

Searches for the Higgs-like boson decaying into bottom quarks in the WH channel

Christian Böser¹ on behalf of the CMS collaboration

¹ *Institut für Experimentelle Kernphysik, Karlsruhe Institute of Technology (KIT), Germany*

Abstract. The most important finding of the LHC so far was the discovery of the Higgs-like boson at 125 GeV in 2012. We present the most recent results of the search for the Higgs-like boson decaying into bottom quarks, when produced in association with a W boson. Only events where the leptonically decaying W boson and the Higgs boson possess large transverse momenta are selected. The full proton-proton collision data recorded by the CMS detector in 2011 and 2012 at 7 and 8 TeV respectively, corresponding to an integrated luminosity of 24/fb, is used for the search.

1 Introduction

In 2012 the ATLAS and CMS collaborations reported the discovery of a new boson with a mass of about 125 GeV [1, 2]. The discovery was mainly driven by the decay channels into either γ , W or Z boson pairs. So far, all properties of the new particle are in good agreement with the predicted ones for the standard model Higgs boson. If the new particle is the Higgs boson, it decays predominantly into a bottom quark pair. The discovery of this decay channel could establish the interaction of the Higgs boson with fermions and offers a direct test for its coupling to quarks. However, due to the huge multijet background a Higgs boson signal can hardly be extracted in the di-b-quark final state. Requiring also a specific production mechanism for the Higgs boson improves the ratio of expected signal over background events significantly. The presented analysis focusses on the VH channel, where the Higgs boson is produced in association with a leptonically decaying W or Z boson. Triggering on the additional leptons (or missing energy) in the final state suppresses a significant amount of background, making the search possible.

2 Analysis strategy

In this note the search for the process $H \rightarrow b\bar{b}$ is presented for the three channels $W(\ell\nu)H$, $W(\mu\nu)H$, $W(\tau\nu)H$ using proton-proton collision data recorded by the CMS detector [3]. Figure 1 shows the representative Feynman diagram of the processes. For the $W(\mu\nu)H$ and $W(\ell\nu)H$ channels data with a center-of-mass energy of 7 TeV and 8 TeV is analyzed corresponding to an integrated luminosity of 5.0 fb^{-1} and 19.0 fb^{-1} , respectively. For the $W(\tau\nu)H$ final state, only 8 TeV data with an integrated luminosity of 18.3 fb^{-1} is included in the search. In the following sec-

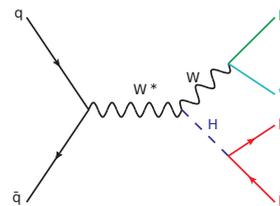


Figure 1. Tree-level Feynman diagram for the process $pp \rightarrow WH \rightarrow \ell\nu b\bar{b}$.

tions selected features of the analysis are highlighted. For a much more detailed description see [4].

2.1 Event selection

For every event the dijet system with the highest transverse momentum is chosen to form the Higgs boson candidate. One major challenge of this analysis is to handle the copious backgrounds, namely W boson production in association with additional jets and top quark pair, single top, diboson and multijet production. For a proper reduction of background event rates in the signal region several selection criteria are applied. We are selecting only events where the Higgs and the W boson are traveling back-to-back in the transverse plane and possess high transverse momenta ($p_T(H/W) > 100 \text{ GeV}$). In addition, an isolated lepton is required. In order to assure that the two jets from the Higgs boson candidate (Higgs jets) stem most likely from b quarks the multivariate b-tagging algorithm “Combined Secondary Vertex” (CSV) is used. For the electron and muon channels three different analysis bins corresponding to different ranges of the W boson’s transverse momentum (low, intermediate and high $p_T(W)$ bin) are chosen. Due to the lower amount of available data events after the selection in the $W(\tau\nu)H$ channel only one category is used. The detailed selection criteria for the different bins can be found in Table 1.

Variable	W($\ell\nu$)H	W($\tau\nu$)H
$p_T(j_1)$	> 30	> 30
$p_T(j_2)$	> 30	> 30
$p_T(jj)$	> 100	> 120
$m(jj)$	< 250	< 250
$p_T(V)$	100 – 130 (130 – 180, > 180)	> 120
CSV _{max}	> 0.40	> 0.40
CSV _{min}	> 0.40	> 0.40
N_{al}	= 0	= 0
E_T^{miss}	> 45	> 80
$\Delta\phi(E_T^{miss}, \ell)$	< $\pi/2$	–
Tightened Lepton Iso.	< 0.075 (–, –)	–
$p_T(\tau)$	–	> 40
$p_T(track)$	–	> 20

Table 1. Selection criteria for each channel. Entries marked with “–” indicate that the variable is not used in the given channel. For W($e\nu$)H and W($\mu\nu$)H entries in parenthesis indicate the selection for the intermediate- and high- $p_T(V)$ regions if different.

Process	Low p_T	Intermediate p_T	High p_T
W0b	$1.03 \pm 0.01 \pm 0.05$	$1.02 \pm 0.01 \pm 0.07$	$1.04 \pm 0.01 \pm 0.07$
W1b	$2.22 \pm 0.25 \pm 0.20$	$2.90 \pm 0.26 \pm 0.20$	$2.46 \pm 0.33 \pm 0.22$
W2b	$1.58 \pm 0.26 \pm 0.24$	$1.30 \pm 0.23 \pm 0.14$	$0.77 \pm 0.25 \pm 0.08$
$t\bar{t}$	$1.03 \pm 0.01 \pm 0.04$	$1.02 \pm 0.01 \pm 0.15$	$1.00 \pm 0.01 \pm 0.11$

Table 2. Data/MC scale factors in 8 TeV data in the different analysis bins including the statistical uncertainty from the fit and a systematic uncertainty accounting for possible data/MC shape differences in the discriminating variables.

2.2 Background control regions

By introducing slightly different cuts, control regions (CR) enriched with the main background processes are defined orthogonally to the previously denoted signal region. A $t\bar{t}$ enhanced control region is achieved by requiring additional jets in the event. Inverting the b-tagging requirements for the Higgs jets yields the W+light enhanced CR. Further, the W+heavy control region is obtained by applying a mass veto in the range of the Higgs boson mass. The three control regions are employed to correct simulation yield estimates for several of the main backgrounds, namely top quark pair production and W boson production in association with zero, one and two b jets. This is achieved by a simultaneous fit to data in discriminating variables separately for each analysis bin. For $t\bar{t}$ and the W+light control regions the invariant dijet mass, and for the W+heavy CR the CSV output of the Higgs jet with the second highest transverse momentum are used. The resulting scale factors can be found in Table 2. For W+1b the fitted estimate is about two times larger than the prediction from simulation. This discrepancy is also found by other studies and arises due to mismodeling of the generator parton shower when a gluon splits to a bottom quark pair. All other scale factors are within their uncertainties well compatible with unity. The control regions are also used to validate variables important in the analysis. Overall a good agreement between data and simulation is found.

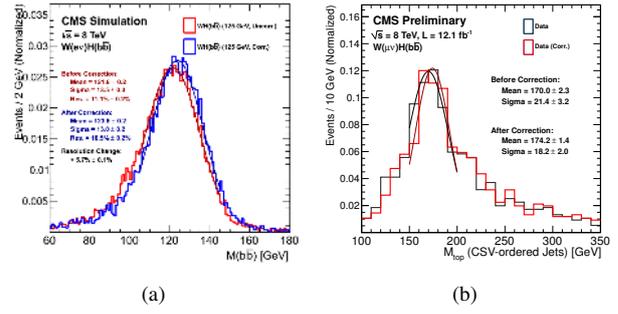


Figure 2. The dijet mass distribution before and after applying the regression in signal MC (a) and a validation of the regression performance using the reconstructed top quark mass (b).

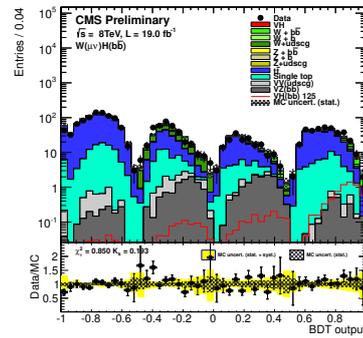


Figure 3. Multi-BDT outputs for the muon channel. Post-fit normalization factors are applied.

2.3 Jet energy regression

Reconstructed jets stemming from b quarks usually have a worse energy resolution compared to light flavor and gluon induced jets. This is due to the fact, that in 20% of all cases neutrinos are present in the B hadron decay leading to missing energy within the jet. A regression technique is performed in order to compute correction factors for individual b jets. We train a dedicated Boosted Decision Tree (BDT) algorithm on the true transverse momenta of b quarks in simulated signal events using jet properties, b-tag and soft lepton information as input variables. Applying the resulting regression weights not only corrects the transverse momenta and improves the resolution of the b jets, it also increases the resolution of the reconstructed dijet mass as shown in figure 2(a) and hence the search sensitivity.

All input variables of the regression are validated in the control regions. Additionally, the regression performance is confirmed in the invariant mass distribution of hadronically decaying top quarks in a $t\bar{t}$ sample as shown in figure 2(b).

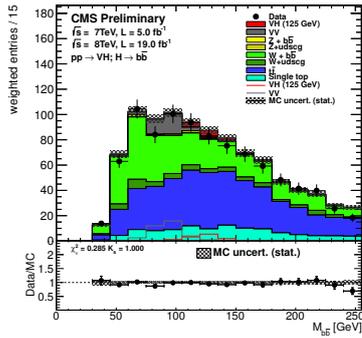


Figure 4. Combined $M(jj)$ distribution for all WH analysis bins.

2.4 Signal extraction

In order to discriminate signal from background events a multivariate analysis is performed. For all channels, boosted decision trees are trained versus all occurring backgrounds with a large set of kinematic variables of the Higgs and W boson, b-tag information of the two Higgs jets and features of the additional jets in the events as inputs. For the $W(\tau\nu)H$ channel the resulting BDT distribution is used as final discriminant. For the electron and muon channels additional BDTs are trained to separate signal events individually from $t\bar{t}$, W +light and diboson events. By successively applying cuts on these outputs one can categorize the events into four regions. For all events the nominal BDT output is transferred to the corresponding category leading to final discriminants. In Figure 3 the final discriminant is shown for the high p_T bin in the $W(\mu\nu)H$ channel and one can clearly indicate the four different categories, where the three left ones are seen as $t\bar{t}$, W +jets and diboson like regions and the rightmost bin is signal enhanced. This categorization of events helps to increase the search sensitivity.

In addition to the multivariate approach, a simpler cut-based analysis is performed using the dijet mass spectrum. This analysis has a lower search sensitivity compared to the BDT analysis and serves as a cross check. The combined invariant mass distribution for the WH channel is shown in Figure 4. In the following only results from the BDT analyses are presented.

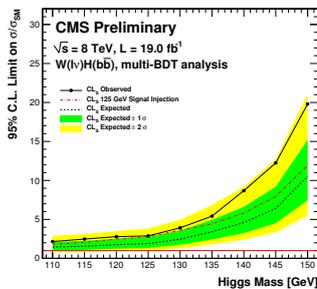


Figure 5. Expected and observed 95% upper limits on the ratio of $W(\ell\nu)H(bb)$ production for the BDT analysis.

3 Results

The final results are calculated using a simultaneous fit to the final discriminants in each channel. In Figure 5 the resulting upper limits for several mass hypotheses are presented for the combined $W(e\nu)H$, $W(\mu\nu)H$ and $W(\tau\nu)H$ channels. The plot shows an excess between one and two standard deviations over the whole mass range. For a Higgs boson mass of 125 GeV we report a local significance in the WH channel of 1.0σ , where 1.0σ was expected if a Higgs boson was present. The fitted signal strength is $\mu = 1.1$.

In Figure 6 the signal strength for the $W(\ell\nu)H$ process is shown together with the $Z(\nu\nu)H$ and $Z(\ell\ell)H$ channels [4]. We find a good compatibility between the signal strengths of all channels. For the combined VH channel a local significance for a 125 GeV Higgs boson of 2.1σ is found with a fitted signal strength of $\mu = 1.0 \pm 0.5$ [4].

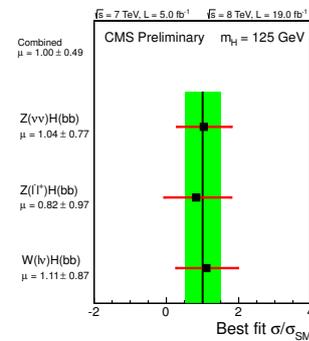


Figure 6. Most likely value of the production cross section for a 125 GeV Higgs boson, relative to the standard model cross section, for each channel.

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**, 1-29 (2012).
- [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**, 30-61 (2012).
- [3] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3**, S08004 (2008).
- [4] CMS Collaboration, “Search for the standard model Higgs boson produced in association with W or Z bosons, and decaying into bottom quarks (LHCP 2013)”, CMS Physics Analysis Summary **HIG-13-012** (2013).