15th Annual Workshop
Physics at the Terascale

First measurement of the tt production cross section at 13.6 TeV

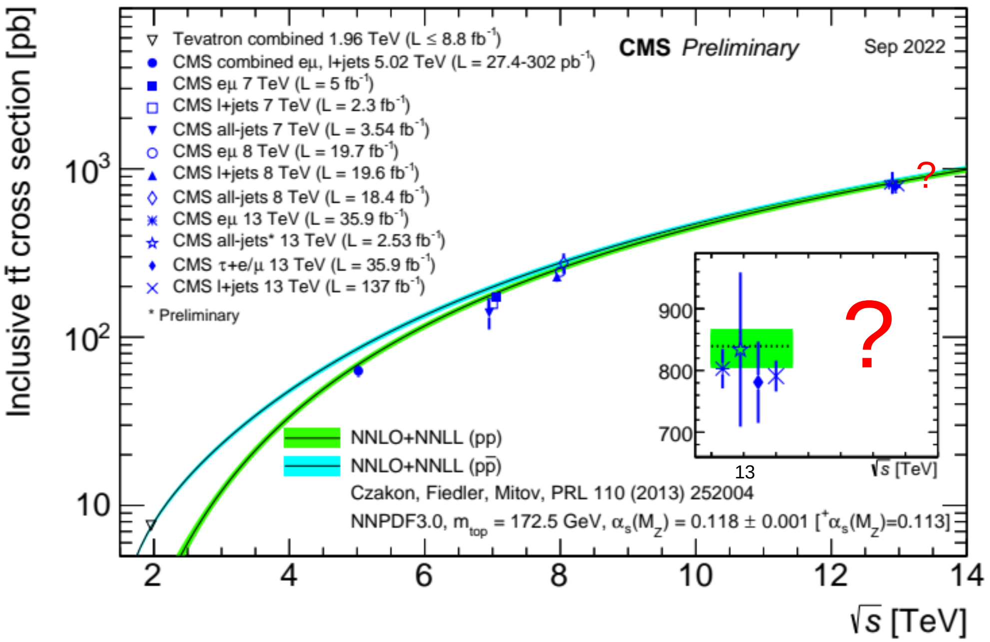
Laurids Jeppe on behalf of the CMS collaboration

29.11.2022 | CMS-PAS-TOP-22-012

Introduction

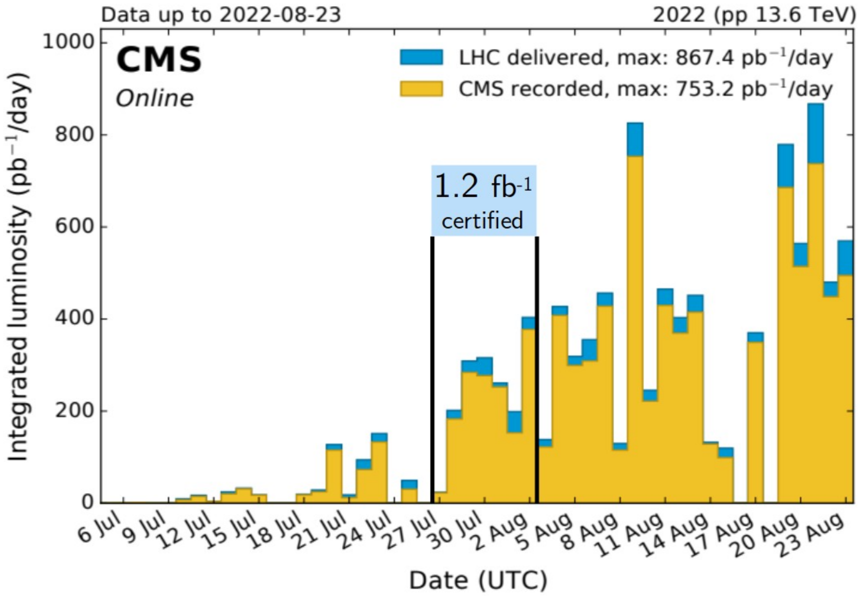
- Run 3 is ongoing!
  - ~ 39 fb^{-1} already delivered by LHC
  - ~ 36 fb^{-1} recorded by CMS
- New center-of-mass energy of 13.6 TeV

- σ_{tt} expected to rise to **921 pb** (from 834 pb in Run 2)
- Early opportunity to…
  - Explore physics at the new energy frontier
  - Check CMS performance in Run 3!



Dataset

- 1.20 fb^{-1} +/- 6% of certified data
  - Collected from July 27th to August 3rd
  - Luminosity value: from emittance scans, cross-checked via Z boson counting



Analysis strategy

- New technique designed for early data
  - ... but can also be adapted for future high-precision measurements

- Profile likelihood fit to constrain exp. uncertainties *in situ* where possible
  - Channel combination constrains lepton ID & b tag efficiencies

    → split into five channels: ee, μμ, eμ, e+jets, μ+jets
  - Advantage for early data: no need to rely on some time-consuming general-purpose calibration

Lepton selection

- Overall offline scale factors (SF) depend on lepton kinematics, but these variables are not needed for a simple cross section measurement → efficiencies enter in acceptance only
- Synchronize selection cuts between dilepton, lepton+jets channels
- In this case, dependence on lepton kinematics integrates out, and offfline efficiencies ϵ_μ and ϵ_e factorize:

- ee channel yield	$\sim \epsilon_e^2$	– e+jets channel yield	$\sim \epsilon_e$
- eμ channel yield	$\sim \epsilon_e \epsilon_\mu$	– μ+jets channel yield	$\sim \epsilon_\mu$
- μμ channel yield	$\sim \epsilon_\mu^2$		

- Channel combination distinguishes the effect of lepton ID efficiencies from σ_{tt}
- Lepton scale factors can be estimated in situ in the fit - no need for general-purpose efficiency studies

Object selection

- **Leptons:** $p_T > 35$ GeV, $|\eta| < 2.4$
  - tight cut-based ID, ported from Run 2 (70% signal efficiency)
- **Jets**: AK4 jets, $p_T > 30$ GeV, $|\eta| < 2.4$
  - b-tagging: DeepJet algorithm
- $m_{ll} > 20$ GeV
- No use of MET, no kinematic reconstruction

dilepton

- 2 leptons
- opposite sign
- At least 1 jet

ee, μμ only:

- At least 1 b-jet
- cut Z window

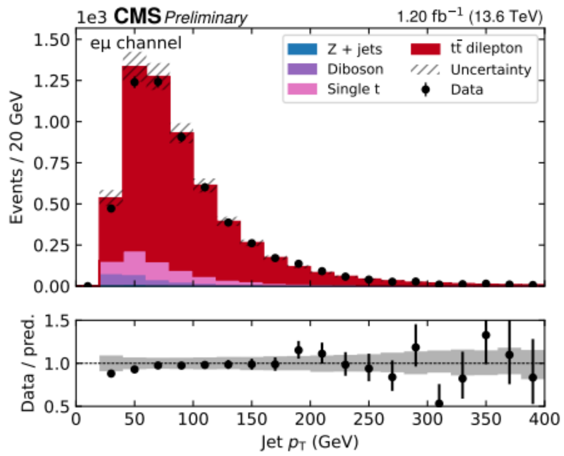
lepton+jets

- 1 lepton
- At least 3 jets
- At least 1 b-jet

Corrections and backgrounds

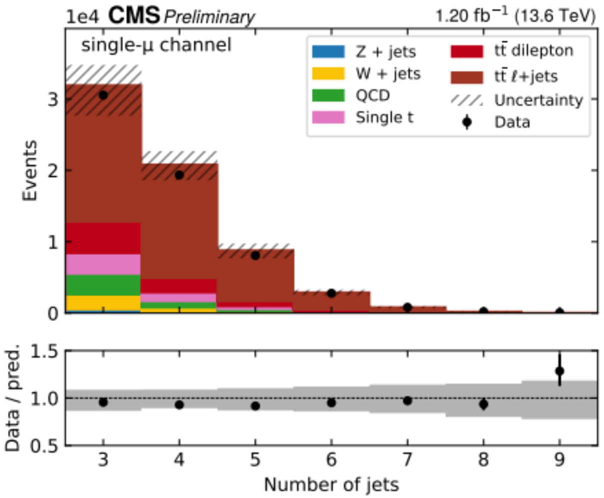
Jet energy correction

- Lepton+jets channel: define hadronic W with two leading non-btagged jets
- Use dijet mass to check & correct jet p_T agreement



Nonprompt / QCD background

- Data-driven method using lepton isolation & 1 jet sideband
- Relevant for lepton+jets

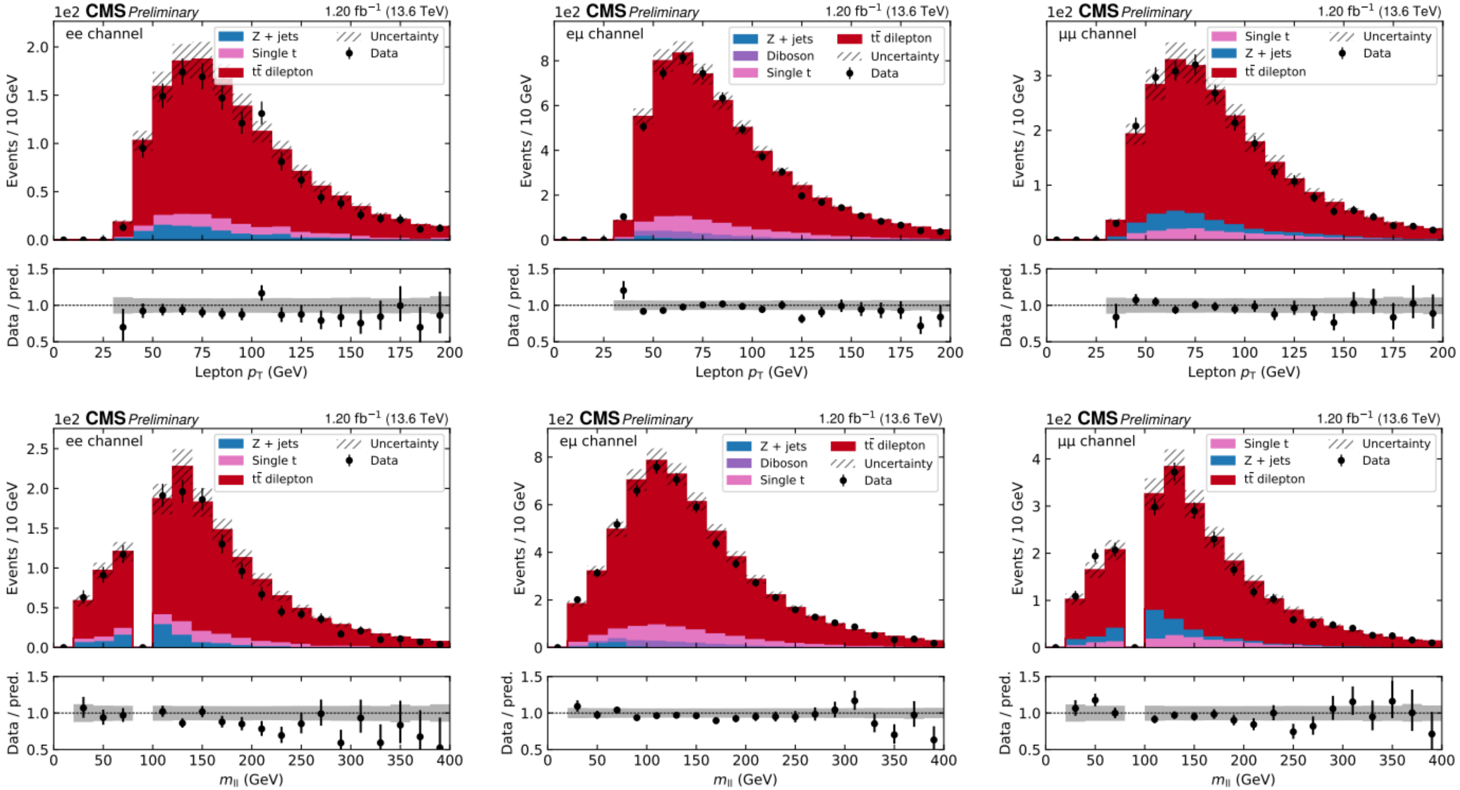


Control plots: dilepton

ee | eμ | μμ

leading lepton p_T

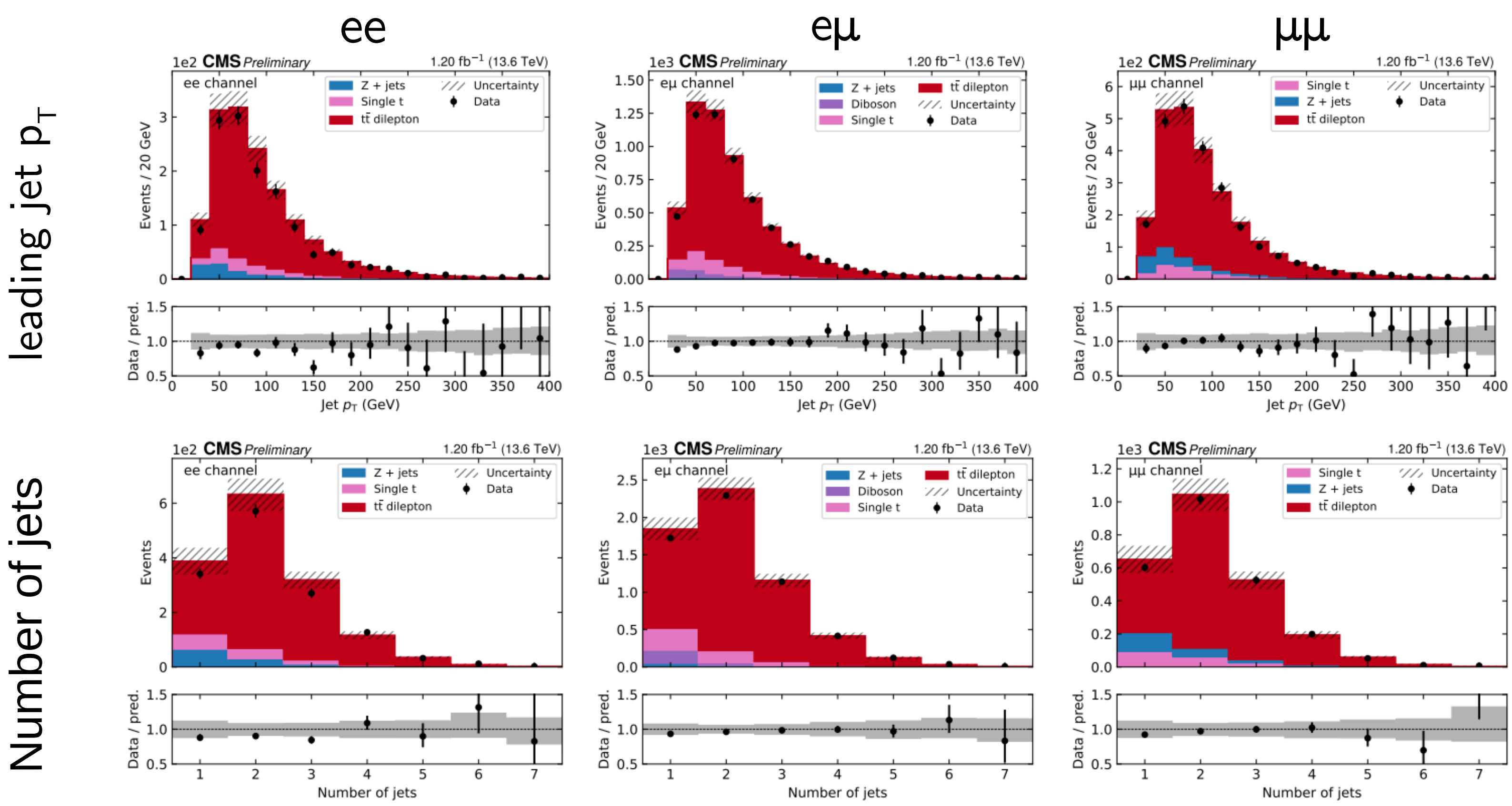
$m_{\ell\ell}$



Post-fit lepton SF applied

Note: we are looking for slopes and major mismodelings

Control plots: dilepton



Post-fit lepton SF applied

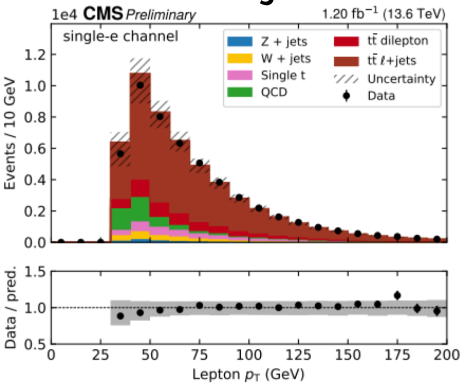
Note: we are looking for slopes and major mismodelings

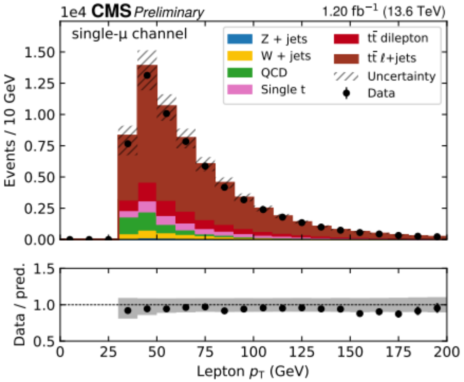
Control plots: lepton+jets

e+jets

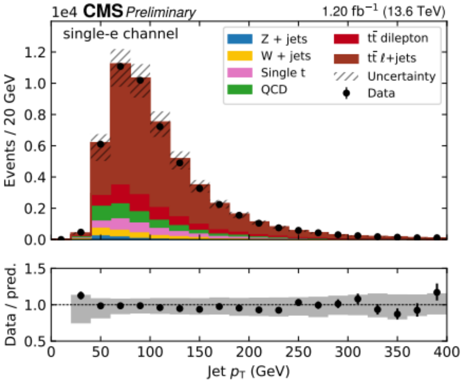
μ+jets

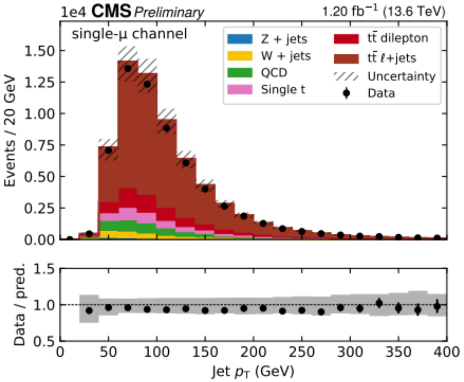
lepton p_T





leading jet p_T

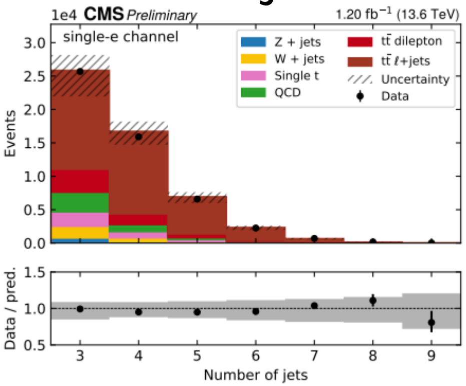


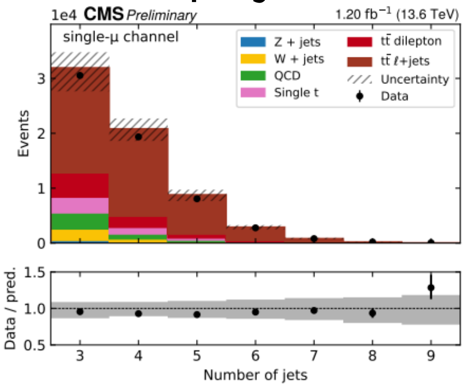


e+jets

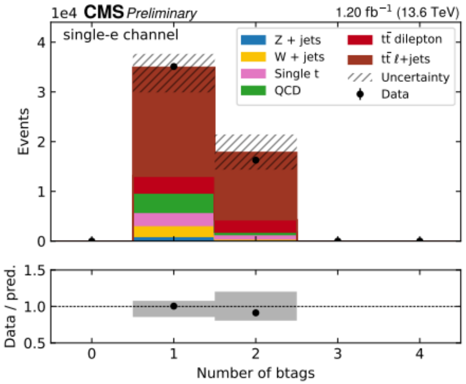
μ+jets

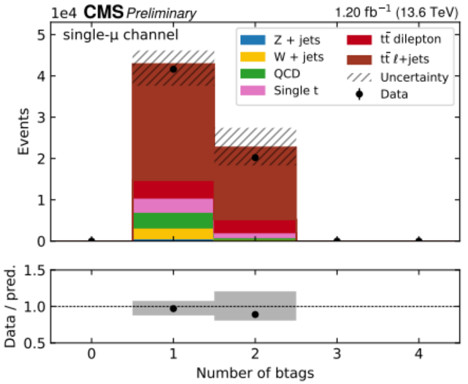
Number of jets





Number of b-tags





Analysis binning

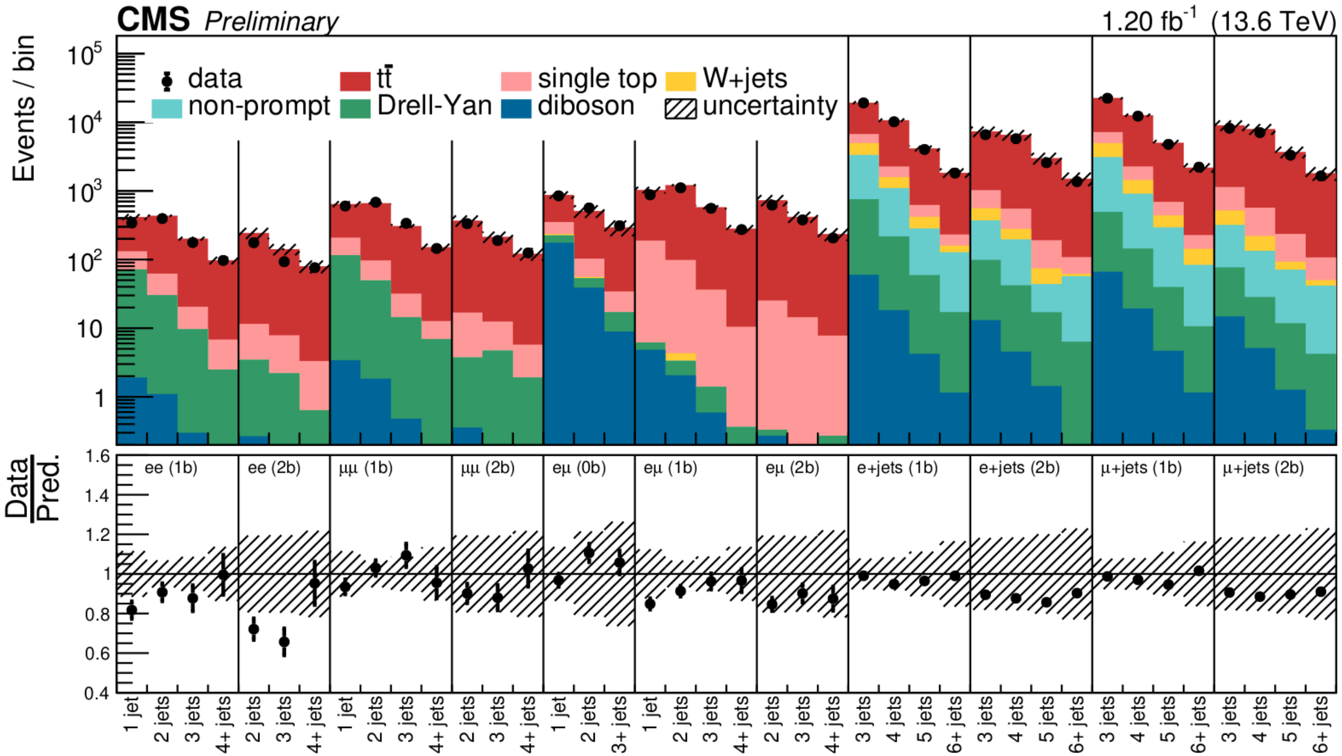
- Channels defined by
  - lepton content

  → further separated by
  - b jet content

  → coarsely binned in
  - N_{jet}

- Final binning shown (pre-fit)

Note: no lepton & b-jet SF applied, no lepton uncertainties



Statistical Methods

- Likelihood fit

$$\mathcal{L} = \prod_{\text{bin}} \mathcal{L}_{\text{bin}} \ , \quad \mathcal{L}_{\text{bin}} = \Gamma\left[n_{\text{obs}}^{\text{bin}} \middle| r\, s^{\text{bin}}(\{\theta_i\}) + b^{\text{bin}}(\{\theta_i\}) \right] \times \prod_i p_i(\theta_i)$$

  - Statistical fluctuations:
    $\rightarrow$ Poisson distribution
  - Normalization uncertainties:
    $\rightarrow$ log-normal distribution
  - Binned shape effects:
    $\rightarrow$ Template morphing
  - Lepton scale factors:
    $\rightarrow$ Floating parameters (flat pdf)

$$\Gamma[n|\lambda] = \frac{\lambda^n e^{-\lambda}}{n!}$$

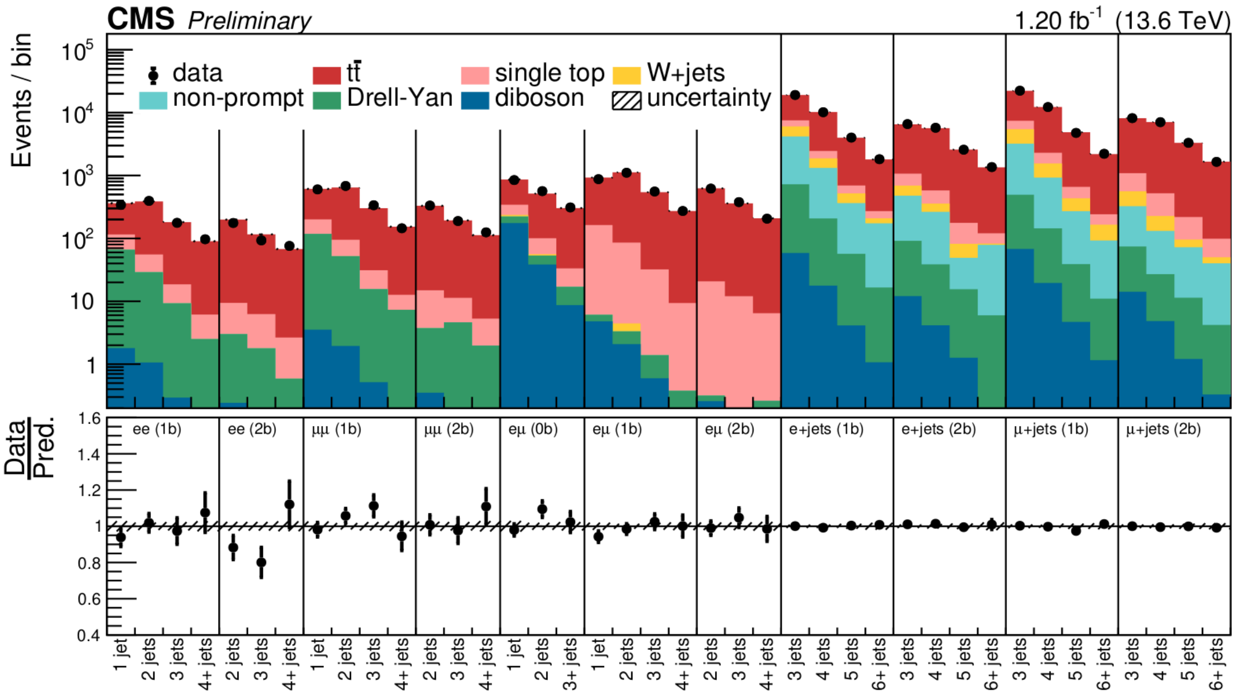
$$\begin{aligned} s &= \text{signal} \\ b &= \text{background} \\ \{\theta_i\} &= \text{nuisances} \\ p_i(\theta_i) &= \text{penalties} \end{aligned}$$

Jet calibrations:

- Only preliminary calibrations available for 2022 data
- We use a coarse calibration based on hadronic W mass for the main fit
- Run 2 jet uncertainties + difference w.r.t preliminary calibrations added as an external uncertainty

Result

$$\sigma_{t\bar{t}} = 887^{+43}_{-41}(\text{stat} + \text{syst}) \pm 53(\text{lumi})\text{pb}$$



Dominant uncertainties

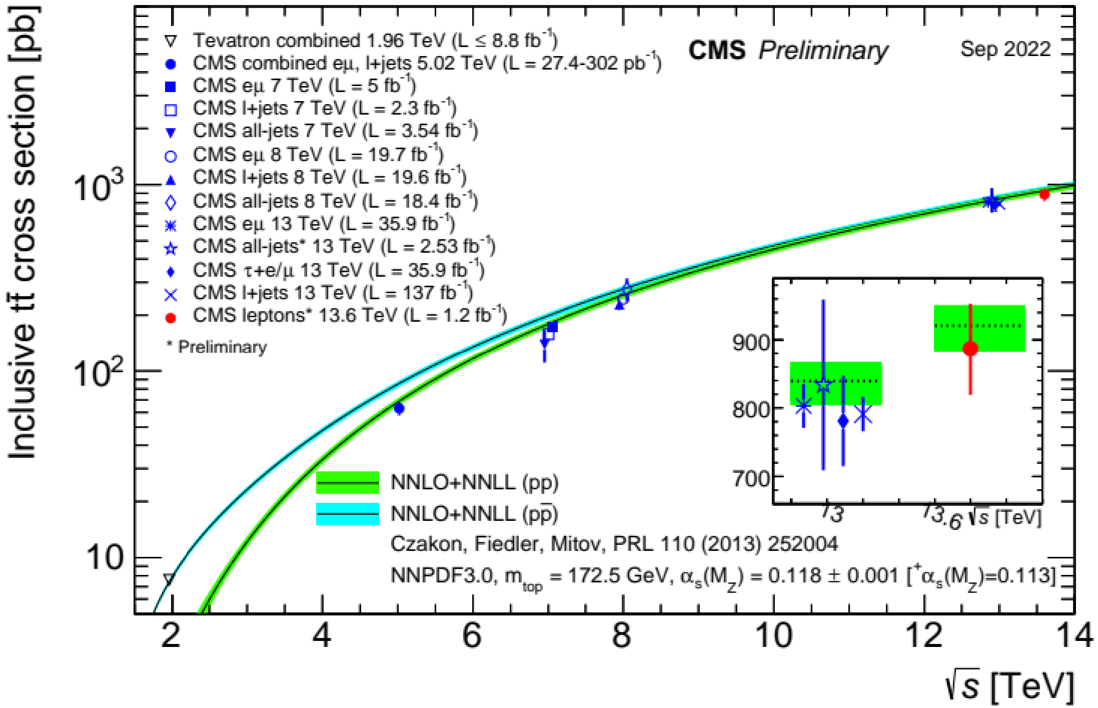
Source	Uncertainty (%)
Lepton ID SF	3.4
Jet energy scale	1.6
b tagging SF	1.5
ME/PS matching	1.1
Drell-Yan background	0.9
Pileup	0.7
combined likelihood fit	4
Jet calibration (external)	2
luminosity (external)	6

$$\sigma_{t\bar{t}} = 887^{+68}_{-67}\text{pb}$$

Theory: $\sigma_{t\bar{t}} = 921^{+18}_{-16}\text{pb}$

Conclusion

- First look at top quark physics at the new energy frontier!
- Novel measurement technique using multiple channels to constrain efficiencies in situ
- Published: CMS-PAS-TOP-22-012
- Aim for paper soon!



$$\sigma_{t\bar{t}} = 887^{+43}_{-41}(\mathrm{stat+syst}) \pm 53(\mathrm{lumi})\mathrm{pb}$$

Backup

Corrections and backgrounds

Pileup

- Experimental reweighting approach to 3 pileup-related variables:
  - Number of good vertices
  - Tracker energy flux
  - Calorimeter energy flux

Drell-Yan normalization

- DY background depends on b-jet multiplicity
- Check against data-driven estimate from inside the Z window
- Correction consistent with unity

Trigger efficiency

- Tag & Probe for single lepton triggers
  - Tag: loose ID lepton (90% efficiency) passing trigger
  - Probe: tight ID lepton (70% efficiency), as in selection
  - Coarsely binned in p_T and $|\eta|$
- Trigger efficiencies ≈ 95% for electrons and 90% for muons
- Scale factors are applied to the rest of the analysis

Non-prompt (NP) normalization

- QCD/non-prompt background modeled via ABCD method in lepton+jets channels
- Lepton isolation: invert iso requirement included in lepton ID to define sideband / control region (CR)
- Use 1-jet bin as orthogonal CR to apply ABCD method

  $\rightarrow$ Derive fake rate in 1 jet CR, apply to shape from iso sideband

- Simplest model w/ constant fake rates for e and μ:

$$N_{NP}^{SR} = (N_{data}^{CR} - N_{MC}^{CR}) \times \frac{(N_{data}^{SR,1j} - N_{MC}^{SR,1j})}{(N_{data}^{CR,1j} - N_{MC}^{CR,1j})}$$

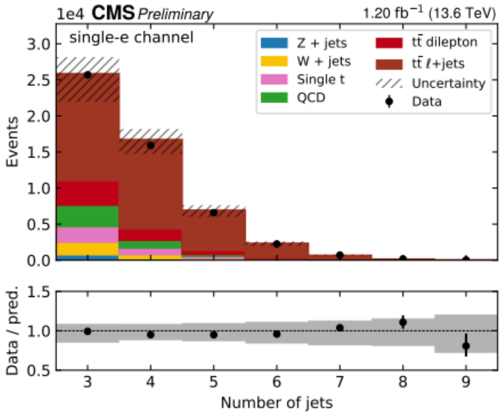
- But: signal contamination in CR!

Non-prompt (NP) normalization

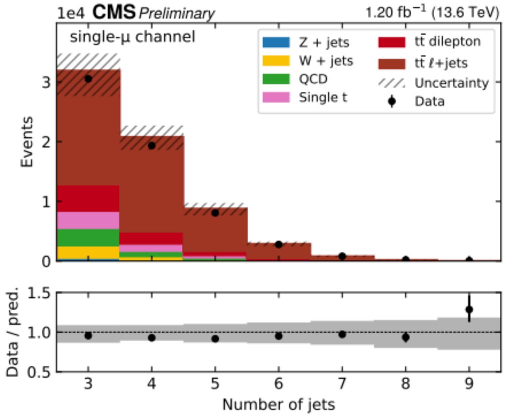
- Signal contamination might introduce circularity since cross section is unknown
- Use modified ABCD method to get around this:
  - Take only ratio $f_{B/A}$ of signal in CR/SR from MC
  - Use yield in SR and ratio to eliminate signal CR yield
- Final formula:

$$N_A^{QCD} = \frac{N_B^{tot} - N_B^{MCbg} - f_{B/A}(N_A^{tot} - N_A^{MCbg})}{1 - f_{B/A} R_{fake}} R_{fake}$$

electrons



muons



Drell Yan normalization

- Compare events with $m_{\ell\ell}$ inside/outside Z-peak sideband: $R_{in/out} = N_{in}/N_{out}$
- Use weaker assumption than standard $R_{in/out}$

$$\frac{(R^{\geq 1b}_{in/out})_{data}}{(R^{\geq 1b}_{in/out})_{MC}} = \frac{(R^{0b}_{in/out})_{data}}{(R^{0b}_{in/out})_{MC}}$$

$$\text{SF} = \frac{(N^{\geq 1b}_{out})_{data}}{(N^{\geq 1b}_{out})_{MC}} = \frac{(N^{\geq 1b}_{in})_{data}}{(N^{\geq 1b}_{in})_{MC}} \frac{(R^{0b}_{in/out})_{MC}}{(R^{0b}_{in/out})_{data}}.$$

- Drell Yan data content:

$$N_{data} = N^{\ell\ell}_{data} - 0.5 N^{e\mu}_{data} k_{\ell\ell}, \quad \text{where} \quad k_{ee} = \frac{1}{k_{\mu\mu}} = \sqrt{\frac{N^{ee}_{data}}{N^{\mu\mu}_{data}}}$$

Uncertainties

Externalized: lumi, Winter22 JES

Included in Likelihood fit:

Dominant experimental

Dominant theory

	implementation	treatment	tt	Single t	Drell Yan	diboson	W+jets
lepton ID	unconstrained	correlated	✓	✓	✓	✓	✓
JES	shape	correlated	✓	✓	✓	✓	✓
b tag SF	shape	correlated	✓	✓	✓	✓	✓
light mistag SF	shape	correlated	✓	✓	✓	✓	✓
Pileup	shape	correlated	✓	✓	✓	✓	✓
Trigger SF	shape	correlated	✓	✓	✓	✓	✓
PDF ($+\alpha_S$)	normalized shape	correlated	✓	✓	✓	✓	✓
ME scale	normalized shape	uncorrelated	✓	✓	✓	✓	✓
PS scale	normalized shape	correlated	✓	✓	✓	✓	✓
h_damp	dedicated sample	signal only	✓	–	–	–	–
BG cross sec*	normalization	uncorrelated	✓	15%	20%	30%	30%

*also 20% on NP background

b tag scale factors

- No b tag scale factors applied to nominal. Instead: estimate in situ
- b tag efficiency varies between data & MC as a function of jet kinematics
- However, our binning integrates over kinematics
  - Get template for flat b tag variation from first principles
  - The expected bin content follows a simple binomial distribution

$$\epsilon_b^{N_{\mathrm{b-tag}}} (1 - \epsilon_b)^{N_{\mathrm{b-jet}} - N_{\mathrm{b-tag}}}$$

$N_{\text{b-tag}}$ = number of b tagged jets,

$N_{\text{b-jet}}$ = number of true b-jets in event acceptance region

- 10% uncertainty on efficiency: intentionally wide

b tag scale factors

- The distribution of $N_{b\text{-jet}}$ is taken from MC to derive templates
- This method takes into account that in many cases, the "missing" b-jets in tt events are outside acceptance
- It also handles well samples which do not generally have 2 truth-level b jets
- Same "binomial" strategy applied for light jet mis-tagging