Heavy flavour modelling in top-related analyses at ATLAS

Andrea Knue

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

Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

Outline

Motivation

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

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

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: $t\bar{t}H(H \to b\bar{b})$
- Exotics searches
- Summary
- Discussion

$t\bar{t}b\bar{b}$ production @ 8 TeV

▸ Eur. Phys. J. C (2016) 76:11

- want to understand and test QCD
- best predictions from fixed order calc. @ NLO
- but: large uncertainties on predictions

  $\hookrightarrow$ 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 $t\bar{t}$+$b/c/l$

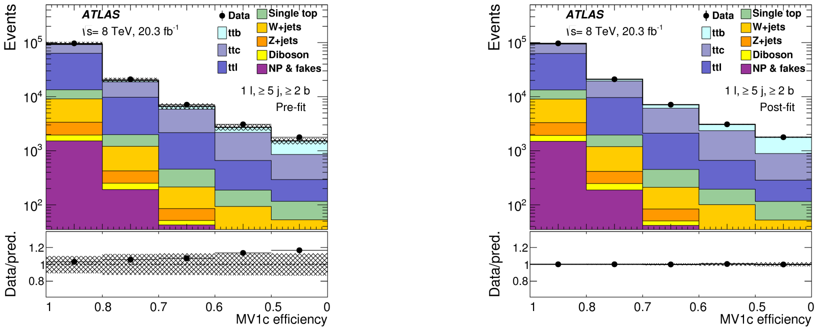
Fiducial $t\bar{t} + b\bar{b}$ cross-section

- ➔ 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:

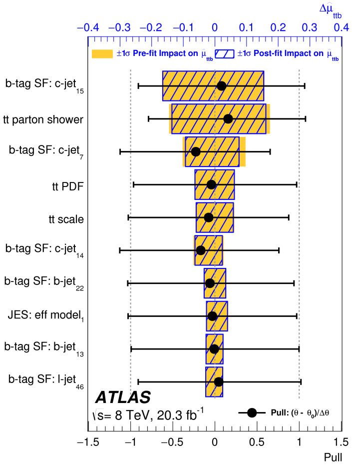
➔ profile likelihood fit of templates to data

➔ allows to strongly constrain the uncertainties (right plot)

➔ insignificant uncertainties are excluded from the fit

Impact of systematic uncertainties

Source	$\sigma^{\mathrm{fid}}_{ttb}$ Lepton-plus-jets uncertainty (%)	$\sigma^{\mathrm{fid}}_{ttb}$ ttb $e\mu$ uncertainty (%)	$\sigma^{\mathrm{fid}}_{ttbb}$ Cut-based uncertainty (%)	$\sigma^{\mathrm{fid}}_{ttbb}$ Fit-based uncertainty (%)	R_{ttbb} Fit-based uncertainty (%)
Total detector	+17.5 −14.4	+11.6 −8.0	±14.5	+11.9 −13.1	+10.9 −12.5
Jet (combined)	+3.9 −2.7	+10.1 −6.1	±5.5	+6.0 −8.5	+8.7 −10.7
Lepton	±0.7	+1.0 −0.5	±2.0	+2.4 −2.7	+0.8 −1.6
b−tagging effect on b−jets	+4.4 −4.0	+3.6 −3.1	±12.9	+9.4 −9.0	+6.0 −5.8
b−tagging effect on c−jets	+16.2 −13.4	+4.0 −3.6	±1.7	±1.4	+1.2 −1.3
b−tagging effect on light jets	+3.1 −2.0	+1.9 −2.0	±4.3	+3.3 −2.9	+2.2 −1.9
Total tt modelling	+13.1 −13.7	+23.8 −16.1	±23.8	±21.7	±16.1
Generator	+1.1 −1.4	+23.3 −15.1	±16.9	±17.4	±12.4
Scale choice	±4.3	+1.1 −2.7	±14.2	±9.5	±6.0
Shower/hadronisation	+11.4 −12.1	+3.0 −3.4	±8.2	±8.7	±7.1
PDF	+4.7 −4.5	±3.3	±3.3	±0.8	±4.1
Removing/doubling ttV and ttH	±0.4	+1.1 −0.9	±1.5	+3.1 −2.7	+3.0 −2.6
Other backgrounds	±0.8	+0.9 −0.8	±1.6	+3.5 −3.3	±2.5
MC sample size	< 1	< 1	±9.6	±7.4	±7.4
Luminosity	±2.8	±2.8	±3.2	±2.9	±0.1
Total systematic uncertainty	+25.5 −19.2	+30.5 −19.9	±29.5	+26.4 −26.9	+21.1 −21.9
Statistical uncertainty	±7.1	+19.2 −17.9	±18.4	±24.6	±25.2
Total uncertainty	+26.5 −20.5	+36.0 −26.8	±35.2	+36.1 −36.4	+32.9 −33.4



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 μ: PS and scale variations

Comparison of generator setups

Sample	Generator	Shower	$m_{\rm b}$ [GeV]	Comments
		b-quarks in ME, NLO+PS		
$t\bar{t}+b\bar{b}$	aMC@NLO	Pythia8	4.8	BDDP scale
$t\bar{t}+b\bar{b}$	aMC@NLO	Pythia8	4.8	$H_{\rm T}/4$ scale
$t\bar{t}+b\bar{b}$	Powhel	Pythia8	0.0	$H_{\rm T}/2$ scale
		$t\bar{t}+b\bar{b}$ @ LO+PS		
$t\bar{t}$+(0-3) p.	Madgraph5	Pythia6	4.8	
		b-quarks only in PS		
$t\bar{t}$ (a)	Pythia8	Pythia8	4.8	enhanced $g\to b\bar{b}$
$t\bar{t}$ (b)	Pythia8	Pythia8	4.8	suppr. $g\to b\bar{b}$
$t\bar{t}$ (c)	Pythia8	Pythia8	4.8	enhanced $g\to b\bar{b}$
$t\bar{t}$	Powheg	Pythia6	0.0	

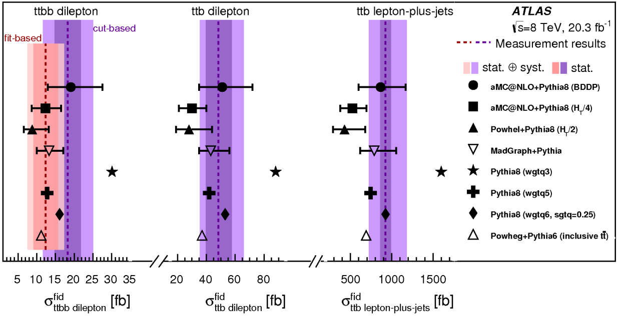
Options for $g\to b\bar{b}$ splitting

a) wgtq3: mass dep. terms in splitting kernel,

b) wgtq5: no mass dep. terms, sgtq=1: small $g\to b\bar{b}$ rate

c) wgtq6: mass dep. terms, sgtq=0.25: high $g\to b\bar{b}$ rate

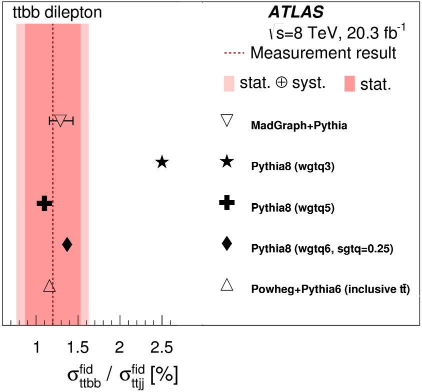
Results for fiducial cross-sections



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
  $\hookrightarrow$ extreme option with highest rate of $g \to b\bar{b}$: strongly disfavoured

Results ratio: Comparison for diff. $g \to b\bar{b}$ rates



$\to$ enhanced $g \to b\bar{b}$ rate (mass. dep. terms)

$\to$ reduced $g \to b\bar{b}$ rate (no mass. dep.)

$\to$ enhanced $g \to b\bar{b}$ rate (mass. dep. terms)

➔ results sensitive to modelling of gluon splitting in PS

➔ model with highest $g \to b\bar{b}$ rate overestimates the cross section by factor of 2!

Outline

- Motivation
- $t\bar{t} + b(\bar{b})$ measurement
- SM search: $t\bar{t}H(H \to b\bar{b})$
- Exotics searches
- Summary
- Discussion

Motivation: Why search for $t\bar{t}H$?

➔ after discovery of the Higgs Boson:

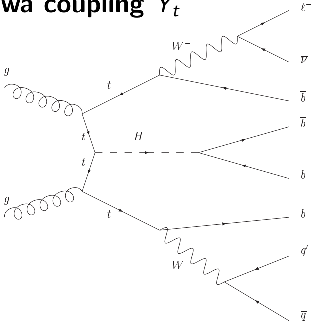
↪ what are its properties?

↪ is it really the SM particle?

➔ important: directly measure the **top-Higgs Yukawa coupling** Y_t

➔ top quark heaviest fermion:

↪ Y_t largest: ≈ 1



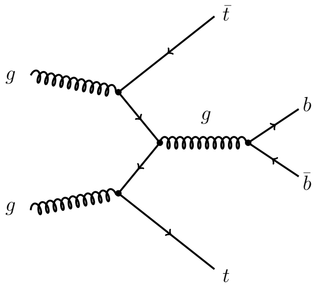
➔ any deviation would be sign for BSM processes

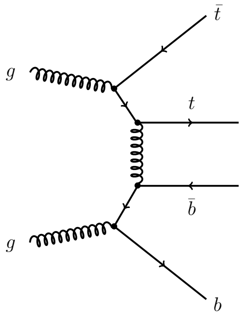
➔ but: challenging search

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

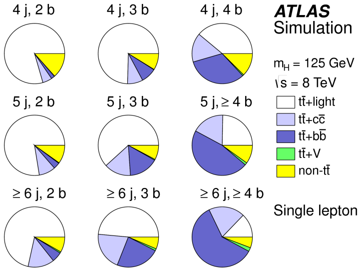
➔ use signal-depleted regions to constrain bkg and unc

➔ combined nuisance parameter fit to all regions

↪ analysis here: 8 TeV, 20.3 fb^{-1} ▸ Eur. Phys. J. C (2015) 75:349

Lepton+jets channel

Eur. Phys. J. C (2015) 75:349



Event selection

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

Discriminating variables

- signal-depleted regions: H_T^{had}
- signal-enriched regions: MVA output

  $\hookrightarrow$ input to MVA are 10 variables per region

➔ note: if for $g \rightarrow b\bar{b}$ the b's are within the same jet, events can migrate between different jet/tag categories

How to define different $t\bar{t}$+HF processes?

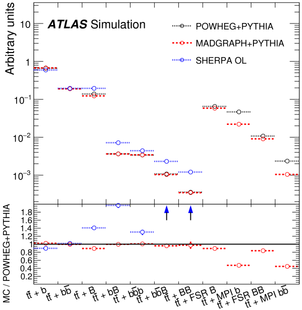
- take antiKt (R=0.4) particle level jets ($p_{\mathrm{T}} > 15$ GeV, $|\eta| < 2.5$)
- use truth matching to hadrons with $\Delta R < 0.4$, $p_{\mathrm{T}} > 5$ GeV
- now check additional jets in the event:

  $\hookrightarrow$ if add. jet contains one B/D hadron: $t\bar{t} + b$

  $\hookrightarrow$ if add. jet contains more than one B/D hadron: $t\bar{t} + B$

  $\hookrightarrow$ 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



➔ Sherpa does not include $t\bar{t}+b\bar{b}$ production via MPI and FSR

Samples used for analysis:

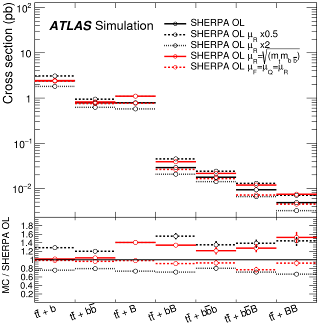
➔ Powheg+Pythia6: $t\bar{t}+b\bar{b}$ only from PS

➔ Madgraph+Pythia6: $t\bar{t}$+(0–3) partons (MLM matching)

➔ Sherpa+OpenLoops: (4FS, massive b-quarks)

$\hookrightarrow$ 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

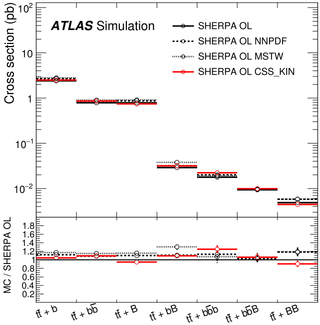


Other variations

- ➔ impact of MSTW larger than NNPDF
- ➔ impact of shower recoil visible

Scale variations

- ➔ large impact from μ_R variation
- ➔ large impact from functional form of μ_R

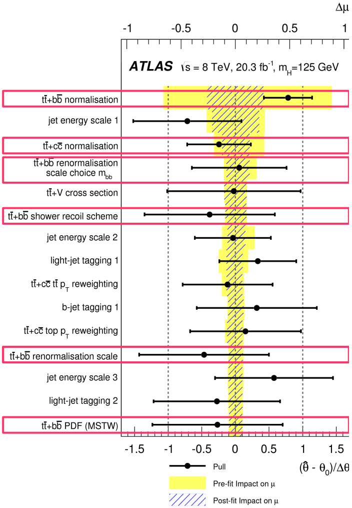


Systematic uncertainties and impact on result

- 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 $\mu_{\mathrm{R}} = \sqrt{m_t m_{b\bar{b}}}$
  - ➔ vary μ_{R} up and down
  - ➔ change func. form of μ_F and μ_Q
  - ➔ shower recoil scheme
  - ➔ PDF choice in Sherpa
  - ➔ missing MPI and FSR



Outline

Motivation

↓

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

↓

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

↓

Exotics searches

↓

Summary

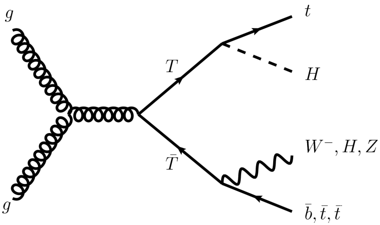
↓

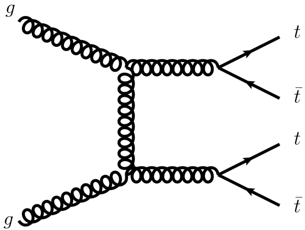
Discussion

VLQ and four-top quark searches @ 8TeV

JHEP 08 (2015) 105

- $T\bar{T} \to Ht + X$ 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

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

$\geq$6 j, $\geq$4 b, high $M_{bb}^{\min\Delta R}$

	Pre-fit				Post-fit		
	Signal	$t\bar{t}$+light-jets	$t\bar{t}+c\bar{c}$	$t\bar{t}+b\bar{b}$	$t\bar{t}$+light-jets	$t\bar{t}+c\bar{c}$	$t\bar{t}+b\bar{b}$
Luminosity	±2.8	±2.8	±2.8	±2.8	±2.6	±2.6	±2.6
Lepton efficiencies	±1.5	±1.5	±1.5	±1.5	±1.5	±1.4	±1.5
Jet energy scale	±4.4	±15	±11	±12	±8.7	±6.4	±6.7
Jet efficiencies	–	±4.0	±2.2	±1.9	±2.7	±1.5	±1.3
Jet energy resolution	±0.1	±4.4	±3.8	±0.5	±3.1	±2.6	±0.4
b-tagging efficiency	±13	±5.6	±5.4	±9.3	±4.6	±4.6	±6.6
c-tagging efficiency	±1.6	±5.8	±12	±3.1	±5.6	±11	±2.9
Light-jet tagging efficiency	±0.6	±20	±5.7	±2.0	±17	±5.1	±1.8
High-p_T tagging efficiency	±4.8	±0.7	±1.7	±1.6	±0.6	±1.3	±1.2
$t\bar{t}$: reweighting	–	±13	±15	–	±10	±10	–
$t\bar{t}$: parton shower	–	±28	±17	±6.2	±13	±11	±4.0
$t\bar{t}$+HF: normalisation	–	–	±50	±50	–	±32	±18
$t\bar{t}$+HF: modelling	–	–	±17	±12	–	±16	±10
Theoretical cross sections	–	±6.3	±6.3	±6.3	±4.6	±4.6	±4.6
Total	±15	±42	±61	±55	±22	±30	±15

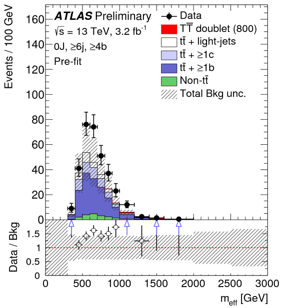
$\geq$6 j, $\geq$4 b, high $M_{bb}^{\min\Delta R}$

	Pre-fit				Post-fit		
	Signal	$t\bar{t}$+light-jets	$t\bar{t}+c\bar{c}$	$t\bar{t}+b\bar{b}$	$t\bar{t}$+light-jets	$t\bar{t}+c\bar{c}$	$t\bar{t}+b\bar{b}$
Luminosity	±2.8	±2.8	±2.8	±2.8	±2.7	±2.7	±2.7
Lepton efficiencies	±1.6	±1.4	±1.5	±1.7	±1.4	±1.5	±1.6
Jet energy scale	±5.6	±14	±14	±11	±13	±14	±11
Jet efficiencies	±3.1	±3.3	±1.0	±0.9	±3.2	±0.9	±0.8
Jet energy resolution	±0.1	±6.0	±1.1	±1.9	±4.5	±0.9	±1.5
b-tagging efficiency	±16	±7.6	±9.2	±16	±3.9	±5.2	±7.5
c-tagging efficiency	±1.0	±6.1	±15	±3.0	±5.8	±14	±2.8
Light-jet tagging efficiency	–	±19	±6.3	±2.4	±18	±5.8	±2.3
High-p_T tagging efficiency	±11	±2.7	±5.3	±5.0	±1.9	±3.8	±3.6
$t\bar{t}$: reweighting	–	±15	±16	–	±14	±15	–
$t\bar{t}$: parton shower	–	±22	±35	±26	±14	±33	±24
$t\bar{t}$+HF: normalisation	–	–	±50	±50	–	±44	±30
$t\bar{t}$+HF: modelling	–	–	±27	±24	–	±28	±21
Theoretical cross sections	–	±6.3	±6.2	±6.3	±5.9	±5.9	±5.9
Total	±21	±38	±73	±65	±24	±46	±27

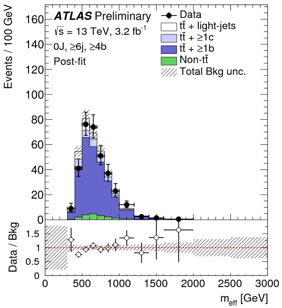
- $t\bar{t}t\bar{t}$ production (left)
- large uncertainties on $t\bar{t} + b\bar{b}$ PS and modelling
- post-fit uncertainties: strongly reduced

VLQ and $t\bar{t}t\bar{t}$ @ 13 TeV ▸ ATLAS-CONF-2016-013

– prefit –



– postfit –



➔ similar analysis strategy and HF modelling/uncertainties as for 8 TeV

➔ see large discrepancy between data and MC

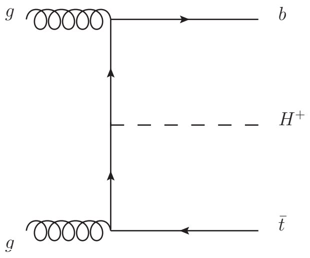
➔ combined likelihood fit to 11 channels:

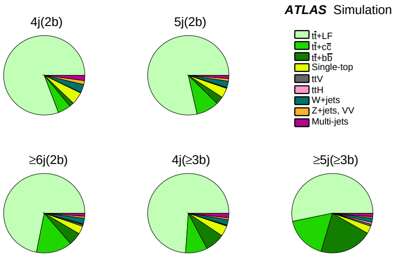
↪ $t\bar{t} + b\bar{b}$ contribution increases about a factor of 2!

Search for $H^+ \rightarrow tb$ production @ 8 TeV

JHEP 03 (2016) 127

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





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

Outline

Motivation

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

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

Exotics searches

Summary

Discussion

Summary

- heavy-flavour production in association with top-quarks is difficult to predict

  $\hookrightarrow$ need experimental measurements for better understanding (need higher stats)

- $t\bar{t} + b\bar{b}$ is important background for a lot of searches

  $\hookrightarrow$ problem: have no HF enriched, signal free control region

  $\hookrightarrow$ 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

$\downarrow$

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

$\downarrow$

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

$\downarrow$

Exotics searches

$\downarrow$

Summary

$\downarrow$

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 resonable 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?