Measurements of the Higgs boson properties in hadronic final states at CMS

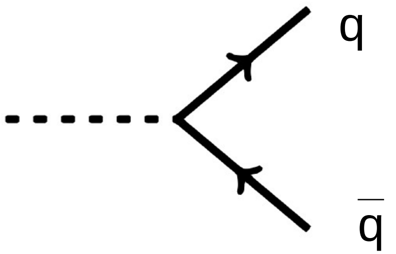
Leonardo Giannini
on behalf of the CMS collaboration

SNS, INFN Pisa

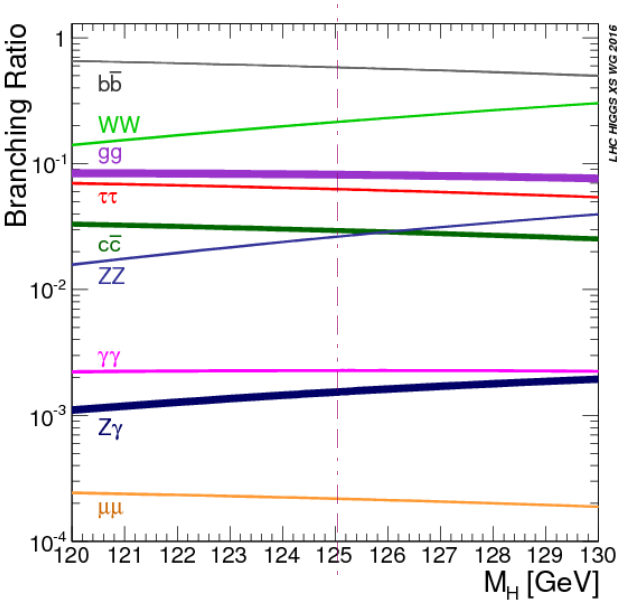
EPS-HEP - 10-17 July 2019, Ghent

Hadronic final states → Higgs to $b\bar{b}$

- H → $q\bar{q}$ via Yukawa coupling
- H→ $b\bar{b}$ has a large branching ratio (~58%) – but also very large backgrounds at hadron colliders
  - After a long experimental history → finally observed in 2018
- Ongoing efforts to search for H → $c\bar{c}$ which has non-negligible branching ratio (~3%), but experimentally much more difficult

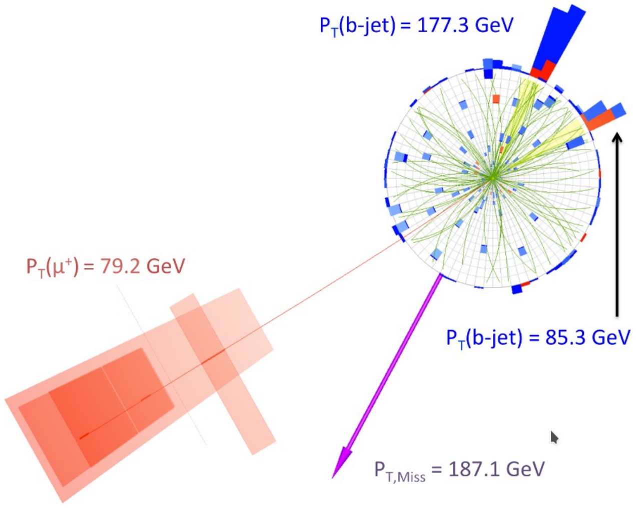


- H→ $b\bar{b}$:
  - VH production mode is the most sensitive:
    - Testing at the same time H($b\bar{b}$) Yukawa and VH vertex
  - large branching ratio allows to probe specific phase space regions
    - $t\bar{t}$H production
    - High p_T tail of the Higgs spectrum



- Focus of this talk: H→ $b\bar{b}$ results with 2017 data

VH, H → $b\bar{b}$



- **"Old" strategy:**
  - 3 channels – 0/1/2 leptons targeting Z(νν), W(lν) Z(ll), - leptons and MET used also at trigger level
  - Vector boson + Vector boson boost (~100 GeV) reduce greatly the background
  - Jet b-tagging + b jet energy regression fundamental tools to select higgs candidates and improve the mass resolution
  - Multivariate analysis to isolate signal from V+(HF) jets and Top backgrounds

NEW

- Deep Neural Networks (DNNs) for jet b-tagging, improved b-jet energy regression, S vs B classifiers
- Kinematic fit in the 2 leptons channel

Higgs mass resolution

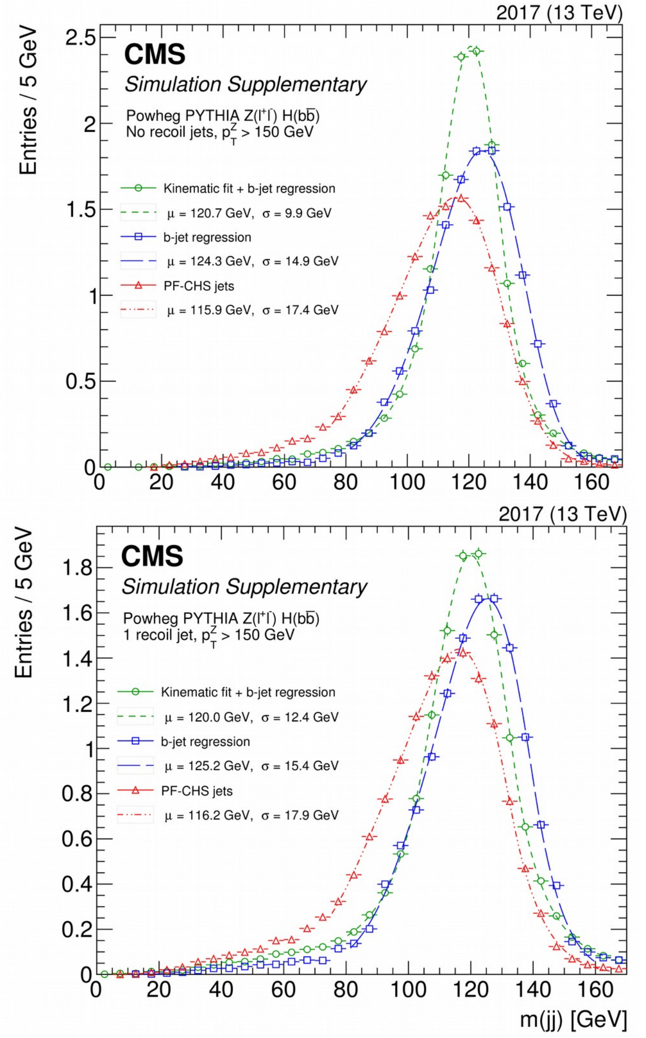
- 15% baseline dijet mass resolution

- **DNN b jet regression** (0/1/2 leptons):
  - Trained on b jets from $t\bar{t}$ simulation
  - Recovers b jet missing energy ($B \rightarrow l\nu$)
  - Uses jet shape/lepton/SV information → works as dedicated b jet calibration:

    ~10% improvement in jet resolution

    ~20% relative resolution improvement in analysis phase space
  - Improvement verified in data in Z+ b-jets control region (+data/MC scale factor)

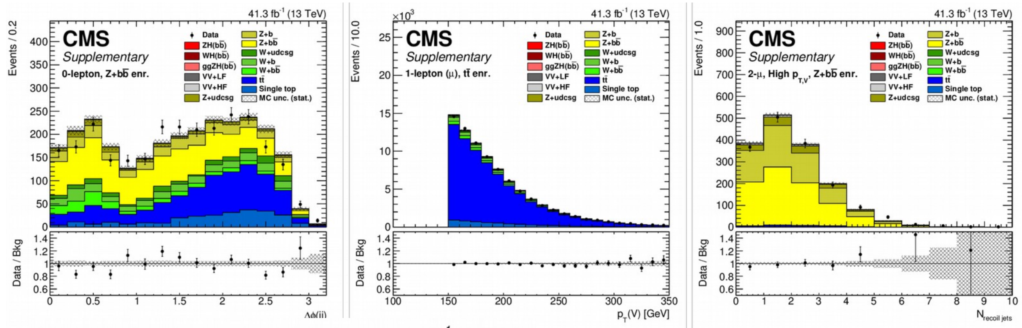
- **Kinematic fit** (2 leptons):
  - Lepton and jet 4-vector components adjusted within the experimental resolution with 2 constraints: Z mass, true MET = 0
  - Dijet mass resolution <10 GeV in the most sensitive category

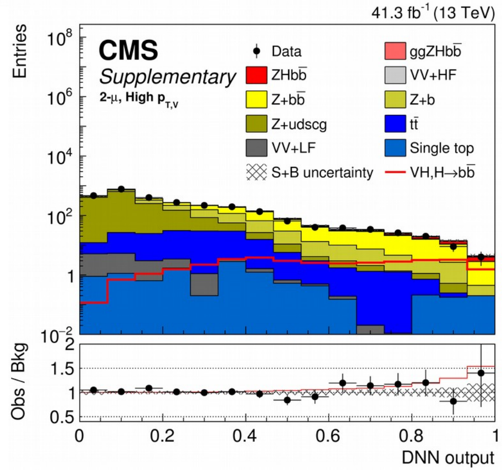


Multivariate analysis

- **Multivariate discriminator** per channel
  - DNN introduced with 2017 data
  - Makes use of the most discriminating variables (dijet mass, V p_T, Δφ(VH), …)
  - About 10-15 variables – choice based on 2016 analysis
- **Background estimate** → MC templates – normalization from data:
  - Need to isolate background sources → dedicated $t\bar{t}$, V+jets, V+HF jets control regions
  - DNN used also to categorize events in the V+HF control regions → better control of V+bb background

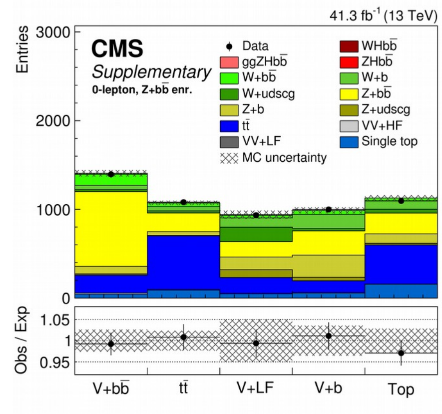
Input features





S vs B classifier (2-leptons)

Background categorization (0-lepton)



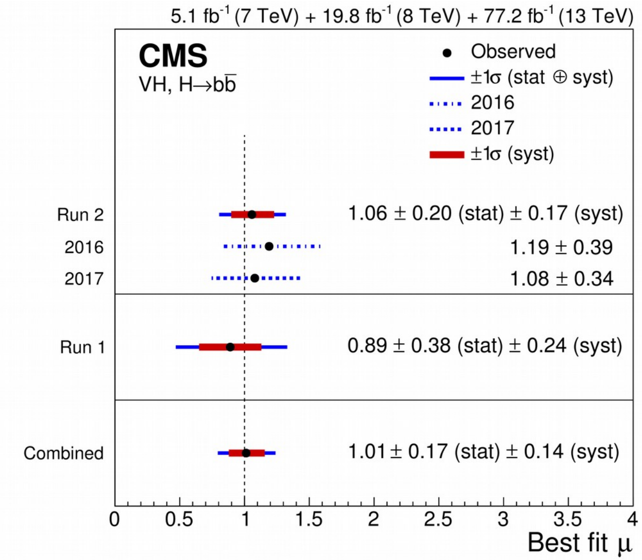
Results

2017- 41.5 fb^{-1}

$$\mu = 1.08\,^{-0.26}_{+0.26}\,(\text{stat})\,^{+0.23}_{-0.20}\,(\text{syst})$$

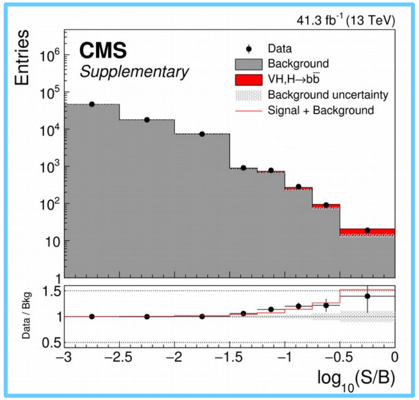
observed (expected) significance of 3.3 (3.1) σ

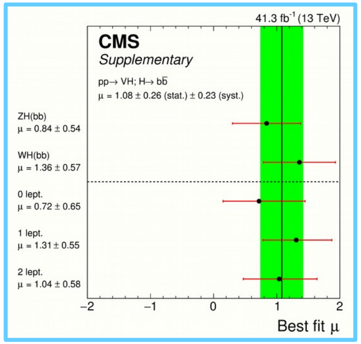
- 2017 data → improved sensitivity wrt to 2016 analysis at 13 TeV
- *Uncertainty still dominated by statistical contribution, but e.g. the MC statistical uncertainty is not negligible*
- Combination with 2016+Run1 results
  - 4.8(4.9) σ for VH($b\bar{b}$) only
  - Adding $t\bar{t}$H($b\bar{b}$) we reached the observation – 5.6 (5.5) σ → *next topic*



	Significance (σ)		
Data set	Expected	Observed	Signal strength
2017			
0-lepton	1.9	1.3	0.73 ± 0.65
1-lepton	1.8	2.6	1.32 ± 0.55
2-lepton	1.9	1.9	1.05 ± 0.59
Combined	3.1	3.3	1.08 ± 0.34
Run 2	4.2	4.4	1.06 ± 0.26
Run 1 + Run 2	4.9	4.8	1.01 ± 0.23

Channels combined





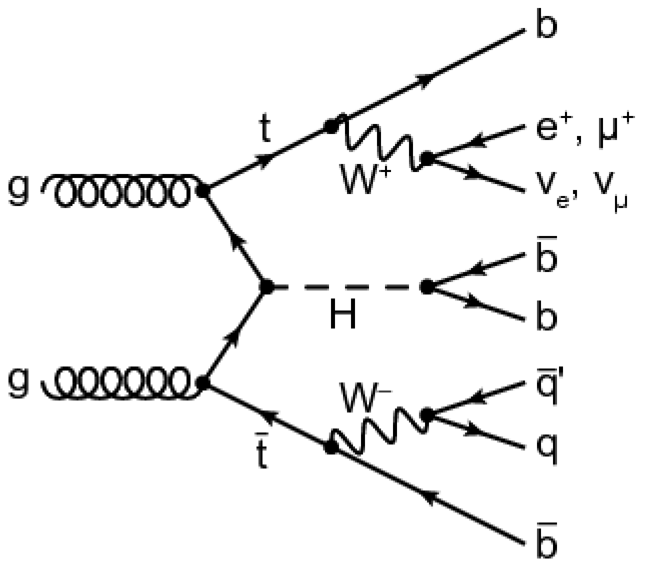
More in C. Martin Perez's talk

CMS-PAS-HIG-18-030

New result! 2017 data

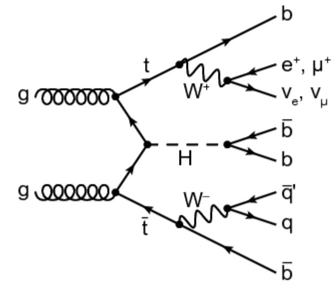
$t\bar{t}H$, $H \to b\bar{b}$

- The $t\bar{t}H$ production mode allows to probe directly the **Higgs-top Yukawa**
- **$t\bar{t}H(bb)$** is an important channel in the $t\bar{t}H$ combined measurement
- 3 final states considered defined by the top decays:
  - Dileptonic
  - Semileptonic
  - Fully-Hadronic
- Sensitivity driven by semileptonic + dileptonic final states



- $t\bar{t}$+jets backgrounds everywhere
- QCD multijet production in the fully hadronic channel
- Ambiguity in Higgs reconstruction due to combinatorics

$t\bar{t}H$, $H \to b\bar{b}$



Fully hadronic

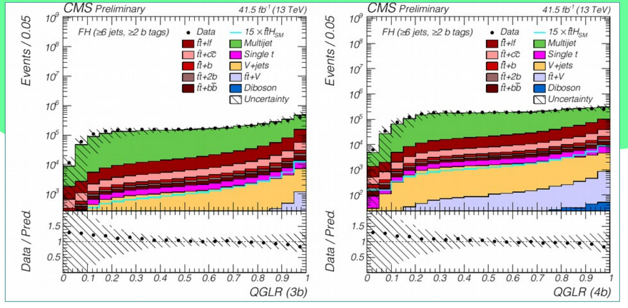
multijet triggers: 6 jets + 1 b-tag, 6 jets + 2 b-tags, 4 jets + 3 b-tags

Dijet mass compatible with W for at least a pair of non b-tagged jet

“quark vs gluon likelihood” (QGL) discriminator for all non b-tagged jets and average Δη selection between jets to reduce the QCD background

Analysis categories:
(7-8-9 jets) x (3, 4+ b-tags)

QCD multijet backgrounds estimated from data using #b-tags and QGL of jets



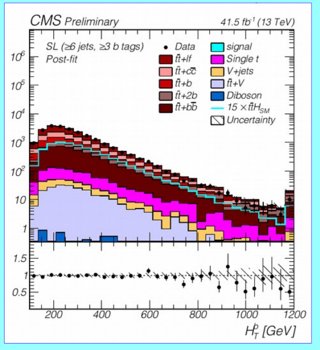
Semileptonic

Single lepton or single lepton +HT trigger

MET

1 isolated lepton

Events categorization:
(4, 5, 6+ jets) x (3+ b-tags)



Dileptonic

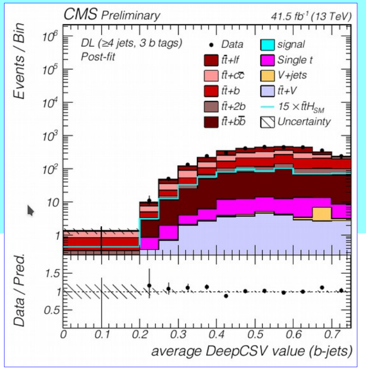
Single lepton or dilepton trigger

MET

2 opposite charge isolated leptons

Outside of Z mass window in case of same flavour

Analysis categories:
(3 jets, >3 jets) x (2,3,4+ b-tags)*



*if # Jets allows

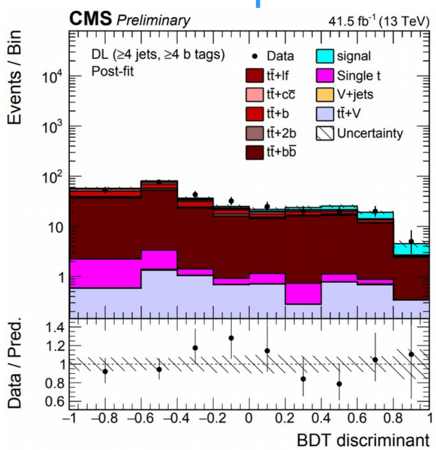
Multivariate analysis

- Different techniques used to build discriminators:
  - Fully hadronic:
    - $t\bar{t}H(b\bar{b})$ vs $t\bar{t}b\bar{b}$ with Matrix Element Method (**MEM**)
  - Semileptonic:
    - Multi-category **DNN**
  - Dileptonic:
    - Boosted decision tree (**BDT**)

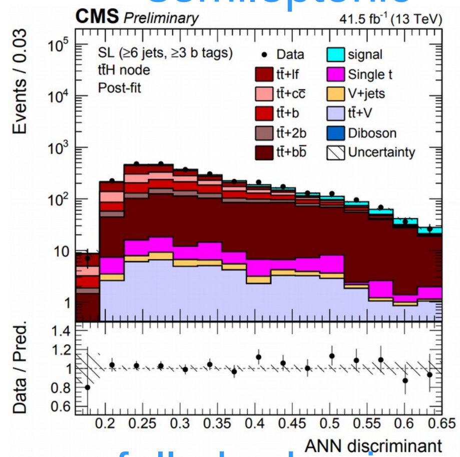
  Matrix element used as input in channels with leptons

- Simultaneous fit of all the channels

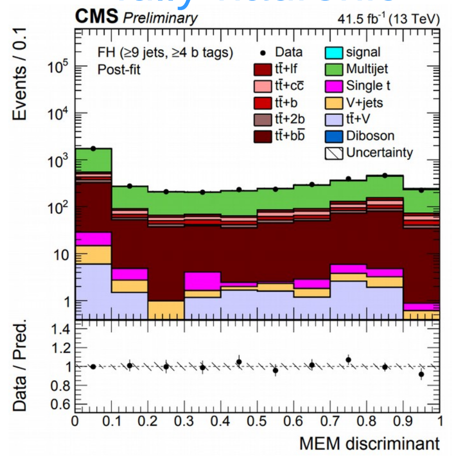
dileptonic



semileptonic



fully-hadronic



Post-fit distributions of the discriminators

*input background distributions estimated from MC simulation (**tt+jets**) and data (**QCD**)*

More in C. Martin Perez's talk

Results

$\mu = 1.49\ ^{-0.20}_{+0.21}\ (\text{stat})\ ^{+0.39}_{-0.35}\ (\text{syst})$

observed (expected) significance of 3.7 (2.6) σ

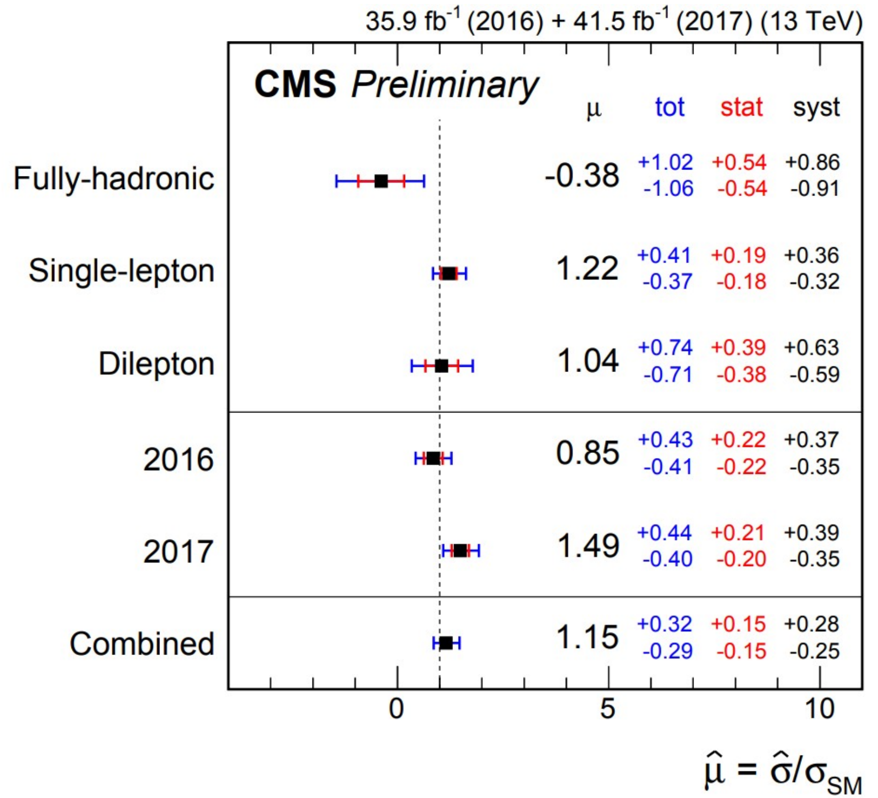
2017- 41.5 fb^{-1}

$\mu = 1.15\ ^{-0.15}_{+0.15}\ (\text{stat})\ ^{-0.25}_{+0.28}\ (\text{syst})$

observed (expected) significance of 3.9 (3.5) σ

2016+17- 77.4 fb^{-1}

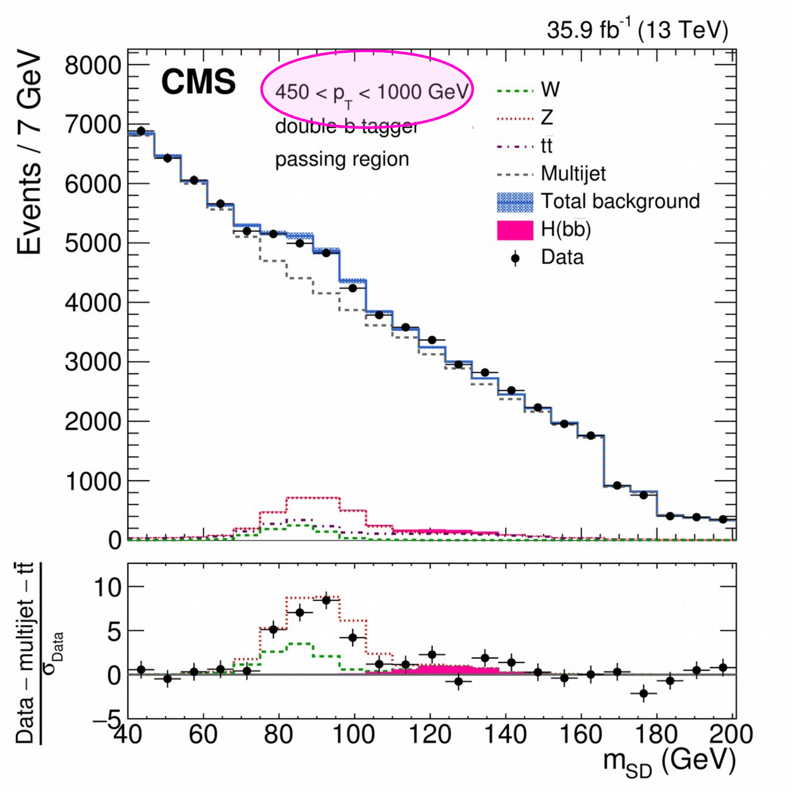
First evidence for $t\bar{t}H, H \rightarrow b\bar{b}$



- systematic uncertainty, especially theoretical (70% of the total), dominant at this stage

Boosted H → $b\bar{b}$

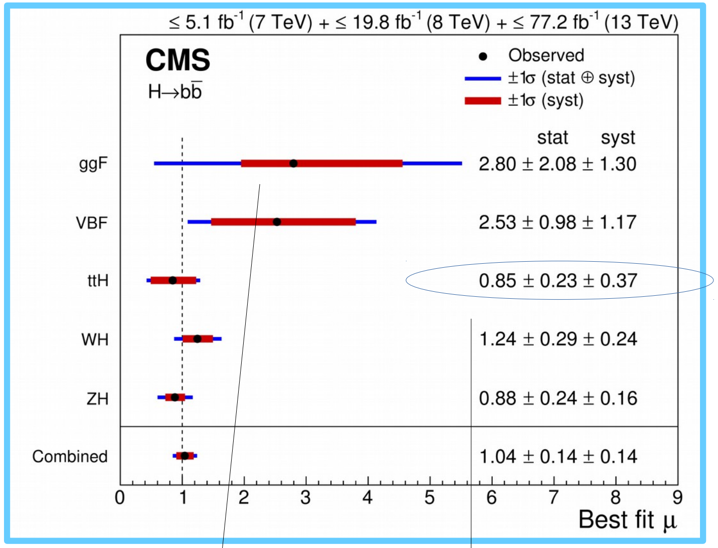
- **Higgs p_T >~ 500 GeV** to be sensitive to the inclusive Higgs production
  - The high p_T tail can be sensitive also to **deviations from SM**
- Using substructure and dedicated bb-tagging techniques
- Data-driven background estimate
- 2016 result
  - Observation of boosted Z($b\bar{b}$)
  - Higgs significance: 1.5(0.7) σ with statistical uncertainty larger than the systematic uncertainties
- **Targeting full Run2 analysis**



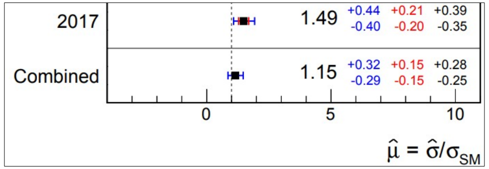
Conclusions

- **H→ $b\bar{b}$ had great succes with CMS Run 2 data**:
  - 2016 dataset:
    - Evidence for H($b\bar{b}$) in VH mode
    - H($b\bar{b}$) as input to ttH combination
    - Inclusive search for boosted H($b\bar{b}$)
  - 2017 data:
    - Observation of H($b\bar{b}$) with an improved VH($b\bar{b}$) analysis
    - First Evidence of $t\bar{t}$H($b\bar{b}$)
- Possible thanks to great **b-tagging** and **multivariate analysis techniques**
- More to come with legacy Run2 results
  - New strategies in order to reduce the systematic uncertainties under investigation
  - Efforts to go for differential measurements

H(bb) combined result summer 2018



Update: '16+'17 data



Boosted ggF to be updated

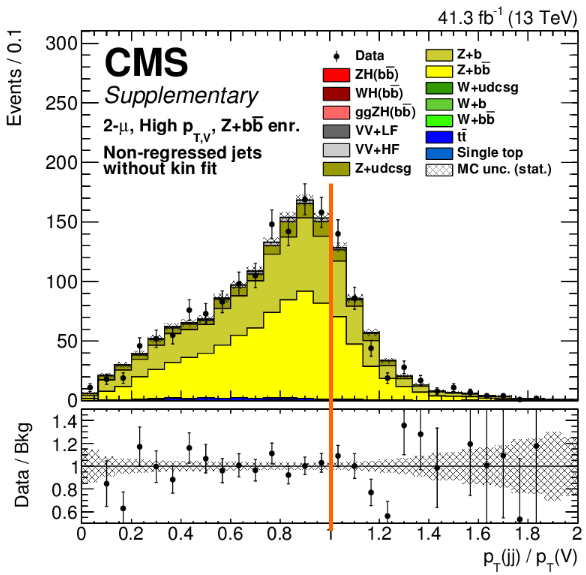
Thank you for your attention!

Additional Material

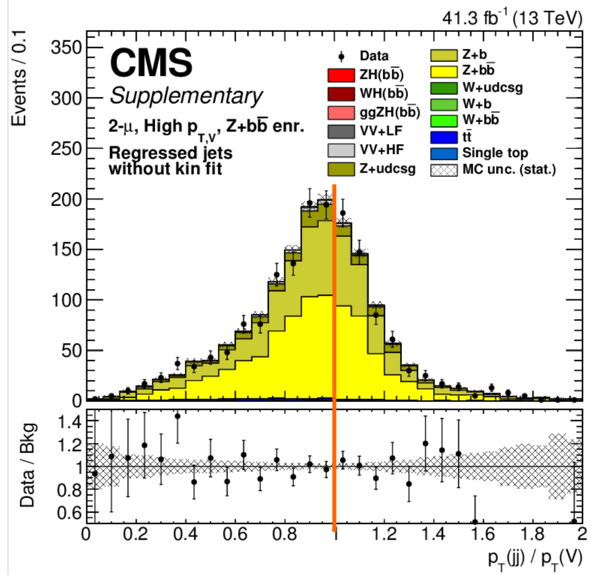
Bibliography

- “Measurement of ttH production in the H → bb decay channel in 41.5 fb^{-1} of proton-proton collision data at $\sqrt{s}$ = 13 TeV”, CMS-PAS-HIG-18-030
  - https://cds.cern.ch/record/2675023
- “Observation of Higgs Boson Decay to Bottom Quarks” Phys. Rev. Lett. 121, 121801 (2018)
  - Supplementary material url: http://cms-results.web.cern.ch/cms-results/public-results/publications/HIG-18-016/index.html
- “Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair” Phys. Rev. Lett. 120 (2018) 071802
  - Supplementary material url: http://cms-results.web.cern.ch/cms-results/public-results/publications/HIG-17-010/index.html

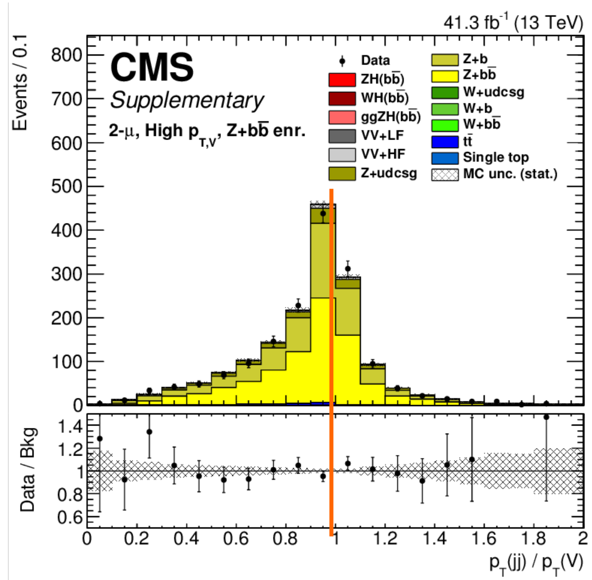
Z+jj pT balance – Z+HF control region



CMS jet calibration



Regression



Kinematic fit

Results VH(bb)

Uncertainty source	$\Delta\mu$	
Statistical	$+0.26$	-0.26
Normalization of backgrounds	$+0.12$	-0.12
Experimental	$+0.16$	-0.15
b-tagging efficiency and misid	$+0.09$	-0.08
V+jets modeling	$+0.08$	-0.07
Jet energy scale and resolution	$+0.05$	-0.05
Lepton identification	$+0.02$	-0.01
Luminosity	$+0.03$	-0.03
Other experimental uncertainties	$+0.06$	-0.05
MC sample size	$+0.12$	-0.12
Theory	$+0.11$	-0.09
Background modeling	$+0.08$	-0.08
Signal modeling	$+0.07$	-0.04
Total	$+0.35$	-0.33

	Significance (σ)		
Data set	Expected	Observed	Signal strength
2017			
0-lepton	1.9	1.3	0.73 ± 0.65
1-lepton	1.8	2.6	1.32 ± 0.55
2-lepton	1.9	1.9	1.05 ± 0.59
Combined	3.1	3.3	1.08 ± 0.34
Run 2	4.2	4.4	1.06 ± 0.26
Run 1 + Run 2	4.9	4.8	1.01 ± 0.23

Results ttHbb

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	$+0.32/-0.29$

	$\hat{\mu}$ ± tot (±stat ± syst)	significance obs (exp)
FH 3 b-tags	$1.36^{+3.57}_{-5.36}\ \left(^{+1.68}_{-1.69}\ ^{+3.15}_{-5.09}\right)$	$0.3\,\sigma$ $(0.2\,\sigma)$
FH 4 b-tags	$-1.54^{+1.41}_{-1.45}\ \left(^{+0.91}_{-0.90}\ ^{+1.08}_{-1.13}\right)$	— $(0.7\,\sigma)$
FH combined	$-1.69^{+1.43}_{-1.47}\ \left(^{+0.83}_{-0.83}\ ^{+1.16}_{-1.22}\right)$	— $(0.7\,\sigma)$
SL 4 jets	$1.73^{+2.25}_{-2.21}\ \left(^{+0.88}_{-0.87}\ ^{+2.07}_{-2.04}\right)$	$0.8\,\sigma$ $(0.5\,\sigma)$
SL 5 jets	$0.73^{+0.98}_{-0.97}\ \left(^{+0.47}_{-0.46}\ ^{+0.86}_{-0.86}\right)$	$0.8\,\sigma$ $(1.0\,\sigma)$
SL $\geq$ 6 jets	$2.05^{+0.76}_{-0.69}\ \left(^{+0.31}_{-0.31}\ ^{+0.69}_{-0.62}\right)$	$3.0\,\sigma$ $(1.6\,\sigma)$
SL combined	$1.84^{+0.62}_{-0.56}\ \left(^{+0.26}_{-0.26}\ ^{+0.56}_{-0.50}\right)$	$3.3\,\sigma$ $(1.9\,\sigma)$
DL 3 jets	$-2.35^{+4.40}_{-2.65}\ \left(^{+2.13}_{-2.06}\ ^{+3.85}_{-1.66}\right)$	— $(0.2\,\sigma)$
DL $\geq$ 4 jets	$1.57^{+1.02}_{-0.98}\ \left(^{+0.55}_{-0.53}\ ^{+0.86}_{-0.82}\right)$	$1.6\,\sigma$ $(1.0\,\sigma)$
DL combined	$1.62^{+0.90}_{-0.85}\ \left(^{+0.50}_{-0.48}\ ^{+0.76}_{-0.70}\right)$	$1.9\,\sigma$ $(1.2\,\sigma)$
FH+SL+DL combined	$1.49^{+0.44}_{-0.40}\ \left(^{+0.21}_{-0.20}\ ^{+0.39}_{-0.35}\right)$	$3.7\,\sigma$ $(2.6\,\sigma)$
FH+SL+DL combined 2016+2017	$1.15^{+0.32}_{-0.29}\ \left(^{+0.15}_{-0.15}\ ^{+0.28}_{-0.25}\right)$	$3.9\,\sigma$ $(3.5\,\sigma)$

CMS b-tagging