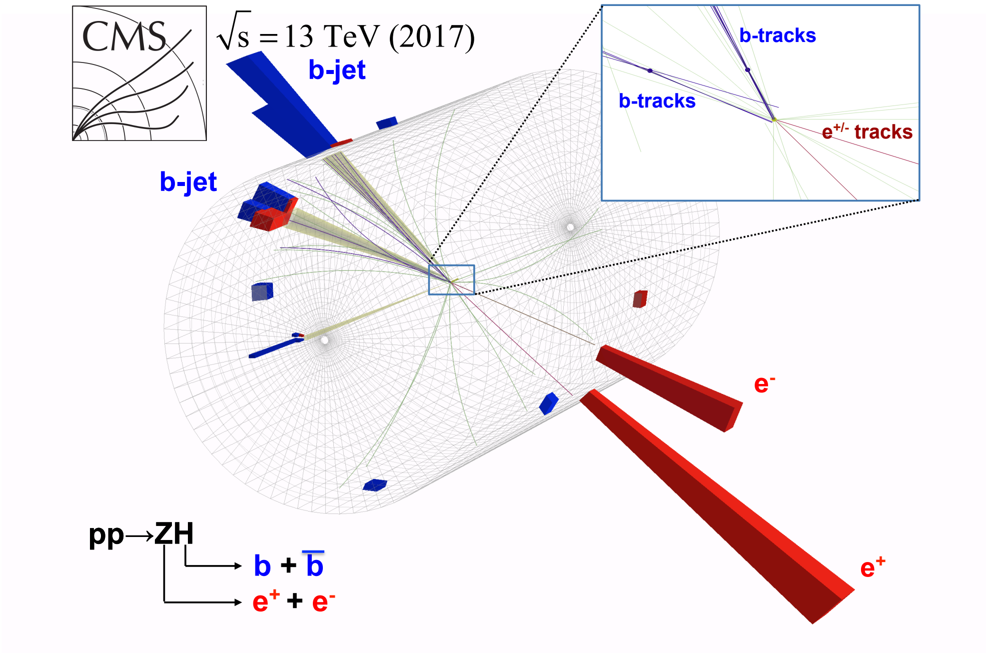
Higgs decay to b-quarks at CMS

A. de Wit (DESY)

Outline

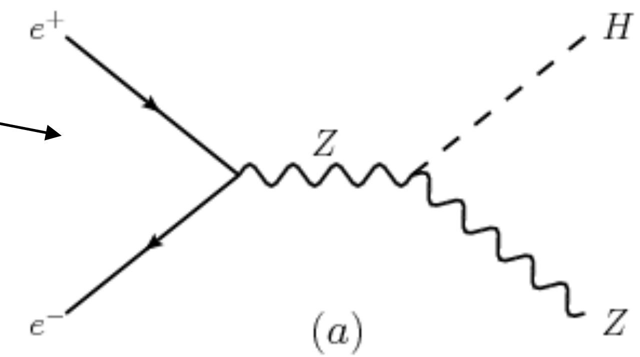
- Introduction
- LHC & CMS
  - Key ingredients
- CMS H➔$b\bar{b}$ analyses
  - ttH, ggH
  - VH
  - Combination
- H➔$b\bar{b}$ prospects at HL-LHC
- Summary & outlook

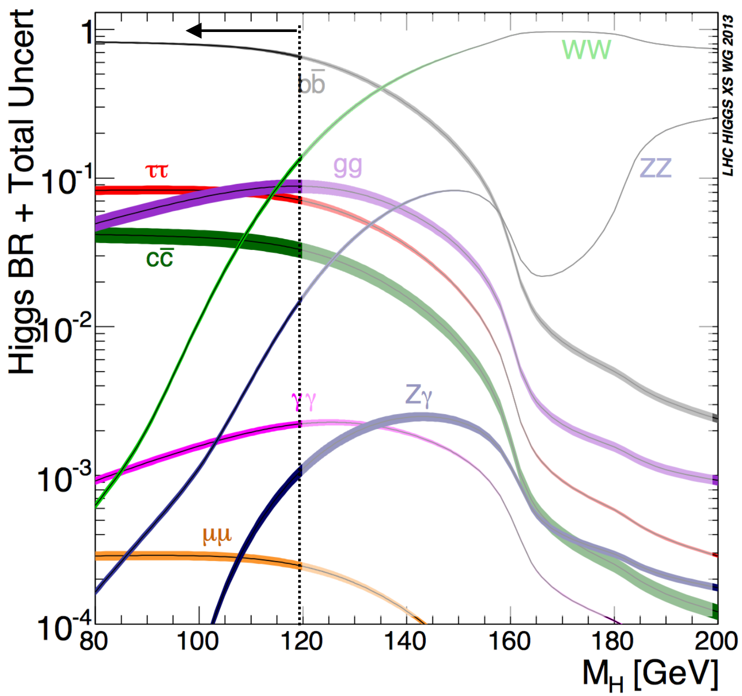


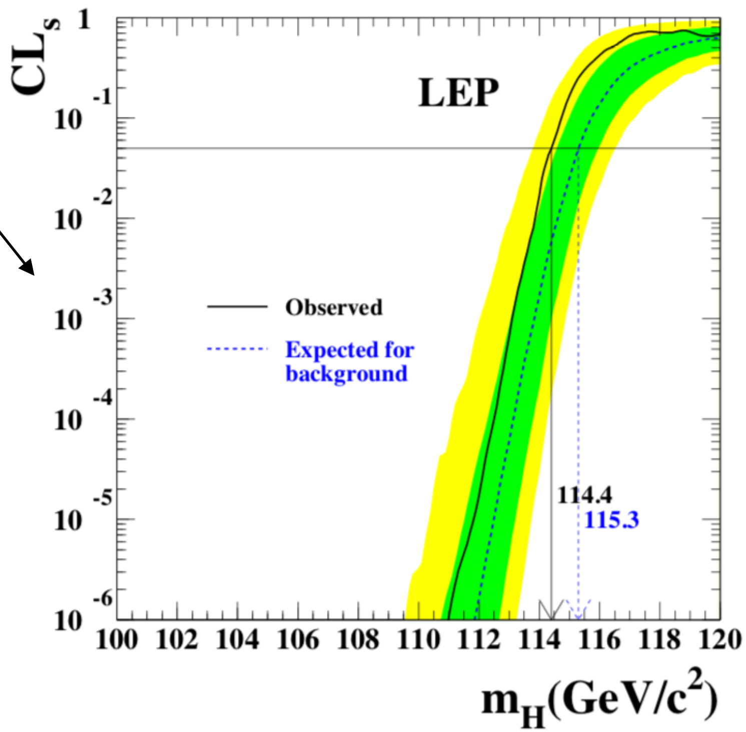
Phys.Lett.B565:61-75,2003

Higgs boson searches at LEP

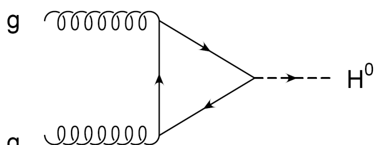
- LEP: e^+e^- collider ➔ main Higgs boson production mode: ZH
- Collisions at CM energies of 189-209 GeV: Higgs boson masses up to ~120 GeV accessible ➔ largest branching ratio into $b\bar{b}$
- **$m_H \leq$ 114.4 GeV excluded at 95% CL**



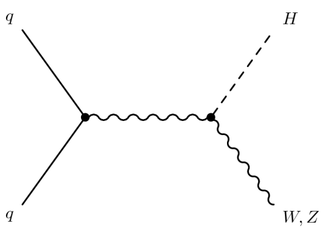




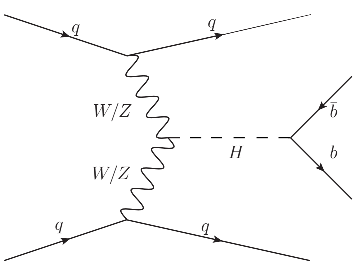
Higgs boson production at the Tevatron



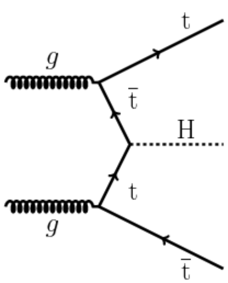
gluon fusion



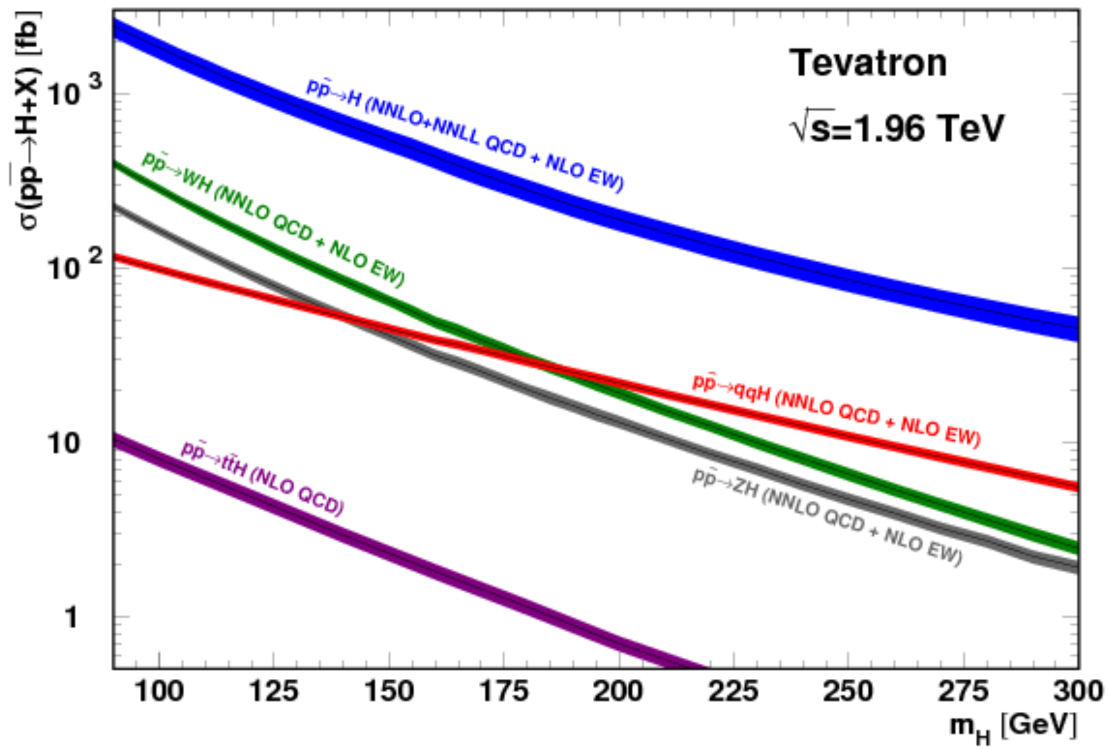
WH, ZH



vector boson fusion

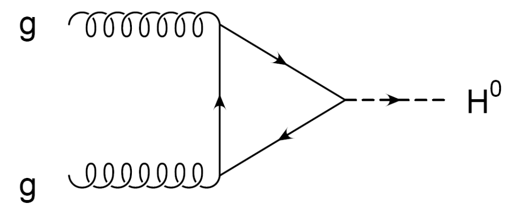


ttH

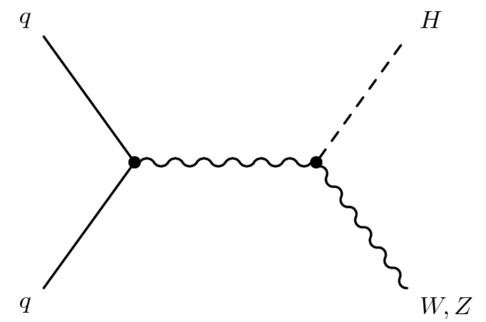


Higgs boson searches at Tevatron

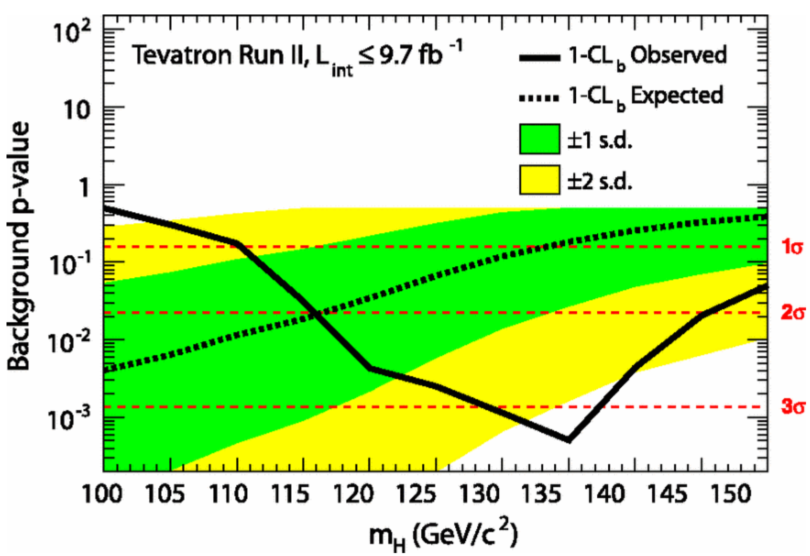
- Tevatron: $p\bar{p}$ collider: main production mode **gluon fusion** (as at LHC).



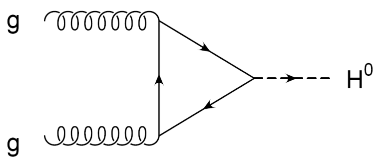
- However, CDF and D∅ experimental sensitivity dominated by **VH(bb)**



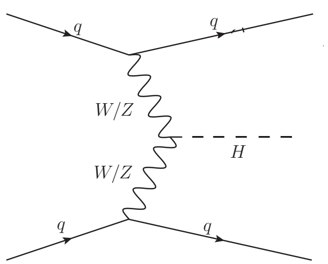
- On July 2nd 2012, evidence for VH(bb) announced by CDF and D∅



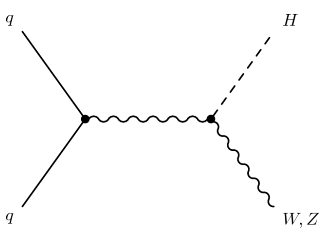
Higgs boson production at the LHC



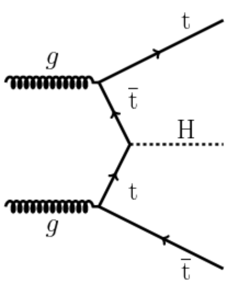
gluon fusion
(~87%)



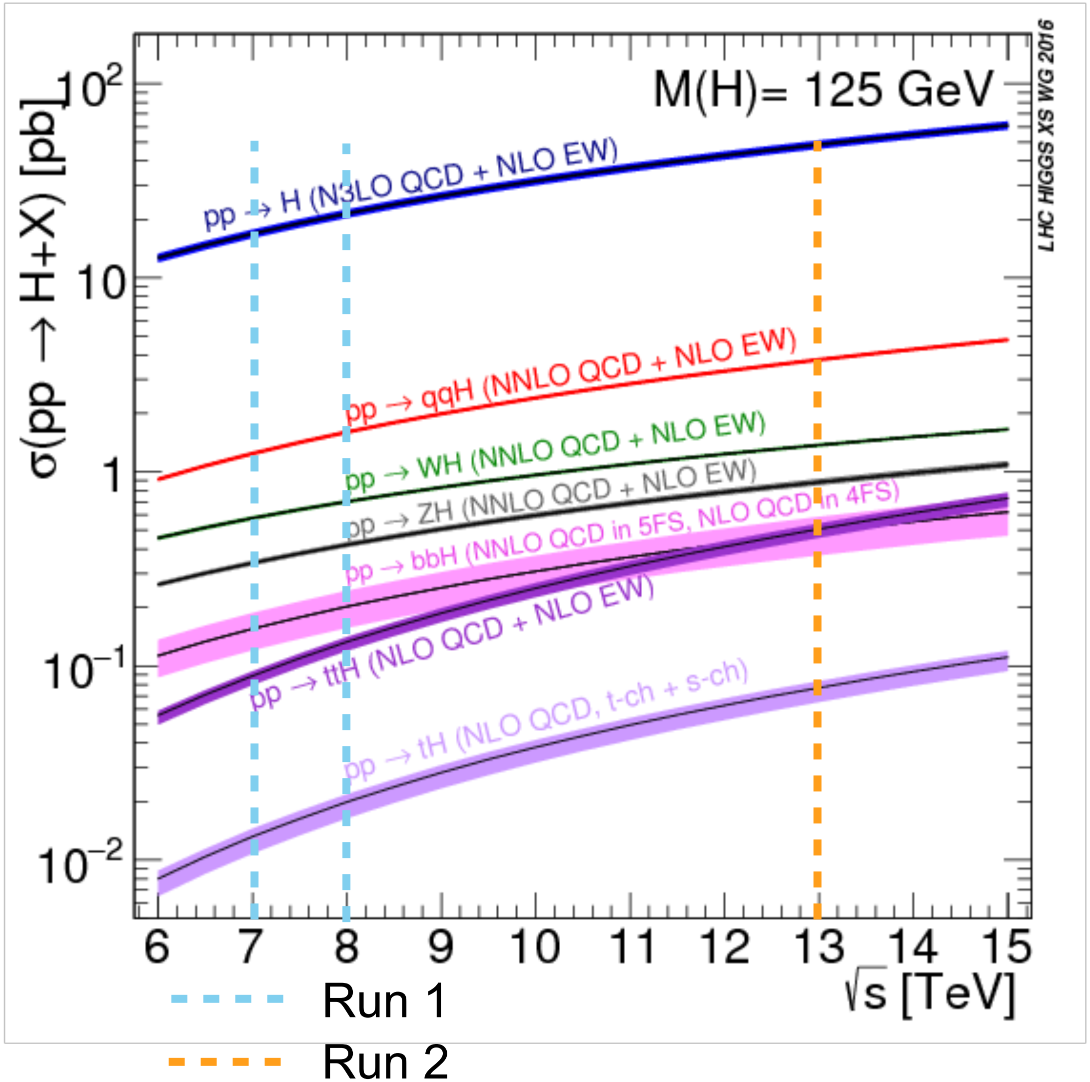
vector boson fusion
(~7%)



WH, ZH
(~4%)



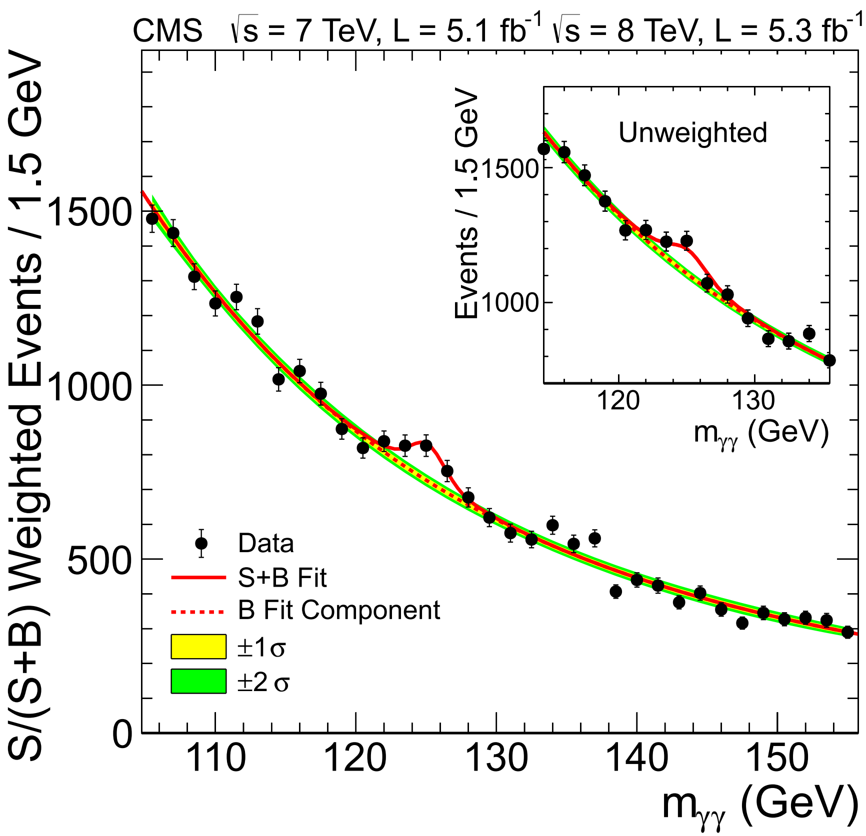
ttH
(~1%)



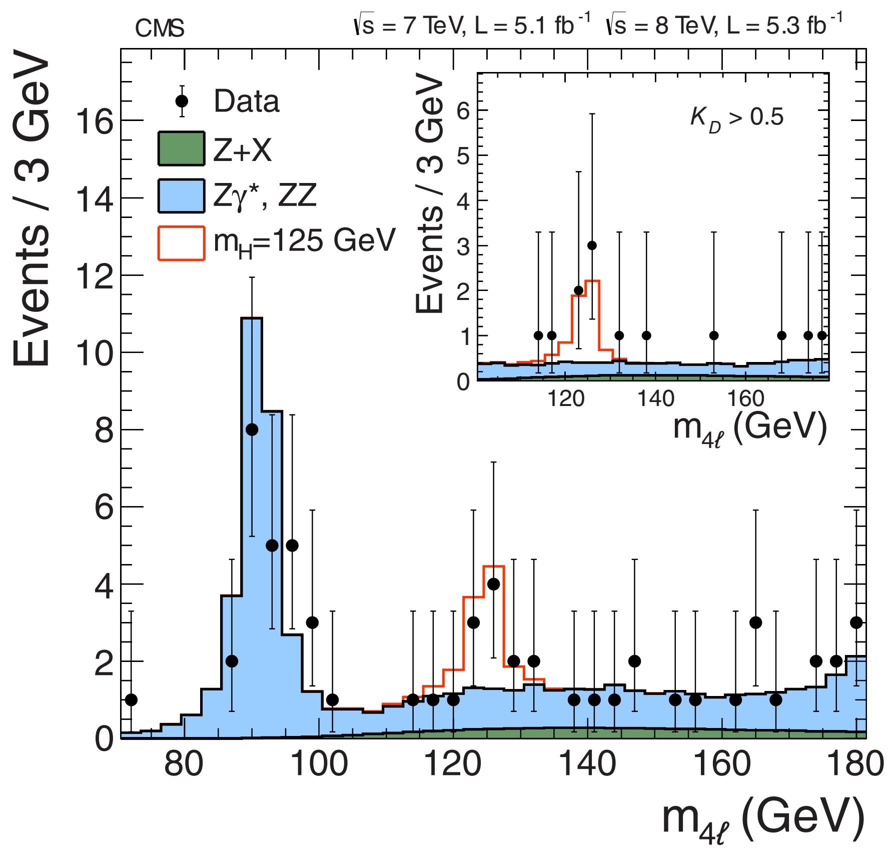
Higgs boson discovery at the LHC

- 4^{th} July 2012: ATLAS and CMS announce discovery of a particle with a mass of around 125 GeV
- Driven by H➔γγ (excellent energy resolution) and H➔ZZ➔4l (small background)

H➔γγ

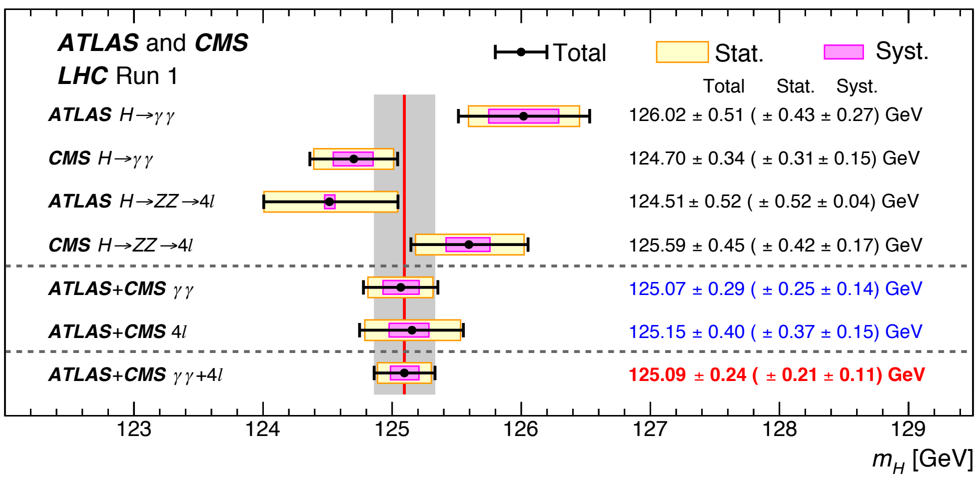


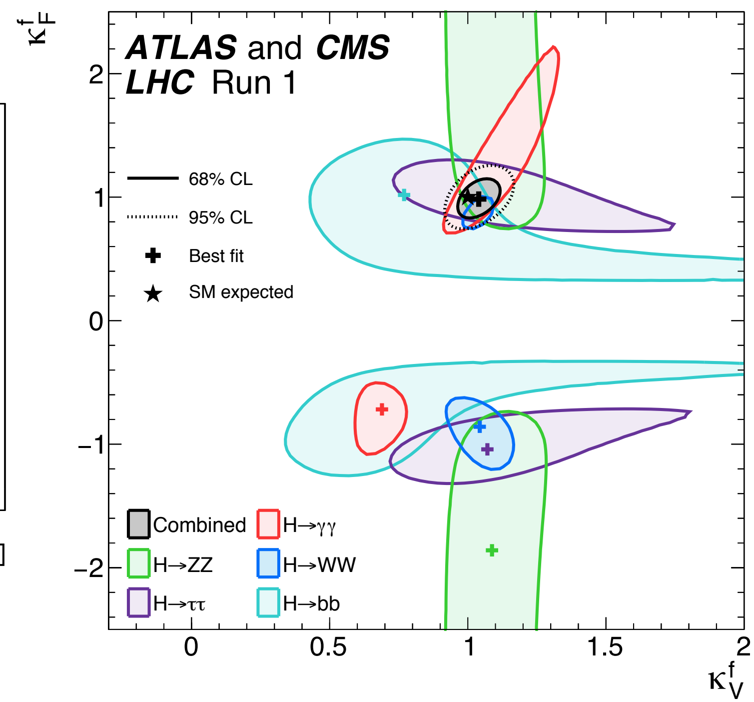
H➔ZZ➔4l



Properties (Run-1)

- Precise mass measurement: **m_H=125.09±0.21(stat.)±0.11(syst) GeV**
- Couplings compatible with the SM expectation
  - Bosonic couplings and coupling to τ leptons established
- Spin/parity also in agreement with SM

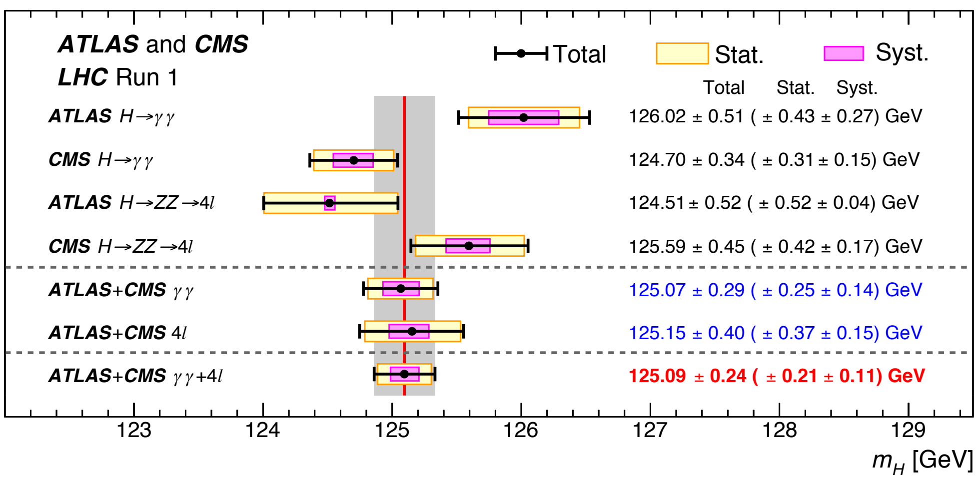


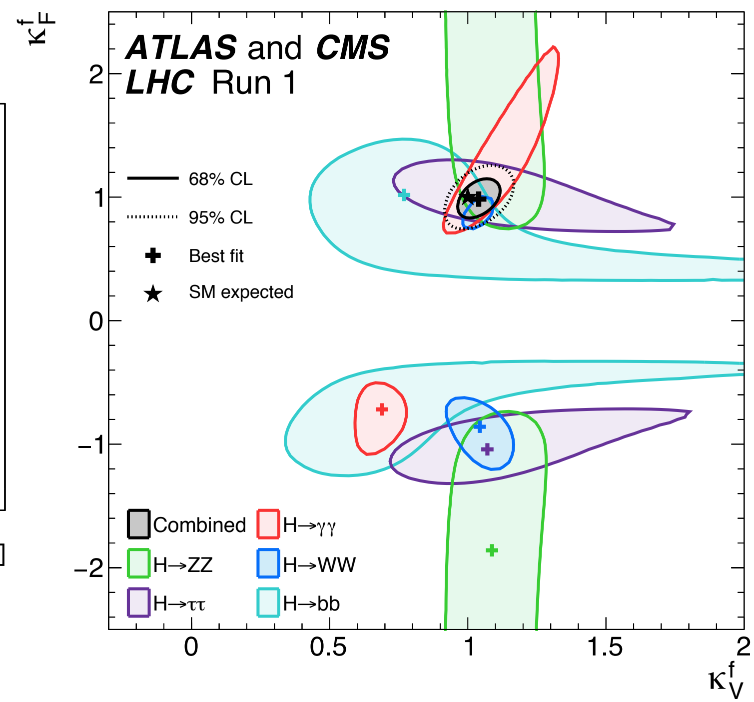


Properties (Run-1)

- Precise mass measurement: **m_H=125.09±0.21(stat.)±0.11(syst) GeV**
- Couplings compatible with the SM expectation
  - Bosonic couplings and coupling to τ leptons established
- Spin/parity also in agreement with SM

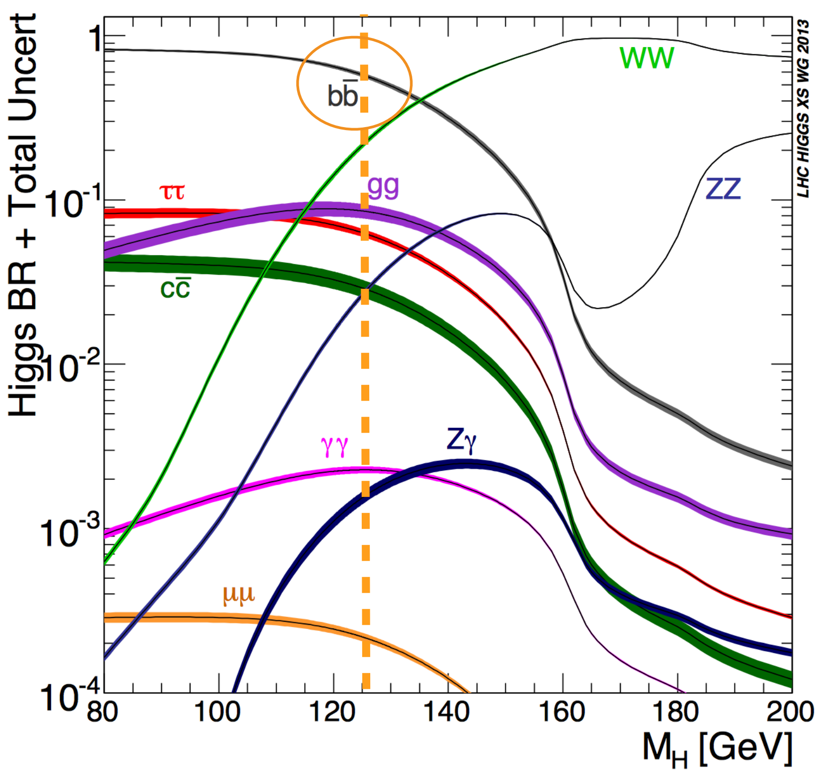
So far, the Higgs boson looks SM-like!

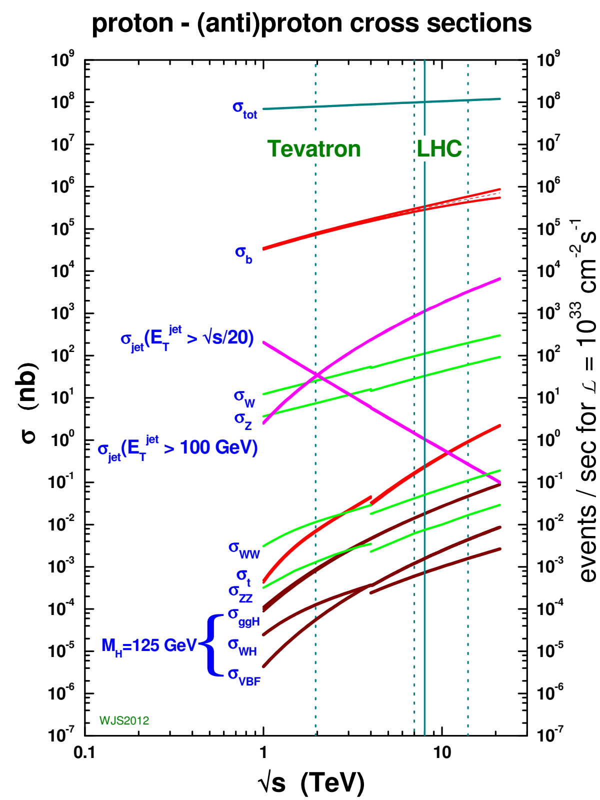




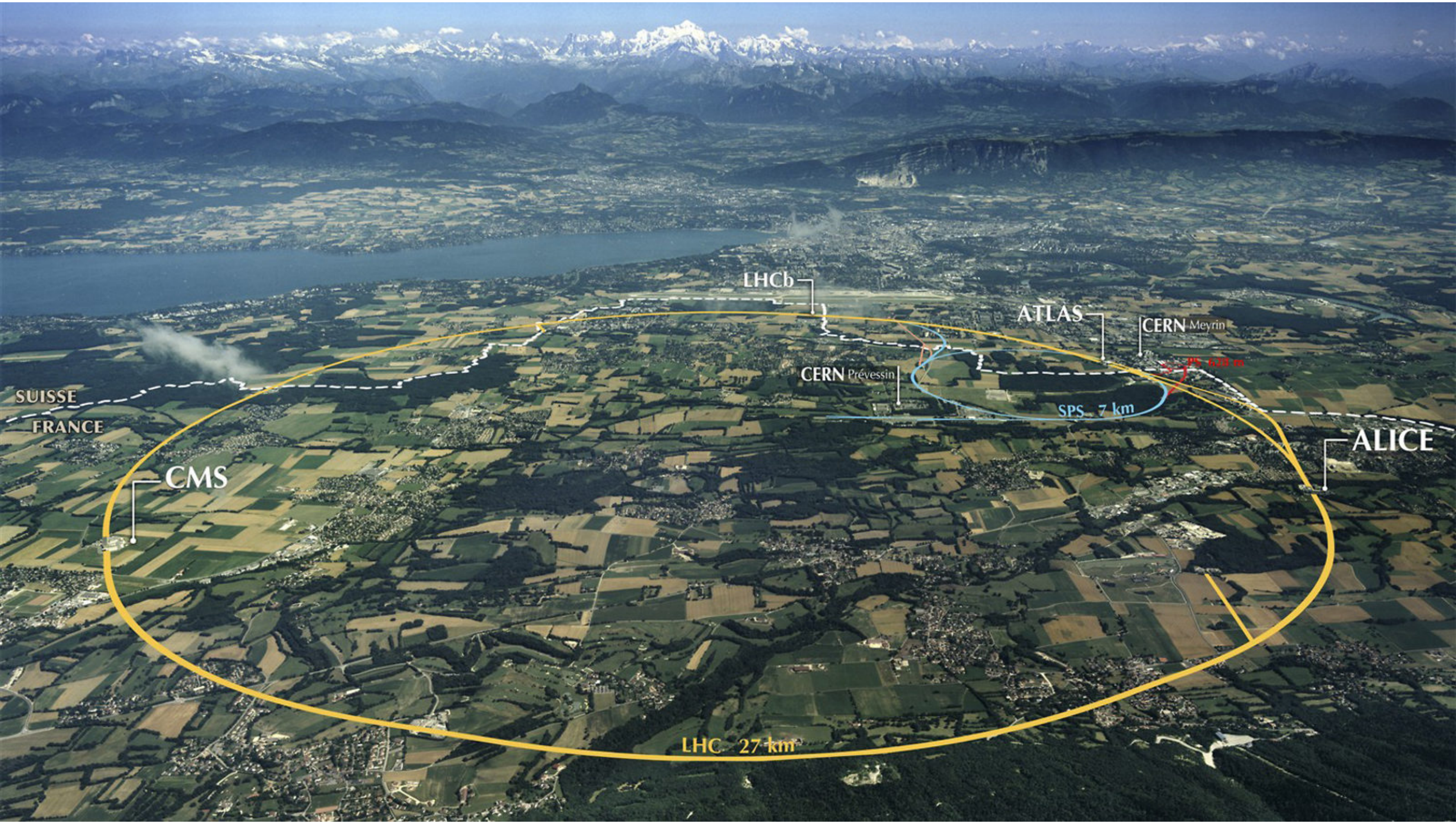
Why H→b$\bar{b}$ at the LHC?

- H→b$\bar{b}$ has by far the largest branching fraction (58%) at m_H=125 GeV
    - Largest contribution to the total Higgs boson width
    - Access to down-type quark coupling
- **Large multijet background** makes the search for this channel very challenging
- **Di-jet final state →** mass resolution not as good as in di-γ or multi-lepton final state



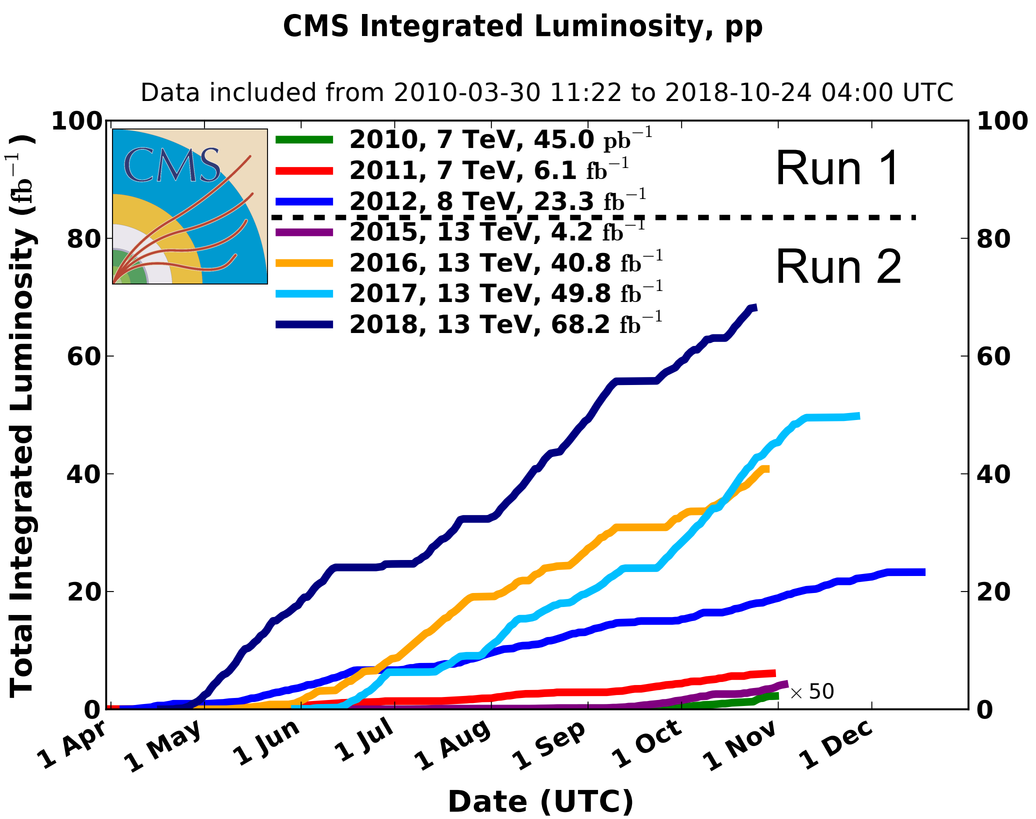


The Large Hadron Collider

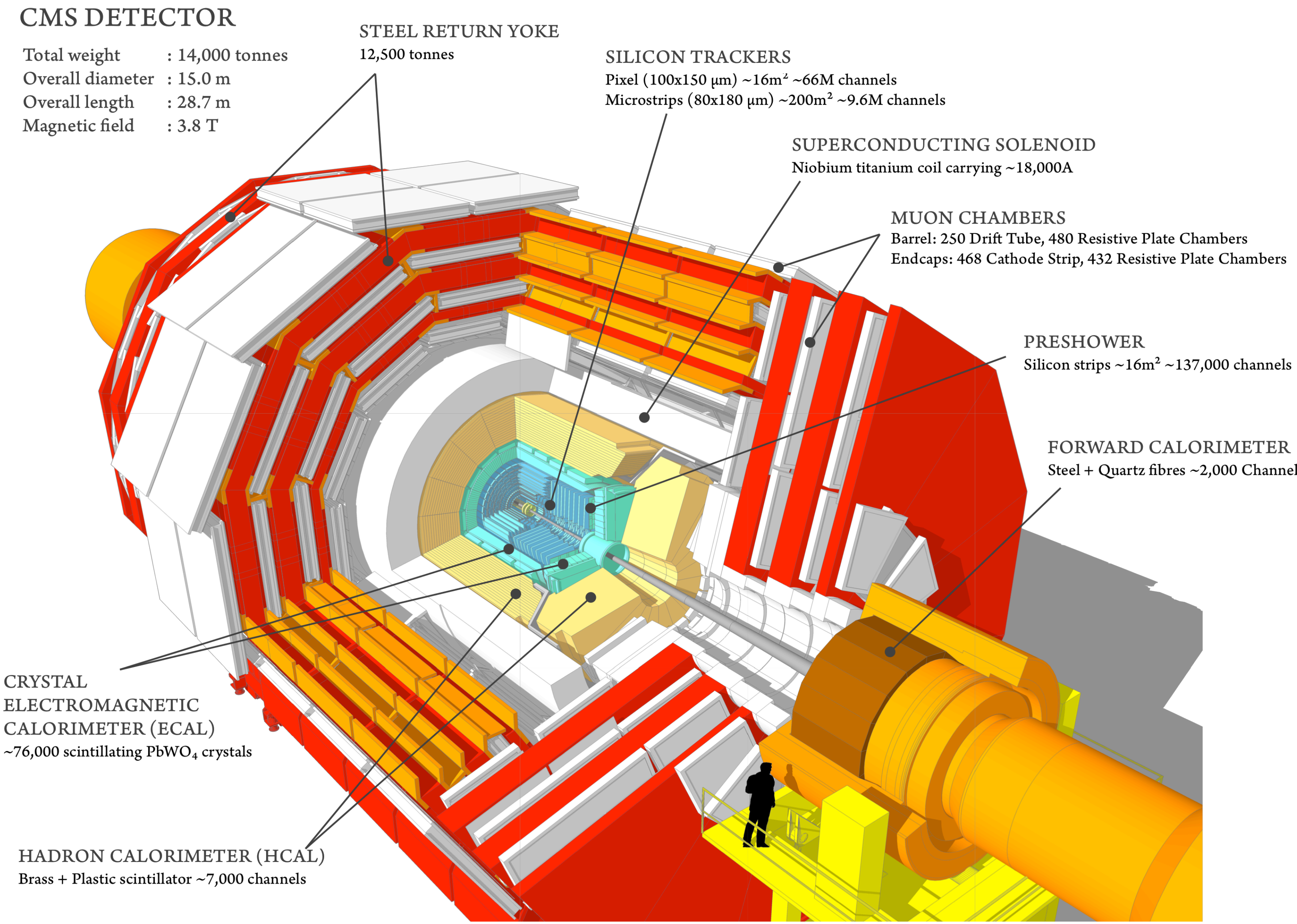


LHC performance

- **Excellent LHC performance**
  - Remarkable machine availability
- **163 fb^{-1} delivered to CMS in Run 2 alone**
  - The goal of 150 fb^{-1} in Run 1 and Run 2 was reached a long time ago
- 2016 + 2017 datasets for both CMS and ATLAS: ~ 80 fb^{-1}

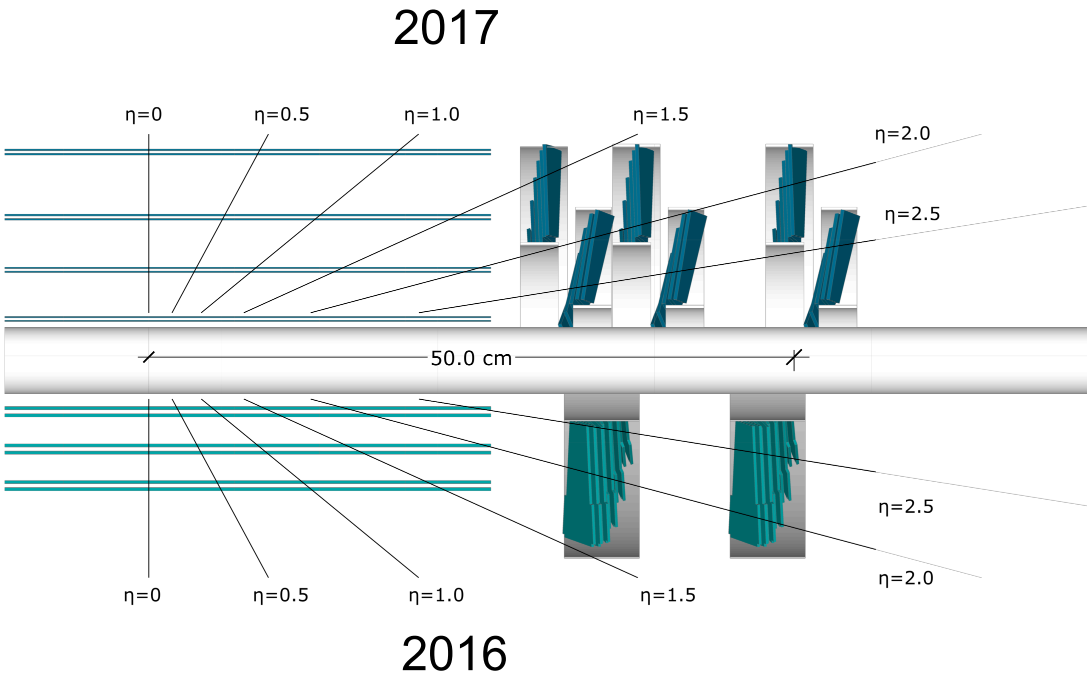


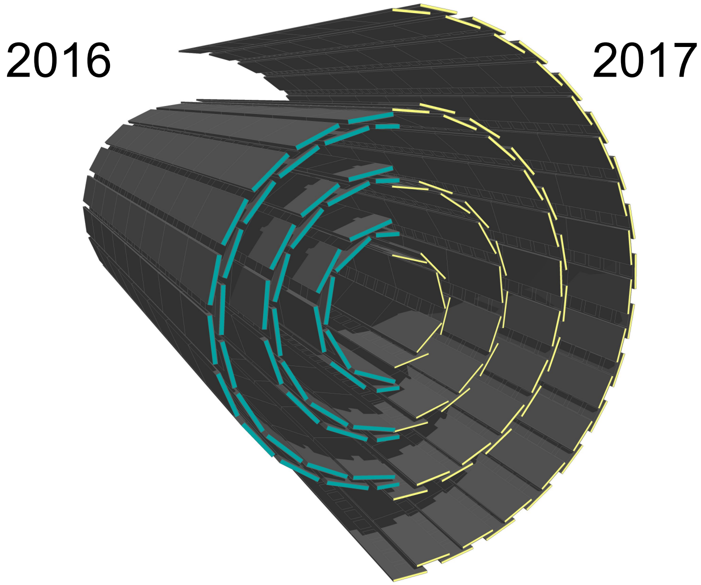
The CMS detector



Key ingredients: silicon tracker

- Efficient identification of b-jets relies on secondary decay vertices ➔ need the silicon pixel tracker at the heart of CMS!
- Pixel tracker was upgraded at the end of 2016- early 2017, adding a fourth layer of pixels in the barrel and improving the coverage in the forward region.

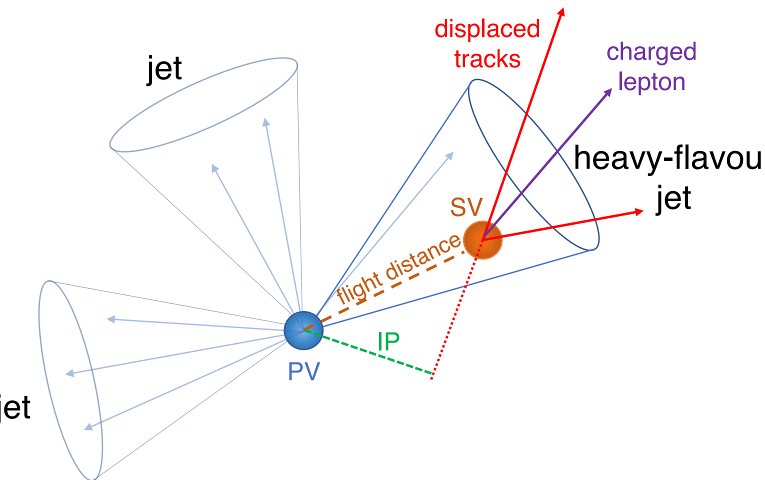


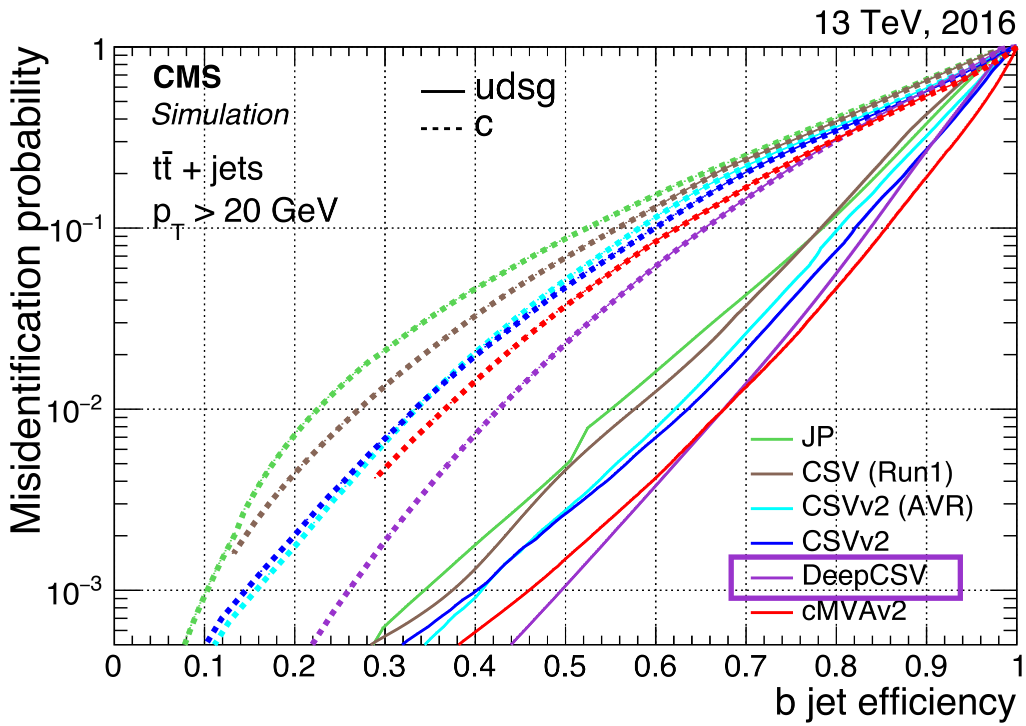


JINST 13 (2018) P05011

Key ingredients: b-tagging

- b-quarks: finite lifetime so decay after O(100)μm ➔ **displaced vertex!**
  - Vital to distinguish b-jets from light quark/gluon jets
- B-taggers in CMS use multivariate techniques to maximally exploit the available information

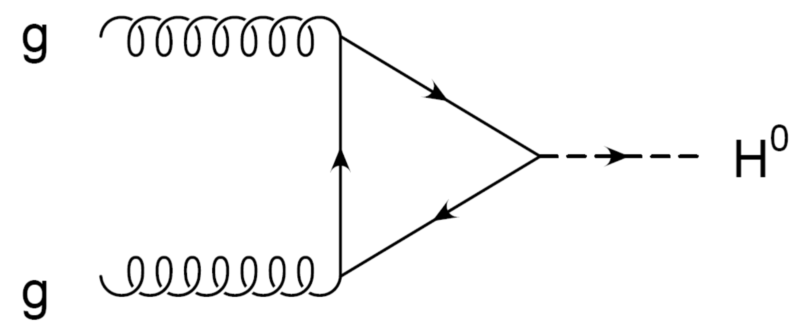


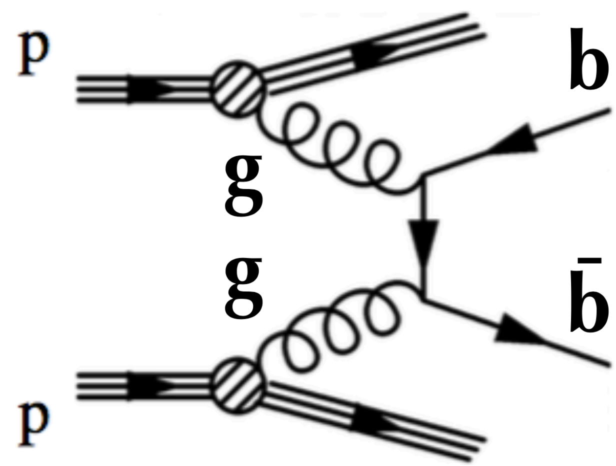


ggH, $H \rightarrow b\bar{b}$

Overview

- Largest production cross section, but suffers from overwhelming background from heavy flavour multijet production
  - Up to recently a search for ggH production in the $H \rightarrow b\bar{b}$ decay channel would have been deemed impossible
- At high H p_T the two b-jets are likely to merge into a single 'fat' jet ➔ **exploit di-b-jet substructure to make an inclusive $H \rightarrow b\bar{b}$ search at high H p_T possible**

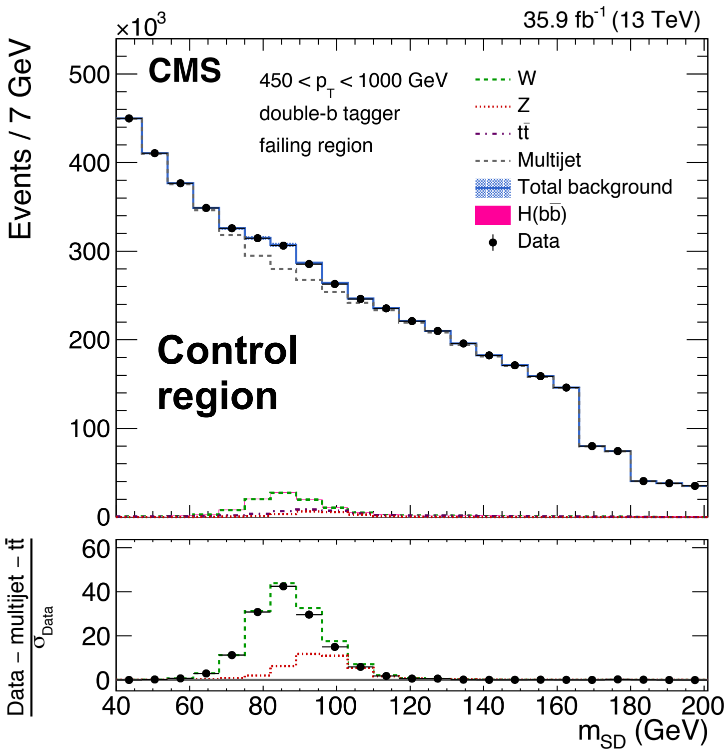


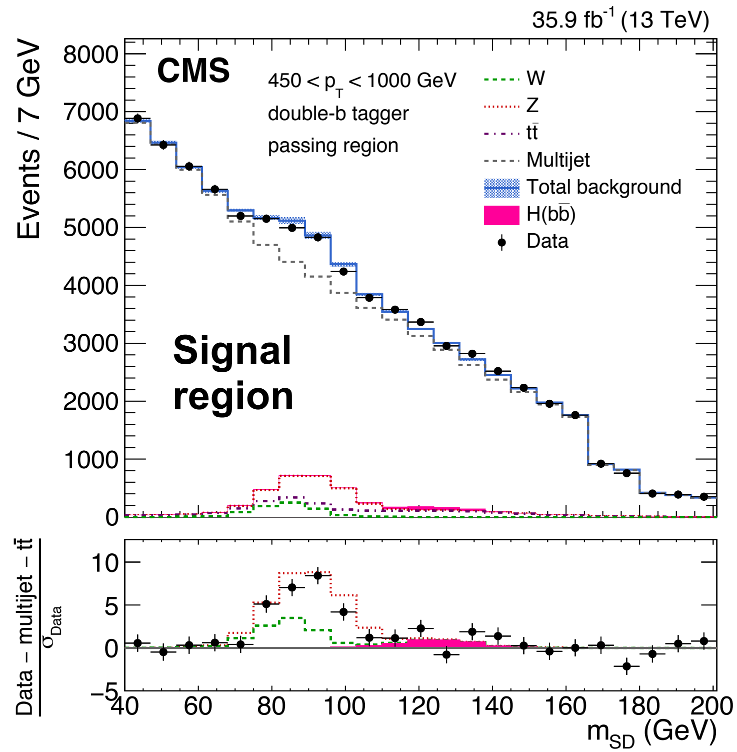


PRL 120(2018) 071802

Inclusive (ggH) H→b$\bar{b}$

- Analysis performed using 2016 dataset
- Require a high p_T (>450 GeV) wide-cone jet and **exploit 2-prong jet substructure** and **b-tagging information of the subjects** to reduce multijet background
- **Backgrounds** mostly from multijet production, with smaller contributions from tt, W and Z production
- Result: $\mu_H = 2.3 \pm 1.5\ (\text{stat.})\ ^{+1.0}_{-0.4}\ (\text{syst})$

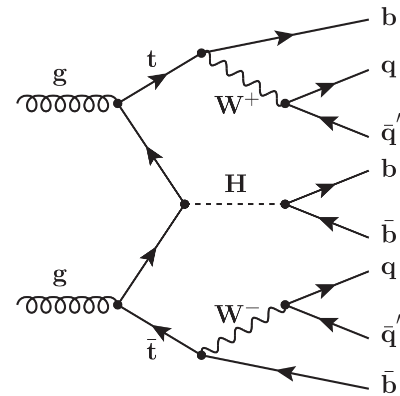




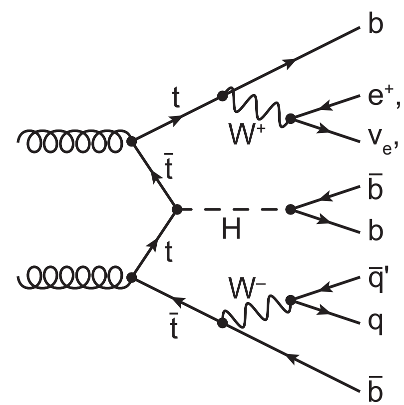
$t\bar{t}H$, $H \rightarrow b\bar{b}$

Overview

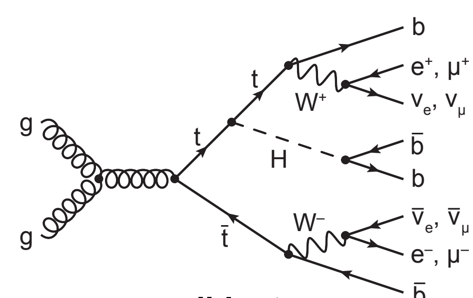
- Small production cross section
- Large backgrounds
  - Combinatoric
  - $t\bar{t}+b\bar{b}$, with associated large theory uncertainties
- Make use of many possible W decay modes:
  - Fully hadronic } Higher rate
  - Semi-leptonic
  - Fully leptonic } Higher purity



Fully hadronic



single-lepton



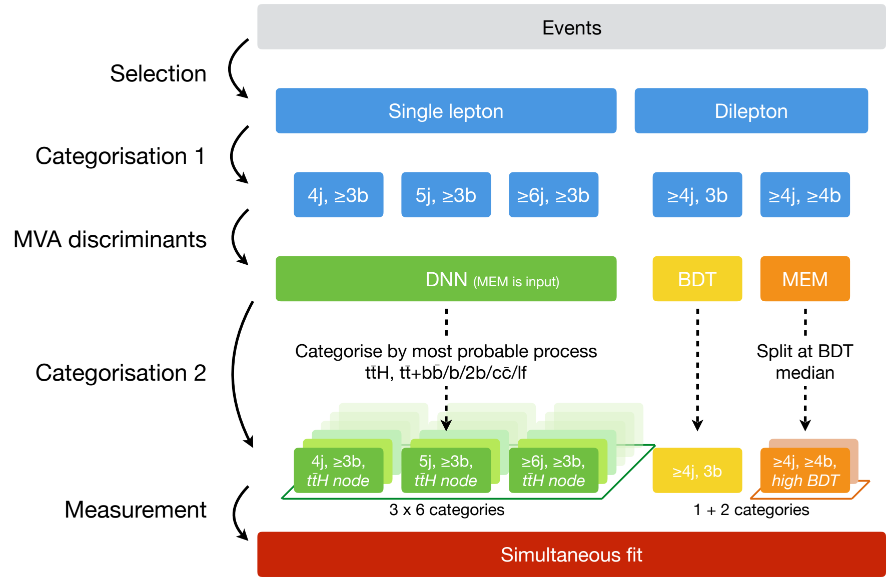
di-lepton

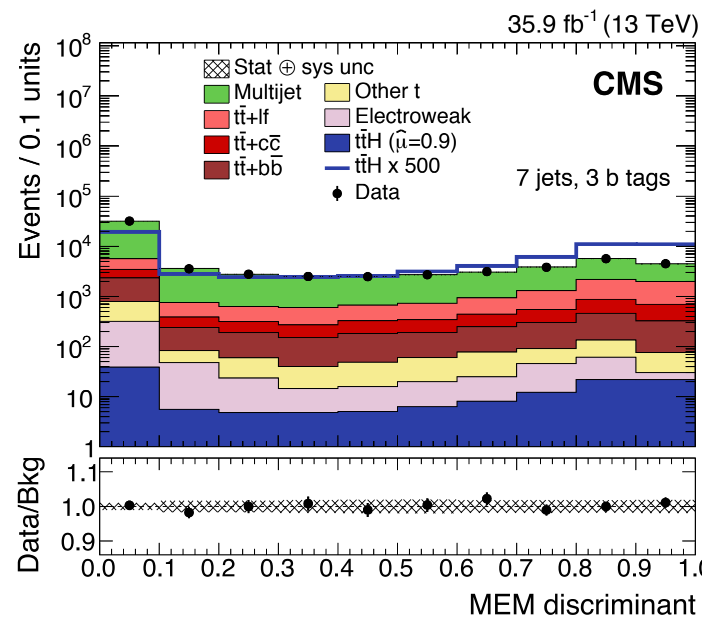
JHEP 06 (2018) 101; arXiv:1804.03682 [sub'd to JHEP]

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

Strategy

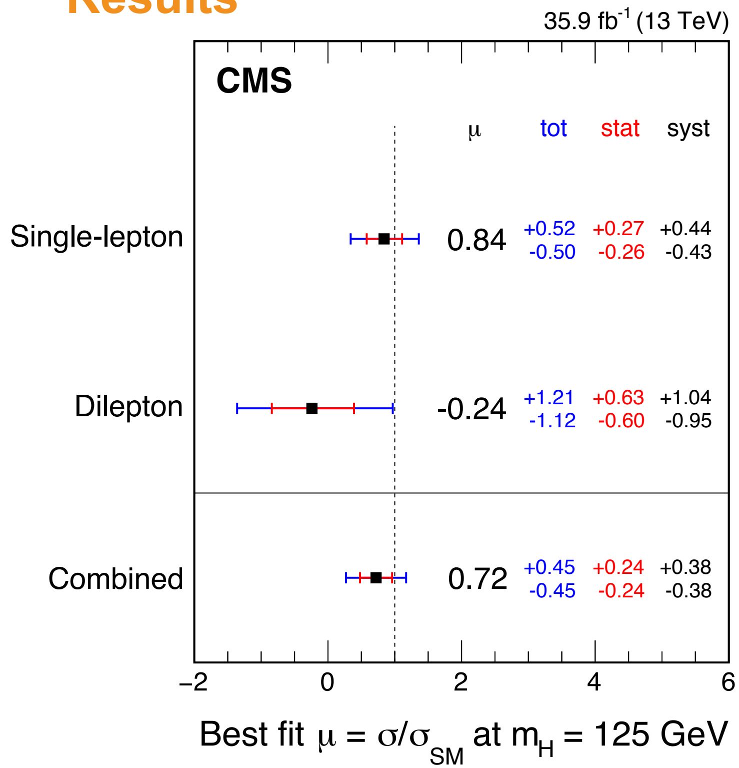
- **Categorise** by number of jets and b-tagged jets
  - **Single-lepton**: further subcategorisation using DNN multiclassification
  - **Di-lepton:** Low/High BDT subcategories
- **Backgrounds** mostly estimated from simulation
  - Dominant background in leptonic channels from $t\bar{t}$+jets production
  - Dominant background in hadronic channel from multijet production ➔ estimated from control regions
- **Signal extraction** performed with a fit to **multivariate discriminator** distribution or distribution of **matrix element** classifier

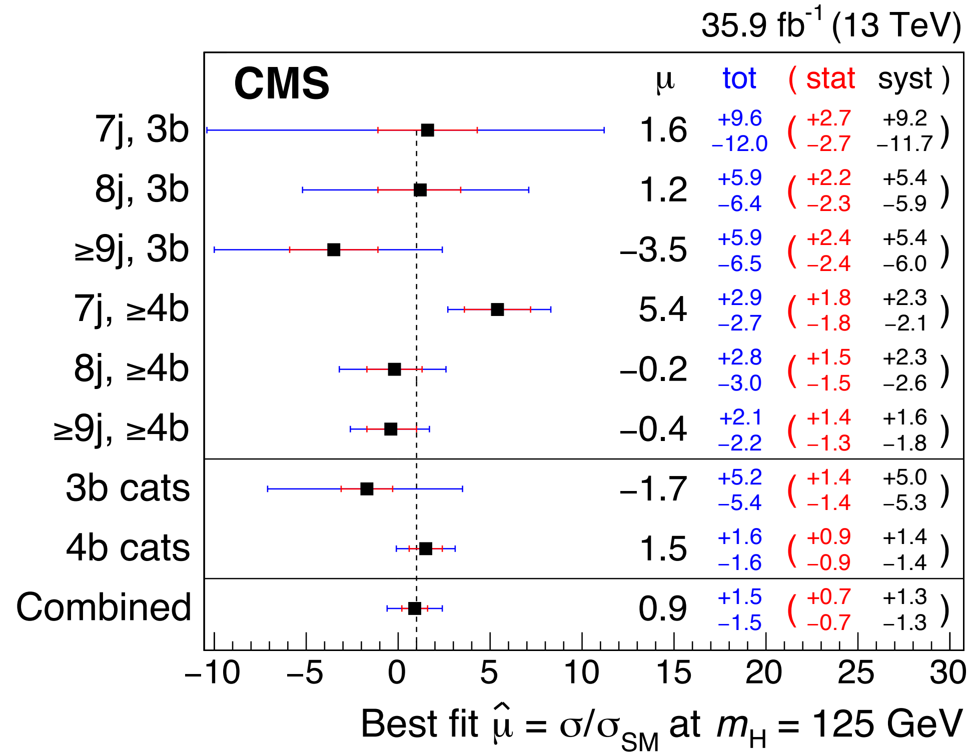




$t\bar{t}H$, $H \rightarrow b\bar{b}$

Results



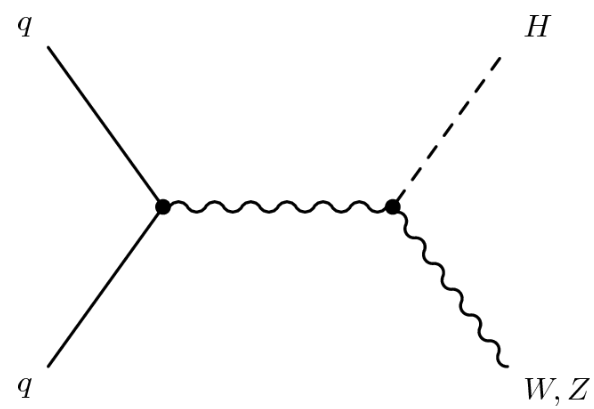


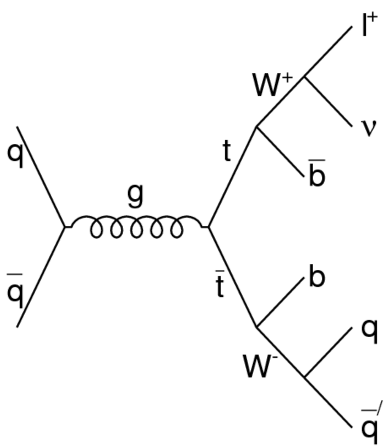
- Leptonic channel: $\mu=0.72\pm0.45$
- Hadronic channel: $\mu=0.9\pm1.5$
- Limiting uncertainties: $t\bar{t}$+HF modelling, QCD multijet modelling, b-tagging uncertainties

VH, H→b$\bar{b}$

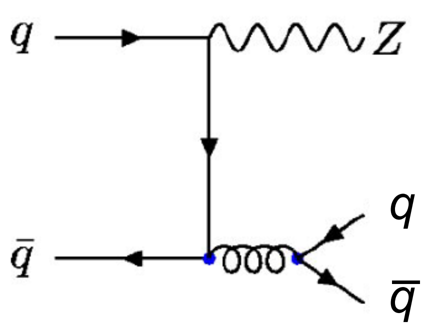
Overview

- Higgs boson produced in association with a vector boson
  - Leptonically decaying vector boson gives a clean signature to tag ➔ helpful for online selection
  - Much reduced background from multijet production
- Most sensitive channel for H➔b$\bar{b}$ despite relatively small production cross section
- **Backgrounds:**

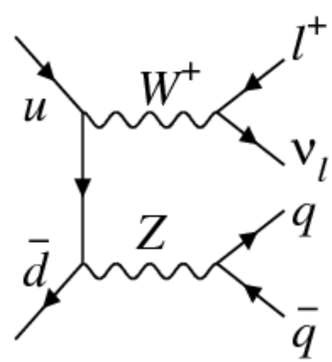




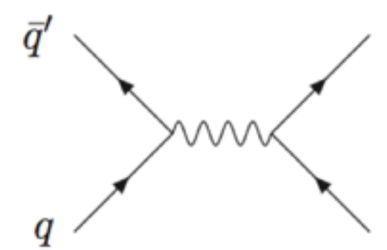
t$\bar{t}$



W,Z + jets



di-boson



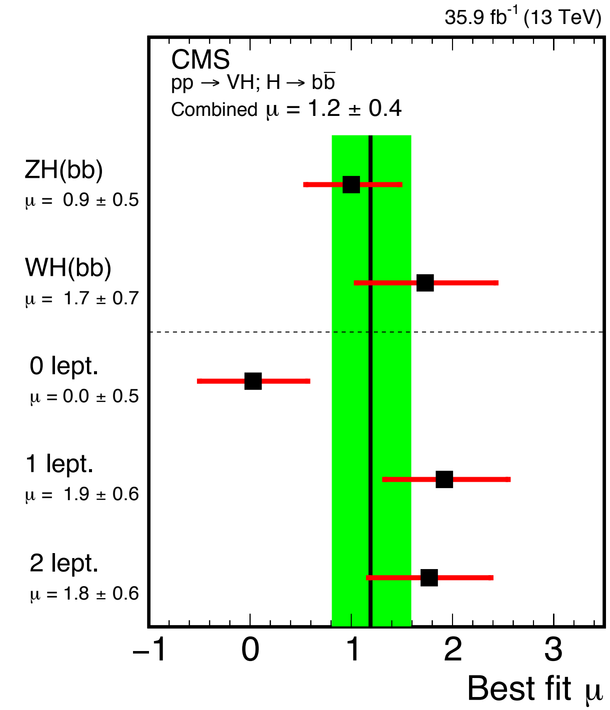
single top

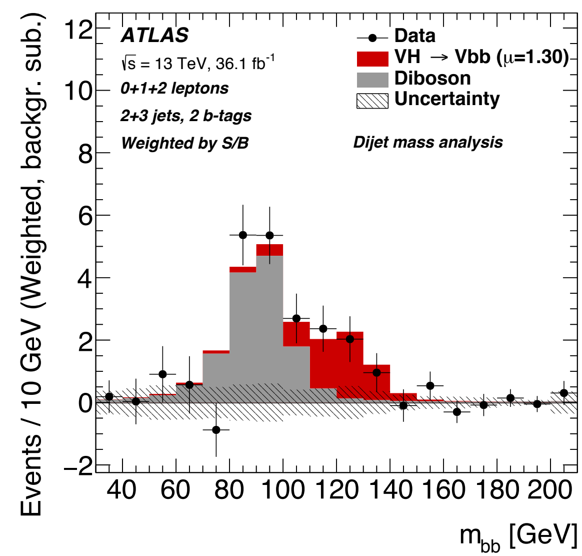
Previous VH, H→b$\bar{b}$ results

- Evidence for VH, H→b$\bar{b}$ established by both ATLAS and CMS using data collected during 2016

	Signal strength	Significance observed (expected)
ATLAS Run 1 [1]	$0.52^{+0.40}_{-0.37}$	1.4σ (2.6σ)
CMS Run 1 [2]	1.0 ± 0.5	2.1σ (2.1σ)
CMS + ATLAS Run 1 [3] (all Hb$\bar{b}$)	$0.70^{+0.29}_{-0.27}$	2.6σ (3.7σ)
ATLAS 2015+2016 [4]	$1.2^{+0.42}_{-0.36}$	3.5σ (3.0σ)
CMS 2016 [5]	$1.19^{+0.40}_{-0.38}$	3.3σ (2.8σ)

[1] JHEP 01 (2015) 069
[2] PRD 89 (2014) 012003
[3] JHEP 08 (2016) 045
[4] JHEP 12 (2017) 024
[5] PLB 780 (2018) 501



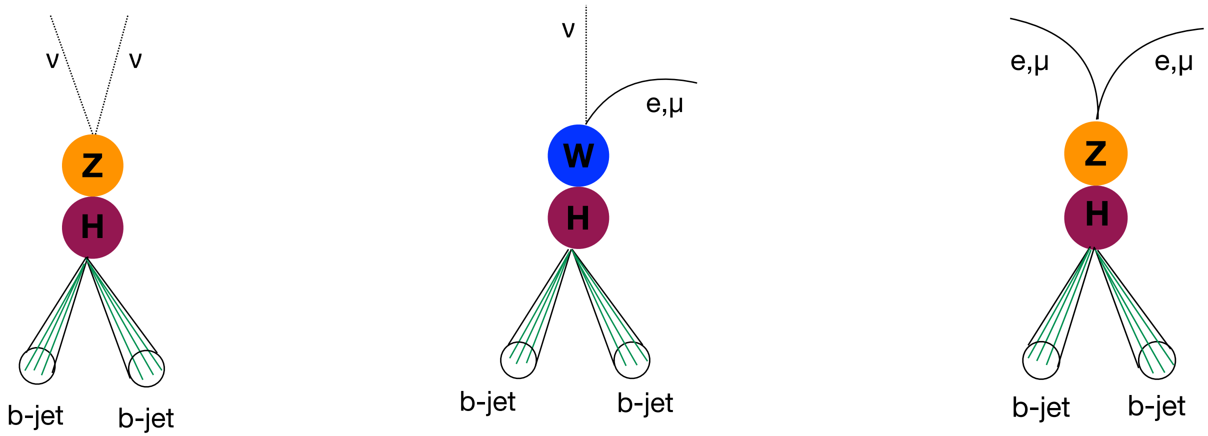


PRL 121(2018) 121801

VH, H→b$\bar{b}$

Analysis strategy

- **Select events** with 0, 1, or 2 electrons or muons, consistent with a W or Z decay, and 2 b-tagged jets



0-lepton **1-lepton** **2-lepton**

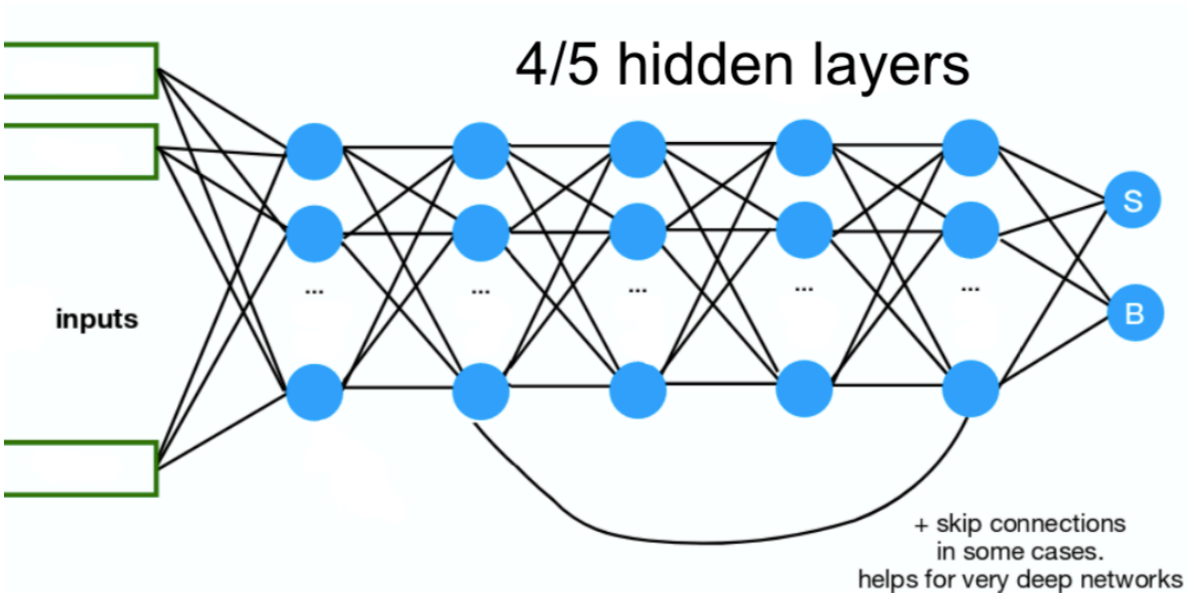
- b-jets and vector boson produced back-to-back, plus increased sensitivity for enhanced Higgs boson p_T: **categorise** based on V p_T

	0-lepton	1-lepton	2-lepton	
V p_T	> 170 GeV	> 150 GeV	50-150 GeV	> 150 GeV

- **Improve** the di-b mass resolution
  - Multivariate regression techniques
  - Kinematic fit
- Use a **deep neural network (DNN)** to separate signal and background
- **Fit for signal** using the DNN as final discriminant, simultaneously fitting control regions to constrain the major backgrounds.

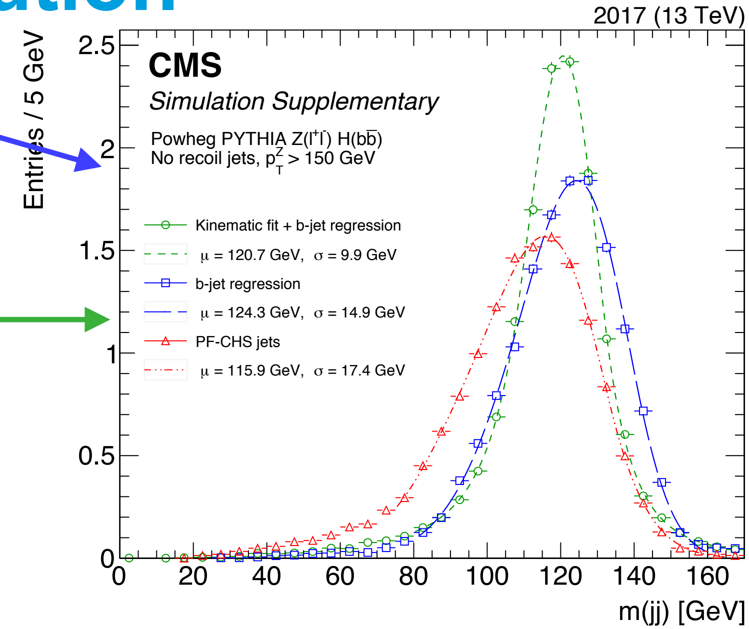
Deep neural network (DNN)

- Used in this analysis for
  - b-jet energy regression
  - Signal vs background discrimination
  - Discrimination between different types of background in some control regions

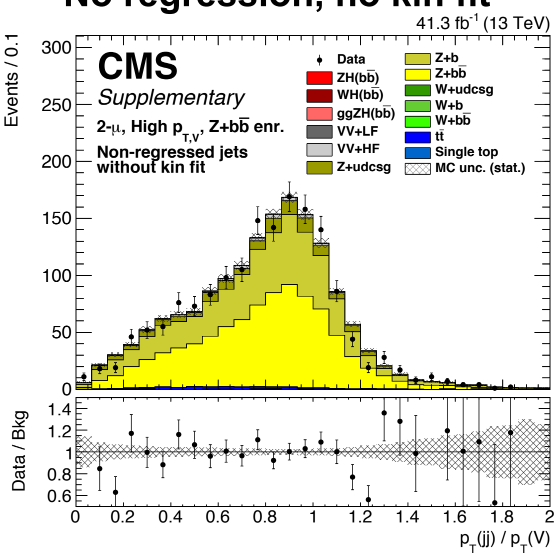


Improving di-b mass resolution

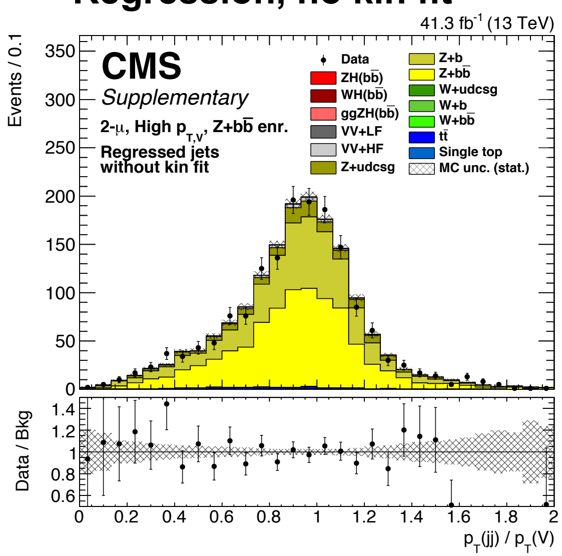
- In all channels, di-b mass resolution improved with DNN b-jet energy regression
- In 2-lepton channel, additional improvement from use of kinematic fit
- Validation using $p_T(jj)/p_T(ll)$ in 2-lepton Z+HF control region ➔ data described well by simulation after all techniques



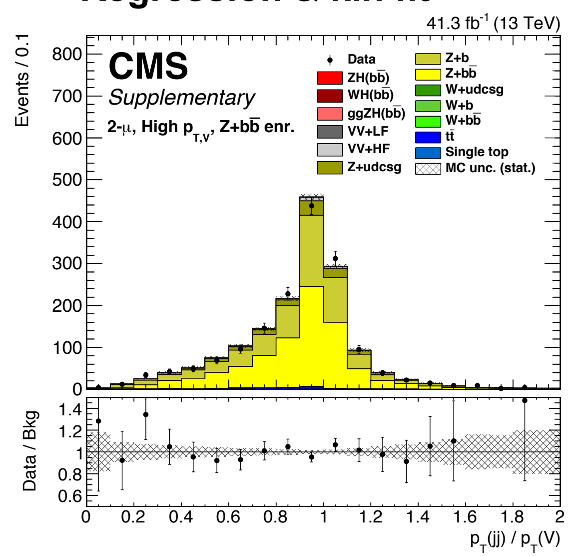
No regression, no kin fit



Regression, no kin fit

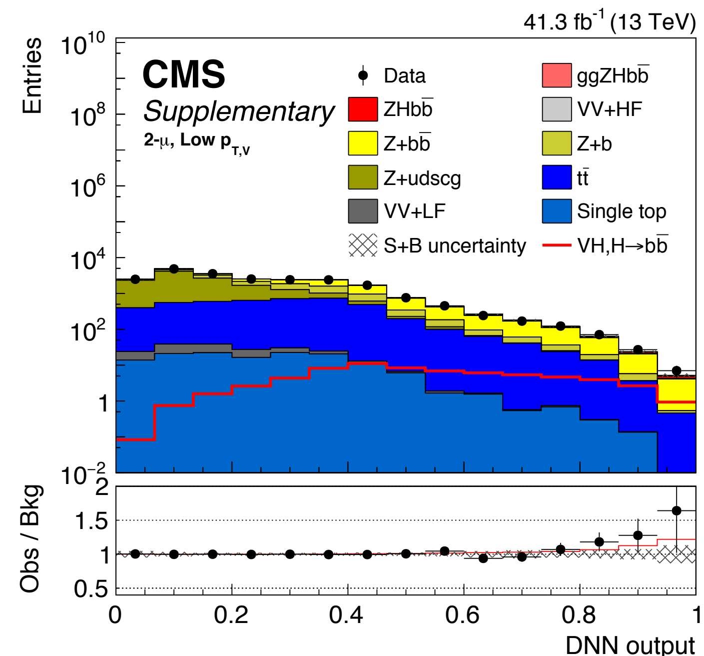


Regression & kin fit

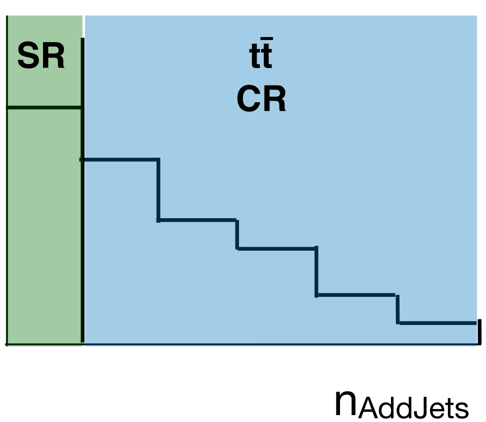


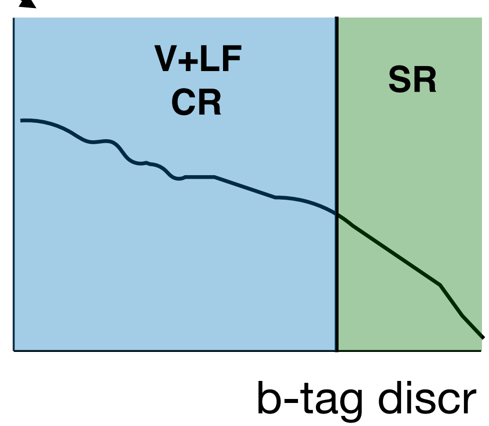
Backgrounds

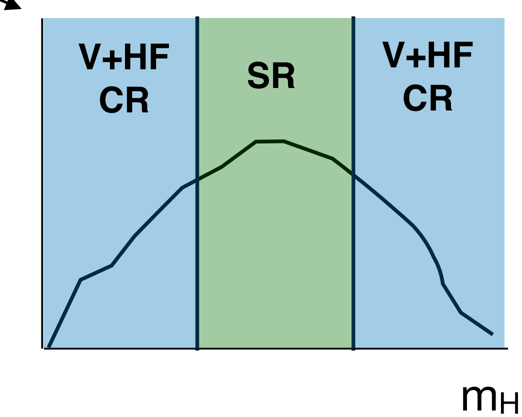
- Many different backgrounds play a role. Largest: V+jets, $t\bar{t}$
- Use dedicated control regions to constrain the normalisation of these major backgrounds
- Control regions are defined by inverting selections that define the signal regions to create control regions enhanced in
$t\bar{t}$, **V+light jets,** **V+b-jets**



2μ low V p_T signal region

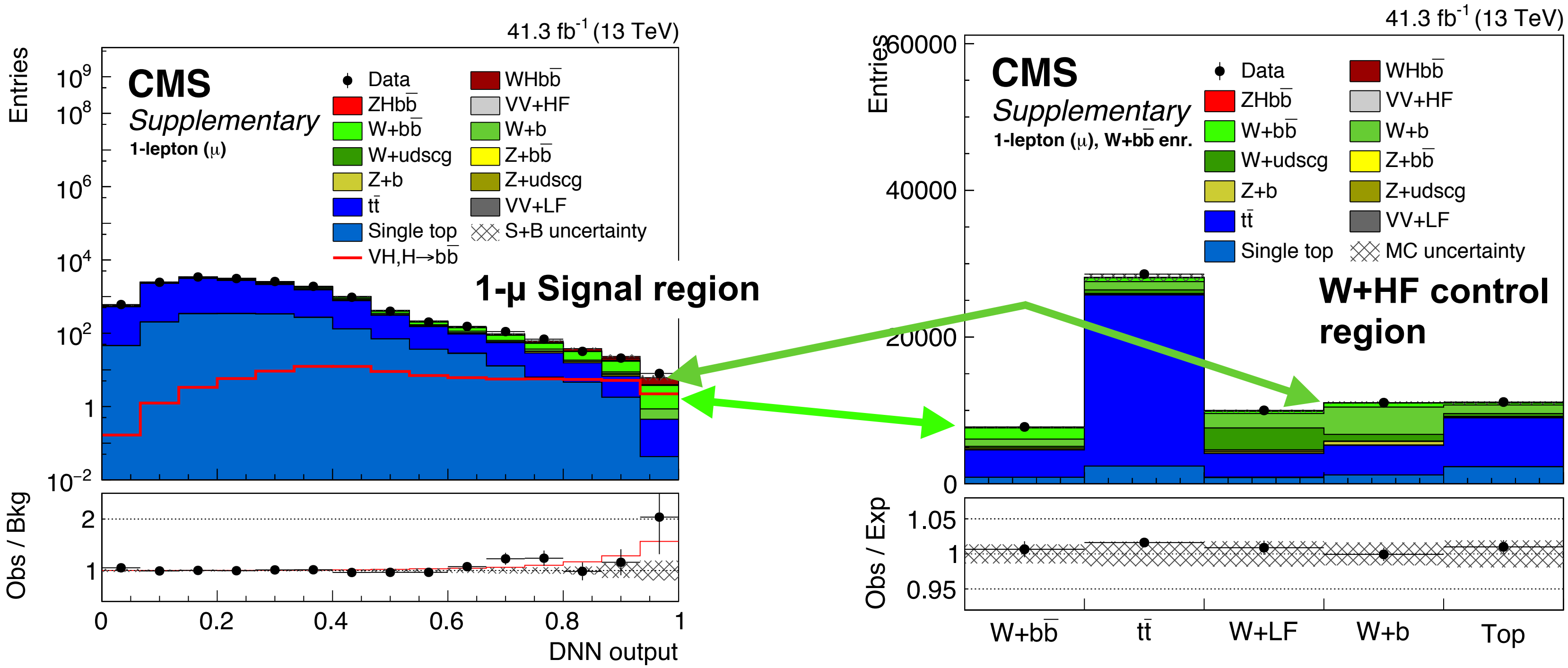






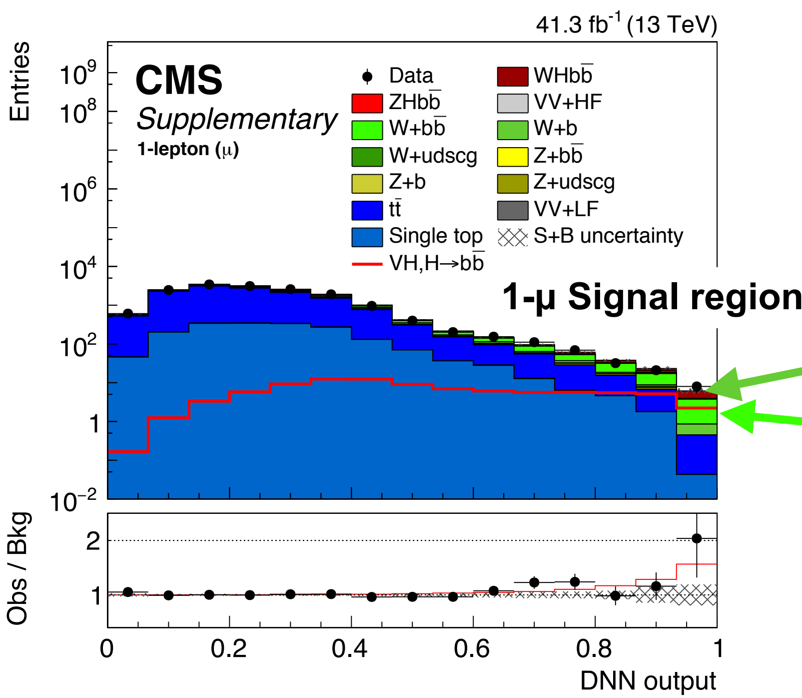
Background normalisation

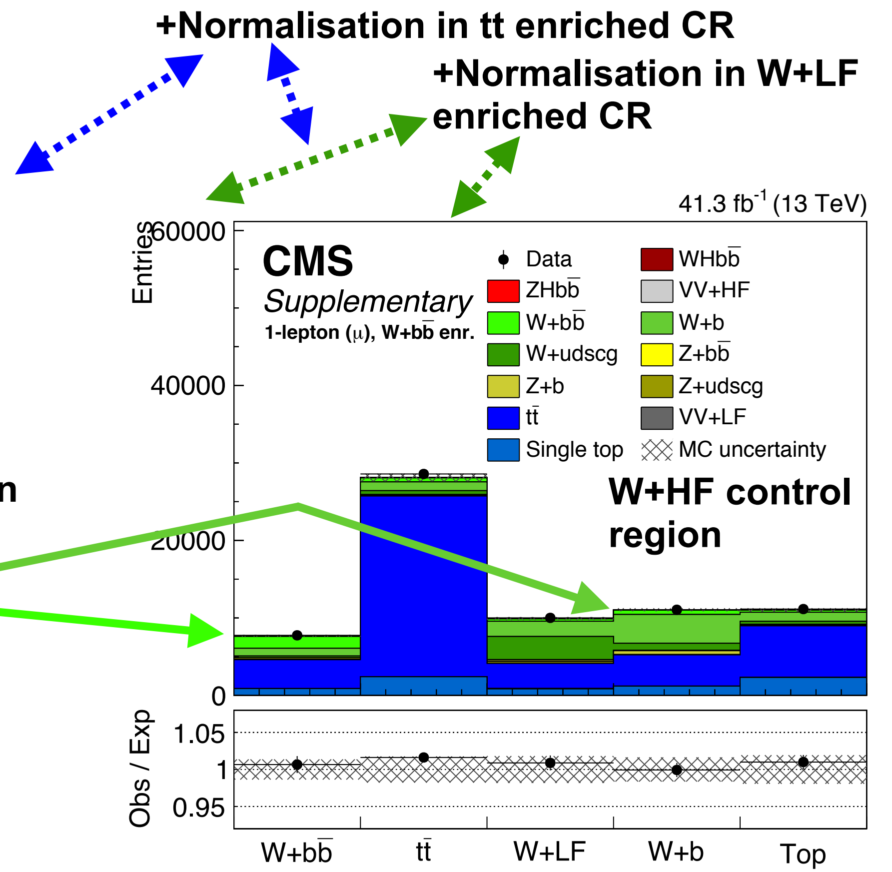
- **Example: constraining background normalisations in 1-lepton channel**
  - **$t\bar{t}$** and **W+LF** background: single-bin control regions
  - **W+HF** background: DNN multi-classifier to distinguish between background components
- Similar strategy in 0-lepton channel, 2-lepton channel Z+HF control region is more pure and we fit the b-tagging discriminator distribution in 2 bins.



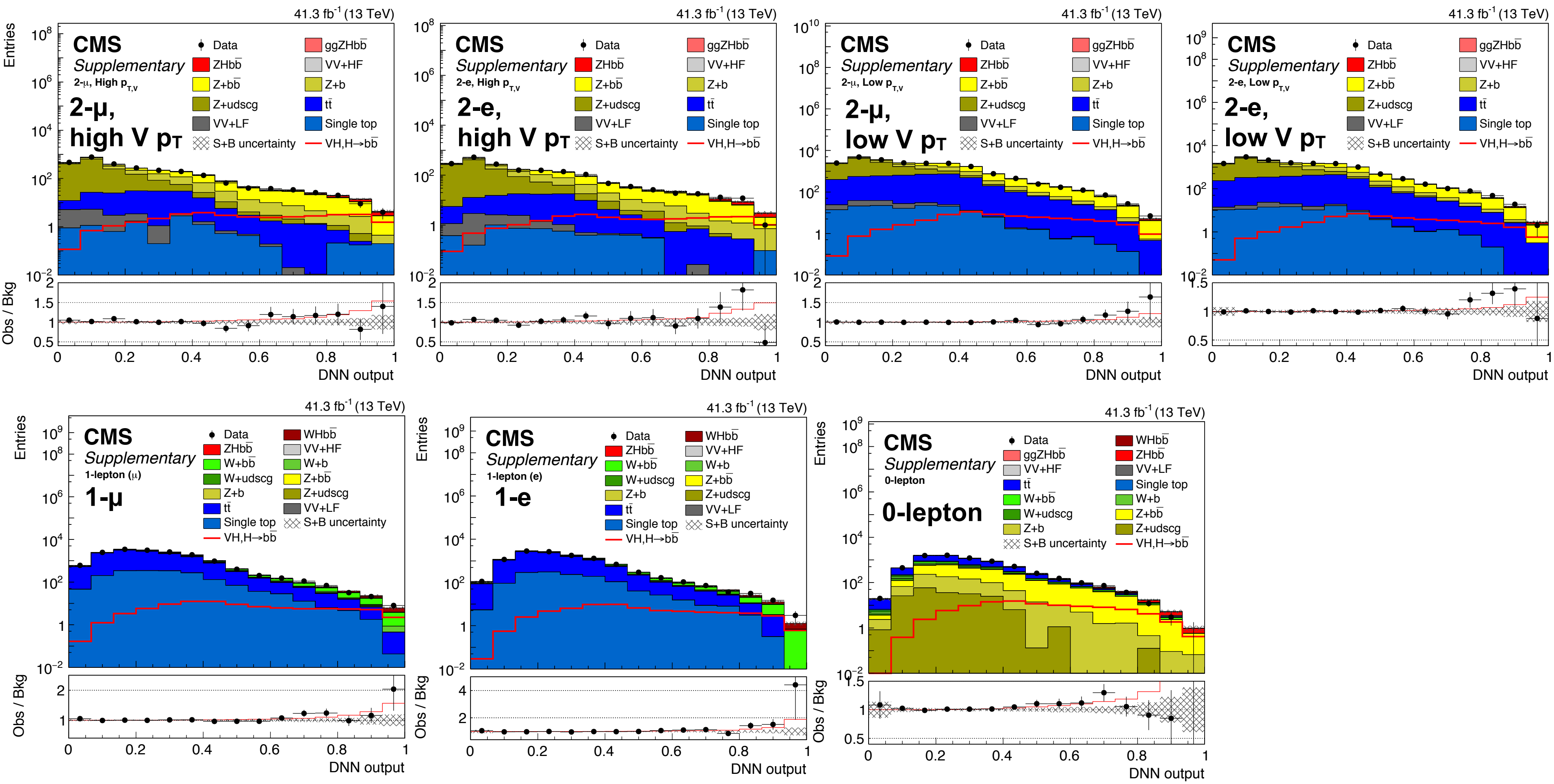
Background normalisation

- **Example: constraining background normalisations in 1-lepton channel**
  - **$t\bar{t}$** and **W+LF** background: single-bin control regions
  - **W+HF** background: DNN multi-classifier to distinguish between background components
- Similar strategy in 0-lepton channel, 2-lepton channel Z+HF control region is more pure and we fit the b-tagging discriminator distribution in 2 bins.

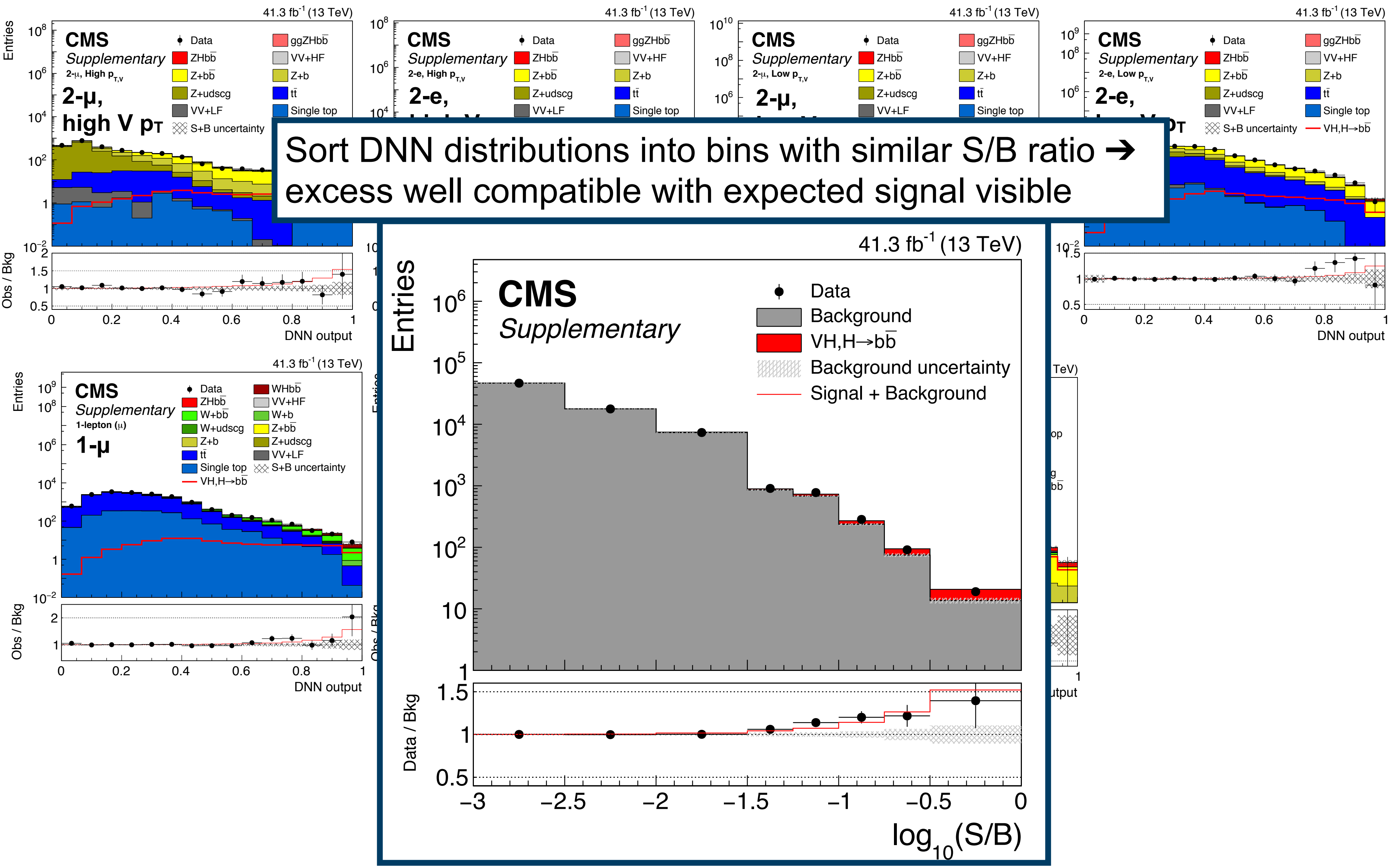




A closer look at the signal regions



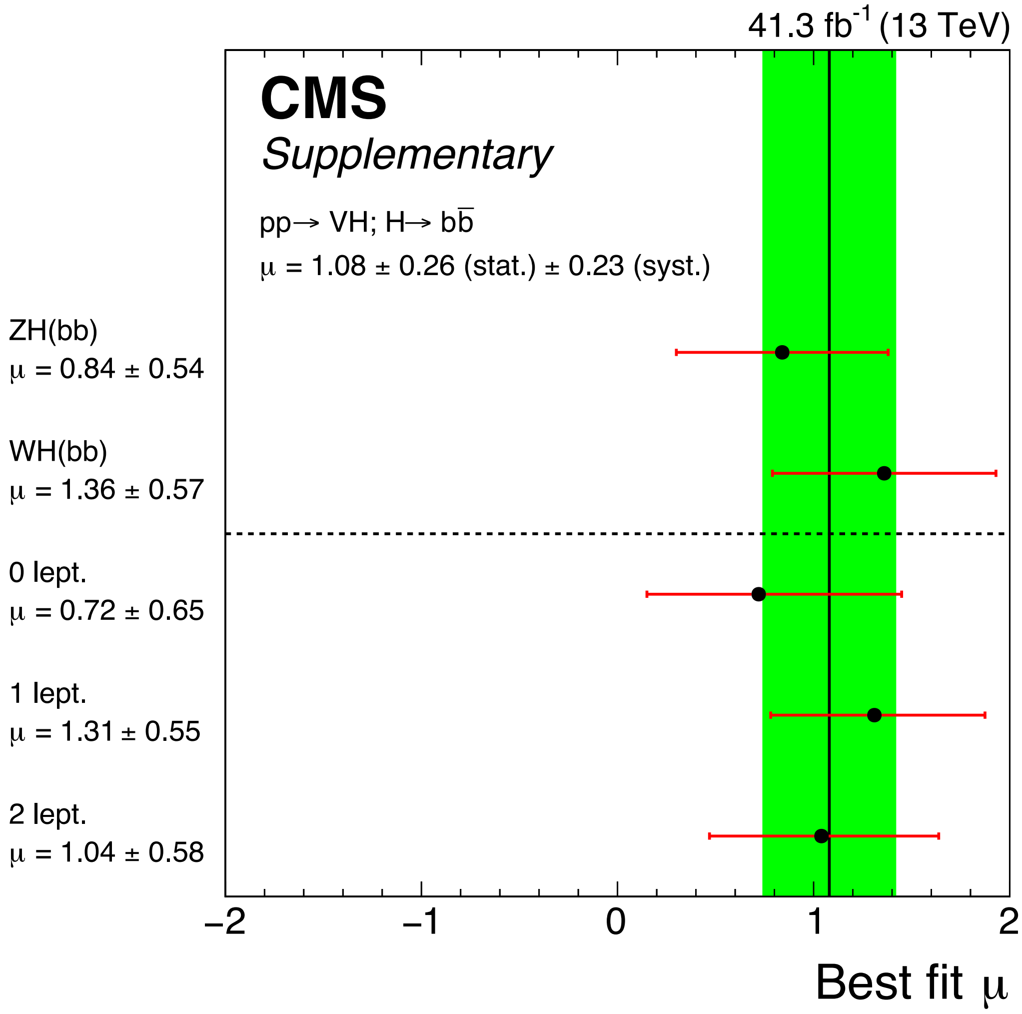
A closer look at the signal regions



Sort DNN distributions into bins with similar S/B ratio ➔ excess well compatible with expected signal visible

Results of analysis on 2017 dataset

- **Result of 2017 analysis: μ=1.08 ± 0.26 (stat.) ± 0.23 (syst.)**
  - Compatible with SM expectation



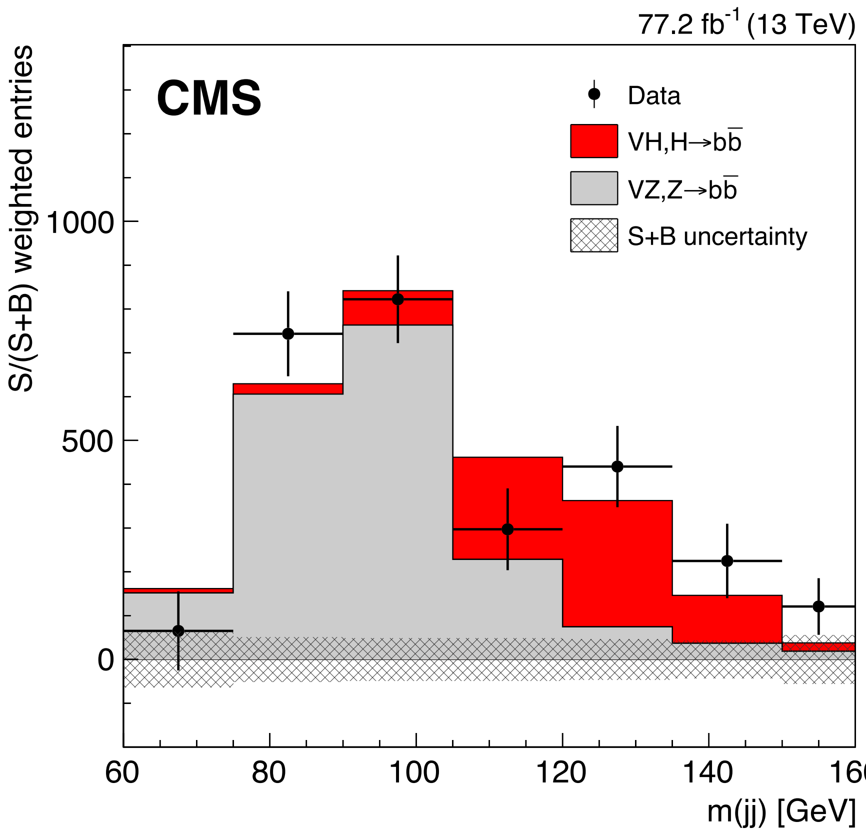
Systematic uncertainties

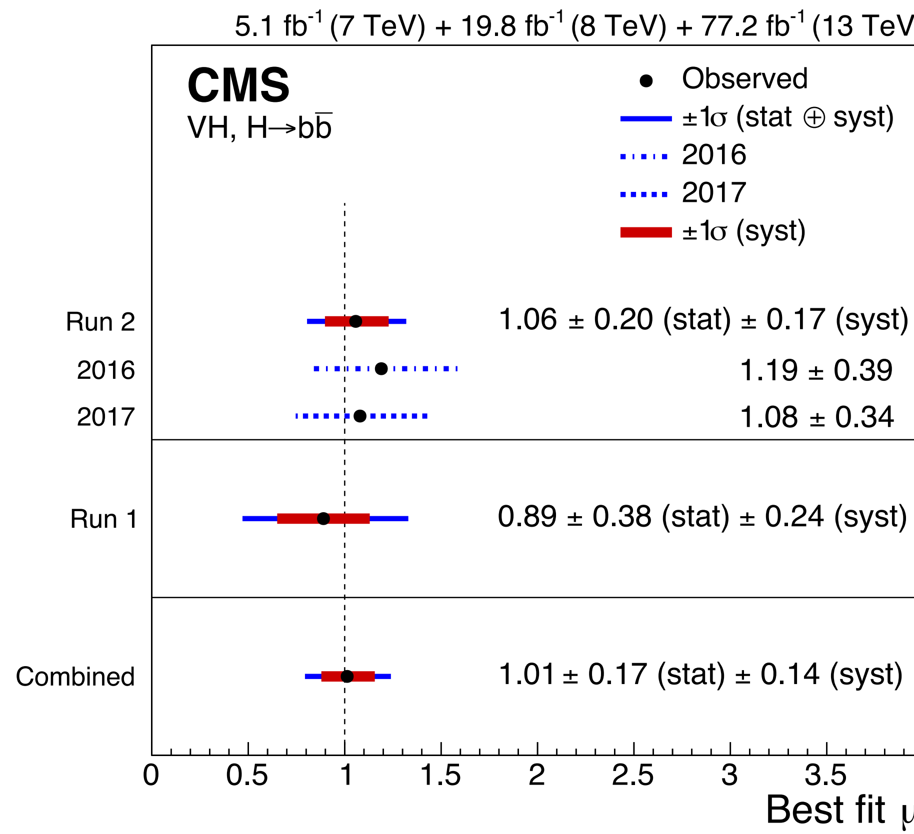
- **Total uncertainty ~0.34, statistical and systematic component of similar order**
- Major sources of systematic uncertainty: background normalisation, size of simulated samples, b-tagging

Uncertainty source	$\Delta\mu$		
Statistical	+0.26	−0.26	**Statistical: 0.26**
Normalization of backgrounds	+0.12	−0.12	
Experimental	+0.16	−0.15	**Systematic: 0.23**
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	

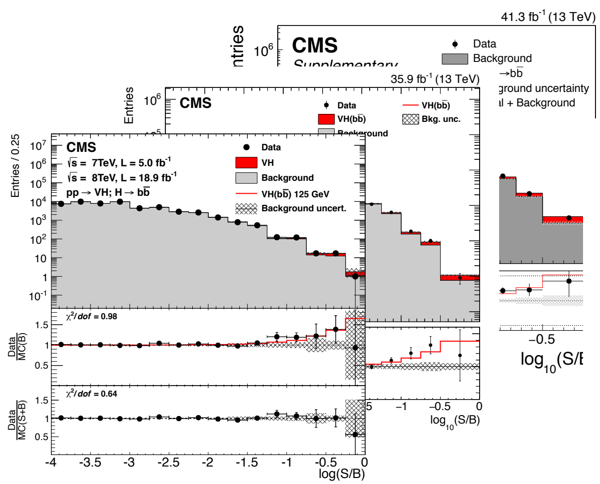
Results

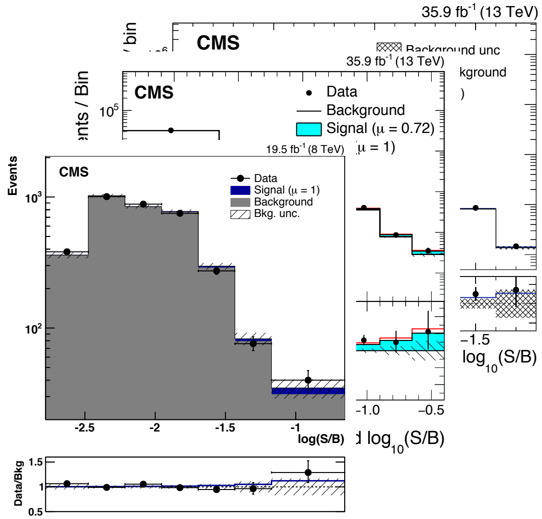
- Combination of Run 1 + Run 2 VH($b\bar{b}$) analyses
  - **μ=1.01 ± 0.17 (stat) ± 0.14 (syst)**
  - Significance: 4.8σ obs (4.9σ exp)

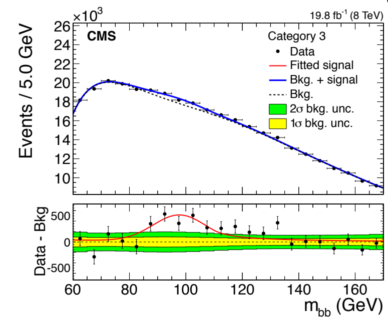


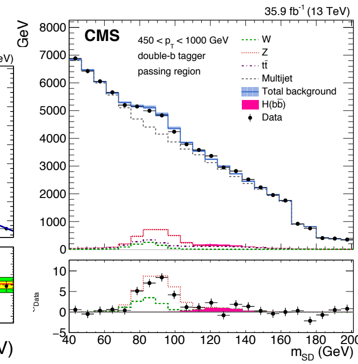


Combination of H➔bb̄ analyses





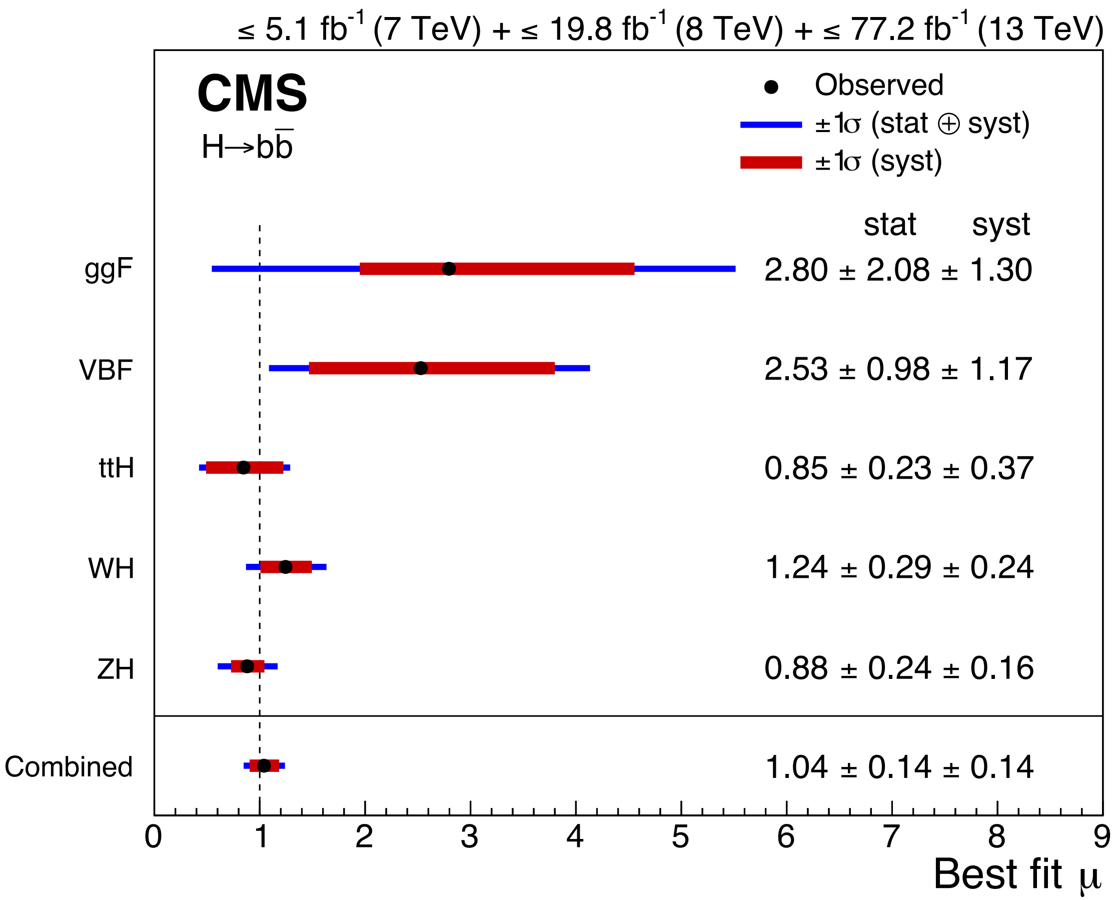




Run-1 & 2016 & 2017 VH ✚ Run-1 & 2016 tt̄H ✚ Run-1 VBF ✚ 2016 ggH

- Combine all published H➔bb̄ results from CMS in Run-1 and Run-2
- Theoretical uncertainties correlated between different measurements (signal, background) and data-taking periods (signal)
- Experimental uncertainties considered correlated between measurements and data-taking periods in some cases.

Combination of H→b$\bar{b}$ analyses

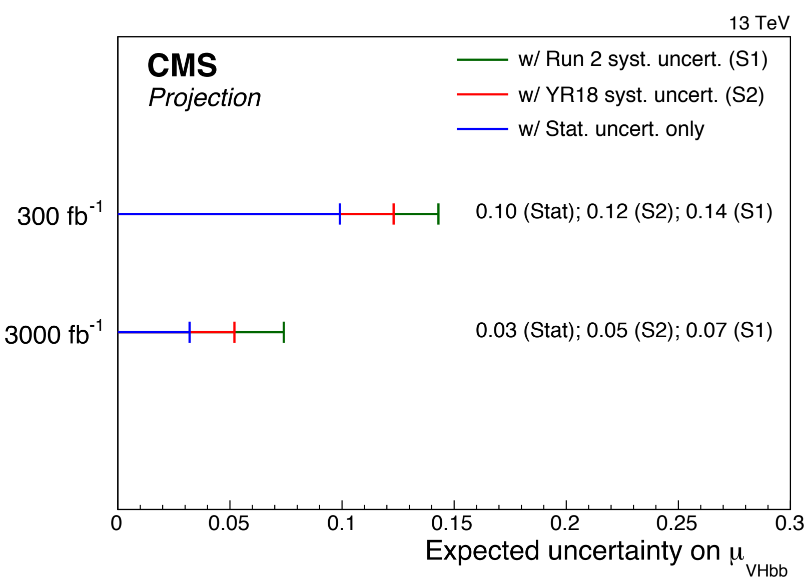


- Combine VH(b$\bar{b}$) results with:
  - Run 1 + 2016 t$\bar{t}$H(b$\bar{b}$)
  - 2016 boosted ggH(b$\bar{b}$)
  - Run 1 VBF (b$\bar{b}$)
- **μ=1.04 ± 0.14 (stat) ± 0.14 (syst)**
- observed (expected) significance 5.6σ (5.5σ)

Sensitivity driven by VH(b$\bar{b}$) analyses, with t$\bar{t}$H also making a sizeable contribution

VH($b\bar{b}$) at HL-LHC

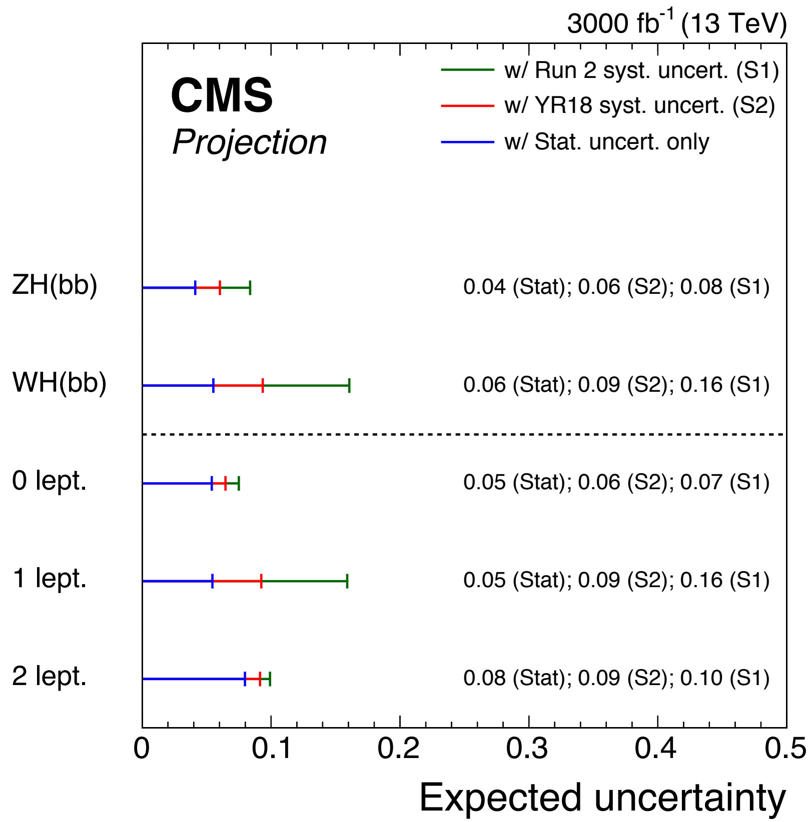
- HL-LHC: high-luminosity running of LHC, starting 2026. Expected L^{int}=3000 fb^{-1}
- Project existing analyses to study possible sensitivity at HL-LHC.
- Several scenarios for uncertainties considered:
  - **w/Run 2 syst. uncert (S1):** Uncertainties are unchanged from current implementation
  - **w/YR18 syst. uncert (S2):** Most experimental uncertainties scale down with the square root of the integrated luminosity, until a lower limit is reached. Theoretical uncertainties are assumed halved
  - Also consider the results without any systematic uncertainties applied

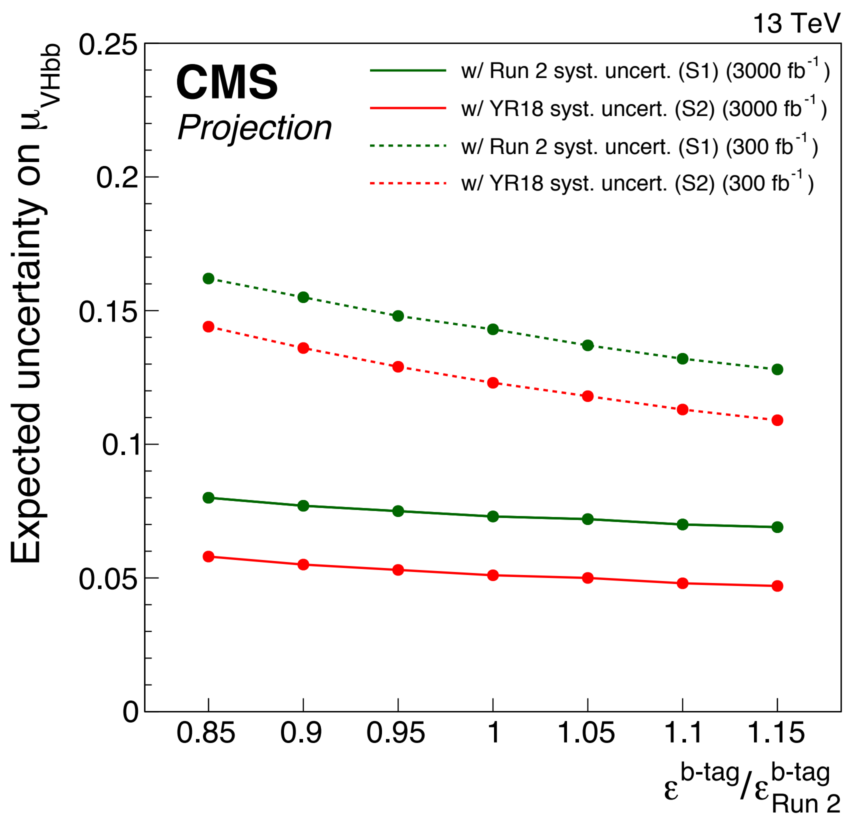


	S1	S2
Total uncertainty	7.3%	5.1%
Signal theory uncertainty	5.4%	2.6%
Inclusive	4.6%	2.2%
Acceptance	2.7%	1.3%
Background theory uncertainty	2.8%	2.3%
Experimental uncertainty	2.6%	2.2%
b-tagging	2.2%	2.0%
JES and JER	0.7%	0.6%
Statistical uncertainty	3.2%	3.2%

Precision limited by theoretical uncertainties

VH($b\bar{b}$) at HL-LHC





All channels contribute equally in the projection ➔ **need to ensure performance of triggers under high PU conditions.**

Change in μ uncertainty as a function of changes in the b-tagging efficiency with respect to the current values ➔ **change is non-negligible**

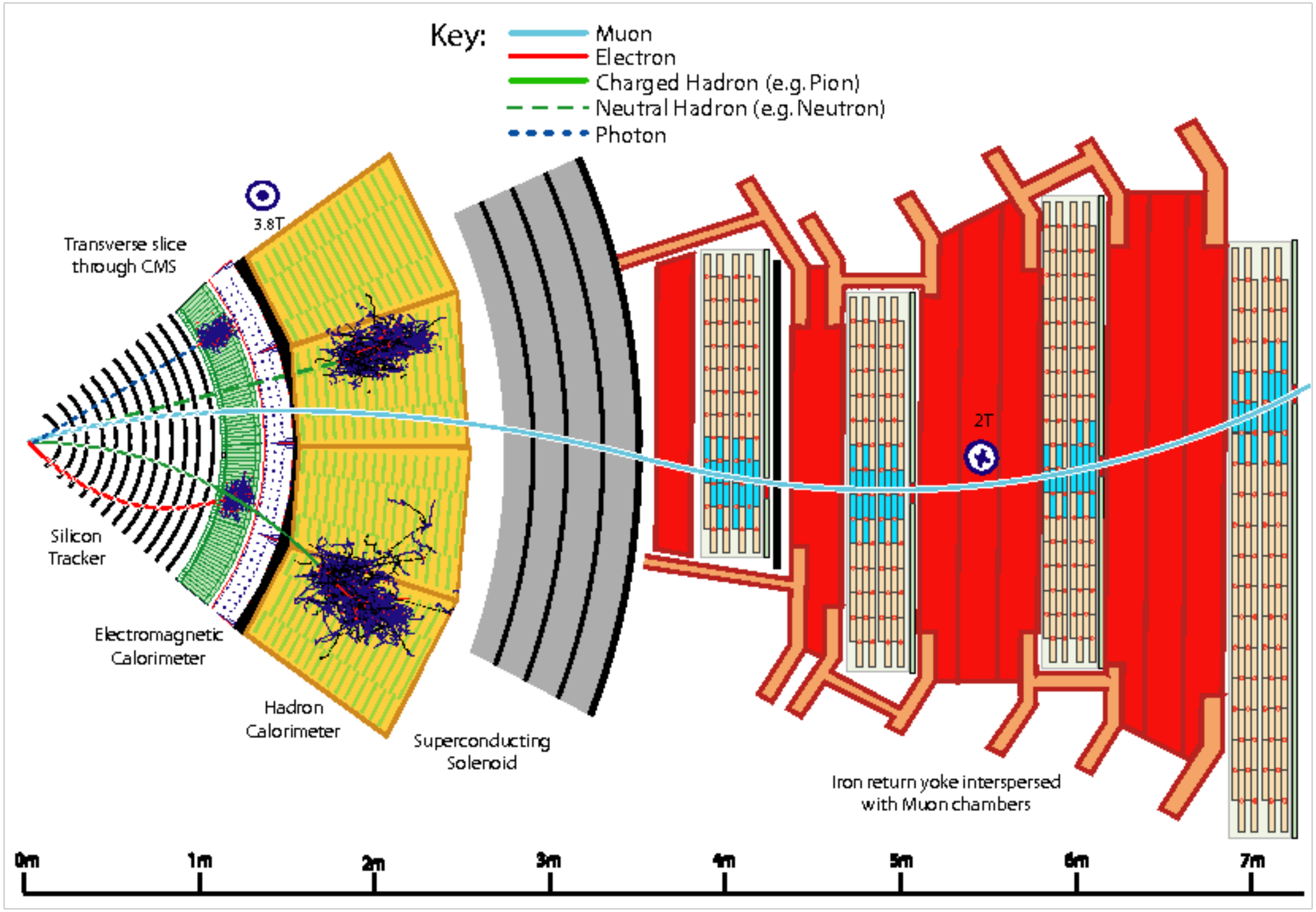
Summary & Outlook

- Presented latest CMS results on VH($b\bar{b}$) and combinations of H→$b\bar{b}$ analyses
  - **We are entering the precision era**
- Results of projections of the analysis to HL-LHC luminosities also shown
  - Work from experimental and theoretical community can lead to **5% precision in signal strength measurements**

- In the shorter term many other interesting measurements to be done
  - Differential cross section measurements
  - Simplified template cross section measurements
  - Searches for anomalous couplings in the VH vertex
  - … and more!
- **Who knows what might be around the corner**

Backup

Particle identification



Cross-check of analysis strategy

- **Extraction of VZ, Z➔b$\bar{b}$**
- Used to validate MVA methods: re-train DNN to separate VZ, Z➔bb from other backgrounds, extract signal strength
- **μ=1.05±0.22: compatible with SM expectation**