Inclusive search for boosted Higgs bosons using H→ bb̄ decays with the CMS experiment

Caterina Vernieri[*][†]

Fermi National Accelerator Laboratory, Batavia, IL, USA,

E-mail: caterina.vernieri@cern.ch

We present the first search for the standard model Higgs boson (H) produced with large transverse momentum (p_T) via gluon fusion and decaying to a bottom quark-antiquark pair (bb̄). The search is performed using a data set of pp collisions at $\sqrt{s} = 13$ TeV collected with the CMS experiment at the LHC, corresponding to an integrated luminosity of 35.9 fb^{-1}. A highly Lorentz-boosted Higgs boson decaying to bb̄ is reconstructed as a single, large radius jet and is identified using jet substructure and dedicated b tagging techniques. The method is validated with the first observation of the Z→ bb̄ process in the single-jet topology, with a local significance of 5.1 standard deviations (5.8 expected).

---

[*]Speaker.

[†]on behalf of the CMS Collaboration

1. Motivations

At a mass (m_H) of 125 GeV the standard model (SM) Higgs boson decay mode into a bottom quark-antiquark pair (bb̄) dominates the total width (∼58%) [1]. While the decay of the Higgs boson to vector bosons has been observed in different channels (ZZ, γγ,WW) [2, 3, 4], the direct couplings of the H to down-type quarks, remains to be firmly established [5, 6, 7, 8].

The current measurements constrain indirectly the couplings to the up-type top quark, since the dominant Higgs boson production mechanism is gluon fusion induced by top-quark loop. The measurement of the H(bb̄) decay represents a direct test of whether the observed boson interacts as expected with the quark sector and provides the unique final test of the direct coupling of the Higgs boson to down-type quarks, an essential aspect of the nature of the newly discovered boson. To date, the most precise constraints on the couplings to down-type quarks are provided by the CMS and ATLAS experiments, that have recently announced the first evidence for the Higgs boson decay into b quarks and for its production in association with a vector boson [6, 8]. No one has yet to consider the gluon fusion (ggF) production mode, despite being the dominant production mechanism at the LHC (87%). The search for ggF H(bb̄) historically has been deemed impossible [9, 10] because of the overwhelming irreducible background from QCD production of b quark, which is roughly 7 order of magnitude larger than the signal. We present the first inclusive search for H→ bb̄ based on a data set of pp collisions at $\sqrt{s} = 13$ TeV collected with the CMS detector [11] at the LHC in 2016 and corresponding to an integrated luminosity of 35.9 fb^{-1} [12]. The main experimental difficulties for this search originate from the large cross section for background multijet events and the restrictive trigger requirements needed to reduce the data recording rate. Therefore, we require events to have a high-p_T H, above 450 GeV. Being this search performed at very high p_T it could also potentially be sensitive to new physics variation to the couplings that would enhance or reduce the Higgs boson production cross section [13, 14], if the scale of new physics is larger than the electroweak scale.

2. ggF H modeling at high p_T

Computing the differential cross section in H p_T for ggF H $p_T > 450$ GeV poses a number of challenges. At low H p_T, the dominant contributions come from the application of higher order corrections which are large for loop-induced processes. The dominant correction at values of the H p_T greater than approximately twice the mass of the top quark (m_t) originates from the resolved top quark loop (finite top mass correction) [15]. The resolved top quark loop induces a deficit in the production of Higgs bosons at high p_T relative to the case where the loop is unresolved, known as the effective field theory (EFT) or $m_t \to \infty$ approximation.

In the interest of comparing with other CMS results, the POWHEG generator with H matrix elements up to 1 jet is used and tuned with the *h-fact* parameter set to 104.13 GeV. The resulting tuned H generation is normalized to the inclusive next-to-next-to-next-to-leading order (N^{3}LO) accuracy. In addition, an alternative approach is considered to get the highest order possible differential H p_T spectrum [16, 17, 18], while preserving the finite top mass correction [19]. To account for both the effects of higher order corrections and the resolved top loop, a multi-correction ap-

proach is adopted [21, 22], that can be summarized as:

$$\text{ggF H(NNLO}+\text{m}_\text{t}) = \text{Powheg(1 jet } m_t) \times \frac{\text{MG LO } 0-2 \text{ jet } m_t}{\text{Powheg(1 jet } m_t)} \times \frac{\text{NLO 1 jet } m_t}{\text{LO 1 jet } m_t} \times \frac{\text{NNLO 1 jet } m_t \rightarrow \infty}{\text{NLO 1 jet } m_t \rightarrow \infty}$$

(2.1)

Samples are generated at LO for the 0, 1, and 2 jet H production with MADGRAPH (MG) using the loop$_\text{SM}$ model [23] showered with CKKW-L scheme [24, 25, 26]. This spectrum is then corrected by the approximate NLO to LO ratio, obtained by expanding in powers of $1/m_t^2$ and it is found to be 2.0 ± 0.5 and roughly constant as a function of p$_\text{T}$. The effective NNLO to NLO ratio [19] in the infinite top quark mass approximation is found to be 1.25 ± 0.15 and is also roughly constant across p$_\text{T}$ [27, 16, 28]. For H p$_\text{T}$ > 450 GeV, the correction to the default POWHEG is found to be 1.27 ± 0.38, resulting in a cross section of 31.7 ± 9.5 fb for ggF H→ bb̄. The Higgs boson generator-level p$_\text{T}$ distribution is shown in Fig. 1, to compare the POWHEG prediction and the one derived to account of higher order corrections and the finite top mass loop. An uncertainty of 30% to the overall correction is estimated from the comparison of different predictions obtained by using: (i) a merging scale of 100 instead of 20 GeV, (ii) the inclusive two-jet ggF process generation, (iii) the MADGRAPH5_AMC@NLO effective field theory approximation [19, 20] normalized to the inclusive N^{3}LO cross section.
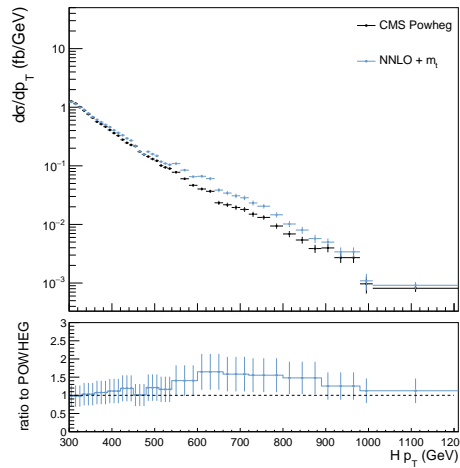

Figure 1: Generator level H p$_\text{T}$ distribution for the gluon fusion production mode. The CMS default POWHEG sample and the corrected spectrum to account for both higher order and finite top mass effects are compared [12].

3. Higgs boson tagging

The angular separation of the two b from H decay can be approximated by $2m_H/\text{p}_\text{T}$. Thus, at high p$_\text{T}$ the products of hadronization of two b quarks are merged in one single large cone size (fat) jet. Higgs boson candidates decaying to bb̄ are reconstructed as jet (H jet) using the anti-k$_\text{T}$ algorithm [29] with a distance parameter of 0.8 (AK8 jets), and specific techniques are used to mitigate effects from multiple p-p interactions per collision (pileup). The pileup per particle

identification (PUPPI) algorithm [30] assigns a weight to each particle prior to jet clustering based on the likelihood of the particle originating from the hard interaction. Mainly three properties are then exploited to isolate H jet candidate from background of q/g jets: the jet mass, b tagging and the composite nature of the jet using substructure.

The jet mass should be compatible with the Higgs boson mass hypothesis and provide good discrimination from quark- and gluon-initiated jets. Soft-drop grooming is applied to remove soft and wide-angle radiation [31, 32]. Grooming tends to push the jet mass scale of the background to lower values while preserving the hard scale of the heavy resonance. The soft drop jet mass (m_{SD}) peaks at the H mass for signal events and reduces the masses of jets from background. The N_2^1 variable [33], which is based on a ratio of 2-point and 3-point generalized energy correlation functions (ECFs) [34], is exploited to determine how consistent a jet is with having a two-prong substructure. However, any selection on N_2^1 shapes the jet mass distributions differently depending on the p_{T} of the jet. Therefore a mass-decorrelated procedure is applied [35] to achieve a constant QCD background efficiency of 26% across the entire mass and p_{T} range considered in this search. The chosen percentile maximizes the sensitivity to the H(bb̄) signal.

In order to select events in which the H jet is most likely to contain two b quarks, we use the double-b tagger algorithm [36]. Several observables that characterize the distinct properties of b hadrons and their flight directions in relation to the jet substructure are used as input variables to this multivariate algorithm, to distinguish between H jets and QCD jets. In this search, an H jet is considered double-b tagged if its double-b tag discriminator value is above a threshold corresponding to a 1% misidentification rate for QCD jets and a 33% efficiency for H(bb̄) jets. By design the mistag rate is approximately flat across the p_{T} range, and it is a critical point for this search.

4. Event Selection

We exploit the H tagging tools to perform the first search for ggF H(bb̄). We look for a single high-p_{T} H jet, recoiling against some other object, like a narrow jet radiated in the initial state, although no assumption or requirement on such additional object is actually made in the analysis. Combinations of several online selections are used, all requiring the total hadronic transverse energy in the event (H_{T}) or jet p_{T} to be above a given threshold. The online selection is fully efficient at selecting events offline with at least one AK8 jet with $p_{\mathrm{T}} > 450$ GeV and $\eta < 2.5$. The leading (in p_{T}) jet in the event is assumed to be the Higgs boson candidate, the H jet, and the substructure and b tagging requirements are applied. Events with (without) a double-b tagged H jet define the passing (failing) region. In the passing region, the gluon fusion process dominates, although other Higgs boson production mechanisms contribute: VBF (12%), VH (8%), ttH̄ (5%). They are all taken into account when extracting the Higgs boson yield.

5. Background modeling

The contribution of tt̄ production to the total SM background is estimated to be less than 3%. It is obtained from simulation corrected with scale factors derived from a tt̄-enriched control sample in which an isolated muon is required.

The contributions of the W and Z+jets processes is about 5% and it is modeled from simulation. During the signal extraction we measure the Z+jets normalization while the W+jets normalization is allowed to vary within its systematic uncertainties.

The main background in the passing region, QCD multijet production, has a nontrivial jet mass shape that is difficult to model parametrically and dependent on jet p_T so we constrain it using the signal-depleted failing region. Since the double-b tagger discriminator and the jet mass are largely uncorrelated, the passing and failing regions have similar QCD jet mass distributions, and their ratio, the "pass-fail ratio" $R_{p/f}$, is expected to be nearly constant as a function of jet mass and p_T. To account for the residual difference between the shapes of passing and failing events, $R_{p/f}$ is parametrized as a polynomial in ρ and p_T, which is determined from a simultaneous fit to the data in passing and failing regions across the whole jet mass range.

6. Results

A binned maximum likelihood fit to the observed m_{SD} distributions in the range 40 to 201 GeV is performed using the sum of the H(bb̄), W, Z, tt̄, and QCD multijet contributions. The fit is done simultaneously in the passing and failing regions. The production cross sections relative to the SM cross sections (signal strengths) for the Higgs and the Z bosons, μ_H and μ_Z, respectively, are extracted from the fit. Figure 2 shows the m_{SD} distributions in data for the passing and failing regions with measured SM background and H(bb̄) contributions. Contributions from W and Z boson production are clearly visible in the data. Table 1 summarizes the measured signal strengths
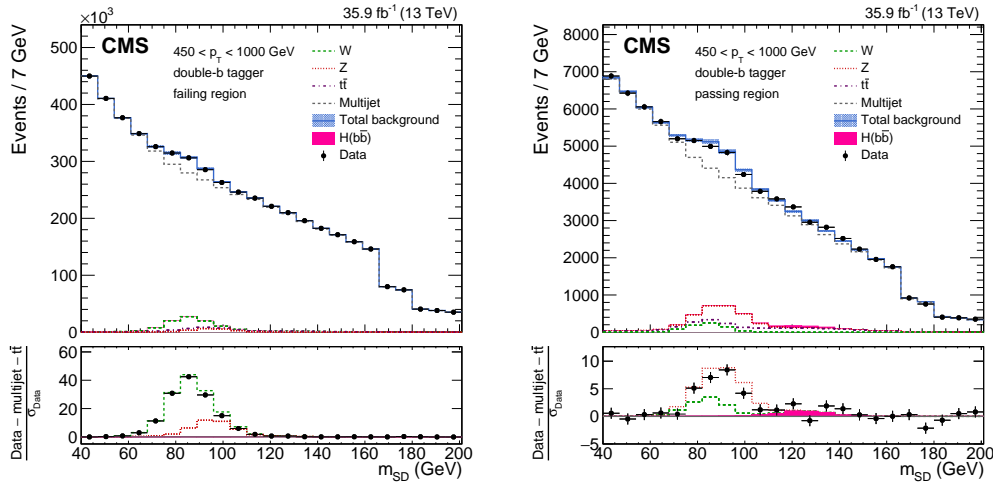

Figure 2: The m_{SD} distributions in data for the failing (left) and passing (right) regions. In the bottom panel, the ratio of the data to its statistical uncertainty, after subtracting the non-resonant backgrounds, is shown [12].

and significances for the Higgs and Z boson processes. In particular, they are also reported for the case in which no corrections to the Higgs boson p_T spectrum are applied.

Table 1: Fitted signal strength, expected and observed significance of the Higgs and Z boson signal.

	H	H no $p_T$corr.	Z
Observed signal strength	$2.3^{+1.8}_{-1.6}$	$3.2^{+2.2}_{-2.0}$	$0.78^{+0.23}_{-0.19}$
Expected significance	0.7σ	0.5σ	5.8σ
Observed significance	1.5σ	1.6σ	5.1σ

7. Conclusions

In summary, an inclusive search for the standard model Higgs boson with $p_T > 450$ GeV decaying to bottom quark-antiquark pairs and reconstructed as a single, large-radius jet is presented. The Z+jets process is observed for the first time in the single-jet topology with a significance of 5.1σ. The Higgs boson production is measured with an observed (expected) significance of 1.5σ (0.7σ) when including Higgs boson p_T spectrum corrections accounting for higher-order and finite top quark mass effects. The measured cross section times branching fraction for the gluon fusion H(bb̄) production for $p_T > 450$ GeV and $\eta < 2.5$ is $74 \pm 48(\text{stat})^{+17}_{-10}(\text{syst})$ fb, which is consistent with the SM prediction within uncertainties. This search looks at previously unexplored regions of phase space and opens a new strategy to search for H(bb̄) and probe BSM contributions to the Higgs boson production cross section at very high p_T.

References

[1] LHC Higgs Cross Section Working Group CERN-2011-002, arXiv:1101.0593 (2011)

[2] CMS Collaboration, Phys. Rev. D 89 (2014) 092007

[3] CMS Collaboration, EPJC 74 (2014) 3076

[4] CMS Collaboration, JHEP 1401 (2014) 096

[5] CMS Collaboration, Nature Phys. 10 (2014) 3005

[6] ATLAS Collaboration, arXiv:1708.03299 (2017) submitted to JHEP

[7] CDF and D0 Collaborations, Phys. Rev. D 88 (2013) 052014

[8] CMS Collaboration, arXiv:1709.07497 (2017) submitted to PLB

[9] CDF Collaboration, CDF-PUB-11228 (2017)

[10] ATLAS Collaboration, Phys. Lett. B 738 (2014)

[11] CMS Collaboration, JINST 3 (2008) S08004

[12] CMS Collaboration, arXiv:1709.05543 (2017) submitted to PRL

[13] M. Grazzini, A. Ilnicka, M. Spira and M. Wiesemann, JHEP 1703 (2017) 115

[14] F. Maltoni, E. Vryonidoua and C. Zhangb, JHEP 10 (2016) 123

[15] U. J. Baur and E. W. N. Glover, Nucl. Phys. B 339 (1990) 38

[16] R. Boughezal, F. Caola, K. Melnikov, F. Petriello and M. Schulze, JHEP 06 (2013) 072

[17] R. Boughezal, J. M. Campbell, R. K. Ellis, C. Focke, W. Giele, X. Liu, F. Petriello, and C. Williams, arXiv:1605.08011 (2016)

PoS(EPS-HEP2017)349

[18] X. Chen, T. Gehrmann, N. Glover, and M. Jaquier, arXiv:1604.04085 (2016)

[19] T. Neumann, and C. Williams, Phys. Rev. D 95 (2017) 014004

[20] D. de Florian, C. Grojean, F. Maltoni, C. Mariotti, A. Nikitenko, M. Pieri, P. Savard, M. Schumacher and R. Tanaka, arXiv:1610.07922 (2016)

[21] M. Buschmann, D. Goncalves, S. Kuttimalai, M. Schonherr, F. Krauss, and T. Plehn, JHEP 02 (2015) 038

[22] R. Frederix, S. Frixione, E. Vryonidou and M. Wiesemann, JHEP 08 (2016) 006

[23] V. Hirschi and O. Mattelaer, JHEP 10 (2015) 146

[24] S. Catani, F. Krauss, R. Kuhn, B. R. Webber, JHEP 11 (2001) 063

[25] L. Lonnblad, JHEP 05 (2002) 046

[26] M. Czakon, P. Fiedler and A. Mitov, Phys. Rev. Lett. 110 (2013) 252004

[27] X. Chen, T. Gehrmann, E. W. N. Glover, and M. Jaquier, Phys. Lett. B 740 (2015)

[28] R. Boughezal, F. Caola, K. Melnikov, F. Petriello, and M. Schulze, Phys. Rev. Lett. 115 (2015) 082003

[29] M. Cacciari, G. P. Salam, and G. Soyez, JHEP 04 (2008) 063

[30] D. Bertolini, P. C. Harris, M. Low, and N. Tran, JHEP 10 (2014) 059

[31] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, JHEP 09 (2013) 029

[32] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, JHEP 05 (2014) 146

[33] I. Moult, L. Necib, and J. Thaler, JHEP 12 (2016) 153

[34] A. J. Larkoski, G. P. Salam and J. Thaler, JHEP 06 (2013) 108

[35] J. Dolen, P. C. Harris, S. Marzani, S. Rappoccio, and N. Tran, JHEP 05 (2016) 156

[36] CMS Collaboration, CMS-BTV-15-002 (2015), https://cds.cern.ch/record/2195743?ln=it