中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences
Beijing

Higgs Hunting - 29 to 31 Jul 2019, Paris, France

$t\bar{t}H$ production studies in final states with multileptons at 13 TeV with CMS

Binghuan Li

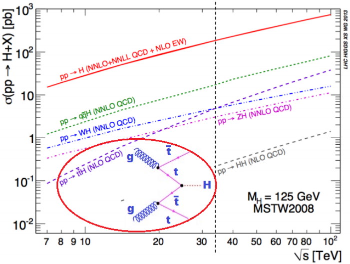
On behalf of the CMS Collaboration

Motivations for studying the $t\bar{t}H$ process

- After the new boson discovery (Higgs-like), focus on measuring its properties $\rightarrow$ couplings to fermions, gauge bosons, and itself
- Top quark is the most strongly-coupled SM particle ($Y_t \approx 1$)
- Directly probe top-Higgs coupling through $t\bar{t}H$ production mechanism
- Large deviation from SM Higgs coupling could be explained by some beyond SM models with enhanced $t\bar{t}H$ production without changing the Higgs BR (Vector-like, heavy top partner, Compositeness, RS, little Higgs)

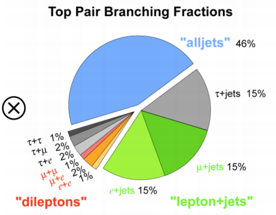
The $t\bar{t}H$ measurement

Production



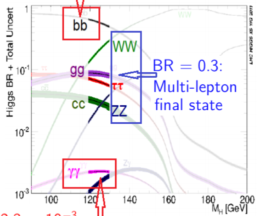
$\otimes$

Decay



$\otimes$

BR $= 0.58$: dominant mode but large bkg

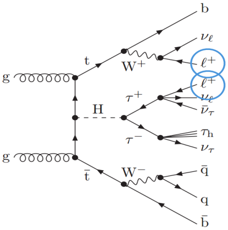


BR $= 0.3$: Multi-lepton final state

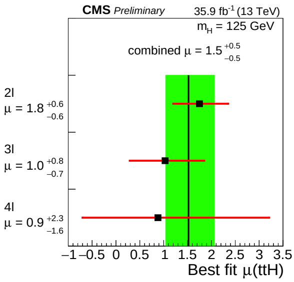
BR $= 2.3 \times 10^{-3}$: tiny but with clean resonant signature

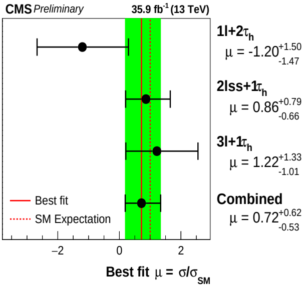
Challenges

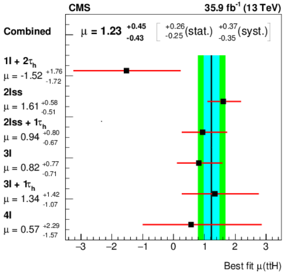
- Low signal cross-section
- Complicated final states with high-jet/b-tag multiplicity
- This talk focuses on three decay channels $H \to WW$, $H \to ZZ$, $H \to \tau\tau$



CMS results with 2016 dataset







channels without τ_h	channels with τ_h	combination lepton+τ_h
[PAS-HIG-17-004]	[PAS-HIG-17-003]	[HIG-17-018]

Combination lepton + τ_h channels

$$\mu_{t\bar{t}H} = 1.23^{+0.45}_{-0.43}[^{+0.36}_{-0.34}(stat)^{+0.29}_{-0.27}(syst)]$$

$\mu_{t\bar{t}H}$, reaches 3.2 σ obs. (2.8 σ exp.)

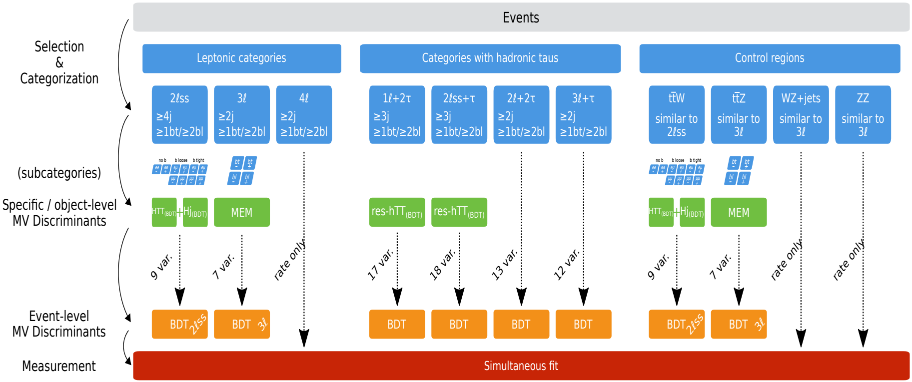
$t\bar{t}H, H \rightarrow$MultiLeptons [PAS-HIG-18-019]

Seven categories according to the lepton and τ_h multiplicity

- 0 τ_h
  - 2 l ss
  - 3 leptons
  - 4 leptons
- $\geq 1\tau_h$
  - 1 lepton + 2 τ_h
  - 2 l ss + 1 τ_h
  - 2 lepton + 2 τ_h
  - 3 leptons + 1 τ_h

- lepton identification through a dedicated BDT
- charge requirements suppress $t\bar{t}$
- remaining backgrounds
  - irreducible ttW, ttZ, WZ, ZZ
  - misidentified jets, misidentified leptons

Analysis workflow



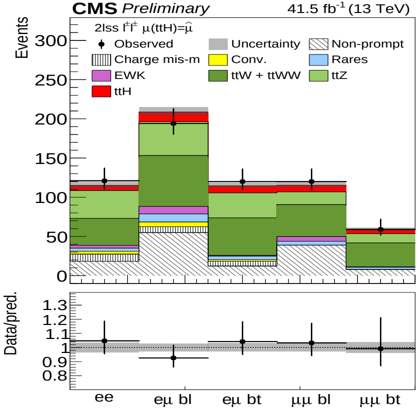
The signal rate μ is computed for each of the categories individually and for thier combination

Use rate or BDT depending on categories

Background Characterization

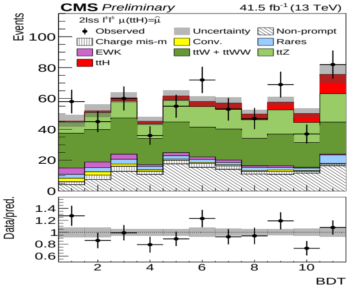
2lss sub-categories

- $t\bar{t}W/Z$, WZ, ZZ
  irreducible, taken from simulation
  constraint by dedicated control regions
- Flips
  Electron charge mismeasured
  Mis-ID rate reduced by a factor of 2 comapred to 2016 (new pixel detector)
- Fakes
  misidentified non prompt lepton
  jet misidentified as τ_h

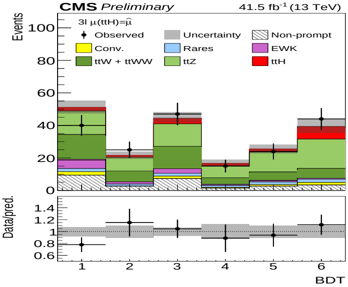


Signal extraction in categories without τ_h

2lSS



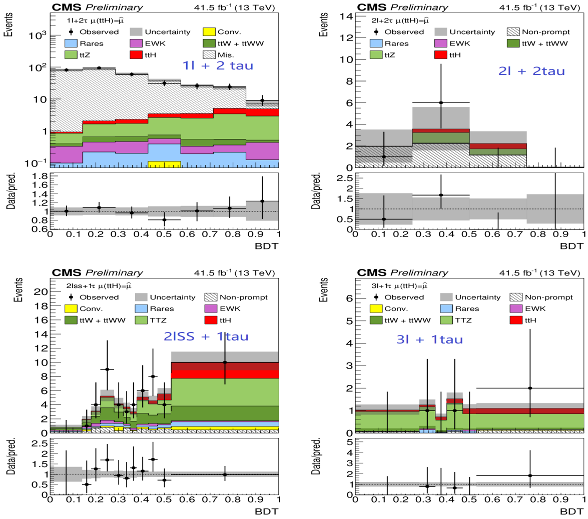
3 leptons



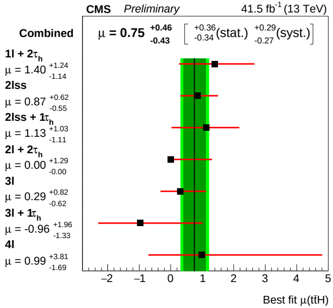
- The 2l and 3l signal are extracted from a combined multivariate discriminator obtained by a training against $t\bar{t}$ and $t\bar{t}W, Z$
- Mapped into 1D distribution based on S/B likelihood ratio
- The 4l, due to low statistics, is estimated using rate only

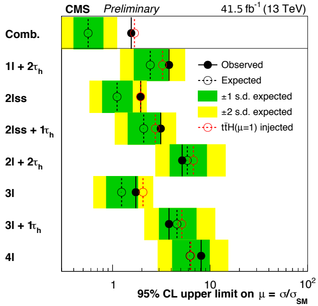
Signal extraction in categories with τ_h

- 1 BDT per category
- sig: $t\bar{t}H$
  bkg: $t\bar{t}$ and $t\bar{t}W, Z$
- NEW
  - trained using xgBoost algorithm
  - hadronic top tagger as input of BDT



Results of $t\bar{t}H$, $H \rightarrow$MultiLeptons





Category	Signal Strength $\pm 1\sigma$ Measured	Signal Strength $\pm 1\sigma$ Expected
$1\ell + 2\tau_h$	$1.40^{+1.24}_{-1.14}$	$1.00^{+1.14}_{-0.93}$
$2\ell ss$	$0.87^{+0.62}_{-0.55}$	$1.00^{+0.53}_{-0.49}$
$2\ell ss + 1\tau_h$	$1.13^{+1.03}_{-1.11}$	$1.00^{+0.93}_{-0.80}$
$2\ell + 2\tau_h$	$0.00^{+1.29*}_{-0.00}$	$1.00^{+2.63*}_{-1.56}$
3ℓ	$0.29^{+0.82}_{-0.62}$	$1.00^{+0.59}_{-0.52}$
$3\ell + 1\tau_h$	$-0.96^{+1.96}_{-1.33}$	$1.00^{+1.91}_{-1.37}$
4ℓ	$0.99^{+3.31}_{-1.69}$	$1.00^{+2.41}_{-1.72}$
Combined	$0.75^{+0.46}_{-0.43}$	$1.00^{+0.39}_{-0.35}$
Combined with 2016 analysis	$0.96^{+0.34}_{-0.31}$	$1.00^{+0.30}_{-0.27}$

	Observed limit	Expected limit ($\mu = 0$)	Expected limit ($\mu = 1$)
$1\ell + 2\tau_h$	3.8	$2.4^{+1.3}_{-0.8}$	3.3
$2\ell ss$	2.0	$1.1^{+0.5}_{-0.3}$	1.9
$2\ell ss + 1\tau_h$	3.1	$2.1^{+1.0}_{-0.7}$	2.8
$2\ell + 2\tau_h$	5.2	$5.8^{+3.4}_{-2.0}$	6.8
3ℓ	1.7	$1.2^{+0.6}_{-0.4}$	2.1
$3\ell + 1\tau_h$	3.8	$4.6^{+2.7}_{-1.6}$	5.1
4ℓ	8.1	$6.2^{+3.6}_{-2.1}$	6.4
Combined	1.6	$0.8^{+0.3}_{-0.2}$	1.7
Combined with 2016 analysis	1.6	$0.6^{+0.2}_{-0.2}$	1.5

Summary

- CMS has performed measurement of a Higgs boson in associated with a top quark pair in multilepton final states with 13 TeV data recorded in 2017 (41.5 fb^{-1})
- Machine learning and matrix element methods are used to optmize the analysis sensitivity
  $\mu_{t\bar{t}H}$: Best-fit value of $0.75^{+0.46}_{-0.43}$, with observed (expected) significance of $1.7\sigma(2.9\sigma)$
- These results are combined with those previously obtained from data recored in 2016 (35.9 fb^{-1})
  $\mu_{t\bar{t}H}$: Best-fit value of $0.96^{+0.34}_{-0.31}$, with observed (expected) significance of $3.2\sigma(4.0\sigma)$
- All results are consistent with the SM expectation
- Result of full run II dataset (137.2 fb^{-1}) to come

Input variables of BDTs

Category	$1\ell + 2\tau_h$	$2\ell ss + 1\tau_h$	$2\ell + 2\tau_h$	$3\ell + 1\tau_h$	$2\ell ss$		3ℓ	
					$t\bar{t}$	$t\bar{t}V$	$t\bar{t}$	$t\bar{t}V$
Leading ℓ cone p_T	X		X	X		X		X
Trailing ℓ cone p_T		X		X		X		X
Minimum of ΔR(leading ℓ, j)	X	X	X	X	X	X	X	X
Minimum of ΔR(trailing ℓ, j)		X			X	X	X	X
ΔR(leading ℓ, trailing ℓ		X		X				
Transverse Mass of leading ℓ	X	X			X	X	X	X
Transverse Mass of trailing ℓ		X						
Maximum $\lvert\eta\rvert$ of ℓ collection		X		X	X	X	X	X
Signal leading $\ell\times$ signal trailing ℓ			X					
Average of ΔR(jj)	X	X	X					
Number of jets ($p_T > 25$ GeV)		X		X	X	X	X	X
Number of loose b-jets	X		X					
Mass of leading medium b-jet pair		X						
Mass of leading loose b-jet pair				X				
E_T^{miss}	X	X		X				
res-hTT	X	X						
Hadronic t p_T	X	X						
$\mathcal{D}_{\rm thad}^{\rm max}$					X			
$\mathcal{D}_{\rm Hj}^{\rm max}$						X		
Leading τ_h p_T	X	X	X	X				
Trailing τ_h p_T	X		X					
Mass of leading τ_h + trailing τ_h	X		X					
ΔR(leading τ_h, trailing τ_h)	X		X					
$cos(\theta)^*$(leading τ_h, trailing τ_h)	X		X					
Minimum of ΔR(leading τ_h, j)	X	X		X				
Minimum of ΔR(trailing τ_h, j)	X							
Minimum of $\Delta R(\tau_h, j)$			X					
Mass of leading ℓ+ leading τ_h				X				
Mass of trailing ℓ+ leading τ_h		X		X				
ΔR(leading ℓ, leading τ_h)	X	X						
ΔR(trailing ℓ, leading τ_h)		X						
$\Delta R(\ell, \tau_h)$ for same-sign pair of (ℓ, τ_h)	X							
Average of $\Delta R(\ell, \tau_h)$			X					
MEM							X	X
Number of variables	17	18	13	12	6	8	6	8

yield

Category	$2\ell ss$	3ℓ	4ℓ	$1\ell + 2\tau_h$	$2\ell + 2\tau_h$	$3\ell + 1\tau_h$	$2\ell ss + 1\tau_h$
$t\bar{t}H$	43.0 ± 7.1	18.8 ± 4.8	0.7 ± 0.3	6.6 ± 3.6	0.9 ± 0.5	1.0 ± 0.4	5.1 ± 2.1
$t\bar{t}W + t\bar{t}WW$	218.5 ± 13.7	51.0 ± 5.3	0.13 ± 0.03	1.1 ± 0.3	< 0.05	0.5 ± 0.1	13.1 ± 2.4
tH	2.4 ± 0.1	0.9 ± 0.1	< 0.05	0.3 ± 0.1	< 0.05	< 0.05	0.5 ± 0.0
$WZ + ZZ$	< 0.05	12.0 ± 1.7	0.15 ± 0.10	1.5 ± 0.8	< 0.05	0.1 ± 0.0	2.8 ± 2.0
$t\bar{t}Z/\gamma^*$	138.2 ± 7.6	74.1 ± 6.3	3.9 ± 0.6	11.6 ± 2.6	1.6 ± 0.5	4.5 ± 0.7	15.4 ± 2.4
Misidentified	132.1 ± 10.0	26.8 ± 4.0	< 0.05	299.6 ± 19.1	5.3 ± 2.2	0.3 ± 0.3	5.3 ± 2.2
Conversions	11.6 ± 3.0	6.6 ± 1.3	< 0.05	0.3 ± 0.1	< 0.05	< 0.05	< 0.05
Signal flip	22.8 ± 2.3	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05
Other	26.7 ± 3.9	9.7 ± 2.2	< 0.05	1.2 ± 0.5	0.06 ± 0.04	0.3 ± 0.2	3.2 ± 1.1
SM expectation	595.3 ± 20.6	200.0 ± 10.8	5.0 ± 0.7	322.0 ± 19.6	7.9 ± 2.3	6.7 ± 0.9	45.3 ± 5.1
Observed data	614	195	6	324	7	4	53

Systematic uncertainty correlation

Source	Uncertainty [%]	$\Delta\mu/\mu$ [%] (2017)	$\Delta\mu/\mu$ [%] (Comb.)	Correlations
Theoretical sources	≈ 8	8	9	Correlated
e, μ selection efficiency	3–5	4	3	Correlated
τ_h selection efficiency	5	3	5	Correlated
τ_h energy calibration	1.2	1	2	Correlated
b tagging efficiency	2–15 [**?**]	10	5	Correlated
Jet energy calibration	2–15 [**?**]	3	3	Correlated
Fake background yield	$\approx$ 30–50	17	9	Un-correlated

Hjet Tagger in $t\bar{t}H, H \rightarrow$MultiLeptons w/o τ_h

- Hj Tagger : BDT
- Objective : Identifiying a Jet decayed from W from Higgs
- Sig: correctly matched Jet in $t\bar{t}H$
  bkg: all jets in $t\bar{t}V$
- Variables :
  jet ID (b-tagging, qg likelihood)
  geometric (dR to l) properties
- take the highest BDT score jet