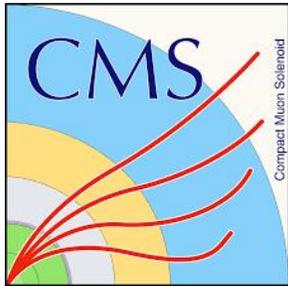


# Standard Model $H \rightarrow \gamma\gamma$ analysis

Sijing Zhang

On behalf of the CMS Collaboration

FCPPL, 24-27 April 2019, Shanghai



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



# Outline

- ✦ ttH analysis
- ✦ STXS analysis
- ✦ Conclusions

# *ttH (H → $\gamma\gamma$ ) analysis*

# Introduction

- This analysis is based on **2017 data** ( $41.5 \text{ fb}^{-1}$ ):
  - The results are combined with ttH ( $H \rightarrow \gamma\gamma$ ) 2016 results in [CMS-PAS-HIG-16-040 \(JHEP 11 \(2018\) 185\)](#)
- Documentation:
  - [CMS-PAS-HIG-18-018](#)

Available on the CERN CDS information server

**CMS PAS HIG-18-018**

---

## CMS Physics Analysis Summary

---

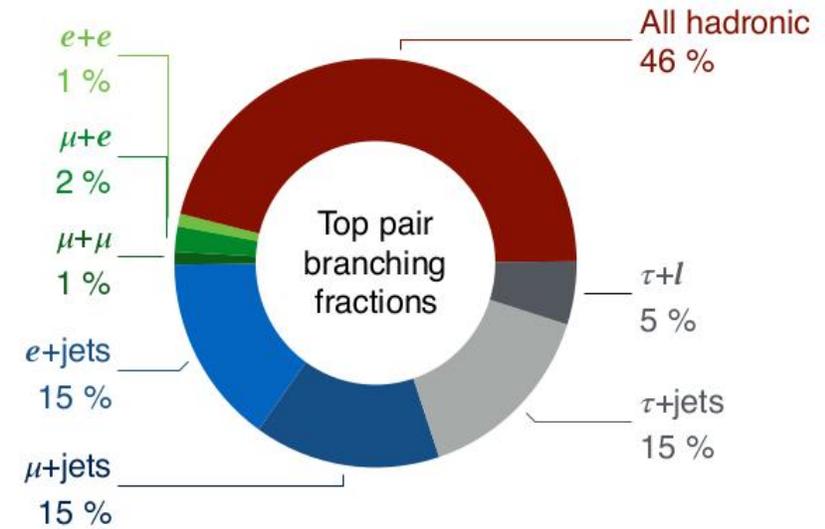
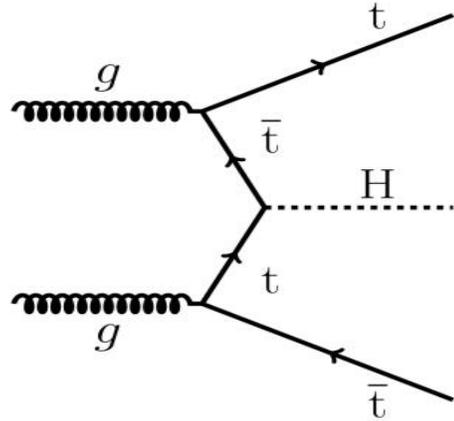
Contact: [cms-pag-conveners-higgs@cern.ch](mailto:cms-pag-conveners-higgs@cern.ch)

2018/11/28

Measurement of the associated production of a Higgs boson and a pair of top-antitop quarks with the Higgs boson decaying to two photons in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$

The CMS Collaboration

# Introduction



- The associated production of a Higgs boson and a top quark-antiquark pair ( $t\bar{t}H$  production) is a direct probe of the top-Higgs coupling.
- In this analysis we are only considering Higgs decaying into two photons.
- $H \rightarrow \gamma\gamma$  is identified by selecting two high  $p_T$  isolated photons.

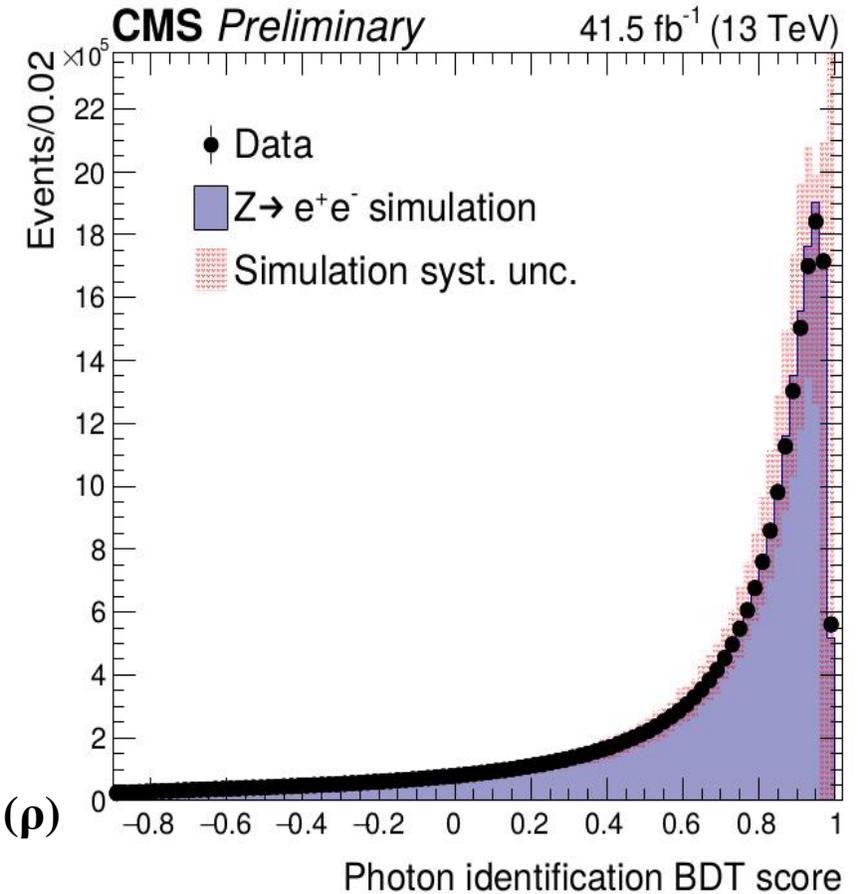
- The top pair decays:
  - Hadronic:  $t\bar{t} \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}q\bar{q}q\bar{q}$
  - Leptonic:  $t\bar{t} \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}l^+\nu_l\nu_l$
  - semi-leptonic:  $t\bar{t} \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}l^+\nu_l q\bar{q}$

# Analysis Strategy

- Select two **isolated** high  $p_T$  **photons**
- Find the **primary vertex** of the decay
  - the **CMS 0th vertex selection** is used instead of the  $H \rightarrow \gamma\gamma$  specific vertex.
- Depending on the top quark decay mode, events forming exclusive hadronic and leptonic channels are selected.
  - **Hadronic:**  $n_{\text{lepton}} = 0$
  - **Leptonic:**  $n_{\text{lepton}} > 0$
- A new **ttH Leptonic MVA** is introduced in this analysis
  - The MVAs are optimized maximizing the expected significance estimate.
- The signal and background shapes are fit to **diphoton invariant mass spectrum:**
  - The signal model is fit to **all simulated Higgs production modes**
  - the background model is **extracted from data**
- Results are presented as **signal strength modifiers**  $\mu$ :
  - The principal result is the **ttH signal strength**

# Photon Selection

- Trigger selection:
  - The events with two isolated high  $p_T$  photons are selected with **diphoton triggers on data**
  - double-photon trigger path based on **transverse energy**, **ratio** of hadronic to electromagnetic energy in the calorimeters, **electromagnetic shower shapes** and **isolation variables**, lower bound on  $m_{\gamma\gamma}$
- Preselection:
  - Similar to trigger requirements, but more stringent
- Photon Identification:
  - Multivariate approach to reject fake photon candidates (mainly from  $\pi_0$  mesons produced in jets)
  - **Shower shape** and **isolation observables**, **median energy density ( $\rho$ )**
  - **BDT output** provides an estimate of the per-photon quality



# Events Tagging

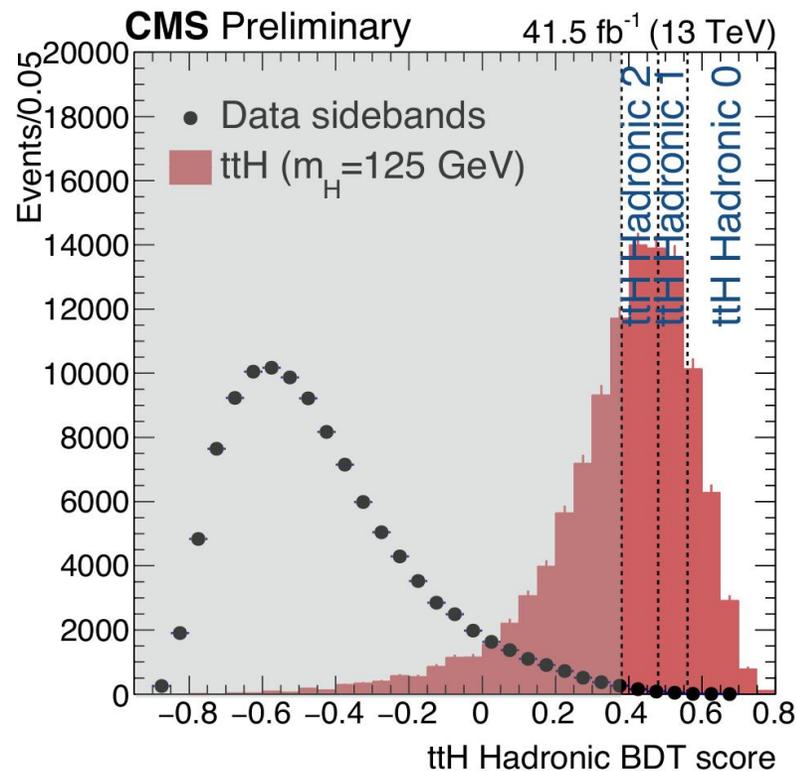
📌 Higgs accompanied by two b quarks and two W bosons:

- Hadronic (additional jets) Category:

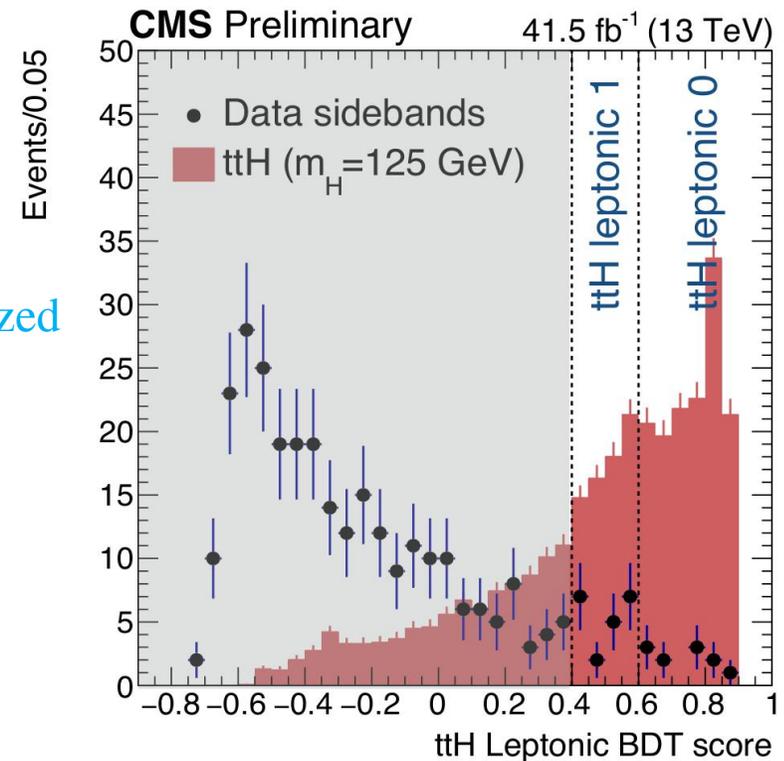
- **Hadronic 0**:  $0.56 < \text{ttHMVA}$
- **Hadronic 1**:  $0.48 < \text{ttHMVA} < 0.56$
- **Hadronic 2**:  $0.38 < \text{ttHMVA} < 0.48$

- Leptonic (additional leptons) Category:

- **Leptonic 0**:  $0.6 < \text{ttHMVA}$  for single lepton  
 $0.5 < \text{ttHMVA}$  for dilepton
- **Leptonic 1**:  $0.4 < \text{ttHMVA} < 0.6$  for single lepton



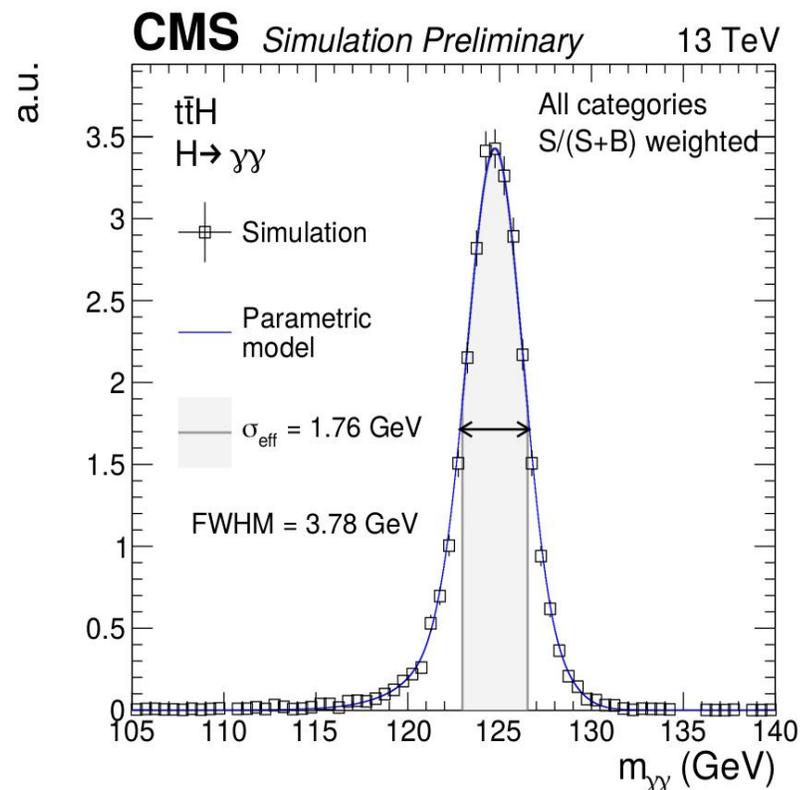
Simulation is normalized to data sideband.



# Signal and Background Parametrization

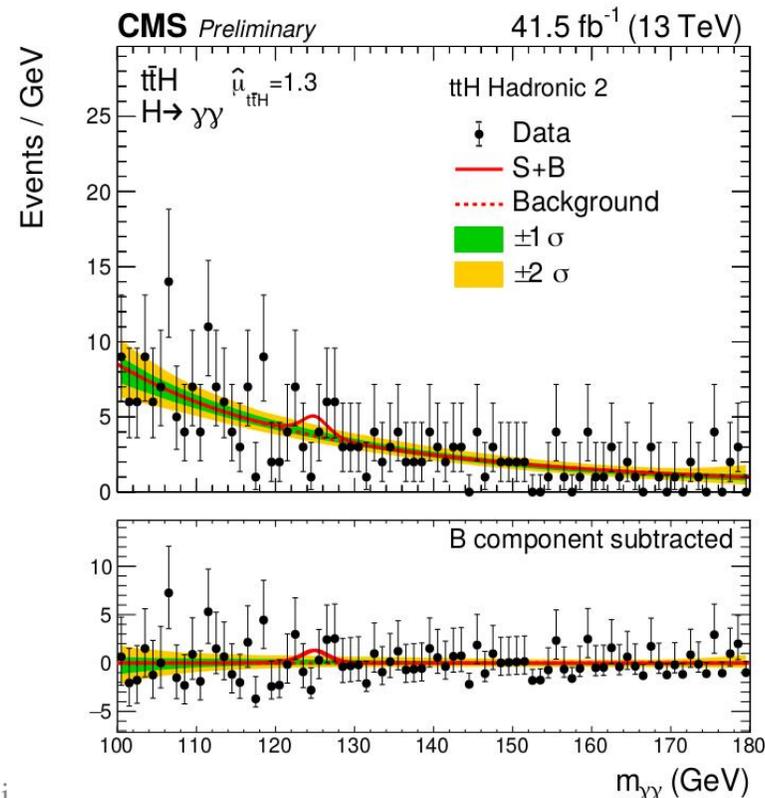
## Parametrized model of Higgs boson mass shape

- Obtained from simulation
- MC tuning and data/MC efficiency scale factors applied
- Parametric signal model with shape parameters linear functions of  $m_H$
- obtained from simultaneous fit to 120, 125, and 130 GeV mass points
- each process  $\times$  category  $\times$  right/wrong vertex treated separately



## Background model extracted from data

- Different functional forms used for each category
- Choice of function treated as a discrete nuisance parameter

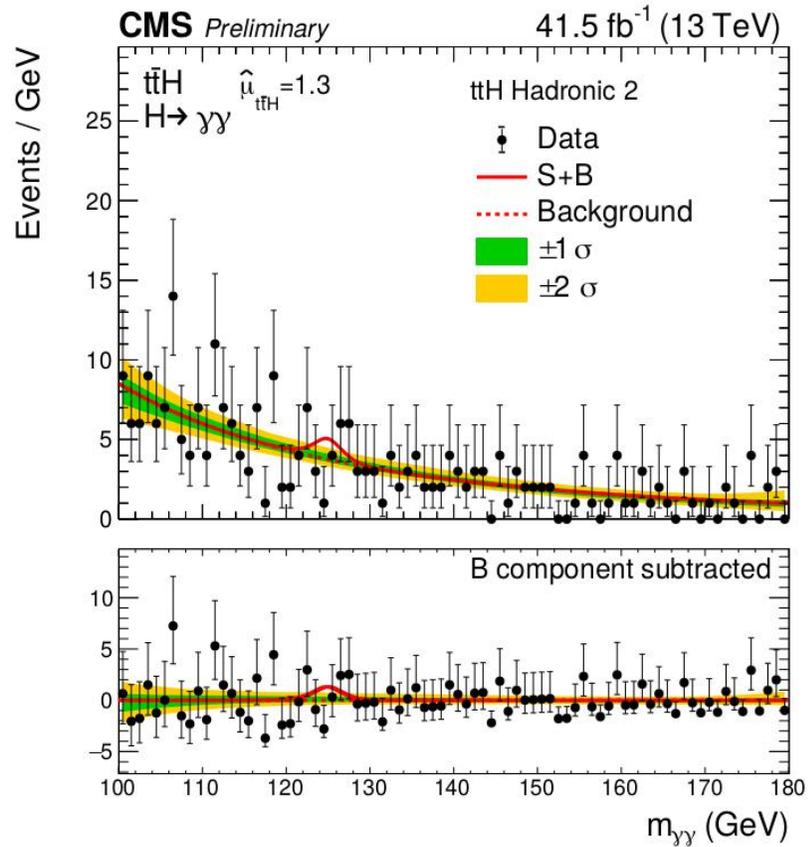


# Systematic uncertainties

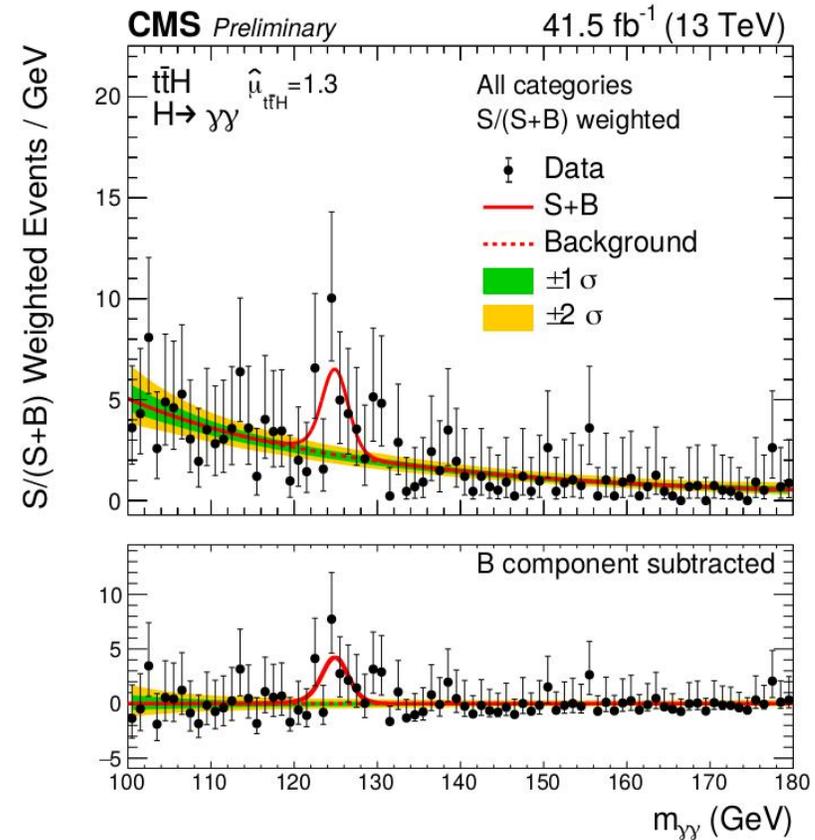
- The same strategy as CMS-PAS-HIG-16-040 is followed.
- The dominant theoretical uncertainties:
  - QCD scale: 9%
  - PDF: 5%
  - Strong coupling constant: 3%
  - $H \rightarrow \gamma\gamma$  branching fraction: 2%
  - $ggH$  contamination: 2%
- The dominant experimental uncertainties:
  - Photon identification: 6%
  - Jet energy scale resolution: 4%
  - Shape of the  $b$  discriminant: 3%
  - Integrated luminosity: 2.3%

# Mass Spectra

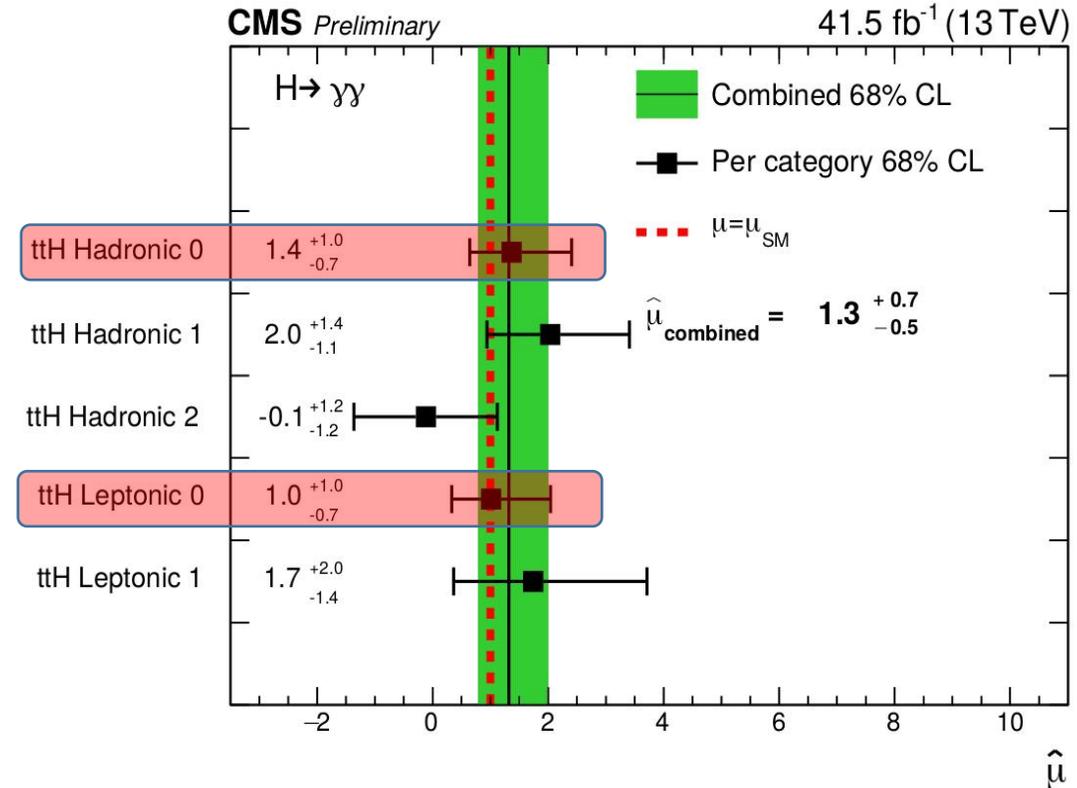
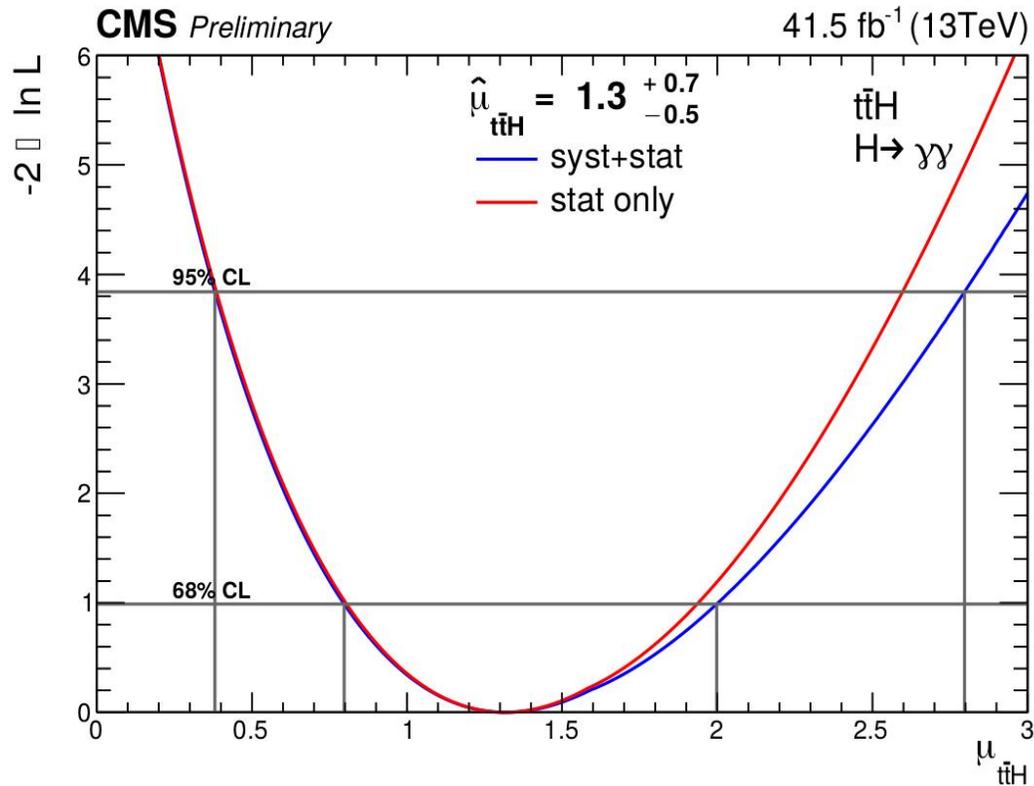
Category ttH Hadronic 2



All Categories weighted by sensitivity



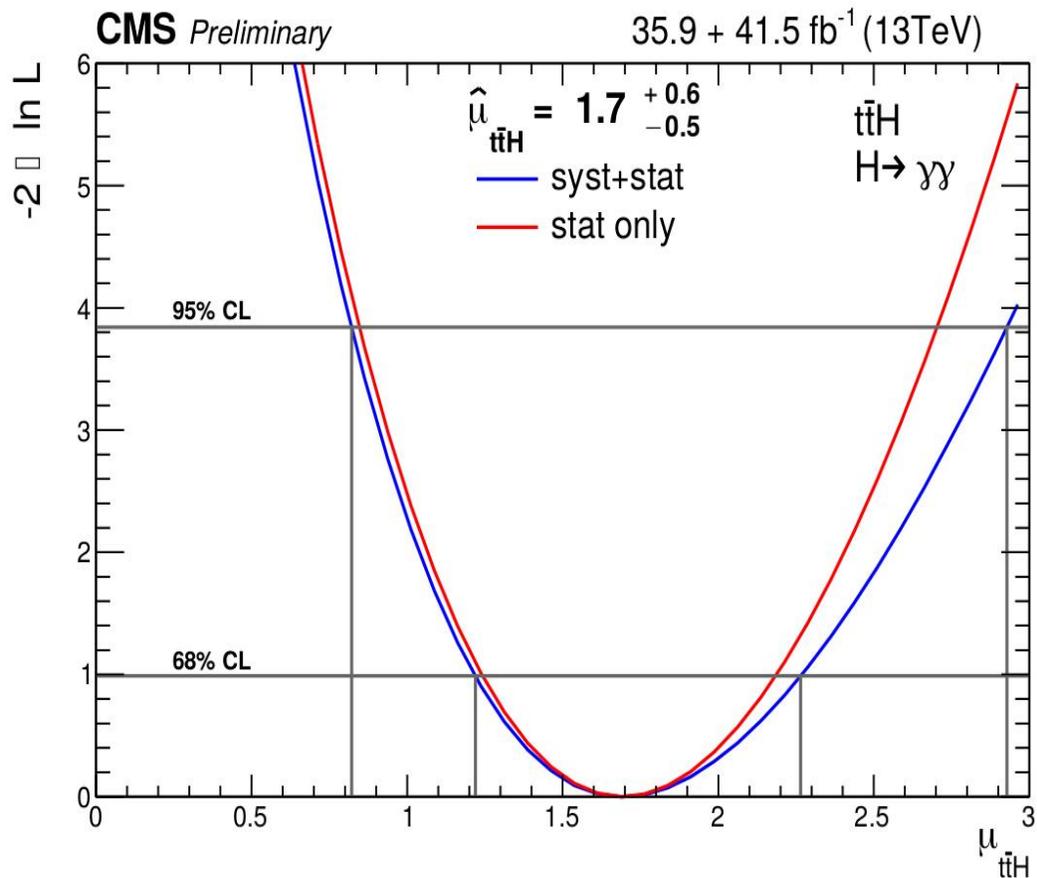
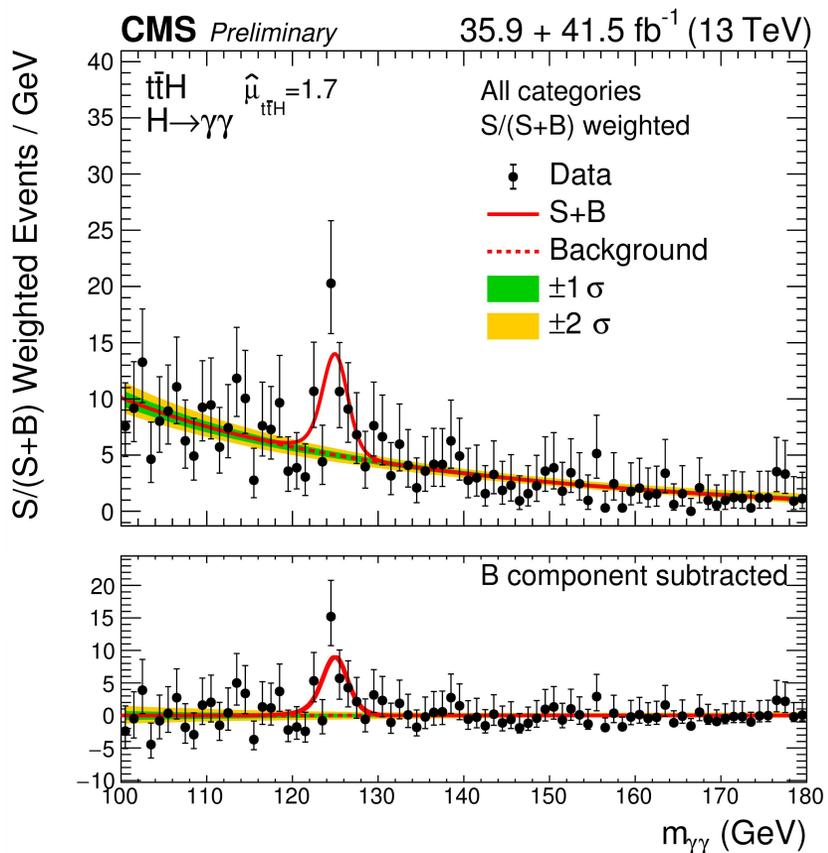
# Likelihood scans and per category strength



• The observed  $\mu_{ttH} = 1.3^{+0.7}_{-0.5} = 1.3^{+0.6}_{-0.5}(\text{stat.}) \pm^{+0.3}_{-0.1}(\text{syst.})$

- The  $\mu$  value of each category is presented.
- The **zeroth categories** (ttH Hadronic 0, ttH Leptonic 0) have the highest sensitivity.

# Combination with 2016 ttH results



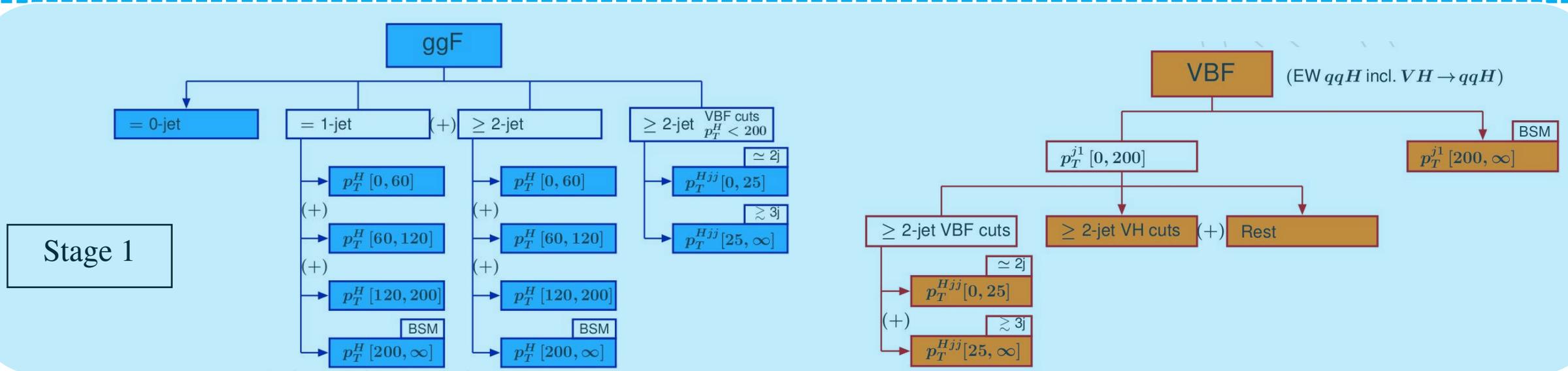
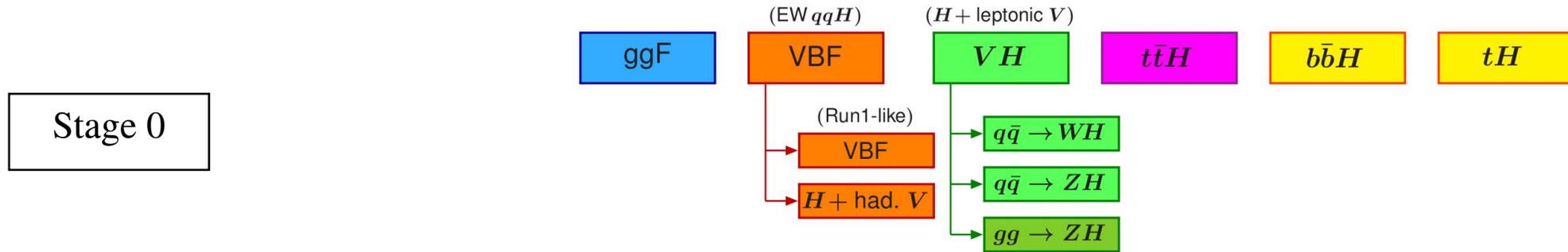
- The **diphoton mass distribution plot of 2016 and 2017 data** combined is presented with the signal and the background models.
  - Green** and **yellow** regions show the one standard and two standard deviations from the background model respectively.

- Combination of 2016 (3.3 obs., 1.5 exp.) and 2017 results
 
$$\mu_{t\bar{t}H} = 1.7^{+0.6}_{-0.5} = 1.7^{+0.5}_{-0.4} \text{ (stat.) } \pm^{+0.3}_{-0.1} \text{ (syst.)}$$
- Combined expected significance: 2.7 σ ,observed significance: 4.1 σ.

# *STXS analysis*

# STXS Framework

- Simplified Template Cross Sections:
  - The measured cross-sections per Higgs boson production mode in different regions of phase space



# Introduction

- This analysis is based on  $35.9 \text{ fb}^{-1}$  2016 +  $41.5 \text{ fb}^{-1}$  2017 datasets:
  - 2016 ingredients very similar to 2016 paper: CMS-PAS-HIG-16-040) (JHEP 11 (2018) 185)
  - 2017 ingredients very similar to the recent ttH analysis (CMS-PAS-HIG-18-018)
- Documentation:
  - *CMS-PAS-HIG-18-029*

Available on the CERN CDS information server

CMS PAS HIG-18-029

---

## CMS Physics Analysis Summary

---

Contact: cms-pag-conveners-higgs@cern.ch

Measurements of Higgs boson production via gluon fusion and vector boson fusion in the diphoton decay channel at  $\sqrt{s} = 13 \text{ TeV}$

The CMS Collaboration

# Analysis Strategy

- The analysis is targeting the ggH and VBF phase space regions only
  - no sensitivity to the stage 1 VH bins, and ttH is not split at stage 1
  - therefore **do not include ttH or VH categories** in this analysis
  - the excluded processes are constrained to the SM prediction.
- Techniques similar to in the past: define categories targeting (now more finely binned) production modes, reject background using BDTs
- ggH categories are defined with cuts on the equivalent reconstructed quantities of the defining generator level variables
  - $p_T(\gamma\gamma)$  very well-measured
  - background rejection using the same **diphoton BDT** as in previous versions of the analysis
- VBF categorisation based on the **dijet BDT**
  - Improved - **data-driven** replacement for backgrounds with low MC statistics
  - Final categories defined using cuts on **dijet BDT and diphoton BDT**

# Categorisation

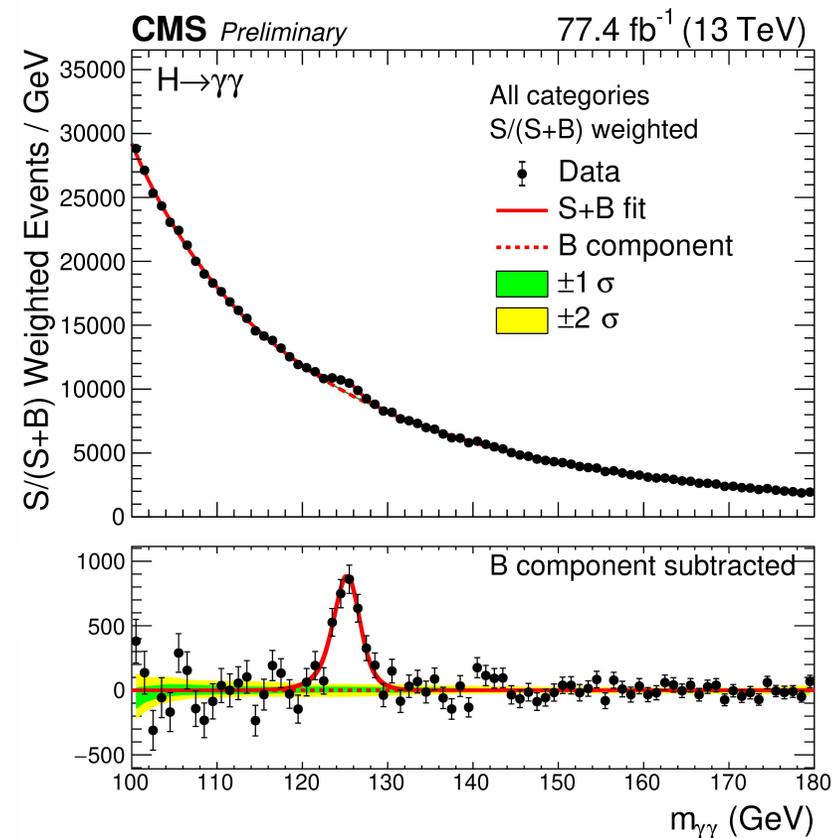
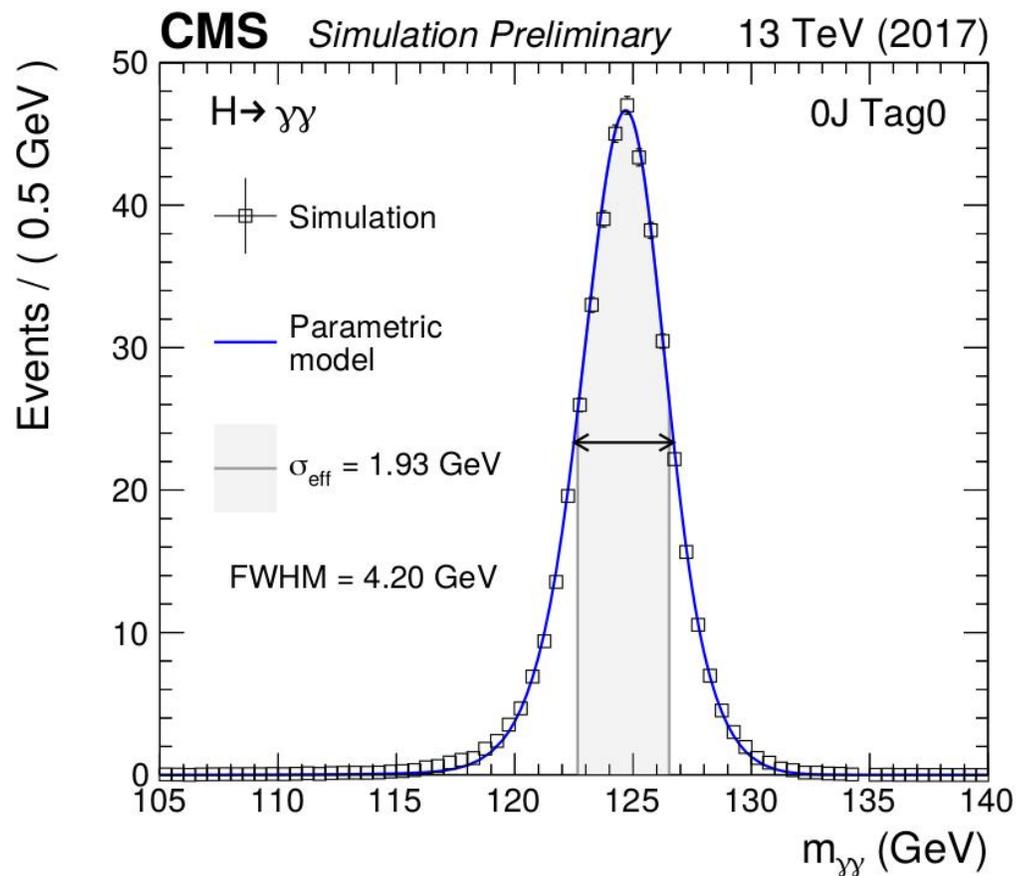
- ggH categorisation:
  - first a target bin is assigned based on the reconstructed  $p_T$  (H) and nJets
  - then the diphoton BDT boundaries are chosen independently for each bin
- VBF categorisation:
  - 2016 analysis: used “**combined BDT**”, having the diphoton BDT and dijet BDT as inputs, then the diphoton BDT boundaries are chosen independently for each bin
  - 2017 analysis: used **simple cuts on diphoton and dijet BDTs** instead of “combined BDT”

# Signal and Background Parametrization

- Signal Parametrization:
  - same signal modeling procedure as for  $ttH$  ( $H \rightarrow \gamma\gamma$ ) analysis

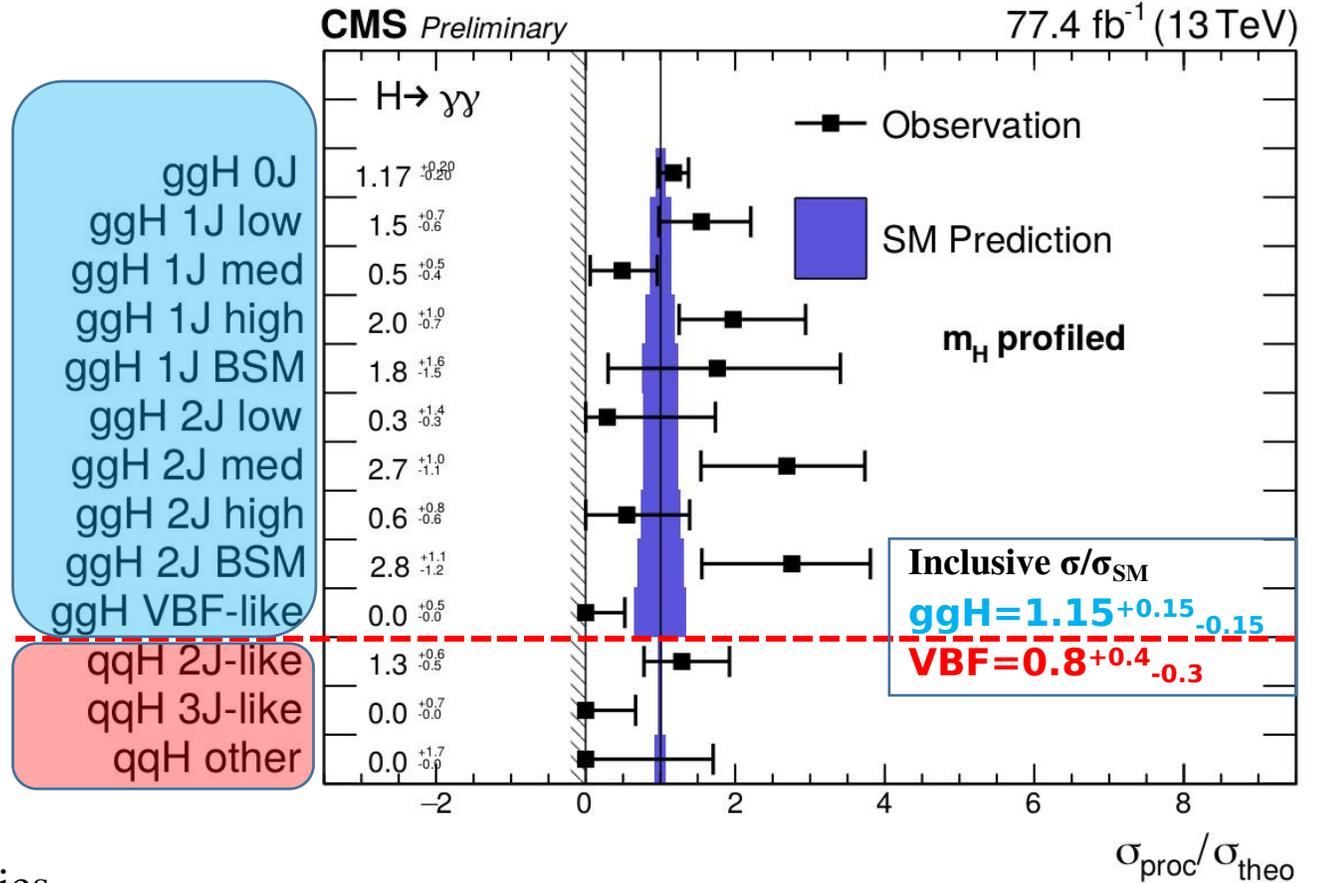
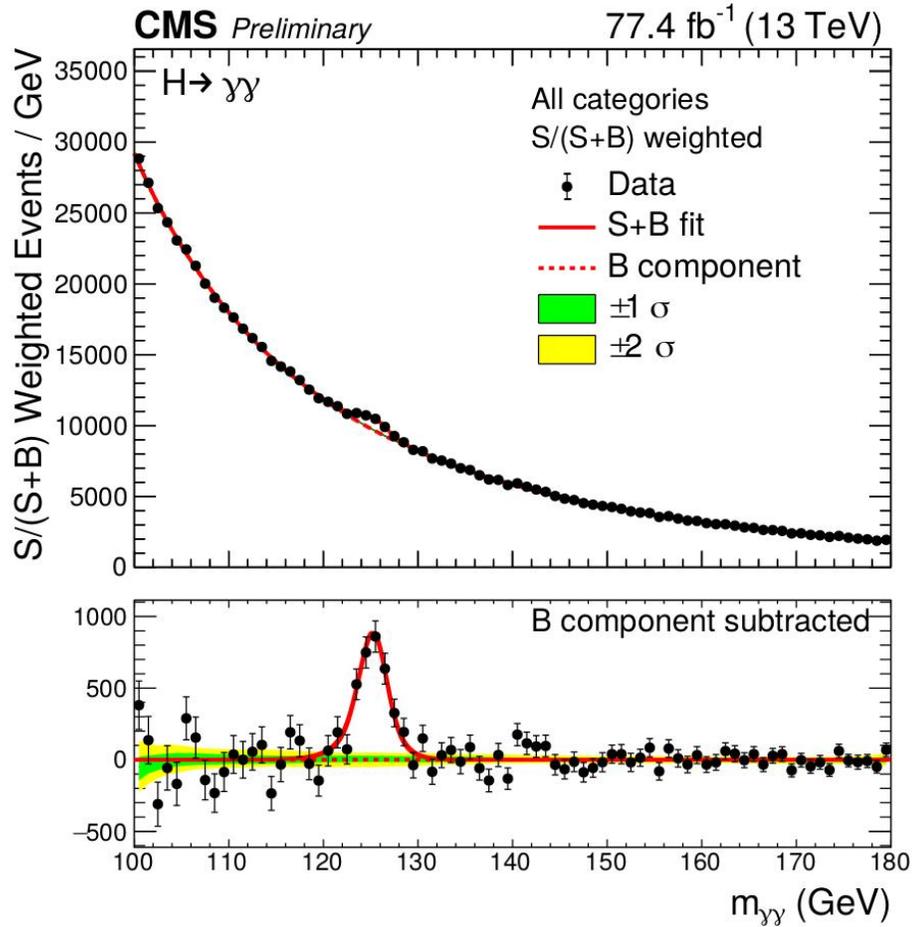
- Background Parametrization:

- Data-driven background model using the envelope method
- Treat choice of background function as discrete nuisance parameter



# Results

★ Combination of 2016 and 2017 datasets



- Mass spectrum showing the weighted sum of all categories

# Conclusions

- The ttH ( $H \rightarrow \gamma\gamma$ ) analysis based on  $41.5 \text{ fb}^{-1}$  proton-proton collision data collected in 2017 at CMS is presented
  - The results are combined with ttH ( $H \rightarrow \gamma\gamma$ ) 2016 published results:  
CMS-PAS-HIG-16-040 (JHEP 11 (2018) 185)
  - best fit  $\mu_{\text{ttH}} = 1.7^{+0.6}_{-0.5}$  with observed significance of  $4.1 \sigma$  (expected  $2.7\sigma$ ).
- The measurements at stage 1 of the STXS framework at CMS have been presented:
  - ggH: Inclusive  $\sigma/\sigma_{\text{SM}} = 1.15^{+0.15}_{-0.15}$
  - VBF: Inclusive  $\sigma/\sigma_{\text{SM}} = 0.8^{+0.4}_{-0.3}$
- All results are found to be in agreement with the standard model expectations.

*Thanks for your attention!*

# *BackUp*

# ttH Analysis — Identification of other objects

- **Jets** are reconstructed using anti- $k_T$  algorithm with a radius parameter of 0.4.
  - Jets are selected by requiring tight jet ID and  $p_T > 25$  GeV in  $|\eta| < 2.4$ .
- **b-jets** are tagged using the centrally defined DeepCSV algorithm,
  - medium working point is chosen to quantify the b-jet multiplicity.
- Electrons are identified using the standard CMS electron ID with  $p_T > 10$  GeV and  $|\eta| < 2.5$ , muons are required to have  $p_T > 10$  GeV and  $|\eta| < 2.4$ .
- All leptons are required not to overlap with photons by imposing  $\Delta R > 0.2$ . Jets are also required not to overlap with photons and leptons by imposing  $\Delta R > 0.4$ .
- Data/MC scale factors are applied following the standard CMS values.

# ttH Analysis — Vertex Selection

- CMS 0th vertex selection:
  - “central CMS vertex selection algorithm”
  - saves times as no special retraining or validation is needed for the vertex selection
  - this choice has no impact on the resolution or on the background invariant mass distributions
  
- $H \rightarrow \gamma\gamma$  specific vertex:
  - vertex assignment considered as correct within 1 cm of the diphoton interaction point
  - followed by another BDT estimating the probability to correctly choose the vertex
  - this probability is then used as an input of the diphotonMVA discriminating variable

# ttH Analysis — Photon Identification

- Input variables:

Shower shape variables:  $E_{2 \times 2} / E_{5 \times 5}$ ,  $cov_{i\eta i\phi}$ ,  $\sigma_{i\eta i\eta}$ ,  $R_9$ ,  $\sigma_\eta$ ,  $\sigma_\phi$ , Preshower  $\sigma_{RR}$

Isolation variables: PF Photon ISO, PF Charged ISO (selected vertex), PF Charged ISO (worst vertex)

Other variables:  $\rho$ , Supercluster  $\eta$ , Supercluster  $E_{RAW}$

# ttH Analysis — ttH MVA

- **new ttH Leptonic MVA:**

- Input variables are selected from photons, lepton, jet, b jet and MET.

- photon variables:

- leading and subleading photon  $p_T/m_{\gamma\gamma}$ ;
- leading and subleading photon  $\eta$ ;
- leading and subleading photon identification score;
- leading and subleading photon azimuthal angle difference  $\Delta\phi(\gamma\gamma)$ ;
- leading and subleading photon pixel seed veto (PSV): requires the absence of a track seed in the pixel detector matching the photon direction. The pixel seed veto has an efficiency of 85% for signal events and of about 50% for background event, since help in rejecting events where an electron following a top quark decay is misidentified as a photon;

- jet variables:

- number of jets;
- transverse momentum of the three highest  $p_T$  jets;
- pseudorapidity of the three highest  $p_T$  jets;

- b jet variables:

- number of jets b-tagged with the medium working point of the DeepCSV b-tagger;
- b-tag score of the two jets with the largest value of the b-tag score;

- leptonic variables:

- transverse momentum of the lepton; if more than one lepton is present, the highest  $p_T$  lepton is chosen;
- pseudorapidity of the lepton; if more than one lepton is present, the highest  $p_T$  lepton is chosen;
- missing transverse momentum.

- **ttH Hadronic MVA:**

- more variables were added in order to increase the sensitivity
- input variables (30):

- the number of jets;
- the transverse momenta  $p_T$ , pseudorapidities  $\eta$  and btag scores of the four leading jets;
- the first, second and third maximum btag scores;
- the transverse momenta for both photons, rescaled for the diphoton mass,  $p_T^{1,(2)}/m_{\gamma\gamma}$ ;
- the pseudorapidities of both photons' superclusters,  $\eta^{1,(2)}$ ;
- the  $\phi$  of both photons' superclusters,  $\phi^{1,(2)}$ ;
- the presence of a seed in the pixel detector for both photons;
- the minimum and maximum values of the identification BDT score for both photons: maxPhoID, minPhoID;
- the diphoton rapidity  $Y_{\gamma\gamma}$ ;
- the diphoton transverse momentum rescaled for the diphoton mass,  $p_T^{\gamma\gamma}/m_{\gamma\gamma}$ .
- the sum of the transverse momenta of all jets;
- the missing transverse energy  $MET$ .

# ttH Analysis — Photon Preselection

- Select the events with diphoton pairs:
  - both photons in the ECAL acceptance range ( $|\eta| < 2.5$  and  $1.44 < |\eta| < 1.57$ )
  - $p_T$  of leading (subleading) photon  $> 35$  (25) GeV
  - conversion safe electron veto
  - $100 \text{ GeV} < m_{\gamma\gamma} < 180 \text{ GeV}$
- In addition, both photons are required to satisfy  $R_9 > 0.8$  or  $I_{ch} < 20 \text{ GeV}$  or  $I_{ch}/p_T < 0.3$ . The other requirements are shown in the following table:

	$R_9$	H/E	$\sigma_{\eta\eta}$	$I_{ph}$	$I_{tk}$
Barrel	$> 0.85$	$< 0.08$	-	-	-
	$[0.50, 0.85]$	$< 0.08$	$< 0.015$	$< 4.0$	$< 6.0$
Endcaps	$> 0.90$	$< 0.08$	-	-	-
	$[0.80, 0.90]$	$< 0.08$	$< 0.035$	$< 4.0$	$< 6.0$

- $\sigma_{\eta\eta}$  – the lateral extension of the shower
- $I_{ch}$  – charged-hadron isolation conversion safe electron veto
- $I_{ph}$  – photon isolation
- $I_{tk}$  – track isolation in a hollow cone

# STXS Analysis — Systematics

- Most of the uncertainties are already defined by previous analyses
- But the jet energy scale is very important for this analysis
- Previously, was implemented as multiple nuisances representing migrations between Untagged and VBF tags, and within VBF tags
  - “conservative” approach inherited from Run 1
- However jets are now also used in the ggH phase space
  - single nuisance is standard implementation → try this first
  - checked that these are not highly constrained in the fit ✓
- Otherwise, the leading experimental systematic is the photon IDMVA
- Theory systematics: do not include uncertainty on the cross-section itself
  - this differentiates the STXS measurement from a signal strength
  - still need to account for the effect on the analysis efficiency x acceptance

# STXS Analysis — DiPhoton BDT

- Most of the uncertainties are already defined by previous analyses
- Input variables:
  - the transverse momenta for both photons, rescaled for the diphoton mass,  $p_T^{1,(2)} / m_{\gamma\gamma}$ ;
  - the pseudorapidities of both photons,  $\eta^{1(2)}$ ;
  - the cosine of the angle between the two photons in the transverse plane,  $\cos(\Delta\phi)$ ;
  - the identification BDT score for both photons;
  - the per-event relative mass resolution estimate, under the hypothesis that the mass has been reconstructed using the correct primary vertex ( $\sigma_{rv}$ );
  - the per-event relative mass resolution estimate, under the hypothesis that the mass has been reconstructed using an incorrect primary vertex ( $\sigma_{wv}$ );
  - the per-event probability estimate that the correct primary vertex has been used to reconstruct the mass, based on the event-level vertex selection MVA as described in Section in 7 ( $p_{vtx}$ ).
- Framework changes from from TMVA to **XGBoost**
- Trained with all signal vs **all background**

# STXS Analysis — Dijet BDT

- Dijet BDT has always been used to discriminate between ggH and VBF
- Inputs mostly related to jet kinematics:
  - the transverse momenta of the leading and subleading photons divided by the invariant mass of the diphoton candidate:  $p_T^{\gamma_1} / m_{\gamma\gamma}$  and  $p_T^{\gamma_2} / m_{\gamma\gamma}$
  - the transverse momenta of the leading and subleading jets:  $p_T^{j_1}$  and  $p_T^{j_2}$
  - the dijet invariant mass,  $m_{j_1j_2}$
  - the difference in pseudo-rapidity between the two jets,  $\Delta\eta_{j_1j_2}$
  - the difference in azimuthal angle between the dijet and the diphoton,  $\Delta\phi_{(j_1j_2,\gamma\gamma)}$ .
  - centrality variable defined as,

$$C_{\gamma\gamma} = \exp\left(-\frac{4}{(\eta_1 - \eta_2)^2} \left(\eta_{\gamma\gamma} - \frac{\eta_1 + \eta_2}{2}\right)^2\right) \quad (7)$$

where  $\eta_1, \eta_2$ , and  $\eta_{\gamma\gamma}$  are the pseudo-rapidities of the two jets, and the diphoton.

- the difference in azimuthal angle between the two leading jets  $\Delta\phi_{jj}$
- the minimum distance between a leading or subleading jet and leading or subleading photon  $\min \Delta R(\gamma, jet)$ .