Physics Letters B 780 (2018) 501–532

Contents lists available at ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Evidence for the Higgs boson decay to a bottom quark–antiquark pair

The CMS Collaboration *

CERN, Switzerland

ARTICLE INFO

Article history:
Received 21 September 2017
Received in revised form 17 February 2018
Accepted 20 February 2018
Available online 27 February 2018
Editor: M. Doser

Keywords:
CMS
Physics
Higgs

ABSTRACT

A search for the standard model (SM) Higgs boson (H) decaying to $\mathrm{b\bar{b}}$ when produced in association with an electroweak vector boson is reported for the following processes: $\mathrm{Z(\nu\nu)H}$, $\mathrm{W(\mu\nu)H}$, $\mathrm{W(e\nu)H}$, $\mathrm{Z(\mu\mu)H}$, and Z(ee)H. The search is performed in data samples corresponding to an integrated luminosity of 35.9 $\mathrm{fb^{-1}}$ at $\sqrt{s} = 13$ TeV recorded by the CMS experiment at the LHC during Run 2 in 2016. An excess of events is observed in data compared to the expectation in the absence of a $\mathrm{H \to b\bar{b}}$ signal. The significance of this excess is 3.3 standard deviations, where the expectation from SM Higgs boson production is 2.8. The signal strength corresponding to this excess, relative to that of the SM Higgs boson production, is 1.2 ± 0.4. When combined with the Run 1 measurement of the same processes, the signal significance is 3.8 standard deviations with 3.8 expected. The corresponding signal strength, relative to that of the SM Higgs boson, is $1.06^{+0.31}_{-0.29}$.

1. Introduction

The ATLAS and CMS Collaborations reported in 2012 the discovery of a new boson with a mass near 125 GeV using data from the Large Hadron Collider (LHC) at CERN [1–3]. Significant signals have been observed in channels where the boson decays into $\gamma\gamma$, ZZ, WW, or $\tau\tau$ [4–13]. The measured production and decay rates and spin-parity properties of this boson [14–20] are compatible with those of the standard model (SM) Higgs boson (H) [21–26].

The $\mathrm{H \to b\bar{b}}$ decay tests directly the Higgs boson coupling to fermions, and more specifically to down-type quarks, and has not yet been established experimentally. In the SM, for a Higgs boson mass $m_\mathrm{H} = 125$ GeV, the branching fraction is approximately 58% [27], by far the largest. An observation in this channel is necessary to solidify the Higgs boson as the source of mass generation in the fermion sector of the SM [28,29].

At the Tevatron $\mathrm{p\bar{p}}$ collider the sensitivity of the SM Higgs boson search, for masses below 130 GeV, was dominated by its production in association with a weak vector boson (VH production) and its decay to $\mathrm{b\bar{b}}$ [30]. The combined searches from the CDF and D0 Collaborations resulted in an excess of events with a local significance, at $m_\mathrm{H} = 125$ GeV, of 2.8 standard deviations, with an expected value of 1.6. For the $\mathrm{H \to b\bar{b}}$ search at the LHC, the following Higgs boson production processes have been considered: in association with a top quark pair [31–34], through vector boson fusion [35,36], through VH production [37,38], and, more recently, through gluon fusion [39]. The process with the largest sensitivity is VH production.

The combined searches for $\mathrm{H \to b\bar{b}}$ by the ATLAS and CMS Collaborations in Run 1, at $\sqrt{s} = 7$ and 8 TeV, evaluated for a Higgs boson mass of 125.09 GeV, resulted in a significance of 2.6 standard deviations, with 3.7 standard deviations expected [18]. The corresponding signal strength, relative to the SM expectation, is $\mu = 0.7 \pm 0.3$. The significance from the individual search by the ATLAS (CMS) experiment is 1.7 (2.0) standard deviations, with 2.7 (2.5) standard deviations expected, and a signal strength $\mu = 0.6 \pm 0.4$ ($\mu = 0.8 \pm 0.4$).

Recent results by the ATLAS Collaboration [40] in the search for $\mathrm{H \to b\bar{b}}$ through VH production at $\sqrt{s} = 13$ TeV, with data corresponding to an integrated luminosity of 36.1 $\mathrm{fb^{-1}}$, report a significance of 3.5 standard deviations, corresponding to a signal strength of $\mu = 1.20^{+0.42}_{-0.36}$. The combination with the results from the same search in Run 1 [37] yields a significance of 3.6 standard deviations and a signal strength $\mu = 0.90^{+0.28}_{-0.26}$.

This article reports on the search with the CMS experiment for the decay of the SM Higgs boson to bottom quarks, $\mathrm{H \to b\bar{b}}$, when produced through the $\mathrm{pp \to VH}$ process, where V is either a W or a Z boson. This search is performed with data samples from Run 2 of the LHC, recorded during 2016, corresponding to an integrated luminosity of 35.9 $\mathrm{fb^{-1}}$ at $\sqrt{s} = 13$ TeV. The following five processes are considered in the search: $\mathrm{Z(\nu\nu)H}$, $\mathrm{W(\mu\nu)H}$, $\mathrm{W(e\nu)H}$, $\mathrm{Z(\mu\mu)H}$, and Z(ee)H. The final states that predominantly correspond to these processes, respectively, are characterized by

* *E-mail address:* cms-publication-committee-chair@cern.ch.

https://doi.org/10.1016/j.physletb.2018.02.050
0370-2693/

the number of leptons required in the event selection, and are referred to as the 0-, 1-, and 2-lepton channels.

Throughout this article the term "lepton" (denoted ℓ) refers solely to muons and electrons, but not to taus. The leptonic tau decays in WH and ZH processes are implicitly included in the W($\mu\nu$)H, W(eν)H, Z($\mu\mu$)H, and Z(ee)H processes. Background processes originate from the production of W and Z bosons in association with jets from gluons and from light- or heavy-flavor quarks (W+jets and Z+jets), from singly and pair-produced top quarks (single top and $\mathrm{t\bar{t}}$), from diboson production (VV), and from quantum chromodynamics multijet events (QCD).

Simulated samples of signal and background events are used to optimize the search. For each channel, a signal region enriched in VH events is selected together with several control regions, each enriched in events from individual background processes. The control regions are used to test the accuracy of the simulated samples' modeling for the variables relevant to the analysis. A simultaneous binned-likelihood fit to the shape and normalization of specific distributions for the signal and control regions for all channels combined is used to extract a possible Higgs boson signal. The distribution used in the signal region is the output of a boosted decision tree (BDT) event discriminant [41,42] that helps separate signal from background. For the control regions, a variable that identifies jets originating from b quarks, and that discriminates between the different background processes, is used. To validate the analysis procedure, the same methodology is used to extract a signal for the VZ process, with $\mathrm{Z} \to \mathrm{b\bar{b}}$, which has a nearly identical final state to VH with $\mathrm{H} \to \mathrm{b\bar{b}}$, but with a production cross section of 5 to 15 times larger, depending on the kinematic regime considered. Finally, the results from this search are combined with those of similar searches performed by the CMS Collaboration during Run 1 [18,36,38].

This article is structured as follows: Sections 2–3 describe the CMS detector, the simulated samples used for signal and background processes, and the triggers used to collect the data. Sections 4–5 describe the reconstruction of the detector objects used in the analysis and the selection criteria for events in the signal and control regions. Section 6 describes the sources of uncertainty in the analysis, and Section 7 describes the results, summarized in Section 8.

2. The CMS detector and simulated samples

A detailed description of the CMS detector can be found elsewhere in Ref. [43]. The momenta of charged particles are measured using a silicon pixel and strip tracker that covers the range $|\eta| < 2.5$ and is immersed in a 3.8 T axial magnetic field. The pseudorapidity is defined as $\eta = -\ln[\tan(\theta/2)]$, where θ is the polar angle of the trajectory of a particle with respect to the direction of the counterclockwise proton beam. Surrounding the tracker are a crystal electromagnetic calorimeter (ECAL) and a brass and scintillator hadron calorimeter (HCAL), both used to measure particle energy deposits and both consisting of a barrel assembly and two endcaps. The ECAL and HCAL extend to a range of $|\eta| < 3.0$. A steel and quartz-fiber Cherenkov forward detector extends the calorimetric coverage to $|\eta| < 5.0$. The outermost component of the CMS detector is the muon system, consisting of gas-ionization detectors placed in the steel flux-return yoke of the magnet to measure the momenta of muons traversing through the detector. The two-level CMS trigger system selects events of interest for permanent storage. The first trigger level, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events in less than 3.2 μs. The high-level trigger software algorithms, executed on a farm of commercial processors, further reduce the event rate using information from all detector subsystems. The variable $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is used to measure the separation between reconstructed objects in the detector, where ϕ is the angle (in radians) of the trajectory of the object in the plane transverse to the direction of the proton beams.

Samples of simulated signal and background events are produced using the Monte Carlo (MC) event generators listed below. The CMS detector response is modeled with GEANT4 [44]. The signal samples used have Higgs bosons with $m_\mathrm{H} = 125$ GeV produced in association with vector bosons. The quark-induced ZH and WH processes are generated at next-to-leading order (NLO) using the POWHEG [45–47] v2 event generator extended with the MiNLO procedure [48,49], while the gluon-induced ZH processes (denoted ggZH) are generated at leading-order (LO) accuracy with POWHEG v2. The MADGRAPH5_aMC@NLO [50] v2.3.3 generator is used at NLO with the FxFx merging scheme [51] for the diboson background samples. The same generator is used at LO accuracy with the MLM matching scheme [52] for the W+jets and Z+jets in inclusive and b-quark enriched configurations, as well as the QCD multijet sample. The $\mathrm{t\bar{t}}$ [53] production process, as well as the single top quark sample for the t-channel [54], are produced with POWHEG v2. The single top quark samples for the tW- [55] and s-channel [56] are instead produced with POWHEG v1. The production cross sections for the signal samples are rescaled to next-to-next-to-leading order (NNLO) QCD + NLO electroweak accuracy combining the VHNNLO [57–59], VH@NNLO [60,61] and HAWK v2.0 [62] generators as described in the documentation produced by the LHC Working Group on Higgs boson cross sections [63], and they are applied as a function of the vector boson transverse momentum (p_T). The production cross sections for the $\mathrm{t\bar{t}}$ samples are rescaled to the NNLO with the next-to-next-to-leading-log (NNLL) prediction obtained with TOP++ v2.0 [64], while the W+jets and Z+jets samples are rescaled to the NLO cross sections using MADGRAPH5_aMC@NLO. The parton distribution functions (PDFs) used to produce the NLO samples are the NLO NNPDF3.0 set [65], while the LO NNPDF3.0 set is used for the LO samples. For parton showering and hadronization the POWHEG and MADGRAPH5_aMC@NLO samples are interfaced with PYTHIA 8.212 [66]. The PYTHIA8 parameters for the underlying event description correspond to the CUETP8M1 tune derived in Ref. [67] based on the work described in Ref. [68].

During the 2016 data-taking period the LHC instantaneous luminosity reached approximately 1.5×10^{34} cm^{-2} s^{-1} and the average number of pp interactions per bunch crossing was approximately 23. The simulated samples include these additional pp interactions, referred to as pileup interactions (or pileup), that overlap with the event of interest in the same bunch crossing.

3. Triggers

Several triggers are used to collect events with final-state objects consistent with the signal processes in the channels under consideration.

For the 0-lepton channel, the quantities used in the trigger are derived from the reconstructed objects in the detector identified by a particle-flow (PF) algorithm [69] that combines the online information from all CMS subsystems to identify and reconstruct individual particles emerging from the proton-proton collisions: charged hadrons, neutral hadrons, photons, muons, and electrons. The main trigger used requires that both the missing transverse momentum, $p_\mathrm{T}^\mathrm{miss}$, and the hadronic missing transverse momentum, $H_\mathrm{T}^\mathrm{miss}$, in the event be above a threshold of 110 GeV. Online $p_\mathrm{T}^\mathrm{miss}$ is defined as the magnitude of the negative vector sum of the transverse momenta of all reconstructed objects identified by the PF algorithm, while $H_\mathrm{T}^\mathrm{miss}$ is defined as the magnitude of the negative vector sum of the transverse momenta of all reconstructed