PROCEEDINGS OF SCIENCE

# Search for the standard model Higgs boson produced in vector boson fusion and decaying to bottom quarks using the Run1 and 2015 Run2 data samples with the CMS experiment

**Nadezda Chernyavskaya** *[*] [†]*

*Swiss Federal Institute of Technology in Zurich (ETHZ), Zurich, Switzerland*
*E-mail:* nadezda.chernyavskaya@cern.ch

A search for the standard model Higgs boson is presented in the Vector Boson Fusion production channel with decay to bottom quarks with the CMS experiment at the CERN LHC. A data sample comprising 2.3 $\mathrm{fb}^{-1}$ of proton-proton collision at $\sqrt{s} = 13$ TeV collected during the 2015 running period has been analyzed. Production upper limits at 95% Confidence Level are derived for a Higgs boson mass of 125 GeV, as well as the fitted signal strength relative to the expectation for the standard model Higgs boson. Results are also combined with the ones obtained with Run1 $\sqrt{s} = 8$ TeV data collected in 2012.

*Fourth Annual Large Hadron Collider Physics*
*13-18 June 2016*
*Lund, Sweden*

[*] Speaker.
[†] on behalf of the CMS Collaboration





PoS(LHCP2016)186

## 1. Introduction

In the Standard Model (SM)[1] the electroweak symmetry breaking is explained by the Brout-Englert-Higgs[2, 3] mechanism responsible for the electroweak gauge bosons to acquire mass. The mechanism predicts the existence of a Higgs scalar boson, which observation at the mass of 125 GeV was announced by the CMS[4] and ATLAS[5] experiments at the Large Hadron Collider (LHC) in 2012. After the Higgs boson discovery, the main efforts are aimed to measure precisely its properties and couplings and to compare them with the ones predicted by SM. A SM Higgs boson with a mass of 125 GeV has a dominant decay branching ratio of 58% to b-quarks. At LHC a search for the Higgs boson in the most abundant production channel Gluon Fusion (GF) is not pursuable in proton-proton collisions due to overwhelming QCD background. Thus Vector Boson fusion (VBF), the production channel with the second largest cross section and a distinct topology, is the one exploited in the analysis presented here[6, 7].
In the VBF process a quark of each one of the colliding protons radiates a W or Z boson that subsequently interacts or fuses. The main feature of the VBF process $qqH \to qqb\bar{b}$ is the presence of four energetic jets in the final state. Two heavy flavor jets are expected from the decay of the Higgs boson to bottom quarks pair in the central region of the detector. Two additional light flavor jets originate from two valence quarks from each of the colliding protons and are scattered away from the beam line in the VBF process in forward and backward direction. The dominant background is QCD production of multijets events. Other backgrounds arise from hadronic decays of Z or W bosons produced in association with additional jets, hadronic decays of top quark pairs, and hadronic decays of singly produced top quarks. The contribution of the Higgs boson in GF processes with at least two associated jets is included in the expected signal yield.

## 2. Trigger

The data used for this analysis were collected using two trigger strategies that were specifically designed for VBF $qqH \to qqb\bar{b}$ search. The CMS trigger system has the first level trigger(L1) and the high-level trigger(HLT). Both paths use common L1 and initial HLT events selection. The total collected integrated luminosity used in this analysis was 2.32 $\mathrm{fb}^{-1}$ for both paths collected in 2015 at $\sqrt{s}$ = 13 TeV.

- **Common part**. The L1 path requires the presence of at least three jets with $p_T$ above decreasing thresholds $p_T^1$ = 84 GeV, $p_T^2$ = 68 GeV, $p_T^3$ = 48 GeV. Only one of the two leading jets can be in the forward region with $2.6 < |\eta| \leq 5.2$, the other two jets are required to be central ($|\eta| \leq 2.6$). The common HLT event selection requires the presence of at least four jets with $p_T$ >92, 76, 64, 15 GeV. At least one of the event jets must fulfil minimum HLT b-tagging requirements[8, 9].
- **SingleB**. This path selects events that pass the common HLT selection. The most b-tagged jet is labelled as a b-jet and, among the remaining three, the two jets with the largest pseudorapidity opening are labelled as the two q-jets, and the second b-jet is identified as the remaining jet. The flowing final cuts are applied : $\Delta\eta_{qq} > 4.1$, $m_{qq} > 460$ GeV, $\Delta\phi_{bb} < 1.6$.
- **DoubleB**. This path selects events that pass the common HLT selection and have a second b-tag, among the six $p_T$-leading jets. The two most b-tagged jets are labelled as b-jets and



PoS(LHCP2016)186

the remaining two $p_T$-leading ones are labelled as the two q-jets. The flowing final cuts are applied : $\Delta\eta_{qq}$> 1.2, $m_{qq}$ > 200 GeV.

## 3. Event reconstruction and selection

The offline analysis uses reconstructed charged-particle tracks and candidates of a Particle-Flow (PF) algorithm[10]. Jets are reconstructed by clustering the PF candidates with the anti-kT algorithm with distance parameter 0.4. Jets that are likely to originate from the hadronization of b quarks are identified with the CSV b-tagger[8, 9]. The events used in the offline analysis are required to have at least four reconstructed jets and the most probable b-jet and VBF-tagging q-jet candidates are searched among the seven $p_T$ leading ones. For the SingleB sample, the distinction between the two jet types is done by means of a multivariate discriminant that takes into account the jet b-tag, $\eta$ and $p_T$ values and their rankings. The offline event selection follows the trigger paths requirements. Selected events are divided into two sets: SingleB and DoubleB and the selection requirements are shown in Table 1. After all selection requirements, 1.6% of the simulated VBF signal events end up in the SingleB sample, and 0.6% end up in DoubleB.

| | SingleB | DoubleB |
|---|---|---|
| Trigger | one b-tagged jet | two b-tagged jets |
| jets $p_T$ | $p_{T,0,1,2,3}$ > 92, 76, 64, 30 GeV | |
| jets $\|\eta\|$ | < 4.7 | |
| b-tag | no cut | $CSV_{0,1} > 0.5$ |
| $\Delta\phi_{bb}$ | $< 1.6$ | $< 2.4$ |
| VBF topology | $m_{qq} > 460GeV$, $\|\Delta\eta_{qq}\| > 4.1$ | $m_{qq} > 200GeV$, $\|\Delta\eta_{qq}\| > 1.2$ |
| Veto | None | Events that belong to SingleB |

**Table 1:** Summary of selection requirements for the two data samples.



## 4. Signal properties

- **Jet transverse-momentum regression**. The $b\bar{b}$ mass resolution is improved by the means of regression technique. The calibration is carried out for individual b jets that takes into account the jet composition properties. A multivariate technique, Boosted Decision Tree(BDT), is trained on simulated ditop events and a validation in data is done with samples of $Z \to ll$ events with one or two b-tagged jets. With this technique the dijet invariant mass resolution is improved by 7%.
- **Discrimination between quark-and gluon-originated jets**. To discriminate between quark and gluon jets the minor Root-Mean-Square of the distribution of jet constituents in the $\eta - \phi$ plane[11] is used. VBF signal events, dominated by quark jets, have narrower jets, while the background and GF events are enriched in wider gluon jets.
- **Soft QCD activity**. QCD processes have a strong color flow whilst VBF signal has a suppressed one. To discriminate between the signal and the QCD background soft tracks multiplicity is used as a discriminating variable[6, 7].

## 5. Search for a Higgs boson

In order to separate the overwhelmingly large QCD background from the Higgs boson signal, all the discriminating features are exploited by the means of BDT multivariate discriminant. The input variables are not correlated to the dynamics of the bb-system. According to the BDT outputs, seven categories are defined to maximize the signal sensitivity.

The QCD $m_{b\bar{b}}$ spectrum shape is assumed to be the same in all BDT categories of the same set of events. In reality, a small correction is needed to take into account residual differences. For this transfer functions are exploited (linear function in SingleB and quadratic in DoubleB). The fit model for the Higgs boson signal is given by equation :

$$f_i(m_{bb}) = N_{i,\mathrm{qcd}} \cdot R_i(m_{bb}) \cdot Q(m_{bb};\vec{p}) + N_{i,\mathrm{top}} \cdot T_i(m_{bb};k_{\mathrm{JES}},k_{\mathrm{JER}}) + N_{i,Z} \cdot Z_i(m_{bb};k_{\mathrm{JES}},k_{\mathrm{JER}}) + \mu_H \cdot N_{i,\mathrm{H}} \cdot H_i(m_{bb};k_{\mathrm{JES}},k_{\mathrm{JER}})$$

where the subscript i denotes the category and $\mu_H$, $N_{i,\mathrm{qcd}}$ are free parameters for the signal strength and the QCD event yield. $N_{i,\mathrm{H}}$, $N_{i,\mathrm{Z}}$, $N_{i,\mathrm{top}}$ are the expected yields for the Higgs boson signal, the Z +jets, and the top quark background respectively. The shape of the top quark background $T_i$ is taken from the simulation and described by a broad Gaussian. The Z/W+jets background $Z_i$ and the Higgs boson signal $H_i$ shapes are taken from the simulation and are parametrized as a Crystal ball function of a polynomial. The position and the width of the Gaussian core of the MC templates are allowed to vary within their uncertainties by the factors $k_{JES}$ and $k_{JER}$, respectively. Finally, the QCD shape is described by a polynomial B, common within the categories of each set, and a multiplicative transfer function $K_i$. The parameters are determined by the fit, which is performed simultaneously in all categories in each set. For SingleB, the polynomial is of fifth order, while for DoubleB it is of fourth order.

## 6. Results

The $m_{bb}$ distributions in data are fitted simultaneously in all categories under two hypotheses: background only and background plus a Higgs boson signal. The fit is a binned likelihood fit incorporating the systematic uncertainties as nuisance parameters. The limits on the signal strength are computed with the asymptotic CLs method[12, 13]. For a 125 GeV Higgs boson signal the observed 95% confidence level upper limit is 3.0 times the standard model expectation, compared to 5.0 expected in absence of a signal. The results are combined with the previous CMS search on the Run I proton-proton collision data at $\sqrt{s}$ = 8 TeV[14]. The combination is based on the likelihood ratio test statistics. For $m_H$ = 125 GeV the Run 2 data yields alone a fitted signal strength of $\mu = \sigma/\sigma_{SM} = 3.7^{+2.4}_{-2.5}$ which is compatible with the SM Higgs boson prediction $\mu$ = 1 at the 3% level. The combination of Run 1 and Run 2 results yields an observed (expected) upper limit of 3.4 (2.3) times the SM prediction, and a signal strength $\mu$ = 1.3±1.2 with a observed (expected) significance of 1.2 (0.9) standard deviations. Figure 1 shows the likelihood scans of the expected and observed best-fit signal strength separately with Run 1$\sqrt{s}$ = 8 TeV data, with Run 2 $\sqrt{s}$ = 13 TeV data, and for their combination.



PoS(LHCP2016)186

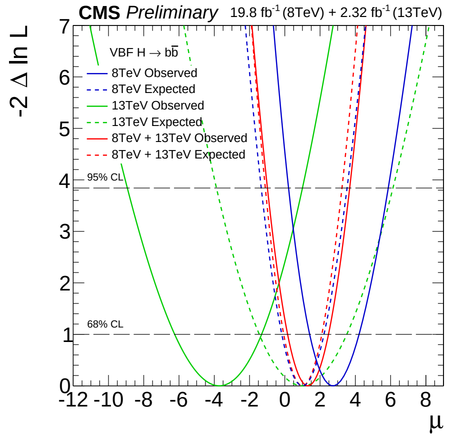


**Figure 1:** Observed and SM-expected likelihood profile of the signal strength $\mu = \sigma/\sigma_{SM}$ with $m_H$ = 125 GeV, using Run 1 $\sqrt{s}$ = 8 TeV data, Run 2 $\sqrt{s}$ = 13 TeV data, and for the combination of 8 TeV and 13 TeV data.

## References

[1] A. Salam, J. C. Ward, *"On A Gauge Theory Of Elementary Interactions,"* Nuovo Cim. **19** (1961) 165.

[2] F. Englert, R. Brout, *"Broken Symmetry and the Mass of Gauge Vector Mesons,"* Phys. Rev. Lett. **13** (1964) 321.

[3] P. W. Higgs, *"Broken Symmetries and the Masses of Gauge Bosons,"* Phys. Rev. Lett. **13** (1964) 508.

[4] S. Chatrchyan *et al.* [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 [arXiv:1207.7235 [hep-ex]].

[5] G. Aad *et al.* [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 [arXiv:1207.7214 [hep-ex]].

[6] CMS Collaboration, *"VBF H to bb using the 2015 data sample,"* CMS-PAS-HIG-16-003.

[7] V. Khachatryan *et al.* [CMS Collaboration], *"Search for the standard model Higgs boson produced through vector boson fusion and decaying to $b\bar{b}$,"* Phys. Rev. D **92** (2015) no.3, 032008 [arXiv:1506.01010 [hep-ex]].

[8] S. Chatrchyan *et al.* [CMS Collaboration], *"Identification of b-quark jets with the CMS experiment,"* JINST **8** (2013) P04013 [arXiv:1211.4462 [hep-ex]].

[9] CMS Collaboration, *"Identification of b quark jets at the CMS Experiment in the LHC Run 2,"* CMS-PAS-BTV-15-001.

[10] CMS Collaboration, *"Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET,"* CMS-PAS-PFT-09-001.

[11] CMS Collaboration, *"Performance of quark/gluon discrimination in 8 TeV pp data,"* CMS-PAS-JME-13-002.

[12] T. Junk, *"Confidence level computation for combining searches with small statistics,"* Nucl. Instrum. Meth. A **434** (1999) 435

[13] 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 [arXiv:1007.1727 [physics.data-an]].

[14] V. Khachatryan *et al.* [CMS Collaboration], *"Search for the standard model Higgs boson produced through vector boson fusion and decaying to $b\bar{b}$,"* Phys. Rev. D **92** (2015) no.3, 032008 [arXiv:1506.01010 [hep-ex]].