Heavy quarks and the Higgs: $t\bar{t}b\bar{b}$ with ATLAS

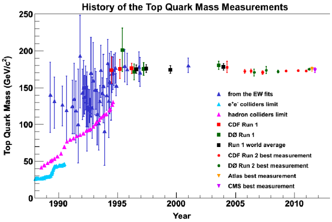
Tom Neep

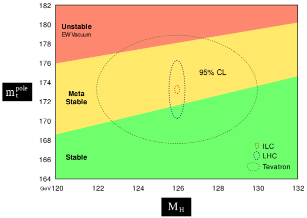
October 9, 2019

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement no 714893 (ExclusiveHiggs)

The top quark

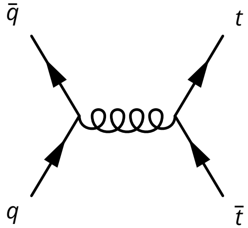
- The top quark is the **heaviest** known fundamental particle
- It has a mass of $\approx$ 173 GeV. **40 times the mass of the next heaviest quark!**
- Similar to that of a **gold atom**.
- Discovered in **1995** at the Tevatron by the CDF and D0 experiments
- “Rediscovered” at the LHC in 2010
- **Unique** amongst the quarks – **decays before hadronisation**
- Yukawa coupling of $\mathcal{O}(1)$. Some special relationship with the Higgs?

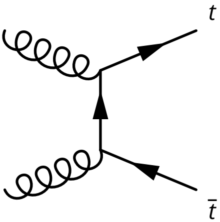




Top quark pair production

- Top quarks produced most often in **pairs**
- At the $\sqrt{s} = 13$ TeV around 90% of $t\bar{t}$ pairs are produced via $gg \rightarrow t\bar{t}$ and the remaining 10% by $q\bar{q} \rightarrow t\bar{t}$
- $\sigma_{t\bar{t}} \approx 830$ pb (NNLO+NNLL QCD) $\rightarrow \approx$ **10 $t\bar{t}$ pairs / s** at a luminosity of $10^{34} \text{cm}^{-2}\ \text{s}^{-1}$





- Large number of tops allows us to make **precise cross-section measurements**
- Many **new physics models** enhance the $t\bar{t}$ cross-section
- Large number of $t\bar{t}$ pairs allows us to measure $\boldsymbol{t\bar{t} + X}$ where X can be $H, W, Z, \gamma, b\bar{b}$ and possibly one day even $t\bar{t}$

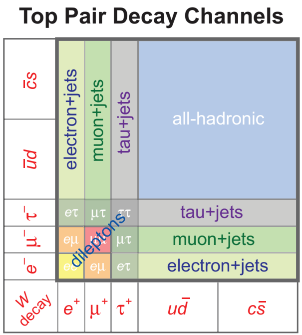
Top quark decay

- The top quark decays nearly 100% of the time to a *b*-quark and a *W*-boson

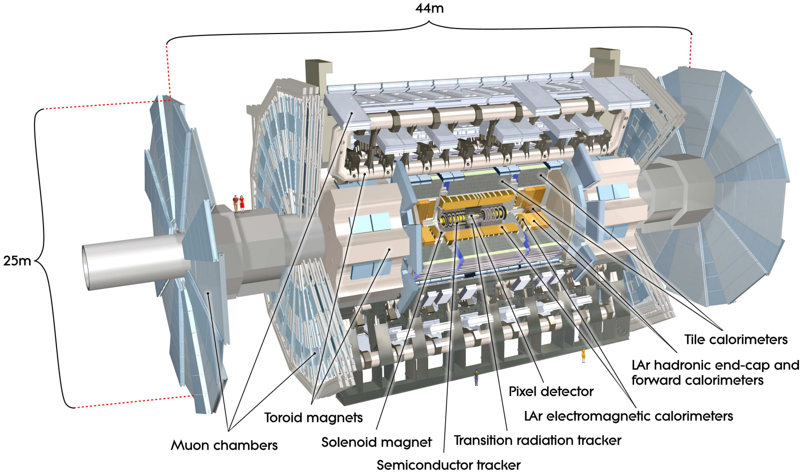
$$t\bar{t} \rightarrow W^+W^-b\bar{b}$$

- $t\bar{t}$ decays are therefore categorised based on how each of the two *W*s decay

- Three main channels
  1. All-hadronic
  2. Dilepton
  3. Semi-leptonic
- The $t\bar{t}$ final state can include electrons, muons, taus, neutrinos (not detected) and jets (including *b*-jets).
- We need to make use of the entire ATLAS detector to make measurements!

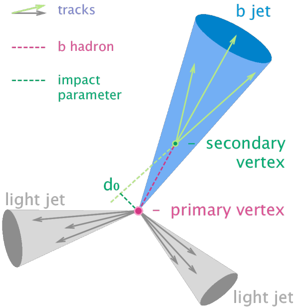


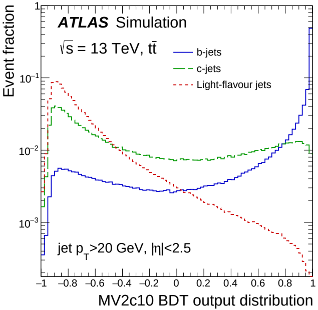
The ATLAS detector



b-tagging

- ***b*-tagging** is crucial for top physics
- Exploit large **impact parameters**, **seconday vertices** and $\boldsymbol{b \to c}$ **decay chains**
- Information is combined using a **Boosted Decision Tree** to identify jets containing b-hadrons





ϵ_b [%]	light-jet mistag	c-jet mistag
60	1550	35
70	380	12
77	135	6
85	35	3

NEW: $t\bar{t}$ cross-section in $e\mu$ events

ATLAS-CONF-2019-041

- The most accurate $t\bar{t}$ cross-section measurements have been made in the $e\mu$ channel
- This is a very clean channel with only small backgrounds
- "Simple" technique, count the number of b-tagged jets

$$N_1 = L\sigma_{t\bar{t}}\epsilon_{e\mu}2\epsilon_b(1 - C_b\epsilon_b) + N_1^{\mathrm{bkg}}$$

$$N_2 = L\sigma_{t\bar{t}}\epsilon_{e\mu}C_b\epsilon_b^2 + N_2^{\mathrm{bkg}}$$

L: Integrated luminosity

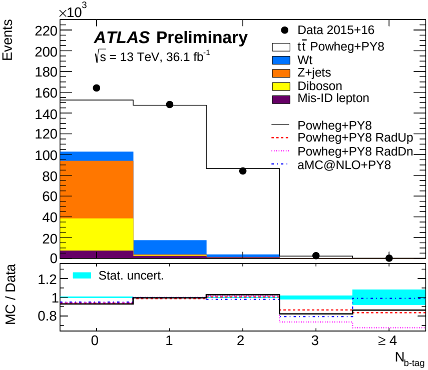
$\sigma_{t\bar{t}}$: $t\bar{t}$ cross-section

$\epsilon_{e\mu}$: Efficiency for event to have one electron and one muon ($\approx 1\%$)

ϵ_b: Efficiency to tag and select a b-jet

C_b: b-tagging correlation ≈ 1

$N_{1,2}^{\mathrm{bkg}}$: Number of background events with 1/2 b-tags



NEW: $t\bar{t}$ cross-section in $e\mu$ events

ATLAS-CONF-2019-041

Uncertainty source (%)	$\Delta\sigma_{t\bar{t}}/\sigma_{t\bar{t}}$
Data statistics	0.44
$t\bar{t}$ mod.	0.97
Lept.	0.59
Jet/b	0.21
Bkg.	0.78
Analysis systematics	1.39
Integrated luminosity	1.90
Beam energy	0.23
Total uncertainty	2.40

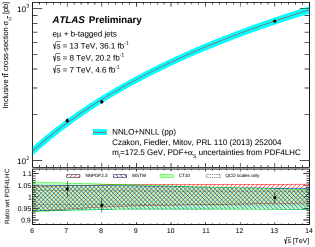
- Uncertainties are statistical, systematic, luminosity and beam energy
- The total uncertainty is dominated by the luminosity uncertainty
- $t\bar{t}$ and background modelling are the next largest

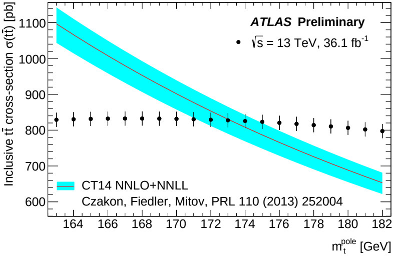
NEW: $t\bar{t}$ cross-section in $e\mu$ events

ATLAS-CONF-2019-041

Result: $\sigma_{t\bar{t}} = 826.4 \pm 3.6 \pm 11.5 \pm 15.7 \pm 1.9$ pb (2.4%)

- Analysis has now been performed at 7, 8 and 13 TeV
- All results are consistent with the SM (NNLO+NNLL QCD) prediction
- The measurement is more precise than the prediction!





Some Birmingham involvement!

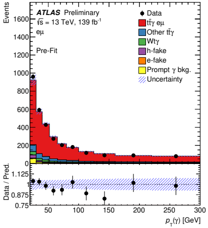
EB chair: Miriam Watson

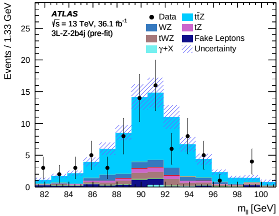
Top cross-section convener: TN

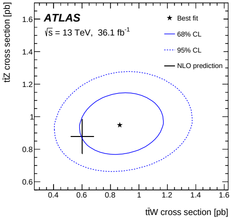
$$m_t^{\text{pole}} = 173.1 \pm 1.0(\text{exp.})^{+1.8}_{-2.1}(\text{theory}) \text{ GeV}$$

$t\bar{t} + X$

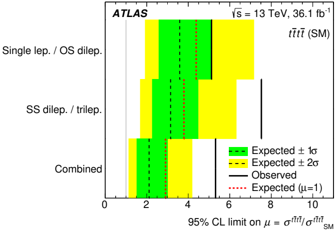
- ATLAS has also measured $t\bar{t}$ production in association other particles





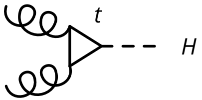


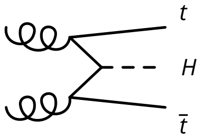
- $t\bar{t} + \gamma$ and $t\bar{t} + Z$ give very clean signals
- Can start to measure differential distributions
- $t\bar{t}W$ more challenging
- Searches for $t\bar{t}t\bar{t}$ ongoing, but will likely need more data for evidence

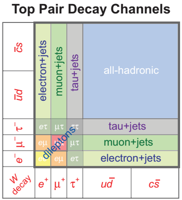


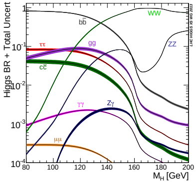
The Higgs and the top

- It is important that we make the most of the LHC and study the Higgs as comprehensively as possible
- The top Yukawa coupling can be probed through loops but also directly in Higgs production in association with top quarks ($t\bar{t}H$)



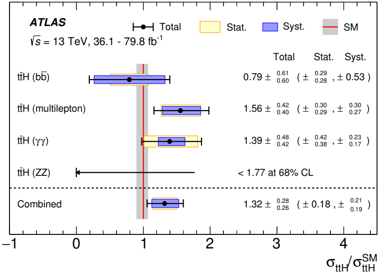


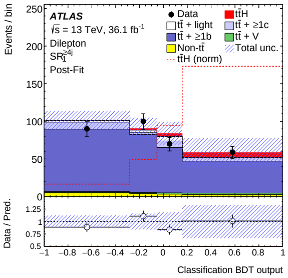




- The $t\bar{t}H$ process can decay to a large number of different final states. The more we measure the better!
- $H \to b\bar{b}$ is the dominant decay – can we measure $t\bar{t}H(H \to b\bar{b})$?

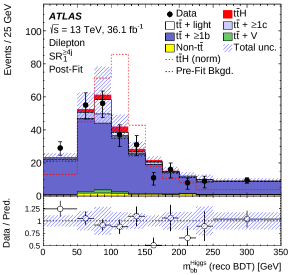
- ATLAS observed $t\bar{t}H$ production last year
- Sensitivity comes mainly from the $H \to \gamma\gamma$ and **multilepton** channels ($H \to \tau\tau$ and $H \to WW^*$)
- $H \to b\bar{b}$ not competitive, despite large branching ratio
- The sensitivity of the $t\bar{t}H(b\bar{b})$ channel is limited by systematic uncertainties on the QCD $t\bar{t}b\bar{b}$ background





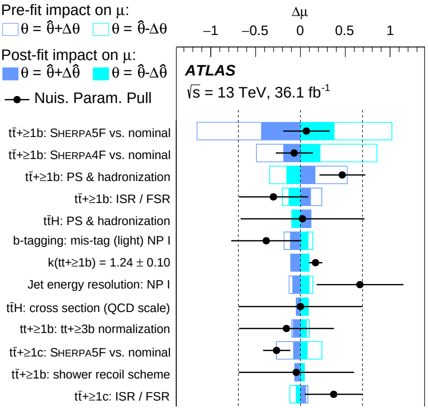
- Measuring $t\bar{t}H(bb)$ is extremely challenging
- Final state with four b-jets – need to determine which jets are from $H \to b\bar{b}$ and which are from $t \to Wb$
- Use MVA techniques to reconstruct the system and to separate signal from background

- Background is completely dominated by $t\bar{t}b(\bar{b})$



$t\bar{t}H(b\bar{b})$ uncertainties

Phys. Rev. D 97 (2018) 072016

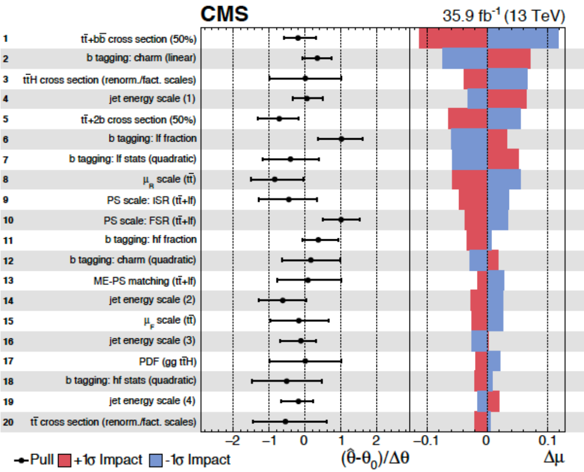


- $k(t\bar{t}+\geq 1b)$ is a normalization parameter
- Measure 1.24 ± 0.10 (uncertainty statistical)

- The four uncertainties with the largest impact on the limit are all $t\bar{t}b$ related
- A better understanding of $t\bar{t}b\bar{b}$ crucial for this channel to be useful

CMS uncertainties

JHEP 03 (2019) 026



CMS result: $\mu = 0.72 \pm 0.24 \pm 0.38$ ATLAS result: $\mu = 0.79 \pm 0.29 \pm 0.53$

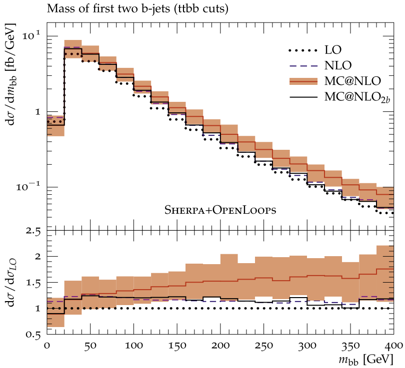
$t\bar{t}b\bar{b}$ predictions

- Clearly, an improved understanding of $t\bar{t}b\bar{b}$ is need for searches/measurements of $t\bar{t}H(b\bar{b})$
- Many other searches have large $t\bar{t}b\bar{b}$ backgrounds (SUSY, four top production . . .)

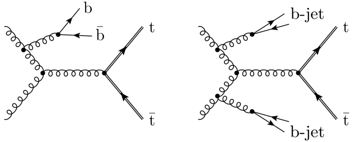
- Predicting $t\bar{t}b\bar{b}$ is challenging
- Massive b-quarks in the matrix element, large scale differences (m_t versus m_b)
- NLO predictions of $t\bar{t}b\bar{b}$ started to arrive about five years ago
- Some surprising results . . .

$t\bar{t}b\bar{b}$ predictions

- NLO $t\bar{t}b\bar{b}$ production with massive b-quarks in the matrix element using the four flavour scheme



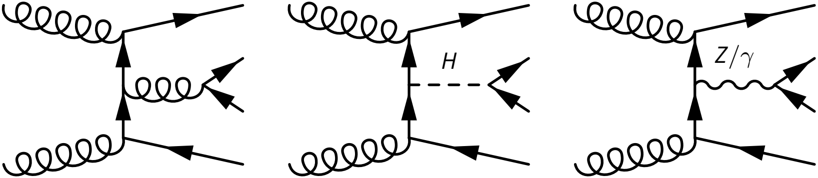
- The effect of $g \to b\bar{b}$ splitting in the parton shower is important (MC@NLO vs. MC@NLO$_{2b}$)
- The contribution of the right diagram below is surprisingly large



- Parton shower effects still important at NLO
- Cross-section uncertainties range from 20-40% (depending on fiducial cuts)

Measuring $t\bar{t}b\bar{b}$

- Experimental input is required to move our understanding of $t\bar{t}b\bar{b}$ forward
- ATLAS has performed a measurement of $t\bar{t}$ with additional heavy-flavour jets at 13 TeV, using data collected in 2015 & 2016
- Fiducial cross-sections measured
- We **DO NOT** attempt to identify which *b*-jets are from the top quarks and which are considered "additional"
- The measurement therefore includes "QCD" $t\bar{t}b\bar{b}$, $t\bar{t}H$ and $t\bar{t}Z$



Analysis outline

1. Select an inclusive (≥ 2 *b*-jets) sample of $t\bar{t}$ events
2. Categorise simulated $t\bar{t}$ events based on the "flavours" of the jets in the event. Use these to create templates from $t\bar{t}$ simulation
3. Fit the templates to data in a discriminating variable
4. From the results of this fit, measure inclusive and differential fiducial cross-sections

---

The analysis is performed in two channels

- ℓ+jets
- $e\mu$

Analysis selection

- Both channels require the ATLAS detector to be fully operational
- A primary vertex with at least two tracks
- Single electron/muon triggers with $p_{\mathrm{T}} > 20(26)$ GeV for muons and $p_{\mathrm{T}} > 24(26)$ GeV for electrons in 2015 (2016)

ℓ+jets

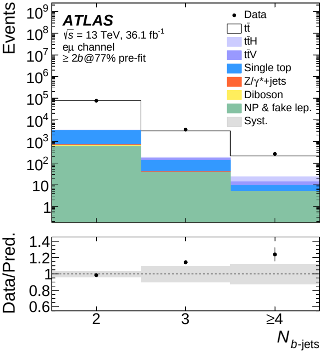
- 1 $\ell(e/\mu)$ with $p_{\mathrm{T}} > 27$ GeV
- ≥ 5 jets with $p_{\mathrm{T}} > 25$ GeV, $|\eta| < 2.5$
- ≥ 2 tagged at the 60% *b*-tagging efficiency WP

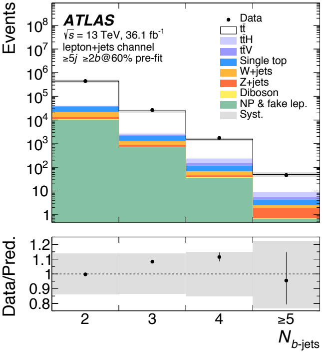
$e\mu$

- 1 e and 1 μ with $p_{\mathrm{T}} > 27$ GeV
- $Q^{e} \cdot Q^{\mu} = -1$
- ≥ 2 jets with $p_{\mathrm{T}} > 25$ GeV, $|\eta| < 2.5$
- ≥ 2 jets tagged at the 77% *b*-tagging efficiency WP

Number of *b*-tags after pre-selection

JHEP 04 (2019) 046



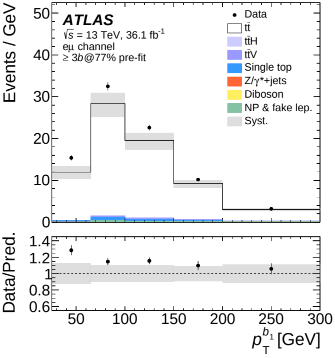


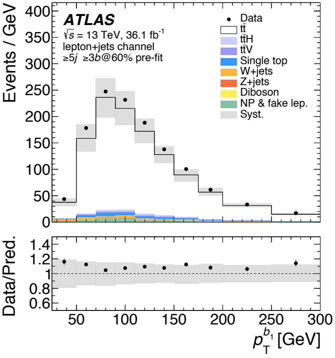
- After pre-selection there is a slope in the data / MC ratio in the number of *b*-jets distribution
- The number of events with ≥ 3 *b*-jets is under-estimated
- We want to identify the cause of this. Is it due to modelling or an experimental effect (flavour tagging?)?

Leading b-jet p_{T}

JHEP 04 (2019) 046

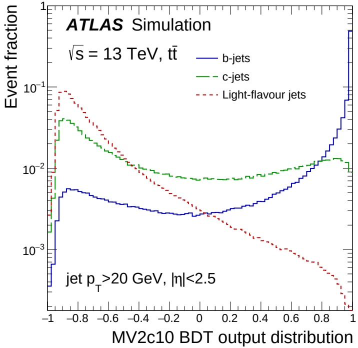
- We can also look at jet variables
- Here we can clearly see the ”purity” of the sample we are dealing with
- Non-$t\bar{t}$ backgrounds are small contributions in both channels

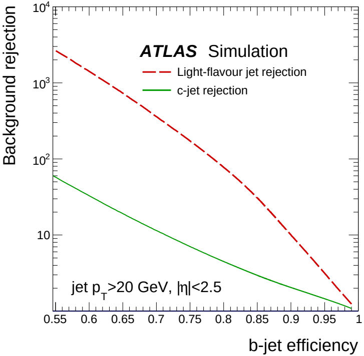




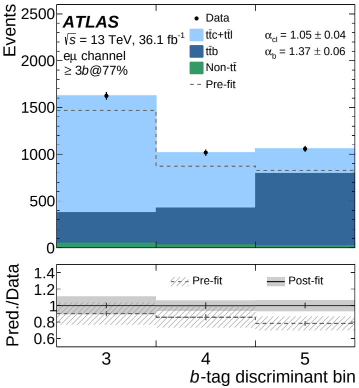
JHEP 08 (2018) 89

- To discriminate between the various cases, we use the output of the *b*-tagging variable
- The output of the *b*-tagging algorithm is split into five bins, each of which is calibrated
- The tightest *b*-tagging working point (5) is 60% efficiency and has a light(*c*)-jet rejection rate of $\approx 1550(35)$



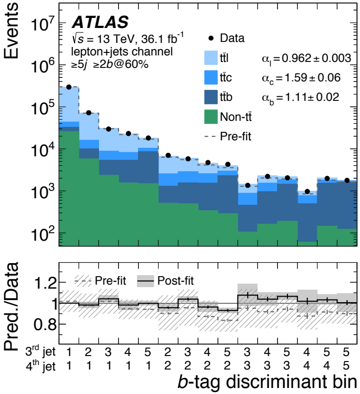


- In the $e\mu$ channel the fit is performed using the b-tagging discriminant of the jet with the third largest value of D_{MV2}
- $t\bar{t}c$ and $t\bar{t}l$ templates are combined
- A systematic uncertainty is included by varying the normalisation of the $t\bar{t}c$ template by $\pm 40\%$ before combining with the $t\bar{t}l$ template
- The best fit value scales the $t\bar{t}b$ template by ≈ 1.4



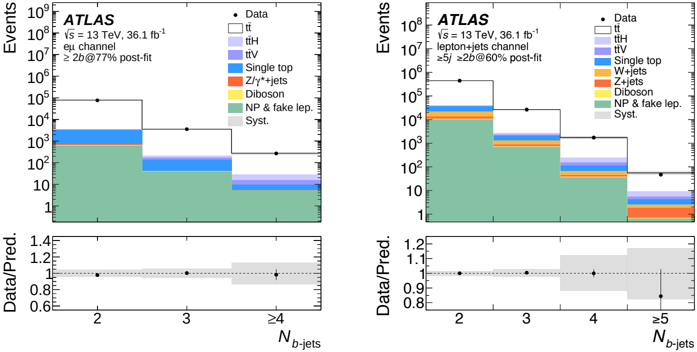
ℓ+jets channel

JHEP 04 (2019) 046



- In the ℓ+jets channel there are always at least four jets and so the fit is performed using the *b*-tagging discriminants of the two jets with the third and fourth largest values of D_{MV2}
- 2D fit flattened to 1D in the figure
- $t\bar{t}c$ and $t\bar{t}l$ templates are treated separately
- The best fit value scales the $t\bar{t}b$ template by $\approx$ 1.1
- Can see that the final bin in the distribution, which is equivalent to four very tight *b*-tags, is very pure in $t\bar{t}b$ events

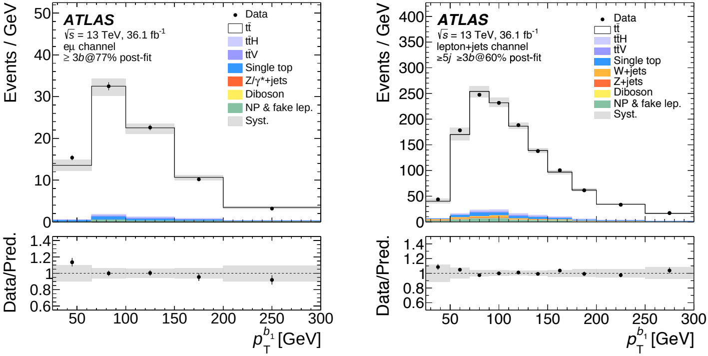
Applying the correction factors



- We can apply these correction factors back to our poorly modelled distributions
- There is a clear improvement in the agreement between the data and the prediction

Applying the correction factors

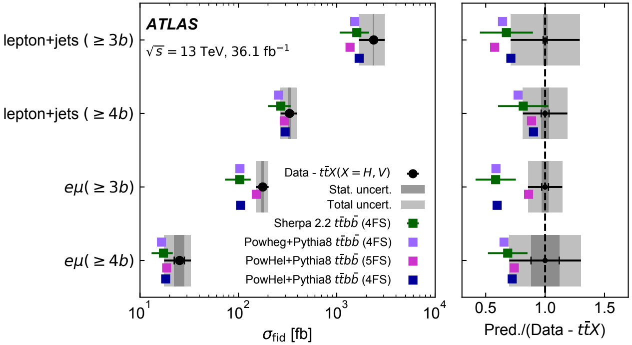
JHEP 04 (2019) 046



- We can apply these correction factors back to our poorly modelled distributions
- There is a clear improvement in the agreement between the data and the prediction

$t\bar{t}b$ fiducial cross-sections

JHEP 04 (2019) 046



- Using these correction factors we can measure $t\bar{t} + b$-jet fiducial cross-sections
- Data here refers to the measured cross-section
- The $t\bar{t}H$ and $t\bar{t}Z$ components are subtracted from the results to allow for easy comparison with QCD $t\bar{t}b\bar{b}$ predictions
- Measured fiducial cross-sections generally **larger** than $t\bar{t}b\bar{b}$ predictions by $\approx 1\sigma$
- **Confirms** what had been seen in related analyses ($t\bar{t}e\mu$, $t\bar{t}H$)
- Uncertainties range from 13 – 28 %

Uncertainties

Source	Fiducial cross-section phase space			
	$e\mu$		$\ell + \text{jets}$	
	$\geq 3b$ unc. [%]	$\geq 4b$ unc. [%]	$\geq 5j, \geq 3b$ unc. [%]	$\geq 6j, \geq 4b$ unc. [%]
Data statistics	2.7	9.0	1.7	3.0
Detector+background total syst.	8.5	14	18	12
$t\bar{t}$ modelling total syst.	10	20	21	12
Total	13	26	28	17

- Largest uncertainties due to b-tagging (mistagging light and c-jets) and $t\bar{t}$ modelling
- Improving these areas important to understand $t\bar{t} + b$-jets in more detail

Differential measurements

JHEP 04 (2019) 046

- In addition, several differential measurements are made
- Use 3 b-jets in the $e\mu$ channel and 4 b-jets in the ℓ+jets channel
- “Simple” variables are chosen
- $N_{\mathrm{b-jets}}$ in the $e\mu$ channel
- The b-jet p_{T}s

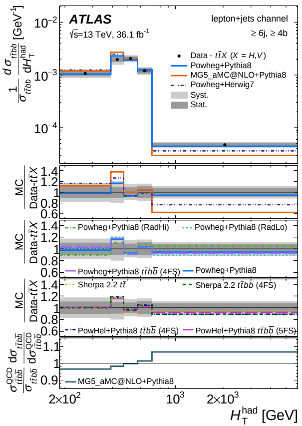
$$p_{\mathrm{T}}^{b,1}, p_{\mathrm{T}}^{b,2}, p_{\mathrm{T}}^{b,3}, p_{\mathrm{T}}^{b,4}$$

- The scalar sum of jet p_{T} and lepton p_{T} (used in $t\bar{t}H(b\bar{b})$ MVAs)

$$H_{\mathrm{T}}^{\mathrm{jets}} = \sum_{i \in \mathrm{jets}} p_{\mathrm{T}}^{i}, \quad H_{\mathrm{T}} = H_{\mathrm{T}}^{\mathrm{jets}} + p_{\mathrm{T}}^{\ell}$$

- Properties of the bb system (both the leading two b-jets and the closest two b-jets)

$$p_{\mathrm{T}}^{bb}, m_{bb}, \Delta R_{bb}$$



- The shapes of distributions are generally reasonably described

Summary and outlook for $t\bar{t}b\bar{b}$

- $t\bar{t} + b$-jets measured by ATLAS in two channels
- $t\bar{t} + b$-jet cross-sections measured to be larger than what is predicted
- **Results confirm** what has been **hinted** at in other analyses
- **No major shape differences** seen in distributions

- We have **four times more data** on disk than has been analysed
- As with other $t\bar{t}$ measurements, the very clean $e\mu$ channel will become the most useful channel
- We need to understand **which variables** in the $t\bar{t}H(b\bar{b})$ MVA cause large modelling uncertainties!
- Attempt to assign jets from top decays in next measurement?
- **MC improvements** are on the way!

The view from the top

- The top quark may be a window to new physics
- Testing how it interacts with other particles, particularly the Higgs, is important
- The latest $t\bar{t}$ measurements from ATLAS are a challenge to theorists!
- The future (and present) of top cross-section measurements is **differential** and **associated**!!