PROCEEDINGS OF SCIENCE

Measurements of Higgs boson properties in hadronic final states at CMS

Leonardo Giannini* **on behalf of the CMS Collaboration**
SNS & INFN Pisa
E-mail: leonardo.giannini@cern.ch

The most recent CMS Higgs boson (H) physics results, with the Higgs decaying to a bottom quark-antiquark pair ($b\bar{b}$), are presented. The focus is on the analysis of data collected at $\sqrt{s} = 13$ TeV in 2017, corresponding approximately to 41 fb^{-1}.
The analysis targeting the associated production of a Higgs boson and a vector boson (VH), with $H \to b\bar{b}$, yields an observed significance of 3.3 standard deviations (σ) above the background-only hypothesis, with an expected significance of 3.1σ. The corresponding measured signal strength is μ = 1.08 $\pm$ 0.34. When combined with previous VH($b\bar{b}$) measurements using data collected at $\sqrt{s} = 7$, 8 and 13 TeV, the observed (expected) significance is 4.8 (4.9) σ, corresponding to μ = 1.01 $\pm$ 0.22. Finally, the combination of this result with searches for $H \to b\bar{b}$ in other production modes by the CMS experiment, including $t\bar{t}H(b\bar{b})$ with 2016 data, results in a significance of 5.6 (5.5) σ, and $\mu = 1.04 \pm 0.20$.
The analysis targeting the $t\bar{t}H(b\bar{b})$ process using 2017 data is also covered. The significance for this production mode is of 3.7 (2.6) σ , corresponding to $\mu = 1.49 \pm 0.42$. In combination with the previous result using 2016 data, a significance of 3.9 (3.5) σ is obtained, with $\mu = 1.15 \pm 0.30$.

European Physical Society Conference on High Energy Physics - EPS-HEP2019 -
10-17 July, 2019
Ghent, Belgium

*Speaker.

PoS(EPS-HEP2019)342

1. Introduction

The only hadronic decay of the Higgs boson we can currently probe with sensitivity at the level of the Standard Model (SM) expectation at the LHC is the decay to $b\bar{b}$. The $b\bar{b}$ decay has a few experimental advantages: in the SM, for a 125 GeV Higgs boson, the $H \to b\bar{b}$ decay has the largest branching fraction ($\sim 58\%$). Additionally, hadronic jets originating from b quarks' hadronization can be effectively tagged, thus removing a large amount of the SM backgrounds. A precise measurement of the $H \to b\bar{b}$ decay is important, as it allows to measure directly the Yukawa coupling of the Higgs boson to a down-type quark, and provides a necessary test of the hypothesis that the Higgs field is the source of mass generation for the fermions.

However, all the searches for $H \to b\bar{b}$ need to target either single production modes, or very specific kinematic regimes, or both to be sensitive. This is mandatory at the trigger level, as selecting final states based on the presence of two b jets only is not feasible, and necessary also in the offline selection in order to further reject the backgrounds. Multivariate techniques are then fundamental to optimize the searches.

At the LHC the most sensitive production process in the search for the $H \to b\bar{b}$ decay is when the Higgs boson is produced in association with a vector boson (VH). The vector boson decay into leptons is used both at the trigger level and in the offline event selection. Another important production mode with good sensitivity is the $t\bar{t}H$ one. The $t\bar{t}H(b\bar{b})$ measurement in combination with other final states allows to precisely measure the coupling of the Higgs boson to the top quark. Other searches for $H \to b\bar{b}$ have been demonstrated to be possible. In particular, the search for inclusively produced $H(b\bar{b})$ with large transverse momentum (p_T) [1] is worth mentioning. In this case the tiny phase space selected allows both to remove the backgrounds and to access the very high p_T tail of the Higgs production spectrum.

In the following sections the two analyses targeting $VH(b\bar{b})$ [2] and $t\bar{t}H(b\bar{b})$ [3] using CMS 2017 data are presented. A detailed description of the CMS experiment can be found in reference [4]. The data was collected at $\sqrt{s} = 13$ TeV and corresponds approximately to 41 fb^{-1}. Both the new results are combined with the previous ones by CMS. Previously, the ATLAS [5] and CMS [6] Collaborations reported evidence for $VH(b\bar{b})$ with observed (expected) significances of 3.6 (4.0) σ and 3.8 (3.8) σ, respectively. The results were obtained using data collected during Run 1 at $\sqrt{s} = 7$ and 8 TeV, correspoding to approximately 20 fb^{-1}, and Run 2 at $\sqrt{s} = 13$ TeV, corresponding to approximately 36 fb^{-1}. Previous searches for $t\bar{t}H(b\bar{b})$ [7, 8, 9] reached observed (expected) significances of up to 1.6 (2.2) σ, and were used as input for the $t\bar{t}H$ combination, leading to the observation of the $t\bar{t}H$ production mode [10, 11].

2. VH with $H \to b\bar{b}$

The $VH(b\bar{b})$ analysis targets five final states: $Z(\nu\nu)$, $W(\mu\nu)$, $W(e\nu)$, $Z(\mu\mu)$, and $Z(ee)$ + $H(b\bar{b})$. Hence three mutually exclusive categories with 0, 1, or 2 reconstructed charged leptons are defined. The online selection is based on the presence of leptons or p_T^{miss}, where p_T^{miss} is the module of the negative vectorial p_T sum of all the particles. Offline, events with 0 leptons and large p_T^{miss}, one central isolated lepton and p_T^{miss}, or two opposite charge central isolated leptons, are selected. The leptons and p_T^{miss} are used to reconstruct the vector boson candidates and a boost of the vector boson is required to reject the backgrounds. Two central hadronic jets are also required. The jets selected to build the Higgs boson candidate and reconstruct the dijet invariant mass (m(jj)) are selected as the two most b-tagged jets according to the DeepCSV algorithm [12], which combines track and secondary vertex information via a deep neural network (DNN).

After all event selection criteria are applied, the m(jj) resolution is approximately 15%. The m(jj) resolution is improved by applying a multivariate regression, implemented via a DNN trained on simulated b jets. The usage of a DNN allowed to train on larger statistics and to add inputs describing the shape and composition of the jet, together with properties of secondary vertices and displaced tracks in the jet, thus improving on the previously otimized regression used in [6]. Additionally, in the 2-lepton channel, as no genuine p_T^{miss} from the hard-scattering process is expected, a kinematic fit is used to constrain the full event kinematics and further improve the mass resolution. After these improvements, the average m(jj) resolution is in the 10 - 13% range, depending on the channel, the p_T of the reconstructed Higgs boson and the number of extra jets. The signal invariant mass distributions for the best categories in the 2-lepton channel are shown in figure 1.

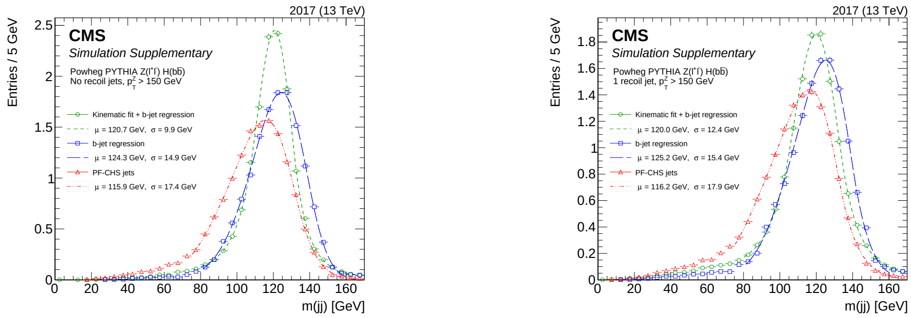

Figure 1: m(jj) distributions for simulated signal samples in the 2-lepton channel for events with no exta jets (left) and with one recoiling jet (right). The dstributions are shown before (red) and after (blue) the regression application, and after the kinematic fit procedure (green) is used on top of them. A Bukin fit is performed. The fitted mean and width are displayed on the figure [13].

The main background processes include the production of W and Z bosons in association with jets (V+jets), production of top quark pairs ($t\bar{t}$) and single top quarks, diboson production (WW, WZ, ZZ). For each channel, a signal region enriched in $VH(b\bar{b})$ events is selected together with several control regions, each enriched in events from individual background processes. 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 the signal.

The score of a DNN, which discriminates signal from background, is fitted in signal regions. The DNNs are trained separately for each channel using simulated samples for signal and all background processes. The inputs are discriminating variables exploiting the full event topology and including the m(jj). The control regions are chosen to be enriched in $t\bar{t}$ and for the V+jets, both with heavy and light flavor jets. The normalization only is fitted in the $t\bar{t}$ and V + light flavor jets control regions. In the V + heavy flavor jets, the minimum DeepCSV score is fitted in two bins in the 2-lepton channel, while the score of a yet another dedicated DNN (DNNHF) is used in the 0- and 1-lepton channels. The DNNHF exploits the same input variables as the signal region DNN, but is trained independently and with multiple output classes to distinguish the $t\bar{t}$, single top, and V+jets background processes. Both the DNN and the DNNHF for the 1-lepton channel post-fit distributions, with 1 electron in the final state, are shown as an example in figure 2.

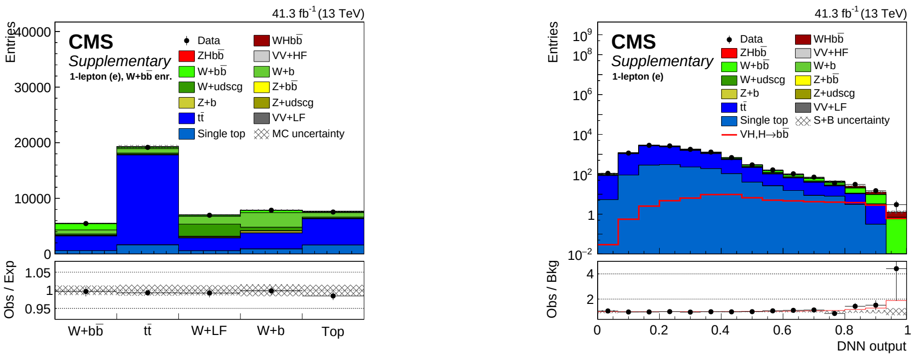

Figure 2: Post-fit distributions of the DNNHF fit variable for 2017 analysis in the 1-lepton (e) channel in control region and of the DNN in signal region. The DNNHF bins include events with the maximum score for each output class [13].

For the 2017 data, the observed significance is 3.3σ above the background-only hypothesis, while 3.1σ is expected for the SM Higgs boson. The corresponding measured signal strength is $\mu = 1.08 \pm 0.34$, where the uncertainty is a combination of statistical and systematic components. The dominant sources of uncertainty are the background normalization and the simulated sample size, followed by other experimental uncertainties.

The result is combined with CMS results using 2016 data and Run 1 data for VH($b\bar{b}$) only. The observed (expected) significance is 4.8 (4.9) σ, corresponding to $\mu = 1.01 \pm 0.22$. Figure 3 shows the distribution of events in all channels in bins of $\log_{10}(S/B)$ for the combined Run 1 and Run 2 datasets, where S/B is the signal to background ratio for the final discriminators (the DNNs for 2017) according to the fit result. Figure 3 (right) summarizes the signal strength results for both the inputs and the combination. A global combination of CMS measurements of the H($b\bar{b}$) decay is performed: the observed (expected) signal significance is 5.6 (5.5) σ, and the measured signal strength is $\mu = 1.04 \pm 0.20$. This result represents he observation of the H($b\bar{b}$) decay by the CMS Collaboration. The results of the fit are summarized in figure 5 (left).

3. $t\bar{t}$H with H $\rightarrow b\bar{b}$

The analysis targets final states with $t\bar{t}$ and b jets. Three different channels are considered depending on decays of the W bosons from the top quarks. The fully-hadronic channel, where both W bosons decay into quarks, the single-lepton channel, where one W boson decays into a charged

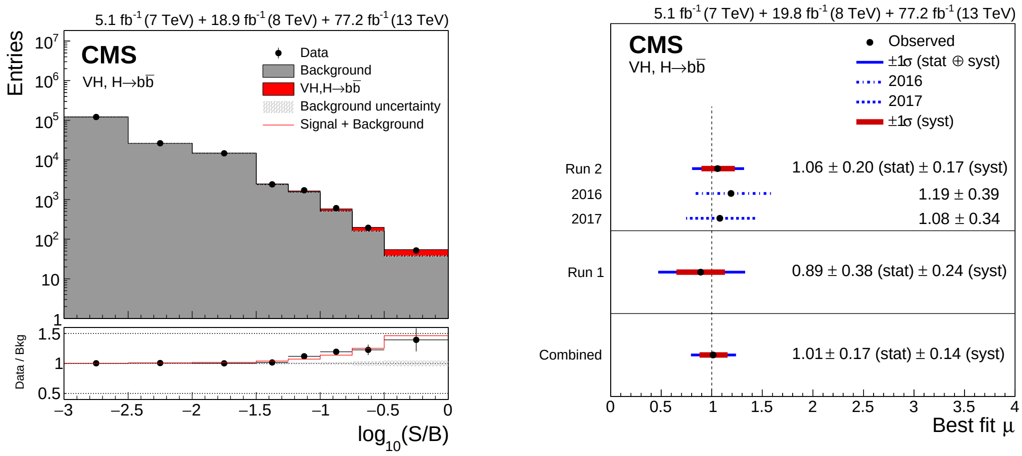

Figure 3: Distributions of data and simulated events in bins $\log_{10}(S/B)$ according to the results of the fits. All events in the VH($\mathrm{b\bar{b}}$) signal regions of the combined Run 1 and Run 2 data sets are included. The lower panel shows the data to background ratio with the total background uncertainty. The red line shows the signal+background divided by the background yield (left). Best fit value of the signal strength μ for the fit of the VH($\mathrm{b\bar{b}}$) process in the Run 1 and Run 2 data sets. 2016 and 2017 results, the Run 2 combination, and the Run 1 result are also shown. The 1σ systematic (red) and total (blue) uncertainties are indicated (right), [2].

lepton (e or μ) and a neutrino and the other W boson decays into quarks, and the dilepton channel, where both W bosons decay into a charged lepton (e or μ) and a neutrino. The dominant background contributions arise from QCD multijet production in the fully-hadronic channel and from $\mathrm{t\bar{t}}$+jets production in all channels. The latter includes $\mathrm{t\bar{t}}$ + light-flavor jets and $\mathrm{t\bar{t}}$ + heavy flavor jets. The $\mathrm{t\bar{t}}$+bb background, in particular, is irreducible. The sensitivity of the analysis is also limited by a combinatorial background due to multiple b jets in the final state, and no unambiguous m(jj) is reconstruction possible, so multivariate analysis techniques are necessary.

Events are selected in each channel based the number of isolated leptons and the number of jets and b-tagged jets. Within each channel, events are further categorized based on the jet and b-tagged jet multiplicity. Multivariate discriminators are trained for each channel and category. The multivariate discriminators used are boosted decision trees (BDT) for the dilepton channel, DNNs for the semileptonic channel, and the Matrix Element Method (MEM) output in the fully hadronic channel. The MEM, based on the evaluation of the leading-order $\mathrm{t\bar{t}H(b\bar{b})}$ and $\mathrm{t\bar{t}b\bar{b}}$ matrix elements, is used as input also for the BDTs and the DNNs in the other channels. The signal is extracted in a simultaneous binned-likelihood fit of the discriminators in all the categories, where the categories with low signal purity are useful for constraining backgrounds. Backgrounds are modelled using simulated samples as in VH($\mathrm{b\bar{b}}$) and using data-driven methods for the QCD multijet production only.

The post-fit distributions are shown in figure 4 for the most sensitive categories. The fit results in combined significance of 3.7 (2.6) σ, corresponding to $\mu = 1.49 \pm 0.42$. The result constitutes the first evidence for the $\mathrm{t\bar{t}H(b\bar{b})}$ process. In combination with the previous result using 2016 data, a significance of 3.9 (3.5) σ is obtained, with $\mu = 1.15 \pm 0.30$, as summarized in figure 5 (right).

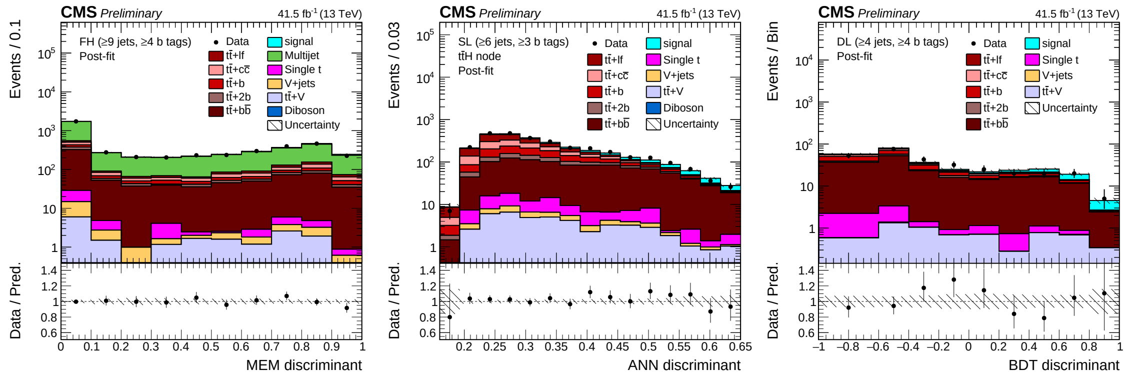

Figure 4: Final discriminators shapes in the categories with the highest sensitivity in fully-hadronic, semi-leptonic, and dilepton channels after the fit to data. The hatched uncertainty bands include the total uncertainty of the fit model. The lower plots show the ratio of the data to the signal+background (post-fit) prediction [3].

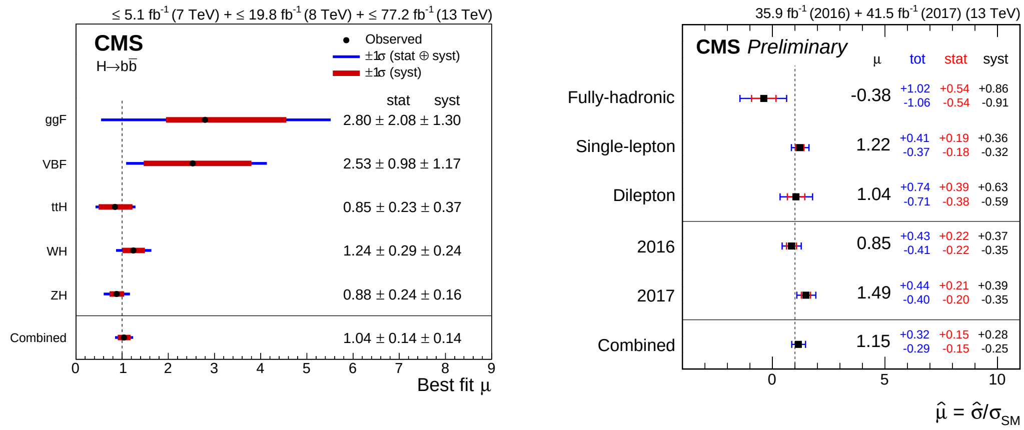

Figure 5: Best fit value of the H($b\bar{b}$) signal strength μ with its 1σ systematic (red) and total (blue) uncertainties by production mode and combined (left). The fit comes from the H($b\bar{b}$) obsevation paper [2]. Best fit values of the signal strength μ obtained in the combined $t\bar{t}$H($b\bar{b}$) fit of the 2016 and 2017 datasets per channel and in combination (right) [3].

The total uncertainty is dominated by systematic components. The largest contributions originate from the theoretical uncertainties, where the $t\bar{t}$ + heavy flavor jets modelling uncertainties give a major contribution.

4. Conclusions

The analysis of 2017 CMS data allowed for a significant progress in the sensitivity to both the VH($b\bar{b}$) and $t\bar{t}$H($b\bar{b}$) processes. Both analyses were improved with respect to the previous works. As both analyses rely on multivariate techniques, DNNs were introduced and applied successfully to most of the final states.

The analysis targeting the associated production of a Higgs boson and a vector boson (VH), with $\mathrm{H} \to \mathrm{b\bar{b}}$, yields an observed significance of 3.3σ above the background-only hypothesis, with an expected significance of 3.1σ. The corresponding measured signal strength is μ = 1.08 $\pm$ 0.34. When combined with previous VH($\mathrm{b\bar{b}}$) measurements, the observed (expected) significance is 4.8 (4.9) σ, corresponding to μ = 1.01 $\pm$ 0.22. In combination with searches for $\mathrm{H} \to \mathrm{b\bar{b}}$ in other production modes by CMS, including $\mathrm{t\bar{t}H(b\bar{b})}$ with 2016 data, results in a significance of 5.6 (5.5) σ, and $\mu = 1.04 \pm 0.20$. With this result, the goal of observing of the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay is achieved by the CMS Collaboration.

In the analysis targeting the $\mathrm{t\bar{t}H(b\bar{b})}$ a significance for this production mode of 3.7 (2.6) σ is obtained, corresponding to $\mu = 1.49 \pm 0.42$. The result constitutes the first evidence for the $\mathrm{t\bar{t}H(b\bar{b})}$ process. In combination with the previous result using 2016 data, a significance of 3.9 (3.5) σ is obtained, with $\mu = 1.15 \pm 0.30$.

References

[1] CMS Collaboration, *Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair*, Phys. Rev. Lett. **120** (2018) 071802 [1709.05543].

[2] CMS Collaboration, *Observation of Higgs boson decay to bottom quarks*, Phys. Rev. Lett. **121** (2018) 121801 [1808.08242].

[3] CMS Collaboration, *Measurement of* $\mathrm{t\bar{t}H}$ *production in the* $\mathrm{H} \to \mathrm{b\bar{b}}$ *decay channel in* $41.5\,\mathrm{fb}^{-1}$ *of proton-proton collision data at* $\sqrt{s} = 13\,\mathrm{TeV}$, *https://cds.cern.ch/record/2675023* (2019) .

[4] CMS Collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[5] ATLAS Collaboration, *Evidence for the* $H \to b\bar{b}$ *decay with the ATLAS detector*, *JHEP* **12** (2017) 024 [1708.03299].

[6] CMS Collaboration, *Evidence for the Higgs boson decay to a bottom quark-antiquark pair*, Phys. Lett. **B780** (2018) 501 [1709.07497].

[7] ATLAS Collaboration, *Search for the standard model Higgs boson produced in association with top quarks and decaying into a* $b\bar{b}$ *pair in pp collisions at* $\sqrt{s}$ = *13 TeV with the ATLAS detector*, Phys. Rev. **D97** (2018) 072016 [1712.08895].

[8] CMS Collaboration, *Search for* $\mathrm{t\bar{t}H}$ *production in the* $\mathrm{H} \to \mathrm{b\bar{b}}$ *decay channel with leptonic* $\mathrm{t\bar{t}}$ *decays in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **03** (2019) 026 [1804.03682].

[9] CMS Collaboration, *Search for* $\mathrm{t\bar{t}}$*H production in the all-jet final state in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **06** (2018) 101 [1803.06986].

[10] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. **B784** (2018) 173 [1806.00425].

[11] CMS Collaboration, *Observation of* $\mathrm{t\bar{t}}$*H production*, Phys. Rev. Lett. **120** (2018) 231801 [1804.02610].

[12] CMS Collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *JINST* **13** (2018) P05011 [1712.07158].

[13] CMS Collaboration, *Supplementary material of CMS-HIG-18-016*, *http://cms-results.web.cern.ch/cms-results/public-results/publications/HIG-18-016/index.html* .

PoS(EPS-HEP2019)342