TOP2020 – IPPP Durham

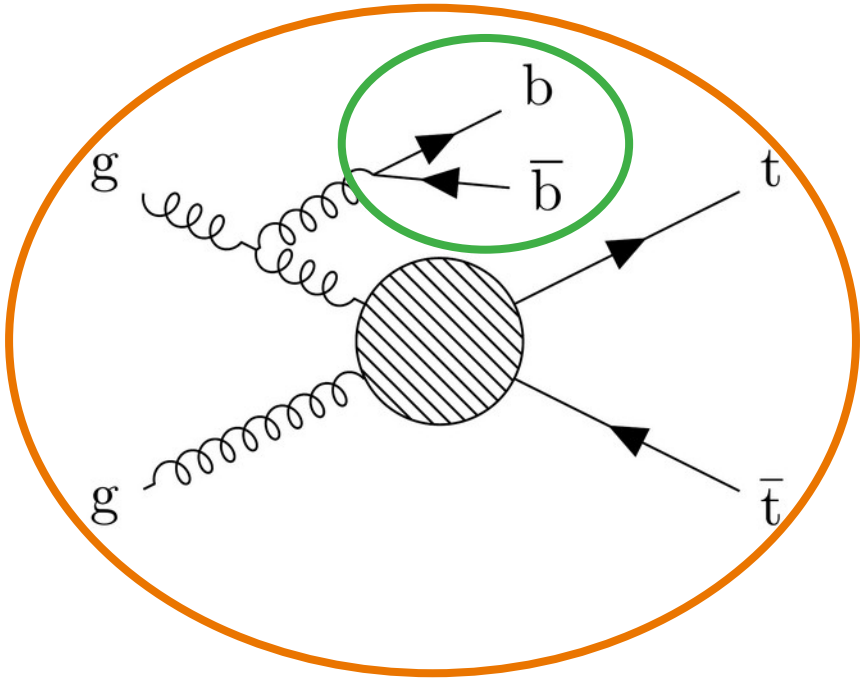
15th September 2020

Top quark production with heavy-flavour jets

Sébastien Wertz, for the ATLAS and CMS collaborations

$t\bar{t}$ + heavy flavour jets

- $t\bar{t}b\bar{b}$ (and to a lesser extent $t\bar{t}c\bar{c}$) production @LHC: unique modelling challenges!
- Currenly different approaches:
  - Inclusive $t\bar{t}$ ME (NLO) + PS, 5FS
    - (Powheg, MG5_aMC@NLO) x (Pythia, Herwig)
  - Multi-leg merged $t\bar{t}$ + jets ME + PS, 5FS
    - MG5_aMC@NLO + Pythia, Sherpa
  - $t\bar{t}b\bar{b}$ ME (NLO) + PS, 4FS
    - Powheg+OL+Pythia, Sherpa+OL



(Varying) sensitivity to:

- Scale choice ($t\bar{t} \leftrightarrow b\bar{b}$ gap)
- Parton shower modelling
- NLO+PS matching

- Spread in the predictions
- Large intrinsic uncertainties

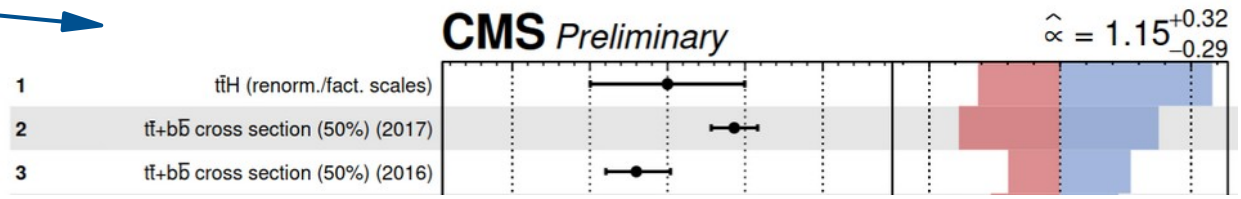
The future? [F. Siegert, ZPW20]

- Consistent merging of 5FS $t\bar{t}$+jets and 4FS $t\bar{t}b\bar{b}$
- Better tuning of scales

ME = Matrix Element, PS = Parton Shower
5FS = 5-flavour scheme, massless b+c quarks (inside PDFs)
4FS = 4-flavour scheme, massive b quarks, massless c quarks

$t\bar{t}$ + heavy flavour jets: why the fuss?

- Stress-test for perturbative QCD predictions → intrinsic value!
- Background for other high-interest processes, in particular:
  - $t\bar{t}H(b\bar{b})$
  - Four tops



PAS-HIG-18-030

Observation of $t\bar{t}H$, evidence for $t\bar{t}H(b\bar{b})$

- $t\bar{t}b\bar{b}$ background modelling = **leading source of uncertainty**

Phys. Lett. B 784 (2018) 173
Phys.Rev.D 97 (2018) 072016
Phys. Rev. Lett. 120 (2018) 231801

ATLAS: evidence for $t\bar{t}t\bar{t}$ ($\mu=2.0^{+0.8}_{-0.6}$) 2007.14858

- **50% uncertainty** on irreducible $t\bar{t}(V/H)$ + b jets background

See also CMS: Eur. Phys. J. C 80 (2020) 75, JHEP 11 (2019) 082

- Large, additional uncertainties on $t\bar{t}(X)$ + b jets (taken from measurements)

Flavour tagging

Caveat: results shown today not based on most recent b/c taggers!

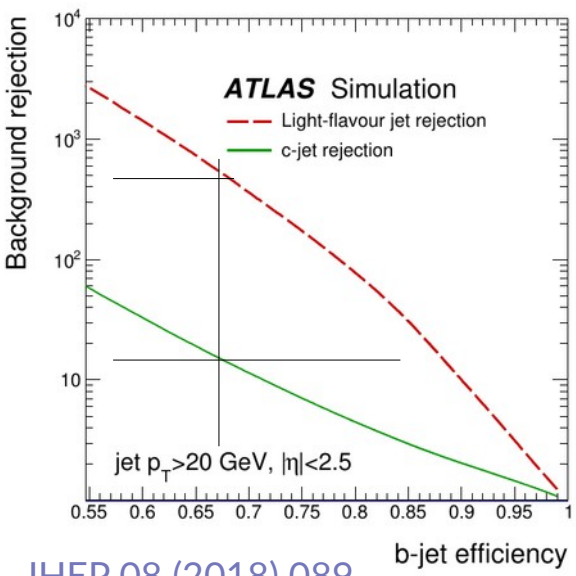
- Mandatory tools to disentangle $t\bar{t}$+HF from overwhelming $t\bar{t}$+light jets!
- B/D hadrons "long" lifetime, large mass → displaced tracks, secondary vertices in jets
  - Rely on exceptional performance of pixel detectors

B tagging:

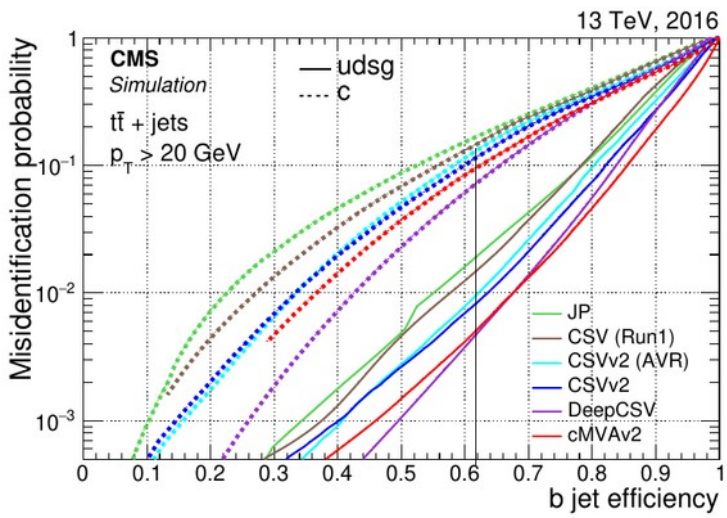
- ATLAS: BDT ("MV2c10")
- CMS: NN ("CSVv2")

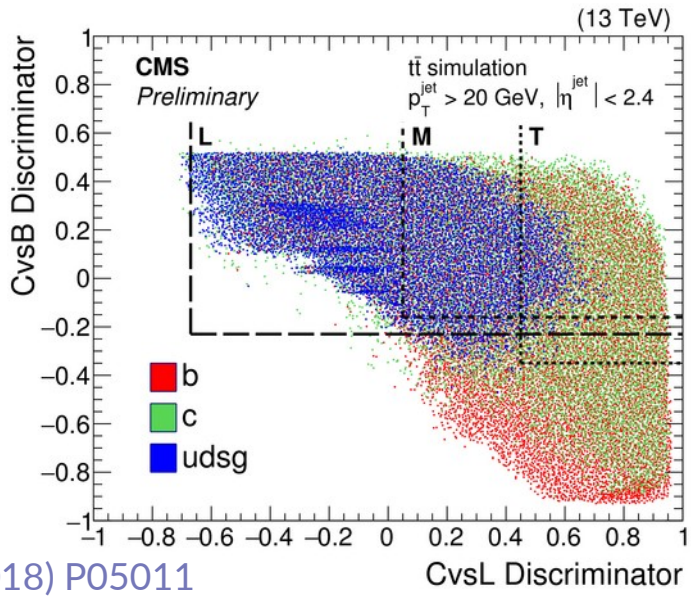
C tagging:

- c jets "in between" b and light jets
- CMS: DNN ("DeepCSV")
  → CvsL and CvsB discriminators



JHEP 08 (2018) 089





JINST 13 (2018) P05011

State of the art for full-Run 2 analyses: RNN(+CNN)-based taggers ("deepJet", "DL1r")

Flavour tagging calibration

Caveat: results shown today not based on most recent b/c taggers!

- Correct algorithm performance in the simulation
- Quantify uncertainty in the performance
- All $t\bar{t}$+HF measurements rely on correcting shape of discriminators
  - 5 bins (ATLAS), pseudo-continuous (CMS)

	ATLAS	CMS	Limiting factors
Heavy flavour	$t\bar{t}$ events (2l) Δ≈2-10% Tag and probe	Δ≈1-5%	$t\bar{t}$ modelling (light radiation) Soft physics (fragmentation, gluon splitting, ...)
	c mistag in $t\bar{t}$ (1l) Δ≈6-22%	No dedicated c mistag scale factors Iterative procedure	
Light flavour	Z+jets events		HF contamination Flipped tagger extrapolation Tracking modelling
	"Flipped" taggers Δ≈15-75%	"anti-tag" and probe Δ≈10-20%	
Uncertainty model provided	Eigenvector decomposition (>100)	Simplified: 8 + 26 (JES)	

$t\bar{t}b\bar{b}$ cross section: all-jet final state

Phys. Lett. B 803 (2020) 135285

2016 data, 36 fb^{-1}

- Fully-hadronic channel: 8 jets, of which 4 b jets
  → largest branching fraction (45%) & fully reconstructible final state
- Select events with ≥ 8 jets, ≥ 2 b-tagged jets
- Suffers from:
  - Combinatorial self-background

<u>"Permutation" BDT:</u>

- Trained to identify jets from $t\bar{t}$ decays
- ~60% correct (if all jets reconstructed)
- Keep permutation with highest score

$t\bar{t}b\bar{b}$ cross section: all-jet final state

Phys. Lett. B 803 (2020) 135285

2016 data, 36 fb^{-1}

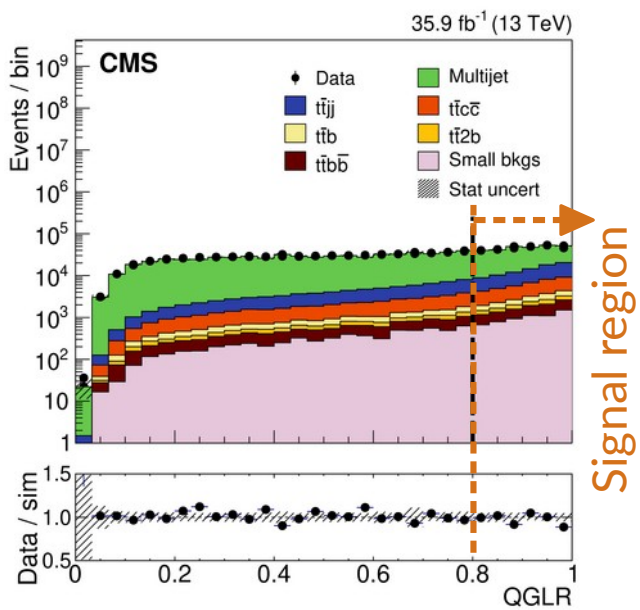
- Fully-hadronic channel: 8 jets, of which 4 b jets
  → largest branching fraction (45%) & fully reconstructible final state
- Select events with ≥ 8 jets, ≥ 2 b-tagged jets
- Suffers from:
  - QCD multijet background

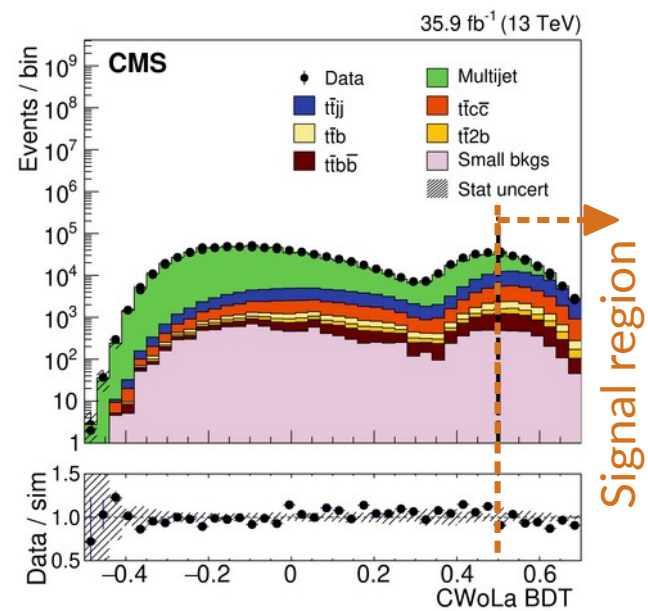
Quark-gluon likelihood ratio (QGLR):

- Based on quark ↔ gluon jet discr.
- QCD multijet: more gluon jets

QCD rejection BDT:

- Uses Classification Without Labels (CWoLa) 1708.02949
- Trained using data with =7 jets

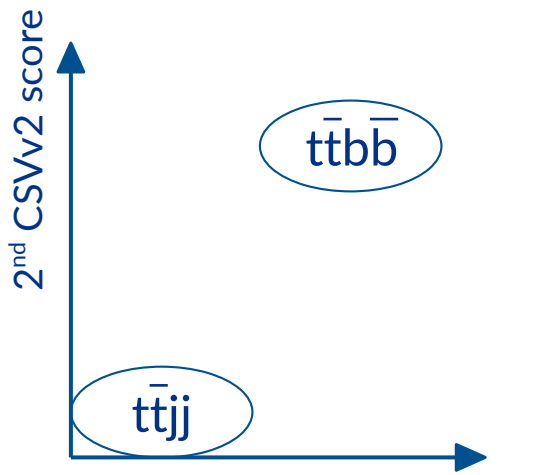




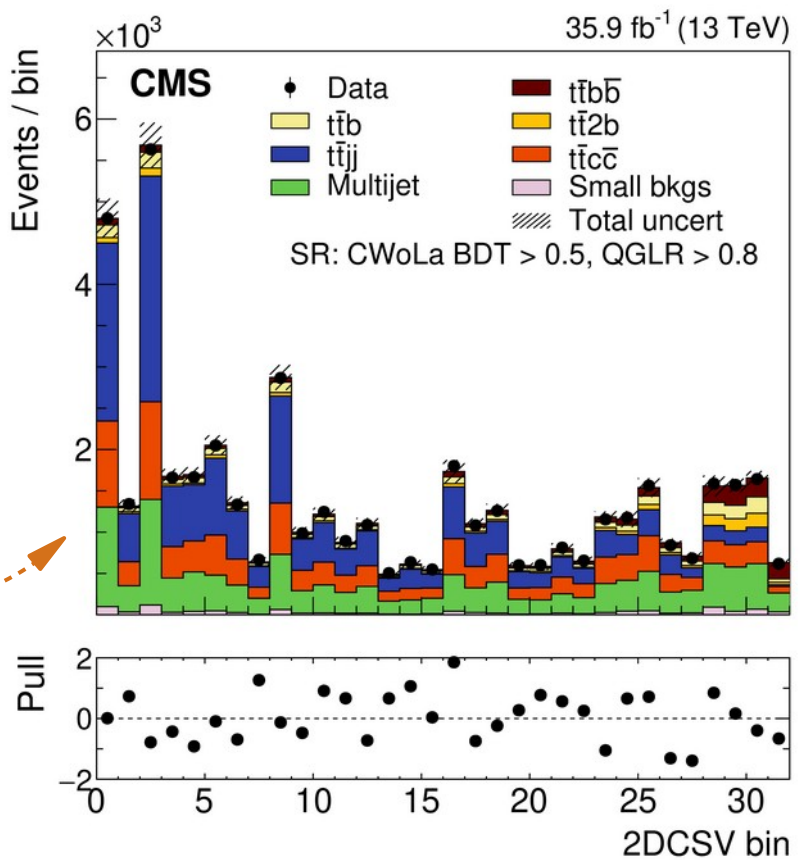
$t\bar{t}b\bar{b}$ cross section: all-jet final state

Phys. Lett. B 803 (2020) 135285

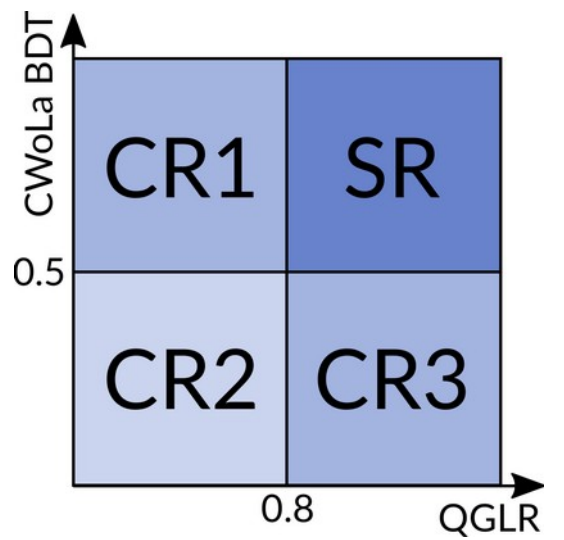
- Extract signal using b tagging (CSVv2) scores of two <u>extra</u> jets with highest score:



Unrolled to 1D



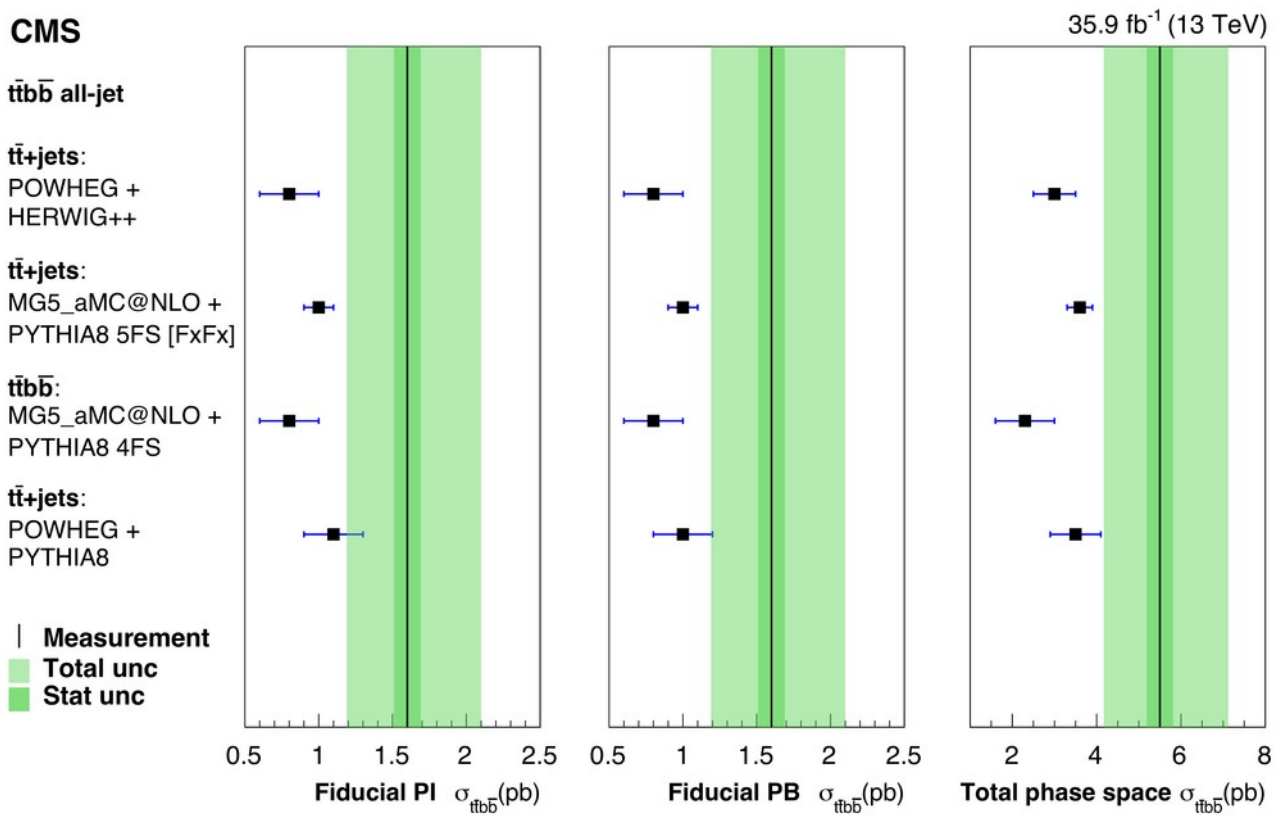
- Estimate QCD contribution: "bin-wise ABCD"
- QGLR and BDT uncorrelated
- For each bin: $N^{SR} = N^{CR3}\, N^{CR1}/N^{CR2}$



$t\bar{t}b\bar{b}$ cross section: all-jet final state

Phys. Lett. B 803 (2020) 135285

- Cross section reported in fiducial (two definitions) and total phase space
- ≈ 30% precision, dominated by systematics
  - Quark-gluon likelihood, b tagging calibrations
  - Signal + background modelling: μ_R/μ_F, PS ISR/FSR scales ($t\bar{t}$ Powheg+P8)
- **Under-prediction** of all generators by factor ~1.5



$t\bar{t}b\bar{b}$ cross section: leptonic final states

JHEP 07 (2020) 125

2016 data, 36 fb^{-1}

Dilepton channel

2 leptons, ≥ 4 jets, ≥ 2 b-tagged jets

Lepton+jets channel

1 lepton, ≥ 6 jets, ≥ 2 b-tagged jets

Identify additional jets → disentangle $t\bar{t}b\bar{b}$ from $t\bar{t}jj$

Take 3rd and 4th largest **CSVv2 scores**

→ 85% (23%) correct for $t\bar{t}jj$ ($t\bar{t}b\bar{b}$)

Kinematic fit:

- Reconstruct $t\bar{t}$ system
- p_T^{miss}; top and W mass constraints

→ take 2 largest **CSVv2 scores** among remaining jets

→ 40% (12%) correct for $t\bar{t}jj$ ($t\bar{t}b\bar{b}$)

+ combinatorics
- statistics

2D template fits → $R(t\bar{t}b\bar{b}/t\bar{t}jj)$, $\sigma(t\bar{t}jj)$

(independently in 2 channels)

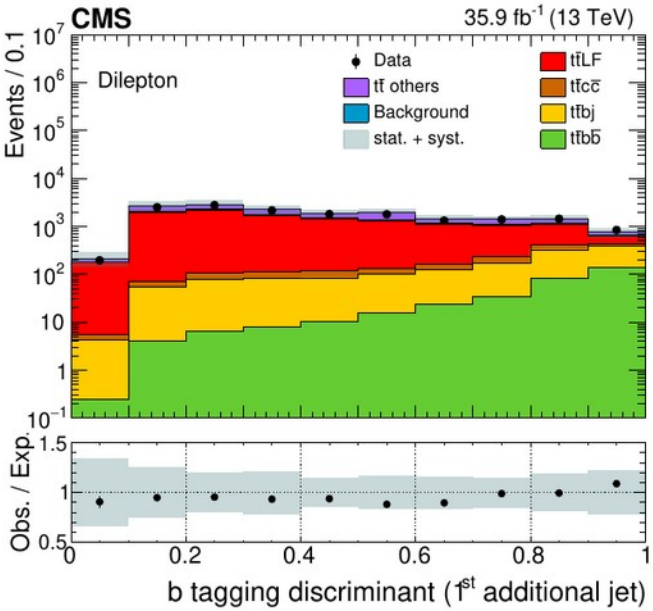
- combinatorics
+ statistics

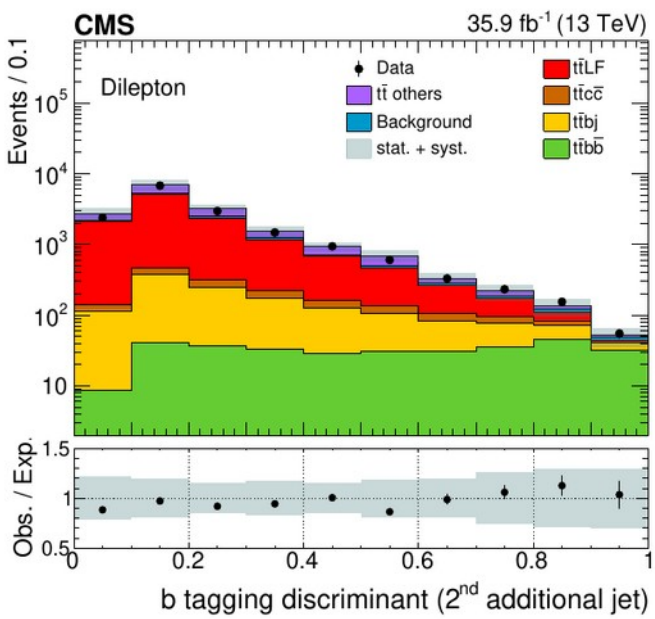
$t\bar{t}b\bar{b}$ cross section: leptonic final states

JHEP 07 (2020) 125

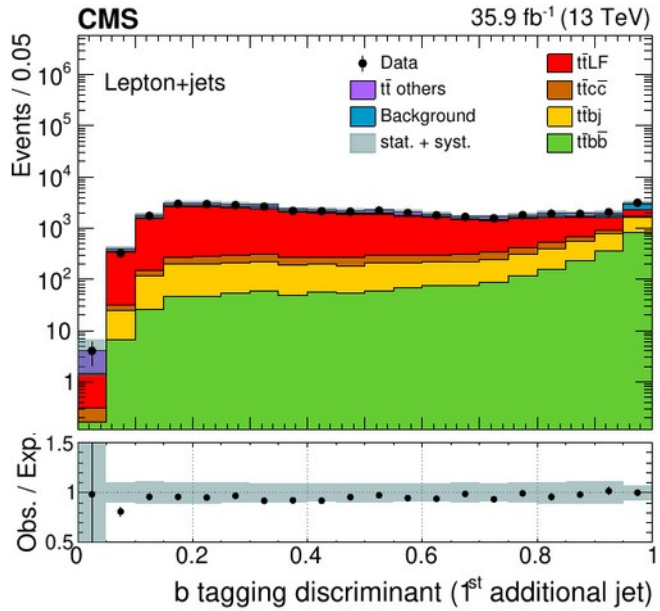
2016 data, 36 fb^{-1}

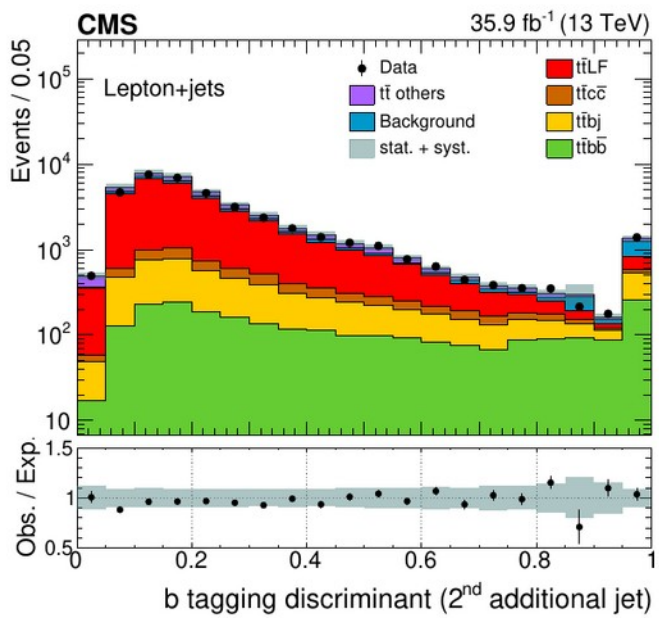
Dilepton





Lepton+jets





Leading CSVv2

vs.

Subleading CSVv2

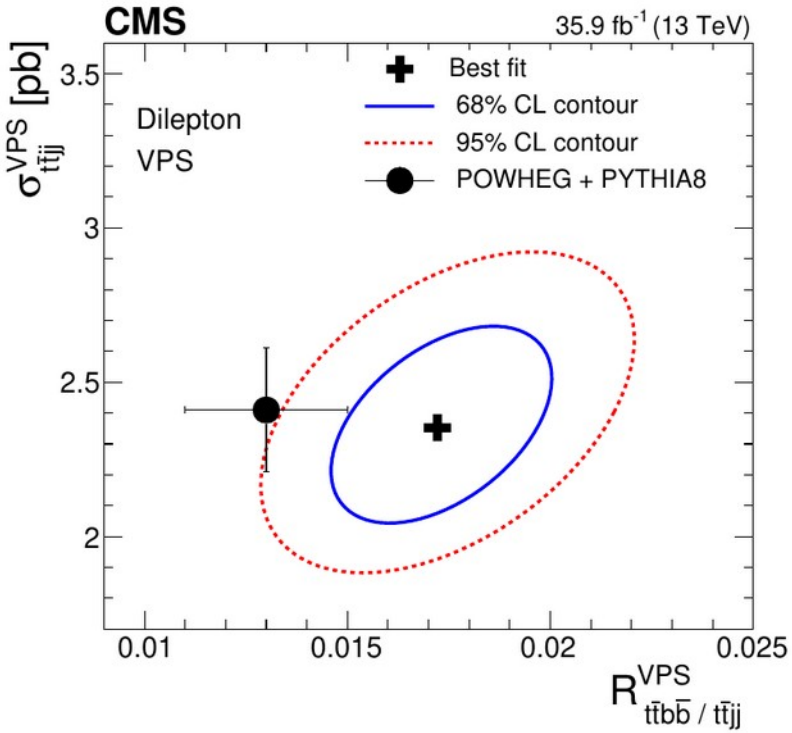
Fit R($t\bar{t}b\bar{b}$/$t\bar{t}jj$), σ($t\bar{t}jj$), using **2D** templates

$t\bar{t}b\bar{b}$ cross section: leptonic final states

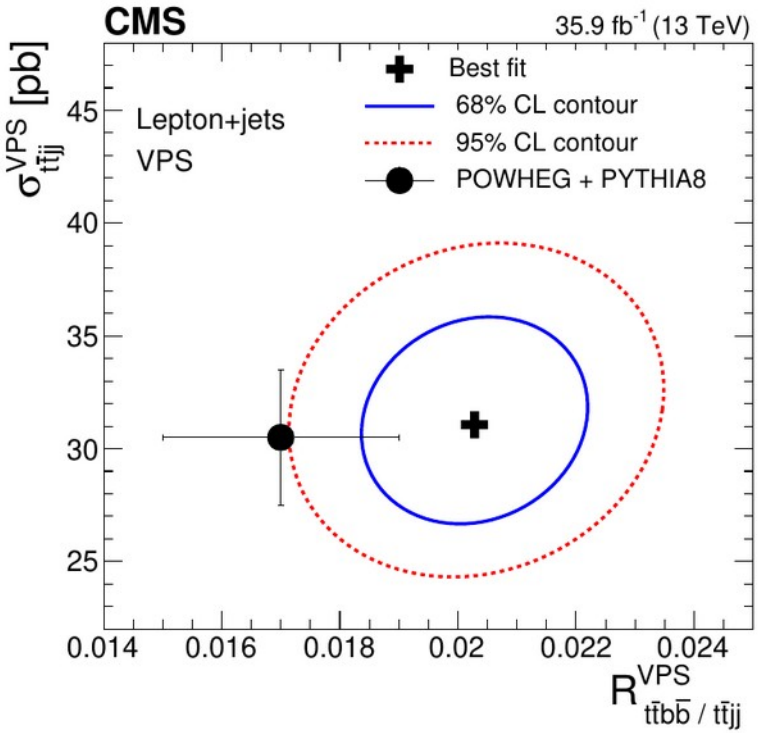
JHEP 07 (2020) 125

2016 data, 36 fb^{-1}

Dilepton



Lepton+jets

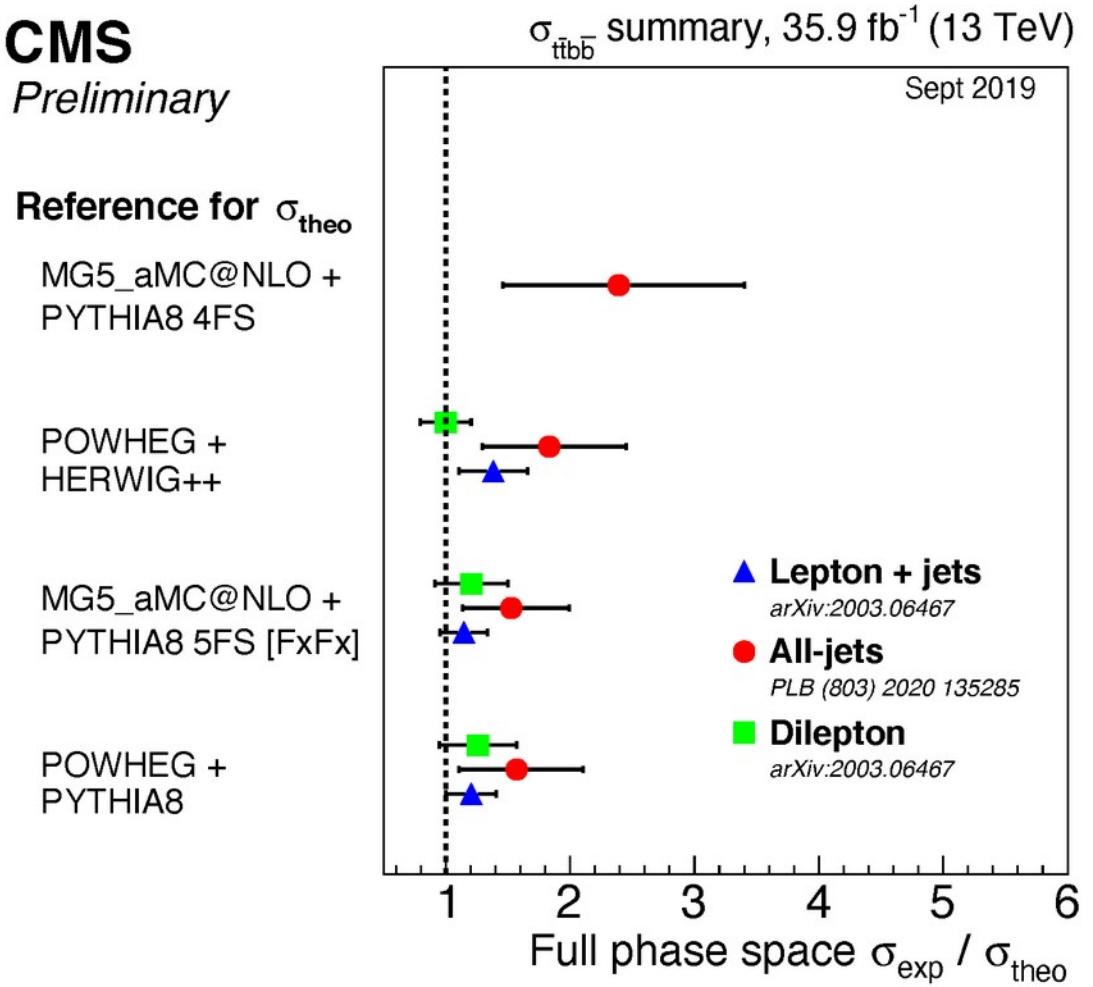


12-13% precision on $\sigma(t\bar{t}b\bar{b})$ (1L/2L)

- Good agreement for $\sigma(t\bar{t}jj)$, slight underprediction for $\sigma(t\bar{t}b\bar{b})$
- Cross sections reported in visible phase space (VPS), and unfolded to full phase space
- Note: different phase-space definitions (30 vs. 20 GeV b jet p_T threshold) in two channels
- Dominant uncertainties: b tagging, $t\bar{t}$ modelling ($t\bar{t}$ Powheg+P8)

$t\bar{t}b\bar{b}$ cross section: CMS summary

- Consistent trend of under-prediction for $\sigma(t\bar{t}b\bar{b})$
- Just within uncertainties (exp + th)

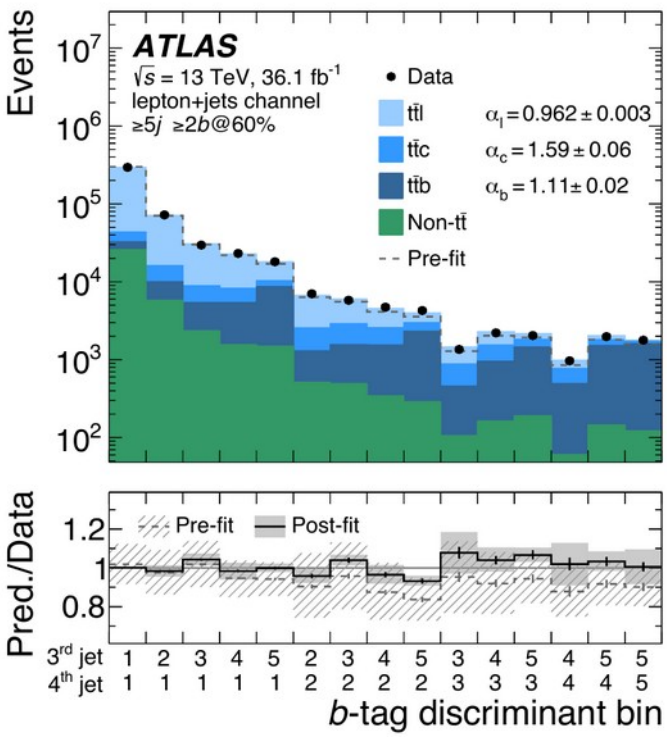


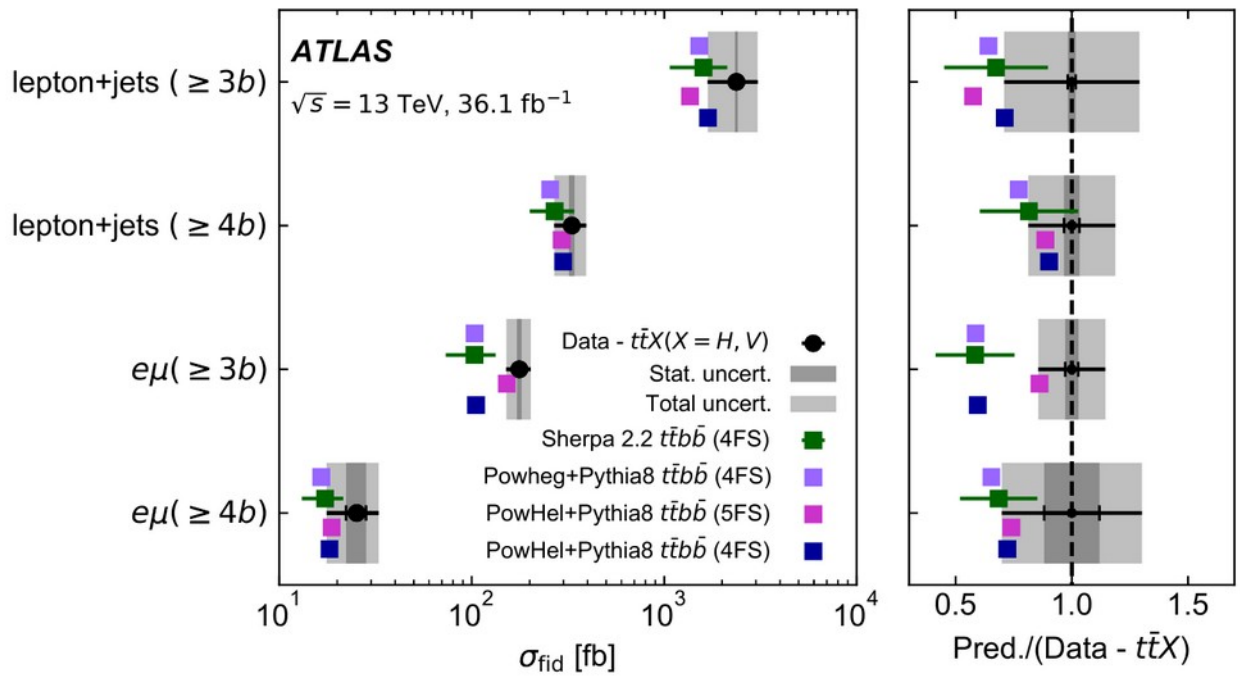
$t\bar{t}b\bar{b}$ cross sections: leptonic final states

JHEP 04 (2019) 046

2015-6 data, 36 fb^{-1}

- l+jets and dilepton (eμ) channels
- Fit normalization of $t\bar{t}$+light, $t\bar{t}$+c, $t\bar{t}$+b using 3rd (vs. 4th) b tagging discriminator values
- Fiducial cross sections in ≥3b and ≥4b phase spaces
  - "Agnostic" wrt extra b origin – subtract $t\bar{t}H(b\bar{b})$, $t\bar{t}Z(b\bar{b})$ from measured values
  - Comparison with several generators → overall under-prediction of $\sigma(t\bar{t}+b(b))$





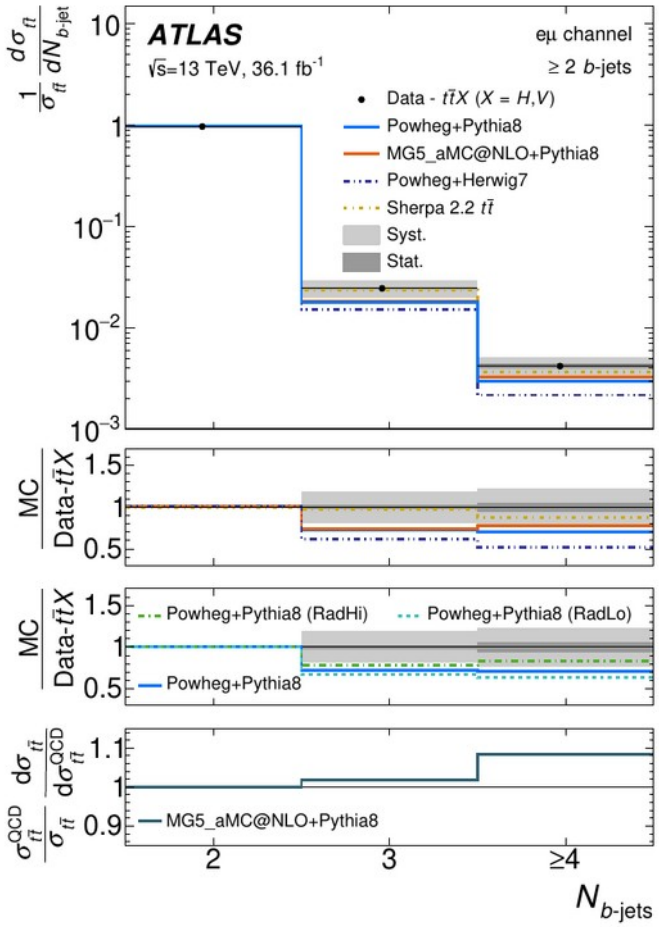
≈ 17% precision on $\sigma(t\bar{t}b\bar{b})$ in l+jets

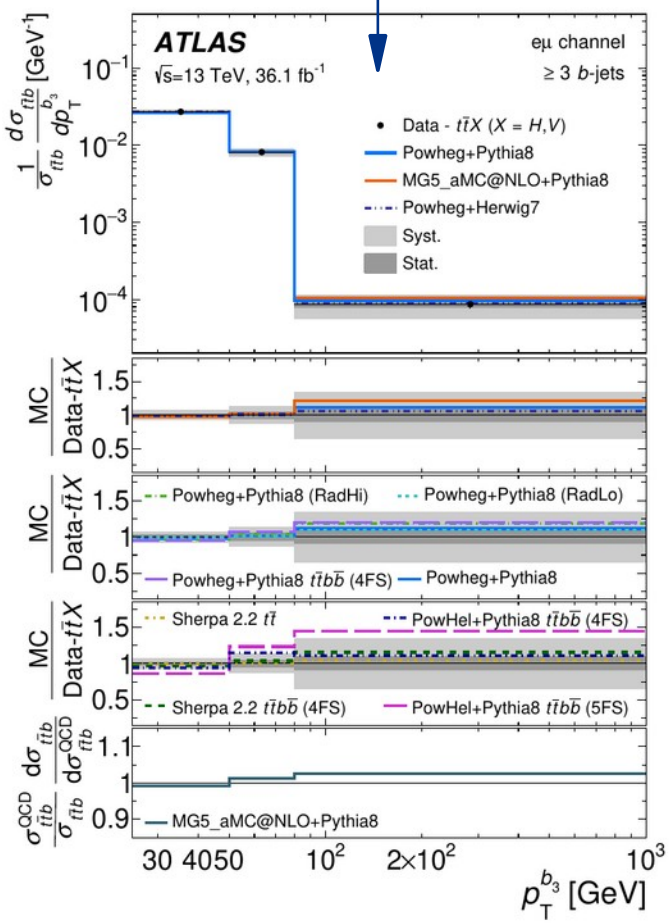
$t\bar{t}b\bar{b}$ differential cross sections

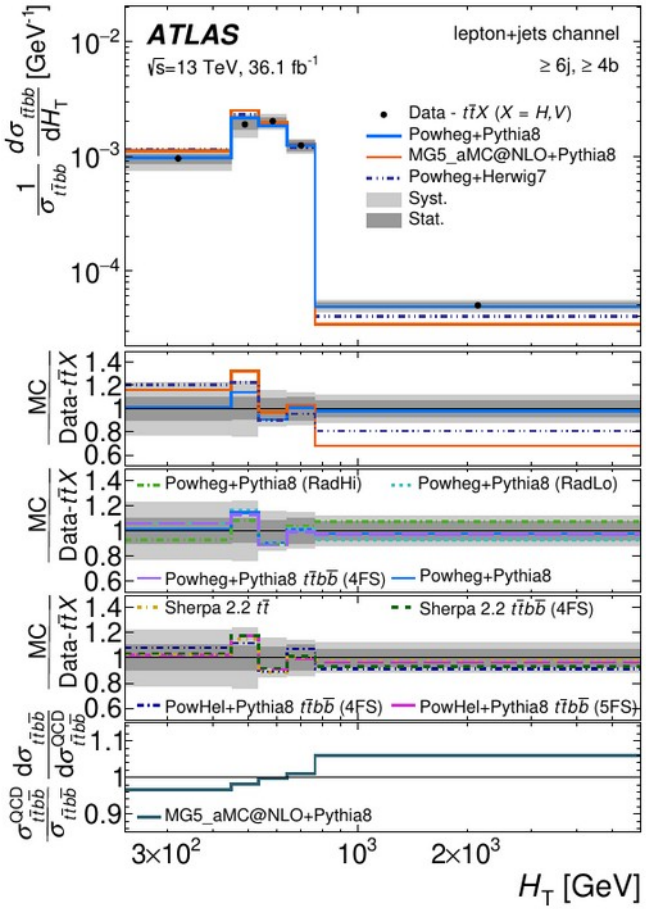
JHEP 04 (2019) 046

2015-6 data, 36 fb^{-1}

- Unfold several normalized differential distributions to particle level
  - Only in ≥3b volume for eμ channel – lower statistics
  - Kinematics of p_T-ordered b jets → subleading jets = proxy for add. b jets





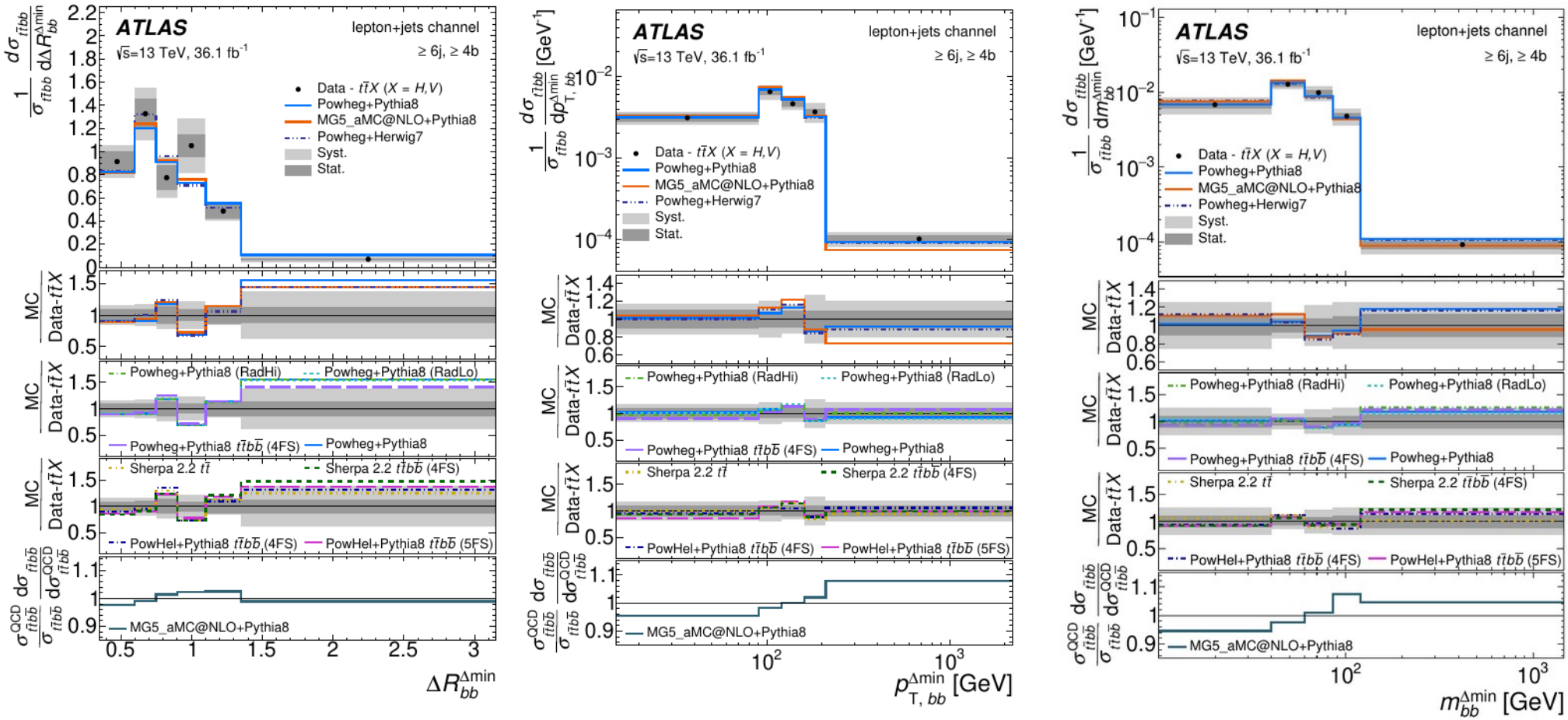


$t\bar{t}b\bar{b}$ differential cross sections

JHEP 04 (2019) 046

2015-6 data, 36 fb^{-1}

- Unfold several normalized differential distributions to particle level
  - $b\bar{b}$ pair kinematics of two closest (in ΔR) b jets → proxy for add. b jets (gluon splitting)

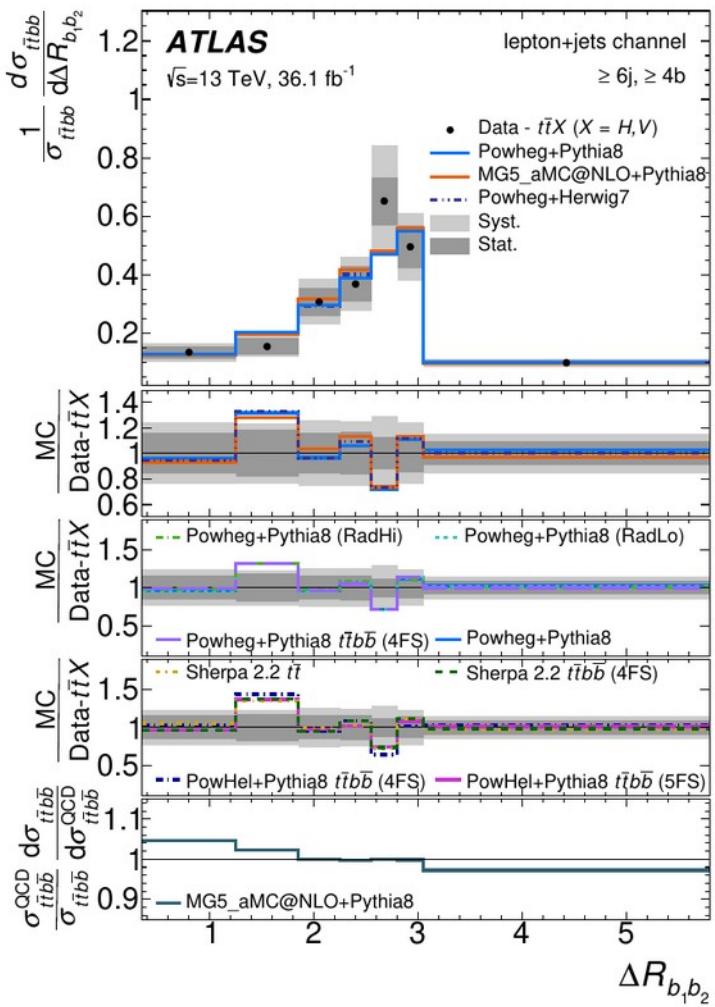


$t\bar{t}b\bar{b}$ differential cross sections

JHEP 04 (2019) 046

2015-6 data, 36 fb^{-1}

- Unfold several normalized differential distributions to particle level
  - $b\bar{b}$ pair kinematics of two leading b jets → proxy for b jets from top decays



- Dominant uncertainties: signal modelling, b tagging, JES
- Comparisons with generators accounting for bin-to-bin correlations
- Overall **good agreement** with generators
  - Some discrepancies with PowHel 5FS $t\bar{t}b\bar{b}$, MG5_aMC@NLO FxFx
- Sizable statistical uncertainties in some phase space regions

- All results available as Rivet routines!

$t\bar{t}c\bar{c}$ cross section: a first!

TOP-20-003

NEW!

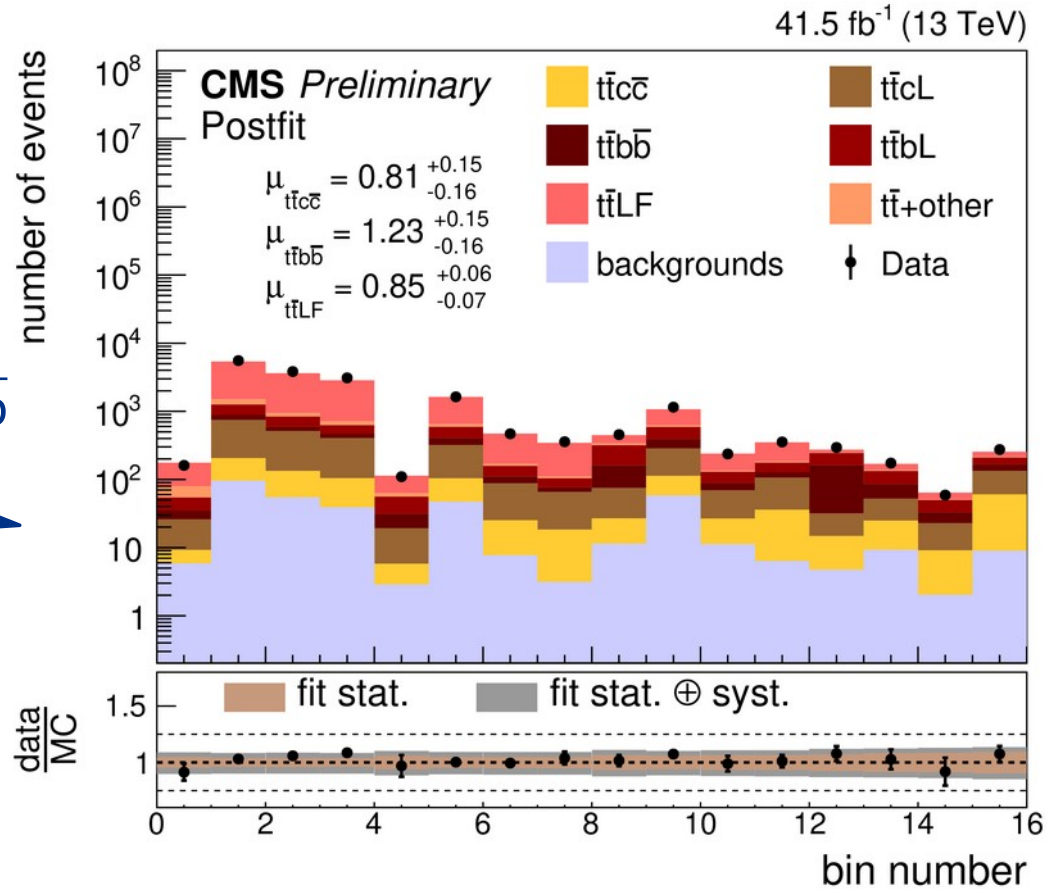
2017 data, 41 fb^{-1}

- First simultaneous measurement of $t\bar{t}$LF, $t\bar{t}c\bar{c}$, $t\bar{t}b\bar{b}$ cross sections
- Rely on DeepCSV: c jet tagger**s**

- Permutation NN for additional jets

- Multiclass NN: disentangle $t\bar{t}$LF, $t\bar{t}c\bar{c}$, $t\bar{t}b\bar{b}$

- Measurement: template fit of $\sigma(t\bar{t}LF)$, $\sigma(t\bar{t}c\bar{c})$, $\sigma(t\bar{t}b\bar{b})$
- Also report in terms of $R_c = \sigma(t\bar{t}c\bar{c})/\sigma(t\bar{t}LF)$, $R_b = \sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}LF)$



$t\bar{t}c\bar{c}$ cross section: a first!

TOP-20-003

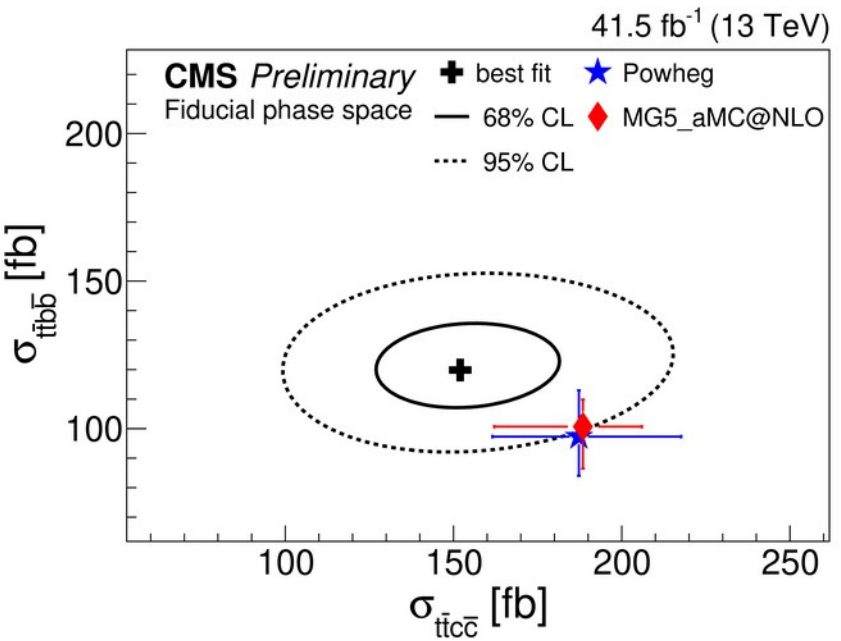
NEW!

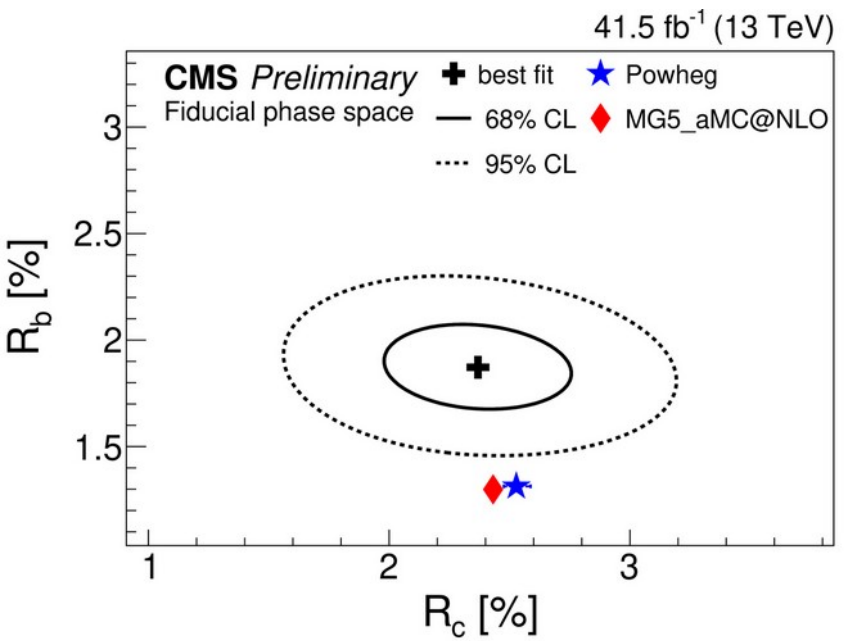
2017 data, 41 fb^{-1}

- Measurement precision: $\sigma(t\bar{t}c\bar{c})$: 15% (stat), 13% (syst)
  R_c: 15% (stat), 11% (syst)

  ≈ 20%
- Both **$\sigma(t\bar{t}c\bar{c})$** and $\sigma(t\bar{t}LF)$ **slightly overpredicted** → good agreement in R_c
- R_b slightly underpredicted (consistent with other results)
- Dominant uncertainties:
  - $t\bar{t}$ modelling
  - c tagging calibration
  - Jet energy scale

Many more details at Seth's talk!





Conclusions

Comprehensive results on $t\bar{t}$ + heavy-flavour jets from ATLAS and CMS

- Under-prediction of $\sigma(t\bar{t}+b(b))$ but reasonable description of differential spectra
- First-ever direct measurement of $t\bar{t}c\bar{c}$!

Iterative feedback loop:

Improvements in modelling

More precise + detailed measurements

Improved precision in $t\bar{t}H(b\bar{b})$ and four tops

+ future differential $t\bar{t}$+HF measurements will profit from

- Increased statistics (full Run 2)
- Increased flavour-tagging performances
- More advanced FT calibration techniques

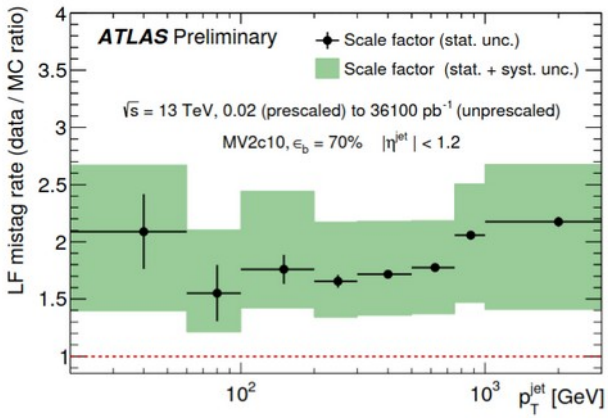
  → Stay tuned!

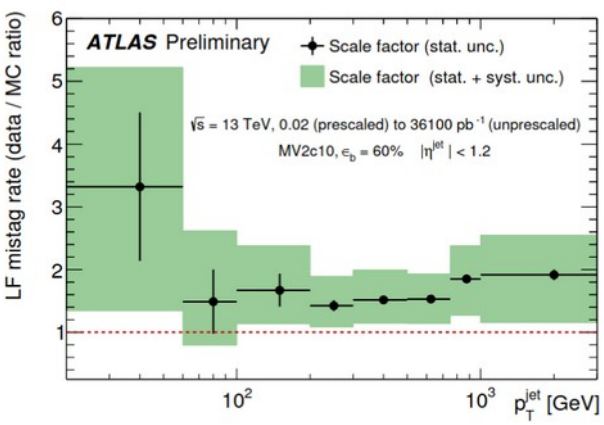
Thank you!

Back-up

Flavour tagging: SFs and uncertainties

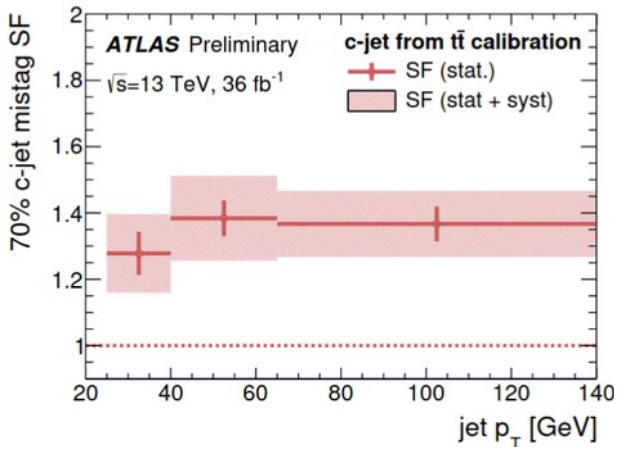
- Light mistag:

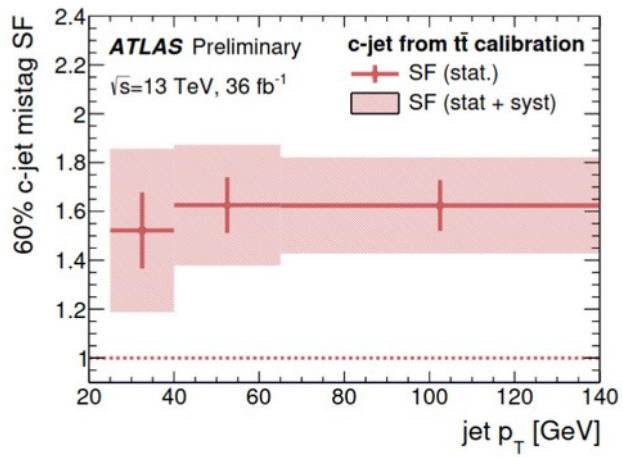




ATLAS-CONF-2018-006

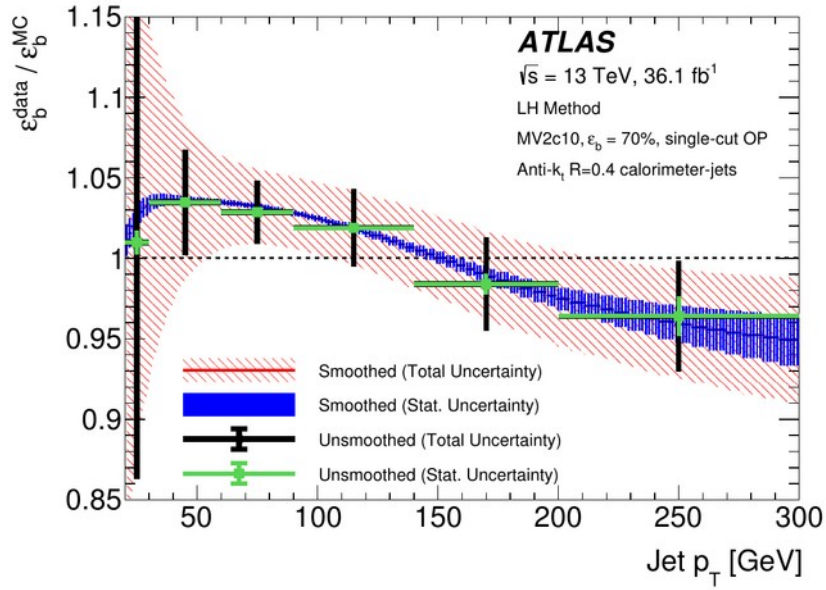
- c mistag:





ATLAS-CONF-2018-001

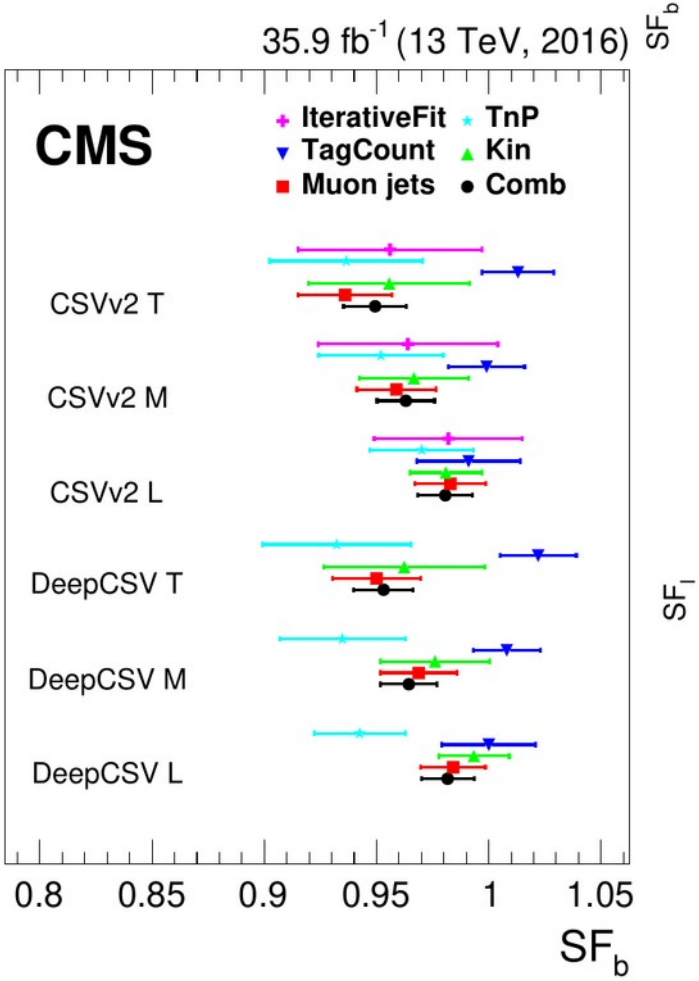
- b tag:



JHEP 08 (2018) 089

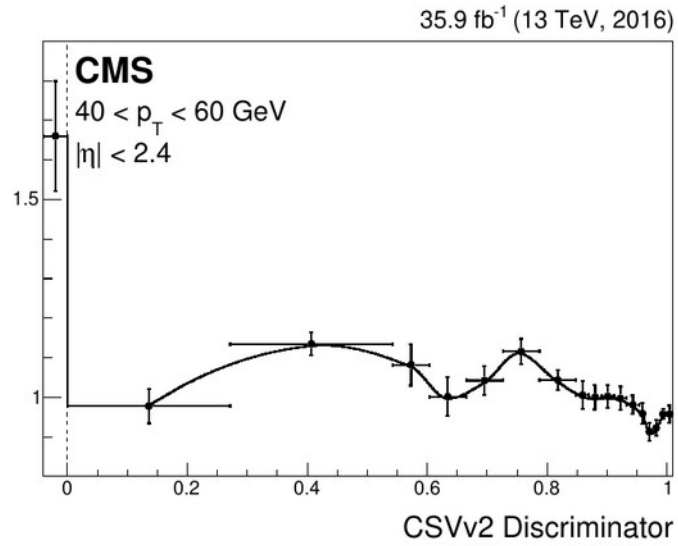
Flavour tagging: SFs and uncertainties

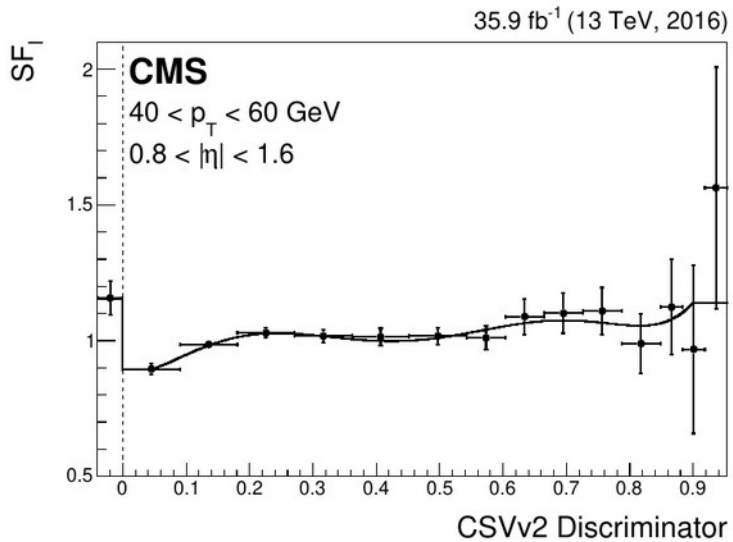
- b tag (fixed working points):



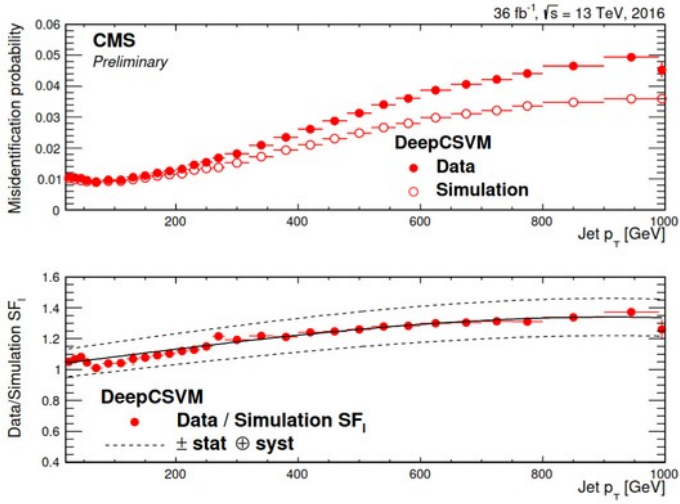
JINST 13 (2018) P05011

- Continous scale factors:





- light mistag (fixed working points):



CMS-DP-2017-012

$t\bar{t}b\bar{b}$ cross section: all-jet final state

Phys. Lett. B 803 (2020) 135285

- Cross section measured for three phase spaces:

Fiducial, Parton Independent (PI):

- ≥ 8 jets with pT > 20 Gev, of which ≥ 6 with pT > 30 GeV
- ≥ 4 b jets

Fiducial, Parton Based (PB):

- ≥ 8 jets with pT > 20 Gev, of which ≥ 6 with pT > 30 GeV
- ≥ 4 b jets, of which ≥ 2 not coming from top decays

Total:

≥ 2 b jets with pT > 20 GeV, not coming from top decays

- Event selection:
  - ≥ 8 jets with pT > 30 Gev, of which ≥ 6 with pT > 40 GeV
  - ≥ 2 b-tagged jets
  - HT > 500 GeV
  - P(Chi2) < 1e-6 $\chi^2 = (M(\mathrm{j}_1,\mathrm{j}_3,\mathrm{j}_4) - m_\mathrm{t})^2/\sigma_\mathrm{t}^2 + (M(\mathrm{j}_3,\mathrm{j}_4) - m_\mathrm{W})^2/\sigma_W^2 + (M(\mathrm{j}_2,\mathrm{j}_5,\mathrm{j}_6) - m_\mathrm{t})^2/\sigma_\mathrm{t}^2 + (M(\mathrm{j}_5,\mathrm{j}_6) - m_\mathrm{W})^2/\sigma_W^2$

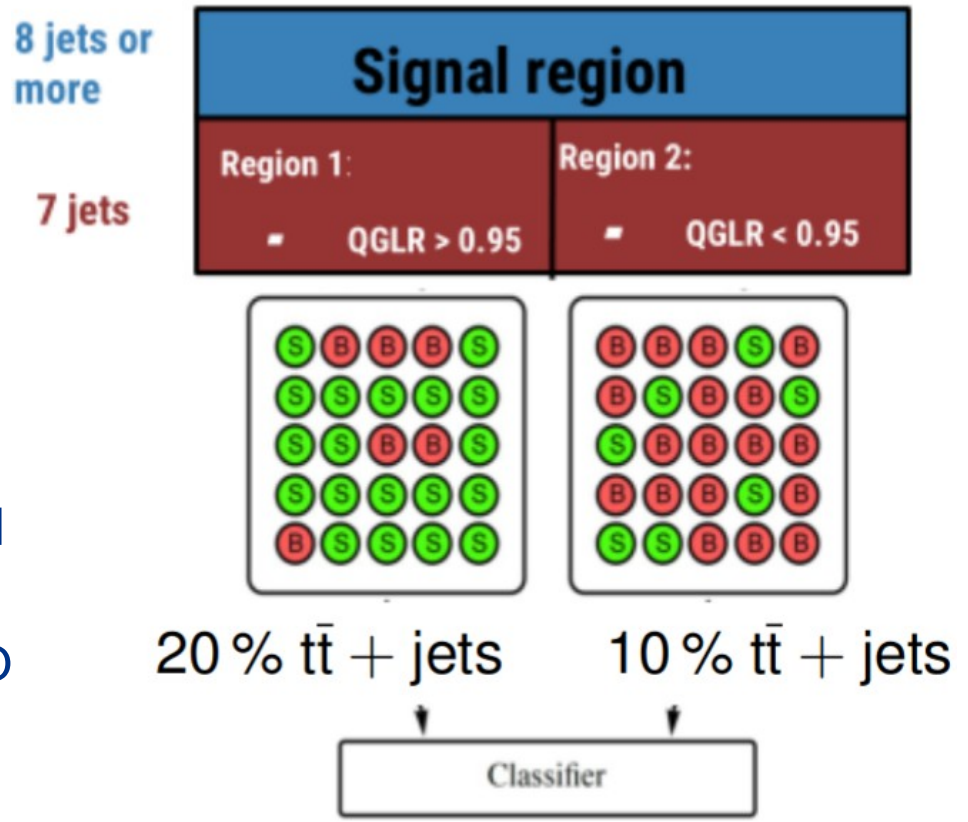
- QGLR: L(4, 0)/(L(4, 0) +L(0, 4)) ↔ on average 4 quark jets in tt events

$$L(N_\mathrm{q}, N_\mathrm{g}) = \sum_{\mathrm{perm}} \left(\prod_{k=i_1}^{i_{N_\mathrm{q}}} \prod_{k=i_{N_\mathrm{q}+1}}^{i_{N_\mathrm{q}+N_\mathrm{g}}} f_\mathrm{q}(\zeta_k) f_\mathrm{g}(\zeta_m) \right)$$

$t\bar{t}b\bar{b}$ cross section: all-jet final state

Phys. Lett. B 803 (2020) 135285

- CWoLa method: 1708.02949
  - 1 signal, 1 background
  - 2 regions with different signal purity
  - Treat 2 regions as “signal” and “background” in training
  - Classifier converges to discriminator for actual signal and background
  - Condition: region definition uncorrelated with input variables
  - Here: signal = tt+jets, background = QCD



$t\bar{t}b\bar{b}$ cross section: all-jet final state

Phys. Lett. B 803 (2020) 135285

Source	Fiducial, parton-independent (%)	Fiducial, parton-based (%)
Simulated sample size	$^{+15}_{-11}$	$^{+15}_{-11}$
Quark-gluon likelihood	$^{+13}_{-8}$	$^{+13}_{-8}$
b tagging of b quark	± 10	± 10
JES and JER	$^{+5.1}_{-5.2}$	$^{+5.0}_{-5.4}$
Integrated luminosity	$^{+2.8}_{-2.2}$	$^{+2.4}_{-2.2}$
Trigger efficiency	$^{+2.6}_{-2.1}$	$^{+2.5}_{-2.2}$
Pileup	$^{+2.3}_{-2.0}$	$^{+2.2}_{-1.9}$
μ_R and μ_F scales	$^{+13}_{-9}$	$^{+13}_{-9}$
Parton shower scale	$^{+11}_{-8}$	$^{+11}_{-8}$
UE tune	$^{+9.0}_{-5.3}$	$^{+9.0}_{-5.2}$
Colour reconnection	± 7.2	± 7.1
Shower matching (h_{damp})	$^{+4.3}_{-2.8}$	$^{+3.8}_{-2.7}$
$t\bar{t}c\bar{c}$ normalization	$^{+3.2}_{-4.4}$	$^{+2.9}_{-4.5}$
Modelling of p_T of top quark	± 2.5	± 2.4
PDFs	$^{+2.2}_{-2.0}$	$^{+2.2}_{-2.0}$
Total	$^{+28}_{-23}$	$^{+28}_{-23}$

$t\bar{t}b\bar{b}$ cross section: leptonic final states

JHEP 07 (2020) 125

2016 data, 36 fb^{-1}

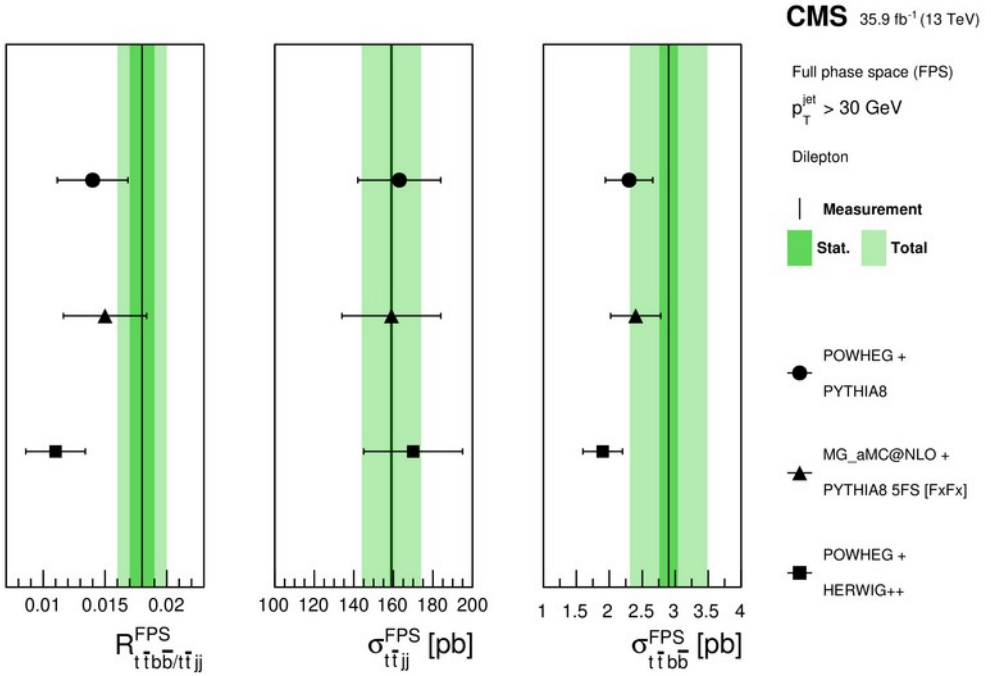
Source	$R^{VPS}_{t\bar{t}b\bar{b}/t\bar{t}jj}$ [%]		$\sigma^{VPS}_{t\bar{t}jj}$ [%]	
	Dilepton	Lepton+jets	Dilepton	Lepton+jets
Lepton uncertainties				
Trigger	<0.1	0.2	1.0	0.5
Lepton identification	0.6	0.2	1.1	1.3
Lepton energy scale	—	<0.1	—	0.1
Jet uncertainties				
Jet energy resolution (JER)	0.4	0.3	0.3	0.7
Jet energy scale (JES)	1.5	1.2	2.9	3.6
b tagging uncertainties				
c-flavor b tag (lin.)	2.2	2.0	1.0	0.3
c-flavor b tag (quad.)	0.7	1.2	0.3	0.2
Heavy-flavor b tag	4.0	0.1	0.5	0.9
Heavy-flavor b tag (lin.)	0.9	0.4	1.5	0.5
Heavy-flavor b tag (quad.)	2.0	0.3	1.5	0.8
Light-flavor b tag	4.9	0.9	5.5	4.9
Light-flavor b tag (lin.)	0.1	0.2	0.3	1.1
Light-flavor b tag (quad.)	0.7	0.7	0.1	1.4
Theoretical uncertainties				
Initial-state radiation (ISR)	1.0	2.2	2.5	1.2
Final-state radiation (FSR)	0.8	0.7	2.5	5.9
ME-PS matching	0.5	<0.1	1.8	1.9
Underlying event tune (UE)	1.5	1.5	0.4	1.4
μ_F/μ_R scales (ME)	0.1	0.4	0.1	1.4
top-p_T	0.2	0.4	1.6	0.3
Ratio $R^{MC}_{t\bar{t}bj/t\bar{t}b\bar{b}}$	1.4	0.2	1.3	0.7
Other uncertainties				
Pileup	0.7	0.2	1.3	0.1
Backgrounds	0.3	2.0	0.7	1.2
Simulated sample size	1.5	2.8	0.1	2.2
Luminosity	0.2	0.5	2.6	3.1
Total systematic	8.0	5.5	8.8	10.0
Statistical	5.8	5.6	0.9	0.6

$t\bar{t}b\bar{b}$ cross section: leptonic final states

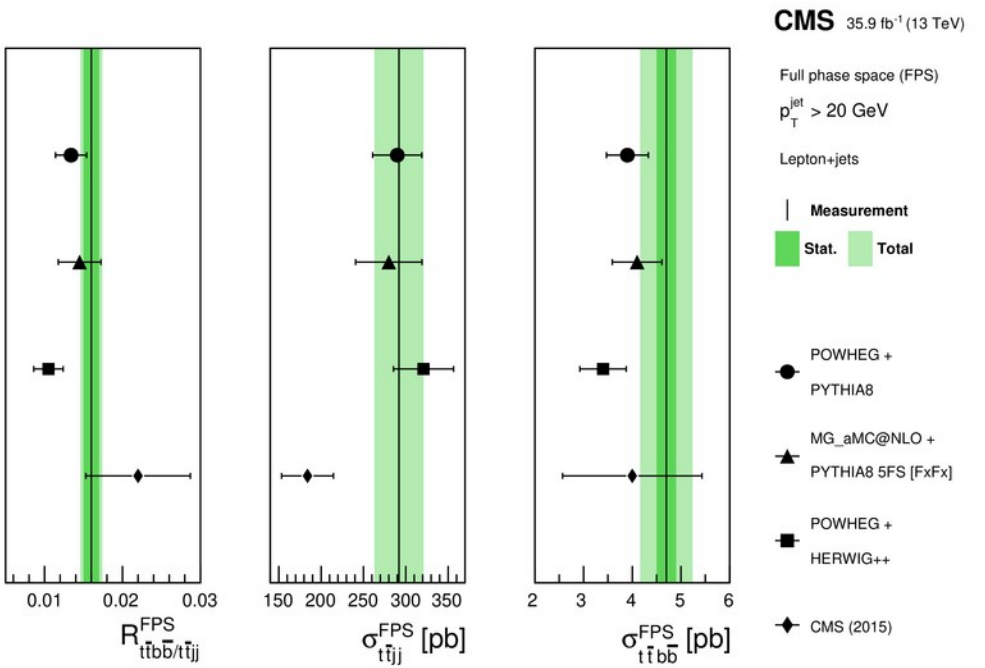
JHEP 07 (2020) 125

2016 data, 36 fb^{-1}

Dilepton



Lepton+jets

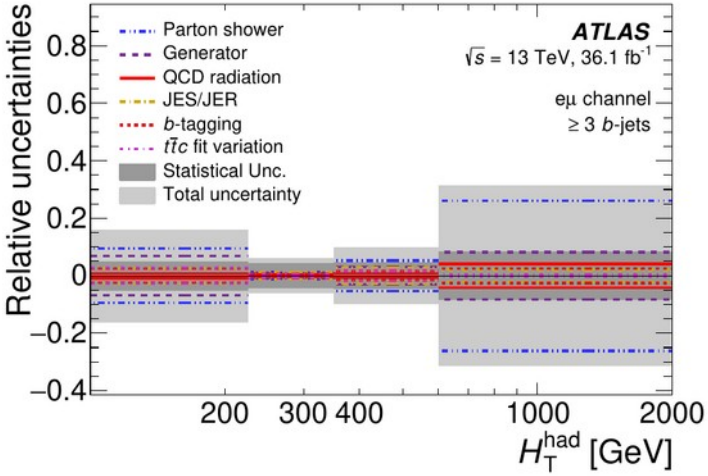


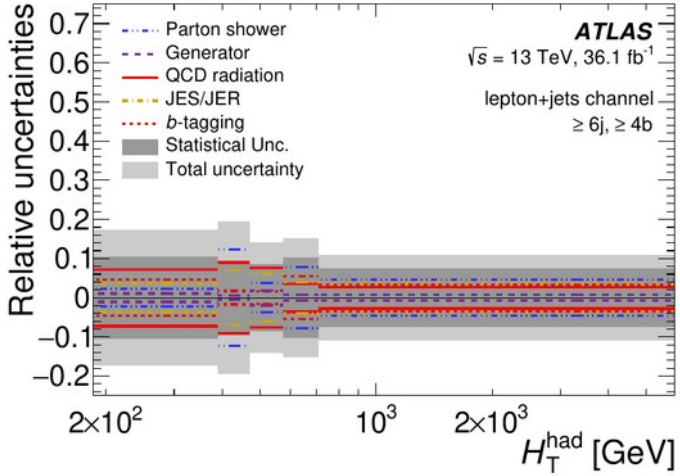
$t\bar{t}b\bar{b}$ differential cross sections

JHEP 04 (2019) 046

- Iterative unfolding, 4 iterations
- Statistical uncertainties: Unfold 10 000 replicas, Poisson-fluctuated data event weights
- Systematic uncertainties:
  - Experimental: in replicas, smear data event weights according to MC uncertainty
  - Modelling: unfold distributions of alternative models using corrections obtained with nominal model, take relative differences between unfolded alternative and true alernative

Source	Fiducial cross-section phase space			
	$e\mu$		lepton + jets	
	$\geq 3b$	$\geq 4b$	$\geq 5j, \geq 3b$	$\geq 6j, \geq 4b$
	unc. [%]	unc. [%]	unc. [%]	unc. [%]
Data statistics	2.7	9.0	1.7	3.0
Luminosity	2.1	2.1	2.3	2.3
Jet	2.6	4.3	3.6	7.2
b-tagging	4.5	5.2	17	8.6
Lepton	0.9	0.8	0.8	0.9
Pile-up	2.1	3.5	1.6	1.3
$t\bar{t}c$ fit variation	5.9	11	—	—
Non-$t\bar{t}$ bkg	0.8	2.0	1.7	1.8
Detector+background total syst.	8.5	14	18	12
Parton shower	9.0	6.5	12	6.3
Generator	0.2	18	16	8.7
ISR/FSR	4.0	3.9	6.2	2.9
PDF	0.6	0.4	0.3	0.1
$t\bar{t}V/t\bar{t}H$	0.7	1.4	2.2	0.3
MC sample statistics	1.8	5.3	1.2	4.3
$t\bar{t}$ modelling total syst.	10	20	21	12
Total syst.	13	24	28	17
Total	13	26	28	17





$t\bar{t}c\bar{c}$ cross section

TOP-20-003

NEW!

2017 data, 41 fb^{-1}

- 2D shape calibration of c-taggers CvsB and CvsL
- Permutation neural network → identify additional jets
  - 2 additional c (b) jets found in 50% (30%) of cases
- Multi-class neural network → disentangle $t\bar{t}$LF, $t\bar{t}c\bar{c}$, $t\bar{t}b\bar{b}$
  - 6 inputs: 2x CvsL+CsvB scores, ΔR of 2 additional jets, permutation NN
  - Outputs P($t\bar{t}$LF), P($t\bar{t}c\bar{c}$), P($t\bar{t}b\bar{b}$)

→ measurement: 2D template fit of $\Delta_b^c = \frac{P(t\bar{t}c\bar{c})}{P(t\bar{t}c\bar{c}) + P(t\bar{t}b\bar{b})}$, $\Delta_L^c = \frac{P(t\bar{t}c\bar{c})}{P(t\bar{t}c\bar{c}) + P(t\bar{t}LF)}$