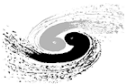
Associated production of tt̄ and heavy flavor at CMS

Fabio Iemmi

on behalf of the CMS Collaboration

Institute of High Energy Physics (IHEP) – Chinese academy of sciences

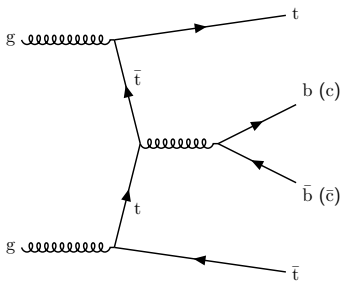
European Physical Society Conference on High Energy Physics – July 29, 2021



Institute of High Energy Physics
Chinese Academy of Sciences

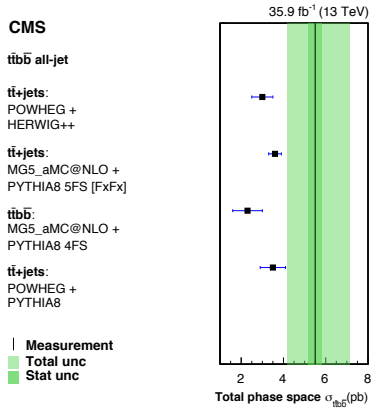
Why $t\bar{t}$+HF?

- $t\bar{t}$+jets production is abundant at the LHC
- **$t\bar{t}$+$b\bar{b}$** is very **challenging to model**
  - Very different energy scale between $t\bar{t}$ and $b\bar{b}$ systems
- **$t\bar{t}$+$b\bar{b}$** is a **background for many** interesting **searches**
  - Irreducible background in $t\bar{t}H(b\bar{b})$
  - Background for $t\bar{t}t\bar{t}$ searches
- Development of **new charm taggers** [JINST 13 (2018) P05011] made **$t\bar{t}$+$c\bar{c}$ measurements possible**
- Thanks to high integrated luminosity delivered by LHC, **$t\bar{t}t\bar{t}$ searches are now possible too**



Why $t\bar{t}$+HF?

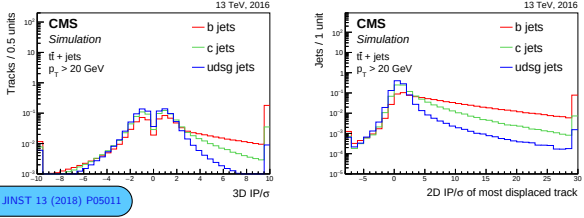
- Some degree of **tension with SM** observed **in previous $t\bar{t}$+$b\bar{b}$** cross section **measurements**
- $t\bar{t}$+$b\bar{b}$ cross section in the all-jets final state
  - Phys. Lett. B 803 (2020) 135285
  - 35.9 fb^{-1} of data
- Use a combination of **MVA methods** to
  - **Reject** dominant **QCD** background
  - **Discriminate between jets** from top quark decays **and additional jets**



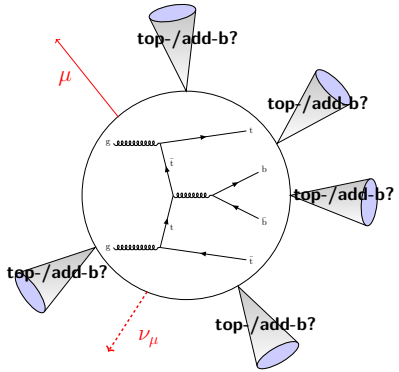
- Predictions from different generators smaller than measured values by a factor 1.5–2.4 (1–2 standard deviations)

Challenges in $t\bar{t}$+HF searches

- **Jet-parton assignment**
  - Final states with **high jet multiplicities**
  - How to correctly match jets with partons is not trivial
- **HF tagging**
  - Relies on mass, $c\tau$ of hadrons, track multiplicity, impact parameters, [...]
  - Particularly challenging for charm jets



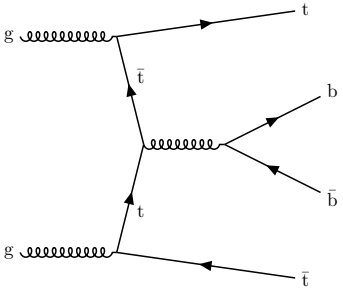
JINST 13 (2018) P05011



- Is a b-jet coming from a top decay or an additional b quark?

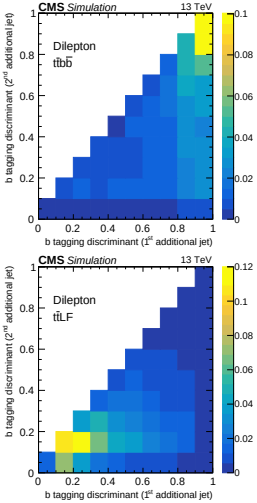
$t\bar{t}$+jets and $t\bar{t}$+$b\bar{b}$ cross sections (JHEP 07 (2020) 125)

- **Goal**: measure $\sigma_{t\bar{t}b\bar{b}}$, $\sigma_{t\bar{t}jj}$ and their ratio $R_{t\bar{t}b\bar{b}/t\bar{t}jj} = \sigma_{t\bar{t}b\bar{b}}/\sigma_{t\bar{t}jj}$
- Use 35.9 fb^{-1} of data
- Perform separate measurements in **dilepton** (DL) and **single-lepton** (SL) **$t\bar{t}$ final states**
- Visible ("fiducial") phase space (VPS) different for the two final states: additional info for testing theory predictions
- **Events** in VPS **split in categories** based on jet flavors ($t\bar{t}b\bar{b}$, $t\bar{t}bj$, $t\bar{t}c\bar{c}$, $t\bar{t}LF$)
- **Main backgrounds**: $t\bar{t}$ events not falling in any $t\bar{t}$+jj category, single-top quark events



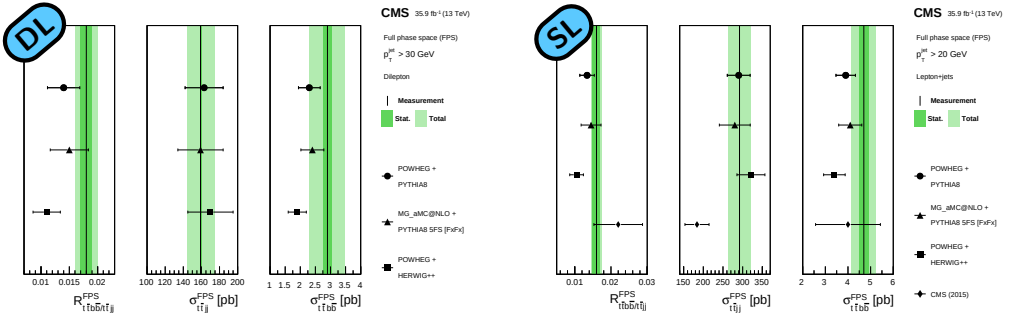
Signal extraction

- DL channel: **jet-parton assignment using b tagging** discriminant
- SL channel: **jet-parton assignment using kinematic fit** constrained to $t\bar{t}$ production
- Extract signal with a ML fit to the b tagging discriminant of additional jets
- **DL and SL channels** fitted **independently**
- Within channels, **categories** are **fitted simultaneously**
- Get $\sigma_{t\bar{t}jj}$ and $R_{t\bar{t}b\bar{b}/t\bar{t}jj}$ from fit; then compute $\sigma_{t\bar{t}b\bar{b}}$ as their product



Results

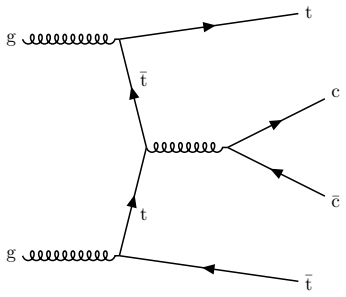
- **Results** in VPS extrapolated to **full phase space** using acceptance corrections
- **Dominant systematics**: FSR, ME-PS matching, b tagging efficiency



- POWHEG + HERWIG++ prediction **slightly lower than measured** values
- $\sigma_{t\bar{t}b\bar{b}}$, $\sigma_{t\bar{t}jj}$ and $R_{t\bar{t}b\bar{b}/t\bar{t}jj}$ measured with precisions between 9 – 18%

$t\bar{t}$+$c\bar{c}$ cross section (arXiv:2012.09225, submitted to PLB)

- **Goal**: measure $\sigma_{t\bar{t}b\bar{b}}$, $\sigma_{t\bar{t}c\bar{c}}$, $\sigma_{t\bar{t}LL}$ and their ratio to $\sigma_{t\bar{t}jj}$
- **First measurement** ever **of $\sigma_{t\bar{t}c\bar{c}}$ and $R_c = \sigma_{t\bar{t}c\bar{c}}/\sigma_{t\bar{t}jj}$**
- Use 41.5 fb^{-1} of data
- Measurement performed in the **DL $t\bar{t}$ final state**
- **Events** in VPS **split in categories** based on jet flavors ($t\bar{t}b\bar{b}$, $t\bar{t}bL$, $t\bar{t}c\bar{c}$, $t\bar{t}cL$, $t\bar{t}LL$)
- **Main backgrounds**: Drell-Yan events, single-top quark events



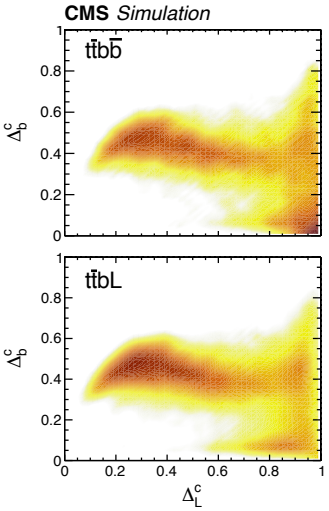
Signal extraction

- **HF tagging**: combine multiclass output of DeepCSV to obtain **CvsL and CvsB discriminators**
- **Jet-parton assignment trough DNN**, trained to identify correct permutation of jets
- Train 2nd multiclass **DNN to predict probabilities** $P(t\bar{t}b\bar{b})$, $P(t\bar{t}bL)$, $P(t\bar{t}c\bar{c})$, $P(t\bar{t}cL)$, $P(t\bar{t}LL)$
- Extract signal with ML fit to

$$\Delta_{\mathrm{b}}^{\mathrm{c}} = \frac{P(t\bar{t}c\bar{c})}{P(t\bar{t}c\bar{c}) + P(t\bar{t}b\bar{b})}$$

$$\Delta_{\mathrm{L}}^{\mathrm{c}} = \frac{P(t\bar{t}c\bar{c})}{P(t\bar{t}c\bar{c}) + P(t\bar{t}LL)}$$

for each category and lepton flavor

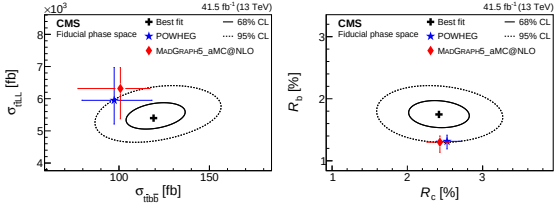


Results

- **Results** in VPS extrapolated to **full phase space** using acceptance corrections
- **Dominant systematics**: c-tagging scale factor, JES, ME-PS matching

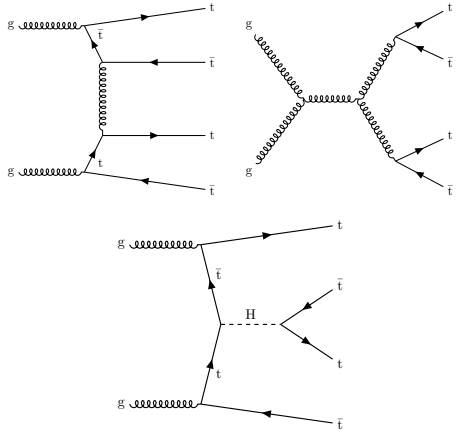
	Result	POWHEG	MadGraph5_aMC@NLO
$\sigma_{\mathrm{t\bar{t}c\bar{c}}}$ [pb]	$8.0 \pm 1.1 \pm 1.3$	9.1 ± 1.8	8.9 ± 1.5
R_{c} [%]	$2.69 \pm 0.36 \pm 0.32$	2.81 ± 0.20	2.72 ± 0.19

- Also perform **2D likelihood scans**
- Measurements and predictions in **agreement within 1–2 σ**
- $\sigma_{\mathrm{t\bar{t}b\bar{b}}}$ and $R_{\mathrm{b}} = \sigma_{\mathrm{t\bar{t}b\bar{b}}}/\sigma_{\mathrm{t\bar{t}jj}}$ slightly above the predictions (in agreement with JHEP 07 (2020) 125)



$t\bar{t}t\bar{t}$ cross section (Eur. Phys. J. C 80 (2020) 75)

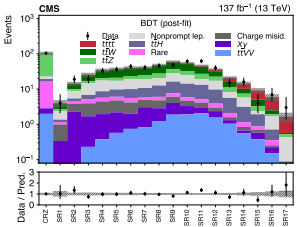
- **Goal**: measure $\sigma_{t\bar{t}t\bar{t}}$
- **Very rare** SM process ($\sigma^{SM}_{t\bar{t}t\bar{t}} \simeq 12$ fb), **yet unobserved** at the LHC
- Use 137 fb^{-1} of data (full RunII analysis)
- **Targeting** the **two-same-sign-leptons** (2SSL) **and multileptons** (ML) final states
- Measurement is **interpreted in several BSM scenarios** (Type-II 2HDM, simplified dark matter models, EFT)
- **Main backgrounds**: $t\bar{t}V$, $t\bar{t}H$, DL and SL $t\bar{t}$ events

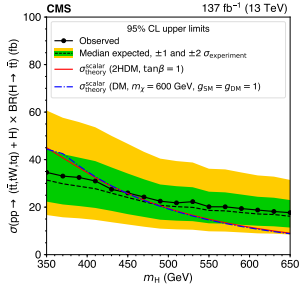


Results

- **BDT to separate $\mathrm{t\bar{t}t\bar{t}}$ from background**; discretize the output to get 17 signal regions (SRs)
- Define $\mathrm{t\bar{t}Z}$-enriched CR (CRZ)
- Extract signal with a fit to SRs and CRZ
- **Dominant systematics**: JES, JER, b tagging scale factors
- $$\boxed{\sigma^{\mathrm{obs}}_{\mathrm{t\bar{t}t\bar{t}}} = 12.6^{+5.8}_{-5.2}\ \mathrm{fb}}$$

  observed (expected) significance of 2.6 (2.7) σ
- Set **limits on the production of heavy scalar** H
- Exclude the mass range $350 < \mathrm{m_H} < 470$ GeV
  - More than **100 GeV improvement** wrt previous CMS measurements





Conclusions

- **$t\bar{t}$+HF measurements** are **important** from both theoretical and experimental point of view
  - Comparison with measurements can give **guidance to improve the theoretical predictions**
  - **$t\bar{t}$+HF** production **background for many** interesting **searches**
  - **$t\bar{t}t\bar{t}$ measurements** can lead to several **BSM interpretations**
- **CMS is pursuing $t\bar{t}$+HF measurements** in many directions
  - First measurement of $\sigma_{t\bar{t}c\bar{c}}$ and $R_c = \sigma_{t\bar{t}c\bar{c}}/\sigma_{t\bar{t}jj}$
  - Precise measurement of $\sigma_{t\bar{t}b\bar{b}}$, $\sigma_{t\bar{t}jj}$ and $R_{t\bar{t}b\bar{b}/t\bar{t}jj}$
  - $t\bar{t}t\bar{t}$ seraches with BSM interpretations
- More measurements to come, **stay tuned!**

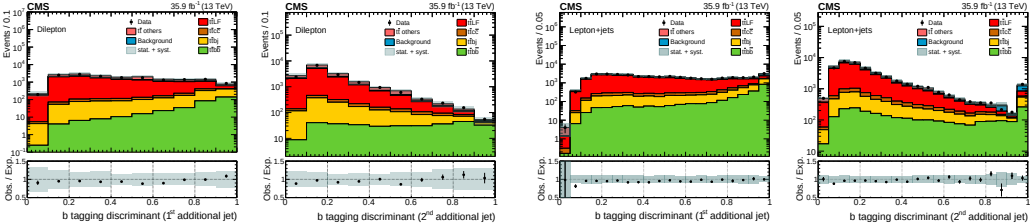
Backup slides
TBD

$t\bar{t}$+jets and $t\bar{t}$+$b\bar{b}$ xsecs: jet-parton assignment

- DL channel
  - Use simulation to study b tagging properties of jets
  - First- and second-leading jets in b tagging score come from top quark decay in 85 (23)% of selected $t\bar{t}jj$ ($t\bar{t}b\bar{b}$) events
  - Thus, **third- and fourth-leading jets are considered to be additional jets**
- SL channel
  - Higher jet multiplicity, b tag based jet-parton assignment impossible
  - **Use a kinematic fit** instead
  - **Inputs**: jets and lepton 4-momenta, btag information, $p_{\mathrm{T}}^{\mathrm{miss}}$
  - **Constraints**: production of 2 top quarks and two W bosons
  - **Compute a χ^2-like function** for each permutation of jets and **select** the one with **lowest χ^2**
  - Jets not included in the solution are additional jets

$t\bar{t}$+jets and $t\bar{t}$+$b\bar{b}$ xsecs: additional jets b tagging

- Amongst additional jets, the two leading jets in b tagging score are kept

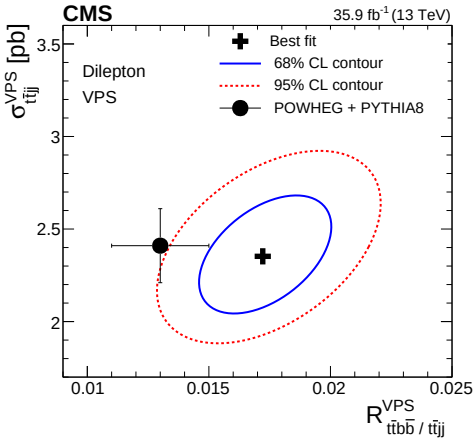


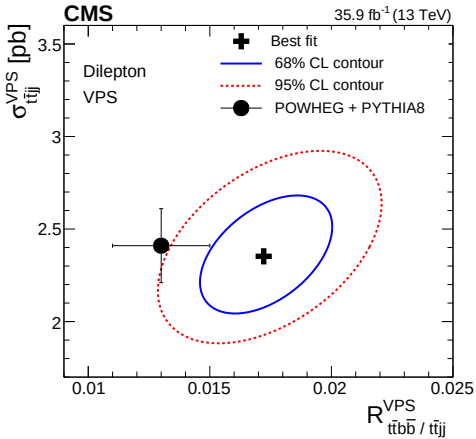
$t\bar{t}$+jets and $t\bar{t}$+$b\bar{b}$ xsecs: systematic uncertainties

Source	$R^{\mathrm{VPS}}_{t\bar{t}b\bar{b}/t\bar{t}jj}$ [%]		$\sigma^{\mathrm{VPS}}_{t\bar{t}jj}$ [%]	
	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 btag (lin.)	2.2	2.0	1.0	0.3
c-flavor btag (quad.)	0.7	1.2	0.3	0.2
Heavy-flavor btag	4.0	0.1	0.5	0.9
Heavy-flavor btag (lin.)	0.9	0.4	1.5	0.5
Heavy-flavor btag (quad.)	2.0	0.3	1.5	0.8
Light-flavor btag	4.9	0.9	5.5	4.9
Light-flavor btag (lin.)	0.1	0.2	0.3	1.1
Light-flavor btag (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 $R^{\mathrm{MC}}_{t\bar{t}bj/t\bar{t}b\bar{b}}$	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	8.0	5.5	8.8	10.0

$t\bar{t}$+jets and $t\bar{t}$+$b\bar{b}$ xsecs: fit results

- ML fit extracts $\sigma_{t\bar{t}jj}$ and $R_{t\bar{t}b\bar{b}/t\bar{t}jj}$





$t\bar{t}$+jets and $t\bar{t}$+$b\bar{b}$ xsecs: $\sigma_{t\bar{t}b\bar{b}}$

	$R_{t\bar{t}b\bar{b}/t\bar{t}jj}$	$\sigma_{t\bar{t}jj}$ [pb]	$\sigma_{t\bar{t}b\bar{b}}$ [pb]
	Dilepton channel (VPS)		
POWHEG + PYTHIA8	0.013 ± 0.002	2.41 ± 0.21	0.032 ± 0.004
Measurement	$0.017 \pm 0.001 \pm 0.001$	$2.36 \pm 0.02 \pm 0.20$	$0.040 \pm 0.002 \pm 0.005$
	Dilepton channel (FPS)		
POWHEG + PYTHIA8	0.014 ± 0.003	163 ± 21	2.3 ± 0.4
MG_aMC@NLO + PYTHIA8 5FS [FxFx]	0.015 ± 0.003	159 ± 25	2.4 ± 0.4
POWHEG + HERWIG++	0.011 ± 0.002	170 ± 25	1.9 ± 0.3
Measurement	$0.018 \pm 0.001 \pm 0.002$	$159 \pm 1 \pm 15$	$2.9 \pm 0.1 \pm 0.5$
	Lepton+jets channel (VPS)		
POWHEG + PYTHIA8	0.017 ± 0.002	30.5 ± 3.0	0.52 ± 0.06
Measurement	$0.020 \pm 0.001 \pm 0.001$	$31.0 \pm 0.2 \pm 2.9$	$0.62 \pm 0.03 \pm 0.07$
	Lepton+jets channel (FPS)		
POWHEG + PYTHIA8	0.013 ± 0.002	290 ± 29	3.9 ± 0.4
MG_aMC@NLO + PYTHIA8 5FS [FxFx]	0.014 ± 0.003	280 ± 40	4.1 ± 0.4
POWHEG + HERWIG++	0.011 ± 0.002	321 ± 36	3.4 ± 0.5
Measurement	$0.016 \pm 0.001 \pm 0.001$	$292 \pm 1 \pm 29$	$4.7 \pm 0.2 \pm 0.6$

$t\bar{t}$+$c\bar{c}$ cross section: jet-parton assignment

- Consider all permutations of 4 jets passing selection
- DNN-based choice of best permutation
- Input variables: jet kinematic variables, b and c tagging discriminators, angular separations and invariant masses of jet pairs
- In assignment of b jets to top quark decays, it does not matter which jet is matched to t or $\bar{\mathrm{t}}$
- Thus, three output probabilities: probability for correct assignment (P^{+}), probability for correct matching but b jets from $t\bar{t}$ reversed ($P^{\times}$), probability for a wrong assignment (P^{-})
- Select permutation with highest value of

$$\max\left(\frac{P^{+}}{P^{+}+P^{-}}, \frac{P^{\times}}{P^{\times}+P^{-}}\right)$$

$t\bar{t}$+$c\bar{c}$ cross section: charm jet identification

- Use DeepCSV multiclass tagger to predict probability for a jet to contain a single b hadron ($P(\mathrm{b})$), two b hadrons ($P(\mathrm{bb})$), one or more c hadrons ($P(\mathrm{c})$), no b or c hadrons ($P(\mathrm{udsg})$)
- Define two discriminators as

$$\mathrm{CvsL} = \frac{P(\mathrm{c})}{P(\mathrm{c}) + P(\mathrm{udsg})}$$

$$\mathrm{CvsB} = \frac{P(\mathrm{c})}{P(\mathrm{c}) + P(\mathrm{b}) + P(\mathrm{bb})}$$

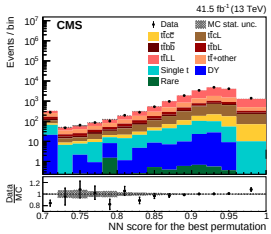
- Correct the CvsL and CvsB distribution to reproduce the distributions observed in data
- Define three control regions enriched in $t\bar{t}$, W+c and DY+jets events and obtain scale factors with iterative procedure

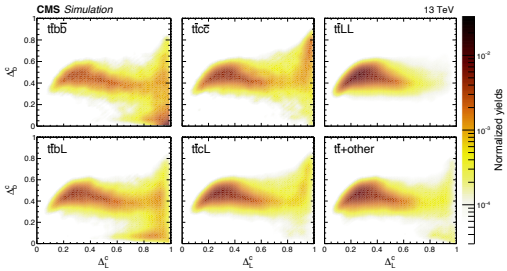
$t\bar{t}$+$c\bar{c}$ cross section: event based NN discriminator

- Train a neural network using CvsL, CvsB of two additional jets, angular separation between additional jets, NN score for best permutation of jets
- NN predicts 5 probabilities: $P(t\bar{t}b\bar{b})$, $P(t\bar{t}bL)$, $P(t\bar{t}c\bar{c})$, $P(t\bar{t}cL)$, $P(t\bar{t}LL)$
- Combine them to

$$\Delta_{\mathrm{b}}^{\mathrm{c}} = \frac{P(t\bar{t}c\bar{c})}{P(t\bar{t}c\bar{c}) + P(t\bar{t}b\bar{b})}$$

$$\Delta_{\mathrm{L}}^{\mathrm{c}} = \frac{P(t\bar{t}c\bar{c})}{P(t\bar{t}c\bar{c}) + P(t\bar{t}LL)}$$





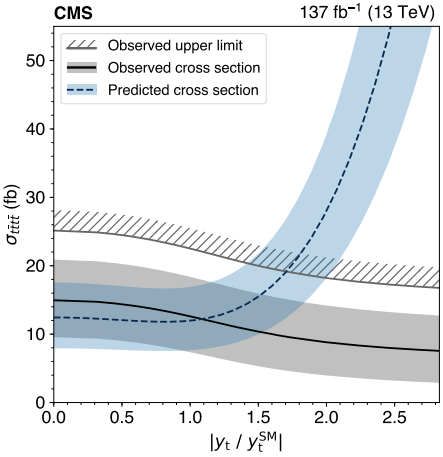
$t\bar{t}$+$c\bar{c}$ cross section: systematic uncertainties

Sources	Systematic uncertainty (%)				
	$\Delta\sigma_{t\bar{t}c\bar{c}}$	$\Delta\sigma_{t\bar{t}b\bar{b}}$	$\Delta\sigma_{t\bar{t}LL}$	ΔR_c	ΔR_b
Jet energy scale	7.9	4.0	5.4	3.5	4.6
Jet energy resolution	2.1	0.8	1.1	2.7	1.4
c tagging calibration	8.1	3.9	1.8	8.0	4.4
Lepton identification and isolation	1.2	1.2	1.1	0.3	0.4
Trigger	2.0	2.0	2.0	<0.1	<0.1
Pileup	0.7	0.7	0.4	1.0	0.4
Total integrated luminosity	2.3	2.3	2.3	<0.1	<0.1
μ_R and μ_F scales in ME	6.2	7.4	1.5	5.7	7.5
PS scale	0.8	0.9	0.2	0.5	0.6
PDF	0.5	0.1	0.1	0.4	0.1
ME-PS matching	7.5	5.7	4.0	2.7	1.4
Underlying event	1.9	2.0	0.9	0.8	1.0
b fragmentation	0.4	3.6	0.3	0.4	3.6
$t\bar{t}bL(cL)/t\bar{t}b\bar{b}(c\bar{c})$ and $t\bar{t}$+other/$t\bar{t}LL$	2.4	1.7	1.2	1.9	1.4
Efficiency (theoretical)	2.0	2.0	2.0	<0.1	<0.1
Simulated sample size	4.4	3.0	1.3	4.3	3.1
Background normalization	0.6	<0.1	0.4	0.3	0.4
Total	15.1	12.6	8.3	12.0	10.3

$t\bar{t}t\bar{t}$ analysis: constraints on y_t

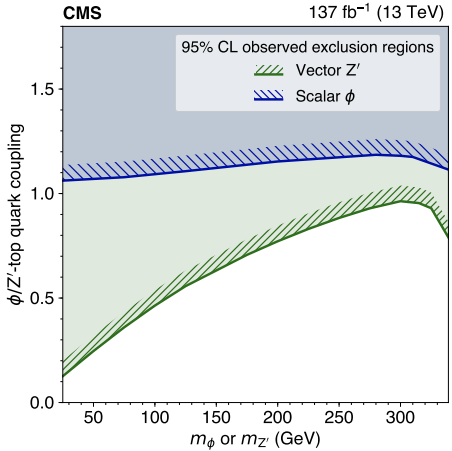
- Account for the dependence of backgrounds on y_t by rescaling the $t\bar{t}H$ cross section by $|y_t/y_t^{SM}|$
- As a result, $\sigma_{t\bar{t}t\bar{t}}$ depends on $|y_t/y_t^{SM}|$
- Compare the observed upper limit with central (upper, lower) theoretical predictions:

$$|y_t/y_t^{SM}| < 1.7\ (1.4, 2.0)$$



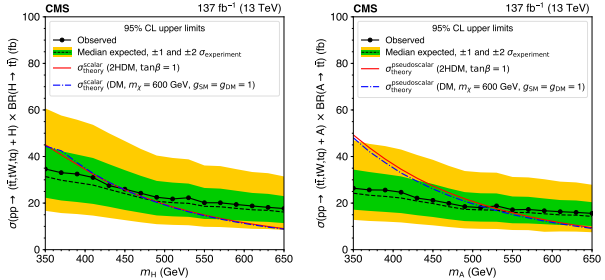
$t\bar{t}t\bar{t}$ analysis: off-shell effects of new particles with $m < 2m_t$

- BSM models predict the presence of virtual scalar (ϕ) or vector (Z') with $m < 2m_t$ coupling with top quarks
- At LO, these particles affect acceptance by $\approx 10\%$
- Add 10% uncertainty on acceptance and calculate upper limit on $\sigma_{t\bar{t}t\bar{t}}$
- Compare upper limit with models including ϕ and Z', get limits on masses and couplings of these particles



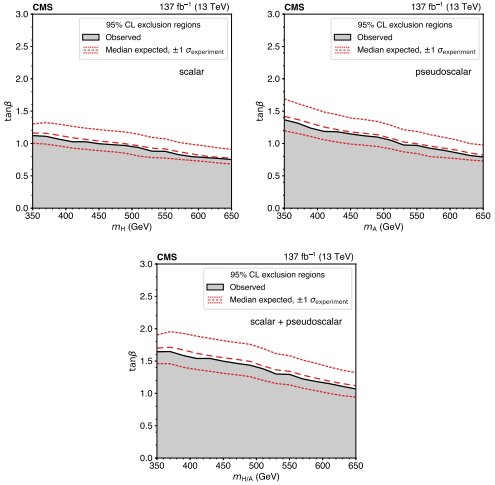
$\mathrm{t\bar{t}t\bar{t}}$ analysis: on-shell effects of new particles with $\mathrm{m} > 2\mathrm{m_t}$

- BSM models predict the presence of new scalar (H) or pseudoscalar (A) particles with $\mathrm{m} > 2\mathrm{m_t}$ coupling with top quarks
- Consider $\mathrm{tqH/A}$, $\mathrm{tWH/A}$, $\mathrm{t\bar{t}H/A}$
- Treat SM $\mathrm{t\bar{t}t\bar{t}}$ as a background



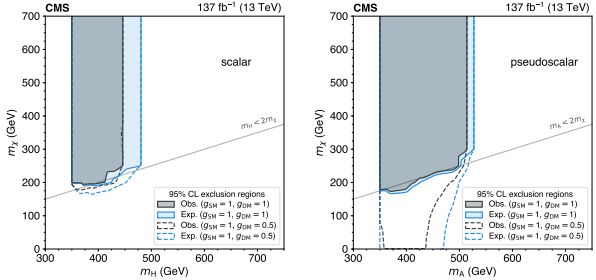
$t\bar{t}t\bar{t}$ analysis: simplified dark matter model

- Simplified dark matter model predicts presence of a Dirac fermion dark matter candidate χ in addition to $\mathrm{H/A}$
- Couplings of $\mathrm{H/A}$ and χ to fermions depend on $\mathrm{g_{SM}}$ and $\mathrm{g_{DM}}$ respectively
- Assuming $\mathrm{g_{SM}} = \mathrm{g_{DM}} = 1$ and taking $\mathrm{m_{H/A}} < 2\mathrm{m}_\chi$
- Get limits on $\tan\beta$ and $\mathrm{m_{H/A}}$
- Exclude $\tan\beta$ up to 0.8 – 1.6 depending on assumptions



$t\bar{t}t\bar{t}$ analysis: simplified dark matter model

- Relaxing $m_{H/A} < 2m_{\chi}$ condition
- Set limits as a function of $m_{H/A}$ and $2m_{\chi}$ for $g_{DM} = 1$ and two different values of g_{SM}



$t\bar{t}t\bar{t}$ analysis: systematic uncertainties

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