Overview and status of $t\bar{t}H$ measurements at the LHC

Marino Missiroli (DESY-CMS)

DESY HEP Students Seminar

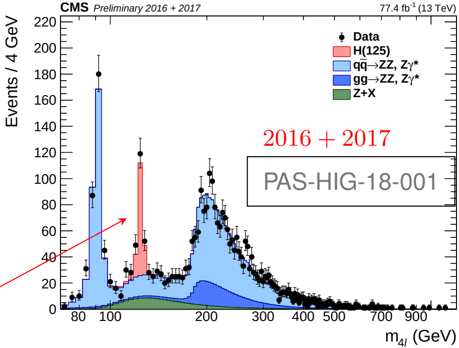
February 11, 2019

Outline

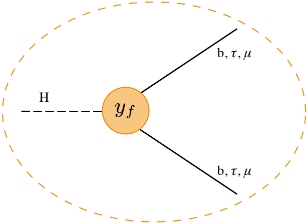
- why $t\bar{t}H$? and how to look for it?
- overview of $t\bar{t}H$ searches, channel by channel
- combination of $t\bar{t}H$ searches
- quick look to the future
- summary

The SM Higgs boson

- *The* milestone of the LHC Run-1
  - discovery of a new boson with $m \sim 125$ GeV by ATLAS and CMS
- first discovered in bosonic decay modes: $H \to \gamma\gamma$, $H \to ZZ^* \to 4\ell$
- thus far consistent with SM Higgs

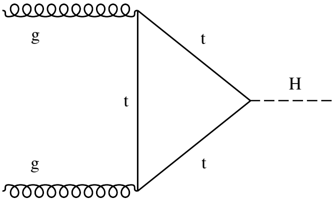


- key to the SM Higgs: coupling to fermions
  - $Hf\bar{f}$ Yukawa interaction leads to fermion masses
  - y_f coupling strength proportional to m_f



Higgs coupling to the top quark

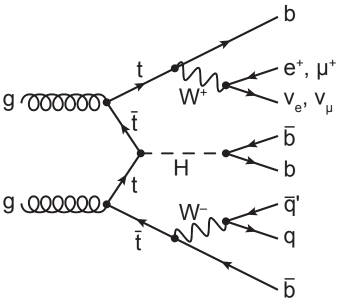
- Top quark: the elementary particle with the largest mass
- Top-Higgs Yukawa coupling: the largest, close to 1 $(m_t \sim v/\sqrt{2})$



- Gluon Fusion (ggF) mechanism
  - main Higgs production mode at the LHC
  - SM: largest contribution from top quark loop
  - Higgs discovery at the LHC and $\sigma(pp \to H)$ so far consistent with SM
    $\to$ indirect evidence of top-Higgs coupling
  - new physics could also be “hiding” in the loop
    $\to$ need to directly measure top-Higgs coupling

Higgs coupling to the top quark

- Top quark: the elementary particle with the largest mass
- Top-Higgs Yukawa coupling: the largest, close to 1 $(m_t \sim v/\sqrt{2})$



- $t\bar{t}H$ production
  $\rightarrow$ direct measurement of top-Higgs coupling
  - rare Higgs production mode at the LHC
  - one of the challenges: $t\bar{t}+X$ bkgs

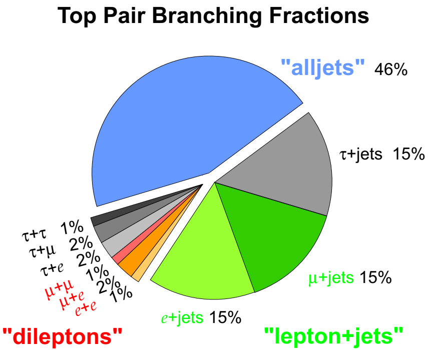
σ_{SM} ($\sqrt{s} = 13$ TeV) [*]	
$t\bar{t}$	831.8 pb
$t\bar{t}+b\bar{b}$	3.5 pb
$t\bar{t}V$	1.5 pb
$t\bar{t}H$	0.5 pb
tH	0.07 pb

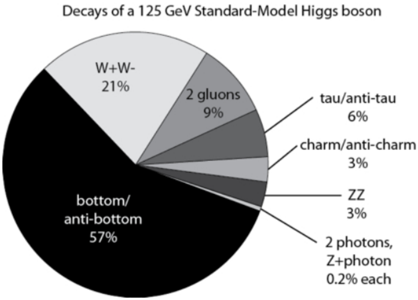
[*] approx. values

$t\bar{t}H$ final states

“what to look for . . .

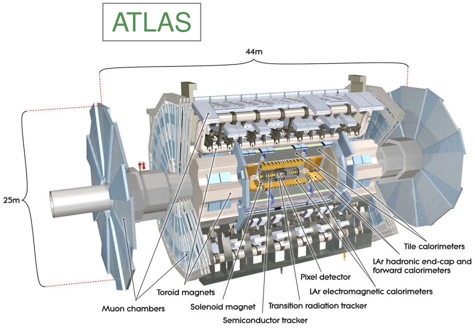
- $t\bar{t}H$ searches $\Rightarrow$ many different final states
- combination of all possible decay modes of a $t\bar{t}$ pair and a Higgs

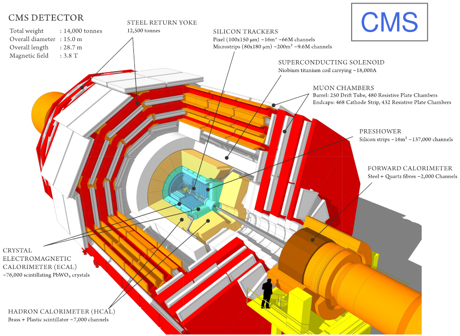




ATLAS and CMS

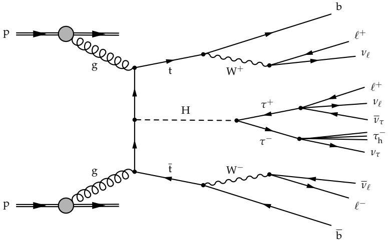
- $t\bar{t}H$ searches $\Rightarrow$ a testbed for general-purpose detectors at the LHC
- all physics objects enter the game: γ, e, μ, τ, (b-)jets, MET



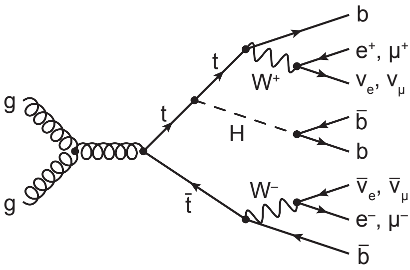


Grouping $t\bar{t}H$ searches

- $t\bar{t}H$ multilepton:
  - targets $H \to WW^*$, $H \to ZZ^*$, $H \to \tau^+\tau^-$ decays
  - events with $\geqslant 3$ charged leptons, including hadronic τ decays
  - *keys*: lepton identification, control over fake-lepton bkgs



- $t\bar{t}H$ with $H \to b\bar{b}$ decays:
  - 0, 1 or 2 leptons + many jets, with up to 4 b-tags
  - *keys*: jet b-tagging, separation of $t\bar{t}H(b\bar{b})$ signal from $t\bar{t} + b\bar{b}$ bkg



- $t\bar{t}H$ with $H \to \gamma\gamma$ and $H \to 4\ell$ decays:
  - high purity, but lowest signal yields

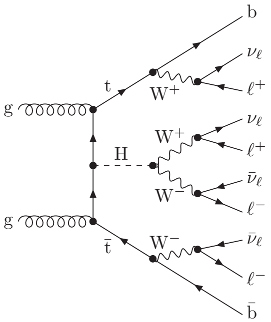
$t\bar{t}H \to$ multilepton

ATLAS : HIGG-2017-02

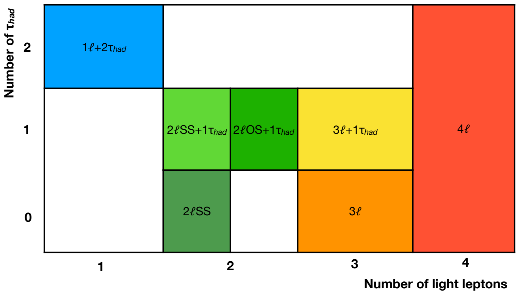
CMS : HIG-18-019

$t\bar{t}H$ multilepton

- targets $H \to WW^*, ZZ^*, \tau\tau$ decays
- lepton channels: (with $\geqslant 2$ jets and $\geqslant 1/2$ b-tags)



slightly different categs for ATLAS and CMS (2ℓOS+1τ, $2\ell + 2\tau$)



- Main backgrounds:
  - $t\bar{t} + V$ and VV: shape from MC, normalization from CRs
  - fake-lepton bkg (from $t\bar{t}$ + jets): fully data-driven

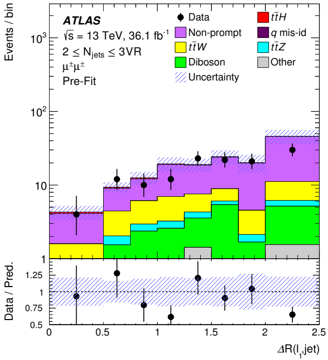
$t\bar{t}H$ multilepton / lepton-ID and fake-ℓ bkg

- lepton identification is key
  $\Longrightarrow$ not too loose (large fake-ℓ bkg), not too tight (low signal yield)

ATLAS loose isolation cut, with p_T^ℓ-dependent iso-cone

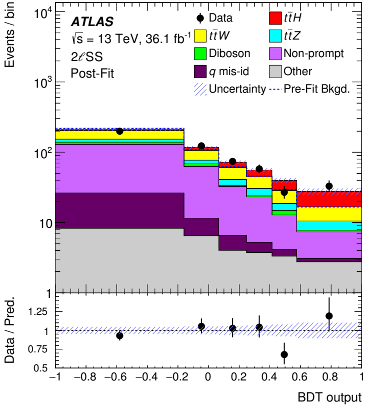
CMS BDT-based using lepton properties + energy deposits around lepton

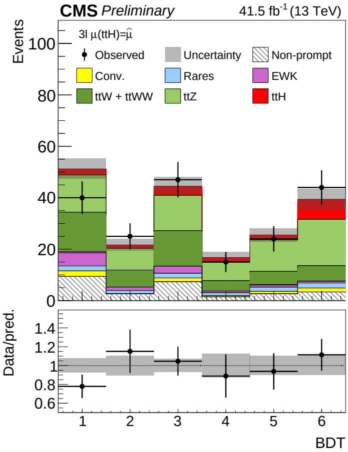
- data-driven fake-ℓ bkg estimate:
  - separate estimate for e, μ, τ_{h}, and for each SR
  - 1st CR used to parametrize fake-ℓ mistag wrt lepton kinematics
  - fake-ℓ mistag applied to Data from 2nd CR, to estimate non-prompt ℓ bkg in SRs



$t\bar{t}H$ multilepton / SR discriminants

- final discriminants: outputs of BDTs based on event kinematics

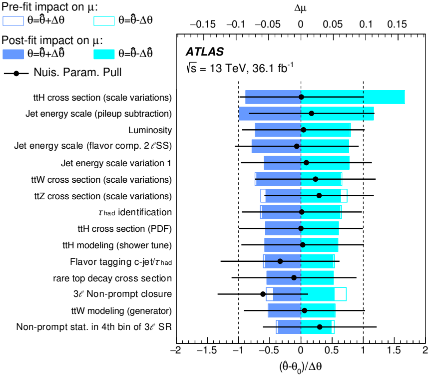




$t\bar{t}H$ multilepton / systematics

- main systematics:
  - lepton ID, jet energy scale, b-tagging eff.
  - fake-ℓ bkg and MC modeling of $t\bar{t}+V$ and $t\bar{t}H$

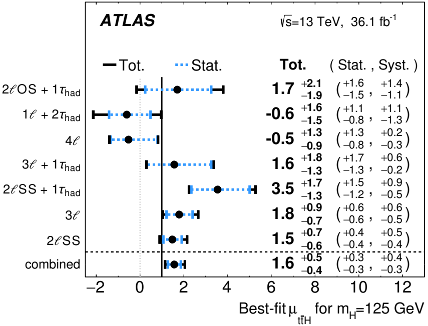
impact: $\delta\mu_{\text{fit}}$ induced by 1 single syst unc.

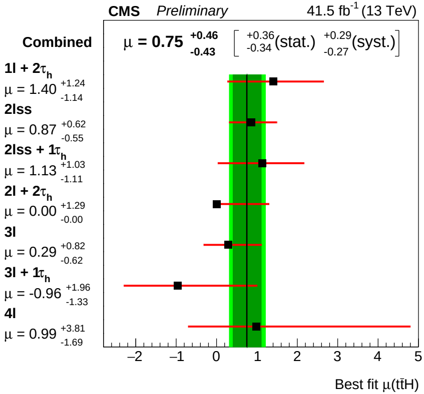


CMS

Source	Uncertainty [%]	$\Delta\mu/\mu$ [%] (2017)
Theoretical sources	≈ 8	8
e, μ selection efficiency	3–5	4
τ_h selection efficiency	5	3
τ_h energy calibration	1.2	1
b tagging efficiency	2–15 [48]	10
Jet energy calibration	2–15 [56]	3
Fake background yield	$\approx$ 30–50	17

$t\bar{t}H$ multilepton / results





- evidence of $t\bar{t}H$ in multilepton final states

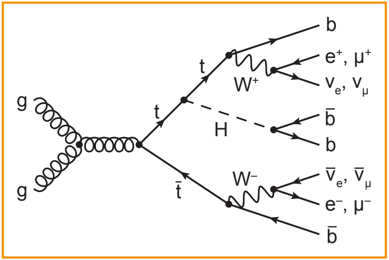
$t\bar{t}H$ multilepton	Significance: Obs (Exp)	
ATLAS (2016)	4.1σ	(2.8σ)
CMS (2017)	1.7σ	(2.9σ)
CMS (2016+2017)	3.2σ	(4.0σ)

$t\bar{t}H$ with $H \rightarrow b\bar{b}$

ATLAS : HIGG-2017-03

CMS : HIG-17-022, HIG-17-026

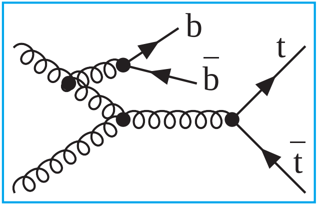
$t\bar{t}H(b\bar{b})$



- 3 channels based on $t\bar{t}$ decay mode
  - single-lepton (SL), dilepton (DL)
  - [CMS-only] fully-hadronic (FH)

[will focus on SL and DL (the 2 most sensitive channels)]

- main bkg: $t\bar{t}$ + jets, modeled with MC (POWHEG)
  - irreducible $t\bar{t} + b\bar{b}$ bkg with large modeling unc.



Goal #1 construct samples with high $t\bar{t} + b\bar{b}$ purity

- jet b-tagging is key, to reduce $t\bar{t}$ + light bkg

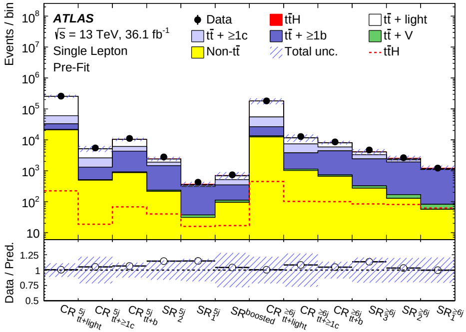
Goal #2 construct discriminants to distinguish $t\bar{t} + b\bar{b}$ and $t\bar{t}H(b\bar{b})$

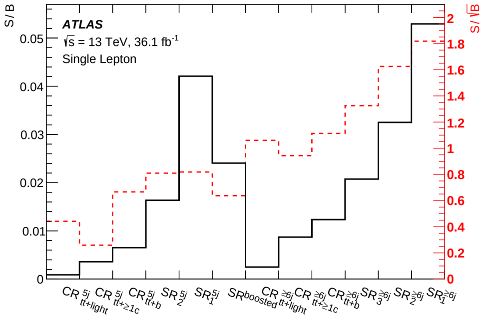
- $H(b\bar{b})$ candidate cannot be uniquely identified due to b-jet combinatorics $\Rightarrow$ MVA techniques

$t\bar{t}H(b\bar{b})$ / categories: pre-fit yields and $S/\sqrt{B}$

CMS : SRs based on N_{jets} and $N_{\text{b-tags}}$ (1-btag Working Point)

ATLAS : SRs and CRs based on N_{jets} and b-tag purity (4 b-tag WPs)

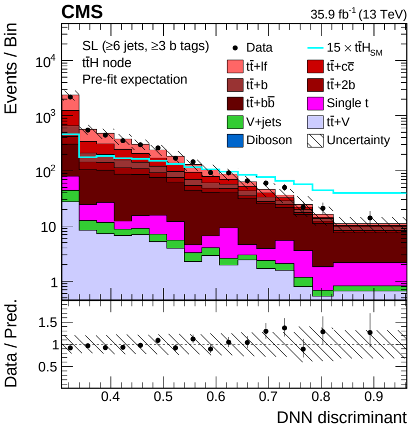


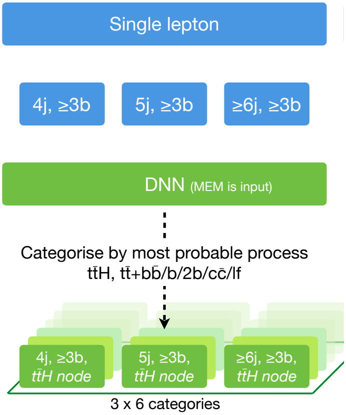


$t\bar{t}H(b\bar{b})$ / final discriminants in SRs

ATLAS : combination of BDT methods (jet-parton assignment + event classification)

CMS : (SL) multi-class DNN to separate $t\bar{t}+\mathrm{lf}$, $t\bar{t}+\mathrm{hf}$ and $t\bar{t}H$; (DL) BDT

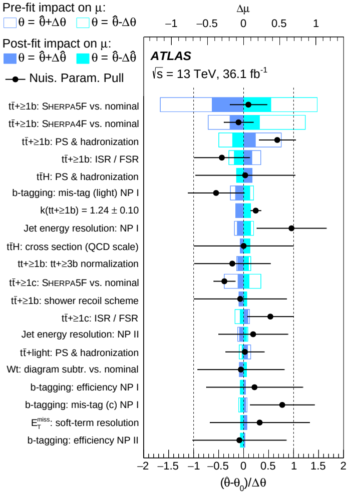




$t\bar{t}H(b\bar{b})$ / systematics

- main systematics:

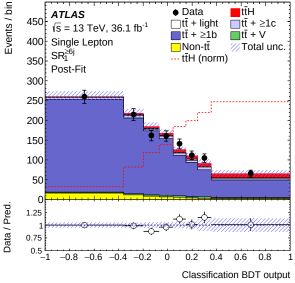
JES | b-tagging eff. | MC modeling of $t\bar{t}+\mathrm{hf}$ and $t\bar{t}H$ | MC statistics

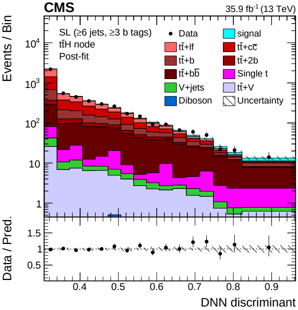


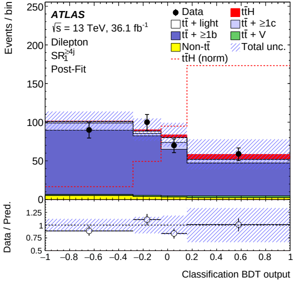
CMS

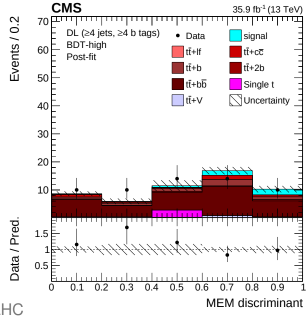
Uncertainty source	$\pm\Delta\mu$ (observed)	$\pm\Delta\mu$ (expected)
Total experimental	$+0.15/-0.16$	$+0.19/-0.17$
b tagging	$+0.11/-0.14$	$+0.12/-0.11$
jet energy scale and resolution	$+0.06/-0.07$	$+0.13/-0.11$
Total theory	$+0.28/-0.29$	$+0.32/-0.29$
$t\bar{t}$+hf cross section and parton shower	$+0.24/-0.28$	$+0.28/-0.28$
Size of the simulated samples	$+0.14/-0.15$	$+0.16/-0.16$
Total systematic	$+0.38/-0.38$	$+0.45/-0.42$
Statistical	$+0.24/-0.24$	$+0.27/-0.27$
Total	$+0.45/-0.45$	$+0.53/-0.49$

$t\bar{t}H(b\bar{b})$ / post-fit discriminants

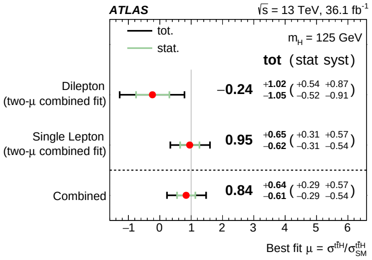


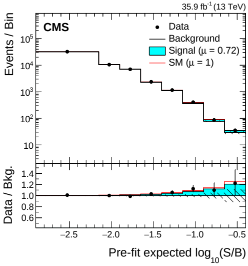


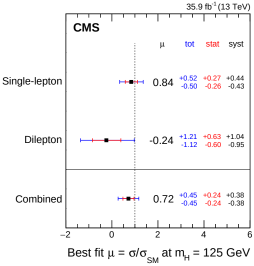




$t\bar{t}H(b\bar{b})$ / **results**







- small excess of events compatible with SM $t\bar{t}H$
- best-fit signal strength:
  - systematic unc. > stat. unc. (but stat. unc. still far from negligible)

$t\bar{t}H(b\bar{b})$	Significance: Obs	(Exp)
ATLAS (2016)	1.4σ	(1.6σ)
CMS (2016)	1.6σ	(2.2σ)

$t\bar{t}H$ with $H \to \gamma\gamma$

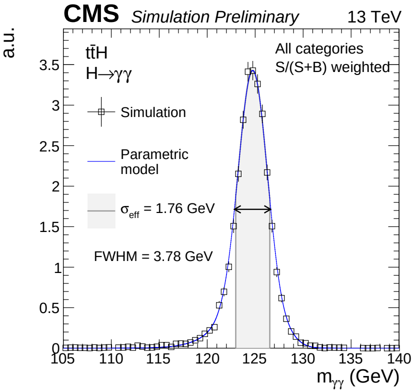
ATLAS : HIGG-2018-13

CMS : HIG-18-018

$t\bar{t}H(\gamma\gamma)$

$\gamma\gamma$: a golden channel, but sensitivity to $t\bar{t}H$ production requires a lot more data (compared to $gg \to H$)

- $\sigma_{t\bar{t}H} \cdot \mathrm{BR}(H \to \gamma\gamma) \simeq (0.5\ \mathrm{pb})(0.2\%) \simeq 1\ \mathrm{fb}$
- $\gamma\gamma$ pair: a well-defined Higgs candidate with good mass resolution ($\delta m_{\gamma\gamma} \sim 2$ GeV)

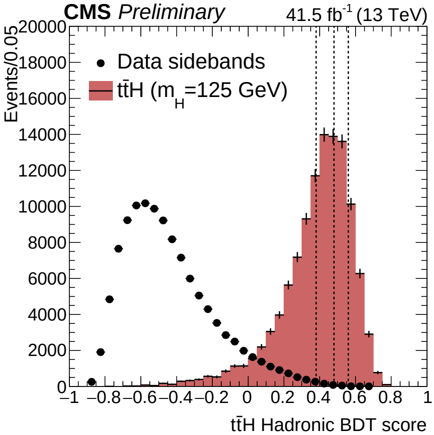


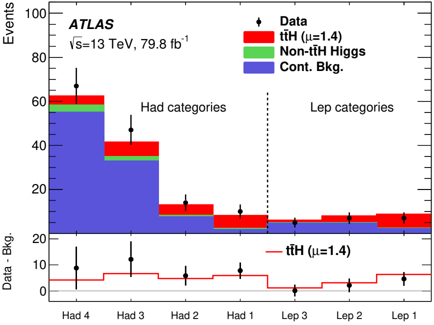
[similar analysis strategies for ATLAS and CMS]

- Event selection: $\gamma\gamma$ + (0, 1 or 2) leptons + jets
  - 2 Event-BDTs trained with MC to tag $t\bar{t}H$-leptonic and $t\bar{t}H$-hadronic events
- Signal extraction: $m_{\gamma\gamma}$
  - bkg and signal parameterized with analytical functions

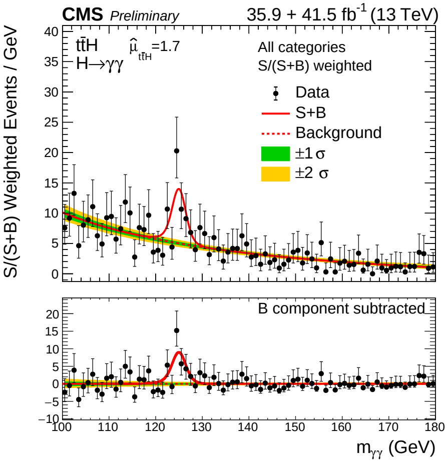
$t\bar{t}H(\gamma\gamma)$ / ttH-Lep and ttH-Had categories

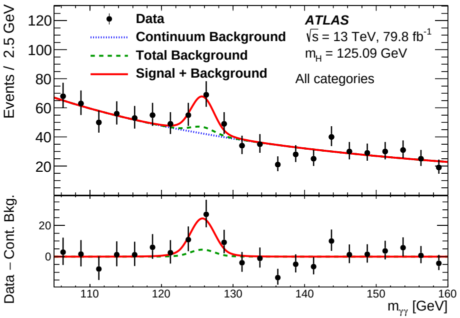
- ttH-BDT inputs: γ, ℓ and jets kinematics, b-tag scores (no info on $m_{\gamma\gamma}$)



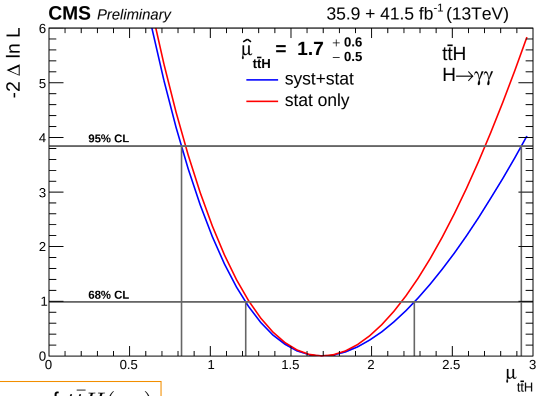


$t\bar{t}H(\gamma\gamma)$ / $m_{\gamma\gamma}$ fits





$t\bar{t}H(\gamma\gamma)$ / results



evidence of $t\bar{t}H(\gamma\gamma)$

- limited by statistics
- main systematics: $t\bar{t}H$ modeling, photon-ID, JEC, b-tagging eff.

$t\bar{t}H(\gamma\gamma)$	μ_{fit}	Obs. Signif.
ATLAS (2016+2017)	$1.39^{+0.48}_{-0.42}$	4.1σ
CMS (2017)	$1.3^{+0.7}_{-0.5}$	3.1σ
CMS (2016+2017)	$1.7^{+0.6}_{-0.5}$	4.1σ

$t\bar{t}H$ with $H \to ZZ^* \to 4\ell$

ATLAS : HIGG-2018-13

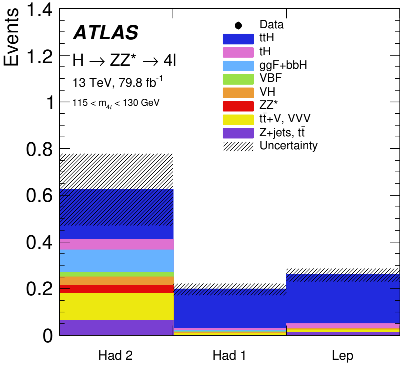
CMS : HIG-18-001

$t\bar{t}H(4\ell)$

$ZZ^* \to 4\ell$: another golden channel, but even more limited in statistics than $t\bar{t}H(\gamma\gamma)$

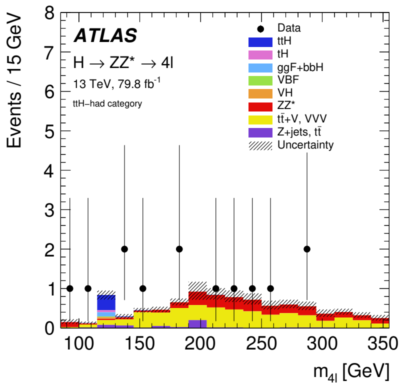
- $\sigma_{t\bar{t}H} \cdot$ BR$(H \to 4\ell) \simeq (0.5 \text{ pb})(0.02\%) \simeq 0.1$ fb
- Higgs candidate: 4ℓ system

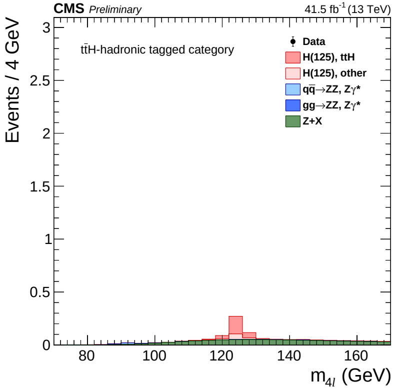
[similar analysis strategies for ATLAS and CMS]



- Event selection:
  - standard 4ℓ selection + requirements on add. leptons and/or jets to construct ttH-Lep and ttH-Had categories
- with $\sim 80 \text{ fb}^{-1}$: order of 1 $t\bar{t}H(4\ell)$ event expected

$t\bar{t}H(4\ell)$ / results





$t\bar{t}H(ZZ^* \to 4\ell)$	$\mu_{t\bar{t}H}$
ATLAS (2016+2017)	< 1.77 (68% CL)
CMS (2017)	< 0.93 (68% CL)
CMS (2016+2017)	< 0.53 (68% CL)

$t\bar{t}H$ combination

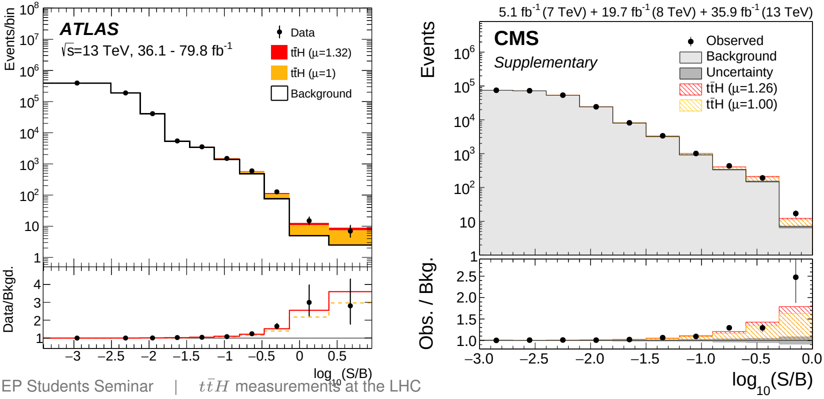
ATLAS : HIGG-2018-13

CMS : HIG-17-035

$t\bar{t}H$ combination

- Apr-Jun 2018: each experiment combined all available $t\bar{t}H$ searches
  - Run-1 ($8\ \mathrm{TeV}$) data + Run-2 data ($13\ \mathrm{TeV}$) up to 2016
    [ATLAS also included 2017 data for $\gamma\gamma$ and 4ℓ]
- Plot: bins of all searches ranked by $\log_{10}(S/B)$

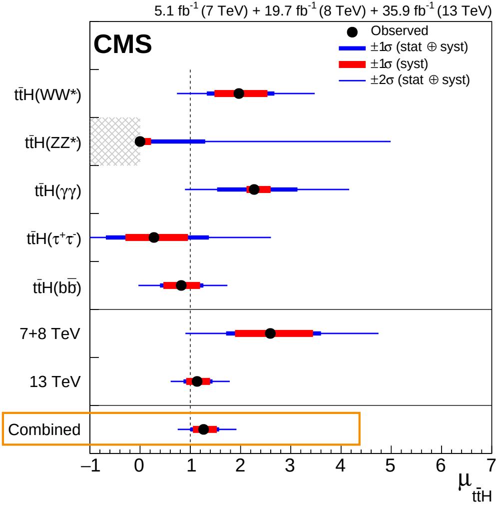
clear excess of events compatible with SM $t\bar{t}H$



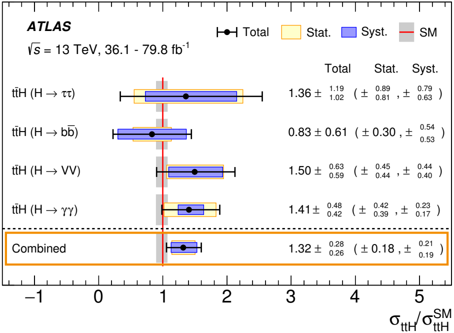
$t\bar{t}H$ observation at CMS

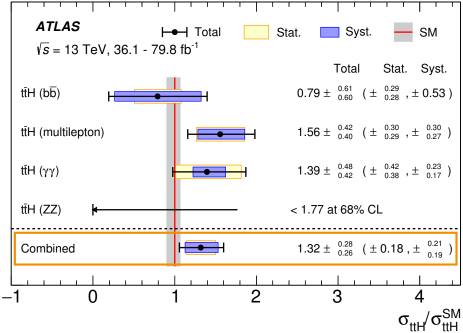
Obs. (Exp.) Significance: 5.2σ (4.2σ)

Parameter	Best fit	Uncertainty: Stat	Expt	Thbgd	Thsig
$\mu^{7+8\,\text{TeV}}_{t\bar{t}H}$	$2.59^{+1.01}_{-0.88}$	$^{+0.54}_{-0.53}$	$^{+0.53}_{-0.49}$	$^{+0.55}_{-0.49}$	$^{+0.37}_{-0.13}$
	$\left(^{+0.87}_{-0.79}\right)$	$\left(^{+0.51}_{-0.49}\right)$	$\left(^{+0.48}_{-0.44}\right)$	$\left(^{+0.50}_{-0.44}\right)$	$\left(^{+0.14}_{-0.02}\right)$
$\mu^{13\,\text{TeV}}_{t\bar{t}H}$	$1.14^{+0.31}_{-0.27}$	$^{+0.17}_{-0.16}$	$^{+0.17}_{-0.17}$	$^{+0.13}_{-0.12}$	$^{+0.14}_{-0.06}$
	$\left(^{+0.29}_{-0.26}\right)$	$\left(^{+0.16}_{-0.16}\right)$	$\left(^{+0.17}_{-0.16}\right)$	$\left(^{+0.13}_{-0.12}\right)$	$\left(^{+0.11}_{-0.05}\right)$
$\mu_{t\bar{t}H}$	$1.26^{+0.31}_{-0.26}$	$^{+0.16}_{-0.16}$	$^{+0.17}_{-0.15}$	$^{+0.14}_{-0.13}$	$^{+0.15}_{-0.07}$
	$\left(^{+0.28}_{-0.25}\right)$	$\left(^{+0.15}_{-0.15}\right)$	$\left(^{+0.16}_{-0.15}\right)$	$\left(^{+0.13}_{-0.12}\right)$	$\left(^{+0.11}_{-0.05}\right)$



$t\bar{t}H$ observation at ATLAS





Analysis	Integrated luminosity [fb^{-1}]	$t\bar{t}H$ cross section [fb]	Obs. sign.	Exp. sign.
$H \to \gamma\gamma$	79.8	$710\ ^{+210}_{-190}$ (stat.) $^{+120}_{-90}$ (syst.)	$4.1\ \sigma$	$3.7\ \sigma$
$H \to$ multilepton	36.1	790 ± 150 (stat.) $^{+150}_{-140}$ (syst.)	$4.1\ \sigma$	$2.8\ \sigma$
$H \to b\bar{b}$	36.1	$400\ ^{+150}_{-140}$ (stat.) $\pm\ 270$ (syst.)	$1.4\ \sigma$	$1.6\ \sigma$
$H \to ZZ^* \to 4\ell$	79.8	<900 (68% CL)	$0\ \sigma$	$1.2\ \sigma$
Combined (13 TeV)	36.1−79.8	670 ± 90 (stat.) $^{+110}_{-100}$ (syst.)	$5.8\ \sigma$	$4.9\ \sigma$
Combined (7, 8, 13 TeV)	4.5, 20.3, 36.1−79.8	–	$6.3\ \sigma$	$5.1\ \sigma$

Looking ahead

Ways to improve

A non-comprehensive list

- all $t\bar{t}H$ analyses will profit from more data
  - $\gamma\gamma$ and 4ℓ limited by statistics
  - $t\bar{t}H(b\bar{b})$: stronger constraint on $t\bar{t}+\mathrm{hf}$ bkgs from bins w/ low S/B
  - multilepton: constraints on $t\bar{t}V$, and more stats for fake-ℓ bkg estimate
- upgraded detectors
  - *Examples*: CMS Phase-1 Pixel installed in Jan 2017, Phase-2 upgrades
- improved object reconstruction methods
  - *Example*: jet b-tagging (next slide)
- improvements to analysis techniques
  - *Example*: MVA discriminants
    (S/B discriminants in multilepton and $t\bar{t}H(b\bar{b})$)

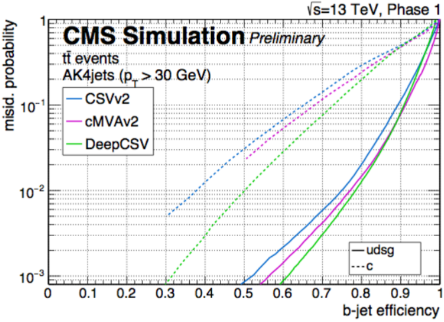
New b-tagging methods in CMS

Near future

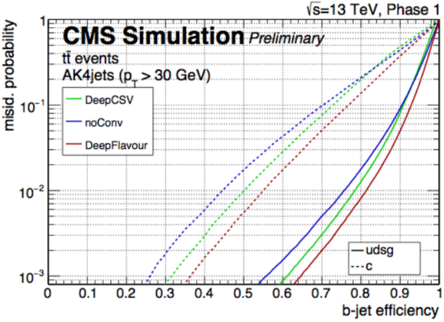
- application of latest ML techniques to b-tagging algorithms

Refs: CMS-DP-2017-005, CMS-DP-2017-013

CSVv2 → DeepCSV



DeepCSV → DeepFlavour

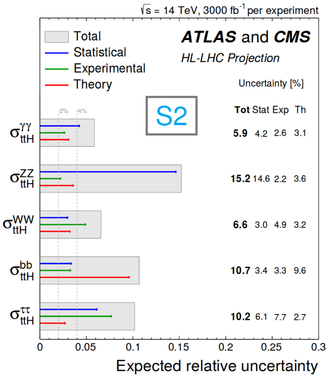


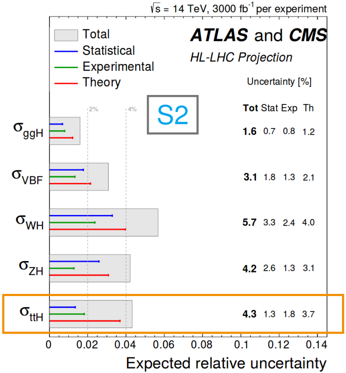
[CSV = Combined Secondary Vertex; b-tagging algorithm most widely used in CMS up to 2016]

HL-LHC Projections

Not-so-near future

- ATLAS and CMS projections for physics analyses at HL-LHC:
  - Higgs Chapter: Higgs Physics at the HL-LHC and HE-LHC
- $\Delta\sigma_{t\bar{t}H} = 4.3\%$ (S2) projected for ATLAS+CMS at $3000\ \text{fb}^{-1}$ per experim.
  - S2: assumes lumi-scaling of some syst, and reduced theory unc.





Summary

- observation of $t\bar{t}H$ production by ATLAS and CMS
  - required the analysis of many different final states, using both Run-1 and Run-2 data
  - observed $t\bar{t}H$ signal strengths in good agreement with SM, measured with approx. 20% uncertainty
- measurements with full Run-2 data set ($\sim 140\ \mathrm{fb}^{-1}$) underway
  - some analyses already updated with 2017 data ($\sim 80\ \mathrm{fb}^{-1}$)
  - evidence of $t\bar{t}H$ signal in (some) individual decay channels
  - increased sensitivity expected with more data and improvements to reconstruction/analysis methods (e.g. b-tagging)
- HL-LHC projections:
  - realistic goal to reduce $\Delta\sigma_{t\bar{t}H}$ to $O(5\%)$ by the end of Phase-2