PROCEEDINGS OF SCIENCE

Higgs boson production in association with a top quark pair at the LHC

Daniele Madaffari*
IFIC - Instituto de Física Corpuscular
E-mail: daniele.madaffari@cern.ch
On beahalf of the ATLAS and CMS collaborations.

The review of the results of the latest searches of Higgs boson production in association with a top quark-antiquark pair is presented. The searches are performed using a dataset of proton-proton collisions at centre-of-mass energy of 13 TeV recorded by the ATLAS and CMS detectors at the Large Hadron Collider. The data analysed corresponds to 36.1 fb^{-1} and 35.9 fb^{-1} respectively for the two experiments. The review covers studies targeting Higgs boson decay to pair of photons, W bosons, Z bosons, τ leptons and bottom quarks.

An Alpine LHC Physics Summit (ALPS2018)
15-20 April, 2018
Obergurgl, Austria

*Speaker.

PoS(ALPS2018)019

1. Introduction

In the Standard Model (SM) the Higgs boson has a Yukawa type interaction with fermions, with a coupling proportional to the fermion mass. One of the key parameters of the SM is y_t, the Yukawa coupling of the Higgs boson to the top-quark, the heaviest known particle. This quantity can be measured in model-dependent ways through loop mediated processes, which can have contributions from physics beyond the SM. The measurement of Higgs boson production in association with a top quark-antiquark pair ($t\bar{t}H$) gives a model-independent access to y_t. The $t\bar{t}H$ process contributes to $\sim 1\%$ of the total Higgs production at the Large Hadron Collider (LHC) with: $\sigma_{t\bar{t}H} = 0.5$ pb [1], for centre-of-mass energy of 13 TeV and Higgs mass of 125.09 GeV. As a counterpart, the presence of the two top quarks opens the possibility of a large spectra of signal topologies to be studied.

The most recent updates on the searches for the $t\bar{t}H$ process performed by the ATLAS [2] and CMS [3] experiments are reported in this review. The searches are performed using a dataset of proton-proton collisions at $\sqrt{s} = 13$ TeV produced at the LHC. The data analysed corresponds to 36.1 fb^{-1} and 35.9 fb^{-1} respectively for the two experiments. The review is organised in sections, each focused on studies targeting a specific Higgs boson decay. The combinations of these searches, performed independently by the two experiments, are also presented.

2. Higgs decay into a pair of b-quarks

The search for the Higgs decay to a b quark-antiquark pair benefits from a large branching ratio[1], but suffers from an irreducible background difficult to model: the associated production of top and b quark-antiquark pairs ($t\bar{t}+b\bar{b}$). The normalisation of this background currently can not be constrained with more than 35% accuracy by studies in data [4] and theoretical calculations [5][6]. ATLAS has performed dedicated studies to control as good as possible this background. The $t\bar{t}$ background contribution was divided, using generator-level information, in three categories according to the flavour of additional quarks present in the event: $t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$ and $t\bar{t}$ + light. The relative contribution in different sub-categories of the $t\bar{t}+b\bar{b}$ component was compared between the baseline $t\bar{t}$ Monte Carlo (MC) generator and an NLO $t\bar{t}+b\bar{b}$ sample. The predicted fractions in each sub-categories can be seen in Figure 1. This reflects in dedicated systematic uncertainties affecting shape and yield of this background. For detailed description of MC generators and parameters used in this study see [7].

ATLAS and CMS performed analysis targeting signal topologies where at least one of the two W bosons from the top quark decays into a lepton and a neutrino [7][8]. Here and in the following, unless differently specified, with lepton is intended only electron or muon. CMS studied also the all-jet final state [9], where both the W bosons decay to quarks.

2.1 Leptonic final states

Events are selected requiring one or two leptons at trigger level. Both experiments separate events in channels according to the number of reconstructed leptons. Events with exactly one lepton enter the single-lepton channel, the ones with two, the di-lepton channel. These are further divided as function of the multiplicity of jets and b-tagged jets[1]. ATLAS defines in the single-lepton channel also a region targeting boosted topologies.

Different discriminating variables are defined and optimised in each region. ATLAS uses a Boosted Decision Tree (BDT) as final discriminant in the signal-enriched regions of the single-lepton channel. The BDTs are trained against $t\bar{t}$ + jet background and have as inputs kinematic variables, the output of a dedicated BDT for event reconstruction, likelihood discriminants and Matrix Element Method (MEM) discriminants . The latter is used to improve rejection of $t\bar{t}+b\bar{b}$ background. In the regions dominated by the $t\bar{t}+c\bar{c}$ background the scalar sum of transverse momentum of jets (H_T) is used as final discriminant, in any other regions the yield is used instead. In the single-lepton channel CMS uses the output of a Deep Neural Network (DNN) to categorise events in each jet-multiplicity region. The DNNs, trained independently, separate the events according to 6 physics hypothesis: $t\bar{t}H$, $t\bar{t}+b\bar{b}$, $t\bar{t}$ +2b, $t\bar{t}$ +b, $t\bar{t}+c\bar{c}$ and $t\bar{t}$ + light-flavour. The output of the DNNs are also used as discriminant in the final fit. In di-lepton channel regions the output of a BDT, trained against $t\bar{t}$, is used as final discriminant. Output of MEM discriminating against $t\bar{t}+b\bar{b}$ is used instead in the most signal-enriched region of this channel. Input variables are similar for the two multivariate techniques: kinematic properties of objects and event-shape variables.

The signal strength $\mu = \sigma_{t\bar{t}H}/\sigma^{SM}_{t\bar{t}H}$ is extracted after a fit to the data. The best-fit values are:

$$\begin{aligned}
\mu_{\text{ATLAS}} &= 0.84 \pm 0.29 \text{ (stat.) } {}^{+0.57}_{-0.52} \text{ (syst.)} = 0.84^{+0.64}_{-0.61}, \\
\mu_{\text{CMS}} &= 0.72 \pm 0.24 \text{ (stat.) } {}^{+0.38}_{-0.38} \text{ (syst.)} = 0.72^{+0.45}_{-0.45}.
\end{aligned}$$

The largest limitation to the sensitivity is coming from the systematic uncertainty on the modelling of $t\bar{t}+b\bar{b}$ background. Other relevant systematics are the statistical uncertainty of the MC event samples and of the data-driven backgrounds and b-tagging efficiency. Results are compatible with SM prediction.

2.2 All-jets final states

This channel benefits from the high branching ratio of the top-quarks to an all-jets final state. As a counterpart, it is limited by the presence of non resonant multi-jet background, mainly originated from strong interactions, and identified as quantum chromodynamic (QCD) multi-jet production. To reject events originated by such process a quark-gluon likelihood (QGL) discriminant is used. This separates events rich in gluon jets, typical of QCD process, from the ones containing light-flavour jets from hadronic decay of W bosons.

Trigger selection is based on the presence of b-tagged jets, H_T, and number of jets. Events are categorised according to jet and b-tagged jet multiplicity. Further selection is applied as function of the combination in a likelihood ratio of the QGL output for all the non-b-tagged jets in the event, shown in Figure 1. The QCD multi-jet background is estimated from data at low b-tag multiplicity after applying corrections as function of p_T, η and ΔR between jets and b-tagged jets in the event. The estimation is validated in a data sample orthogonal to the estimation one. MEM is used to discriminate $t\bar{t}H$ against $t\bar{t}+b\bar{b}$, with one different signal hypothesis in each region.

[1]Jets identified as originating from a b-quark.

The fit to the data gives the signal strength: $\mu = 0.9 \pm 0.7$ (stat.) ± 1.3 (syst.) $= 0.9 \pm 1.5$. The dominating systematic uncertainties are the QCD yields, free floating in the fit, and the $t\bar{t}+b\bar{b}$ normalisation, with a prior uncertainty of 50%. Results are compatible with the SM.

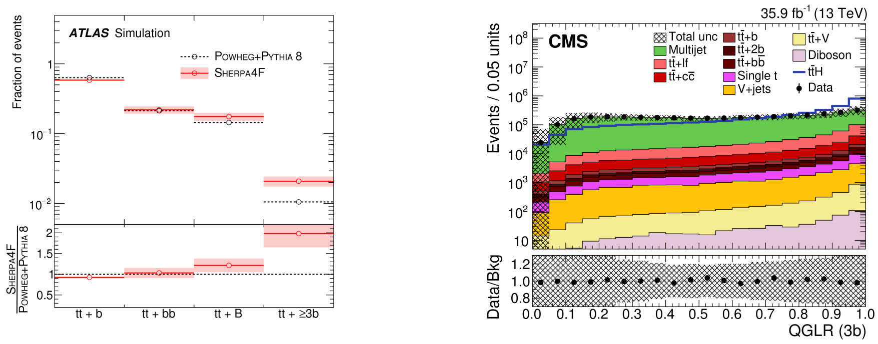

Figure 1: *Left*: Relative fractions of different $t\bar{t}+b\bar{b}$ subcategories previous the event selection. Comparison is between the baseline $t\bar{t}$ sample (POWHEG +PYTHIA 8) and the NLO $t\bar{t}+b\bar{b}$ sample (SHERPA4F). For a detailed description of MC generators' parameters and the definition of the subcategories see [7]. *Right*: Distribution in data and simulation of the quark-gluon likelihood ratio. Multijet background yield is scaled to match the one in data [9].

3. Higgs decay to WW^*, ZZ^* and $\tau\tau$

ATLAS and CMS experiments [10][11] model the main irreducible background, the associated production of a top quark-antiquark pair with a W or Z bosons ($t\bar{t}V$), using MC. A combination of multi-variate and data-driven techniques is used to reject the main reducible background: the mis-identification of τ and lepton candidates. The analyses consider only hadronic τ decays (τ_{had}).

Events are divided in categories according to the number of leptons (ℓ) and τ_{had}. Both collaborations use BDTs to discriminate $t\bar{t}H$ against $t\bar{t}V$ in all regions. Exceptions are the regions requesting 4 leptons, here the yield is used due to the small statistics collected. Moreover, CMS uses a MEM against $t\bar{t}Z$, instead of a BDT, in the 2 leptons with same-sign + 1 τ_{had} region.

The signal strengths measured after the fit to the data are:

$$\begin{aligned} \mu_{\mathrm{ATLAS}} &= 1.6 \pm 0.3 \text{ (stat.) } {}^{+0.4}_{-0.3} \text{ (syst.)} = 1.6^{+0.5}_{-0.4}, \\ \mu_{\mathrm{CMS}} &= 1.23^{+0.26}_{-0.25} \text{ (stat.) } {}^{+0.37}_{-0.35} \text{ (syst.)} = 1.23^{+0.45}_{-0.43}. \end{aligned}$$

These results have observed significances over the background-only hypothesis of 4.1 σ and 3.2 σ respectively. This represents evidence of $t\bar{t}H$ production in these final states. Summary of the results is shown in Figure 2. Limiting factors of the analyses are the uncertainties on τ identification, jet energy scale, mid-identified lepton backgrounds modelling, $t\bar{t}H$ and $t\bar{t}V$ modelling. Many channels are also limited by data statistics.

4. Higgs decay to two photons

At the time of this review the latest results for $t\bar{t}H$ in this channel are only available as part of general $H \to \gamma\gamma$ analyses [12][13]. This Higgs decay benefits from a clear signature given by

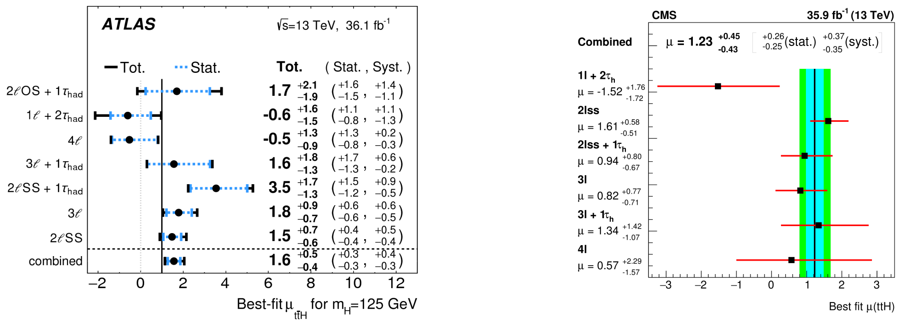

Figure 2: Summary of the observed best-fit values of the signal strength in each of the analysis regions and their combinations for ATLAS [10] (*Left*) and CMS [11] (*Right*). *SS* stands for same-sign leptons, *OS* for opposite-sign leptons.

the presence of the two photons while suffering from the small branching fraction. The signal is identified as a peak in the two-photons mass. ATLAS uses a double sided Crystal Ball function to parametrise the signal shape, while CMS uses a combination of Gaussian functions. Both experiments have regions tailored for $t\bar{t}H$ process, in particular two and nine for CMS and ATLAS respectively. These regions target both leptonic and hadronic top quark decays. Four of the ATLAS regions target also tH process. BDTs are implemented in the hadronic regions to increase background rejection.

The results obtained after the fit to the data are:

$$\mu_{\mathrm{ATLAS}} = 0.5 \pm 0.6,$$
$$\mu_{\mathrm{CMS}} = 2.2^{+0.9}_{-0.8}.$$

The two analyses are limited by data statistics. The systematic uncertainties related to photon energy scale and shower shape have the largest impact on the sensitivity.

5. Combination

ATLAS presented the combination of $t\bar{t}H$ searches at $\sqrt{s} = 13$ TeV [10]. These include the searches described in previous sections of this review and also the results from $t\bar{t}H$ regions in the general $H \to ZZ^* \to 4\ell$ analysis [14]. The best-fit signal strength is: $\mu = 1.17 \pm 0.3$, with a significance of 4.2 σ. Summary of the results in shown in Figure 3. The largest limitation to the sensitivity are coming from systematic uncertainties in $t\bar{t} + b\bar{b}$ modelling in $H \to b\bar{b}$ analysis, $t\bar{t}H$ cross-section modelling, non-prompt light-lepton and fake τ_{had} estimations.

CMS combined $t\bar{t}H$ searches performed in data collected at $\sqrt{s} = 13$ TeV, with those performed at $\sqrt{s} = 7$ and 8 TeV, corresponding to 5.1 and 19.7 fb^{-1} respectively [15]. The measured signal strength is :$\mu = 1.26^{+0.31}_{-0.26}$ with a significance above the background-only-hypothesis of 5.2 σ, as can be seen in Figure 3. This corresponds to the first observation of the $t\bar{t}H$ process at the LHC. The systematic uncertainties with largest impact are, similarly to the ATLAS combination, $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ normalisations and $t\bar{t}H$ cross-section modelling.

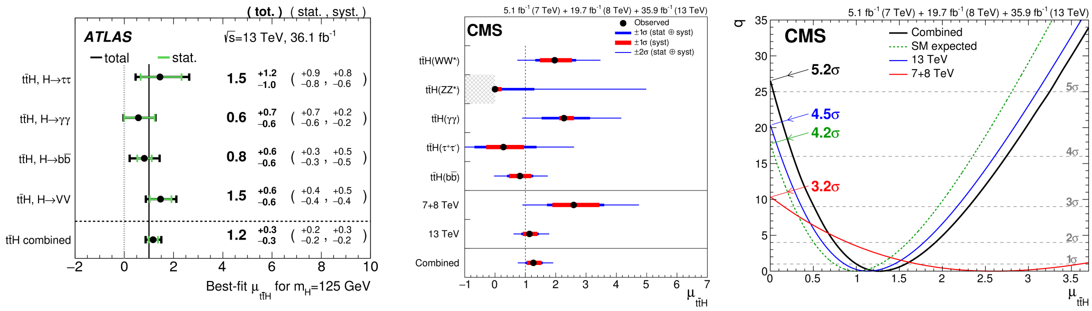

Figure 3: *Left*: Signal strength for each Higgs decay mode and overall result for the ATLAS combination [10]. *Centre*: Signal strength for each Higgs decay mode, for combined $\sqrt{s} = 7$ and 8 TeV and 13 TeV alone and overall result for the CMS combination [15]. *Right*: Test statistic q, described in [15], as function of signal strength for combined $\sqrt{s} = 7$ and 8 TeV and 13 TeV alone and overall CMS combination.

6. Conclusions

This review presents the latest searches of Higgs boson production in association with a top quark-antiquark pair performed by ATLAS and CMS experiments at LHC. A wide variety of final states is considered in order to maximise the sensitivity and reduce the impact of systematic uncertainties. A combination of the results of these analyses, independently for each experiment, is also presented. ATLAS combined the $\sqrt{s} = 13$ TeV analyses and measured a signal strength $\mu_{ttH} = 1.2^{+0.3}_{-0.3}$. This result excludes the background-only hypothesis with a significance of 4.2 σ, corresponding to the evidence of the $t\bar{t}H$ process at the LHC. CMS combined the $\sqrt{s} = 7$, 8, and 13 TeV analyses and measured a signal strength $\mu_{ttH} = 1.26^{+0.31}_{-0.26}$. The excess of events over the background-only hypothesis has a significance of 5.2 σ, corresponding to the observation of the $t\bar{t}H$ process at the LHC. All the reported results are compatible with the SM expectation.

In the near future the inclusion of the data collected in 2017 and 2018 will make the statistically limited channels more relevant in the combinations. The increased amount of data can also allow to use new strategies for background control. Moreover, work is ongoing to lessen the impact of theoretical and experimental systematic uncertainties.

References

[1] LHC Higgs Cross Section Working Group collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 1610.07922.

[2] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *Journal of Instrumentation* **3** (2008) S08003.

[3] CMS Collaboration, *The CMS experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) S08004.

[4] CMS collaboration, *Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s}$ = 13 TeV*, Phys. Lett. **B776** (2018) 355 [1705.10141].

[5] G. Bevilacqua, M. V. Garzelli and A. Kardos, *$t\bar{t}b\bar{b}$ hadroproduction with massive bottom quarks with PowHel*, 1709.06915.

[6] T. JežO, J. M. Lindert, N. Moretti and S. Pozzorini, *New NLOPS predictions for $t\bar{t}+b$-jet production at the LHC*, 1802.00426.

[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* $H \to \mathrm{b\bar{b}}$ *decay channel with leptonic* $\mathrm{t\bar{t}}$ *decays in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, 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*, 1803.06986.

[10] ATLAS collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Phys. Rev. **D97** (2018) 072003 [1712.08891].

[11] CMS collaboration, *Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying* τ *leptons at* $\sqrt{s} = 13$ *TeV*, 1803.05485.

[12] ATLAS collaboration, *Measurements of Higgs boson properties in the diphoton decay channel with 36* fb^{-1} *of pp collision data at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, 1802.04146.

[13] CMS collaboration, *Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, 1804.02716.

[14] ATLAS collaboration, *Measurement of the Higgs boson coupling properties in the* $H \to ZZ^* \to 4\ell$ *decay channel at* $\sqrt{s}$ = 13 *TeV with the ATLAS detector*, JHEP **03** (2018) 095 [1712.02304].

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