Chapter 3
Search for Standard Model Higgs Boson Production in Association with Top Quark Pairs at CMS

Ram Krishna Dewanjee

Abstract After the discovery of a new boson of mass 125 GeV (LHC. Phys. Lett. B716:30, 2012 [1]), one of the main goals of the LHC is to precisely measure its properties. Within the current experimental uncertainties, the properties of this boson are compatible with the expectation for the standard model (SM) Higgs boson. However, due to lack of sufficient data in Run-1, some of the properties of this particle are yet to be measured. The Yukawa coupling between this boson with the top (t) quark is one such crucialy important property. Many beyond SM (BSM) theories predict deviations of this coupling from SM value as evidence for new physics. SM Higgs (H) production in association with top quarks ($t\bar{t}H$) allows a direct measurement of this coupling. In this paper, results of searches for $t\bar{t}H$process are presented in final states involving bottom (b) quarks, photons (γ), leptons (e/μ), and (hadronically decaying) tau leptons (τ_h) using Run-2 luminosity collected in 2016 (35.9 fb^{-1}) at $\sqrt{s}$ = 13 TeV by the CMS experiment (CERN LHC. J. Instrum. 08: S08004, 2008 [2]).

3.1 Introduction

Within the SM framework, the Higgs boson is responsible for the dynamic generation of masses of all SM particles via the Brout–Englert–Higgs mechanism. In the fermion sector, Higgs interacts via yukawa couplings which are proportional to the fermion masses. Top quark being the heaviest SM fermion, may play a yet unknown role in the electroweak symmetry breaking. This makes the top-Higgs yukawa coupling an important probe for BSM physics. Although indirect constraints on its value (via gluon fusion and the $H \rightarrow \gamma\gamma$ loop contribution) are available, its direct measurement is possible only via studying top quark associated Higgs production ($t\bar{t}H$and tH). This article overviews all major Run-2 $t\bar{t}H$searches: $\boldsymbol{t\bar{t}H(H \rightarrow b\bar{b})}$

R. K. Dewanjee (✉)
Laboratory of High Energy and Computational Physics, KBFI, Rävala pst 10, 10143 Tallinn, Estonia
e-mail: ram.krishna.dewanjee@cern.ch; dewanjeeramkrishna5@gmail.com

P. K. Behera et al. (eds.), *XXIII DAE High Energy Physics Symposium*, Springer Proceedings in Physics 261,
https://doi.org/10.1007/978-981-33-4408-2_3

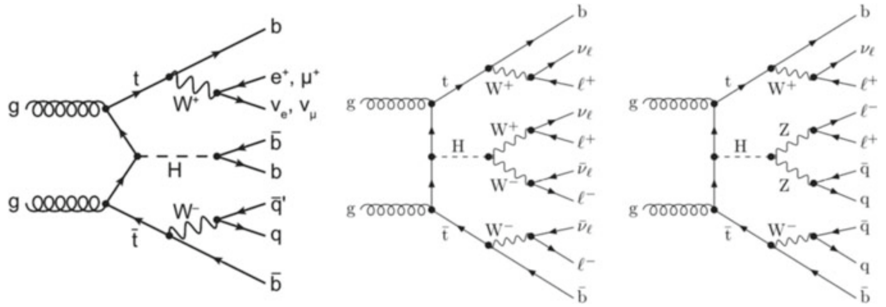

Fig. 3.1 LO Feynman diagrams for top quark associated Higgs boson production ($t\bar{t}H$) with Higgs decaying to pair of b-quarks (left), W-bosons (middle) and Z-bosons (right)

[3, 4], $\boldsymbol{t\bar{t}H(H \to \gamma\gamma)}$ [5] and $\boldsymbol{t\bar{t}H}$ **Multi-lepton** [6] search (which combines $\boldsymbol{t\bar{t}H(H \to VV^*)}$[1] [7] and $\boldsymbol{t\bar{t}H(H \to \tau\tau)}$ [8] searches) (Fig. 3.1).

3.2 Analysis Strategy and Event Categorization

The general analysis strategy of all $t\bar{t}H$searches involve categorization to seperate events into categories of varying sensitivity. Signal over background discrimination in each category is then enhanced via. Dedicated shape analysis of a usually multivariate (MVA) discriminator. Finally, a binned maximum likelihood fit on the distribution of this discriminator is performed for signal extraction and limit computation. The major categories employed by the $t\bar{t}H$searches are as follows:

1. $\boldsymbol{t\bar{t}H(H \to b\bar{b})}$: This search benefits from the large SM $H \to b\bar{b}$ branching fraction ($\sim$57%) but is limited by systematic uncertainity on the (irreducible) $t\bar{t} + b\bar{b}$ background. It is divided into two primary channels (based on the hadronic or leptonic decays of the top quarks):

(a) **Hadronic channel**: Events are first selected using dedicated multi-Jet triggers. They are then divided into six categories depending on jet and b-tagged jet multiplicity. These are (7Jets, 3b-tagged), (7Jets, $\geq$4b-tagged), (8Jets, 3b-tagged), (8Jets, $\geq$4b-tagged), ($\geq$9Jets, 3b-tagged), and ($\geq$9Jets, $\geq$4b-tagged). MEM[2] was employed in all the categories to distinguish signal ($t\bar{t}H$) from background ($t\bar{t} + b\bar{b}$) and was also used for signal extraction.

(b) **Leptonic channel**: Events are first selected using single lepton triggers and divided into two main categories (each of which is further split by lepton flavor).

[1] V = W/Z-boson.

[2] Matrix element method.

* **Single Lepton category**: Events are split into (4Jets, ≥3b-tagged), (5Jets, ≥3b-tagged), and (≥6Jets, ≥3b-tagged) categories. DNNs[3] are trained to seperate $t\bar{t}H$signal from $t\bar{t} + X$ background (where $X = b\bar{b}/c\bar{c}$/2b/b and light flavors).
* **Di-Lepton category**: Events are divided into (≥4Jets, 3b-tagged) and (≥4Jets, ≥4b-tagged) categories with BDTs[4] trained to distinguish $t\bar{t}H$signal from $t\bar{t}$ background in both. While the former uses this BDT itself for signal extraction the latter gives this as input to MEM discrimintaor.

2. $\boldsymbol{t\bar{t}H(H \rightarrow \gamma\gamma)}$: Events are first selected using asymmetric[5] di-photon triggers. This search is itself a channel under the general SM $H \rightarrow \gamma\gamma$ search and utilizes special BDT for tagging $t\bar{t}H$multi-jet events. This BDT uses the following variables as inputs:

   - Number of jets ($p_T > 25$ GeV).
   - Leading jet p_T.
   - 2 jets with the highest b-tagging score.

Contributions of other SM (non $t\bar{t}H$) $H \rightarrow \gamma\gamma$ processes e.g., Gluon fusion, VBF[6] are treated as background by this BDT. Additional BDTs are used for photon identification and di-photon vertex assignment. This search too is split into Leptonic and Hadronic channels targetting semi-leptonic and hadronic top decays.

(a) Leptonic channel: Events enter this channel if they satisfy the following kinematic selections:

* ≥1 Lepton ($p_T > 20$ GeV) non-overlapping with any photon.
* ≥2 Jets, ($p_T > 25$ GeV, $|\eta| < 2.4$) non-overlapping with any photon or lepton.
* ≥1 Medium b-tagged Jet.
* Di-photon BDT score > 0.11.

(b) Hadronic channel: This comprises events passing the following kinematic selections:

* ≥3 Jets, ($p_T > 25$ GeV, $|\eta| < 2.4$) non-overlapping with any photon.
* ≥1 Loose b-tagged Jet.
* No Lepton in the event (passing the lepton selections of the Leptonic channel described above).
* High score (>0.75) on the $t\bar{t}H$multi-jet tagging BDT (described above).
* Di-photon BDT score > 0.4.

In addition to this, di-photon invariant mass dependent p_T cuts are applied to selected photons in both channels to get a distortion free di-photon mass spectrum.

---

[3] Deep Learning Neural Networks.

[4] Boosted Decision Trees.

[5] $E_T^{\gamma 1} > 30$ GeV, $E_T^{\gamma 2} > 18$ GeV, Loose ECAL based photon identification.

[6] Vector boson fusion.

3. ***$t\bar{t}H$* Multi-lepton**: In this search, events are first selected using lepton or lepton $+\tau_h$ triggers. They are then required to have $\geq$2 loose b-tagged jets[7] out of which $\geq$1 is medium b-tagged. A special BDT trained on simulated $t\bar{t}H(t\bar{t})$ events as signal (background) is used to distinguish "prompt" leptons (produced by W/Z/leptonic τ decays) from "non-prompt" leptons (produced in b-hadron decays, decays-in-flight, and photon conversions). Leptons passing (failing) it are called tight (loose) leptons in this analysis. An additional di-lepton invariant mass cut ($m_{ll} > 12$ GeV) is applied to all channels to reject phase space dominated by low mass SM di-lepton resonances (e.g., J/ψ, Υ etc.) which are not well modeled in simulation. The major channels of this search (which are also split by lepton flavor) are

(a) **2 Lepton Same-Sign channel (2*l*SS)**: Events containing exactly 2 same charge tight leptons ($p_T^{lep1/2} > 25/15$ GeV) and ≥ 4 Jets. Additional selections are applied to reduce backgrounds due to conversions and lepton charge mis-identification.

(b) **3 Lepton channel (3*l*)**: Events containing exactly 3 leptons ($p_T^{lep1/2/3} >$ 25/15/15 GeV) with additional selections for rejection of backgrounds due to $Z \rightarrow ll$ events and conversions.

(c) **4 Lepton channel (4*l*)**: Events passing exactly the same selections as 3l channel but now having an additional requirement of a fourth lepton ($p_T > 10$ GeV) in the event.

(d) **1 Lepton + 2τ_h channel (1*l*+2τ_h)**: Events containing exactly 1 tight lepton ($p_T^{e/\mu} > 25/20$ GeV, $|\eta| < 2.1$) and 2 opposite charge tight τ_h leptons ($p_T > 30$ GeV each). Events should also contain $\geq$ 3 Jets.

(e) **2 Lepton Same-Sign + 1τ_h channel (2*l*SS+1τ_h)**: Events containing exactly 2 same charge tight leptons ($p_T^{lep1/2} > 25/15$ GeV for electrons, $p_T^{lep1/2} > 25/10$ GeV for muons) and 1 tight τ_h ($p_T > 30$ GeV) with sign opposite to that of the leptons. Events should also contain $\geq$ 3 Jets. Additional selections were applied to reduce background due to charge mis-identification and $Z \rightarrow ll$.

(f) **3 Lepton + 1τ_h channel (3*l*+1τ_h)**: Events containing $\geq$ 3 tight leptons ($p_T^{lep1/2/3} > 20/10/10$ GeV) and $\geq$ 1 tight τ_h ($p_T > 30$ GeV). Sum of the charges of the leptons and τ_h must be zero. Additional selections were applied to reduce background due to charge mis-identification and $Z \rightarrow ll$.

Events containing τ_h are vetoed in the pure leptonic channels ((a), (b), and (c)) to keep them orthogonal to the lepton + τ_h channels ((d), (e), and (f)) For the 2lSS and 3l channels, a pair of BDTs (one each for training against $t\bar{t}V$ and $t\bar{t}$ backgrounds vs the $t\bar{t}H$signal) was used for enhanced sensitivity. The BDTs for 2lSS included discriminators for "hadronic top tagging". For signal extraction, the BDT pairs were mapped onto one dimension using "Likelihood based clustering". The minimum invariant mass of the opposite sign di-lepton pair was used for signal extraction for the 4l channel. BDT (trained to distinguish $t\bar{t}H$from $t\bar{t}$) was used as final discriminator in 1l+2τ_h channel. MEM (useful in separating $t\bar{t}H$from $t\bar{t}V$ and $t\bar{t}$ backgrounds)

[7] $p_T > 25$ GeV, $|\eta| <$2.4, non-ovelapping with any lepton/τ_h.

was used for signal extraction in $2l$SS+1τ_h channel. BDT pair (one each trained against $t\bar{t}V$ and $t\bar{t}$) mapped into a one-dimensional discriminant was used for signal extraction in $3l$+1τ_h channel.

3.3 Backgrounds

Backgrounds in all the $t\bar{t}H$ searches can be classified into three major categories depending on their source/origin.

1. **Reducible/Fake backgrounds**: These backgrounds arise predominantly due to Jets faking the final state particles (γ/leptons/τ_h or b-jets) used in the searches. They are mostly either estimated from data or from simulation (with their yield and shape corrected from background rich sidebands in data). In the $t\bar{t}H$ Multi-Lepton search, the probability of a Jet to pass tight selections is measured in a multi-jet enriched sideband in data (measurement region). This probability is then used to reweight another data sideband having the same selections as the signal region but with relaxed lepton identification requirements (application region). This reweighted sideband is then used to estimate fake background in all channels.
2. **Charge Flip background**: This background is caused due to mis-identification of lepton charge inside the detector (due to inelastic scattering or missing hits in the tracker). It is measured from data via. "Tag and Probe" method in $Z \rightarrow ll$ events in bins of lepton p_T and $|\eta|$. It is measured to be $\sim 10^{-3}$ for electrons and was found to be negligible for muons.
3. **Ir-reducible backgrounds**: This background is caused by genuine physical processes having the same final state particles as in the signal process. They are usually estimated via. Simulation and their modeling validated using control regions in data. Examples include $t\bar{t}V$ and $WW/WZ/ZZ$ backgrounds.

3.4 Results

Results of the above mentioned $t\bar{t}H$ searches with the data collected by CMS in 2016 (35.9 fb^{-1} at $\sqrt{s} = 13$ TeV) were combined with the CMS Run-1 dataset[8] to obtain a "5 σ discovery" of $t\bar{t}H$ process (Fig. 3.2) [9]. For the 2017 dataset,[9] public results were available only for the $t\bar{t}H$ Multi-lepton search at the time of the symposium [10]. A new category ($2l + 2\tau_h$) was added by this search on top of the existing ones and the combined (2016 + 2017) signal strength from the $t\bar{t}H$ Multi-lepton search alone equals $0.96^{+0.34}_{-0.31}$ ($1.00^{+0.30}_{-0.27}$) observed (expected) in units of the SM expectation.

[8] 5.1 (19.7) fb^{-1} collected at $\sqrt{s} = 7$ (8) TeV.
[9] 41.4 fb^{-1} at $\sqrt{s} = 13$ TeV.

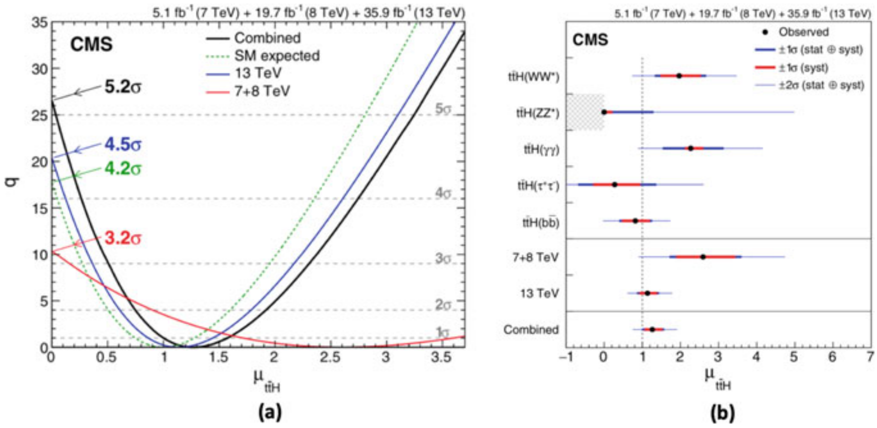

Fig. 3.2 Results of the CMS Run-1+Run-2 $t\bar{t}H$ combination. Likelihood scan of the signal strength (**a**) and Channel-wise signal strength split by era (**b**)

References

1. CMS collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B716, 30 (2012)
2. CMS collaboration, The CMS experiment at the CERN LHC. J. Instrum. **3**(08), S08004 (2008)
3. CMS collaboration, Search for ttH production in the all-jet final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. J. High Energy Phy. (6) (2018). arXiv:1803.06986
4. CMS collaboration, Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-protoncollisions at $\sqrt{s} = 13$ TeV. J. High Energy Phys. (3), 26 (2019). arXiv:1804.03682
5. CMS collaboration, Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV. J. High Energy Physics. (11), 185 (2018). arXiv:1804.02716
6. 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. J. High Energy Phys. (8), 66 (2018). arXiv:1803.05485
7. CMS collaboration, Search for Higgs boson production in association with top quarks in multilepton final states at $\sqrt{s} = 13$ TeV, CMS Physics Analysis Summary CMS-PAS-HIG-17-004
8. CMS collaboration, Search for the associated production of a Higgs boson with a top quark pair in final states with a τ lepton at $\sqrt{s} = 13$ TeV, CMS Physics Analysis Summary CMS-PAS-HIG-17-003
9. CMS collaboration, Observation of $t\bar{t}H$ production. Phys. Rev. Lett. **120**(23), 231801. arXiv:1804.02610
10. CMS collaboration, Measurement of the associated production of a Higgs boson with a top quark pair in final states with electrons, muons and hadronically decaying τ leptons in data recorded in 2017 at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-HIG-18-019