PROCEEDINGS OF SCIENCE

Measurements of $t\bar{t}H$ and tH associated production at CMS

Sergio Sánchez Cruz[a,b,*] **(on behalf of the CMS Collaboration)**

[a] *Universidad de Oviedo (formerly), Oviedo, Spain*

[b] *University of Zurich, Zurich, Switzerland*

E-mail: sergio.sanchez.cruz@cern.ch

We present measurements of the $t\bar{t}H$ and tH associated production rate performed by the CMS Collaboration in a sample of pp collision events at $\sqrt{s} = 13$ TeV. The analyses are performed in the diphoton and multilepton channels by categorizing events according to the lepton and jet multiplicity, and to multivariate classifiers that discriminate signals from the corresponding background processes. The observed results are consistent with the SM expectations, achieving sensitivities close to five standard deviations with respect to the background-only hypothesis for $t\bar{t}H$ production.

40th International Conference on High Energy physics - ICHEP2020
July 28 - August 6, 2020
Prague, Czech Republic (virtual meeting)

*Speaker

https://pos.sissa.it/

PoS(ICHEP2020)080

1. Introduction

The discovery of the Higgs (H) boson in 2012 by the ATLAS [1] and CMS [2, 3] experiments was one of the main missing pieces to complete the Standard Model (SM) of particle physics. While the properties of the recently discovered particle resemble the expectations by the SM, their precise study remains of paramount importance.

This contribution focuses on the Yukawa interaction (y_t) between the H boson and the the top quark, the heaviest particle in the SM. Although deviations of this interaction from its SM prediction can affect the H boson production rate via gluon fusion and its decay rate into two photons, these interactions are mediated by loop diagrams, which could receive contributions from Beyond SM (BSM) particles that could mask deviations of the top quark Yukawa interaction. The $t\bar{t}H$ and tH processes involve this coupling at tree level, so this effect cannot be present. In tH production, diagrams proportional to the Higgs coupling to the W boson (g_W) and y_t interfere destructively in the SM. However if the sign of one of these couplings is flipped, the interference becomes constructive, enhancing the cross section of this process by an order of magnitude. Therefore the study of tH production provides sensitivity to the sign of y_t.

2. Measurements in the $H \rightarrow \gamma\gamma$ signatures

Despite of the low branching fraction of the Higgs boson decay into photons, diphoton signatures provide a clean final state that allows to resolve the Higgs boson system in $t\bar{t}(H \rightarrow \gamma\gamma)$ candidate events almost without ambiguity, and with little contribution from other SM processes. A measurement of $t\bar{t}H$ production [4] is performed in events with two photons and in categories with and without leptons, targeting the hadronic and semi-leptonic decays of the top quarks. Multivariate methods are used to discriminate between signal and backgrounds. The signal production rate is measured by performing a likelihood fit to the diphoton invariant mass, obtaining a signal strength of $1.38^{+0.36}_{-0.29}$ and an observed (expected) significance of 6.6 (4.7) standard deviations.

3. Measurements in the multilepton signatures

Multilepton signatures show two or more leptons in the final state, and target decay modes of the Higgs boson into WW^*, ZZ^* and $\tau\tau$, and semileptonic or hadronic decays of each one of the top quarks. These signatures are typically characterized by a moderate contribution from other SM backgrounds and signal has a significantly large branching ratio. The analysis described in this contribution is fully documented in [5].

3.1 Event selection

The expected signal signature are final state with multiple leptons, one (two) b jets in tH ($t\bar{t}H$) events and several light jets produced in the top quark or H boson decays. tHq events additionally feature a light flavor jet that may be emitted in the forward direction, due to the presence of the expectator quark. Events in this measurement are selected and classified in ten disjoint signal regions, according to the lepton (e and μ) and semi-hadronically decaying τ (τ_h) multiplicity: $2\ell SS + 0\tau_h$, $3\ell + 0\tau_h$, $2\ell SS + 1\tau_h$, $2\ell OS + 1\tau_h$, $1\ell + 2\tau_h$, $4\ell + 0\tau_h$, $3\ell + 1\tau_h$, $2\ell + 2\tau_h$, $1\ell + 1\tau_h$ and

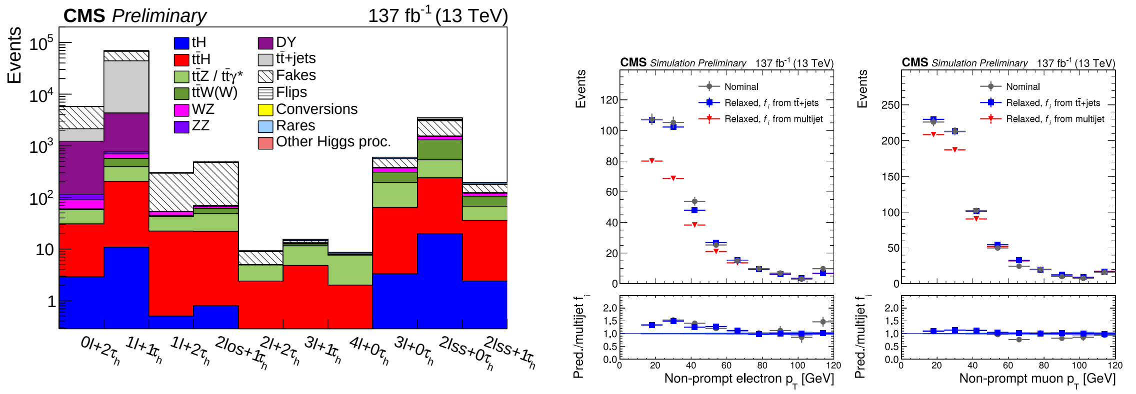

Figure 1: Number of expected signal and background event in each one of the signal regions (left). Transverse momentum distribution of nonprompt electrons (middle) and muons (right) in simulated $\mathrm{t\bar{t}}$ events. The distribution of the nominal samples (gray) is compared to the prediction of the MP method with MPs computed in $\mathrm{t\bar{t}}$ (blue) and multijet (red) simulated samples [5].

$0\ell + 2\tau_h$. The charge of the leptons and τ_h are required to be consistent with the expected final state in $\mathrm{t\bar{t}H}$ and tH events. Additionally, in the $2\ell\mathrm{SS} + 0\tau_h$category the two leptons are required to be of the same sign, in order to profit from this rare SM topology, that enhances the signal-to-background ratio by a large factor, while keeping roughly half of the signal dilepton events. Events are also required to have a jet and b-tagged jet multiplicity consistent with the expected $\mathrm{t\bar{t}H}$ final state. This requirement is relaxed in the $2\ell\mathrm{SS} + 0\tau_h$, $3\ell + 0\tau_h$and $2\ell\mathrm{SS} + 1\tau_h$categories, in which sensitivity to tH production is expected, and all events with a b-tagged jet and a light (possibly forward) jet are accepted. The signal and background contribution to the defined signal regions is shown in Fig. 1, as obtained from the background estimation methods described in Sec. 3.2.

The categories defined have a significant contribution from signal events, but backgrounds are still the dominant contribution. In order to gain sensitivity to the signals, events are classified according to the event topology and by means of multivariate classifiers. A dedicated classifier has been built and optimized for each category with a suitable choice of input variables. These input variables include 3-momenta of the reconstructed objects, angular distances and invariant masses of groups of objects, as well as multivariate discriminators aiming to reconstruct the H boson and top quark decay products.

The $2\ell\mathrm{SS} + 0\tau_h$, $3\ell + 0\tau_h$and $2\ell\mathrm{SS} + 1\tau_h$implement artificial neural networks (ANN) that aim to classify between the different signal and background species. These ANNs feature a number of output nodes that return a quantity that can be interpreted as the probability for an event to have been produced by a given process. In the $3\ell + 0\tau_h$and $2\ell\mathrm{SS} + 1\tau_h$categories the ANNs implemented have three nodes, aiming to discriminate $\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}H}$ events and backgrounds, while in the $2\ell\mathrm{SS} + 0\tau_h$category the ANN employed features a fourth node, to discriminate $\mathrm{t\bar{t}W}$. One sub-category is built for each node of the ANN, containing events for which that score is the highest. Events may then be classified according to the lepton flavor and to the b-tagged jet multiplicity. Finally, events are classified according to the score of the most probable node.

The rest of the categories are not expected to profit from this multi classification approach because they do not have enough tH contribution or they are dominated by background. In those

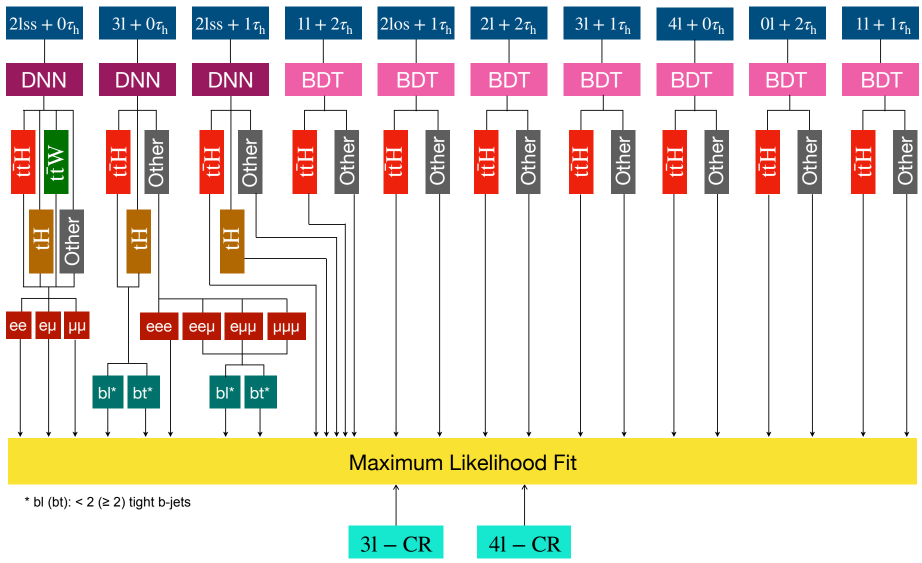

Figure 2: Overview of the categories used for the signal extraction [5].

cases, events are classified according to a boosted decision tree (BDT) that discriminates the $t\bar{t}H$ signal from background. Figure 2 describes the categorization of events employed in the analysis.

3.2 Background estimation

The classification described in the previous section allows to obtain regions purer in signal events. However, a precise background estimation remains crucial for the analysis.

Reducible backgrounds are dominated by the contribution from nonprompt leptons and misidentified τ_h, and electron charge flips. The contribution from these processes in the signal regions is estimated using data driven method. There is a residual contribution from photon conversions to electrons, that is estimated using samples of simulated events.

The contribution from nonprompt leptons is highly suppressed by means of an MVA discriminator developed for this analysis and that makes use of impact parameter, isolation and b tagging variables to discriminate prompt and nonprompt leptons. Misidentified τ_h are also suppressed by employing a selection based in the DeepTau algorithm [6]. Nonpromt leptons and misidentified τ_h are estimated using the misidentification probability (MP) method, that infers their contribution in the signal region from their contribution to a sideband constructed by relaxing the lepton and τ_h selection criteria. The MP for a nonprompt lepton passing these relaxed criteria to pass the selection on the MVA discriminator is determined from a data sample of multijet events. This is equivalently done for misidentified τ_h in a sample of $t\bar{t}$+jets events.

Figure 1 shows the validity of the MP method in $t\bar{t}$ simulated events, in which their prediction in the signal region is compared to the estimation of the MP method applied to these simulations in the relaxed region. The method is performed with MPs computed in $t\bar{t}$ and multijet simulated events, separately for electrons and muons, showing a very good closure for the first ones and acceptable for the latter ones.

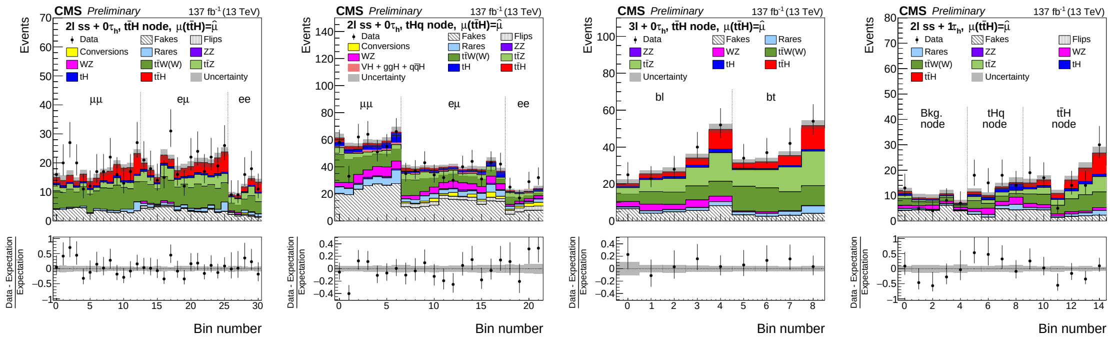

Figure 3: Observed events in the 2ℓSS + $0\tau_h$ $\mathrm{t\bar{t}H}$ (left), 2ℓSS + $0\tau_h$ tHq (center left), 3ℓ + $0\tau_h$ $\mathrm{t\bar{t}H}$ (center right) and 2ℓSS + $1\tau_h$(right) categories [5].

Electron charge flips are also estimated in a data driven way in the regions requiring two same-sign leptons, using the opposite-sign region as a sideband. The flip rate has been determined in a $\mathrm{Z} \to \mathrm{ee}$ data sample.

Irreducible backgrounds are dominated by $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}W}$ in the main signal regions. $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}W}$ events simulated at next-to-leading-order (NLO) using the MadGraph5_amc@nlo [7] generator. These samples are produced with one additional parton with multijet merging. The modeling of $\mathrm{t\bar{t}W}$ events includes α^3 and $\alpha_s\alpha^3$ electroweak corrections [8, 9], also simulated MadGraph5_amc@nlo. Generated events are then interfaced to pythia 8 citepythia to model the parton shower, the hadronization and the underlying event. The normalization of the simulated samples is an unconstrained nuisance parameter of the fit, allowing to constrain this quantity from data. Two dedicated control regions are used to constrain $\mathrm{t\bar{t}Z}$ in events with three or four leptons in the final state, while the contribution from $\mathrm{t\bar{t}W}$ is constrained from the $\mathrm{t\bar{t}W}$ subcategory in the 2ℓSS + $0\tau_h$category.

Other contributions from vector boson pair production or rare processes ($\mathrm{t\bar{t}t\bar{t}}$ or $\mathrm{tZ\bar{q}}$) are also relevant. Top pair production and DY+jets events are irreducible in the $1\ell + 1\tau_h$and $0\ell + 2\tau_h$regions. All these processes are simulated using simulations and normalized to the most precise cross sections.

3.3 Results and signal extraction

Signals are extracted by performing a maximum likelihood fit to data in all the subcategories of signal and control regions. This signal extraction assumes the distributions of the signal to behave according to the SM expectations, allowing only variations of the production rate of the processes as BSM effects. The likelihood fit implements the systematic uncertainties as nuisance parameters of the fit. Uncertainties on the parameters of interest are estimated using a *t*-statistic similar to the one employed in [11].

Figure 3 shows the distribution of events in some of the signal regions after the likelihood fit is performed, showing a good compatibility between data and the fitted model. The regions enriched in signal events also show a clear presence of the fitted $\mathrm{t\bar{t}H}$ signal.

The fitted signal strength is $0.92 \pm 0.19\,(\mathrm{stat})^{+0.17}_{-0.13}\,(\mathrm{syst})$ for the $\mathrm{t\bar{t}H}$ signal and $5.7 \pm 2.7\,(\mathrm{stat}) \pm 3.0\,(\mathrm{syst})$ for the tH signal, compatible with the SM expectations. The postfit value for the unconstrained nuisance parameters that represent the $t\bar{t}$W and $t\bar{t}$Z production rates are $\mu_{t\bar{t}W}$ = 1.43 ± 0.21 (stat+syst) and $\mu_{t\bar{t}Z}$ = 1.03 ± 0.14 (stat+syst). The $t\bar{t}$Z production rate is in clear agreement with the model, while the $t\bar{t}$W production rate is slightly above the expectations. The observed (expected) significance for $t\bar{t}$H production is 4.7 (5.2) standard deviations, while it is 1.4 (0.3) for the tH signal.

3.4 Interpretation

The observed yields can also be interpreted in terms of inference in the coupling modifier of the Higgs boson to the top quark (κ_t) and to the W and Z bosons (κ_V). The compatibility of the observed data is compared for a grid of (κ_t, κ_V) hypotheses, computing a likelihood score. The effect of deviations of κ_t and κ_V with respect to the unity in the kinematic distribution of tH events and the H boson branching fractions are taking into account.

The likelihood scan allows to construct confidence level regions in the (κ_t, κ_V) plane and confidence intervals in κ_t. The results allow to constrain κ_t to be within $-0.9 < \kappa_t < -0.7$ or $0.7 < \kappa_t < 1.1$ and 95% confidence level.

4. Conclusions

The observation of the Yukawa interactions between the top quark and the Higgs boson are crucial to establish the validity of the SM of particle physics. These couplings can be explored at tree level by performing measurements of $t\bar{t}$H and tH production, which can be done with high precision in the diphoton and multilepton final states. These measurements have been performed by the CMS Collaboration and are shown in this contribution. The results obtained are consistent with the SM expectations and achieve sensitivities for $t\bar{t}$H production above five standard deviations.

The author thanks Programa Severo Ochoa del Principado de Asturias for funding this work.

References

[1] ATLAS Collaboration, *Phys.Lett.* **B** 716 (2012) 1-29

[2] CMS Collaboration, *JINST* **3** S08004 (2008)

[3] CMS Collaboration, *Phys. Lett.* **B** 716 (2012) 30

[4] CMS Collaboration, *Phys. Rev. Lett.* , **125** 061801 (2020)

[5] CMS Collaboration, CMS-PAS-HIG-19-008

[6] CMS Collaboration, CMS-DP-2019-033

[7] Alwall et al., *JHEP* **07** (2014) 079

[8] Frederix, Tsinikos, *Eur. Phys. J.* C **80** (2020) 803

[9] Dror et al., *JHEP* **01** (2016) 071

[10] Sjöstrand et al., *Comput. Phys. Commun.* **191** (2015) 159

[11] ATLAS and CMS Collaborations, ATL-PHYS-PUB-2011-011, CMS-NOTE-2011-005.