Search for $ttH$ production in the $H \to bb$ channel

Claire David (DESY)

SM@LHC 2018, Berlin  
Young Scientist Talk  
April 11th, 2018
ttH in the $H \rightarrow bb$ channel: motivation and challenges

Direct measurement of top Yukawa coupling

any deviation ⇒ hint to new physics!

Highest branching ratio
$BR( H \rightarrow bb )$: 58 %

Final states

1 or 2 leptons (electron or muon) opposite sign

Large multiplicity of jets & b-jets

5 b-tag working points

<table>
<thead>
<tr>
<th>b-jet → efficiencies</th>
<th>none</th>
<th>loose</th>
<th>medium</th>
<th>tight</th>
<th>very-tight</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>85%</td>
<td>77%</td>
<td>70%</td>
<td>60%</td>
<td></td>
</tr>
</tbody>
</table>

⇒ to build regions rich in $tt+$light / $tt+c / tt+b / ttH$

Events are split according to $N_{jets}$ and $N_{b-jets}$

Challenges

→ $ttH$ production only 1% of Higgs total cross-section
→ complex combinatorics for Higgs/top reconstruction
→ main background: $tt+jets$ ⇒ difficult to model, especially $tt+bb$
Background composition in signal and control regions

Single lepton channel

Case ≥ 6 jets *

- CR tt+light
- CR tt + ≥1c
- CR tt + ≥1b

(mostly Single-top & fakes)

Dedicated control regions for
\( tt+light, \ tt+ ≥1c, \ tt+ ≥1b \)

Main background: \( tt + \text{jets} \)
- very difficult to model
- irreducible \( tt+bb \) ⇒ 30 × signal
  more next slide

Max S/B = 5.3% in SR1\( ^{≥6j} \)

* Other regions:
1L 5 jets: 3 CR + 3 CR
2L 3 jets: 2 CR
2L ≥4 jets: 2 CR + 3 SR

See backups

Low purity of \( ttH \) signal
⇒ use of multivariate techniques
→ 2-stage approach
→ separate \( ttHbb \) from \( tt+bb \)

DESY. Claire David

Search for \( ttH \) in the \( H \rightarrow bb \) channel

SM@LHC 2018
April 11th, 2018

3
Fit model, results and main uncertainties

- simultaneous profile likelihood fit in all $1\ell + 2\ell$ regions (100 bins)
- fitted variables: BDT output (SR), $H_T = \Sigma p_T^{jets}$ (CR)
- floating parameters: $tt + \geq 1c$ & $tt + \geq 1b$ normalization factors

POI: signal strength

$$\mu = \frac{\sigma_{\text{measured}}}{\sigma_{\text{SM}}}$$

Example post-fit SR1$^{\geq 6j} 1\ell$

Main uncertainties

- **tt$+ \geq 1b$ modelling**
  - Large differences
  - Powheg+Pythia8 & Sherpa 5F
  - Limited MC statistics
  - Low stat in $tt$ sample for $ttbb$ phase-space
- **b-tagging**
- **Jet uncertainties**
  - Total systematics
  - Limited sensitivity!

Results (two-$\mu$ combined fit)

- Sensitivity dominated by $1\ell$ channel

Excess above background: 1.4 $\sigma$ (obs) 1.6 $\sigma$ (exp)

Upper limit: $\mu_{ttH} < 2.0$ at 95% CL
Additional material
Jet b-tagging working points

MV2c10 tagger output

ATLAS Simulation Preliminary
\( \sqrt{s} = 13 \text{ TeV}, \, \tilde{t}\tilde{t} \)

\( b \) jets
\( c \) jets
Light-flavour jets

b-jet efficiencies →

<table>
<thead>
<tr>
<th>none</th>
<th>loose</th>
<th>medium</th>
<th>tight</th>
<th>very-tight</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td>85%</td>
<td>77%</td>
<td>70%</td>
<td>60%</td>
</tr>
</tbody>
</table>

pure b-jets

ATL-PHYS-PUB-2016-012
Background composition in signal and control regions

Single lepton channel

5 jets

- CR $t\bar{t}$ + light
- CR $t\bar{t}$ + $\geq$1c
- CR $t\bar{t}$ + $\geq$1b

Control regions

- CR $t\bar{t}$ + light
- CR $t\bar{t}$ + $\geq$1c
- CR $t\bar{t}$ + $\geq$1b

Signal regions

- SR$_{S2}^5$
- SR$_{S1}^5$
- SR$_{boosted}$

ATLAS

$\sqrt{s} = 13$ TeV, 36.1 fb$^{-1}$

Single Lepton

Search for $t\bar{t}H$ in the $H \rightarrow bb$ channel

DESY. Claire David

SM@LHC 2018

April 11th, 2018
Background composition in signal and control regions

Dilepton channel

3 jets

≥ 4 jets

Low purity in ttH signal $\rightarrow$ 1.8% to 5.4 %

CR enriched in tt+light (2), tt+c (1), tt+b (1)
**tt + jets background modelling**

**Monte Carlo nominal sample**

- **Powheg + Pythia8 (PP8)**
  - A14 tune, NNPDF3.0 PDF set
  - $h_{\text{damp}} = 1.5 \times m_t$

  Tuned to 13 TeV data (unfolding @ particle level)

**Monte Carlo alternative samples for ttbar systematics**

- Radiation up/down: Specific Rad Up/Down PP8
- Parton shower: Powheg + Herwig7 (PH7)
- Generator: Sherpa 5 Flavor Scheme

**PDF Flavor Scheme**

- 4 FS ($m_b > 0$)
- 5 FS ($m_b = 0$)

**tt+bb Matrix Elements**

- covers full b-quark phase space
- Can’t describe collinear $g \rightarrow bb$ splittings
- only $tt+ \geq 1b$ prediction
- inclusive predictions $tt+\text{light} / tt+c / tt+b$
- how to merge with $tt+\text{light}$ prediction?

**Available:**

- Sherpa+OpenLoops 4F
- PP8, PH7, Sherpa 5F

**ATL-PHYS-PUB-2017-007**
\( \text{tt} + \geq 1\text{b background} \)

\[ \text{Relative contributions of } \text{tt} + \geq 1\text{b sub-categories} = \text{tt} + \text{b} \quad \text{tt} + \text{bb} \quad \text{tt} + \text{B} \quad \text{tt} + \geq 3\text{b} \]

Reweighted to match Sherpa+OpenLoops (4F) predictions

\[ \text{Same done for ttbar systematics samples: Sherpa 5F, PP8 radiation up/down, Powheg + Herwig7} \]

Nominal, inclusive \( \text{tt+light} / \text{tt+c} / \text{tt+b} \) (5FS)

Exclusive \( \text{tt+} \geq 1\text{b prediction} \)

Predicts more jets than 5 FS scheme

In high b-jets category, difference not covered by 4FS uncertainty

Additional uncertainty added to cover the difference:

“\( \text{tt+} \geq 1\text{b: Sherpa 4F vs nominal} \)”

(next slide)
Multivariate analysis: 2 stage strategy

**Stage 1: Event Reconstruction**

- **Reconstruction BDT**
  - $m_{H}^{\text{recBDT}}$
  - Matching reconstructed jets & b-jets
  - $\rightarrow$ reconstruct $ttHbb$ system
  - with 22 kinematic variables
  - $\varepsilon_{\text{reco Higgs}} \sim 48\%$ (32% no $m_H$ info)

- **Likelihood discriminant**
  - Probabilities from product of 1D pdfs (from invariant mass / angular distributions)
  - from all possible combinations
  - No correlations

- **Matrix Element Method**
  - $\text{MEM}_{i1} = \log_{10}(L_{i}) - \log_{10}(L_{i})$
  - Likelihoods computed at parton level
  - Transfer functions from detector $\rightarrow$ parton level
  - no training on MC samples needed

**Stage 2: Classification**

- **ttbb $\leftarrow$ separate $\rightarrow$ ttH**

- **General kinematic variables**
  - Invariant masses: $m_{bb}$, $m_{..}$
  - Angular separation: $\Delta R_{bb}$, $\Delta R_{bb..}$

- **b-tagging discriminant**
  - $3^{rd}, 4^{th}, 5^{th}$ b-jet WP

- **Outputs of Stage 1**
  - powerful discriminants
  - Only variables validated against data

- **Classification BDT output**
  - High scores $\equiv$ highest S/B
  - $\Rightarrow$ final discriminant

---

**Toolkit**

- **BDT class**

**Example in:**

- $\Delta R_{bb}$, $\Delta R_{bb..}$

---

**DESY**. Claire David

Search for $ttH$ in the $H \rightarrow bb$ channel

SM@LHC 2018

April 11th, 2018
Uncertainties

Main uncertainties on signal strength $\mu$

**tt+ ≥1b modelling**  \(±0.46\)
- Choice of NLO generator
- Large differences PP8 vs Sherpa 4FS

**Limited MC statistics**  \(±0.30\)
- Low stat in $t\bar{t}$ sample for $ttbb$ phase-space

**b-tagging**  \(±0.16\)

**Jet uncertainties**  \(±0.14\)

Total systematics

\[\Delta\mu = +0.57 - 0.54\]

Total statistics  \(±0.29\)

Results combined:

\[\Delta\mu = 0.84 +0.64 - 0.61\]

Sensitivity dominated by $1\ell$ channel