Associated Higgs production with top quarks

GK Krintiras on behalf of CMS Collaboration
Living dangerously

In SM, **Yukawa couplings** are determined uniquely by the particles' mass

We measure $m_t = 172.44(0.4\%) \rightarrow$ ends up to $|y_t| = 0.990(0.3\%)$, i.e. **almost 1**


If desert exists up to $M_{Planck}$

For a non dimensional parameter that's at least interesting

*Degrassi * et al., JHEP08, 098 (2012)
Associated Higgs (H) production with top (t) quarks

Measure $y_t$ through observables independent from $m_t$

- sensitive already at tree-level
- relatively small event rate

$ttH$ production: $\sim 1/100$ of $ggH$

$Probes$ the modulus of $y_t$

$tH$ production: $\sim 1/30$ of $ttH$

$Probes$ the relative sign of $y_t$
Distinct experimental signatures of $ttH$ and $tH$

- Searches are complementary, and could be
- grouped by the interplay of efficiency & purity over the studied:
  - **Hadronic** event activity
    - $H \rightarrow bb$, $H \rightarrow \tau \tau$
  - **Leptonic** event activity
    - $H \rightarrow WW$, $H \rightarrow \ell \ell$
  - **Bosonic** event activity
    - $H \rightarrow \gamma \gamma$, $H \rightarrow ZZ$

Probes the modulus

Probes the relative sign

Remember!
So, what's currently the status of analyses that study the modulus?

- Three pivotal updates with 36 fb-1 of data
  - all in multilepton final state

**HIG-17-003**

- $ttH(\tau_l/h\tau_h)$

**HIG-17-004**

- $H(\tau_l/h\tau_h)$

**HIG-17-004**

- 2$\ell$s

**HIG-17-004**

- +4$\ell$ category

**HIG-17-003**

- $ttH(WW/ZZ*/\tau\tau\ell\ell)$

**HIG-17-004**

- L with hadronic tau
So, what's currently the status of analyses that study the modulus?

- Three pivotal updates with 36 fb-1 of data
  - all in **multilepton** final state

- No m₄ₜ overlap between HIG-17-004 and HIG-16-041

---

HIG-16-041

$$H \rightarrow ZZ \rightarrow 4\ell$$

$$m_{4\ell} \in [118, 130]$$

<table>
<thead>
<tr>
<th>Signal</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bkg.</td>
<td>0.3</td>
</tr>
<tr>
<td>Data</td>
<td>0.0</td>
</tr>
</tbody>
</table>

13 fb-1

---

**CMS**

$$\text{ttH production}$$

Run 2

$$\text{Preiminary}$$

$$\tau_{n} + X$$

$$\mu^+\mu^-$$

Let's fill it!
Treating the $ttH(WW/ZZ*/\tau\tau\tau)$ signal with care

- Irreducible bkg. from $tt + V$ bosons
- Two BDTs ($ttH$ vs $ttV$ and $ttH$ vs $tt$) built per lepton category
  - 1D bins of 2D space spanned by the two BDTs
- Reducible bkg. from non-prompt $l$ and $l^\pm$ confusion from control regions

- hadronic top “tagger”
- $H\rightarrow WW(qq)$ “tagger”

HIG-17-004

Binning to ~ flatten bkg.
Treating the $ttH(\bbbar/ZZ*/\tau\tau\tau)$ signal with care

- Irreducible bkg. from $tt + V$ bosons
- Two BDTs ($ttH$ vs $ttV$ and $ttH$ vs $tt$) built per lepton category
  - 1D bins of 2D space spanned by the two BDTs
- Reducible bkg. from non-prompt $l$ and $l^\pm$ confusion
  - from control regions

HIG-17-004
Treating the $ttH(\text{WW/ZZ}*/\tau\tau\tau)$ signal with care

- Measurement limited by **statistics** and **systematics**
  - (almost) **equally dominated** by theo. and exp. uncertainties
- Best fit $ttH$ yield $1.5\pm0.5 \times$ the SM prediction at $3.3\sigma(2.5\sigma)$ level
  - no visible impact from the addition of the 2015 dataset (2.7 fb-1)
  - compatible with SM; cross-checked with floating $ttV$ in the fit
The first $ttH(\tau_l/\tau_h)$ reconstruction in Run2

- $1l + \tau_h\tau_h$, $2lSS + \tau_h$ and $3l + \tau_h$ categories
- MVAs to increase the sensitivity per category
- bkg. similar to the “multileptons” case
- irreducible from MC; reducible from data

$2\ell_{ss} + 1\tau_h$

$H \rightarrow WW(qq)$ reconstructed

$H \rightarrow WW(qq)$ not reconstructed

$1\ell + 2\tau_h$

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

CMS Preliminary

- Observed
- $t\bar{t}H, H \rightarrow \tau\tau$
- $t\bar{t}H, H \rightarrow WW/ZZ$
- $t\bar{t}Z$
- $t\bar{t}W$

Electroweak
Rares
Fakes
Flips
Uncertainty

$2\ell_{ss} + 1\tau_h$
no-missing-jet

$2\ell_{ss} + 1\tau_h$
missing-jet

$3\ell + 1\tau_h$

Merged 2D BDT bin number $\Delta_{DMVA}$
The first $ttH(\tau_{l/h} \tau_{h})$ reconstruction in Run2

- Measurement limited by **statistics** and **systematics**
  - dominated both by irreducible and reducible bkg.
- Significance of $ttH$ production at $1.4\sigma (1.8\sigma)$ level
  - upper limit set at $2.0 (2.2) \times$ the SM $ttH$ production rate

![Graphs showing $ttH$ production and significance levels](image-url)
How to deal with the overwhelming bkg. in $ttH(\bb)$

- Important to resolve jet combinatorics in jets/b-tags categories
  - MVAs in a 2D fit,
    - improves MC prediction, and
    - reduces uncertainties
    - correlations well modeled

**Conditions**

- BDT for $ttH$ vs $tt + \text{jets}$
- MEM for $ttH$ vs $tt + bb$ hypotheses

**Significance**

- 12.9 fb$^{-1}$ (13 TeV)

- Before any fit

- After the 2D fit

**Tables and Figures**

- CMS Preliminary
- Dilepton
- Lepton+jets
- Combined

- $\mu$, tot. stat. syst.

- Best fit $\sigma/\sigma_{SM}$ at $m_H = 125$ GeV

**HIG-16-038**
Where the sensitivity for the $ttH(γγ)$ stands?

- Look for **small but clean** $m_{γγ}$ peak with
  - an expected photon energy resolution of $\sim 1\%$
  - separately in **leptonic** (95% pure) & **hadronic** (80% pure) $tt$ decays
- limited by **statistics**

**Background**: functional forms driven by data

**Signal**: $n$Gaussian model fit to MC
Best fit at Run1 favors the SM solution (★) within 5σ (!)

But, what if beyond SM Physics is hidden in loops (?)

a direct way to lift the sign ambiguity is tH production

- unitarity and renormalizability are spoiled in $c_V \neq c_F$ deviations
- i.e. enhancement of amplitude $\rightarrow$ higher production rate
First derived limits with Run2 data: \( tH(bb) \)

- Jet assignment crucial
- MVA for \( tH \) vs \( tt \) hypotheses
- A second MVA for \( tH \) extraction
- first time done for multiple \( k_t/k_v \) scenarios
Key improvements in $ttH$; obviously more data

- Multileptons & dedicated $\tau_h$ searches
  - improved handling of reducible backgrounds
  - better signal extraction and validation of $ttV$
- $bb$ searches
  - improvement of the signal extraction
  - modeling of $tt+(b-)\text{ jets}$
- $\gamma\gamma$ searches
  - statistics limited

- $bb$ sensitivity comparable with Run I dataset
- rest of channels on the pipeline
Treating the $ttH(\mathbf{WW}/ZZ^*/\tau\tau\tau)$ signal with care

• **irreducible background from theory predictions**
  – For the main $ttV$ backgrounds, NLO QCD+EWK cross section (YR4), NLO QCD+PS MC (uncertainties $\sim 10\%$)
  – CMS also performed cross check with $ttW$ & $ttZ$ floating, including some CRs for $ttW$ & $ttZ$ in the fit

• **reducible background from non-prompt leptons estimated from data:**
  – Events in SR but with loose-but-not-tight leptons are weighted by lepton mis-identification probabilities
  – Mis-identification probabilities derived from multi-jet events
  – Sizeable relative systematic uncertainty, $O(30\%)$
Treating the $ttH(\text{WW/ZZ*/}\tau\tau\tau)$ signal with care

- maximum $|\eta|$ of the two leading leptons
- jet multiplicity
- minimum $\Delta R$ separation between each of the two leading leptons and a jet
- transverse mass of the leading lepton and the $E_T^{\text{miss}}$
- maximum score among jet permutations of a BDT discriminator that aims at re-construction hadronic top decays (for $ttH$ vs $tt$, $\ell^\pm\ell^\pm$ only);
- maximum score among jet permutations of a BDT discriminator that aims at tagging jets from Higgs decay products (for $ttH$ vs $ttV$, $\ell^\pm\ell^\pm$ only);
- highest and lowest $p_T$ of the selected leptons (for $ttH$ vs $ttV$ only);
- matrix element weights for signal and irreducible backgrounds, combined in one likelihood ratio variable ($3\ell$ only).
The first $ttH(\tau_l/\tau_h)$ reconstruction in Run2

$1\ell+2\tau_h$
BDT trained against ttbar

- The invariant mass and angular separation of the two reconstructed $\tau_h$.
- The transverse momenta of the two reconstructed $\tau_h$.
- The observable $H_T^{\text{miss}}$, computed according to Eq. (3).
- The average angular separation between any pair of jets.
- The multiplicity of jets, with and without b-tagging criteria applied.

$2\ell_{ss}+1\tau_h$
MEM with $ttH$ and $ttV$ hypotheses

two sub-categories: with / without missing jet

- The transverse momenta of the leading lepton and of the third lepton.
- The maximum $|\eta|$ of the two leading leptons.
- The multiplicity of jets.
- The angular separation of the leading and of the subleading lepton with respect to the nearest jet.
- The transverse mass of the leading lepton and the missing transverse energy vector.
- The observable $H_T^{\text{miss}}$.
- The average angular separation between any pair of jets.

$3\ell+1\tau_h$
2BDTs: against $ttV$ and ttbar
Hadronic $\tau$ reconstruction and identification in Run 2

**Hadron + strip (HPS) algorithm**
- Seeded by reconstructed PF jets
- Neutral pions: strips 0.05 x 0.020 in $\eta$-$\Phi$
- Look into jet constituents, decay mode finding
  - a single charged particle without any strips: $h^\pm$;
  - combination of one charged particle and one strip: $h^\pm \pi^0$;
  - combination of a single charged particle with two strips: $h^\pm \pi^0 \pi^0$;
  - combination of three charged particles: $h^+ h^- h^\pm$.

**Dynamic strip reconstruction**
- Widen strip size in the case of bremsstrahlung or $\tau_h$ nuclear interaction, depends on $p_T$

**MVA based discriminator against jets**
- Use isolation sums computed within a cone of 0.3, optimised for $ttH$ background hadronic environment
How to deal with the overwhelming bkg. in \( ttH(bb) \)

- Important to resolve jet combinatorics in jets/b-tags categories
- MVAs in a 2D fit

**Signal enriched categories**

- First, a BDT that reduces bkg. (\(\equiv\)) and leaves signal almost intact (\(\equiv\)) in each sub-category
  - (generic) discrimination of \( ttH \) vs \( tt + \) jets
  - BDT input variables and its structure optimized
How to deal with the overwhelming bkg. in $ttH(\bb)$

- Measurement limited by **systematics**
  - **dominated** by $tt+(b-)$ jets bkg.
- Upper limit set at $1.5(1.7) \times$ the SM $ttH$ production rate
  - improved Run1 limit by a factor 2
Where the sensitivity for the $ttH(\gamma\gamma)$ stands?

- Look for **small but clean** $m_{\gamma\gamma}$ peak with
- an expected photon energy resolution of $\sim 1\%$
- separately in **leptonic** (95% pure) & **hadronic** (80% pure) $tt$ decays

**Background**: functional forms driven by data
- each form treated as a discrete nuisance parameter

**Signal**: $n$Gaussian model fit to MC that eases
- the necessary tuning to the simulation
- corrections for the relevant efficiencies to be applied
- spline interpolation; for (7) mass hypotheses (within [120,130] GeV)
Where the sensitivity for the $ttH(\gamma\gamma)$ stands?

- Measurement currently dominated by statistical power
- **Independently** from other production mechanisms
- Simultaneously (grouped) in fermionic and bosonic modes

### Event Categories

<table>
<thead>
<tr>
<th>Event Categories</th>
<th>SM$^{125\text{GeV}}$ Higgs boson expected signal $\sigma_{eff}$</th>
<th>$\sigma_{HM}$ (GeV$^{-1}$)</th>
<th>Bkg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untagged Tag 0</td>
<td>11.92</td>
<td>79.10 %</td>
<td>7.60 %</td>
</tr>
<tr>
<td>Untagged Tag 1</td>
<td>128.78</td>
<td>85.98 %</td>
<td>7.38 %</td>
</tr>
<tr>
<td>Untagged Tag 2</td>
<td>220.12</td>
<td>91.11 %</td>
<td>5.01 %</td>
</tr>
<tr>
<td>Untagged Tag 3</td>
<td>258.50</td>
<td>92.35 %</td>
<td>4.23 %</td>
</tr>
<tr>
<td>VBF Tag 0</td>
<td>9.35</td>
<td>29.47 %</td>
<td>69.97 %</td>
</tr>
<tr>
<td>VBF Tag 1</td>
<td>15.55</td>
<td>41.91 %</td>
<td>52.50 %</td>
</tr>
<tr>
<td>TTH Hadronic Tag</td>
<td>2.42</td>
<td>16.78 %</td>
<td>1.28 %</td>
</tr>
<tr>
<td>TTH Leptonic Tag</td>
<td>1.12</td>
<td>1.09 %</td>
<td>0.08 %</td>
</tr>
<tr>
<td>Total</td>
<td>647.77</td>
<td>87.93 %</td>
<td>7.29 %</td>
</tr>
</tbody>
</table>
Difficulty to build models for negative Yukawa

- negative Yukawa coupling would point to new physics
- hard to explain, models are tightly constrained/have problems
- example: dim-6 operator

\[ \mathcal{L}_{\text{eff}}^{(1)} = \frac{Y^3}{\Lambda^2} \left( H^\dagger H - \frac{v^2}{2} \right) \left( H \bar{q}_3 u_3^c + h.c. \right) \]

- does not change top mass
- gives a contribution to the top Yukawa coupling

\[ y_t^{\text{eff}} = \frac{\sqrt{2} m_t}{v} + \frac{Y^3}{\sqrt{2}} \frac{v^2}{\Lambda^2} \]


- for \( y_t^{\text{eff}} < 0 \), the cutoff scale \( \Lambda \) must be close to electroweak scale
- need new sub-TeV particles to generate operator
- tightly constrained
- possibilities are vector-like fermions, charged scalars