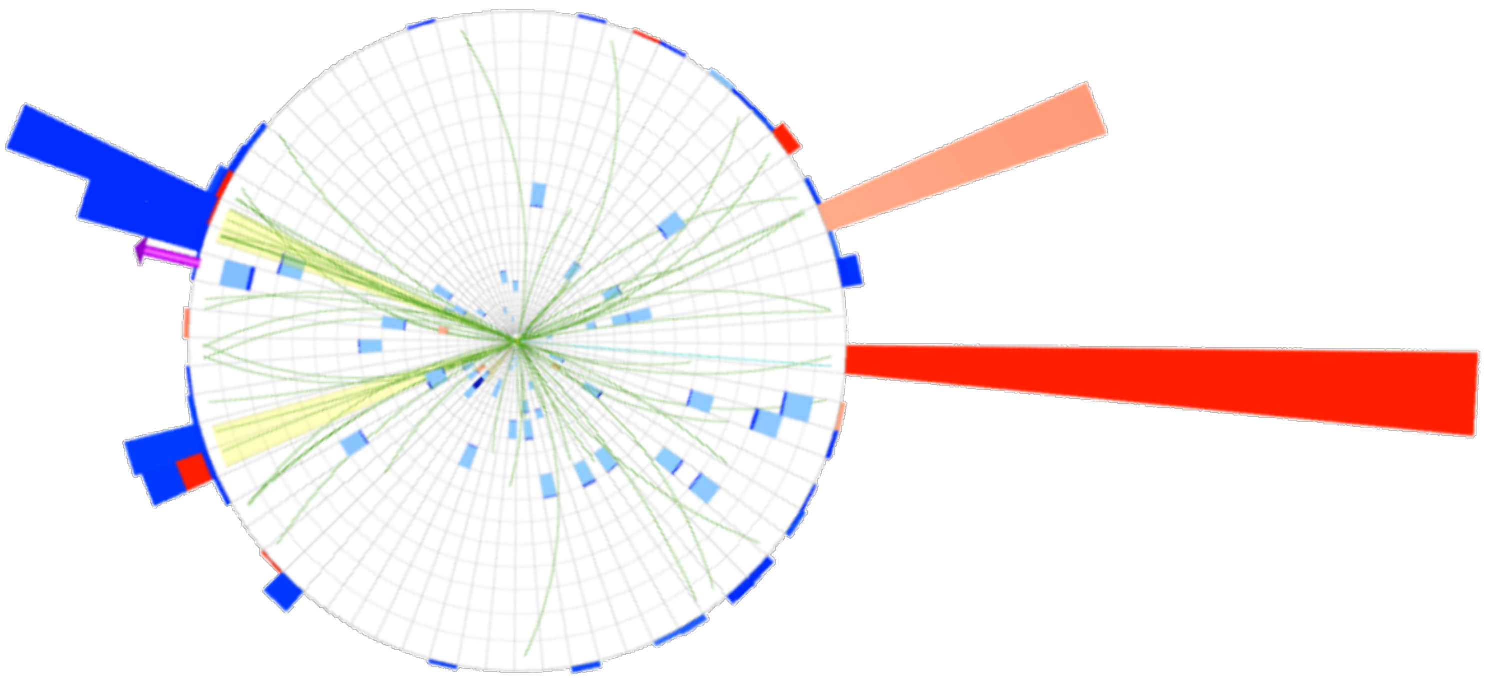
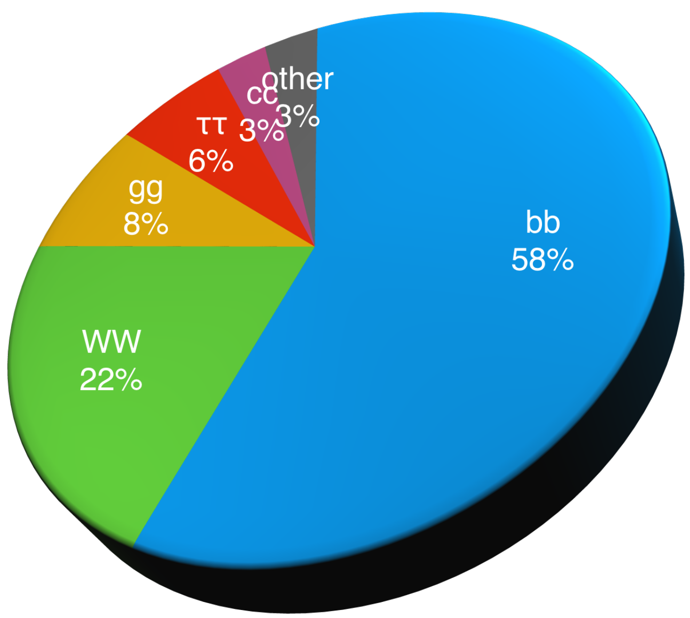
# Observation of Hbb at CMS

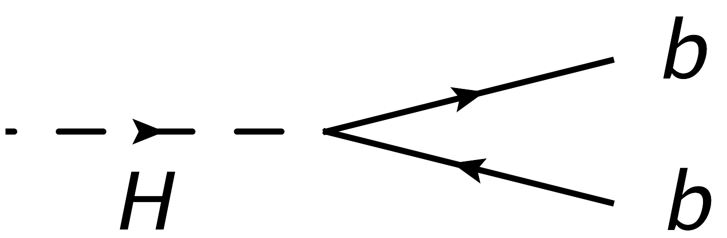
Pirmin Berger (ETH Zürich), on behalf of the CMS Collaboration



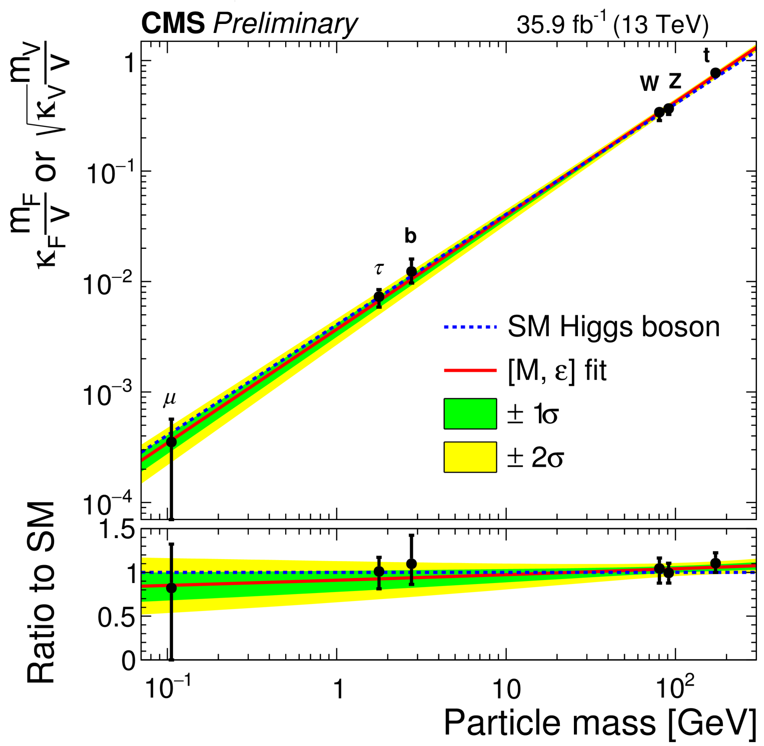
# Higgs Boson decay to $b\bar{b}$ quark pair



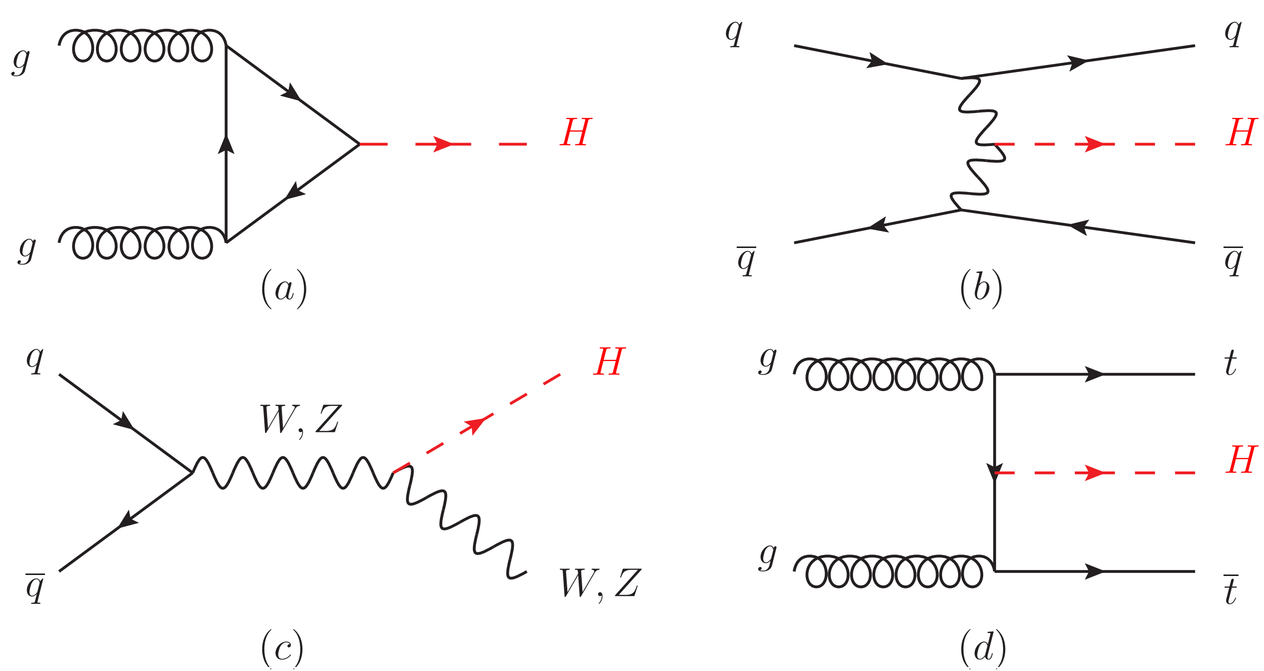

decay branching ratio




test of coupling of Higgs to down type quarks



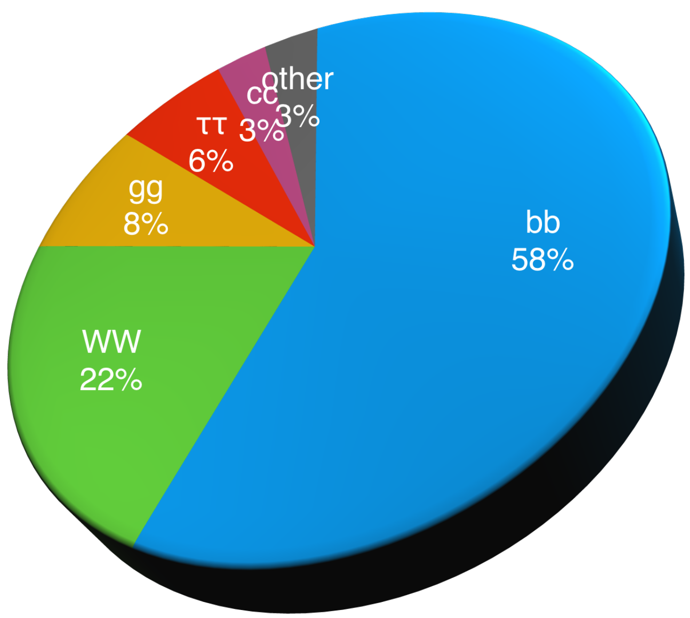

# Major Higgs production modes at LHC



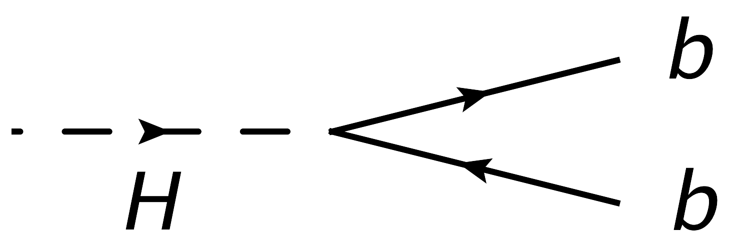

from PDG

**all 4 modes used for combination**

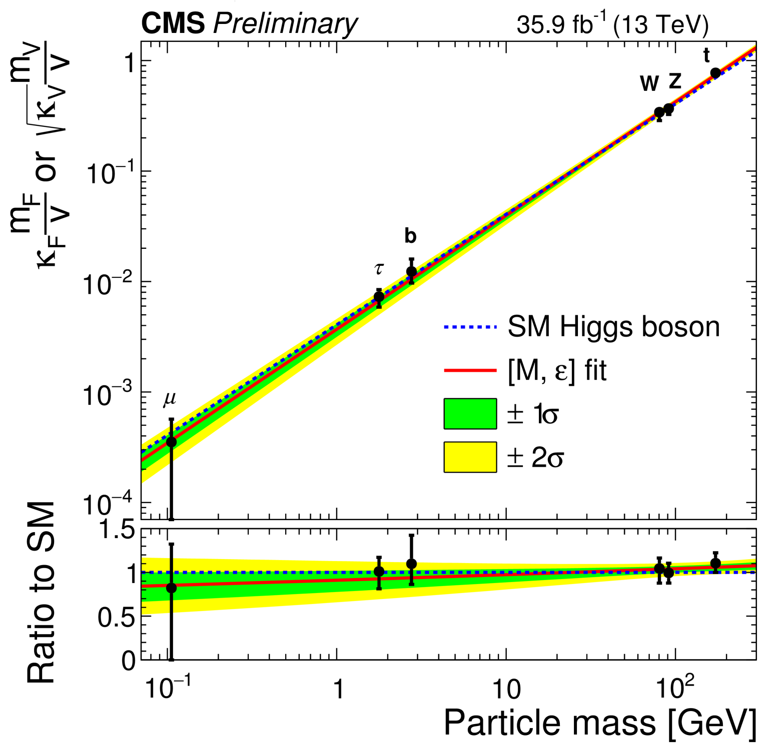
## Higgs Boson decay to bb̄ quark pair




decay branching ratio

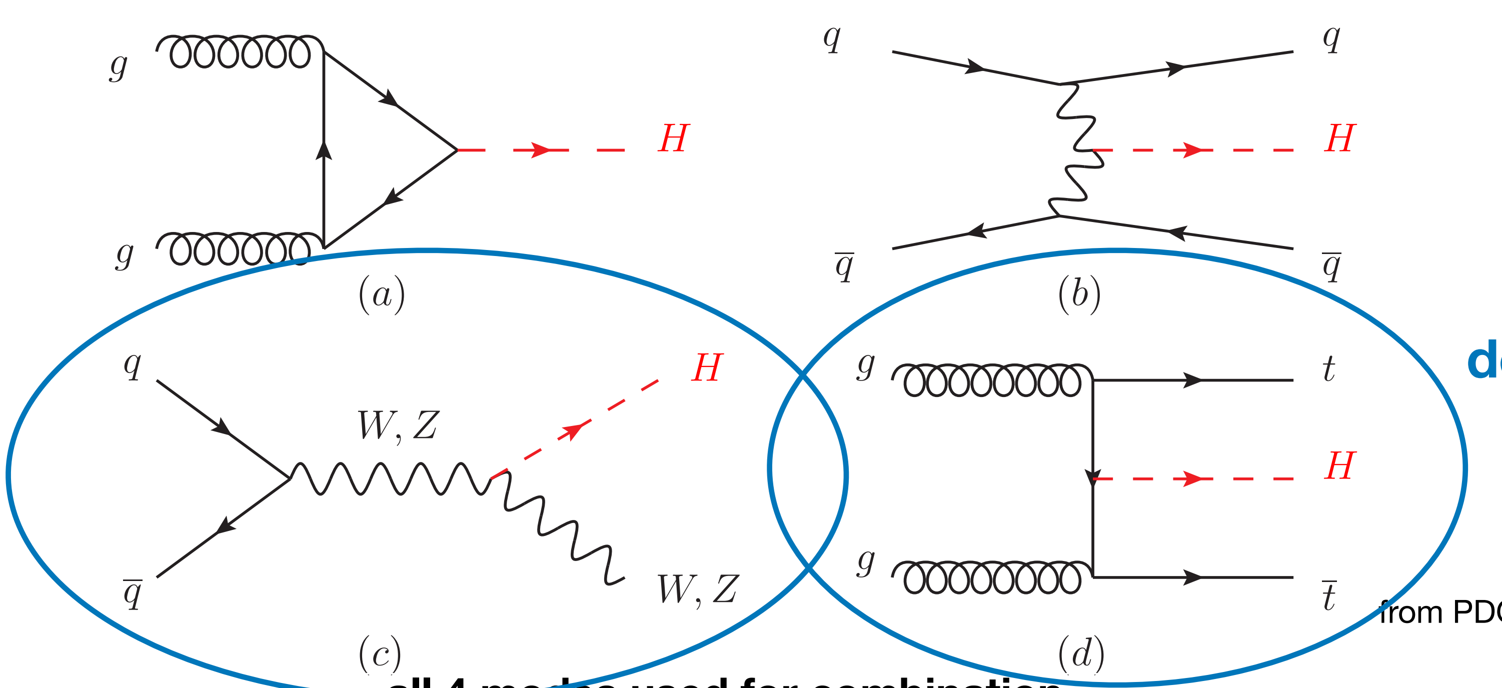



test of coupling of Higgs to down type quarks




## Major Higgs production modes at LHC

**this talk will focus on new VHbb results**

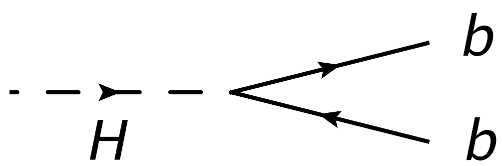


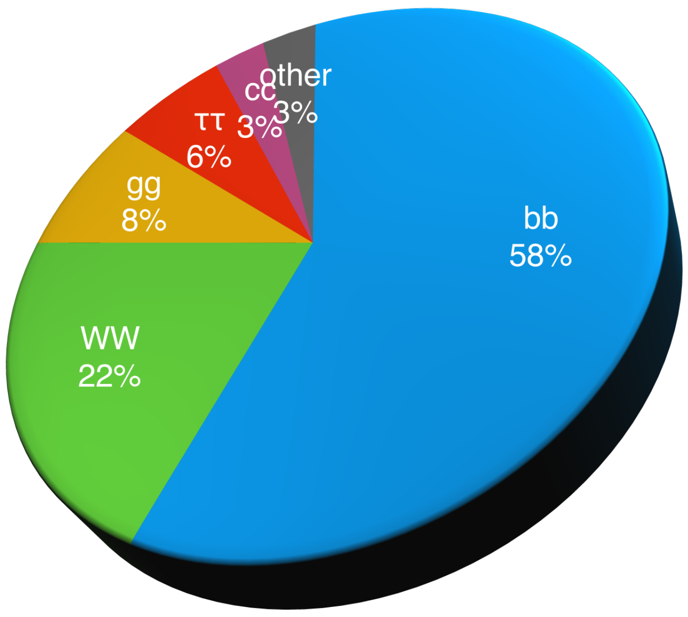

**dedicated ttH talk later!**

from PDG

**all 4 modes used for combination**

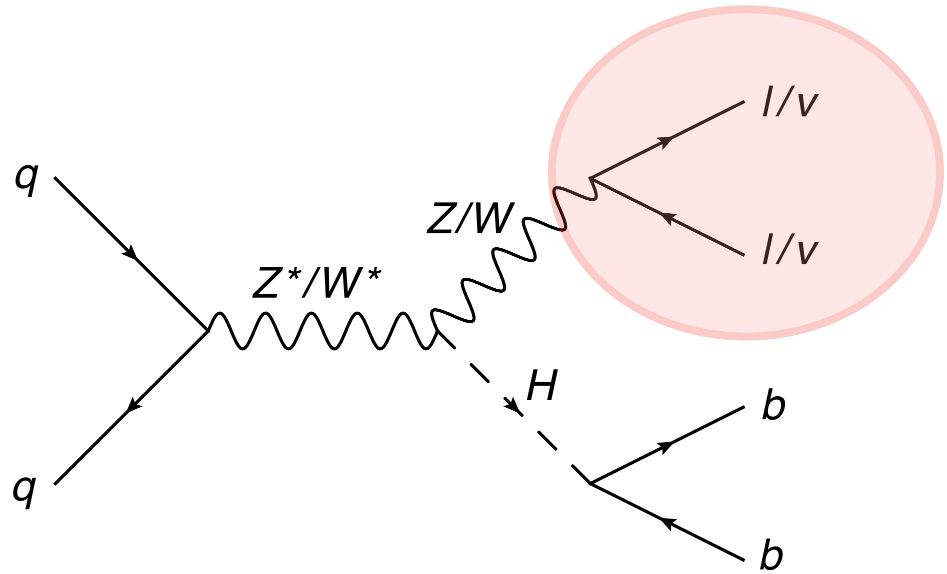
- Higgs boson has largest **branching ratio** for decay to $b\bar{b}$ quark pair




- challenges:
  - large **background** consisting of QCD multi jet events with two b-jets
- VH production and ttH production mode most significant to study Hbb decay

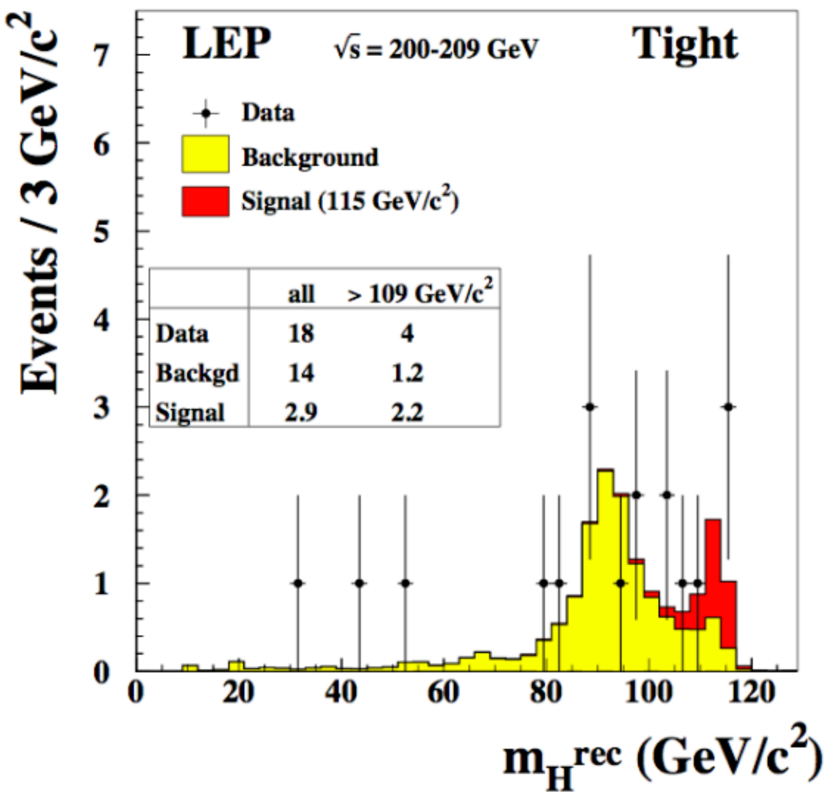
**Associated production with a Vector Boson (VH)**

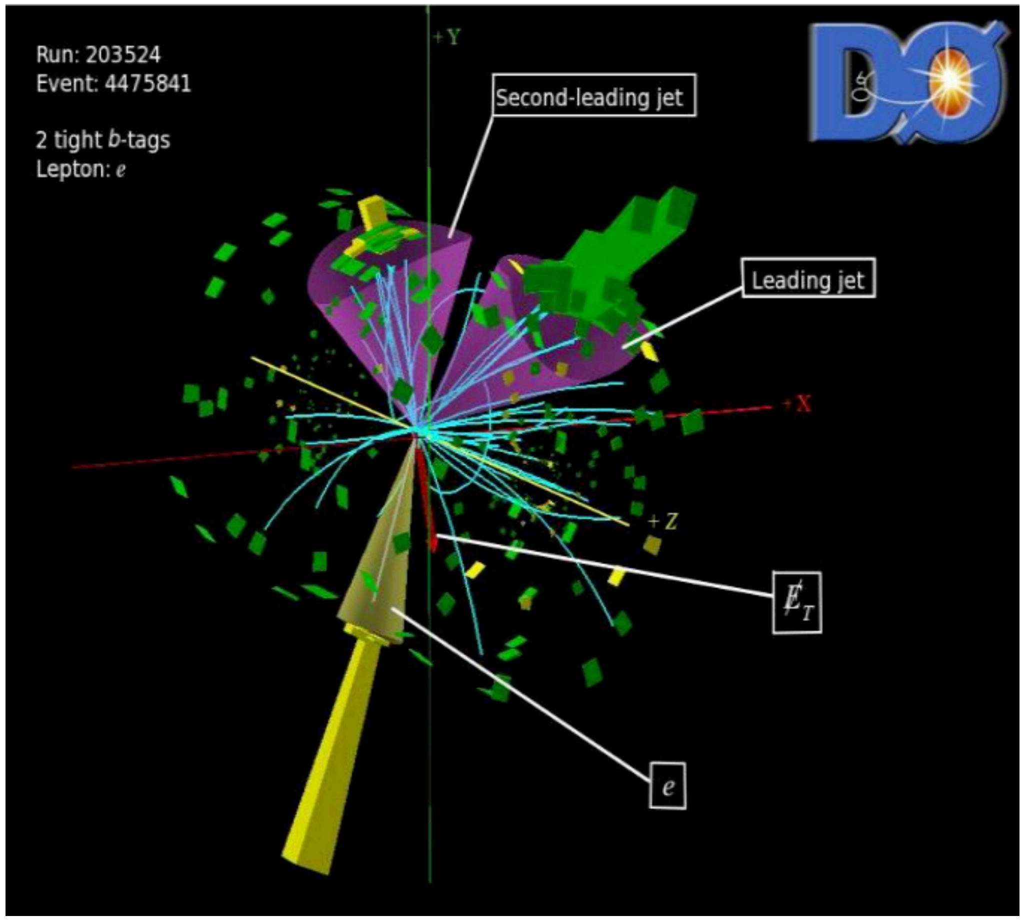



- reduces the background (from QCD) since it requires additional leptons with invariant mass of vector boson or missing energy
- can trigger leptons/$E_{T,miss}$
- 3 channels
  - 0 leptons
  - 1 lepton
  - 2 leptons

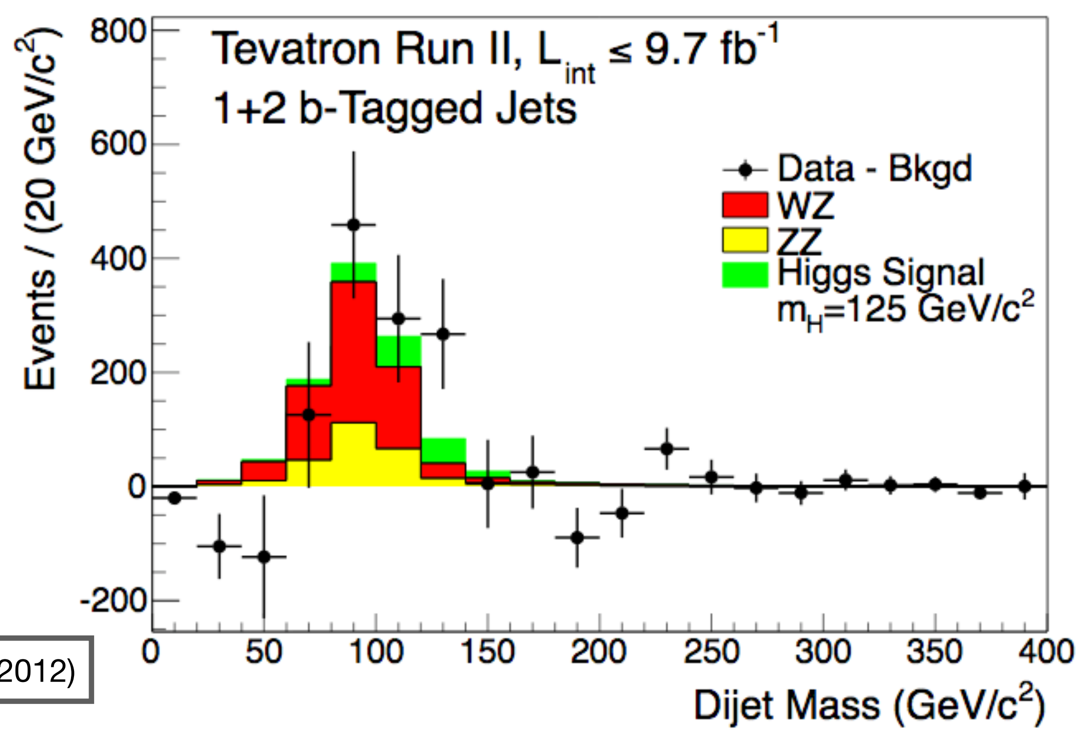
# History of VHbb searches

**started already at LEP**




**CDF+D0 combined**



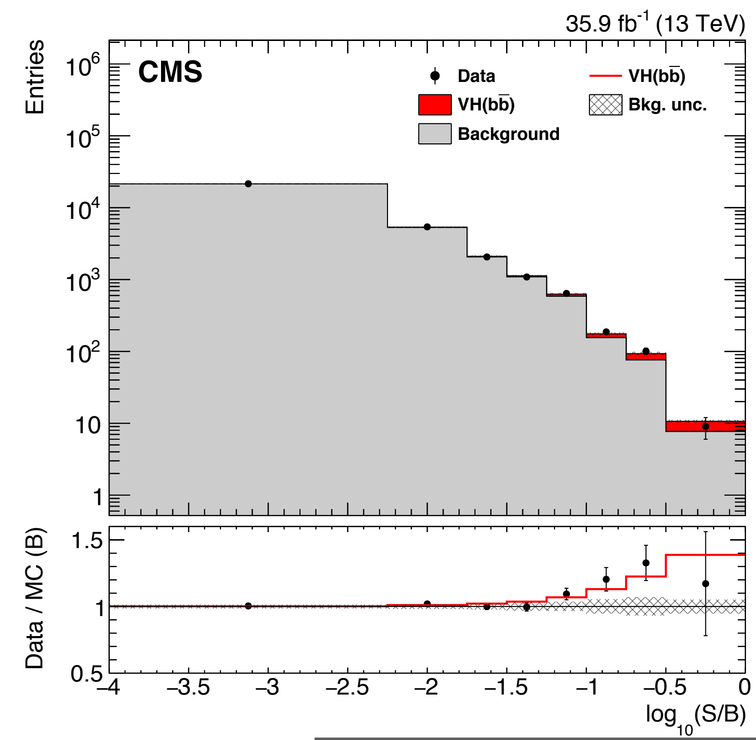

**VHbb most sensitive channel in Tevatron Higgs search**

**2012:**
observed significance: 2.8
for $m_H$ = 125 GeV

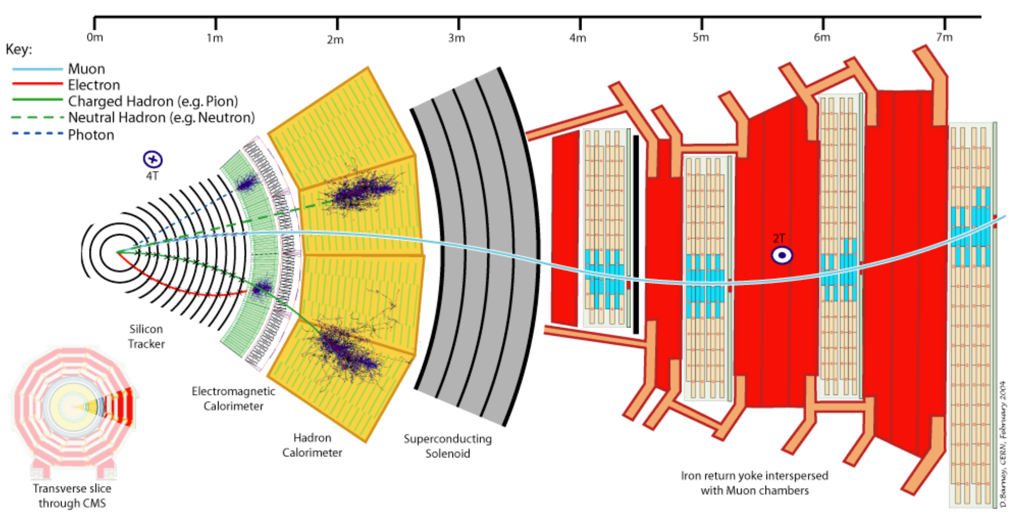
PRL109 071804 (2012)

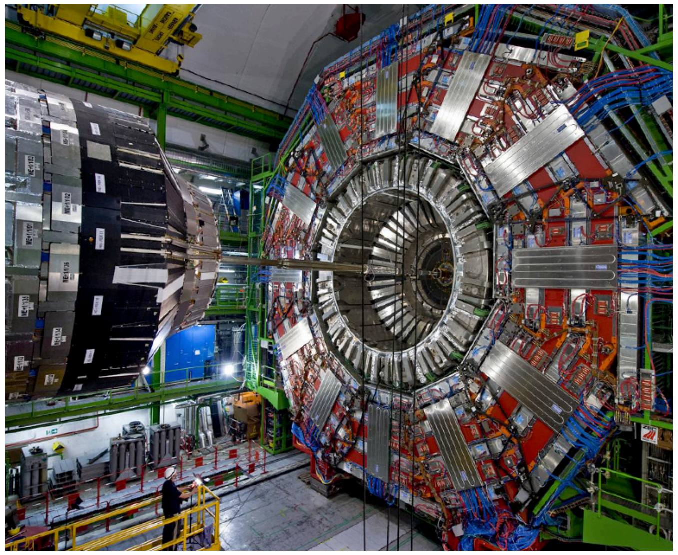
## LHC era

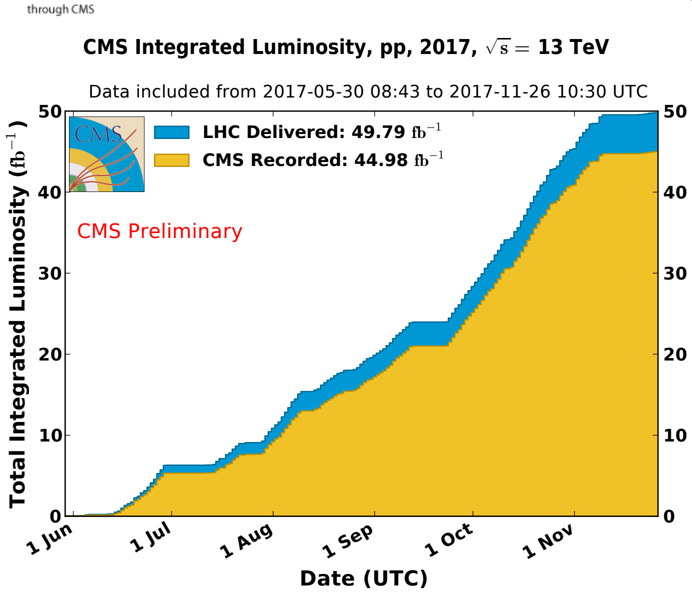
- VHbb analysis is done at ATLAS and CMS
- started with 7 & 8 TeV in Run 1
  - ATLAS: observed (expected) significance: 1.4 (2.6)
  - CMS:  observed (expected) significance: 2.1 (2.1)
- evidence for VHbb at the LHC established by ATLAS and CMS in 2017:
  - using ~36 $fb^{-1}$ of 2016 dataset at 13 TeV
  - combined with Run 1



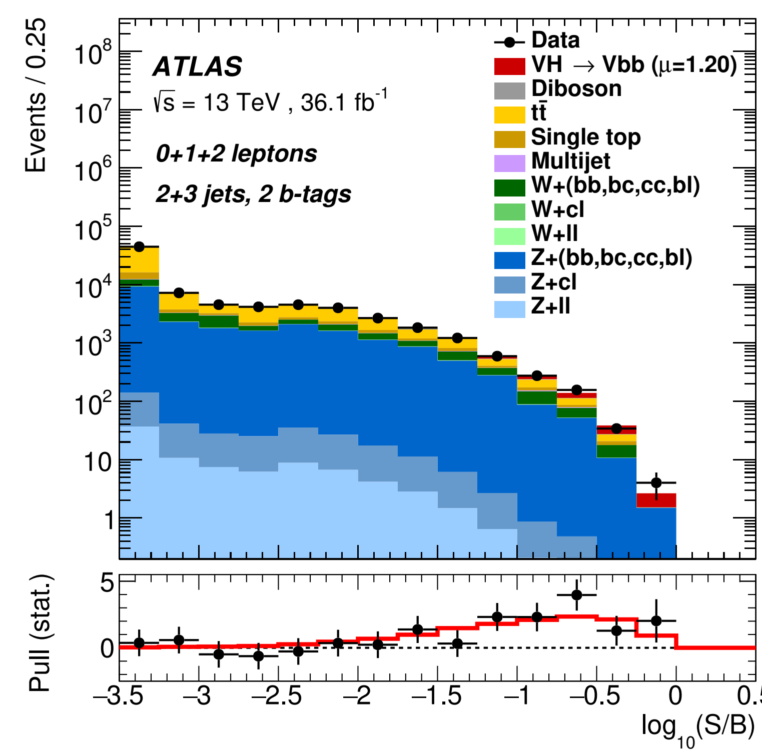

Phys. Lett. B *780* (2018) *501*



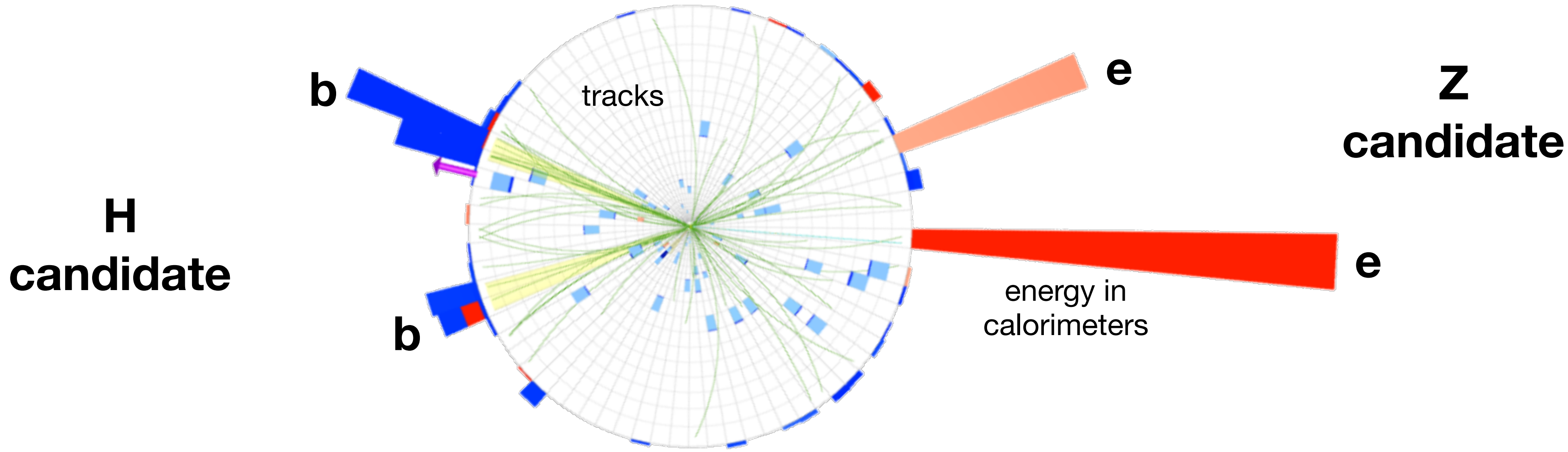


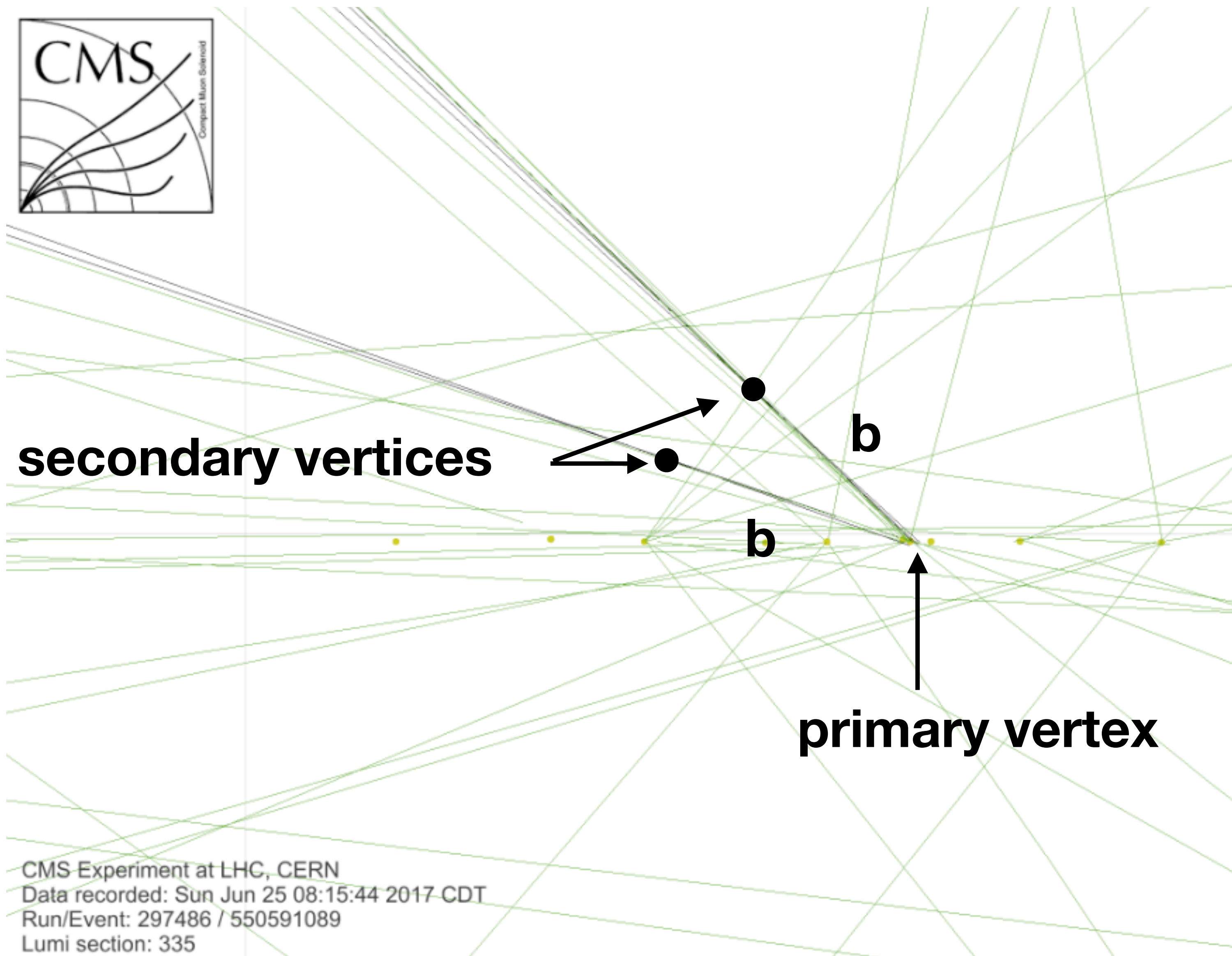

- great machine performance of LHC in 2017 => observation of (V)Hbb in reach!



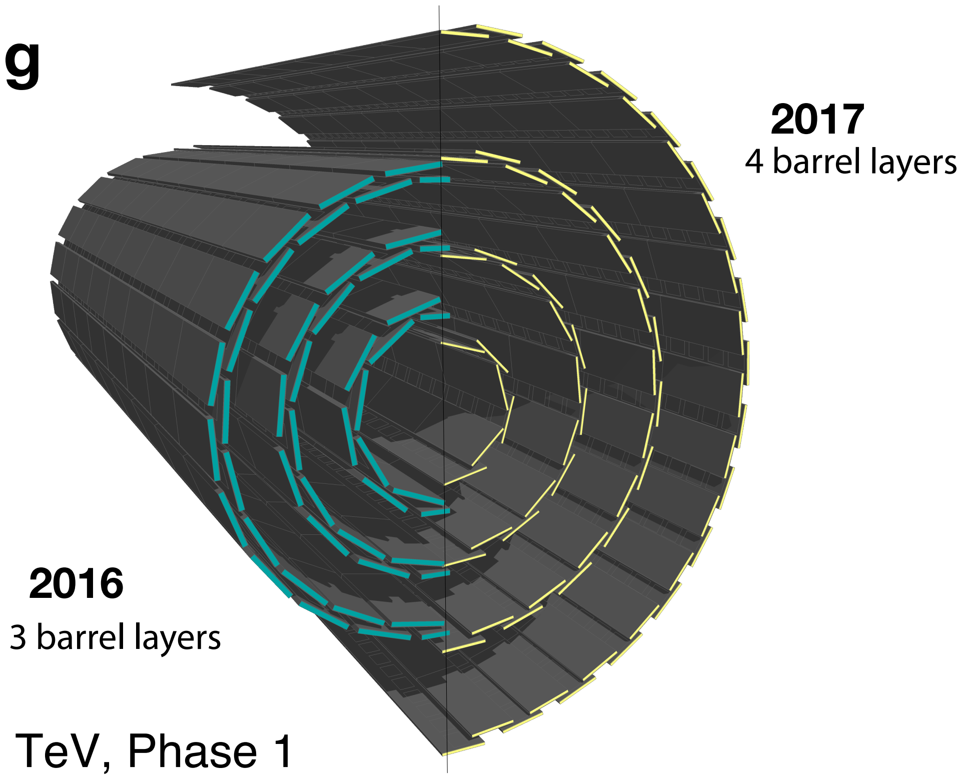

JHEP12 (2017) 024

# Object reconstruction



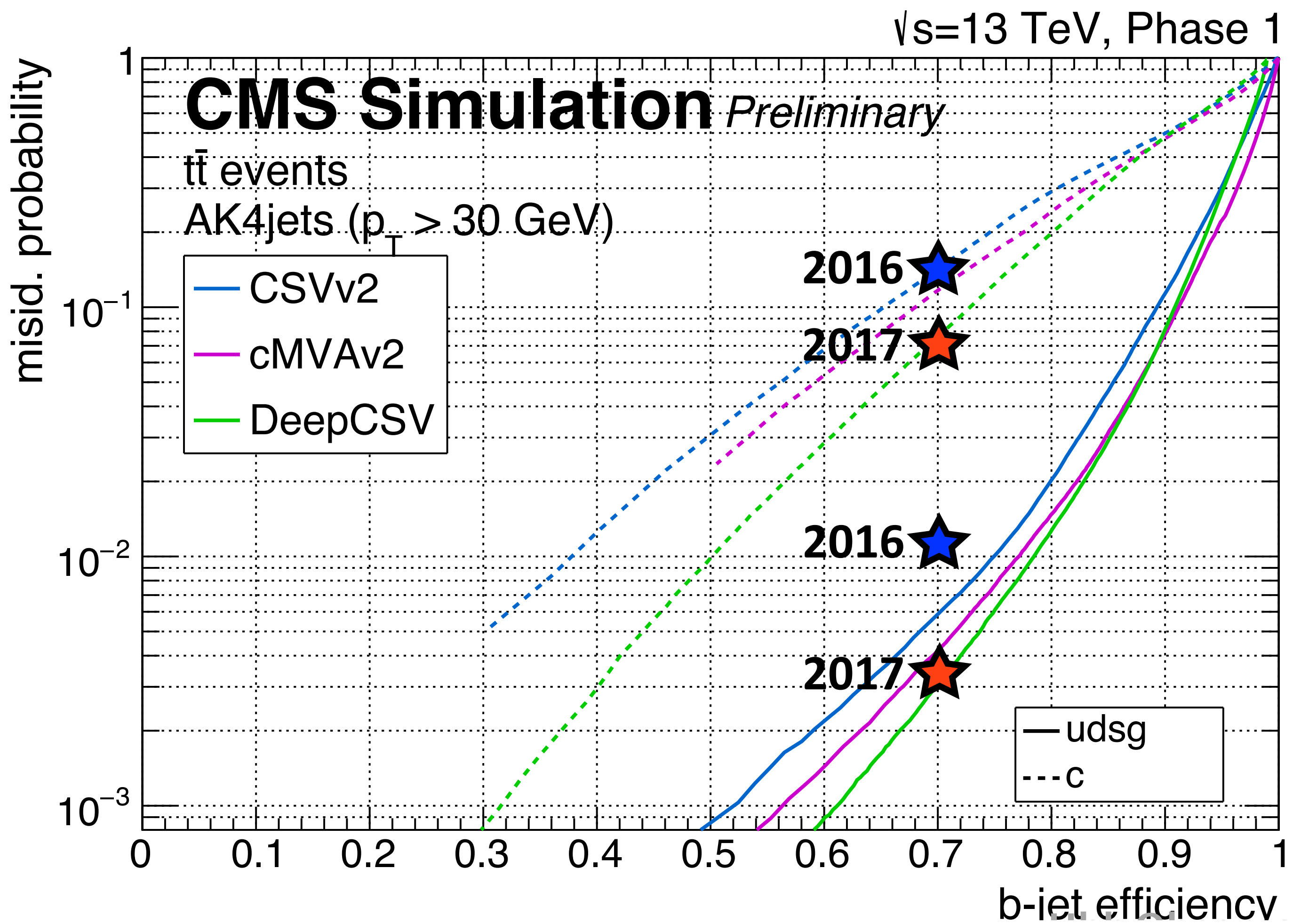



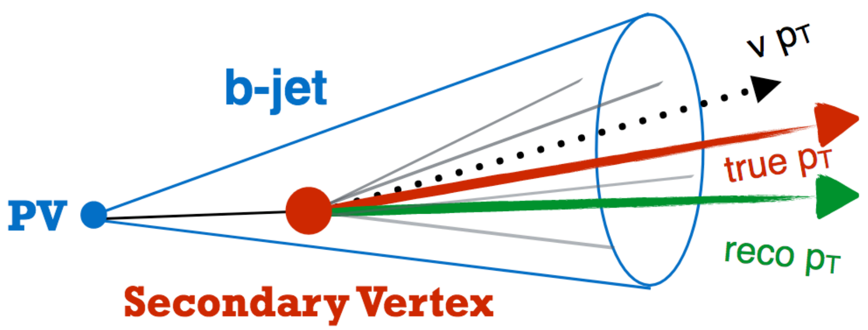
## improvements in b-tagging




- identification of b-jets plays important role in search for Hbb coupling!

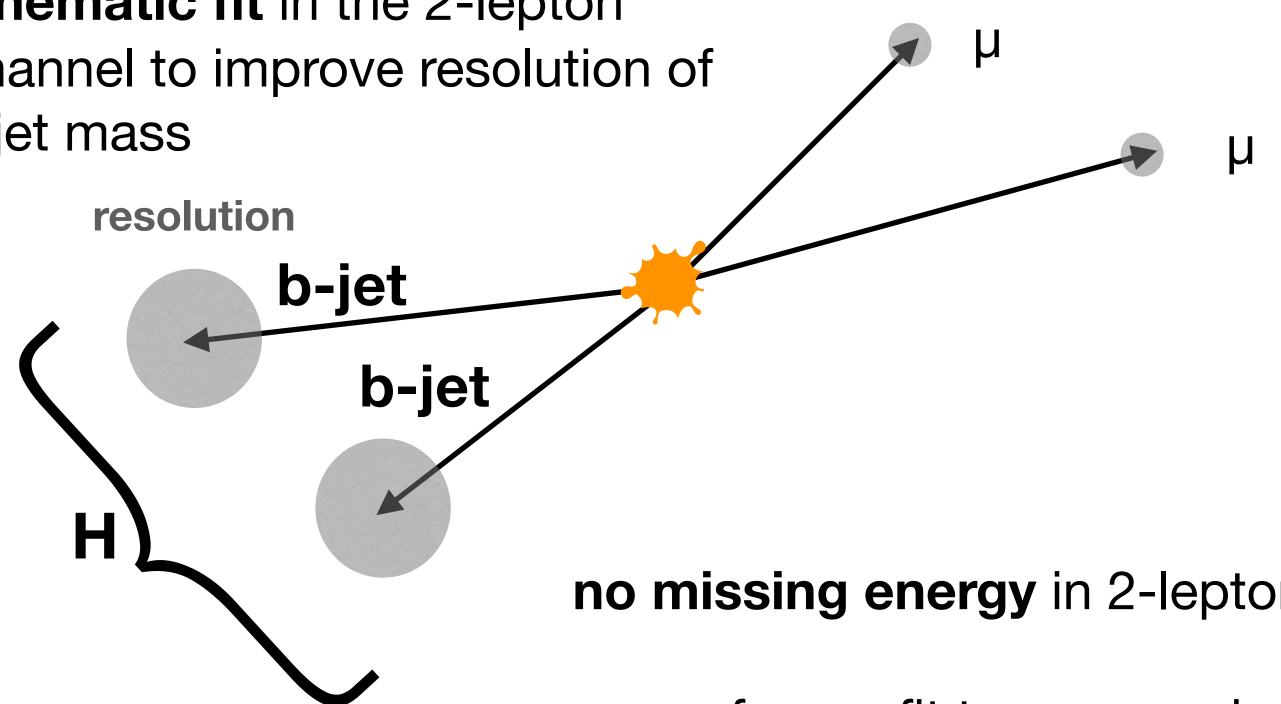
- new pixel detector with 4 instead of 3 layers
- new b-tagging algorithm: DeepCSV
    - ~ 70% efficiency for < 1% mis-tag rate for light (udscg) jets

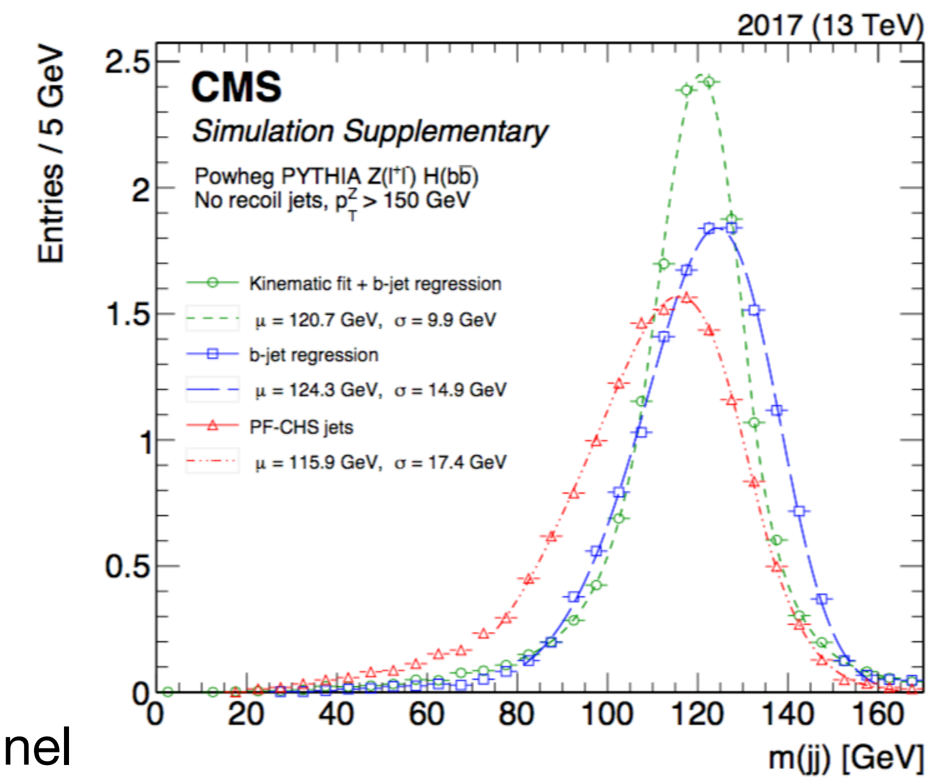





- **DNN regression for b-jets** (AK04) to recover the escaping neutrino in semi-leptonic decays
- FSR jet recovery: add back jets in $\Delta R < 0.8$ cone to Higgs candidate

- **kinematic fit** in the 2-lepton channel to improve resolution of dijet mass




**no missing energy** in 2-lepton channel

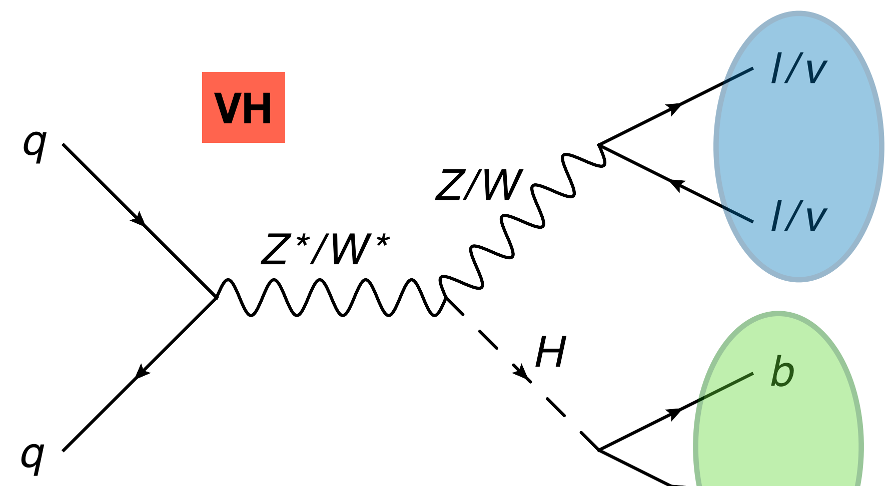
=> perform a fit to use good resolution on leptons to improve resolution of b-jets and therefore resolution of $m_H$

- **vector boson** reconstructed with leptons/MET

# Signal and backgrounds

## signal

**VH**



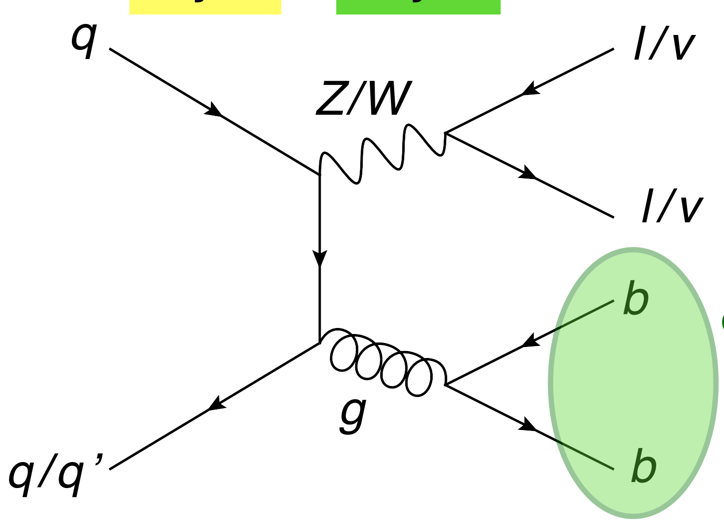

**+ ggZH contribution**

**invariant mass has a peak at $m_V$**

**invariant mass has a peak at $m_H$ ~ 125 GeV**

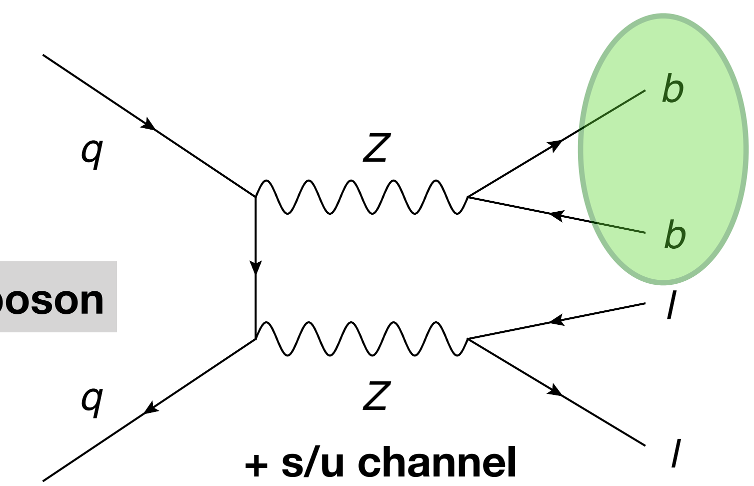
## main backgrounds

**Z+jets** **W+jets**




**falling mass distribution without peak, softer $p_T$**

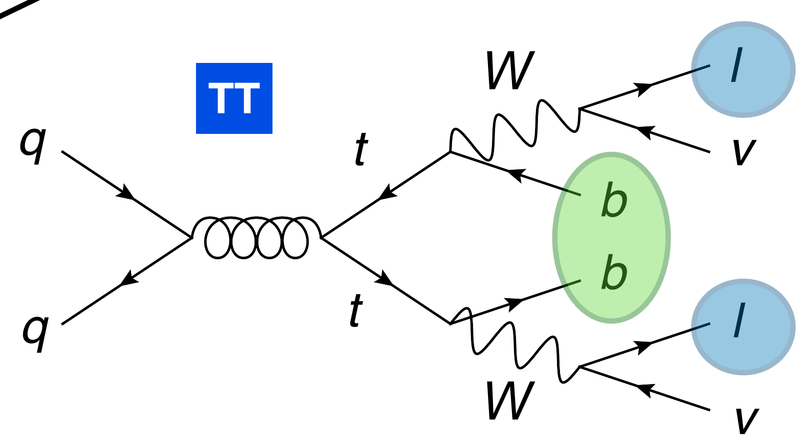
**Di-boson**




**+ s/u channel**

**invariant mass has a peak at $m_V$**

**TT**



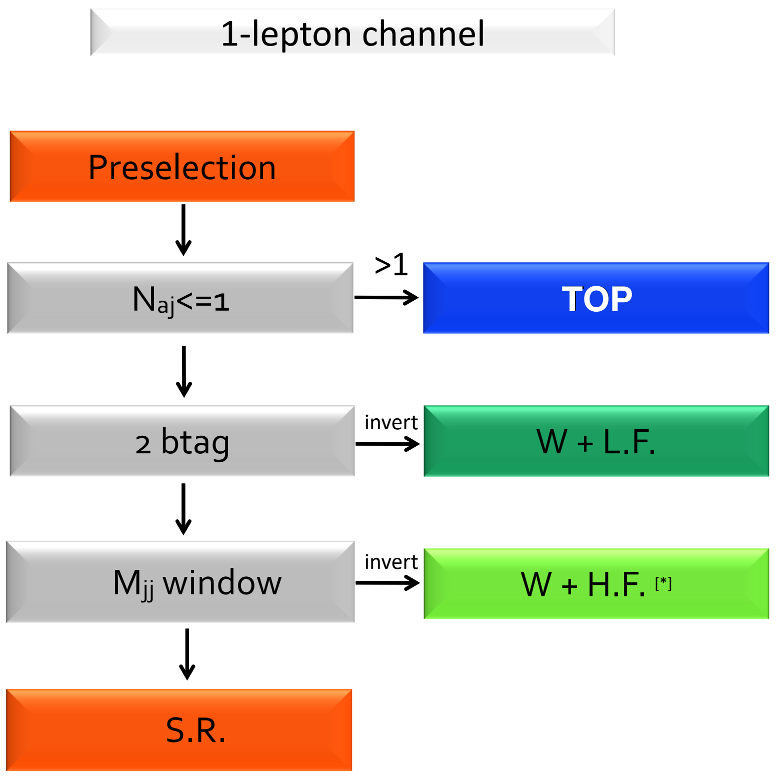

**invariant mass has flat distribution**

**if W decays hadronically, more additional jets**

in each of the three channels define:

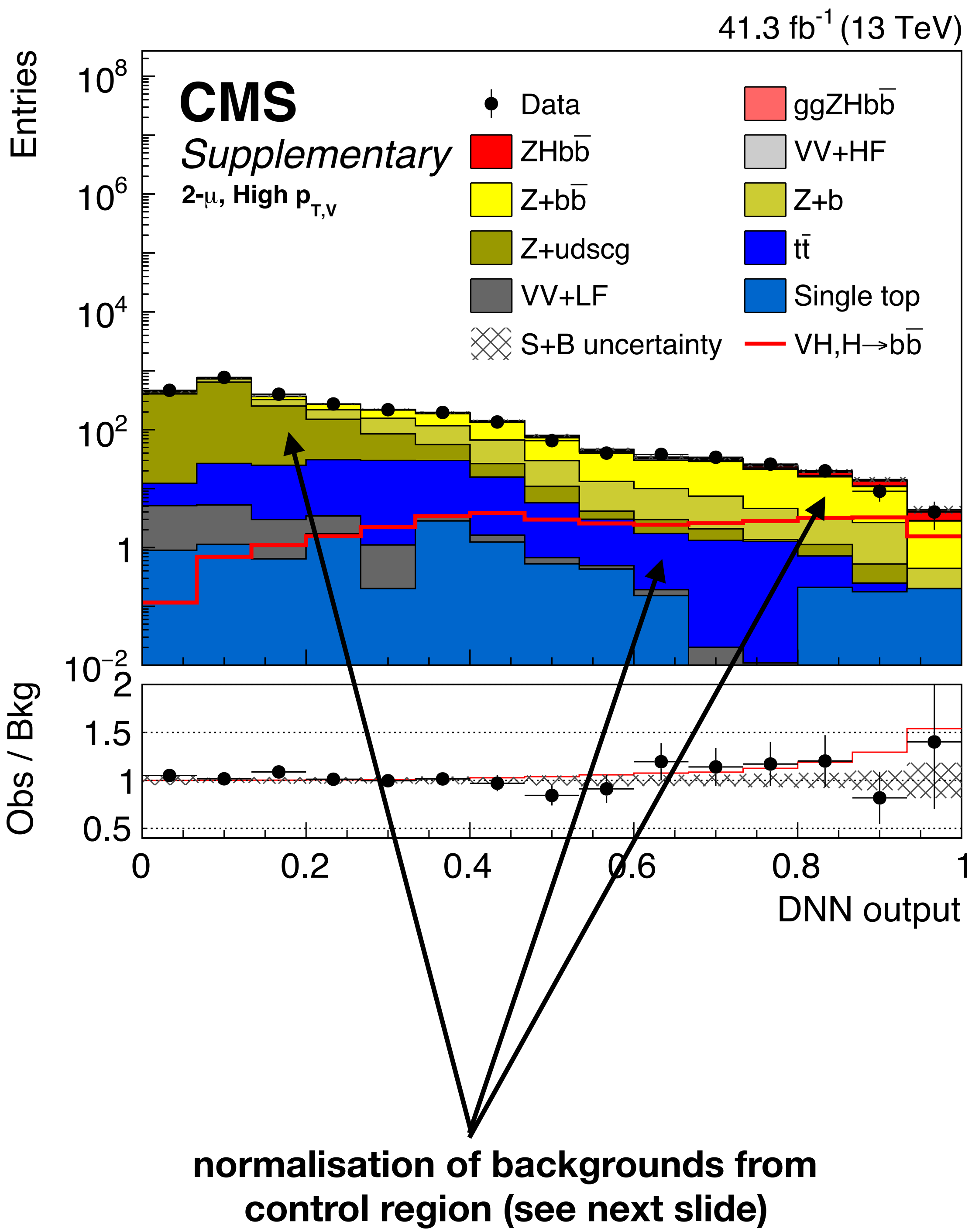
- a **signal region:**
  - to extract the signal strength
- **control regions for TT and V+jets** for
  - validation of DATA/MC agreement
  - constrain normalisation of backgrounds in a simultaneous fit of signal and control regions

## simplified selection scheme




- signal regions still full of background, S/B < 5%
- use MVA discriminator to increase sensitivity

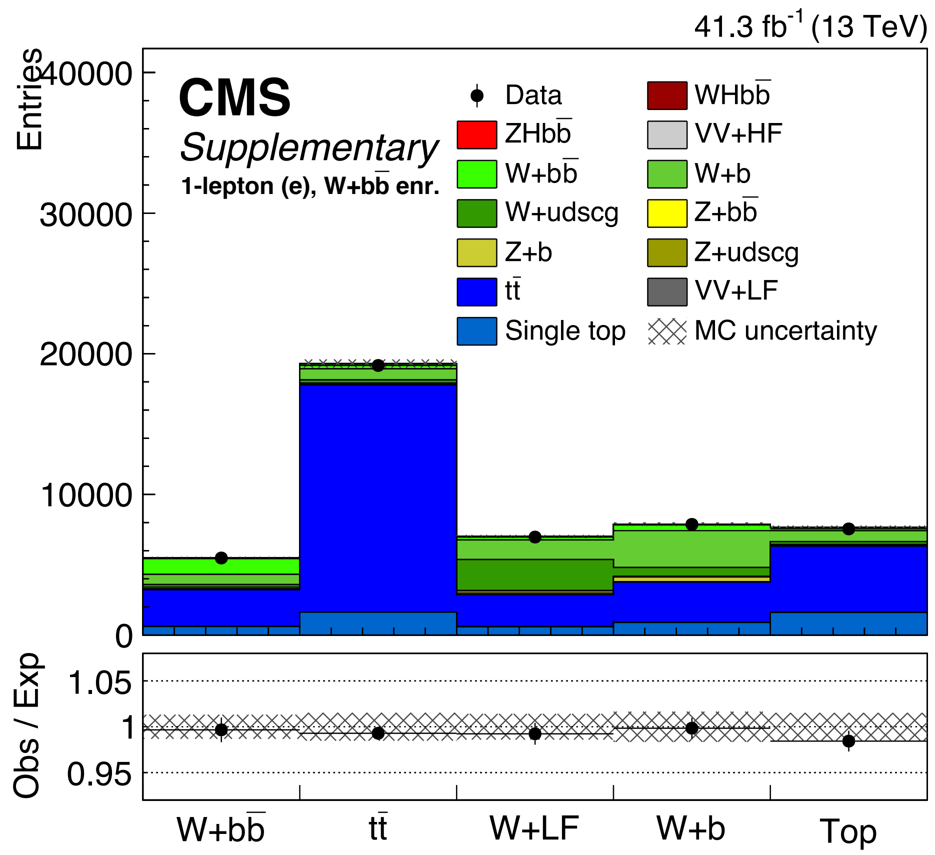
- **DNN signal/background discriminator** trained in the signal regions

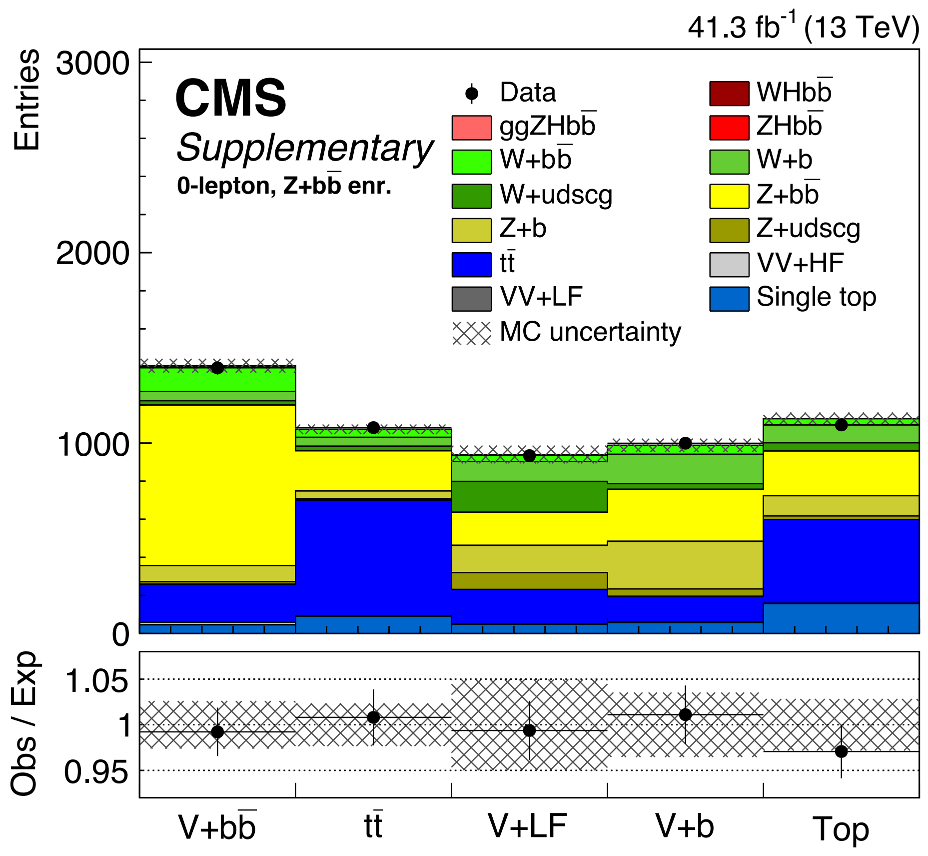



- 16 selected input features validated in data/simulation agreement
  - b-tagging (DeepCSV) score of two Higgs candidate jets
  - kinematic variables: $p_T(H)$, $m_H$, $m_V$, $\Delta R(V, H)$, $E_{T,miss}$ …
  - additional jet information, soft activity
- trained in the signal region only

# Multi-background MVA

- **multi-class DNN**
  - separate the background processes in the **heavy flavour control region**
  - same set of features as the signal/background discriminator
  - trained in heavy flavour control regions

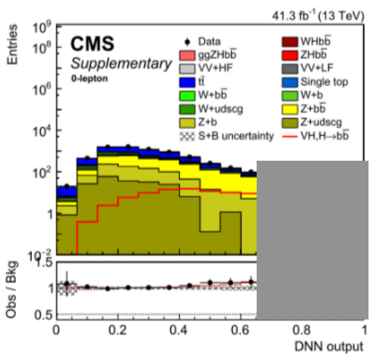


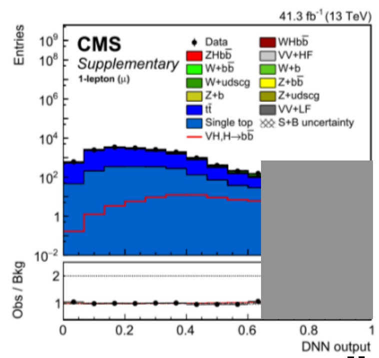
- **what do we fit?**

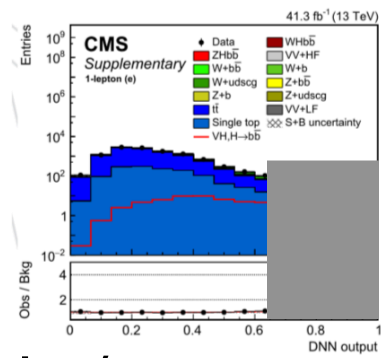
## Signal regions for the 3 channels

**0-lepton**



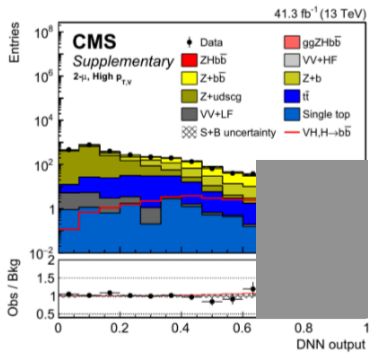
**1-lepton**

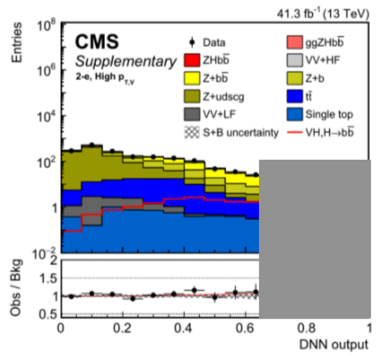


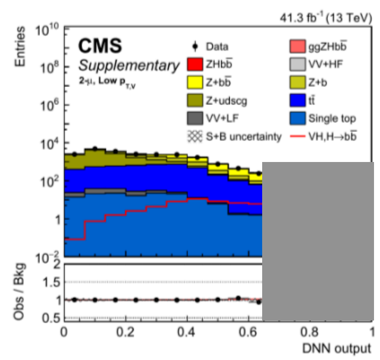


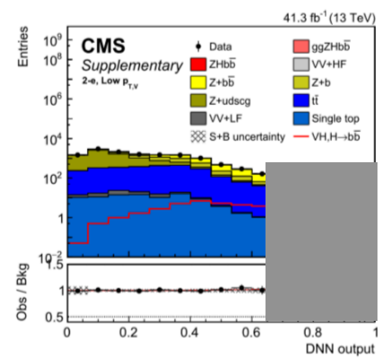
**split in e/μ**

**2-lepton**



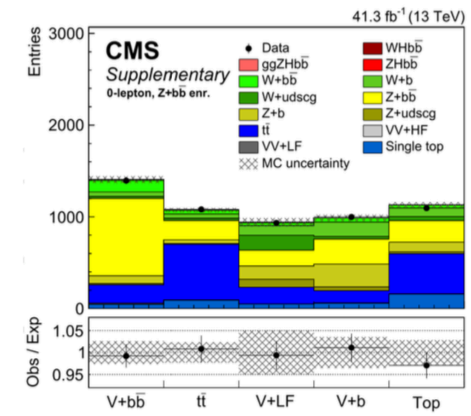




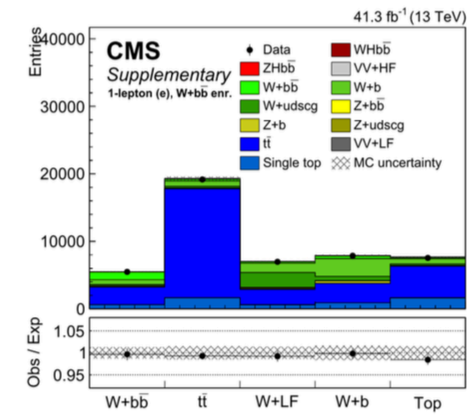


**split in e/μ and two categories of $p_T(V)$**

## Heavy flavour control regions with multi-output classifier

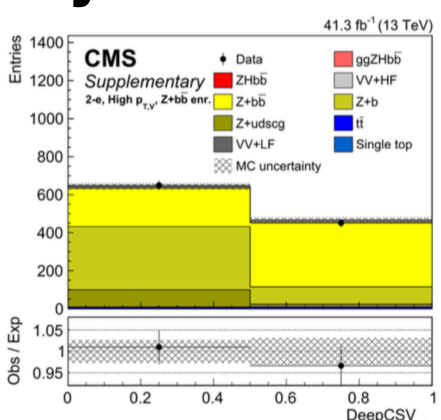


**0-lepton**



**1-lepton**

## Heavy flavour control regions for 2-lepton channel



**2-lepton**

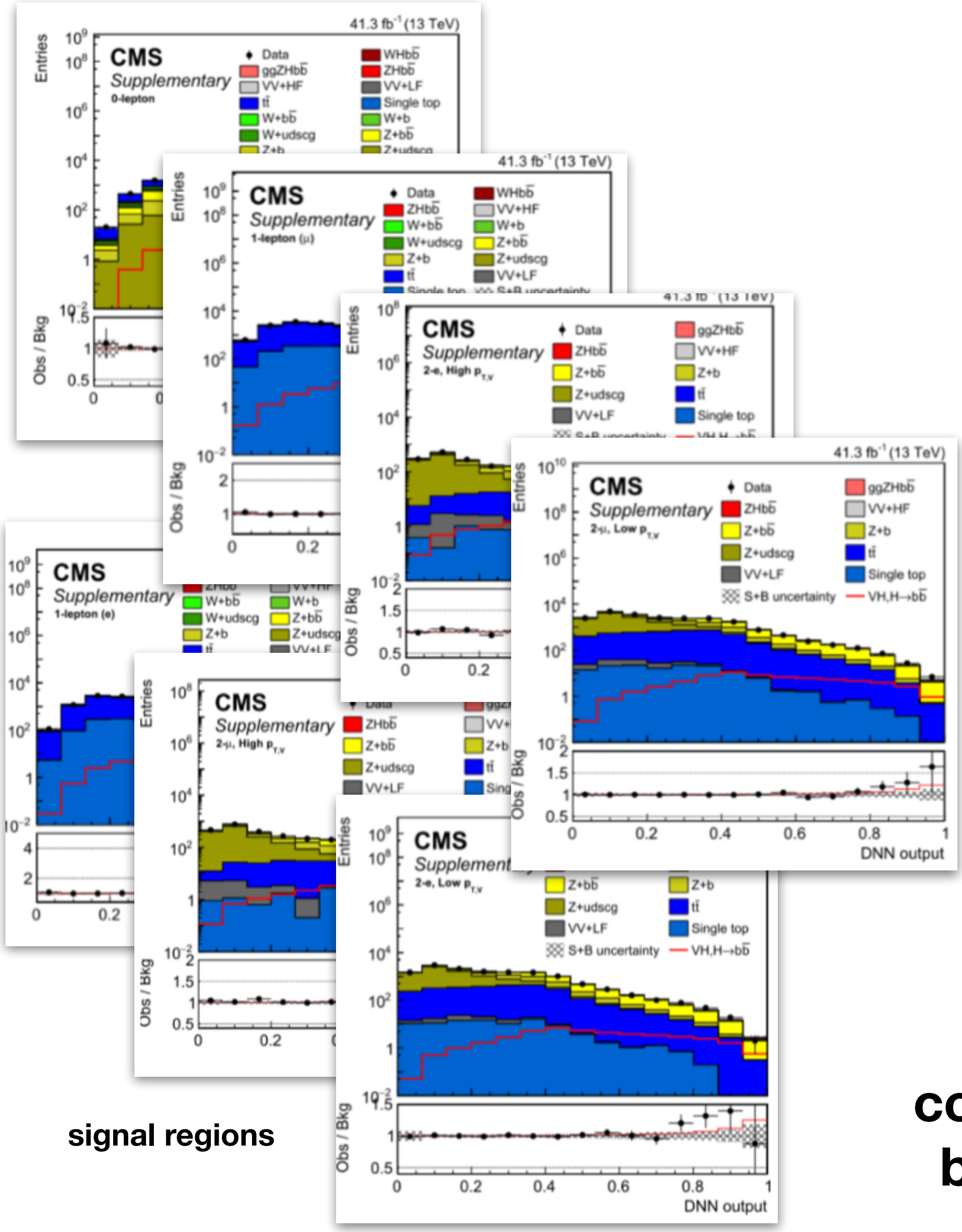
**for 2-lepton channel:**
2-bins of b-tag score instead
(is enough)

**split in e/μ and two categories of $p_T(V)$**

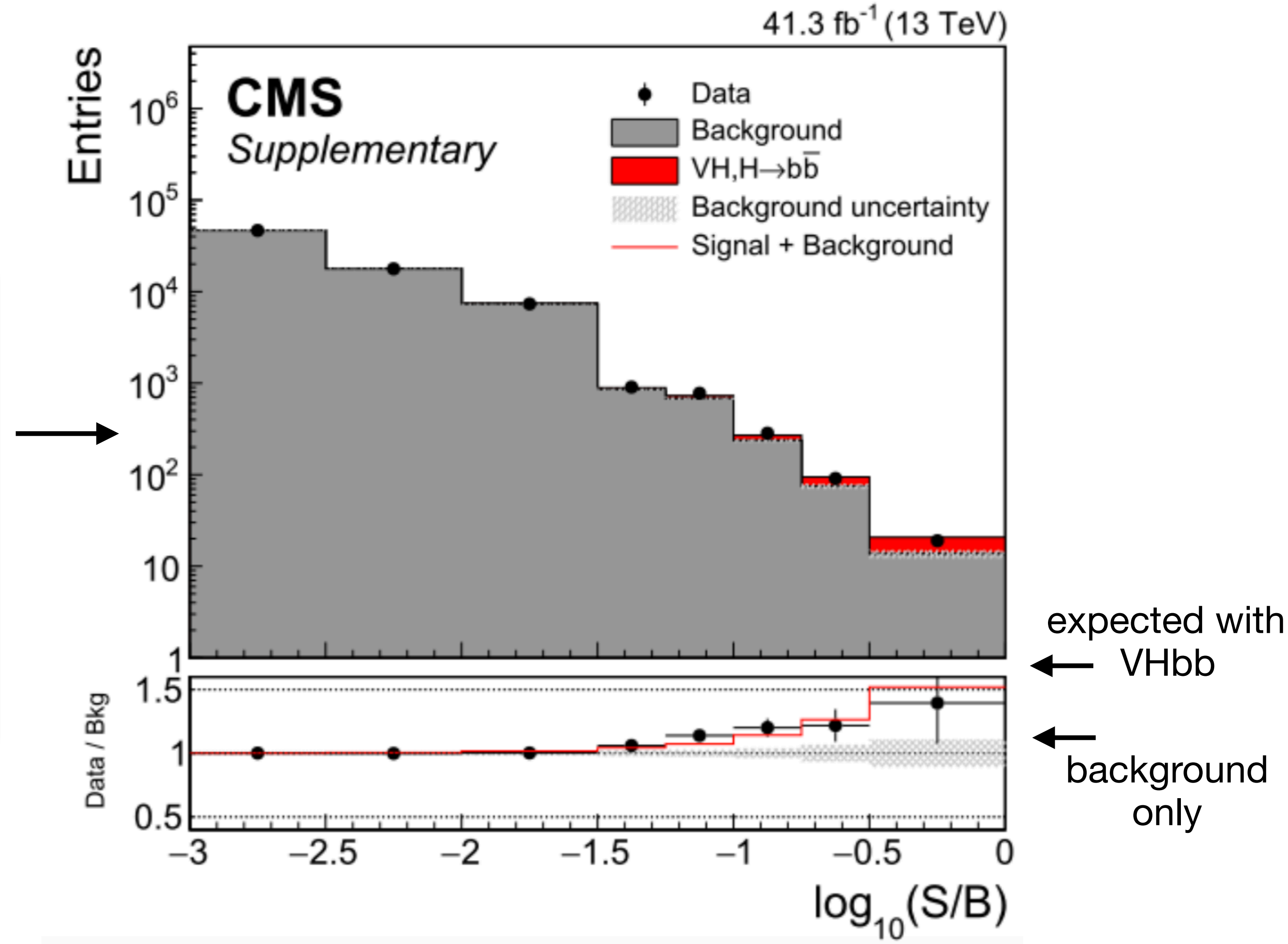
## Light flavour and TT control regions

- fit normalisation only for remaining 16 control regions

- ~250 sources of systematic uncertainties also included in the fit




**2017 only**

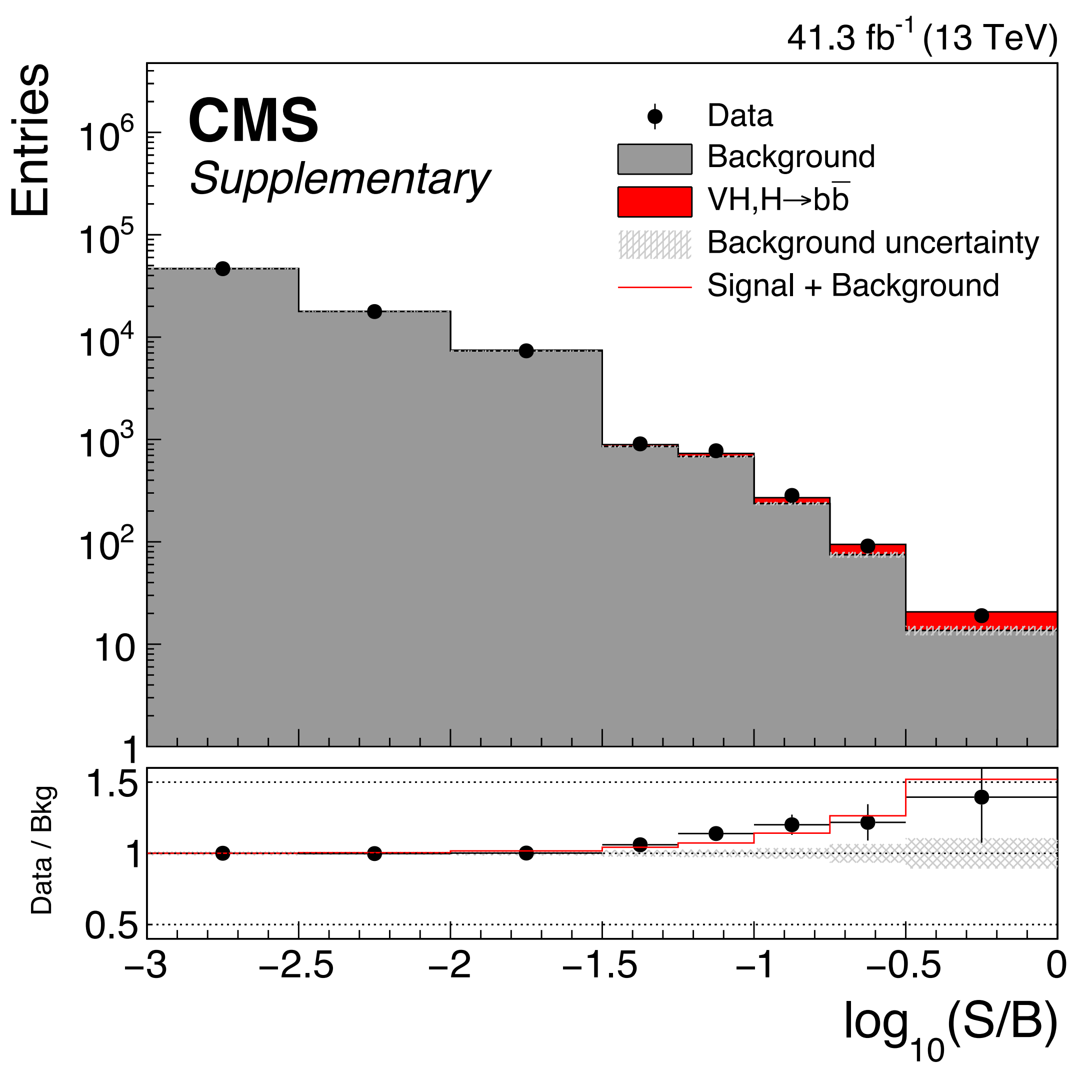


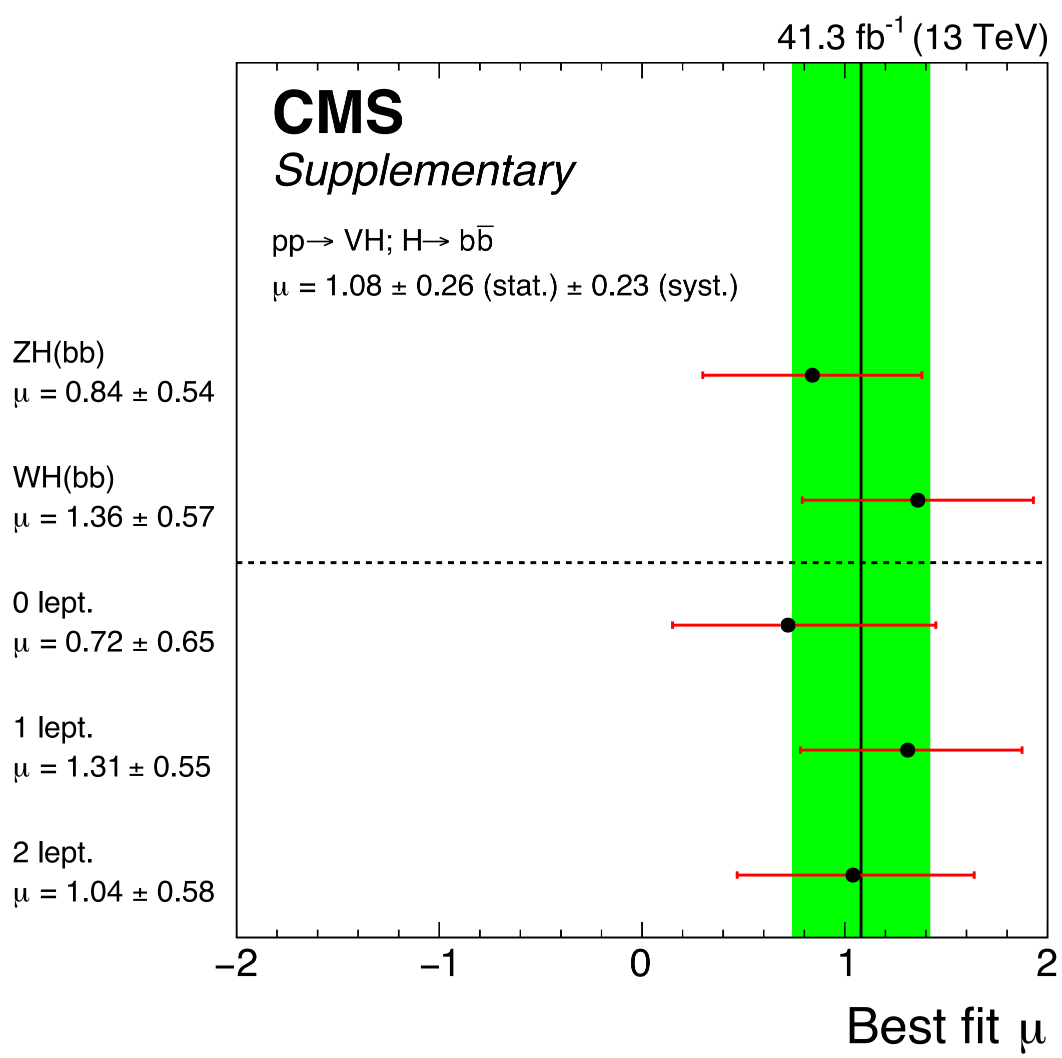

**signal regions**

**combine result of all 7 signal regions by combining bins with similar S/B**

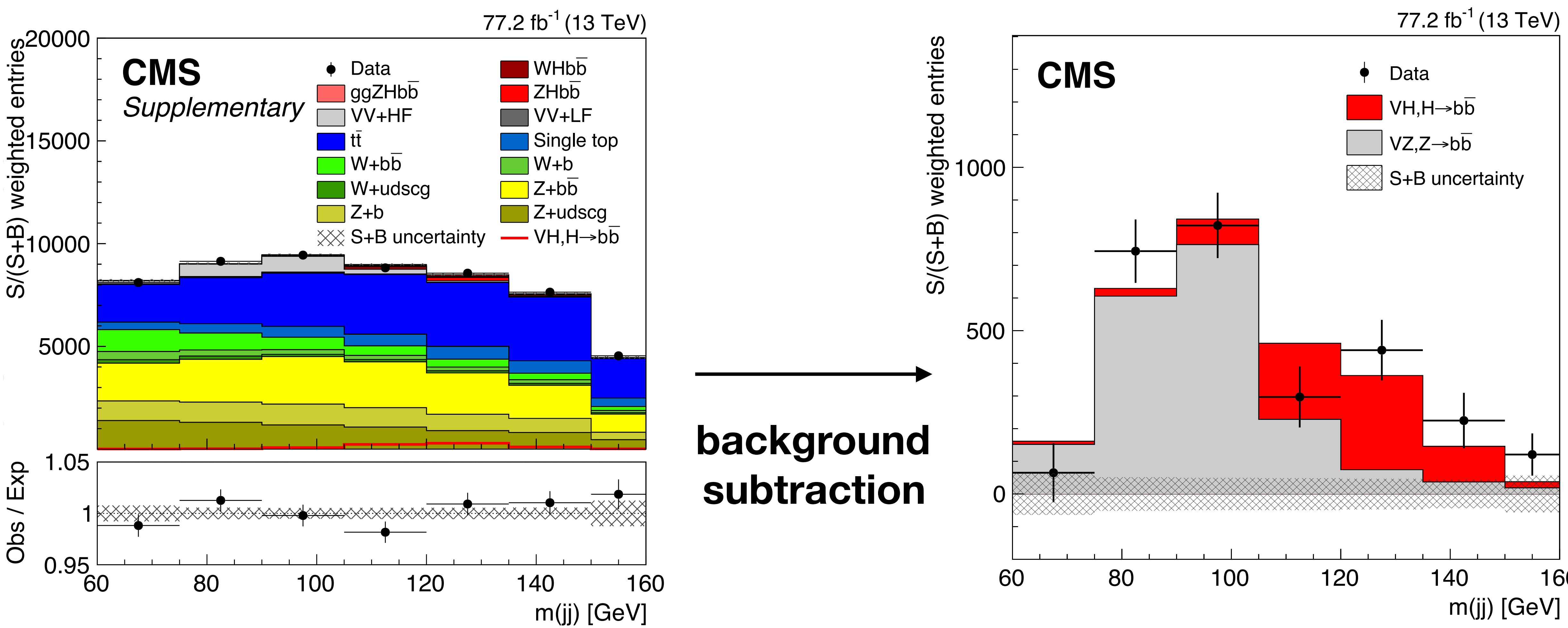
# Results for 2017 data only

| Data set | Significance ($\sigma$) Expected | Observed | Signal strength |
|---|---|---|---|
| 2017 | | | |
| 0-lepton | 1.9 | 1.3 | $0.73 \pm 0.65$ |
| 1-lepton | 1.8 | 2.6 | $1.32 \pm 0.55$ |
| 2-lepton | 1.9 | 1.9 | $1.05 \pm 0.59$ |
| Combined | 3.1 | 3.3 | $1.08 \pm 0.34$ |



- instead of DNN score one can also fit $m_{jj}$
- less sensitive, but more directly interpretable variable used in the fit




- MVA discriminator evaluated with mass and correlated variables set to mean value
- 4 categories in this MVA discriminator with increasing S/B used in the fit

# Systematic uncertainties

- in combination of Run1 with 2016 and 2017 data, systematic uncertainties of similar impact as statistical uncertainties

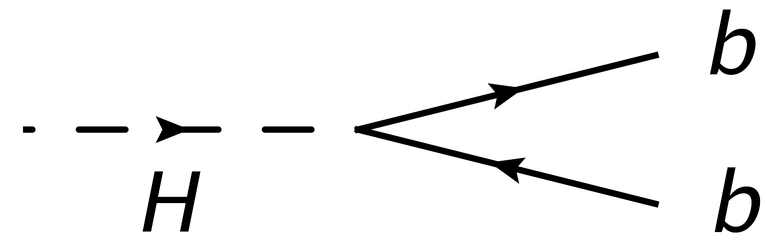
Run1 + 2016 + 2017: μ= 1.01 ± 0.17 (stat) ± 0.14 (syst)

- uncertainties (for 2017 analysis only):

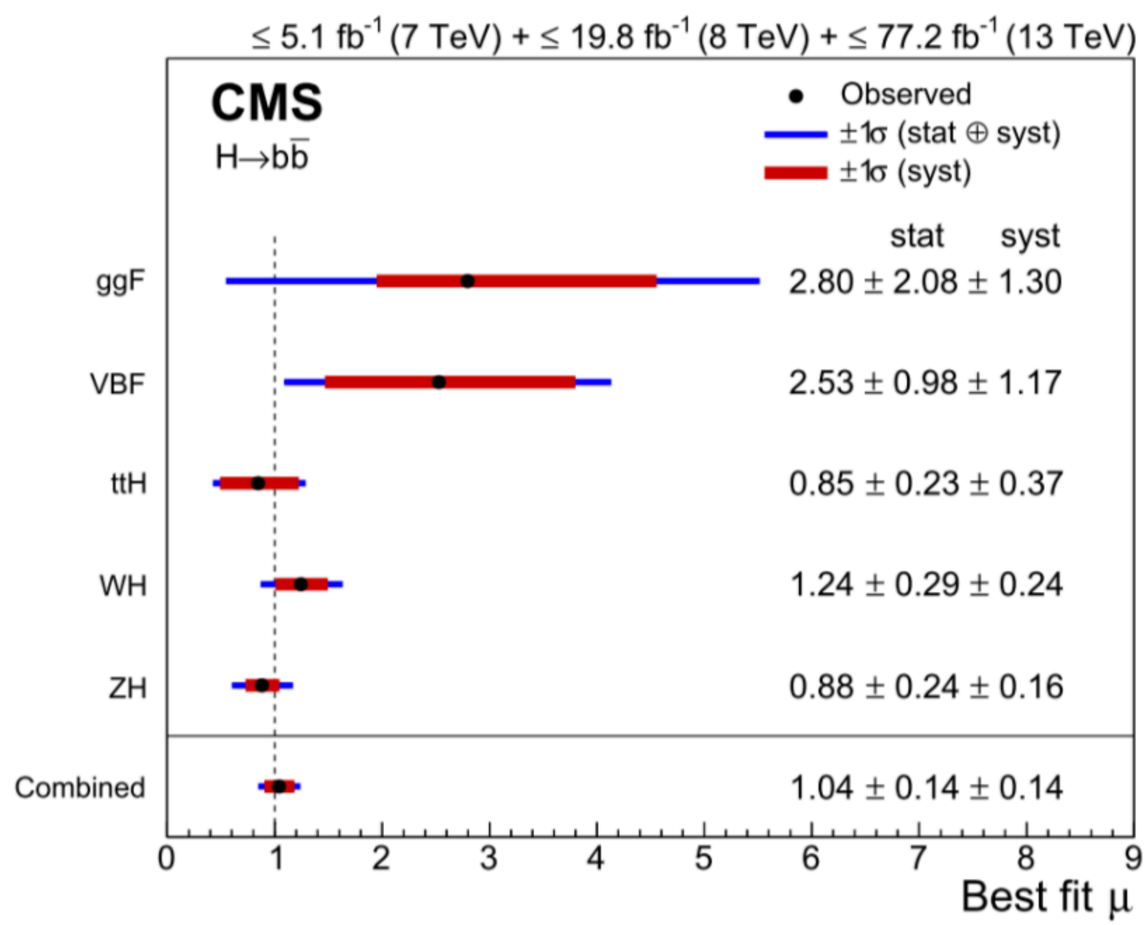
| Uncertainty source | $\Delta\mu$ | |
|---|---|---|
| Statistical | +0.26 | −0.26 |
| Normalization of backgrounds | +0.12 | −0.12 |
| Experimental | +0.16 | −0.15 |
| b-tagging efficiency and misid | +0.09 | −0.08 |
| V+jets modeling | +0.08 | −0.07 |
| Jet energy scale and resolution | +0.05 | −0.05 |
| 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 |

# Combination with other production channels

remember we wanted to look at the decay H->bb




- can combine results of VH(bb) with CMS results for other production channels:
  - ttH (H in association with $t\bar{t}$ pair)
  - VBF (vector boson fusion)
  - boosted ggH (gluon fusion)




<u>significance</u>

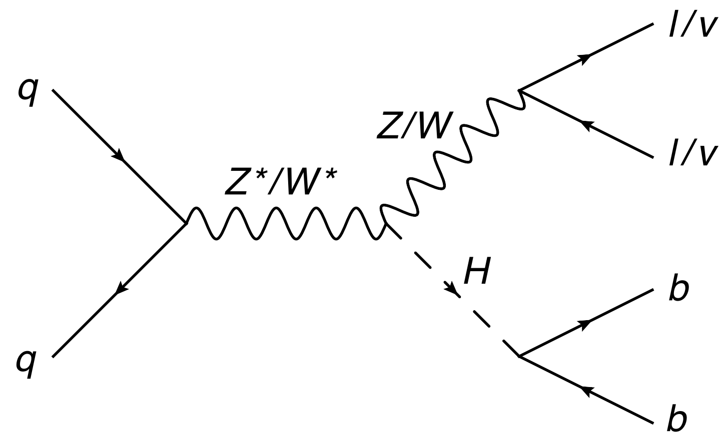
5.5 σ expected

**5.6 σ observed**

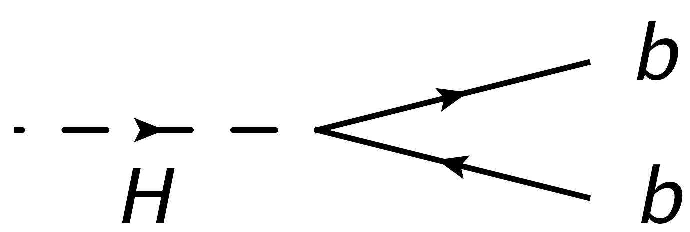
**Observation of the H → bb decay by the CMS Collaboration**

# Summary and outlook

- VHbb analysis of 2017 data completed
  - observed (expected) significance of 3.3 (3.1) σ
  - $\mu = 1.08 \pm 0.34$
- combination of all **VHbb** data from CMS is at 4.8 (4.9) σ
- $\mu = 1.01 \pm 0.22$, compatible with Standard Model




- combination with other Higgs production channels done
  - **first observation (5.6 σ) of Hbb at CMS!**



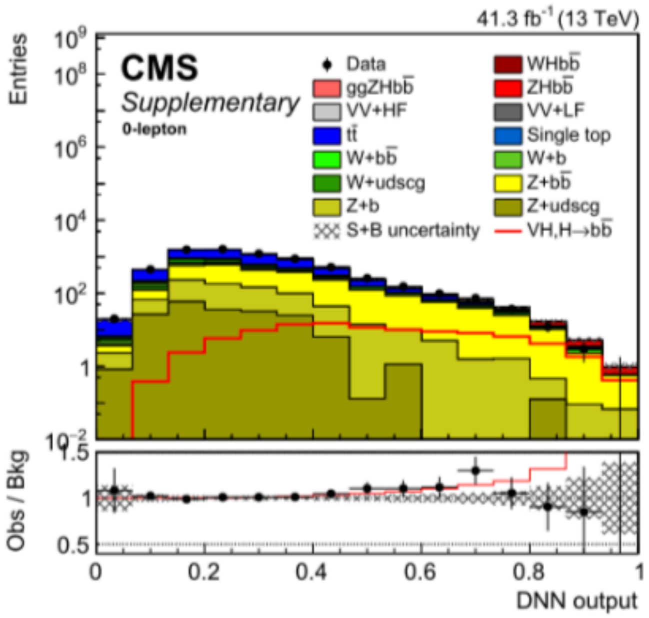

- future plans include analysis of full Run 2 data (include 2018) and a **differential analysis** (in pT, number of jets, STXS) to increase sensitivity for certain BSM models

# Backup

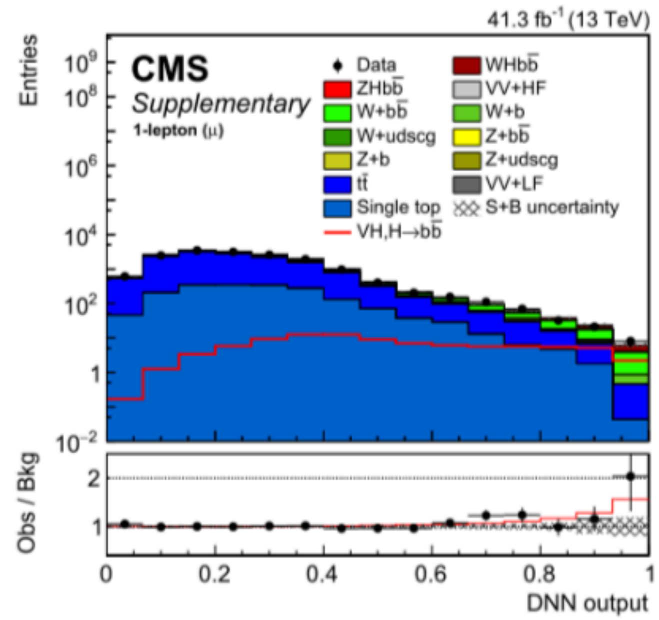
# DNN post-fit

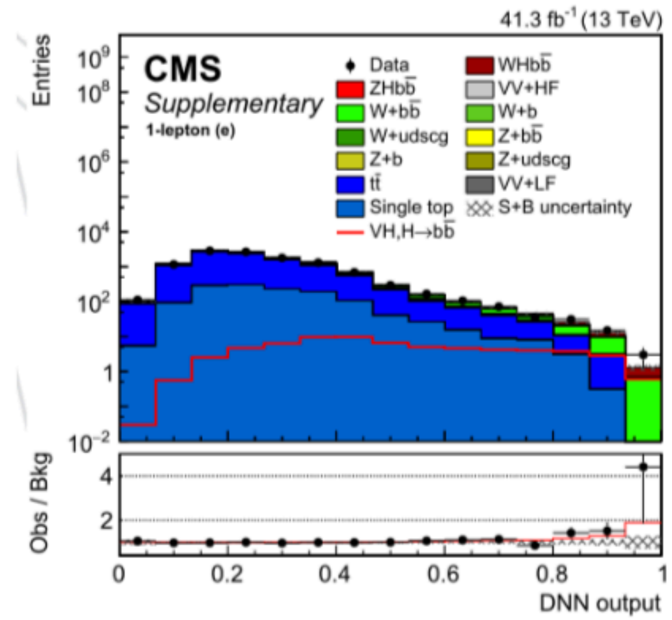
- **result of the fit:**

## 0-lepton

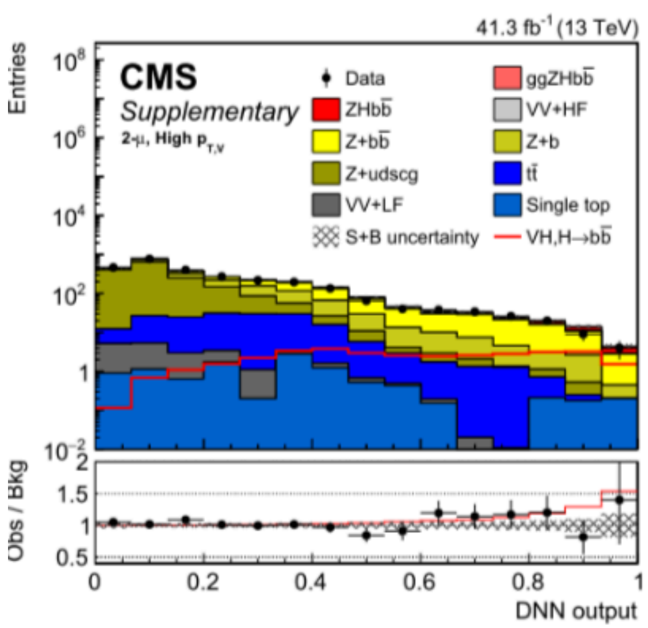


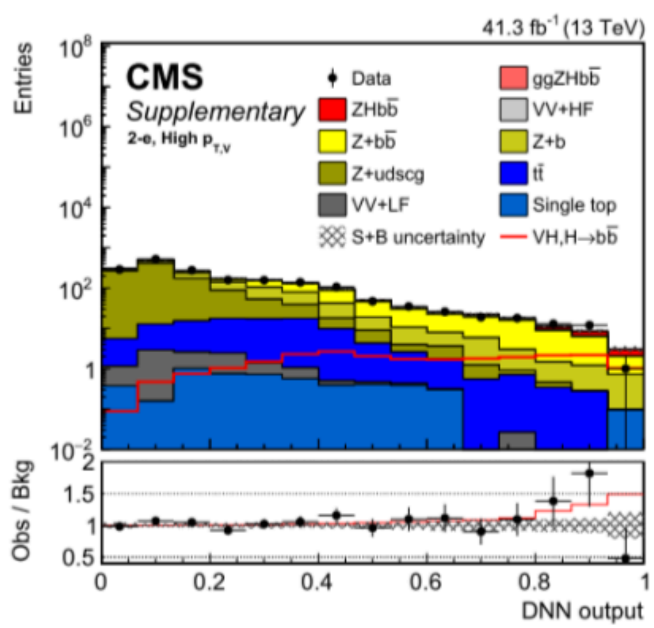

## 1-lepton

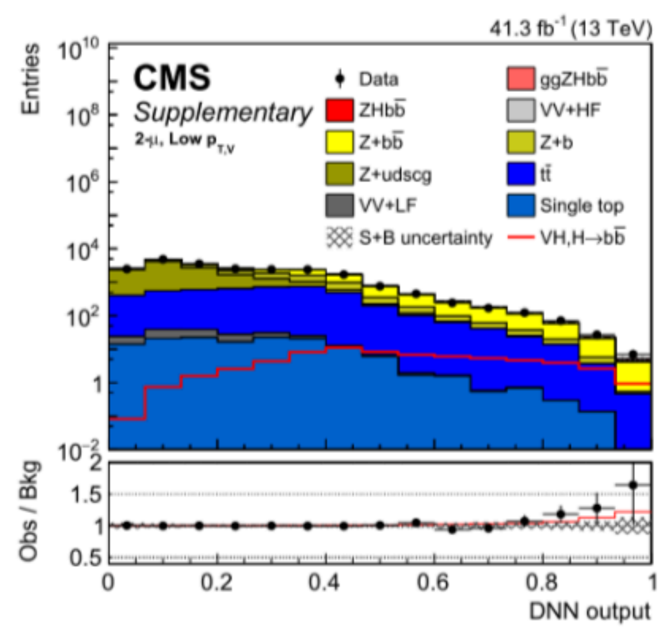


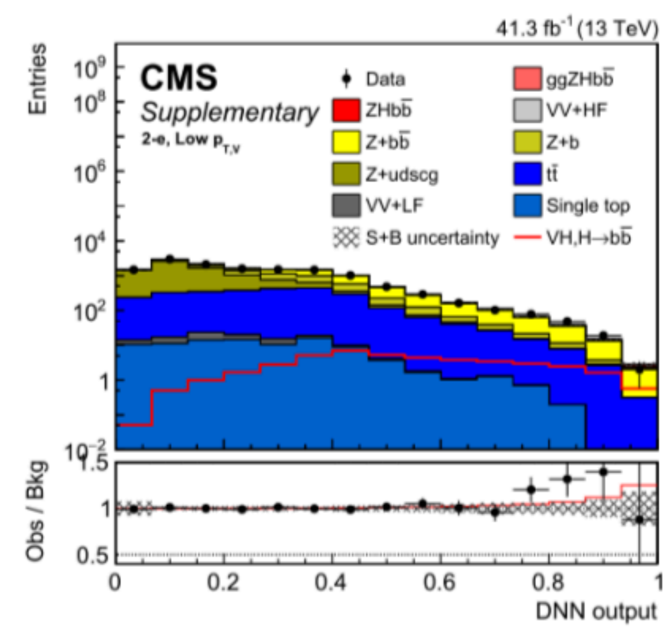




## 2-lepton

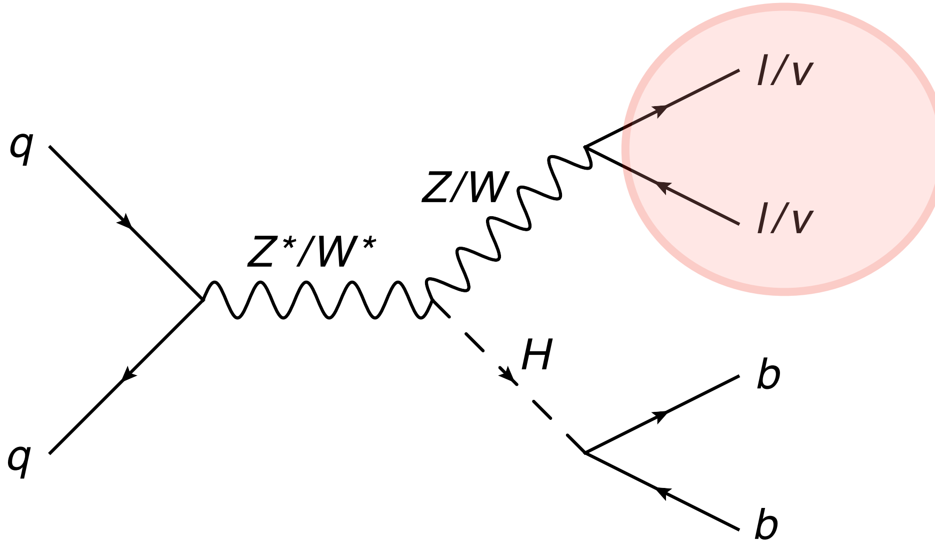


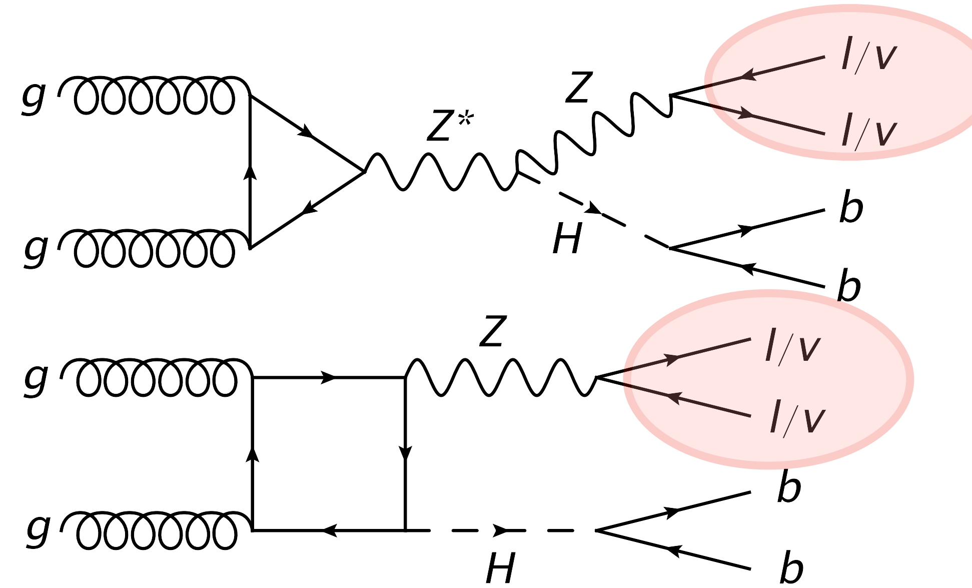



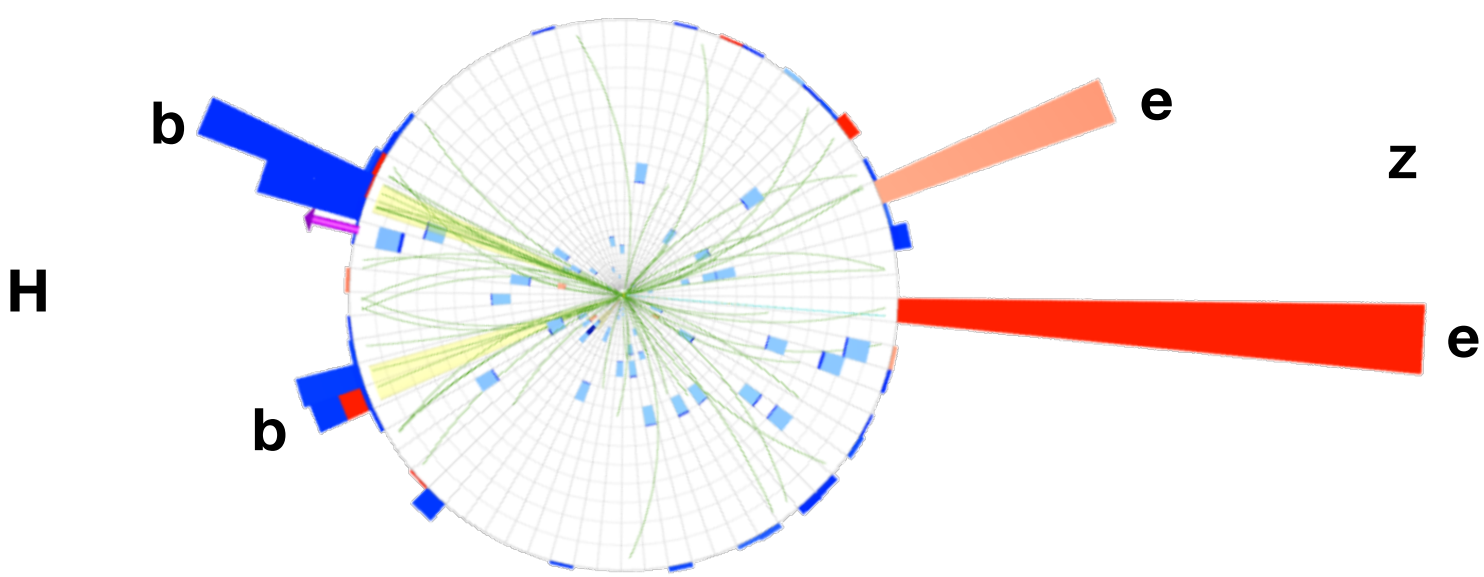



# Associated production with a vector boson (VH)



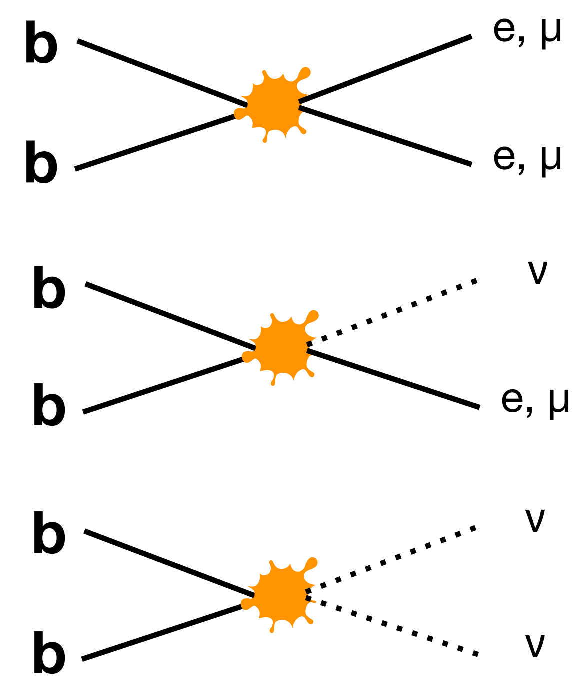




l = e or μ




VHbb 2-lepton candidate event display

**3-channels**



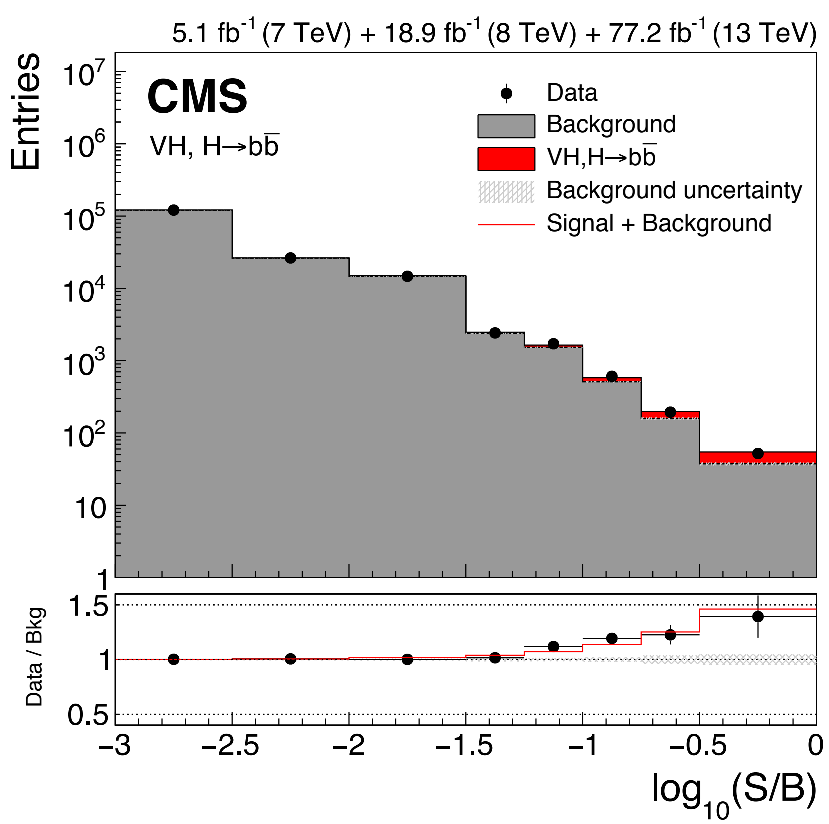

**2-lepton**

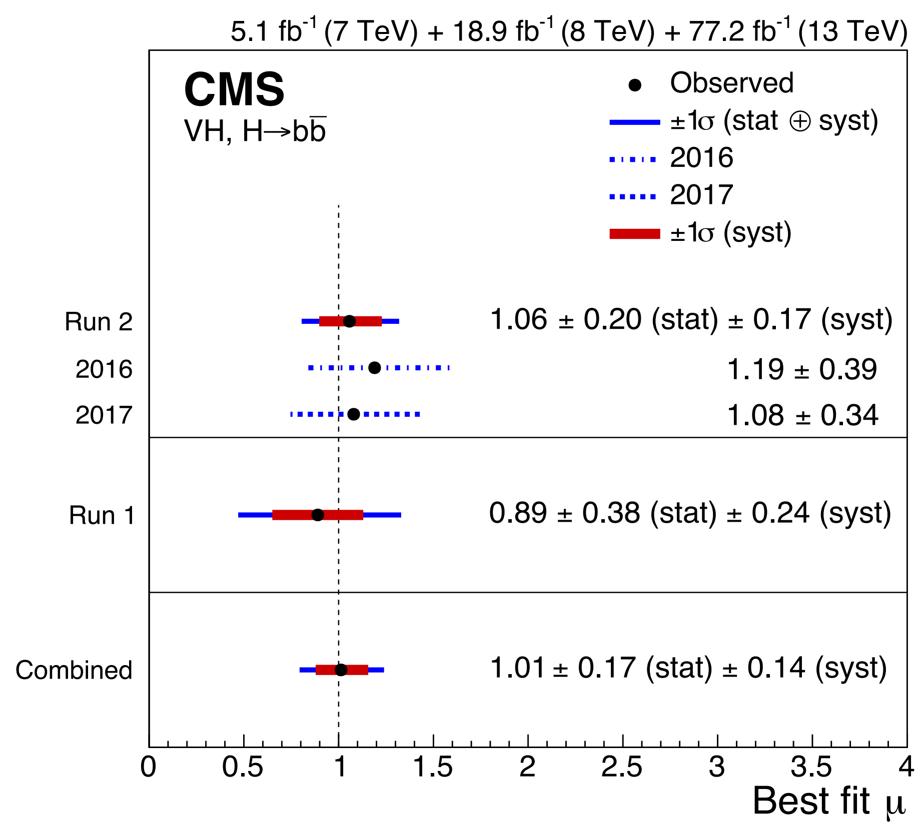
**1-lepton**

**0-lepton**

# Combination with previous runs

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



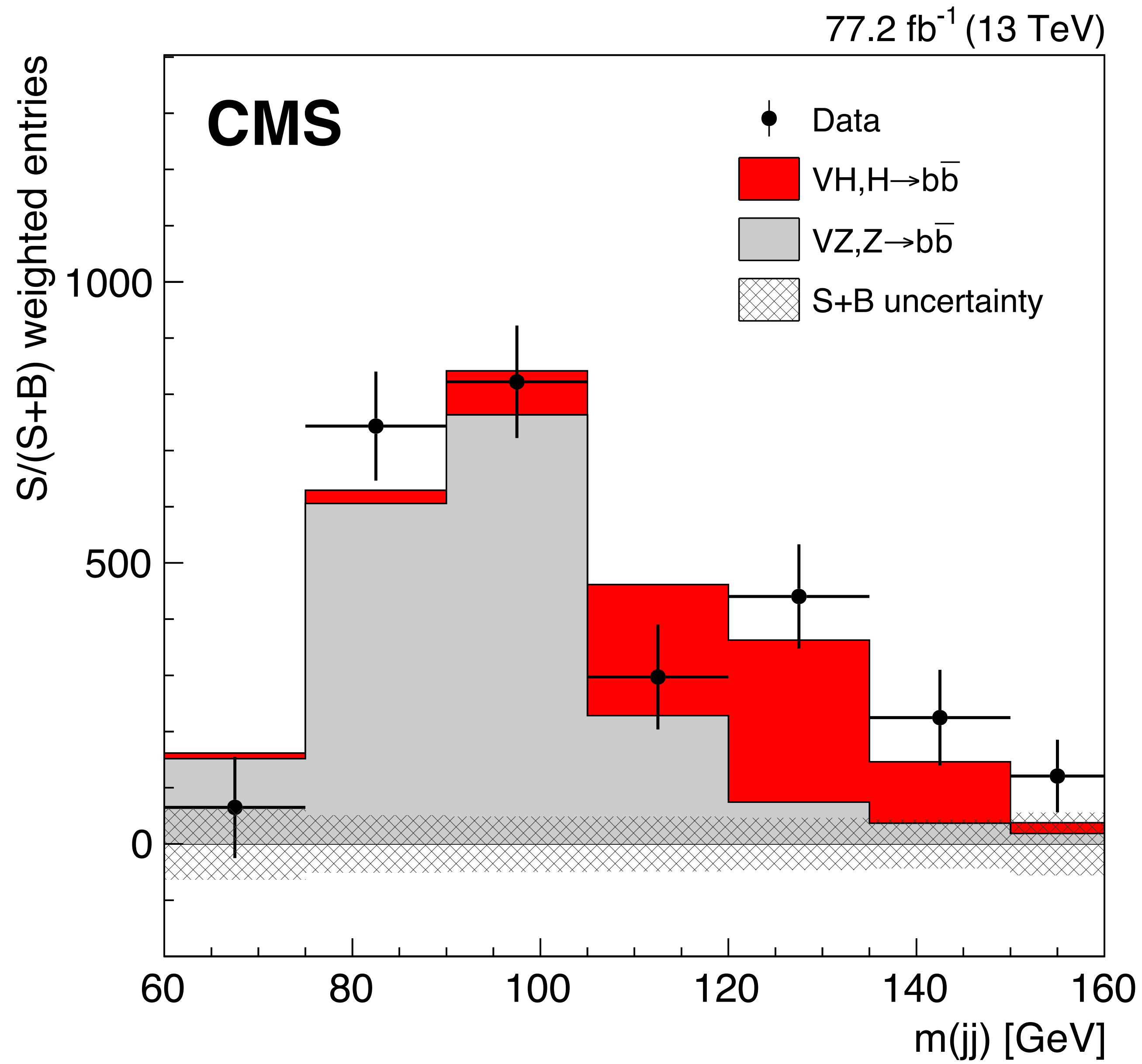



Figure 2: Dijet invariant mass distribution for events weighted by $S/(S+B)$ in all channels combined in the 2016 and 2017 data sets. Weights are derived from a fit to the $m$(jj) distribution, as described in the text. Shown are data (points) and the fitted VH signal (red) and VZ background (grey) distributions, with all other fitted background processes subtracted. The error bar for each bin represents the pre-subtraction $1\sigma$ statistical uncertainty on the data, while the grey hatching indicates the $1\sigma$ total uncertainty on the signal and all background components.

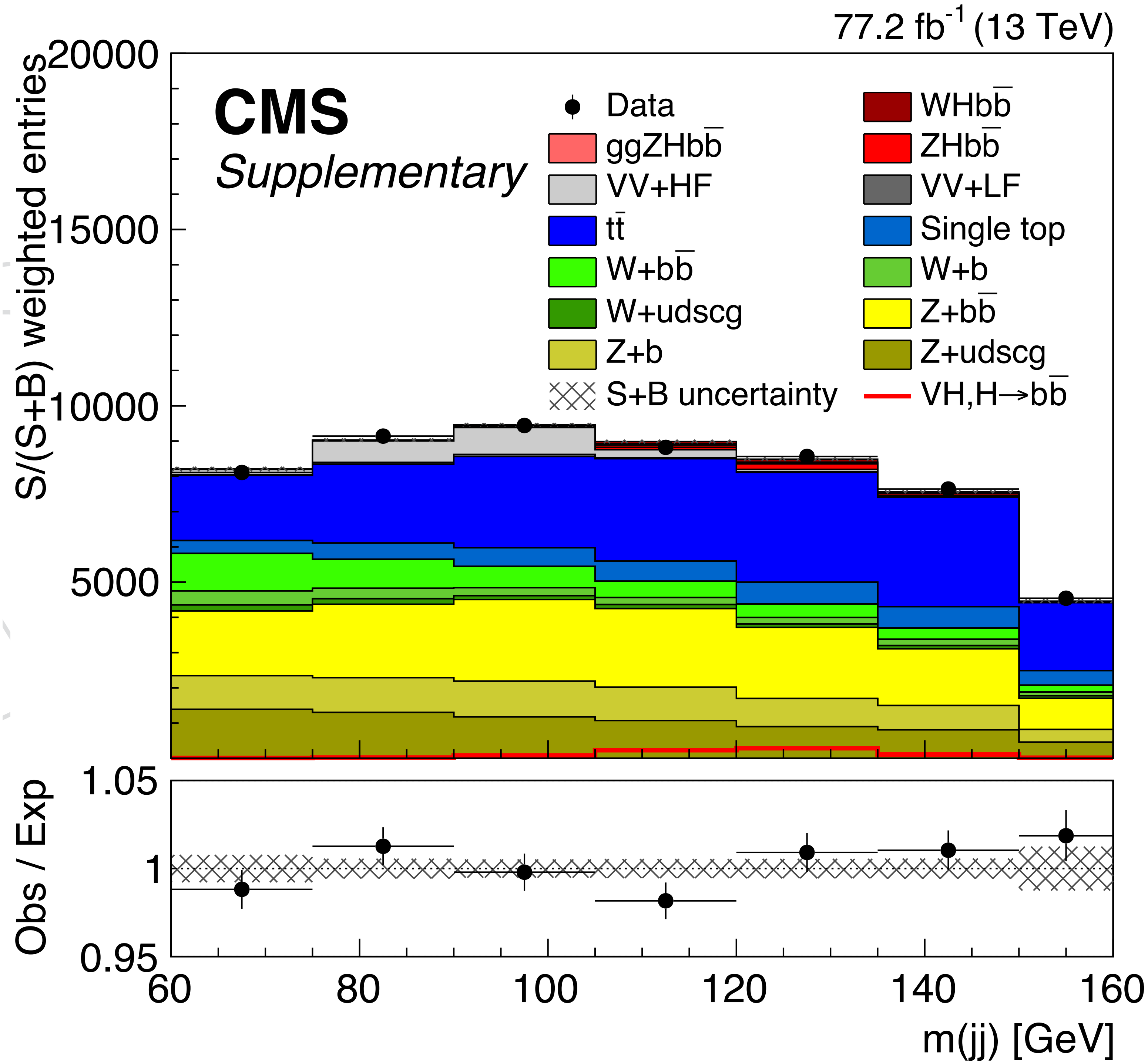
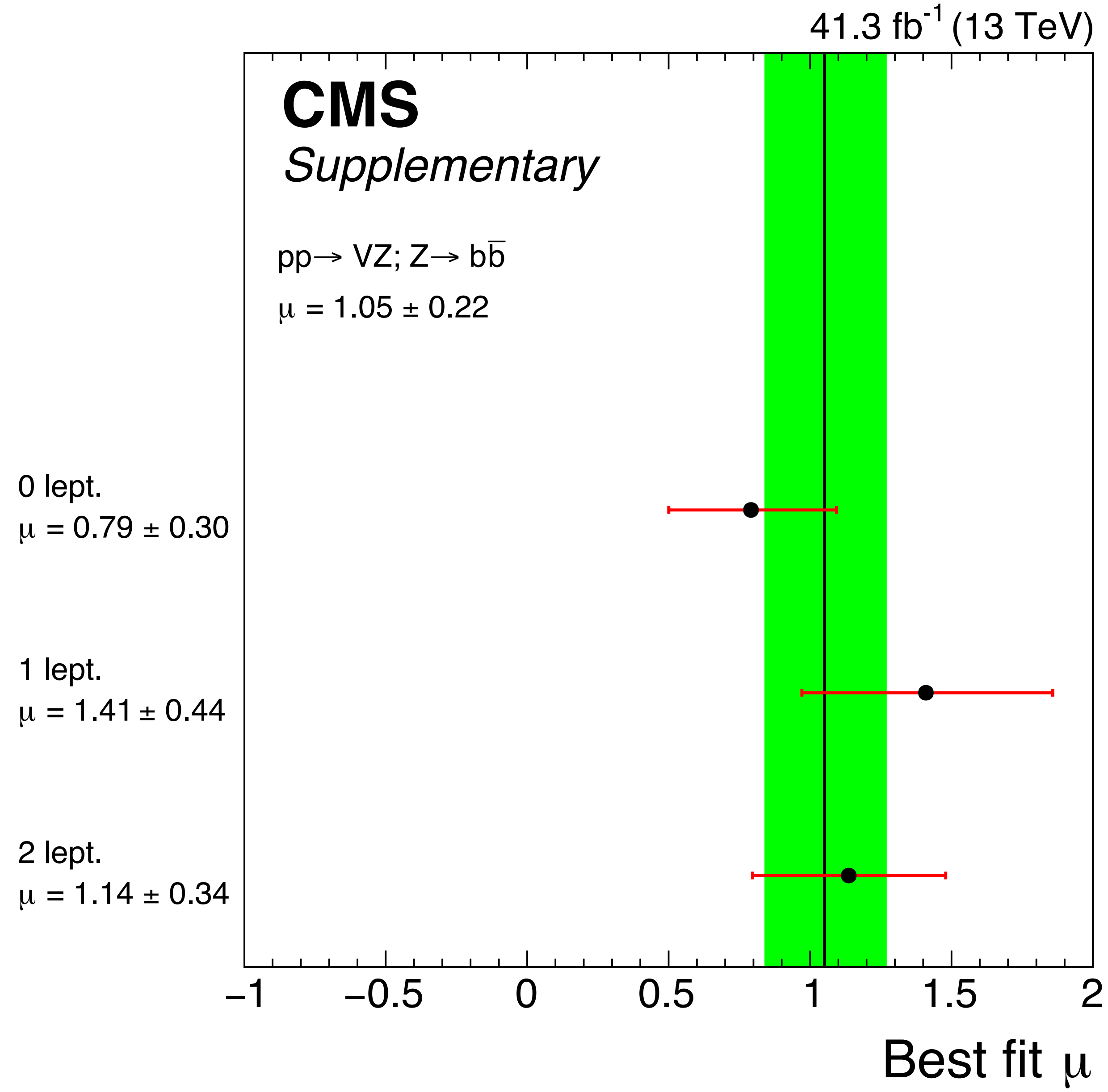


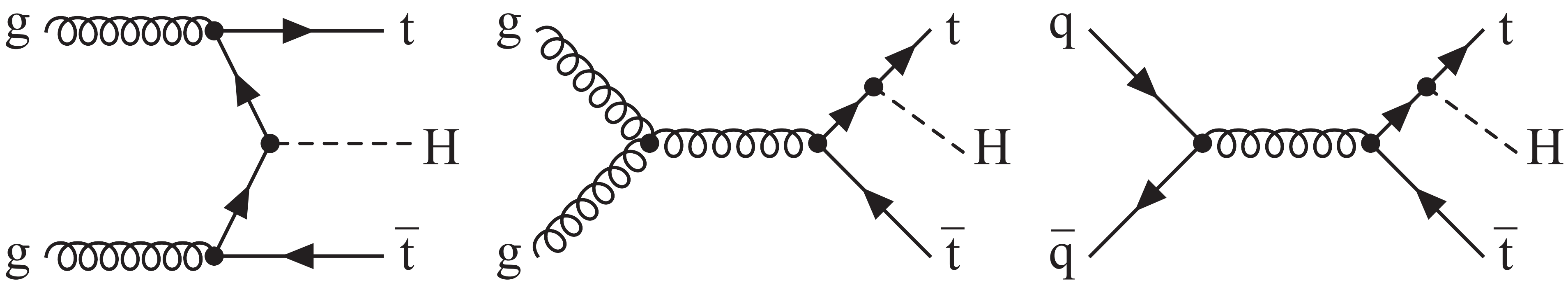

Figure 3: Dijet invariant mass distribution for events weighted by $S/(S+B)$ in all channels combined in the 2016 and 2017 data sets. Weights are derived from a fit to the $m(\text{jj})$ distribution, as described in the text. Shown are data (points) and the fitted VH signal (red) and VZ background (grey) distributions, as well as all other backgrounds.

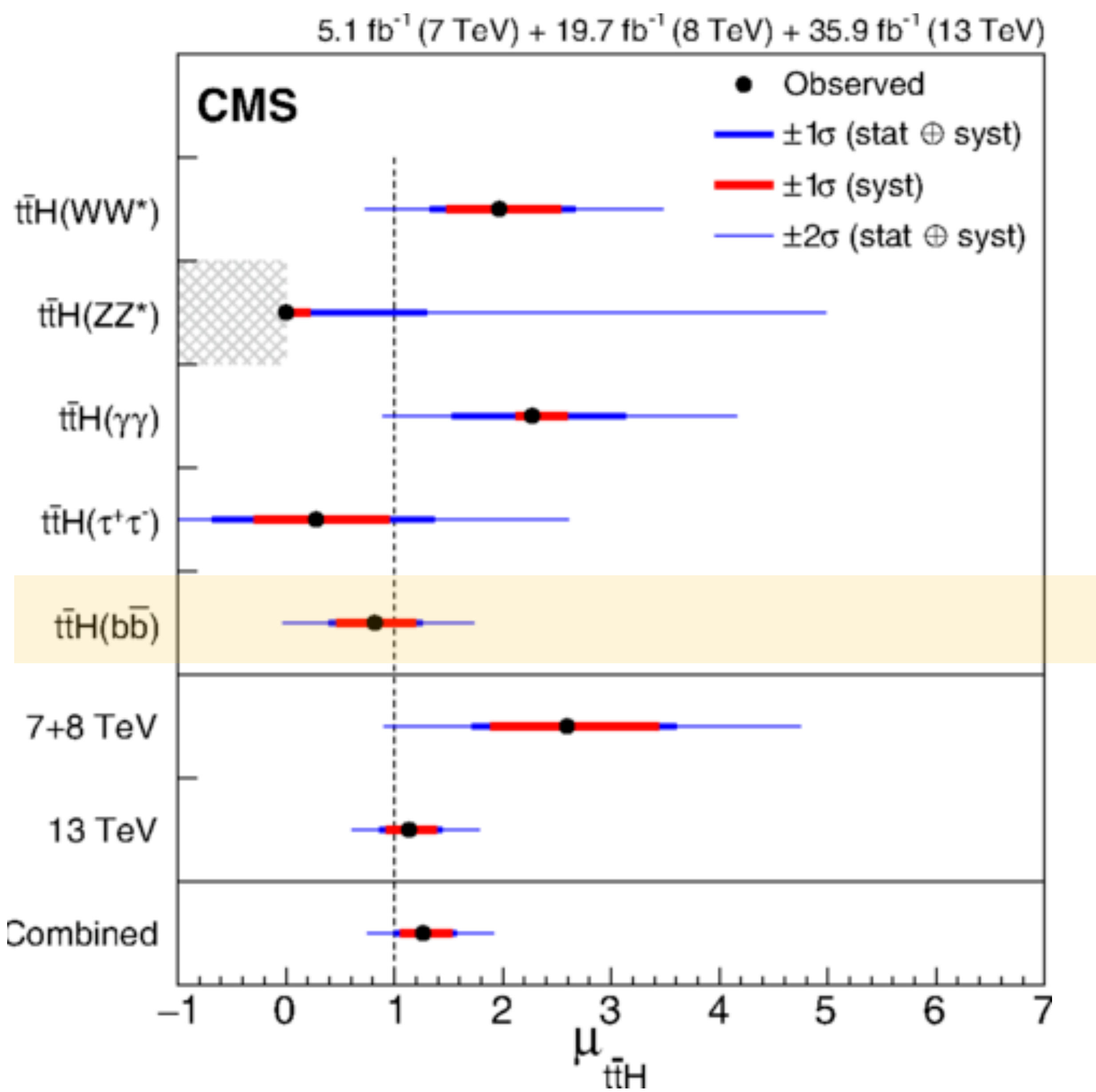



- analysis for VZ with Z -> bb is done as cross-check to validate tools used for VHbb analysis
- signal has same bb final state (apart from invariant mass) but has higher cross-section
- observed (expected) significance of 5.2 (5.0) $\sigma$ , signal strength: $\mu$ = 1.05 ± 0.22

# CMS-HIG-17-035

Phys. Rev. Lett. 120 (2018) 231801



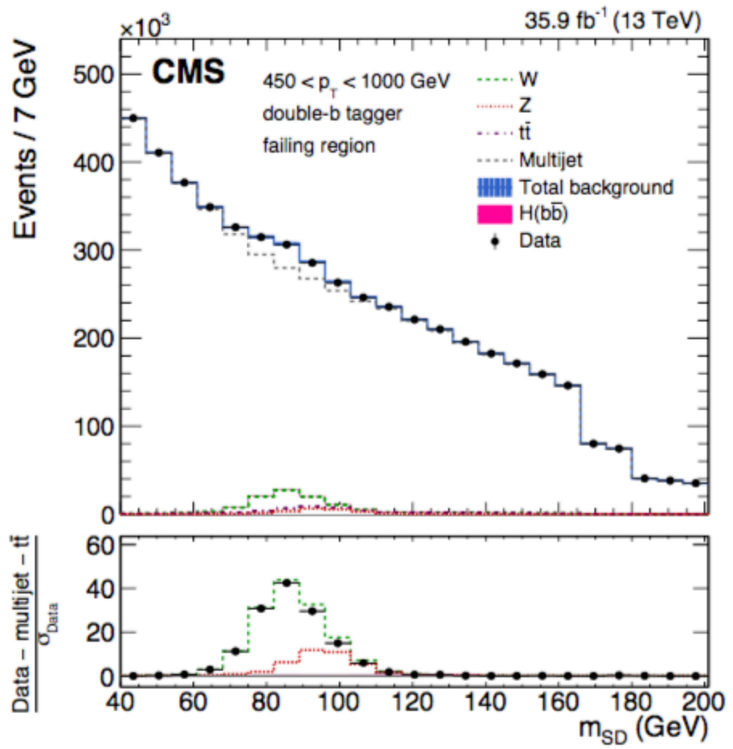

| Parameter | Best fit | Uncertainty: Statistical | Experimental | Background theory | Signal theory |
|---|---|---|---|---|---|
| $\mu_{t\bar{t}H}^{WW^*}$ | $1.97^{+0.71}_{-0.64}$ | $^{+0.42}_{-0.41}$ | $^{+0.46}_{-0.42}$ | $^{+0.21}_{-0.21}$ | $^{+0.25}_{-0.12}$ |
| | $\left(^{+0.57}_{-0.54}\right)$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

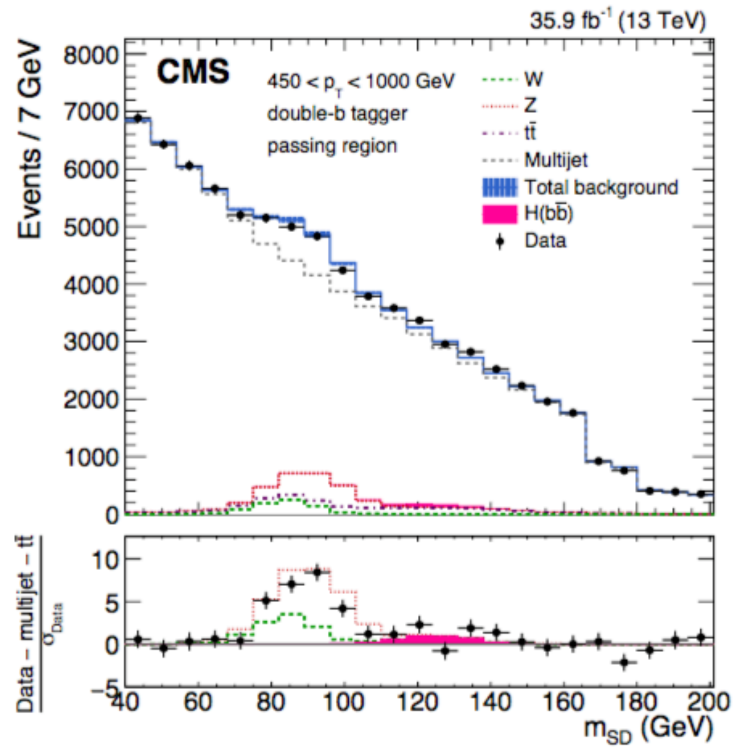
**CMS-HIG-17-010**

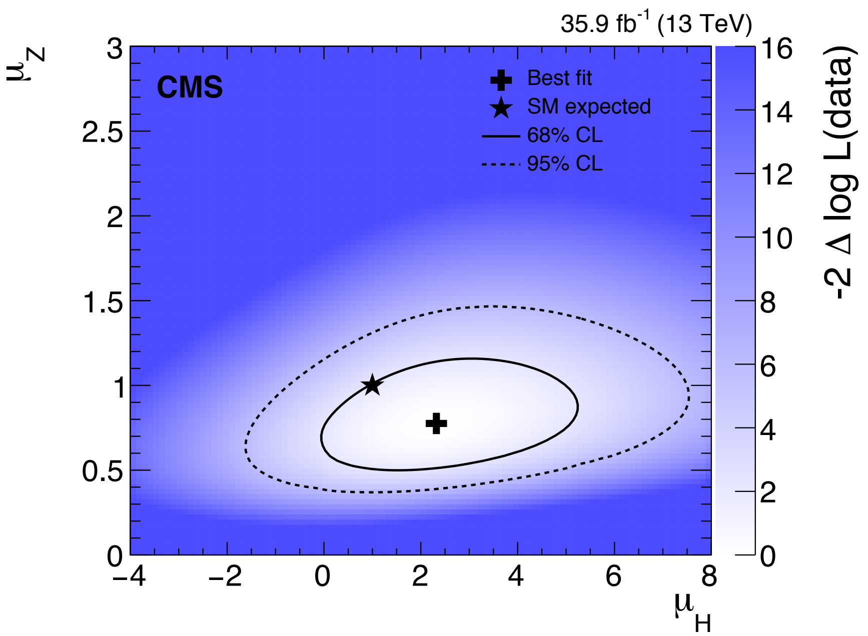
10.1103/PhysRevLett.120.071802

## Abstract

An inclusive search for the standard model Higgs boson (H) produced with large transverse momentum ($p_T$) and decaying to a bottom quark-antiquark pair ($b\bar{b}$) is performed using a data set of pp collisions at $\sqrt{s} = 13$ TeV collected with the CMS experiment at the LHC. The data sample corresponds to an integrated luminosity of 35.9 fb$^{-1}$. A highly Lorentz-boosted Higgs boson decaying to $b\bar{b}$ is reconstructed as a single, large radius jet and is identified using jet substructure and dedicated b tagging techniques. The method is validated with $Z \to b\bar{b}$ decays. The $Z \to b\bar{b}$ process is



**CMS-HIG-14-004**

10.1103/PhysRevD.92.032008

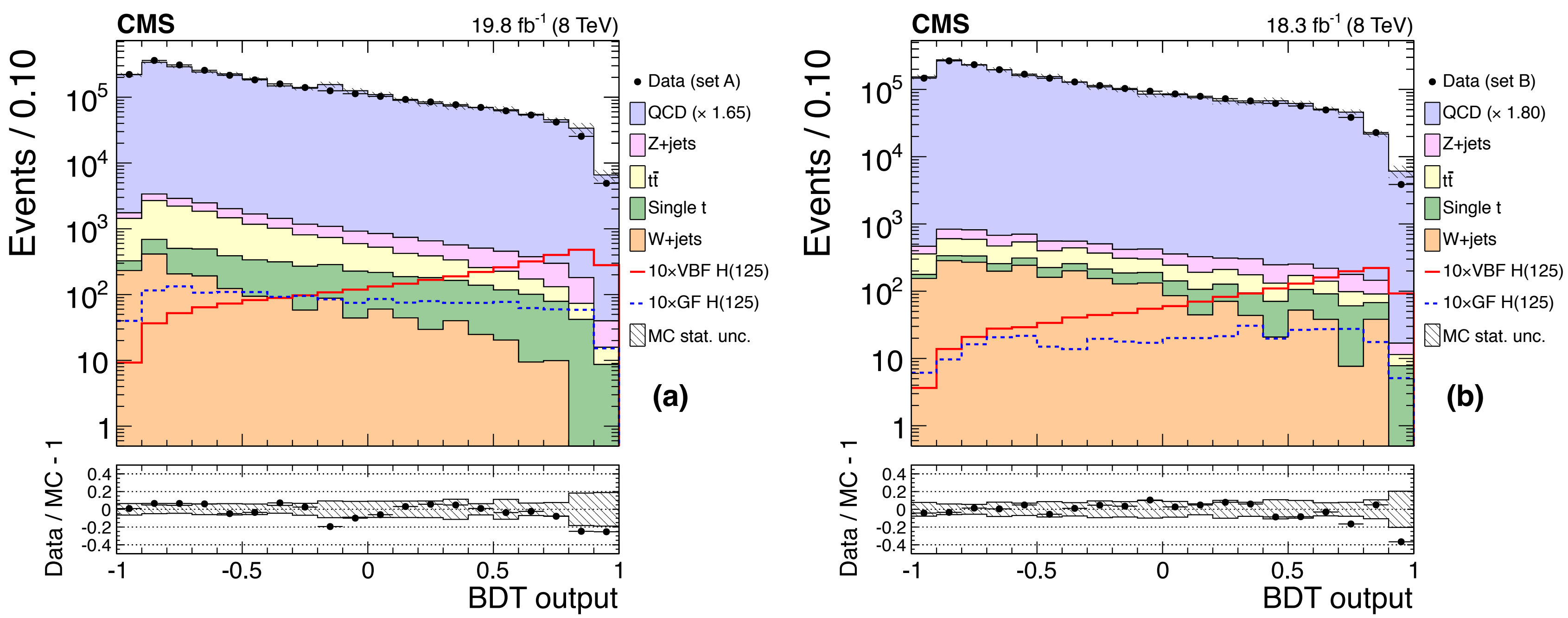



Table 3: Definition of the event categories and corresponding yields in the $m_{bb}$ interval $[80, 200]$ GeV, for the data and the MC expectation. The BDT output boundary values refer to the distributions shown in Fig. 8.

| BDT boundary values | set A | | | | set B | | |
|---|---|---|---|---|---|---|---|
| | Cat. 1 | Cat. 2 | Cat. 3 | Cat. 4 | Cat. 5 | Cat. 6 | Cat. 7 |
| | −0.6 – 0.0 | 0.0 – 0.7 | 0.7 – 0.84 | 0.84 – 1.0 | −0.1 – 0.4 | 0.4 – 0.8 | 0.8 – 1.0 |
| Data | 546121 | 321039 | 32740 | 10874 | 203865 | 108279 | 15151 |
| Z +jets | 2038 | 1584 | 198 | 71 | 435 | 280 | 45 |
| W+jets | 282 | 135 | 4 | <1 | 225 | 92 | 17 |
| $t\bar{t}$ | 2818 | 839 | 45 | 14 | 342 | 169 | 21 |
| Single t | 960 | 633 | 64 | 25 | 194 | 159 | 30 |
| VBF $m_H(125)$ | 53 | 140 | 58 | 57 | 33 | 57 | 31 |
| GF $m_H(125)$ | 53 | 51 | 8 | 5 | 9 | 10 | 2 |