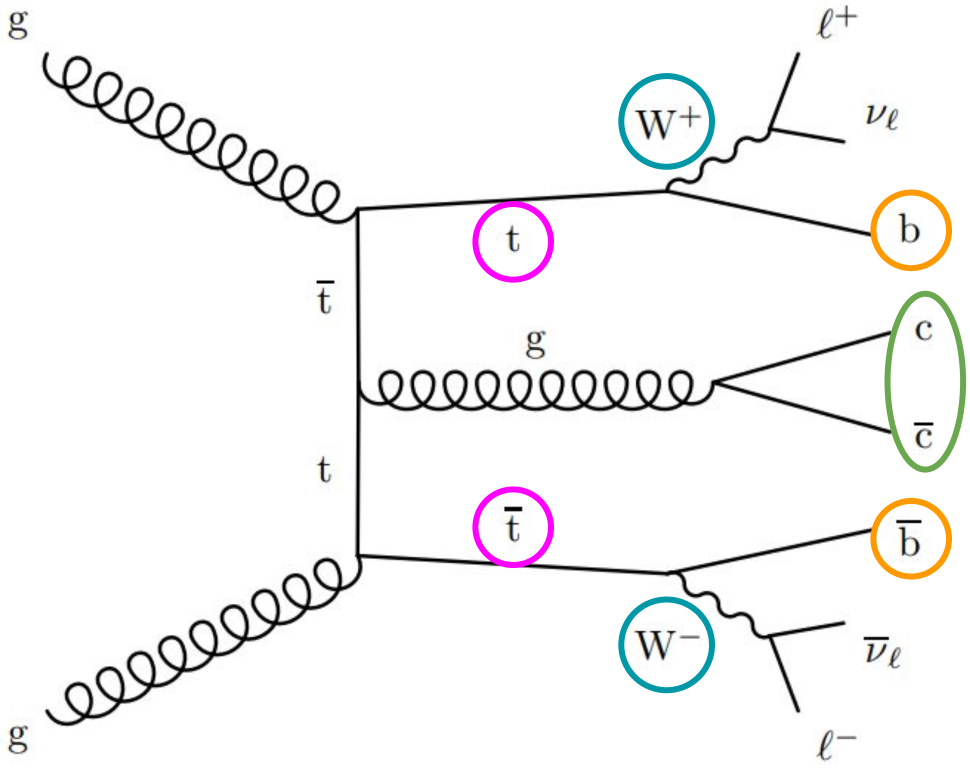
Top-quark pair production with heavy-flavor jets

LHCP

June 3rd-7th 2024, Boston University

Ana Luísa Carvalho on behalf of the ATLAS and CMS Collaborations

Anatomy of tt production with additional heavy-flavor jets



Adapted from PLB 820 (2021)

Top quark

- **2 b-jets** from top quark decay
- **2 W-bosons**

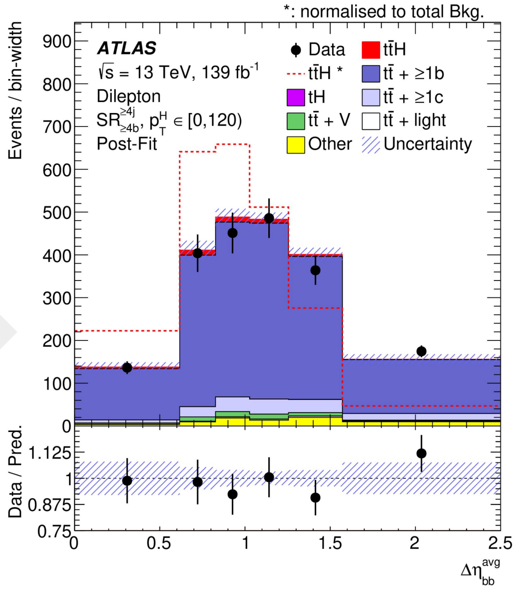
Both decay to leptons =2 charged leptons **Dilepton channel**	One decays to leptons, the other to quarks =1 charged lepton **Single-lepton channel**

Extra-jets

- Mostly from gluon splitting
- Can originate from heavy boson

Important and challenging

- Important test of quantum chromodynamics (QCD) perturbative calculations
- Leading backgrounds in searches for BSM physics and measurements of other SM processes: ttH(H→bb) and 4-top production
  - Irreducible background in ttH(H→bb), extremely hard to isolate



JHEP 06 (2022) 97

The challenges

Theoretical

- Different energy scales associated with the tt and bb/cc pair
- Non-negligible mass of the b-quark

Experimental

- Identification of b- and c-jets comes with inefficiencies and large uncertainties
- Measurements have to at least partially rely on MC simulation

At a glance

- Focusing on measurements performed at $\sqrt{s}$ = 13 TeV
- CMS tt+b-jets measurement in all-hadronic final state (PLB 803 (2020))
  - Very different phase space with ≥8 jets ⇒ Not covered in this talk

Measurement	ATLAS	CMS
tt+b-jets Inclusive and differential cross-sections	JHEP 04 (2019) 046 • $\mathscr{L}$ = 36.1 \fb • Dilepton (eμ) and single-lepton	JHEP 05 (2024) 042 • $\mathscr{L}$ = 138 \fb • Single-lepton
tt+c-jets Inclusive cross-section	Not yet	PLB 820 (2021) • $\mathscr{L}$ = 45.1 \fb • Dilepton

tt + b-jets | Fiducial phase space

- Fiducial phase space regions target different aspects of tt+b-jets production: ttb, ttbb, ttbj, ttbbj

Truth-level objects (ATLAS/CMS)	ATLAS dilepton ($e\mu$)	ATLAS single-lepton	CMS single-lepton
Leptons pT>25/29 GeV (electrons) pT>25/26 GeV (muons)	= 1 electron and = 1 muon with opposite charge	= 1 electron or muon	= 1 electron or muon
Jets and b-jets anti-kT R=0.4 pT>25 GeV	≥ 3 b-jets ≥ 4b-jets	≥ 5 jets, ≥ 3 b-jets ≥ 6 jets, ≥ 4 b-jets	≥ 5 jets, ≥ 3 b-jets ≥ 6 jets, ≥ 4 b-jets ≥ 6 jets, ≥ 3 b-jets, ≥ 3 light-jets ≥ 7 jets, ≥ 4 b-jets, ≥ 3 light-jets

tt + b-jets | Background estimation

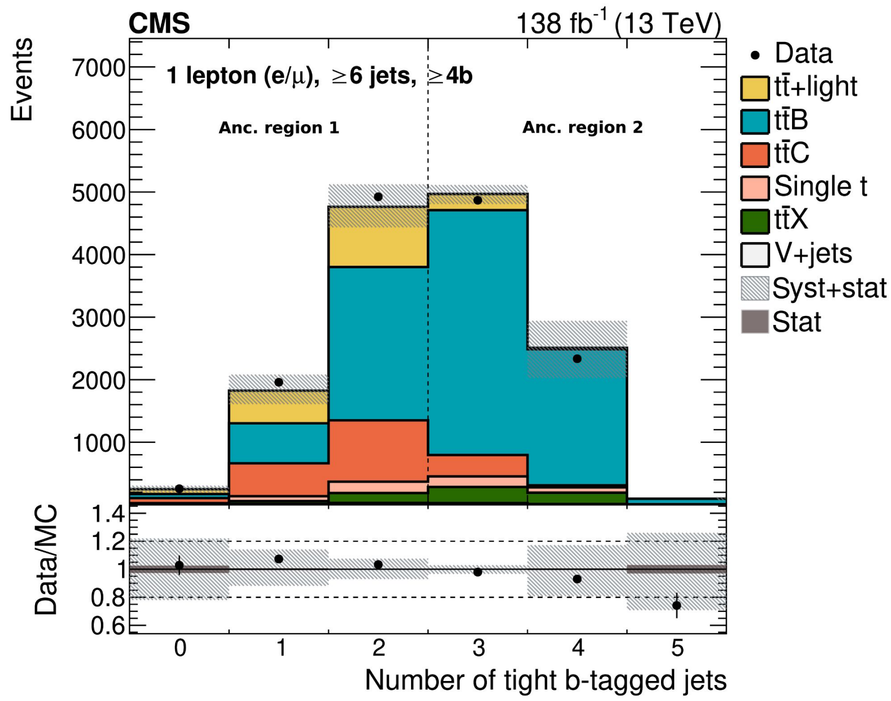
- **tt+jets MC simulation:** massive b-quarks in matrix element calculation but not in PDFs (4 flavor-scheme) or massless b-quarks from parton shower but included in PDFs (5 flavor-scheme)?

Process	ATLAS	CMS
tt+b-jets	ttbar 5FS Powheg+Pythia8 $\mu_F=\mu_R=m_{T,t}$	ttbb 4FS Powheg+OpenLoops+Pythia8 $\mu_F=¼\ HT, \mu_R=½\ \Pi_i\ m_{Ti}^{¼}$
tt+c-jets **tt+light**	ttbar 5FS Powheg+Pythia8 $\mu_F=\mu_R=m_{T,t}$	

tt + b-jets | CMS signal extraction strategy

- Likelihood fit to data for each variable in signal (>3 b-jets) and control regions (≤2 b-jets)
  - Particle-level unfolding ⇒ Inclusive fiducial cross-section extracted directly from the fit
  - Systematic uncertainties included as nuisance parameters in the likelihood
- Dedicated variables for extra-jets ⇒ DNN is trained to identify the most likely jet-parton assignments

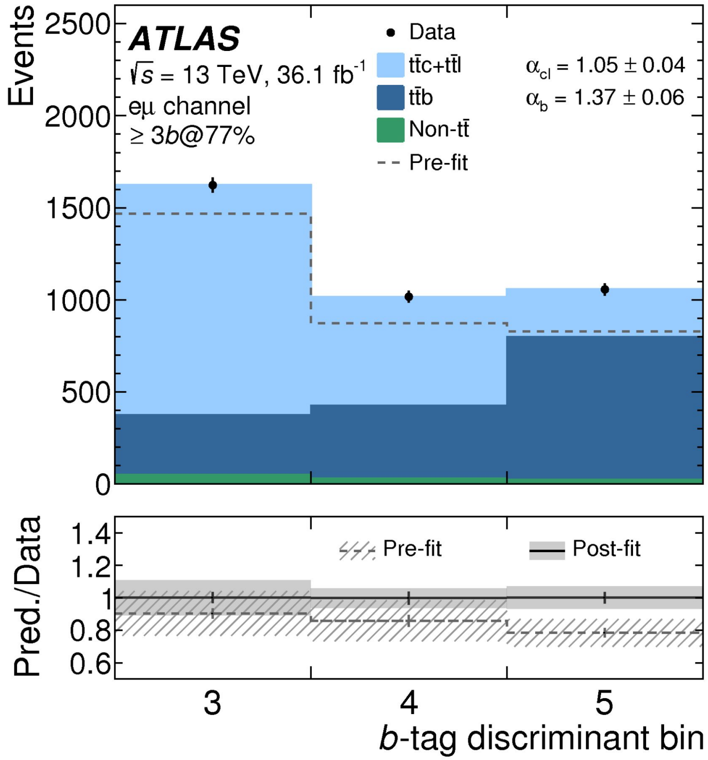
- **Dominant systematic uncertainties:** b-tagging, μ_R, μ_F, parton shower scales (ISR, FSR)



tt + b-jets | ATLAS signal extraction strategy

- Likelihood fit to data in regions with ≥3 b-jets (dilepton) or ≥5 jets, ≥ 2b-jets (single-lepton)
  - Extract normalization factors for tt+≥1b, tt+≥1c and tt+light
- In each fiducial region, non-ttbar background subtracted from data at detector level, in each bin
- Particle-level unfolding to obtain fiducial cross-sections
  - Data-driven correction factors for flavour composition included

- **Dominant systematic uncertainties:** b-tagging, ttc fit variation (dilepton), parton shower (Powheg+H7), generator (varied in the unfolding)



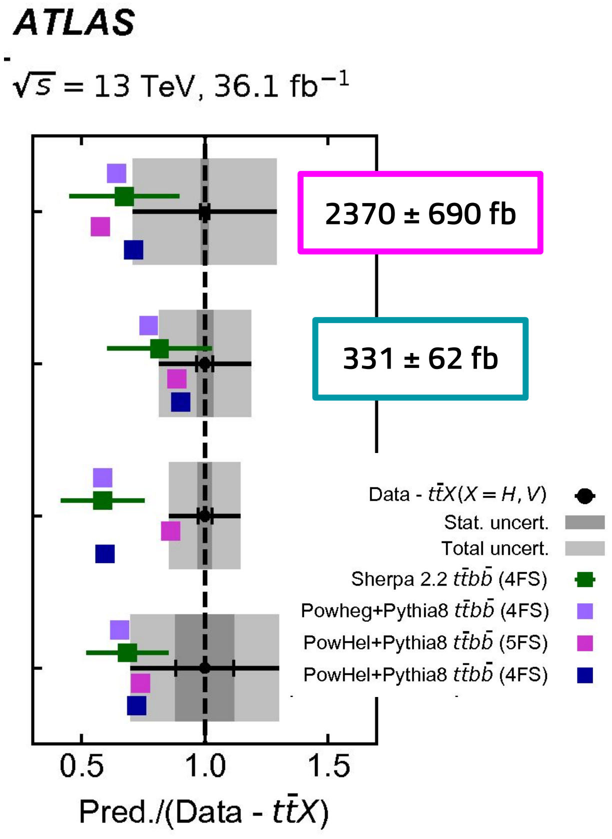
tt + b-jets | Inclusive cross-section results

Single-lepton ≥5j ≥3b

Single-lepton ≥6j ≥4b

Dilepton ≥3b

Dilepton ≥4b

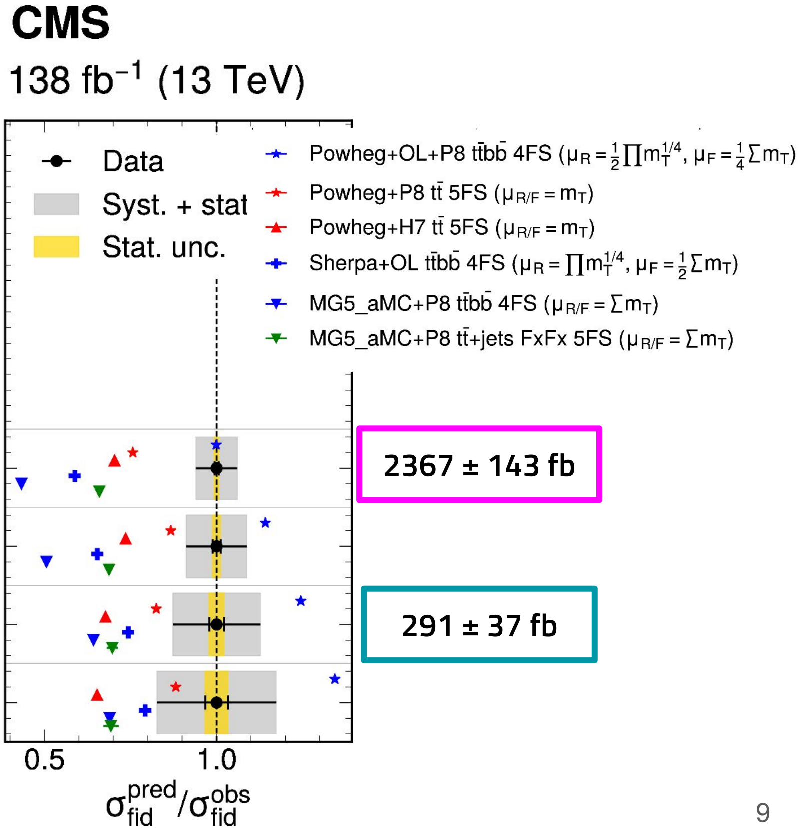


≥5j ≥3b

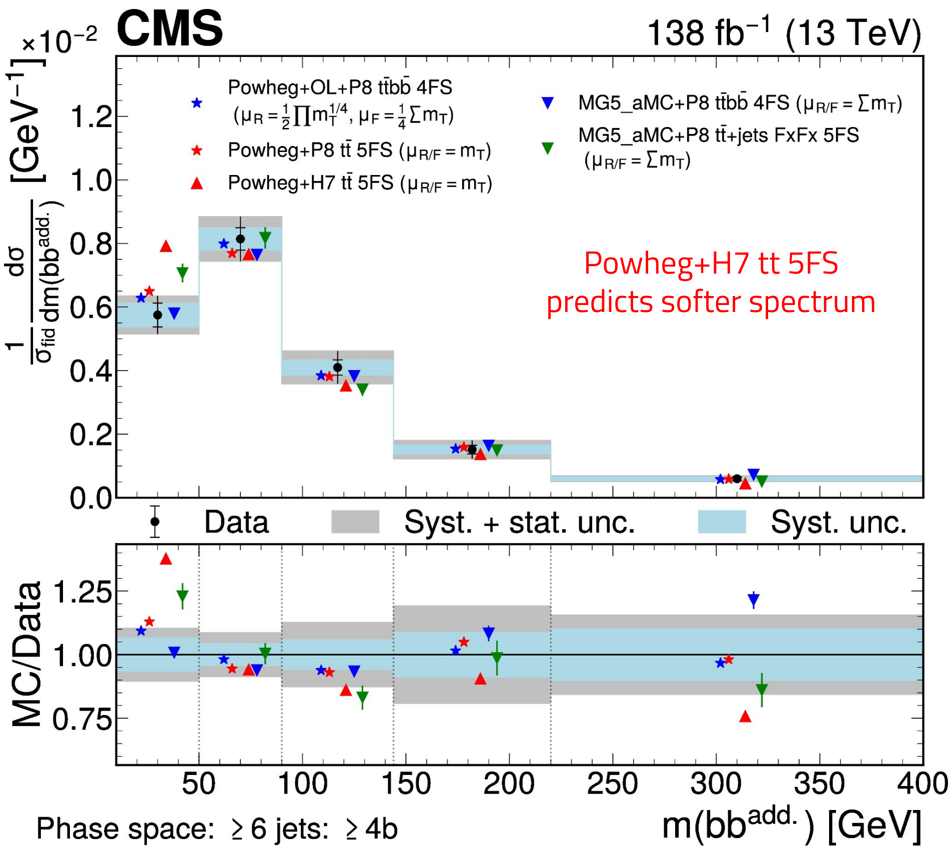
≥6j ≥3b ≥3l

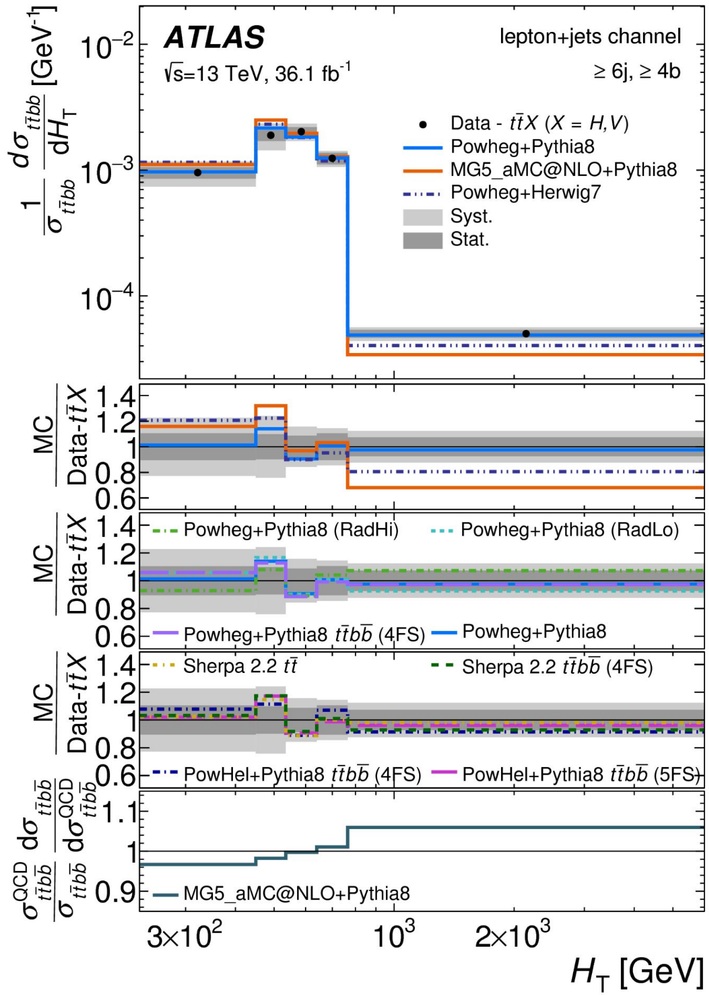
≥6j ≥4b

≥7j ≥4b ≥3l



tt + b-jets | Differential cross-section results





tt + c-jets | Strategy

- Event selection: =2 leptons with opposite charge, ≥4 jets, b-jets from top decay, MET>30 GeV (ee or $\mu\mu$)
- Signal and background estimated from MC simulation: ttbar inclusive Powheg Pythia8 5FS @NLO in QCD
  - Cross-section: $\sigma_{t\bar{t}} = 832^{+39.9}_{-45.8}$ pb (Phys. Lett. B 710 612)

Input

Multivariate method	Purpose	Output
Multi-class neural network*	Identify b- and c-jets	P(b), P(bb), P(c), P(light)
Neural network	Assign jets to partons	Correct permutation
Multi-class neural network	Separate signal from backgrounds	P(ttcc), P(ttcL), P(ttbb), P(ttbL), P(ttLL)

Input

*Calibrated to data in control regions

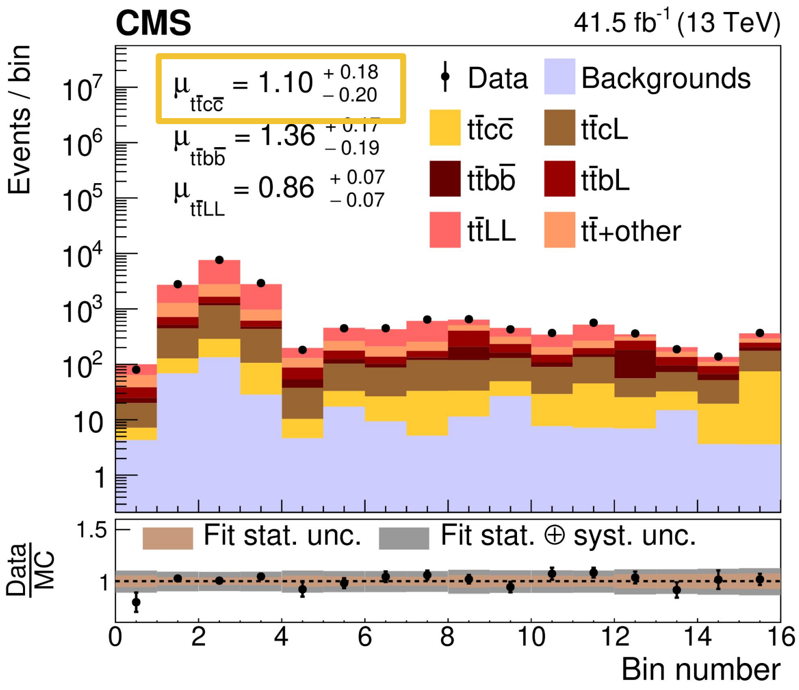
- Discriminants used in the fit to data $\Delta^{c}_{b} = \frac{P(t\bar{t}c\bar{c})}{P(t\bar{t}c\bar{c}) + P(t\bar{t}b\bar{b})}$ $\Delta^{c}_{L} = \frac{P(t\bar{t}c\bar{c})}{P(t\bar{t}c\bar{c}) + P(t\bar{t}LL)}$

tt + c-jets | Signal extraction strategy and results

- Fit performed in fiducial and generator phase spaces
  - Fiducial: =2 leptons with opposite charge from W decays, ≥2 b-jets from top, ≥2 extra jets
  - Generator: ≥2 extra jets

σ(ttcc) [pb]	Measurement	PP8
Fiducial	0.207±0.025±0.027	0.187±0.038
Generator	10.1±1.2±1.4	9.1±1.8

- **Dominant systematic uncertainties:** c-tagging calibration and ME-PS matching



One-dimensional representation of the two-dimensional discriminants

Closing words

- Top quark pair production with additional heavy flavor jets is a dominant background in SM analysis and searches for new physics
    - Striking example is ttH(H→bb)
- Very different energy scales associated with the top- and b/c-jet pair ⇒ **Hard to model**
- Identifying (tagging) heavy flavor jets has associated inefficiencies and uncertainties ⇒ **Hard to measure**
- Despite improvements in the simulation of ttbb the agreement between data and MC is still not optimal
    - **Large impact from different renormalization and factorization scales**
- First measurements of tt+c-jets appearing now ⇒ Need for development of theoretical predictions

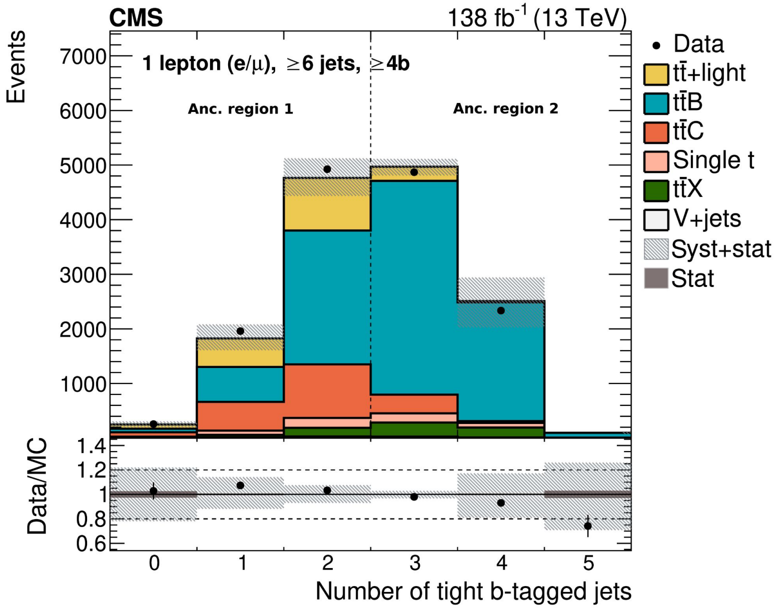
- Only one of the results shown today uses the full Run 2 dataset
    - **Final word on the tt+b-jets and tt+c-jets cross-sections still to come**

Thank you for your attention

Backup

tt + b-jets | CMS signal extraction strategy

- Likelihood fit to data for each variable in signal (>3 b-jets) and control regions (≤2 b-jets)
  - Response matrices included in the definition of likelihood ⇒ **Inclusive fiducial cross-section extracted directly from the fit**
  - Systematic uncertainties included as nuisance parameters in the likelihood
- Dedicated variables for extra-jets ⇒ DNN is trained to identify the most likely jet-parton assignments
- **Dominant systematic uncertainties:** b-tagging, μ_R, μ_F, parton shower scales

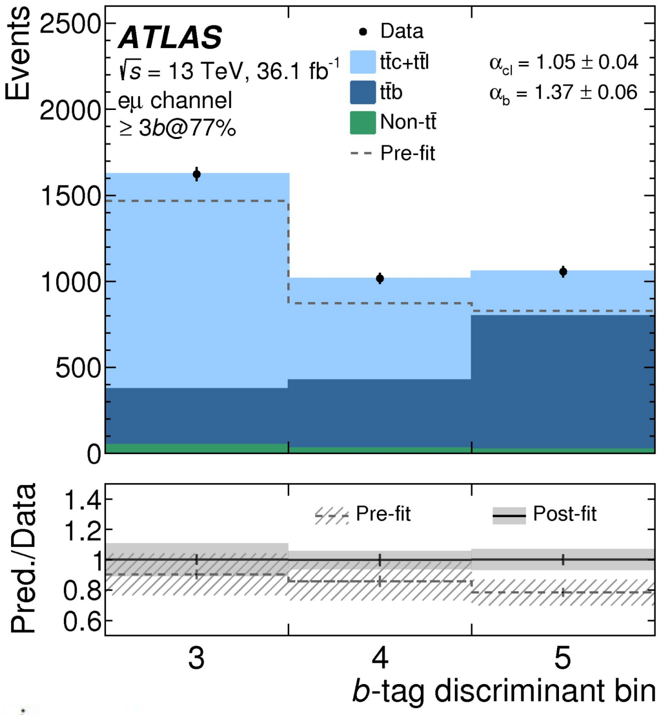


$$S_{e,i}(\vec{\mu},\vec{\alpha}) = \mu_{\text{fid}} \sum_{j=1}^{n} \mu_j M^{e}_{ij}(\vec{\alpha})$$

Encodes probability that a simulated event in **generator-level bin j** is reconstructed and selected in **detector-level bin i**

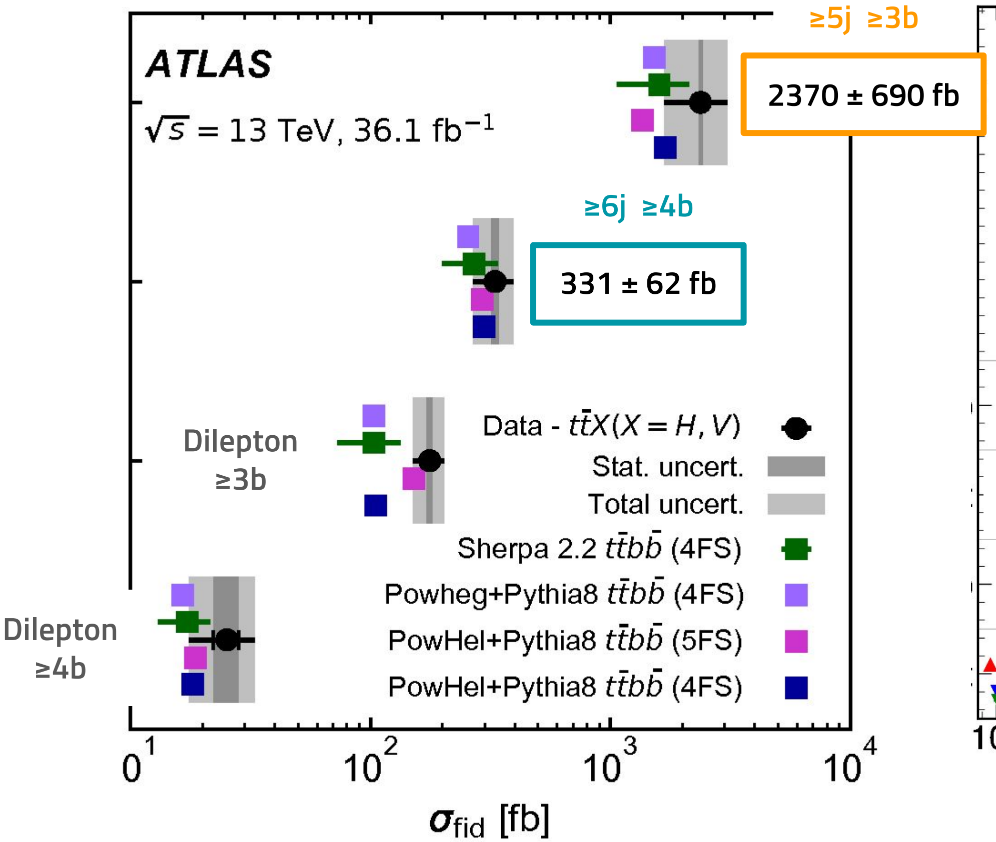
tt + b-jets | ATLAS signal extraction strategy

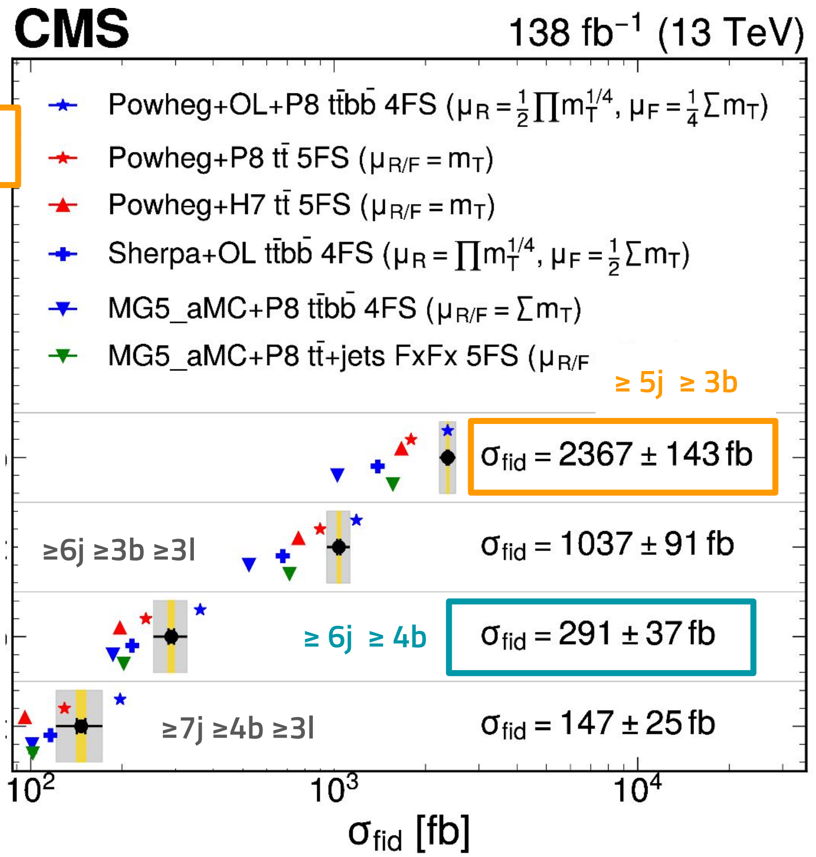
- Likelihood fit to data in regions with ≥3 b-jets (dilepton) or ≥5 jets, ≥ 2b-jets (single-lepton)
  - Normalization factors for tt+≥1b, tt+≥1c and tt+light
- In each fiducial region, non-ttbar background subtracted from data at detector level, in each bin
- Corrections applied to data in fiducial regions
  - Mis-tagged events
  - Fiducial acceptance correction
  - Detector and particle level matching correction
  - Reconstruction efficiency
- **Dominant systematic uncertainties:** b-tagging, ttc fit variation (dilepton), parton shower, generator



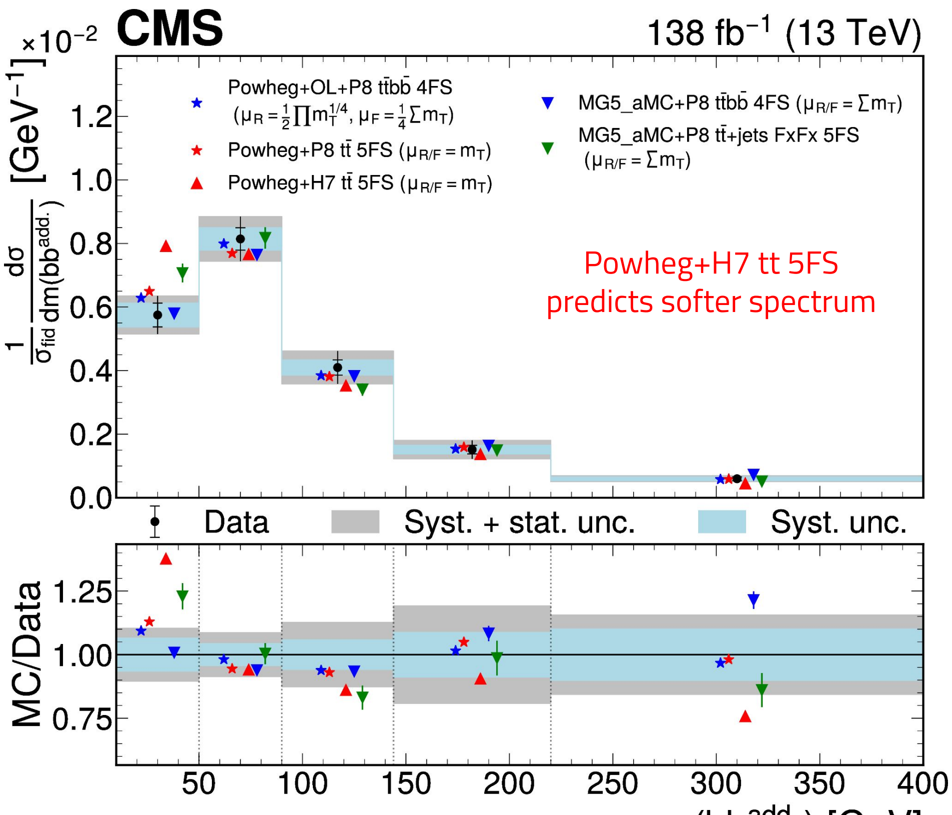
$$\frac{d\sigma^{\text{fid}}}{dX^i} = \frac{N^i_{\text{unfold}}}{\mathcal{L}\,\Delta X^i} = \frac{1}{\mathcal{L}\,\Delta X^i\, f^i_{\text{eff}}} \sum_j \mathcal{M}^{-1}_{ij}\, f^j_{\text{matching}}\, f^j_{\text{accept}}\, \boxed{f^j_{t\bar{t}b}}\,(N^j_{\text{data}} - N^j_{\text{non-}t\bar{t}\text{-bkg}})$$

tt + b-jets | Inclusive cross-section results

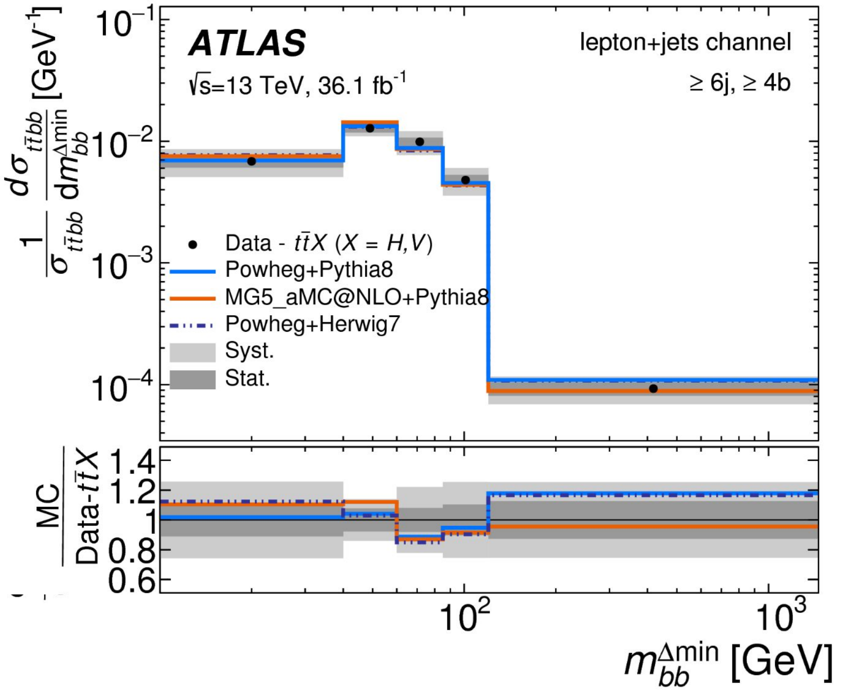




tt + b-jets | Differential cross-section results

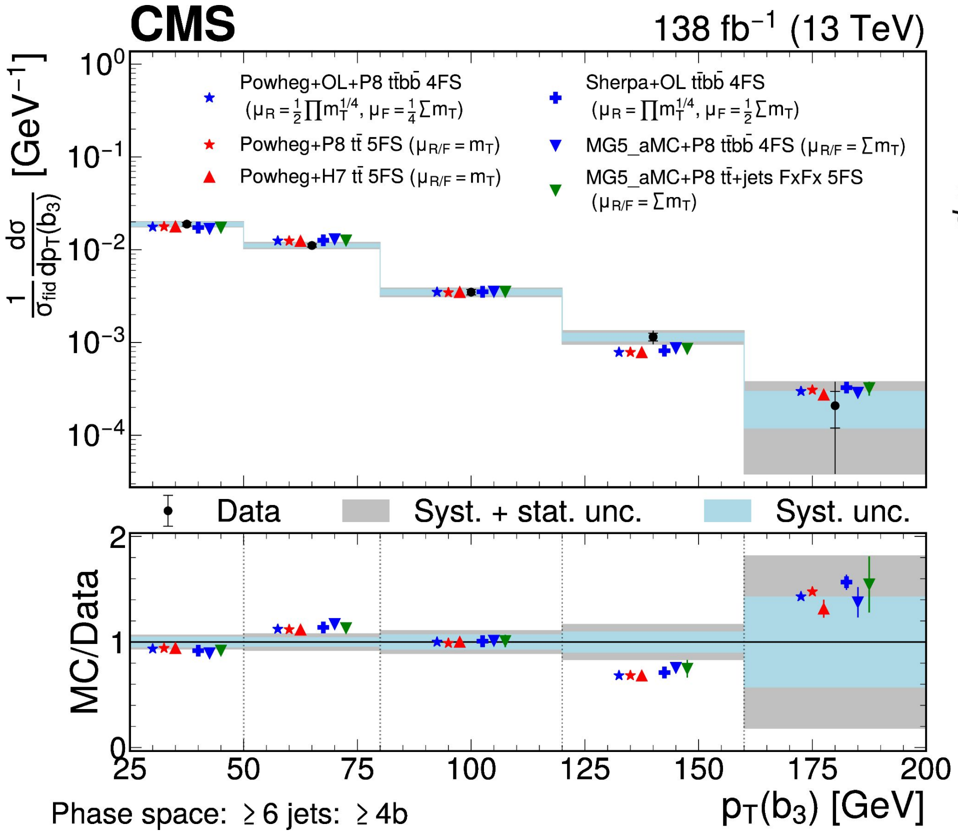


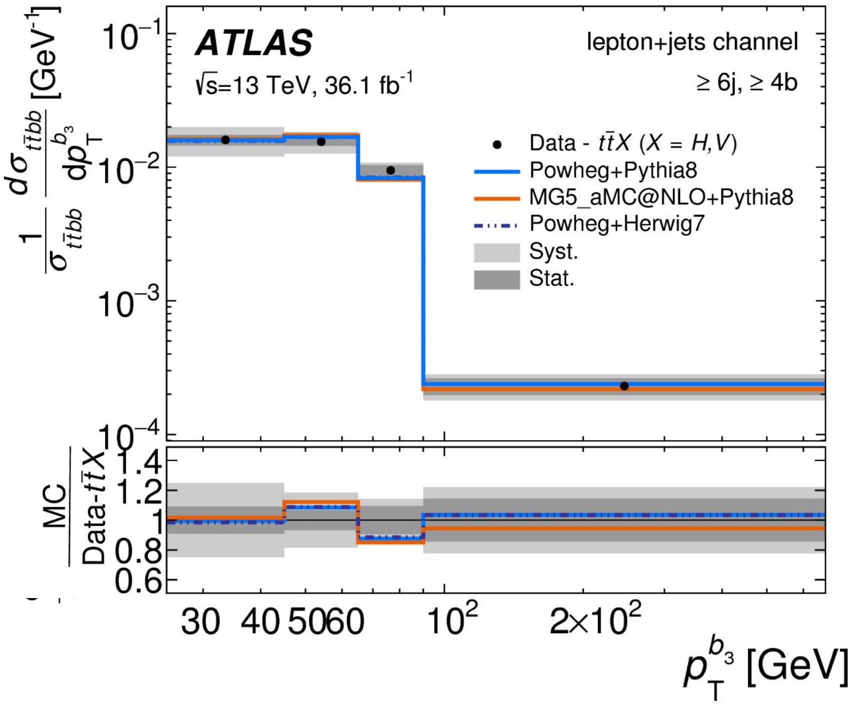
Phase space: ≥ 6 jets: ≥ 4b



- Pair of b-jets with smaller angular distance ⇒ Gluon splitting

tt + b-jets | Differential cross-section results





- 3rd leading b-jet most likely from gluon splitting