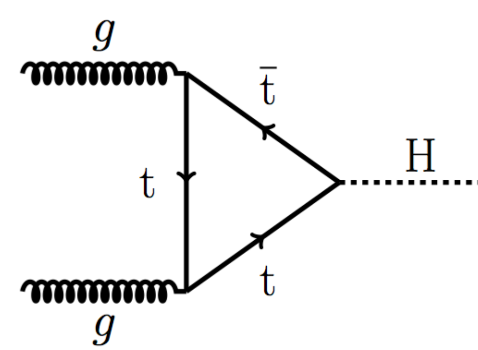
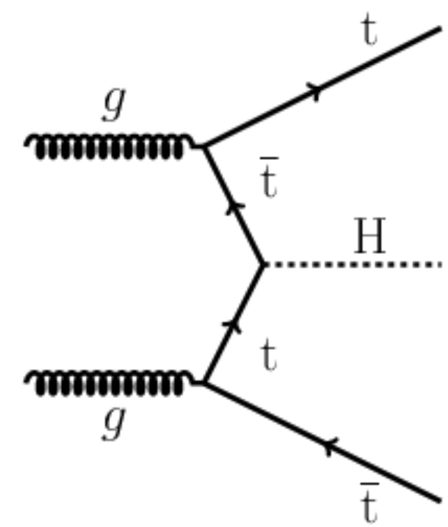
Searches for ttH production at CMS

Arun Nayak
Institute of Physics
Bhubaneswar, India
(For CMS collaboration)

DIS – 2018
16th – 20th April, Kobe, Japan

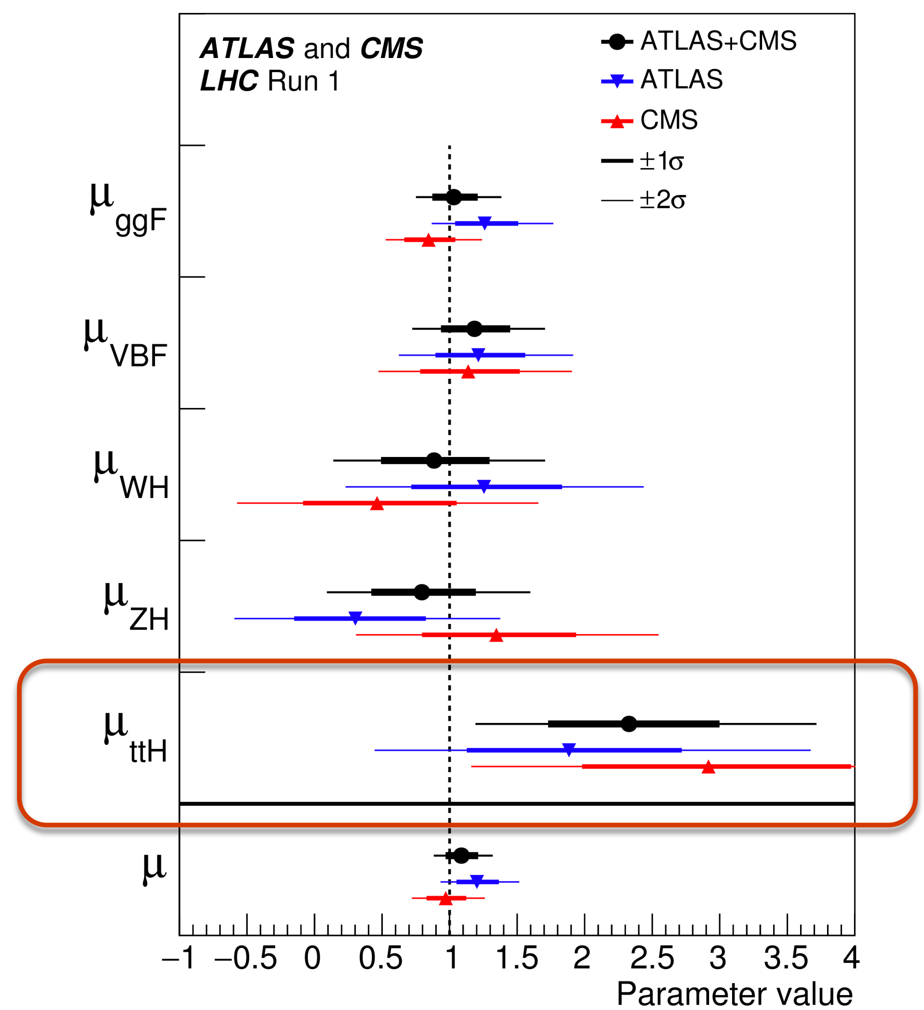
Introduction





Probing Higgs to Top coupling (y_t):

- Production via gluon fusion
  - assumes no BSM coupling
- **Associated production with top-quark pair**
  - Direct measurement
  - Larger increase in signal than backgrounds from 8 to 13 TeV



Challenge:

- Large backgrounds: tt+jets, tt+bb, ttW, ttZ etc..
  - σ(ttH) ~ 510 fb, σ(ttbar) ~830 pb @13 TeV
- Large combinatorics of leptons and jets from top decay

Analysis Strategies:

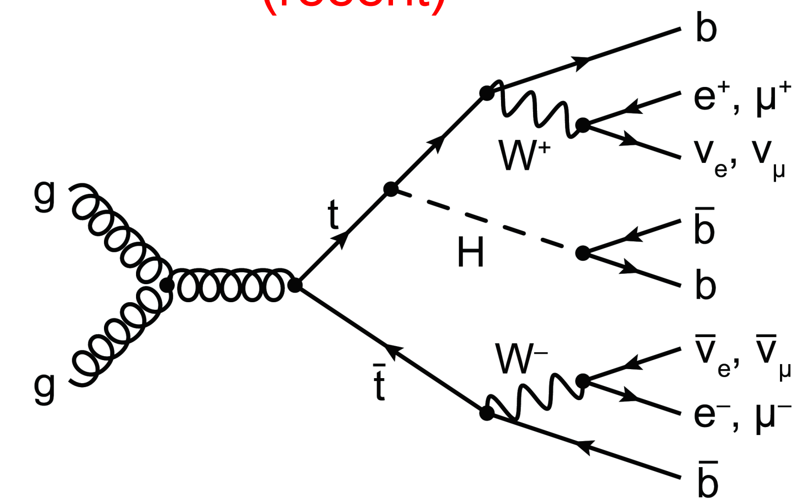
- ttbar like selections with additional searches for Higgs decay products
- Event categorization based on top quark (W boson) and Higgs decay modes
- MVA techniques OR Matrix-Element-Methods used to extract signal

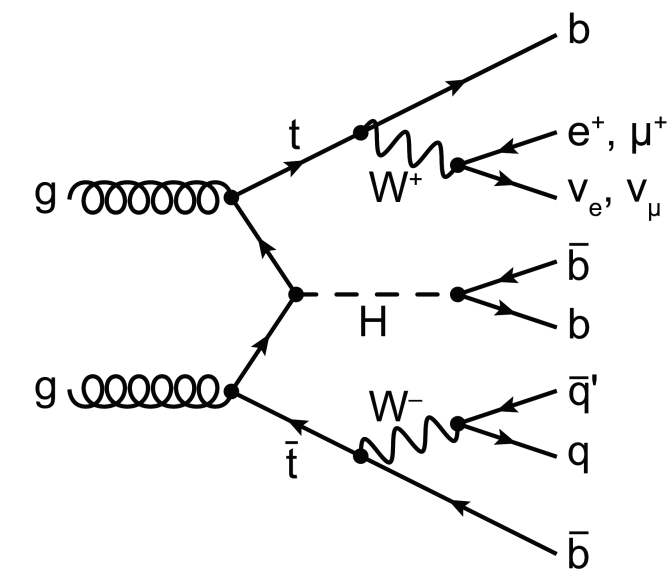
Measurements @ CMS

Variety of final states studied:

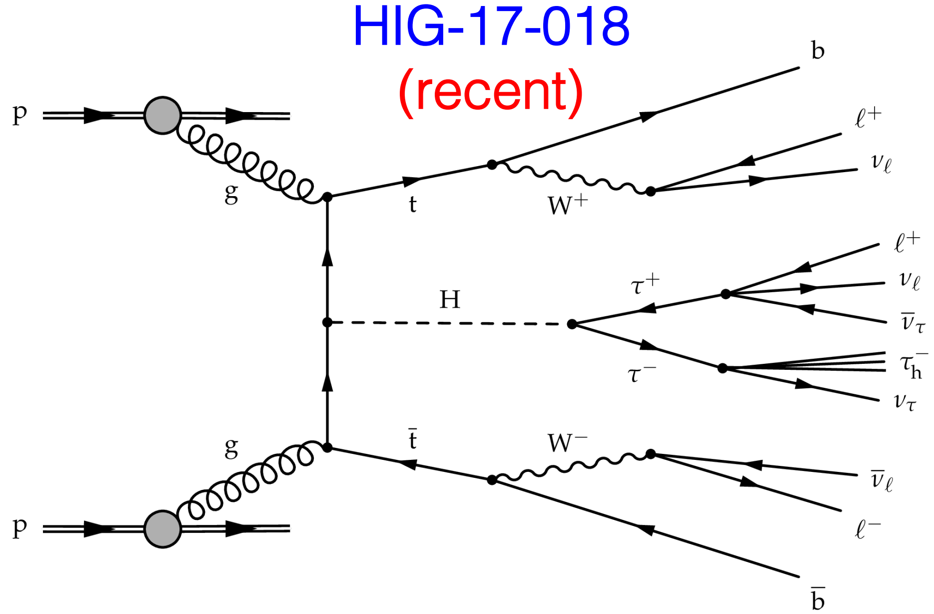
- tt+ bb (H→bb): Large branching fraction, but higher background rates
  - At least one W decaying to leptons (semilepton, dilepton)
  - Both W decaying to jets (fully hadronic)
- tt+leptons (H→WW*, ZZ*, ττ): smaller production rate, relative lower backgrounds
- tt+γγ (H→γγ): clean final state, but very small rate

HIG-17-026
(recent)

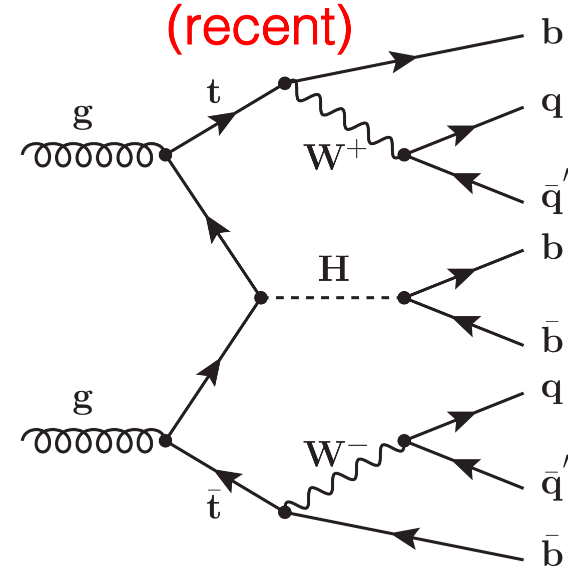




HIG-17-018
(recent)



HIG-17-022
(recent)

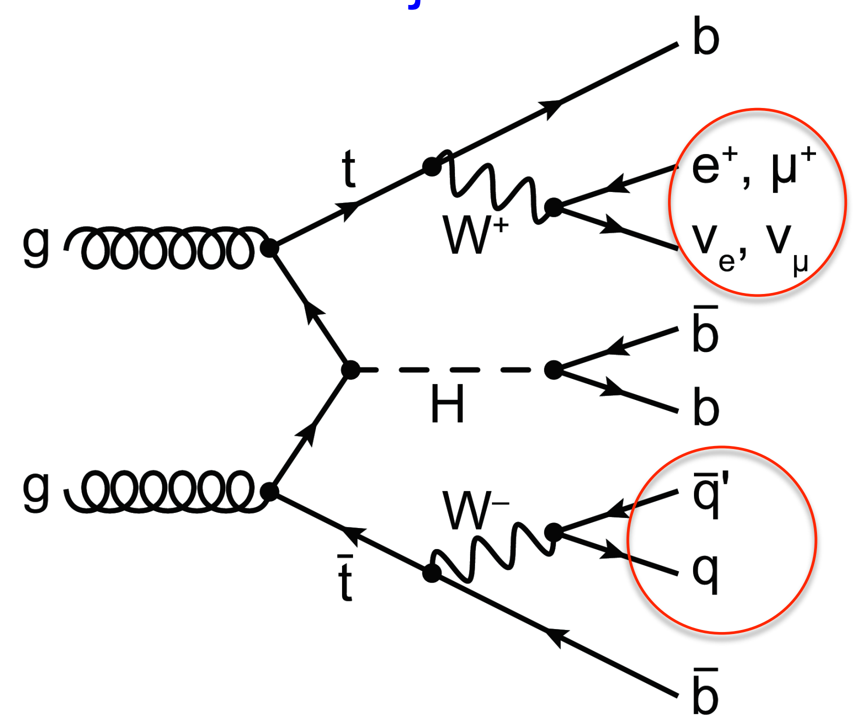


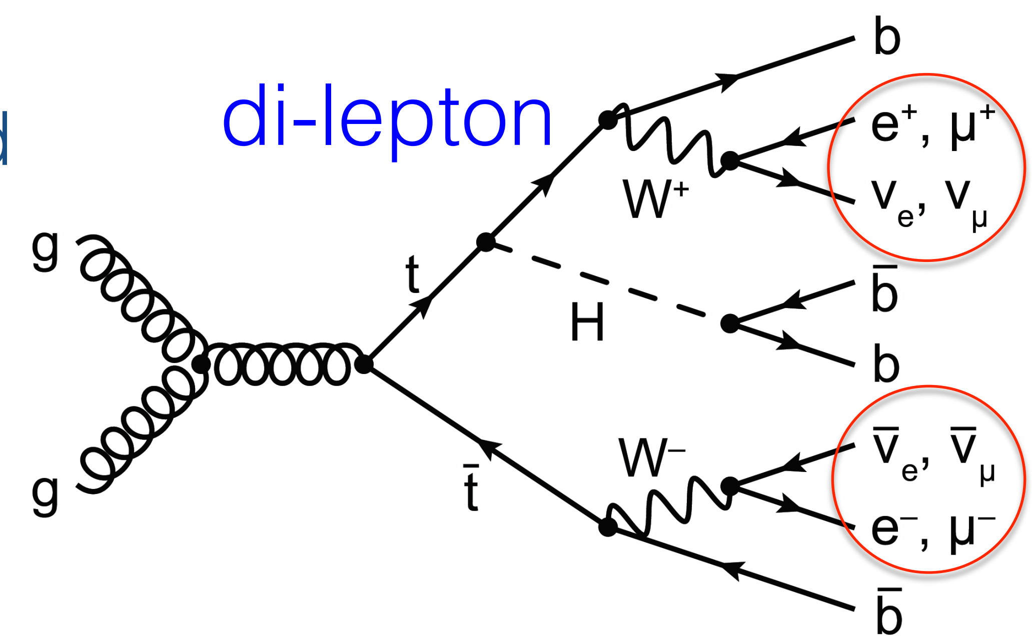
ttH, H → bb (leptonic W decay)

HIG-17-026

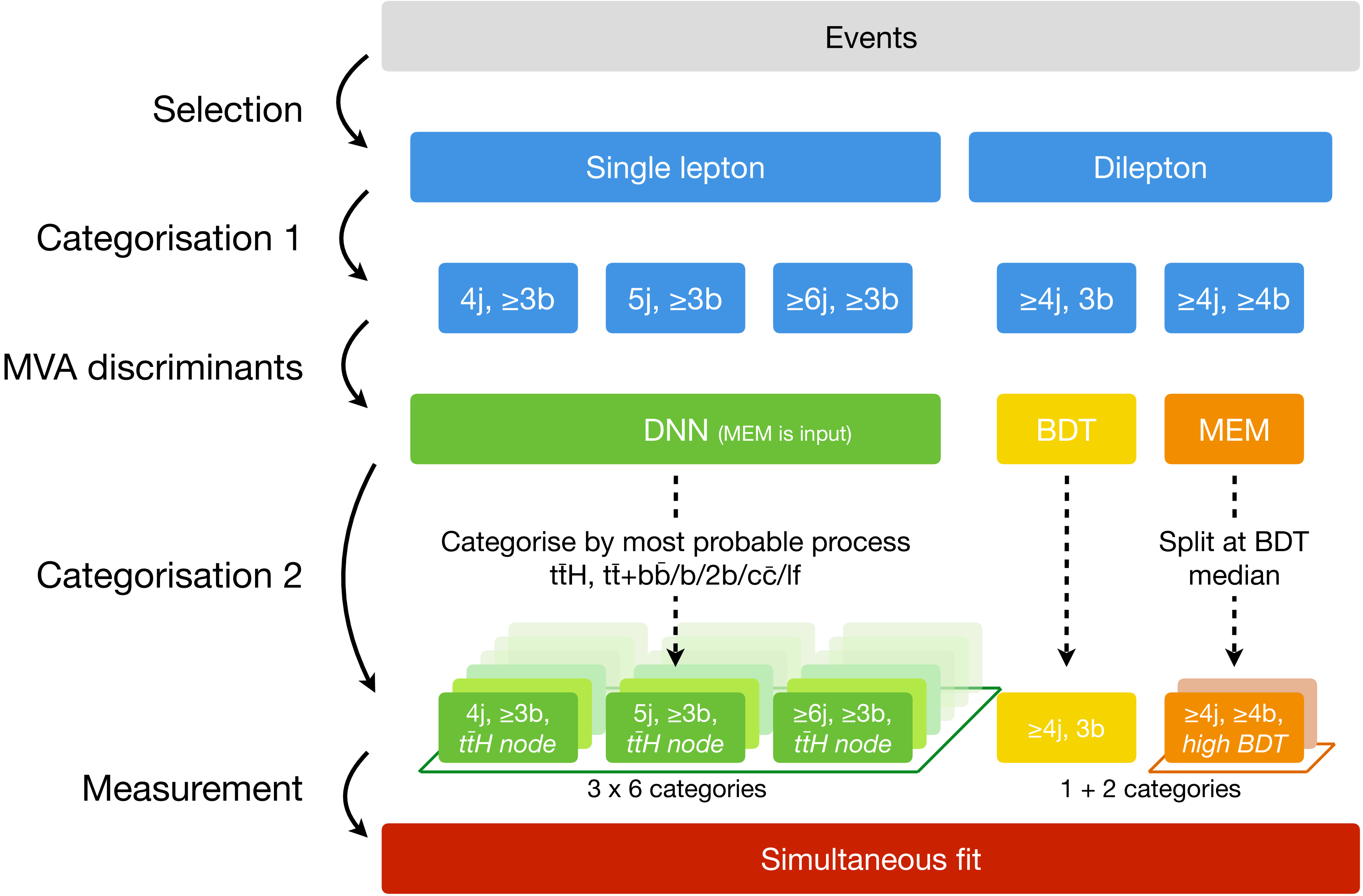
- Analysis channels:
  - ℓ+jets: tt→$\ell\nu$qq'bb, H→bb
  - dilepton: tt→$\ell\nu\ell\nu$bb, H→bb
- At least 4 jets, with at least 3 b-tagged (≥4 jets, ≥3 b-tagged)
- Poor H→bb mass resolution
- Exploiting Matrix-Element methods and Machine learning techniques to discriminate signal from background
- Major backgrounds: tt+jets (especially tt+bb)

ℓ+jets





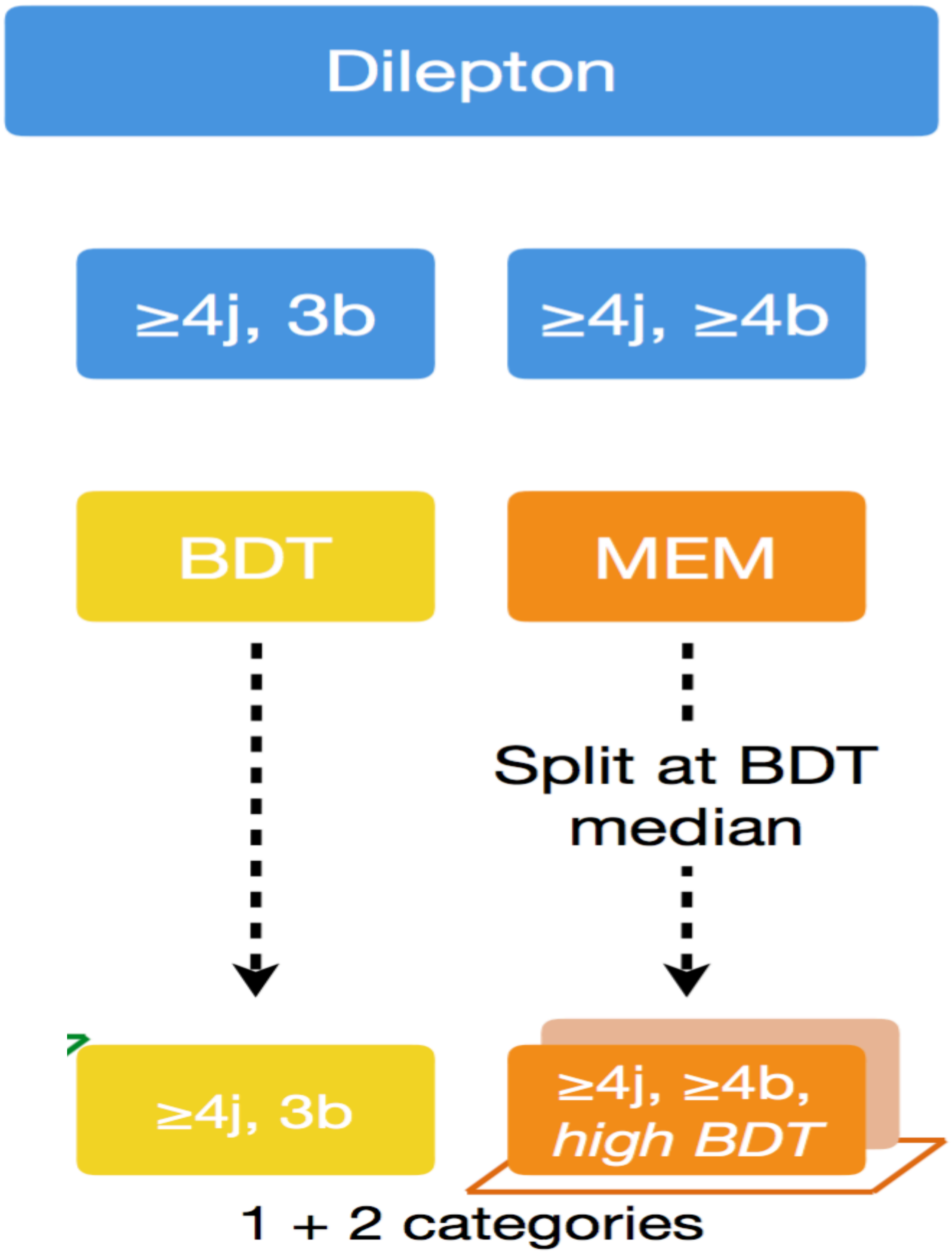
Analysis Strategy



Analysis Strategy

CMS-PAS-HIG-17-026

Dilepton channel: tt→$\ell\nu\ell\nu$bb, H→bb

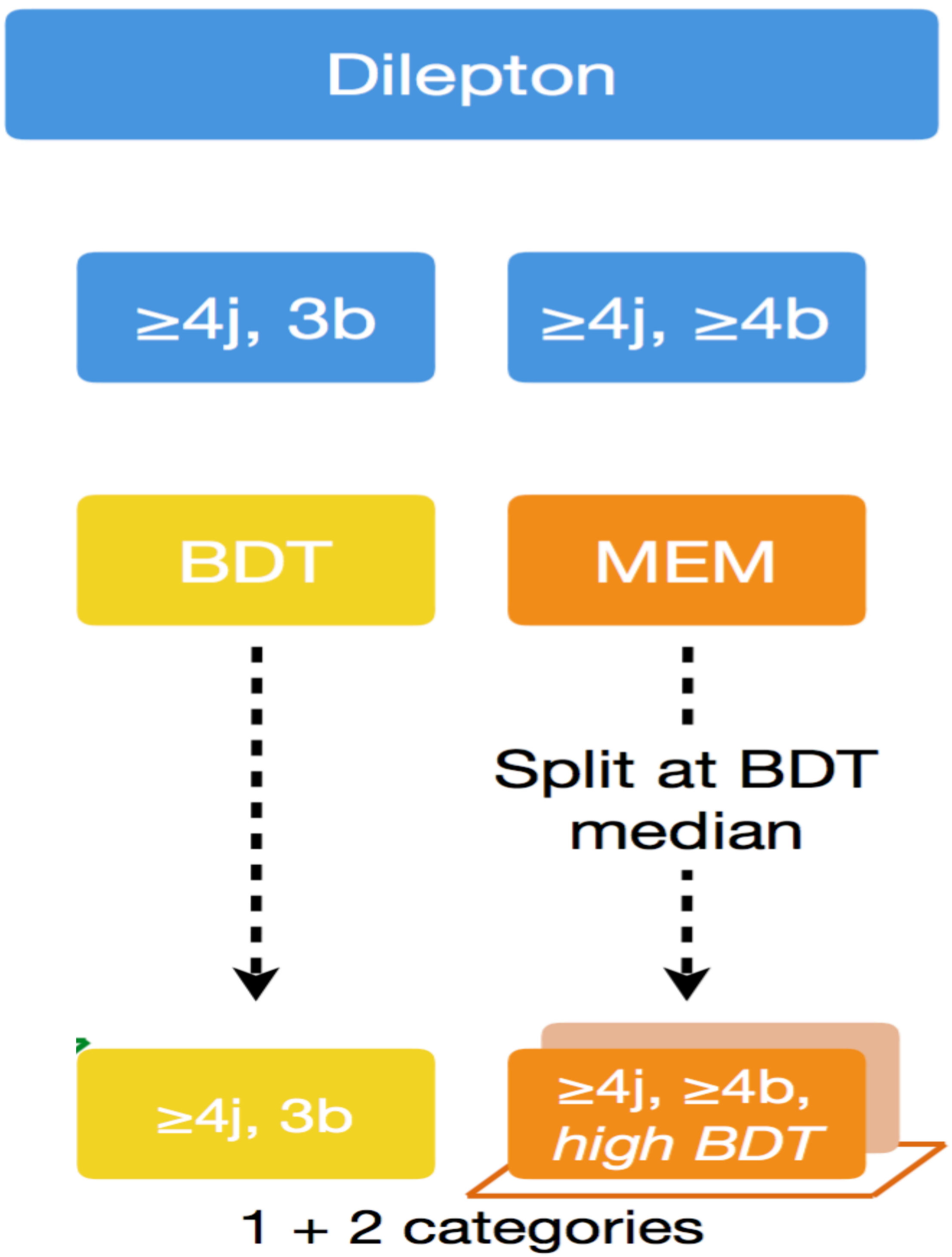


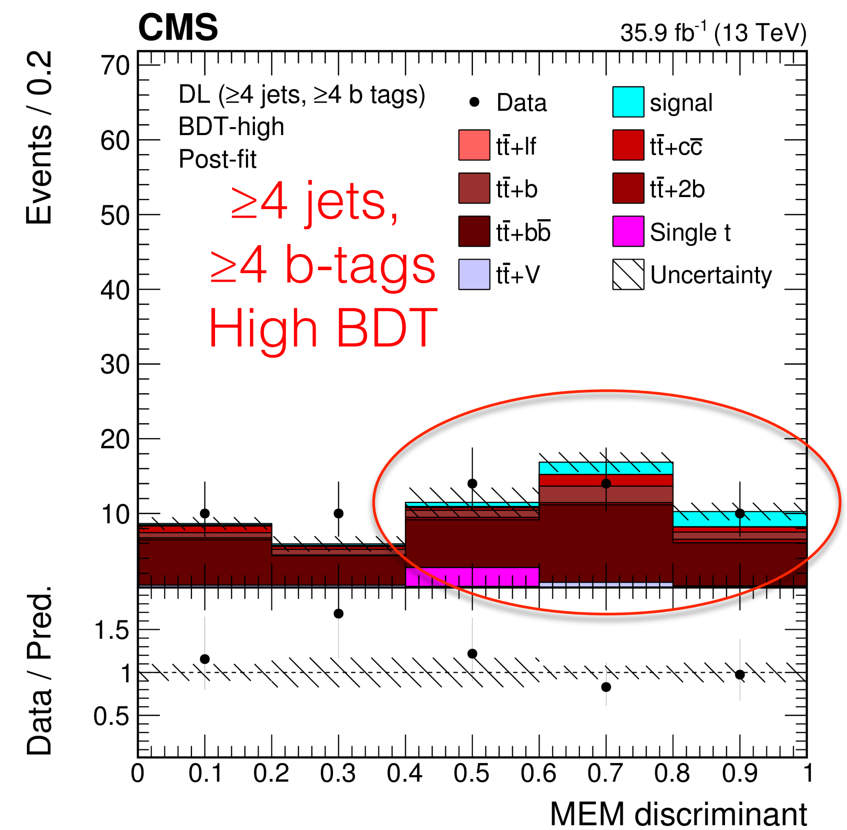
- MVA (BDT) and Matrix-Element methods used to discriminate signal from background
- BDT trained for each of the two (b-tagged) jet categories
  - Inputs: object kinematics, event shapes, b-tag discriminants
- BDT as final discriminant in 3 b-tag category
  - Discrimination against tt+jets
- >=4 b-tags category splitted to two sub-categories based on BDT output
- MEM as final discriminant in each of the two sub-categories
  - Discrimination against tt+bb

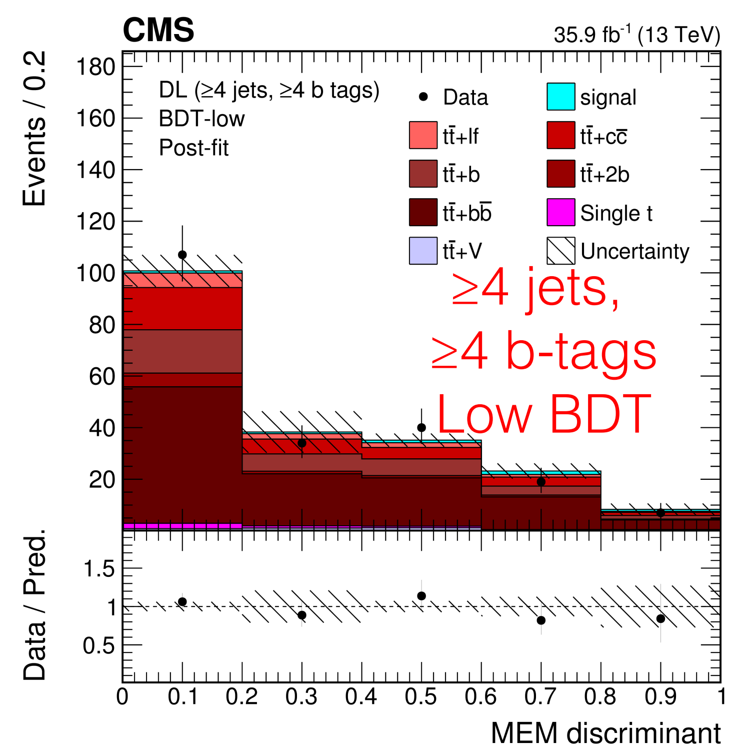
Analysis Strategy

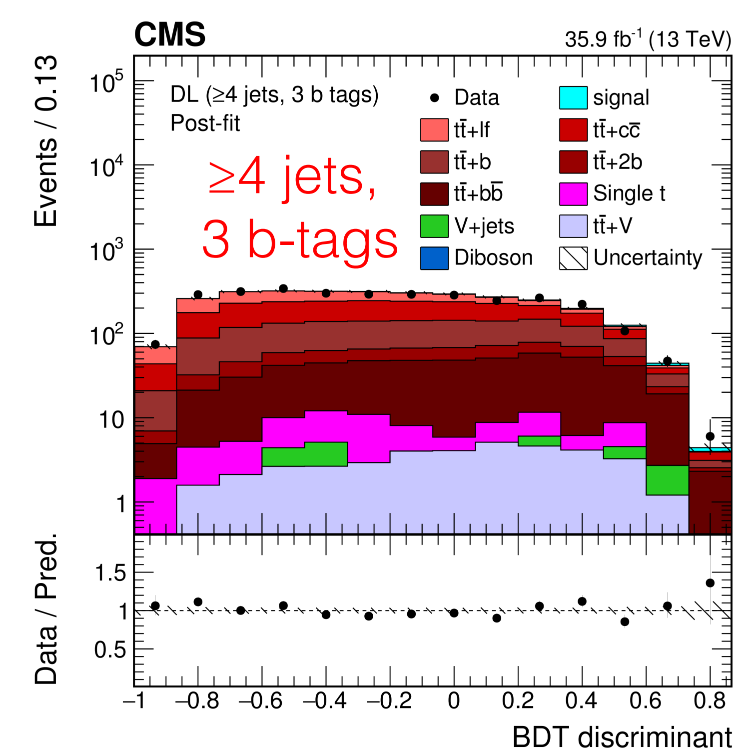
HIG-17-026

Dilepton channel: tt→ℓνℓνbb, H→bb





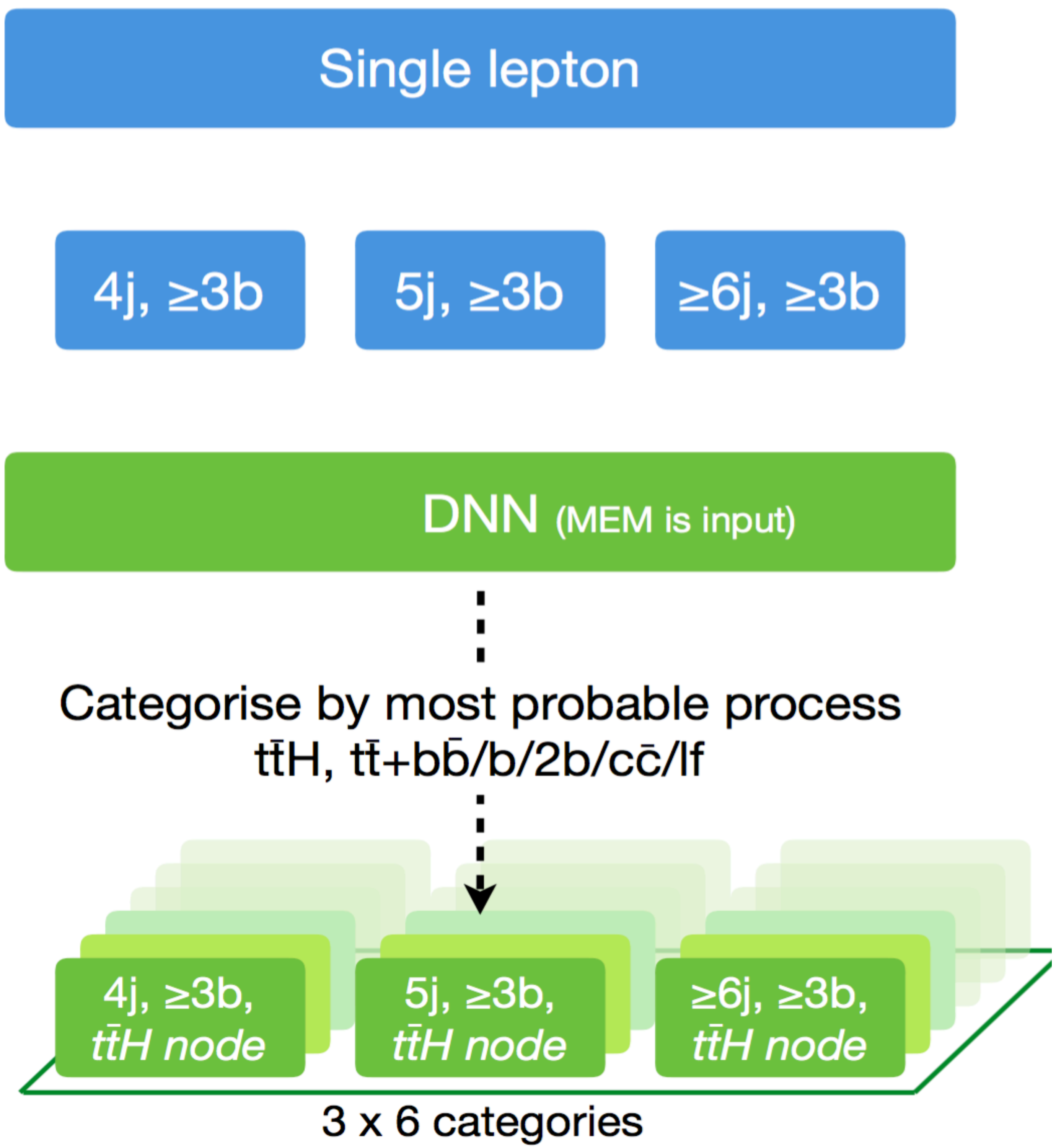




Analysis Strategy

HIG-17-026

ℓ+jets channel: tt→$\ell\nu$qq'bb, H→bb



- Events divided to 3 jet-multiplicity categories
- Multi-classification Deep-Neural Network (DNN) for each jet-multiplicity category
  - Same input variables as BDT + **MEM discriminant**
- DNN associates a set of probabilities describing the probability of the events being either signal ttH or one of the tt+X background like.
- In each jet-process categories, the DNN output distribution of the node that matches the process category is used as final discriminant

Analysis Strategy

ℓ+jets channel: tt→ℓνqq'bb, H→bb

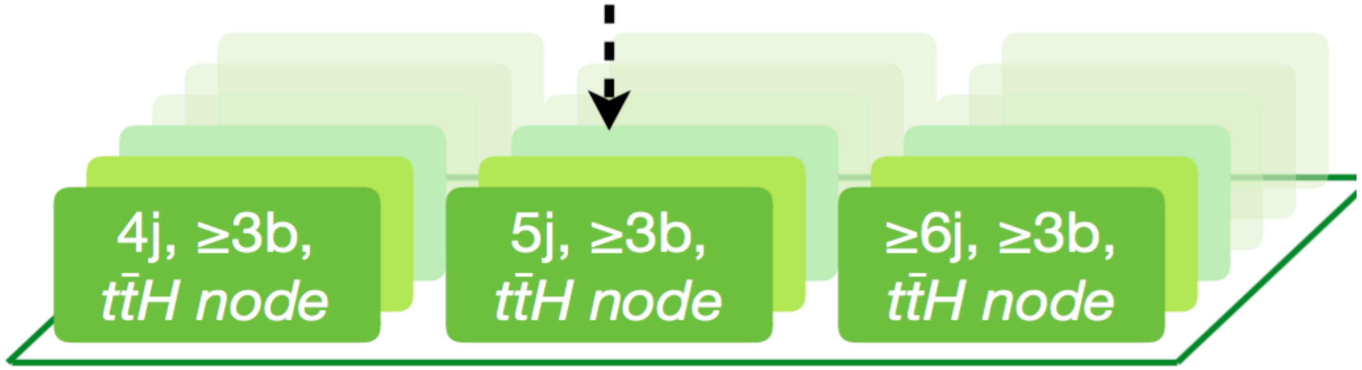
HIG-17-026

Single lepton

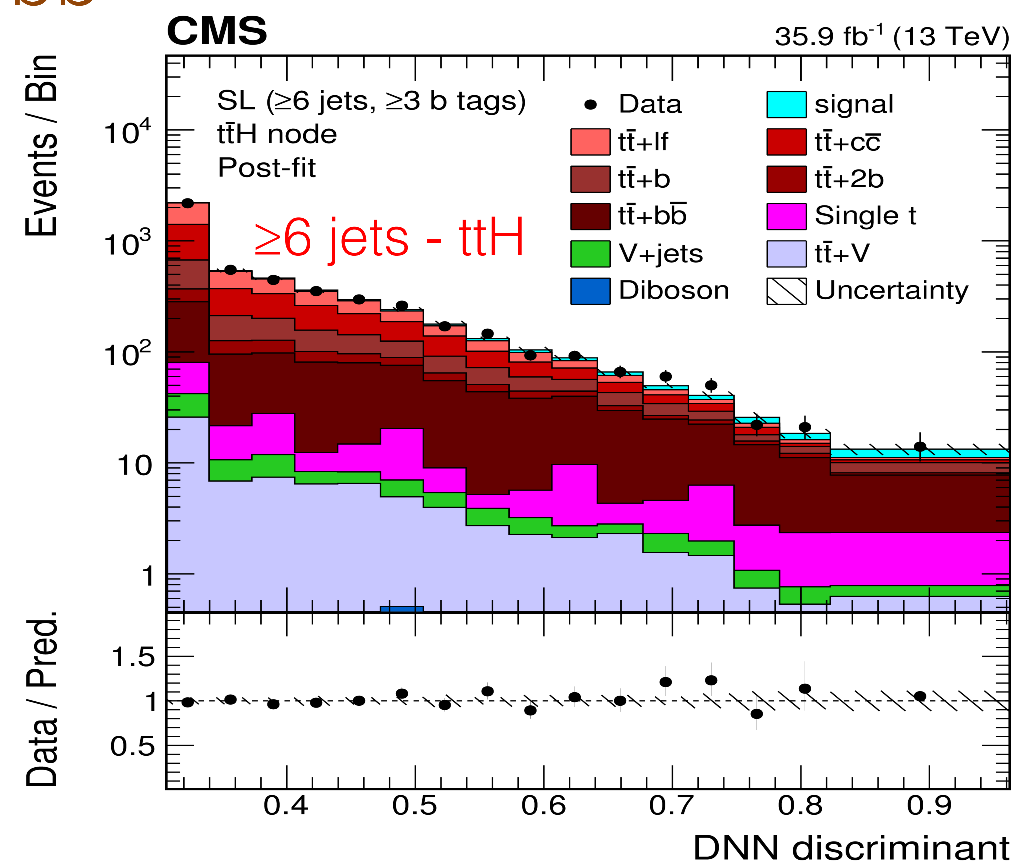
4j, ≥3b | 5j, ≥3b | ≥6j, ≥3b

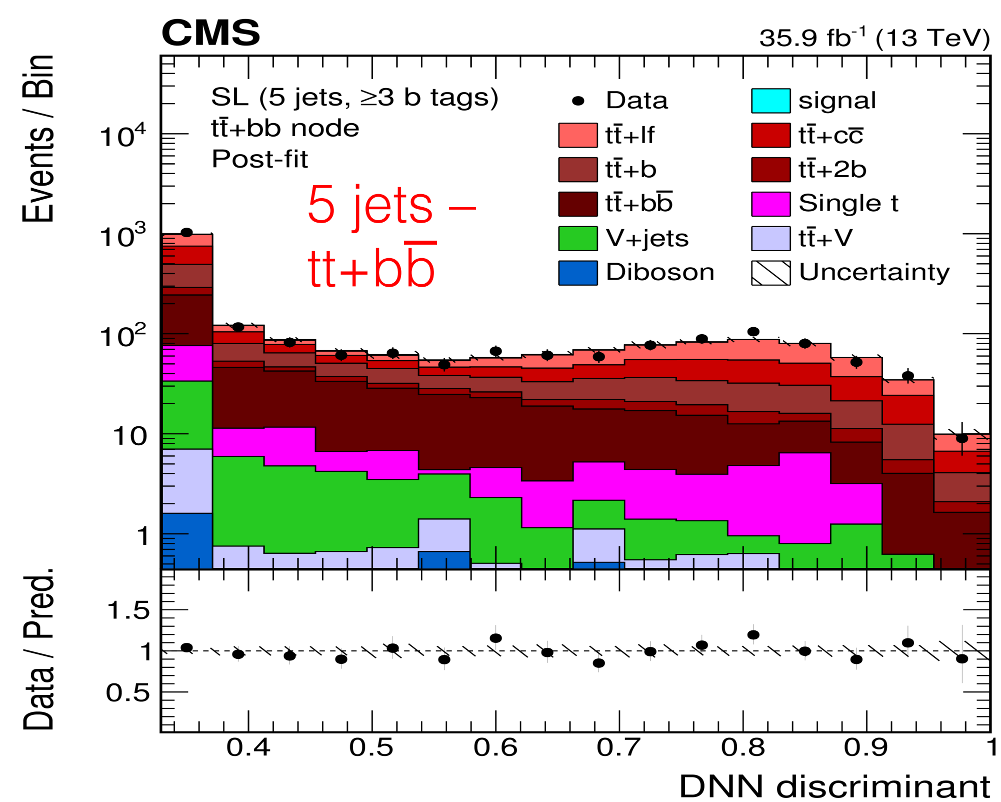
DNN (MEM is input)

Categorise by most probable process
tt̄H, tt̄+bb̄/b/2b/cc̄/lf



3 x 6 categories





ttH, H → bb (leptonic) results

Combined 2ℓ and ℓ+jets channels

HIG-17-026, Submitted to JHEP

Channel	95% CL upper limit observed	95% CL upper limit expected	Best-fit μ ±tot (±stat ± syst)
Single-lepton	1.75	$1.03^{+0.44}_{-0.29}$	$0.84^{+0.52}_{-0.50}$ $\left(^{+0.27}_{-0.26}\ ^{+0.44}_{-0.43}\right)$
Dilepton	2.34	$2.48^{+1.17}_{-0.76}$	$-0.24^{+1.21}_{-1.12}$ $\left(^{+0.63}_{-0.60}\ ^{+1.04}_{-0.95}\right)$
Combined	1.51	$0.92^{+0.39}_{-0.26}$	$0.72^{+0.45}_{-0.45}$ $\left(^{+0.24}_{-0.24}\ ^{+0.38}_{-0.38}\right)$

Best fit

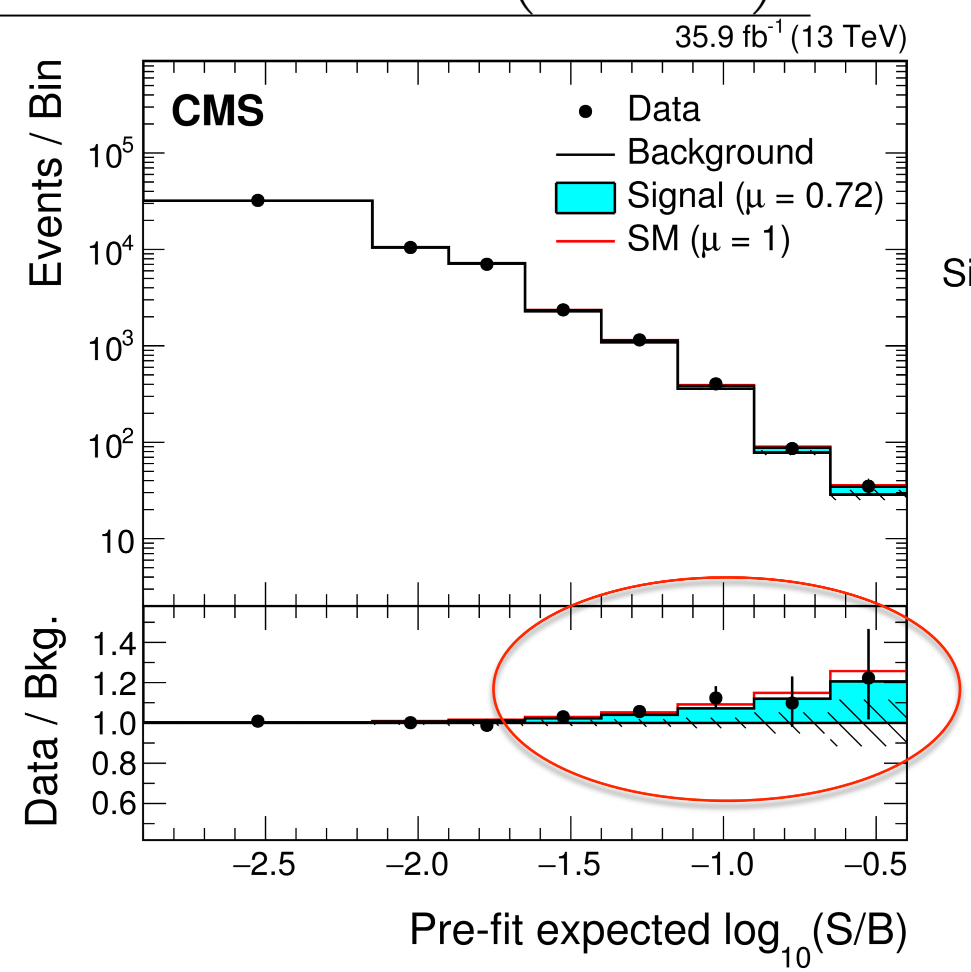
$\mu = 0.72 \pm 0.45$

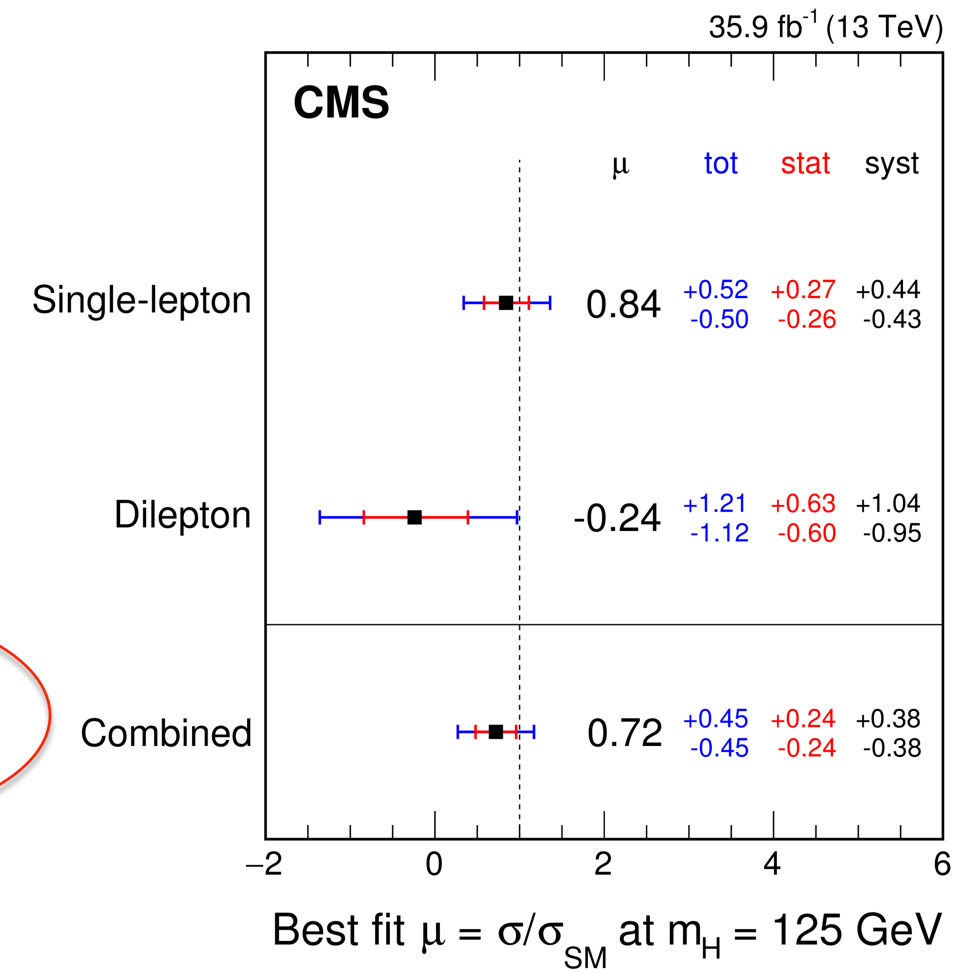
Observed (expected)

Significance = 1.6 (2.2) σ

Major systematic uncertainties:

- tt + HF prediction
- Jet energy scale & b-tagging

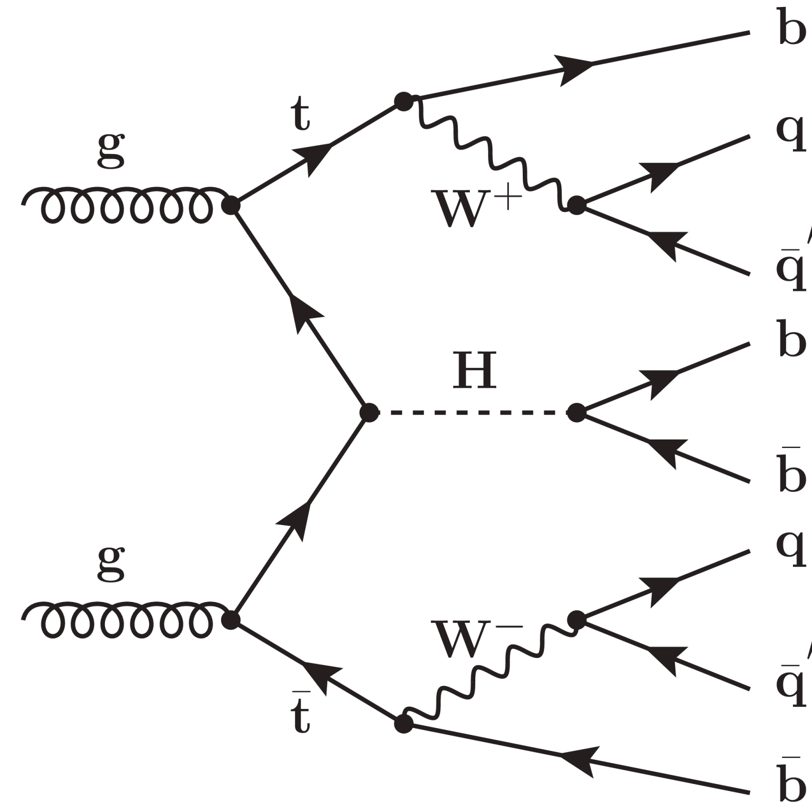




ttH, H → bb (Fully Hadronic)

HIG-17-022

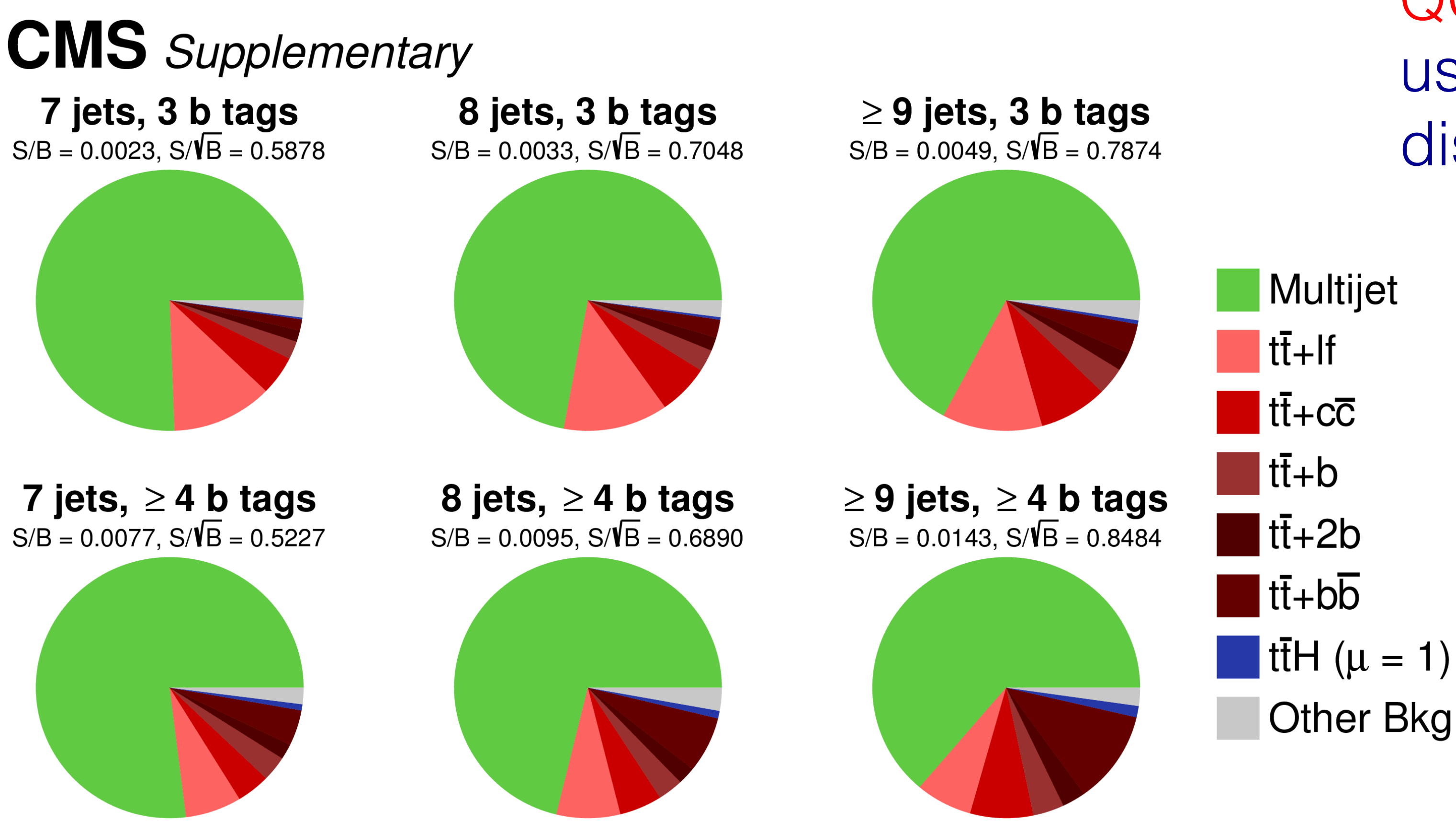
- Trigger: ≥ 6 jets, large H_T, ≥1 or 2 b-tagged jets
- Major Challenge:
  - Large backgrounds from QCD multi-jets, tt+jets, and the irreducible tt+bb
  - Large combinatorics of jets
  - Poor $b\bar{b}$ mass resolution
- However,
  - Larger signal contribution
  - Possibility to fully reconstruct the event



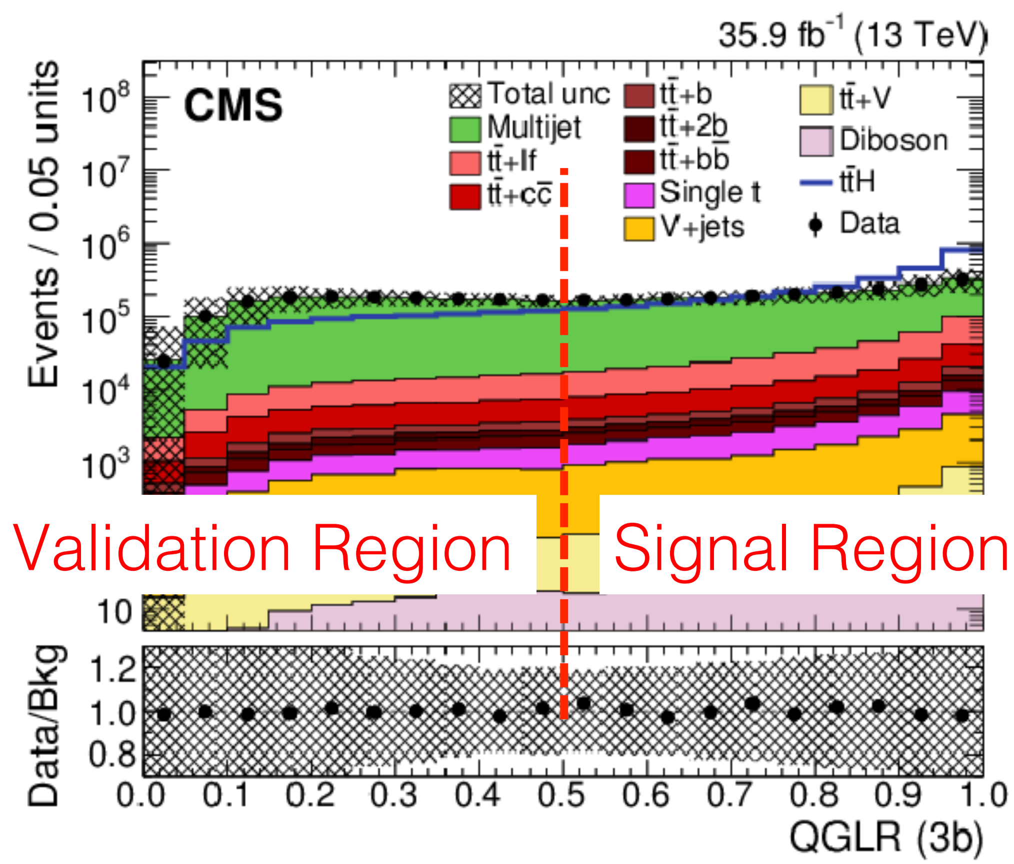
- A quark-gluon discriminant is used to differentiate quarks jets from gluon jets
  - Discrimination against QCD multijets
- Dedicated Matrix-Element (MEM) discriminants to discriminate signal against tt+jets and tt+bb

Analysis Strategy

Events divided in 8 exclusive categories



QCD multijet events are suppressed using a Quark – Gluon Likelihood (QGL) discriminant



Signal Regions: 7, 8, ≥9 jets, 3, ≥4 b-tags

Control region for QCD bkg estimation :
≥ 7 jets, 2 b-tagged

b-tagging using CSVM discriminator

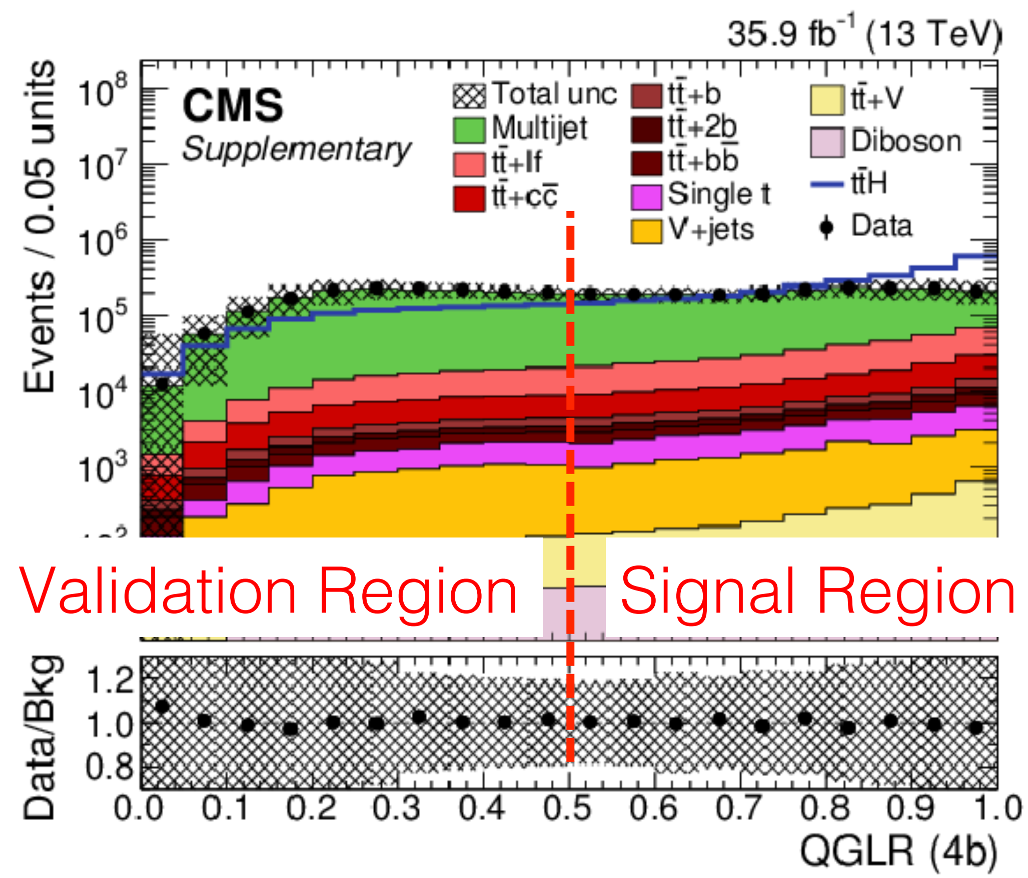
Analysis Strategy

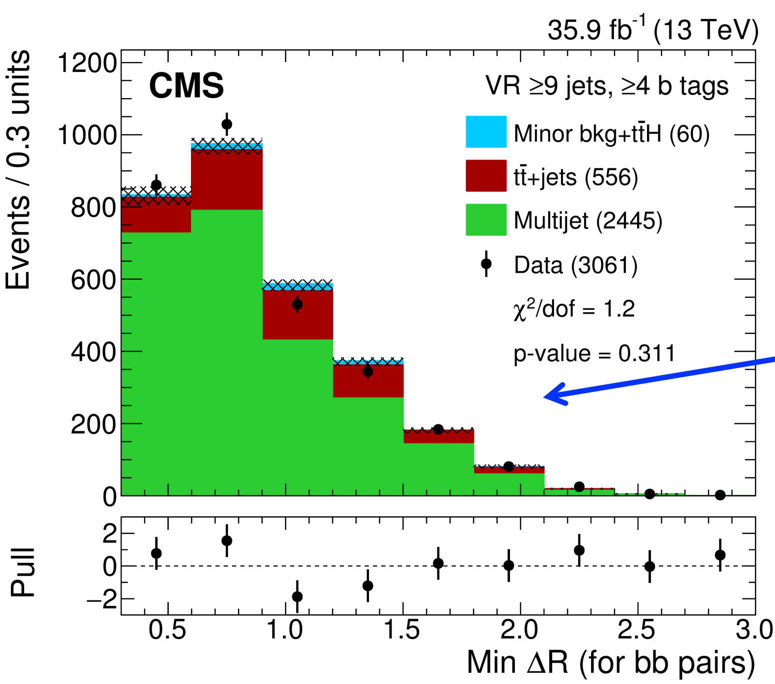
HIG-17-022

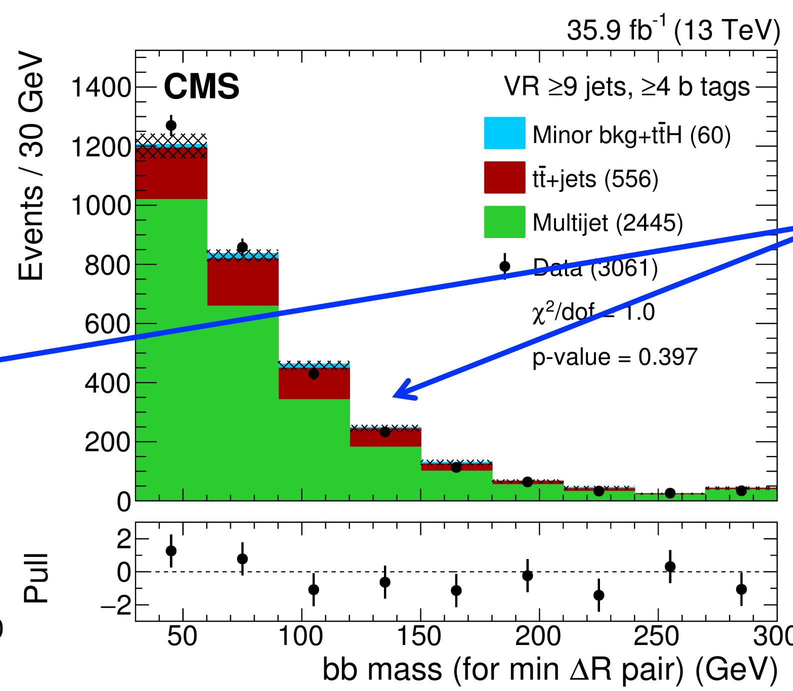
Control region for QCD bkg estimation

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

QCD multijet events are suppressed using a Quark – Gluon Likelihood (QGL) discriminant



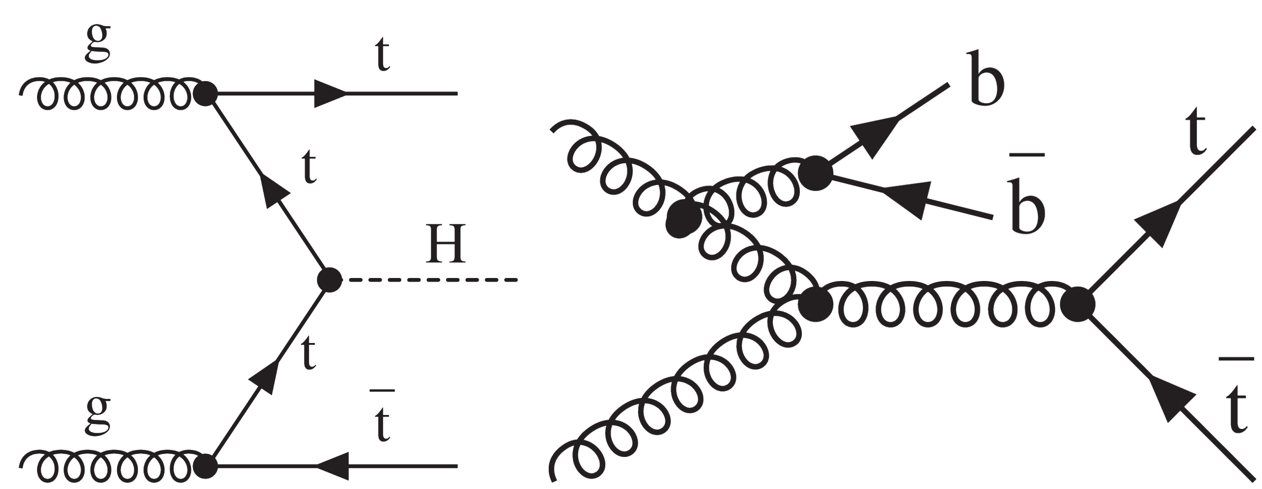


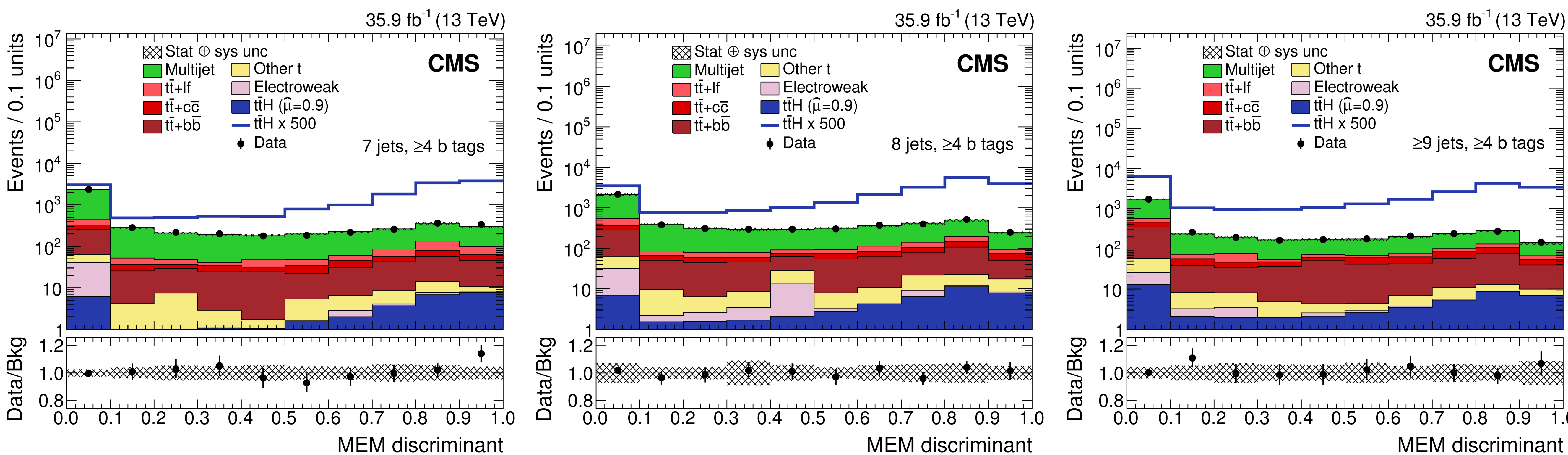


Signal Extraction

HIG-17-022

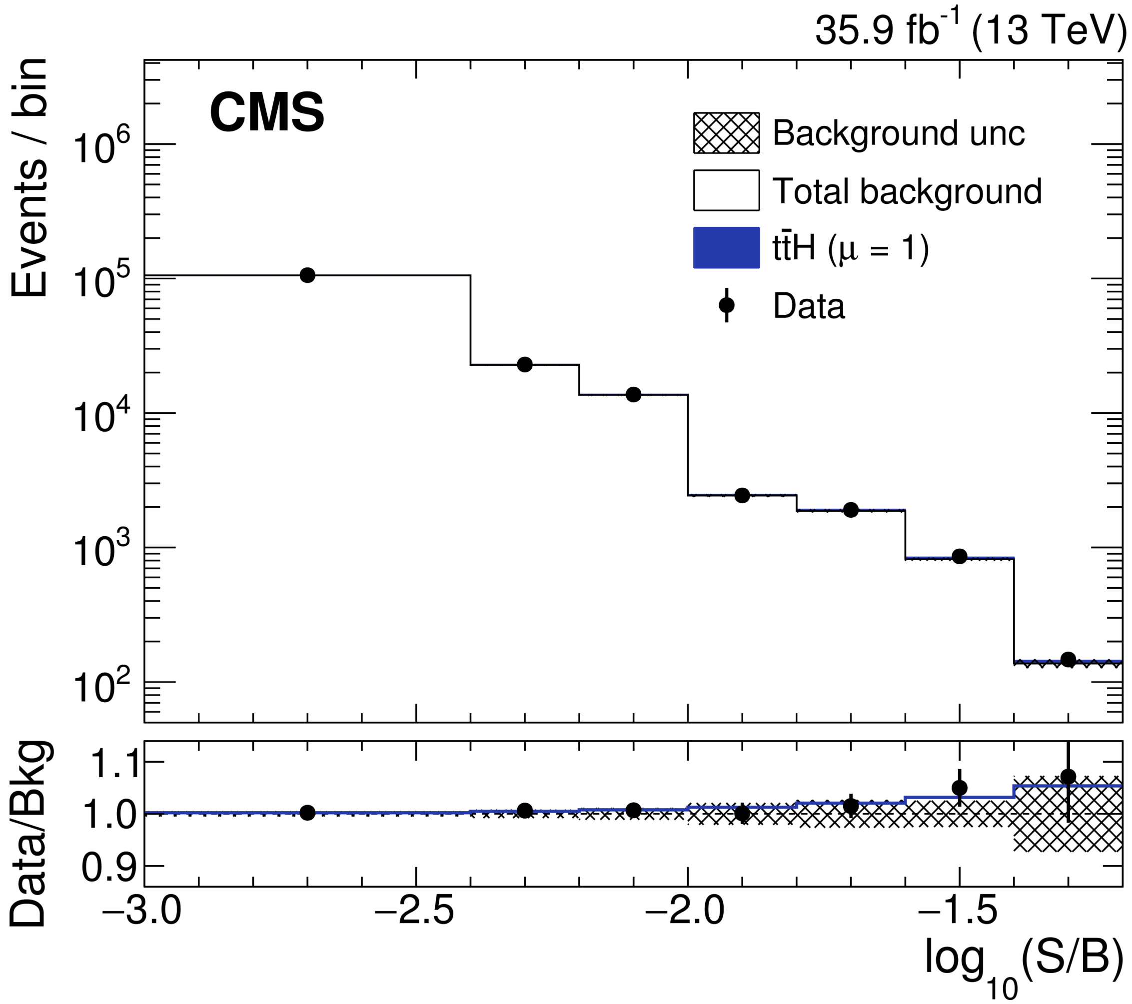
- Matrix-Element (MEM) discriminant to discriminate signal against backgrounds
- Constructed from LO matrix elements for the ttH signal and tt+bb backgrounds
- Also performs well against the tt+lf jets and QCD multi-jets backgrounds

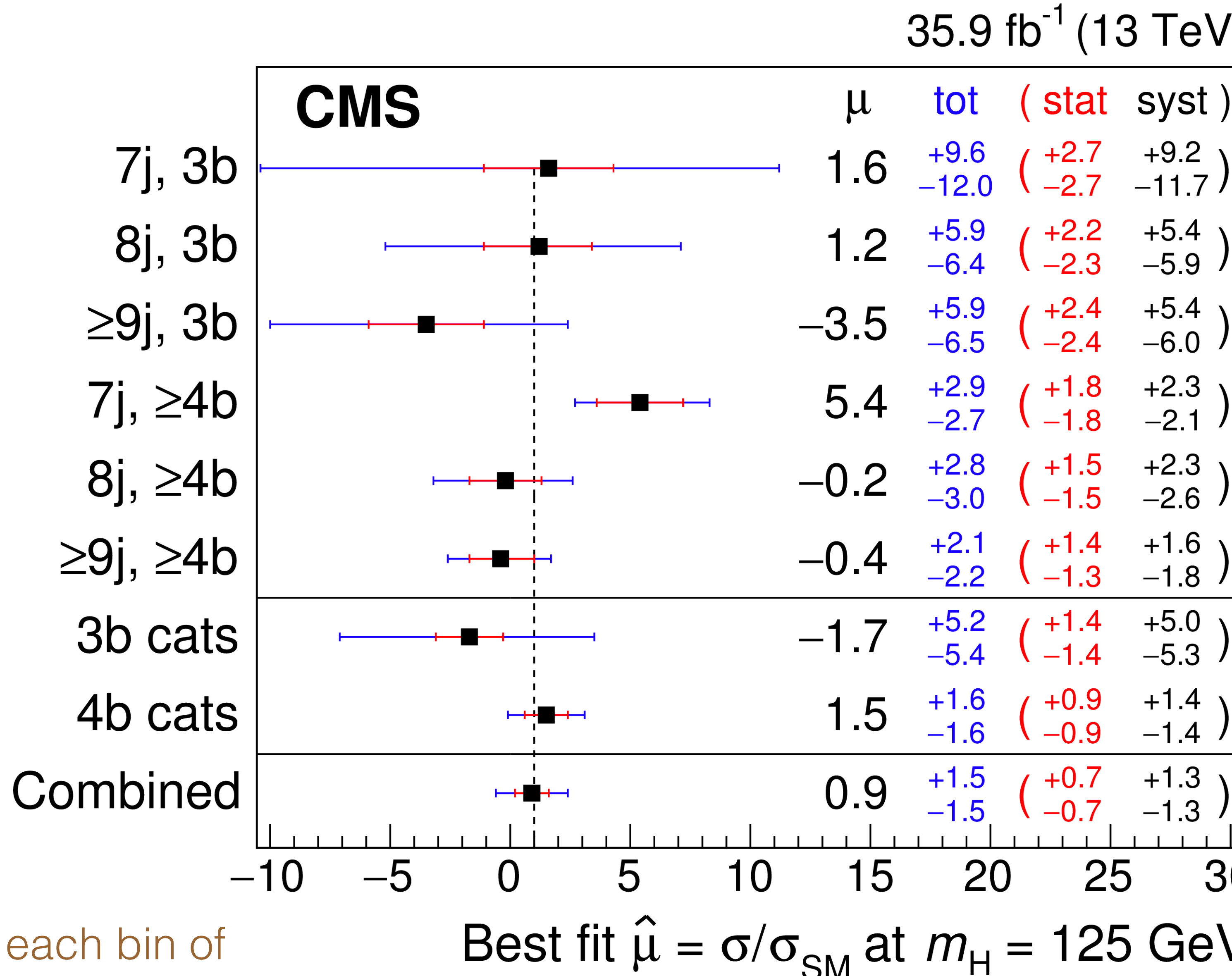




Results

HIG-17-022
Submitted to JHEP



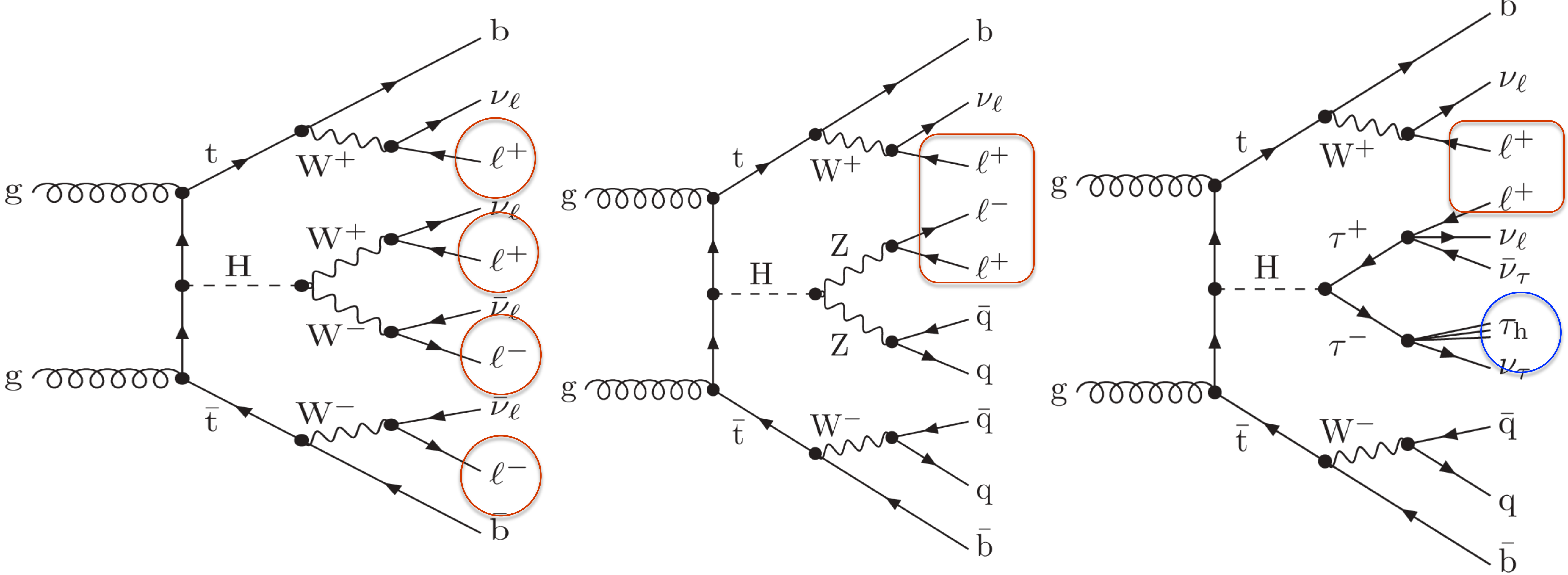


The ratio of the signal to background yields in each bin of the six MEM discriminant histograms, obtained from a combined fit constrained to the SM cross section of μ = 1

- Consistent with the SM expectation, driven by ≥4b categories
- Major systematic uncertainties: Multijet estimation, tt+HF prediction, b-tagging and JES etc..

ttH, H → multi-leptons

HIG-17-018



- Multi-lepton final states constitute Higgs decay to W^+W^-, ZZ, and $\tau^+\tau^-$
- Events categorized based on number of leptons and τ_h candidates
- BDT and MEM discriminants to discriminate signal from backgrounds

Analysis Strategy

HIG-17-018

Search performed in 6 exclusive event categories:

- One lepton and two τ_h ($1\ell + 2\tau_h$) [two τ_h to be of opposite charge]
- Two same-sign leptons and zero τ_h ($2\ell ss$)
- Two same-sign leptons and one τ_h ($2\ell ss + 1\tau_h$) [τ_h charge is opposite to ℓ]
- Three leptons and zero τ_h (3ℓ) [| Σcharge | = 1]
- Three leptons and one τ_h ($3\ell + 1\tau_h$)) [| Σcharge | = 0]
- Four leptons (4ℓ) [| Σcharge | = 0]

ℓ = e or μ

Additional Event Selections:

- 1 tight or 2 loose b-tagged jets
- At least 2 to 4 jets, depending on event category
- Cut on L_D ($0.6 \times p_T^{miss} + 0.4 \times H_T^{miss}$) and/or Z-veto depending on categories, lepton flavour etc..

Analysis Strategy

HIG-17-018

Backgrounds:

- ttV, di-boson (irreducible) and tt+jets (misidentified leptons) are major backgrounds
- Irreducible backgrounds are modeled using MC simulation, and validated in dedicated control regions
- Background from misidentified leptons and/or τ_h is measured from data using a fake factor method, applied to each category separately
  - Fake probabilities are measured in dedicated control regions as function of p_T and η.
- The background with misidentified lepton charge, to 2ℓss & 2ℓss + $1\tau_h$ categories, are also measured from data similarly
  - Charge misidentification probabilities are measured in Z→ee and Z→$\mu\mu$ events

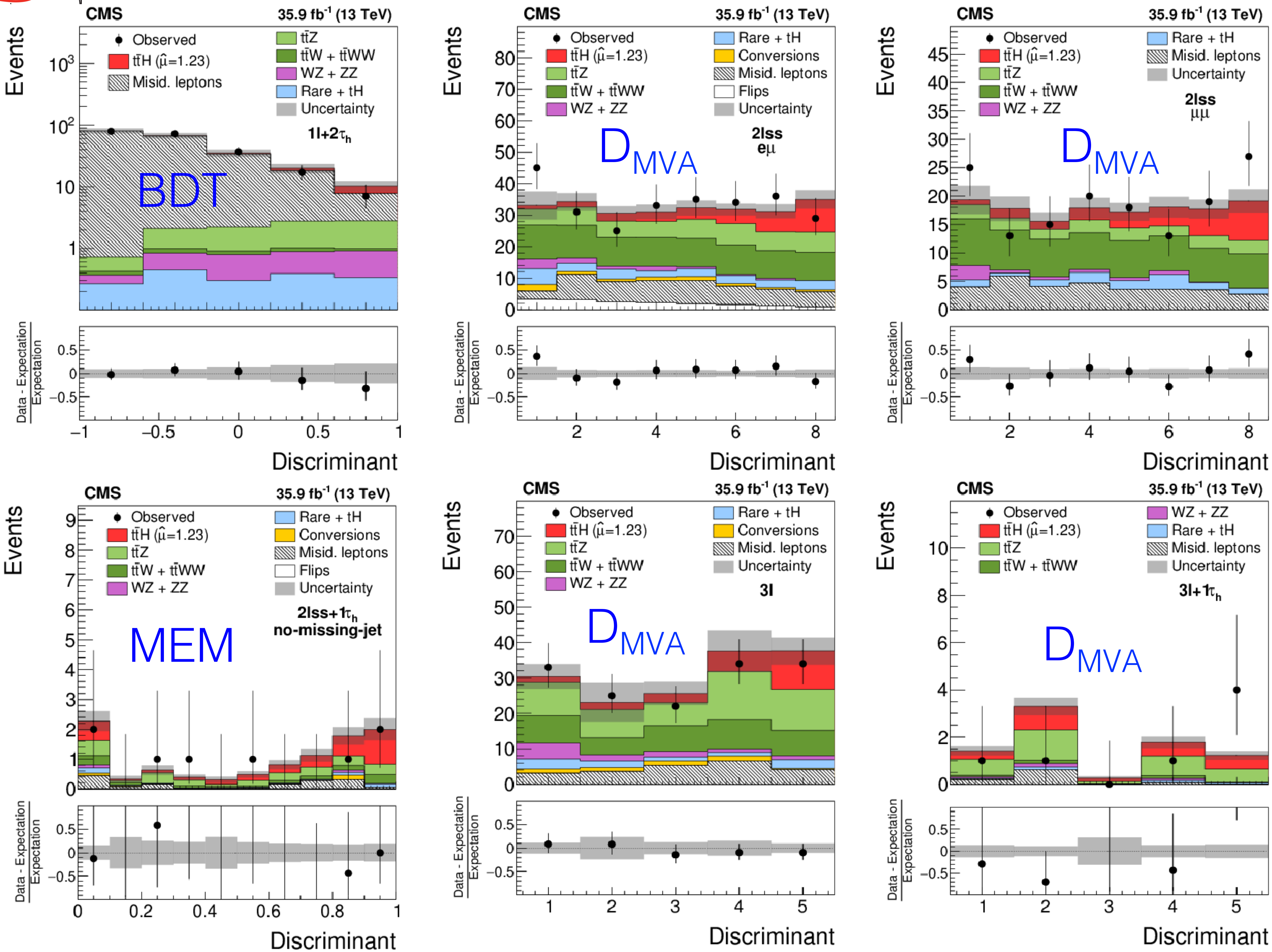
Signal Extraction:

- 2ℓss + $1\tau_h$: MEM discriminant to separate signal from ttZ and tt+jets
- 1ℓ + $2\tau_h$: 1 BDT against tt+jets
- 2ℓss, 3ℓ, 3ℓ+$1\tau_h$: Separate BDTs for ttH vs ttV and ttH vs tt+jets
  - Matrix-Element weights as input to BDT in 3ℓ category
  - 2 BDTs are mapped to a single discriminant (D_{MVA}), used for final signal extraction
- 4ℓ: simple counting due to small statistics

Simultaneous Maximum Likelihood fit to discriminating variables in all categories

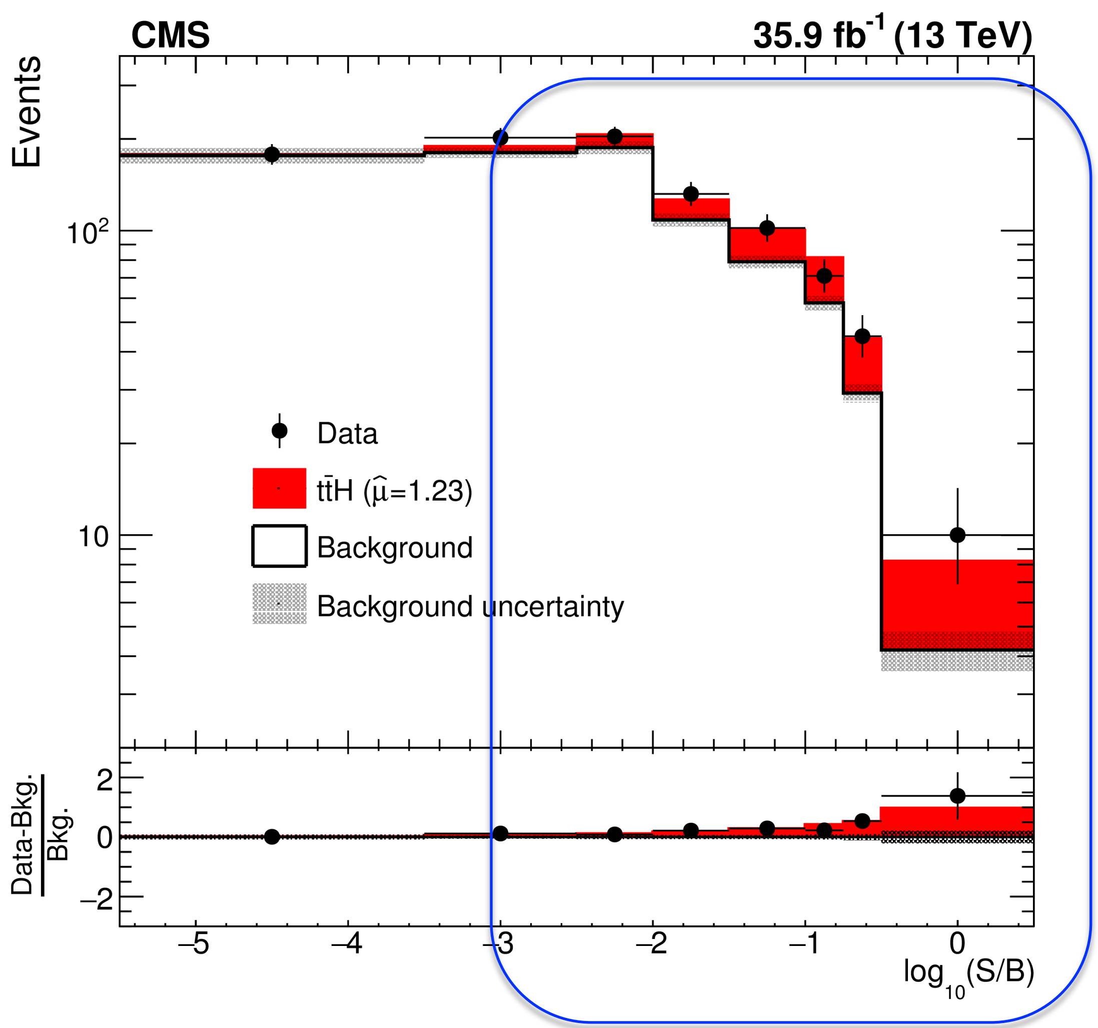
Analysis Strategy

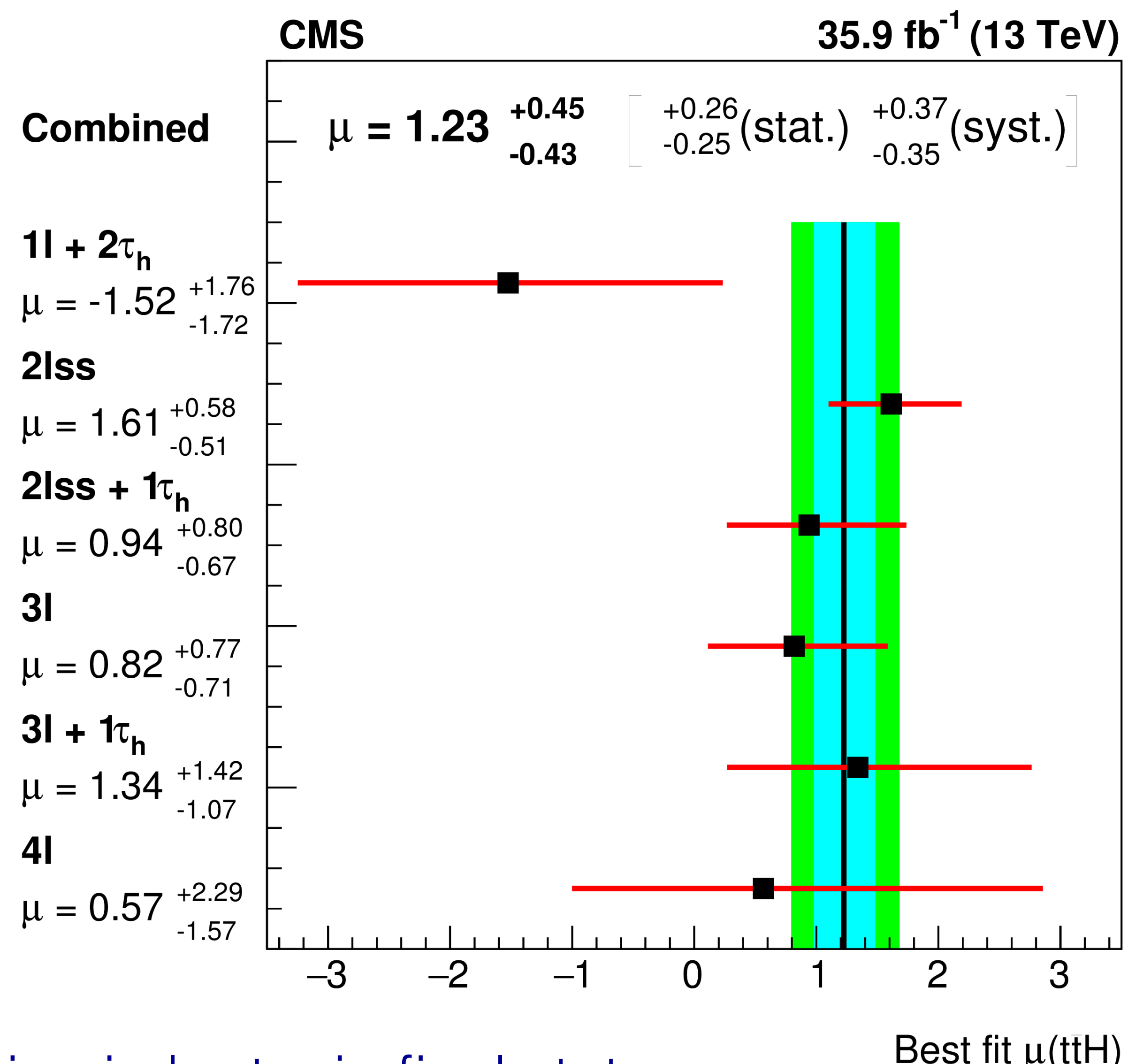
HIG-17-018



ttH, H → multi-leptons (Results)

HIG-17-018, Submitted to JHEP





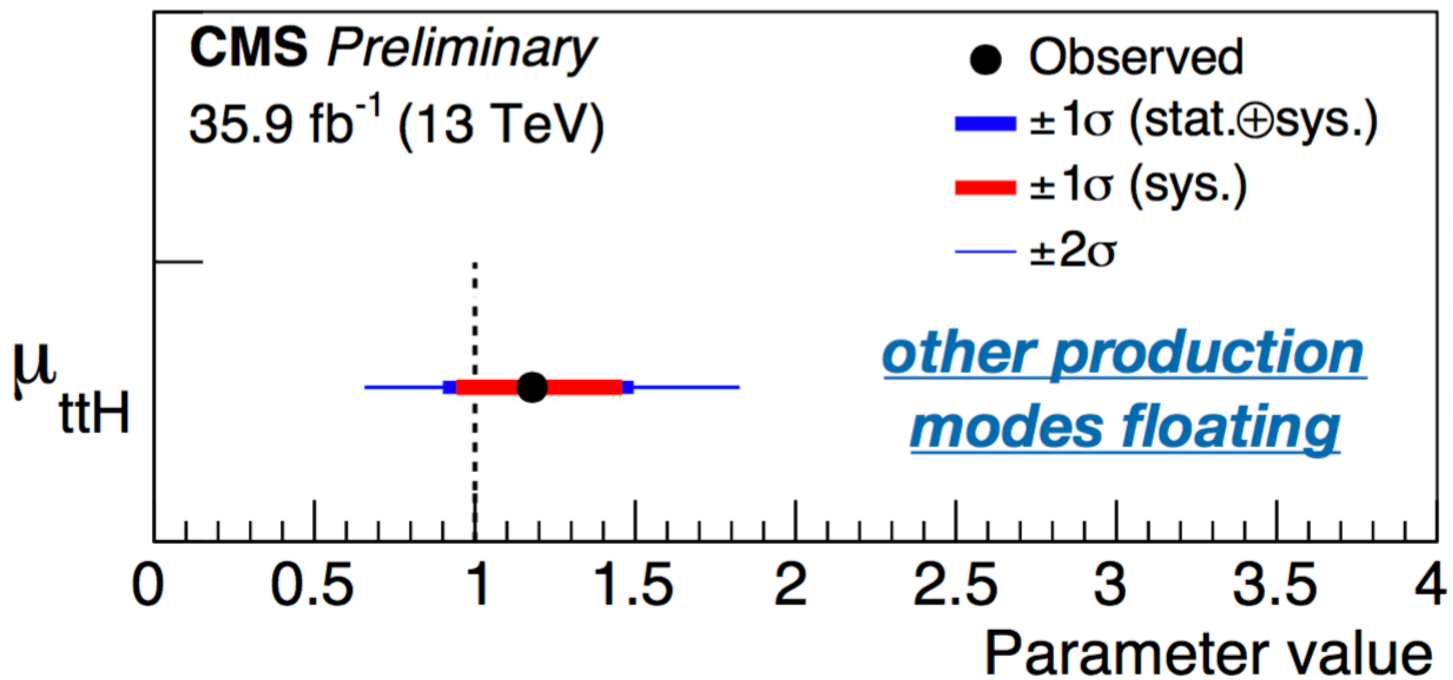
Evidence for the ttH production in leptonic final states:
3.2σ (obs) / 2.8σ (exp) significance

Cross check analysis with ttV as signal, with normalization constrained using control regions: $\mu = 1.04^{+0.50}_{-0.36}$, significance = 2.7σ

ttH Combination

CMS-PAS-HIG-17-031

Combination of ttH analyses, along with other Higgs measurements, for 13 TeV data



ttH		
Best fit value	Uncertainty Stat.	Syst.
$1.18\ ^{+0.31}_{-0.27}$	$^{+0.16}_{-0.16}$	$^{+0.26}_{-0.21}$
$\left(^{+0.28}_{-0.25}\right)$	$\left(^{+0.16}_{-0.16}\right)$	$\left(^{+0.23}_{-0.20}\right)$

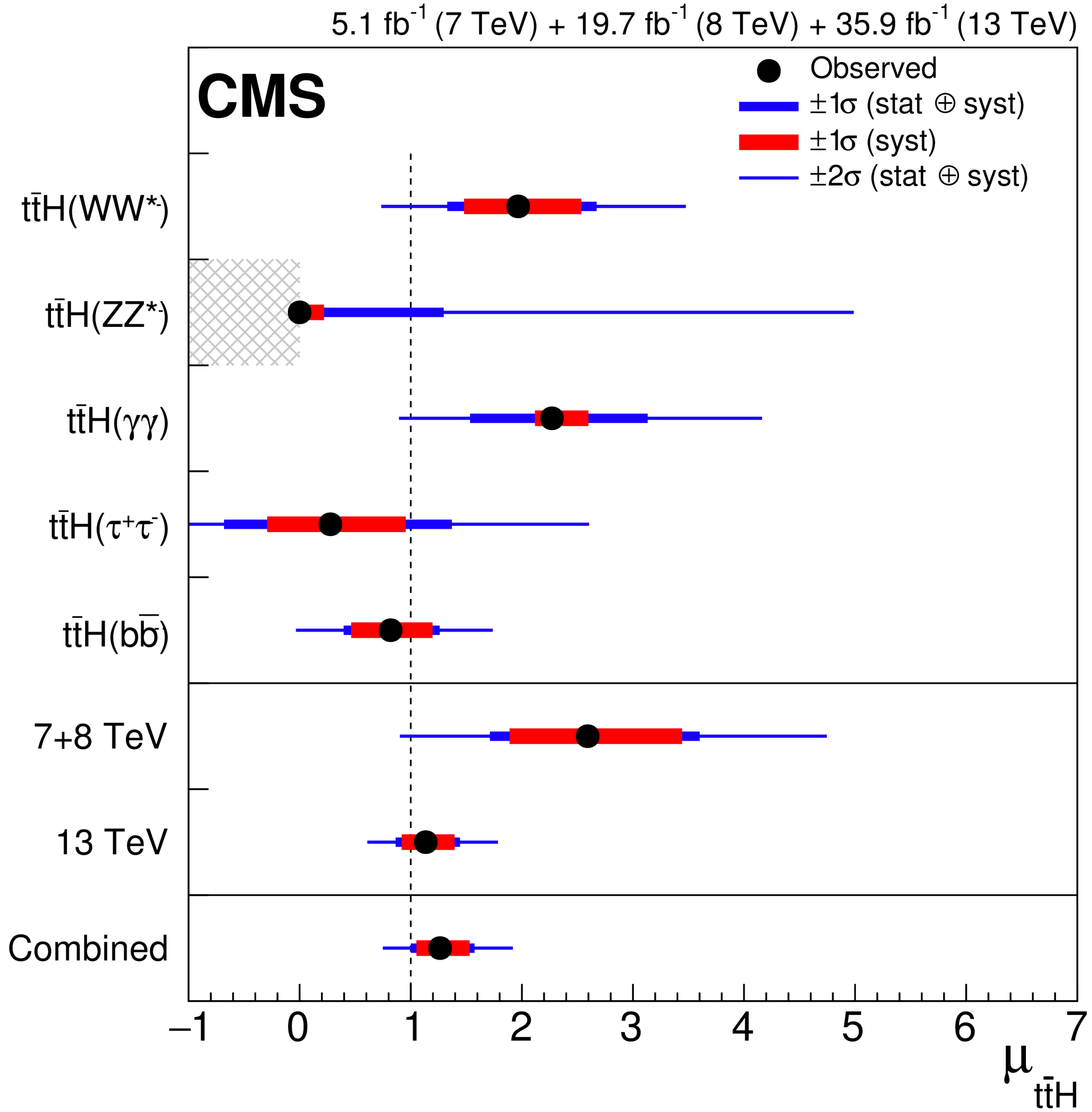
ttH ($\Delta\mu_{ttH}$) production cross section modifier from per-production mode fit

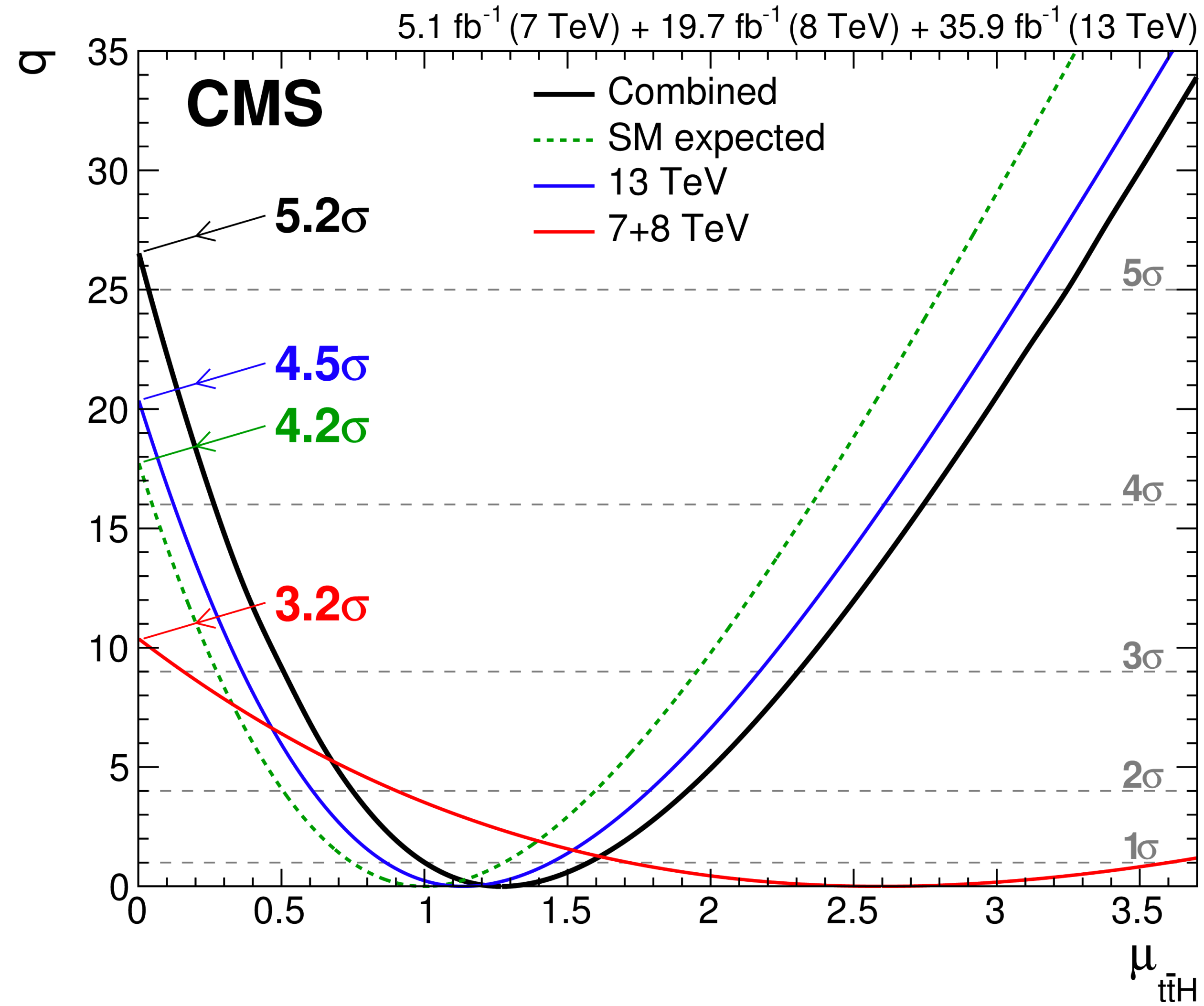
$\Delta\mu_{ttH} \approx 30\%$

Observation of ttH production

Combination of measurements in 7, 8, & 13 TeV data

Significance: 5.2σ (4.2σ Exp)





HIG-17-035

Submitted to PRL (arXiv: 1804.02610)

Summary

- Results presented for ttH searches with 35 fb^{-1} of pp collision data @ 13 TeV
  - Lots of improvement in analysis techniques compared to run-1
  - Addition of new challenging final states: fully hadronic mode, final states with hadronic decaying τ leptons
  - Top-Higgs couplings constrain to about ~15% with direct measurements
  - Working on further improvements in analysis sensitivity
- Observation of ttH production, combining 7, 8, and 13 TeV analyses

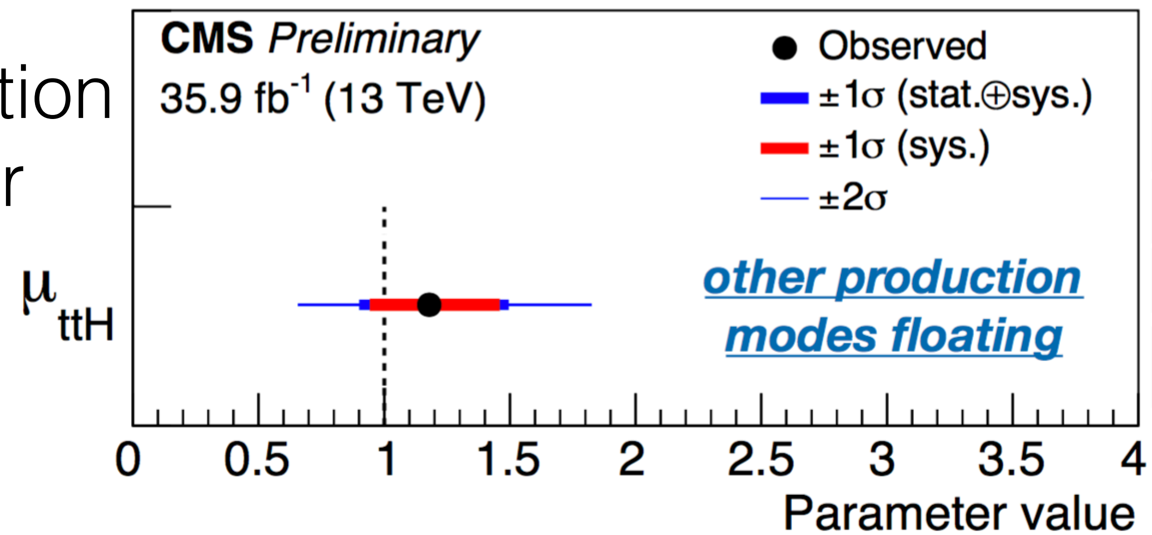
BACKUP

ttH Combination

CMS-PAS-HIG-17-031

Combination of ttH analyses, along with other Higgs measurements, for 13 TeV data

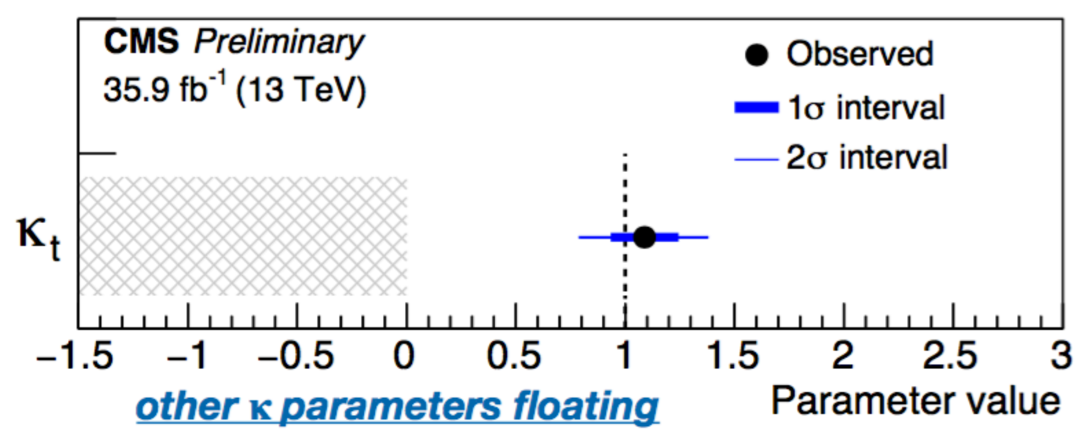
ttH+tH ($\Delta\mu_{ttH}$) production cross section modifier from per-production mode fit



ttH			
Best fit value		Uncertainty Stat.	Syst.
1.18	$^{+0.31}_{-0.27}$ $\left(^{+0.28}_{-0.25}\right)$	$^{+0.16}_{-0.16}$ $\left(^{+0.16}_{-0.16}\right)$	$^{+0.26}_{-0.21}$ $\left(^{+0.23}_{-0.20}\right)$

$\Delta\mu_{ttH} \approx 30\%$

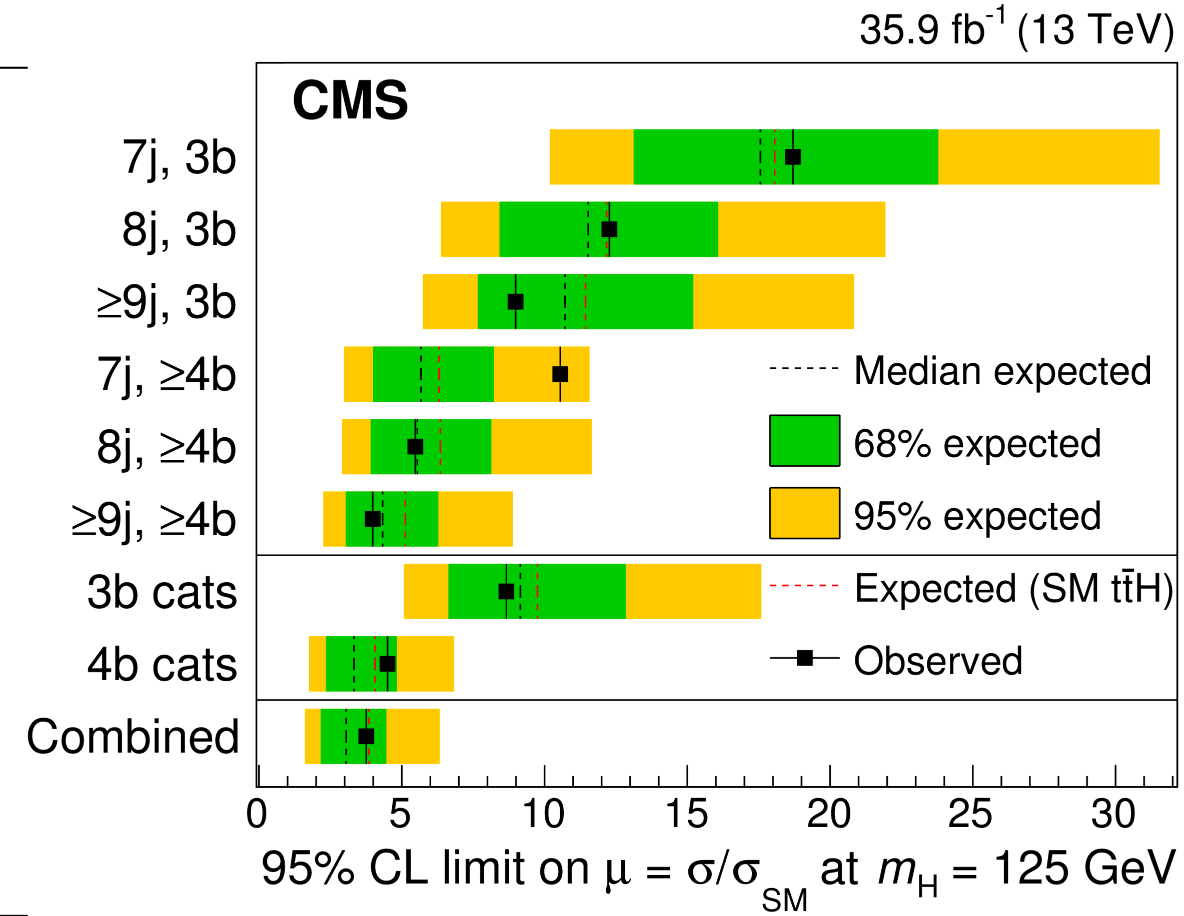
top coupling modifier
from **κ-framework** fit
in the unresolved loops assumption:



Best fit		Uncertainty Stat.	Syst.
1.09	$^{+0.14}_{-0.14}$ $\left(^{+0.14}_{-0.15}\right)$	$^{+0.08}_{-0.08}$ $\left(^{+0.08}_{-0.09}\right)$	$^{+0.12}_{-0.12}$ $\left(^{+0.12}_{-0.12}\right)$

ttH, H → bb (Fully Hadronic) Results

Category	Best fit $\hat{\mu}$ and uncertainty: $\hat{\mu}$	total	(stat syst)	Observed UL	Expected UL
7j, 3b	1.6	$^{+9.6}_{-12.0}$	$\left(^{+2.7}_{-2.7}\ ^{+9.2}_{-11.7}\right)$	18.7	$17.6^{+6.2}_{-4.4}$
8j, 3b	1.2	$^{+5.9}_{-6.4}$	$\left(^{+2.2}_{-2.3}\ ^{+5.4}_{-5.9}\right)$	12.3	$11.5^{+4.6}_{-3.1}$
≥9j, 3b	−3.5	$^{+5.9}_{-6.5}$	$\left(^{+2.4}_{-2.4}\ ^{+5.4}_{-6.0}\right)$	9.0	$10.7^{+4.5}_{-3.1}$
7j, ≥4b	5.4	$^{+2.9}_{-2.7}$	$\left(^{+1.8}_{-1.8}\ ^{+2.3}_{-2.1}\right)$	10.6	$5.7^{+2.6}_{-1.7}$
8j, ≥4b	−0.2	$^{+2.8}_{-3.0}$	$\left(^{+1.5}_{-1.5}\ ^{+2.3}_{-2.6}\right)$	5.5	$5.5^{+2.6}_{-1.6}$
≥9j, ≥4b	−0.4	$^{+2.1}_{-2.2}$	$\left(^{+1.4}_{-1.3}\ ^{+1.6}_{-1.8}\right)$	4.0	$4.3^{+1.9}_{-1.3}$
Combined	0.9	$^{+1.5}_{-1.5}$	$\left(^{+0.7}_{-0.7}\ ^{+1.3}_{-1.3}\right)$	3.8	$3.1^{+1.4}_{-0.9}$

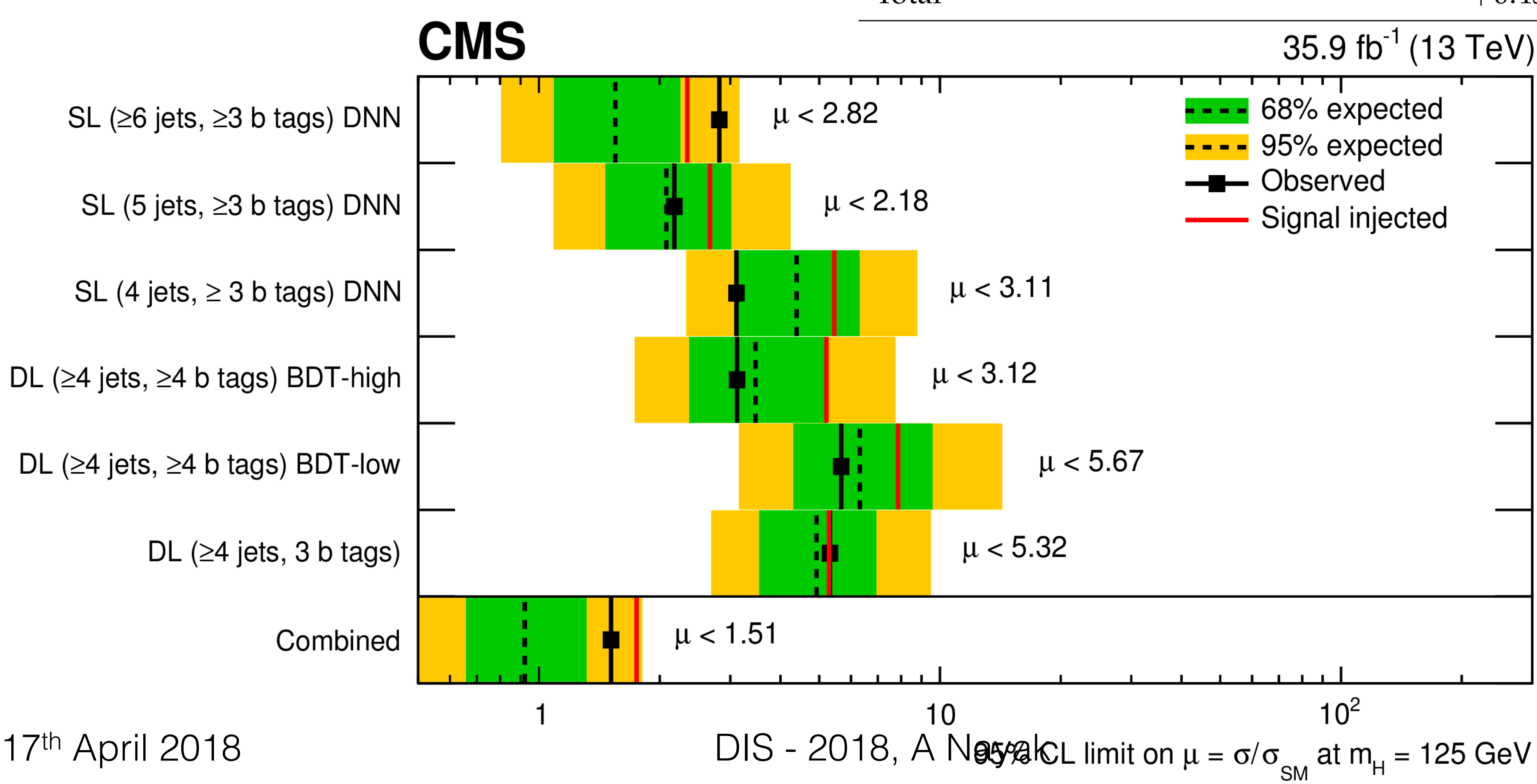


ttH, H → bb (Fully Hadronic) Results

Source	Range of uncertainty (%)	Distribution	Process			
			$t\bar{t}H$	Multijet	$t\bar{t}$+jets	Other
Experimental uncertainties						
Integrated luminosity	2.5	No	✓	*	✓	✓
Trigger efficiency	1–2	Yes	✓	*	✓	✓
Pileup	0.2–5	Yes	✓	*	✓	✓
JES	3–11	Yes	✓	*	✓	✓
JER	2–11	Yes	✓	*	✓	✓
b tagging	4–40	Yes	✓	*	✓	✓
QGL reweighting	4–11	Yes	✓	*	✓	✓
Top quark p_{T} reweighting	1–2	Yes	—	*	✓	—
Multijet estimation						
CSVL correction	—	Yes	—	✓	—	—
MEM first bin	—	Yes	—	✓	—	—
H_{T} reweighting	—	Yes	—	✓	—	—
Normalization	∞	No	—	✓	—	—
Theoretical uncertainties						
$t\bar{t}+b\bar{b}$ normalization	50	No	—	*	✓	—
$t\bar{t}$+2b normalization	50	No	—	*	✓	—
$t\bar{t}$+b normalization	50	No	—	*	✓	—
$t\bar{t}+c\bar{c}$ normalization	50	No	—	*	✓	—
$\mu_{\mathrm{F}}/\mu_{\mathrm{R}}$ scales for signal	6–9	No	✓	—	—	—
$\mu_{\mathrm{F}}/\mu_{\mathrm{R}}$ scales for background	1–13	No	—	*	✓	✓
PDFs	2–4	No	✓	*	✓	✓
Simulated sample size	2–40	Yes	✓	*	✓	✓

ttH, H → bb (Leptonic) Results

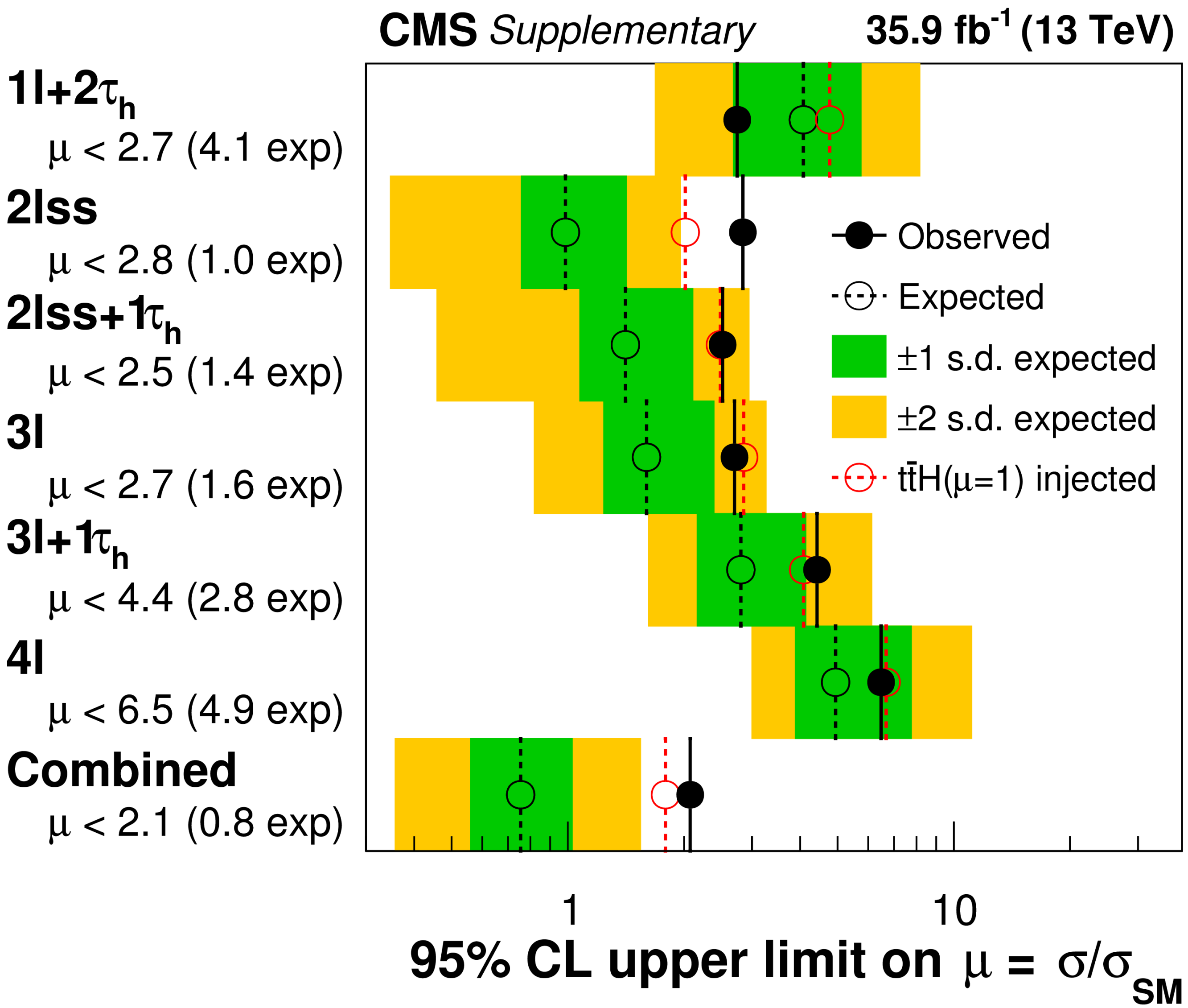
Uncertainty source	±Δμ (observed)	±Δμ (expected)
Total experimental	+0.15/−0.16	+0.19/−0.17
b tagging	+0.11/−0.14	+0.12/−0.11
jet energy scale and resolution	+0.06/−0.07	+0.13/−0.11
Total theory	+0.28/−0.29	+0.32/−0.29
$t\bar{t}$+hf cross section and parton shower	+0.24/−0.28	+0.28/−0.28
Size of the 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

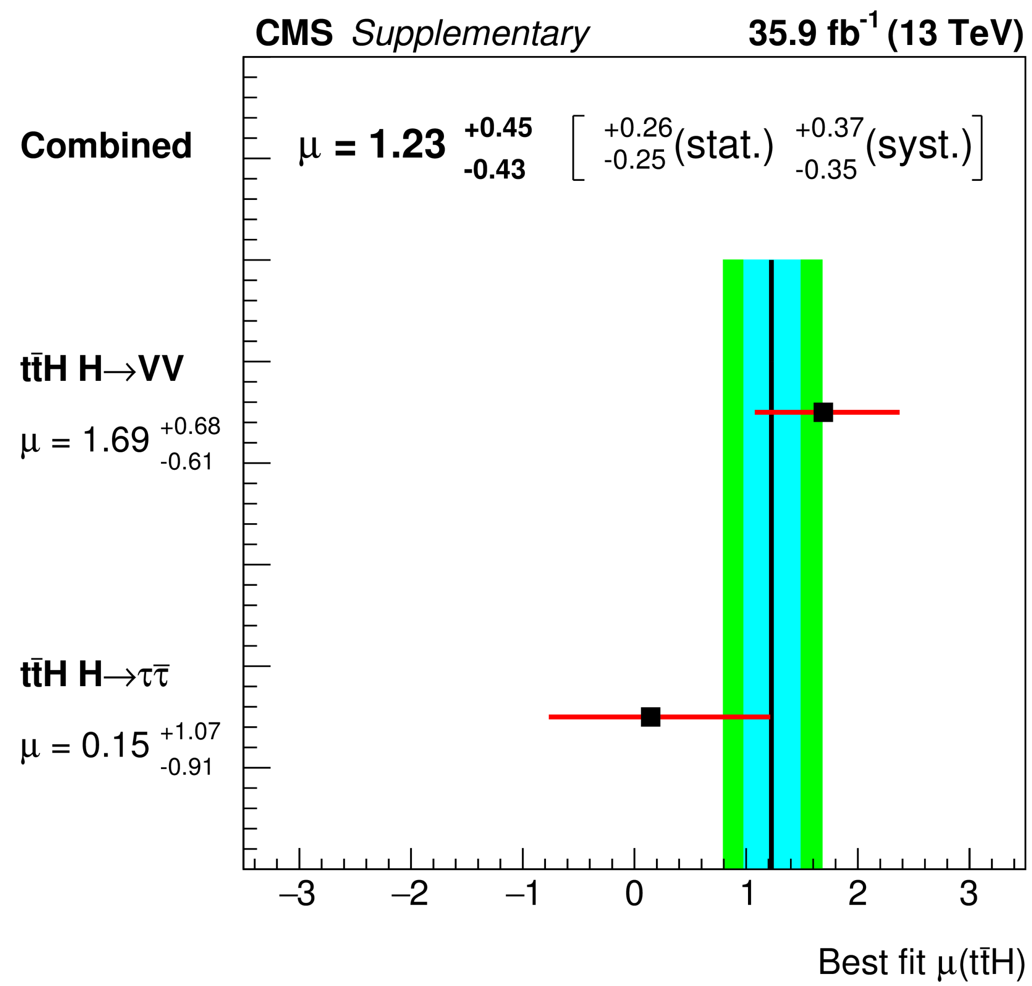


ttH, H → bb (Leptonic) Results

Source	Type	Remarks
Integrated luminosity	rate	Signal and all backgrounds
Lepton identification/isolation	shape	Signal and all backgrounds
Trigger efficiency	shape	Signal and all backgrounds
Pileup	shape	Signal and all backgrounds
Jet energy scale	shape	Signal and all backgrounds
Jet energy resolution	shape	Signal and all backgrounds
b tag hf fraction	shape	Signal and all backgrounds
b tag hf stats (linear)	shape	Signal and all backgrounds
b tag hf stats (quadratic)	shape	Signal and all backgrounds
b tag lf fraction	shape	Signal and all backgrounds
b tag lf stats (linear)	shape	Signal and all backgrounds
b tag lf stats (quadratic)	shape	Signal and all backgrounds
b tag charm (linear)	shape	Signal and all backgrounds
b tag charm (quadratic)	shape	Signal and all backgrounds
Renorm./fact. scales ($t\bar{t}$H)	rate	Scale uncertainty of NLO $t\bar{t}$H prediction
Renorm./fact. scales ($t\bar{t}$)	rate	Scale uncertainty of NLO $t\bar{t}$ prediction
Renorm./fact. scales ($t\bar{t}$+hf)	rate	Additional 50% rate uncertainty of $t\bar{t}$+hf predictions
Renorm./fact. scales (t)	rate	Scale uncertainty of NLO single t prediction
Renorm./fact. scales (V)	rate	Scale uncertainty of NNLO W and Z prediction
Renorm./fact. scales (VV)	rate	Scale uncertainty of NLO diboson prediction
PDF (gg)	rate	PDF uncertainty for gg initiated processes except $t\bar{t}$H
PDF (gg $t\bar{t}$H)	rate	PDF uncertainty for $t\bar{t}$H
PDF ($q\bar{q}$)	rate	PDF uncertainty of $q\bar{q}$ initiated processes ($t\bar{t}$+W,W,Z)
PDF (qg)	rate	PDF uncertainty of qg initiated processes (single t)
μ_R scale ($t\bar{t}$)	shape	Renormalisation scale uncertainty of the $t\bar{t}$ ME generator, independent for additional jet flavours
μ_F scale ($t\bar{t}$)	shape	Factorisation scale uncertainty of the $t\bar{t}$ ME generator, independent for additional jet flavours
PS scale: ISR ($t\bar{t}$)	rate	Initial state radiation uncertainty of the PS (for $t\bar{t}$ events), jet multiplicity dependent rate uncertainty, independent for additional jet flavours
PS scale: FSR ($t\bar{t}$)	rate	Final state radiation uncertainty (for $t\bar{t}$ events), jet multiplicity dependent rate uncertainty, independent for additional jet flavours
ME-PS matching ($t\bar{t}$)	rate	NLO ME to PS matching, *hdamp* [?] (for $t\bar{t}$ events), jet multiplicity dependent rate uncertainty, independent for additional jet flavours
Underlying event ($t\bar{t}$)	rate	Underlying event (for $t\bar{t}$ events), jet multiplicity dependent rate uncertainty, independent for additional jet flavours
NNPDF3.0NLO ($t\bar{t}$H, $t\bar{t}$)	shape	Based on the NNPDF replicas, same for $t\bar{t}$H and additional jet flavours
Bin-by-bin event count	shape	Statistical uncertainty of the signal and background prediction due to the limited sample size

ttH, H → multi-leptons (Results)



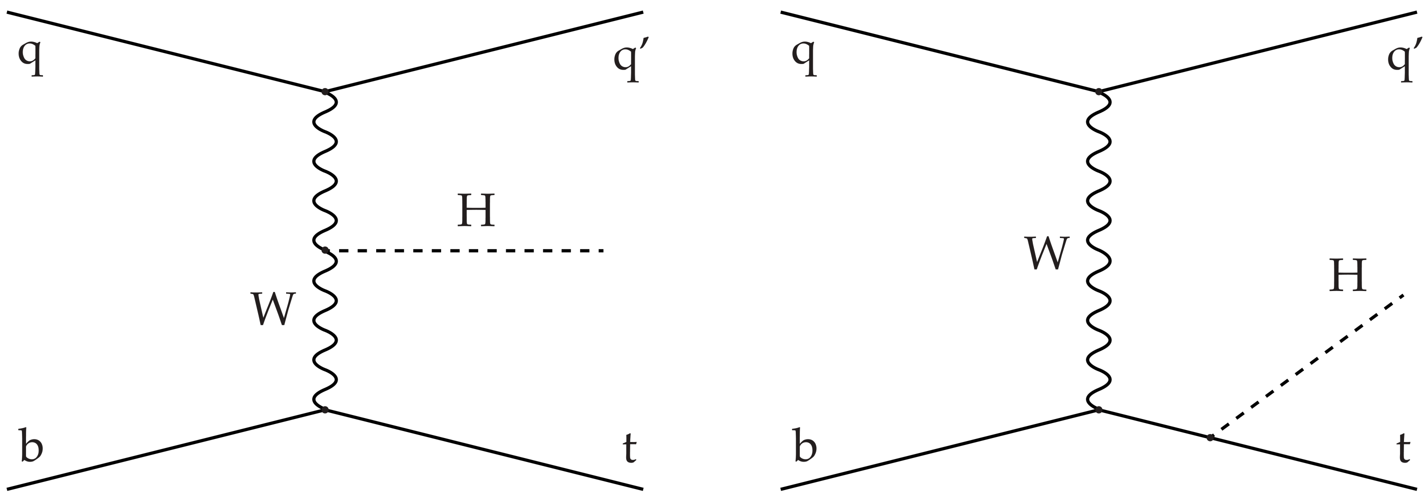


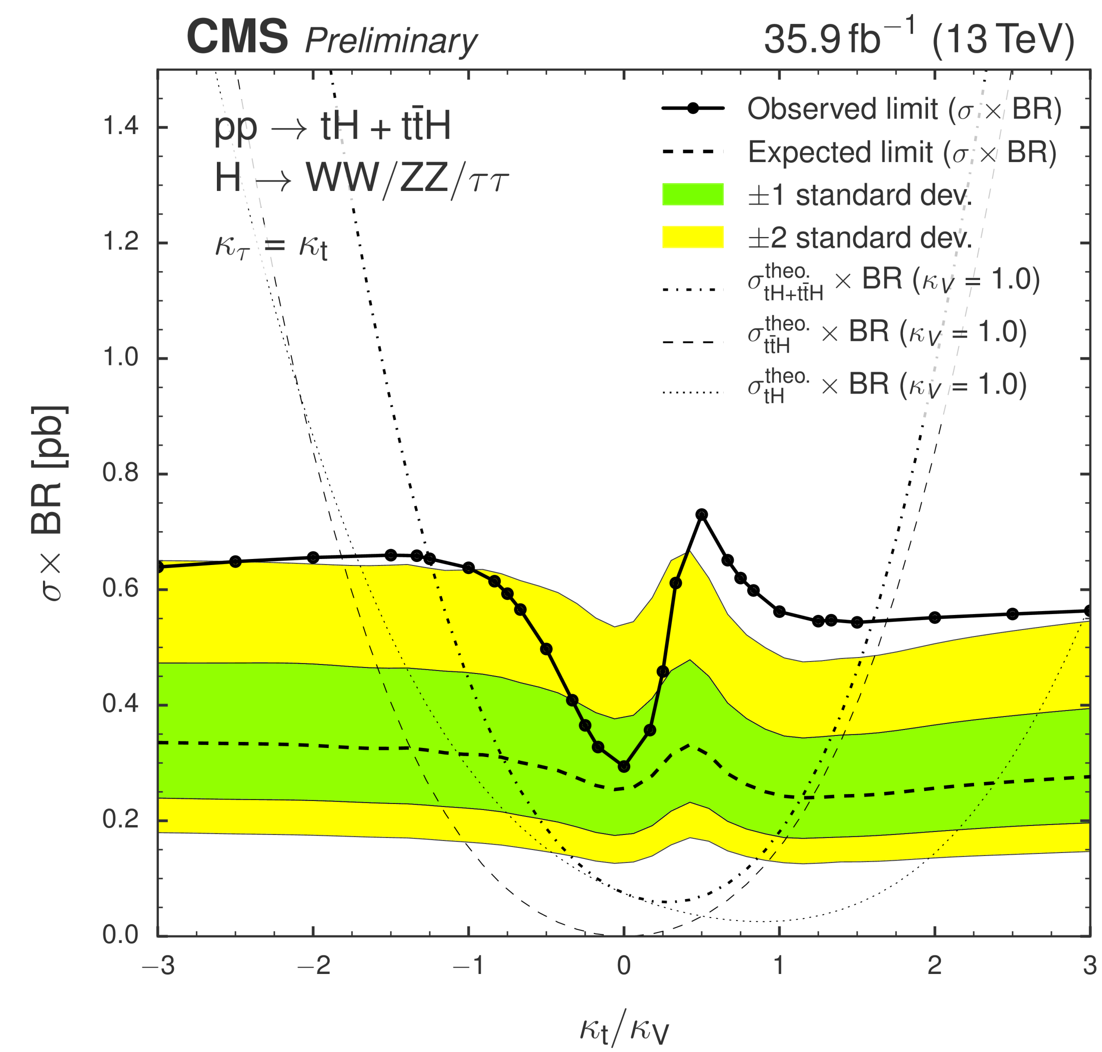
ttH, H → multi-leptons (Results)

Category	Observed limit on μ	Expected limit ($\mu = 0$)	Expected limit ($\mu = 1$)
$1\ell + 2\tau_h$	2.7	$4.1^{+1.7}_{-1.4}$	$4.8^{+2.0}_{-1.9}$
$2\ell ss$	2.8	$1.0^{+0.4}_{-0.2}$	$2.0^{+0.7}_{-0.3}$
$2\ell ss + 1\tau_h$	2.5	$1.4^{+0.7}_{-0.3}$	$2.5^{+0.9}_{-0.5}$
3ℓ	2.7	$1.6^{+0.8}_{-0.4}$	$2.9^{+1.1}_{-0.4}$
$3\ell + 1\tau_h$	4.4	$2.8^{+1.3}_{-0.6}$	$4.1^{+1.5}_{-0.7}$
4ℓ	6.5	$4.9^{+2.8}_{-1.1}$	$6.7^{+2.5}_{-0.8}$
Combined	2.1	$0.8^{+0.3}_{-0.2}$	$1.7^{+0.5}_{-0.5}$

Source	Uncertainty [%]	$\Delta\mu/\mu$ [%]
e, μ selection efficiency	2–4	11
τ_h selection efficiency	5	4.5
b tagging efficiency	2–15 [?]	6
Reducible background estimate	10–40	11
Jet energy calibration	2–15 [?]	5
τ_h energy calibration	3	1
Theoretical sources	≈10	12
Integrated luminosity	2.5	5

tHq, H → leptons





Destructive interference in SM and Constructive interference in inverted top coupling (ITC) scenarios

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]