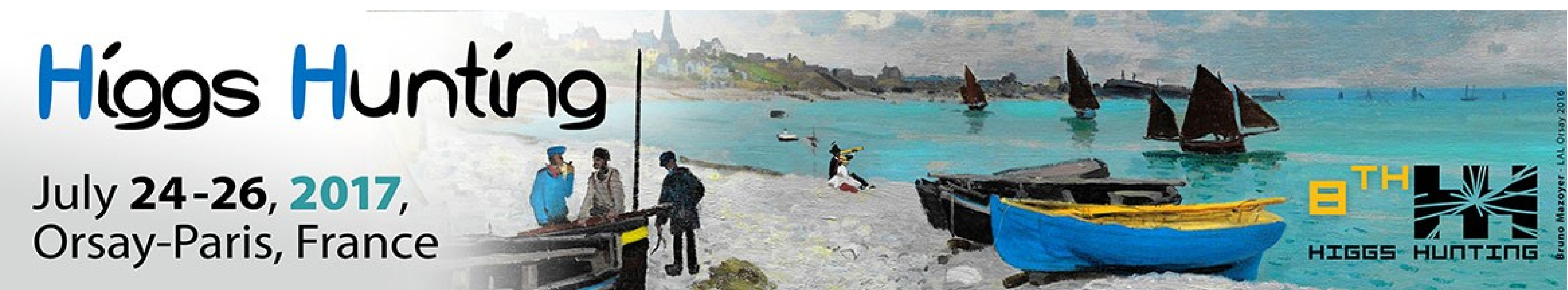
$t\bar{t}H$ and tH searches in CMS

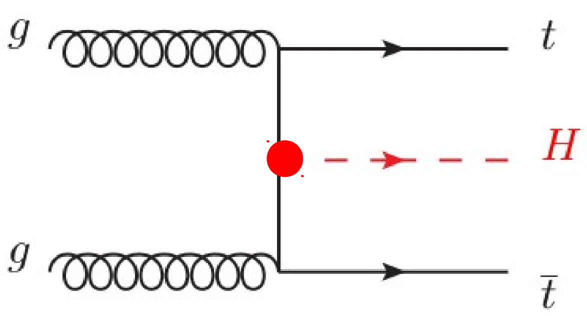
Inna Kucher

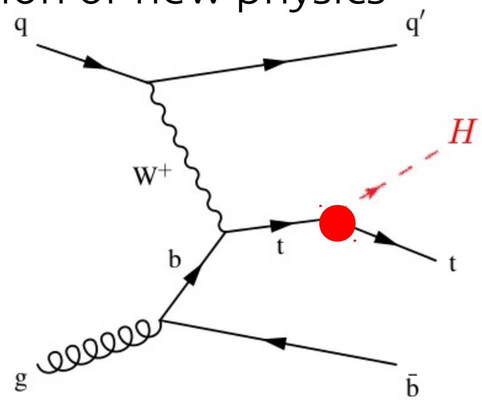
On behalf of CMS collaboration

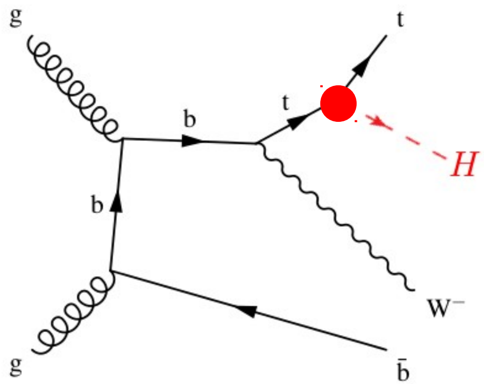


$t\bar{t}H$ and tH: why so relevant now?

- In SM the top quark Yukawa coupling is the strongest one ($y_t \propto m_t/v \approx 1$)
- **ttH:** the only direct access to the top Yukawa coupling (indirect from ggF)
- **tH:** top Yukawa coupling sign (interference between κ_{V} vs κ_{t})
- Deviations from SM → indication of new physics

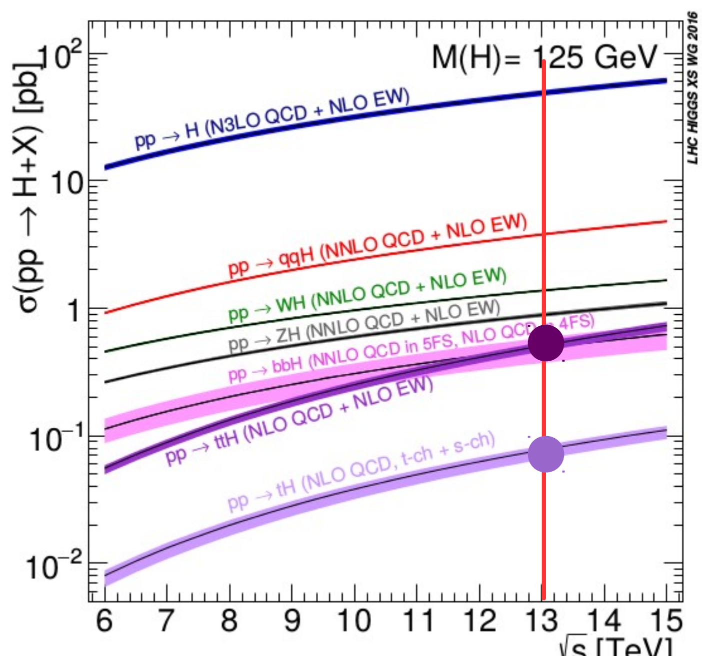






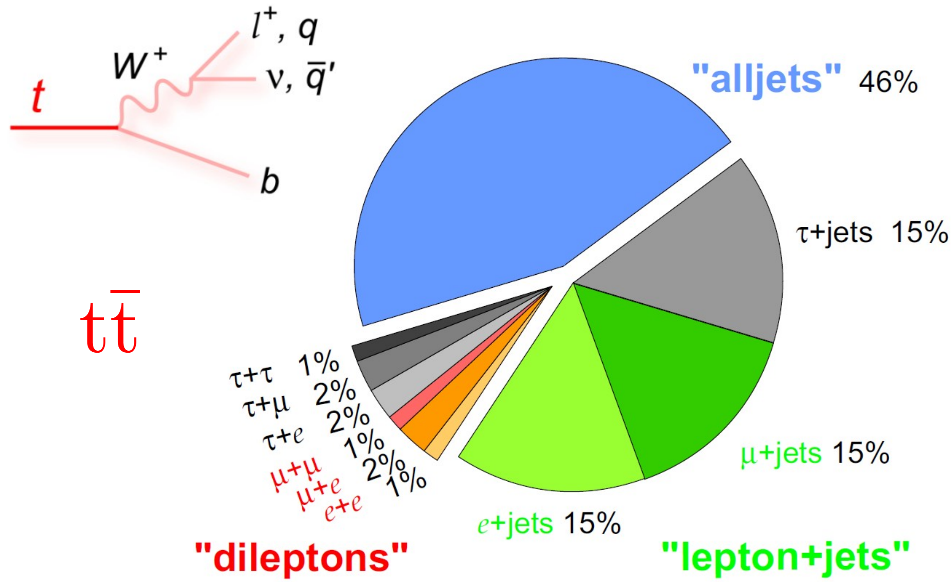
tH: t-channel and tW associated production, the s-channel is negligible

- LHC Run II ($\sqrt{s} = 13\ \mathrm{TeV}$) wrt Run I ($\sqrt{s} = 7-8\ \mathrm{TeV}$):
  - $t\bar{t}H$ and **tH** cross-section increases x4,
  - x2 other main Higgs productions,
  - x3.5 $t\bar{t}$ production.

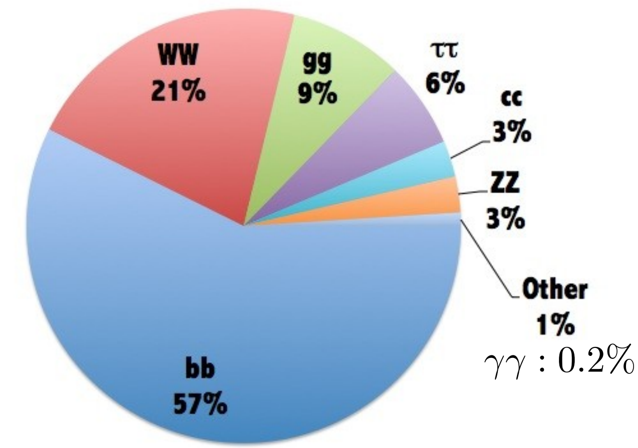


Searches in CMS

$t\bar{t}$



+ Higgs boson



$t\bar{t}H/tH,\ H \to b\bar{b}$

- High BR, complex final state, large backgrounds

$t\bar{t}H/tH,\ H \to \text{multileptons}\ (WW^*,\ ZZ^*,\ \tau_{lep}\tau_{lep})$

- Intermediate BR, low background

$t\bar{t}H, H \to \tau_{had}\tau_{any}(\text{also } WW^* \text{ and } ZZ^*)$

- Complex final state, significant 'fake' background

$t\bar{t}H,\ H \to \gamma\gamma, 4l$

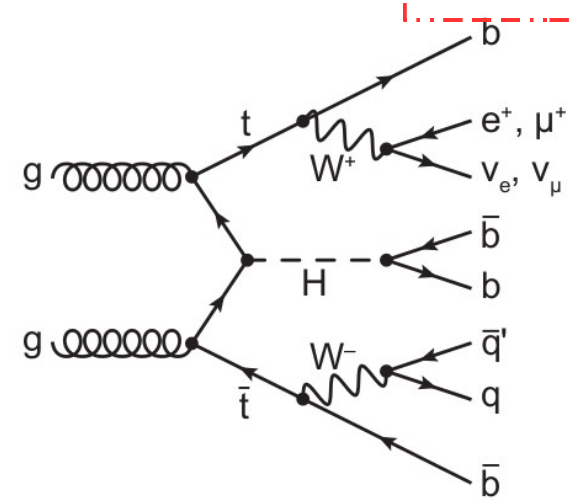
- Tiny BR, clean final state → small systematic uncertainty

General strategies and results are shown

$t\bar{t}H$

$t\bar{t}H,\ H \to b\bar{b}$

HIG-16-038



- $t\bar{t}$+jets overwhelming background
  (irreducible $t\bar{t} + b\bar{b}$: large theoretical uncertainties)

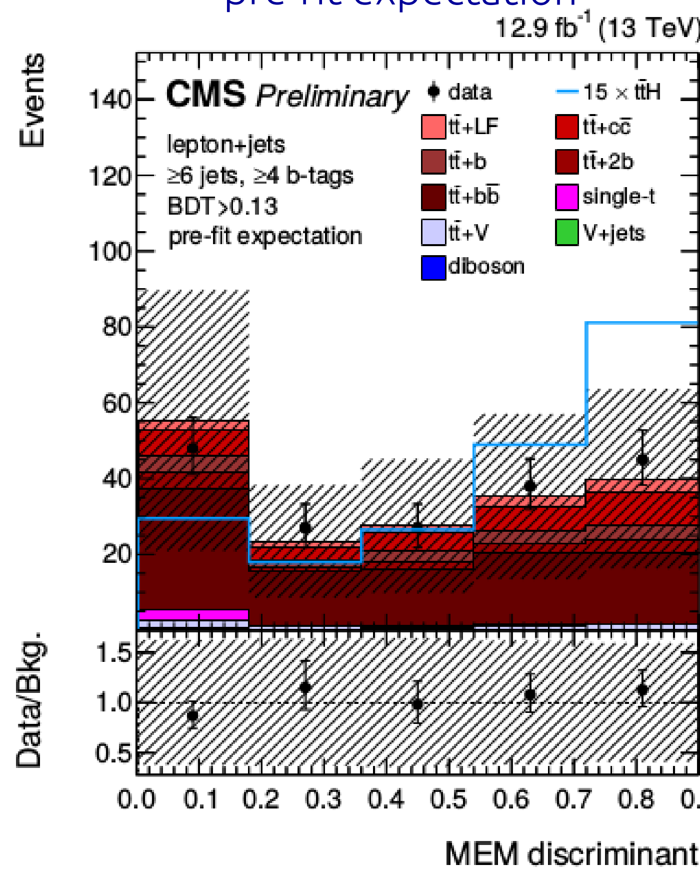
- Lepton+jets (4 cat.), dilepton (3 cat.):
  - Lepton triggers
  - =1, =2 opposite-sign (OS) leptons
  - ≥4, ≥2 jets
  - ≥3 b-tags

- Discrimination signal from background with BDT:
  - kinematic properties of jets, leptons, b-tag probability, inv. masses, and angular correlations between leptons and jets
  - Categories with high number of jets/b-jets are divided into sub-cat. depending on BDT output

  and MEM (final discriminant):
  - vs $t\bar{t} + b\bar{b}$

- Extracting limit by combined fit to data in all categories

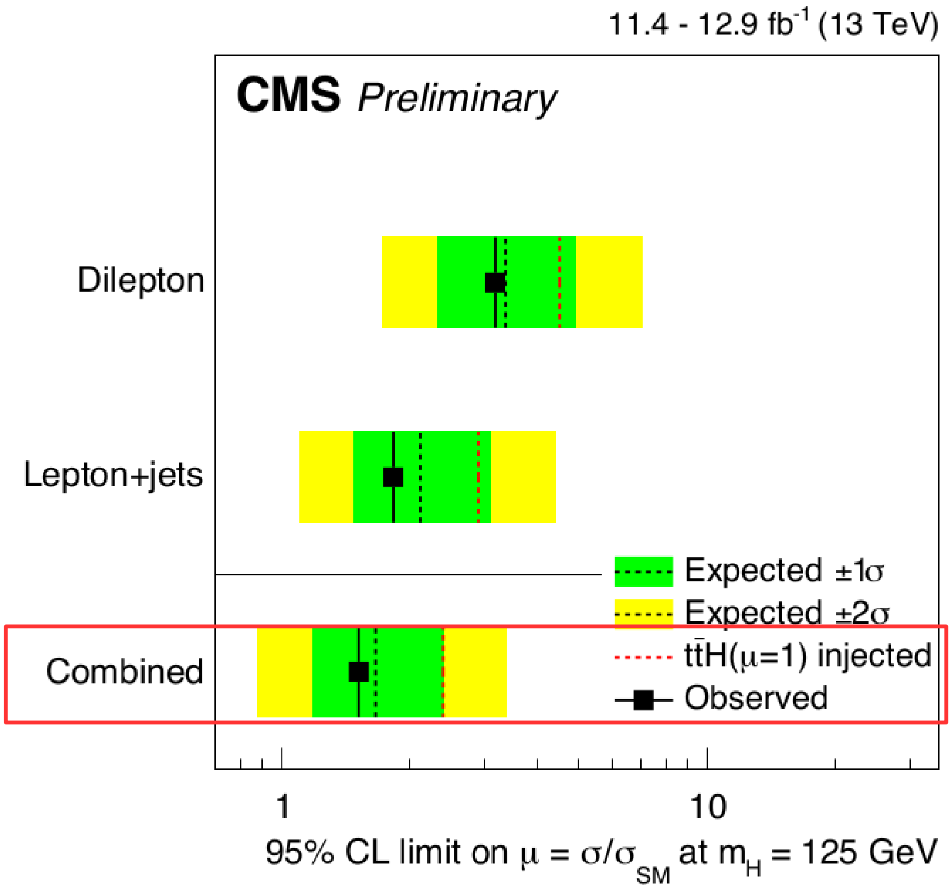
pre-fit expectation

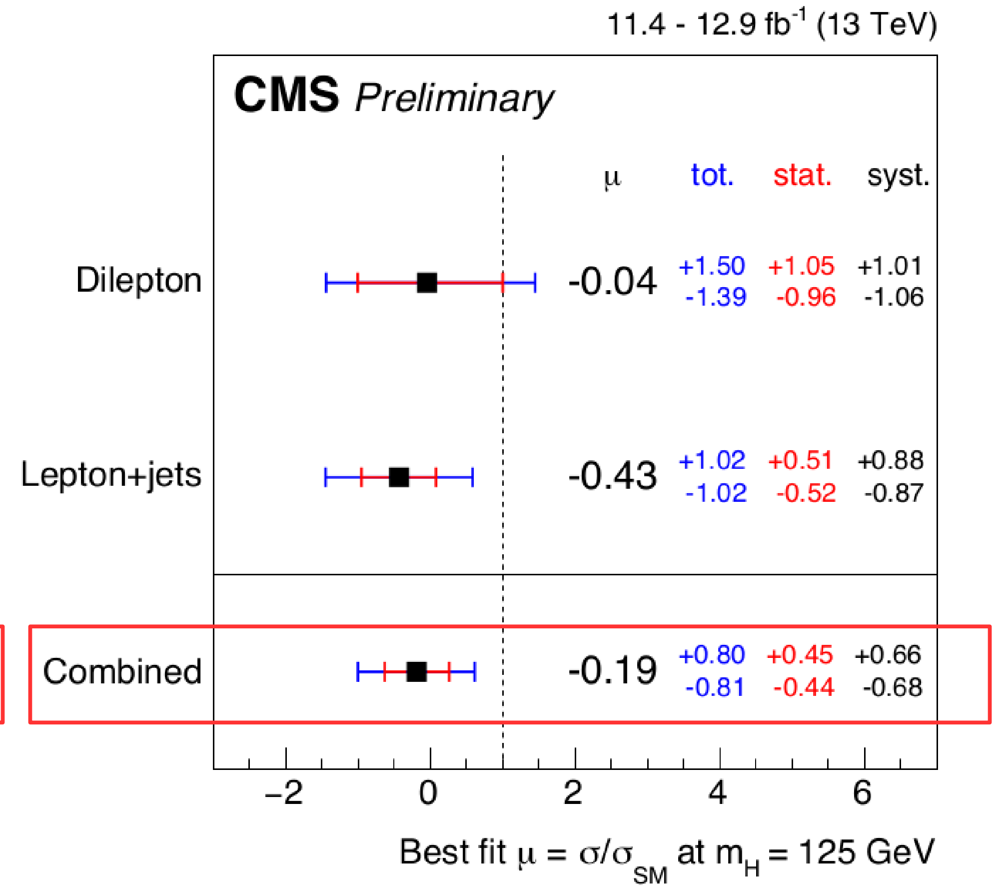


$t\bar{t}H, \; H \to b\bar{b}$

HIG-16-038

<u>Result</u> based on 11.4-12.9 fb^{-1} (2016)



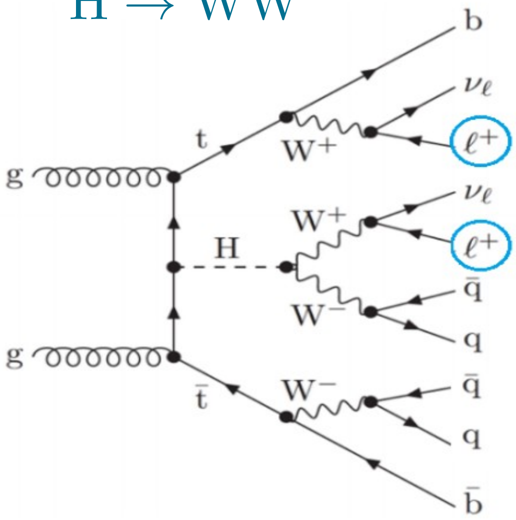


Obs.(exp.) upper limit : $\mu < \mathbf{1.5} (1.7)$ at the 95% CL

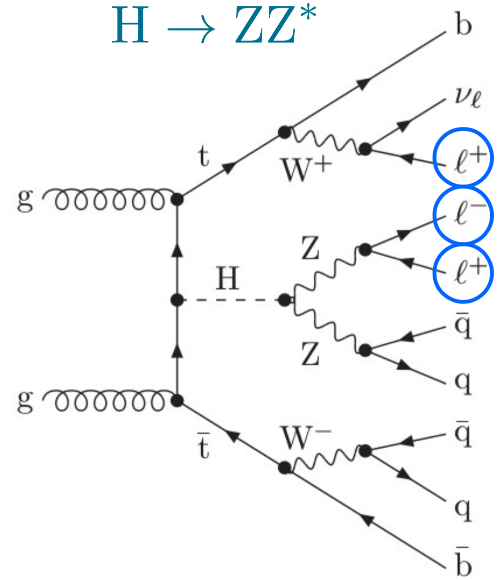
$t\bar{t}H,\ H \rightarrow$ multileptons $(WW^*,\ ZZ^*,\ \tau_{lep}\tau_{lep})$

HIG-17-004

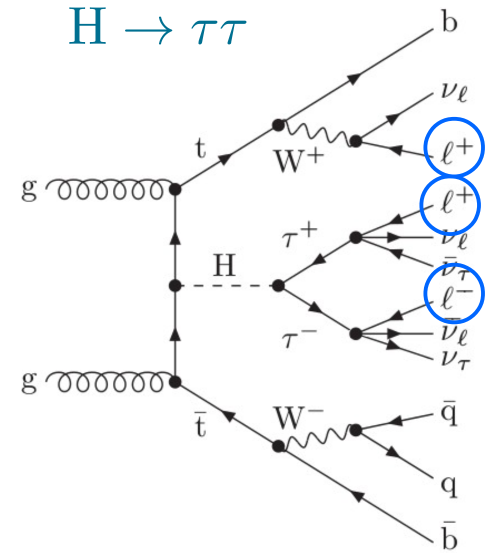
$H \rightarrow WW^*$



$H \rightarrow ZZ^*$



$H \rightarrow \tau\tau$



Event selection

- ≥2 same-sign leptons (2LSS)
- 4 jets

- =3, ≥4 leptons
- 2 jets

- Z veto, based on m_{ll}, MET and H_T
- B-jets: ≥ 1 jet passing medium WP

  ≥ 2 jets passing loose WP
- τ_h veto

Further categorization in 2LSS and 3L channels using:

- Lepton flavor
- B-tag information

Main backgrounds

- Irreducible: $t\bar{t}V$, diboson

  estimated from MC
- Reducible: mis-IDs non-prompt leptons/jets ($t\bar{t}$, single top, W+jets), charge flip in 2LSS (ee)

  estimated from data

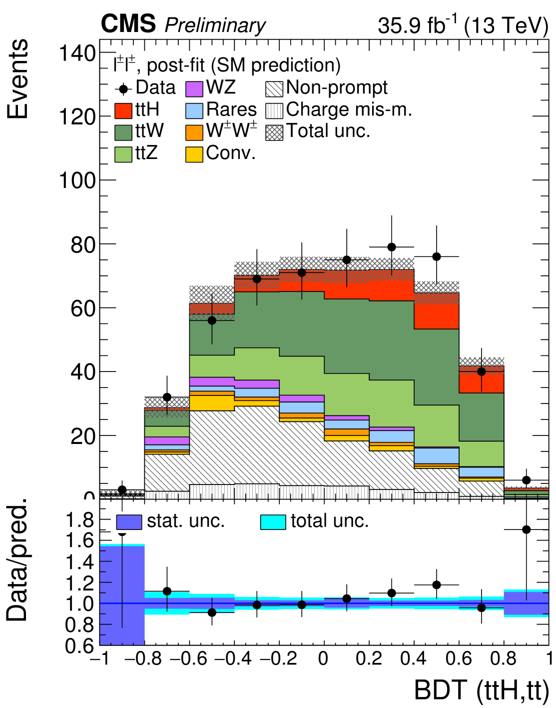
μ±μ±	e±μ±	e±e±	3L
b tight +-	b tight +-	+-	b tight +-
b loose +-	b loose +-		b loose +-

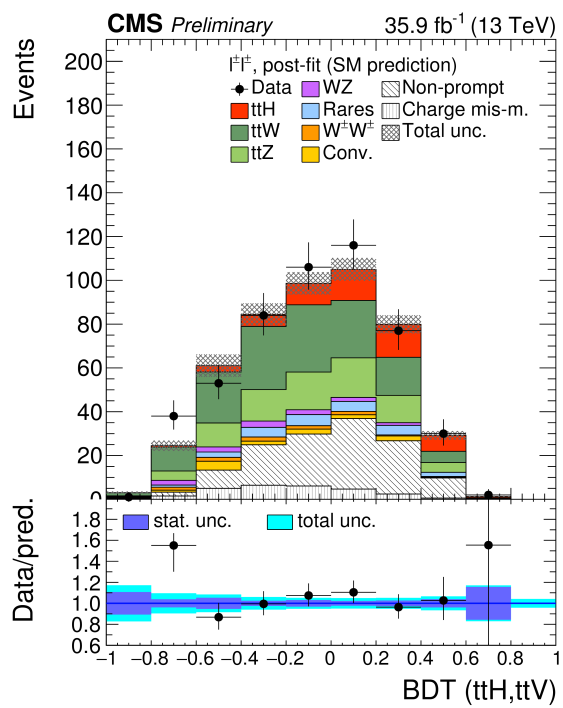
$t\bar{t}H$, $H \rightarrow$ multileptons $(WW^*,\ ZZ^*,\ \tau_{lep}\tau_{lep})$

HIG-17-004

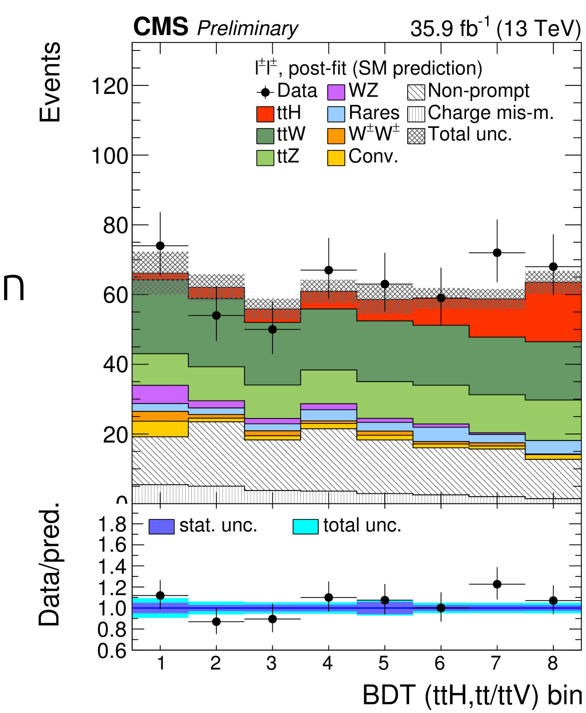
Signal extraction:

- 2 kinematics BDT ($t\bar{t}H$ vs $t\bar{t}/t\bar{t}V$) using :
  - η of leptons, jet multiplicity, distance between lepton and jet, m_T, E_T^{miss}
  - Against $t\bar{t}$: jet BDT discriminator score that is used to identify jets from top decay
  - Against ttV:
    - for 2LSS : jet BDT discriminator score that is used to identify jets from Higgs
    - for 3L : MEM weight $w_{i,\alpha}(\Phi') = \frac{1}{\sigma_\alpha}\int d\Phi_\alpha \cdot \delta^4\left(p_1^\mu + p_2^\mu - \sum_{k\geq 2} p_k^\mu\right) \cdot \frac{f(x_1,\mu_F)f(x_2,\mu_F)}{x_1 x_2 s} \cdot \left|\mathcal{M}_\alpha(p_k^\mu)\right|^2 \cdot W(\Phi'|\Phi_\alpha)$
- 2 kinematics BDT outputs after fit to data, combined from different sub-categories, are shown for 2LSS category:





Combination



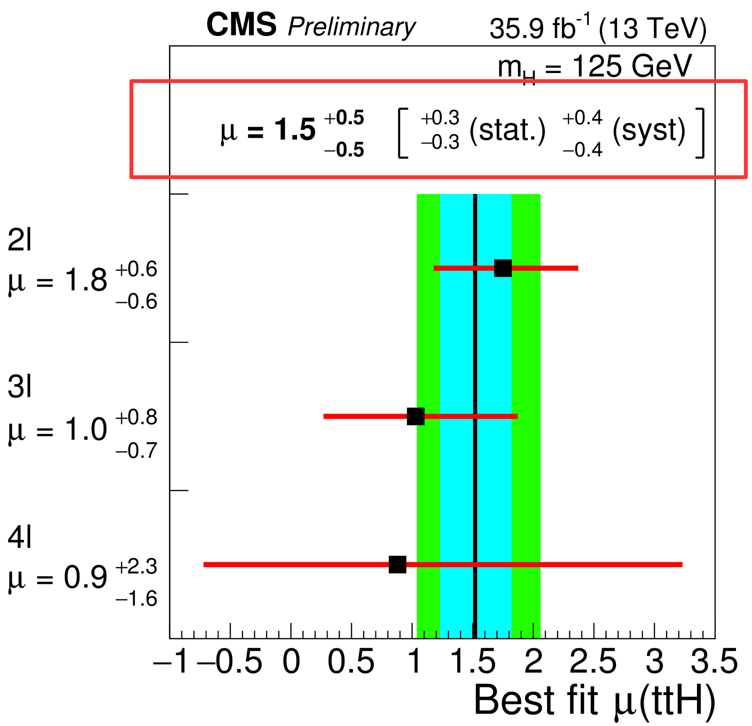
$t\bar{t}H,\ H \to$ multileptons $(WW^*,\ ZZ^*,\ \tau_{lep}\tau_{lep})$

HIG-17-004

Main sources of systematic uncertainties:

- Lepton selection efficiency
- Fake rate measurement for background estimate

Result based on 35.9 fb^{-1}(2016) and combination with 2.7 fb^{-1} (2015)



Category	Observed μ fit $\pm 1\sigma$	Expected μ fit $\pm 1\sigma$
Same-sign di-lepton	1.7 (-0.5) $(+0.6)$	1.0 (-0.5) $(+0.5)$
Three lepton	1.0 (-0.7) $(+0.8)$	1.0 (-0.7) $(+0.8)$
Four lepton	0.9 (-1.6) $(+2.3)$	1.0 (-1.6) $(+2.4)$
Combined (2016 data)	1.5 (-0.5) $(+0.5)$	1.0 (-0.4) $(+0.5)$
Combined (2015 data) [42]	0.6 (-1.1) $(+1.4)$	1.0 (-1.1) $(+1.3)$
Combined (2015+2016 data)	1.5 (-0.5) $(+0.5)$	1.0 (-0.4) $(+0.5)$

Obs.(exp.) significance: 3.3σ (2.4σ) wrt background-only hypothesis

$t\bar{t}H$, $H \to \tau_{had}\tau_{any}$ (also WW^* and ZZ^*)

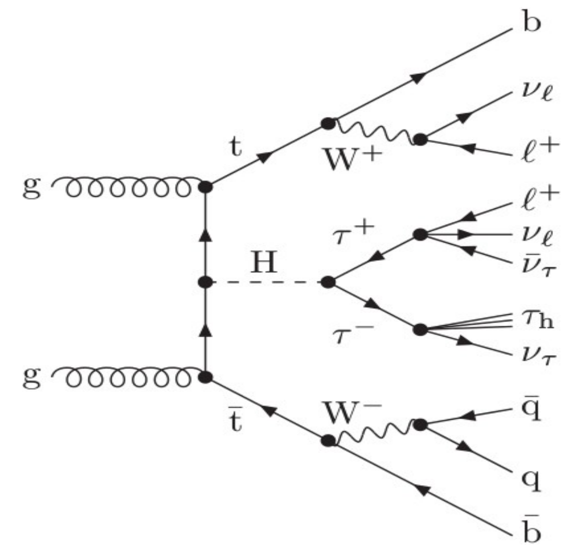
HIG-17-003

τ_h identification:

- Dedicated MVA based discriminant (isolation, impact parameter, τ lifetime)

Categories:

- $1l + 2\,\tau_h$ (+3 jets)
- $2lss + 1\,\tau_h$ (+3 jets)
- $3l + 1\,\tau_h$ (+2 jets)

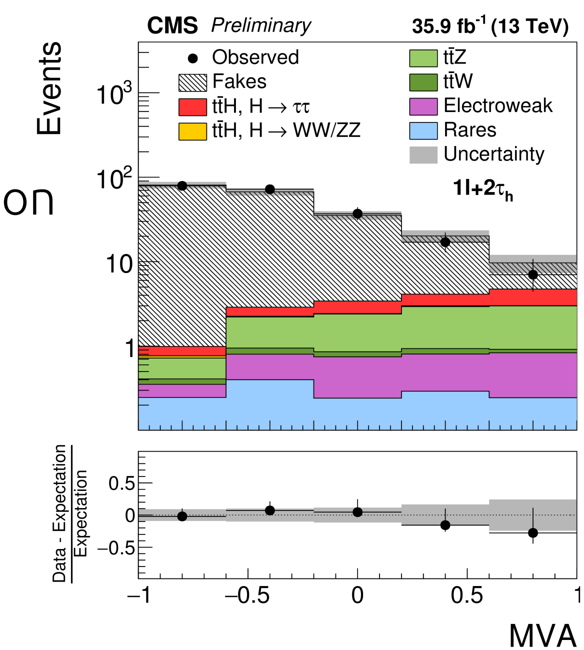


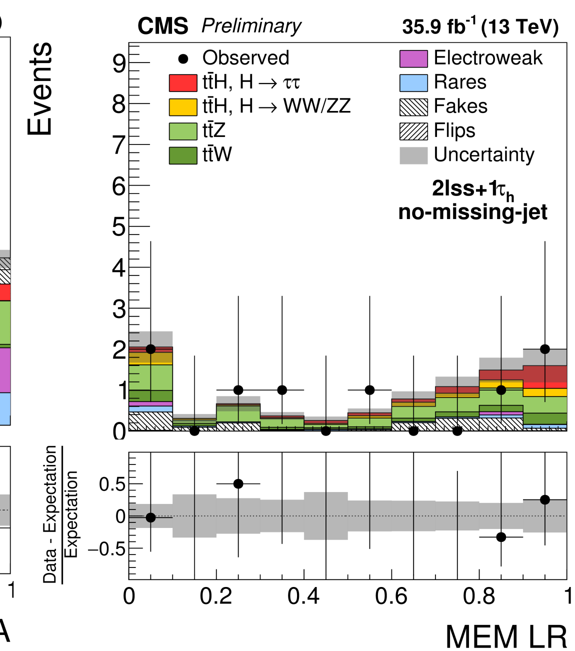
Main backgrounds:

- irreducible (*from MC*): $t\bar{t}V$, diboson
- reducible (*from data*): fakes leptons and τ_h, charge flips in $2lss + 1\,\tau_h$

<u>Signal extraction:</u>

- Maximum likelihood fit to the distribution of discriminating observables
- Discriminants to separate signal from $t\bar{t}V/t\bar{t}$ backgrounds in each category
- MEM is used in $2lss + 1\,\tau_h$ category to discriminate from $t\bar{t}Z/t\bar{t}$





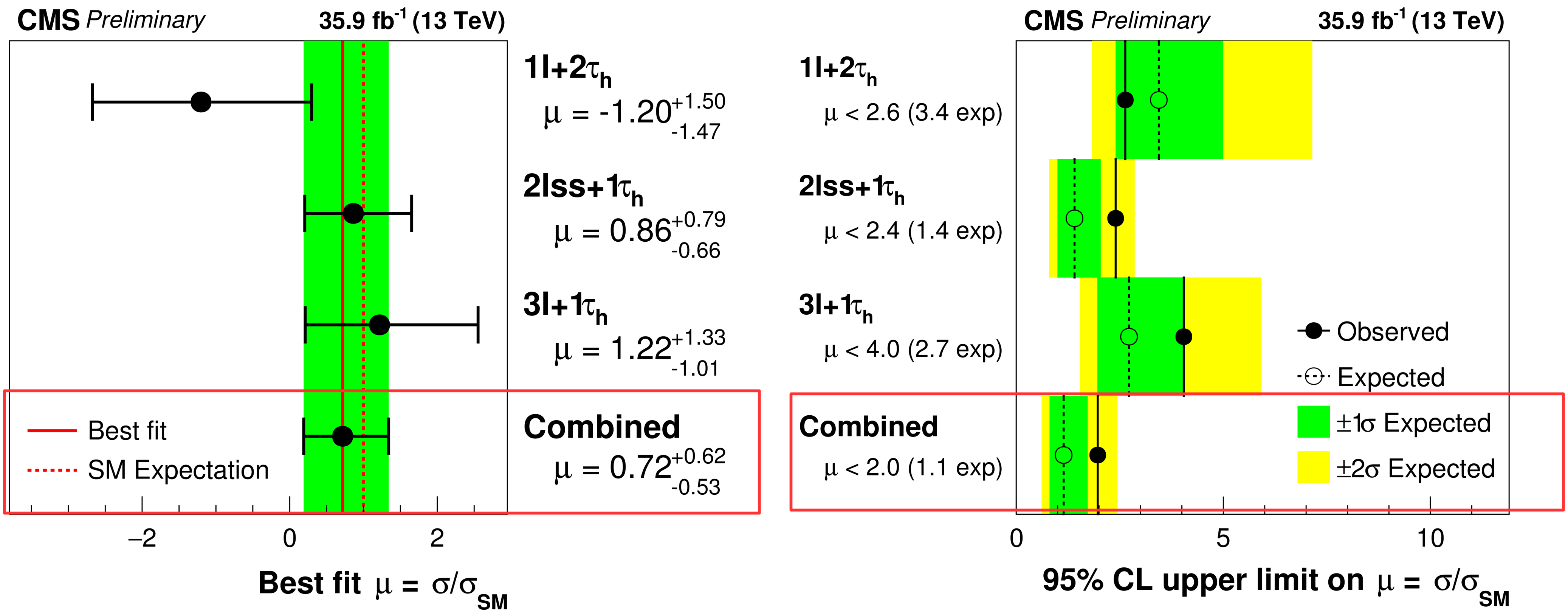
$t\bar{t}H$, $H \rightarrow \tau_{had}\tau_{any}$ (also WW^* and ZZ^*)

HIG-17-003

Main sources of systematic uncertainties:

- Charge mis-measurement, lepton and τ_h identification efficiency
- Jet and τ_h energy scale, b-tag efficiency, fake background

Result based on 35.9 fb^{-1} (2016)



Obs.(exp.) significance: 1.4σ (1.8σ) wrt background-only hypothesis

$\mathrm{t\bar{t}H,\ H \rightarrow \gamma\gamma}$

HIG-16-040

Part of the $\mathrm{H \rightarrow \gamma\gamma}$ analysis: (see Jonatan Piedra Gomez talk)

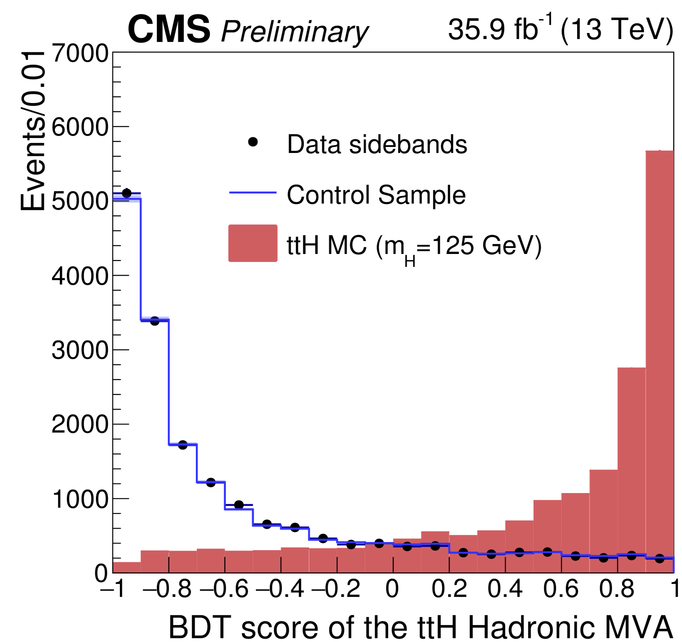
- events with high-energy isolated photon pairs
- $\mathrm{t\bar{t}H}$, VH and VBF production modes identified by additional final state objects
- signal and background extraction from fit of the $\mathrm{m_{\gamma\gamma}}$ distribution

Main backgrounds: $\gamma\gamma,\ \gamma + \mathrm{jets},\ \mathrm{t\bar{t}} + \gamma\gamma,\ \mathrm{t\bar{t}} + \mathrm{jets}$

Two categories:

Hadronic: $\mathrm{t\bar{t} \rightarrow bq\bar{q}'\bar{b}q\bar{q}'}$

- no leptons (e/μ)
- ≥ 3 jets
- ≥ 1b-jet (loose WP)
- $\mathrm{t\bar{t}H}$ BDT cut (info: jet multiplicity, b-tagging, lead. jet pT)
- Diphoton BDT cut



Leptonic: $\mathrm{t\bar{t} \rightarrow l\nu_l b\bar{b}l'\nu_{l'}}$ or $\mathrm{t\bar{t} \rightarrow l\nu_l b\bar{b}q\bar{q}'}$

- ≥ 1 isolated lepton
- ≥ 2 jets
- ≥1 b-jet (medium WP)
- Diphoton BDT cut

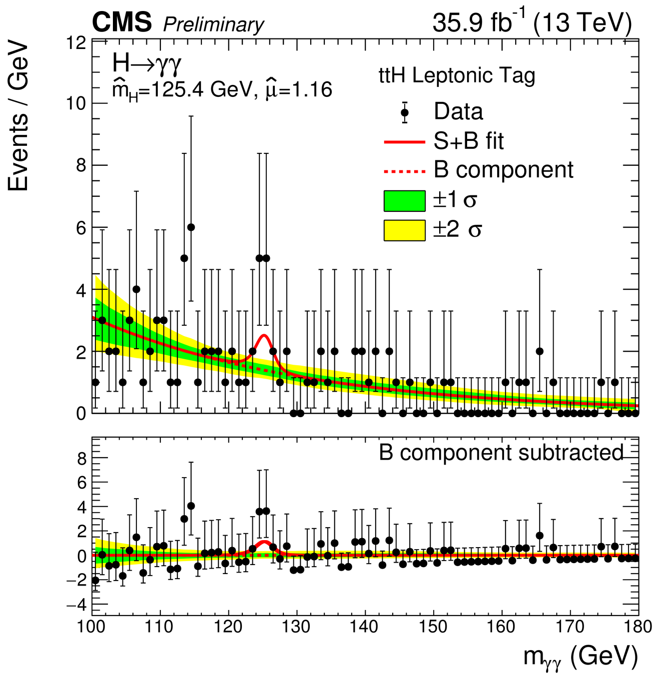
Expected signal yields: good signal purity

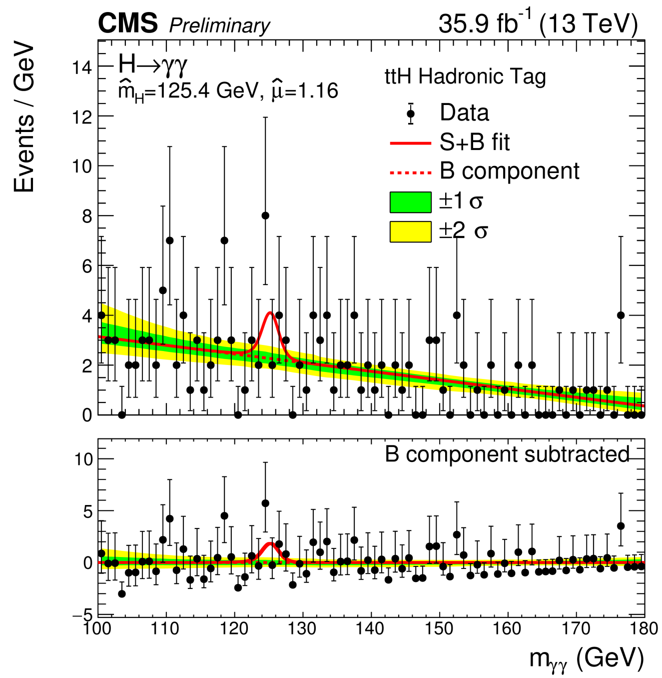
Category	SM 125GeV Higgs boson expected signal													Bkg
	Total	ggH	VBF	$\mathrm{t\bar{t}H}$	$\mathrm{b\bar{b}H}$	tHq	tHW	WH lep	ZH lep	WH had	ZH had	σ_{eff}	σ_{HM}	(GeV^{-1})
$\mathrm{t\bar{t}H}$ Hadronic	5.85	10.99 %	0.70 %	77.54 %	2.02 %	4.13 %	2.02 %	0.09 %	0.05 %	0.63 %	1.82 %	1.48	1.30	2.40
$\mathrm{t\bar{t}H}$ Leptonic	3.81	1.90 %	0.05 %	87.48 %	0.08 %	4.73 %	3.04 %	1.53 %	1.15 %	0.02 %	0.02 %	1.60	1.35	1.50

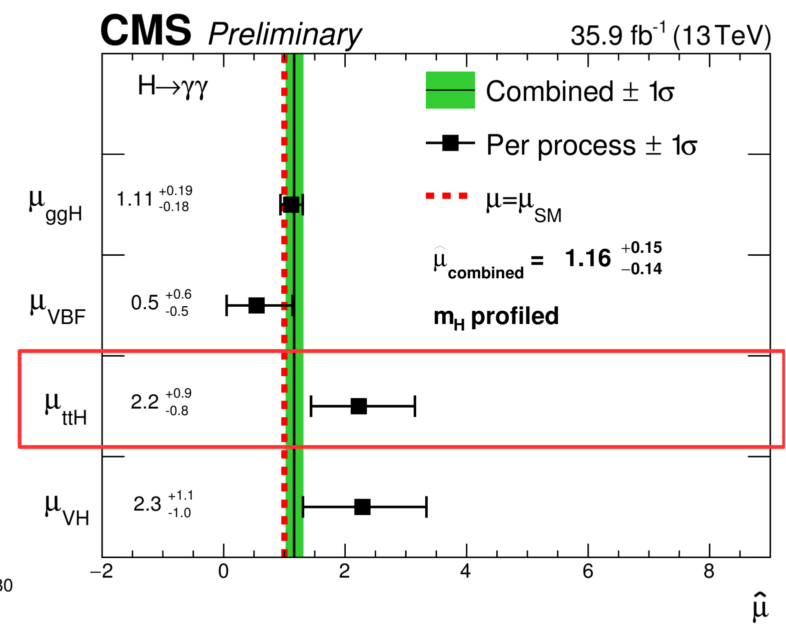
$t\bar{t}H,\ H \to \gamma\gamma$

HIG-16-040

Result based on $35.9\ \mathrm{fb}^{-1}$ (2016)







Uncertainties are dominated by statistical component

Obs.(exp.) significance: **3.3σ** (1.5σ) wrt background-only hypothesis

$\mathrm{t\bar{t}H},\ \mathrm{H} \to \mathrm{ZZ}^* \to 4\mathrm{l}$

HIG-16-041

submitted to JHEP

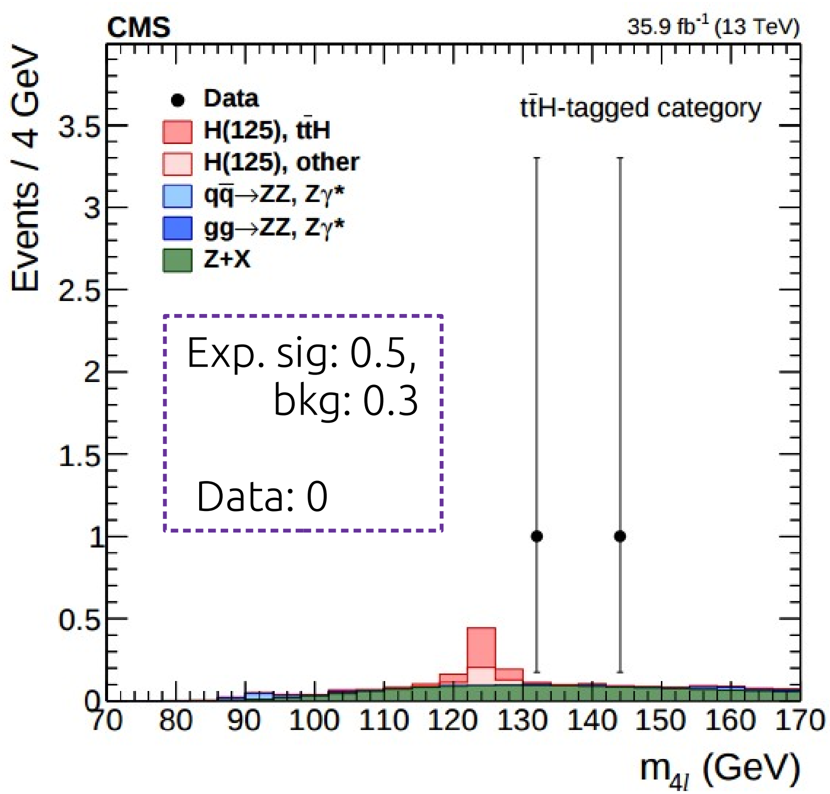
Part of the H → ZZ*→ 4l analysis (see Jonatan Piedra Gomez talk)

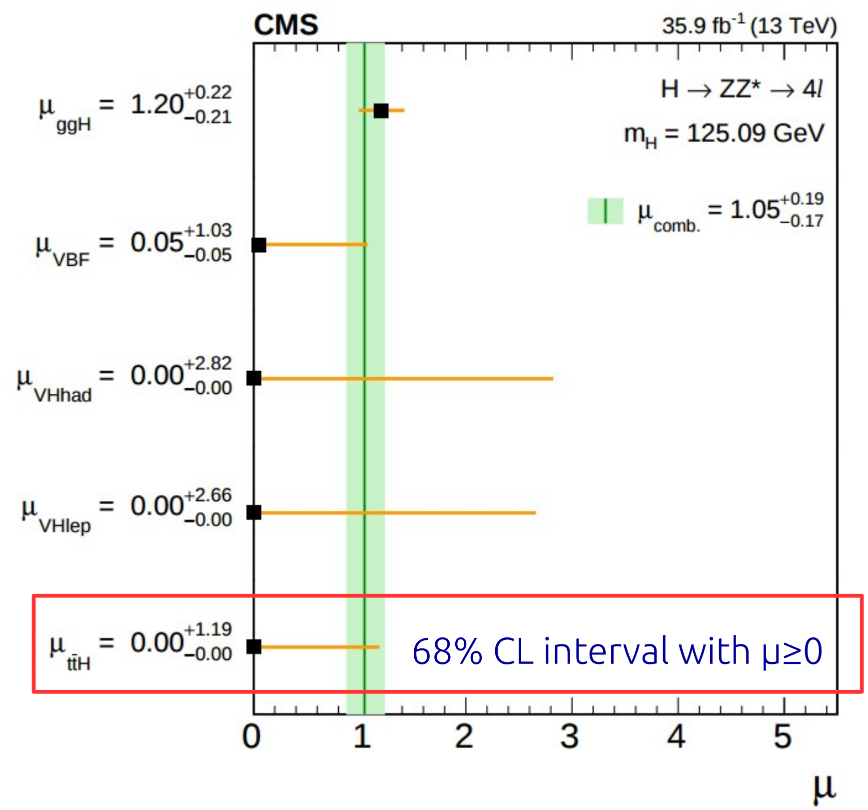
Triggers require a pair of electrons or muons, or electron-muon pair

Main backgrounds: $\mathrm{q\bar{q}} \to \mathrm{ZZ};\ \mathrm{Z+jets},\ \mathrm{t\bar{t}+jets},\ \mathrm{WW+jets},\ \mathrm{WZ+jets}$ (noted Z+X)

Event selections:

- ≥ 4 jets
- ≥ 1b-jet or ≥ 1 additional lepton (e/μ)





$t\bar{t}H$ Summary

	$\mu_{t\bar{t}H} = \sigma_{t\bar{t}H}/\sigma_{SM}$	significance/upper limit	Reference
$b\bar{b}$	-0.2 ± 0.8	$\mu < 1.5(1.7)@95\%CL$	HIG-16-038
multilept	1.5 ± 0.5	$3.3\sigma(2.4\sigma)$	HIG-17-004
$\tau_{had}\tau_{any}$	$0.7^{+0.6}_{-0.5}$	$1.4\sigma(1.8\sigma)$	HIG-17-003
$\gamma\gamma$	$2.2^{+0.9}_{-0.8}$	$3.3\sigma(1.5\sigma)$	HIG-16-040
$4l$	$0.00^{+1.2}_{-0.0}$		HIG-16-041

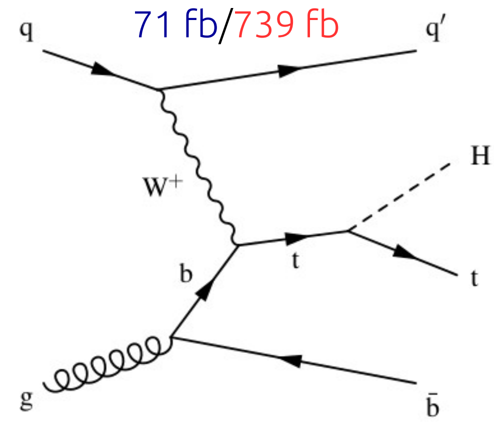
* Luminosity: 12.9 fb^{-1} and 35.9 fb^{-1}

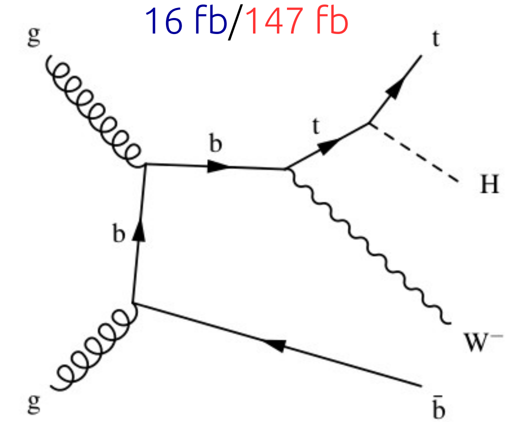
ttH combination: after bb analysis with 35.9 fb^{-1} is available tH

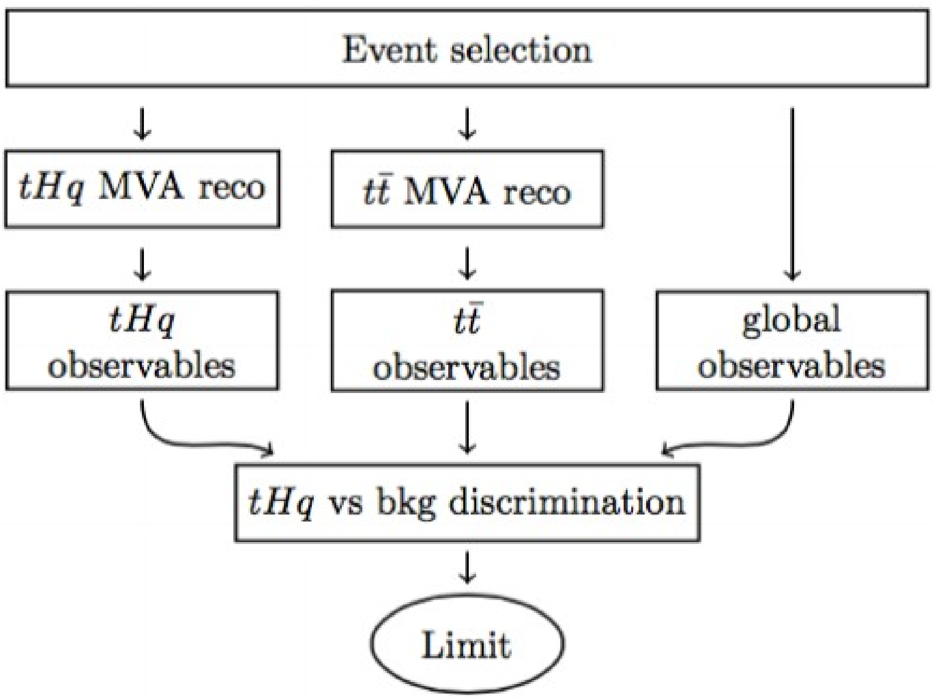
tH, $H \to b\bar{b}$

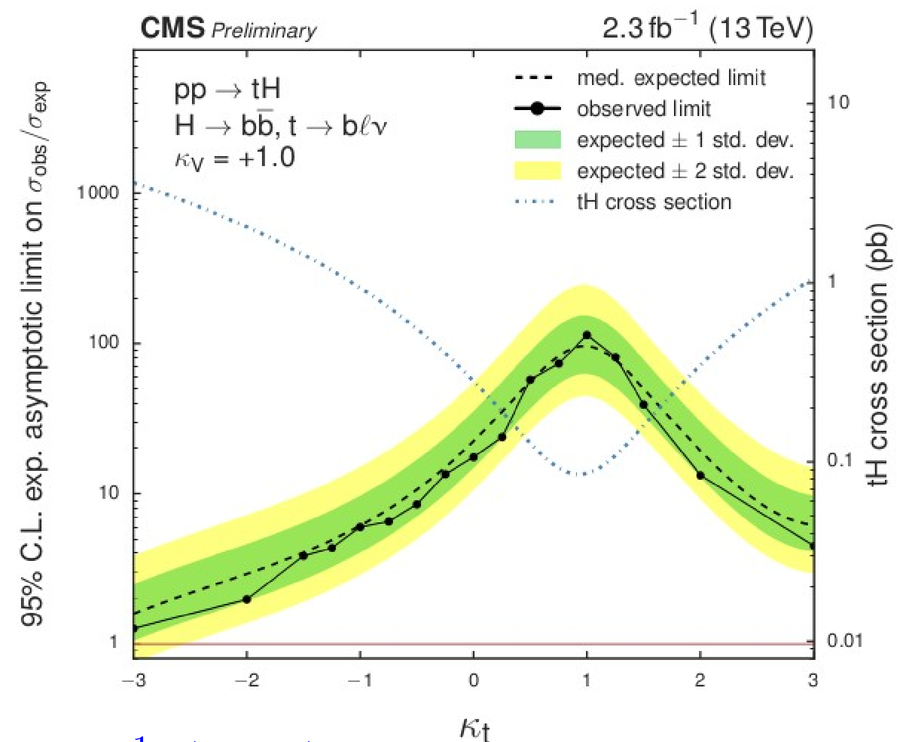
HIG-16-019

- tHq and tHW with $\kappa_t = 1$ and $\kappa_t = -1$
- Event selection ($t \to Wb \to l\nu b$):
  - ≥3, ≥4 b-tags + ≥1 jet
  - =1 lepton (μ/e)
  - MET > 35 GeV, > 45 GeV
- Main background: $t\bar{t}$









Upper limits based on 2.3 fb^{-1} (2015):

$$\kappa_t = 1 : \sigma_{obs}(\sigma_{exp}) < 113.7 \times \sigma_{SM}(98.6 \times \sigma_{SM})$$

$$\kappa_t = -1 : \sigma_{obs}(\sigma_{exp}) < 6.0 \times \sigma_{SM}(6.4 \times \sigma_{SM})$$

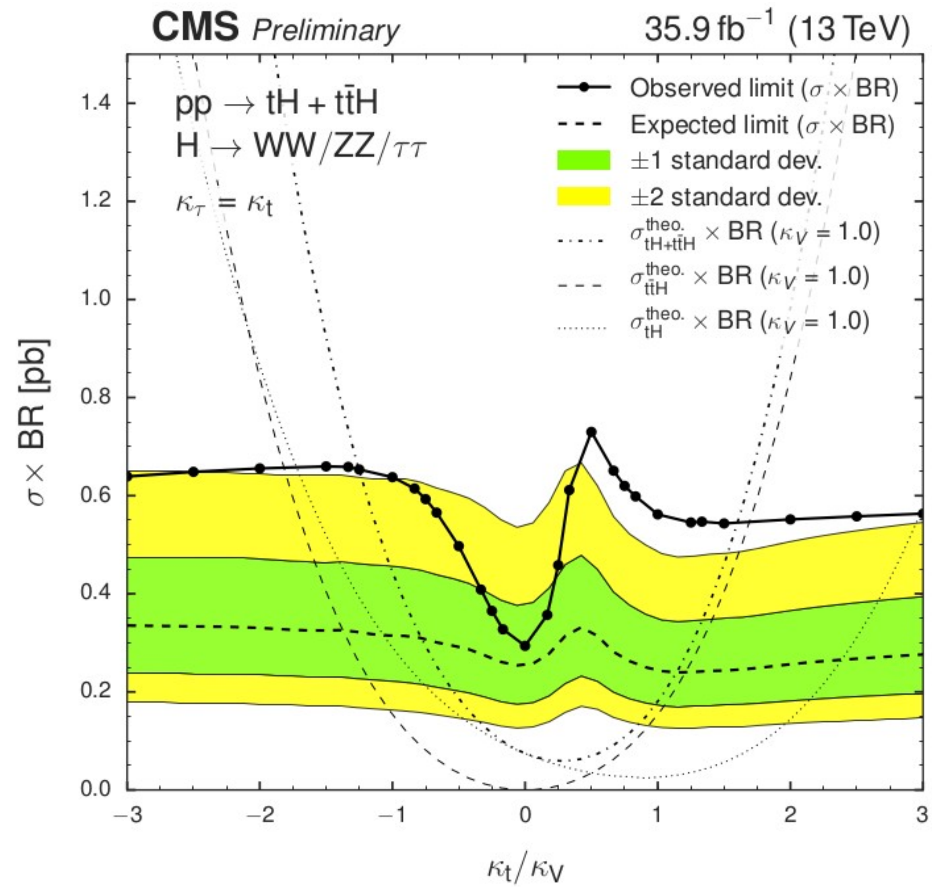
tH, $H \to$ multileptons (ZZ^*, $\tau\tau$, WW^*)

HIG-17-005

Follows closely $t\bar{t}H$, $H \to$ multileptons analysis strategy.

tHq and tHW with $\kappa_t/\kappa_V = 1$ and $\kappa_t/\kappa_V = -1$

- Event selection ($t \to Wb \to l\nu b$):
  - 3 l / 2lss ($e\mu/\mu\mu$)
  - 1, 2 jets
  - 1 b-jet
- Signal discrimination (tHq vs $t\bar{t}V/t\bar{t}$) using:
  - lepton kinematics
  - jet multiplicities
  - forward jet kinematics
- Limit on common signal strength for (tHq+tHW+ttH) as function of κ_t/κ_V
- ttH is included as signal since its cross section varies as κ_t^2



Allowed : $-1.25 < \kappa_t < 1.6$ for $\kappa_V = 1$

Result based on 35.9 fb^{-1} (2016):

	$\hat{\mu}$	significance
κ_t/κ_V = 1	1.8 $\pm$0.3 $(stat.) \pm 0.6$ $(syst.)$	$\mathbf{2.7}\sigma(1.5\sigma)$
κ_t/κ_V = -1	0.7 $\pm$0.4	$\mathbf{1.7}\sigma(2.5\sigma)$

Summary

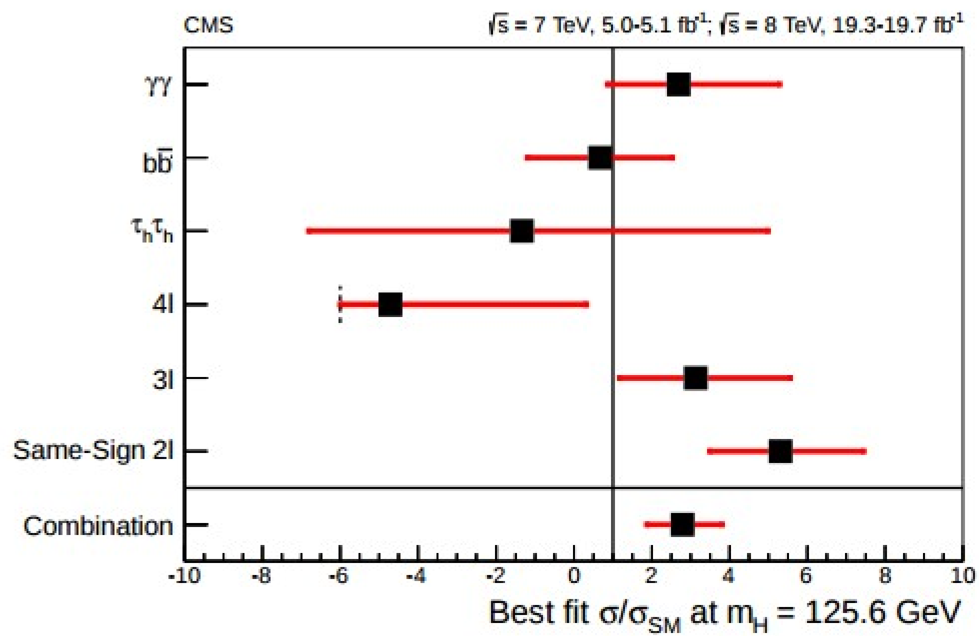
- $t\bar{t}H$ is the Higgs-physics analysis to directly access the top quark Yukawa coupling, it has gained a important role in Run II LHC Higgs-physics analyses.
- tH analysis can give a hint of the top Yukawa coupling sign. The first limits with Run II data are shown.
- With a larger dataset it should be possible to have a clear evidence for ttH production at Run II:
  - 2017 data taking @ LHC is ongoing,
  - Work on-going to reduce the systematic uncertainties in bb and multilepton channels,
  - Clear peak in $\gamma\gamma$ and $4l$.

Bonus slides

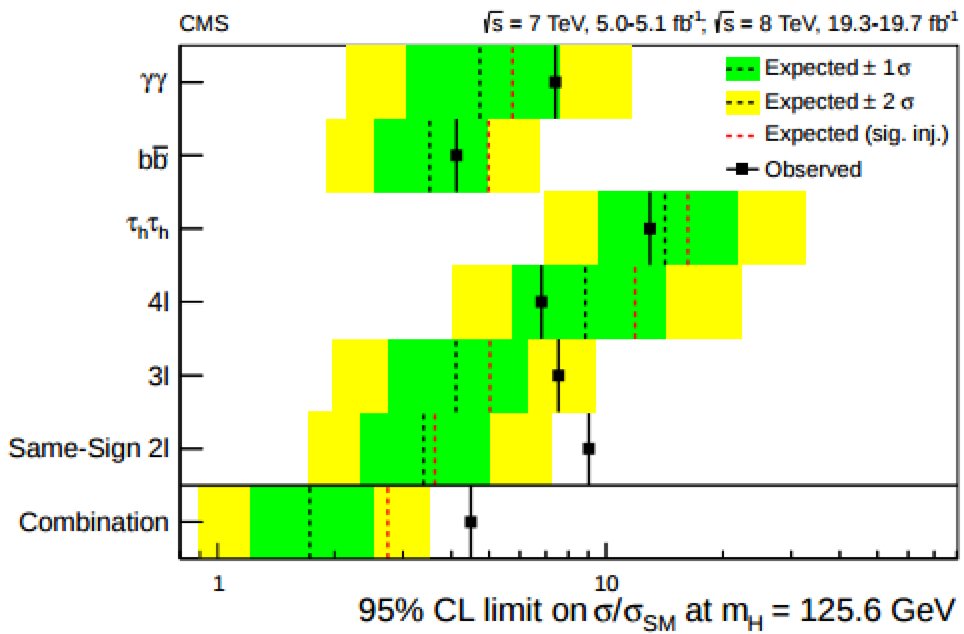
$t\bar{t}H$ Run I combination

JHEP09(2014)087

The searches have been done for the H → bb, ττ (hadronic), γγ, WW, and ZZ decay modes.



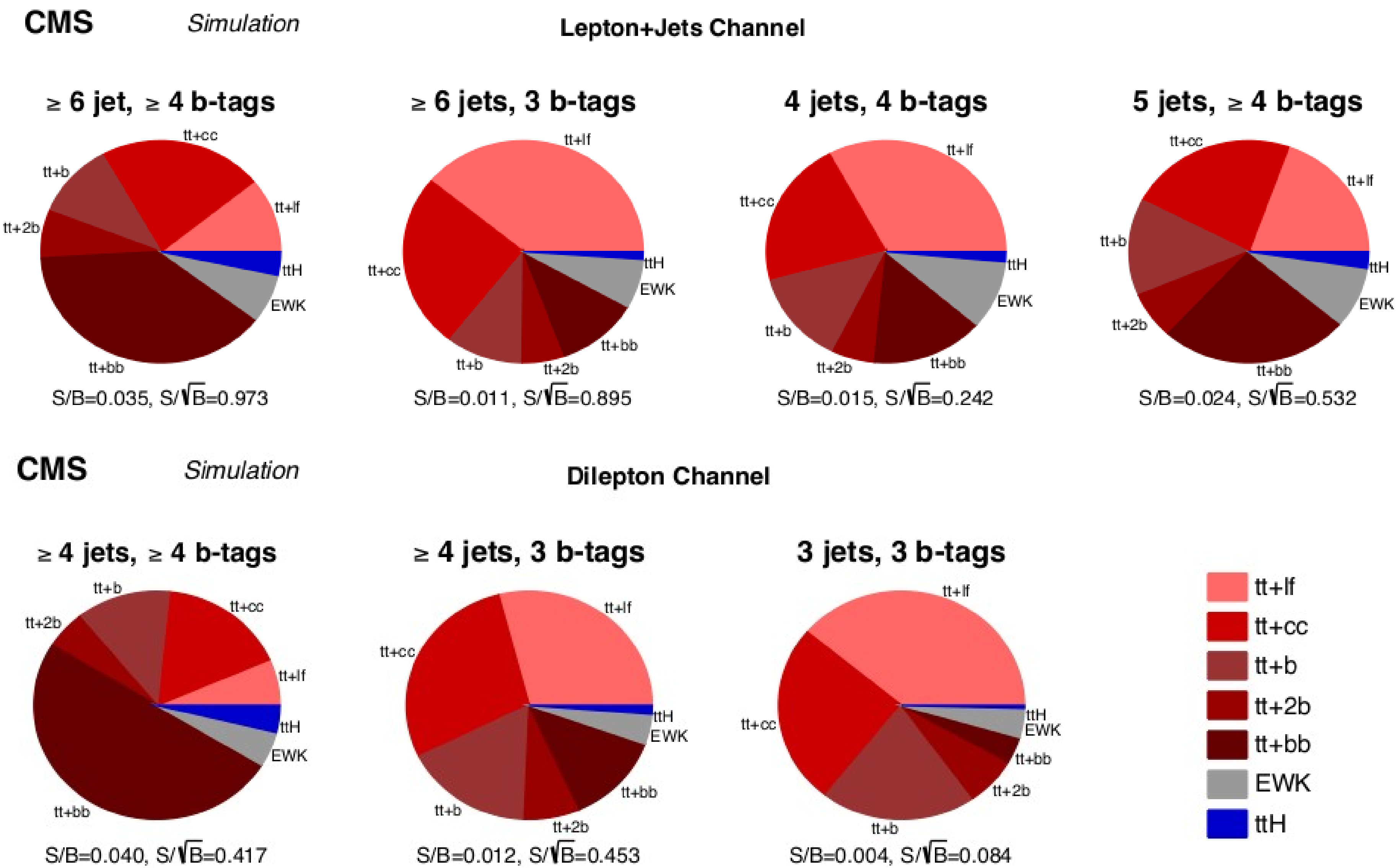
$$\mu_{t\bar{t}H} = 2.8 \pm 1.0$$



$$95\%\ \mathrm{CL\ limit} = 4.5$$

$t\bar{t}H,\ H \to b\bar{b}$: event classification

HIG-16-038



$t\bar{t}H,\ H \to b\bar{b}$: BDT

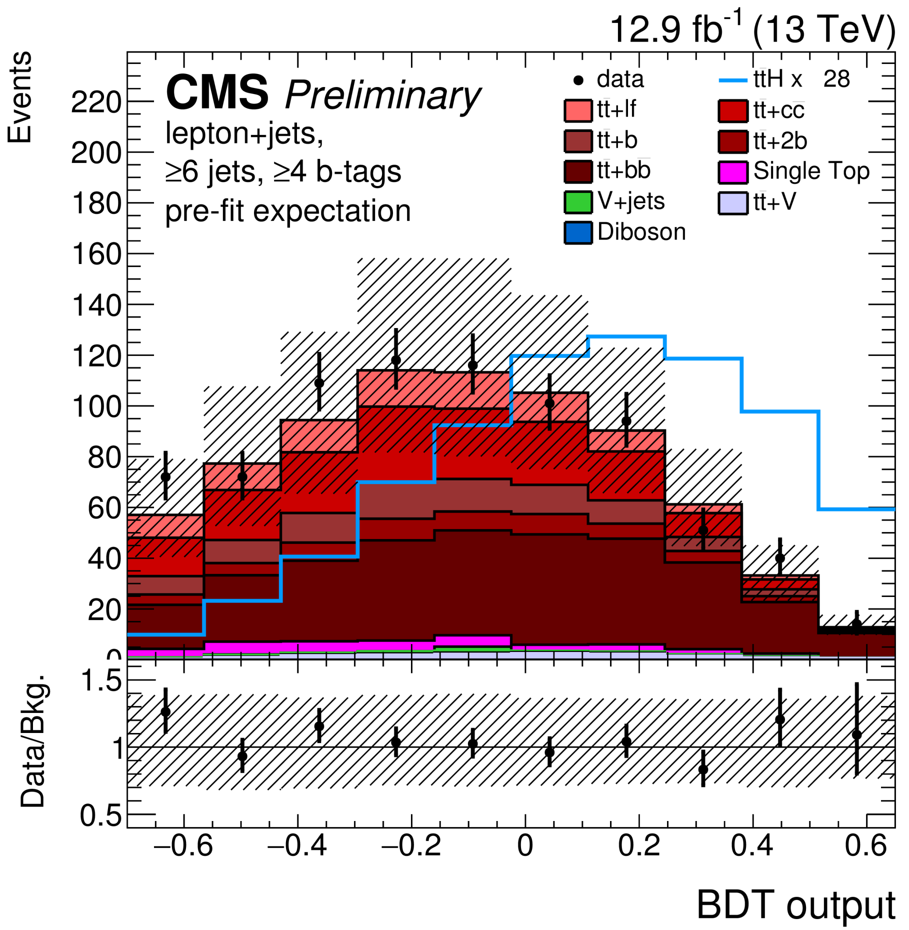
HIG-16-038

Gradient boosting decision trees:

- Train for ttH(bb) vs sum-of-backgrounds
- Individual training and variable selection for each jet/b-tag category
- Only consider well modeled variables

Ex: BDT input in Lepton+jets ≥6 jets ≥4 b-tags:

- best Higgs mass
- M(tag, tag) closest to 125 GeV
- M(jets, lepton, MET)
- 4th and 5th highest b-tag discriminator score
- ∑ p T (jets, lepton, MET)



$t\bar{t}H,\ H \to b\bar{b}$: MEM $t\bar{t}H,\ H \to b\bar{b}$ vs $t\bar{t}b\bar{b}$

Construct per-event signal/background probability using full kinematic information in an analytical approach

Numerical integration

Momentum conservation

Resolution function (allow ISR)

$$w(\vec{y}|\mathcal{H}) = \sum_{i=1}^{N_a} \int \frac{dx_a dx_b}{2x_a x_b s} \int \prod_{k=1}^{8} \left(\frac{d^3\vec{p}_k}{(2\pi)^3 2E_k} \right) (2\pi)^4 \delta^{(E,z)} \left(p_a + p_b - \sum_{k=1}^{8} p_k \right) \mathcal{R}^{(x,y)} \left(\vec{\rho}_T, \sum_{k=1}^{8} p_k \right) \times$$
$$g(x_a, \mu_F) g(x_b, \mu_F) |\mathcal{M}_{\mathcal{H}}(p_a, p_b, p_1, \ldots, p_8)|^2 W(\vec{y}, \vec{p}),$$

Parton density functions

LO Scattering amplitude (Open Loops)

Detector transfer function

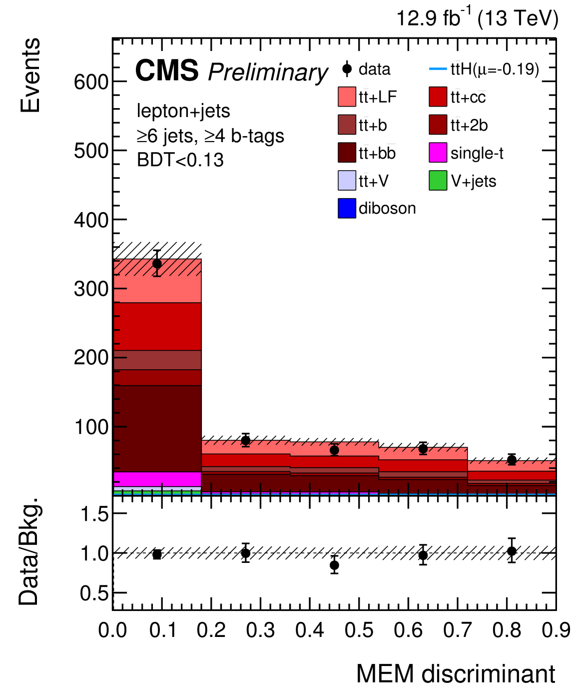
$$P_{s/b} = \frac{w(\vec{y}|t\bar{t}H)}{w(\vec{y}|t\bar{t}H) + k_{s/b} w(\vec{y}|t\bar{t}+b\bar{b})}$$

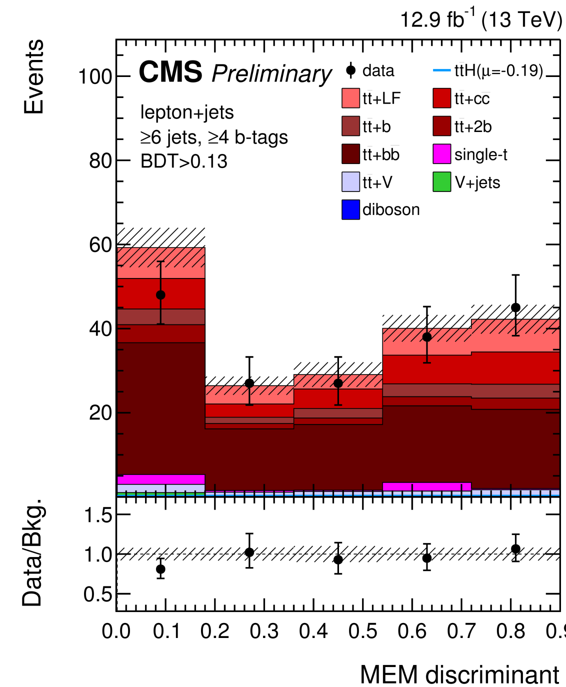
$t\bar{t}H,\ H \rightarrow b\bar{b}$: leptons + jets/dileptons fit

HIG-16-038

Leptons+jets:

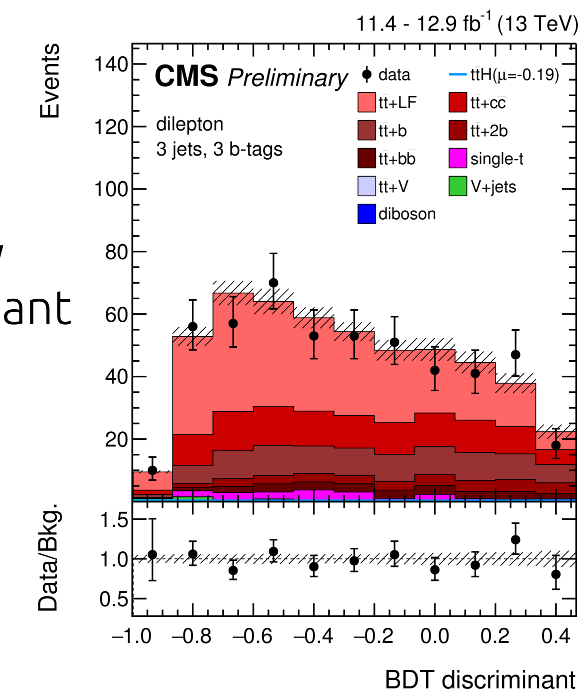
- Divide events by jet/b-tag multiplicity
- Subdivide by low/high BDT
- Fit MEM discriminant

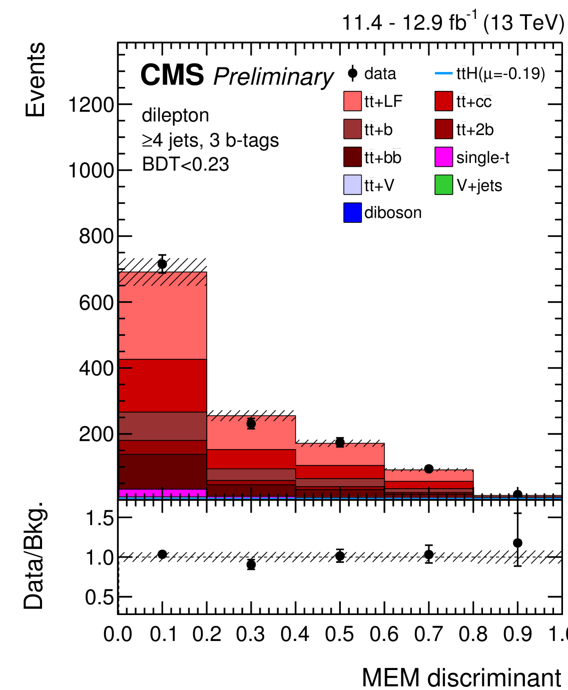




Dileptons:

- Divide events by jet/b-tag multiplicity
- For the case “3 jets” Fit BDT discriminant
- For the case “≥4 jets”:
  - Subdivide by low/high BDT
  - Fit MEM discriminant





$t\bar{t}H$, $H \to$ multileptons (WW^*, ZZ^*, $\tau_{lep}\tau_{lep}$)

HIG-17-004

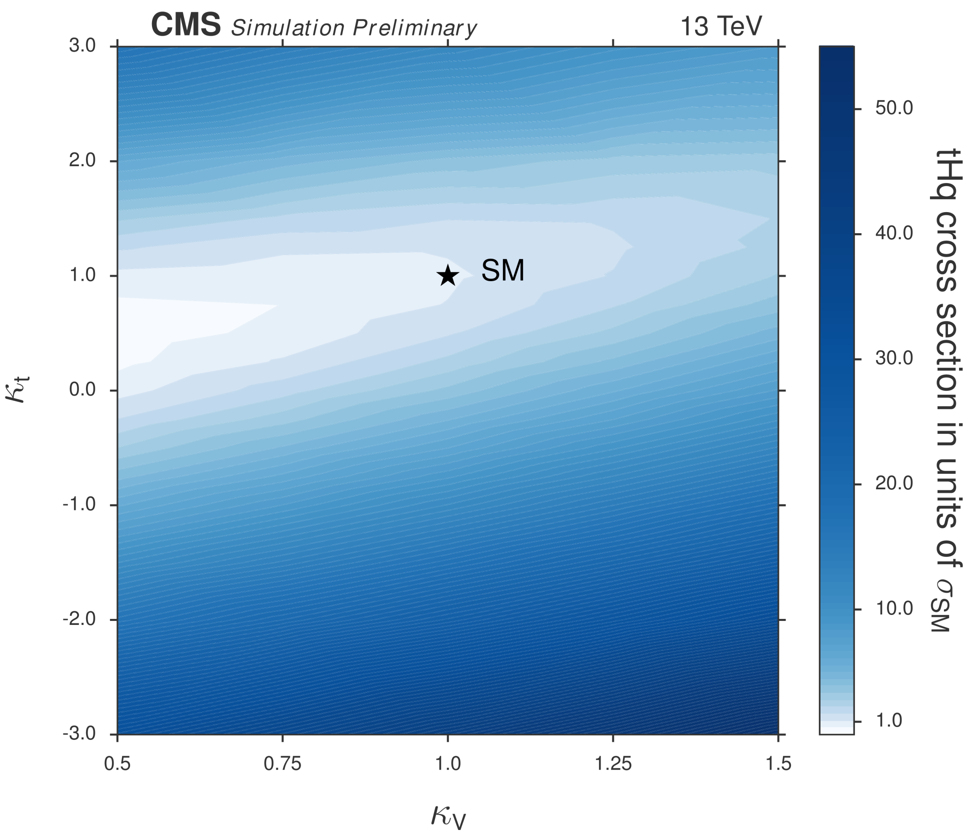
Limits with 2016 dataset:

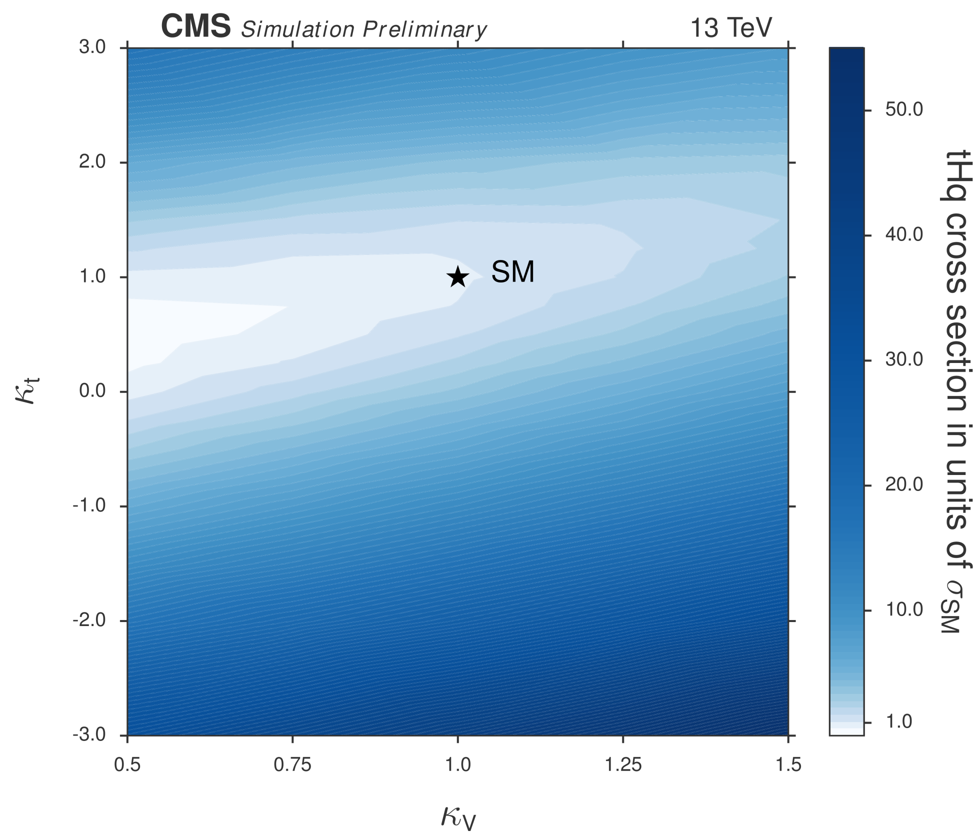
Category	Observed limit	Expected limit $\pm 1\sigma$
Same-sign di-lepton	2.8	0.9 (−0.3) (+0.4)
Three lepton	2.5	1.4 (−0.4) (+0.7)
Four lepton	5.9	4.9 (−1.7) (+3.1)
Combined	2.5	0.8 (−0.2) (+0.3)

tH, H → $b\bar{b}$

HIG-16-019

- Cross sections in the κ_t – κ_V plane at 13 TeV for tHq (left) and tHW (right) production.

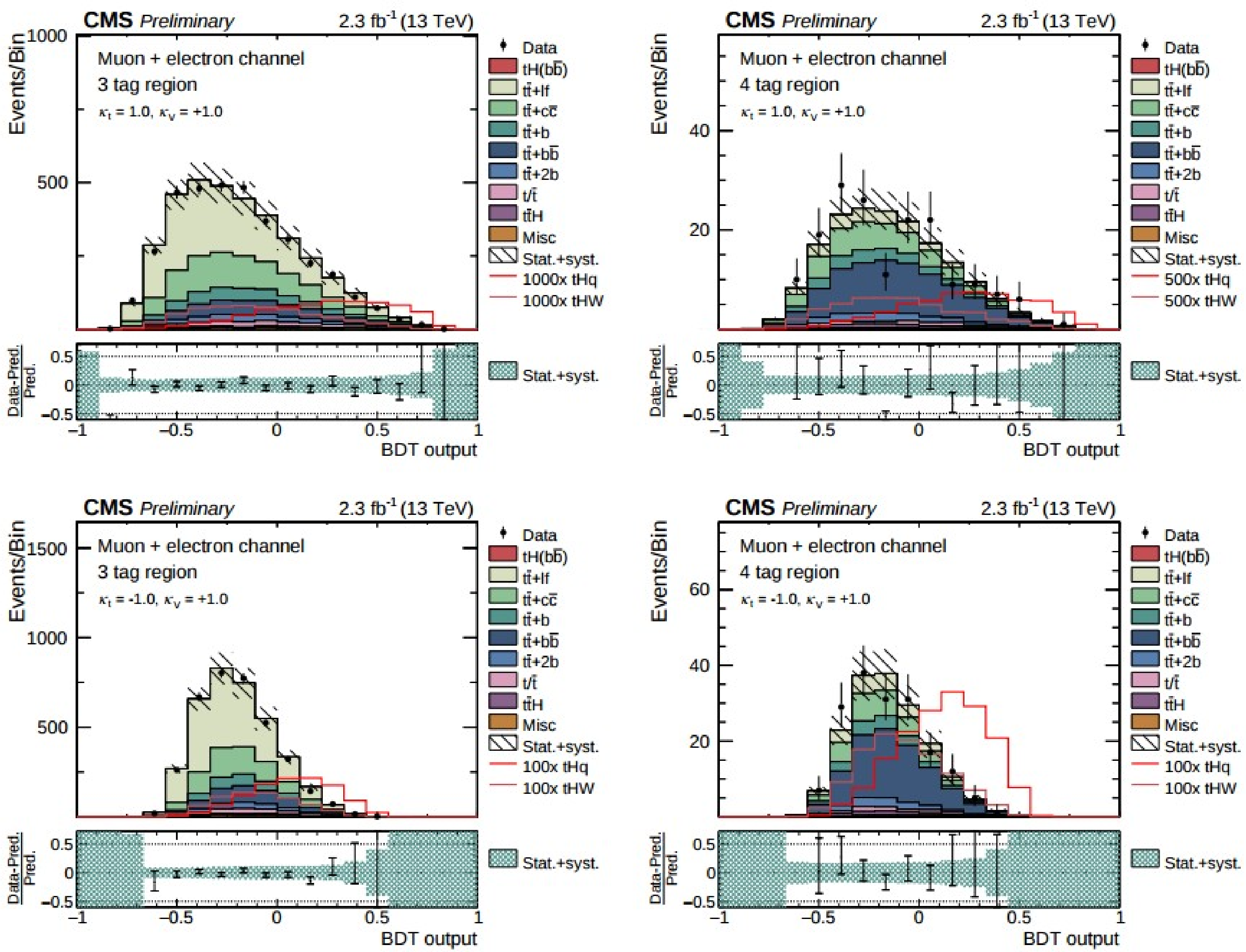




$tH,\ H \to b\bar{b}$

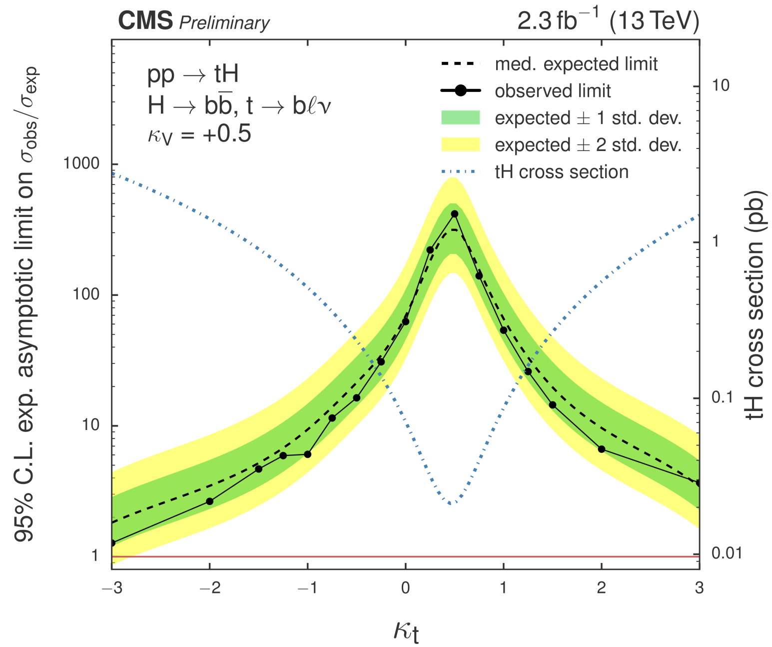
HIG-16-019

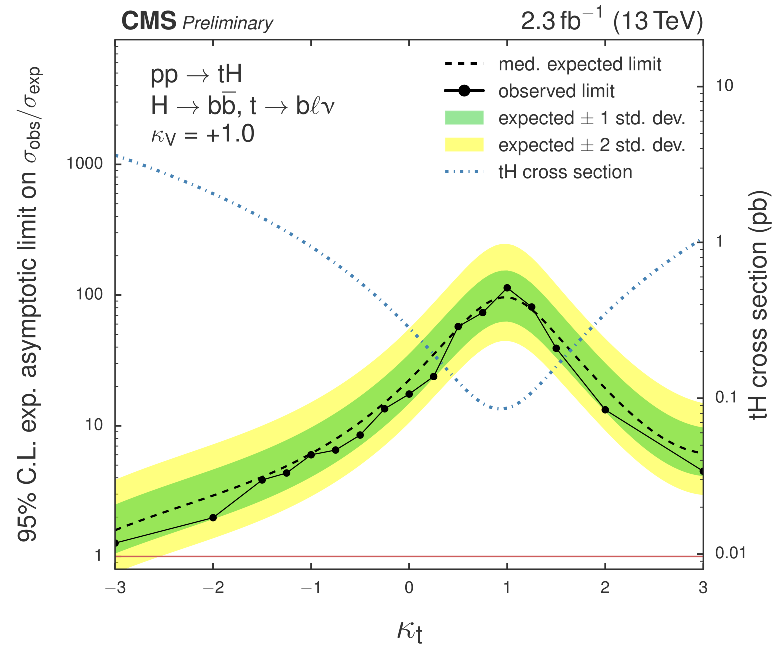
- Postfit distributions of the classification BDT response in the 3 tag (left) and 4 tag (right) regions for the SM (top) and ITC (bottom) coupling scenarios.

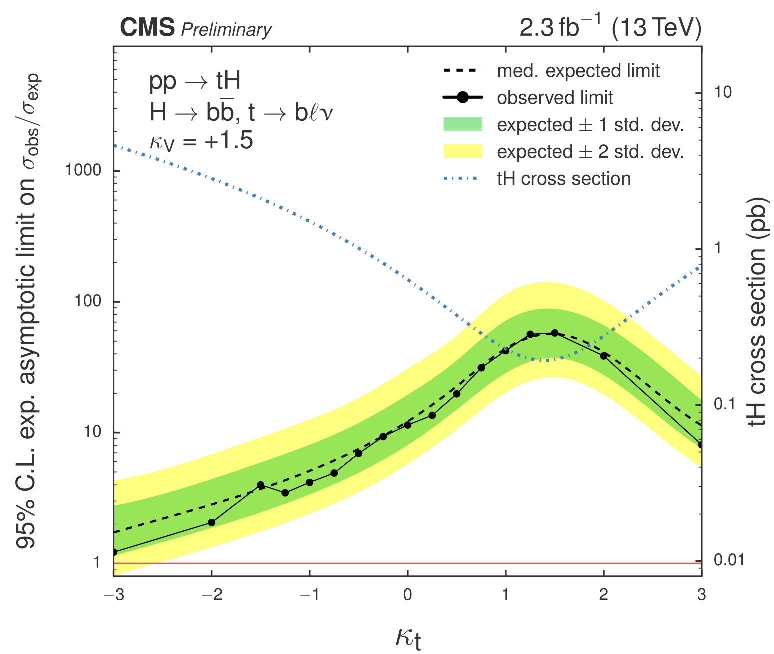


tH, $H \to b\bar{b}$

HIG-16-019







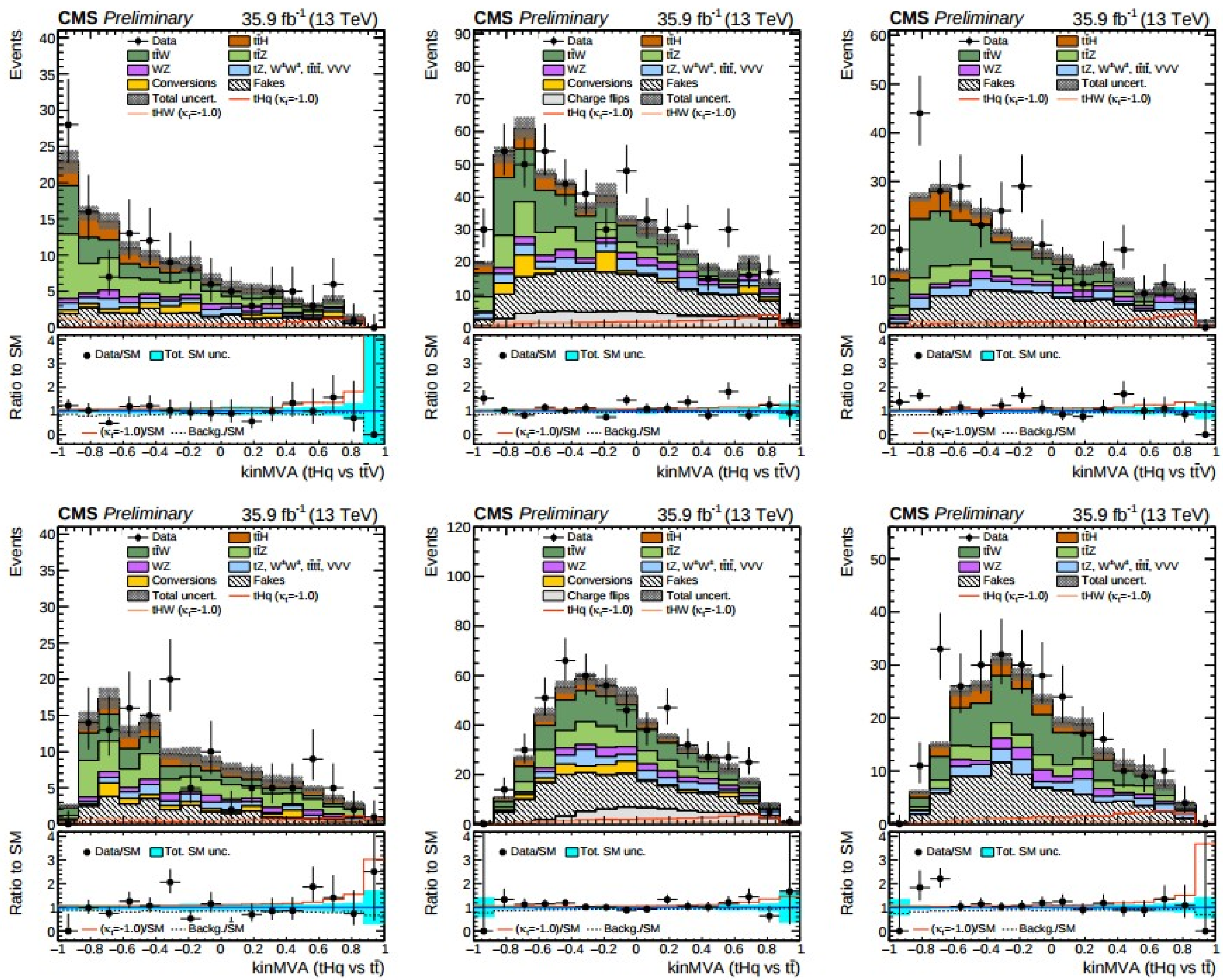
$tH,\ H \to b\bar{b}$

HIG-16-019

	Region	Observed Limit	Expected Limit		
			Median	$\pm 1\sigma$	$\pm 2\sigma$
SM scenario	3 tag	124.0	114.3	[73.6, 184.4]	[52.0 , 295.2]
	4 tag	195.8	174.6	[112.9, 287.4]	[78.8 , 464.4]
	Combination	**113.7**	**98.6**	**[64.0 , 159.2]**	**[45.3 , 254.8]**
ITC scenario	3 tag	7.4	7.4	[4.9 , 11.6]	[3.5 , 17.8]
	4 tag	9.2	10.0	[6.5 , 16.3]	[4.5 , 26.3]
	Combination	**6.0**	**6.4**	**[4.2 , 10.1]**	**[3.0 ,15.7]**

tH, H $\rightarrow$ multileptons (ZZ^*, $\tau\tau$, WW^*)

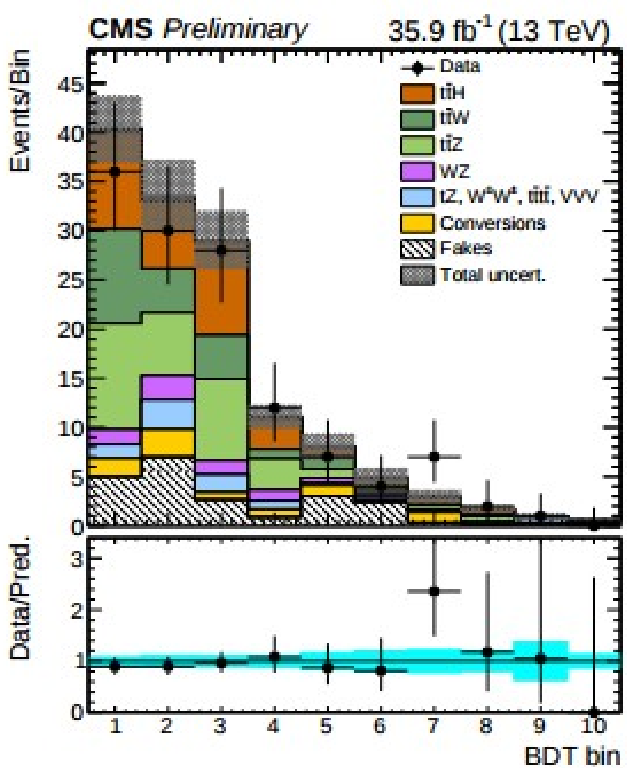
Pre-fit BDT classifier outputs, for the three-lepton channel (left), eμ (center), and μμ (right), for training against ttV (top row) and against tt (bottom row). In the box below each distribution, the ratio of the observed and predicted event yields is shown. The shape of the two tH signals for κt = –1.0 is shown, normalized to their respective cross sections for κt = –1.0, κV = 1.0.

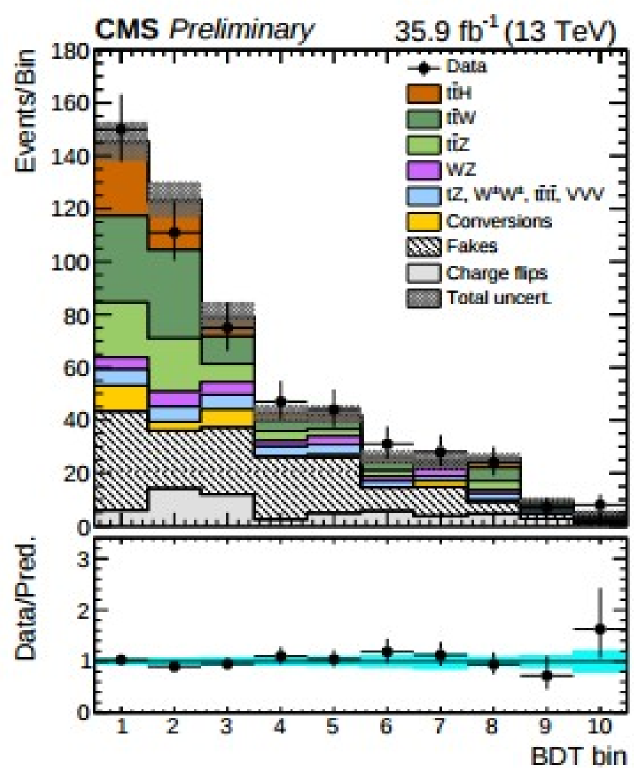


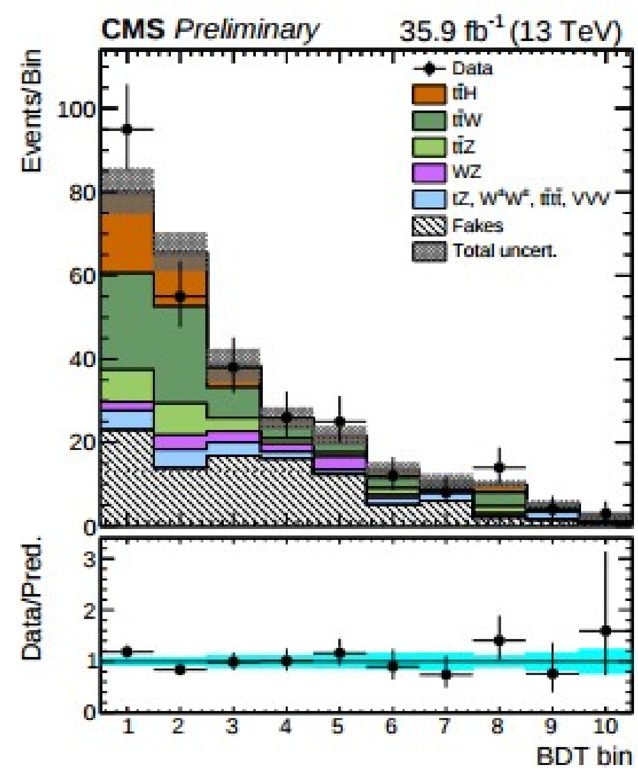
$tH,\ H \to$ multileptons $(ZZ^*,\ \tau\tau,\ WW^*)$

HIG-17-005

Post-fit categorized BDT classifier outputs as used in the maximum likelihood fit for the three-lepton channel (left), eμ (center), and μμ (right), for 35.9 /fb. In the box below each distribution, the ratio of the observed and predicted event yields is shown.







tH, H $\rightarrow$ multileptons (ZZ^*, $\tau\tau$, WW^*)

HIG-17-005

Expected and observed 95% C.L. upper limits on the tH + ttH production cross section times H → WW + ττ + ZZ branching ratio for a scenario of inverted couplings (κ_t/κ_V = −1.0, top rows) and for a standard-model-like signal (κ_t/κ_V = 1.0, bottom rows), in pb.

Scenario	Channel	Obs. Limit (pb)	Exp. Limit (pb) Median	±1σ	±2σ
$\kappa_t/\kappa_V = -1$	$\mu\mu$	1.00	0.58	[0.42, 0.83]	[0.31, 1.15]
	$e\mu$	0.84	0.54	[0.39, 0.76]	[0.29, 1.03]
	$\ell\ell\ell$	0.70	0.38	[0.26, 0.56]	[0.19, 0.79]
	Combined	**0.64**	**0.32**	[0.22, 0.46]	[0.16, 0.64]
$\kappa_t/\kappa_V = 1$ (SM-like)	$\mu\mu$	0.87	0.41	[0.29, 0.58]	[0.22, 0.82]
	$e\mu$	0.59	0.37	[0.26, 0.53]	[0.20, 0.73]
	$\ell\ell\ell$	0.54	0.31	[0.22, 0.43]	[0.16, 0.62]
	Combined	**0.56**	**0.24**	[0.17, 0.35]	[0.13, 0.49]