

Studies of the $t\bar{t}H$ production at 13 TeV

Predrag Cirkovic^{1,a} on behalf of the CMS Collaboration

¹*Institute of Physics Belgrade, University of Belgrade*

Abstract. The latest results of searches for the standard model Higgs boson produced in association with a top quark-antiquark pair ($t\bar{t}H$), where Higgs decays into photons, bottom quark-antiquark pair or leptons via WW , ZZ and $\tau\tau$ are presented. The analyses have been performed using the 13 TeV pp collisions data recorded by the CMS experiment in 2015 and part of 2016. The results are presented in the form of the best fit to the signal strength ($\mu = \sigma/\sigma_{SM}$) measured with respect to the Standard Model prediction and its expected and observed 95% CL upper limits.

1 Introduction

Following the discovery of a Higgs boson at the LHC, exploring its properties such as production and decay rates, as well as its couplings with other Standard Model (SM) particles has become the main focus of interest. Higgs boson mass (m_H) has been measured in both experiments CMS and ATLAS and amounts to 125.09 GeV [1]. Out of the four main production mechanisms (gluon-gluon fusion - ggH, vector boson fusion - VBF, Higgs associated to a vector boson - VH and Higgs in association with a $t\bar{t}$ pair - $t\bar{t}H$), this search focuses on the process which involves coupling of the Higgs boson to fermions ($t\bar{t}H$). The final signal strength measurement is obtained combining the results of measurements in several Higgs decay channels: $H \rightarrow b\bar{b}$ (which occurs in approximately 58% of the cases), $H \rightarrow WW^*$ (in 20% of the cases), and $H \rightarrow \tau^+\tau^-$, ZZ and $\gamma\gamma$ (in less than 10%). Direct measurement of top Yukawa coupling is performed in channels where Higgs is associated with $t\bar{t}$ pair or a single top quark. Since this particular channel has a small cross section comparing to gluon-gluon fusion, where this coupling cannot be measured directly, one can use the results obtained with direct measurement to put constrains on contributions from Beyond the Standard Model (BSM) processes.

2 Previous results and motivation

The analysis of the $t\bar{t}H$ production has been already performed with LHC Run I data in both experiments (CMS and ATLAS), and the relevant signal strength measurements are listed in the Tab. 1.

Comparing the signal strength measurements obtained for each of the Higgs production mechanisms with Run I CMS and ATLAS data [2], one can see that $t\bar{t}H$ signal strength has the largest uncertainties comparing to the other production modes, mostly due to the low cross section of the $t\bar{t}H$ process. First 13 TeV combined measurement of the signal strength and its 95% confidence level upper limit obtained with 2015 CMS data is shown on the Fig. 2.

^ae-mail: predrag.cirkovic@cern.ch

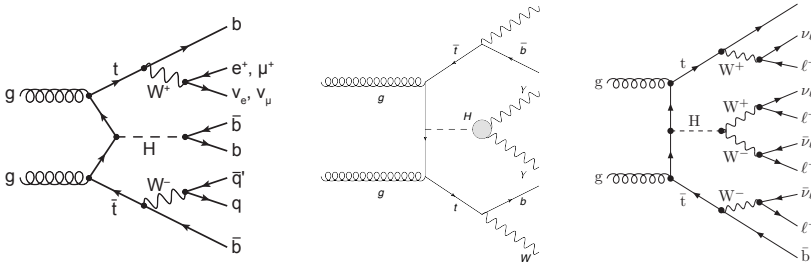


Figure 1. Feynman diagrams showing the $t\bar{t}H \rightarrow b\bar{b}$ decay that includes subsequent decays of the top quark-antiquark pair in the lepton+jets channel (left), the $t\bar{t}H$ with the Higgs boson decaying into a pair of photons, i.e. $t\bar{t}H \rightarrow \gamma\gamma$ (middle) and one possible Feynman diagram for $t\bar{t}H$ production, where the Higgs boson decays to WW^* , producing the final state with four leptons (right).

Table 1. CMS and ATLAS $t\bar{t}H$ signal strength measurements obtained using the LHC Run I data.

Run I	μ (CMC)	μ (ATLAS)
$H \rightarrow b\bar{b}$	$0.7^{+1.9}_{-1.9}$	$1.5^{+1.1}_{-1.1}$
$H \rightarrow \gamma\gamma$	$2.7^{+2.6}_{-1.8}$	$1.4^{+2.2}_{-1.4}$
$H \rightarrow$ leptons	$3.7^{+1.7}_{-1.4}$	$2.1^{+1.4}_{-1.2}$
Combined	$2.9^{+1.0}_{-0.9}$	$1.9^{+0.8}_{-0.7}$
Total	$2.3^{+0.7}_{-0.6}$	

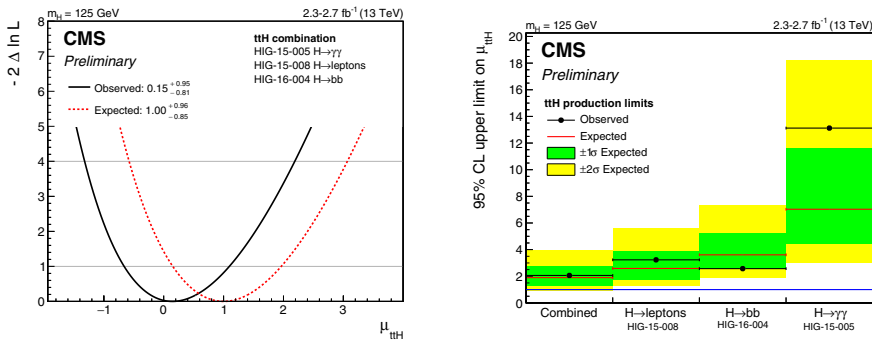


Figure 2. Combined signal strength measurement of the $t\bar{t}H$ production process, $\mu_{t\bar{t}H}$, via decays of the Higgs boson to $b\bar{b}$ [3], $\gamma\gamma$ ($t\bar{t}H$ -tag categories only) [5] and multilepton [7] using $2.3 - 2.7\text{fb}^{-1}$ of data collected at $\sqrt{s} = 13$ TeV (left). Corresponding 95% confidence level upper limits on the rate of $t\bar{t}H$ production relative to the SM expectation, $\mu_{t\bar{t}H} = 1$ (right).

3 $t\bar{t}H(H \rightarrow b\bar{b})$

The strategy of the $t\bar{t}H$ with $H \rightarrow b\bar{b}$ analysis is based on selection of events compatible with $H \rightarrow b\bar{b}$ and one of the $t\bar{t}$ decays: lepton+jets or dilepton. Selected events are then categorized by number of jets and number of b-tags into 13 distinct categories with different background compositions (Fig. 3 and 4). The next step is to build discriminators for each of the categories and for this pur-

pose two different approaches are used: physics motivated matrix element method (MEM) and a machine learning method based on the boosted decision tree algorithm (BDT). The main reason why the MEM is used is that some of the hadronically decaying background components, such as $t\bar{t} + b\bar{b}$, are difficult to distinguish from the $t\bar{t}H \rightarrow b\bar{b}$ signal. Both methods are used in one of the following ways: pure BDT, MEM included as an input to the BDT, and two-dimensional MEM and BDT. For each of the categories, one of the approaches is used according to the best performance: BDT-only for ≥ 6 jets, 2 b-tags (lepton+jets) and all dilepton categories, 2D MEM-BDT for ≥ 4 jets, 4 b-tags, ≥ 5 jets, ≥ 4 b-tags and ≥ 6 jets, ≥ 4 b-tags and BDT including MEM for the rest of the categories.

On the Fig. 5, one can see what the shape of a discriminator looks like in one of the categories (e.g. lepton+jets category having 5 jets and 3 b-tags). A good agreement between distributions of the prediction of signal plus various background components and the data is achieved. Such distributions are used to extract the signal with a maximum likelihood fit combining data in all the categories.

CMS Simulation

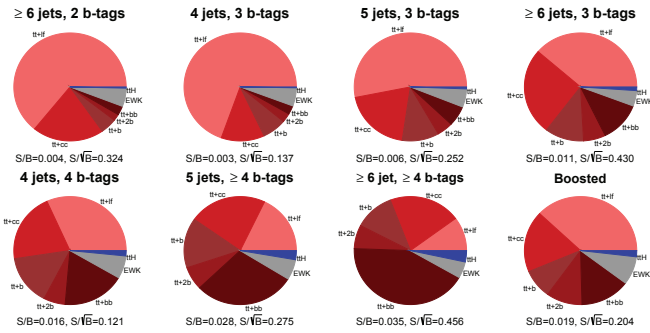


Figure 3. Fraction of processes contributing in the lepton+jets analysis categories [3].

CMS Simulation

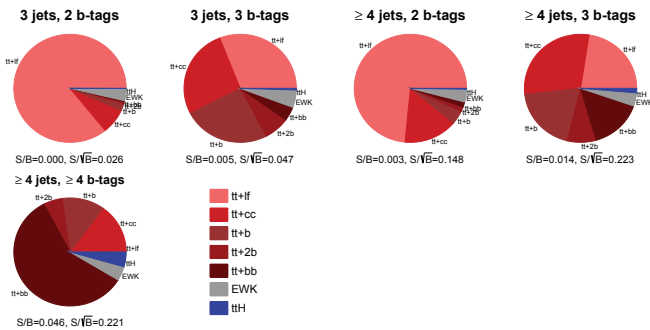


Figure 4. Fraction of processes contributing in the dilepton analysis categories [3].

The results of the $t\bar{t}H$ with $H \rightarrow b\bar{b}$ analysis with first 13 TeV data are presented in form of the measured signal strength for each of the categories and their combination (Fig. 6, left) and the corresponding 95% C.L. upper limits, both expected and observed (Fig. 6, right). The relevant numbers are listed in the Tab. 2.

4 $t\bar{t}H(H \rightarrow \gamma\gamma^{(*)})$

Strategy of the search for $t\bar{t}H$ with $H \rightarrow \gamma\gamma$ is based on the inclusive $H \rightarrow \gamma\gamma$ production measurement [5, 6]. This analysis provides clean final state signatures since it exploits one of the advantages

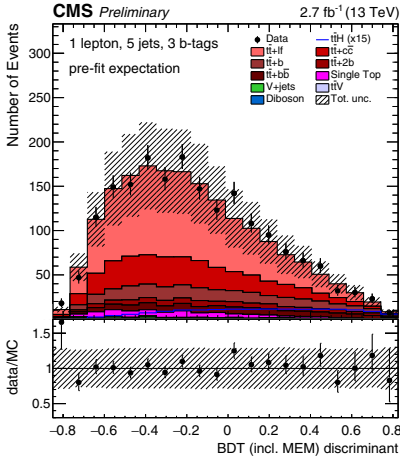


Figure 5. One of the final discriminant shapes, the 5 jets 3 b-tag category, in the lepton+jets channel before the fit to data [3]. The expected background contributions (filled histograms) are stacked, and the expected signal distribution (line) for a Higgs-boson mass of $m_H = 125$ GeV is superimposed. Each contribution is normalized to the integrated luminosity of 2.7 fb^{-1} , and the signal contribution is additionally scaled by a factor of 15 for better readability. The distribution in data (markers) is also shown.

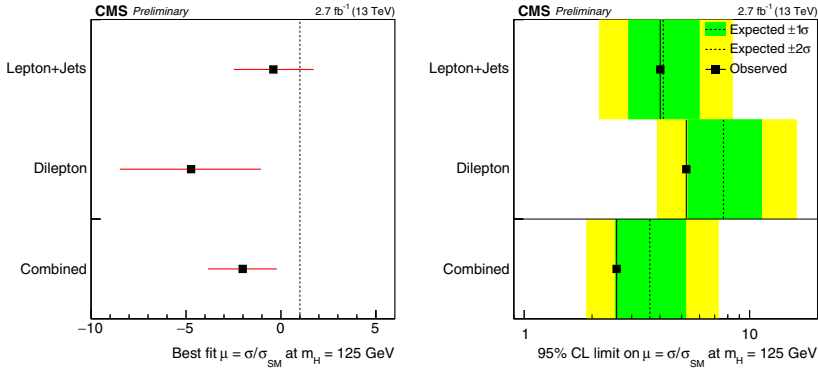


Figure 6. Best-fit values of the signal strength modifiers μ with their $\pm 1\sigma$ confidence intervals (left). Median expected and observed 95% C.L. upper limits on μ (right). The expected limits are displayed together with $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals [3].

Table 2. Best-fit value of the signal strength modifier μ and expected and observed 95% C.L. upper limits (UL) in the lepton+jets and dilepton channels as well as the combined results [3]. The one standard deviation ($\pm 1\sigma$) confidence intervals of the best-fit values and the expected limits are also quoted.

Channel	Best-fit μ	Observed UL	Expected UL
Lepton+jets	$-0.4^{+2.1}_{-2.1}$	4.0	$4.1^{+1.8}_{-1.2}$
Dilepton	$-4.7^{+3.7}_{-3.8}$	5.2	$7.7^{+3.6}_{-2.3}$
Combined	$-2.0^{+1.8}_{-1.8}$	2.6	$3.6^{+1.6}_{-1.1}$

of the CMS detector [4], which is ability to measure diphoton mass with good resolution. Main backgrounds here are $t\bar{t} + \gamma\gamma$ and $t\bar{t} + \text{fake photons}$. Main event categories are chosen according to $t\bar{t}$ decay: semi leptonic and fully leptonic (both merged into one leptonic category) and hadronic $t\bar{t}$ decays. The $t\bar{t}H$ tagger performs with high efficiencies in both hadronic ($\sim 77\%$) and leptonic ($\sim 95\%$) category [6].

Signal is extracted using a parametrized model of the Higgs boson mass shape obtained from simulation, while the combined signal and background models for these two categories are shown on the Fig. 8. The model used to describe the background is extracted from data with the discrete profiling method, which is designed to estimate the systematic uncertainty associated with choosing a particular analytic function to fit the background distribution. The method treats the choice of the background function as a discrete parameter in the likelihood fit to the data. The resulting systematic uncertainty is then calculated in an analogous way to systematic uncertainties associated with other contributions [6].

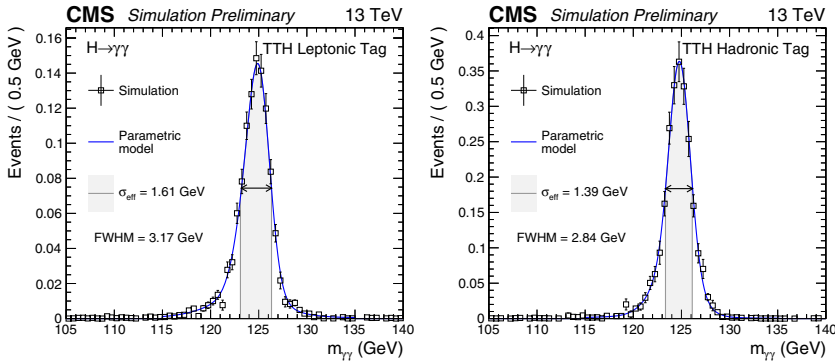


Figure 7. Parametrized signal shapes for the $t\bar{t}H$ category, for a simulated $H \rightarrow \gamma\gamma$ signal sample with $m_H = 125$ GeV [6]. The black points represent weighted simulation events and the blue lines are the corresponding models. Also shown are the σ_{eff} value (half the width of the narrowest interval containing 68.3% of the invariant mass distribution), FWHM and the corresponding interval.

The results of the $t\bar{t}H$ production measurements performed in $H \rightarrow \gamma\gamma$ are summarized in plots shown on Fig. 9.

5 $t\bar{t}H \rightarrow \text{multilepton}$

The third complementary analysis channel is the $t\bar{t}H$ decaying to leptons [7, 8]. The strategy of $t\bar{t}H \rightarrow \text{multilepton}$ is to select events compatible with coupled $t\bar{t}$ and Higgs decaying to a pair of vector bosons (WW^* or ZZ^*), or a pair of τ leptons, producing four, three or two leptons with same sign of the charge in the final state, respectively. Dominant backgrounds are $t\bar{t} + \text{jets}$ (misidentified as prompt leptons) and $t\bar{t}V$. Selection of the signal events is done applying a certain set of criteria, such as: trigger and kinematic requirements, linear discriminant based on MET, charge quality, Z-veto and conversion veto, as well as the jet and b-tag multiplicity requirements.

Multivariate method for lepton identification (lepton MVA) is used to identify the leptons from $t\bar{t}$ and Higgs decays (i.e. prompt leptons), and final 2D MVA is used for signal extraction. 2D BDT discriminant is obtained using simulated $t\bar{t}H$ trained against two main background components: $t\bar{t}$ and

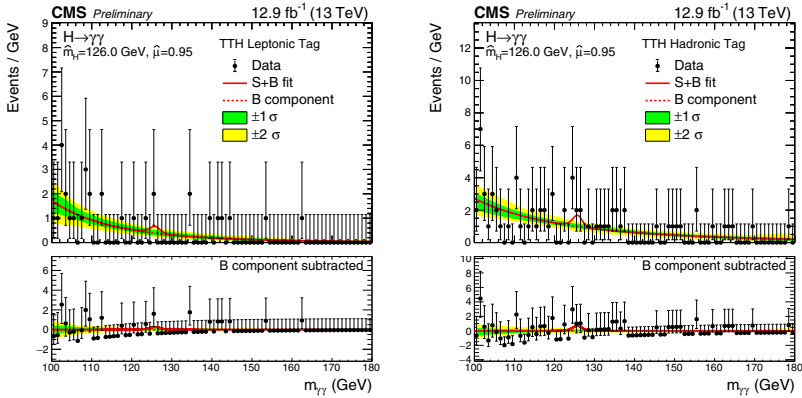


Figure 8. Data points (black) and signal plus background model fits $t\bar{t}H$ category is shown [6]. The one standard deviation (green) and two standard deviation bands (yellow) include the uncertainties of the fit. The bottom plots show the residuals after background subtraction.

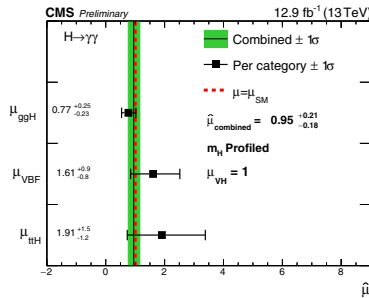


Figure 9. Signal strength modifiers measured for each process (black points) for profiled m_H , compared to the overall signal strength (green band) and to the SM expectation (dashed red line) [6].

$t\bar{t}V$. Two main event categories are two lepton same sign ($2lss$) and at least three leptons ($\geq 3l$). In order to gain sensitivity of the analysis, an additional categorization is performed by lepton flavour (broken down to ee , $e\mu$ and $\mu\mu$ in $2lss$ final state), presence of 2 b-tags ("b-tight" representing the category that has two b-jets in the final state and the complementary "b-loose" category), presence of reconstructed hadronic τ leptons (in $2lss$) and the total lepton charge (motivated by the charge asymmetry present in several background processes), which ultimately amounts to 15 distinct categories in total.

Signal extraction is performed using final discriminators obtained with 2D MVA approach. The MVA outputs, separately trained against $t\bar{t}$ and $t\bar{t}V$ are used to build two dimensional BDT discriminator, which is further split into bins according to minimal uncertainty on the signal strength. The MVA takes into account several kinematic variables (such as maximum pseudorapidity of the leading and sub-leading leptons), jet multiplicity, angular separation between lepton and jet and between jets, and missing transverse energy.

In addition to that, MEM approach has been recently introduced, in order to obtain better discrimination power between $t\bar{t}H$ signal and irreducible $t\bar{t}W$ and $t\bar{t}Z$ backgrounds in category with at least three leptons. 2D MVA has been performed separately in $2lss$ and $\geq 3l$, postfit plots (Figs. 10 and 11) shown good agreement between the data and the MC distributions. The final results are obtained performing simultaneous fit on the binned BDT output distributions.

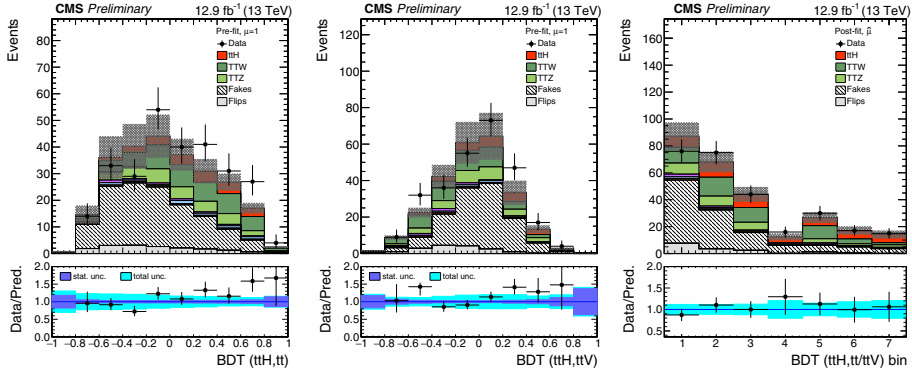


Figure 10. Same-sign dilepton channel: distributions of the BDT classifier outputs shown before the signal extraction fit ($\times 2$), trained against $t\bar{t}$ (left) and $t\bar{t}V$ (middle). Post-fit distributions and uncertainties of the combination of the BDT classifier outputs in the bins used for signal extraction (right) [8].

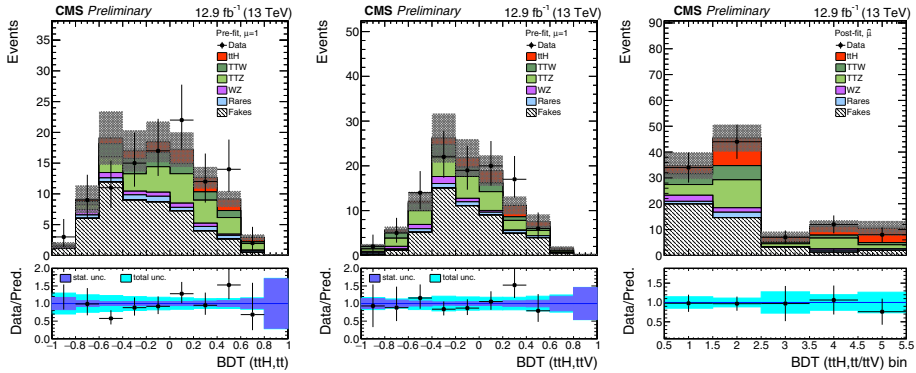


Figure 11. At least three leptons channel: distributions of the BDT classifier outputs shown before the signal extraction fit ($\times 2$), trained against $t\bar{t}$ (left) and $t\bar{t}V$ (middle). Post-fit distributions and uncertainties of the combination of the BDT classifier outputs in the bins used for signal extraction (right) [8].

The final results obtained in $t\bar{t}H$ to multilepton are presented in Fig. 12 and the corresponding numbers are listed in Tab. 3. The numbers represent the signal strength values and their observed and expected limits for each of the final state categories and their combinations. First combination is given for 2016 data only, and the second one for 2015 and 2016 combined.

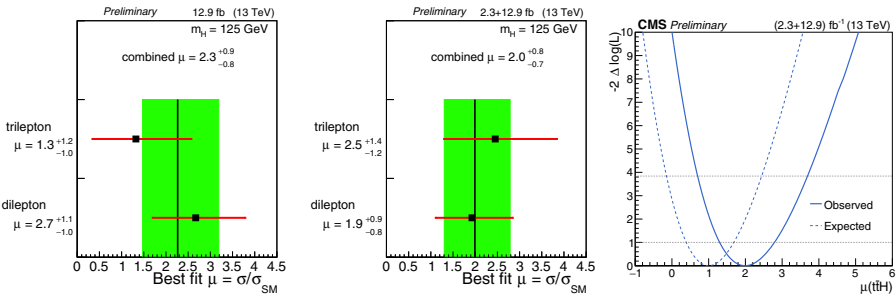


Figure 12. Best fit signal strength for the 2016 analysis alone (left) and the combined 2015 and 2016 analysis (middle), in the dilepton and trilepton channels. Negative log-likelihood as a function of the signal strength for the combined 2015 and 2016 result, as observed in the data and expected from simulation (right) [8].

Table 3. Observed and expected asymptotic 95% C.L. upper limits on and best fit value of the signal strength parameter (μ) [8].

Category	Obs. limit	Exp. limit $\pm 1\sigma$	Best fit $\mu \pm 1\sigma$
Same-sign dileptons	4.6	$1.7^{+0.9}_{-0.5}$	$2.7^{+1.1}_{-1.0}$
Trileptons	3.7	$2.3^{+1.2}_{-0.7}$	$1.3^{+1.2}_{-1.0}$
Combined categories	3.9	$1.4^{+0.7}_{-0.4}$	$2.3^{+0.9}_{-0.8}$
Combined with 2015 data	3.4	$1.3^{+0.6}_{-0.4}$	$2.0^{+0.8}_{-0.7}$

6 Summary

Analysis has been performed in order to measure $t\bar{t}H$ production rate at the center of mass energy of 13 TeV, targeting $t\bar{t}H$, $H \rightarrow b\bar{b}$ with 2015 CMS data only, and $t\bar{t}H$, $H \rightarrow \gamma\gamma$ and $t\bar{t}H \rightarrow$ multilepton with 2015 and part of 2016 CMS data. The results, both separate and combined, are compatible with the SM predictions within the present uncertainties. The goal of Run II is to perform top Yukawa coupling measurement, which requires collecting higher integral luminosity data. Excellent operation conditions of both LHC and CMS, followed by an increasing trend in data taking have been preserved during 2016 with the objective to collect more than 30 fb^{-1} by the end of the year.

References

- [1] CMS Collaboration, JHEP **1306**, 081 (2013)
- [2] CMS Collaboration, CMS PAS **HIG-16-022**, (2016)
- [3] CMS Collaboration, CMS PAS **HIG-16-004**, (2016)
- [4] CMS Collaboration, JINST **3** S08004 (2008)
- [5] CMS Collaboration, CMS PAS **HIG-15-005**, (2015)
- [6] CMS Collaboration, CMS PAS **HIG-16-020**, (2016)
- [7] CMS Collaboration, CMS PAS **HIG-15-008**, (2015)
- [8] CMS Collaboration, CMS PAS **HIG-16-022**, (2016)