

$t\bar{t}$ and $t\bar{t}H$ measurements at the CMS experiment

Philip Keicher for the CMS Collaboration
Higgs Hunting 2023, September 12th, 2023

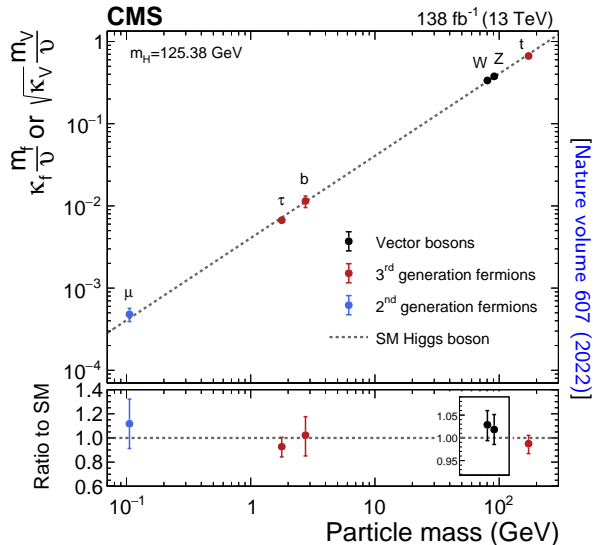


SPONSORED BY THE



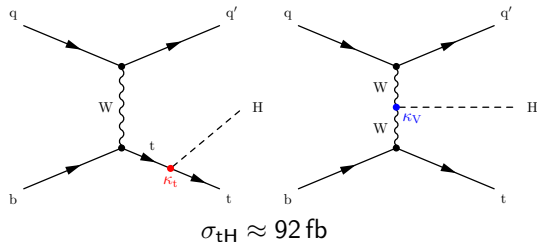
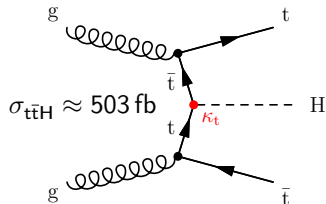
Why tH and $t\bar{t}H$?

- **Higgs Physics** crucial for **checks of the Standard Model** and **searches for physics beyond**
- This talk: precision measurements of coupling Higgs bosons to other particles

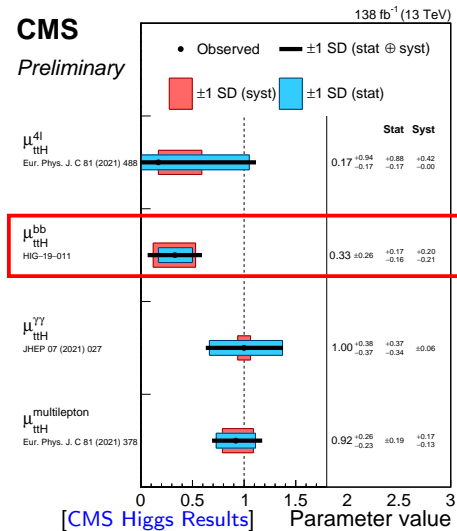


Why tH and $t\bar{t}H$?

- **Higgs Physics** crucial for **checks of the Standard Model** and **searches for physics beyond**
- This talk: precision measurements of coupling Higgs bosons to other particles
- **Direct access** to κ_t in $t\bar{t}H$ production
- **Additional sensitivity** to sign of κ_t and κ_V via tH production
- Also **sensitive to CP -odd** contributions in Top-Higgs Yukawa coupling
- **Challenge:** small cross sections of the processes w.r.t. background



Current status at CMS

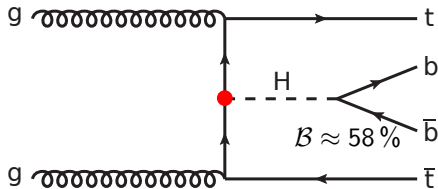


- Analyze tH and $t\bar{t}H$ production in **many final states** with full Run-II dataset
- Focus of this talk:
New $tH + t\bar{t}H, H \rightarrow b\bar{b}$ measurement
[CMS-PAS-HIG-19-011] [LHC Seminar, J. Steggemann]

$tH + t\bar{t}H, H \rightarrow b\bar{b}$ setup

[CMS-PAS-HIG-19-011]

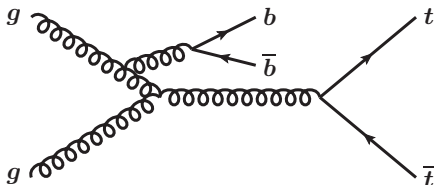
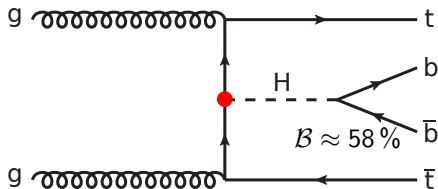
- Profit from **large branching fraction**
- **Fully reconstructable Higgs** final state
- Analysis **divided according to $t\bar{t}$ final states**
 - fully hadronic (FH): 0 leptons
 - semi-leptonic (SL): 1 lepton
 - di-leptonic (DL): 2 leptons



$t\bar{t}H + t\bar{t}H, H \rightarrow b\bar{b}$ setup

[CMS-PAS-HIG-19-011]

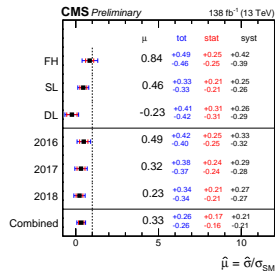
- Profit from **large branching fraction**
- **Fully reconstructable Higgs** final state
- Analysis **divided according to $t\bar{t}$ final states**
 - fully hadronic (FH): 0 leptons
 - semi-leptonic (SL): 1 lepton
 - di-leptonic (DL): 2 leptons
- **Main challenges:**
 - FH: **QCD multi-jet background**
→ **data-driven estimation**
 - DL+SL: **$t\bar{t}$ + jets background**
[Talk by A. Calandri]
→ **dedicated MC** for irreducible background
→ infer normalization **from data**



$t\bar{t}H + t\bar{t}H, H \rightarrow b\bar{b}$ statistical interpretations

[CMS-PAS-HIG-19-011]

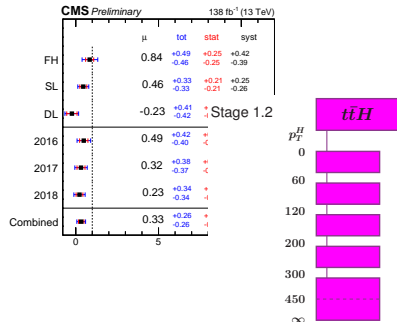
- Perform **five interpretations**
 - inclusive $t\bar{t}H(b\bar{b})$ signal strength $\mu_{t\bar{t}H}$



$t\bar{t}H, H \rightarrow b\bar{b}$ statistical interpretations

[CMS-PAS-HIG-19-011]

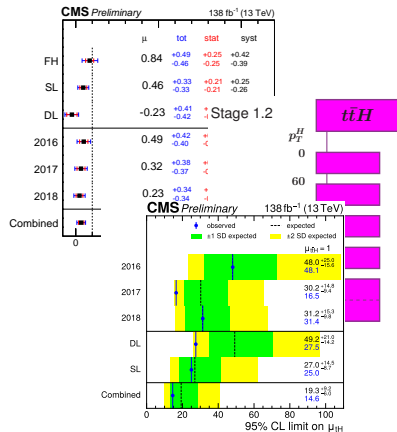
- Perform **five interpretations**
 - inclusive $t\bar{t}H(b\bar{b})$ signal strength $\mu_{t\bar{t}H}$
 - **New: Simplified Template Cross Section (STXS)**



$t\bar{t}H, H \rightarrow b\bar{b}$ statistical interpretations

[CMS-PAS-HIG-19-011]

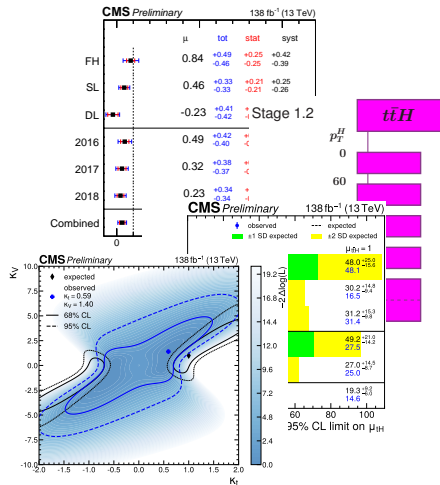
- Perform **five interpretations**
 - inclusive $t\bar{t}H(b\bar{b})$ signal strength $\mu_{t\bar{t}H}$
 - **New: Simplified Template Cross Section (STXS)**
 - Upper limit on $tH(b\bar{b})$ production



tH + tH-bar, H -> b b-bar statistical interpretations

[CMS-PAS-HIG-19-011]

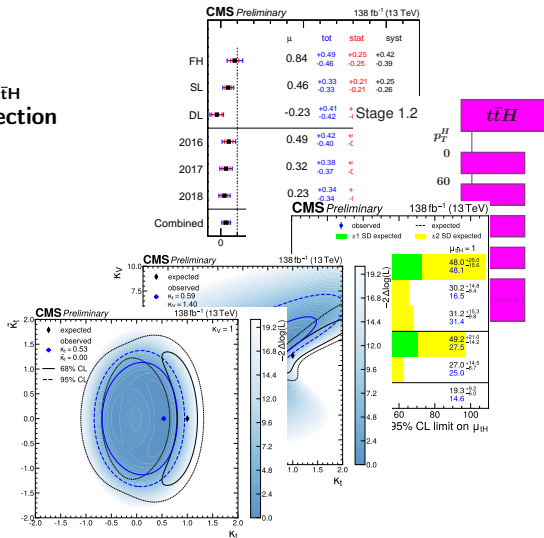
- Perform **five interpretations**
 - inclusive tH(tH-bar) signal strength $\mu_{t\bar{t}H}$
 - New: Simplified Template Cross Section (STXS)**
 - Upper limit on tH(b b-bar) production
 - Measurement of κ_t and κ_V



$t\bar{t}H + t\bar{t}H, H \rightarrow b\bar{b}$ statistical interpretations

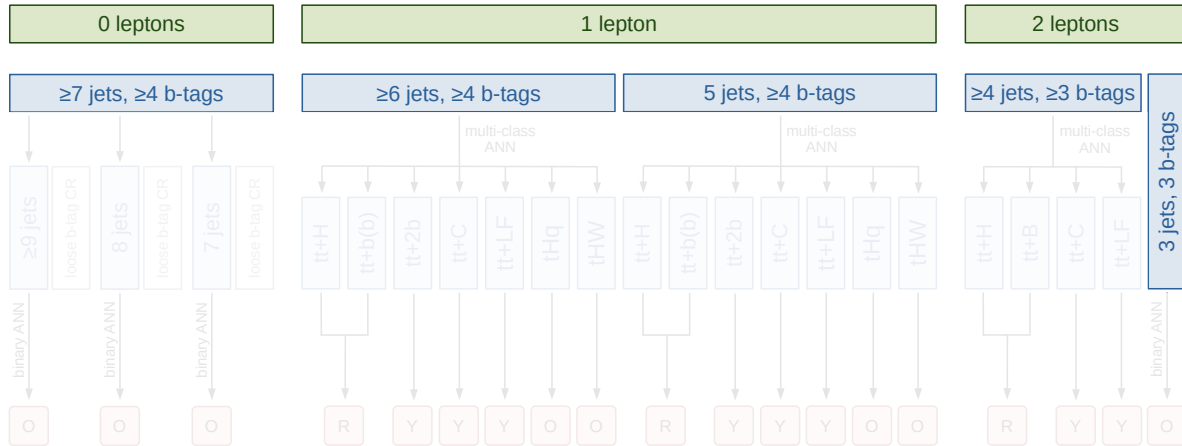
[CMS-PAS-HIG-19-011]

- Perform **five interpretations**
 - inclusive $t\bar{t}H(b\bar{b})$ signal strength $\mu_{t\bar{t}H}$
 - **New: Simplified Template Cross Section (STXS)**
 - Upper limit on $tH(b\bar{b})$ production
 - Measurement of κ_t and κ_V
 - CP measurement



Inclusive $\mu_{t\bar{t}H}$ Strategy

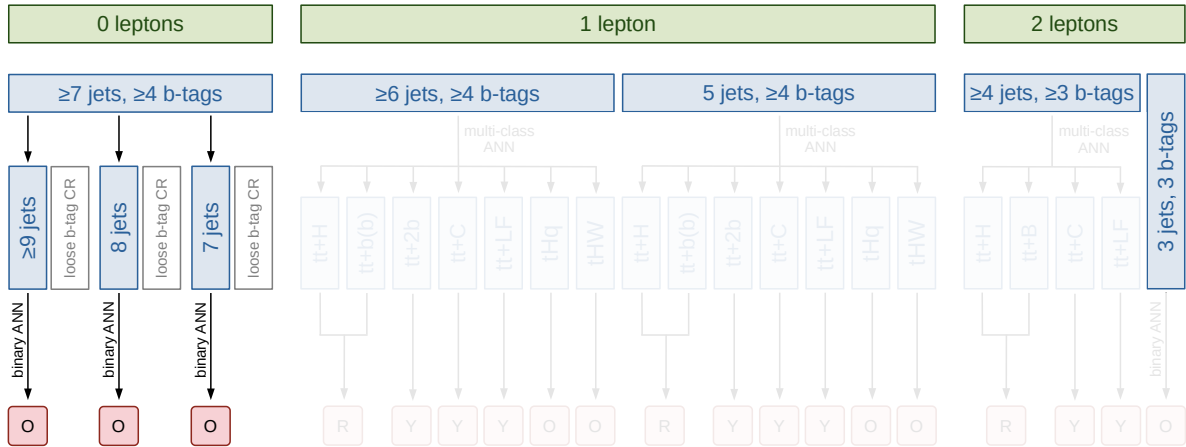
[CMS-PAS-HIG-19-011]



□ Distribution in template fit, event yield (Y), ANN output (O), likelihood ratio of ANN outputs (R)

Inclusive $\mu_{t\bar{t}H}$ Strategy

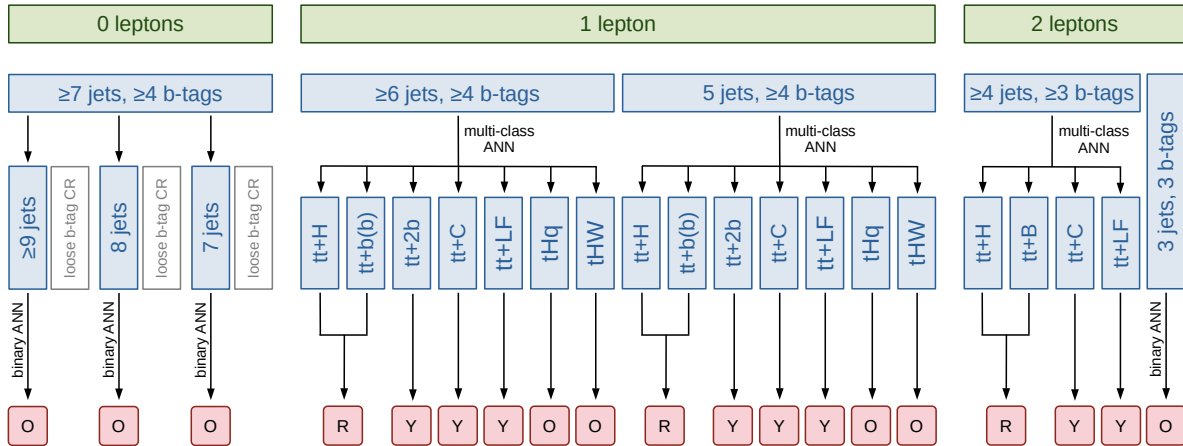
[CMS-PAS-HIG-19-011]



□ Distribution in template fit, event yield (Y), ANN output (O), likelihood ratio of ANN outputs (R)

Inclusive $\mu_{t\bar{t}H}$ Strategy

[CMS-PAS-HIG-19-011]



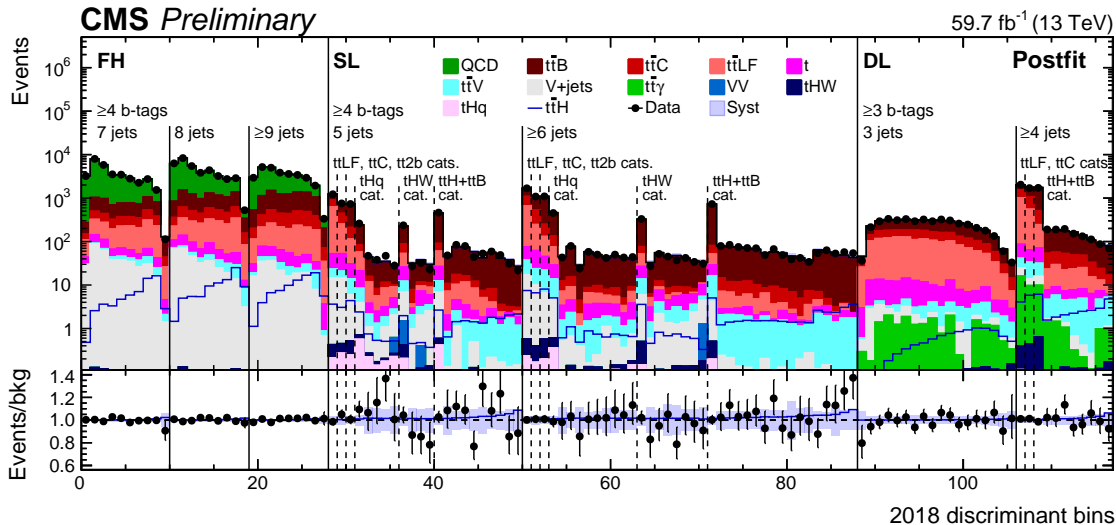
O Distribution in template fit, event yield (Y), ANN output (O), likelihood ratio of ANN outputs (R)

$$R_{SL} = \frac{o(t\bar{t}H)}{o(t\bar{t}H)+o(t\bar{t}+b(b))+o(t\bar{t}+2b)}$$

$$R_{DL} = \frac{o(t\bar{t}H)}{o(t\bar{t}H)+o(t\bar{t}+B)}$$

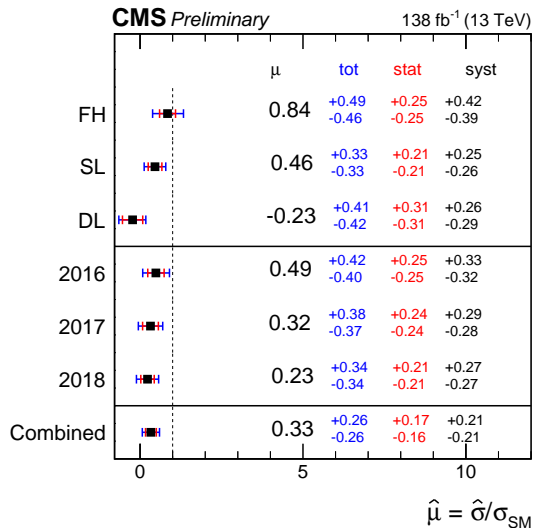
Final discriminant distribution

[CMS-PAS-HIG-19-011]



Inclusive $\mu_{t\bar{t}H}$ Results

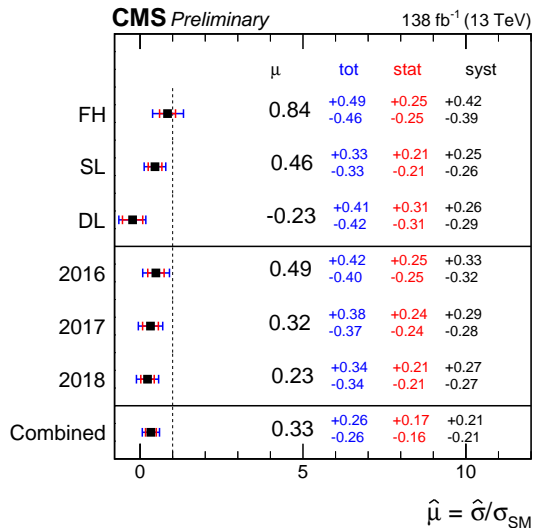
[CMS-PAS-HIG-19-011]



- Compatible with [ATLAS result](#): $\mu_{t\bar{t}H} = 0.35^{+0.36}_{-0.34}$

Inclusive $\mu_{t\bar{t}H}$ Results

[CMS-PAS-HIG-19-011]



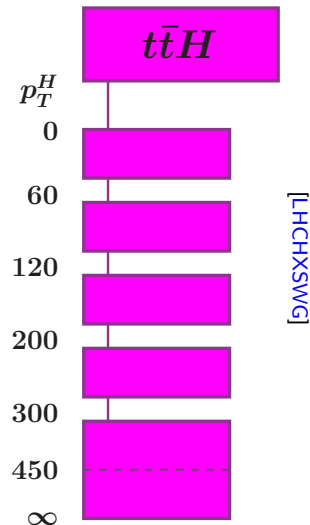
Uncertainty source	$\Delta\mu_{t\bar{t}H}$ (observed)	$\Delta\mu_{t\bar{t}H}$ (expected)
Total experimental	+0.10/ - 0.10	+0.11/ - 0.10
jet energy scale and resolution	+0.08/ - 0.07	+0.09/ - 0.09
b tagging	+0.07/ - 0.06	+0.06/ - 0.02
luminosity	+0.02/ - 0.02	+0.01/ - 0.01
Total theory	+0.16/ - 0.16	+0.18/ - 0.14
t \bar{t} + jets background	+0.15/ - 0.16	+0.12/ - 0.11
signal modelling	+0.06/ - 0.01	+0.13/ - 0.06
Size of the simulated event samples	+0.13/ - 0.12	+0.10/ - 0.10
Total systematic	+0.20/ - 0.21	+0.23/ - 0.19
Statistical	+0.17/ - 0.16	+0.17/ - 0.17
background normalisation	+0.13/ - 0.13	+0.13/ - 0.13
t $\bar{t}B$ and t $\bar{t}C$ normalisation	+0.12/ - 0.12	+0.12/ - 0.12
QCD normalisation	+0.01/ - 0.01	+0.01/ - 0.01
Total	+0.26/ - 0.26	+0.28/ - 0.25

STXS interpretation strategy

[CMS-PAS-HIG-19-011]

- Measure XS in 5 p_T^H bins
- Perform **reconstruction of bins with χ^2 (FH)**
or **neural networks (SL/DL)**

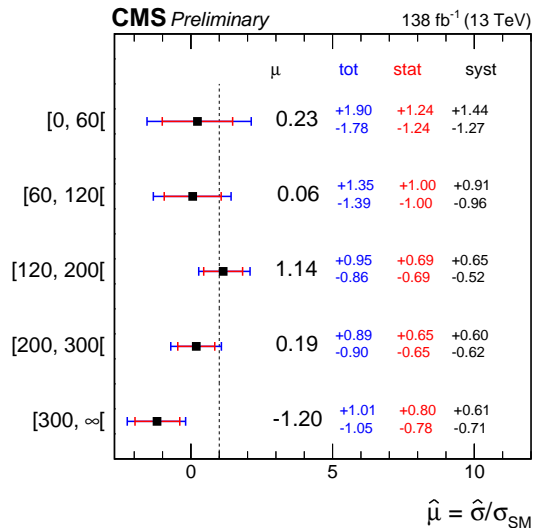
Stage 1.2



STXS interpretation strategy

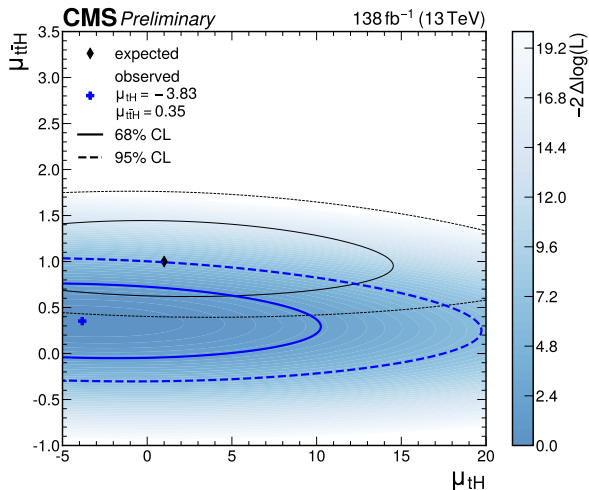
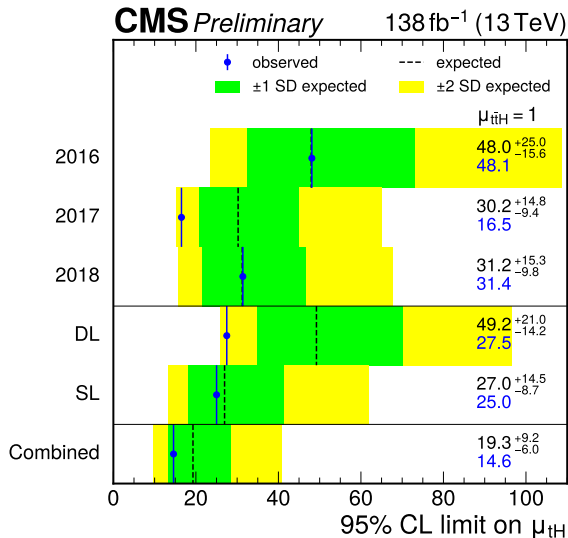
[CMS-PAS-HIG-19-011]

- Measure XS in 5 p_T^H bins
- Perform **reconstruction of bins with χ^2 (FH)** or **neural networks (SL/DL)**
- Obtain varying degrees of sensitivity
- Result **completely compatible** with inclusive $\mu_{t\bar{t}H}$ measurement



Upper limits for $t\bar{H}(b\bar{b})$ production

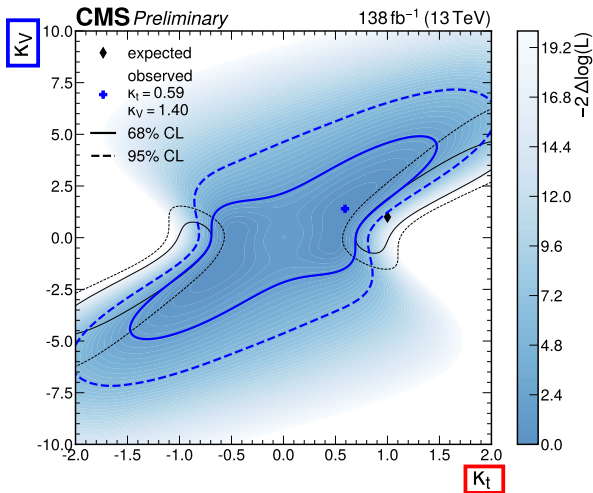
[CMS-PAS-HIG-19-011]



Coupling interpretation

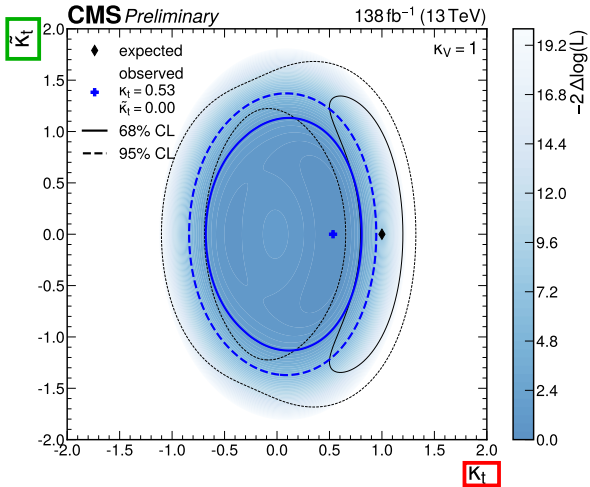
[CMS-PAS-HIG-19-011]

- Re-parameterize analysis in terms of κ_t and κ_V
 - $t\bar{t}H \propto \kappa_t^2$
 - $tH \propto \kappa_t, \kappa_V$ via interference
- Fully compatible with other interpretations



CP measurement

[CMS-PAS-HIG-19-011]



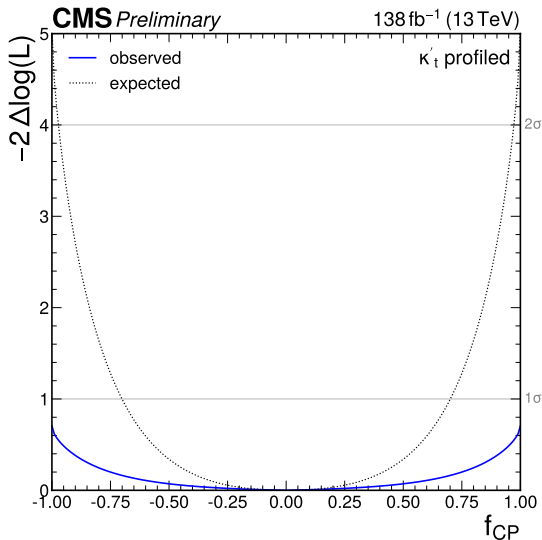
- Extend Top-Higgs Yukawa Lagrangian with CP-odd component

$$\mathcal{L} \propto \bar{\psi}_t (\kappa_t + i\gamma_5 \tilde{\kappa}_t) \psi_t H$$

→ both $t\bar{t}H$ and $tH \propto \tilde{\kappa}_t^2$

\mathcal{CP} measurement

[CMS-PAS-HIG-19-011]



- Extend Top-Higgs Yukawa Lagrangian with \mathcal{CP} -odd component

$$\mathcal{L} \propto \bar{\psi}_t (\kappa_t + i\gamma_5 \tilde{\kappa}_t) \psi_t H$$

→ both $t\bar{t}H$ and $tH \propto \tilde{\kappa}_t^2$

- Additional parameterization in κ_t - $\tilde{\kappa}_t$ plane

$$f_{\mathcal{CP}} = \frac{\tilde{\kappa}_t^2}{\kappa_t^2 + \tilde{\kappa}_t^2} \cdot \text{sign} \left(\frac{\tilde{\kappa}_t}{\kappa_t} \right)$$

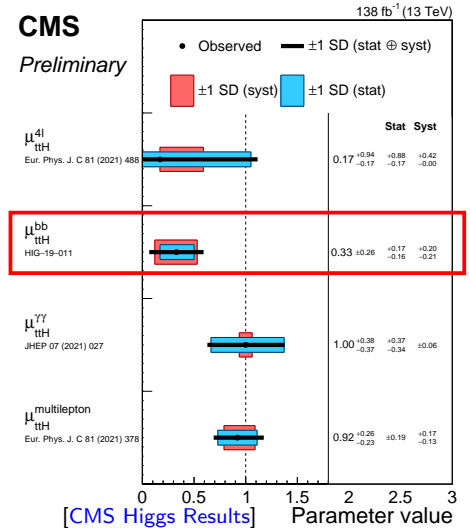
$$\kappa_t = \kappa_t' \sqrt{1 - |f_{\mathcal{CP}}|}$$

$$\tilde{\kappa}_t = \kappa_t' \sqrt{|f_{\mathcal{CP}}|}$$

→ Exclude \mathcal{CP} -odd component at almost 1σ

Conclusions

- $t\bar{t}H$ analyses especially interesting for **direct tests of Top-Higgs Yukawa coupling**
- Presented **new $t\bar{t}H, H \rightarrow b\bar{b}$ analysis** with full Run-II statistics [[CMS-PAS-HIG-19-011](#)]
 - Updated background modeling and analysis techniques
- $t\bar{t}H$ analyses **continue to be interesting in the future**



Backup

Kinematic selection criteria

[CMS-PAS-HIG-19-011]

	FH channel	SL channel	DL channel
Number of leptons	0	1	2
Sign and flavour of leptons	—	e^\pm, μ^\pm	$e^+e^-, \mu^\pm e^\mp, \mu^+\mu^-$
Min. p_T of leading electron (GeV)	—	29/30/30	25
Min. p_T of leading muon (GeV)	—	26/29/26	25
Min. p_T of additional leptons (GeV)	—	—	15
Max. p_T of additional leptons (GeV)	15	15	—
Max. $ \eta $ of leptons	2.4	2.4	2.4
Min. $m_{\ell\ell}$ (GeV)	—	—	20
$m_{ee/\mu\mu}$ (GeV)	—	—	< 76 or > 106
Min. number of jets	7	5	3
Min. p_T of jets (GeV)	30	30	30
Min. p_T of 6 th jet (GeV)	40	—	—
Max. $ \eta $ of jets	2.4	2.4	2.4
Min. number of b-tagged jets	2	4	3
m_{qq} (GeV)	> 30 and < 250	—	—
Min. H_T (GeV)	500	—	—
Min. p_T^{miss} (GeV)	—	20	40

FH data-driven QCD estimation

[CMS-PAS-HIG-19-011]

	$60(72) \text{ GeV} < m_{qq} < 100(90) \text{ GeV}$	$30 \text{ GeV} < m_{qq} < 60(72) \text{ GeV}$ or $100(90) \text{ GeV} < m_{qq} < 250 \text{ GeV}$
2 b tags	<i>training region (TR)</i>	<i>validation region (VR-TR)</i>
≥ 4 loose b tags	QCD events for ANN training	input variable validation
3 b tags	<i>evaluation region (ER)</i>	<i>validation region (VR-ER)</i>
≥ 4 loose b tags	discriminant shape for QCD	discriminant shape for QCD
≥ 4 b tags	<i>signal region (SR)</i>	<i>validation region (VR-SR)</i>
	analysis region	comparison of QCD shape with data

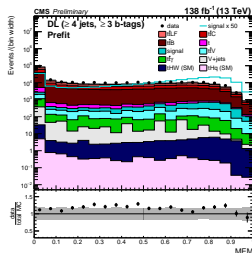
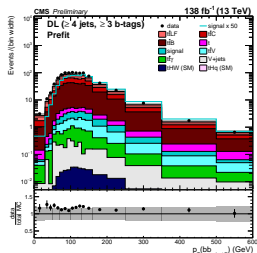
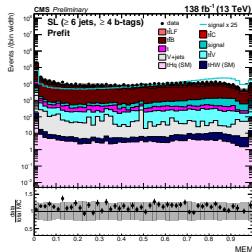
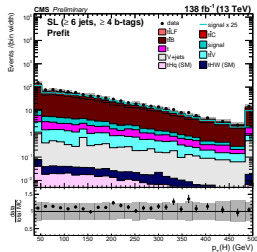
ANN input features

[CMS-PAS-HIG-19-011]

Observable	(≥ 2 jets, ≥ 2 b tags)	FH (3 jets, ≥ 4 b tags)	SL (7 jets, ≥ 4 b tags)	DL (≥ 4 jets, ≥ 2 b tags)
MEM	matrix element method discriminant	✓	✓	✓
BLR	b tagging likelihood ratio discriminant	✓	✓	✓
$\ln \left(\frac{1+MEM}{1-MEM} \right)$	transformed b tagging likelihood ratio discriminant		✓	✓
$p_2(\bar{t})$	p_2 of second leading jet, ranked in p_1		✓	✓
$p_3(\bar{t})$	p_3 of third leading jet, ranked in p_1			✓
$p_7(\bar{t})$	p_7 of seventh leading jet, ranked in p_1	✓		
$p_2(b^i)$	p_2 of b^i , $i=1-4$, leading b-tagged jet, ranked in p_1			✓
$q(\bar{t})$	q of b^i , $i=1-2$, leading jet, ranked in b tagging discriminant value	✓	✓	
$d_b(j)$	average b tagging discriminant value of all jets		✓	✓
$d_b(b)$	average b tagging discriminant value of all b-tagged jets		✓	✓
$d_b^3(j)$	third highest b tagging discriminant value of all jets		✓	✓
$\text{Var}(d_b(j))$	variance of b tagging discriminant values of all jets		✓	✓
$\langle \Delta R(bb) \rangle$	average of ΔR between two b-tagged jets			✓
$\langle \Delta R(jj) \rangle$	average of ΔR between two jets	✓	✓	
$\min \Delta R(jj)$	minimum of ΔR between two jets	✓	✓	✓
$\max \Delta R(jj)$	maximum of ΔR between two jets	✓	✓	
$\langle \Delta \eta(bb) \rangle$	average of $\Delta \eta$ between two b-tagged jets		✓	✓
$\langle \Delta \eta(jj) \rangle$	average of $\Delta \eta$ between two jets	✓	✓	✓
$\langle m(b) \rangle$	average invariant mass of all b-tagged jets		✓	✓
$\langle m(j) \rangle$	average invariant mass of all jets		✓	✓
$m(b\bar{b}_{\text{min}, \Delta R})$	invariant mass of pair of b-tagged jets closest in ΔR		✓	✓
$m(j\bar{b}_{\text{min}, \Delta R})$	invariant mass of pair of jet and b-tagged jet closest in ΔR			✓
$m(\bar{t}\bar{t}_{12 \text{ GeV}})$	invariant mass of pair of jets with mass closest to 125 GeV	✓		
$m(\bar{b}\bar{b}_{\text{max}, m})$	maximum invariant mass of pairs of b-tagged jets	✓	✓	✓
$m(\bar{b}\bar{b}_{\text{max}, p_1})$	invariant mass of jet and pair of b-tagged jets with highest p_1			✓
$\langle p_1(j) \rangle$	average p_1 of all jets		✓	✓
$\langle p_1(b) \rangle$	average p_1 of all b-tagged jets		✓	✓
$p_1(\bar{b}\bar{b}_{\text{min}, \Delta R})$	p_1 of pair of b-tagged jets closest in ΔR		✓	✓
$p_1(\bar{b}\bar{b}_{\text{max}, \Delta R})$	p_1 of pair of jets closest in ΔR		✓	✓
$p_1(\bar{b}\bar{t}_{\text{min}, \Delta R})$	p_1 of pair of jet and b-tagged jet closest in ΔR			✓
$H_1(j)$	scalar sum of p_1 of all jets		✓	✓
$H_1(b)$	scalar sum of p_1 of all b-tagged jets		✓	✓
$N(j)$	number of jets		✓	✓
$N(b^{\text{loose}})$	number of jets with loose b tag		✓	✓
$d_b(b_{\text{top}})^t$	b tagging discriminant value of b jet from t quark decay from tHq reconstruction		✓	✓
$ \eta(q^{\text{light}}) ^t$	$ \eta $ of light-quark jet from tHq reconstruction		✓	✓
$m(\bar{t}_{\text{top}})^t$	inv. mass of leptonically decaying t quark from tH reconstruction		✓	✓
BDT ^t	reconstruction BDT output for tHq, tH, tt hypotheses		✓	✓
A, S	event aplanarity and sphericity [76]	✓	✓	✓
$H^{(i)}$	p^i , $i=0-5$, Fox-Wolfman moment [77]	✓	✓	✓
$H^{(i)}/H_1^{(i)}$	ratio of Fox-Wolfman moments, $i=1-4$	✓	✓	✓

Observable construction

[CMS-PAS-HIG-19-011]



- Generally observe **good data/MC agreement**
 - But: **no strong separation power**
- ⇒ **Utilize Artificial Neural Networks**

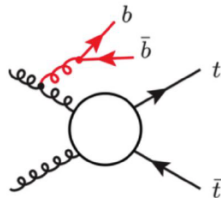
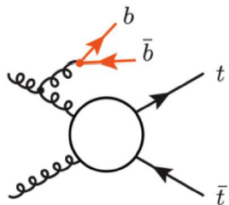
$t\bar{t}$ + jets modeling

Largest background in the leptonic channels. Divided into three sources depending on flavour of additional particle-level jets with $p_T > 20$ GeV and $|\eta| < 2.4$

- ttB : ≥ 1 additional b jet \Rightarrow largest contribution
- ttC : ≥ 1 additional c jet but no b jet
- ttLF : all other events

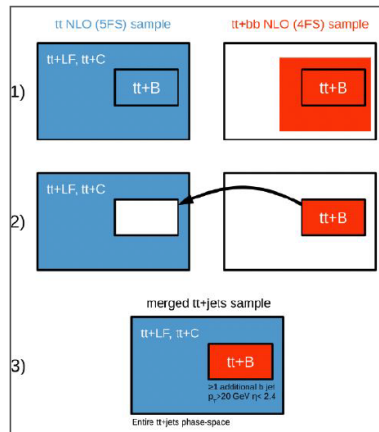
ttB is difficult to model and to measure, XS underpredicted by 20-30%

- New in this analysis: state-of-art MC simulation for ttB at NLO (4FS)
 - additional b jets from **matrix element** instead of **parton shower**
 - acceptance better described
 - better coverage by scale uncertainties



$t\bar{t}$ + jets merging

- $t\bar{t}B$ prediction from $t\bar{t}bb$ NLO simulation (4FS)
- $t\bar{t}C$ and $t\bar{t}LF$ taken from the POWHEG $t\bar{t}$ NLO sample (5FS)
 - $t\bar{t}B$ normalised to the expected yield from 5FS sample
 - $t\bar{t}B$ and $t\bar{t}C$ normalisation freely floating in the fit
 - measured in control regions built with ANN multiclassifiers in leptonic channels



Uncertainty model

Theory

- QCDscale
- pdfscale
- ME+PS matching
- Underlying event
- ISR/FSR
- μ_R/μ_F
- PDF
- 100% additional gluon splitting uncertainty

Experiment

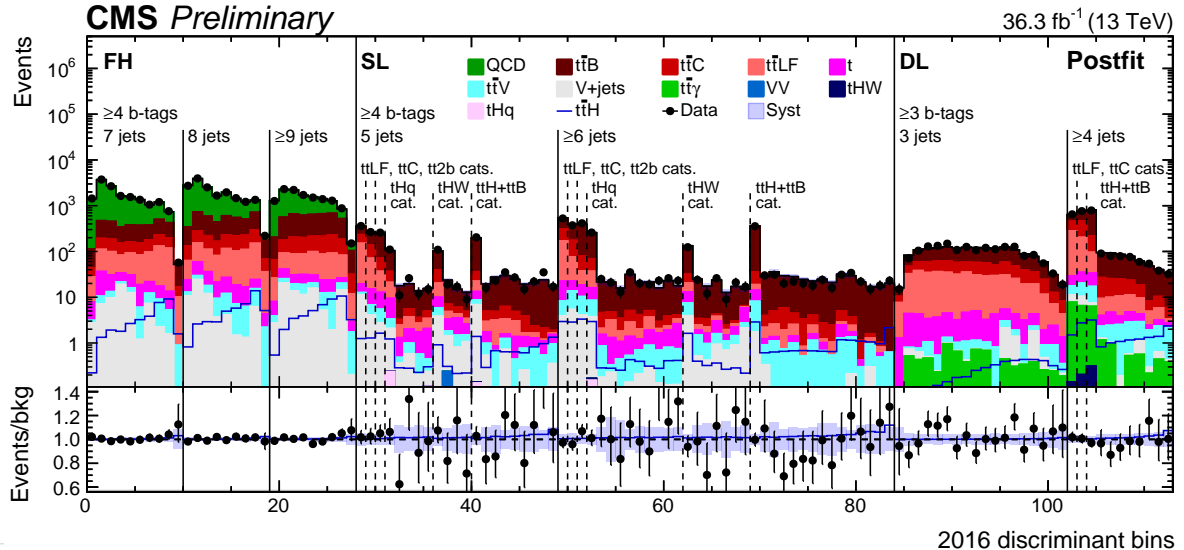
- btag (lf, hf, cf)
- lepton trigger eff.
- lepton ID eff.
- Jet energy scale + resolution
- TF Loose
- L1 Prefire
- Luminosity

Misc

- Size of MC samples
- freely-floating norms for $t\bar{t} + b\bar{b}$, $t\bar{t} + c\bar{c}$ and data-driven QCD estimation

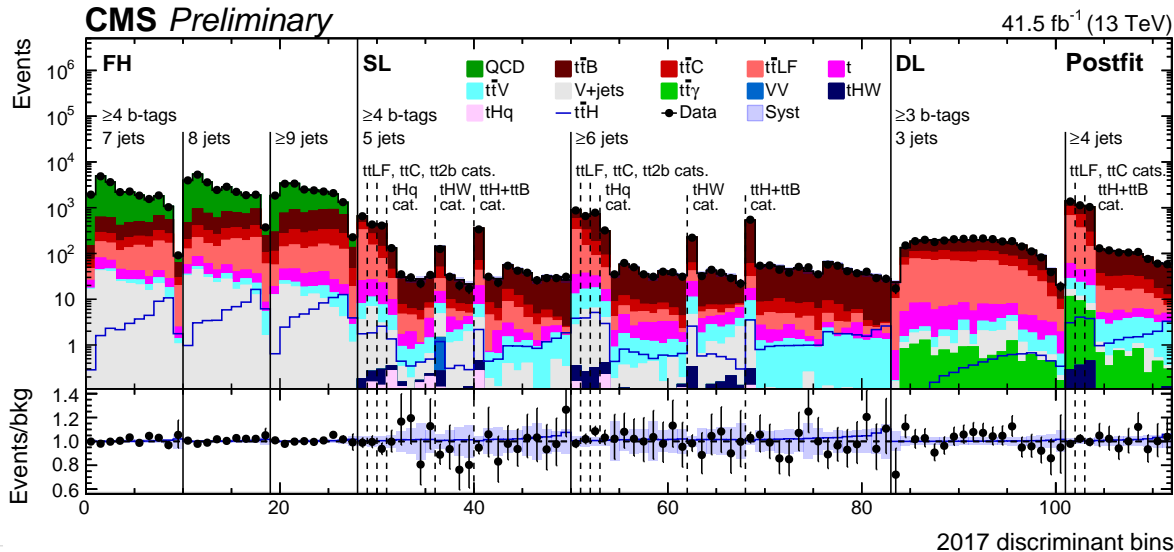
Final discriminant distribution - 2016

[CMS-PAS-HIG-19-011]



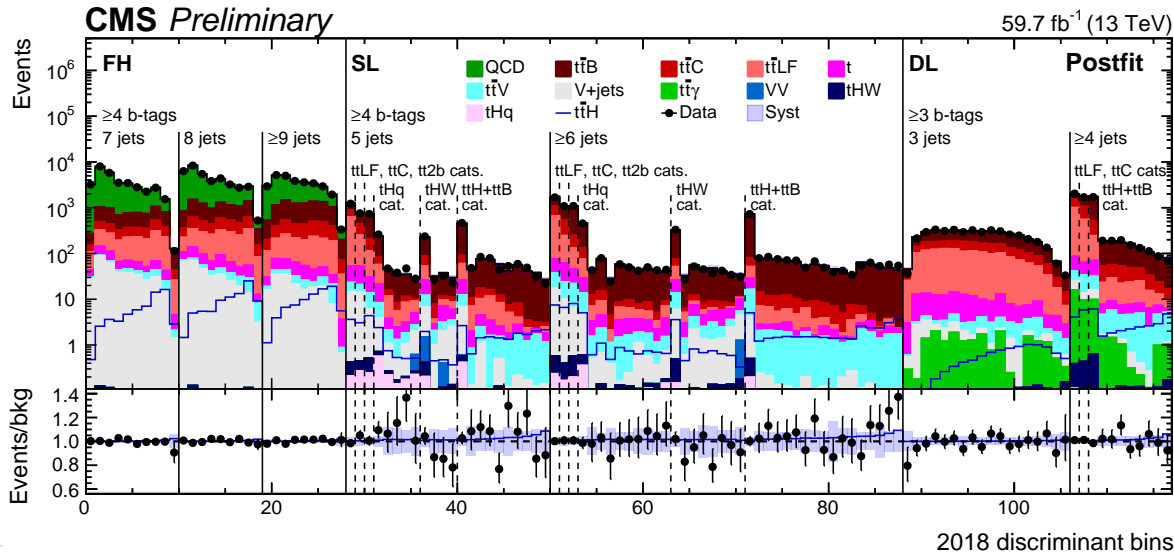
Final discriminant distribution - 2017

[CMS-PAS-HIG-19-011]



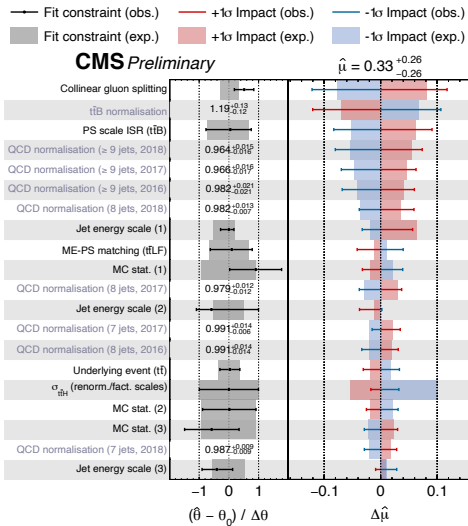
Final discriminant distribution - 2018

[CMS-PAS-HIG-19-011]



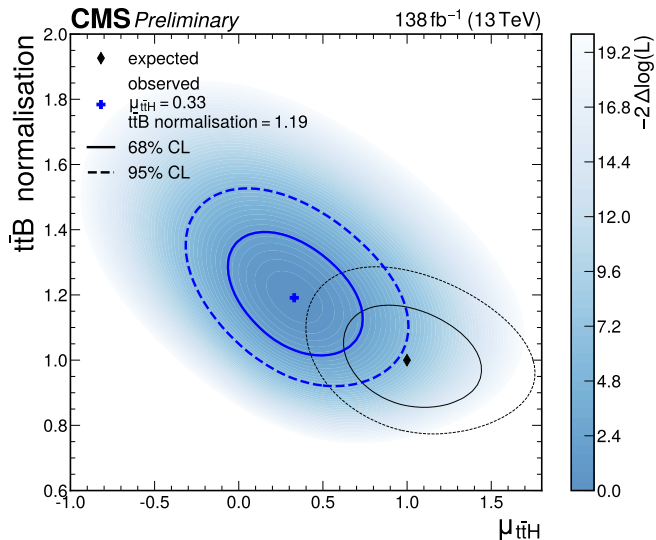
$t\bar{t}H(bb)$ measurement: Impacts

[CMS-PAS-HIG-19-011]



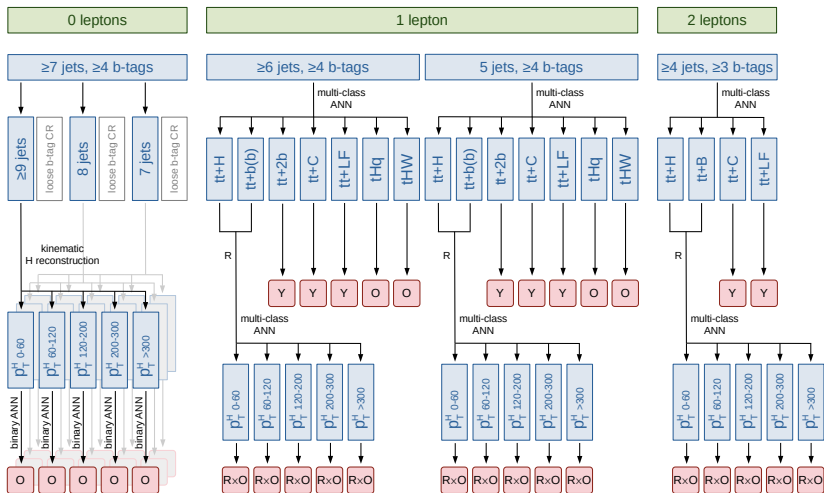
$t\bar{t}H(b\bar{b})$ measurement: Comparison to $t\bar{t} + B$ norm

[CMS-PAS-HIG-19-011]



STXS strategy

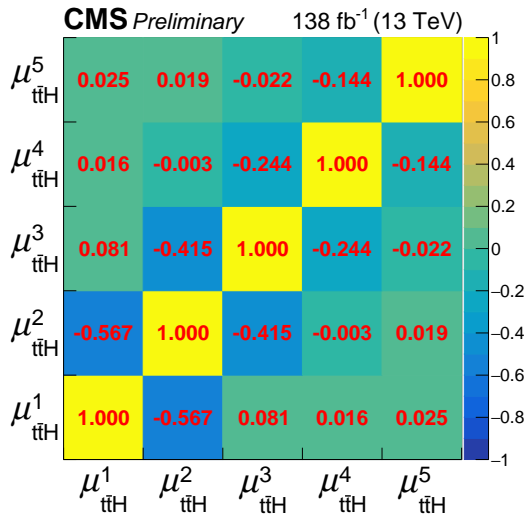
[CMS-PAS-HIG-19-011]



Legend: ■ Distribution in template fit, event yield (Y), ANN output (O), likelihood ratio of ANN outputs (R)

STXS POI correlations

[CMS-PAS-HIG-19-011]



Statistical model for coupling interpretation

- Introduce additional signal templates for tH and t \bar{t} H for inter-/extrapolation
- t \bar{t} H template calculated with:

$$c_{t\bar{t}H}(\kappa_t, \kappa_V, \tilde{\kappa}_t) = A \cdot \kappa_t^2 + B \cdot \tilde{\kappa}_t^2$$

with pure \mathcal{CP} -even template A (SM) and pure \mathcal{CP} -odd template B

- tH template calculated with:

$$c_{tH}(\kappa_t, \kappa_V, \tilde{\kappa}_t) = A \cdot \kappa_t(\kappa_t - \kappa_V) + B \cdot \tilde{\kappa}_t^2 + C \cdot \kappa_V(\kappa_V - \kappa_t) + (D + A + C) \cdot \kappa_t \kappa_V$$

with templates A ($\kappa_t = 1, \tilde{\kappa}_t = 0, \kappa_V = 0$), B ($\kappa_t = 0, \tilde{\kappa}_t = 1, \kappa_V = 0$), C ($\kappa_t = 0, \tilde{\kappa}_t = 0, \kappa_V = 1$) and term $D + A + C$ (SM, $\kappa_t = 1, \tilde{\kappa}_t = 0, \kappa_V = 1$)

- Everything else same as in tH and t \bar{t} H interpretations

Compatibility to previous results

The p-values are calculated by evaluating the postfit likelihood of HIG-19-011 at the nominal signal strength measured in the previous analysis. **Caveat: these are approx. numbers (!)**, assuming uncorrelated datasets among analyses and neglecting the uncertainty on the signal strength of the analysis we compare to. The compatibility to the SM (last row) is not subject to such approximations.

Reference:		Tested against:		Compatibility	
Analysis	Combination	Analysis	Combination	p-value	in S.D.
HIG-19-011	SL+DL 2016	HIG-17-026	SL+DL 2016	0.41	0.8 σ
HIG-19-011	FH+SL+DL 2016+2017	PAS-HIG-18-030	FH+SL+DL 2016+2017	0.06	1.9 σ
HIG-19-011	FH+SL+DL 2017	PAS-HIG-18-030	FH+SL+DL 2017	0.02	2.4 σ
HIG-19-011	FH+SL+DL 2016	PAS-HIG-18-030	FH+SL+DL 2016	0.53	0.6 σ
HIG-19-011	FH+SL+DL 2016+2017+2018	Standard Model hypothesis		0.02	2.4 σ