Latest CMS results on Higgs boson production in association with top quarks ($t\bar{t}H$)

Saranya Ghosh  
(DPhP, IRFU, CEA Paris-Saclay),

For the CMS collaboration

EPS-HEP 2017 : Venice, Italy : 7TH JULY 2017
INTRODUCTION

- Production of the Higgs boson in association with a pair of top quarks (ttH) probes the **top quark Yukawa coupling**
  - Direct handle on the ttH vertex at tree level: gluon fusion involves a loop
  - Cross-section of ttH is \(~508\) fb at 13 TeV, roughly 4 times the value at 8 TeV, increased potential for discovery

- Searches for ttH at CMS in Run 2: based on decay channel of the Higgs
  - Multi-leptonic (H→ZZ*, H→WW*, H→ττ): higher rate, multi-lepton final state with low background
  - τH (H→ττ, also H→ZZ*, H→WW*): One τ decaying hadronically, significant ‘fake’ background
  - H→γγ: small branching ratio, but clean final state (low systematic uncertainties). Excellent diphoton mass resolution
  - H→b̅b: high branching ratio, but complex multi-jet final state
**t\bar{t}H MULTI-LEPTONIC**

- Multi-lepton final states from $H \rightarrow WW^*$, $H \rightarrow ZZ^*$, $H \rightarrow \tau\tau$ (for $\tau$ leptonic decays)

- Event selection:
  - **Same-sign dilepton channel (2LSS):**
    - 2 leptons (e or $\mu$) with same sign: ee, $\mu\mu$, $e\mu$
    - ≥ 4 jets

  - **Three lepton channel (3L):**
    - 3 leptons
    - ≥ 2 jets

  - **Four lepton channel (4L):**
    - ≥ 4 leptons
    - ≥ 2 jets

- **B Tag jets:** at least 1 jet passing medium WP or 2 jets passing loose WP of B tag algorithm

- **Z veto:** based on $m_{\ell\ell}$, $E_T^{\text{miss}}$, $H_T^{\text{miss}}$

- **$\tau_H$ veto:** veto events containing reconstructed $\tau_H$; for orthogonality with analysis for $t\bar{t}H$ in the $\tau_H$ channel

- 2LSS and 3L categories have further sub-categories based on charge of leptons, B Tag criteria
**t\(\bar{t}\)H Multi-Leptonic: Signal Extraction**

<table>
<thead>
<tr>
<th>(2LSS)</th>
<th>(3L)</th>
<th>(4L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t\bar{t}W)</td>
<td>51.0 (\pm) 0.6 (stat.) (\pm) 6.9 (syst.)</td>
<td>72.8 (\pm) 0.7 (stat.) (\pm) 10.2 (syst.)</td>
</tr>
<tr>
<td>(t\bar{t}Z/\gamma^*)</td>
<td>17.7 (\pm) 0.8 (stat.) (\pm) 2.9 (syst.)</td>
<td>47.3 (\pm) 1.6 (stat.) (\pm) 9.0 (syst.)</td>
</tr>
<tr>
<td>WZ</td>
<td>4.2 (\pm) 0.6 (stat.) (\pm) 4.1 (syst.)</td>
<td>7.0 (\pm) 0.8 (stat.) (\pm) 6.8 (syst.)</td>
</tr>
<tr>
<td>Rare SM bkg.</td>
<td>4.2 (\pm) 1.5 (stat.) (\pm) 3.0 (syst.)</td>
<td>13.3 (\pm) 1.9 (stat.) (\pm) 9.3 (syst.)</td>
</tr>
<tr>
<td>WWss</td>
<td>3.5 (\pm) 0.6 (stat.) (\pm) 2.5 (syst.)</td>
<td>4.1 (\pm) 0.6 (stat.) (\pm) 3.2 (syst.)</td>
</tr>
<tr>
<td>Conversions</td>
<td>120.3 (\pm) 2.5 (stat.) (\pm) 11.7 (syst.)</td>
<td>231.2 (\pm) 4.3 (stat.) (\pm) 13.3 (syst.)</td>
</tr>
<tr>
<td>Non-prompt leptons</td>
<td>38.7 (\pm) 1.6 (stat.) (\pm) 20.5 (syst.)</td>
<td>61.8 (\pm) 2.0 (stat.) (\pm) 13.0 (syst.)</td>
</tr>
<tr>
<td>All backgrounds</td>
<td>199.7 (\pm) 2.5 (stat.) (\pm) 21.7 (syst.)</td>
<td>392.0 (\pm) 4.3 (stat.) (\pm) 33.3 (syst.)</td>
</tr>
<tr>
<td>t(\bar{t})H signal</td>
<td>20.1 (\pm) 0.5 (stat.) (\pm) 2.1 (syst.)</td>
<td>27.9 (\pm) 0.5 (stat.) (\pm) 3.0 (syst.)</td>
</tr>
<tr>
<td>Data</td>
<td>150</td>
<td>268</td>
</tr>
</tbody>
</table>

**Background estimation:**

- **Signal like final states:** \(t\bar{t}V, Diboson\) : estimated from MC simulation

- **Others:** Non-prompt leptons (largely from \(t\bar{t}\)), charge mis-measured leptons : Data driven estimation

**Signal extraction:**

- Multivariate BDT discriminants trained in simulated events to separate the signal from \(t\bar{t}V\) backgrounds and also non-prompt \(t\bar{t}\) backgrounds (not for statistically limited 4L)

- 2 separate BDTs trained using kinematical observables.
  - \(\eta\) of leptons, jet multiplicity, distance between lepton & jet, \(m_T, E_T\)
  - Against \(t\bar{t}\) : 2LSS : jet BDT discriminator score that is used to identify jets from top decay
  - Against \(t\bar{t}V\) : leading, trailing lepton \(p_T\)

  for **2LSS** : jet BDT discriminator score that is used to identify jets from Higgs decay

  for **3L category** : MEM weight :

  \[
  w_{i\alpha}(\Phi') = \frac{1}{\sigma_\alpha} \int d\Phi_\alpha \delta^4 \left( p_T^i + p_T^\nu - \sum_{k=2}^{n} p_T^\nu_k \right) \cdot \frac{f(x_1, \mu_T^i)f(x_2, \mu_T^\nu)}{x_1 x_2 s} \cdot \left| M_\alpha(p_T^\nu) \right|^2 \cdot W(\Phi'|\Phi_\alpha)
  \]
• Distribution of discriminating BDT outputs after fit to data, combined from different sub-categories, for 2LSS and 3L categories
Results based on 35.9 fb\(^{-1}\) of data collected during 2016 and combination with 2.3 fb\(^{-1}\) collected in 2015.

- Main sources of systematic uncertainties:
  - Lepton selection efficiency, modelling of non-prompt backgrounds, charge mis-identification, diboson background modelling, theoretical uncertainties

- Signal significance observed (expected) : 3.3\(\sigma\) (2.5\(\sigma\))
**t\bar{t}H TAU-HADRONIC (τ_H)**

- **Tau-lepton decaying hadronically in final state:** $H \rightarrow \tau\tau$ (also $H \rightarrow WW^*$, $H \rightarrow ZZ^*$)

- **Event selection:**
  
  - $1\ell + 2\tau_H$:
    - 1 lepton (e or $\mu$)
    - 2 $\tau_H$ opposite charge, ‘tight’ WP of $\tau_H$ ID, $p_T$ > 30, 20 GeV
    - $\geq 3$ hadronic jets
  
  - $3\ell + 1\tau_H$:
    - $\geq 3$ leptons of
    - 1 $\tau_H$ sum of charge with $\ell = 0$, ‘medium’ WP of $\tau_H$ ID, $p_T$ > 20 GeV
    - $\geq 2$ hadronic jets
  
  - $2\ell ss + 1\tau_H$:
    - 2 leptons of same sign
    - 1 $\tau_H$ opposite charge wrt $\ell$, ‘medium’ WP of $\tau_H$ ID, $p_T$ > 20 GeV
    - $\geq 3$ hadronic jets
    - sub-categories:
      - no-missing-jet: full reconstruction of $t\bar{t}H \rightarrow b\bar{W}bW\tau\tau \rightarrow b\bar{J}J\ell\nu\bar{\nu}\ell\nu\tau\bar{\nu}$
      - missing-jet: one jet missing: outside acceptance overlapping

**τ_H identification:** Dedicated MVA based discriminant (isolation, impact parameter, $\tau$ lifetime) with “hadrons plus strips” (HPS) algorithm to identify

\[
\begin{align*}
\tau^+ & \rightarrow h^+\nu_{\tau} \\
\tau^\mp & \rightarrow h^\pm\pi^0\nu_{\tau} \\
\tau^\pm & \rightarrow h^\pm\pi^0\pi^0\nu_{\tau} \\
\tau^\pm & \rightarrow h^\pm h^\mp h^\pm\nu_{\tau}
\end{align*}
\]
### Expected and observed yields after the selection in the different categories

<table>
<thead>
<tr>
<th>Process</th>
<th>$1\ell + 2\tau_H$</th>
<th>$3\ell + 1\tau_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}H$, $H \rightarrow \tau\tau$</td>
<td>$2.84 \pm 1.35$</td>
<td>$1.01 \pm 0.65$</td>
</tr>
<tr>
<td>$t\bar{t}H$, $H \rightarrow WW$</td>
<td>$0.07 \pm 0.04$</td>
<td>$0.63 \pm 0.29$</td>
</tr>
<tr>
<td>$t\bar{t}H$, $H \rightarrow ZZ$</td>
<td>$0.02 \pm 0.01$</td>
<td>$0.09 \pm 0.04$</td>
</tr>
<tr>
<td>$t\bar{t}Z$</td>
<td>$4.07 \pm 0.56$</td>
<td>$3.78 \pm 0.62$</td>
</tr>
<tr>
<td>$t\bar{t}W$</td>
<td>$0.21 \pm 0.05$</td>
<td>$0.24 \pm 0.05$</td>
</tr>
<tr>
<td>Electroweak</td>
<td>$1.10 \pm 1.05$</td>
<td>$0.32 \pm 0.05$</td>
</tr>
<tr>
<td>Fake</td>
<td>$20.98 \pm 3.87$</td>
<td>$1.07 \pm 0.34$</td>
</tr>
<tr>
<td>Other</td>
<td>$0.54 \pm 0.23$</td>
<td>$0.24 \pm 0.08$</td>
</tr>
<tr>
<td>Total expected background</td>
<td>$26.91 \pm 3.84$</td>
<td>$5.65 \pm 0.85$</td>
</tr>
<tr>
<td>SM expectation</td>
<td>$29.85 \pm 4.07$</td>
<td>$7.38 \pm 1.10$</td>
</tr>
<tr>
<td>Observed data</td>
<td>$24$</td>
<td>$7$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Process</th>
<th>$2\ell ss + 1\tau_H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}H$, $H \rightarrow \tau\tau$</td>
<td>$1.38 \pm 0.89$</td>
</tr>
<tr>
<td>$t\bar{t}H$, $H \rightarrow WW$</td>
<td>$1.03 \pm 0.47$</td>
</tr>
<tr>
<td>$t\bar{t}H$, $H \rightarrow ZZ$</td>
<td>$0.06 \pm 0.03$</td>
</tr>
<tr>
<td>$t\bar{t}Z$</td>
<td>$3.07 \pm 0.46$</td>
</tr>
<tr>
<td>$t\bar{t}W$</td>
<td>$1.10 \pm 0.15$</td>
</tr>
<tr>
<td>Electroweak</td>
<td>$0.21 \pm 0.19$</td>
</tr>
<tr>
<td>Fake</td>
<td>$1.66 \pm 0.52$</td>
</tr>
<tr>
<td>Charge flip</td>
<td>$0.05 \pm 0.01$</td>
</tr>
<tr>
<td>Other</td>
<td>$0.50 \pm 0.20$</td>
</tr>
<tr>
<td>Total expected background</td>
<td>$6.59 \pm 0.88$</td>
</tr>
<tr>
<td>SM expectation</td>
<td>$9.06 \pm 1.33$</td>
</tr>
<tr>
<td>Observed data</td>
<td>$8$</td>
</tr>
</tbody>
</table>

### Background estimation:

- **Irreducible**: $t\bar{t}V$, Di-boson+jets (estimated from MC, validated in data)

- **Reducible**: Non-prompt leptons (largely from $t\bar{t}$), charge mis-measured leptons: Data driven estimation. Fake $\tau_H$: data driven estimate in $1\ell+2\tau_H$ category, simulation based estimate in other categories.

### Signal extraction:

- Signal extraction through maximum likelihood fit to the distribution of discriminating observable

- Multivariate BDT discriminants to separate the signal from $t\bar{t}$, $t\bar{t}V$ backgrounds, separately in each category

- Additionally MEM based discriminant used in $2\ell ss+1\tau_H$ category to discriminate signal from $t\bar{t}V$ backgrounds
**Results based on 35.9 fb\(^{-1}\) of data collected during 2016**

**Main sources of systematic uncertainties:**
- Affects yield: Charge mis-measurement, lepton & \(\tau_H\) efficiency, theory uncertainty
- Affects shape: Jet & \(\tau_H\) energy scale, b-tag efficiency, fake background estimation, modelling of signal and background in simulation

**Signal significance observed (expected): 1.4\(\sigma\) (1.8\(\sigma\))**
A part of the general $H \rightarrow \gamma\gamma$ analysis

- Events with two high $p_T$ isolated photons selected
- Narrow peak around $m_H$ on top of the falling $m_{\gamma\gamma}$ distribution
- Different production modes ($t\bar{t}H$, VBH, VH) identified based on additional final state objects
- Signal, background extraction from fit to $m_{\gamma\gamma}$ distribution

Overview of the $H \rightarrow \gamma\gamma$ analysis:

- Event categorisation for different production modes of Higgs based on
  - additional final state objects to tag production modes
  - mass resolution and kinematics for the 'Untagged' categories
**ttH with H\rightarrow\gamma\gamma : t\bar{t}H CATEGORIES**

- 2 categories corresponding to t\bar{t}H based on the decay of the top quarks

**TTH Leptonic Tag:**
\[ t\bar{t} \rightarrow b\nu_l b\bar{q}q' \text{ or } t\bar{t} \rightarrow b\nu_l b'\nu_l \]

- Cut based approach:
  - \( \geq 1 \) lepton (muon or electron)
  - \( \geq 2 \) jets
  - \( \geq 1 \) B-tagged jet
  - Diphoton BDT cut

**TTH Hadronic Tag:**
\[ t\bar{t} \rightarrow bq\bar{q}' \]

- Loose selection:
  - \( \geq 3 \) jets
  - \( \geq 1 \) B-jet (loose WP)
  - 0 leptons

- MVA with inputs:
  - number of jets
  - lead jet \( p_T \)
  - Max B-tag discriminant value
  - Second max B-tag value

- Diphoton BDT cut

**Expected signal for t\bar{t}H categories:** very pure in t\bar{t}H contribution

<table>
<thead>
<tr>
<th>Event Categories</th>
<th>SM 125 GeV Higgs boson expected signal</th>
<th>Bkg (GeV(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>ttH Hadronic</td>
<td>5.85 10.99% 0.70% 77.54% 2.02% 4.13% 2.02% 0.09% 0.05% 0.63% 1.82% 1.48</td>
<td>2.40</td>
</tr>
<tr>
<td>ttH Leptonic</td>
<td>3.81 1.90% 0.05% 87.48% 0.08% 4.73% 3.04% 1.53% 1.15% 0.02% 0.02% 1.60 1.35</td>
<td>1.50</td>
</tr>
</tbody>
</table>
**ttH WITH H→γγ : RESULTS**

- Results based on $35.9 \text{ fb}^{-1}$ of data at 13 TeV collected during 2016

**Signal strength for each process**

- Uncertainties are statistics dominated
- Signal significance observed (expected) : $3.3\sigma$ (1.5σ)
Search for $t\bar{t}H$ at CMS in Run 2 (13 TeV) are being performed in different decay modes of the Higgs, with complex analysis techniques. Latest results are compatible with standard model:

- **Multi-leptonic ($H\rightarrow ZZ^*$, $H\rightarrow WW^*$, $H\rightarrow \tau\tau$):**
  
  $\mu_{t\bar{t}H} = 1.5^{+0.5}_{-0.5}$ (± 0.29 stat. ± 0.24 theo. ± 0.32 exp.)

  - Observed (expected) significance of $3.3\sigma$ ($2.5\sigma$)

- **$\tau_H$ ($H\rightarrow \tau\tau$, also $H\rightarrow ZZ^*$, $H\rightarrow WW^*$):**
  
  $\mu_{t\bar{t}H} = 0.72^{+0.62}_{-0.53}$

  - Observed (expected) significance of $1.4\sigma$ ($1.8\sigma$)

- **$H\rightarrow \gamma\gamma$:**
  
  $\mu_{t\bar{t}H} = 2.2^{+0.9}_{-0.8}$ (stat. dominated)

  - Observed (expected) significance of $3.3\sigma$ ($1.5\sigma$)

- **$H\rightarrow b\bar{b}$:**
  
  $\mu_{t\bar{t}H} = -0.19^{+0.8}_{-0.81}$ ($^{+0.45}_{-0.44}$ stat. $^{+0.66}_{-0.68}$ syst.)

- **Data taking** for 2017 underway, more updates to come from Run 2…

The data says we need more data.
SUMMARY OF RUN-1 RESULTS

- Studies of the $t\bar{t}H$ production in LHC Run-1 at CMS were based on the different Higgs decay channels: $\gamma\gamma$, $b\bar{b}$, $\tau\tau$ (hadronic) and multi-leptonic ($H\rightarrow ZZ^*$, $H\rightarrow WW^*$, $H\rightarrow \tau\tau$ with multi-lepton final states).

- Combination of different Higgs decay channels:
  - $t\bar{t}H$ combination from Run1 CMS measurements:
    - $\mu_{t\bar{t}H} = 2.8 \pm 1.0$
    - 95% CL limit = 4.5
    - Best fit signal strength: $\mu_{t\bar{t}H}$

- 95% CL upper limit $\mu_{t\bar{t}H}$

- Observed $\sigma/\sigma_{SM}$ at $m_H = 125.6$ GeV
CMS AT THE LHC

- Compact Muon Solenoid (CMS) is a general purpose detector at the LHC
  - Multiple layers based around the solenoid magnet to detect different kinds of particles
  - \( ttH \) process produces jets, leptons along with photons, other sub detectors also play crucial role

- LHC operations for proton-proton collisions
  - At center of mass energies of 7 TeV (2010-11) and 8 TeV (2012) during Run 1
  - At center of mass energy of 13 TeV during Run 2 (2015-18)
  - CMS is expected to collect 100 fb\(^{-1}\) by the end of Run 2 (2018) and 300 fb\(^{-1}\) by the end of Phase 1 (2022)

- Analyses presented in this talk: 35.9 fb\(^{-1}\) of data at 13 TeV collected in 2016
Search for $\bar{t}tH$ at CMS in Run 2 (13 TeV) are being performed in different decay modes of the Higgs, with complex analysis techniques. Latest results are compatible with standard model:

- Multi-leptonic ($H \rightarrow ZZ^*, H \rightarrow WW^*, H \rightarrow \tau\tau$): $35.9 \text{ fb} \ (2016) + 2.3 \text{ fb} \ (2015)$
  
  $\mu_{t\bar{t}H} = 1.5^{+0.5}_{-0.5} \ (\pm 0.29 \text{ stat.} \pm 0.24 \text{ theo.} \pm 0.32 \text{ exp.})$

  - observed (expected) significance of $3.3\sigma \ (2.5\sigma)$
  - observed (expected) limits on $\mu_{t\bar{t}H} = 2.5 \ (0.8 \text{ without SM } t\bar{t}H)$

- $\tau_H \ (H \rightarrow \tau\tau, \text{also } H \rightarrow ZZ^*, H \rightarrow WW^*)$: $35.9 \text{ fb} \ (2016)$
  
  $\mu_{t\bar{t}H} = 0.72^{+0.62}_{-0.53}$

  - observed (expected) significance of $1.4\sigma \ (1.8\sigma)$
  - observed (expected) limits on $\mu_{t\bar{t}H} = 2.0 \ (2.2 \text{ with SM } t\bar{t}H; 1.1 \text{ without SM } t\bar{t}H)$

- $H \rightarrow \gamma\gamma$: $35.9 \text{ fb} \ (2016)$
  
  $\mu_{t\bar{t}H} = 2.2^{+0.9}_{-0.8} \ (\text{stat. dominated})$

  - observed (expected) significance of $3.3\sigma \ (1.5\sigma)$

- $H \rightarrow b\bar{b}$: $12.9 \text{ fb} \ (2016)$
  
  $\mu_{t\bar{t}H} = -0.19^{+0.8}_{-0.81} \ (^{+0.45}_{-0.44} \text{ stat.}^{+0.66}_{-0.68} \text{ syst.})$

  - observed (expected) limits on $\mu_{t\bar{t}H} = 1.5 \ (1.7)$
Search for $t\bar{t}H$ at CMS

**t\bar{t}H Tau-Hadronic ($\tau_H$): Signal Extraction**

- **Events**
  - Observed
  - Fakes
  - $t\bar{t}H, H \rightarrow \tau\tau$
  - $t\bar{t}H, H \rightarrow WW/ZZ$

- **Data - Expectation**
  - $1\ell+2\tau_H$

- **MEM LR**
  - $0.1$ to $1$

- **MVA**
  - $0$ to $1$

- **35.9 fb$^{-1}$ (13 TeV)**

- **CMS Preliminary**

- **Events**
  - Electroweak
  - $t\bar{t}H, H \rightarrow \tau\tau$
  - $t\bar{t}H, H \rightarrow WW/ZZ$
  - $t\bar{t}Z$
  - $t\bar{t}W$
  - $1\ell+2\tau_H$

- **Data - Expectation**
  - $0.5$ to $1$

- **MEM LR**
  - $0$ to $1$

- **Events**
  - Electroweak
  - $t\bar{t}H, H \rightarrow \tau\tau$
  - $t\bar{t}H, H \rightarrow WW/ZZ$
  - $t\bar{t}Z$
  - $t\bar{t}W$
  - $3\ell+1\tau_H$

- **Data - Expectation**
  - $0.5$ to $1$

- **MEM LR**
  - $0$ to $1$

- **Events**
  - Electroweak
  - $t\bar{t}H, H \rightarrow \tau\tau$
  - $t\bar{t}H, H \rightarrow WW/ZZ$
  - $t\bar{t}Z$
  - $t\bar{t}W$
  - $2\ell+1\tau_H$
  - no-missing-jet

- **Data - Expectation**
  - $0.5$ to $1$

- **MEM LR**
  - $0$ to $1$

- **Events**
  - Electroweak
  - $t\bar{t}H, H \rightarrow \tau\tau$
  - $t\bar{t}H, H \rightarrow WW/ZZ$
  - $t\bar{t}Z$
  - $t\bar{t}W$
  - $2\ell+1\tau_H$
  - missing-jet

- **Data - Expectation**
  - $0.5$ to $1$

- **MEM LR**
  - $0$ to $1$

- **Saranya Ghosh**
  - EPS-HEP 2017