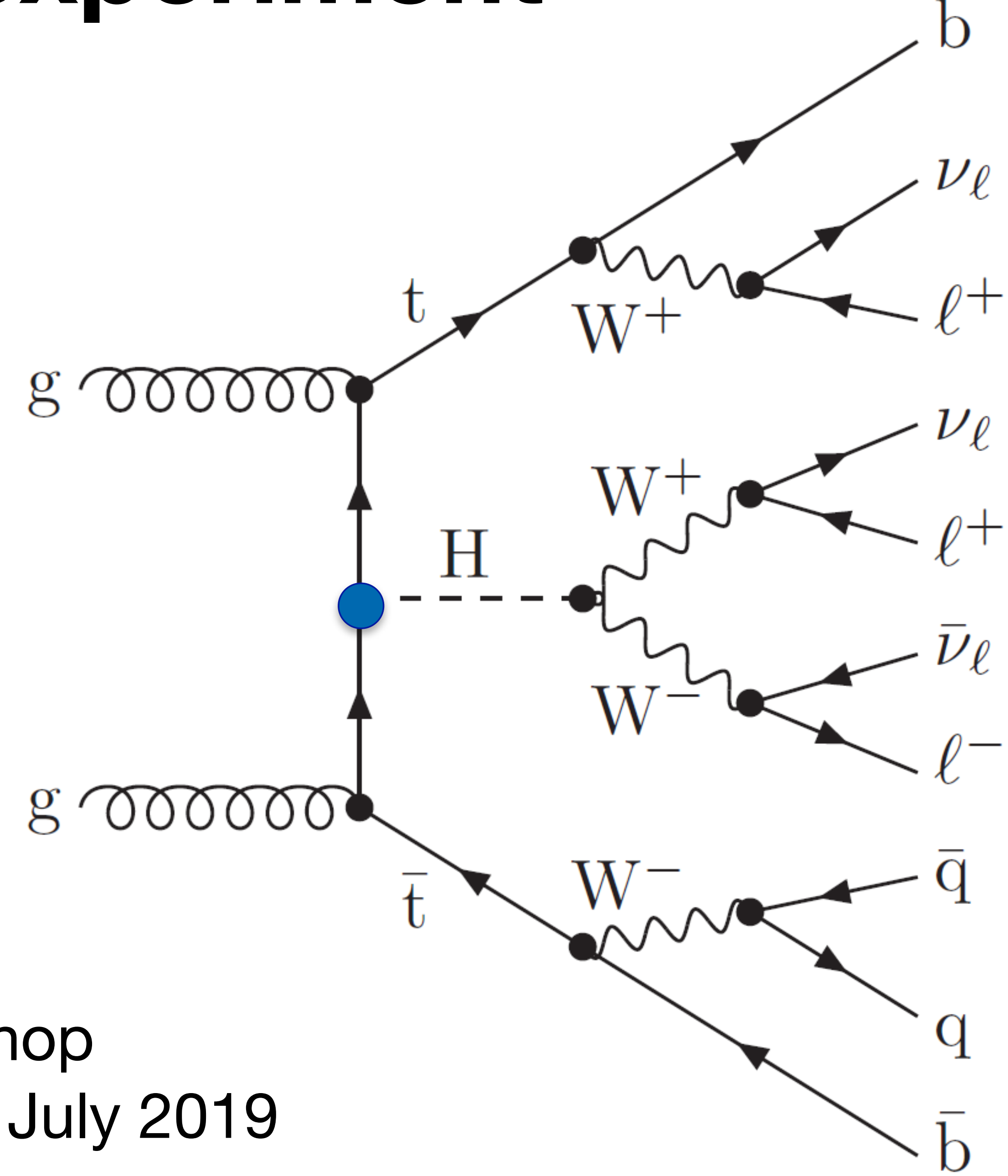
Measurements of Higgs boson production in association with top quarks with the CMS experiment

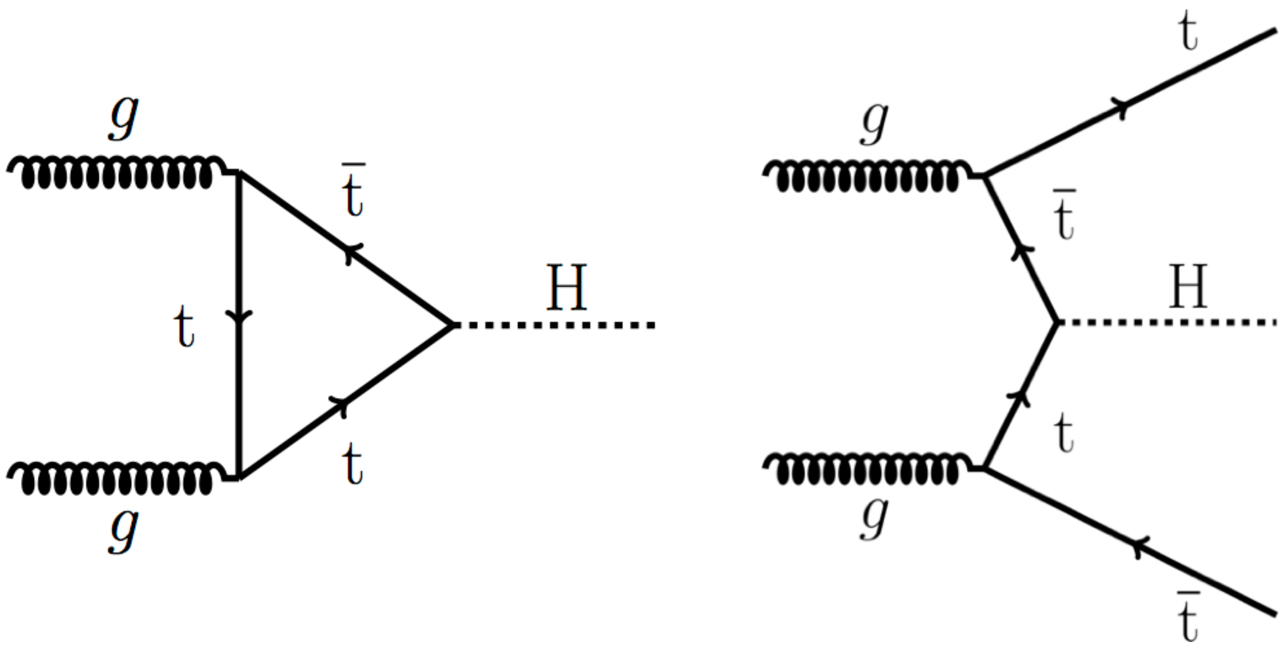


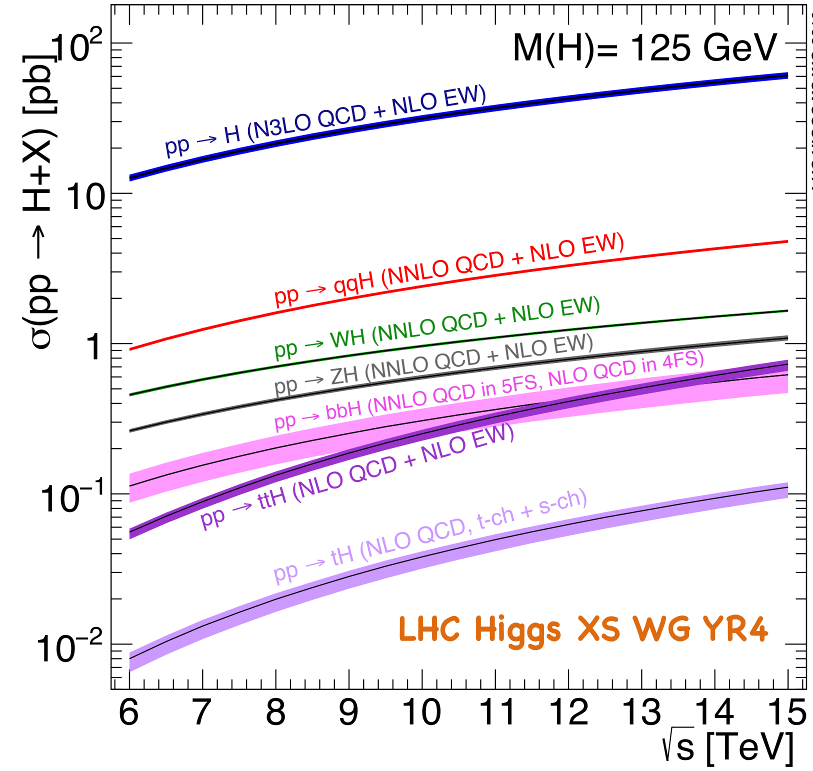
Marco Peruzzi (CERN)
on behalf of the CMS Collaboration

10th Higgs Hunting workshop
Orsay - Paris, France, 29th - 31st July 2019

Introduction

J. High Energ. Phys. (2014) 2014: 87



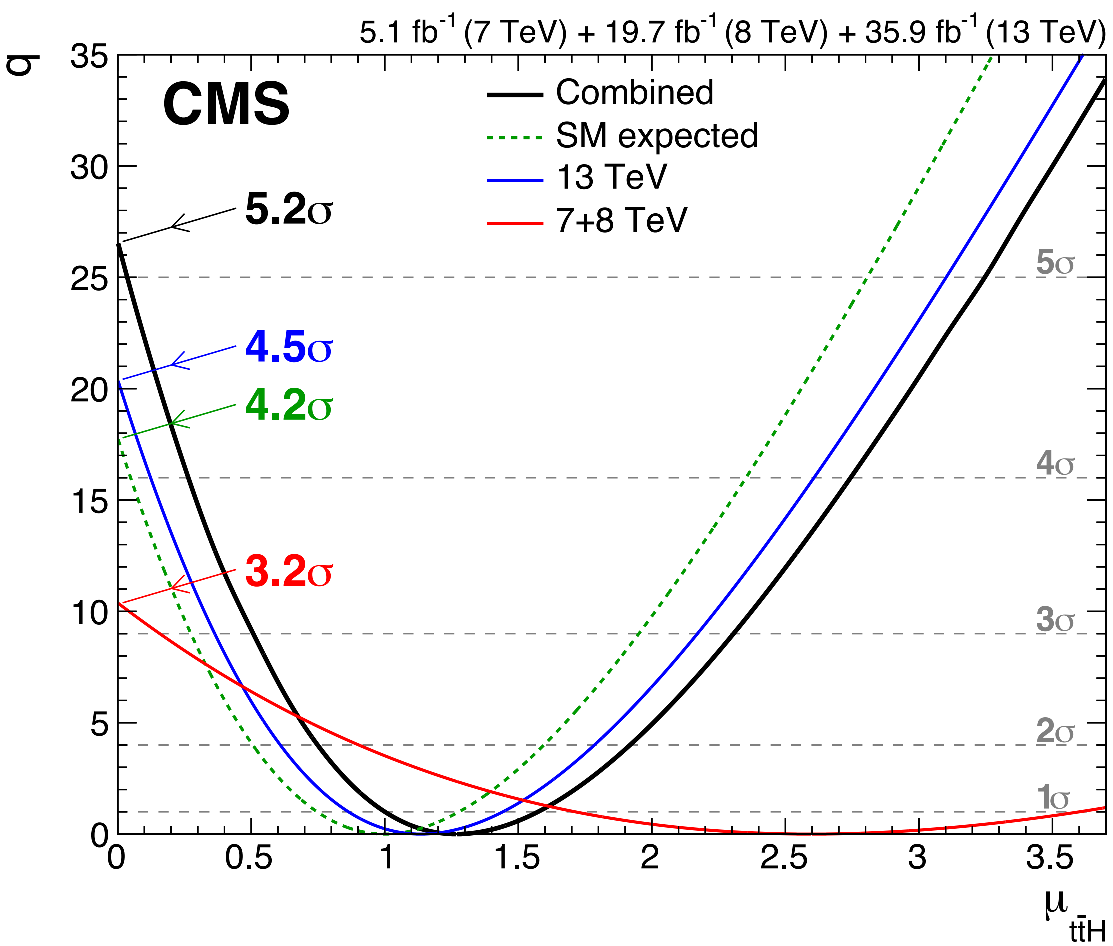


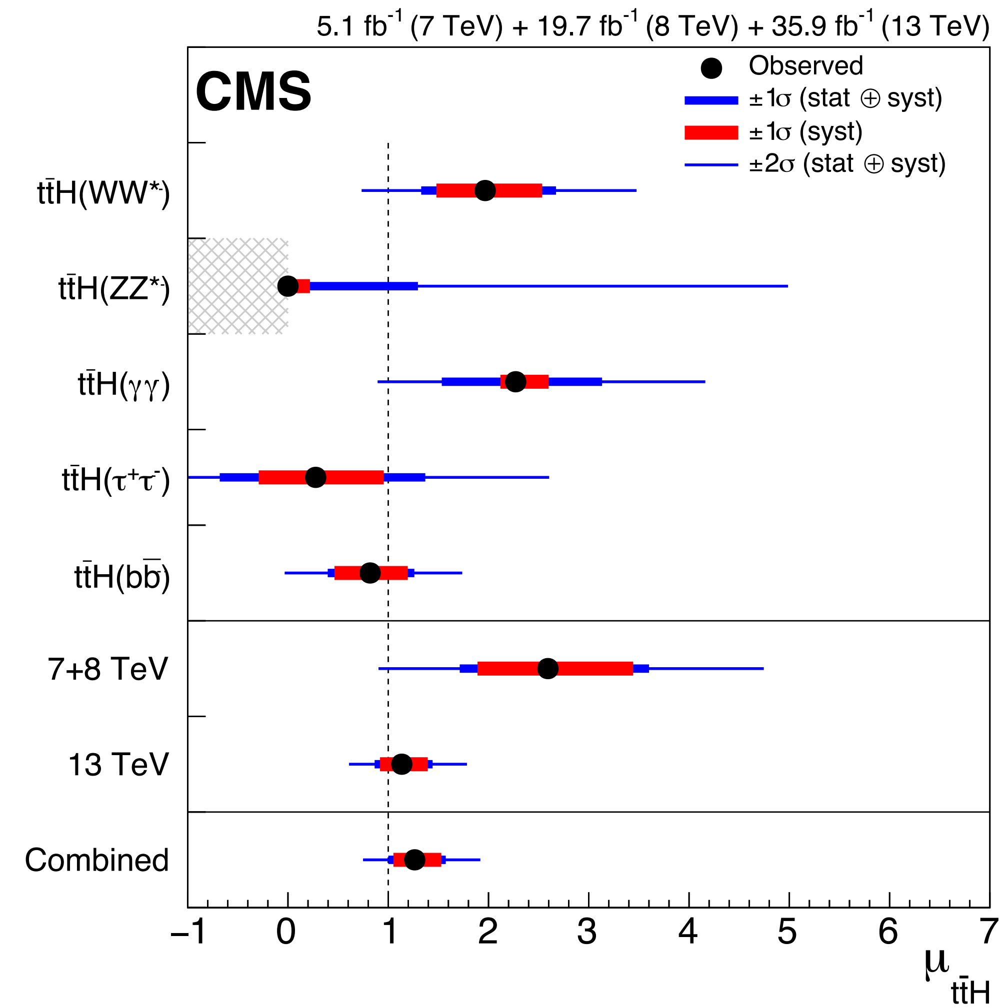
- **top - Higgs coupling** at the LHC:
  - **associated production provides direct sensitivity** (tree level coupling)
  - indirect sensitivity from gluon fusion (without BSM particles in the loop)
- Specific challenging aspects for each Higgs decay channel:
  - **ttH, H → bb: complex hadronic backgrounds (tt+heavy flavour)**
  - **ttH, H → WW*,ZZ*,$\tau\tau$ → leptons: tt+V backgrounds, mis-identified leptons**
  - **ttH, H → γγ , 4ℓ: cleanest, but low decay rate**

ttH measurements in CMS

- ttH production was first established by CMS with Run 1 and 2016 data:

Phys. Rev. Lett. 120 (2018) 231801

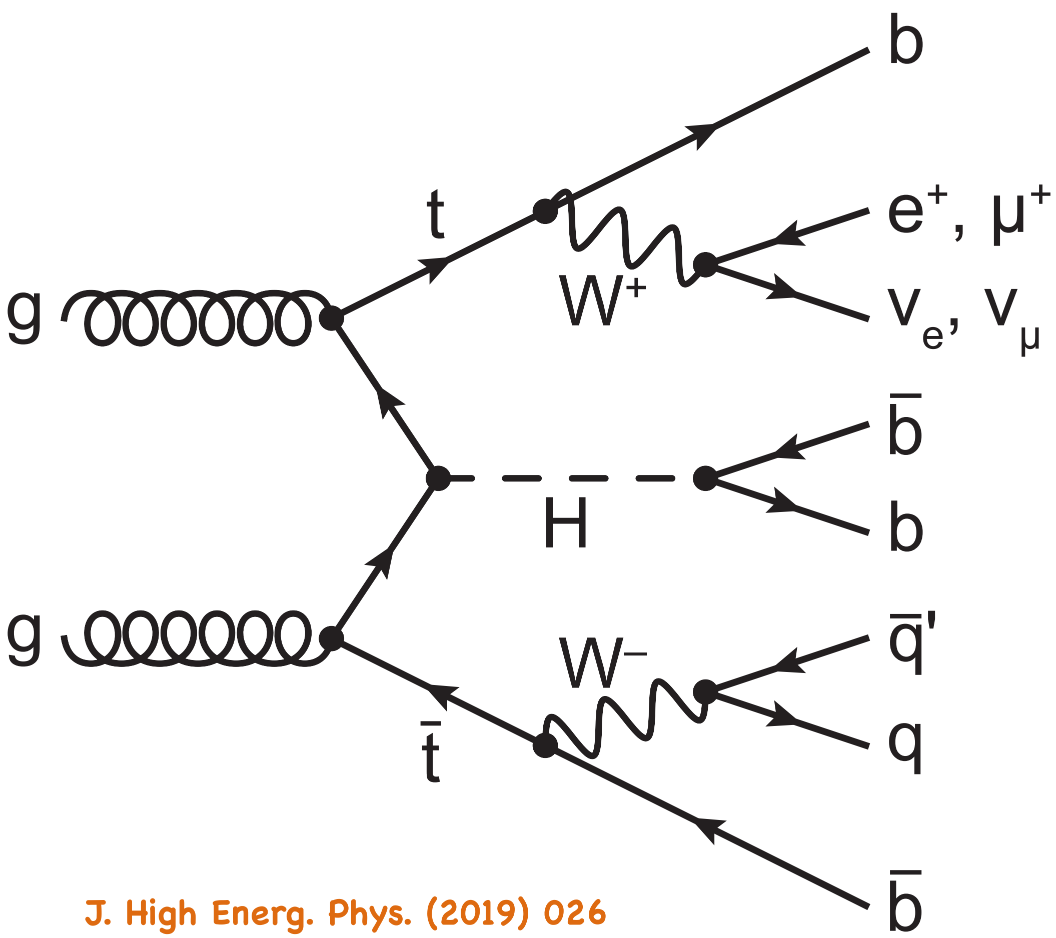




- Searches in **all channels have been improved and extended** since then:
  - **4ℓ**: with the full Run 2 dataset - **137 fb^{-1}**
  - **bb, multi-lepton, γγ**: based on the 2016 and 2017 datasets - **77.4 fb^{-1}**

$t\bar{t}$ + b-jets final states

CMS PAS HIG-18-030



J. High Energ. Phys. (2019) 026

Challenging final state:

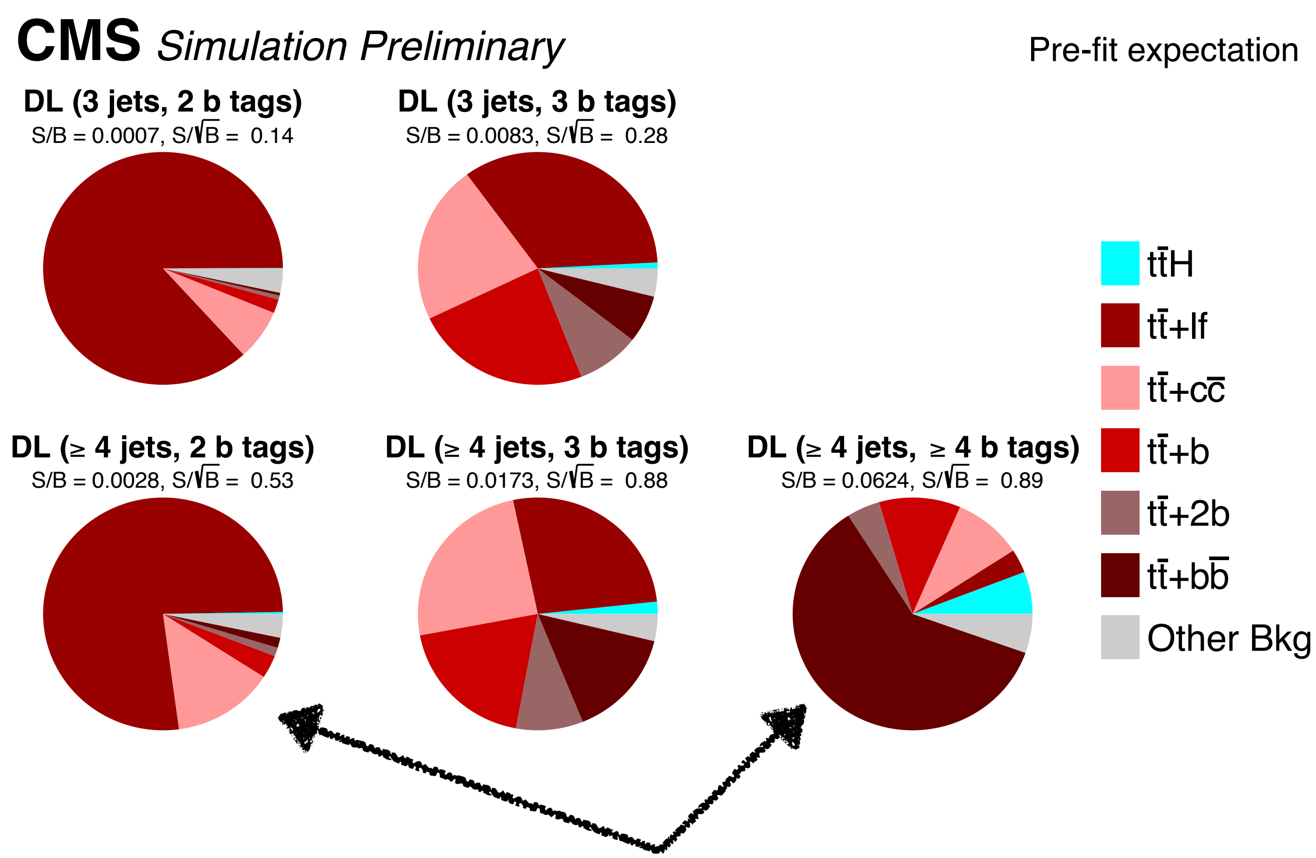
- irreducible background from **$t\bar{t}$ + heavy flavour jets**: modeled with Powheg [**new**: per-event weights for ISR/FSR uncertainties]
- large **jet combinatorics**
- **QCD background** in 0-lepton categories

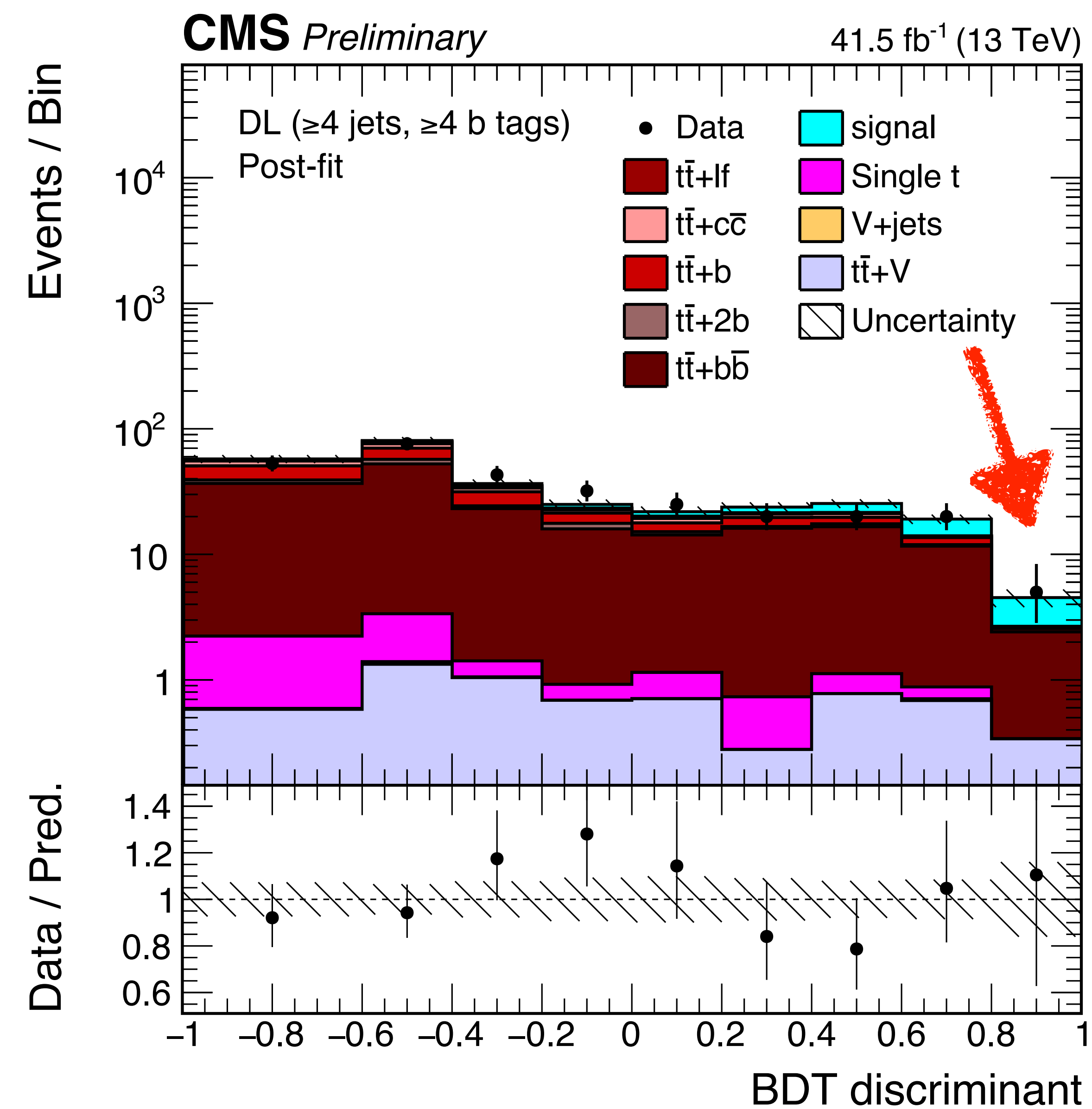
Key ingredients:

- **b-tagging: efficiency improved by ~10%** at 1% mistag probability w.r.t. 2016 (new pixel detector and DeepCSV algorithm)
- advanced methods to reduce backgrounds (**neural networks, matrix element**)

ttH→bb, 2ℓ categories

CMS PAS HIG-18-030



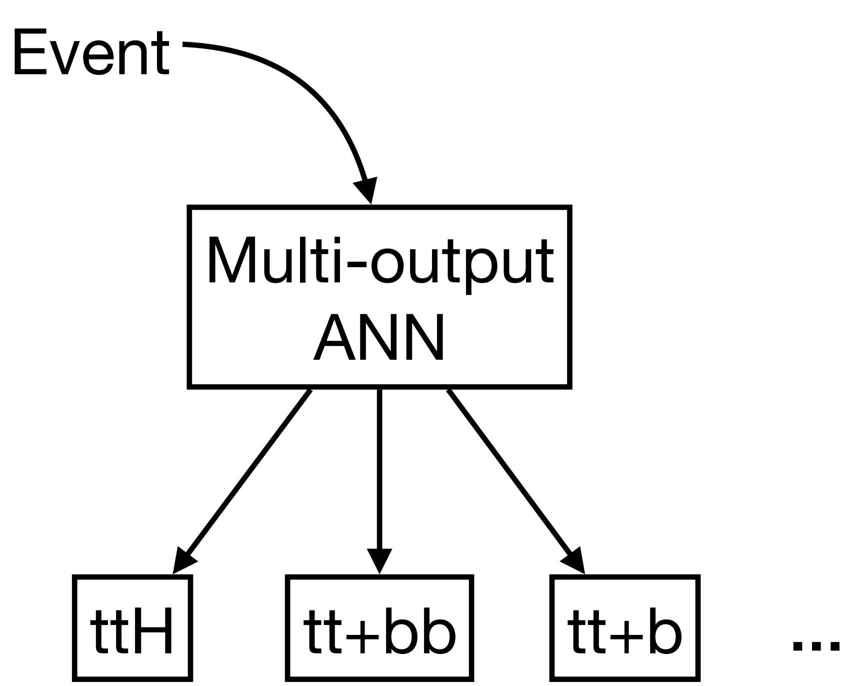


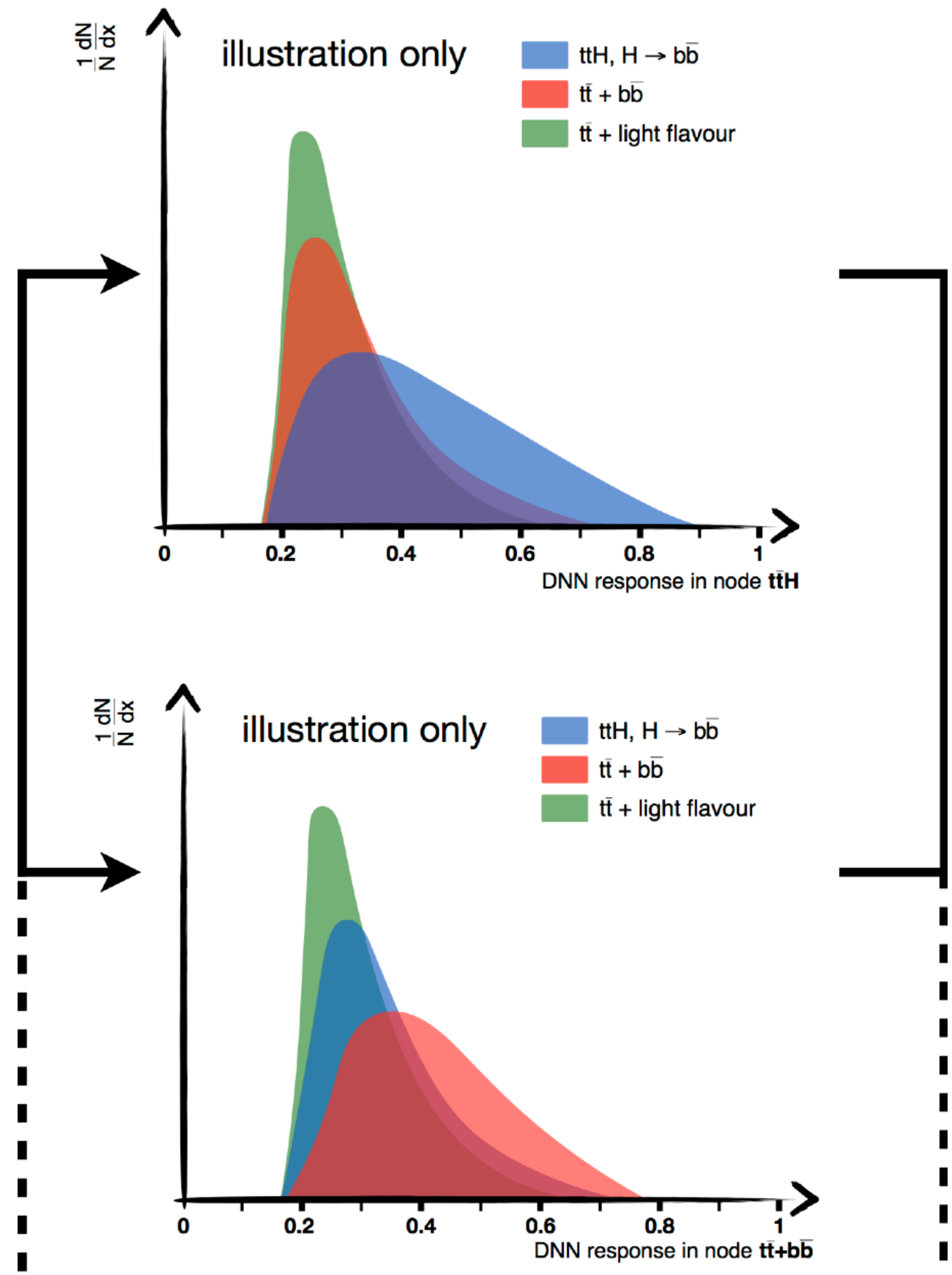
- b-tag categories allow to **separate and constraint different background sub-processes (e.g. tt+bb, tt+light)**
- **BDT discriminant** trained against inclusive tt background:
  - kinematic variables of physics objects, continuous b-tagging information
  - matrix element score, invariant masses and angular separations

ttH→bb, 1ℓ categories

CMS PAS HIG-18-030

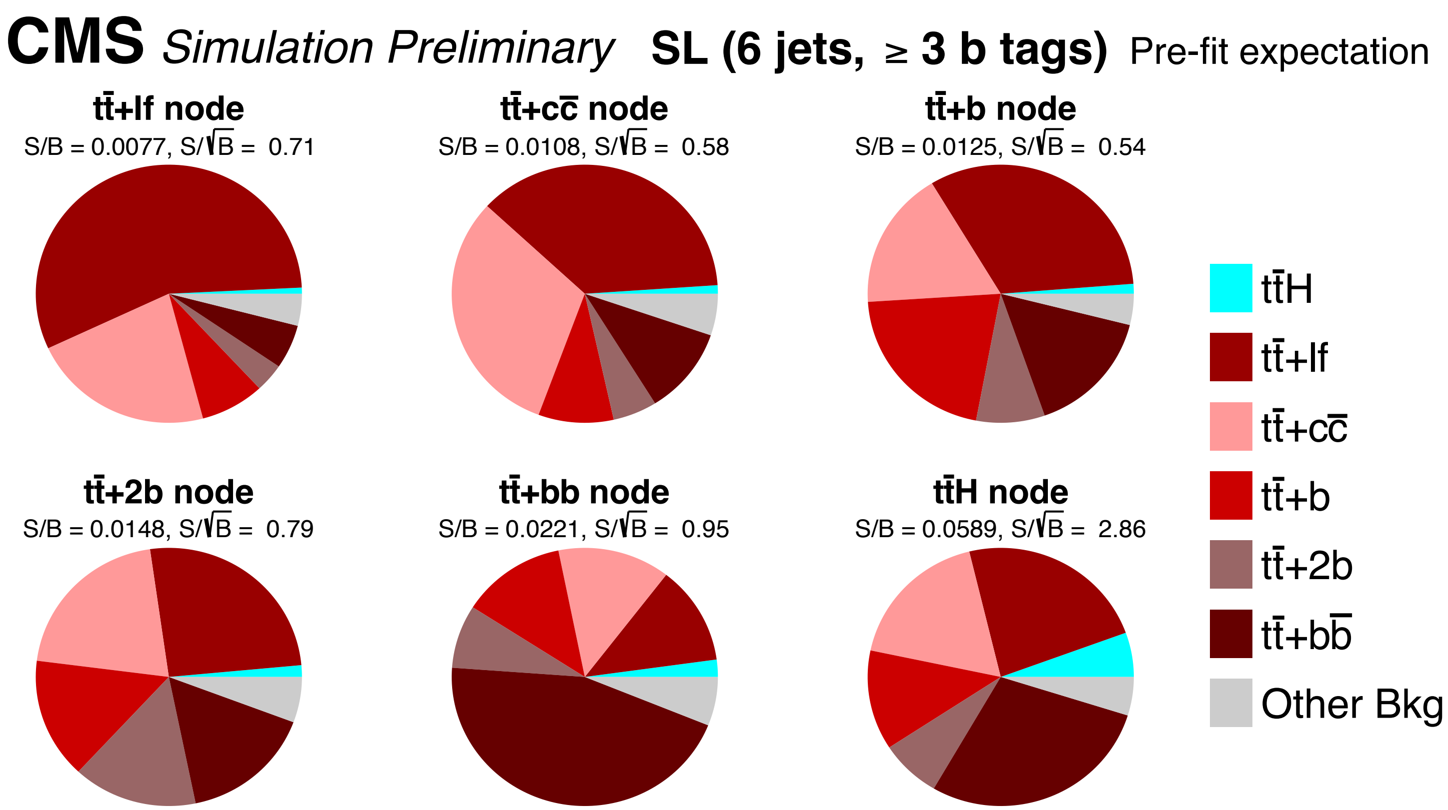
- Event categories: **4, 5, ≥6 jets**
- Multi-classification using an **artificial neural network (ANN)**, with similar inputs as for the 2ℓ case

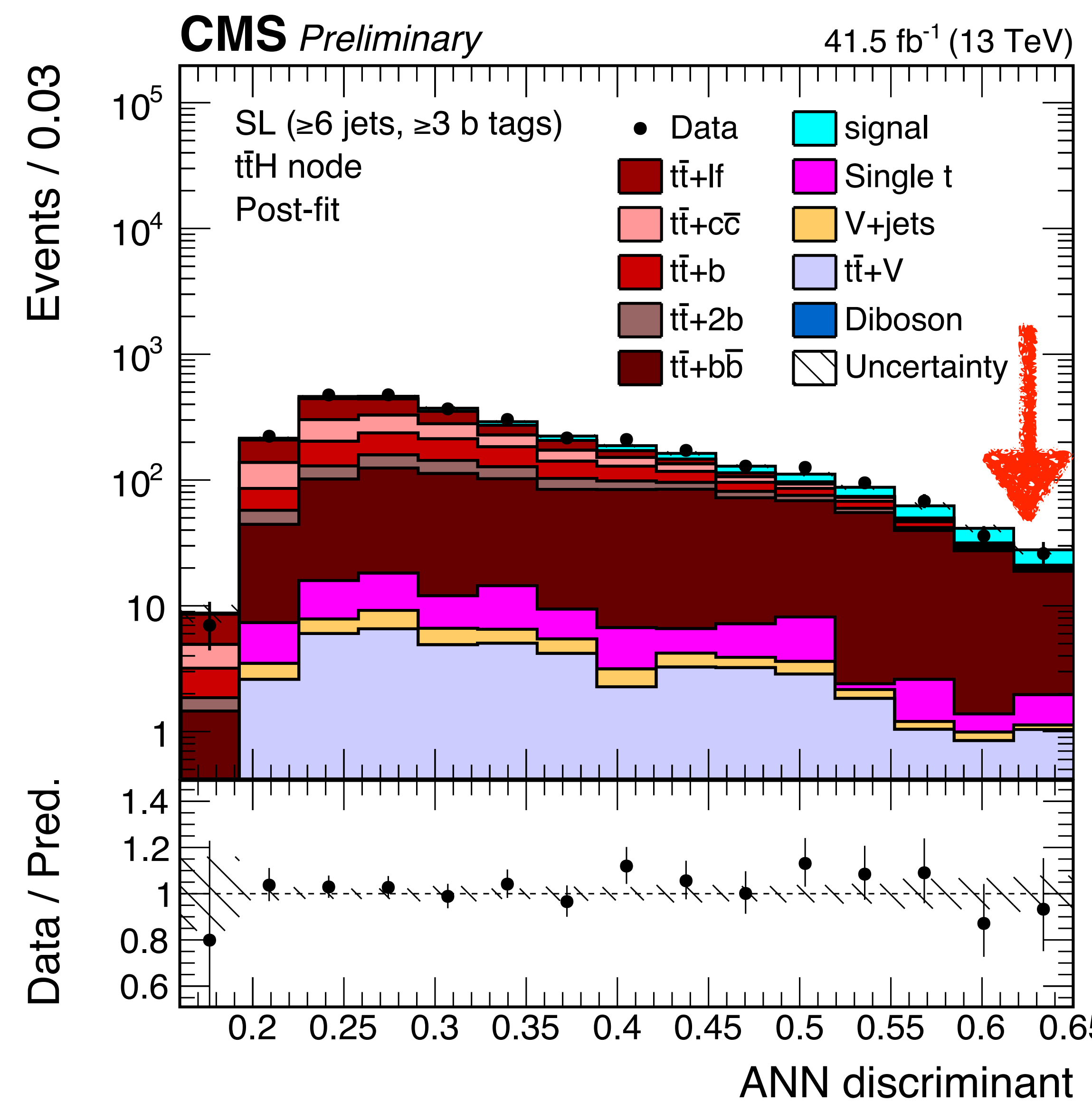




ttH→bb, 1ℓ categories

CMS PAS HIG-18-030



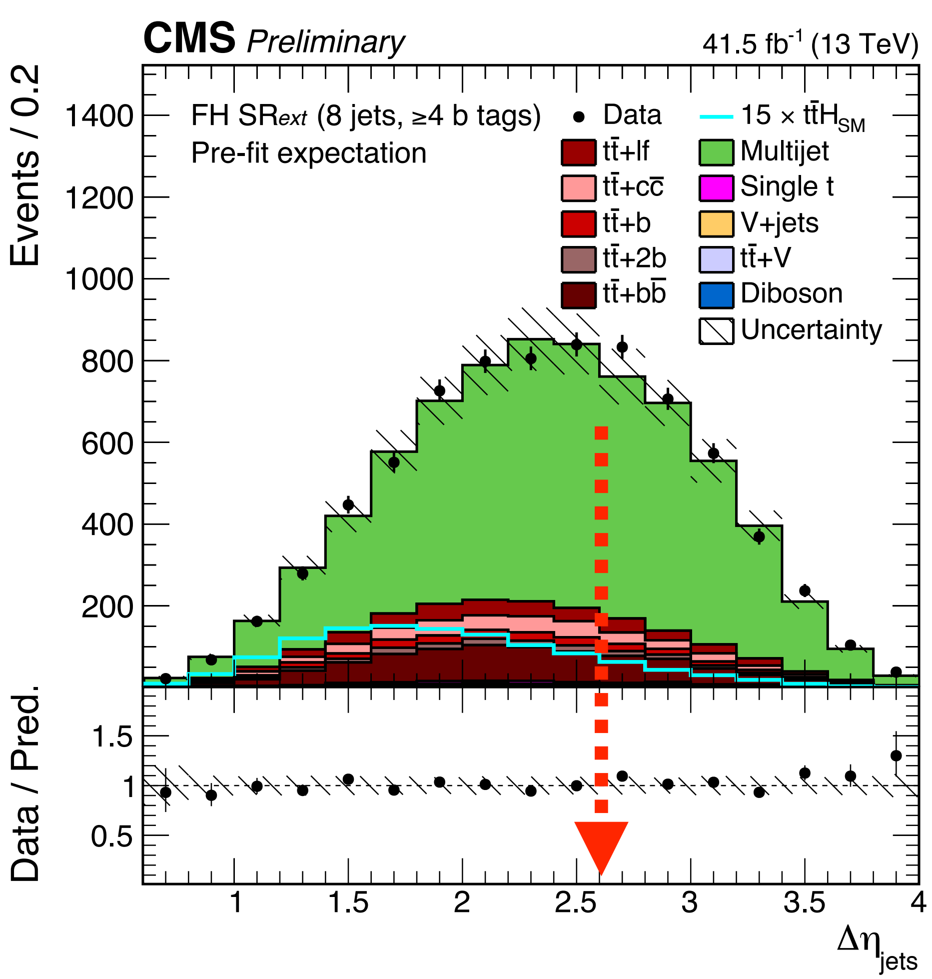


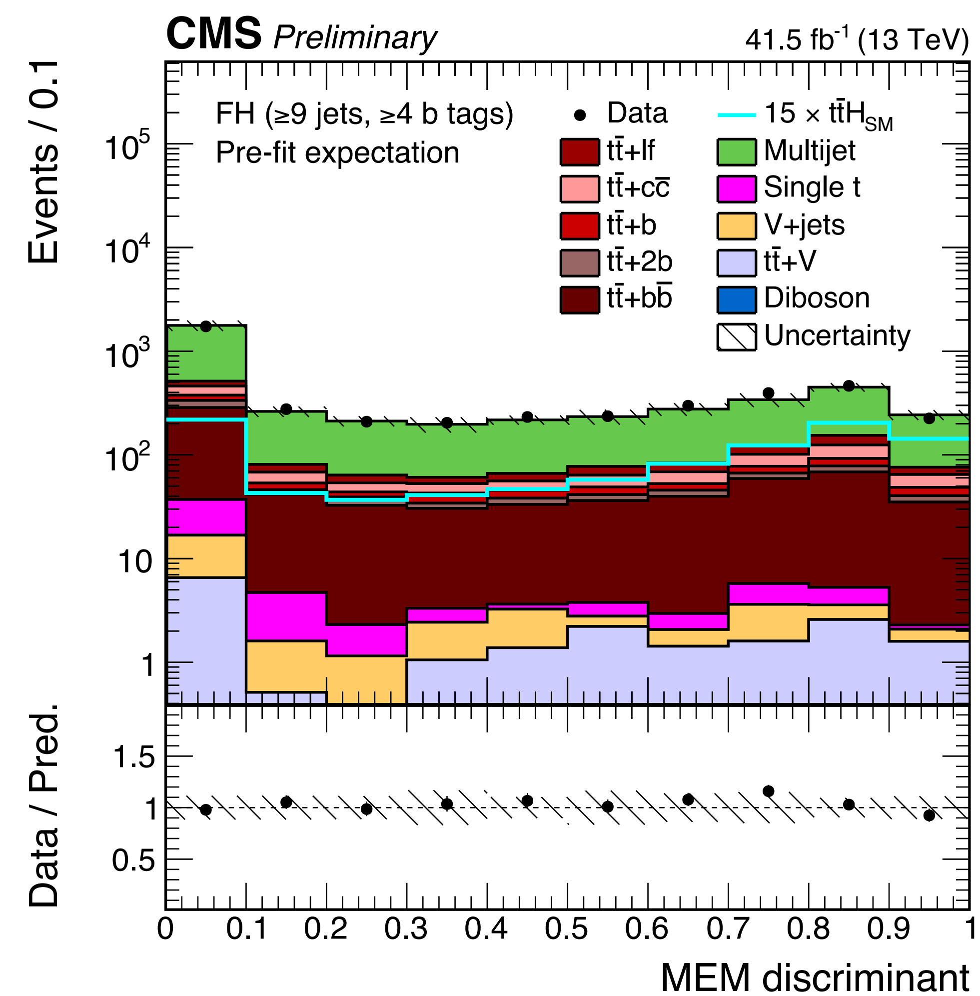
- Scores interpreted as **per-event process probability estimates**
- Events splitted in sub-categories according to the most probable hypothesis

ttH→bb, 0ℓ categories

CMS PAS HIG-18-030

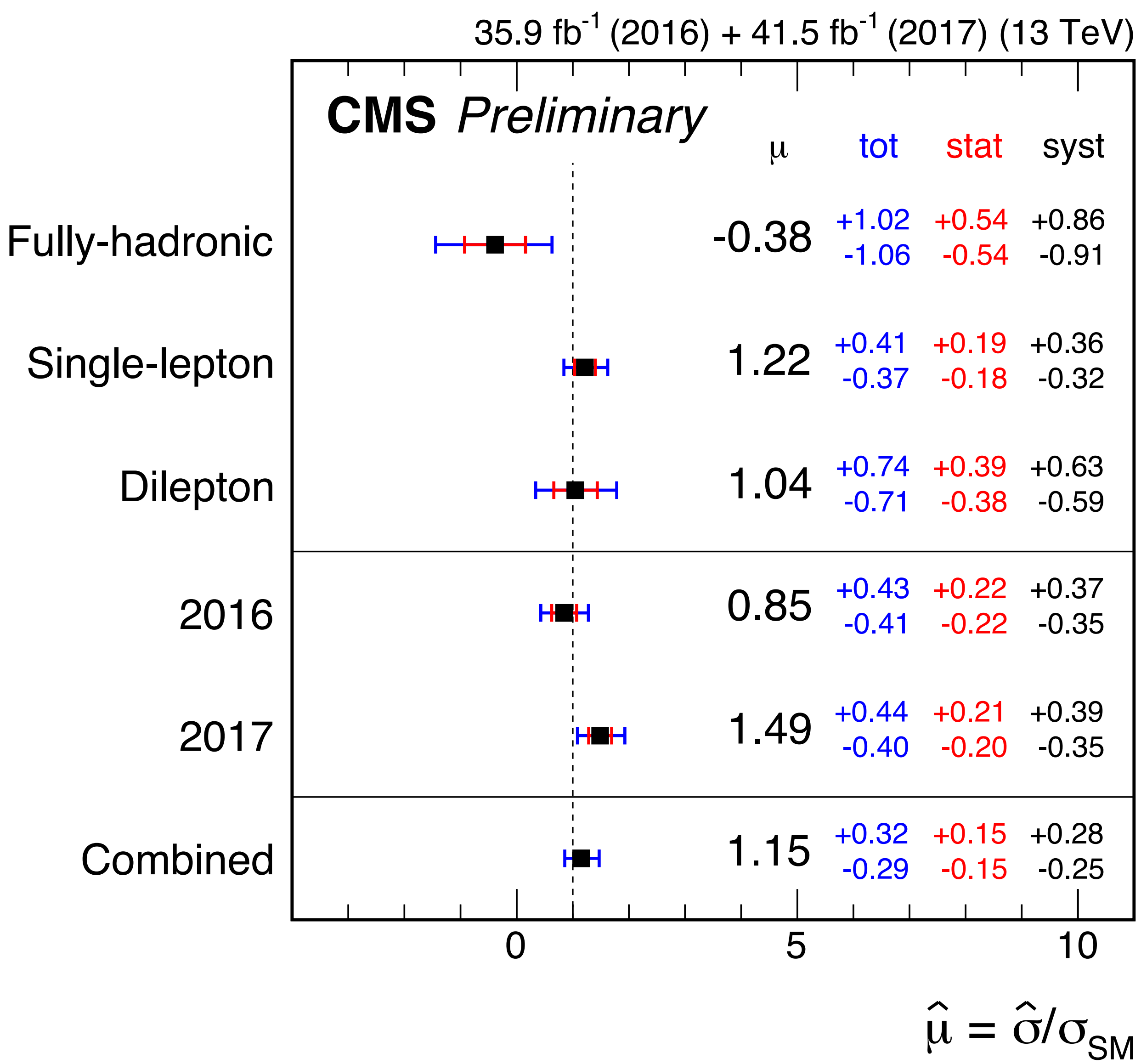
- Handles to reduce the dominant QCD background:
  - discrimination of **quark vs. gluon** jets (likelihood ratio)
  - variable built from **average jet angular separation [new]**
- Signal extracted by a fit to a **matrix element discriminant**
- Background normalized from control regions with looser b-tag requirements





Results

CMS PAS HIG-18-030

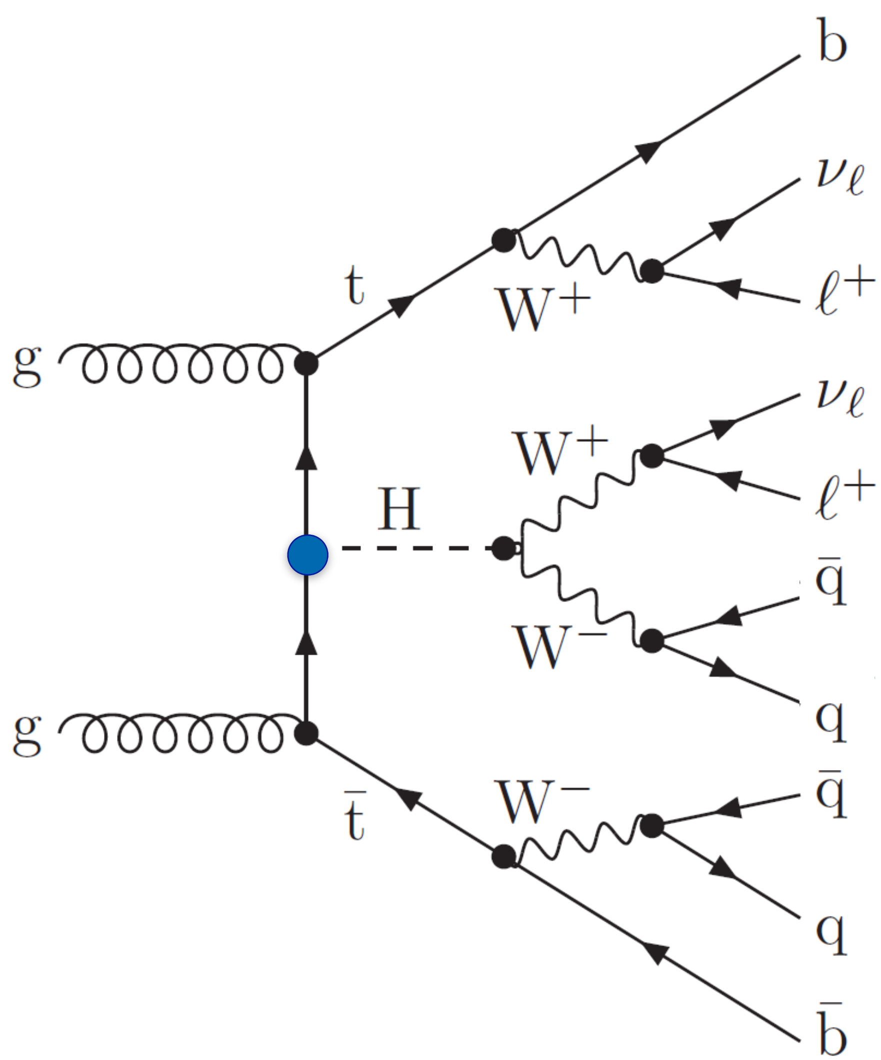


Uncertainty source	$\Delta\hat{\mu}$
Total experimental	$+0.15/-0.13$
b tagging	$+0.08/-0.07$
jet energy scale and resolution	$+0.05/-0.04$
Total theory	$+0.23/-0.19$
signal	$+0.15/-0.06$
$t\bar{t}$+hf modelling	$+0.14/-0.15$
QCD background prediction	$+0.10/-0.08$
Size of simulated samples	$+0.10/-0.10$
Total systematic	$+0.28/-0.25$
Statistical	$+0.15/-0.15$
Total	$+0.32/-0.29$

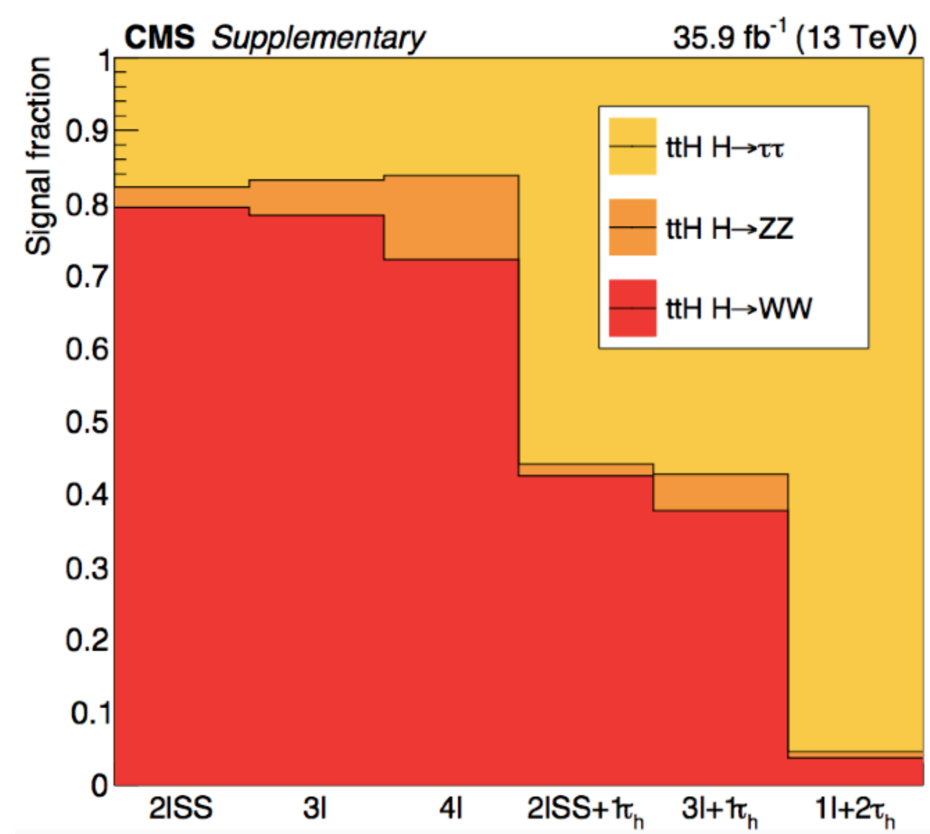
- 2017 result: $\mu = 1.49^{+0.44}_{-0.40}$, 3.7σ significance (2.6σ expected)
- **2016+2017 combination: $\mu = 1.15^{+0.32}_{-0.29}$, 3.9σ significance (3.5σ expected)**
- **Very significant improvements in the control of backgrounds result in an impressive boost of the analysis sensitivity**

ttH leptonic final states

CMS PAS HIG-18-019



- Target **Higgs decays to WW*, ZZ*, $\tau\tau$**
- Channels:
    - **1 lepton + 2 τ_h**
    - **2 same-sign leptons + 0,1,2 τ_h**
    - **3 leptons + 0,1 τ_h**
    - **4 leptons**



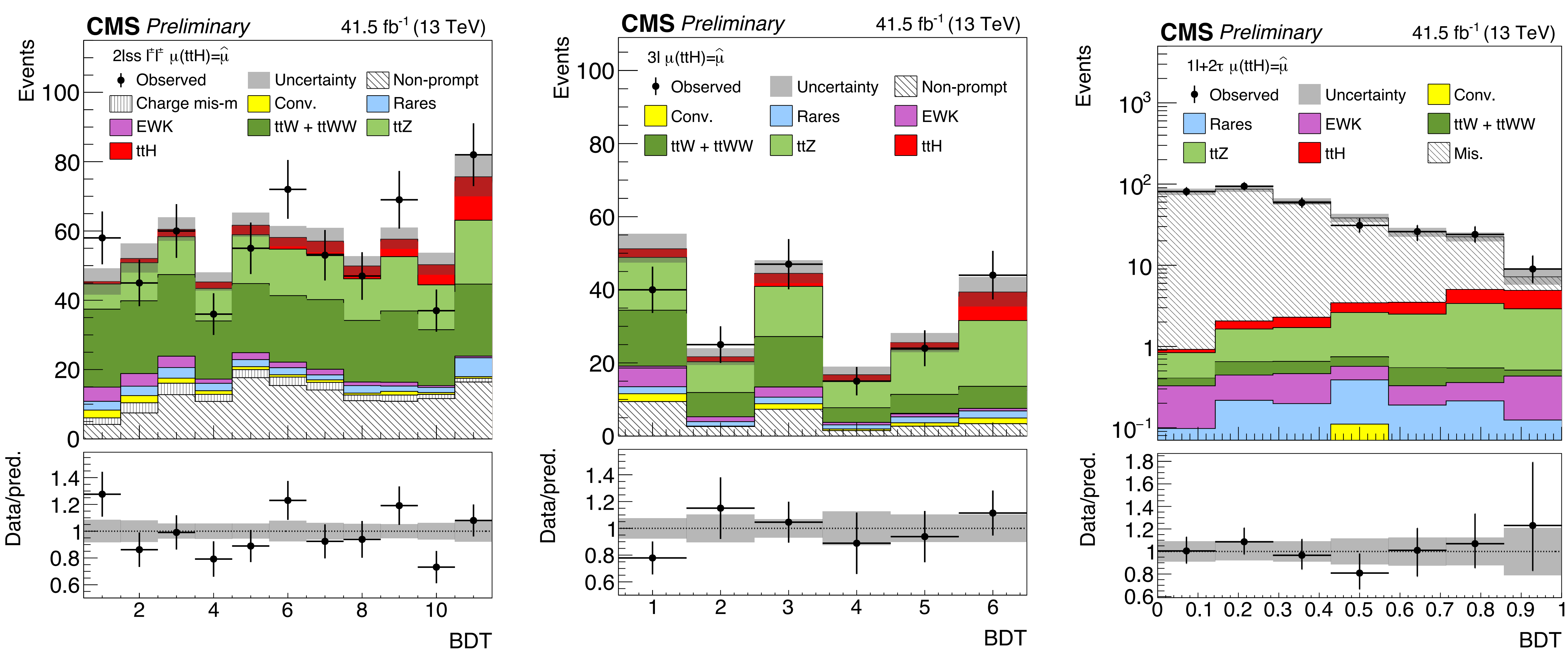
Main selection handles:

- At least 2 loose or 1 medium **b-tagged jets**
- **Jet multiplicity** (≥4 jets in 2-lep, ≥2 jets in 3-lep)

Analysis strategy

CMS PAS HIG-18-019

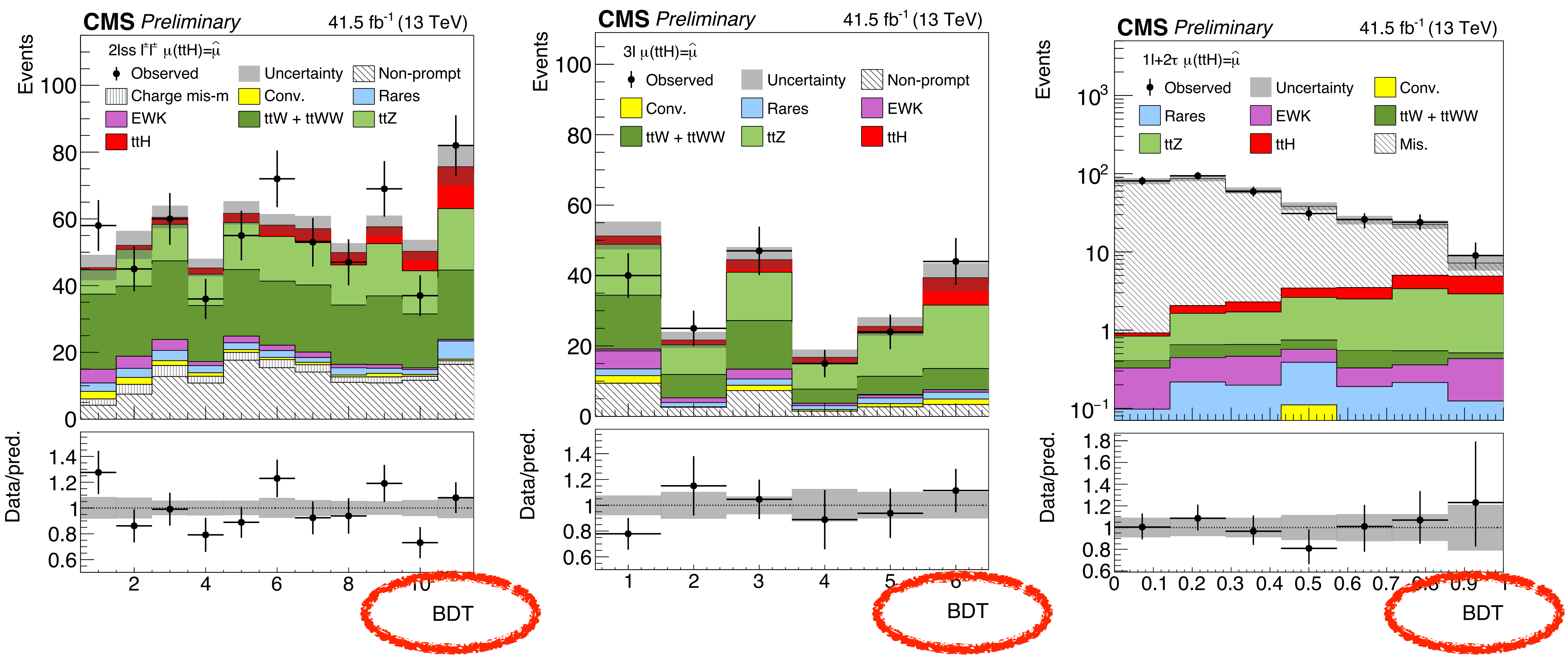
- Categories based on **lepton flavor, b-jet multiplicity and lepton charge**
- Main sources of background:
    - prompt leptons from **ttV**, VV: from simulation normalized in control regions
    - **non-prompt leptons** in tt events: predicted from data



Analysis strategy

CMS PAS HIG-18-019

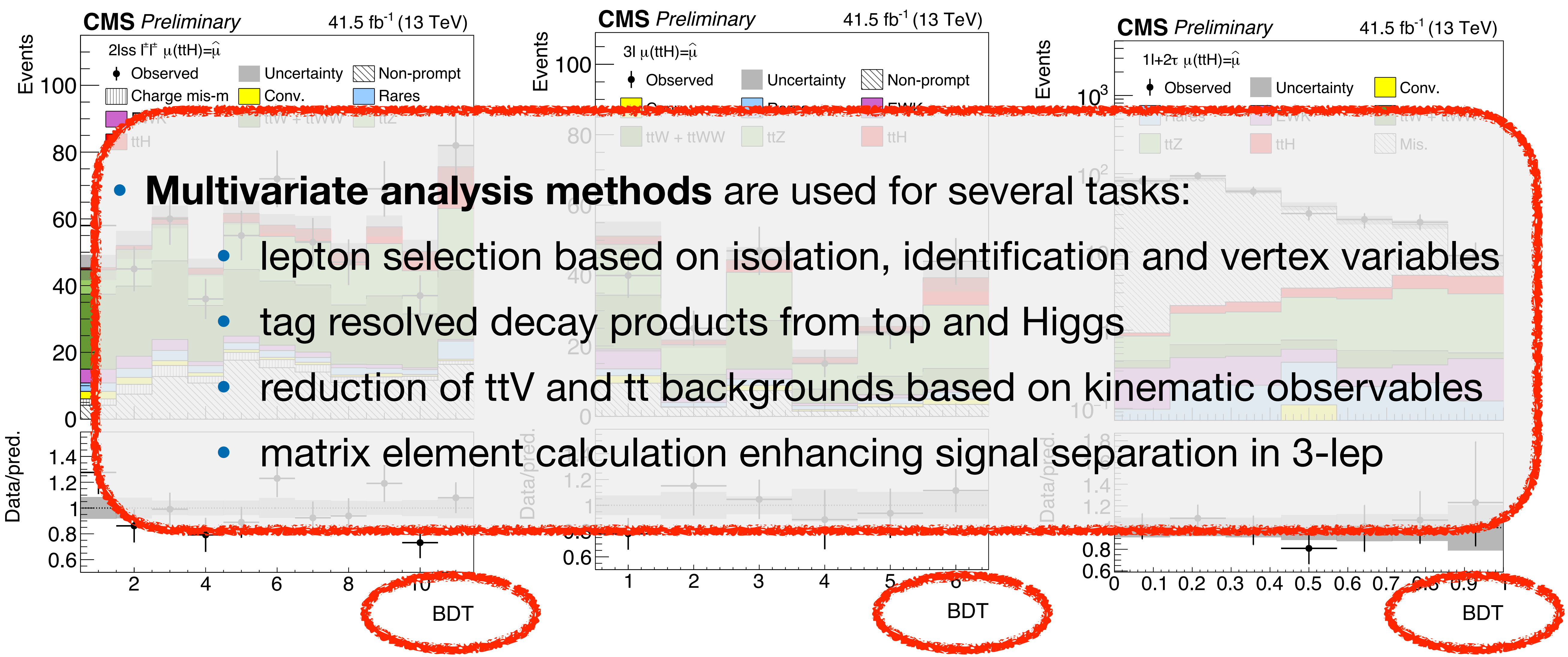
- Categories based on **lepton flavor, b-jet multiplicity and lepton charge**
- Main sources of background:
    - prompt leptons from **ttV**, VV: from simulation normalized in control regions
    - **non-prompt leptons** in tt events: predicted from data



Analysis strategy

CMS PAS HIG-18-019

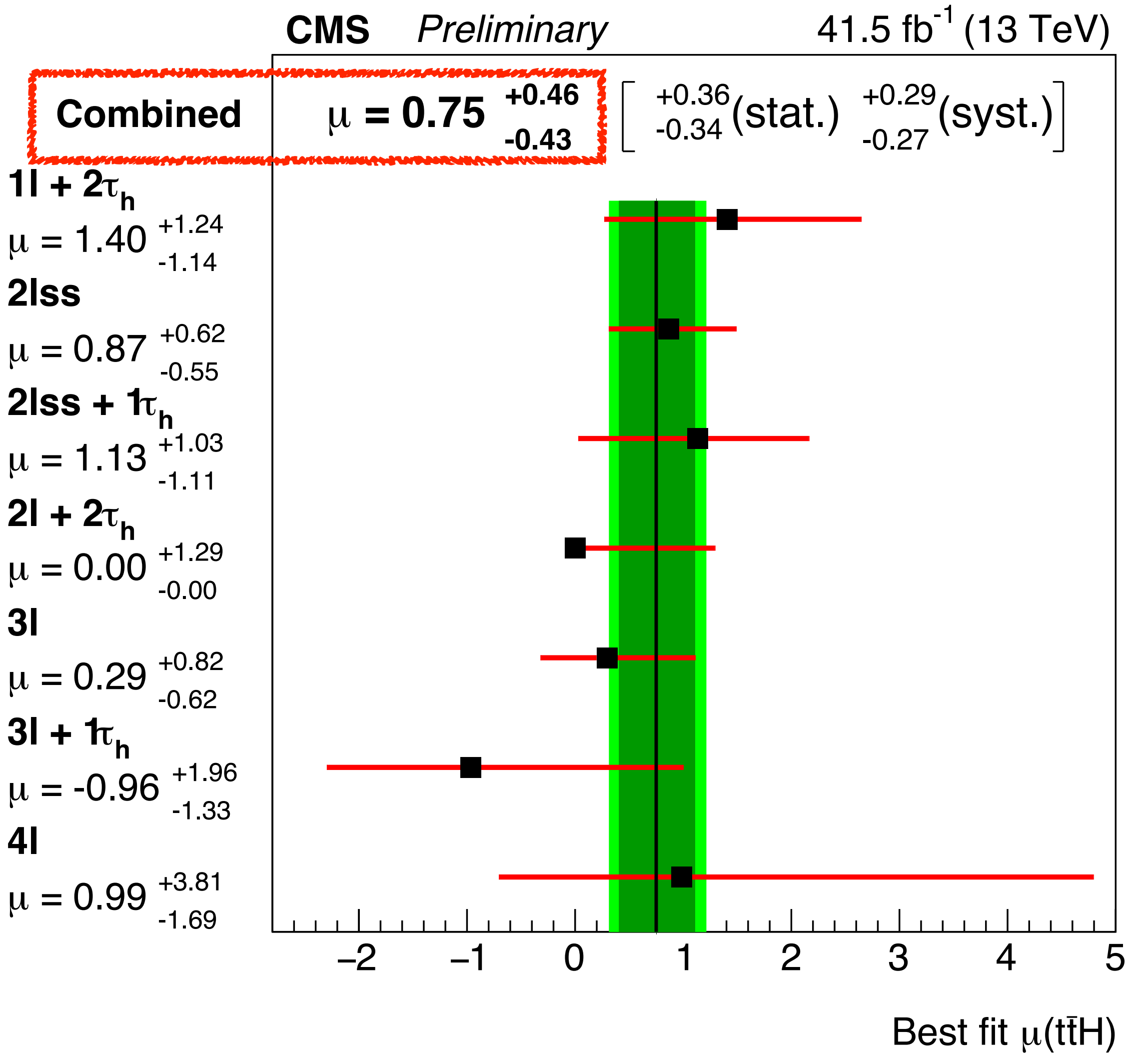
- Categories based on **lepton flavor, b-jet multiplicity and lepton charge**
- Main sources of background:
    - prompt leptons from **ttV**, VV: from simulation normalized in control regions
    - **non-prompt leptons** in tt events: predicted from data



- **Multivariate analysis methods** are used for several tasks:
    - lepton selection based on isolation, identification and vertex variables
    - tag resolved decay products from top and Higgs
    - reduction of ttV and tt backgrounds based on kinematic observables
    - matrix element calculation enhancing signal separation in 3-lep

Results

CMS PAS HIG-18-019



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

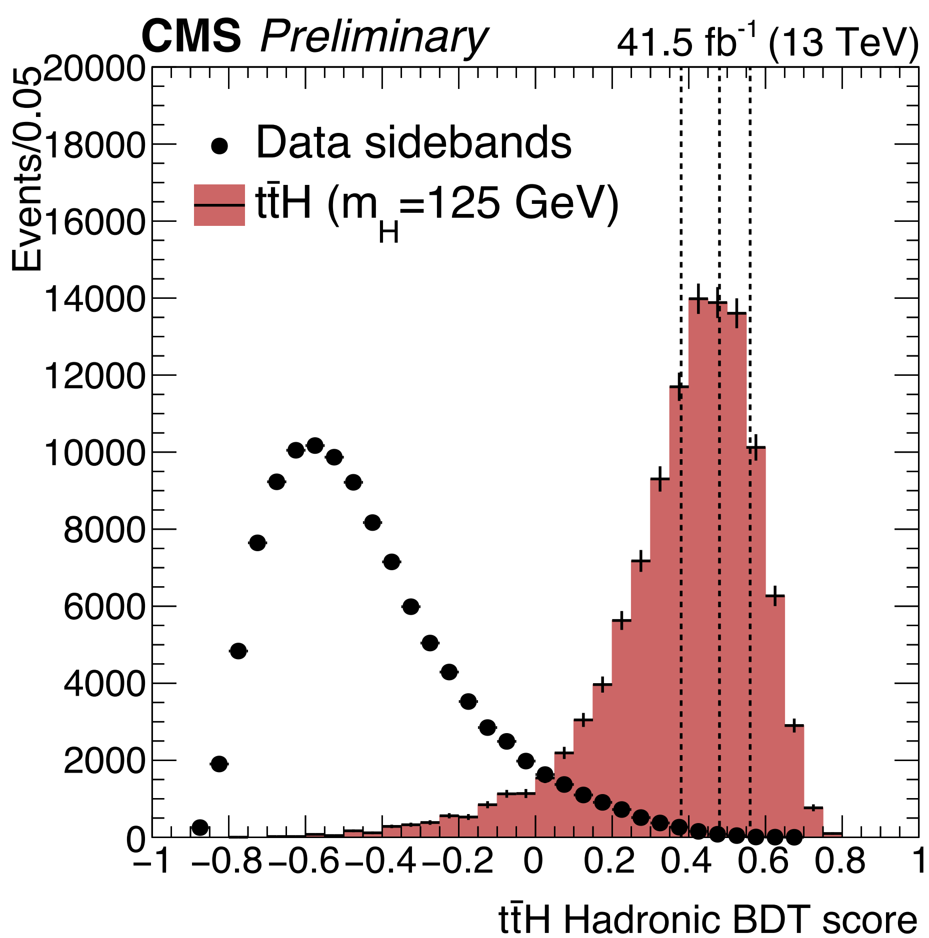
- **ttV and VV normalizations simultaneously fit using dedicated control regions**
- Comparable impacts from theoretical and experimental sources of systematic uncertainty

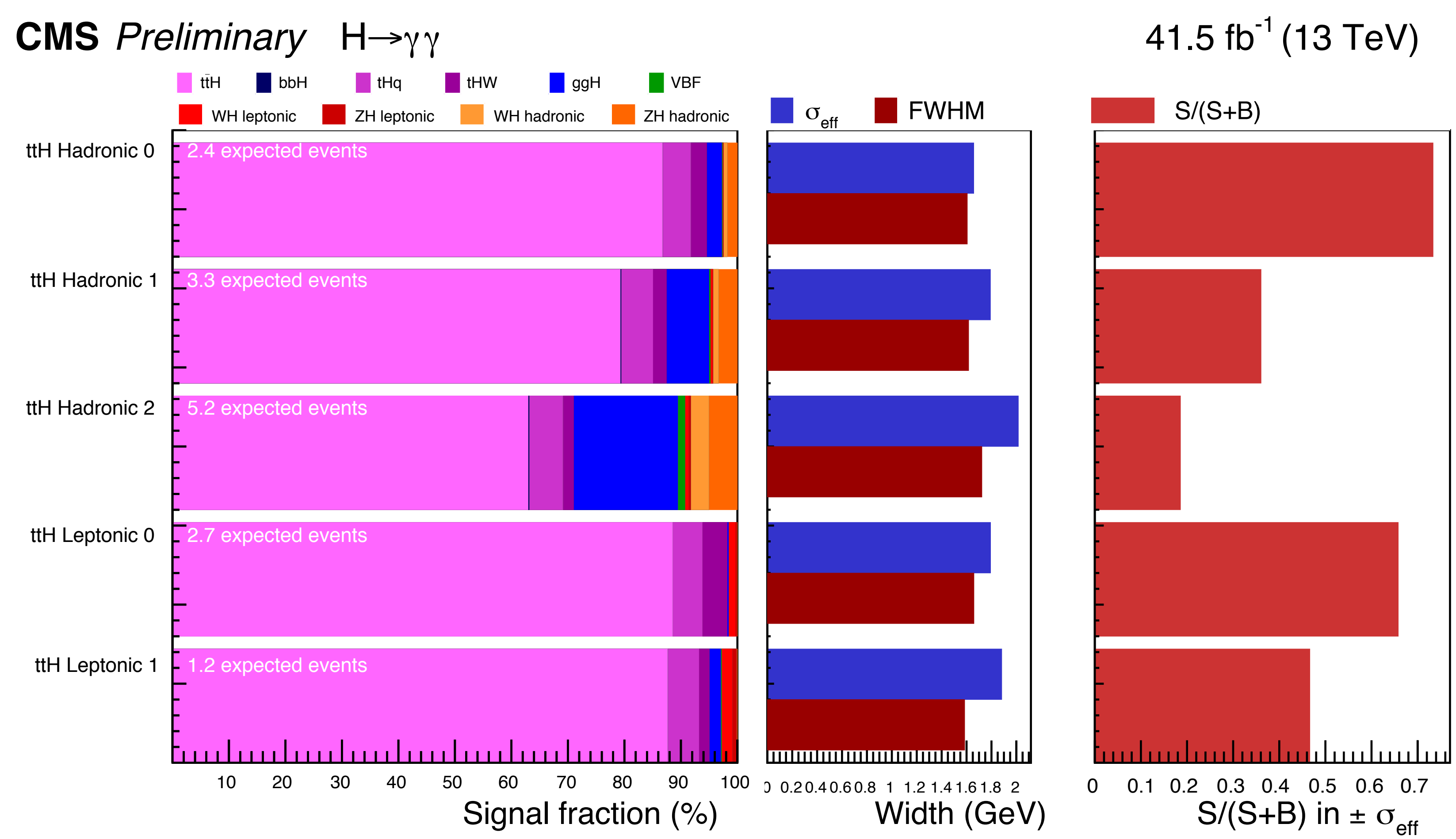
- 2017 result: $\mu = 0.75^{+0.46}_{-0.43}$, 1.7σ significance (2.9σ expected)
- **2016+2017 combination: $\mu = 0.96^{+0.34}_{-0.31}$, 3.2σ significance (4.0σ expected)**
- Post-fit ttV normalization shows a slight excess over the SM expectation (NLO cross section), coherent with dedicated tt+V measurements

ttH, H→γγ

CMS PAS HIG-18-018

- Fit strategy as for the general H→γγ analysis (see A. Kumar's talk yesterday)
- **Categories** targeting **hadronic (≥2j)** and **leptonic (1ℓ, ≥1j)** top decays
- Dedicated **BDT discriminants** to enhance the analysis sensitivity, simultaneously using identification and kinematic variables from leptons and jets

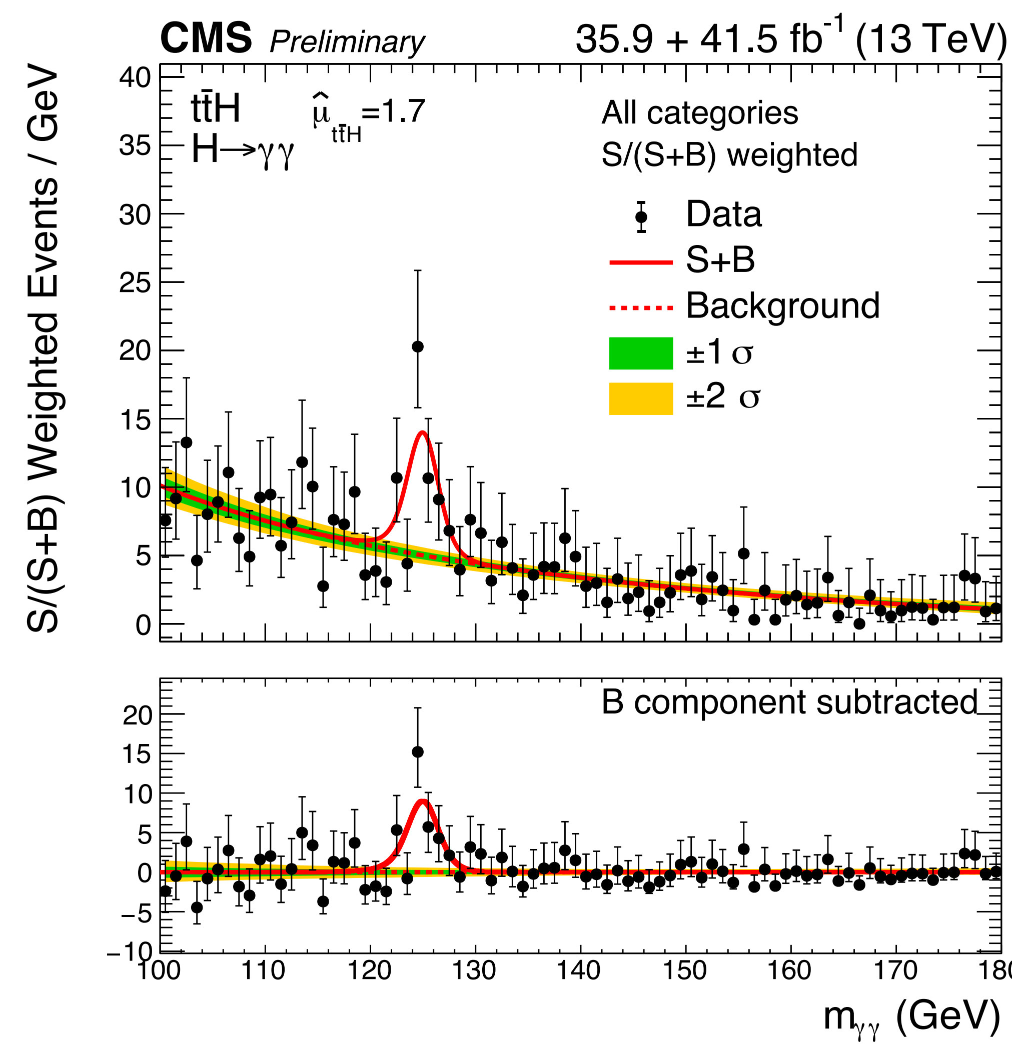


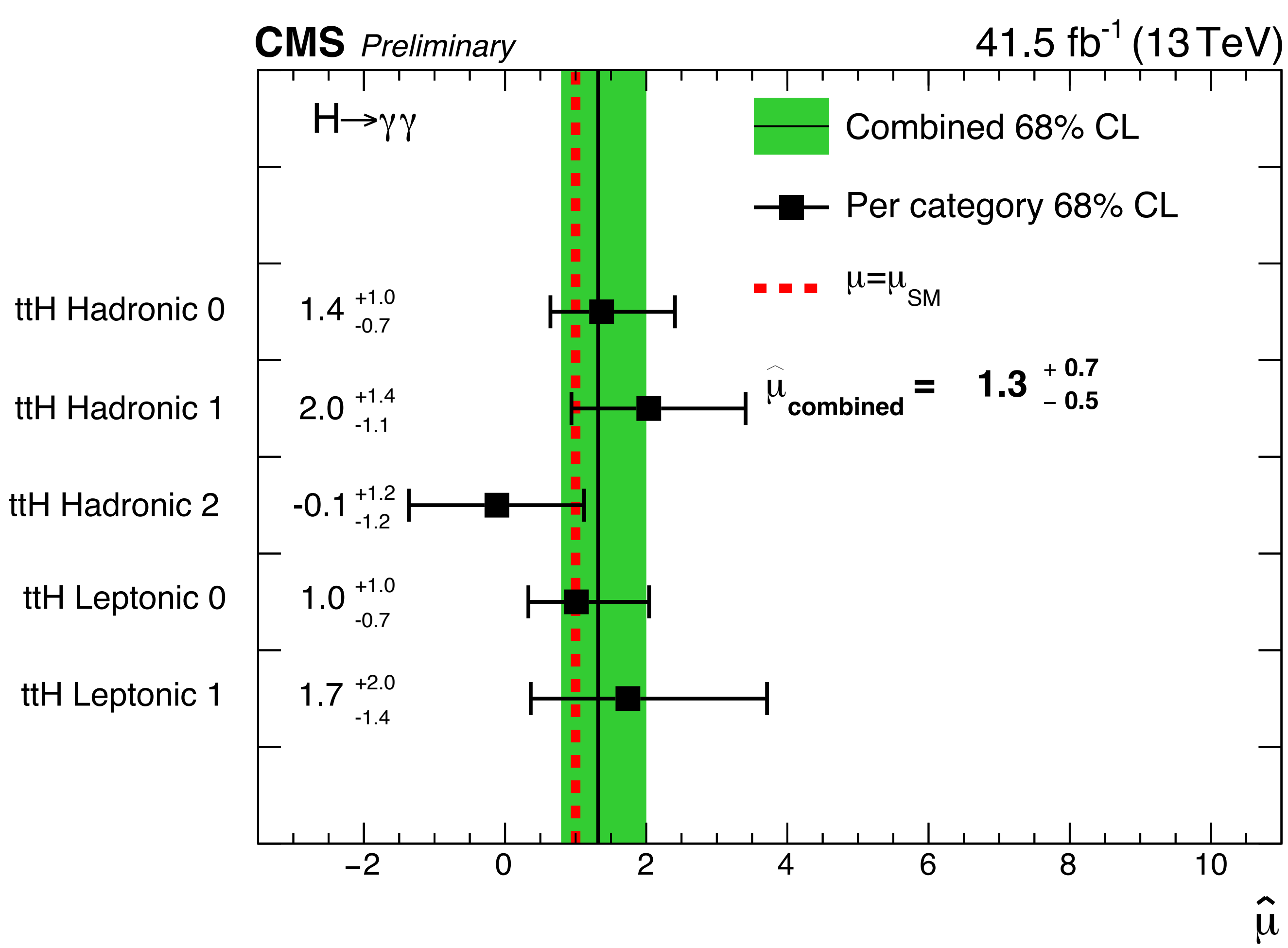


ttH, H→γγ

CMS PAS HIG-18-018

- Diphoton mass spectrum is fit in slices of the BDT discriminant score
- Background shape is determined from data using a discrete profiling method



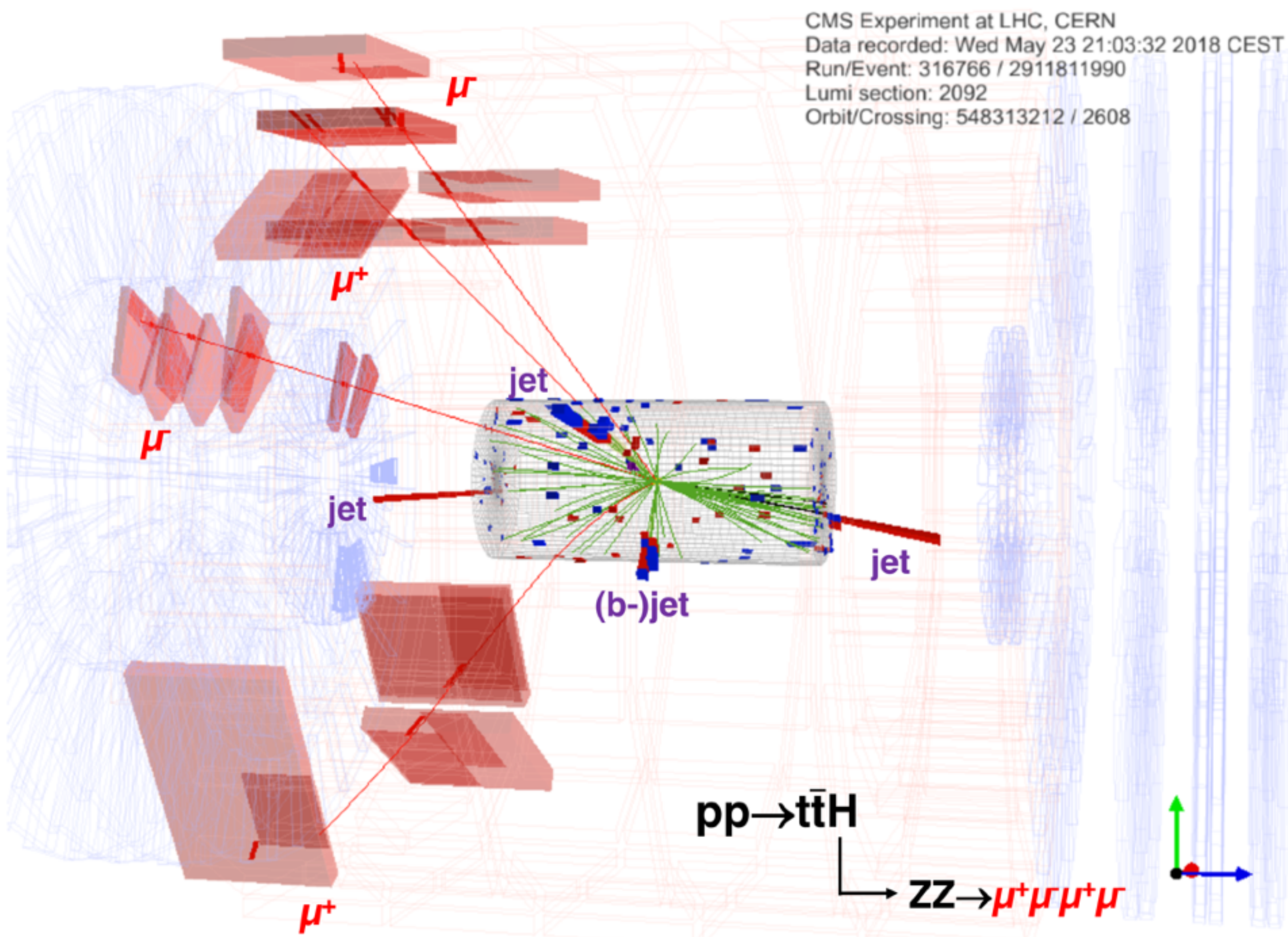


- 2017 result: $\mu = 1.3^{+0.7}_{-0.5}$
  3.1σ significance (2.2σ expected)
- **2016+2017 combination: $\mu = 1.7^{+0.6}_{-0.5}$**
  4.1σ significance (2.7σ expected)

ttH in the 4ℓ final state

CMS PAS HIG-19-001

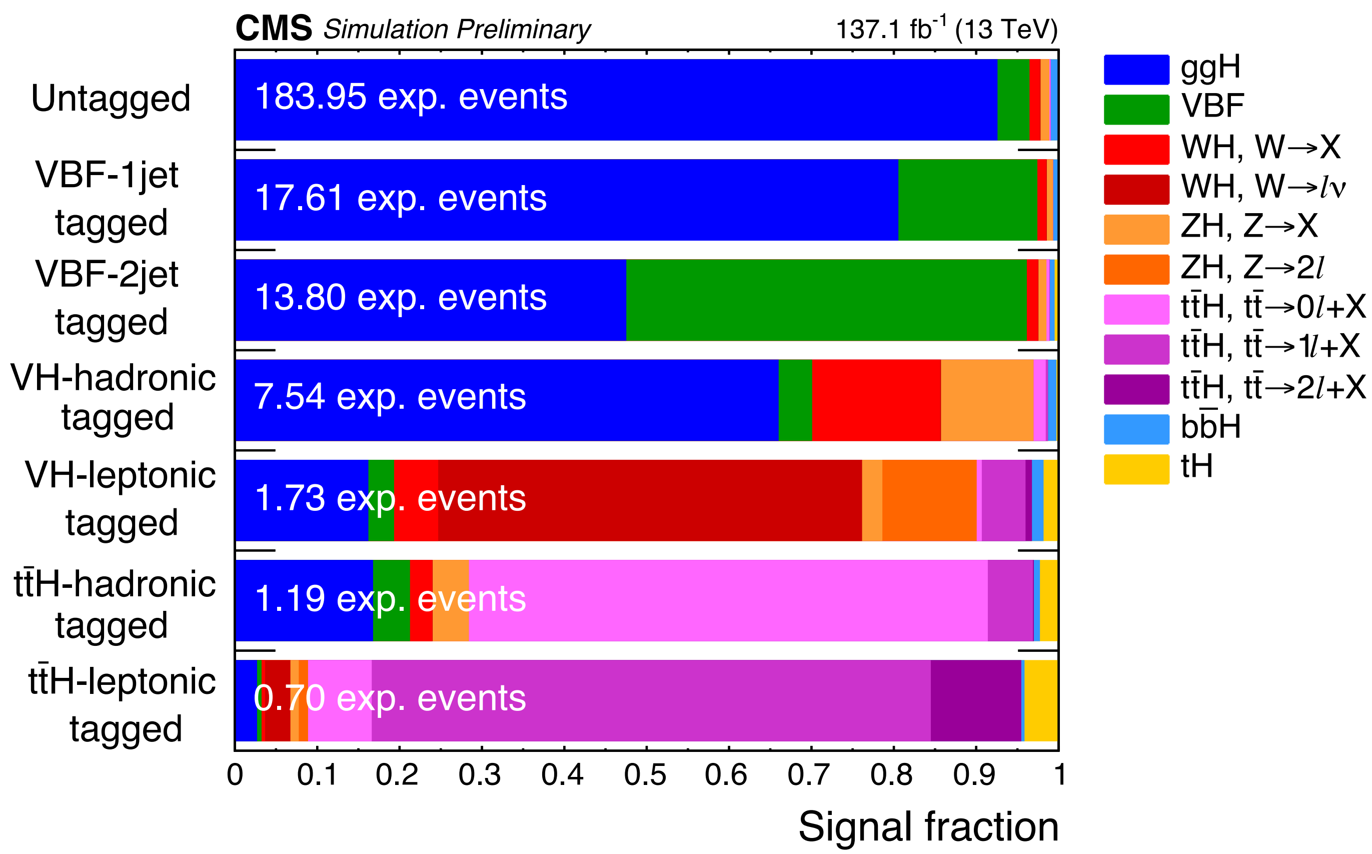
- Dedicated categories as a part of the H→ZZ analysis
- Identify hadronic and leptonic tt decays by requiring at least 4 jets, and b-tagging

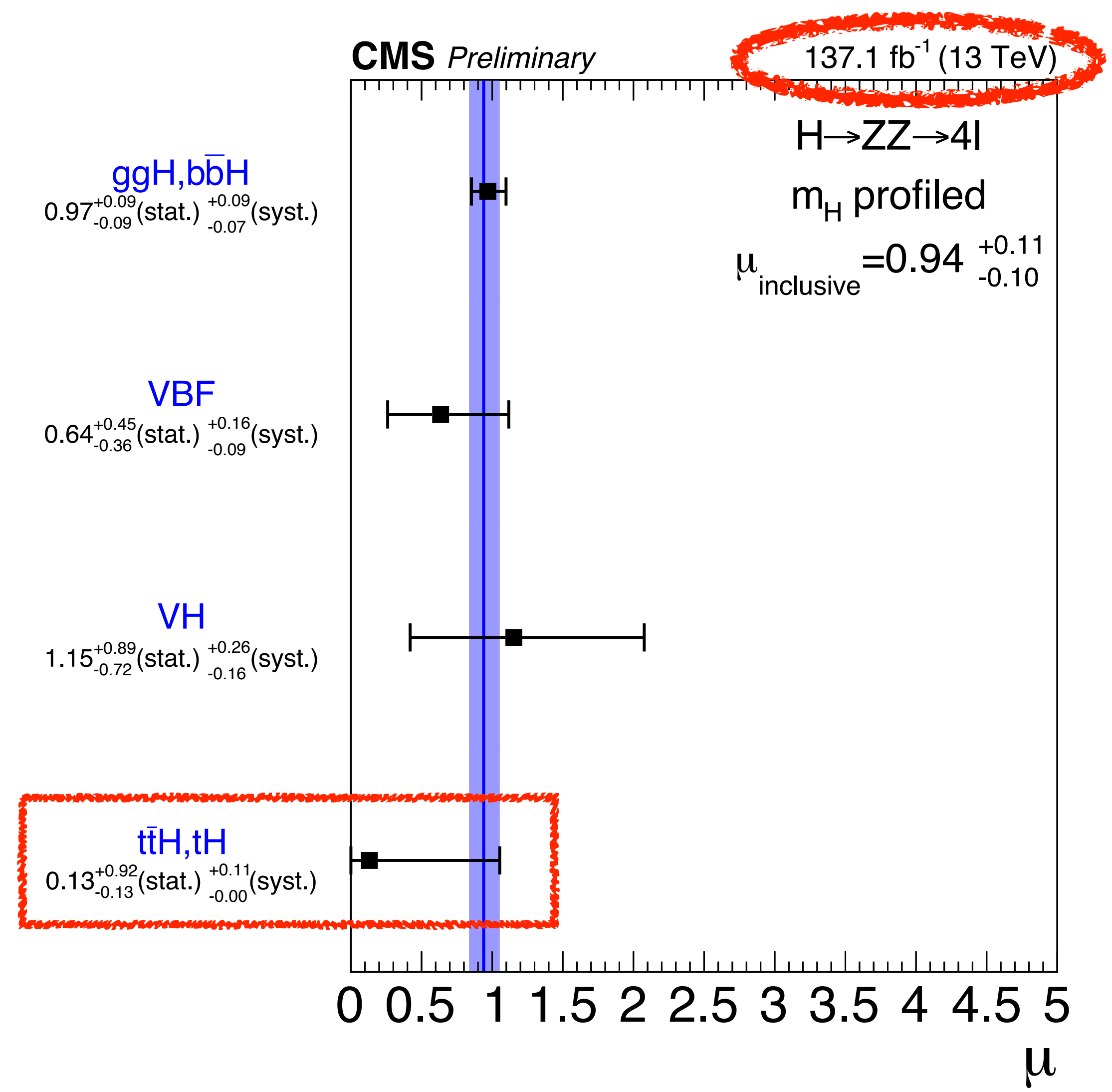


ttH in the 4ℓ final state

CMS PAS HIG-19-001

- Dedicated categories as a part of the H→ZZ analysis
- Identify hadronic and leptonic tt decays by requiring at least 4 jets, and b-tagging

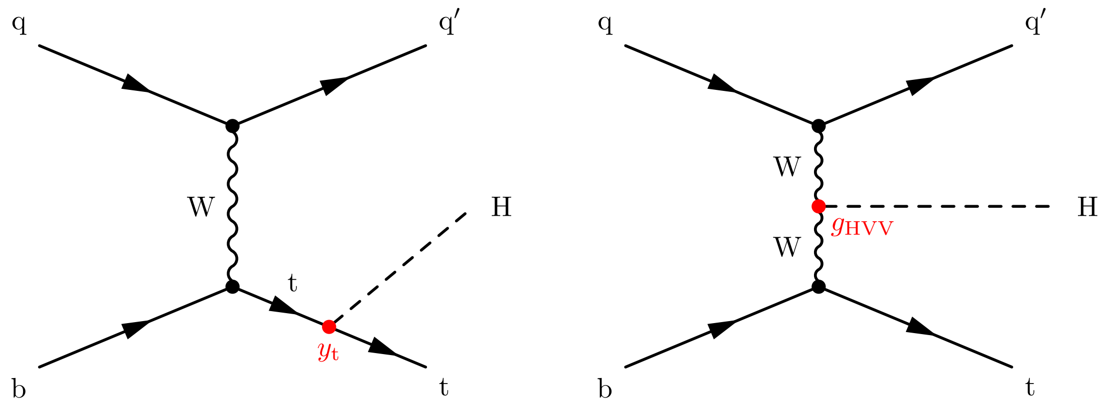




- **Results compatible with the SM expectation**

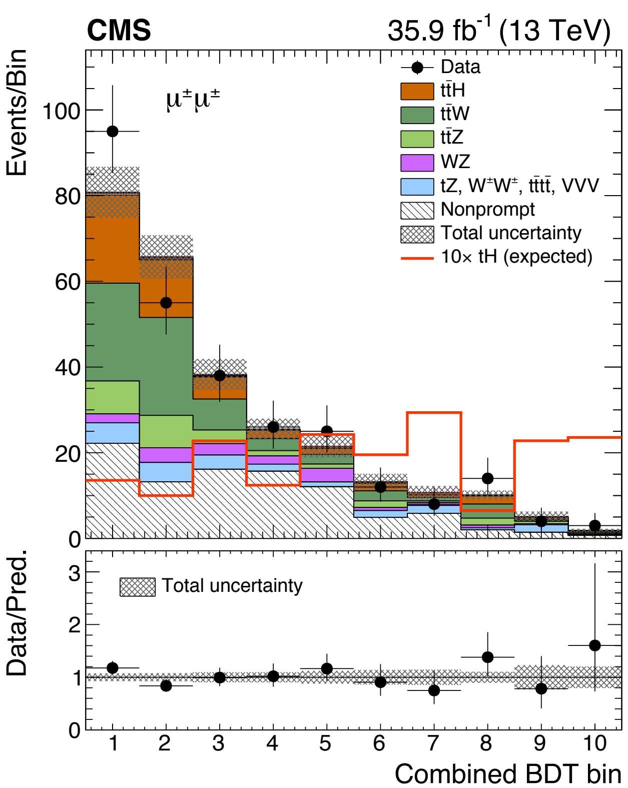
Single top + Higgs

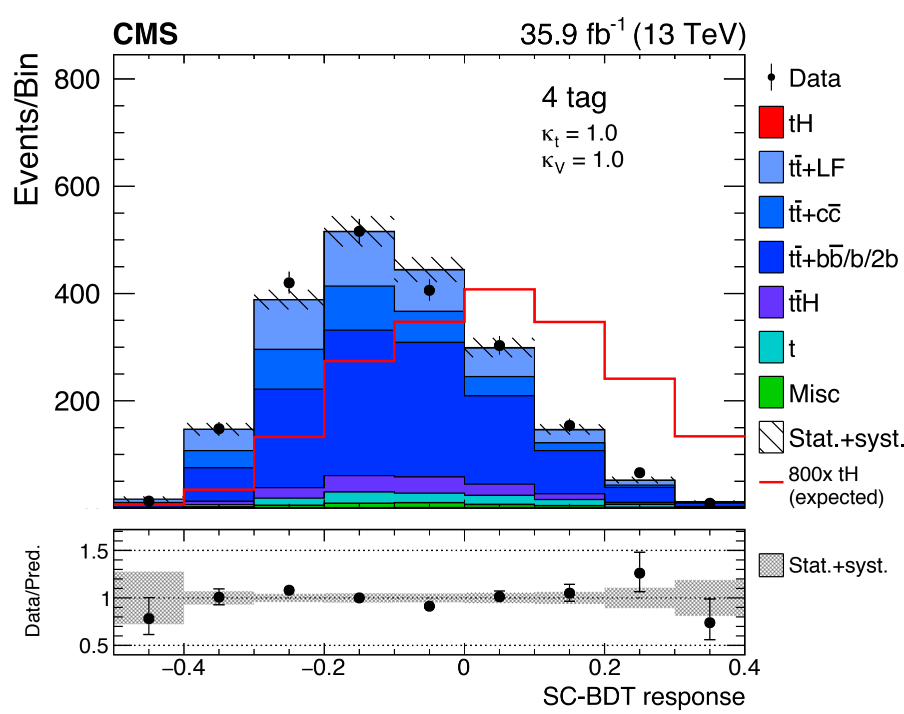
- **tHq production**: sensitive to the **sign of the top-Higgs coupling via interference**



Factor of 10 larger cross section if (κ_T,κ_V) = (-1,1) w.r.t. SM

- Peculiar kinematic features: forward activity, lower jet multiplicity than ttH





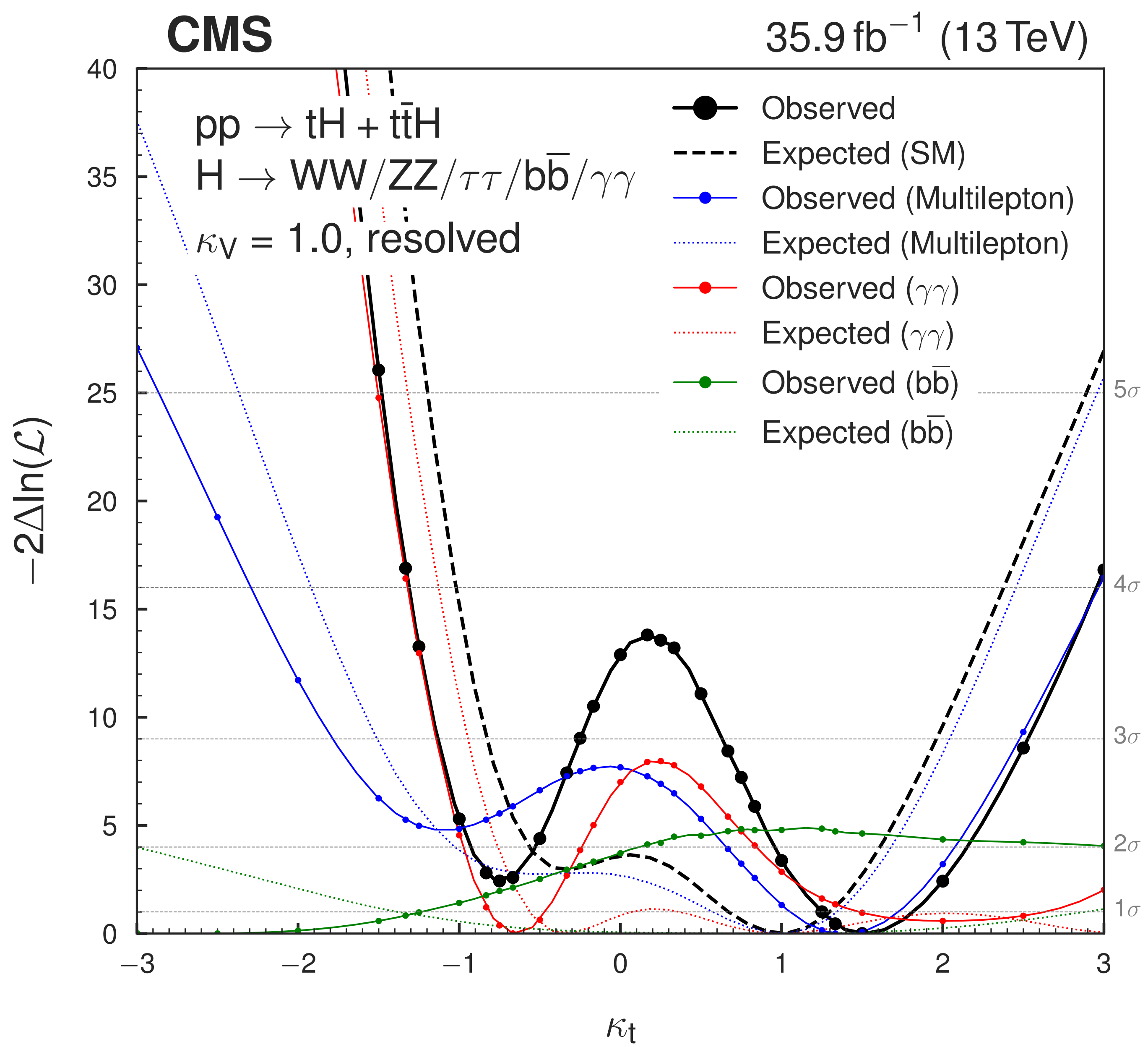
Analysis strategies adapted from ttH leptonic and bb

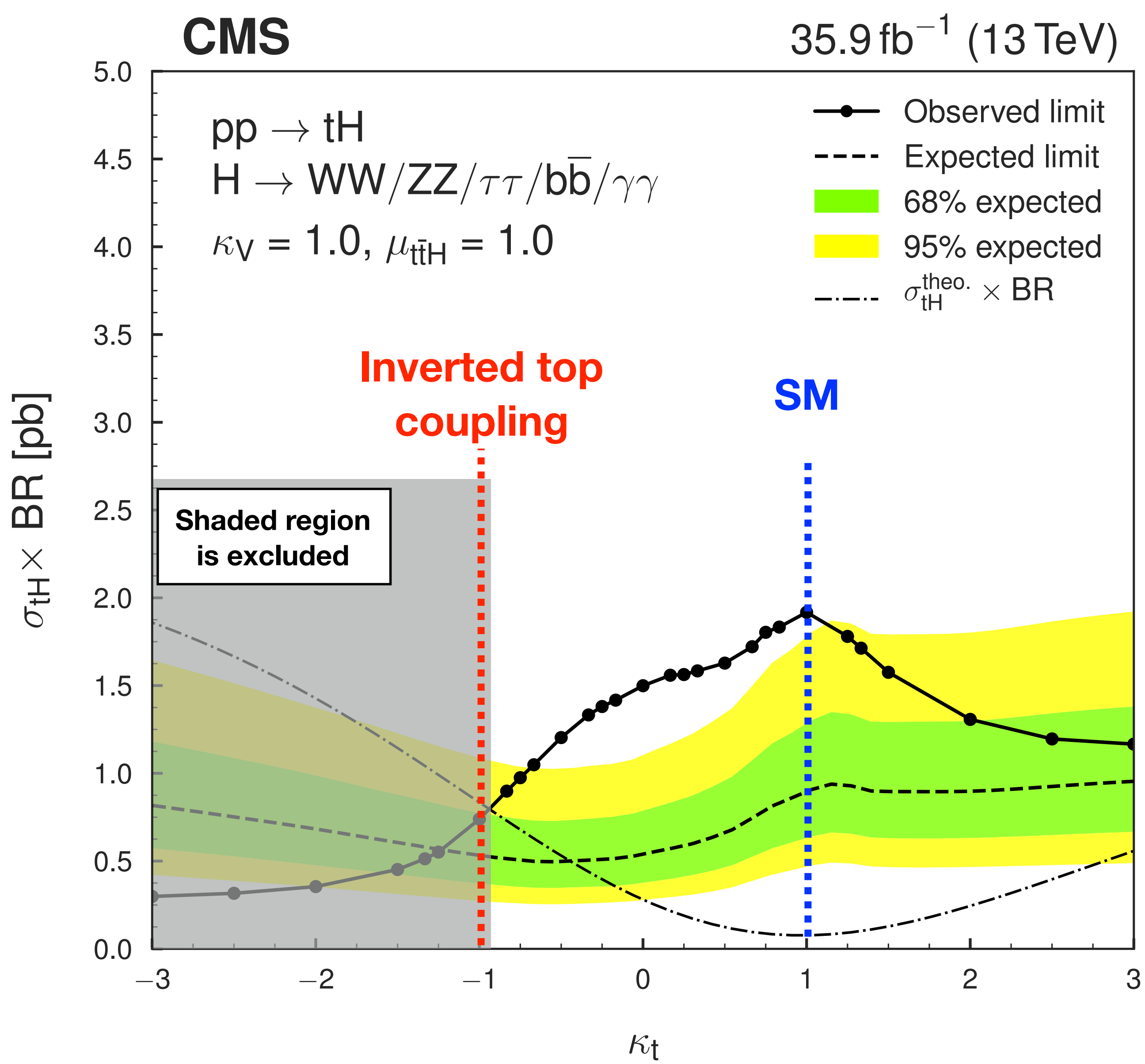
Reinterpretation of H→γγ tagged categories in terms of a tH signal model

Phys. Rev. D 99, 092005 (2019)

Single top + Higgs

Phys. Rev. D 99, 092005 (2019)





- For $\kappa_V = 1$, κ_T ranges outside [–0.9, –0.5] and [1.0, 2.1] are excluded at 95% CL
- Observed (expected) SM-like tH production exclusion at 25 (12) times the SM

Outlook

- CMS has studied the associated production of a Higgs boson with top quarks in all main decay channels, using sophisticated analysis methods
- After the observation of ttH production from their combination, we have now separate **evidence in all main channels** with a partial Run 2 dataset

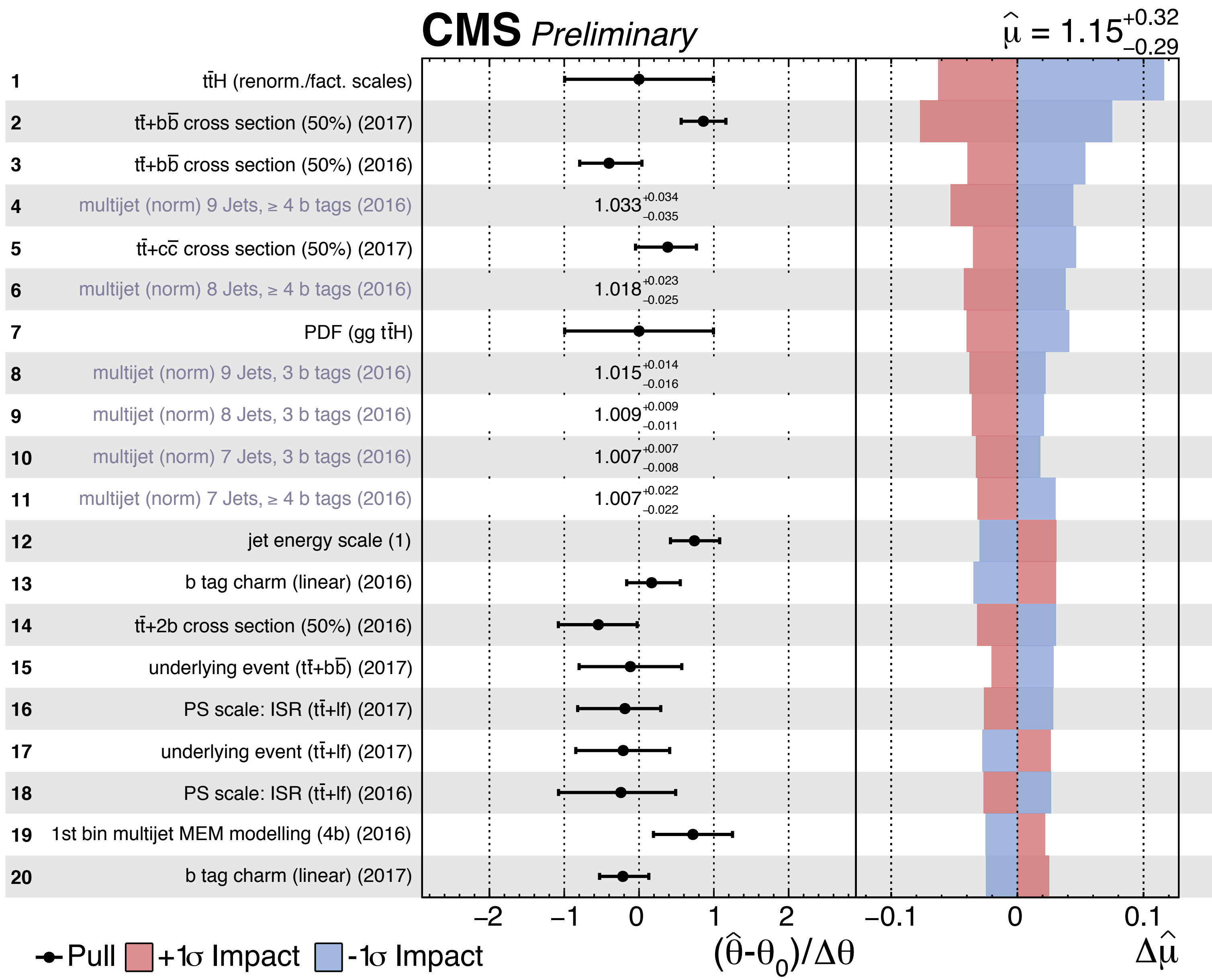
Channel	$\mathscr{L}$	μ	Obs. (exp.) significance
bb	77.4 fb^{-1}	$1.15^{+0.32}_{-0.29}$	3.9σ (3.5σ)
WW* / ZZ* / $\tau\tau$		$0.96^{+0.34}_{-0.31}$	3.2σ (4.0σ)
γγ		$1.7^{+0.6}_{-0.5}$	4.1σ (2.7σ)
4ℓ	137 fb^{-1}	$0.13^{+0.93}_{-0.13}$	-

- All results are **compatible with the expectation for SM Higgs boson couplings**
- As we move forward to the exploration of the full Run 2 dataset, and beyond, we will focus on an even more precise measurement of the coupling (see E.Fontanesi's talk tomorrow), and differential measurements of ttH production

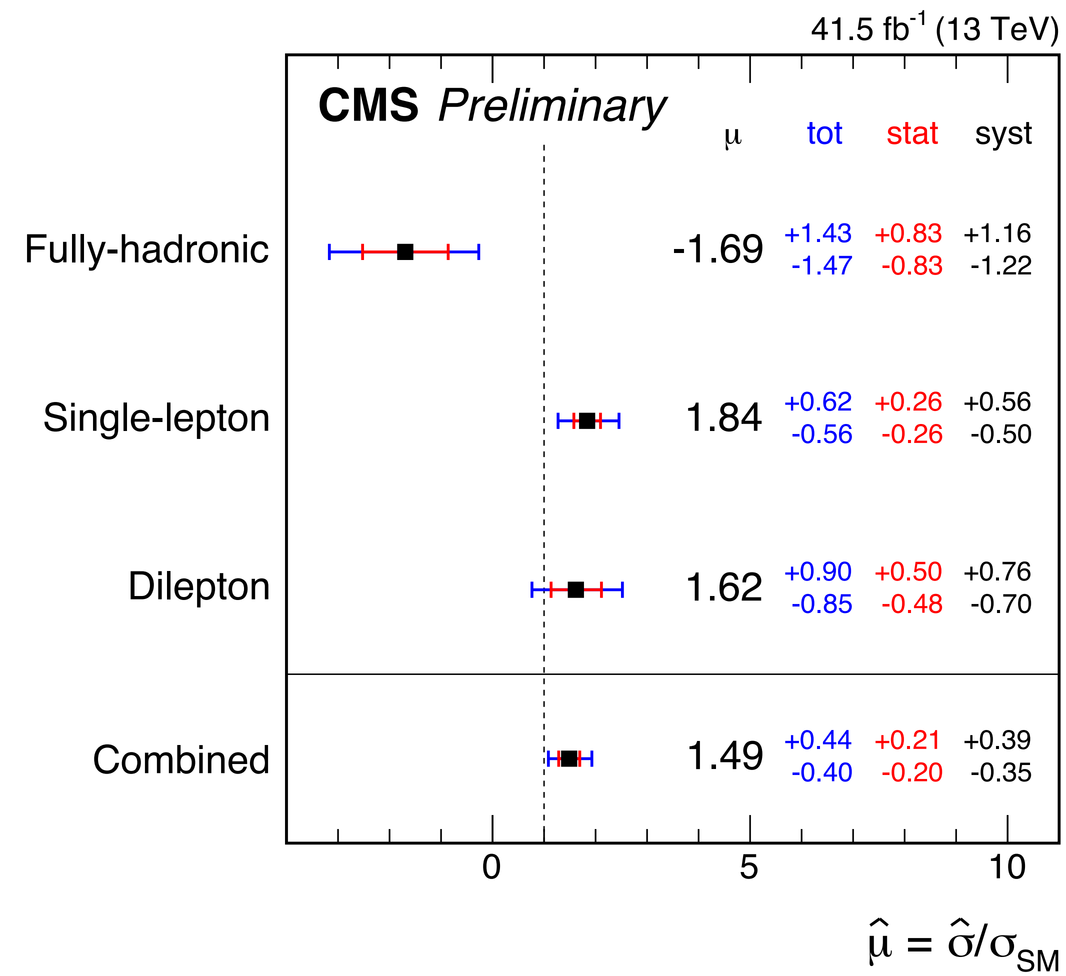
Additional material

ttH, H→bb

CMS PAS HIG-18-030



Impact of systematic uncertainty sources for the 2016+2017 combination



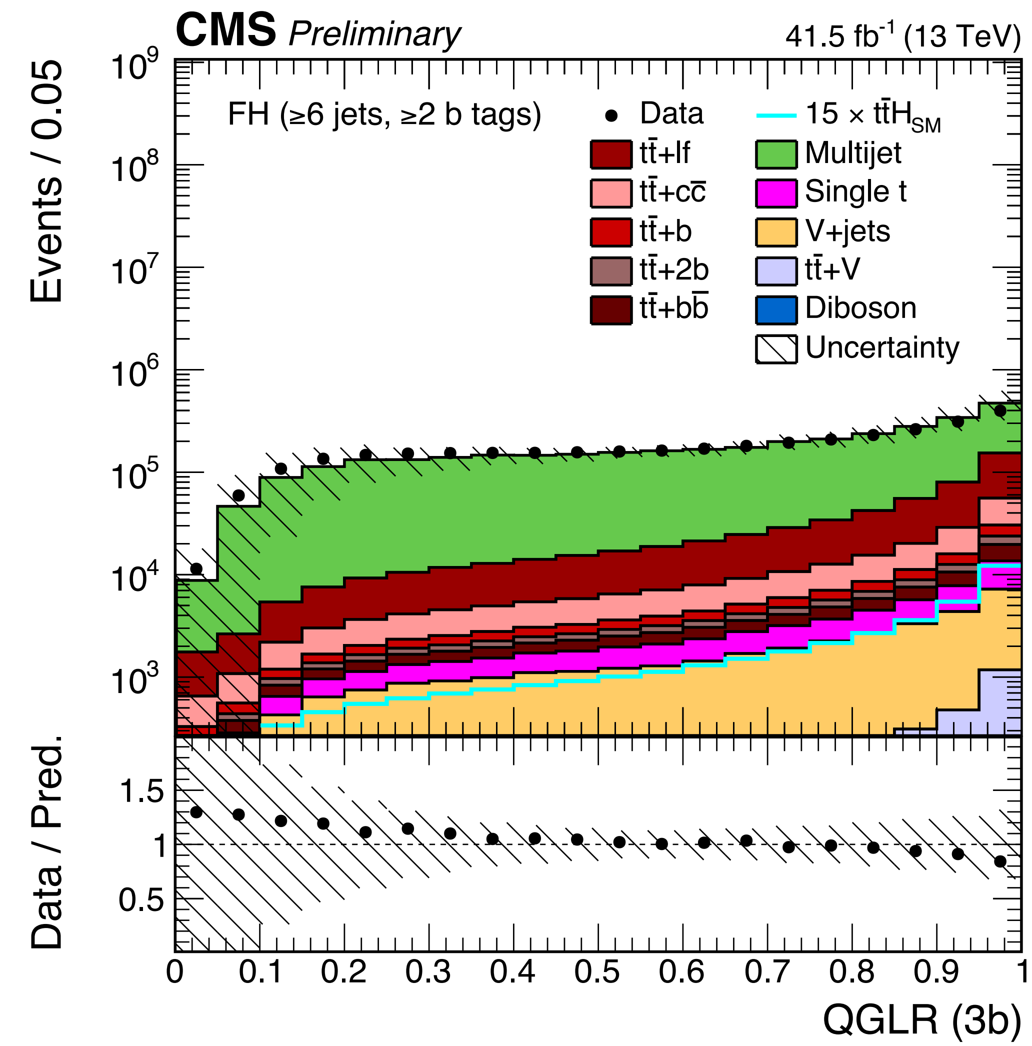
Fit results in the 2017-only analysis

ttH, H→bb

CMS PAS HIG-18-030

	FH channel	SL channel	DL channel
Number of leptons	0	1	2
p_{T} of leptons (e/μ) [GeV]	—	$> 30/29$	$> 25/25$ GeV
p_{T} of additional leptons [GeV]	< 15	< 15	< 15
$\|\eta\|$ of leptons	< 2.4	< 2.4	< 2.4
Number of jets	≥ 6	≥ 4	≥ 2
p_{T} of jets [GeV]	> 40	> 30	$> 30, 30, 20$
$\|\eta\|$ of jets	< 2.4	< 2.4	< 2.4
Number of b-tagged jets	≥ 2	≥ 2	≥ 1
$p_{\mathrm{T}}^{\mathrm{miss}}$	—	> 20 GeV	> 40 GeV

Baseline event selection requirements



	$N_{\mathrm{b\ tag}} = 2$ $N_{\mathrm{b\ tag\ loose}} \geq 3$	$N_{\mathrm{b\ tag}} \geq 3$
QGLR > 0.5	CR (to extract distribution)	SR (final analysis)
QGLR < 0.5	Validation CR (to validate distribution)	VR (comparison with data)

Signal, control and validation region definitions used in the 0-lepton analysis

$$\mathrm{QGLR} = \frac{\mathcal{L}_{\mathrm{qqqq}}}{\mathcal{L}_{\mathrm{qqqq}} + \mathcal{L}_{\mathrm{gggg}}}$$

ttH leptonic

Event selection requirements

CMS PAS HIG-18-019

Selection	$2\ell ss$	$2\ell ss + 1\tau_h$
Targeted $t\bar{t}H$ decays	$t \to b\ell\nu$, $t \to bqq$, $H \to WW \to \ell\nu qq$	$t \to b\ell\nu$, $t \to bqq$, $H \to \tau\tau \to \ell\tau_h + \nu$'s
Trigger	Single- or double-lepton triggers	
Lepton multiplicity	Exactly 2 leptons	
Lepton p_T	$p_T > 25$ / $15\,GeV$	$p_T > 25$ / 15 (e) or $10\,GeV$ (μ)
Lepton η	$\lvert\eta\rvert < 2.5$ (e) or 2.4 (μ)	
τ_h multiplicity	No τ_h (loose WP)	$\geq 1\ \tau_h$ (loose WP) and $< 2\tau_h$ (medium WP)
τ_h p_T	—	$p_T > 20\,GeV$
τ_h η	—	$\lvert\eta\rvert < 2.3$
Charge requirements	2 same-sign leptons and charge quality requirements	$\sum_{\ell,\tau_h} q = \pm 1$
Jet multiplicity	≥ 4 jets	≥ 3 jets
b tagging requirements	≥ 1 tight b-tagged jet or ≥ 2 loose b-tagged jets	
Missing transverse momentum	$L_D > 30\,GeV^{**}$	
Dilepton mass	$m_{\ell\ell} > 12\,GeV^{*}$ and $\lvert m_{ee} - m_Z\rvert > 10\,GeV^{**}$	

Selection	$1\ell + 2\tau_h$	$2\ell + 2\tau_h$
Targeted $t\bar{t}H$ decays	$t \to b\ell\nu$, $t \to bqq$, $H \to \tau\tau \to \tau_h\tau_h + \nu$'s	$t \to b\ell\nu$, $t \to b\ell\nu$, $H \to \tau\tau \to \tau_h\tau_h + \nu$'s
Trigger	Single-lepton or lepton+τ_h triggers	Single-, double-lepton triggers
Lepton multiplicity	Exactly 1 lepton	≥ 2 leptons
Lepton p_T	$p_T > 25$ (e) or $20\,GeV$ (μ)	$p_T > 25$ / 15 (e) or $10\,GeV$ (μ)
Lepton η	$\lvert\eta\rvert < 2.1$	$\lvert\eta\rvert < 2.5$ (e) or 2.4 (μ)
τ_h multiplicity	$\geq 2\ \tau_h$ (medium WP)	
τ_h p_T	$p_T > 30$ / $20\,GeV$	$p_T > 20\,GeV$
τ_h η	$\lvert\eta\rvert < 2.3$	$\lvert\eta\rvert < 2.3$
Charge requirements	$\sum_{\tau_h} q = 0$	$\sum_{\ell,\tau_h} q = 0$
Jet multiplicity	≥ 3 jets	≥ 2 jets
b tagging requirements	≥ 1 tight b-tagged jet or ≥ 2 loose b-tagged jets	
Missing transverse momentum	— — —	No requirement if $N_j \geq 4$ $L_D > 45\,GeV^{\dagger}$ $L_D > 30\,GeV$ otherwise
Dilepton mass	$m_{\ell\ell} > 12\,GeV^{*}$	

* Applied on all pairs of leptons that pass loose selection.

** If both leptons are electrons.

† If the event contains a same-flavor opposite-sign lepton pair and $N_j \leq 3$.

Selection	3ℓ	$3\ell + 1\tau_h$
Targeted $t\bar{t}H$ decays	$t \to b\ell\nu$, $t \to b\ell\nu$, $H \to WW \to \ell\nu qq$ $t \to b\ell\nu$, $t \to bqq$, $H \to WW \to \ell\nu\ell\nu$	$t \to b\ell\nu$, $t \to b\ell\nu$, $H \to \tau\tau \to \ell\tau_h + \nu'$
Trigger	Single-, double- or triple-lepton triggers	
Lepton multiplicity	Exactly 3 leptons	
Lepton p_T	$p_T > 25$ / 15 / $15\,GeV$	$p_T > 20$ / 10 / $10\,Ge$
Lepton η	$\lvert\eta\rvert < 2.5$ (e) or 2.4 (μ)	
τ_h	No τ_h (loose WP)	$\geq 1\ \tau_h$ (loose WP)
τ_h p_T	—	$p_T > 20\,GeV$
τ_h η	—	$\lvert\eta\rvert < 2.3$
Charge requirements	$\sum_{\ell} q = \pm 1$	$\sum_{\ell,\tau_h} q = 0$
Jet multiplicity	≥ 2 jets	
b tagging requirements	≥ 1 tight b-tagged jet or ≥ 2 loose b-tagged jets	
Missing transverse momentum	No requirement if $N_j \geq 4$ $L_D > 45\,GeV^{\dagger}$ $L_D > 30\,GeV$ otherwise	
Dilepton mass	$m_{\ell\ell} > 12\,GeV^{*}$ and $\lvert m_{\ell\ell} - m_Z\rvert > 10\,GeV^{\ddagger}$	
Four-lepton mass	$m_{4\ell} > 140\,GeV^{\S}$	—

Selection	4ℓ
Targeted $t\bar{t}H$ decays	$t \to b\ell\nu$, $t \to b\ell\nu$, $H \to WW \to \ell\nu\ell\nu$ $t \to b\ell\nu$, $t \to b\ell\nu$, $H \to ZZ \to \ell\ell qq$ or $\ell\ell\nu\nu$
Trigger	Single-, double- or triple-lepton triggers
Lepton multiplicity	≥ 4 leptons
Lepton p_T	$p_T > 25$ / 15 / 15 / $10\,GeV$
Lepton η	$\lvert\eta\rvert < 2.5$ (e) or 2.4 (μ)
Charge requirements	—
Jet multiplicity	≥ 2 jets
b tagging requirements	≥ 1 tight b-tagged jet or ≥ 2 loose b-tagged jets
Missing transverse momentum	—
Dilepton mass	$m_{\ell\ell} > 12\,GeV^{*}$ and $\lvert m_{\ell\ell} - m_Z\rvert > 10\,GeV^{\ddagger}$
Four-lepton mass	$m_{4\ell} > 140\,GeV^{\S}$

* Applied on all pairs of leptons that pass loose selection.

† If the event contains a same-flavor opposite-sign (SFOS) lepton pair and $N_j \leq 3$.

‡ Applied to all SFOS lepton pairs.

§ Applied only if the event contains 2 SFOS lepton pairs.

ttH leptonic

CMS PAS HIG-18-019

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 $\|\eta\|$ 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						
$\mathcal{D}_{\mathrm{thad}}^{\max}$					X			
$\mathcal{D}_{\mathrm{Hj}}^{\max}$						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

Input variables of kinematic discriminants used for signal extraction

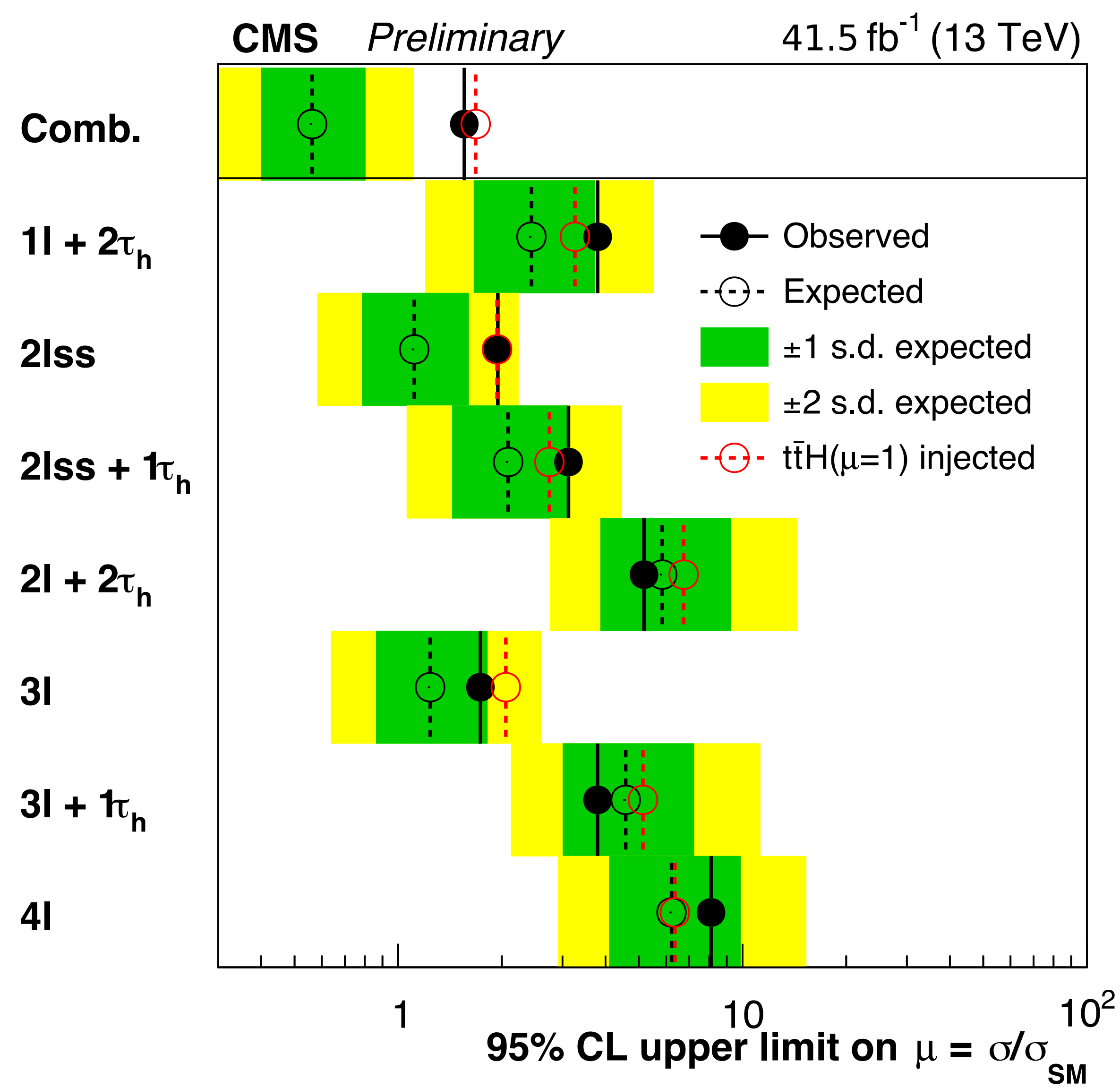
ttH leptonic

CMS PAS HIG-18-019

Category	Signal Strength $\pm 1\sigma$ Measured	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	$0.96^{+0.34}_{-0.31}$	$1.00^{+0.30}_{-0.27}$

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 [48]	10	5	Correlated
Jet energy calibration	2–15 [56]	3	3	Correlated
Fake background yield	$\approx$ 30–50	17	9	Un-correlated

	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



Additional results and summary of systematic uncertainties