

ttH production at 13 TeV (CMS)

Milos Dordevic¹ on behalf of the CMS Collaboration

¹ Vinca Institute of Nuclear Sciences (University of Belgrade), PO Box 522, 11001 Belgrade, Serbia

Abstract. First results of a search for the Standard Model Higgs boson produced in association with a top quark-antiquark pair (ttH) in proton-proton collisions at a center of mass energy of 13 TeV is presented. The decays of Higgs boson into two photons, bottom quark-antiquark pair and multileptons via WW, ZZ and tautau pairs are analysed separately and then combined. The results are presented in terms of the best-fit signal strength relative to the Standard Model prediction and the observed and expected upper limits at the 95% confidence level are extracted.

1 Introduction

The discovery of the Higgs boson at CERN [1, 2] is one of the major achievements in the history of particle physics and now the focus is on the precise measurement of its properties, in particular its couplings to other particles. The top quark and Higgs boson are the heaviest particles yet discovered and the leading experiments of today, such as ATLAS and CMS at CERN, are able to provide a direct measurement of their coupling through the study of the associated production in the ttH process. The indirect measurement of the top-Higgs coupling is also possible from the leading Higgs production mechanism, gluon-gluon fusion, with the assumption of no additional particles (beyond SM) contributing to the interaction loop. Thus, a direct measurement of top-Higgs coupling in the associated production can set constraints on the new physics with respect to the $gg \rightarrow$ Higgs process. The top quark is the most strongly coupled to the Higgs boson of all the SM particles and the top-Higgs Yukawa coupling is predicted to be close to unity. It is the top quark that is expected to be the main responsible for the instability of the Higgs mass against the radiative corrections and it could assume a key role in the electroweak symmetry breaking (EWSB). The top-Higgs Yukawa coupling is also a probe for new physics beyond the SM, tightly related to the EWSB. The higher dimension operators of the top and Higgs field are little tested and particularly sensitive to the new physics associated to the EWSB.

The combined result of the search for $t\bar{t}H$ process, using the 7 and 8 TeV data from LHC Run I, provided the observed significance of 4.4σ and the expected significance of 2.0σ , that is obtained from a combined fit to all the analysed channels, the $t\bar{t}H(\gamma\gamma)$, $t\bar{t}H(b\bar{b})$ and $t\bar{t}H(\text{multilepton})$.

This report presents the first study of ttH production at 13 TeV from the CMS Collaboration, using the dataset recorded in 2015 by the CMS experiment [3]. Three different final states are considered, corresponding to the decays of Higgs boson to pair of photons, pair of bottom quarks or the multilepton final state from the decays to WW, ZZ and $\tau\tau$. These processes, although very rare, represent some of the cleanest signatures for studying the associated top-Higgs production. In addition to its

topology, the ttH cross section increases by a factor of 4 when moving from 8 TeV to 13 TeV of center of mass energy, while the main backgrounds increase only by a factor of 3, thus enhancing the sensitivity at the new collision energy. After being analysed separately, the combination of the results of these three signatures is performed.

2 Analysis and Results

The three ttH analysis at 13 TeV are performed separately, but they have a common dominant background process being the $t\bar{t}$ production in association with other particles (vector bosons and jets). For each of the analysis the irreducible and reducible background yields are dominating the signal after the final selection. The events are further categorised to achieve a better signal to background ratio.

2.1 $t\bar{t}H(\gamma\gamma)$

The $H \rightarrow \gamma\gamma$ channel has a small branching ratio of 0.2%, but offers a very clean final state [5]. The excellent photon identification and energy resolution enables a precise reconstruction of the diphoton invariant mass peak. In the inclusive $H \rightarrow \gamma\gamma$ analysis [5] the events are categorised according to the production mode and the Vector Boson Fusion (VBF) and ttH production are considered. In addition to the inclusive gluon fusion process, in focus here are the diphoton events accompanied by two bottom quarks and two W bosons originating from the decay of top quark pair. Aside of the two photons, the final state may contain leptons and additional jets. The analysis used the $2.7fb^{-1}$ of the CMS data, preselected with diphoton triggers.

The main backgrounds are the irreducible $t\bar{t} + \gamma\gamma$ and $t\bar{t} + \gamma + jet$ which is the reducible component and where the jets are mis-identified as photons. One of the crucial aspects of the analysis is to suppress the mis-identified (fake) photons. A dedicated photon reconstruction and energy calibration is performed, but also the association to the primary vertex is made in the high pileup environment. The events are categorised according to the presence of at least one lepton ("leptonic tag") or no leptons are reconstructed at all ("hadronic tag"). At least two jets of which at least one b-tagged is required in the "leptonic tag" category, while at least five jets with at least one b-tagged are selected for the "hadronic tag" category. The photon identification is performed combining the different information into the Boosted Decision Tree (BDT). Another BDT is used to further select the events with good diphoton mass resolution in each of the two categories.

The signal extraction strategy is the same as in the inclusive $H \rightarrow \gamma\gamma$ analysis, where a diphoton mass resonance is being searched for on top of a non-resonant smoothly falling background. The analytic function is used to fit the signal events in each category and for each simulated Higgs mass point. Parametric fit functions with several functional forms, where a large set of function families are treated as a discrete parameter in a likelihood fit, are used to model the background.

The signal model for the Higgs boson mass of 125 GeV is shown in Fig. 1 for the "hadronic tag" category, while the data with signal plus background model fit is shown in Fig. 2 for the same category. The 1σ (green) and 2σ (yellow) bands are shown for the background component of the fit and include the uncertainty of the fit parameters. A small excess of events around the Higgs mass is observed in the "hadronic tag" category, while for the "leptonic tag" the analysis is still statistically very limited and only three events are passing the final selection, while none of them correspond to the mass of the Higgs boson. The observed signal strength obtained by combining the two categories is $3.8_{-3.6}^{+4.5}$ for the $m_H = 125.09GeV$.

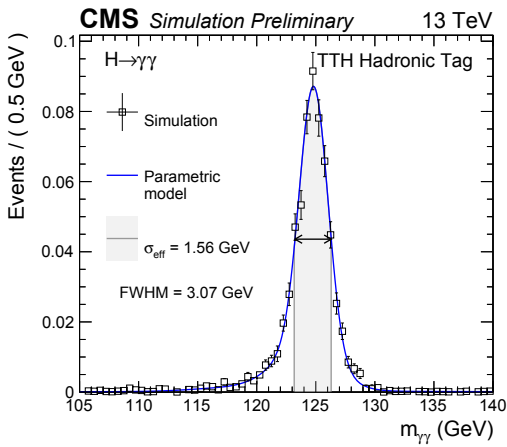


Figure 1. The signal shape parametrization for the "hadronic tag" and Higgs boson mass of 125 GeV. The black points with the uncertainties represent weighted simulation events, while the blue line is the corresponding parametric model.

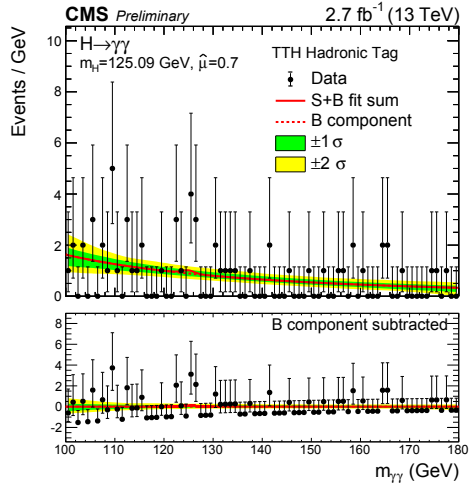


Figure 2. Diphoton mass distribution for the "hadronic tag" category, with the data (black points) and signal plus background fit results (red line), for $m_H = 125.09\text{GeV}$. The pull distribution is obtained after the background is subtracted.

2.2 $t\bar{t}H(b\bar{b})$

This final state profits from the largest branching ratio of 0.56 ± 0.02 for the Higgs boson of 125 GeV [6]. The most dominant background process is the $t\bar{t} + jets$, of which the $t\bar{t} + b\bar{b}$ (with two b-tagged jets) is an irreducible component and it is experimentally very challenging. The limited mass resolution for $H \rightarrow b\bar{b}$ and the presence of jets with similar kinematical properties makes it difficult to assign them properly to the $t\bar{t}$ system or to the Higgs boson itself. The jets that are associated with the top quark pairs are split according to their flavour and are treated independently in the analysis, as different subsamples stem from different physics processes thus having different systematic uncertainties.

The events are categorised in order to enhance the presence of the signal over the dominating background contributions. Two main categories are defined: "lepton+jets" having exactly one lepton in the final state and "dilepton" with exactly two leptons of the opposite sign. The "lepton+jets" category benefits from higher statistics, while still being able to trigger on the presence of one lepton and suppress the QCD background. In the "dilepton" category the contribution of the backgrounds other than $t\bar{t}$ is minimal and the jet combinatorics is reduced as well.

The signal $t\bar{t}H$ events usually have more (b-tagged) jets than the background and thus the events are further categorised based on the number of jets and number of b-tagged jets. As a new feature with respect of the 8 TeV analysis [6], the category with boosted jets from hadronic decays of top quark or $H \rightarrow b\bar{b}$ with large transverse momenta is introduced. The "lepton+jets" category is split into eight subcategories, while the "dilepton" candidate events are further separated into five subcategories. In each of these categories, the machine learning (BDT) and physics motivated (matrix element - MEM) methods, or their combination, are used to achieve the best separation of the signal from the background processes. For example, in the category with 1 lepton, at least 6 jets of which 3 are b-tagged, the BDT is trained using the MEM discriminator as one of its input variables and the output

distribution is shown in Fig. 3. The category with 1 lepton, at least 4 jets of which all 4 are b-tagged is the one for which the MEM is the most suitable and has the best performance, as shown in Fig. 4.

The simultaneous binned maximum-likelihood fit to data in all categories is performed. In the "lepton+jets" category the signal strength is $-4.7^{+3.7}_{-3.8}$, while in the "dileptons" category it is -0.4 ± 2.1 . The combination of the two categories yield with the signal strength of -2.0 ± 1.8 and this result is 1.7σ below the SM expectation, which is compatible with the assumption of the absence of signal.

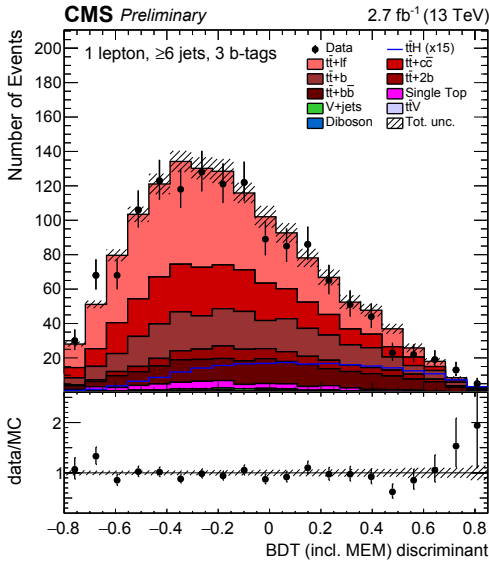


Figure 3. The distribution of the BDT output that includes also the MEM discriminator as one of its input variables, for the 1 leptons, at least 6 jets of which at least 3 are b-tagged. The expected background contributions are normalised to the integrated luminosity and stacked.

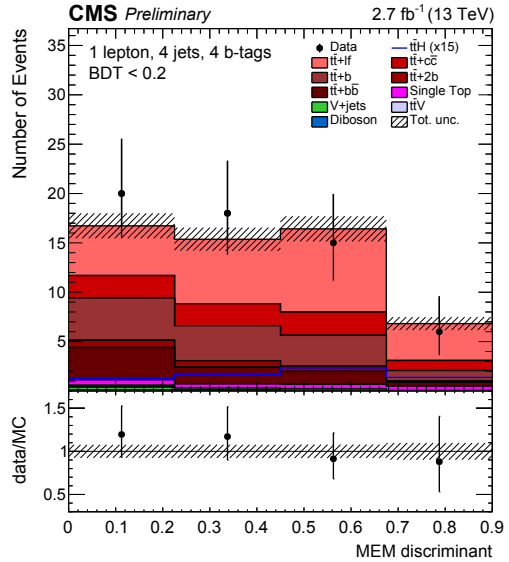


Figure 4. The distribution of the MEM discriminator used exclusively in 1 lepton, at least four jets (all b-tagged) category. The background histograms are stacked and normalised to the luminosity of $2.7 fb^{-1}$.

2.3 $t\bar{t}H(\text{multilepton})$

This analysis considers final states from Higgs decays to WW , ZZ or $\tau\tau$ where at least one of the vector bosons or taus decays to leptons [7]. This channel has a small branching ratio but the presence of the additional one or two leptons from top quark decays leads to clean experimental signatures: two same-sign leptons (electrons or muons) or at least three leptons plus b-tagged jets. There are two main background components: $t\bar{t}W$ and $t\bar{t}Z$ that is irreducible and is estimated from the MC and the reducible $t\bar{t} + jets$ estimated using a dedicated fake rate method from data. The fake rate is estimated from QCD multijet and Z+jets events. Control regions are defined to model the non-prompt lepton background by relaxing the lepton identification requirement and then the events are weighted by the function of the mis-identification probability of the leptons. Charge mis-identification for electrons is evaluated in a sample of electrons from Z decays, that is split in the opposite- and same-sign pairs, where the charge mis-reconstruction probability is measured as a function of $|p_T|$ and $|\eta|$. This probability ranges from 0.03% in barrel and 0.4% in the endcap region of the CMS detector.

The events are preselected with single, double or triple lepton (electron or muon) triggers. At least four jets are required in the same-sign dilepton category; for the three or more leptons event category at least two jets are required. The events are further categorised according to the lepton flavour, by the presence of hadronically decaying tau leptons and by the b-tagging identification criterion. Two BDTs are trained in order to maximise the presence of signal with respect to the $t\bar{t}$ and $t\bar{t}W/Z$ processes. The two-dimensional plane spanning the output of the two BDT trainings is divided into several bins and the signal and background content of each bin is folded to the one-dimensional histogram. The signal extraction is performed by a fit to the distribution of events among these bins. The Fig. 5 and Fig. 6 show the distribution of the output of two-dimensional BDT in the two same-sign and at least three leptons categories, respectively, after the final event selection.

The best fit of the signal strength in the two same-sign lepton and three lepton category is $-0.5^{+1.0}_{-0.7}$ and $5.8^{+3.3}_{-2.7}$, respectively. The combination of the two categories gives a value of $0.6^{+1.4}_{-1.1}$, in agreement with the SM prediction. Slight deficit of events and small excess is observed in the two same-sign and tripleton category, yet not significant for any indication of possible deviation from the SM expectation.

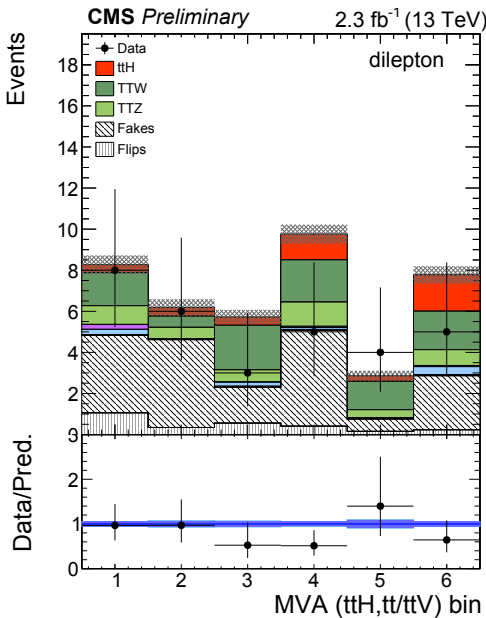


Figure 5. Distribution of the final BDT discriminator, obtained after combining the two BDT trainings against $t\bar{t}$ and $t\bar{t}W/Z$ processes, shown for the two same-sign leptons category.

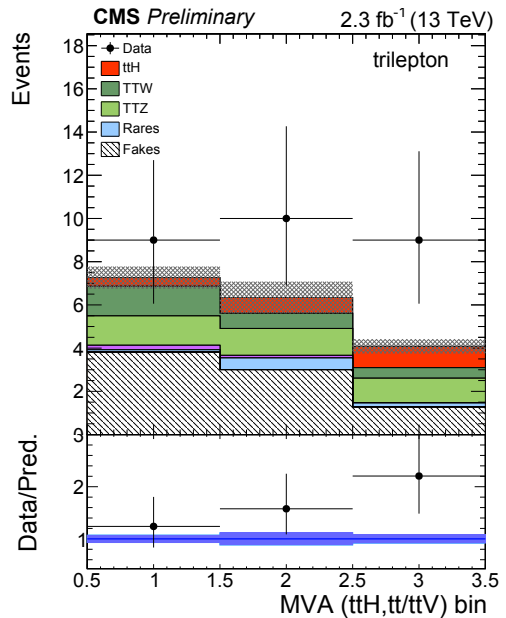


Figure 6. Distribution of the final BDT discriminator, obtained after combining the two BDT trainings against $t\bar{t}$ and $t\bar{t}W/Z$ processes, shown for the at least three leptons category.

2.4 Combination

The three analysis are combined assuming the mass of Higgs boson at 125 GeV and taking into account the correlation of the common systematic uncertainties. The best fit of the signal strength of $0.15^{+0.95}_{-0.81}$ is obtained, in agreement with the SM prediction of $1.00^{+0.96}_{-0.81}$, as shown in Fig. 7. The 95% CL limits on the signal strength in each of the three analysis and also their combination, that has an observed (expected) value of 2.1 (1.9), are shown in Fig. 8.

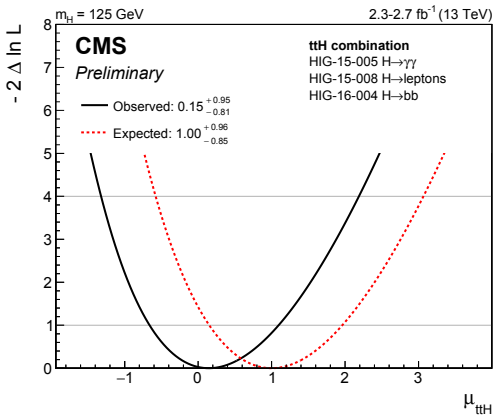


Figure 7. The measurement of the combined signal strength of the $t\bar{t}H$, where the Higgs boson decays to WW , ZZ and $\tau\tau$ are considered.

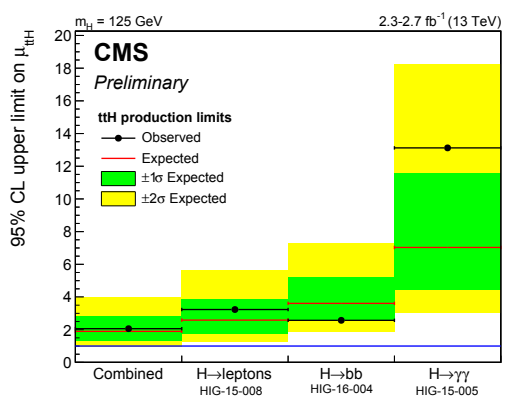


Figure 8. The 95% CL upper limits on the rate of $t\bar{t}H$ production, relative to SM expectation of 1, for each decay mode and the combination.

2.5 Conclusion

The first results of studying the $t\bar{t}H$ production at 13 TeV using the 2015 data recorded by the CMS experiment at the center of mass energy of 13 TeV have been presented in this report. Three independent channels are analysed separately: $t\bar{t}H(\gamma\gamma)$, $t\bar{t}H(b\bar{b})$ and $t\bar{t}H(multilepton)$ and then combined. The combination of the three analyses results with a best fit to the signal strength of $0.15^{+0.95}_{-0.81}$, in agreement with the SM. Results of the individual channels, as well as the combination, are comparable to the Run I in terms of the signal sensitivity, although almost an order of magnitude less data is used, resulting mainly from the increase of the $t\bar{t}H$ production cross section by a factor of 4 at the higher center of mass energy, but also from the various improvements and optimization in each of the analyses.

References

- [1] The ATLAS Collaboration, Phys. Lett. B 716, 1 (2012).
- [2] The CMS Collaboration, Phys. Lett. B 716, 30 (2012).
- [3] The CMS Collaboration, JINST 3 S08004 (2008).
- [4] The ATLAS and CMS Collaborations, ATLAS-CONF-2015-044, CMS-PAS-HIG-15-002.
- [5] The CMS Collaboration, CMS-PAS-HIG-15-005 (2016).
- [6] The CMS Collaboration, CMS-PAS-HIG-16-004 (2016).
- [7] The CMS Collaboration, CMS-PAS-HIG-15-008 (2016).
- [8] CMS public webpage: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/TTHCombMoriond2016>.