Four Top Production at CMS

Nicolas Stylianou *on behalf of the* ***CMS*** *collaboration*

IOP HEPP & APP Annual Conference 2022

University of BRISTOL

HIGH-ENERGY PHYSICS
RESEARCH CENTRE

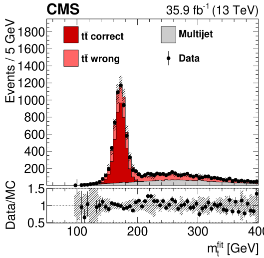
Introduction - Brief History of Top

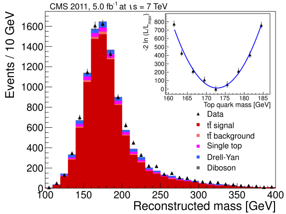
Today
Rarer processes possible

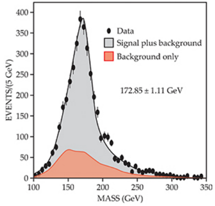
2011
LHC a **top factory**

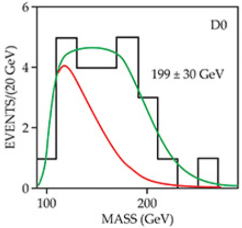
2001-2011
More top candidate events at Tevatron thus **improvements** in precision measurements

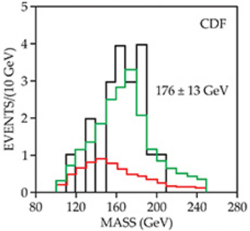
1995
Simultaneous Announcement of **Top Discovery** by CDF and D0 experiments at Fermilab (PRL 74 2626(2632))





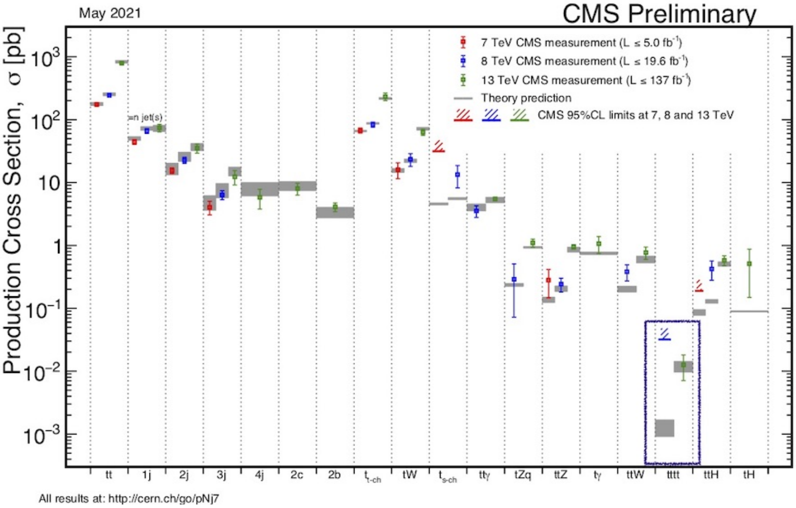






Introduction - Top Processes

A very rare Standard Model process $\sim \mathbf{12^{+2.2}_{-2.5}}$ fb at NLO QCD+Electroweak at 13 TeV (arXiv:1711.02116)

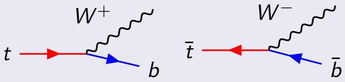


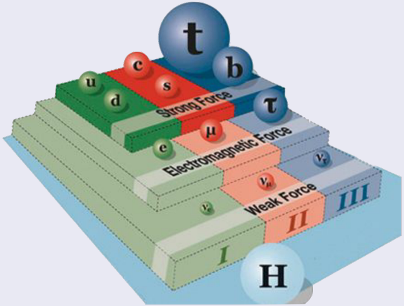
~ 1 **order smaller than** $t\bar{t}H$

Introduction

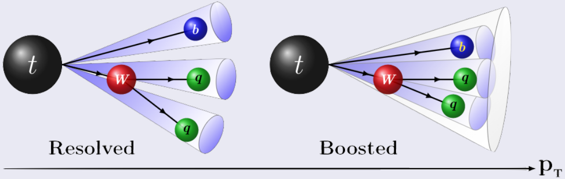
Top Quark

- Most massive elementary particle to date
  - At $\sim$ 40 times the mass of the b-quark
  - Top mass $m_t \sim 173\ \mathrm{GeV/c^2}$
- Only quark to decay before hadronization





- Identification via basic reconstruction is difficult due to background processes
- Needs taggers for resolved and boosted scenarios



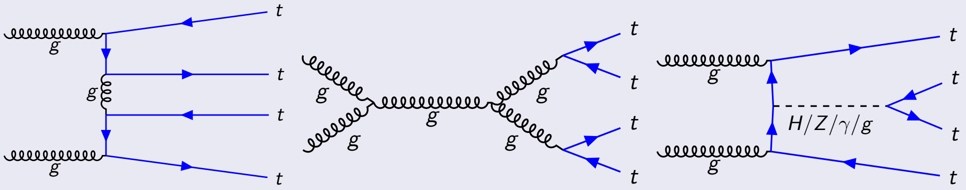
Introduction

Why the Top Quark?

- Standard model predicts top kinematics
- Top physics $\rightarrow$ SM cross check
- Deviations are signs of new physics
- This new physics is at large mass scales $\rightarrow$ making it a good candidate to fix the holes in the SM

Introduction - *Jump A Few Steps Forward*

Standard Model four-top Production Methods



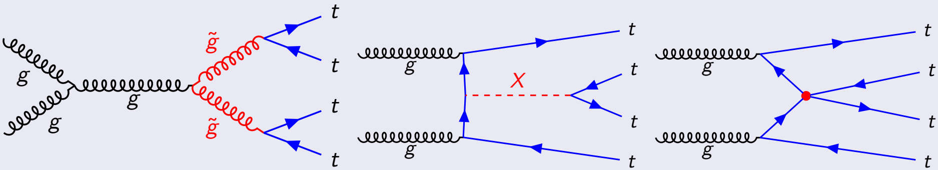
Motivation

- Rare Standard Model process
  - Interesting as a SM-QCD process
    - Includes both 3 point gluon interactions and quark-gluon interaction
  - Provides constraints to the top-Yukawa coupling (y_t)
    - Indicates H-boson must interact with top quarks
- Sensitive to BSM physics

- Leading via QCD $\mathcal{O}(\alpha_S^4)$
- Smaller contributions via Higgs and electroweak $\mathcal{O}(\alpha_S^2 y_t^2)$, $\mathcal{O}(\alpha_S^2 \alpha^2)$
- Predicted $\sigma_{t\bar{t}t\bar{t}}^{SM}$ given at NLO for both QCD and EWK interactions

Introduction

Beyond Standard Model four-top Production Methods



Motivation

- BSM physics:
  - Two-Higgs-Doublet-Model(2HDM): heavy scalar or pseudo-scalar Higgs(H/A) [1]
  - Simplified models of dark matter(DM)
  - Off-shell particles with $m < m_t$
  - Higgs-boson [3] and Top-quark compositeness [4]
  - SUSY (eg. gluino-pair)
  - Extra dimensions (eg. [5,6])
- Interpretation with SM Effective Field Theories (SM-EFTs)

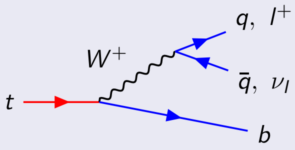
A few possible diagrams are shown, which represent:

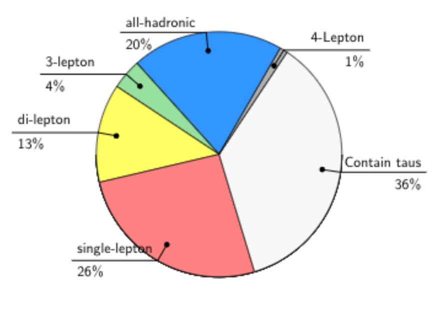
- Rare SUSY example,
- Intermediate scalar and
- Interaction in a SM-EFT

all producing four-tops

Introduction

Possible final states





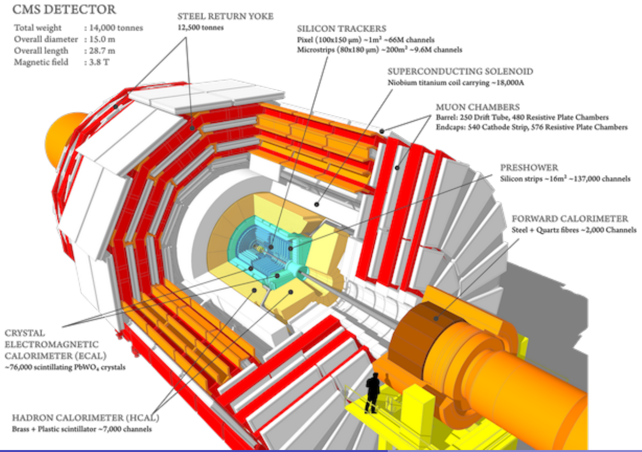
Combination of all final states will prove to yield the best results, but first must be treated separately. Each final state has dominant backgrounds as:

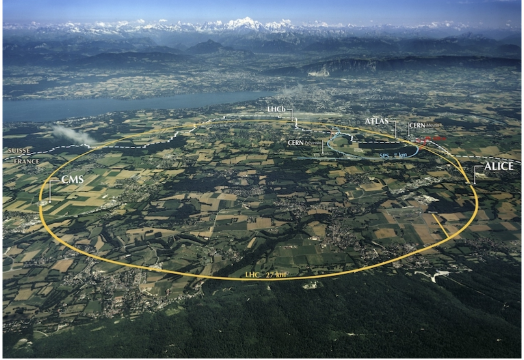
- **0 μ or e:** multijet QCD (difficult as there more backgrounds)
- **1 μ or e:** $t\bar{t} + bb$, $t\bar{t} + cc$
- **2 μ or e:**
  - Same-charge: $t\bar{t}W/Z/H$, dibosons
  - Opposite-charge: $t\bar{t} + bb$, $t\bar{t} + cc$
- **3 μ or e:** $t\bar{t}W/Z/H$, dibosons
- **Contain τ:** Taus have a complicated signature, so we focus on electrons and muons

The CMS detector and the Large Hadron Collider

The LHC:

- Accelerates two beams of protons in opposite directions
- Guides beams to be collided at the interaction points
- Designed to have a 14 TeV centre-of-mass collision energy($\sqrt{s}$)

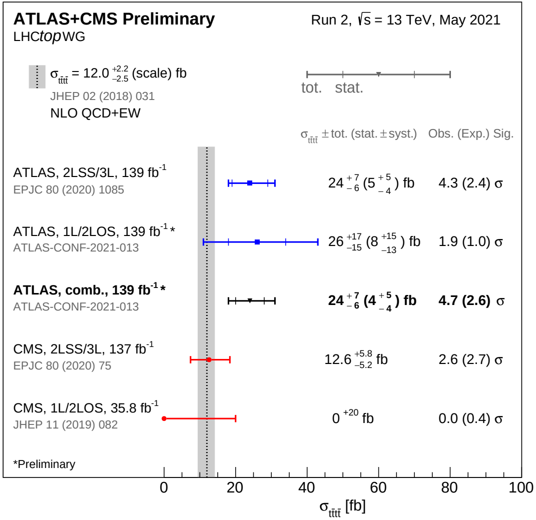




CMS:

- Several sub-detectors essential for four-tops analysis
- During RunII it collected 137 fb^{-1} of data (13 TeV) which is important for rare searches like four-top
- Currently one of the only options available for the four-tops search, the High Energy Frontier experiments (such as ATLAS and CMS)

Overview Four-Top Analyses



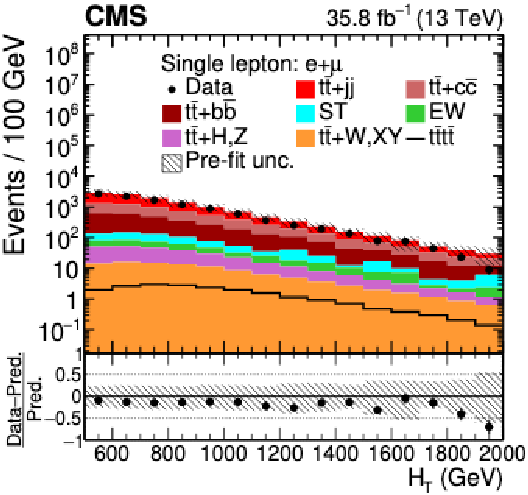
- "Same-Sign-Dilepton + Multilepton ($137\ \mathrm{fb}^{-1}$)" (arXiv:1908.06463)
- "1 Lepton + Opposite-Sign-Dilepton ($36\ \mathrm{fb}^{-1}$)"(arXiv:1906.02805)
- "Projections of sensitivities for tttt production at HL-LHC and HE-LHC"

About four-top analyses

- Final state signature 4 b-jets and $(8 - 2 \times N^{expected}_{leptons})$ jets
- Four-top signal will be at high multiplicities, where signal over background is largest
- Use low multiplicity regions as control, because they have no signal or are depleted of it.

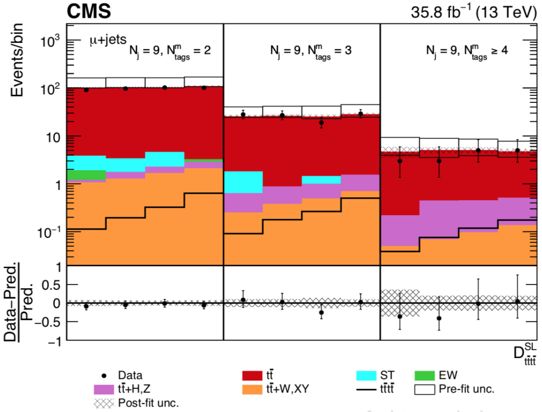
Strategy by analysis dominated by tt+bb/cc

- MC based simultaneous fit of a boosted decision tree, using MC shapes including full theory uncertainties
  - Resolved top-tagging (MVA based)
  - BDT trained on kinematic, b-tagging and resolved top-tagging information.
    - BDT for signal and background discrimination yields discriminant $D^{SL/DL}_{t\bar{t}t\bar{t}}$.
  - Lower jet an tag-jet multiplicity used to constrain large systematic uncertainties during simultaneous fit
  - Strategy is similar to simultaneous fits used for $t\bar{t}$ cross section



Strategy by analysis dominated by tt+bb/cc

- Major Uncertainty Sources:
  - tt+bb(cc) via gluon splitting
  - Renormalisation + factorisation scales
  - Modelling of additional jet radiation in MC
- Fit model consistent with the observed data
- This method is expected to gain better precision with larger datasets when there are more statistics in control regions



Results by analysis dominated by tt+bb/cc

- Even with few $t\bar{t}t\bar{t}$ candidate events $\rightarrow$ a expected limit of $+20^{+10}_{-6}$ fb, which is consistent with SM
- Gives a grand 2016 combination with $\sigma_{t\bar{t}t\bar{t}} = 13^{+11}_{-9}$ fb at 1.1(1.4)s.d. expected(observed) significance.

Channel (36 fb^{-1})	Expected limit μ	Observed limit μ	Expected limit (fb)	Observed limit (fb)	Best fit μ	Best fit $\sigma_{t\bar{t}t\bar{t}}$ (fb)
1L+OS2L	$5.7^{+2.9}_{-1.8}$	5.2	52^{+26}_{-17}	48	$0.0^{+2.2}$	0^{20}
Combination	$2.2^{+1.1}_{-0.7}$	3.6	20^{+10}_{-6}	33	$1.4^{+1.2}_{-1.0}$	13^{+11}_{-9}

Expected and observed 95% CL upper limits on SM tttt production as a multiple of σ_{SM}.The values quoted for the uncertainties in the expected limits indicate the regions containing 68% of the distribution of limits expected under the background-only hypothesis. The expected upper limits are calculated assuming that the data are distributed according to the prediction of the background-only model corresponding to the scenario with signal strength modifier value $\mu = 0$.
Maximum-likelihood signal strength, μ, and cross section estimates. Where uncertainties are 1 $s.d.$ values and include all statistical and systematic uncertainties.

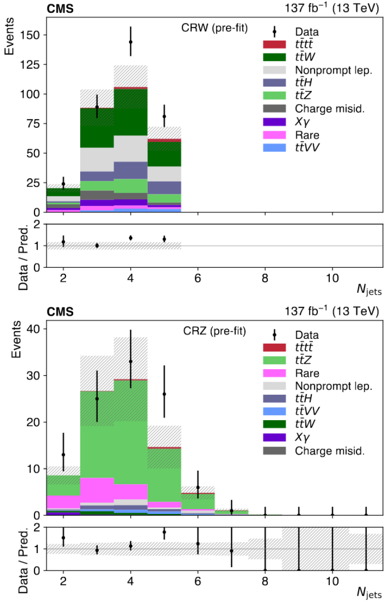
Strategy by analysis dominated by tt+W/Z

On the right you see the regions for the same-sign-dilepton/multi-lepton analysis.

- Dominant backgrounds $t\bar{t}W^{\pm}$ and $t\bar{t}Z/\gamma$, which have dedicated control regions to place constraints on the normalisation of the respective processes

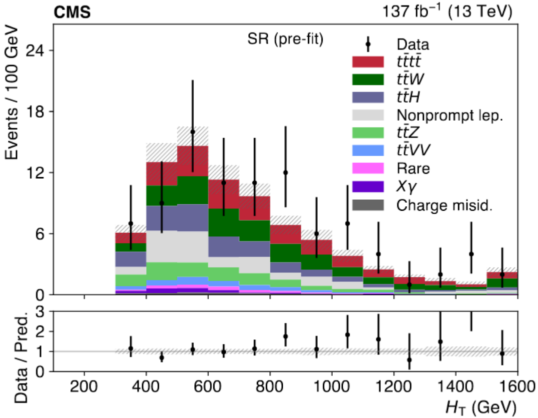
Other less dominant background processes:

- $t\bar{t}H$ (mainly by the $H \to WW$)
- Misidentification backgrounds
  - Charge misidentified prompt lepton
  - Single lepton with additional non-prompt lepton



Strategy by analysis dominated by tt+W/Z

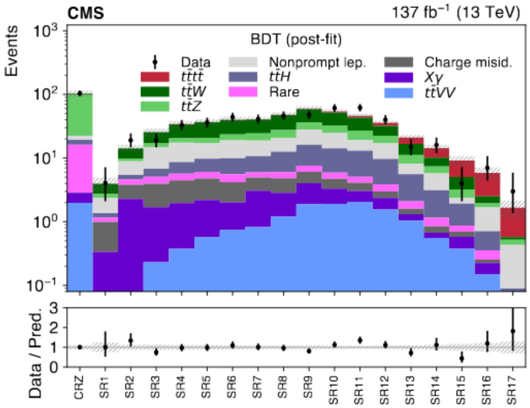
- BDT used to get optimal sensitivity
- Substantial improvement compared to the *cut-and-count* approach
- Uses a simultaneous fit in multiple categories.
- Dominant uncertainties:
  - Modelling SM backgrounds
  - Data-driven charge misidentification estimates
  - Knowledge heavy flavour $t\bar{t} + bb(cc)$



The expected SM $t\bar{t}t\bar{t}$ signal is normalised to its predicted $\sigma_{t\bar{t}t\bar{t}}$

Results by analysis dominated by tt+W/Z

- The values and uncertainties of nuisance parameters are consistent before and after fit
- But $t\bar{t}W$ and $t\bar{t}Z$ normalisation are scaled by 1.3 ± 0.2 by fit, which is in agreement of measurements of these processes by ATLAS and CMS.



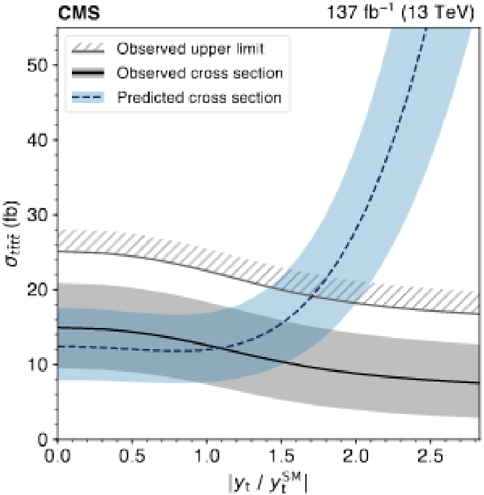
Results by analysis dominated by tt+W/Z

- Significance of BDT analysis is 2.6(2.7) s.d. observed(expected) over background-only hypothesis.
  - $\Delta\sigma_{t\bar{t}t\bar{t}}/\sigma_{t\bar{t}t\bar{t}} \sim 45\%$ uncertainty
  - Agrees well within uncerainties with NLO QCD+EWK value of $\sigma_{t\bar{t}t\bar{t}} = 12.0^{+2.2}_{-2.5}$ fb
  - Observed 95% CL upper limit on $\sigma_{t\bar{t}t\bar{t}}$ is 22.5 fb in BDT analyses
  - The corresponding expected upper limits assuming no SM $t\bar{t}t\bar{t}$, are $8.5^{+3.9}_{2.6}$ fb

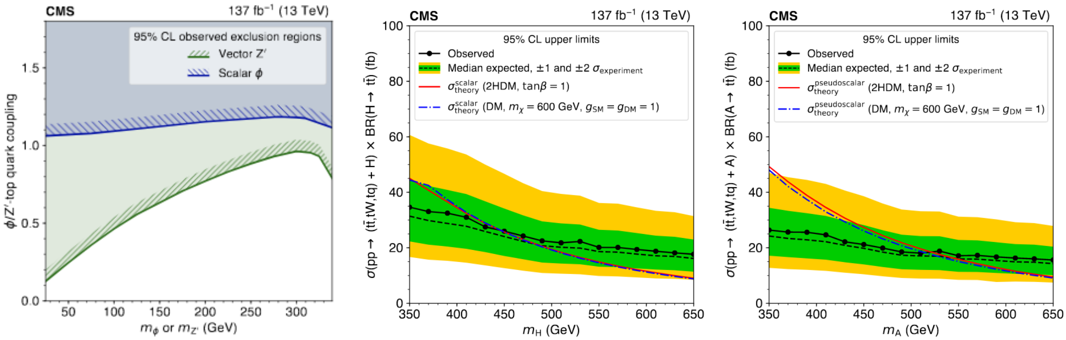
Interpretations by analysis dominated by tt+W/Z

Previously we had mentioned the top-Yukawa coupling. Well, here it is.

- $\sim 20\%$ of the $t\bar{t}t\bar{t}$ LO Feynman diagrams contain H
- y_t substantial impact on $\sigma_{t\bar{t}t\bar{t}}$
- $\sigma_{t\bar{t}t\bar{t}} = 12.6^{+5.8}_{-5.2}$ fb measurement used according to Cao et al.(arXiv:1602.01934)
- ttH is included in background so $\sigma_{t\bar{t}t\bar{t}}$ depends on $|y_t/y_t^{SM}|$
  - Most conservative scenario $|y_t/y_t^{SM}| < 1.7$ at 95% CL.

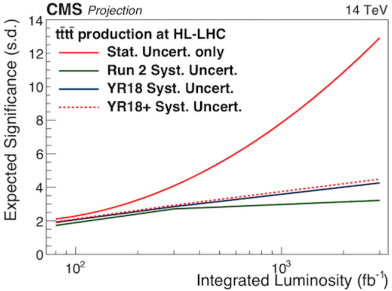


Interpretations by analysis dominated by tt+W/Z



- SM $t\bar{t}t\bar{t}$ as background to BSM $t\bar{t}t\bar{t}$ Z' (or ϕ) enhances, are:
  - Competitive limits at low masses.
  - Complementary to high mass $t\bar{t}$ resonance searches
  - Competitive to direct searches sensitive to SM interference/spectrum modifications.
- Limits on production new particles with $m > 2m_t$, like heavy scalar (H) and pseudoscalar (A), interpreted in terms of 2HDM parameters or in the framework of simplified models of dark matter (DM, m_χ).

Beyond RunII

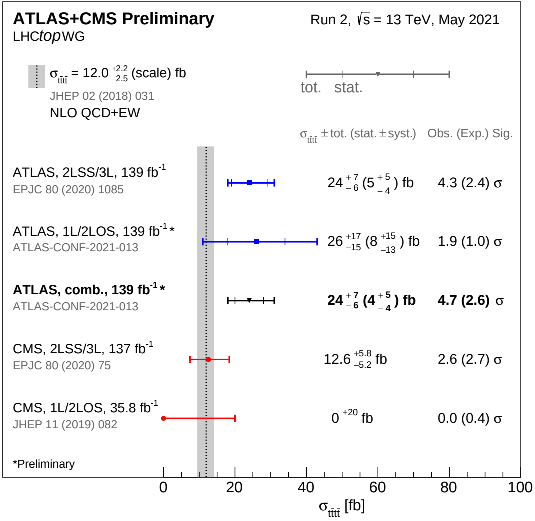


Int. Luminosity	*Stat. only*
300 fb^{-1}	4.09
3 ab^{-1}	12.9

Int. Luminosity	$\sqrt{s}$	*Stat. only* (%)
300 fb^{-1}	14 TeV	$^{+30}_{-28}$
3 ab^{-1}	14 TeV	± 9
3 ab^{-1}	27 TeV	± 2
15 ab^{-1}	27 TeV	± 1

- CMS projection, a scale-up of the 2016 Same-Sign-2L/3L counting experiment analysis for the HL-LHC and HE-LHC
  - Includes EFT projections and various scenarios regarding systematic uncertainties also expected cross section uncertainties at HE-LHC
  - **1% exp uncertainty on cross-section with full HE-LHC sample**
- Conclusion:
  - Multiple evaluation scenarios for systematic uncertainties
  - **Around** $300\ \text{fb}^{-1}$ **of HL-LHC data at** $\sqrt{s} = 14\ \text{TeV}$ **in single channel**
    - Given no systematic uncertainties: evidence for $t\bar{t}t\bar{t}$ (i.e., 5 s.d. significance) possible
    - **Given systematic uncertainties: 3 s.d. significance for** $t\bar{t}t\bar{t}$ **possible**
  - For larger datasets at HL-LHC all scenarios dominated by systematic uncertainties.

Summary



Interpretations

- With large LHC Run 2 dataset, CMS has 2.6(2.7) s.d. excess, and measurement of $\Delta\sigma_{t\bar{t}t\bar{t}}/\sigma_{t\bar{t}t\bar{t}} \sim 45\%$ means CMS can constrain $|y_t/y_t^{SM}| < 1.7$ at 95% CL
- Collaborations are interpreting $\sigma_{t\bar{t}t\bar{t}}$ beyond just the SM value, in EFT and various BSM models.

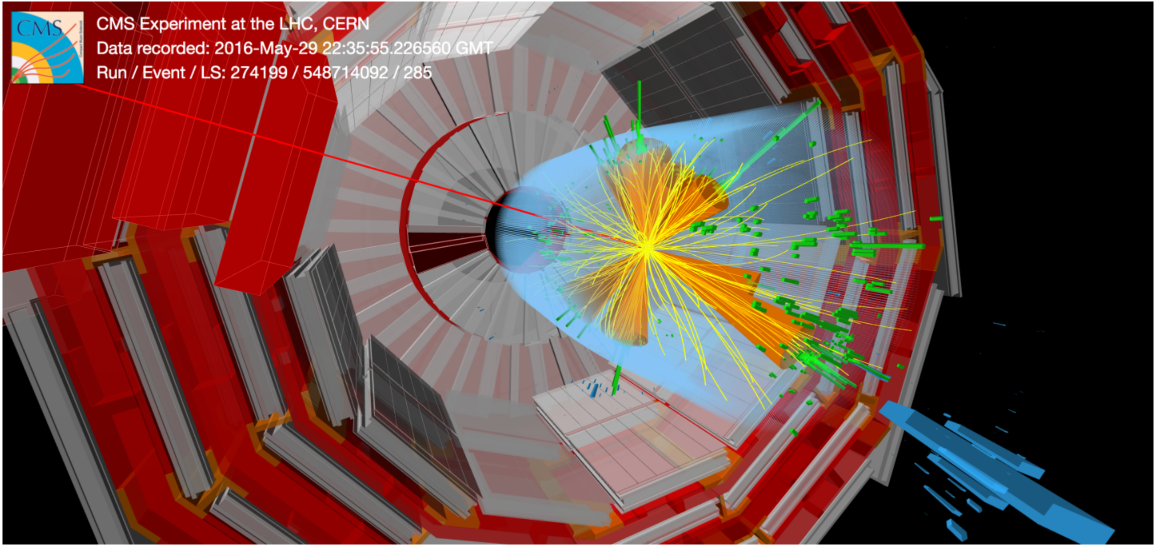
Future HL-LHC

- Single channel observation possible at 3 standard deviations sensitivity

Near Future

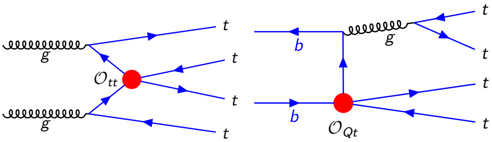
- So we have seen that 3σ possible at 300 fb^{-1} integrated luminosity at $\sqrt{s} = 14\ \mathrm{TeV}$ for a single channel
- Imagine now that we have 137 fb^{-1} at $\sqrt{s} = 13\ \mathrm{TeV}$ (which RunII has) **and** that we merge all the channels.
  Well then, things may get interesting.
- Alas, we cannot disclose anything yet and will leave this as something for everyone to look forward to.

THANKS FOR LISTENING



Backup

Results by analysis dominated by tt+bb/cc



- The EFT Lagrangian:

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \mathcal{L}^{(5)} + \mathcal{L}^{(6)} + \ldots$$

$$\mathcal{L}^{(d)} = \frac{1}{\Lambda^{d-4}} \sum_k C_k^{(d-4)} \mathcal{O}_k^{(d-4)}$$

  - $\mathcal{O}_k^{(n)}$ and $C_k^{(n)}$ are the dimension-n composite operators and their coupling parameters, respectively.
  - Λ related to mass scale of BSM states
- Lowest dimension EFT operators contributing to $pp \rightarrow t\bar{t}t\bar{t}$ have d=6 (operators shown in table)

EFT interpretation

- Cross section limits interpreted in EFT for separate Wilson parameters including marginalisation over all other physical free parameter values.
  - EFT basis recommended by TOP-LHCWG (arxiv:1802.07237)

	Expected C_k/Λ^2 TeV^{-2}	Observed TeV^{-2}
$\mathcal{O}_{tt}^1$	[-2.0, 1.9]	[-2.2, 2.1]
$\mathcal{O}_{QQ}^1$	[-2.0, 1.9]	[-2.2, 2.0]
$\mathcal{O}_{Qt}^1$	[-3.4, 3.7]	[-3.7, 3.5]
$\mathcal{O}_{Qt}^8$	[-7.4, 6.3]	[-8.0, 6.8]

Results by analysis dominated by tt+bb/cc

Assume coupling predominantly to the left-handed third generation quark doublet (Q_L) and the right-handed top quark singlet, only the following operators contribute significantly to the $t\bar{t}t\bar{t}$:

- $\mathcal{O}^1_{tt} = (\bar{t}_R \gamma^\mu t_R)(\bar{t}_R \gamma_\mu t_R)$
- $\mathcal{O}^1_{QQ} = (\bar{Q}_L \gamma^\mu Q_L)(\bar{Q}_L \gamma_\mu Q_L)$
- $\mathcal{O}^1_{Qt} = (\bar{Q}_L \gamma^\mu Q_L)(\bar{t}_R \gamma_\mu t_R)$
- $\mathcal{O}^8_{Qt} = (\bar{Q}_L \gamma^\mu T^A Q_L)(\bar{t}_R \gamma_\mu T^A t_R)$

$\mathcal{O}^1_{Qt}$ and $\mathcal{O}^8_{Qt}$ have identical terms and become identical if differences (terms) are suppressed by PDFs, therefore the two operators have strong correlation.

$$\sigma_{t\bar{t}t\bar{t}} = \sigma^{SM}_{t\bar{t}t\bar{t}} + \frac{1}{\Lambda^2}\vec{C}^T \vec{\sigma^{(1)}} + \frac{1}{\Lambda^4}\vec{C}^T \sigma^{(1)} \vec{C}$$

In order to find $\vec{\sigma}^{(1)}$ and $\sigma^{(2)}$, a system of linear equations has to be solved. It is obtained by substituting linearly-independent vectors C into the equation above. In the cross section calculation, the EFT interactions are implemented in the FEYNRULES package and interfaced with MADGRAPH5_aMC@NLO.

EFT interpretation

cross section limits interpreted in EFT for separate Wilson parameters where only one selected operator has a non-vanishing contribution.

	Exp. C_k/Λ^2 TeV^{-2}	Obs. TeV^{-2}
$\mathcal{O}^1_{tt}$	[-2.0, 1.8]	[-2.1, 2.0]
$\mathcal{O}^1_{QQ}$	[-2.0, 1.8]	[-2.2, 2.0]
$\mathcal{O}^1_{Qt}$	[-3.3, 3.2]	[-3.5, 3.5]
$\mathcal{O}^8_{Qt}$	[-7.3, 6.1]	[-7.9, 6.6]

Why Interest in Higgs?

- Know very little about Higgs boson width Γ_H and its coupling to SM fermions
- Γ_H cannot be measured directly from line-shape of Higgs boson resonance.
    - Can be determined through $gg \to H \to ZZ$ comparison of production rate at the Higgs resonance from that around it
    - Only upper bounds obtained by CMS and ATLAS
    - y_t , it indicates that H-boson must interact with top quarks
- y_t and Γ_H needed to decipher H-boson properties and enlighten us to BSM.

More on Yukawa coupling

- The contribution of $\mathcal{M}_H$ to the $t\bar{t}t\bar{t}$ production cross-section is proportional to y_t^4, and that enables us to measure y_t through the $t\bar{t}t\bar{t}$ channel.
- QCD and EWK gauge interactions of top quarks $\rightarrow$ well established
- y_t might differ from the SM value.

The cross section of the $t\bar{t}t\bar{t}$ is parameterised as:

$$\sigma(t\bar{t}t\bar{t}) = \sigma(t\bar{t}t\bar{t})_{int} + \sigma(t\bar{t}t\bar{t})_{g+Z/\gamma} + \sigma(t\bar{t}t\bar{t})_H$$

where:

- Interference among the three kinds of Feynman diagrams $\sigma(t\bar{t}t\bar{t})_{int} \propto \mathcal{M}_{g+Z/\gamma}\mathcal{M}_H^\dagger + \mathcal{M}_{g+Z/\gamma}^\dagger\mathcal{M}_H$
- $\sigma(t\bar{t}t\bar{t})_{g+Z/\gamma} \propto |\mathcal{M}_g + \mathcal{M}_{Z/\gamma}|^2$
- $\sigma(t\bar{t}t\bar{t})_H \propto |\mathcal{M}_H|^2$

Constraints on $\sigma(t\bar{t}t\bar{t})$ give constraints on y_t

More points on SS2L/3L

- In the boosted decision tree (BDT) analysis, the CRZ is the only control region, and the remaining events are subdivided into 17 SRs by discretizing the discriminant output of a BDT trained to separate tttt events from the sum of the SM backgrounds.
- The BDT classifier utilizes a gradient boosting algorithm to train 500 trees with a depth of 4 using simulation, and is based on 19 variables.
- The observed and predicted yields in the CRs and SRs are used in a maximum likelihood fit with nuisance parameters to measure $\sigma(pp \rightarrow tttt)$