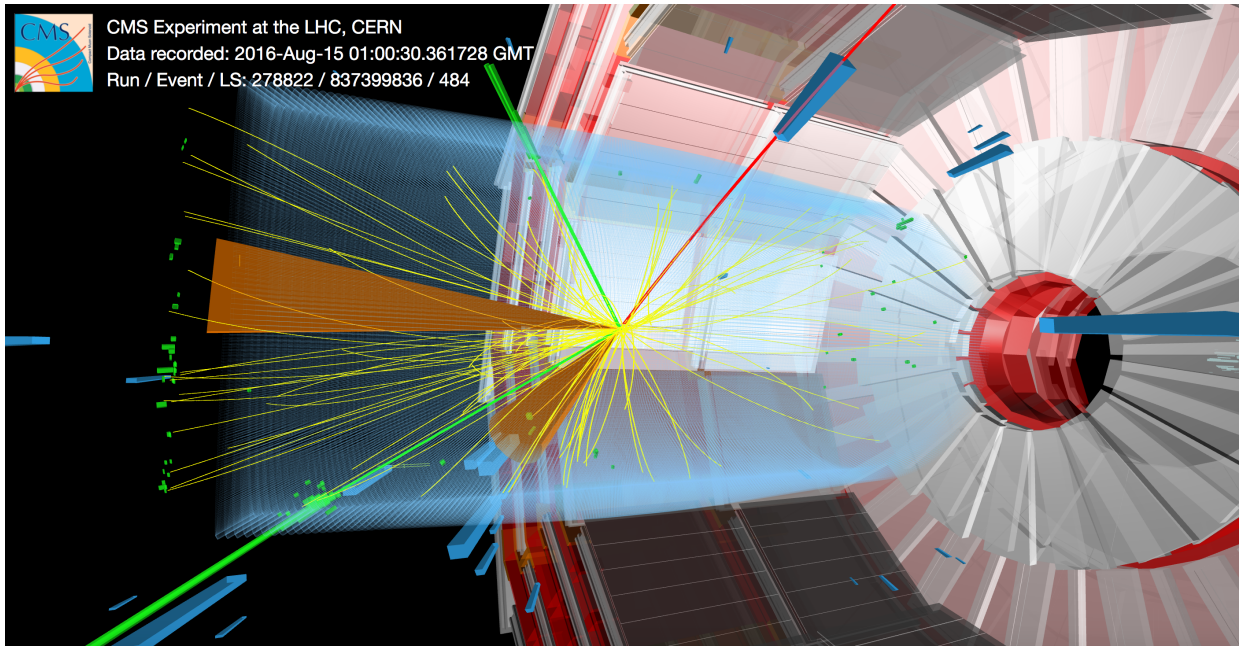
Rare processes with top quarks at CMS

LHCP2020 - 29/05/2020



Event with 2 electrons, 1 muon, 1b-jet and 1 forward jet,
likely to arise from tZq process

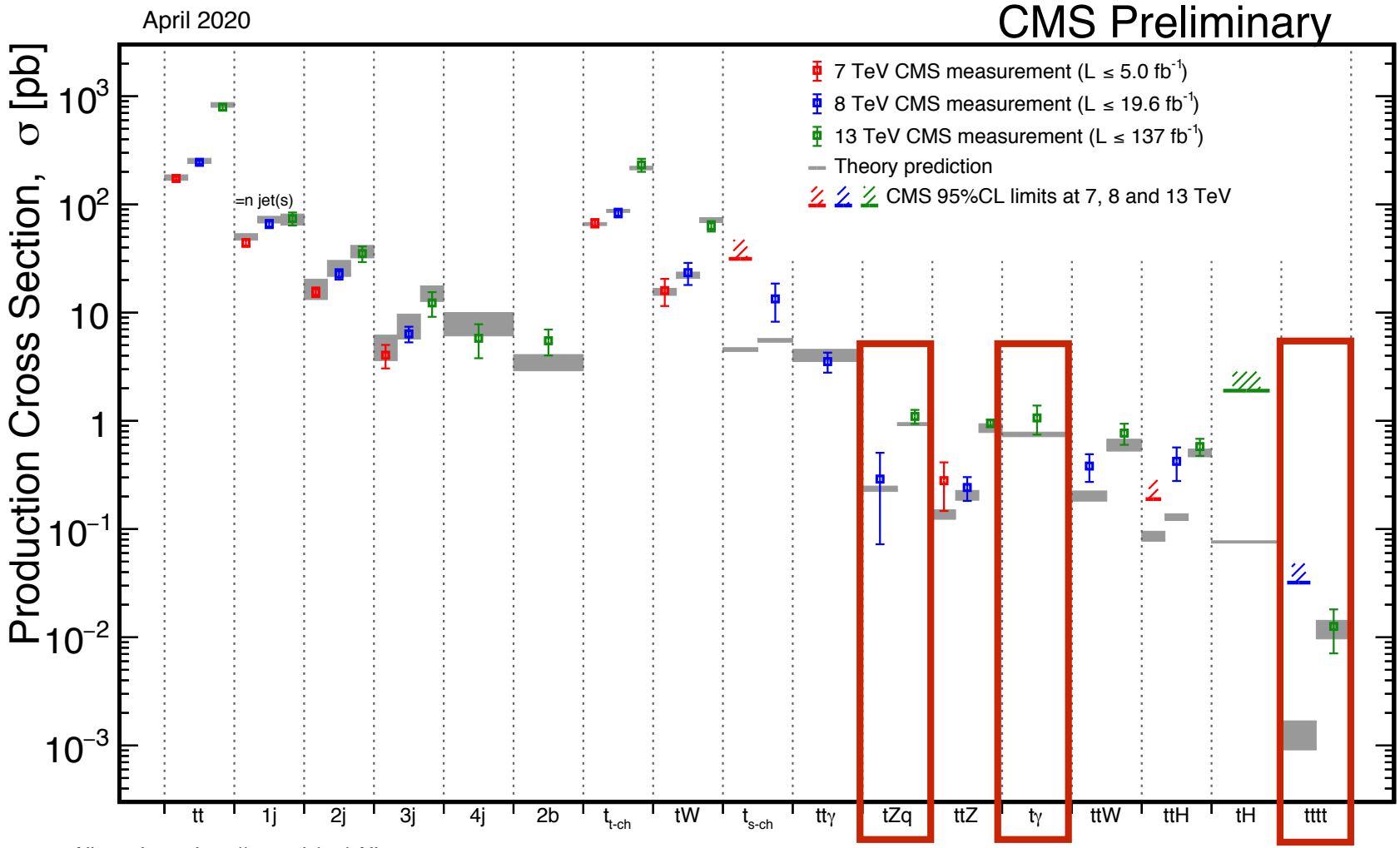
Nicolas Chanon - IP2I Lyon, CNRS/IN2P3 (France)
for the CMS Collaboration

Outline

Run 2 measurement of top quarks produced at low rate in the SM

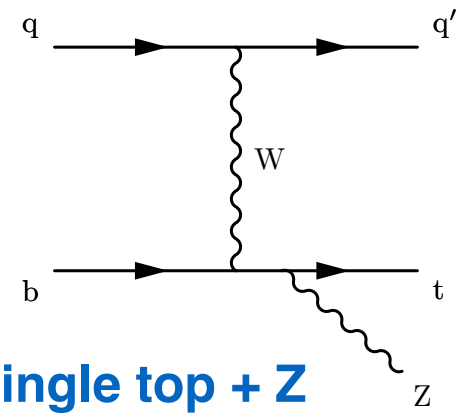
- **Single top + Z** (77.4 fb-1) in **3ℓ final state**
- **Single top + γ** (35.9 fb-1) in **1ℓ final state**
- **Four top quarks:**
  in **1ℓ+jets and 2ℓos** (35.8 fb-1) and **2ℓss and 3ℓ** (137 fb-1)

Latests results

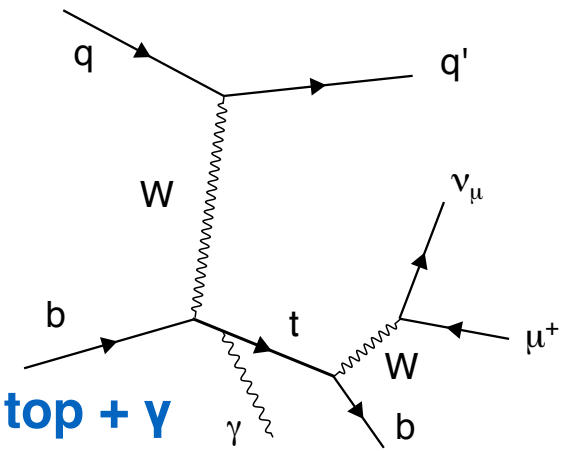


All results at: http://cern.ch/go/pNj7

Seaching for rare top processes: motivations



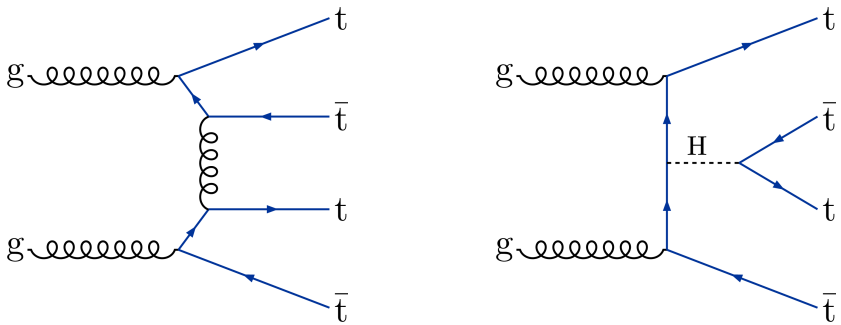
Single top + Z



Single top + γ

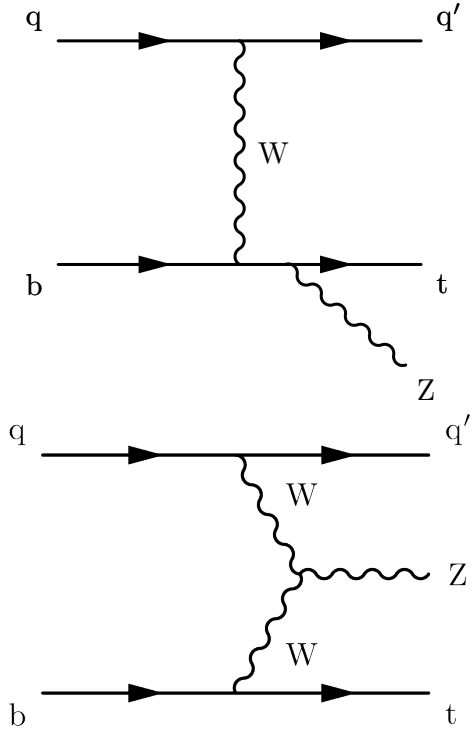
- **Test of SM predictions** (perturbative QCD, EW corrections): tZq and tγq processes are EW produced, four top processes are QCD induced
- **Probe of the top - boson coupling** (tZ in tZq, tγ in tγq, tH in four tops)
- Possible deviations in precision measurements are **sensitive to new physics,** for instance within the framework of the effective field theory

Four top quarks



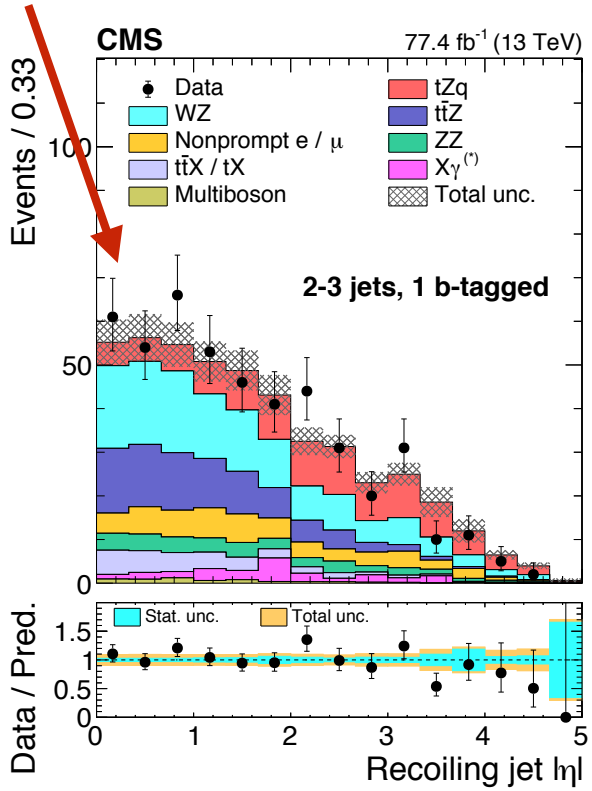
Single top + Z (tZq)

Phys. Rev. Lett. 122 (2019) 132003



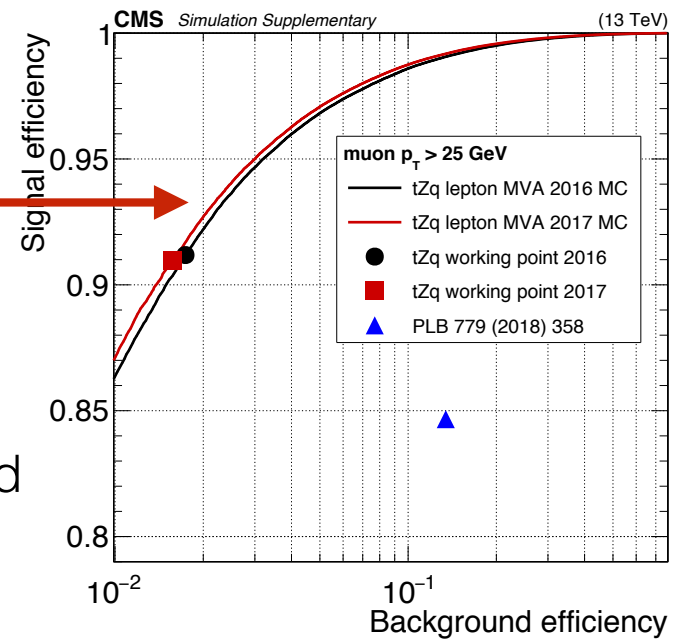
Analysis performed in **3ℓ final state (e or μ) with 2016/2017 data**

- Lepton pT > 25, 15, 10 GeV
- Two leptons (same flavour opposite sign) must have $|m_{\ell\ell} - M_Z| < 30$ GeV
- tZq is produced by EW interaction: make use of the **forward jet** ($|\eta|<4.7$) to discriminate signal from background



Very efficient **lepton identification with a BDT**:

- ~90%efficiency for p_T>25 GeV,
- 4x better background rejection than previous analysis



Single top + Z (tZq)

Phys. Rev. Lett. 122 (2019) 132003

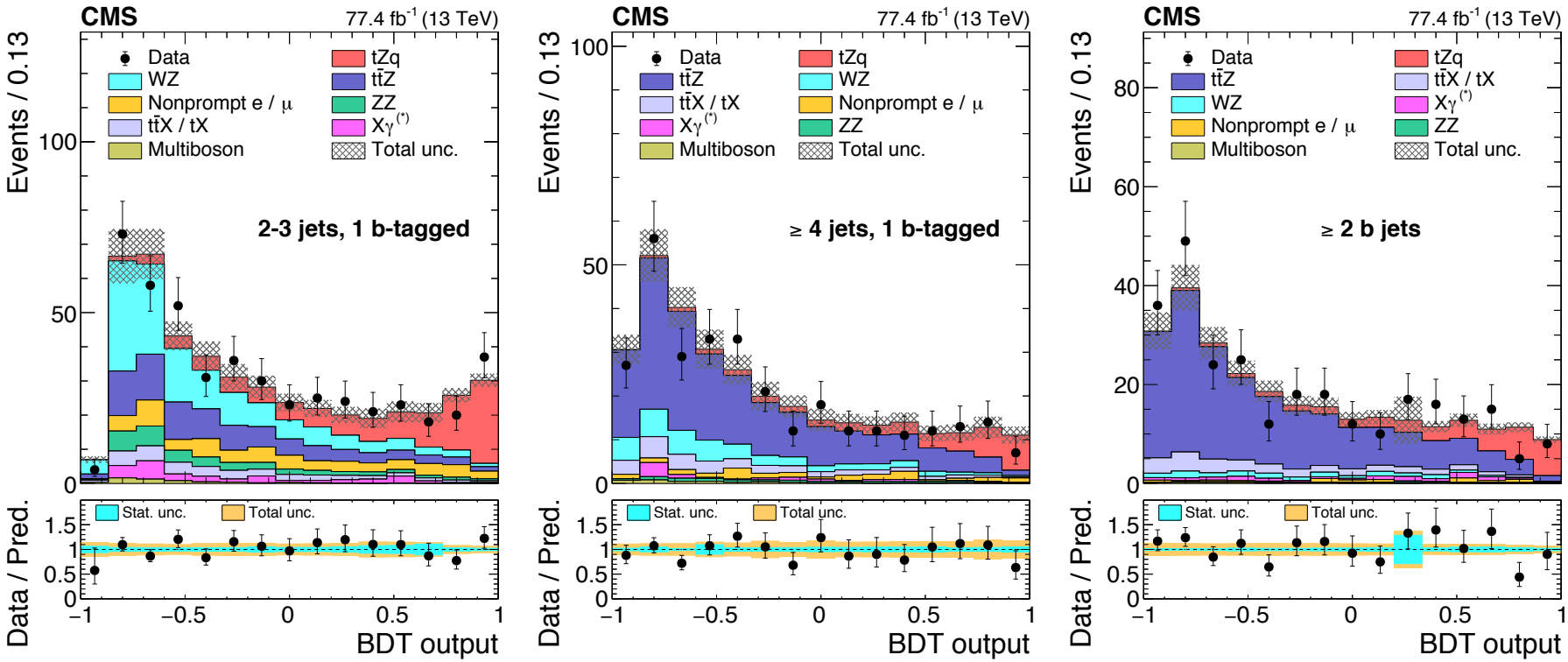
- Remaining **non prompt background** (ttbar and Z+jets): fake rate estimated in QCD jets data, applied to a sample with loosened analysis selection (30% uncertainty)
- WZ (ZZ) control regions: no b-jet, mET>50 GeV and 3 (4) leptons

Signal extraction: simultaneous fit of BDT discriminants in 3 categories

2-3 jets among which **1 b-jet** : Constrains tZq

≥4 jets among which **1 b-jet**: Constrains ttZ

≥2 b-jets: Constrains ttZ



Single top + Z (tZq)

Phys. Rev. Lett. 122 (2019) 132003

$\sigma(pp \to tZq \to t\ell^+\ell^-q) = 111 \pm 13(\text{stat})^{+11}_{-9}(\text{syst})$ fb

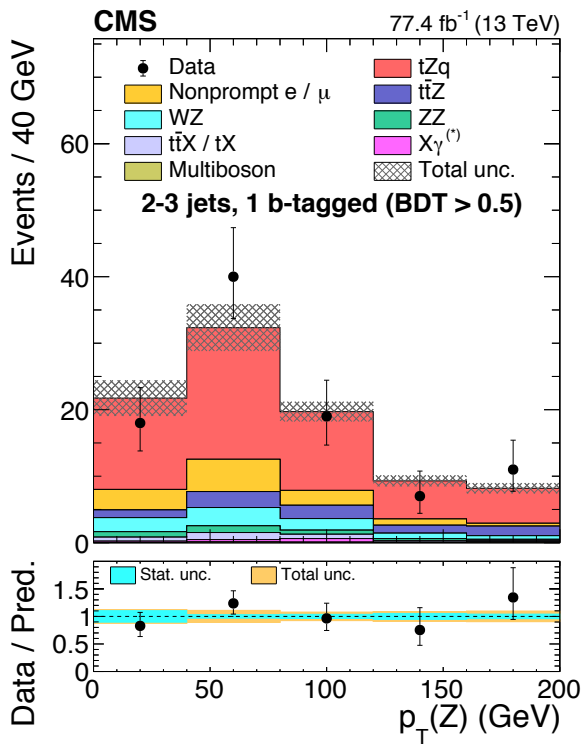
$\sigma^{\text{SM}}(pp \to tZq \to t\ell^+\ell^-q) = 94.2 \pm 3.1$ fb

NLO QCD, $m_{\ell\ell} > 30$ GeV

Observation: 8.2σ (7.7σ expected)

Main uncertainties: Non-prompt background, jet energy scale, lepton efficiency, final state radiation and tZq QCD scale uncertainty

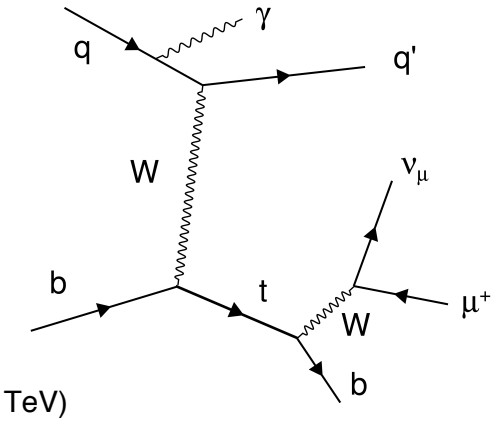
Uncertainty	Impact (%)
Experimental	
lepton selection	3.2
trigger efficiency	1.4
jet energy scale	3.3
b-tagging efficiency	1.7
nonprompt normalization	4.1
$t\bar{t}Z$ normalization	1.0
luminosity	1.7
pileup	1.9
other	1.3
Theoretical	
final-state radiation	2.0
tZq QCD scale	2.0
$t\bar{t}Z$ QCD scale	1.4



Now mature for differential measurements

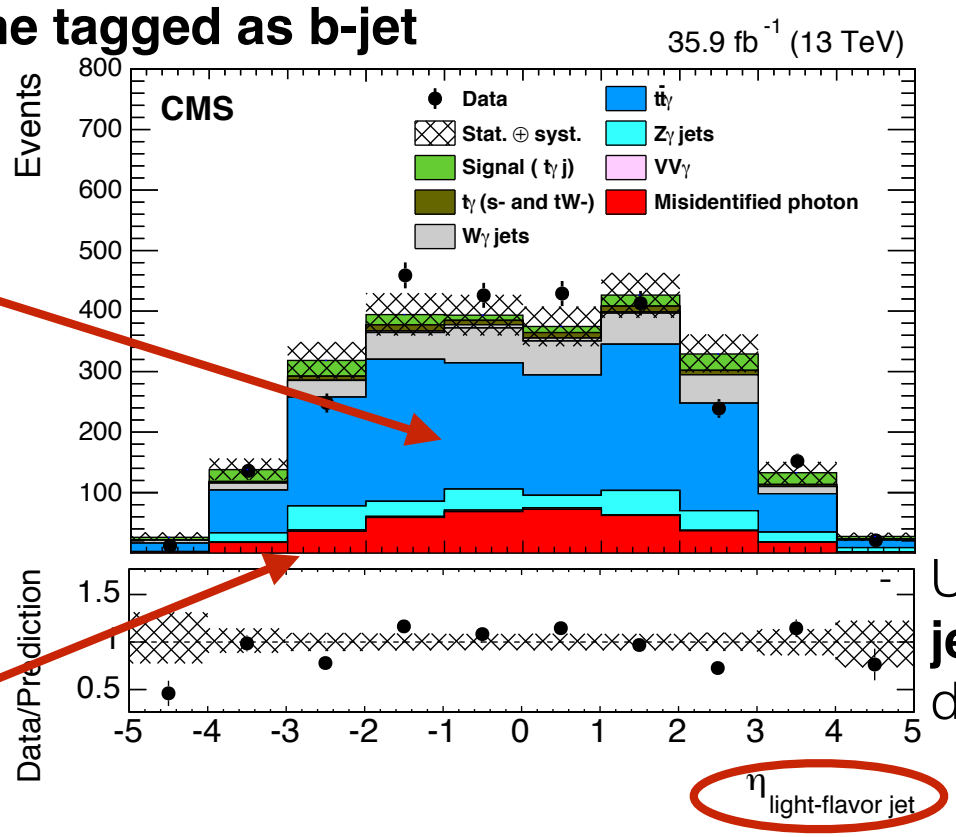
Single top + γ (tγq)

Phys. Rev. Lett. 121 (2018) 221802



Target **muonic** top quark decay, **2016 data**

- At least two jets, with only **one tagged as b-jet**
- **Dominant background tt+γ**



Uses the **forward jet** to increase discrimination

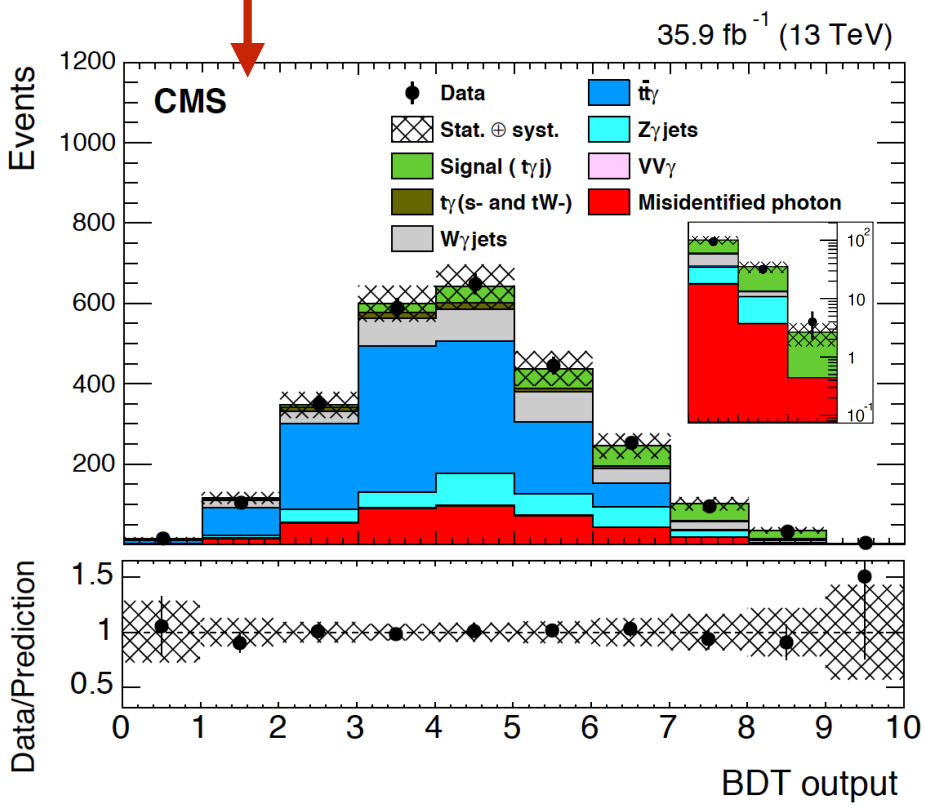
Estimate **non-prompt photon** background with a **fake ratio method**

- relax photon requirement,
- measure probability to mis-identify a non-prompt as a prompt photon

Single top + γ (tγq)

Phys. Rev. Lett. 121 (2018) 221802

Build a **BDT** discriminant using kinematics (most important variable **forward jet η**)



- Template for **tt+γ obtained from data** control region with additional b-jet
- **Simultaneous fit** of **tγq and tt+γ regions**

First evidence for tγq:
Observed significance 4.4σ (3.0σ expected)

Dominated by systematics: jet energy scale, b-tagging, tγq modeling

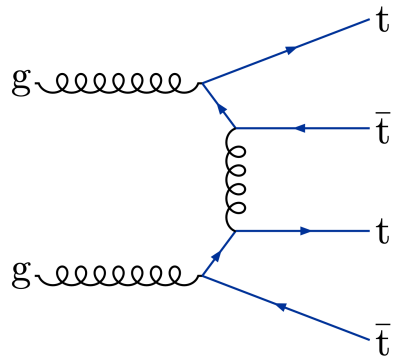
$$\sigma(pp \to t\gamma j)\mathcal{B}(t \to \mu\nu b) = 115 \pm 17(\text{stat}) \pm 30(\text{syst})\ \text{fb}$$

The SM predicted cross section is 81 ± 4 fb

(pTγ > 25 GeV, |η|<1.44)

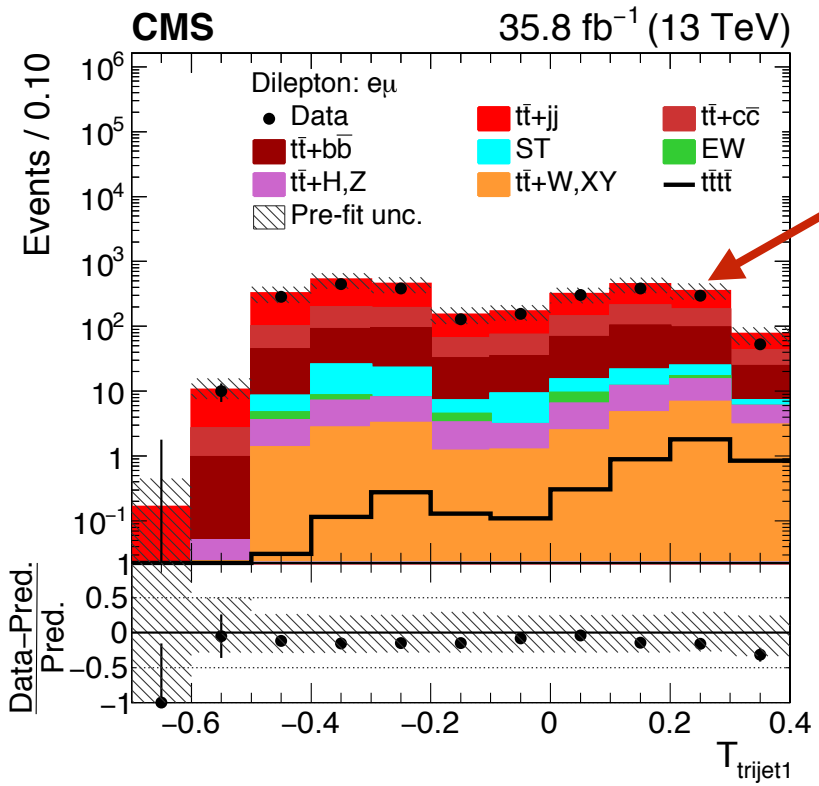
Four tops: 1ℓ and 2ℓ opposite sign

JHEP 11 (2019) 082



Analysis in **opposite sign 2ℓ and 1ℓ+jets final state,** with **2016 data**

- **Single lepton**: 1ℓ + 7 (8) jets in single muon (electron), at least 2 b-tagged jets, HT > 500 GeV, mET > 50 GeV
- **Dilepton**: 4 jets with 2 b-tagged jets, HT > 500 GeV



Hadronic top reconstruction

BDT score for **jet triplets**

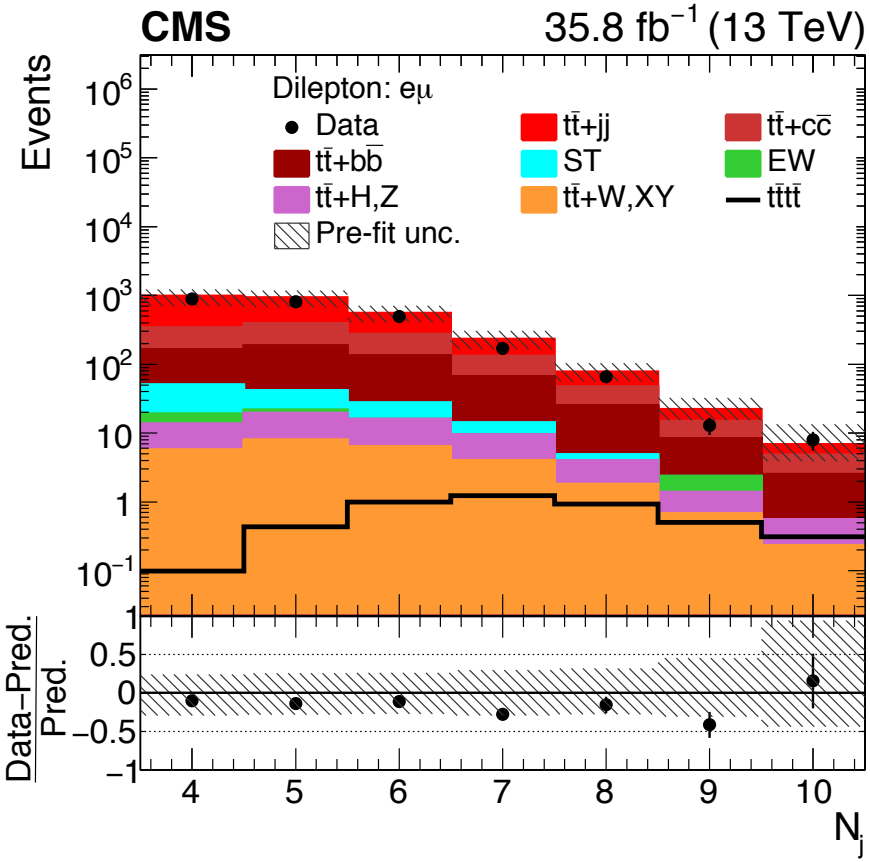
Corrections needed for additional jets:

- **4 tops: jet multiplicity correction** from ttbar data, in a region with 8,9 jets (2,3 b-tags)
- **ttbar+bb reweighed** to 13 TeV measurement

Four tops: 1ℓ and 2ℓ opposite sign

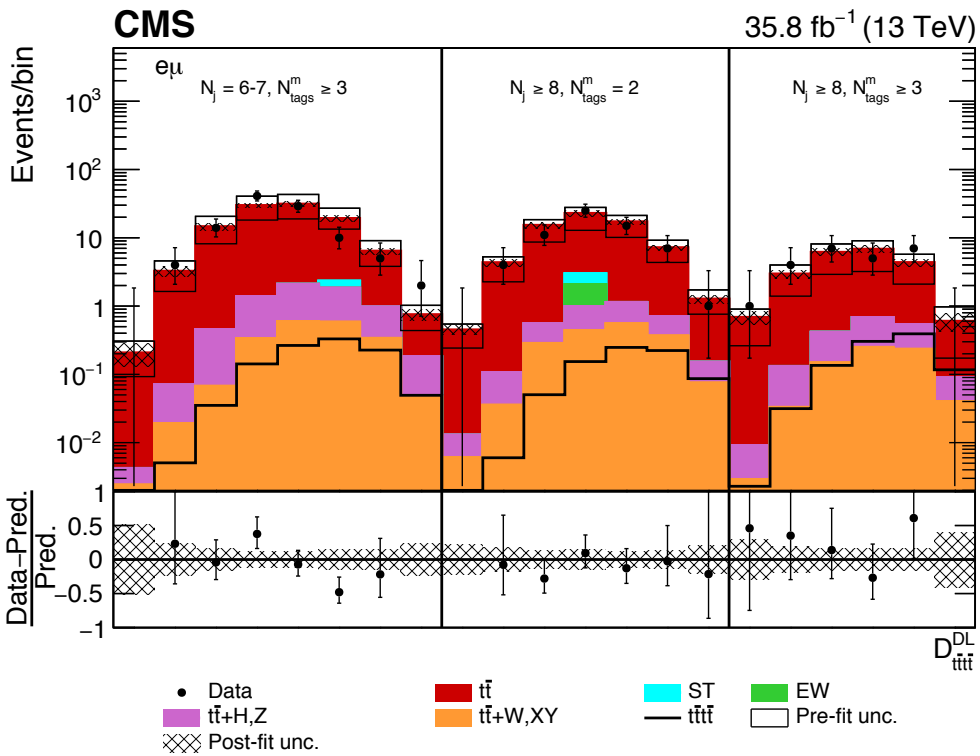
JHEP 11 (2019) 082

Categorize with jets and b-jets multiplicities



Event-BDT:

- BDT score for hadronic top from jet triplets
- hadronic variables (related to the number of jet, jet kinematics, HT, multijet masses, the event sphericity…),
- lepton and b-jets variables



Four tops: 1ℓ and 2ℓ opposite sign

JHEP 11 (2019) 082

Results and combination

- Adding 10% precision to four top multilepton analysis (2016 data)

Channel	Expected limit, μ	Observed limit, μ	Expected limit (fb)	Observed limit (fb)
Single-lepton	$9.4^{+4.4}_{-2.9}$	10.6	86^{+40}_{-26}	97
Dilepton	$7.3^{+4.5}_{-2.5}$	6.9	67^{+41}_{-23}	64
Combined (this analysis)	$5.7^{+2.9}_{-1.8}$	5.2	52^{+26}_{-17}	48
Multilepton	$2.5^{+1.4}_{-0.8}$	4.6	23^{+12}_{-8}	42
Combined (this analysis + multilepton)	$2.2^{+1.1}_{-0.7}$	3.6	20^{+10}_{-6}	33

Dimension 6 EFT interpretation:

- Minimal Flavour Violation: consider **four-quark operators** involving only 3rd generation quarks
- ttbb operator are neglected
- SM kinematics of tttt is assumed, and limits are derived from rate information

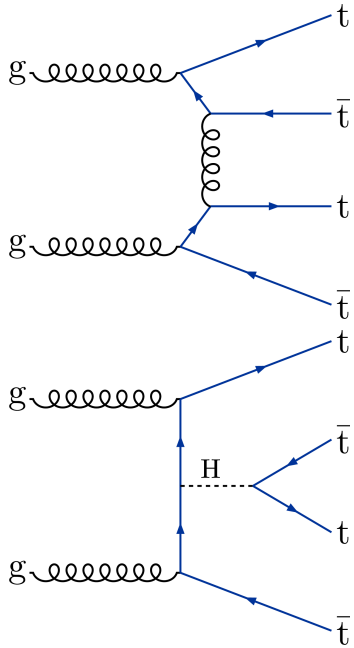
$$L = L_{SM} + \sum_i \boxed{\frac{a_i}{\Lambda^2} O_i} + \sum_j \frac{b_i}{\Lambda^4} O_j + \ldots$$

$$\mathcal{O}^1_{\mathrm{tt}} = (\bar{\mathrm{t}}_{\mathrm{R}}\gamma^\mu \mathrm{t}_{\mathrm{R}})(\bar{\mathrm{t}}_{\mathrm{R}}\gamma_\mu \mathrm{t}_{\mathrm{R}}),$$
$$\mathcal{O}^1_{\mathrm{QQ}} = (\overline{\mathrm{Q}}_{\mathrm{L}}\gamma^\mu \mathrm{Q}_{\mathrm{L}})(\overline{\mathrm{Q}}_{\mathrm{L}}\gamma_\mu \mathrm{Q}_{\mathrm{L}}),$$
$$\mathcal{O}^1_{\mathrm{Qt}} = (\overline{\mathrm{Q}}_{\mathrm{L}}\gamma^\mu \mathrm{Q}_{\mathrm{L}})(\bar{\mathrm{t}}_{\mathrm{R}}\gamma_\mu \mathrm{t}_{\mathrm{R}}),$$
$$\mathcal{O}^8_{\mathrm{Qt}} = (\overline{\mathrm{Q}}_{\mathrm{L}}\gamma^\mu T^{\mathrm{A}} \mathrm{Q}_{\mathrm{L}})(\bar{\mathrm{t}}_{\mathrm{R}}\gamma_\mu T^{\mathrm{A}} \mathrm{t}_{\mathrm{R}})$$

Operator	Expected C_k/Λ^2 (TeV^{-2})	Observed (TeV^{-2})
$\mathcal{O}^1_{\mathrm{tt}}$	$[-2.0, 1.8]$	$[-2.1, 2.0]$
$\mathcal{O}^1_{\mathrm{QQ}}$	$[-2.0, 1.8]$	$[-2.2, 2.0]$
$\mathcal{O}^1_{\mathrm{Qt}}$	$[-3.3, 3.2]$	$[-3.5, 3.5]$
$\mathcal{O}^8_{\mathrm{Qt}}$	$[-7.3, 6.1]$	$[-7.9, 6.6]$

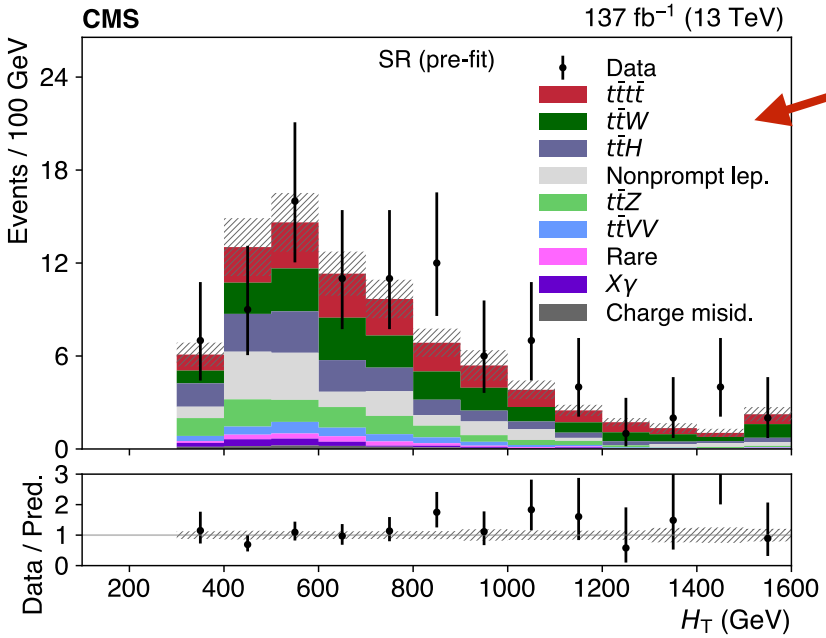
Four tops: multilepton

Eur. Phys. J. C 80 (2020) 75



Analysis in **same sign 2ℓ and ≥3ℓ final state,** with **full Run 2 data**

- Take advantage of the **large jet multiplicity:** require HT>300 GeV



Analysis strategy:

- **New multivariate classifier (BDT):** 17 regions + ttZ region
- **Cut-based cross-check analysis**: Classify events in the number of leptons, jets and b-jets (14 regions), ttZ and ttW control regions

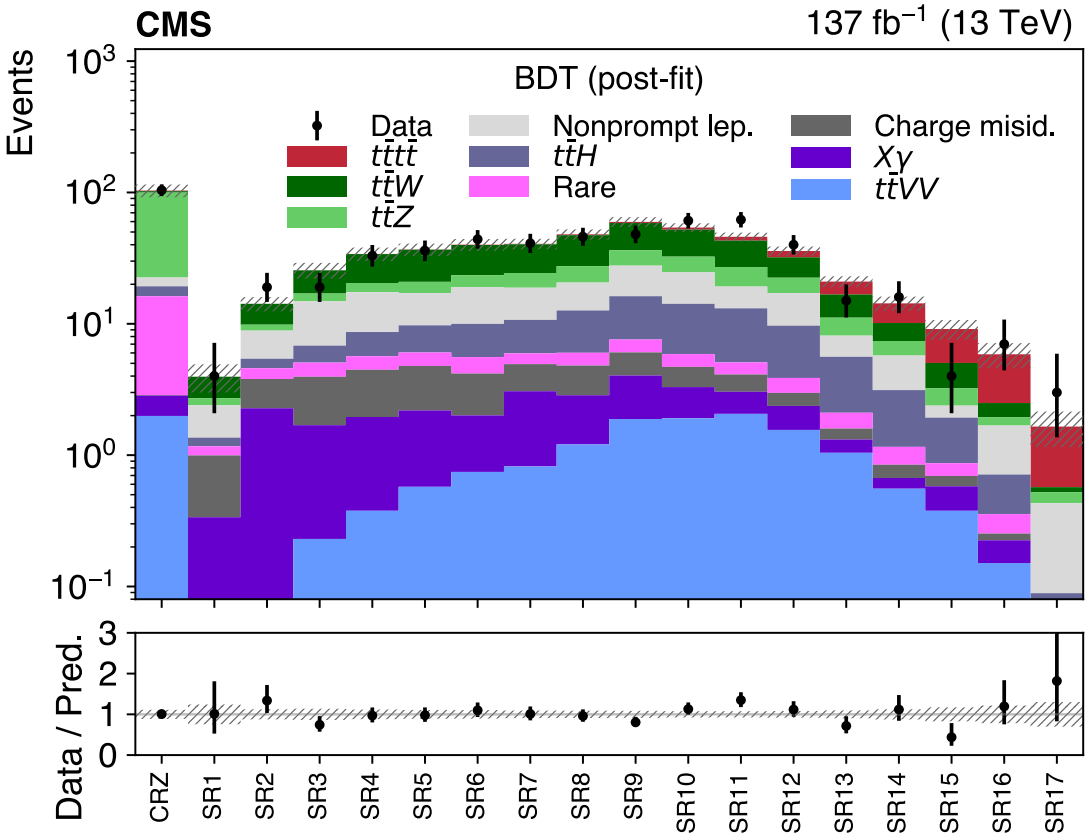
Background estimate:

- **ISR/FSR correction from ttbar data** applied to ttW/Z simulation
- **ttbb/ttjj correction**, resulting in 70% increase in additional bb pair in ttH/W/Z
- **Charge misidentification** (2ℓss only)**:** flip rate from Z→ℓ±ℓ± data
- **Non-prompt lepton background:** similar as in tZq analysis, fake rate from ℓ+jets

Four tops: multilepton

Eur. Phys. J. C 80 (2020) 75

Post-fit BDT distribution:



Source	Uncertainty (%)	Impact on $\sigma(t\bar{t}t\bar{t})$ (%)
Integrated luminosity	2.3–2.5	2
Pileup	0–5	1
Trigger efficiency	2–7	2
Lepton selection	2–10	2
Jet energy scale	1–15	9
Jet energy resolution	1–10	6
b tagging	1–15	6
Size of simulated sample	1–25	<1
Scale and PDF variations †	10–15	2
ISR/FSR (signal) †	5–15	2
$t\bar{t}H$ (normalization) †	25	5
Rare, $X\gamma$, $t\bar{t}VV$ (norm.) †	11–20	<1
$t\bar{t}Z$, $t\bar{t}W$ (norm.) †	40	3–4
Charge misidentification †	20	<1
Nonprompt leptons †	30–60	3
$N_{\text{jets}}^{\text{ISR/FSR}}$	1–30	2
$\sigma(t\bar{t}b\bar{b})/\sigma(t\bar{t}jj)$ †	35	11

Observed: 2.6σ (2.7σ expected)

the $t\bar{t}\,t\bar{t}$ cross section is measured to be $12.6^{+5.8}_{-5.2}$ fb

SM predictions: $\sigma(pp \to t\bar{t}\,t\bar{t}) = 12.0^{+2.2}_{-2.5}$ fb
(NLO QCD + EW correction)

Main systematics: Jet energy scale and resolution, b-tagging, ttbar+bb modeling, ISR/FSR

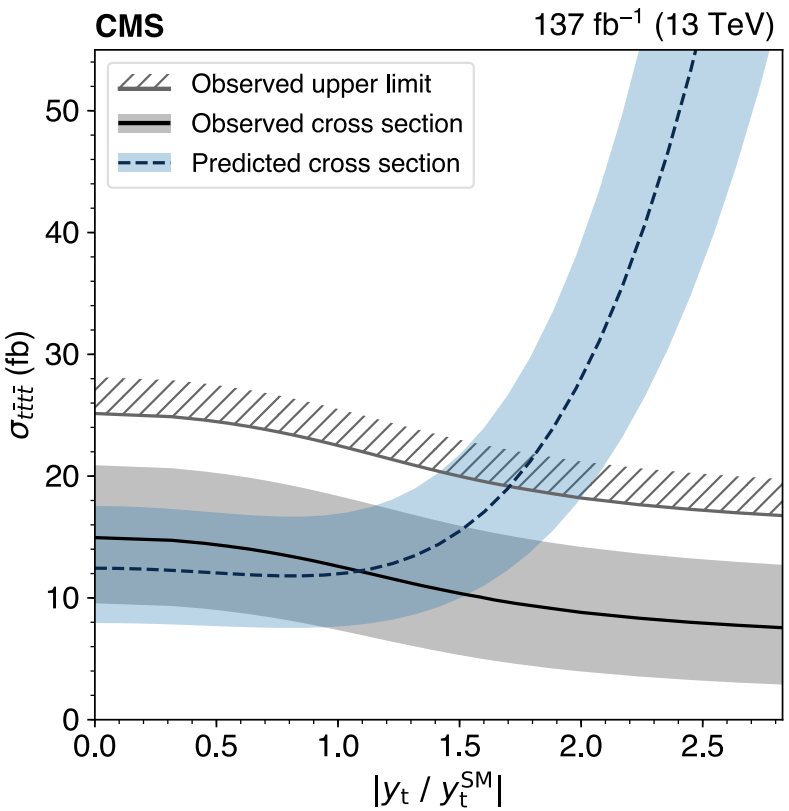
Four tops: multilepton

Eur. Phys. J. C 80 (2020) 75

Constraint on **Top-Higgs coupling** $y_t/y_{t,SM}=\kappa_t$

- Complementary to Higgs measurements
- Includes ttH scaling as y_t^2 in the fit

$|y_t/y_t^{SM}| < 1.7$ 95% confidence level



Constraint on the **Higgs boson oblique parameter $\hat{H}$** *[JHEP09(2019)041]* in the EFT framework:

- Modifies Higgs boson propagator, and coupling to quark and leptons
- Rescale ttH as $(1-\hat{H})^2$

95% CL upper limit of $\hat{H} < 0.12$

Comparable with sensitivity in off-shell Higgs measurements

Several other interpretations are available in the paper (Dark Matter, 2HDM, hMSSM)

Conclusions and perspectives

Processes with top quarks produced at low rate in the SM

- **Observation of tZq process at CMS with >5σ**
  - Main improvement brought in this analysis: reworked lepton identification to further reduce non-prompt background. Now mature for differential measurements

- **tγq analysis:**
  - **First evidence of tγq at 4.4σ** (3.0σ expected): observation is the next obvious goal

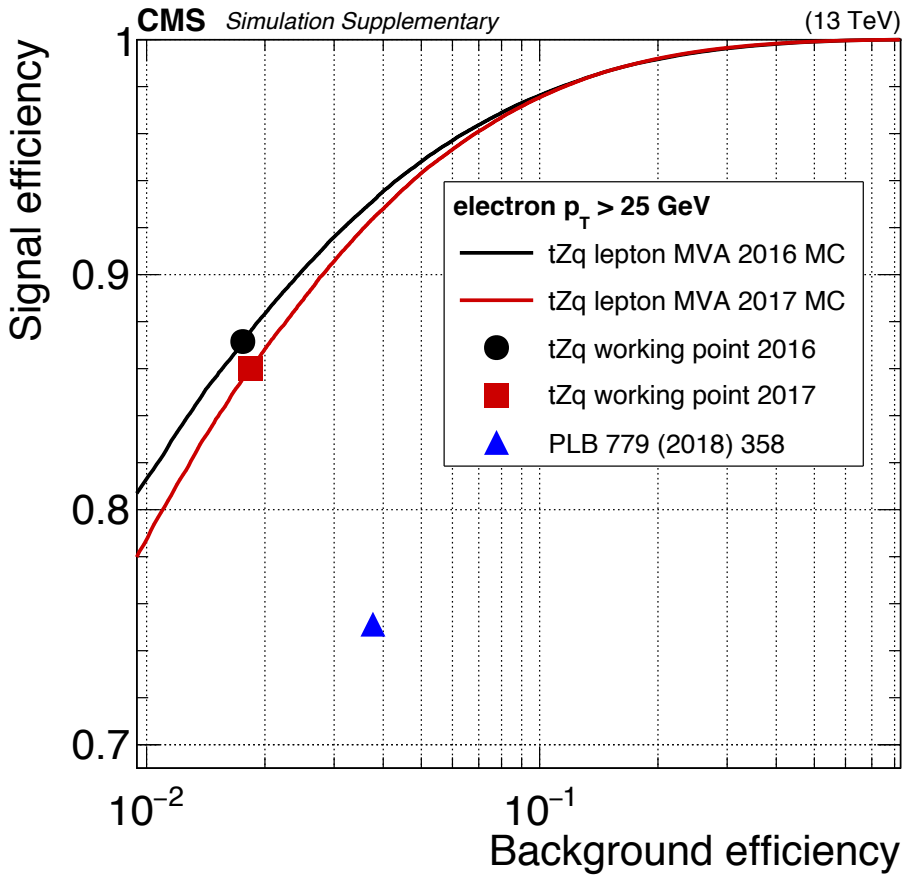
- **Four top quarks**
  - Very rare process, not yet observed
  - **Observed significance 2.6σ** (2.7σ expected) with multilepton analysis: efforts should continue to reach an observation. Maybe at Run 3 ?
  - Very sensitive to new physics, interesting interpretations to be pursued

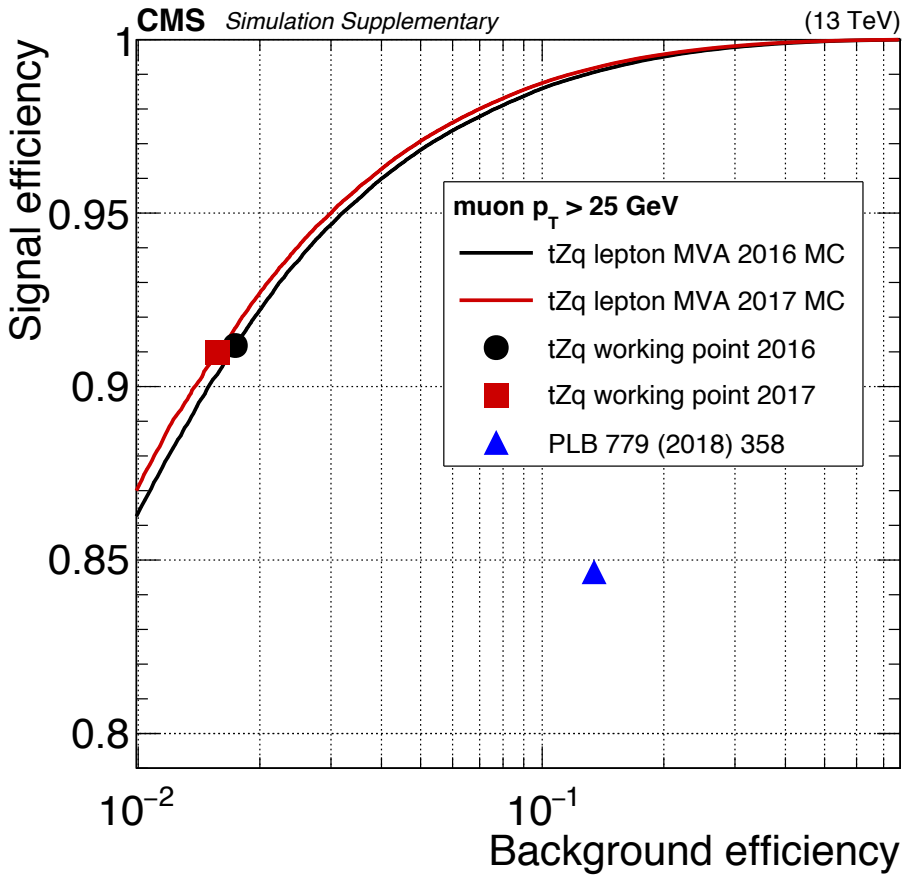
Zoom room for common discussion on rare top processes:
https://cern.zoom.us/j/93618530848?pwd=SHBHY3hZNG50VFFPSndOMFQzTzN6dz09

Back-up slides

Single top + Z (tZq): lepton identification

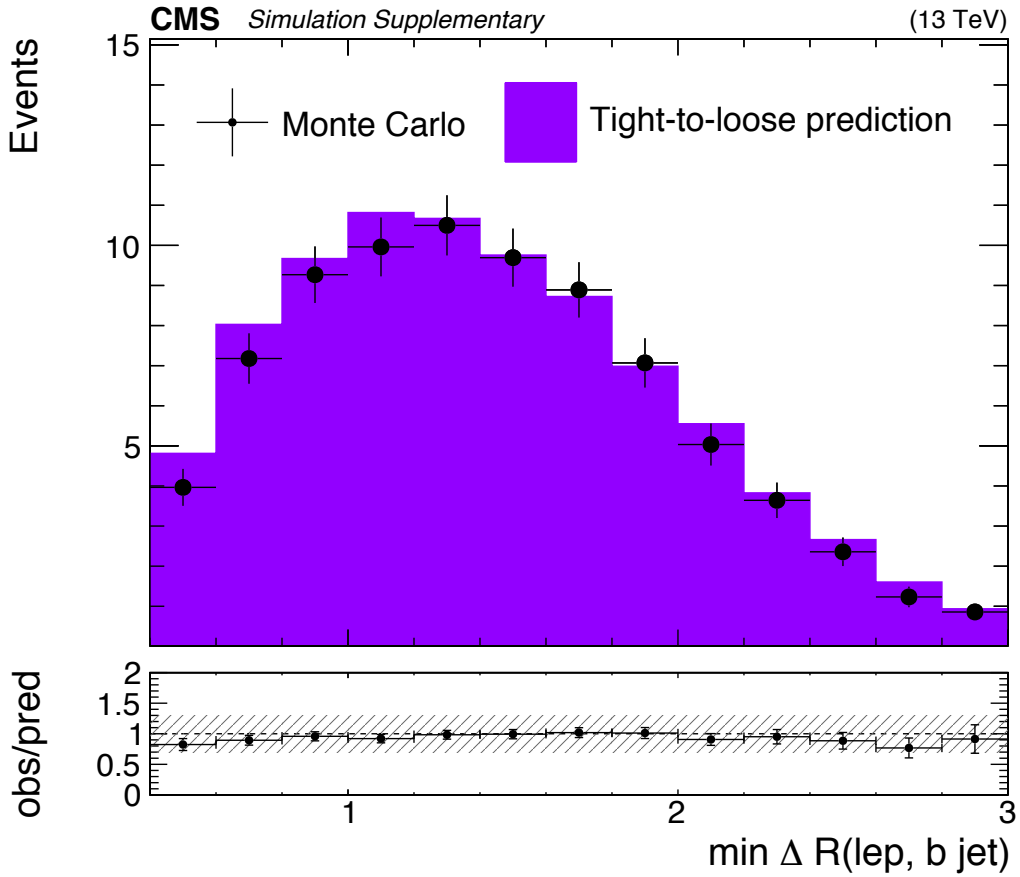
Phys. Rev. Lett. 122 (2019) 132003

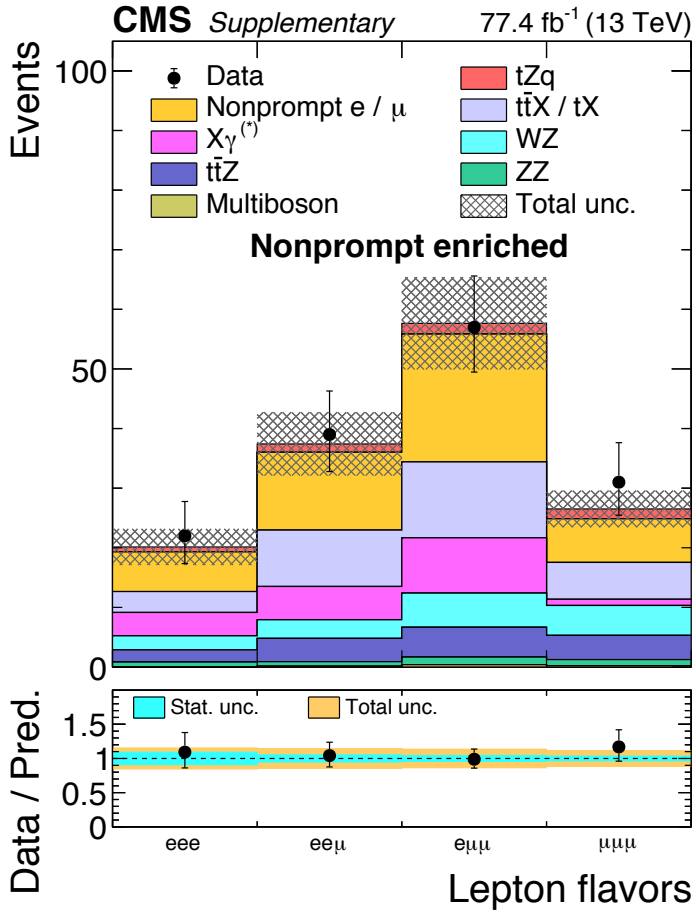




Single top + Z (tZq): non-prompts

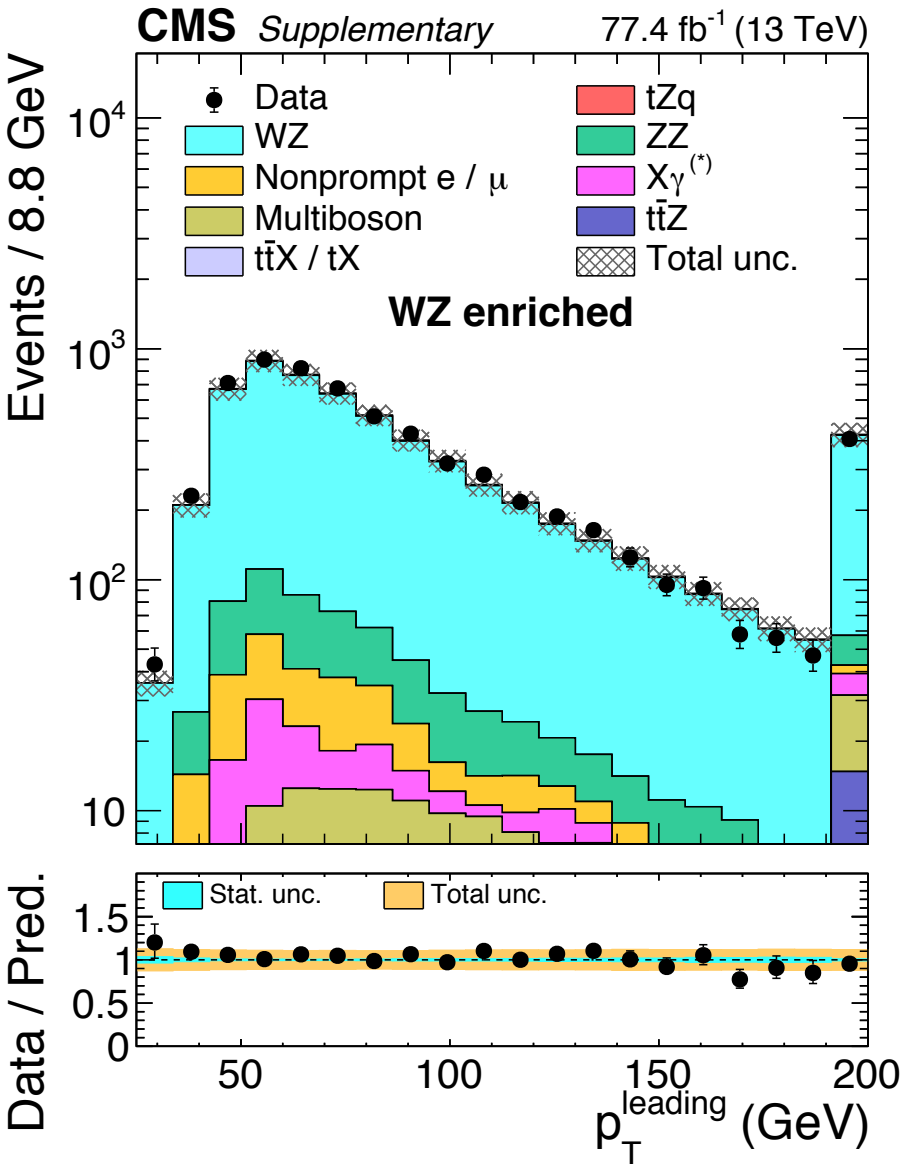
Phys. Rev. Lett. 122 (2019) 132003

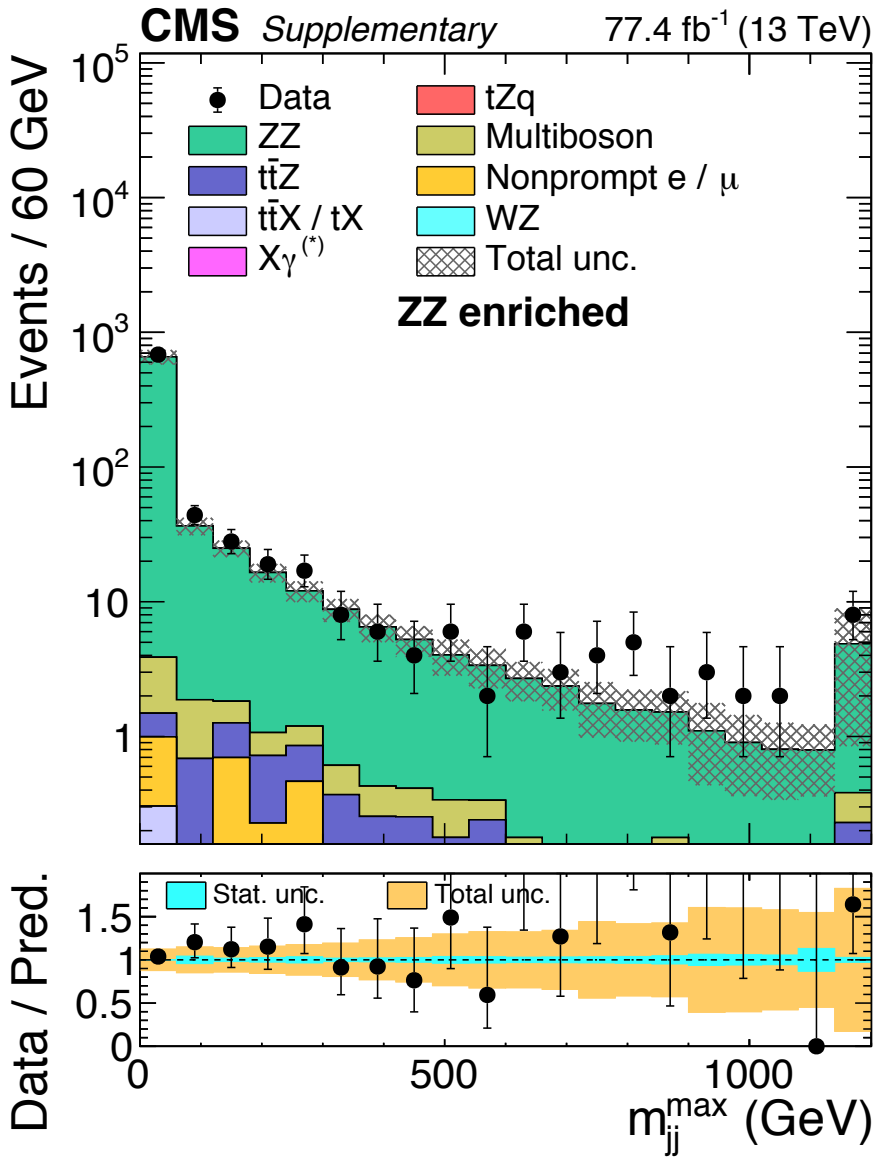




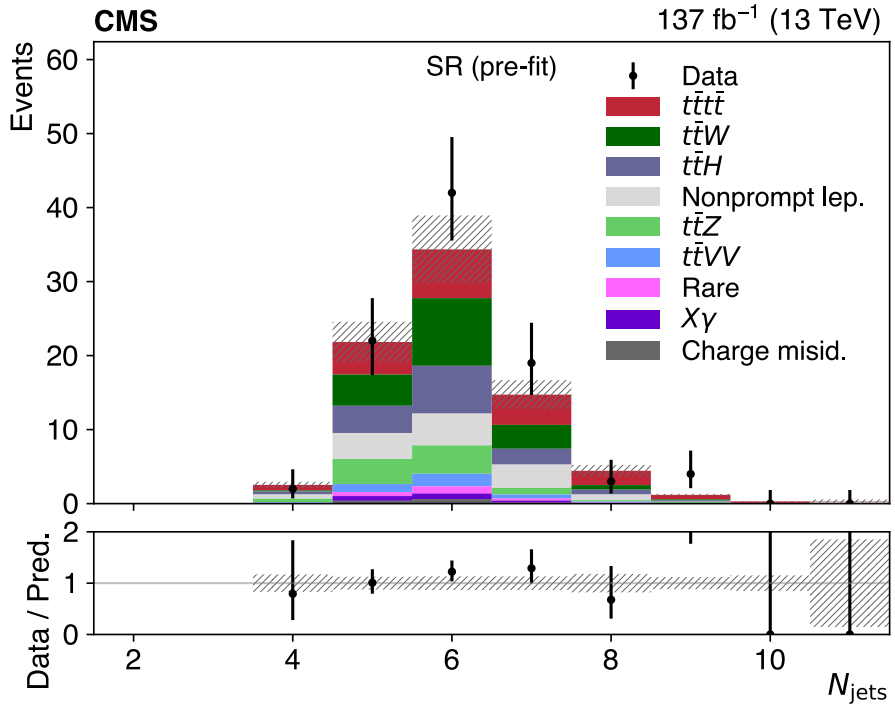
Single top + Z (tZq): diboson regions

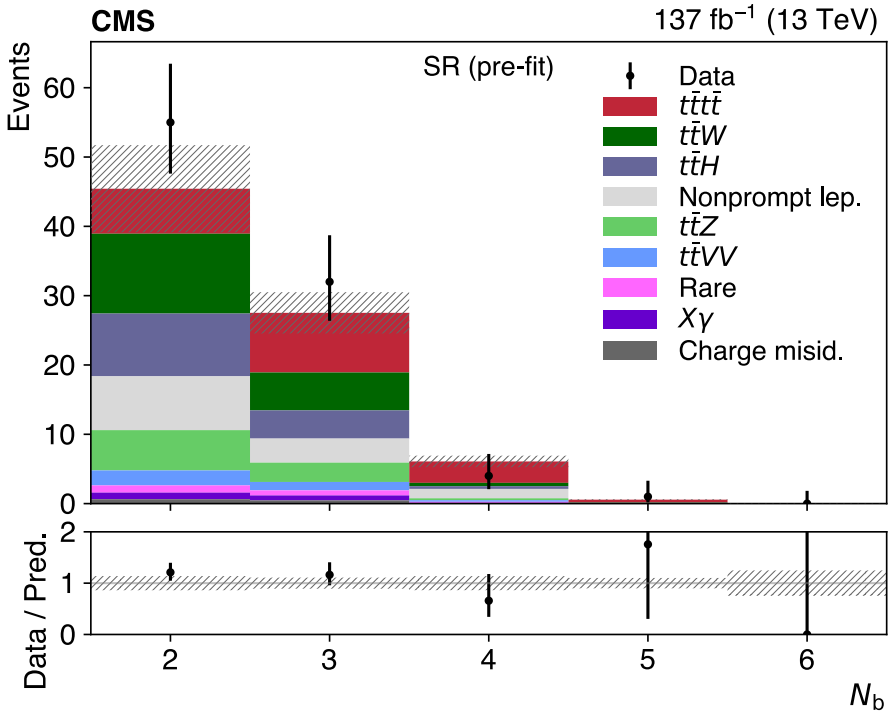
Phys. Rev. Lett. 122 (2019) 132003





Four top multilepton: number of jets





Four top multilepton: BDT details

BDT variables

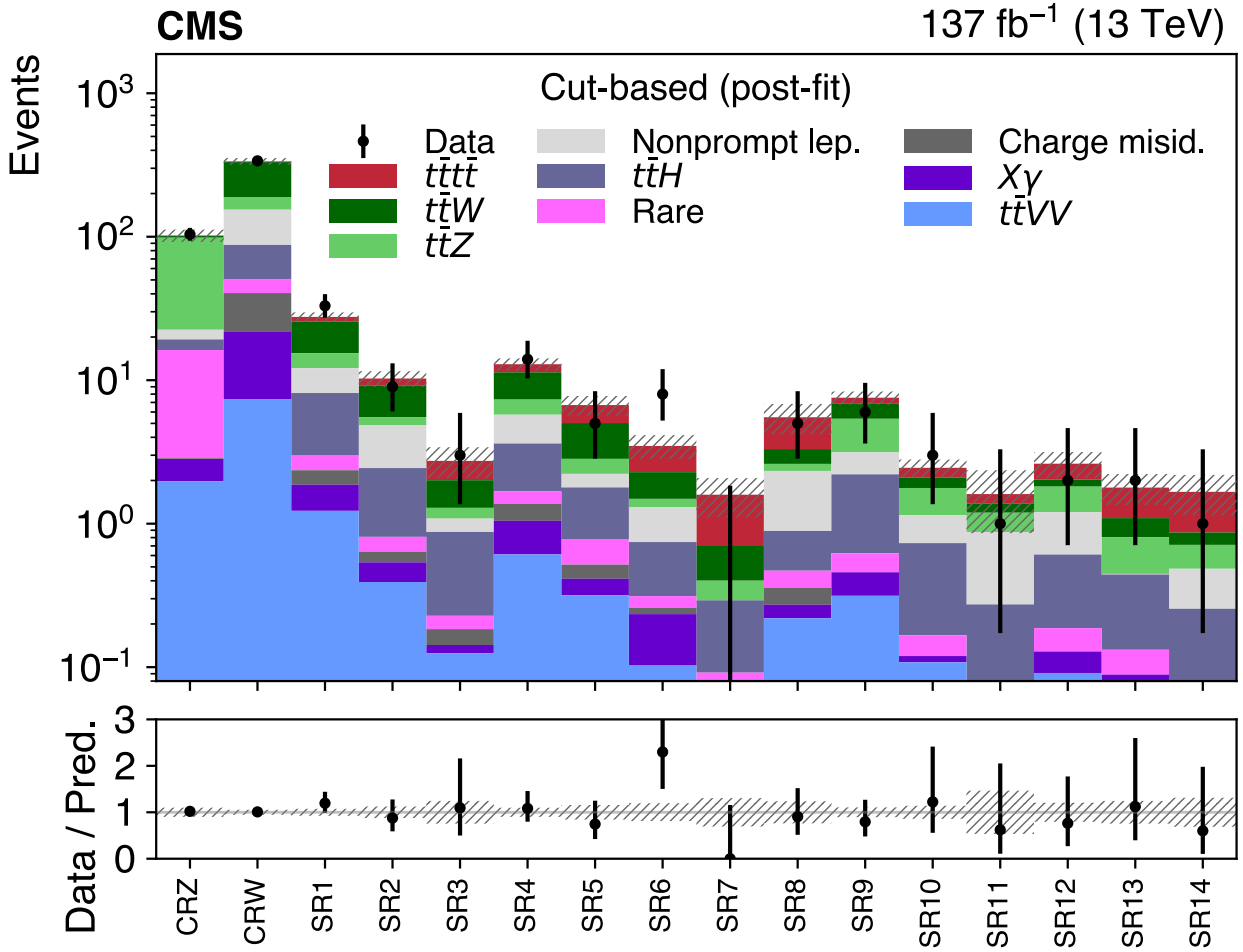
The BDT classifier utilizes a gradient boosting algorithm to train 500 trees with a depth of 4 using simulation, and is based on the following 19 variables: N_{jets}, N_{b}, N_{ℓ}, $p_{\mathrm{T}}^{\mathrm{miss}}$, H_{T}, two alternative definitions of N_{b} based on b tagging working points tighter or looser than the default one, the scalar p_{T} sum of b-tagged jets, the p_{T} of the three leading leptons, of the leading jet and of the sixth, seventh, and eighth jets, the azimuthal angle between the two leading leptons, the invariant mass formed by the leading lepton and the leading jet, the charge of the leading lepton, and the highest ratio of the jet mass to the jet p_{T} in the event (to provide sensitivity to boosted, hadronically-decaying top quarks and W bosons). Three of the most performant input variables, N_{jets}, N_{b}, and N_{ℓ}, correspond to the variables used for the cut-based analysis. Top quark tagging algorithms to identify hadronically

Four top multilepton: Cut-based analysis

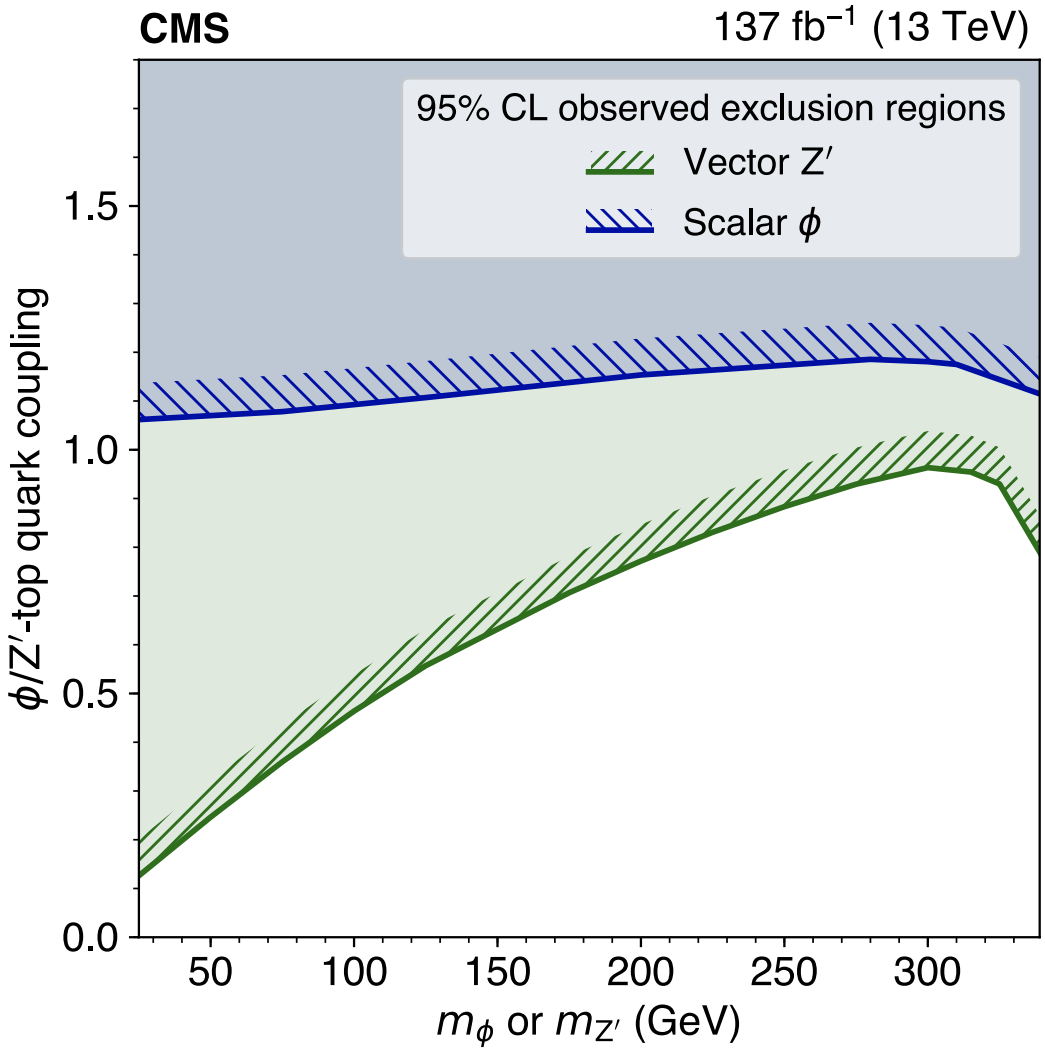
Cut-based categories

N_ℓ	N_b	N_{jets}	Region
2	2	≤ 5	CRW
		6	SR1
		7	SR2
		≥ 8	SR3
	3	5	SR4
		6	SR5
		7	SR6
		≥ 8	SR7
	≥ 4	≥ 5	SR8
≥ 3	2	5	SR9
		6	SR10
		≥ 7	SR11
	≥ 3	4	SR12
		5	SR13
		≥ 6	SR14
Inverted resonance veto			CRZ

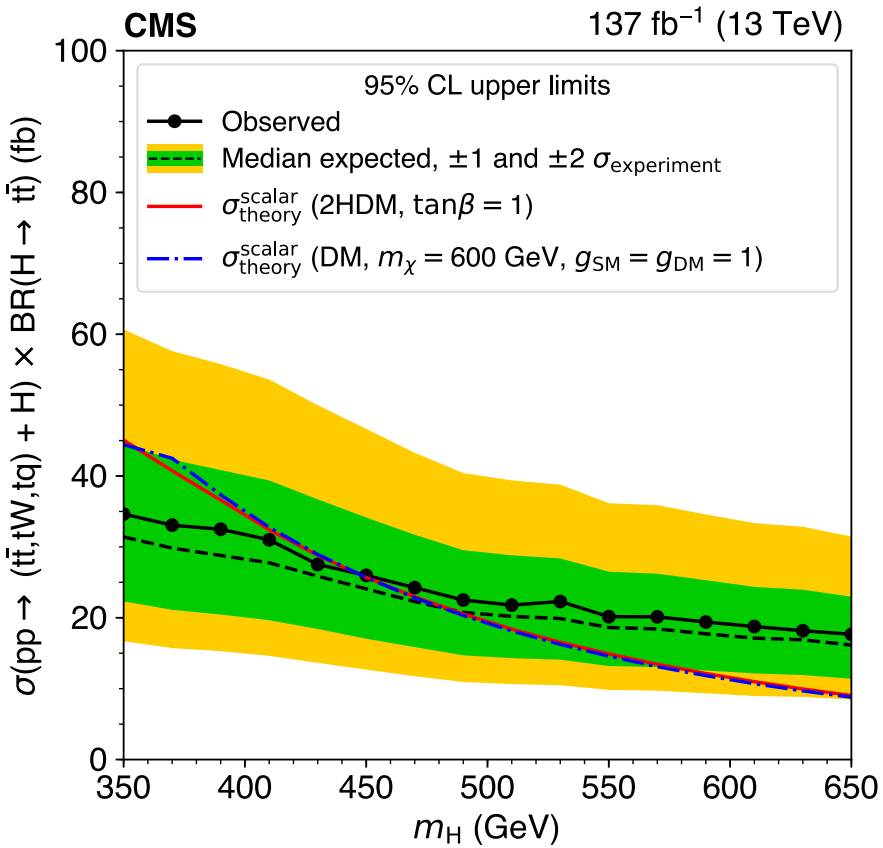
Post-fit yields

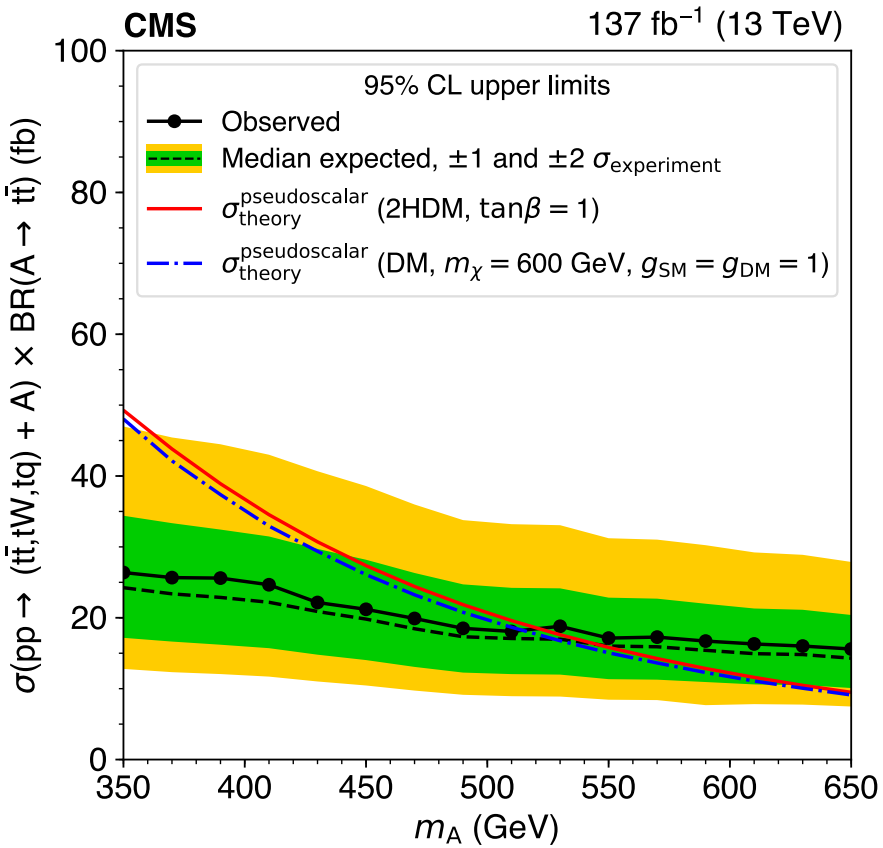


Four top multilepton: interpretations

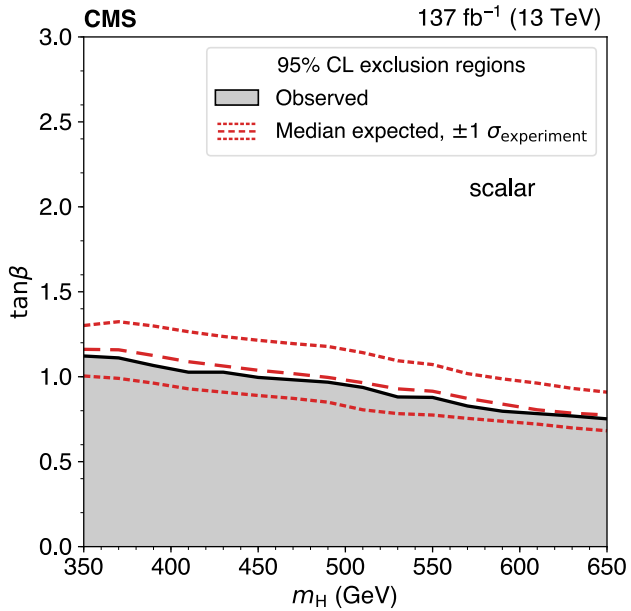


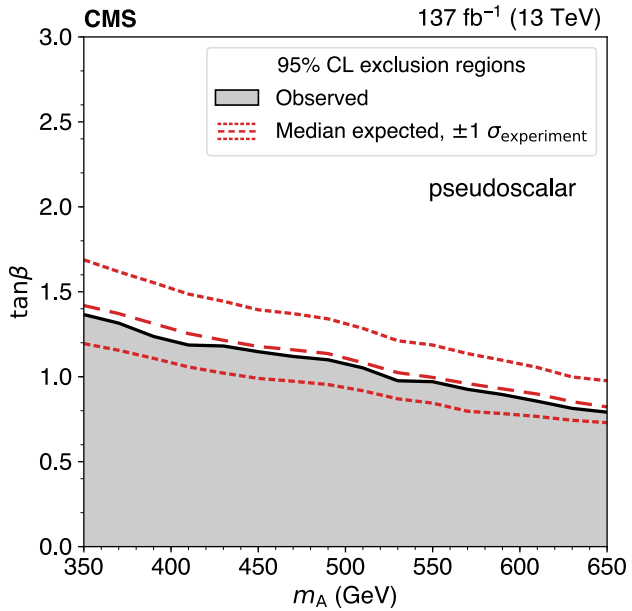
Four top multilepton: interpretations

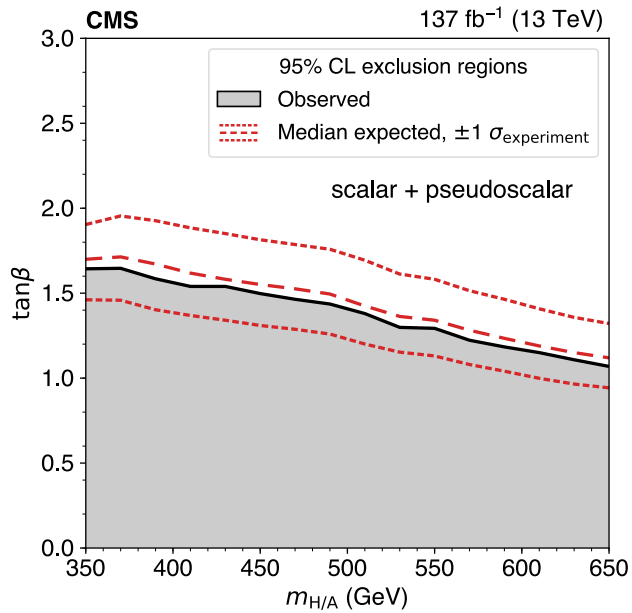




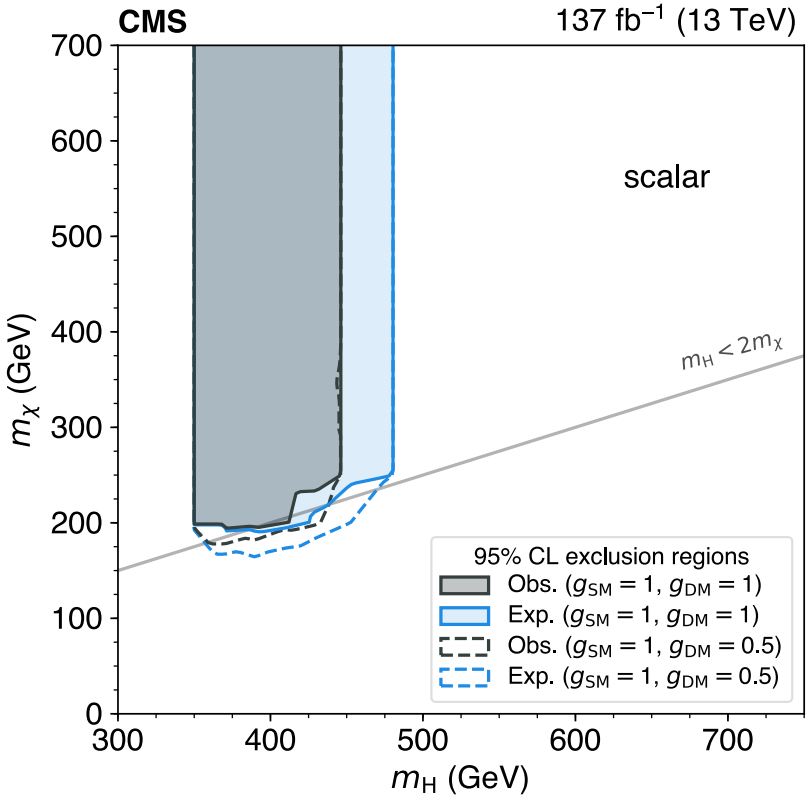
Four top multilepton: interpretations

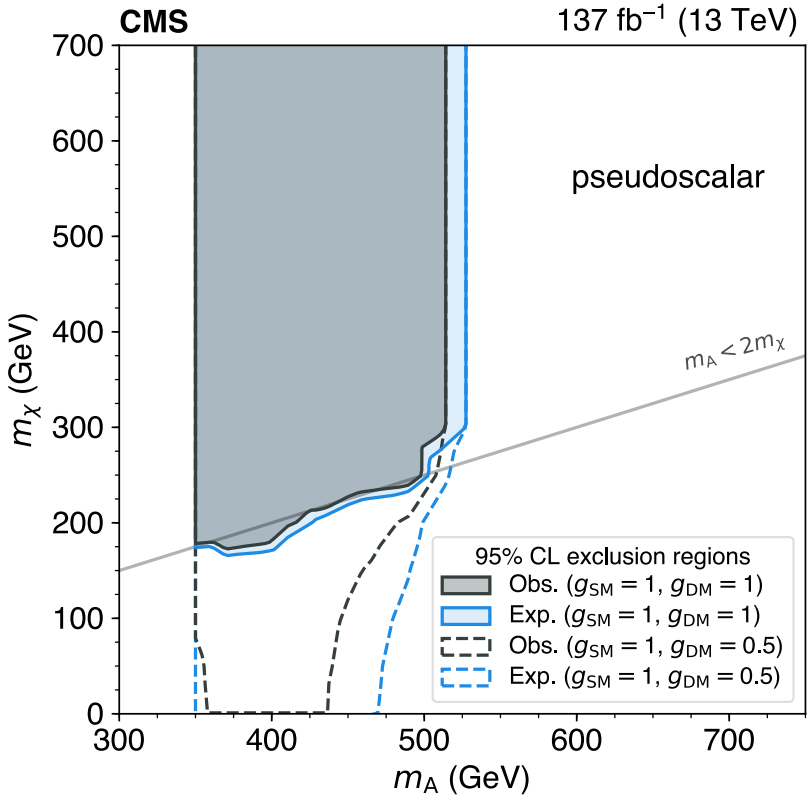




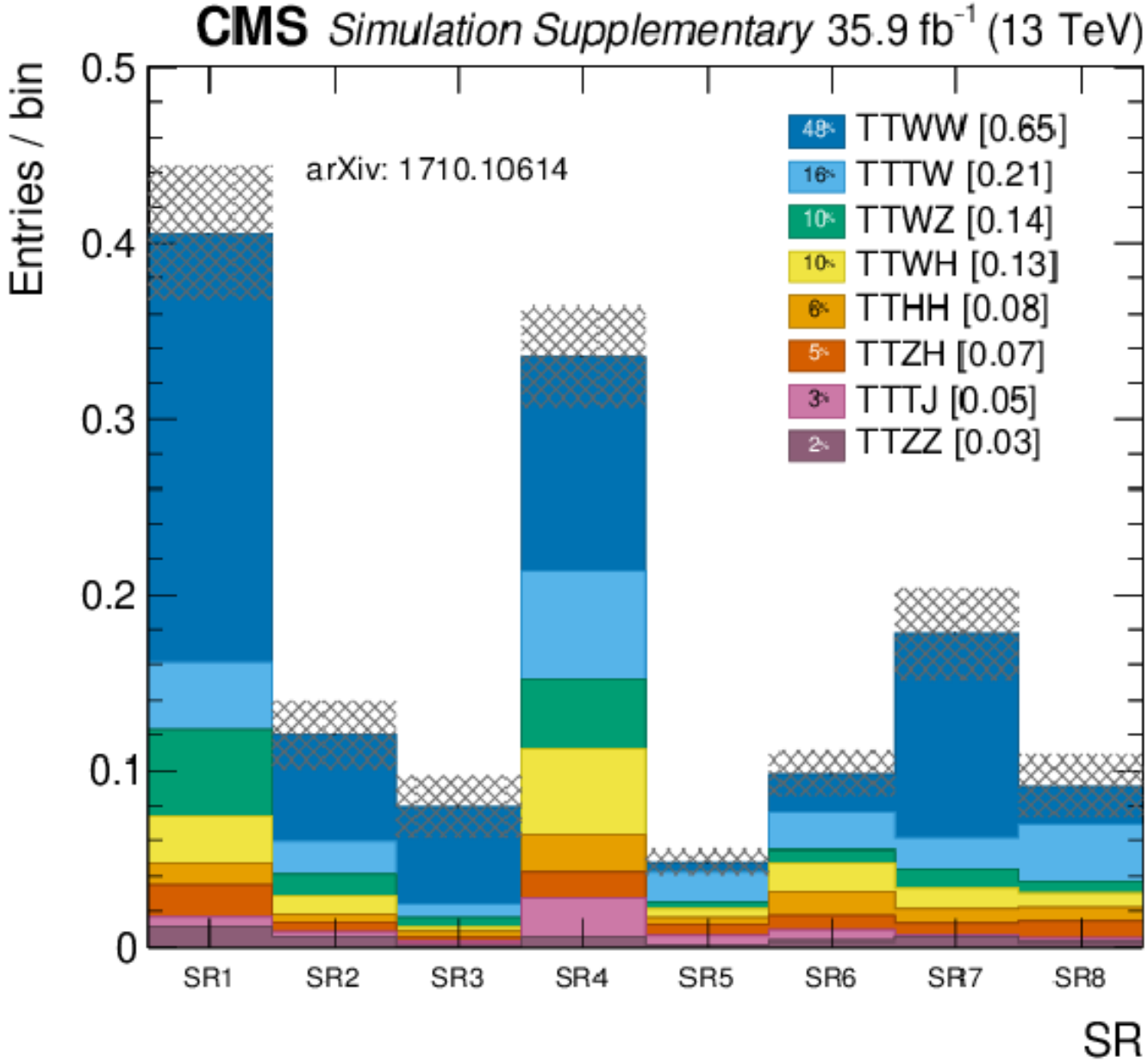


Four top multilepton: interpretations

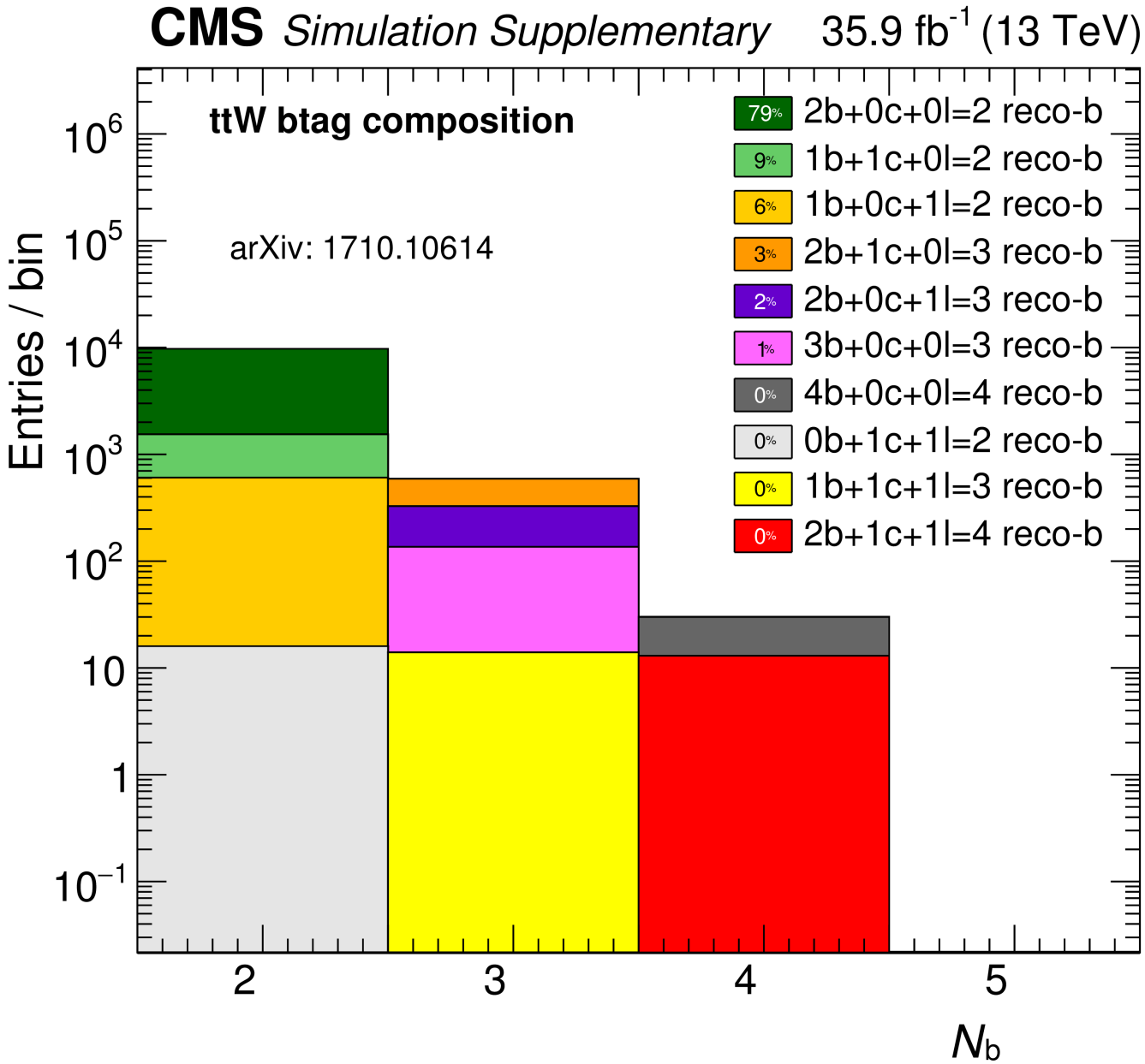




Four top multilepton: rare backgrounds



Four top multilepton: ttW background



Four top 1l and 2lss: BDT details

1. The number of jets present in the event, N_{j}.
2. The scalar sum of the p_{T} of all medium working point b jets in the event, $H_{\mathrm{T}}^{\mathrm{b}}$.
3. The ratio of the sum H_{T} of the four highest p_{T} jets in the event in the single-lepton channel, or the two jets with the highest b tagging discriminant in the dilepton channel, to the H_{T} of the other jets in the event, $H_{\mathrm{T}}^{\mathrm{ratio}}$.
4. The H_{T} sum in the event, subtracting the scalar p_{T} sum of the two highest p_{T} b jets, $H_{\mathrm{T}}^{2\mathrm{m}}$.
5. The transverse momenta of the jets with the third- and fourth-largest p_{T} in the event, $p_{\mathrm{T}}^{\mathrm{j3}}$ and $p_{\mathrm{T}}^{\mathrm{j4}}$.
6. The reduced event mass, $M_{\mathrm{red}}^{\mathrm{h}}$, defined as the invariant mass of the system comprising all the jets in the reduced event, where the reduced event is constructed by removing the jets contained in T_{trijet1} in single-lepton events. In $\mathrm{t\bar{t}}$ events, the reduced event will typically only contain the b jet from the semileptonic top quark decay and jets arising from ISR and FSR. Conversely, a reduced $\mathrm{t\bar{t}t\bar{t}}$ event can contain up to two hadronically decaying top quarks and, as a result, a relatively high reduced event mass.
7. The reduced event H_{T}, $H_{\mathrm{T}}^{\mathrm{x}}$, is defined as the H_{T} of all jets in the single-lepton event selection excluding those contained in T_{trijet1}.

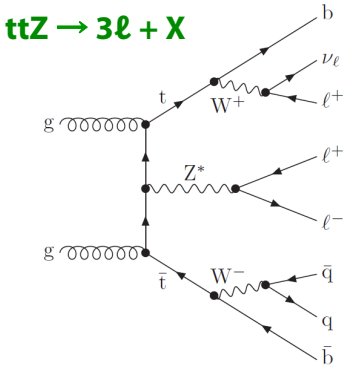
The event topology is characterized by the two variables:

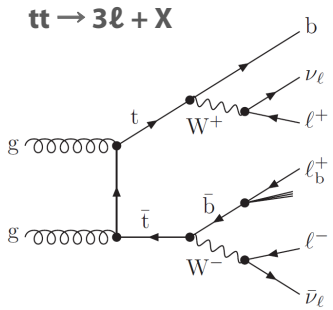
1. Event sphericity, S, [72], calculated from all of the jets in the event in terms of the normalized tensor $M^{\alpha\beta} = \sum_i p_i^{\alpha} p_i^{\beta} / \sum_i |\vec{p_i}|^2$, where α and β refer to the three-components of the momentum of the ith jet. The sphericity is defined as $S = (3/2)(\lambda_2 + \lambda_3)$, where λ_2 and λ_3 are the two smallest eigenvalues of $M^{\alpha\beta}$. The sphericity in $\mathrm{t\bar{t}t\bar{t}}$ events should differ from that in background $\mathrm{t\bar{t}}$ events of the same energy, since the jets in $\mathrm{t\bar{t}}$ events will be less isotropically distributed because of their recoil from sources such as ISR.
2. Hadronic centrality, C, defined as the value of H_{T} divided by the sum of the energies of all jets in the event.

Since all these variables rely only on the hadronic information in the event, sensitivity to the lepton information is provided through the p_{T} and η of the highest p_{T} lepton (or the only lepton for the single-lepton channel) $\left(p_{\mathrm{T}}^{\ell 1}, \eta^{\ell 1}\right)$ and the angular difference $(\Delta R_{\ell\ell})$ between the leptons in dilepton events. The b jet multiplicity is characterized in terms of the number of b jets tagged by the CSVv2 algorithm operating at its loose $(N_{\mathrm{tags}}^{\mathrm{l}})$ and medium $(N_{\mathrm{tags}}^{\mathrm{m}})$ operating points, and the angular separation ΔR_{bb} between the b-tagged jets with the highest CSVv2 discriminants. Finally, the third- and fourth-highest b tagging discriminant values are used as they allow separation between $\mathrm{t\bar{t}}$ +light jets, and genuine additional heavy-flavor jets, as present in $\mathrm{t\bar{t}t\bar{t}}$ events.

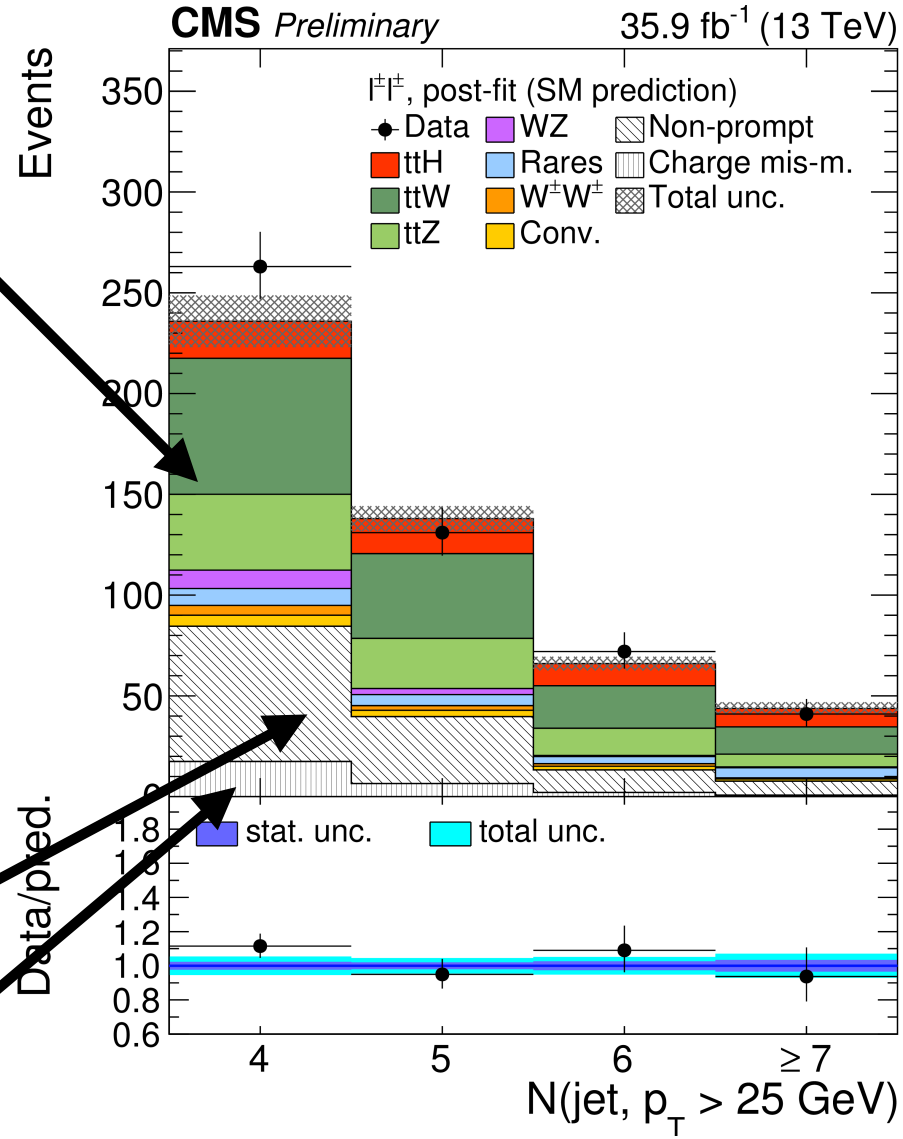
$t\bar{t}H$ multilepton: backgrounds

CMS HIG-17-004





Irreducible: tt+W/Z/γ*
- from Monte Carlo,
- O(10%) uncertainty



Reducible: mainly tt+jets,
- shape obtained from data,
- O(30%) uncertainty

- **Jets faking leptons:** fake rate computed from QCD control region with loosened identification
- **Charge mis-assignment** (2ℓss only)**:** flip rate from $Z \to \ell^\pm\ell^\pm$ data