Heavy flavour modelling in top-related analyses at ATLAS

Andrea Knue

Heavy Flavour Production at the LHC
Durham, 20-22nd April 2016
Outline

Motivation

$t\bar{t} + b(\bar{b})$ measurement

SM search: $t\bar{t}H(H \rightarrow b\bar{b})$

Exotics searches

Summary

Discussion
Motivation

- top-quark events often produced with additional jets @ the LHC
  - want to test pQCD prediction
  - important, often irreducible background for SM and exotics
- will show today three examples:
  1.) SM measurement \( (t\bar{t}b\bar{b}) \)
  2.) SM search \( (t\bar{t}H(b\bar{b})) \)
  3.) Exotics searches (four tops, \( Ht + X, H^+ \))
Outline

Motivation

$t\bar{t} + b(\bar{b})$ measurement

SM search: $t\bar{t}H(H \rightarrow b\bar{b})$

Exotics searches

Summary

Discussion
want to understand and test QCD

best predictions from fixed order calc. @ NLO

but: large uncertainties on predictions

we need experimental measurements!

measurement @ 8TeV: cross-section in fiducial phase space

measure $t\bar{t} + b$, $t\bar{t} + b\bar{b}$ and ratio of $\frac{\sigma(t\bar{t}+b\bar{b})}{\sigma(t\bar{t}+jj)}$
Analysis strategy

**Fiducial $t\bar{t} + b$ cross-section**

→ $e\mu$ & lepton+jets channel

→ two (one) leptons, ≥ 3 (5) jets, ≥ 3 $b$-tagged jets

→ use $b$-tagging discriminant to extract $t\bar{t}+b/c/l$

**Fiducial $t\bar{t} + b\bar{b}$ cross-section**

→ dilepton channel: cut- and fit-based

→ two leptons, ≥ 4 $b$-tagged jets

→ use $b$-tagging discr. to extract $t\bar{t} + b\bar{b}$, $t\bar{t} + bX$, $t\bar{t} + cX$, $t\bar{t} + lX$

**$\sigma(t\bar{t} + b\bar{b})/\sigma(t\bar{t} + jj)$ ratio**

→ dilepton channel, at least 4 particle jets

→ allows to cancel some uncertainties
Pre-/Postfit plots for lepton+jets

Extract cross-sections:

➔ profile likelihood fit of templates to data
➔ allows to strongly constrain the uncertainties (right plot)
➔ insignificant uncertainties are excluded from the fit
Impact of systematic uncertainties

Impact on final result

- dilepton: stat. and syst. of similar size
- lepton+jets: limited by systematics
- systematics dominated by tagging and modelling uncertainties
- impact on $\mu$: PS and scale variations
## Comparison of generator setups

<table>
<thead>
<tr>
<th>Sample</th>
<th>Generator</th>
<th>Shower</th>
<th>$m_b$ [GeV]</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t} + bb$</td>
<td>aMC@NLO</td>
<td>Pythia8</td>
<td>4.8</td>
<td>BDDP scale</td>
</tr>
<tr>
<td>$t\bar{t} + b\bar{b}$</td>
<td>aMC@NLO</td>
<td>Pythia8</td>
<td>4.8</td>
<td>$H_T/4$ scale</td>
</tr>
<tr>
<td>$t\bar{t} + b\bar{b}$</td>
<td>Powheg</td>
<td>Pythia8</td>
<td>0.0</td>
<td>$H_T/2$ scale</td>
</tr>
<tr>
<td>$t\bar{t} + b\bar{b}$ @ LO+PS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t\bar{t} + (0-3)$ p.</td>
<td>Madgraph5</td>
<td>Pythia6</td>
<td>4.8</td>
<td></td>
</tr>
<tr>
<td>$b$-quarks only in PS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}$ (a)</td>
<td>Pythia8</td>
<td>Pythia8</td>
<td>4.8</td>
<td>enhanced $g \rightarrow b\bar{b}$</td>
</tr>
<tr>
<td>$t\bar{t}$ (b)</td>
<td>Pythia8</td>
<td>Pythia8</td>
<td>4.8</td>
<td>suppr. $g \rightarrow b\bar{b}$</td>
</tr>
<tr>
<td>$t\bar{t}$ (c)</td>
<td>Pythia8</td>
<td>Pythia8</td>
<td>4.8</td>
<td>enhanced $g \rightarrow b\bar{b}$</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>Powheg</td>
<td>Pythia6</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

### Options for $g \rightarrow b\bar{b}$ splitting

- **a)** wgtq3: mass dep. terms in splitting kernel,
- **b)** wgtq5: no mass dep. terms, sgtq=1: small $g \rightarrow b\bar{b}$ rate
- **c)** wgtq6: mass dep. terms, sgtq=0.25: high $g \rightarrow b\bar{b}$ rate
Results for fiducial cross-sections

QCD only measurements (subtract $t\bar{t}H$ and $t\bar{t}V$)

- measurement at 8 TeV still limited by statistics
- uncertainty on $t\bar{t} + b$: 25% (lepton+jets) and 32% (dilepton)
- uncertainty on $t\bar{t} + b\bar{b}$: 35% (cut-based) and 36% (fit-based)
- most generators: good agreement with data
  → extreme option with highest rate of $g \rightarrow b\bar{b}$: strongly disfavoured
Results ratio: Comparison for diff. $g \rightarrow b\bar{b}$ rates

$\rightarrow$ enhanced $g \rightarrow b\bar{b}$ rate (mass. dep. terms)
$\rightarrow$ reduced $g \rightarrow b\bar{b}$ rate (no mass. dep.)
$\rightarrow$ enhanced $g \rightarrow b\bar{b}$ rate (mass. dep. terms)

$\rightarrow$ results sensitive to modelling of gluon splitting in PS

$\rightarrow$ model with highest $g \rightarrow b\bar{b}$ rate overestimates the cross section by factor of 2!
Motivation: Why search for $t\bar{t}H$?

→ after discovery of the Higgs Boson:
  ◮ what are its properties?
  ◮ is it really the SM particle?
→ important: directly measure the top-Higgs Yukawa coupling $Y_t$
→ top quark heaviest fermion:
  ◮ $Y_t$ largest: $\approx 1$

→ any deviation would be sign for BSM processes
→ but: challenging search
Why $t\bar{t}H (H \to b\bar{b})$?

**Challenges:**
- largest BR for Higgs decay, **but:**
- irreducible bkg from $t\bar{t}b\bar{b}$
- large uncertainties on $t\bar{t}$+HF

**How to cope with irreducible bkg?**
- exploit as much info as possible
- use a NN to get best possible S/B separation
- use signal-depleted regions to constrain bkg and unc
- combined nuisance parameter fit to all regions

Lepton+jets channel

Event selection

- 1 isolated lepton (25 GeV)
- at least 4 jets, 2 $b$-tagged jets
- MVA based $b$-tagging
- no cut on $\text{MET}/m_T, W$
- ttbar modelling with Powheg

Discriminating variables

- signal-depleted regions: $H_T^{\text{had}}$
- signal-enriched regions: MVA output
  - input to MVA are 10 variables per region
  - note: if for $g \to b\bar{b}$ the $b$’s are within the same jet, events can migrate between different jet/tag categories
How to define different $t\bar{t} + \text{HF}$ processes?

- take antiKt ($R=0.4$) particle level jets ($p_T > 15 \text{ GeV}, \mid \eta \mid < 2.5$)
- use truth matching to hadrons with $\Delta R < 0.4, p_T > 5 \text{ GeV}$
- now check additional jets in the event:
  - $\rightarrow$ if add. jet contains one B/D hadron: $t\bar{t} + b$
  - $\rightarrow$ if add. jet contains more than one B/D hadron: $t\bar{t} + B$
  - $\rightarrow$ if event not categ. as $t\bar{t} + b/B$ and particle jet contains C-hadron: $t\bar{t} + c\bar{c}$
HF categories for different setups

Samples used for analysis:

→ Powheg+Pythia6: $t\bar{t} + b\bar{b}$ only from PS
→ Madgraph+Pythia6: $t\bar{t}+(0–3)$ partons (MLM matching)
→ Sherpa+OpenLoops: (4FS, massive $b$-quarks)

→ expected to model $t\bar{t} + b\bar{b}$ better, reweight Powheg to Sherpa for $t\bar{t} + b\bar{b}$
Sensitivity to variations in Sherpa

**Scale variations**

- large impact from $\mu_R$ variation
- large impact from functional form of $\mu_R$

**Other variations**

- impact of MSTW larger than NNPDF
- impact of shower recoil visible
Systematic uncertainties and impact on result

- PS: Pythia6 vs Herwig
- 50% normalisation $t\bar{t} + b\bar{b}/c\bar{c}$

$t\bar{t} + c\bar{c}$:
- scale and variation
- matching threshold
- Madgraph vs. Powheg

$t\bar{t} + b\bar{b}$:
- using $\mu_R = \sqrt{m_t m_{b\bar{b}}}$
- vary $\mu_R$ up and down
- change func. form of $\mu_F$ and $\mu_Q$
- shower recoil scheme
- PDF choice in Sherpa
- missing MPI and FSR
Outline

Motivation

$t\bar{t} + b(\bar{b})$ measurement

SM search: $t\bar{t}H(H \rightarrow b\bar{b})$

Exotics searches

Summary

Discussion
VLQ and four-top quark searches @ 8TeV

- $T \bar{T} \to Ht + X$ production
- lepton+jets final state
- possible due to allowed FCNC decays
- similar final state as $t\bar{t}H$ shown before

- four-top production: tiny xsec in SM ($\approx 1$ fb-1)
- enhanced in several BSM scenarios
- in EFT: also four-fermion contact interaction
Systematic uncertainties

- $T\bar{T} \rightarrow Ht + X$ search (right):
- $t\bar{t} + b\bar{b}$ fraction in SR: $\approx 50\%$
- same samples/syst. as for $t\bar{t}H$
- large uncertainties on $t\bar{t} + b\bar{b}$

\begin{align*}
\begin{array}{c}
\text{PS and modelling} \\
\text{post-fit uncertainties: strongly reduced}
\end{array}
\end{align*}

\begin{array}{|c|c|c|c|c|}
\hline
\text{Signal} & \text{Pre-fit} & \text{Post-fit} \\
\hline
\text{Luminosity} & \pm 2.8 & \pm 2.8 & \pm 2.8 & \pm 2.8 & \pm 2.6 & \pm 2.6 & \pm 2.6 \\
\text{Lepton efficiencies} & \pm 1.5 & \pm 1.5 & \pm 1.5 & \pm 1.5 & \pm 1.5 & \pm 1.4 & \pm 1.5 \\
\text{Jet energy scale} & \pm 4.4 & \pm 15 & \pm 11 & \pm 12 & \pm 8.7 & \pm 6.4 & \pm 6.7 \\
\text{Jet efficiencies} & - & \pm 4.0 & \pm 2.2 & \pm 1.9 & \pm 3.1 & \pm 2.6 & \pm 0.4 \\
\text{Jet energy resolution} & \pm 0.1 & \pm 4.4 & \pm 3.8 & \pm 0.5 & \pm 4.6 & \pm 4.6 & \pm 6.6 \\
\text{b-tagging efficiency} & \pm 1.3 & \pm 5.6 & \pm 5.4 & \pm 9.3 & \pm 5.6 & \pm 11 & \pm 2.9 \\
\text{c-tagging efficiency} & \pm 1.6 & \pm 5.8 & \pm 12 & \pm 3.1 & \pm 1.7 & \pm 5.1 & \pm 1.8 \\
\text{Light-jet tagging efficiency} & \pm 0.6 & \pm 20 & \pm 5.7 & \pm 2.0 & \pm 0.6 & \pm 1.3 & \pm 1.2 \\
\text{High-pT tagging efficiency} & \pm 1.4 & \pm 0.7 & \pm 1.7 & \pm 1.6 & \pm 0.6 & \pm 1.0 & \pm 0.6 \\
\text{tt\text{:} reweighting} & - & \pm 13 & \pm 15 & - & \pm 10 & \pm 10 & - \\
\text{tt\text{:} parton shower} & - & \pm 28 & \pm 17 & \pm 6.2 & \pm 13 & \pm 11 & \pm 4.0 \\
\text{tt\text{:} HF\text{:} normalisation} & - & - & \pm 50 & \pm 50 & - & \pm 32 & \pm 18 \\
\text{tt\text{:} HF\text{:} modelling} & - & - & \pm 17 & \pm 12 & - & \pm 18 & \pm 10 \\
\text{Theoretical cross sections} & - & \pm 6.3 & \pm 6.3 & \pm 6.3 & \pm 4.6 & \pm 4.6 & \pm 4.6 \\
\hline
\text{Total} & \pm 21 & \pm 38 & \pm 73 & \pm 65 & \pm 24 & \pm 46 & \pm 27 \\
\end{array}
\end{array}
→ similar analysis strategy and HF modelling/uncertainties as for 8 TeV
→ see large discrepancy between data and MC
→ combined likelihood fit to 11 channels:

\[ t\bar{t} + b\bar{b} \text{ contribution increases about a factor of 2!} \]
Search for $H^+ \rightarrow tb$ production @ 8 TeV

- predicted in MSSM or 2HDM
- BR depends on tan $\beta$
- associated top $H^+$ production
- signal region has large HF backgrounds (about 29 %)

\[ 4j(2b) \quad 5j(2b) \quad \geq 6j(2b) \quad 4j(\geq 3b) \quad \geq 5j(\geq 3b) \]

**ATLAS Simulation**

- dominant systematics $\rightarrow$
- tagging/modelling uncertainties:
- larger impact from $t\bar{t} + b\bar{b}$ PS, cross-section

\[ g \rightarrow b \]

\[ H^+ \rightarrow \bar{t} \]

\[ t\bar{t} + b\bar{b} \]
Outline

Motivation

$t\bar{t} + b(\bar{b})$ measurement

SM search: $t\bar{t}H(H \rightarrow b\bar{b})$

Exotics searches

Summary

Discussion
Summary

- heavy-flavour production in association with top-quarks is difficult to predict
  - need experimental measurements for better understanding (need higher stats)

- \( t\bar{t} + b\bar{b} \) is important background for a lot of searches
  - problem: have no HF enriched, signal free control region
  - use special classification to reweight generators to Sherpa+OpenLoops

- see bad data/MC agreement in HF enriched regions in first 13 TeV results

- from experimental side: need more precise measurements of \( t\bar{t} + b(b) \) and \( c(c) \)
Outline

Motivation

$t\bar{t} + b(\bar{b})$ measurement

SM search: $t\bar{t}H(H \rightarrow b\bar{b})$

Exotics searches

Summary

Discussion
Discussion/ “Wishlist”

- $t\bar{t}$+HF MC: what is the best setup with systematic uncertainties for the NLO+PS merging?

- ideal: inclusive sample @ NLO for ttbb/cc/light, with systematics due to parameter variation in the generator

- differences between generators seen for ttbb in the Higgs LHC cross-section WG, need to understand NLO+PS merging and need to come up with reasonable uncertainties

- is Sherpa+OpenLoops currently the best on the market, or are there new developments in the pipeline?

- are there settings/parameters that we can test to understand the underprediction in the signal region?

- will there be an update for $t\bar{t} + c(c)$ in the near future?

- what diff. measurements would theorists like to see?