Search for boosted $tt\bar{t}H(H \rightarrow b\bar{b})$
with the ATLAS detector

Emma Winkels
Higgs Couplings 2018
**Introduction**

- \(t\bar{t}H\) is a rare Higgs boson production mode at the LHC
- Will discuss here the boosted \(H \rightarrow b\bar{b}\) decay channel: [Phys. Rev. D 97, 072016 (2018)](https://doi.org/10.1103/PhysRevD.97.072016)

\[29/11/2018\]

\(t\bar{t}H \sim 1\%\) of total Higgs cross-section

\(H \rightarrow b\bar{b}\) decay BR \(\sim 58\%\)
Motivation

All current Higgs measurements consistent with the Standard Model.

Any deviations in top Yukawa coupling would give indication for new physics.

- $t\bar{t}H$ is direct probe of the top Yukawa coupling whereas e.g. $ggF$ only accesses it via loop effects.
- Boosted channel selects high-$p_T$ events which leads to different kinematics and a simplified combinatorial background compared to the resolved channel.
The boosted analysis

• Using $36 \, fb^{-1}$ of ATLAS data from 2015-2016 at $\sqrt{s} = 13$ TeV
• Target events where both the Higgs and hadronic top are boosted
Analysis overview

The boosted region:

- Gets priority over the resolved single-lepton regions
- Is combined with the full resolved analysis in the final result
Boosted signal region

Preselection to select candidate events:

**Higgs candidate:** large reclustered jet with $\geq 2$ $b$-tagged subjets

**Top candidate:** large reclustered jet with 1 $b$-tagged subjet and $\geq 1$ $b$-veto

$\geq 1$ additional $b$-jet outside the top and Higgs candidates

All jets are constructed with the anti-$k_T$ algorithm.
Small jets have $R=0.4$ and large jets $R=1.0$.

*The subjets are $b$-tagged at the loose (85%) working point.
Boosted signal region composition

Low $S/B \sim 2.4\%$ and background dominated by $t\bar{t} + \geq 1b$ which is very difficult to distinguish from $t\bar{t}H$.

Use multivariate analysis to discriminate between signal and background events.
Multivariate analysis

- Use a Boosted Decision Tree (BDT) to separate signal from background events.
- Eight variables are chosen as input:
  - The Higgs candidate mass
  - The first splitting scale of the top candidate (jet substructure variable)
  - 4 angular variables
  - 2 $b$-tagging variables where all jets are assigned a score depending on their likelihood to be a $b$-jet

<table>
<thead>
<tr>
<th>None</th>
<th>Loose</th>
<th>Medium</th>
<th>Tight</th>
<th>Very tight</th>
</tr>
</thead>
<tbody>
<tr>
<td>b-tagging efficiency</td>
<td>-</td>
<td>85%</td>
<td>77%</td>
<td>70%</td>
</tr>
</tbody>
</table>

Score

1 2 3 4 5
Statistical analysis

• Perform a binned profile likelihood fit simultaneously across all regions to test for presence of signal.

• The boosted region is fitted together with 8 resolved signal regions and 10 resolved control regions.

• In the signal regions, we fit to the BDT.

• Systematics included as nuisance parameters in the fit.

• Extract $\mu_{t\bar{t}H} = \frac{\sigma_{obs}^{t\bar{t}H}}{\sigma_{SM}^{t\bar{t}H}}$.

The control regions are used to constrain systematic uncertainties on backgrounds.
Boosted BDT

The uncertainty is reduced post-fit due to profiling of the nuisance parameters.
More information see p.23 of arXiv 1712.08895.
Uncertainties

The analysis is limited by systematics. Main source is modelling of $t\bar{t} + \geq 1b$.

We compare different Monte Carlo generators to assess uncertainties on matrix element, showering, and radiation variations.

<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>$\Delta \mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t} + \geq 1b$ modeling</td>
<td>$0.46$</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$ modeling</td>
<td>$-0.46$</td>
</tr>
<tr>
<td>$b$-tagging-efficiency and mis-tag rates</td>
<td>$0.16$</td>
</tr>
<tr>
<td>Jet energy scale and resolution</td>
<td>$0.14$</td>
</tr>
<tr>
<td>$t\bar{t}H$ modeling</td>
<td>$0.22$</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1c$ modeling</td>
<td>$0.09$</td>
</tr>
<tr>
<td>$t\bar{t}$ + light modeling</td>
<td>$0.06$</td>
</tr>
<tr>
<td>Luminosity</td>
<td>$0.03$</td>
</tr>
<tr>
<td>Light lepton (e, $\mu$) id., isolation, trigger</td>
<td>$0.03$</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>$0.57$</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1b$ normalization</td>
<td>$0.09$</td>
</tr>
<tr>
<td>$t\bar{t} + \geq 1c$ normalization</td>
<td>$0.02$</td>
</tr>
<tr>
<td>Intrinsic statistical uncertainty</td>
<td>$0.21$</td>
</tr>
<tr>
<td>Total statistical uncertainty</td>
<td>$0.29$</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>$0.64$</td>
</tr>
</tbody>
</table>
Results resolved+boosted

Best-fit value:

\[ \mu_{t\bar{t}H} = 0.84 \pm 0.29^{\text{stat}} +0.57_{-0.54}^{\text{syst}} \]

- Excluded \( \mu_{t\bar{t}H} > 2.0 \) at 95% CL
- Signal seen with obs (exp) significance of 1.4\( \sigma \)(1.6\( \sigma \))

Currently the boosted signal region does not add significant sensitivity to the channel.

- More data → tighter SR selection
- Many ongoing studies for improvements
Conclusions

• Very challenging analysis due to dominant $t\bar{t} + \geq 1b$ background.
• Measured $\mu_{t\bar{t}H}$ consistent with the SM and background-only hypotheses.
• The boosted channel does not currently add significant sensitivity to the resolved analysis.
  • First time it is included in ATLAS; many studies on-going to improve the performance for full Run-2 analysis.
  • More data and better MC statistics will allow for tighter signal region.
• The analysis is systematics limited → need to rethink our approach to dealing with these uncertainties.
BACK-UP
The ATLAS detector

The ATLAS collaboration
>5000 members
38 countries
182 institutes
The ATLAS coordinate system

\[ \eta = -\ln \tan \frac{\theta}{2} \]

\[ \Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2} \]
Granularity of ATLAS calorimeter

The hadronic calorimeter is coarser than the EM calorimeter. If we want to use all calorimeter layers we are limited by coarsest layer ($R_{min}$~0.2). But we can also ignore coarse layers and use only very fine layers.

- EM granularity: $\Delta \eta \times \Delta \phi \approx 0.025 \times \pi / 128$
- Hadronic granularity: $\Delta \eta \times \Delta \phi \approx 0.1 \times \pi / 32$

arXiv 1603.02934
Boosted $t\bar{t}H$

$\theta \approx \frac{2m}{p_T}$

Increasing transverse momentum

100 GeV  200 GeV  300 GeV  400 GeV

Top

Higgs

Combine small jets  Use a single large jet  Use a single small jet

Jets $b$-tagging

- The analysis relies heavily on $b$-tagging: identifying jets containing $b$-hadrons ($b$-jets)
- Using a multivariate $b$-tagging algorithm optimised to select $b$-jets and separate them from $c$-jets, $\tau$-jets, and other jets (light jets)
- MV2c10 uses number of variables to exploit the long lifetime of the $b$-hadron
- Four working points available from this algorithm

<table>
<thead>
<tr>
<th>Working point</th>
<th>Efficiency for $b$-jets</th>
<th>Rejection factor of $c$-jets</th>
<th>Rejection factor of light-jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loose</td>
<td>85%</td>
<td>3.1</td>
<td>33</td>
</tr>
<tr>
<td>Medium</td>
<td>77%</td>
<td>6</td>
<td>134</td>
</tr>
<tr>
<td>Tight</td>
<td>70%</td>
<td>12</td>
<td>381</td>
</tr>
<tr>
<td>Very tight</td>
<td>60%</td>
<td>34</td>
<td>1538</td>
</tr>
</tbody>
</table>
Large jets

Trimming:
• Build large jets from energy clusters in calorimeter
• Extra set of systematics

Reclustering:
• Build large jets from fully calibrated small jets
• No extra systematics
Jet substructure variables

• Jet mass is calculated from the jet constituents (calorimeter energy deposits):

\[ m^{calo} = \sqrt{\left(\sum E_i\right)^2 - \left(\sum p_i\right)^2} \]

• First \( k_T \) splitting scale:

\[ \sqrt{d_{12}} = \min(p_{T,i}, p_{T,j}) \times \frac{\Delta R_{i,j}}{R} \]

where \( i, j \) are the two proto-jets combined at the final step of the \( k_T \) jet algorithm.
Higgs mass reconstruction

Correct Higgs candidate found in 48% of events

Correct Higgs candidate found in 49% of events

Correct Higgs candidate found in 47% of events

ATLAS

\(\sqrt{s} = 13 \text{ TeV}, 36.1 \text{ fb}^{-1}\)

Single Lepton

SR\(_1\)\(^{36}\)

Post-Fit

Dilepton

SR\(_1\)\(^{36}\)

Post-Fit

Boosted

Single Lepton

SR\(_{\text{boosted}}\)

Post-Fit

ATLAS

\(\sqrt{s} = 13 \text{ TeV}, 36.1 \text{ fb}^{-1}\)

Single Lepton

SR\(_{\text{boosted}}\)

Post-Fit

Dilepton

ATLAS

\(\sqrt{s} = 13 \text{ TeV}, 36.1 \text{ fb}^{-1}\)

Dilepton

ATLAS

\(\sqrt{s} = 13 \text{ TeV}, 36.1 \text{ fb}^{-1}\)
Pre-fit impact on $\mu$:
- $\theta = \theta + \Delta \theta$
- $\theta = \theta - \Delta \theta$

Post-fit impact on $\mu$:
- $\theta = \theta + \Delta \theta$
- $\theta = \theta - \Delta \theta$

Nuis. Param. Pull

$b$-tagging: efficiency NP II
$b$-tagging: mis-tag (light) NP I
Jet energy resolution: NP I
Jet energy resolution: NP II
$\mu$ pre-fit impact on $q^D + q = q^D - q = q$
Post-fit impact on $q^D + q = q^D - q = q$

$\mu$ Pre-fit impact on $q^D + q = q^D - q = q$

$\mu$ Post-fit impact on $q^D + q = q^D - q = q$

Nuis. Param. Pull

ATLAS
$ar{t} \bar{s} = 13$ TeV, 36.1 fb$^{-1}$

$\Delta \mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$

$\mu$
$t\bar{t} + \text{jets modelling}$

- Treatment of $t\bar{t} +\text{jets}$ is crucial in the analysis.
- Nominal $t\bar{t}$ MC sample is Powheg+Pythia8, normalised to NNLO+NNLL cross section calculation.
- Separate into $t\bar{t} + \text{light}$, $t\bar{t} + \geq 1c$ and $t\bar{t} + \geq 1b$.
- Dividing $t\bar{t} + \geq 1b$ into 4 more categories, the fractions of which are reweighted to Sherpa sample which has NLO calculation and takes into account $b$-quark mass.
- Large number of systematics covering $t\bar{t}$ modelling.
# $tt\bar{t} + \text{jets modelling systemsatics}$

All sources of systematic uncertainty for $tt\bar{t} + \text{jets modelling}$.

<table>
<thead>
<tr>
<th>Systematic source</th>
<th>Description</th>
<th>$tt\bar{t}$ categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>$tt\bar{t}$ cross-section</td>
<td>Up or down by 6%</td>
<td>All, correlated</td>
</tr>
<tr>
<td>$k(tt\bar{t} + \geq 1c)$</td>
<td>Free-floating $tt\bar{t} + \geq 1c$ normalization</td>
<td>$tt\bar{t} + \geq 1c$</td>
</tr>
<tr>
<td>$k(tt\bar{t} + \geq 1b)$</td>
<td>Free-floating $tt\bar{t} + \geq 1b$ normalization</td>
<td>$tt\bar{t} + \geq 1b$</td>
</tr>
<tr>
<td>SHERPA5F vs. nominal</td>
<td>Related to the choice of NLO event generator</td>
<td>All, uncorrelated</td>
</tr>
<tr>
<td>PS &amp; hadronization</td>
<td>POWHEG+HERWIG 7 vs. POWHEG+PYTHIA 8</td>
<td>All, uncorrelated</td>
</tr>
<tr>
<td>ISR / FSR</td>
<td>Variations of $\mu_{R}$, $\mu_{F}$, $h_{\text{damp}}$ and A14 Var3c parameters</td>
<td>All, uncorrelated</td>
</tr>
<tr>
<td>$tt\bar{t} + \geq 1c$ ME vs. inclusive</td>
<td>MG5_aMC@NLO+HERWIG++: ME prediction (3F) vs. incl. (5F)</td>
<td>$tt\bar{t} + \geq 1c$</td>
</tr>
<tr>
<td>$tt\bar{t} + \geq 1b$ SHERPA4F vs. nominal</td>
<td>Comparison of $tt\bar{t} + bb$ NLO (4F) vs. POWHEG+PYTHIA 8 (5F)</td>
<td>$tt\bar{t} + \geq 1b$</td>
</tr>
<tr>
<td>$tt\bar{t} + \geq 1b$ renorm. scale</td>
<td>Up or down by a factor of two</td>
<td>$tt\bar{t} + \geq 1b$</td>
</tr>
<tr>
<td>$tt\bar{t} + \geq 1b$ resumm. scale</td>
<td>Vary $\mu_{Q}$ from $H_{T}/2$ to $\mu_{\text{CMMP}}$</td>
<td>$tt\bar{t} + \geq 1b$</td>
</tr>
<tr>
<td>$tt\bar{t} + \geq 1b$ global scales</td>
<td>Set $\mu_{Q}$, $\mu_{R}$, and $\mu_{F}$ to $\mu_{\text{CMMP}}$</td>
<td>$tt\bar{t} + \geq 1b$</td>
</tr>
<tr>
<td>$tt\bar{t} + \geq 1b$ shower recoil scheme</td>
<td>Alternative model scheme</td>
<td>$tt\bar{t} + \geq 1b$</td>
</tr>
<tr>
<td>$tt\bar{t} + \geq 1b$ PDF (MSTW)</td>
<td>MSTW vs. CT10</td>
<td>$tt\bar{t} + \geq 1b$</td>
</tr>
<tr>
<td>$tt\bar{t} + \geq 1b$ PDF (NNPDF)</td>
<td>NNPDF vs. CT10</td>
<td>$tt\bar{t} + \geq 1b$</td>
</tr>
<tr>
<td>$tt\bar{t} + \geq 1b$ UE</td>
<td>Alternative set of tuned parameters for the underlying event</td>
<td>$tt\bar{t} + \geq 1b$</td>
</tr>
<tr>
<td>$tt\bar{t} + \geq 1b$ MPI</td>
<td>Up or down by 50%</td>
<td>$tt\bar{t} + \geq 1b$</td>
</tr>
<tr>
<td>$tt\bar{t} + \geq 3b$ normalization</td>
<td>Up or down by 50%</td>
<td>$tt\bar{t} + \geq 1b$</td>
</tr>
</tbody>
</table>
Non-$t\bar{t}$ modelling

- Nominal $t\bar{t}H$ signal model is MG5_aMC@NLO+Pyhtia8 (NLO)
- $t\bar{t}W$, $t\bar{t}Z$: MG5_aMC@NLO+Pythia8 (NLO)
- Single-top ($Wt$, s-channel, t-channel): Powheg+Pythia6 (NLO)
- $w+jets$, $z+jets$: Sherpa 2.2.1 (NLO <3 partons, LO <5 partons)
- $tHqb$: MG5_aMC@NLO+Pythia8 (LO)
- $tWH$: MG5_aMC@NLO+Herwig++ (NLO)
- $t\bar{t}t\bar{t}$, $t\bar{t}WW$: MG5_aMC@NLO+Pythia8 (LO)
- $tZ$: MG5_aMC@NLO+Pythia6 (LO)
- $tZW$: MG5_aMC@NLO+Pythia8 (NLO)