SEARCH FOR ASSOCIATED PRODUCTION OF HIGGS BOSONS AND TOP QUARKS IN MULTILEPTON FINAL STATES AT $\sqrt{s} = 13$ TeV AT CMS

N. CHANON on behalf of the CMS Collaboration
Institut Pluridisciplinaire Hubert Curien, Université de Strasbourg, CNRS/IN2P3, Strasbourg, France

Searches for the production of a standard model (SM) Higgs boson in association with a top quark pair ($t\bar{t}H$) is presented, in the multilepton (e, μ, τ) final states. The dataset analyzed corresponds to an integrated luminosity of 35.9 fb^{-1} of proton–proton collisions at $\sqrt{s} = 13$ TeV by the CMS experiment in 2016. A best fit $t\bar{t}H$ yield of $1.5^{+0.5}_{-0.5}$ times the SM prediction is measured in e, μ final states, and $0.7^{+0.6}_{-0.5}$ times the SM prediction in final states with at least one hadronic τ, with an observed (expected) significance of respectively 3.3σ (2.5σ) and 1.4σ (1.8σ).

1 Introduction

Since the discovery of a new boson with a mass of 125 GeV by the ATLAS and CMS Collaborations [1,2], its properties have been measured extensively and found to be compatible with the SM Higgs boson. The Top Yukawa coupling is still known with poor accuracy. It can be accessed directly, at the tree level, by measuring the production of $t\bar{t}H$, with a cross section of about 1% of the total cross section for the Higgs boson production [3]. At the Run1 of the LHC, the $t\bar{t}H$ production was measured with a significance of 4.4σ (2.0σ expected) by combining data of all channels at ATLAS and CMS [4].

This proceeding presents an analysis with an integrated luminosity of 35.9 fb^{-1} of proton–proton collisions at $\sqrt{s} = 13$ TeV by the CMS experiment [5] in 2016. Searches for the $t\bar{t}H$ process in multilepton final states will be described, targeting the decays $H \to WW^*$, ZZ^* (with subsequent fully leptonic or semi-leptonic decays), and $\tau\tau$, where at least one τ is decaying hadronically. Among the two top quark decays, at least one of them must decay leptonically. The multilepton final states have lower rate than $H \to b\bar{b}$ final state; however these channels have low background, and the irreducible backgrounds are known with a higher precision on the theory uncertainties.

The analysis is divided into two parts: final states targetting electrons and muons [6], and final states with at least one hadronic τ [7].

2 Multilepton (e, μ) final states

The analysis is targeting the following categories: 2 same sign leptons ($2\ell ss$), 3 leptons (3ℓ) or 4 leptons (4ℓ), with no hadronic τ (dedicated analysis for τ_h is presented section in 3). At least 4 jets among which 1 b-tagged jet must be reconstructed in the $2\ell ss$ category. The two leptons are requested to be of same sign to reduce Drell-Yan and $t\bar{t}Z$ contribution. At least 4 jets among which 1 b-tagged jet must be reconstructed in the 3ℓ and 4ℓ category (for the latter, a mass requirement is applied to remove $H \to 4\ell$ events, left to the dedicated $H \to ZZ^* \to 4\ell$ analysis

[8]). The lepton identification was optimized for this analysis, with a selection on the output of a boosted decision tree (BDT) using shape, isolation and overlapping jet information, to reduce the contribution of jets misidentified as leptons in the busy $t\bar{t}H$ environment. The categories are further divided in sub-categories according to tightness of the b-tagging criterion and lepton charge (taking advantage of the charge asymmetry of the $t\bar{t}W^{\pm}$ background at the LHC).

The irreducible background consists in $t\bar{t} + W/Z/\gamma^*$ and is estimated from simulation, with an uncertainty of about 10%. The reducible background, mainly made of $t\bar{t}$+jets, with a jet misidentified as lepton or a lepton with mis-measured charge, is measured from data, with an uncertainty of about 30%. The background of jets mis-identified as leptons is computed from a QCD control region with loosened lepton identification. The background arising from lepton charge mis-assignment is measured from from a same-sign sample by measuring the amount of peaking events (electron only, rate for muon charge mis-assignment is negligible).

In $2\ell ss$ and 3ℓ categories, the signal extraction procedure relies on two BDTs. The first BDT, trained on $t\bar{t}H$ against $t\bar{t}$+jets, includes various kinematic variables, and a dedicated tagger for hadronic top reconstruction in the $2\ell ss$ category. The second BDT, trained on $t\bar{t}H$ against $t\bar{t}W/Z$, includes also kinematic variables as input, as well as a dedicated tagger for jets arising from Higgs semi-leptonic decay in $2\ell ss$ category (where jets tagged by the hadronic top tagger are removed), and a likelihood ratio of $t\bar{t}H$ vs $t\bar{t}W$+$t\bar{t}Z$ hypotheses computed with a Matrix Element Method (MEM) in 3ℓ category. A bidimensionnal distribution of the kinematic BDTs is built, which is mapped into a 1D distribution, grouping bins with similar signal/background ratio such that the expected background yield is similar in each bin. The 1D distribution for $2\ell ss$ category is shown in Fig. 1. A counting experiment is performed in the 4ℓ category given the low expected number of events overall. The signal and background contributions estimated from data and simulation is fit to the data.

The result of the fit is shown in Fig. 1. The best fit signal strength is driven by the $2\ell ss$ category (with highest yield), and found to be 1.5 ± 0.5. The observed (expected) significance is found to be 3.3σ (2.5σ) with 2016 data, and the same result is found combining with the ten times smaller dataset of 2015 data. The behaviour of the fit is checked by allowing the $t\bar{t}W/Z$ contribution to float, and the observed significance found is 3σ. The main systematic uncertainties in the fit are arising from the tight lepton selection and the contribution of jet mis-identified as leptons.

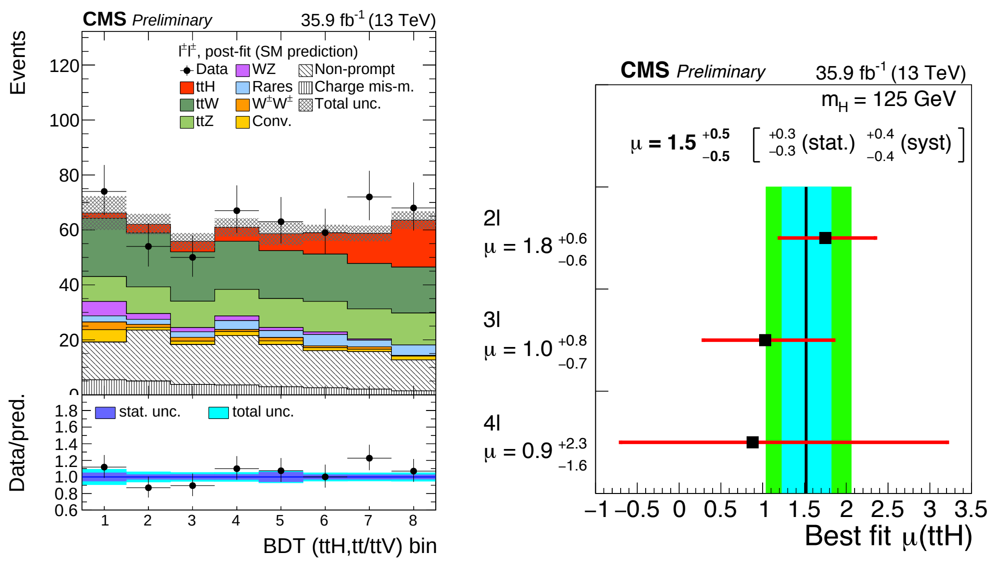

Figure 1 – Combination of the BDT classifier outputs in the bins used for signal extraction, for the same-sign dilepton channel (left). Best fit signal strength for e, μ multilepton final states, by categories (right).

3 Analysis in multilepton final states with at least one hadronic τ

Categories involving hadronic τ final states are orthogonal to multilepton categories defined in section 2. They are treated separately and are described in this section. Three categories are defined: $1\ell{+}2\tau_h$ and $2\ell ss{+}1\tau_h$, requiring at least 3 jets and among which 1 b-tagged jet, and $3\ell{+}1\tau_h$, requiring at least 2 jets with 1 b-tagged jet.

Hadronic τ_h reconstruction is performed with the CMS hadron+strip algorithm [9]. The τ_h reconstruction is seeded by particle flow jets. Neutral pions are recovered using dynamic η-ϕ strips, widened depending on p_T in the case of bremstrahlung or nuclear interaction. A decay mode finding procedure is applied, reconstructing various possible combinations of charged hadrons and strips. The τ_h identification is based on a MVA using isolation sum computed within a cone $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$, optimised for the busy hadronic environment in $t\bar{t}H$ processes.

The background estimate is following a similar strategy as the previously described multilepton (e, μ) analysis. In the $1\ell{+}2\tau_h$ category, the background of jets mis-identified as τ_h is dominant and is measured from data, with a $t\bar{t}$ sample in $e\mu + 1$ jet final state.

The signal extraction procedure is using advanced techniques based on MEM and kinematic BDT. In the $1\ell{+}2\tau_h$ category, a single BDT is trained against the $t\bar{t}$+jets background. The $2\ell ss{+}1\tau_h$ category is using the MEM likelihood ratio of $t\bar{t}H$ against $t\bar{t}Z + t\bar{t}$ hypotheses. Two sub-categories are defined: "no missing jet" if all reconstructed jets can be assigned to the matrix element partons, and "missing jet" if one jet is escaping acceptance or is not reconstructed and cannot be assigned to the matrix element (in which case the phase space is extended to integrate over the missing jet). In the $3\ell{+}1\tau_h$ category, the same 2D BDT procedure as described in section 2 is applied.

The results are shown in Fig. 2. The sensitivity of the analysis is driven by the $2\ell ss{+}1\tau_h$ category. The combined best fit value measured is $0.7^{+0.6}_{-0.5}$. The significance observed is 1.4σ (1.8σ expected). In the absence of signal, the limit at 95% confidence level on the signal strength is observed to be 2.0 (while 1.1 is expected). The main systematic uncertainties are the tight lepton selection, τ_h identification and the background of jets mis-identified as τ_h.

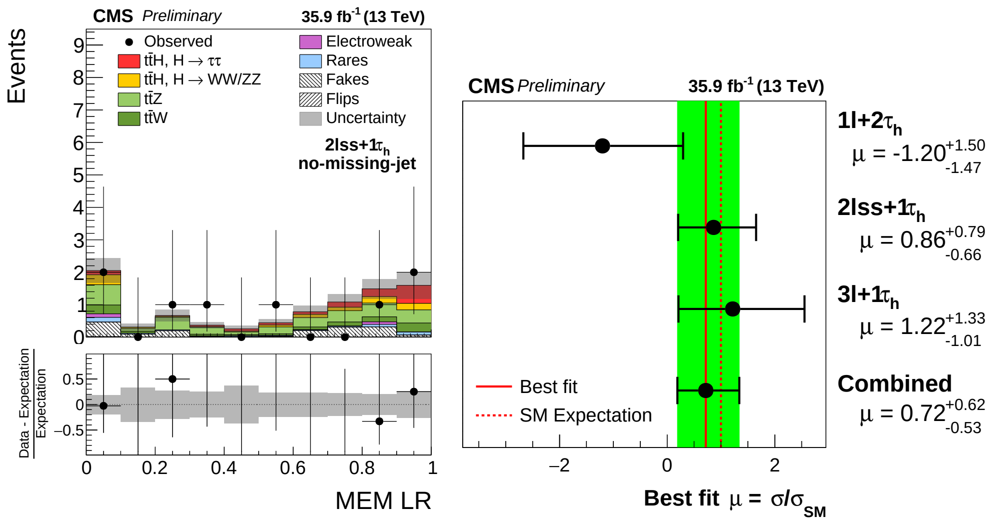

Figure 2 – Distribution in the discriminating observable used for the signal extraction in the "no-missing-jet" sub-category of the $2\ell ss{+}1\tau_h$ category (left). Best fit signal strength in multilepton final states with at least one hadronic τ, by categories (right).

4 Conclusions

A comparison of the best fit signal strength for CMS analyses targeting $t\bar{t}H$ final states can be found on Table 1. The final states with 4ℓ [8], multileptons [6] (described in section 2), and $\tau_h + X$ [7] (section 3) are updated with the full 2016 dataset.

Table 1: Measured best fit signal strength for $t\bar{t}H$ production at CMS.

	$\hat{\mu}$
LHC Run 1 (*JHEP* 08, 045, 2016)	$2.3^{+0.7}_{-0.6}$
CMS Run 2	
$\gamma\gamma$ (HIG-16-040, 35.9 fb^{-1})	$2.2^{+0.9}_{-0.8}$
4ℓ (HIG-16-041, 35.9 fb^{-1})	$0.0^{+1.2}_{-0.0}$
$b\bar{b}$ (HIG-16-038, 12.9 fb^{-1})	-0.2 ± 0.8
multileptons (HIG-17-004, 35.9 fb^{-1})	1.5 ± 0.5
τ_h+X (HIG-17-003, 35.9 fb^{-1})	$0.7^{+0.6}_{-0.5}$

In the multilepton (e, μ) final states, the $t\bar{t}H$ signal is observed with 3.3σ significance (2.5σ expected), and the signal strength is compatible with the standard model. The expected significance with this single channel is already better than the LHC Run1 combination will all channels. The $\tau_h + X$ analysis is performed for the first time at Run 2, with advanced signal extraction techniques, although more data is needed for this channel to achieve comparable sensitivity to the multilepton analysis. More insights on the $t\bar{t}H$ process will be possible at Run 2 where the luminosity might be three times higher than the dataset analyzed in this proceeding.

References

1. CMS Collaboration, *Phys. Lett.* B **716**, 30-61 (2012).
2. ATLAS Collaboration, *Phys. Lett.* B **716**, 1-29 (2012).
3. LHC Higgs Cross Section Working Group, arxiv:1610.07922.
4. ATLAS and CMS Collaborations, *JHEP* **08**, 045 (2016).
5. CMS Collaboration, *JINST* **3** S08004 (2008).
6. CMS Collaboration, PAS HIG-17-004.
7. CMS Collaboration, PAS HIG-17-003.
8. CMS Collaboration, PAS HIG-16-041.
9. CMS Collaboration, PAS TAU-16-002.