Search for resonant Higgs boson pair production in the $b\bar{b}l\nu l\nu$ final state at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

A search for resonant Higgs bosons pair production decaying into $b\bar{b}WW$ with subsequent $WW$ decays into two leptons and two neutrinos, $X \rightarrow HH \rightarrow b\bar{b}WW \rightarrow b\bar{b}l\nu l\nu$, is presented. The analysis is based on a sample of proton-proton collisions at $\sqrt{s} = 13$ TeV at the LHC corresponding to an integrated luminosity of 2.3 fb$^{-1}$. Masses are considered in the range between $m_X = 260$ GeV and 900 GeV. The search focuses on the invariant mass distribution of the $b$-jet pair, searching for a resonant-like excess compatible with the H boson mass in combination with a boosted decision tree discriminant based on kinematic information. Data and predictions from the standard model are in agreement within systematic uncertainties. For mass hypotheses from $m_X = 500$ GeV to $m_X = 900$ GeV, the data are observed (expected) to exclude a production cross-section times branching ratio of a spin-0 particle from 174 to 101 (135 to 75.8) fb.
1 Introduction

The Higgs mechanism is an essential element of the standard model (SM) of particles and their interactions explaining the origin of mass and playing a key role in electroweak symmetry breaking. However, the discovery of a Higgs boson (H) with a mass of around 125 GeV by the ATLAS and CMS experiments [1, 2] opens the question of whether it is the SM Higgs boson or one of typically many Higgs bosons predicted by extensions beyond the standard model.

Extensions of the scalar sector of the SM predict the existence of additional Higgs bosons. A simple scenario is the two Higgs doublet model (2HDM), where a second doublet of complex scalar fields is added to the minimal SM scalar sector lagrangian. The generic 2HDM potential [3] has a large number of degrees of freedom. Assuming the preservation of the electromagnetic gauge symmetry, the CP invariance on the bosonic sector of the theory, the choice of the custodial phase and the suppression of the tree-level flavour-changing neutral currents, the number of free parameters is reduced to six. These free parameters are the mass of the Higgs boson H ($m_H$), the mass of the pseudoscalar A ($m_A$), the masses of the charged Higgs bosons ($m_{H^\pm}$), the mass of the CP-even state X ($m_X$), the ratio of the vacuum expectation values $\tan \beta$ and the mixing angle $\alpha$ between the two CP-even eigenstates. In case the new CP-even state is massive enough ($m_X > 2m_H$) it can decay to a pair of Higgs bosons. Models inspired by warped extra dimensions [4] predict the existence of new heavy particles X (with $m_X > 2m_H$), that can decay to a pair of Higgs bosons. Examples of such particles are radion (spin-0) [5–8] or the first Kaluza-Klein excitation of the graviton (spin-2) [9, 10].

In this document we report on a search for resonant Higgs pair production, $X \rightarrow HH$, where one of the H decays as $H \rightarrow b\overline{b}$, and the other as $H \rightarrow WW \rightarrow l\nu l\nu$ (where l is either an electron or a muon) using LHC proton-proton collision data at $\sqrt{s} = 13$ TeV. The analysis focuses on the invariant mass distribution of the b-jet pair, searching for a resonant-like excess compatible with the H boson mass in combination with a boosted decision tree discriminant based on kinematic information. The dominant background is $t\bar{t}$ production with smaller contributions from Drell-Yan and single top processes. Figure 1 shows the schematic diagram of the resonant HH production with the subsequent Higgs and W boson decays, and the branching ratios of the SM Higgs boson.

This study is performed for the first time using LHC data.

2 CMS detector and simulation

The central feature of the CMS apparatus is a superconducting solenoid, 13 m in length and 6 m in diameter, which provides an axial magnetic field of 3.8 T. The bore of the solenoid is outfitted with various particle detection systems. Charged particle trajectories are measured by silicon pixel and strip trackers, covering $0 < \phi < 2\pi$ in azimuth and $|\eta| < 2.5$, where the pseudo-rapidity $\eta$ is defined as $\eta = -\log \tan(\theta/2)$, with $\theta$ being the polar angle of the trajectory of the particle with respect to the beam direction. A crystal electromagnetic calorimeter (ECAL) and a brass/scintillator hadronic calorimeter (HCAL) surround the tracking volume; in this analysis the calorimetry provides high resolution energy and direction measurements of electrons and hadronic jets. A preshower detector consisting of two planes of silicon sensors interleaved with lead is located in front of the ECAL at $|\eta| < 1.479$. Muons are measured in gas detectors embedded in the steel return yoke outside the solenoid. The detector is nearly hermetic, allowing for energy balance measurements in the plane transverse to the beam directions. A two-tier trigger system selects the most interesting pp collision events for use in physics analysis. A more detailed description of the CMS detector can be found elsewhere [11].
Event selection and background predictions

We collect events with a set of dilepton triggers, which require transverse momentum $p_T > 17$ GeV for the first lepton and $p_T > 12$ GeV (8 GeV) for the second electron (muon). We select events with two oppositely charged leptons ($e^+e^-, \mu^+\mu^-, e^+\mu^+$), in which the electrons (muons) are required to have a $p_T$ greater than 20 GeV and 15 GeV (10 GeV), for the higher and lower $p_T$ lepton, respectively. Muons (electrons) in the pseudo-rapidity range $|\eta| < 2.4$ ($|\eta| < 2.5$) are considered. A dilepton mass requirement of $m_{ll} > 12$ GeV is applied in the
Electrons, reconstructed by associating tracks with ECAL clusters, are identified using information on the cluster shape in the ECAL, track quality, and the matching between the track and the ECAL cluster. Additionally, electrons from photon conversions are rejected. Muons are reconstructed from tracks found in the muon system, associated to tracks in the silicon tracking detectors. They are identified based on the quality of the track fit and the number of associated hits in the different tracking detectors. For both lepton flavours, the impact parameter with respect to the reconstructed vertex with the largest $p_T^2$ sum of associated tracks (primary vertex) has to be below 0.5 mm in the transverse plane and 1 mm along the beam direction. The lepton isolation, defined as the scalar $p_T$ sum of all particle candidates, excluding the lepton, in a cone around the lepton, divided by the lepton $p_T$, is required to be $< 0.04$ ($< 0.15$) for electrons (muons). The lepton objects in the simulations are corrected for the residual differences in between data and simulation.

Jets are reconstructed using a particle flow (PF) technique [20]. Candidates are clustered to form jets using the anti-$k_T$ clustering algorithm [21], implemented in the FASTJET package [22], with a distance parameter of 0.4. Jet energies are corrected for residual non uniformity and non linearity of the detector response using corrections found with collision data [23]. Jets are required to have $p_T > 20$ GeV, $|\eta| < 2.4$, and be separated from identified leptons by a distance of $\sqrt{\Delta \phi^2 + \Delta \eta^2} = \Delta R > 0.3$. The magnitude of the negative vector sum of all PF candidates is referred to as $E_T^{\text{miss}}$. Corrections to the jet energy are propagated to the $E_T^{\text{miss}}$.

To identify jets originating from b quarks, the combined secondary vertex algorithm is used. Jets are considered as b-tagged if they pass the medium working point of the algorithm, which provides around 70% efficiency with a mistag rate less than 1% [24]. Correction factors are applied to the selected jets in simulation to account for the different response of the combined secondary vertex algorithm in between data and simulation.

Among all possible dijet combinations fulfilling the previous criteria we select the two jets with the highest sum of the combined secondary vertex output.

After the final object selection consisting of two opposite sign leptons and two b-tagged jets, four kinematic distributions ($m_{ll}$, $\Delta R_{ll}$, $\Delta R_{jj}$, $\Delta \phi_{ll,jj}$) are selected due to their separation power between signal and background. In all of them, background dominated regions in which the presence of signal is at the percent level, are present. By applying the following set of cuts: $m_Z - m_{ll} > 15$ GeV, $\Delta R_{ll} < 2.2$, $\Delta R_{jj} < 3.1$, and $\Delta \phi_{ll,jj} > 1.5$, 95% of the background is removed while 86-97% of the signal is kept, for signal masses ranging from 400 to 900 GeV. These cuts were applied before further optimization by the boosted decision tree discriminant described in the next section, as they remove significant background with negligible effect on signal. Figure 2 shows the $p_T^{ll}$ and $p_T^{jj}$ distributions for data and simulated events after requiring all the cuts described in this section.

## 4 Signal extraction

Boosted decision trees (BDTs) discriminants are used to further improve the signal-to-background separation. In a phase space dominated by $t\bar{t}$ production, the BDTs exploit information related to object kinematics. The set of variables provided as input to the BDTs is: $m_{ll}$, $\Delta R_{ll}$, $\Delta R_{jj}$, $\Delta \phi_{ll,jj}$, $p_T^{ll}$, $p_T^{jj}$, min($\Delta R_{ll}$), and $M_T$, defined as $M_T = \sqrt{2p_T^{ll}E_T^{\text{miss}}(1 - \cos(\Delta \phi_{ll,E_T^{\text{miss}}}))}$. This set of BDT inputs is similar to the one used in previous CMS studies targeting High Luminosity LHC scenarios [25].
Figure 2: The $p_T^{ll}$ (left) and $p_T^{jj}$ (right) distributions for data and simulated events after requiring two leptons, two b-tagged jets, and $m_Z - m_{ll} > 15$ GeV. The simulated samples are normalized using a maximum likelihood fit, as described in details section 6.

All the lepton channels ($e^+e^-, \mu^+\mu^-, e^\pm\mu^\mp$) are merged during the BDT training. The $t\bar{t}$, Drell-Yan and single top production, dominant SM processes in the selection, are considered as background, and the spin-0 radion as signal. The BDT training and validation is performed in an independent $t\bar{t}$ sample.

While the BDTs are trained using spin-0 signals, their performances are compatible with those using spin-2 signals, allowing one single training for the two different spin hypotheses.

The mass range of the search is defined from $m_X = 260$ GeV to $m_X = 900$ GeV. We train two different BDTs within the mass range for signals with masses $m_X = 400$ GeV and $m_X = 650$ GeV. The BDT trained at $m_X = 400$ GeV is applied to signals with $m_X \leq 450$ GeV, and the BDT trained at $m_X = 650$ GeV is applied to signals with $m_X \geq 450$ GeV. In this way we cover several signal kinematics in the range of interest. While the $m_X = 650$ GeV training could be applied in the whole mass range with some loss in the region between 260 and 450 GeV, the performance of the $m_X = 400$ GeV training is specific to its range of application. The two BDT discriminants are shown in Figure 3.

For each of the BDT outputs we define two regions (low-BDT-scores region and high-BDT-scores region) chosen to maximize the expected sensitivity in the high-BDT-scores region. The boundary on the BDT output defining the regions is 0.1.

The final categories are defined from the two BDT regions in which we apply a cut on the $m_{jj}$ distribution around the Higgs boson mass. The cut on the $m_{jj}$ mass is chosen to maximize the expected sensitivity based on a cut-and-count method. We define the $m_{jj}$ peak region ($m_{jj}$-P) as $95$ GeV < $m_{jj}$ < $135$ GeV. The region outside this window ($m_{jj}$ < $95$ GeV & $m_{jj}$ > $135$ GeV) is called $m_{jj}$ side-band region ($m_{jj}$-SB).

In summary, for each of the two BDTs, four final categories are defined as: high-BDT-scores & $m_{jj}$-P, high-BDT-scores & $m_{jj}$-SB, low-BDT-scores & $m_{jj}$-P, and low-BDT-scores & $m_{jj}$-SB.
5 Systematic uncertainties

The different sources of systematic uncertainties affect both normalization and shape of the background and signal expectations. We consider as experimental sources of systematic uncertainties the lepton selection (identification and isolation), trigger efficiency, jet energy scale, jet energy resolution, jet b-tagging, pileup, parton distributions, renormalization scale, factorization scale and LHC luminosity. In addition, we consider a systematic effect on the \( t\bar{t} \) cross section. The effects on the total yields in signal regions are summarized in Table 1. Shape variations as consequence of the systematic uncertainties are considered when affecting the \( m_{jj} \) distribution.

The impact of the jet energy scale uncertainty is evaluated by shifting the jet energy correction factors for each jet up and down by one standard deviation, in both signal and background predictions. The uncertainties associated with the corrections for the b tagging efficiencies for light-flavor and heavy-flavor jets are evaluated by varying these corrections by one standard deviation. Lepton identification and trigger scale factors are computed with the “tag-and-probe” technique [26] and are applied to the simulated samples to match the performance observed in data.

Theoretical uncertainties on the cross section used to predict the \( t\bar{t} \) background are propagated to the yield estimates. The magnitude of the uncertainties related to the parton distribution functions for each simulated background process, is obtained using the replicas of the NNPDF 3.0 set. The uncertainties related to renormalization and factorization scale are extracted by multiplying or dividing them by two. In addition, we consider modeling uncertainties on the \( t\bar{t} \), Drell-Yan and single-top processes.

Systematic uncertainties due to the finite nature of simulation samples are taken into account.
Table 1: Summary of the systematic uncertainties and their impact range on total yields in all final categories and BDT trainings, for signal $m_X = 400$ GeV, signal $m_X = 650$ GeV, and background.

<table>
<thead>
<tr>
<th>Source</th>
<th>Sig. ($m_X = 400$ GeV)</th>
<th>Sig. ($m_X = 650$ GeV)</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger efficiency</td>
<td>5.1 - 6.0%</td>
<td>6.7 - 7.4%</td>
<td>4.5 - 5.3%</td>
</tr>
<tr>
<td>Jet b-tagging</td>
<td>4.9 - 6.5%</td>
<td>5.7 - 7.3%</td>
<td>5.1 - 6.0%</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>1.6 - 3.0%</td>
<td>0.6 - 3.9%</td>
<td>1.0 - 3.6%</td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td>0.5 - 4.1%</td>
<td>1.8 - 3.5%</td>
<td>0.1 - 2.4%</td>
</tr>
<tr>
<td>Electon ID &amp; ISO</td>
<td>1.3 - 1.6%</td>
<td>1.3 - 1.7%</td>
<td>1.4 - 1.5%</td>
</tr>
<tr>
<td>Muon ID &amp; ISO</td>
<td>0.9 - 1.4%</td>
<td>1.0 - 1.1%</td>
<td>1.2 - 1.5%</td>
</tr>
<tr>
<td>Pileup</td>
<td>0.4 - 1.8%</td>
<td>0.1 - 0.6%</td>
<td>0.5 - 2.2%</td>
</tr>
<tr>
<td>Parton distributions</td>
<td>0.4 - 0.5%</td>
<td>0.2 - 0.5%</td>
<td>0.5 - 0.6%</td>
</tr>
<tr>
<td>QCD scale</td>
<td>0.3 - 0.4%</td>
<td>0.2 - 0.4%</td>
<td>0.8 - 2.4%</td>
</tr>
<tr>
<td>Luminosity</td>
<td></td>
<td></td>
<td>2.7%</td>
</tr>
<tr>
<td>Signal MC stat.</td>
<td>1.4 - 2.4%</td>
<td>0.9 - 3.2%</td>
<td>-</td>
</tr>
</tbody>
</table>

Affecting only $t \bar{t}$ (87.0 - 95.3% of the total bkg.):

<table>
<thead>
<tr>
<th>Source</th>
<th></th>
<th></th>
<th>6.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t \bar{t}$ cross section</td>
<td>-</td>
<td>-</td>
<td>6.5%</td>
</tr>
<tr>
<td>$t \bar{t}$ modeling</td>
<td>-</td>
<td>-</td>
<td>10%</td>
</tr>
<tr>
<td>$t \bar{t}$ MC stat.</td>
<td>-</td>
<td>-</td>
<td>0.6 - 2.3%</td>
</tr>
</tbody>
</table>

Affecting only Drell-Yan (1.8 - 7.1% of the total bkg.):

<table>
<thead>
<tr>
<th>Source</th>
<th></th>
<th></th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drell-Yan modeling</td>
<td>-</td>
<td>-</td>
<td>30%</td>
</tr>
<tr>
<td>Drell-Yan MC stat.</td>
<td>-</td>
<td>-</td>
<td>4.4 - 22.7%</td>
</tr>
</tbody>
</table>

Affecting only single top (2.5 - 4.6% of the total bkg.):

<table>
<thead>
<tr>
<th>Source</th>
<th></th>
<th></th>
<th>20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single top modeling</td>
<td>-</td>
<td>-</td>
<td>20%</td>
</tr>
<tr>
<td>Single top MC stat.</td>
<td>-</td>
<td>-</td>
<td>6.6 - 24.4%</td>
</tr>
</tbody>
</table>

Affecting only other backgrounds (0.4 - 1.4% of the total bkg.):

<table>
<thead>
<tr>
<th>Source</th>
<th></th>
<th></th>
<th>3.5 - 24.6%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other backgrounds MC stat.</td>
<td>-</td>
<td>-</td>
<td>3.5 - 24.6%</td>
</tr>
</tbody>
</table>
6 Results

Selected events are classified into 4 categories, as described in section 4. Two of these categories (low-BDT-scores & \( m_{jj} \)-P, and low-BDT-scores & \( m_{jj} \)-SB) are background-dominated, and help constraining the background normalizations.

Signal efficiencies as a function of the X mass hypothesis in the four final regions defined in the analysis, are shown in figure 4.

Figure 4: Signal efficiencies as a function of the X mass hypothesis in the four final regions defined in the analysis, for spin-0 (left) and spin-2 (right) hypotheses. The dashed line represents the switch in between the two analysis strategies. The markers correspond to efficiencies evaluated on fully-simulated signal samples, while the efficiencies have been interpolated in between. The cyan corresponds to the high-BDT-scores & \( m_{jj} \)-P, the blue corresponds to the high-BDT-scores & \( m_{jj} \)-SB, the red corresponds to the low-BDT-scores & \( m_{jj} \)-P, and the yellow corresponds to the low-BDT-scores & \( m_{jj} \)-SB. The black line, being the sum of all (mutually exclusive) regions, correspond to the full signal acceptance after preselection and as such is continuous.

We first perform a maximum likelihood fit in order to extract best-fit signal cross-section, where all the nuisances parameters described in section 5 are free to float. These cross-sections are all compatible with zero for all the signal mass hypotheses considered in this analysis.

Another maximum likelihood fit is performed to extract post-fit distributions of all distributions shown in this note, along with post-fit uncertainties (figures 2 and 3). The four distributions are fitted simultaneously, the signal strength fixed to zero and all nuisances parameters floating. Post-fit uncertainties are computed from the uncertainties of all the nuisances parameters, using the correlation between all the parameters.

Since no sign of new physics is found, we set limits on the production of resonant Higgs pair production cross-section. Upper limits at 95% confidence level (CL) are computed as function of the X mass hypothesis, taking the data, background, and signal in the four final regions defined in the analysis, using the asymptotic CL \(_s\) method [27]. Limits on the resonant Higgs pair production cross section times branching fraction, \( HH \rightarrow b\bar{b}VV \rightarrow b\bar{b}\ell\ell\nu\) are shown in Figure 5.
Figure 5: Expected and observed 95% CL upper limits on the resonant Higgs pair production cross section times branching fraction for HH → b¯bνlνl, computed using the asymptotic CLs method, for spin-0 (left) and spin-2 (right) hypotheses. The markers correspond to the limit evaluated on the fully-simulated mass points, while in between the signal prediction was interpolated. The change in trend of the observed limits at 450 GeV, represented by the dashed line, corresponds to transition point of the analysis between the two BDTs (see section 4). Theoretical cross-sections corresponding to 2HDM of Type I and Type II are overlaid on the spin-0 plot. Theoretical cross-sections corresponding to WED models for radions and RS1 KK gravitons are also overlaid on the spin-0 and spin-2 plots respectively.

As already demonstrated by the best-fit cross-section, no excess is visible in the whole mass range of the analysis, the observed limits being compatible with the expected ones within 2 standard deviations. The change of trend in the observed limits at 450 GeV corresponds to the transition point of the analysis between the two BDTs, one optimized for low mass resonances, one for high mass resonances.

7 Summary

We have presented a search for resonant Higgs pair production, X → HH, where one of the H decays as H → b¯b, and the other as H → WW → lνlν using LHC proton-proton collision data at √s = 13 TeV, corresponding to an integrated luminosity of 2.3 fb⁻¹. Masses are considered in the range between mX = 260 GeV and 900 GeV.

Data and predictions from the standard model are in agreement within systematic uncertainties. For mass hypotheses from mX = 500 GeV to mX = 900 GeV, the data are observed (expected) to exclude a production cross-section times branching ratio from 174 to 101 (135 to 75.8) fb.

References


