based on the Particle Flow (PF) algorithm, which combines the information coming from all the CMS detector layers to identify each final-state particle, and the corresponding measurements to reconstruct the particle properties on the basis of this identification. In the PF algorithm, physics objects are identified: isolated photons and electrons, muons, missing transverse momentum ($\mathrm{p_T^{miss}}$) and hadronic *jets* [1]. The jet clustering algorithm adopted by the CMS experiment is the *anti-k_t* algorithm. Based on the cone radius, it is possible to identify two jet types: *Small radius jets (SRJ)*, with a radius parameter $R = 0.4$. Such jets usually originate from light flavour quarks and gluons; *Large radius jets (LRJ)*, with a radius parameter $R = 0.8$. They contain decay products from highly energetic heavy particles as top quarks, W, Z and Higgs bosons: such particles have a Lorentz boost and their decay products are highly collimated. So the hadronic decays of such energetic particles are reconstructed as two largely overlapping jets, resulting in a larger, "fat" jet.

The study of large radius jets substructure is crucial for identifying the jet components and the process that lead to the formation of the jet (as highly energetic W/Z/H bosons, top quark decay, or QCD processes), and it represents a key feature to perform SM precision measurements, and BSM searches. Several cut-based jet tagging algorithm have been developed exploiting high-level substructure observables: the *soft-drop mass* ($\mathrm{m_{SD}}$), is the reconstructed jet mass after removing from the jet components soft radiation mostly due to pile-up or underlying events, the *N-subjettiness* (τ_N), which is an indicator of the compatibility of a jet with having N subjets, and energy correlation functions. The advantage of cut-based algorithms on jet substructure observables is the robustness and the easy interpretation. In the last years, machine learning based algorithms have been developing and they demonstrate to have higher discrimination power than cut based ones, since they exploit a larger amount of available jet information. The **DeepAK8** algorithm is a multiclass classifier aiming to identify the hadronically decaying particles, defining five main categories: W/Z/H bosons, top quark and other. These classes are further splitted into sub-categories, corresponding

PoS(ICHEP2022)504

to the decay modes of each particle, for example H $\rightarrow b\bar{b}$ [2]. In the DeepAK8 algorithm, two lists of inputs are defined for each jet and used to feed the algorithm with particles and vertices information, respectively. The background jets selected by DeepAK8 exhibit a modified mass distribution similar to that of the signal. An alternative version of the DeepAK8 algorithm has been developed in order to decorrelate the information on the jet mass while keeping high performances exploiting an adversarial training, the **DeepAK8-MD**.

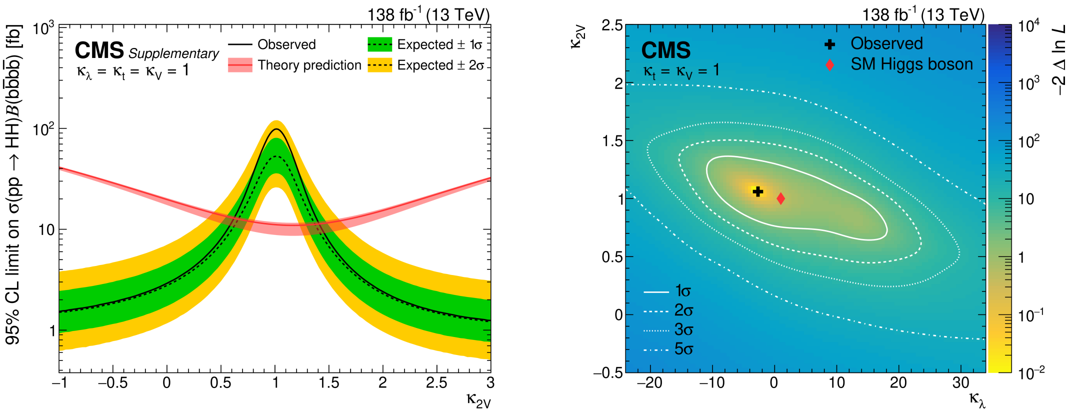
The **ParticleNet** algorithm is based on a graph neural network, aiming to identify hadronic decays of highly Lorentz-boosted top quarks and W, Z, and Higgs bosons, further dividing the final categories according to the decay modes. The algorithm is fed with the same input features as the DeepAK8 one. The key feature of ParticleNet is the edge convolution (EdgeConv) operation and the dynamic graph CNN method, applied on a permutational-invariant set of points (particles), each associated to a feature vector containing low-level information. As for the DeepAK8 algorithm, the ParticleNet tends to modify background jets mass distribution, making it similar to the signal one. Therefore also in this case a mass-decorrelated version of the tagger, the **ParticleNet-MD** algorithm [3] has been introduced.

3. Searches for non resonant HH → $\mathbf{b\bar{b}b\bar{b}}$

The search for non-resonant boosted Higgs boson pair production via ggF and VBF with four bottom quarks in final state is performed using proton-proton collisions at a center-of-mass energy of $\sqrt{s}$ = 13 TeV. The analysis is performed on data collected by the CMS experiment during Run2, corresponding to an integrated luminosity of 138fb^{-1}. It is focused in the phase space of the boosted regime, where Higgs bosons are highly energetic. The Higgs decay products are reconstructed with LRJs. Events are pre-selected applying different triggers setting thresholds on total hadronic transverse energy in the event or on the AK8 jet p_T . Events that pass the trigger selection are required to have at least two AK8 jets with a set of kinematic quality cuts. In this analysis for the first time the ParticleNet-MD discriminator, D_{bb} , is used to identify the $H \rightarrow bb$. The main sources of backgrounds come from the production of a $t\bar{t}$ pair and QCD processes with multi-jets in the final state. These contributions are estimated by fitting the corresponding distributions simultaneously via a maximum likelihood fit to data in Control Regions (CRs) orthogonal to the region where the search is performed. The two main categories - ggF and VBF - present orthogonal topologies: VBF events are selected requiring two additional AK4 jets separated from both Higgs bosons candidate AK8 jets and with large di-jet mass and angular separation. To get ggF subcategories, a BDT trained to discriminate between the HH signal and the main background processes is exploited. A binned maximum likelihood fit to data and the sum of the signal and background contributions has been performed. The fit observables are: the reconstructed HH mass (m_{HH}), D_{bb} , the mass obtained through the ParticleNet mass regression of the sub-leading jet in p_T and BDT distributions. The fit is performed simultaneously in all ggF and VBF categories and in the CRs. The test statistic chosen to determine the signal yield is based on the profile likelihood ratio, considering systematic uncertainties as nuisance parameters and treated according to the frequentist paradigm. The data are found to agree with the background-only hypothesis. As shown in fig 4, the values of κ_λ and κ_{2V} are observed (expected) to be in the ranges $[-9.9, 16.9]([-5.1, 12.2])$and $[0.62, 1.41]([0.66, 1.37])$, respectively [4]. For the first time the κ_{2V} = 0 value has been excluded with a significance of 6.3

PoS(ICHEP2022)504

standard deviations when keeping the values of κ_λ, κ_t and κ_V to the SM ones, as shown in fig 4a. In fig 4b the two-dimensional profile likelihood test statistic ($-2\Delta \ln \mathcal{L}$) scan in data as a function of κ_λ and κ_{2V}.



(a) **(b)**

Figure 4: Fig 4a: the κ_{2V} scan showing the expected and observed 95% CL upper limits on HH production, keeping κ_λ,κ_V and κ_t to the SM values.Fig 4b: the two-dimensional profile likelihood test statistic ($-2\Delta \ln \mathcal{L}$) scan in data as a function of κ_λ and κ_{2V}. The black cross indicates the minimum, while the red diamond marks the SM expectation. Contours enclose the 1, 2, 3, and 5 σ CL regions are shown [4].

This analysis provided a factor of 30 improvement over the previous search for a pair of boosted H→bb jets at CMS, performed on data collected in 2016 corresponding to a total integrated luminosity of 36 fb^{-1}. This result is achieved thanks to the increased statistics and to the introduction of advanced $H \to bb$ identification techniques.

4. Searches for resonant double Higgs production with two b quarks and leptons in the final state

This analysis uses proton-proton collisions at a center-of-mass energy of 13 TeV collected by the CMS detector at the LHC during Run2, corresponding to an integrated luminosity of 138 fb^{-1}. The investigated resonance mass spectrum spans between 0.8 and 4.5 TeV. One of the two Higgs bosons is required to decay into a bottom quark pair (bb) and the other into final states with leptons. This corresponds to three decay channels: $HH \to b\bar{b}WW^* \to b\bar{b}\ell\nu qq$, the single-lepton (SL) channel, $HH \to b\bar{b}WW^* \to b\bar{b}\ell\nu\ell\nu$ and $HH \to b\bar{b}\tau^+\tau^- \to b\bar{b}\ell\nu\ell\nu$, the di-lepton (DL) channels, where ℓ = e, μ. The Higgs bosons are required to be highly energetic, and their Lorentz boost produces a distinct experimental signature with an AK8 jet with a substructure consistent with the decay $H \to b\bar{b}$. In the SL channel also the $W \to qq$ is reconstructed as a merged, AK8 jet. The AK8 jets are identified through the DeepAK8-MD $Z/H \to bb$ tagger, denoted $D_{Z/H \to bb}$, with a working point corresponding to a selection efficiency of ≈ 85% and a misidentification probability to QCD jets smaller than 1%. Events are pre-selected by applying a set of triggers requiring isolated muons or electrons with transverse momentum threshold depending on the year of data taking, total hadronic transverse energy above certain threshold. The Higgs boson candidates are required to have large p_T and angular separation between each other. The HH system mass is reconstructed

PoS(ICHEP2022)504

with different strategies depending on the two main channels. Events are furthermore divided in 8 categories for SL and 4 for DL, for a total of 12 orthogonal regions. In addition, separate criteria are applied to define CRs to provide background processes contributions estimates. The main background sources arise from $t\bar{t}$ production, the production of W (Z/γ^*) bosons leptonically decaying associated with jets for the SL (DL) channel and QCD multi-jets events. The signal and background contribution are estimated through a simultaneous maximum likelihood fit to the 2D distribution of $m_{b\bar{b}}$ and m_{HH}, and systematic uncertainties are included in the fit as nuisance parameters. The fit is performed using one model containing only background processes and another with both background and signal processes. The background-only model fits the data well. The results are interpreted as upper limits at 95% CL on the product of the cross section and branching fraction of resonant HH production, as shown in fig 5 . The introduction of machine-learning based

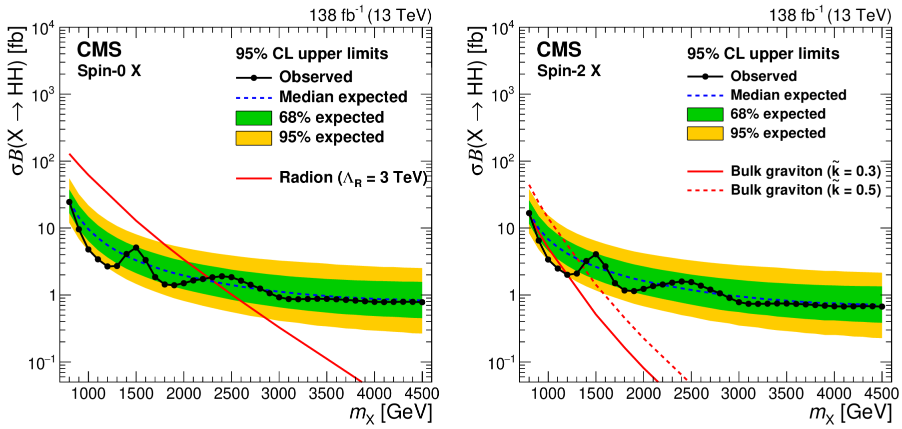


Figure 5: Observed and expected 95% CL upper limits on $\sigma\mathcal{B}(X \to HH)$ for a generic spin-0 (left) and spin-2 (right) boson X, as functions of mass. Example radion and bulk graviton predictions are also shown. The HH branching fraction is assumed to be 25% for radions and 10% for bulk gravitons [5].

jet tagger for boosted H(bb) decays and the increased statistics with respect to the previous analysis, that exploited only 2016 data, allowed to set the most stringent exclusion limits to date for $X \to HH$ signatures with leptons in the final state [5].

References

[1] CMS Collaboration, "The CMS experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, p. S08004, aug 2008.

[2] CMS Collaboration, "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques," *Journal of Instrumentation*, vol. 15, p. P06005, jun 2020.

[3] CMS Collaboration, "Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques," CMS DP Note 2020/002, 2020.

[4] CMS Collaboration, "Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks," CMS Note 2022/003, 2022.

PoS(ICHEP2022)504

[5] CMS Collaboration, "Search for heavy resonances decaying to a pair of Lorentz-boosted Higgs bosons in final states with leptons and a bottom quark pair at $\sqrt{s}$= 13 TeV," *JHEP*, vol. 2205, p. 005, 2022.

PoS(ICHEP2022)504