Searches for $t\bar{t}H$ production at CMS

Aruna Kumar Nayak[*][†]

Institute of Physics, Bhubaneswar, India

E-mail: Aruna.Nayak@cern.ch

Results of searches for the Higgs boson production in association with a pair of top quarks using the proton-proton collision data collected by the CMS detector at LHC are presented. First, results of the searches at centre-of-mass energy of 13 TeV are briefly discussed. Then, the combination of results at centre-of-mass energies of 7, 8, and 13 TeV are presented, which provides an observation of the Higgs boson production in association with a pair of top quarks with a significance of 5.2 standard deviations.

---

[*]Speaker.

[†]On behalf of the CMS Collaboration

1. Introduction

Since the discovery of a Higgs boson by the ATLAS and CMS experiments in 2012 [1, 2], the major physics activity of these experiments at LHC has been centered on the precision measurement of properties of this particle to establish its exact nature. Towards this effort, the measurement of its coupling to the top quark has gained high phenomenological interest due to an extraordinary large value of the top quark mass compared to all other known particles. Though the top quark Yukawa coupling (y_t) is measured indirectly from the Higgs boson production in the gluon fusion process, as shown in Fig. 1 (left), and agrees with the standard model (SM) expectation [3, 4], it can be affected by possible contributions of beyond-the-SM particles to the loop diagram. Hence, the measurement of the Higgs boson production in association with a top quark pair, t̄tH, provides the most precise model-independent estimate of y_t. The diagram for the t̄tH process is shown in Fig. 1 (right).
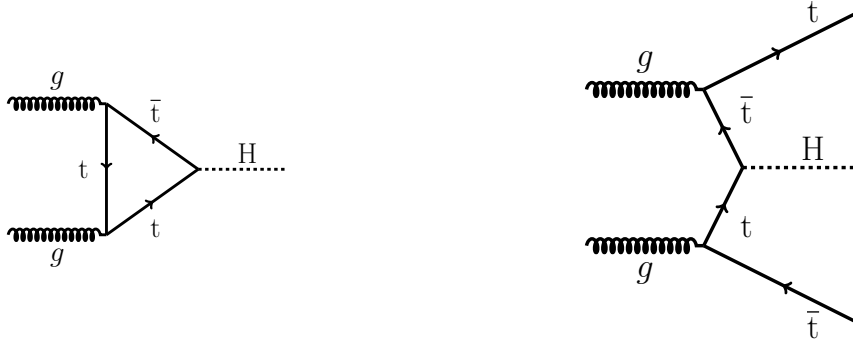

Figure 1: Feynman diagrams for the Higgs boson production via the gluon fusion process (left) and in association with a pair of top quarks (right).

In this article, we briefly discuss the search strategies and results of the t̄tH production in various final states based on the proton-proton collisions recorded at centre-of-mass (CM) energy $\sqrt{s}$ = 13 TeV by the CMS detector [5] during 2016. Results of these searches combined with the same at $\sqrt{s}$ = 7 and 8 TeV are then presented.

2. Search for t̄tH, H → bb̄

This search targets for the final states where the Higgs boson decays to a pair of b quarks [6, 7]. Given the larger branching fraction of H → bb̄, this channel has higher statistics compared to other decays. However, it also suffers from copious contribution of backgrounds. The analyses are separately performed in the final states where the W boson, arising from the decay of a top quark, decays either to a lepton and a neutrino or to a pair of quarks.

2.1 t̄tH, H → bb̄ in semileptonic and leptonic final states

This analysis is further divided into two separate searches: (1) where one W boson decays to a lepton and a neutrino and the other to a pair of quarks, and is named as the ℓ+jets final state, and (2) where both the W bosons decay to leptons, and is called the dilepton final state.

Both the event categories suffer from large contaminations from t̄t+jets, especially from t̄tbb̄. Thus, discriminants based on multivariate methods, such as Boosted Decision Tree (BDT), Deep Neural Network (DNN), and matrix-element method (MEM) are used to separate signal from backgrounds [6].

The dilepton events are required to have at least two electrons, two muons, or one electron and one muon, and at least four jets. The events are further categorised depending on the number of b-tagged jets: = 3b or $\geq$ 4b. A BDT is trained for each of these two categories using input variables such as object kinematics, event shapes, and b-tag discriminants. In the 3b category the BDT distribution is used as the final discriminant to extract signal, while the 4b category is further divided into two sub-categories based on the BDT distribution. In both the sub-categories an MEM discriminant is used to distinguish signal from backgrounds, where the MEM discriminant is constructed from leading-order matrix elements for the t̄tH and t̄tbb̄ processes.

The ℓ+jets events are required to have either an electron or a muon, $\geq$ 4 jets, and $\geq$ 3b-tagged jets. The events are divided into three categories based on the number of jets: 4, 5, and $\geq$ 6 jets. A multi-classification DNN is used in each of these jet multiplicity categories to discriminate signal from backgrounds. A DNN associates a set of probabilities describing the likelihood of an event being either signal t̄tH or one of the t̄t+jets backgrounds. Thus, events in each category are further divided to six sub-categories based on the most probable processes assigned by the DNN, forming a total of 18 jet-process sub-categories. In each sub-category, the DNN output distribution of the node that matches the process category is used as the final discriminant.

The discriminant distribution from all 18 sub-categories are fitted together in a maximum likelihood (ML) fit to extract the signal yield. The best-fit value of the signal strength modifier, which is the ratio of the signal production rate to that expected from the SM ($\mu = \sigma/\sigma_{SM}$), is shown in Fig 2 (left), along with the uncertainties, separately for ℓ+jets and dilepton final states as well as for their combination. Major sources of systematic uncertainties are the predictions for t̄t + heavy-flavour background rates, jet energy scale, and uncertainties on b-tagging efficiency and misidentification probabilities. The significance of the signal is found to be 1.6 standard deviations (σ) compared to an expected significance of 2.2σ.

2.2 t̄tH, H $\rightarrow$ bb̄ in fully hadronic final states

Events where both the W bosons decay to quarks are considered in this study [7]. Given the final state comprises jets only, a special trigger, requiring at least six jets, large H_T, where H_T is the sum p_T of jets, and at least one or two b-tagged jets, is deployed to select events at the online level. The final state suffers from large QCD multijets in addition to t̄t+jets and t̄tbb̄ backgrounds. Since the multijet processes are dominated by gluon jets a quark–gluon likelihood discriminant is constructed to suppress this background. The events are divided into three categories based on the number of jets: 7, 8, and $\geq$ 9 jets. Each jet category is further divided to two sub-categories depending on the number of b-tagged jets, 3b or $\geq$ 4b, resulting in a total of six sub-categories. The contribution and shape of multijet background in the signal region are obtained from a QCD enriched control region. The distributions of MEM discriminants, constructed from leading-order matrix elements for the t̄tH and t̄tbb̄ processses, are used to extract signal.

The best fit values of μ for each category and their combination are shown in Fig. 2 (right), which are consistent with SM expectations. The systematic uncertainties are dominated by the

uncertainties on the estimation of multijet background, prediction of tt̄bb̄ background, b-tagging, and jet energy scale.
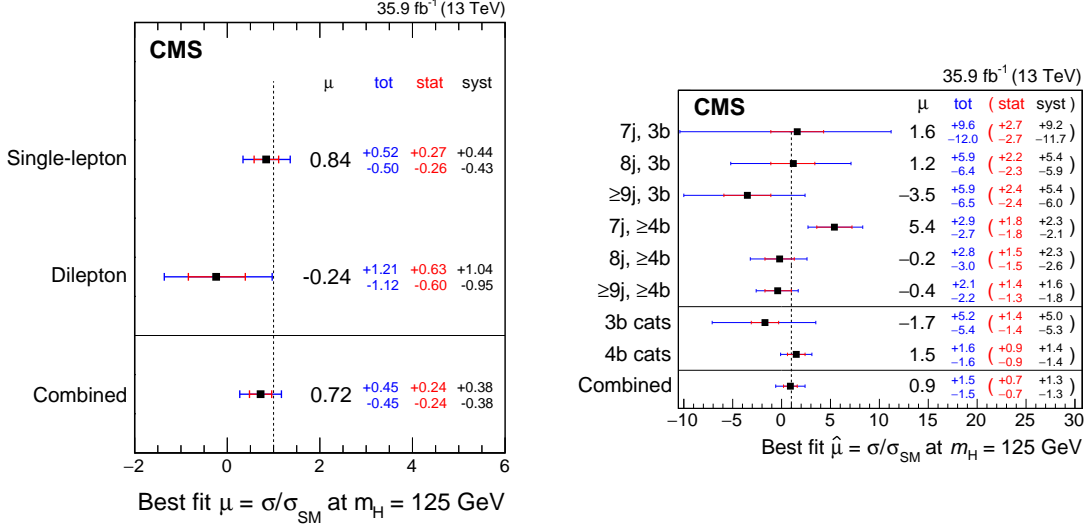

Figure 2: Best fit values of the signal strength modifiers with their 68% expected confidence intervals (outer error bar), also split into their statistical (inner error bar) and systematic components, for the tt̄H, H → bb̄ final states. (Left) For final states with at least one W boson decaying to leptons [6]. (Right) For final states with all W bosons decaying to quarks [7].

3. Search for tt̄H, H → leptons

This final state corresponds to events where the Higgs boson decays to a pair of W bosons (H → WW), Z bosons (H → ZZ), or τ leptons (H → $\tau\tau$), with the subsequent decays of W/Z bosons to electrons, muons, or tau leptons [8].

The events are categorized based on the number of leptons and hadronic τ decay candidates (τ_h). The following six categories are considered: (1) one electron or muon and two τ_h ($1\ell+2\tau_h$), where ℓ and τ_h have opposite charge, (2) two leptons of opposite charge and no τ_h (2ℓss), (3) two leptons of opposite charge and one τ_h (2ℓss+$1\tau_h$), (4) three leptons and no τ_h (3ℓ), where the sum of the charges of 3 leptons is 1, (5) three leptons and one τ_h ($3\ell+1\tau_h$), where the sum of the charges of four candidates is 0, (6) Four leptons (4ℓ), where ℓ represents electron or muon. Additional selection criteria, such as requiring at least two to four jets, of which one to two b-tagged, threshold on missing transverse energy, and veto around the Z boson mass, are applied depending on event categories.

SM processes namely tt̄V (V = W and Z) and diboson (WW, ZZ, and WZ) constitute the major irreducible backgrounds, while tt̄+jets, where jets are misidentified as leptons, is also a major background. The irreducible backgrounds are modeled with Monte Carlo simulation and validated using data in dedicated control regions. Backgrounds with misidentified electron, muon, or τ_h candidates are estimated from data using a probabilistic method, where the jet → ℓ/τ_h misidentification probabilities are measured from data in control regions as a function of p_T and η.

MEM and BDT discriminants are used to extract the signal yield. In some categories, such as 2ℓss, 3ℓ and $3\ell+1\tau_h$, two separate BDT discriminants are used, one trained against the tt̄V

background while the other against t̄t+jets. The outputs of the two BDTs are mapped to a single discriminant using a likelihood method. In the 3ℓ category the MEM discriminant is also used as input to the BDT.

The signal is extracted with a simultaneous ML fit to the distribution of final discriminants in all categories. The best fit value of μ along with its uncertainties are shown in Fig. 3 (left) for the individual category and their combination; they agree with the SM expectation. The distribution of log(S/B) from the pre-fit distributions of final discriminants is shown in Fig. 3 (right), where the excess of events seen in data is compatible with the expected SM signal. The observed significance is 3.2σ against an expected significance of 2.8σ.
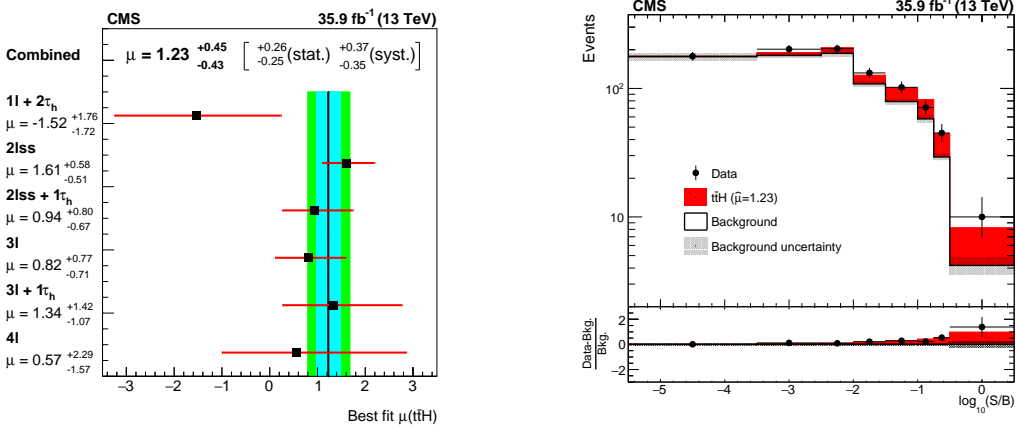

Figure 3: (Left) Best fit values of μ measured in each of the individual categories and for the combination of all categories of the multilepton final state [8]. (Right) Distribution of the logarithm of the ratio between the expected signal and background in each bin of the distributions used for signal extraction [8]. The distributions of signal and background processes are shown for the values of nuisance parameters obtained from the combined ML fit and $\hat{\mu} = 1.23$, corresponding to the best-fit μ value from the ML fit.

4. Combination of t̄tH search results

Results of the searches in different final states, and at the CM energies of 7, 8, and 13 TeV, are combined together [9]. The best-fit values of μ for individual final states as well as their combination are shown in Fig. 4 (left). Results are also presented separately for the searches at 7 & 8 TeV and at 13 TeV. The measurements for different final states and at different CM energies agree well with each other as well as with the combination. Figure 4 (right) shows the variation of test statistic (q) against μ separately for searches at 7 & 8 TeV and at 13 TeV as well as for their combination. The q is minimum at the best-fit value of μ. The excess of events observed in data with respect to SM backgrounds is interpreted in terms of significance. The observed significance is found to be 5.2σ compared to an expected significance of 4.2σ. This constitutes the first observation of the t̄tH process at LHC.

5. Summary

Results of the searches for the Higgs boson production in association with a pair of top quarks,
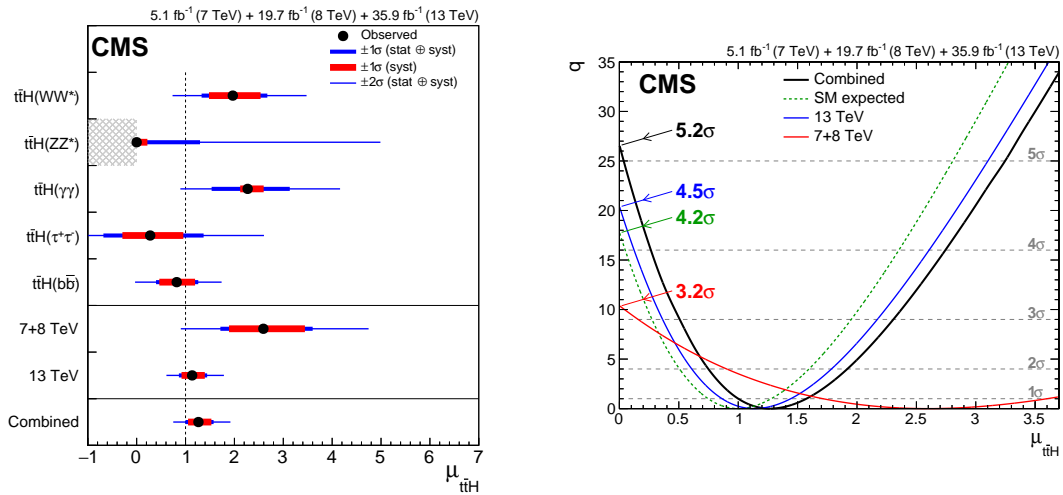
Figure 4: (Left) Best fit values of signal strength for the five individual decay channels as well as for the combined result [9]. (Right) The test statistic as function of signal strength for all decay modes separately for 7+8 TeV and 13 TeV and their combinations [9]. The expected SM result for the overall combination is also shown.

based on pp collision data recorded at 13 TeV by the CMS experiment, are presented. These are combined with the earlier results obtained at 7 and 8 TeV yielding a total significance of 5.2 standard deviations which constitutes the first observation of tt̄H process at the LHC.

References

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

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

[3] CMS Collaboration, JHEP 08 (2016) 045, arXiv:1606.02266 [hep-ex].

[4] CMS Collaboration, CMS-PAS-HIG-17-031, http://cds.cern.ch/record/2308127.

[5] CMS Collaboration, JINST 3 (2008) S08004.

[6] CMS Collaboration, CMS-HIG-17-026, arXiv:1804.03682 [hep-ex].

[7] CMS Collaboration, CMS-HIG-17-022, arXiv:1803.06986 [hep-ex].

[8] CMS Collaboration, CMS-HIG-17-018, arXiv:1803.05485 [hep-ex].

[9] CMS Collaboration, Phys. Rev. Lett. 120 (2018) 231801, arXiv:1804.02610 [hep-ex].