PROCEEDINGS OF SCIENCE

Searching for the standard model Higgs boson produced by vector boson fusion in the fully hadronic four-jet topology with CMS

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 produced by vector boson fusion in the fully hadronic four-jet topology is presented. The analysis is based on 2.3 fb^{-1} of proton-proton collision data at $\sqrt{s}$ = 13 TeV collected by CMS in 2015. Upper limits, at 95% confidence level, on the production cross section times branching fraction of the Higgs boson decaying to bottom quarks, are derived for a Higgs boson mass of 125 GeV. The fitted signal strength relative to the expectation for the standard model Higgs boson is obtained. Results are also combined with the ones obtained with Run1 data at $\sqrt{s}$ = 8 TeV collected in 2012.

The European Physical Society Conference on High Energy Physics
5-12 July 2017
Venice

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

PoS(EPS-HEP2017)688

1. Introduction

In the Standard Model (SM)[1] the electroweak symmetry breaking is explained by the Brout-Englert-Higgs[2, 3, 4] mechanism. This mechanism provides masses to electroweak gauge bosons and predicts the existence of a scalar Higgs boson that was observed at the mass of 125 GeV by the CMS[5, 6] and ATLAS[7] experiments at the Large Hadron Collider (LHC) in 2012.

After the Higgs boson was discovered, the main efforts are aimed at measuring precisely its properties and couplings and comparing them with the SM predictions. The most dominant decay channel, at the observed Higgs mass, is to a pair of bottom quarks with a branching fraction of 58%. At LHC in proton-proton collisions an inclusive search for the Higgs boson in the most abundant channel gluon gluon fusion (ggH) is possible only in extremely boosted regime[8]. The vector boson fusion (VBF) production channel has the second largest cross section and a distinct topology that makes such a search possible. Here a search for the Higgs boson in VBF production channel with the CMS experiment[9] is presented[10, 11].
In the VBF production channel weak bosons radiated by quarks from the colliding protons interact producing the Higgs boson that subsequently decays to a pair of bottom quarks. The prominent feature of the VBF process $qqH \to qqb\bar{b}$ is the presence of two energetic quark jets scattered at a small angle and two heavy-flavour jets in the central region of the detector. The dominant background is QCD production of multijets events. Other background contributions arise from hadronic decays of Z or W bosons produced in association with jets, hadronic decays of top quark pairs and single top quarks. In this analysis the ggH process with two associated jets is included in the expected signal.

2. Trigger

For this VBF $\to qqb\bar{b}$ search data were collected using two specifically designed trigger strategies. The CMS trigger system has the first level trigger(L1) and the high-level trigger(HLT). Both paths use common L1 and initial HLT event selection. The total collected integrated luminosity used in this analysis was 2.3 fb^{-1} for both paths collected in 2015 at $\sqrt{s}$ = 13 TeV.
The L1 path requires the presence of at least three jets with transverse momenta (p$_T$) above decreasing thresholds p$_T^1$ = 84 GeV, p$_T^2$ = 68 GeV, p$_T^3$ = 48 GeV with at most one forward jet with 2.6 $< |\eta| \leq$ 5.2 and two or three central jets with $|\eta| \leq$ 2.6. The common HLT event selection requires the presence of at least four jets with p$_T$ >92, 76, 64, 15 GeV. The HLT b-tagging algorithm[12, 13] is used to classify events in two paths : with one b-tagged jet (SingleB path), or more than one b-tagged jet (DoubleB path). These events are further processed through different selections. Harder cuts on q-jets pseudorapidity separation $\Delta\eta_{qq}$ and invariant mass m$_{qq}$ are applied in the SingleB path.

3. Event reconstruction and selection

The offline analysis uses reconstructed charged-particle tracks and candidates of a Particle-Flow (PF) algorithm[14]. 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[12, 13]. The events 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. In 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, η and p_T values, and their respective rankings among the other jets in the event. The offline event selection follows the trigger paths requirements. Selected events are divided into two sets: SingleB and DoubleB. The selection requirements are shown in Table 1. After all selection requirements the SingleB acceptance is 1.6% and the DoubleB one is 0.6%.

	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	$m_{qq} > 460 GeV$, $\|\Delta\eta_{qq}\| > 4.1$	$m_{qq} > 200 GeV$, $\|\Delta\eta_{qq}\| > 1.2$
Veto	None	Events that belong to SingleB

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

4. Signal properties

Characteristic properties of the VBF $qqH \to qqb\bar{b}$ process can provide a substantial improvement of the overall sensitivity of the search.

- **Jet transverse-momentum regression**. The $b\bar{b}$ mass resolution is improved by the means of regression technique. The regression is carried out on the individual b-jets taking into account the jet composition properties and targets mostly semi-leptonic b-decays that can lead to an energy mismeasurement due to escaping neutrinos. A multivariate Boosted Decision Tree(BDT), is trained on simulated b-jets in ditop events with the inputs including information about the internal jet properties and structure. To validate the regression techniques sample of $Z \to ll$ events with one or two b-tagged jets are used. When regression is applied the p_T balance distribution of the system is improved and the simulation is in agreement with data. Figure 1 shows the reconstructed invariant mass of the b-jet candidates (m_{bb}) before and after the regression for simulated events passing the DoubleB and SingleB selections. With this p_T technique the dijet invariant mass resolution is improved by 7%.
- **Discrimination between quark-and gluon-originated jets**. VBF signal events, dominated by quark jets, have narrower jets, while the background and ggH events are enriched in wider gluon jets. To exploit this property in the signal discrimination the minor Root-Mean-Square of the distribution of jet constituents in the $\eta - \phi$ plane[15] is used.
- **Soft QCD activity**. The color flow is suppressed in the VBF signal events leading to very little additional hadronic activity, while this is not the case for the background QCD produc-

PoS(EPS-HEP2017)688

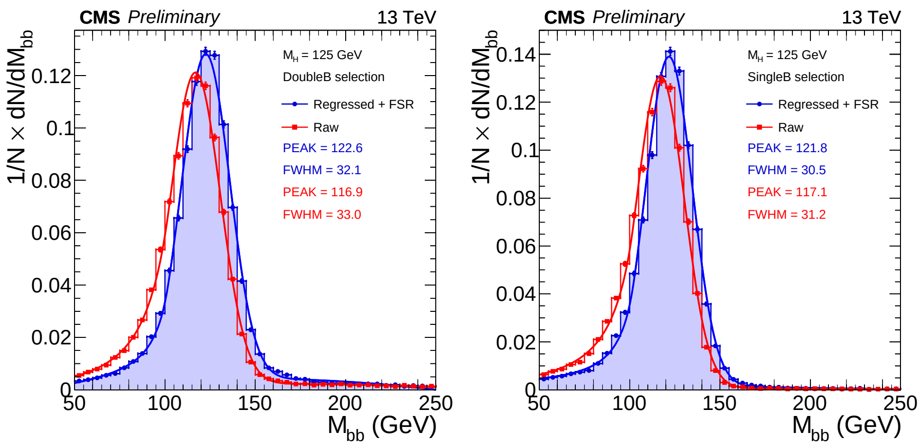

Figure 1: Simulated invariant mass distribution of the two b-jet candidates before and after the jet p_T regression for VBF signal events with the generated Higgs boson signal mass of 125 GeV in DoubleB (left) and SingleB(right) datasets[10].

tion. To discriminate between the signal and the QCD background multiplicity of soft track is used as a discriminating variable[10, 11].

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 a BDT multivariate discriminant. The variables chosen as inputs to the BDT are only weakly correlated to the dynamics of the bb-system. Based on the BDT outputs, seven categories are defined to maximize the signal sensitivity.
By construction the $m_{b\bar{b}}$ spectrum should be independent of the category. In practice, there is a small residual correlation between the BDT output and $m_{b\bar{b}}$, which is taken into account with data-driven transfer functions (linear function in SingleB and quadratic in DoubleB). The fit model for the Higgs boson signal is given by the 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 μ_H, $N_{i,\mathrm{qcd}}$ are free parameters for the signal strength and the QCD event yield. $N_{i,\mathrm{H}}$, $N_{i,Z}$, $N_{i,\mathrm{top}}$ 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. The QCD shape is described by a polynomial B, common within the categories of each set, and a multiplicative transfer function K_i. For SingleB, the polynomial B is of fifth order, for DoubleB it is of fourth order.

6. Results

The invariant mass distributions of the b-jet candidates in data are fitted simultaneously in all categories under two hypotheses: background only and background plus a Higgs boson signal. A binned likelihood fit is performed with the systematic uncertainties incorporated as nuisance parameters. The limits on the signal strength are computed with the asymptotic CLs method[16, 17]. For a 125 GeV Higgs boson signal the observed 95% confidence level upper limit is 3.0 times the SM 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[18]. 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 μ = 1.3±1.2 with a observed (expected) significance of 1.2 (0.9) standard deviations.

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] P. W. Higgs, *"Broken symmetries, massless particles and gauge fields,"* Phys. Lett. **12** (1964) 132.

[5] 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]].

[6] S. Chatrchyan *et al.* [CMS Collaboration], *"Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$ = 7 and 8 TeV,"* JHEP **1306** (2013) 081 [arXiv:1303.4571 [hep-ex]].

[7] 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]].

[8] A. M. Sirunyan *et al.* [CMS Collaboration], *"Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair,"* arXiv:1709.05543 [hep-ex].

[9] S. Chatrchyan *et al.* [CMS Collaboration], *"The CMS Experiment at the CERN LHC,"* JINST **3** (2008) S08004.

[10] CMS Collaboration, *"VBF H to bb using the 2015 data sample,"* CMS-PAS-HIG-16-003, http://cds.cern.ch/record/2160154.

[11] 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]].

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

[13] CMS Collaboration, *"Identification of b quark jets at the CMS Experiment in the LHC Run 2,"* CMS-PAS-BTV-15-001, http://cds.cern.ch/record/2138504.

[14] A. M. Sirunyan *et al.* [CMS Collaboration], *"Particle-flow reconstruction and global event description with the CMS detector,"* arXiv:1706.04965 [physics.ins-det].

[15] CMS Collaboration, *"Performance of quark/gluon discrimination in 8 TeV pp data,"* CMS-PAS-JME-13-002, http://cds.cern.ch/record/1599732.

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

[17] 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]].

[18] 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]].