CMS Physics Analysis Summary

Contact: cms-pag-conveners-b2g@cern.ch

2022/03/11

Nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks

The CMS Collaboration

Abstract

A search for nonresonant Higgs boson (H) pair production via gluon and vector boson (V) fusion is performed in the final state where each H decays to a bottom quarkantiquark pair, using 138 fb⁻¹ of proton-proton collision data collected by the CMS experiment at $\sqrt{s} = 13$ TeV. The analysis targets highly Lorentz-boosted H jets identified using a graph neural network classifier. The total HH production cross section is observed (expected) to be smaller than 9.9 (5.1) relative to the standard model (SM) prediction. The search also yields constraints on the coupling strengths relative to the SM of the H self-coupling, $\kappa_{\lambda} \in [-9.9, 16.9]$, the trilinear VVH coupling, $\kappa_{V} \in [-1.17, -0.79] \cup [0.81, 1.18]$, and the quartic VVHH coupling, $\kappa_{2V} \in [0.62, 1.41]$, at 95% confidence level, excluding $\kappa_{2V} = 0$ for the first time, when other H couplings are fixed to their SM values.

1. Introduction

1 Introduction

The discovery of a Higgs boson by the ATLAS and CMS Collaborations at the CERN LHC in 2012 [1, 2] began a new era in experimental high energy physics. Since its original observation, significant efforts have been made to measure the properties of the Higgs boson, including its self-coupling. The production of a pair of Higgs bosons (HH) is a rare process that provides a unique handle to access the Higgs boson trilinear self-coupling (λ) and the quartic coupling between two Higgs bosons and two vector bosons.

The HH production takes place primarily via gluon fusion (ggF) and vector boson fusion (VBF). In the standard model (SM), the ggF cross section at $\sqrt{s} = 13$ TeV for $m_{\rm H} = 125$ GeV is calculated to be $\sigma_{\rm ggF} = 31.05$ fb [3] and is sensitive to the values of λ and the top quark Yukawa coupling, $y_{\rm t}$. Variations from their SM values are parameterized as $\kappa_{\lambda} = \lambda/\lambda^{\rm SM}$ and $\kappa_{\rm t} = y_{\rm t}/y_{\rm t}^{\rm SM}$. The latter has been measured to be consistent with the SM [4, 5]. The VBF cross section is $\sigma_{\rm VBF} = 1.726$ fb [3] and depends on the strength of the self-coupling (κ_{λ}) and on the interaction of pairs of Z or W bosons (jointly denoted as V) with a single (VVH) and a pair (VVHH) of Higgs bosons, whose values with respect to the SM prediction are parameterized as $\kappa_{\rm V}$ and $\kappa_{\rm 2V}$, respectively. The ATLAS and CMS Collaborations have performed studies of Higgs boson pair production at $\sqrt{s} = 7$, 8, and 13 TeV in the bb $\gamma\gamma$ [6–9], bb $\tau^-\tau^+$ [10, 11], bb bb [12–17], bb VV [18–21] channels, as well as combinations of channels [22–24].

This note reports the results of a search for nonresonant HH production via both the ggF and VBF modes where each Higgs boson decays to a bottom quark-antiquark pair ($b\bar{b}$). It is based on data from proton-proton (pp) collisions at the LHC at a center-of-mass energy of 13 TeV, collected by the CMS experiment in 2016–2018, with an integrated luminosity of 138 fb⁻¹. We select events with both Higgs bosons in the highly Lorentz-boosted regime, i.e. with sufficiently large transverse momentum (p_T) for the decay products of each H to become merged into a single large-radius jet. Events are separated into mutually exclusive categories targeting either ggF or VBF HH production, distinguished by the presence of additional small-radius jets characteristic of the VBF process. A novel graph neural network (GNN) algorithm, ParticleNet [25], is used for the first time to select such large-radius jets, enhancing signal over background in a mass-independent way.

2 The CMS detector

The CMS apparatus [26] is a multipurpose, nearly hermetic detector, designed to trigger on [27, 28] and identify electrons, muons, photons, and charged and neutral hadrons [29–32]. A global reconstruction "particle-flow" (PF) algorithm [33] combines the information provided by the all-silicon inner tracker and by the crystal electromagnetic (ECAL) and brass-scintillator hadron calorimeters, operating inside a 3.8 T superconducting solenoid, with data from gas-ionization muon detectors interleaved with the solenoid return yoke, to build tau leptons, jets, missing transverse momentum, and other physics objects [34–36]. The primary vertex (PV) is taken to be the vertex corresponding to the hardest scattering in the event, evaluated using tracking information alone [37].

3 Simulated samples

Simulated samples for the ggF process are generated at next-to-leading-order (NLO) accuracy using POWHEG 2.0 [38–41]. Nonresonant VBF HH samples are generated at leading-order (LO) accuracy in perturbative quantum chromodynamics (QCD) using MADGRAPH5_AMC@NLO

2.6.5 [42]. A range of samples corresponding to different combinations of the κ_V , κ_{2V} , and κ_λ couplings is generated. Samples for other coupling combinations are constructed as linear combinations of the original generated samples by applying appropriate event weights. The ggF samples are normalized to the next-to-NLO (NNLO) cross section [43–49] corresponding to the coupling values considered. The VBF sample with SM couplings is normalized to the next-to-NNLO (N³LO) cross section, and the same N³LO/LO correction is applied to the VBF samples with modified couplings.

Samples for top quark-antiquark pair production (tt) background process are generated at NLO accuracy using POWHEG v2.0 [38-40, 50, 51] with FxFx jet matching and merging [52], and normalized to the theoretical cross section calculated at NNLO precision using TOP++ v2.0 [53]. The differential tt cross section as a function of top quark $p_{\rm T}$ is corrected to the NNLO QCD + NLO electroweak accuracy [54]. Samples for other background processes, such as the QCD multijet process consisting uniquely of jets produced through the strong interaction, single top quark, Z and W single-boson and diboson production processes, as well as single H production in all relevant production modes, are also generated. Parton showering, hadronization, and the underlying event are modeled by PYTHIA 8.205 [55] with parameters set by the CUETP8M1 [56] and CP5 tunes [57] used for samples simulating the 2016 and 2017–2018 conditions, respectively. The NNPDF 3.0 [58] and 3.1 [59] parton distribution functions (PDFs) are used in the generation of all simulated samples. The GEANT4 [60] package is used to model the response of the CMS detector, and simulated minimum-bias interactions are mixed with the hard interactions in simulated events to model additional pp interactions within the same or nearby bunch crossings (pileup). The simulated events are weighted to match the pileup distribution measured in data.

4 Event reconstruction and selection

The particles reconstructed by the PF algorithm are clustered into jets using the anti- $k_{\rm T}$ algorithm [61] implemented in the FASTJET package [62] with a distance parameter of 0.8 (largeradius jets) or 0.4 (small-radius jets). To mitigate the effect of pileup on small-radius jets, the charged hadron subtraction algorithm [33] is applied. For large-radius jets, the pileup per particle identification algorithm [63, 64] is used to assign a weight to each particle prior to jet clustering based on the likelihood of the particle originating from the PV. Further corrections are applied to the jet energy as a function of jet pseudorapidity (η) and $p_{\rm T}$ to account for detector response nonlinearities [35].

To isolate the HH signal, two high- p_T large-radius jets are required. A combination of several trigger algorithms is used, all requiring the total hadronic transverse energy in the event (H_T) or jet p_T to be above a given threshold. In addition, a minimum threshold on the jet mass is imposed after removing remnants of soft radiation [65] to reduce the H_T or p_T thresholds and improve the signal acceptance. The trigger selection efficiency varies between 10% and 95% for jets with 300 < p_T < 450 GeV and is fully efficient for jets with p_T > 500 GeV. Events are grouped into mutually exclusive ggF and VBF categories. The ggF categories are designed to select SM ggF HH signal events, so the selected events are required to contain at least two large-radius jets with p_T > 300 GeV. The VBF categories are designed to select VBF signal events with a characteristic signature of two additional, opposite small-radius jets in the forward regions. The leading (subleading) large-radius jet is required to have p_T > 500 (400) GeV in order to target non-SM coupling hypotheses that can dramatically increase the VBF cross section, especially for highly-boosted Higgs boson pairs [66].

One of the main challenges for this search is the efficient reconstruction of the pair of H de-

cays, while rejecting light-flavor quarks and gluons. GNN algorithms [25, 67–69] have been shown to achieve state-of-the-art performance in jet classification [70]. For the ParticleNet algorithm [25], the inputs comprise measured properties of jet constituent particles, such as the $p_{\rm T}$, angular separation between the particle and the jet axis, and impact parameter relative to the PV, and information on the secondary vertices [71] within the jet cone, such as the flight distance and the number of associated tracks. The algorithm treats each jet as an unordered set of its constituents, builds constituent-specific representations considering potential correlations between them, and returns a permutation-invariant output corresponding to the probability that the jet belongs to the H \rightarrow bb or QCD multijet class. To achieve independence from the jet mass, the classifier is trained using dedicated simulated samples containing spin-0 particles with a uniform mass spectrum between 15 and 250 GeV and decaying to quark-antiquark pairs, as well as a QCD multijet sample [72]. In addition, jets in each sample are reweighted to obtain uniform distributions in both $p_{\rm T}$ and mass for the training. The ParticleNet algorithm yields significant improvements both in terms of performance and jet mass decorrelation compared to previous jet classifier algorithms [73]. The discriminant $(D_{b\bar{b}})$ is calibrated using data and simulated samples dominated by QCD multijet events, and p_T -dependent simulation-to-data correction factors, typically ranging from 0.9 to 1.1, are extracted and applied to correct the selection efficiency of the H jets. Three working points (WPs) based on $D_{b\bar{b}}$ are used: these *tight, medium,* and *loose* WPs correspond to $H \rightarrow b\overline{b}$ selection efficiencies of \approx 56%, \approx 77%, and \approx 82%, and QCD misidentification rates of \approx 0.3%, \approx 1%, and \approx 2%, respectively.

The search for Higgs boson pairs relies on the ability to accurately reconstruct the mass of the two Higgs boson candidates. The soft-drop (SD) algorithm [74] with angular exponent $\beta = 0$ and soft radiation fraction z = 0.1, also known as the modified mass-drop algorithm [75], is usually applied to the Higgs boson jet candidate to remove soft and wide-angle radiation. The SD algorithm occasionally over-subtracts genuine jet constituents that account for a large fraction of the jet momentum, resulting in a negligible jet mass even for large-mass resonances. To improve the jet mass estimation, we introduce an algorithm based on the ParticleNet GNN architecture to regress the jet mass [76]. The target of the algorithm is the true resonance mass for the heavy spin-0 particles, or the SD mass (m_{SD}) for jets clustered from the generated particle constituents for QCD multijet events. The training is performed with the same samples and a similar event weighting scheme as for jet classification. Data control samples, dominated by t \overline{t} events, are used to calibrate the regressed jet mass m_{reg} . The small corrections to mass scale (< 1%) and resolution ($\approx 3\%$) are then applied to all simulated samples in the rest of the analysis, accompanied by systematic uncertainties that account for the differences in regression performance between the jets originating from W and Higgs boson decays.

Electrons are reconstructed and their momentum is estimated by combining the momentum measurement from the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track [29]. Muons are reconstructed as tracks in the central tracker, consistent with either a track or several hits in the muon chambers, and their momentum is obtained from the curvature of the corresponding track [30]. Electron candidates are reconstructed within the tracker acceptance of $|\eta| < 2.5$, whereas muon candidates are also required to be within the muon chamber acceptance of $|\eta| < 2.4$. The electron and muon candidates are required to have $p_T > 10 \text{ GeV}$. Only tracks consistent with the PV are associated with the electron or muons, and additional identification criteria [29, 30] are applied to improve the purity of the lepton selection. To suppress leptons stemming from bottom or charm hadron decays, the electrons and muons are required to be isolated [77].

Events containing at least two large-radius jets with $|\eta| < 2.4$ and no isolated electrons or muons are selected and grouped into the ggF and VBF categories. For the VBF categories, we require the two large-radius jets to have m_{reg} between 50 and 200 GeV. As the VBF process is characterized by the presence of two forward jets with a large dijet invariant mass and a gap in pseudorapidity, we also require two additional small-radius jets with $p_{\rm T} > 25$ GeV and $|\eta| < 4.7$, referred to as the VBF jets. The pseudorapidity gap $\Delta \eta_{jj}^{VBF}$ between the two VBF jets must be larger than 4.0, and the invariant mass of the two VBF jets m_{ij}^{VBF} must be larger than 500 GeV. The azimuthal angle between the two large-radius jets, $\Delta \phi_{j_1j_2}$, is required to be larger than 2.6, and the $\Delta \eta_{j_1 j_2}$ must be less than 2.0. Three VBF event categories are defined based on the $D_{b\bar{b}}$ scores of the two H candidate jets. Events in the high-purity (HP) category must have both jets passing the tight $D_{b\overline{b}}$ WP; events not in the HP category are categorized as medium purity (MP) if both jets pass the medium $D_{b\overline{b}}$ WP; and events not in the HP or MP categories are categorized as low purity (LP) if both jets pass the loose $D_{b\overline{b}}$ WP. The definitions of the three categories based on the corresponding $D_{b\overline{b}}$ WPs were optimized to provide the best combined sensitivity to VBF HH production, while enabling a robust background estimation in different HH mass ($m_{\rm HH}$) intervals and a reliable $D_{\rm b\overline{b}}$ calibration.

Events not sorted into the VBF categories are considered for the ggF categories. The two selected large-radius jets are ordered by the value of $D_{b\bar{b}}$. The leading- $D_{b\bar{b}}$ jet is required to have $m_{\rm SD} > 50 \,{\rm GeV}$, while the $D_{\rm b\bar{b}}$ -subleading jet is required to have $m_{\rm reg} > 50 \,{\rm GeV}$. A boosted decision tree (BDT) is trained to discriminate between the HH signal and QCD multijet or $t\bar{t}$ background processes. Input variables to the BDT include the $p_{\rm T}$ and substructure variable τ_{32} [73, 78] of the two selected large-radius jets; the mass, η , and $D_{b\overline{b}}$ of the first large-radius jet; the missing transverse momentum; the $p_{\rm T}$, η , and mass $m_{\rm ii}$ of the dijet system; and the ratios p_{T1}/m_{ii} , p_{T2}/m_{ii} , and p_{T2}/p_{T1} . Based on the BDT output score, tight, medium, and loose WPs are defined, corresponding to HH signal selection efficiencies of \approx 23%, \approx 27% and \approx 33%, and QCD and tt misidentification rates of $\approx 1\%$, $\approx 2\%$ and $\approx 12\%$, respectively. Three search region (SR) event categories targeting the ggF production are constructed. Events in the highest signal purity category 1 are required to pass the tight BDT WP and the $D_{b\overline{b}}$ -subleading large-radius jet is required pass the tight $D_{b\overline{b}}$ WP. Events not in category 1 are grouped into category 2 if they pass the tight BDT WP and the second H candidate jet passes the medium $D_{b\bar{b}}$ WP, or if they fail the tight but pass the medium BDT WP and the second jet passes the tight $D_{b\bar{b}}$ WP. Finally, events not in category 1 or 2 are grouped into category 3 if they pass the loose BDT WP and the second jet passes the medium $D_{b\bar{b}}$ WP. The categorization thresholds are chosen to maximize the expected sensitivity to ggF signals.

5 Background estimation

After all selection requirements, the remaining SM background mainly consists of two processes: $t\bar{t}$ and QCD multijet production. These background contributions are estimated by fitting the corresponding distributions simultaneously via a maximum-likelihood fit to data as detailed in the following, similarly to previous CMS searches for boosted H signals [79–81]. Subdominant backgrounds include gluon fusion and VBF H production, $t\bar{t}$ H production, VH production, V+jets production, and diboson production. They are estimated from simulation or a data-driven method in the ggF event categories, and found to be negligible in the VBF event categories.

A control region (CR) enriched in QCD multijet events is selected by changing the requirement

on the $D_{b\overline{b}}$ discriminant. For the VBF categories, this CR is defined by both H jet candidates failing the loose $D_{b\overline{b}}$ WP and only passing a very loose preselection on $D_{b\overline{b}}$. The sidebands of the mass of the p_{T} -subleading large-radius jet are used to obtain transfer factors that account for the ratio of selection efficiencies in the QCD-enriched low- $D_{b\overline{b}}$ CR and the high- $D_{b\overline{b}}$ VBF SRs. A separate transfer factor is derived using the subleading large-radius jet mass sidebands for each m_{HH} interval in each search category (HP, MP, and LP) for each year. For the ggF categories, the QCD-enriched CR is defined by the second large-radius jet failing the medium $D_{b\overline{b}}$ WP and passing the loose BDT WP. The QCD multijet background in the ggF SRs is estimated using the m_{reg} distribution of the $D_{b\overline{b}}$ -subleading jet, assuming a constant transfer factor [81]. We have checked whether a higher-order polynomial dependence on the mass shape is necessary using a Fisher *F*-test [82] and Kolmogorov-Smirnov [83] and saturated model [84, 85] goodness-of-fit tests, but a constant factor is found to be sufficient.

For the VBF categories, an auxiliary sample enriched in semileptonic t \bar{t} events is used to extract two corrections for the t \bar{t} background events: one accounting for the difference in $H \rightarrow b\bar{b}$ misidentification efficiency in t \bar{t} events between data and simulation, measured separately for each $D_{b\bar{b}}$ WP, and the other for the overall normalization of the t \bar{t} process, measured inclusively. We define this sample with a set of selections closely following the definition of the t \bar{t} -enriched region in Ref. [73]. For the ggF categories, the simulation prediction of the BDT and $m_{\rm reg}$ distribution for the top quark background is validated and corrected using a semileptonic t \bar{t} sample with $\tau_{32} < 0.46$ required for two large-radius jets.

The dominant uncertainty in the analysis, accounting for 37% of the uncertainty on the extracted signal yield, is the systematic uncertainty in the QCD multijet background estimate. It is accounted for by profiling the nuisance parameters related to the data-driven background estimation. Subdominant uncertainty sources include the simulation modeling of the $D_{b\bar{b}}$ shape and selection efficiency (18%), the jet energy and mass scale and resolution (16%), the modeling of the top quark background (8%), the theoretical uncertainties from the PDFs, missing higher-order QCD corrections, and initial and final state radiation (14%), trigger efficiency measurement (5%), luminosity measurement (3%), and pileup modeling (2%).

6 Results

A binned maximum likelihood fit to the observed $m_{\rm HH}$, $D_{b\bar{b}}$, $m_{\rm reg}$, and BDT distributions is performed using the sum of the signal and background contributions. Because part of the VBF signal can enter the ggF categories and vice versa, the fit is performed simultaneously in all ggF and VBF categories, including the CRs. The results of the fit in the ggF BDT event categories 1, 2, and 3 are shown in Fig. 1. The results of the fit in the VBF LP, MP, and HP categories are shown in Fig. 2.

The test statistic chosen to determine the signal yield is based on the profile likelihood ratio [86]. Systematic uncertainties are incorporated into the analysis via nuisance parameters and treated according to the frequentist paradigm. The best fit value of each parameter of interest and approximate 1, 2, 3, and 5 standard deviation (σ) confidence level (CL) intervals are extracted following the procedure described in Section 3.2 of Ref. [87]. Figure 3 shows the profile likelihood test statistic scan in data over the (κ_{λ} , κ_{2V}) and (κ_{2V} , κ_{V}) planes.

Upper limits on the HH production cross section at 95% CL based on the CL_s criterion [88, 89] are obtained using asymptotic formulae [90]. The upper limit on the total HH production cross section is observed (expected) to be 9.9 (5.1) relative to the SM prediction. The breakdown



Figure 1: The data and fitted signal and background distributions for the $D_{b\bar{b}}$ -subleading jet regressed mass are shown for the ggF BDT event categories 1 (top), 2 (lower left), and 3 (lower right). The SM HH ($\kappa_{2V} = \kappa_V = \kappa_{\lambda} = 1$) signal is shown scaled to the best fit signal strength $\mu = 3.5$. The lower panel shows ratio of the data and the total prediction, with its uncertainty represented by the hatched fill. The data points are represented with their data statistical uncertainty.

of the observed and expected HH production cross section upper limits by category is shown in Fig. 4. The 95% CL upper limits on κ_{λ} , κ_{2V} , and κ_{V} , assuming all other Higgs boson couplings to be at SM values, are shown in Fig. 5. The observed (expected) 95% CL interval for κ_{λ} is found to be [-9.9, 16.9] ([-5.1, 12.2]), assuming all other Higgs boson couplings to be at their SM values. Similarly, the observed (expected) 95% CL interval for κ_{2V} are found to be [0.62, 1.41] ([0.66, 1.37]). The same observed (expected) interval for κ_{V} is found to be [-1.17, -0.79] \cup [0.81, 1.18] ([-1.17, -0.79] \cup [0.80, 1.18]) The lack of a large observable VBF HH signal establishes a nonzero value of κ_{2V} for the first time, with a significance of 6.3 σ .



Figure 2: The distributions of the invariant mass of the HH system after a background-only fit to the data, for the VBF low-purity, medium-purity, and high-purity categories. The VBF signal for $\kappa_{2V} = 0$, $\kappa_{V} = \kappa_{\lambda} = 1$, is shown in red with the vertical error bar indicating the prefit uncertainty. The lower panel shows ratio of the data and the total background prediction, with its uncertainty represented by the hatched fill. The data points are represented with their data statistical uncertainty.

7 Summary

In summary, a search for nonresonant Higgs boson (H) pair production via gluon fusion and vector boson (V) fusion in the final state with two bottom quark-antiquark ($b\overline{b}$) pairs has been presented. The search focuses on the phase space region where both Hs are highly Lorentzboosted so that each H decay can be reconstructed as a large-radius jet. The analysis is the first to apply a novel algorithm based on graph neural networks to identify the jets that correspond to H \rightarrow bb decays. The data are found to agree with the background-only hypothesis, and an observed (expected) upper limit at the 95% confidence level is set to 9.9 (5.1) relative to the standard model (SM) cross section. This represents a factor of 30 improvement over the previous best search for a pair of boosted $H \rightarrow bb$ jets [80]. Upper limits on the production cross section are presented as a function of the coupling modifier parameters κ_{λ} , κ_{2V} , and κ_{V} , which parametrize the strengths of the H self-coupling, the quartic VVHH coupling, and the trilinear VVH coupling, respectively, relative to their SM values. The values of κ_{λ} , κ_{2V} , and κ_{V} are observed (expected) to be in the ranges [-9.9, 16.9] ([-5.1, 12.2]), [0.62, 1.41] ([0.66, 1.37]), and $[-1.17, -0.79] \cup [0.81, 1.18]$ ($[-1.17, -0.79] \cup [0.80, 1.18]$), when all H couplings except the one being scanned are assumed equal to their SM values. In particular, the search excludes $\kappa_{2V} = 0$ for the first time with a significance of 6.3 standard deviations when $\kappa_{\lambda} = \kappa_{t} = \kappa_{V} = 1$.



Figure 3: Two-parameter profile likelihood test statistic $(-2\Delta \ln \mathcal{L})$ scan in data as a function of κ_{λ} and κ_{2V} (left) and κ_{2V} and κ_{V} (right). The black cross marks the minimum, while the red diamond marks the SM expectation. The gray solid, dashed, dotted, and dash-dotted contours enclose the 1, 2, 3, and 5 σ CL regions, respectively.



Figure 4: Expected and observed 95% CL upper limits on HH production with respect to the SM expectation in the individual ggF and VBF search categories and their combination.



Figure 5: Observed (solid line) and expected (dashed line) 95% CL exclusion limit on the product of the inclusive HH production cross section and the branching fraction into $b\overline{b}b\overline{b}$ as a function of κ_{λ} (upper), κ_{2V} (lower left), and κ_{V} (lower right) is shown with other couplings fixed to the SM values. The green and yellow bands correspond to one and two standard deviations, respectively, around the median expected limit. The red solid line depicts the theoretical prediction for the HH production cross section and the red band shows its uncertainty [45, 46, 91–98]. The crossings of the observed limit and the theoretical cross section indicate the ranges of the coupling values excluded at 95% CL.

References

- [1] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* 716 (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* 716 (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [3] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the nature of the Higgs sector", CERN Yellow Rep. Monogr. 2 (2017) doi:10.23731/CYRM-2017-002, arXiv:1610.07922.
- [4] CMS Collaboration, "Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ", Eur. Phys. J. C **79** (2019) 421, doi:10.1140/epjc/s10052-019-6909-y, arXiv:1809.10733.
- [5] ATLAS Collaboration, "Combined measurements of Higgs boson production and decay using up to 80 fb⁻¹ of proton-proton collision data at √s = 13 TeV collected with the ATLAS experiment", *Phys. Rev. D* 101 (2020) 012002, doi:10.1103/PhysRevD.101.012002, arXiv:1909.02845.
- [6] ATLAS Collaboration, "Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", 2021. arXiv:2112.11876. Submitted to *Phys. Rev. D*.
- [7] CMS Collaboration, "Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **03** (2021) 257, doi:10.1007/JHEP03(2021)257, arXiv:2011.12373.
- [8] CMS Collaboration, "Search for Higgs boson pair production in the γγbb final state in pp collisions at √s = 13 TeV", *Phys. Lett. B* 788 (2019) 7, doi:10.1016/j.physletb.2018.10.056, arXiv:1806.00408.
- [9] ATLAS Collaboration, "Search For Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State using *pp* Collision Data at $\sqrt{s} = 8$ TeV from the ATLAS Detector", *Phys. Rev. Lett.* **114** (2015) 081802, doi:10.1103/PhysRevLett.114.081802, arXiv:1406.5053.
- [10] ATLAS Collaboration, "Search for resonant and non-resonant Higgs boson pair production in the bbτ⁺τ⁻ decay channel in pp collisions at √s = 13 TeV with the ATLAS detector", *Phys. Rev. Lett.* **121** (2018) 191801, doi:10.1103/PhysRevLett.121.191801, arXiv:1808.00336. [Erratum: doi:10.1103/PhysRevLett.122.089901].
- [11] CMS Collaboration, "Search for Higgs boson pair production in the $bb\tau\tau$ final state in proton-proton collisions at $\sqrt{(s)} = 8$ TeV", *Phys. Rev. D* **96** (2017) 072004, doi:10.1103/PhysRevD.96.072004, arXiv:1707.00350.
- [12] CMS Collaboration, "Search for Higgs boson pair production in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", 2022. arXiv: 2202.09617. Submitted to *Phys. Rev. Lett.*

- [13] ATLAS Collaboration, "Search for the $HH \rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **07** (2020) 108, doi:10.1007/JHEP07 (2020) 108, arXiv:2001.05178.
- [14] CMS Collaboration, "Search for nonresonant Higgs boson pair production in the bbbb final state at $\sqrt{s} = 13$ TeV", *JHEP* 04 (2019) 112, doi:10.1007/JHEP04(2019)112, arXiv:1810.11854.
- [15] ATLAS Collaboration, "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **01** (2019) 030, doi:10.1007/JHEP01(2019)030, arXiv:1804.06174.
- [16] ATLAS Collaboration, "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. D* **94** (2016) 052002, doi:10.1103/PhysRevD.94.052002, arXiv:1606.04782.
- [17] ATLAS Collaboration, "Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} = 8$ TeVwith the ATLAS detector", *Eur. Phys. J. C* **75** (2015) 412, doi:10.1140/epjc/s10052-015-3628-x, arXiv:1506.00285.
- [18] ATLAS Collaboration, "Search for non-resonant Higgs boson pair production in the $bb\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **801** (2020) 135145, doi:10.1016/j.physletb.2019.135145, arXiv:1908.06765.
- [19] ATLAS Collaboration, "Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay mode at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP* **04** (2019) 092, doi:10.1007/JHEP04 (2019) 092, arXiv:1811.04671.
- [20] CMS Collaboration, "Search for resonant pair production of Higgs bosons in the *bbZZ* channel in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. D* **102** (2020) 032003, doi:10.1103/PhysRevD.102.032003, arXiv:2006.06391.
- [21] CMS Collaboration, "Search for resonant and nonresonant Higgs boson pair production in the $b\bar{b}\ell \,^{\circ}\ell \,^{\circ}$ final state in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ", *JHEP* **01** (2018) 054, doi:10.1007/JHEP01(2018)054, arXiv:1708.04188.
- [22] CMS Collaboration, "Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ", *Phys. Rev. Lett.* **122** (2019) 121803, doi:10.1103/PhysRevLett.122.121803, arXiv:1811.09689.
- [23] ATLAS Collaboration, "Combination of searches for Higgs boson pairs in *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett. B* **800** (2020) 135103, doi:10.1016/j.physletb.2019.135103, arXiv:1906.02025.
- [24] ATLAS Collaboration, "Searches for Higgs boson pair production in the $hh \rightarrow bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$ channels with the ATLAS detector", *Phys. Rev. D* **92** (2015) 092004, doi:10.1103/PhysRevD.92.092004, arXiv:1509.04670.
- [25] H. Qu and L. Gouskos, "ParticleNet: Jet tagging via particle clouds", Phys. Rev. D 101 (2020) 056019, doi:10.1103/PhysRevD.101.056019, arXiv:1902.08570.
- [26] CMS Collaboration, "The CMS experiment at the CERN LHC", JINST 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

- [27] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ", JINST **15** (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.
- [28] CMS Collaboration, "The CMS trigger system", JINST 12 (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.
- [29] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8 \text{ TeV}$ ", *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- [30] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018) P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.
- [31] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at sqrt(s) = 8 TeV", JINST 10 (2015) P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.
- [32] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", JINST 9 (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [33] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", JINST 12 (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.
- [34] CMS Collaboration, "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_{τ} in pp collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018) P10005, doi:10.1088/1748-0221/13/10/P10005, arXiv:1809.02816.
- [35] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", JINST 12 (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.
- [36] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", JINST 14 (2019) P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.
- [37] CMS Collaboration, "Technical proposal for the Phase-II upgrade of the Compact Muon Solenoid", CMS Technical Proposal CERN-LHCC-2015-010, CMS-TDR-15-02, 2015.
- [38] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", JHEP 11 (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.
- [39] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", JHEP 11 (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.
- [40] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* 06 (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

- [41] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini, "Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM", *JHEP* **02** (2012) 088, doi:10.1007/JHEP02(2012)088, arXiv:1111.2854.
- [42] 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 (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [43] M. Grazzini et al., "Higgs boson pair production at NNLO with top quark mass effects", *JHEP* **05** (2018) 059, doi:10.1007/JHEP05(2018)059, arXiv:1803.02463.
- [44] S. Dawson, S. Dittmaier, and M. Spira, "Neutral Higgs boson pair production at hadron colliders: QCD corrections", *Phys. Rev. D* 58 (1998) 115012, doi:10.1103/PhysRevD.58.115012, arXiv:hep-ph/9805244.
- [45] S. Borowka et al., "Higgs boson pair production in gluon fusion at next-to-leading order with full top-quark mass dependence", *Phys. Rev. Lett.* **117** (2016) 012001, doi:10.1103/PhysRevLett.117.012001, arXiv:1604.06447. [Erratum: doi:10.1103/PhysRevLett.117.079901].
- [46] J. Baglio et al., "Gluon fusion into Higgs pairs at NLO QCD and the top mass scheme", Eur. Phys. J. C 79 (2019) 459, doi:10.1140/epjc/s10052-019-6973-3, arXiv:1811.05692.
- [47] D. de Florian and J. Mazzitelli, "Higgs boson pair production at next-to-next-to-leading order in QCD", *Phys. Rev. Lett.* **111** (2013) 201801, doi:10.1103/PhysRevLett.111.201801, arXiv:1309.6594.
- [48] D. Y. Shao, C. S. Li, H. T. Li, and J. Wang, "Threshold resummation effects in Higgs boson pair production at the LHC", JHEP 07 (2013) 169, doi:10.1007/JHEP07(2013)169, arXiv:1301.1245.
- [49] D. de Florian and J. Mazzitelli, "Higgs pair production at next-to-next-to-leading logarithmic accuracy at the LHC", JHEP 09 (2015) 053, doi:10.1007/JHEP09(2015)053, arXiv:1505.07122.
- [50] T. Ježo et al., "An NLO+PS generator for tī and Wt production and decay including non-resonant and interference effects", Eur. Phys. J. C 76 (2016) 691, doi:10.1140/epjc/s10052-016-4538-2, arXiv:1607.04538.
- [51] S. Frixione, P. Nason, and G. Ridolfi, "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", JHEP 09 (2007) 126, doi:10.1088/1126-6708/2007/09/126, arXiv:0707.3088.
- [52] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", JHEP 12 (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.
- [53] M. Czakon and A. Mitov, "Top++: A program for the calculation of the top-pair cross-section at hadron colliders", *Comput. Phys. Commun.* 185 (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.
- [54] M. Czakon et al., "Top-pair production at the LHC through NNLO QCD and NLO EW", JHEP 10 (2017) 186, doi:10.1007/JHEP10(2017)186, arXiv:1705.04105.

- [55] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", Comput. Phys. Commun. 191 (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.
- [56] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", Eur. Phys. J. C 76 (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [57] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements", Eur. Phys. J. C 80 (2020) 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.
- [58] NNPDF Collaboration, "Parton distributions for the LHC Run II", JHEP 04 (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.
- [59] NNPDF Collaboration, "Parton distributions from high-precision collider data", Eur. Phys. J. C 77 (2017) 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.
- [60] GEANT4 Collaboration, "GEANT4—a simulation toolkit", Nucl. Instrum. Meth. A 506 (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [61] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_T jet clustering algorithm", JHEP 04 (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [62] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", Eur. Phys. J. C 72 (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
- [63] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup per particle identification", JHEP 10 (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.
- [64] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", JINST 15 (2020) P09018, doi:10.1088/1748-0221/15/09/P09018, arXiv:2003.00503.
- [65] D. Krohn, J. Thaler, and L. Wang, "Jet trimming", JHEP 02 (2010) 084, doi:10.1007/JHEP02 (2010) 084, arXiv:0912.1342.
- [66] F. Bishara, R. Contino, and J. Rojo, "Higgs pair production in vector-boson fusion at the LHC and beyond", Eur. Phys. J. C 77 (2017) 481, doi:10.1140/epjc/s10052-017-5037-9, arXiv:1611.03860.
- [67] I. Henrion et al., "Neural message passing for jet physics", in *Deep Learning for Physical Sciences Workshop at the 31st Conference on Neural Information Processing Systems*. https://dl4physicalsciences.github.io, Long Beach, CA, 2017.
- [68] E. A. Moreno et al., "JEDI-net: a jet identification algorithm based on interaction networks", Eur. Phys. J. C 80 (2020) 58, doi:10.1140/epjc/s10052-020-7608-4, arXiv:1908.05318.
- [69] E. A. Moreno et al., "Interaction networks for the identification of boosted H → bb decays", Phys. Rev. D 102 (2020) 012010, doi:10.1103/PhysRevD.102.012010, arXiv:1909.12285.
- [70] A. Butter et al., "The machine learning landscape of top taggers", SciPost Phys. 7 (2019) 014, doi:10.21468/SciPostPhys.7.1.014, arXiv:1902.09914.

- [71] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in p p collisions at 13 TeV", JINST 13 (2018) P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.
- [72] CMS Collaboration, "Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques", CMS Detector Performance Note CMS-DP-2020-002, 2020.
- [73] CMS Collaboration, "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques", JINST 15 (2020) P06005, doi:10.1088/1748-0221/15/06/P06005, arXiv:2004.08262.
- [74] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft drop", JHEP 05 (2014) 146, doi:10.1007/JHEP05 (2014) 146, arXiv:1402.2657.
- [75] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, "Towards an understanding of jet substructure", *JHEP* 09 (2013) 029, doi:10.1007/JHEP09(2013)029, arXiv:1307.0007.
- [76] CMS Collaboration, "Mass regression of highly-boosted jets using graph neural networks", CMS Detector Performance Note CMS-DP-2021-017, 2021.
- [77] CMS Collaboration, "Search for supersymmetry in pp collisions at $\sqrt{s} = 13$ TeV in the single-lepton final state using the sum of masses of large-radius jets", *JHEP* **08** (2016) 122, doi:10.1007/JHEP08(2016)122, arXiv:1605.04608.
- [78] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", JHEP 03 (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.
- [79] CMS Collaboration, "Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13 TeV$ ", *Phys. Lett. B* **781** (2018) 244, doi:10.1016/j.physletb.2018.03.084, arXiv:1710.04960.
- [80] CMS Collaboration, "Search for production of Higgs boson pairs in the four b quark final state using large-area jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **01** (2019) 040, doi:10.1007/JHEP01 (2019) 040, arXiv:1808.01473.
- [81] CMS Collaboration, "Inclusive search for highly boosted Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ", *JHEP* **12** (2020) 85, doi:10.1007/JHEP12(2020)085, arXiv:2006.13251.
- [82] R. A. Fisher, "On the interpretation of χ^2 from contingency tables, and the calculation of P", J. R. Stat. Soc. **85** (1922) 87, doi:10.2307/2340521.
- [83] F. J. Massey Jr., "The Kolmogorov-Smirnov test for goodness of fit", J. Am. Stat. Assoc. 46 (1951) 68, doi:10.1080/01621459.1951.10500769.
- [84] S. Baker and R. D. Cousins, "Clarification of the use of chi square and likelihood functions in fits to histograms", *Nucl. Instrum. Meth.* **221** (1984) 437, doi:10.1016/0167-5087 (84) 90016-4.
- [85] J. Lindsey, "Parametric Statistical Inference". Oxford science publications. Clarendon Press, 1996. ISBN 9780198523598.

- [86] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", Technical Report CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, 2011.
- [87] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV", Eur. Phys. J. C 75 (2015) 212, doi:10.1140/epjc/s10052-015-3351-7, arXiv:1412.8662.
- [88] A. L. Read, "Presentation of search results: the CL_s technique", J. Phys. G 28 (2002) 2693, doi:10.1088/0954-3899/28/10/313.
- [89] T. Junk, "Confidence level computation for combining searches with small statistics", Nucl. Instrum. Meth. A 434 (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.
- [90] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: doi:10.1140/epjc/s10052-013-2501-z].
- [91] S. Borowka et al., "Full top quark mass dependence in Higgs boson pair production at NLO", JHEP 10 (2016) 107, doi:10.1007/JHEP10 (2016) 107, arXiv:1608.04798.
- [92] D. de Florian, I. Fabre, and J. Mazzitelli, "Higgs boson pair production at NNLO in QCD including dimension 6 operators", *JHEP* 10 (2017) 215, doi:10.1007/JHEP10(2017)215, arXiv:1704.05700.
- [93] J. Baglio et al., "gg → HH: Combined uncertainties", Phys. Rev. D 103 (2021) 056002, doi:10.1103/PhysRevD.103.056002, arXiv:2008.11626.
- [94] J. Baglio et al., "The measurement of the Higgs self-coupling at the LHC: theoretical status", JHEP 04 (2013) 151, doi:10.1007/JHEP04 (2013) 151, arXiv:1212.5581.
- [95] R. Frederix et al., "Higgs pair production at the LHC with NLO and parton-shower effects", *Phys. Lett. B* 732 (2014) 142, doi:10.1016/j.physletb.2014.03.026, arXiv:1401.7340.
- [96] L.-S. Ling et al., "NNLO QCD corrections to Higgs pair production via vector boson fusion at hadron colliders", *Phys. Rev. D* 89 (2014) 073001, doi:10.1103/PhysRevD.89.073001, arXiv:1401.7754.
- [97] F. A. Dreyer and A. Karlberg, "Fully differential Vector-Boson Fusion Higgs Pair Production at Next-to-Next-to-Leading Order", *Phys. Rev. D* 99 (2019) 074028, doi:10.1103/PhysRevD.99.074028, arXiv:1811.07918.
- [98] F. A. Dreyer and A. Karlberg, "Vector-Boson Fusion Higgs Pair Production at N³LO", *Phys. Rev. D* 98 (2018) 114016, doi:10.1103/PhysRevD.98.114016, arXiv:1811.07906.