Study of Higgs boson decay to bottom quark pairs with CMS experiment at LHC

Armen Tumasyan

on behalf of CMS collaboration

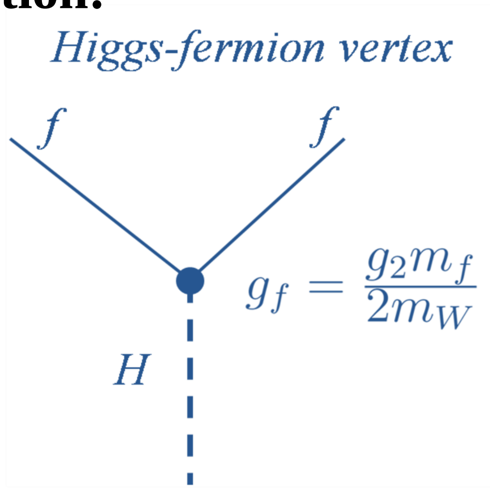
CHEP-Yerevan-2023

A.I. Alikhanyan National Science Laboratory, Yerevan, Armenia

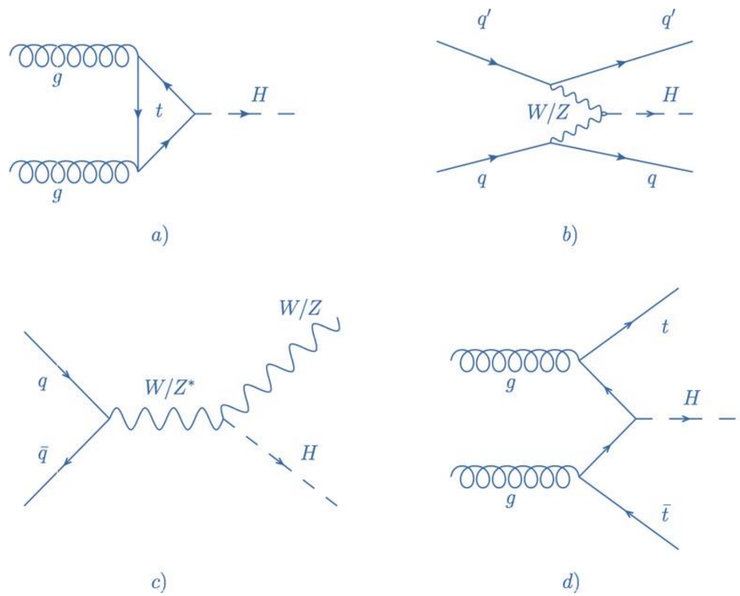
September 11-14, 2023

Introduction

Motivation:



Higgs boson production at LHC



One of the main challenges for LHC Run2 was the Higgs boson observation in 3rd generation fermions decay modes.

$H \rightarrow b\bar{b}$ features

- BR ~ 58 %
- Direct access to *H-b* coupling
- Certain difficulties with event reconstruction and signal extraction

Sensitivity of $H \rightarrow b\bar{b}$ search is highly related to “b-tagging” performance which is in constant development.

$H \rightarrow b\bar{b}$: Observation

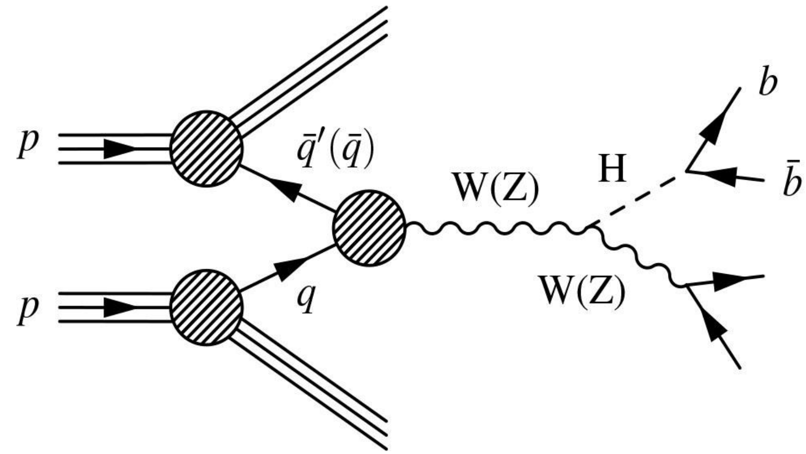
Phys.Rev.Lett. 121, 121801

Main contribution is from **VH**-production mode:
WH or **ZH**, where W/Z are decaying to leptons.

VH(bb)

Event selection and interpretation (2017)

- **0-lepton** channel (p_T^{miss} and H_T^{miss} triggers)
  p_T^{miss} interpreted as p_T^{Z}
- **1-lepton** channel (Single e/μ triggers)
  W-candidate reconstructed from $\boldsymbol{p}_T^{miss}$ and $\boldsymbol{p}_l$
- **2-lepton** channel (Double e/μ triggers)
  Z-candidate reconstructed from leptons pai



- System of two most b-tagged (DeepCSV) jets interpreted as H-candidate.
- Leptons are required to pass isolation condition.
- Jets are required to be within |η| < 2.5, and with p_T > 65,35 (0); 25,25 (1); 20,20 (2) GeV
- Tight/medium b-tagging cuts are applied to higher/lower b-tagged jets, providing 0.1 / 1.0 % missID rate for light-flavour jets and 68/50 % efficiency for b-jets 2018 JINST 13 P05011
- Additional cuts on p_T^{miss}, p_T^{V}, jet-multiplicity in different categories are applied to reduce background contribution

$H \to b\bar{b}$: VH

Following technique was applied to reduce mass (m_{jj}) resolution (from 15% to 10-13%):

- DNN regression and FSR recovery
- In 2-lepton channel p_T-balance between *ll* and *jj* systems was required

DNN classification was applied for signal-background discrimination.

Control regions are used to normalize background processes directly from data.

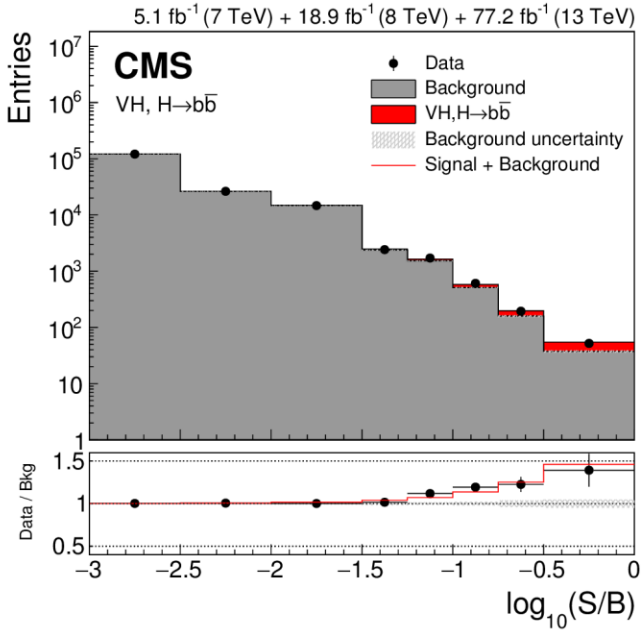
DNN score is used in signal regions to extract signal strength.

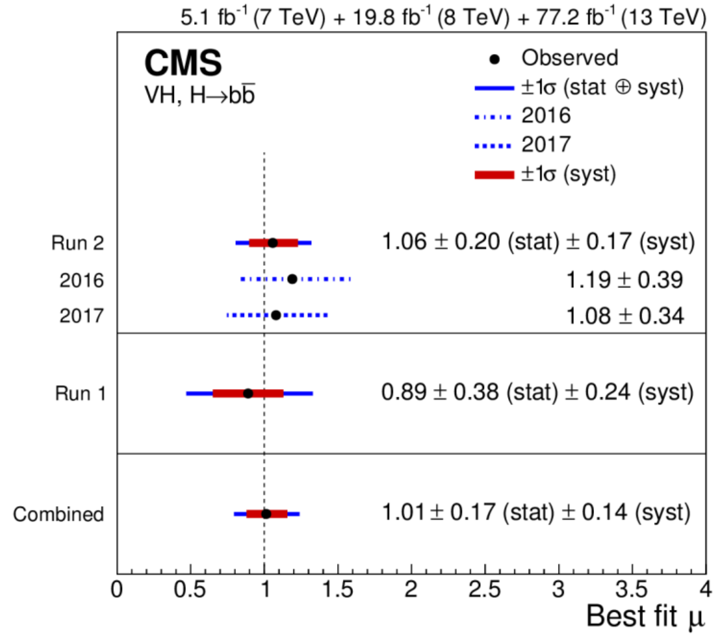
[Phys.Rev.Lett. 121, 121801](#)

Validation on VZ (WZ and ZZ) process

2017 results: observed/expected significance = 3.3 / 3.1 σ, μ=1.08 ±0.34

Combined Run1 and 2016-17 results: observed/expected significance = 4.8 / 4.9 σ, μ=1.01 ±0.22





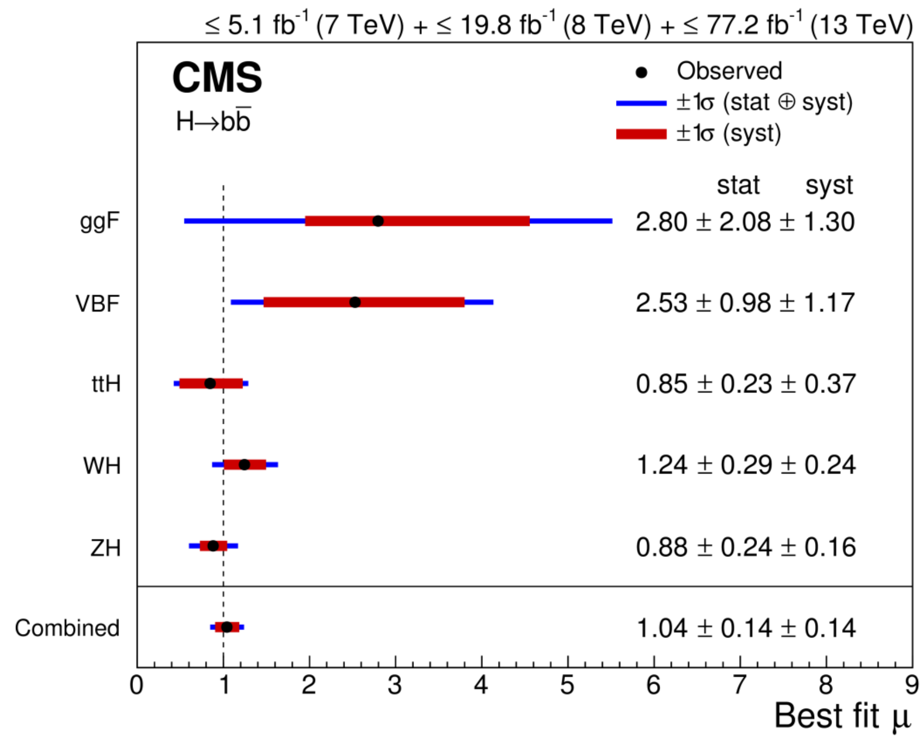
$H \to b\bar{b}$: Observation

Combination of CMS measurements of the $H \to b\bar{b}$ decay including

- VH Phys.Rev.Lett. 121, 121801 Phys.Lett.B 780 (2018) 501–532
- gluon fusion Phys. Rev. Lett. 120, 071802
- vector boson fusion Phys. Rev. D 92 (2015) 032008
- associated production with top quarks Phys. Rev. Lett. 120, 231801

observed/expected significance = 5.6 / 5.5 σ

μ=1.04 ±0.20

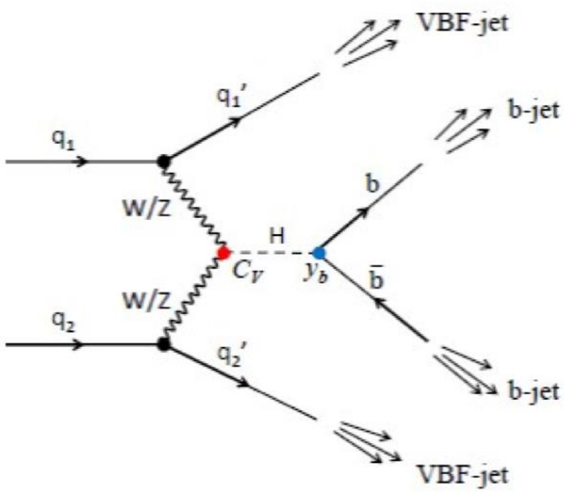


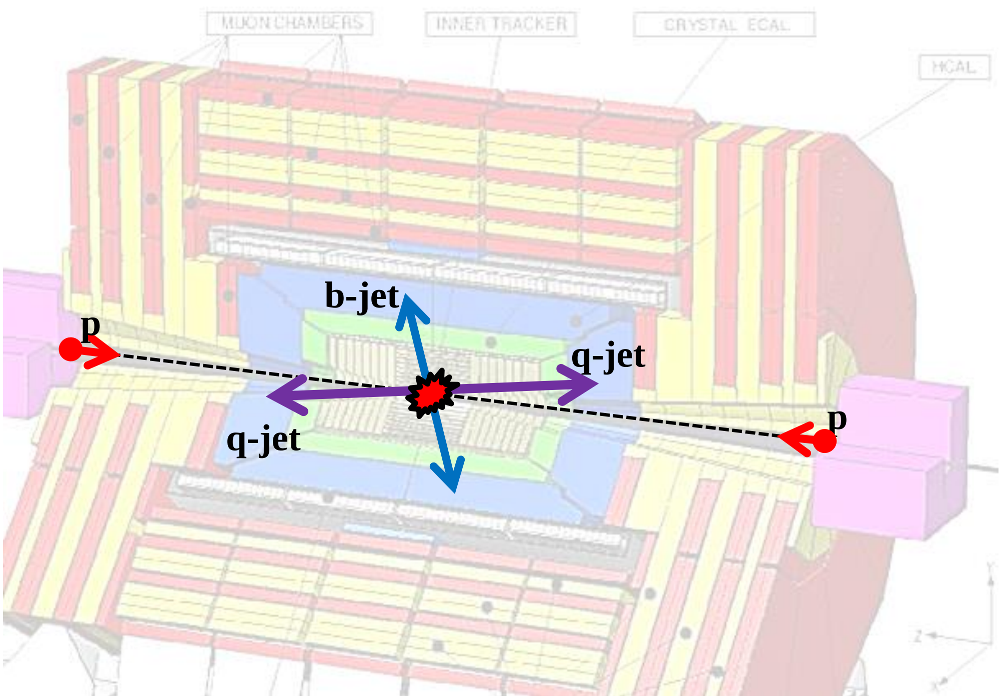
$H \to b\bar{b}$: VBF

CMS 2016 and 2018 data of pp-collisions at 13 TeV corresponding to ~91 fb^{-1} integral luminosity

Two main analysis classes based on two main features of VBF_Hbb process:

- **SingleB** relies on **tight** VBF topology and **loose** b-tagging (≥1 b-tagged jets) CMS-PAS-HIG-22-009
- **DoubleB** relies on **loose** VBF topology and **tight** b-tagging (≥2 b-tagged jets)





Dedicated online triggers (L1 – HLT)

L1: 3Jets (SingleB & DoubleB)

HLT: 4Jets-LooseBTag-TightVBF (SingleB)

4Jets-TightBTag-LooseVBF (DoubleB)

Dominant QCD background

$H \to b\bar{b}$: VBF

Event selection and interpretation

- Four P_T -leading jets with $p_T >$ 95, 80, 65, 30 GeV were considered **(2016)**

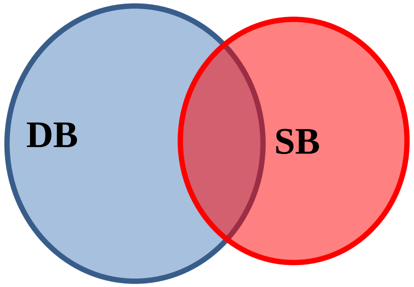
  Four P_T -leading jets with $p_T >$ 110, 90, 80, 30 GeV were considered **(2018)**
- jets are within $|\eta| < 4.7$ and passing *loose* PileUp-condition
- 2 most b-tagged jets with $|\eta| < 2.4$ and passing DeepJet medium WP were selected as b-jets (b1 & b2).

  Here medium WP cut provides 70-80 % b-jet selection efficiency vs ~1% light-flavour missID rate.

CMS-DP-2018-058

- $\Delta\varphi_{bb}$ <1.6 (2.1) in SingleB (DoubleB)
- 2 remaining among 4 p_T -leading jets were selected as q-jets
- $m_{qq} > 500$ (250) GeV, $\Delta\eta_{qq} > 4.2$ (2.5) in SingleB (DoubleB) **(2016)**

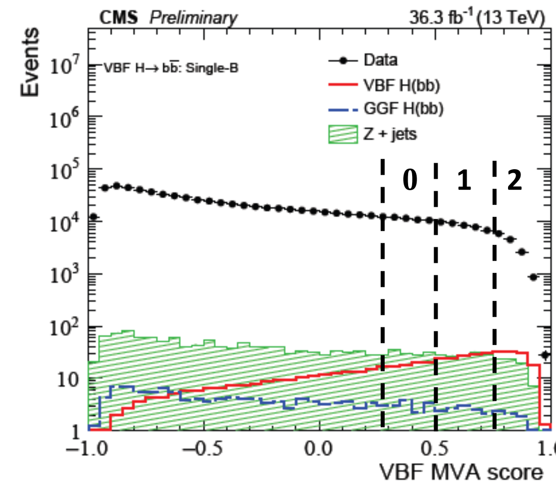
  $m_{qq} > 500$ (250) GeV, $\Delta\eta_{qq} > 3.8$ (2.5) in SingleB (DoubleB) **(2018)**
- Isolated lepton veto: **NO** *e* with $p_T >$ 7 GeV or μ with $p_T >$5 GeV

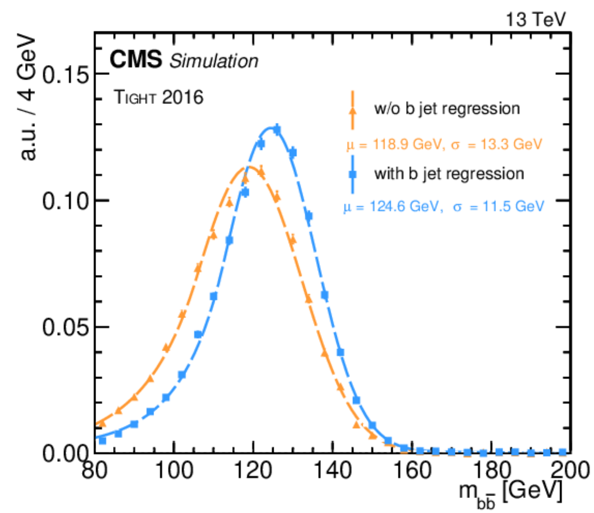


Inclusive SingleB and exclusive DoubleB selections

$H \rightarrow b\bar{b}$: VBF

- **BDT signal-background discrimination**
  - ❑ Binary classification for SingleB
  - ❑ Multiclass classification for DoubleB
- **Zbb category in DoubleB class was used for validation and Z+Jets constrain**
- **Signal and resonant background (Z+Jets) models were extracted from MC (CB + Bernstein pol.)**
- **Data derived QCD model extracted from sideband regions ($80 < m_{bb} < 104$ GeV and $140 < m_{bb} < 200$ GeV). (Exp × Pol.)**
- **Additionally DNN regression was applied to improve mass resolution (~ 15 %)** [Comput. Softw. Big Sci. 4 (2020) 10]
- **Results were extracted from m_{bb} distribution**





$H \to b\bar{b}$: VBF

Results:

- **Inclusive $H \to b\bar{b}$ (VBF+ggF):**

  Signal significance (observed/expected) = 2.6 / 2.9

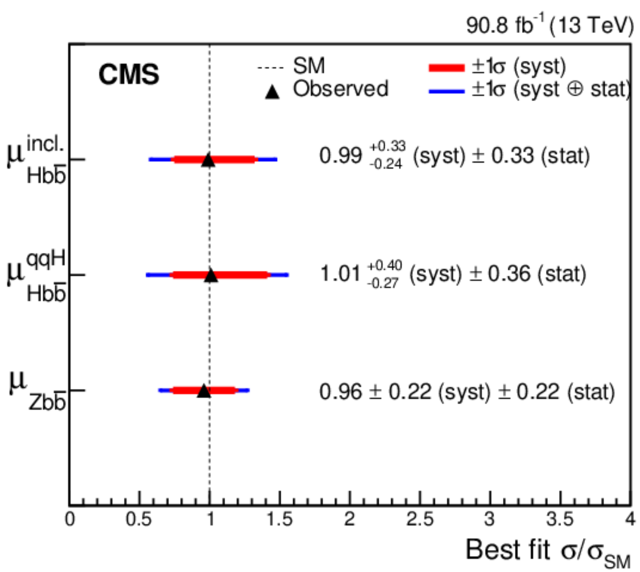
  μ_{Hbb} = 0.99 ± 0.33 (stat.) +0.33 (syst.) – 0.24 (syst.)

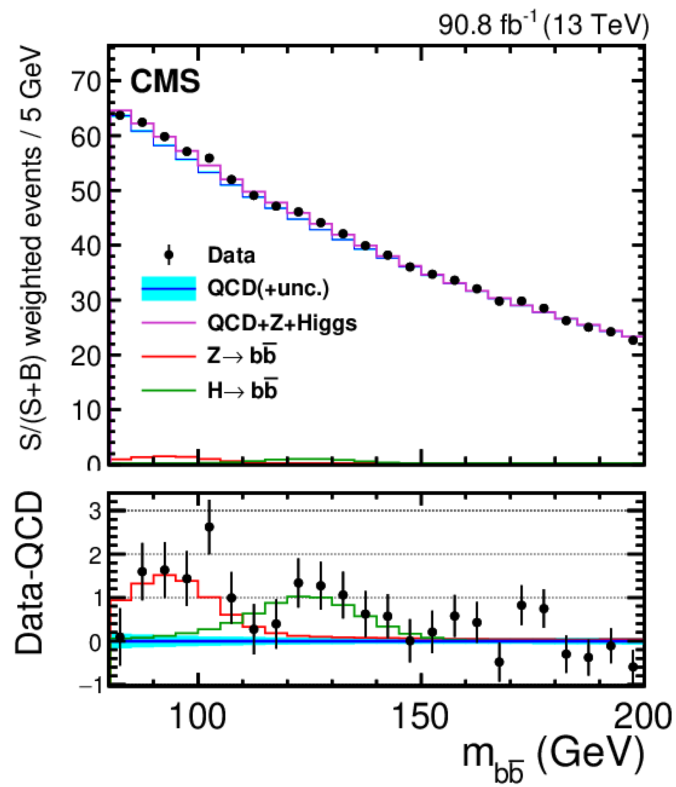
- **Pure VBF** (ggF is constrained within SM-expectation):

  Signal significance (observed/expected) = 2.4 / 2.7

  μ_{Hbb} = 1.01 ± 0.36 (stat.) +0.40 (syst.) – 0.27 (syst.)

CMS-PAS-HIG-22-009



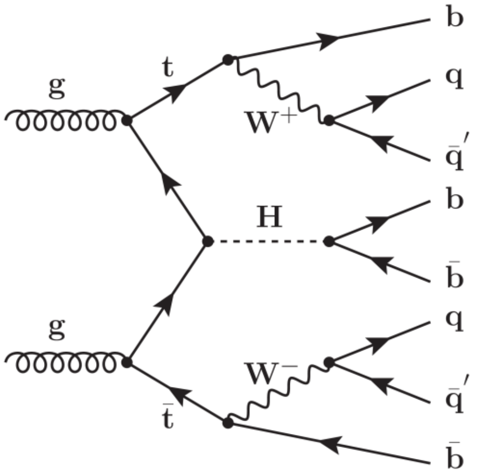


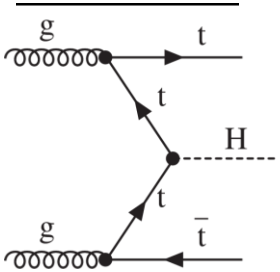
$H \to b\bar{b}$: $t\bar{t}$H all-jet (2016)

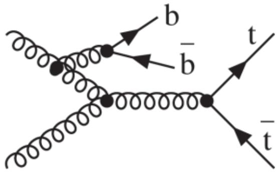
Event (pre)selection and interpretation

JHEP 06 (2018) 101

- Dedicated 6-jet HLTs with b-tag and HT requirements
- $\geq$ 6 jets with $p_T > 40$ GeV and $H_T > 450$ GeV
- $\geq$ 2 medium CSV b-tagged jets, providing ~1% and ~10 % missID rate for light-flavour and c-jets, and ~ 65% efficiency for b-jets   2013 JINST 8 P04013
- Lepton veto
- Categorization by jet (7,8, $\geq$ 9) and b-jet (3, $\geq$ 4) multiplicity
- W-candidates are reconstructed from not b-tagged jets (QGLR > 0.5)
- Background models:
  - ❑ QCD-multijet – extracted from data in control region.
    CR $\rightarrow$ 2 medium b-tagged + $\geq$ 1 loose b-tagged jets
    Validation region $\rightarrow$ QGLR < 0.5
  - ❑ other – MC
- Signal-background discrimination with likelihood technique based on the LO matrix elements using full kinematics: <u>*ttH* vs *tt+bb*</u>





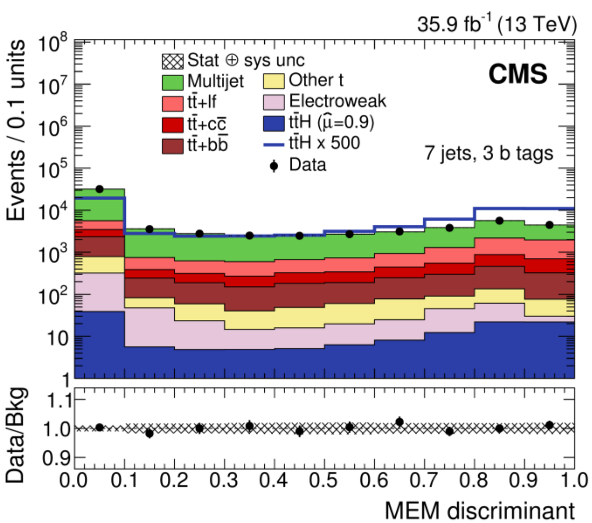


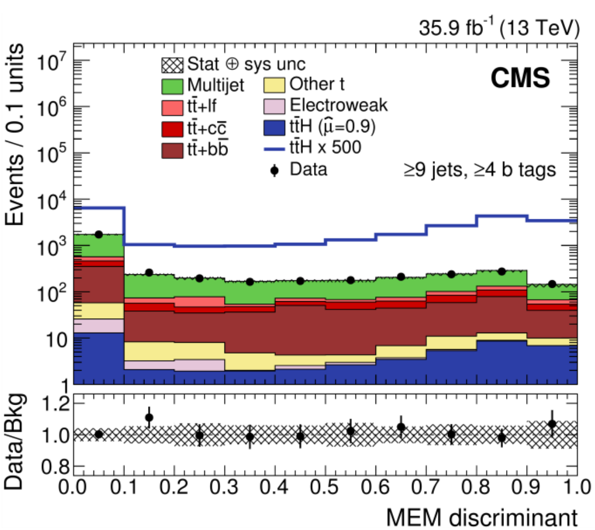
$H \to b\bar{b}$: $t\bar{t}$H all-jet (2016)

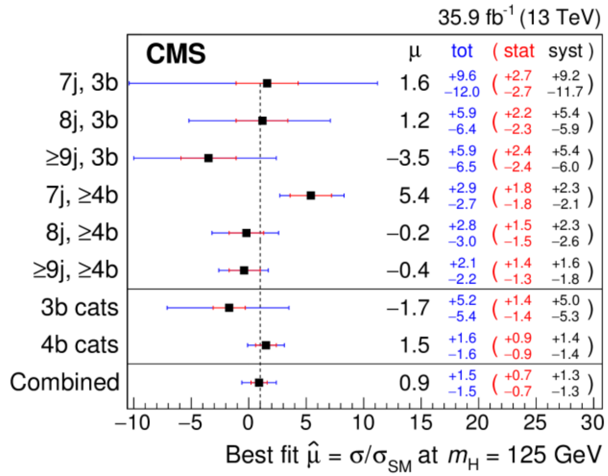
Results are extracted from (matrix element method) MEM distribution

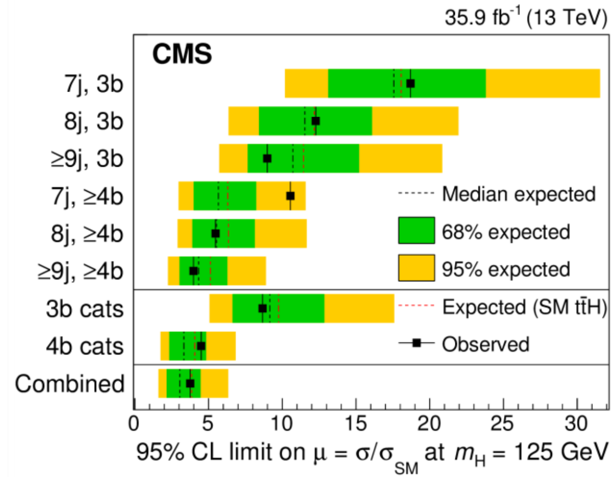
[JHEP 06 (2018) 101](#)

Resutls







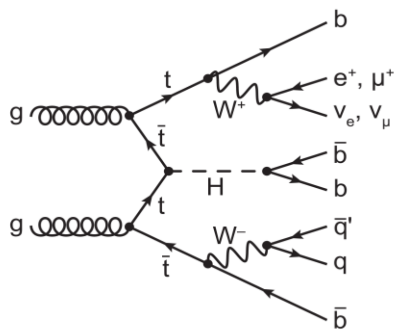


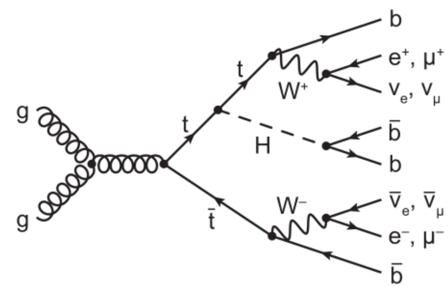
Category	Best fit $\hat{\mu}$ and uncertainty: $\hat{\mu}$	total	(stat syst)	Observed UL	Expected UL
7j, 3b	1.6	$^{+9.6}_{-12.0}$	$\left(^{+2.7}_{-2.7}\ ^{+9.2}_{-11.7}\right)$	18.7	$17.6^{+6.2}_{-4.4}$
8j, 3b	1.2	$^{+5.9}_{-6.4}$	$\left(^{+2.2}_{-2.3}\ ^{+5.4}_{-5.9}\right)$	12.3	$11.5^{+4.6}_{-3.1}$
≥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	$3.1^{+1.4}_{-0.9}$

$H \to b\bar{b}$: $t\bar{t}$H lepton(s)+jets (2016)

Two modes: Single-lepton and Double-lepton

JHEP 03 (2019) 026





- Single HLTs
- One lepton (e/μ) with p_T > 30/26 GeV, |η|<2.1
- Additional lepton veto
- p_T^{miss} > 20 GeV
- Jets with p_T > 30 GeV within |η|<2.4
- ≥ 1 medium CSVv2 b-tagged jets

- Single lepton or Double lepton HLTs
- Two opposite charged leptons (e/μ) with p_T > 25/15 GeV, |η|<2.4, m_{ll} > 20 GeV, if same flavour m_{ll}-region 76-106 GeV is rejected
- Additional lepton veto
- p_T^{miss} > 40 GeV
- 2 p_T-leading jets with p_T > 30 GeV, |η|<2.4, other jets p_T > 20 GeV
- ≥ 2 medium CSVv2 b-tagged jets

QCD-multijet background estimated from low p_T^{miss} region is negligible

Signal and background models are extracted from MC

Categorization by jet and b-jet multiplicity

- 4 jets, ≥ 3 b tags
- 5 jets, ≥ 3 b tags
- ≥ 6 jets, ≥ 3 b tags

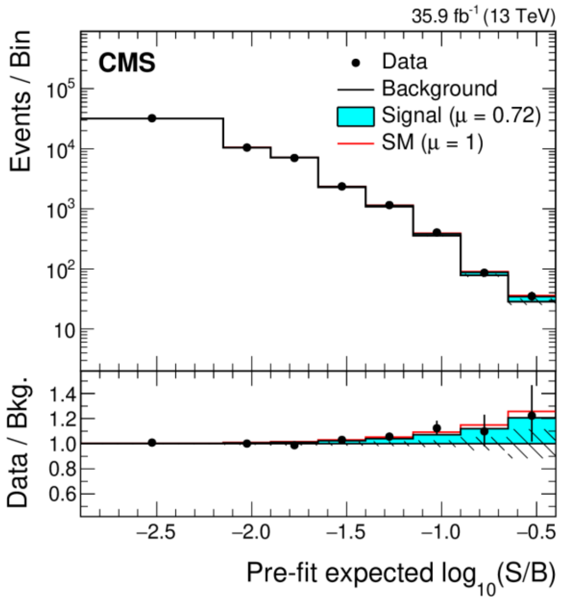
- 4 jets, ≥ 3 b tags
- 5 jets, ≥ 3 b tags

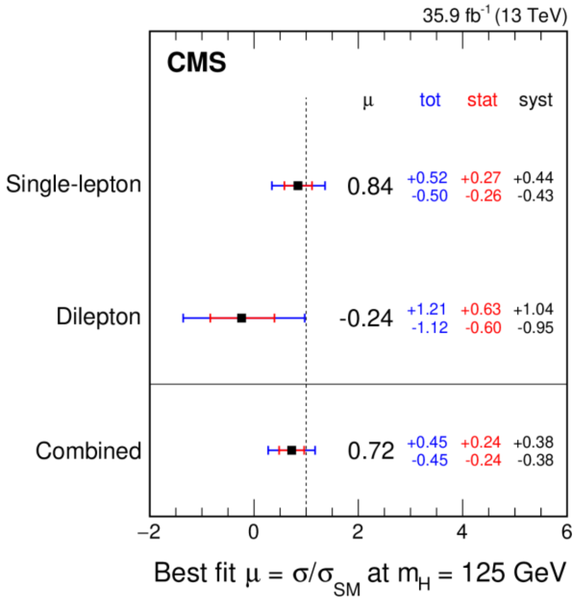
$H \to b\bar{b}$: $t\bar{t}$H lepton(s)+jets (2016)

JHEP 03 (2019) 026

- DNN, MEM signal-background discrimination in Single lepton mode: *tt*H vs *tt*+jets
- BDT signal-background discrimination in Double lepton mode: *tt*H vs *tt*+jets

Results are extracted from discriminants distribution



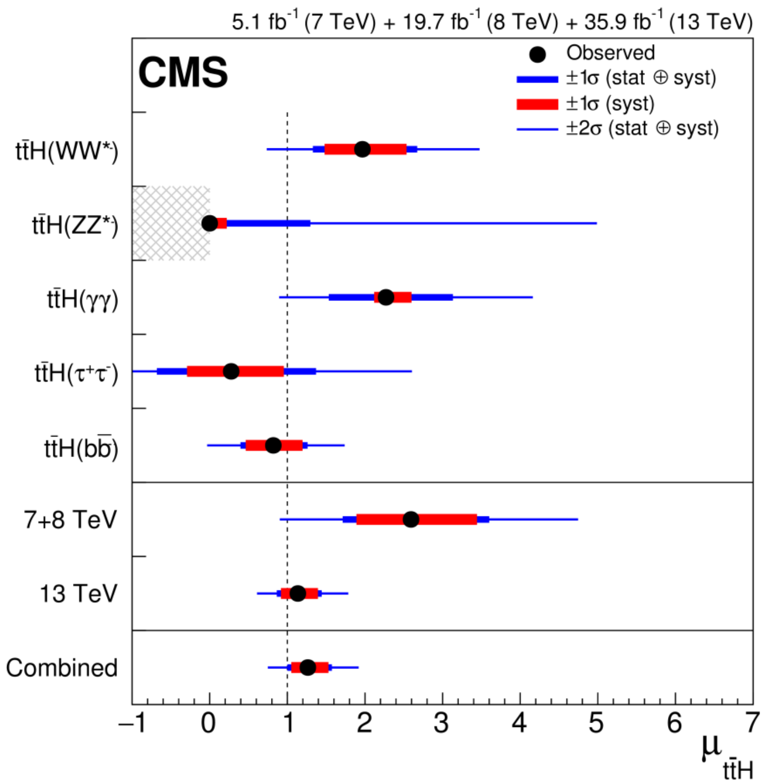


- **Observed (expected) UL on the *tt*H cross section relative to the SM expectations found 1.5 (0.9) at 95% CL**
- **The best fit value of μ is 0.72 ± 0.24 (stat) ± 0.38 (syst).**
- **Observed (expected) significance of 1.6 (2.2) σ above the background-only hypothesis.**

$H \to b\bar{b}$: $t\bar{t}$H combined

[Phys. Rev. Lett. 120, 231801](#)

- Combination of Run1 and 2016 data
- Significant contribution in *tt*H observation



Parameter	Best fit	Uncertainty: Stat	Expt	Thbgd	Thsig
$\mu^{WW^*}_{t\bar{t}H}$	$1.97^{+0.71}_{-0.64}$	$^{+0.42}_{-0.41}$	$^{+0.46}_{-0.42}$	$^{+0.21}_{-0.21}$	$^{+0.25}_{-0.12}$
	$\left(^{+0.57}_{-0.54}\right)$	$\left(^{+0.39}_{-0.38}\right)$	$\left(^{+0.36}_{-0.34}\right)$	$\left(^{+0.17}_{-0.17}\right)$	$\left(^{+0.12}_{-0.03}\right)$
$\mu^{ZZ^*}_{t\bar{t}H}$	$0.00^{+1.30}_{-0.00}$	$^{+1.28}_{-0.00}$	$^{+0.20}_{-0.00}$	$^{+0.04}_{-0.00}$	$^{+0.09}_{-0.00}$
	$\left(^{+2.89}_{-0.99}\right)$	$\left(^{+2.82}_{-0.99}\right)$	$\left(^{+0.51}_{-0.00}\right)$	$\left(^{+0.15}_{-0.00}\right)$	$\left(^{+0.27}_{-0.00}\right)$
$\mu^{\gamma\gamma}_{t\bar{t}H}$	$2.27^{+0.86}_{-0.74}$	$^{+0.80}_{-0.72}$	$^{+0.15}_{-0.09}$	$^{+0.02}_{-0.01}$	$^{+0.29}_{-0.13}$
	$\left(^{+0.73}_{-0.64}\right)$	$\left(^{+0.71}_{-0.64}\right)$	$\left(^{+0.09}_{-0.04}\right)$	$\left(^{+0.01}_{-0.00}\right)$	$\left(^{+0.13}_{-0.05}\right)$
$\mu^{\tau^+\tau^-}_{t\bar{t}H}$	$0.28^{+1.09}_{-0.96}$	$^{+0.86}_{-0.77}$	$^{+0.64}_{-0.53}$	$^{+0.10}_{-0.09}$	$^{+0.20}_{-0.19}$
	$\left(^{+1.00}_{-0.89}\right)$	$\left(^{+0.83}_{-0.76}\right)$	$\left(^{+0.54}_{-0.47}\right)$	$\left(^{+0.09}_{-0.08}\right)$	$\left(^{+0.14}_{-0.01}\right)$
$\mu^{b\bar{b}}_{t\bar{t}H}$	$0.82^{+0.44}_{-0.42}$	$^{+0.23}_{-0.23}$	$^{+0.24}_{-0.23}$	$^{+0.27}_{-0.27}$	$^{+0.11}_{-0.03}$
	$\left(^{+0.44}_{-0.42}\right)$	$\left(^{+0.23}_{-0.22}\right)$	$\left(^{+0.24}_{-0.23}\right)$	$\left(^{+0.26}_{-0.27}\right)$	$\left(^{+0.11}_{-0.04}\right)$
$\mu^{7+8\,\mathrm{TeV}}_{t\bar{t}H}$	$2.59^{+1.01}_{-0.88}$	$^{+0.54}_{-0.53}$	$^{+0.53}_{-0.49}$	$^{+0.55}_{-0.49}$	$^{+0.37}_{-0.13}$
	$\left(^{+0.87}_{-0.79}\right)$	$\left(^{+0.51}_{-0.49}\right)$	$\left(^{+0.48}_{-0.44}\right)$	$\left(^{+0.50}_{-0.44}\right)$	$\left(^{+0.14}_{-0.02}\right)$
$\mu^{13\,\mathrm{TeV}}_{t\bar{t}H}$	$1.14^{+0.31}_{-0.27}$	$^{+0.17}_{-0.16}$	$^{+0.17}_{-0.17}$	$^{+0.13}_{-0.12}$	$^{+0.14}_{-0.06}$
	$\left(^{+0.29}_{-0.26}\right)$	$\left(^{+0.16}_{-0.16}\right)$	$\left(^{+0.17}_{-0.16}\right)$	$\left(^{+0.13}_{-0.12}\right)$	$\left(^{+0.11}_{-0.05}\right)$
$\mu_{t\bar{t}H}$	$1.26^{+0.31}_{-0.26}$	$^{+0.16}_{-0.16}$	$^{+0.17}_{-0.15}$	$^{+0.14}_{-0.13}$	$^{+0.15}_{-0.07}$
	$\left(^{+0.28}_{-0.25}\right)$	$\left(^{+0.15}_{-0.15}\right)$	$\left(^{+0.16}_{-0.15}\right)$	$\left(^{+0.13}_{-0.12}\right)$	$\left(^{+0.11}_{-0.05}\right)$

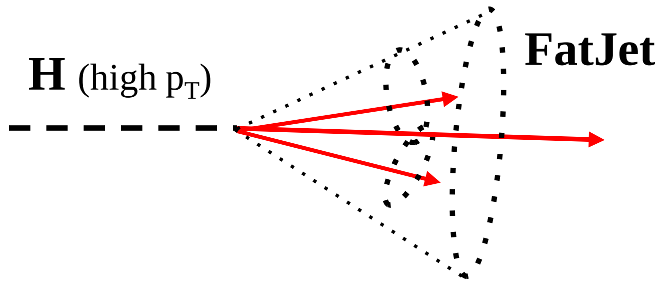
Signal strength: μ_{Hbb} = 0.82 (+0.44 – 0.42) at 68 % CL

Signal strength: μtt_H = 1.26 (+0.31 – 0.26) at 68 % CL

(*tt*H) Observed/expected significance = 5.2/4.2 σ

$H \to b\bar{b}$: ggF boosted

Higgs boson is reconstructed as FatJet (AK08) including products of b and $\bar{b}$:



[Phys. Rev. Lett. 120, 071802](#) **2016**

[JHEP 12 (2020) 085](#) **Full Run2**

Deep double b-tagging for boosted double quark Jets:

[2018 JINST 13 P05011](#) [CMS-DP-2018-046](#)

Event (pre)selection and interpretation

- **Triggers: H_T ($|\eta|<3.0$) and FatJet with TrimmMass requirement removing soft radiation contribution in Jet mass calculation**
- **p_T-leading FatJet is interpreted as "H-jet".**
- **FatJet $p_T > 450$ (525, 500) Gev for 2016 (2017,2018), $|\eta| < 2.5$, SD-requirement, two-prong structure**
- **Lepton veto to reduce EWK background**
- **p_T^{miss} veto (<140) and opposite AK4 b-jet veto to reduce *tt* contribution**
- **FatJet deep double b-tagger defines (medium WP cut) signal enriched (DDBT passing) and control (DDBT failing) regions (**For $H \to b\bar{b}$ signal processes in passing region dominant production mode is ggF (56%), followed by VBF (26%), VH (13%), and ttH (5%)**).**

M_{SD} 47-201 GeV and p_T 450-1200 GeV region is considered

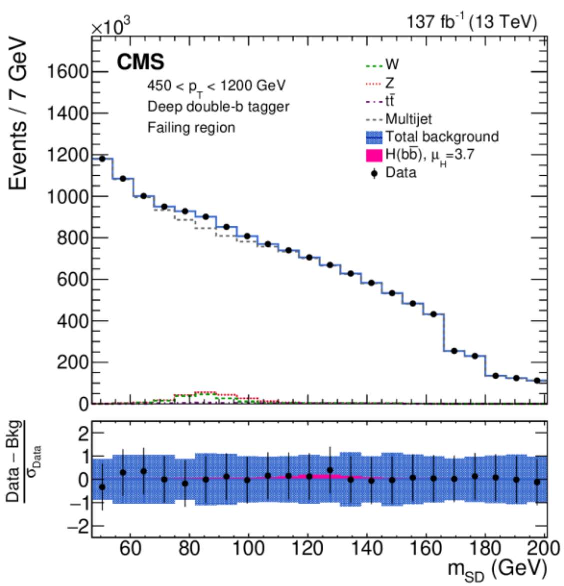
Categories in p_T and M_{SD} bins and DDBT passing/failing

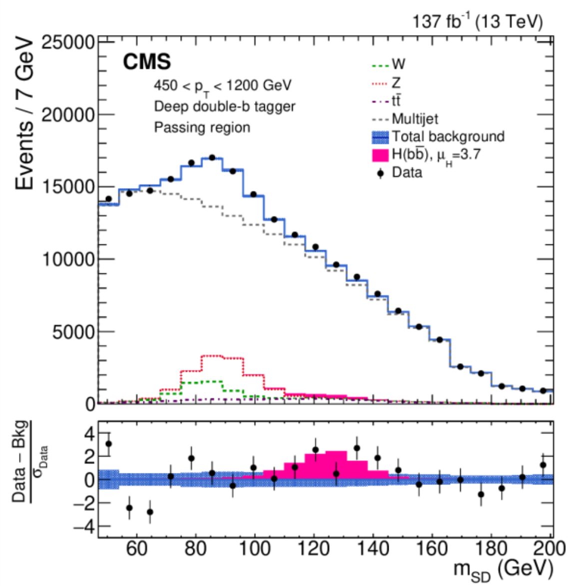
$H \to b\bar{b}$: ggF boosted

JHEP 12 (2020) 085

Signal, V+jets and *tt* background processes are modeled using simulations.
Dominant background is QCD multijet process, which is estimated in background enriched DDBT failed region together with a "pass-fail ratio" function.
Results are extracted from M_{SD} distribution by simultaneous binned maximum likelihood fit in the DDBT passing and failing regions of the (six) p_T categories.
Z+Jets process was used to validate the method. It is measured to be consistent with the SM prediction.

Combining all p_T categories, Full Run2





Signal strength : $\mu_H = 3.7$

$$\pm 1.2(stat)\ ^{+0.8}_{-0.7}(syst)\ ^{+0.8}_{-0.5}(theo)$$

$$\mu_H = 3.7\ ^{+1.6}_{-1.5}$$

Observed/expected significance:

2.5 / 0.7 σ

Summary

- **Overview of CMS studies on Higgs boson decay to bottom quark pairs is presented**
- **Observation of $H \rightarrow b\bar{b}$ decay was reported at Phys.Rev.Lett. 121, 121801, which combines the results with Run1 and 2016 data of various production modes. Main contribution gives the VH production mode.**
- **$H \rightarrow b\bar{b}$ decay channel gave significant contribution in observation of *tt*H production with CMS.**
- **New results of VBF Hbb measurement with 2016 and 2018 data of 13 TeV collisions are presented ($\mu = 0.99^{+0.48}_{-0.41}$, with observed/expected significance 2.6 / 2.9 σ)**
- **For certain cases boosted-H topology can be very beneficial as it is shown in ggF Hbb search. Boosted regime gives big advantage for the searches of some really rare processes such like double H production.**

Thank You