Measurements of the
Higgs boson to bottom quark coupling

Krunal Gedia,
On behalf of the CMS Collaboration

Zurich PhD seminar
8th- 9th September 2020
ETH Zürich

Introduction

Standard model of elementary particles

Fermions/ Matter

mass ↓

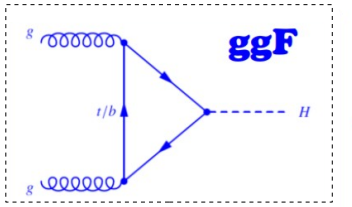
Gen	Quarks	Leptons
I	Up (u), Down (d)	Electron (e), e neutrino
II	Charm (c), Strange (s)	Muon (μ), μ neutrino
III	Top (t), **Bottom (b)**	Tau (τ), τ neutrino

Bosons/Force carriers

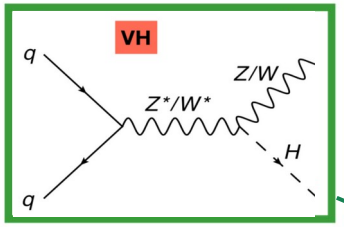
- Z boson
- Photon (γ)
- W boson
- Gluon (g)

Higgs boson

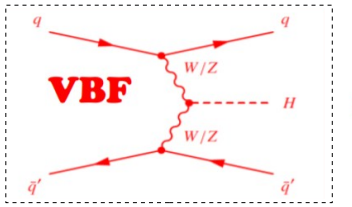
Higgs production modes



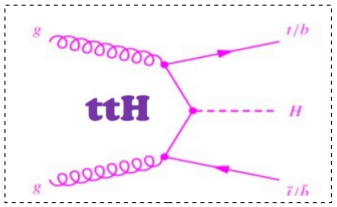
~87%



~4%

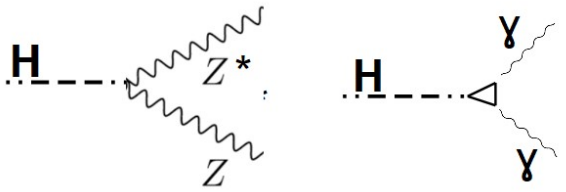


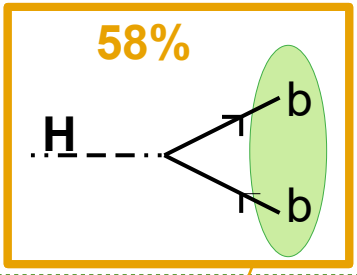
~7%

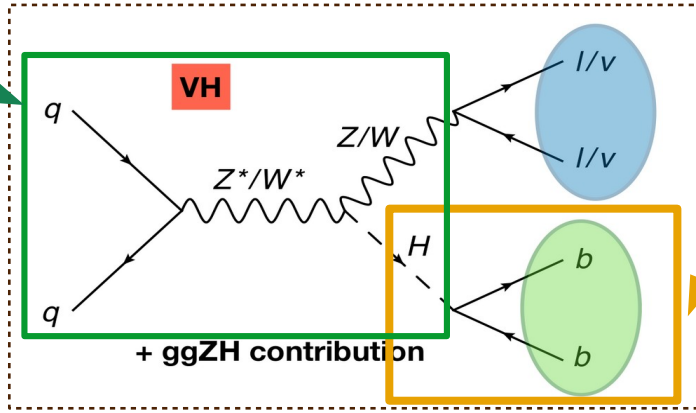


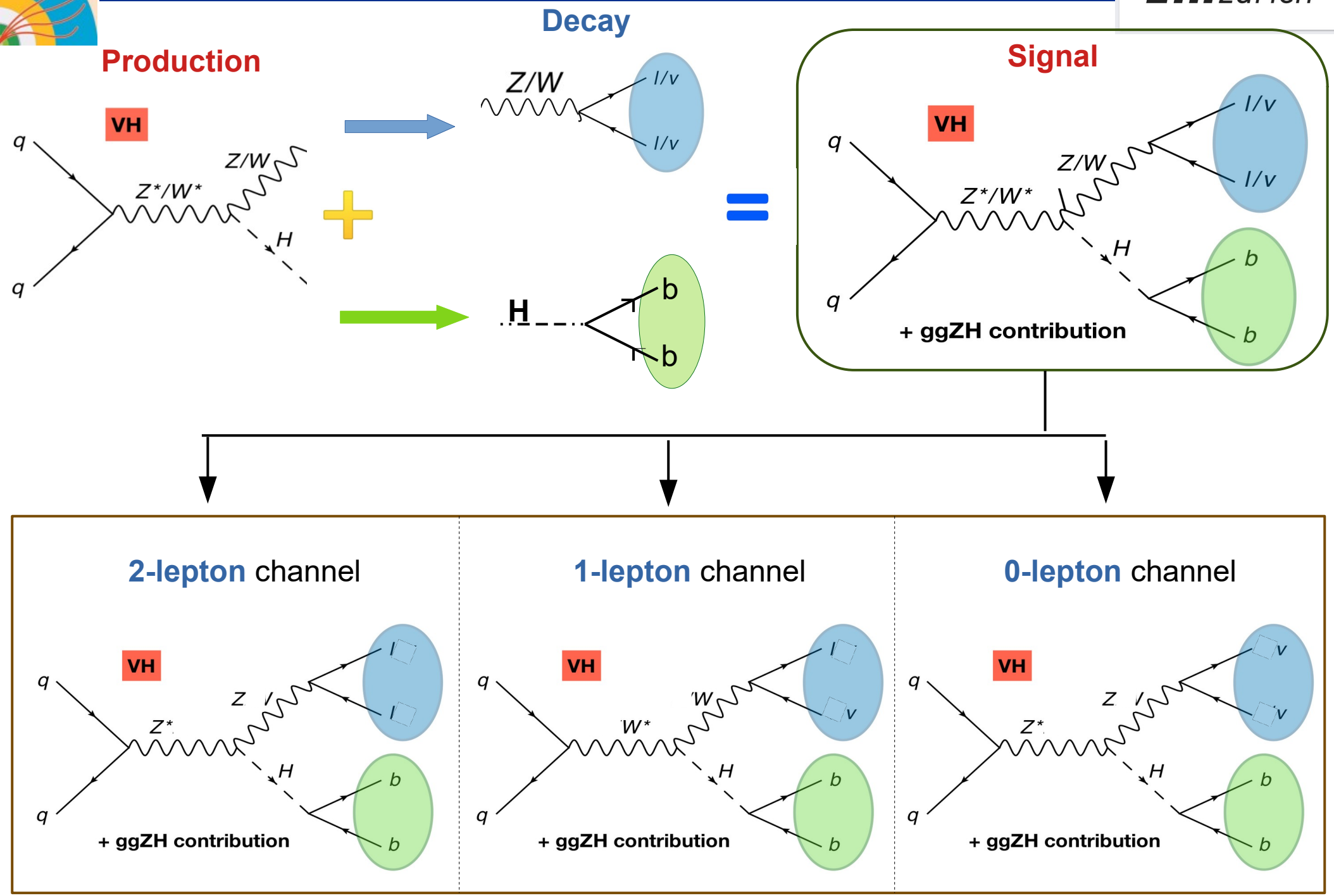
~1%

Higgs decay modes

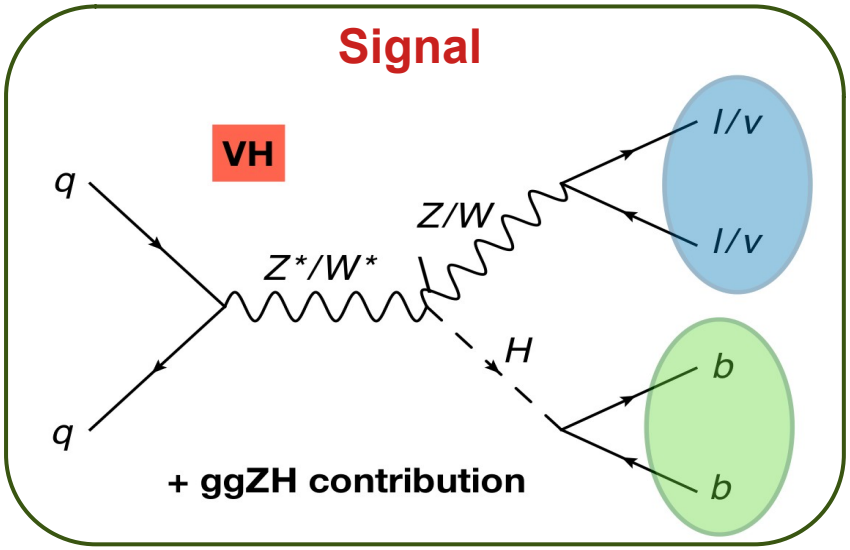




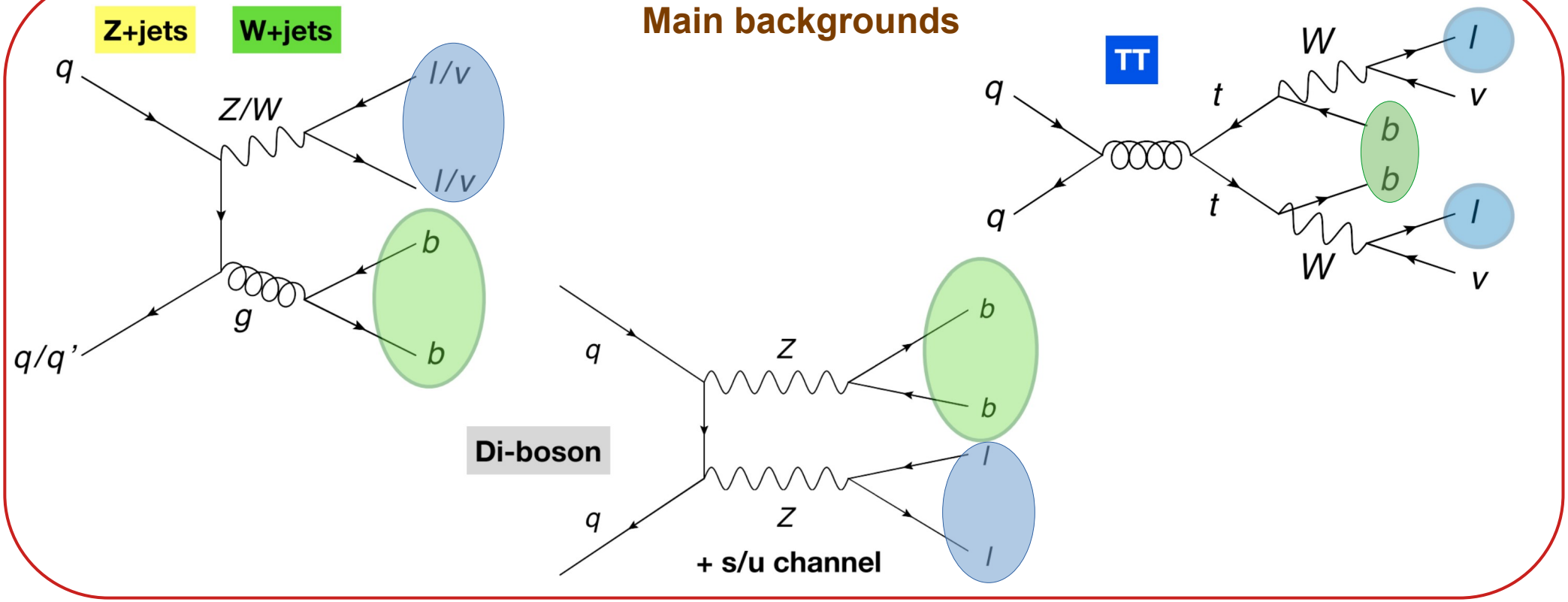



Production
Decay
Signal
VH
q
q
Z*/W*
Z/W
H
Z/W
l/ν
l/ν
H
b
b
+ ggZH contribution
2-lepton channel
1-lepton channel
0-lepton channel
Z*
Z
l
W*
W
ν
+ ggZH contribution

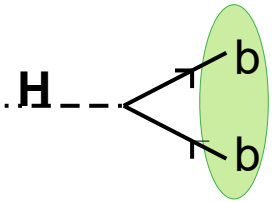
Signal

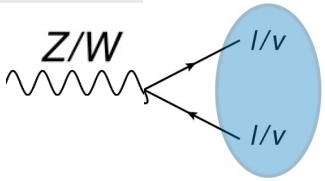


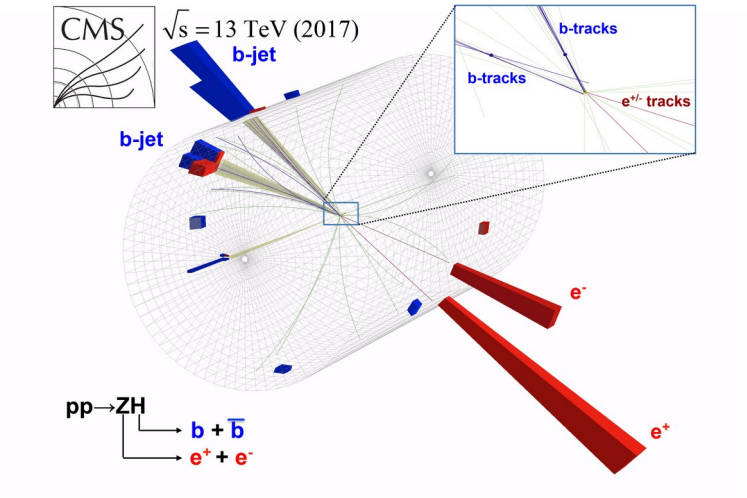
Main backgrounds

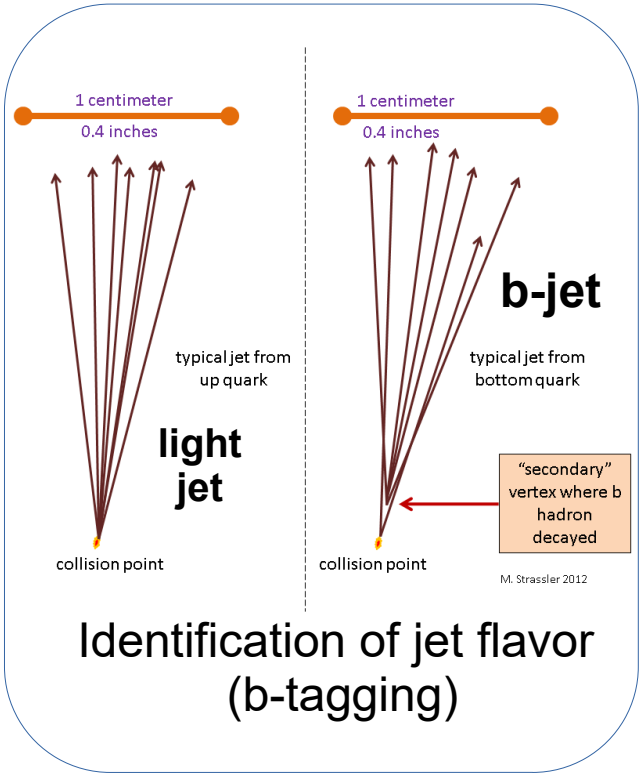


CMS detector

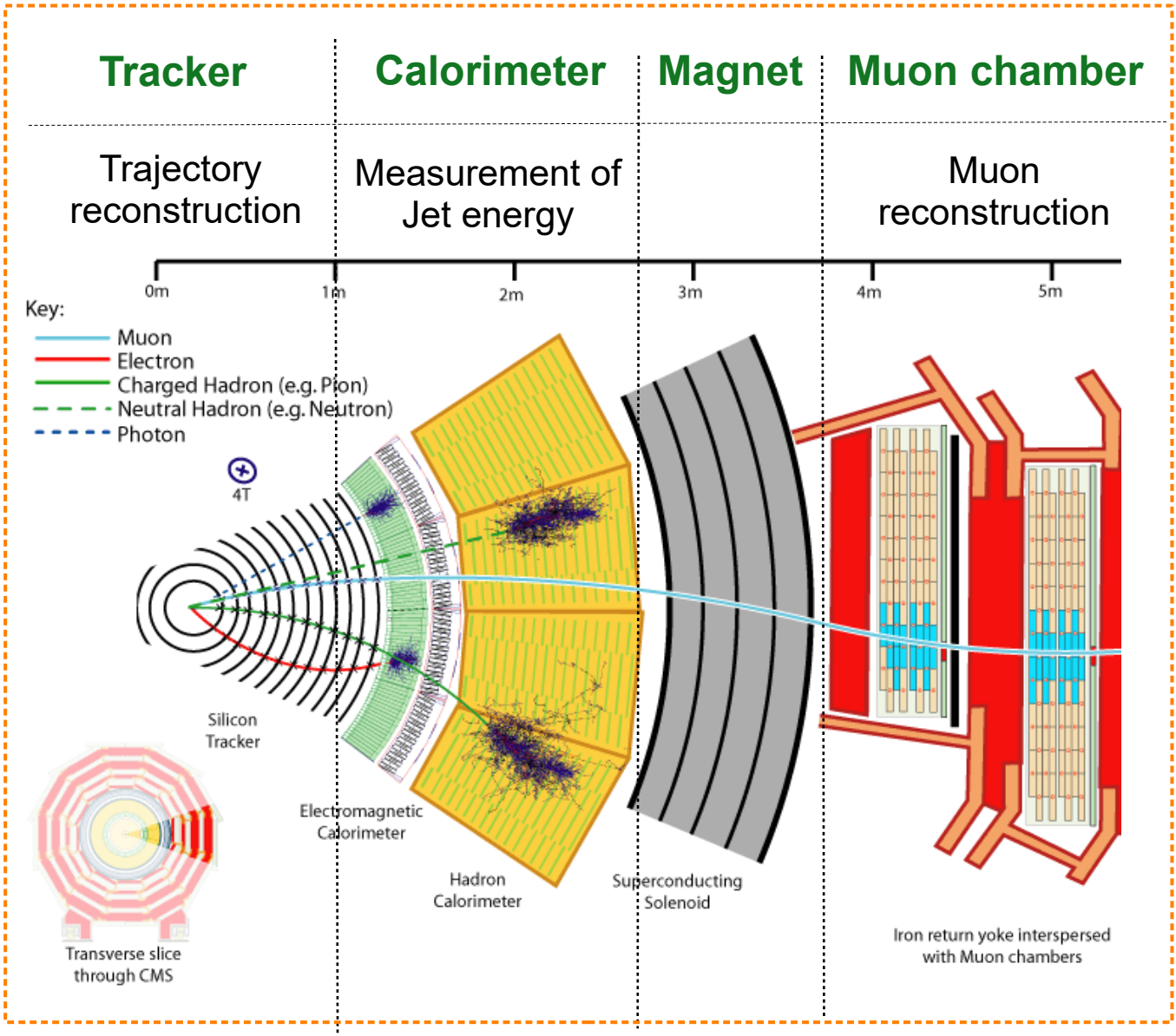








Identification of jet flavor (b-tagging)



Event Selection

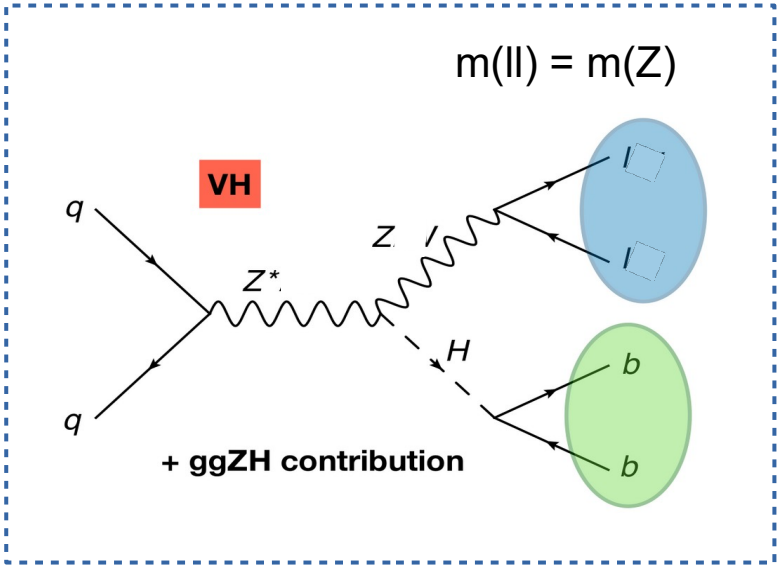
Signal region (SR)

- High signal efficiency.
- Purity (S/B ~ 1 - 5%).
- Used to extract signal strength/significance.

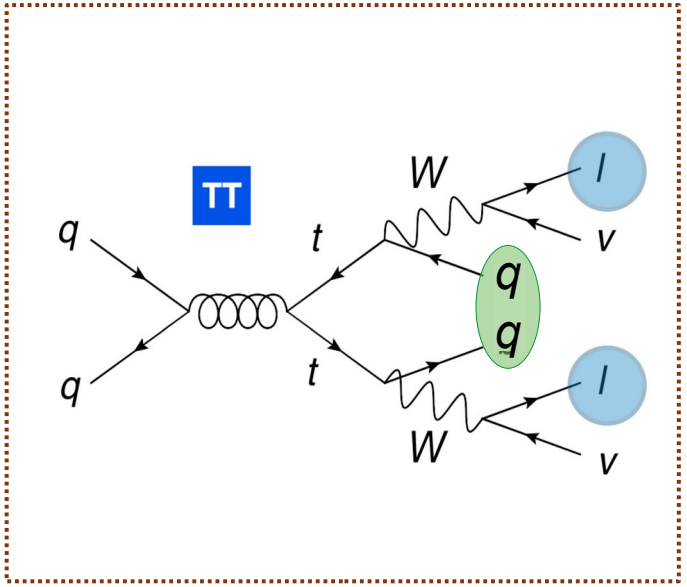
Control region(CR)

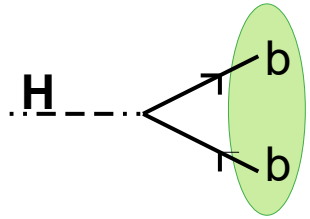
- Enriched in one of the backgrounds.
- **TT CR**, **V+LF CR**, **V+HF CR**.

Signal

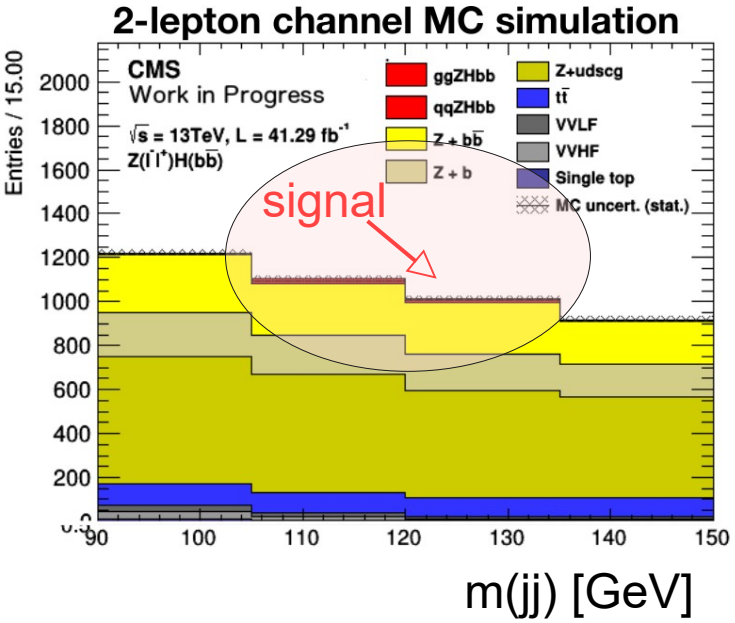


TT CR





m_{jj} in Signal region



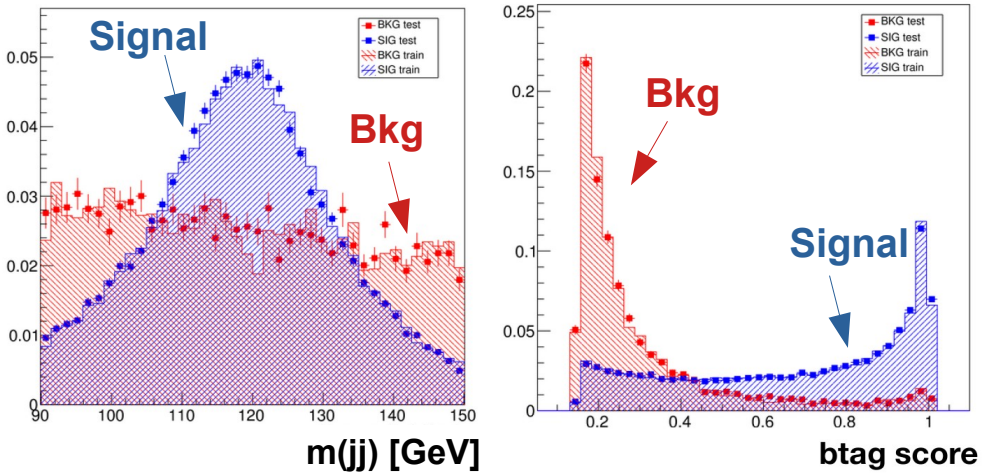
- Signal region still dominated by background
- Mjj not enough to discriminate signal from background.

How to improve the significance of S+B hypothesis?

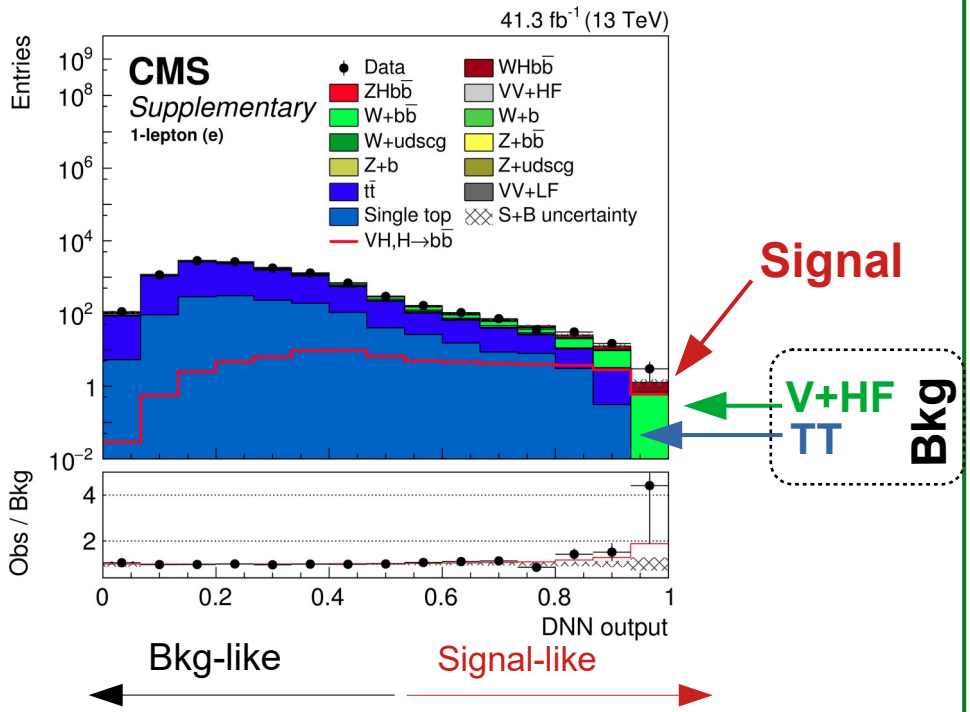
Mjj → MVA variable (trained by many discriminating variables!)

S/B DNN classifier in Signal region

Some of the inputs



DNN score



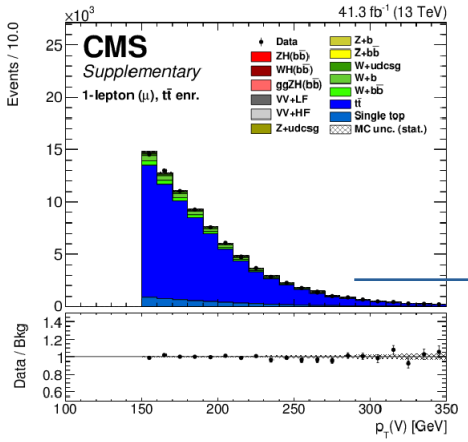
- Now the significance is definitely high!
- But how do we account for the data/MC mis-modelling of background processes?
- For this, we derive background data/MC SF from CRs & extrapolate to SR to extract the significance of the signal.

Simulataneous fit of

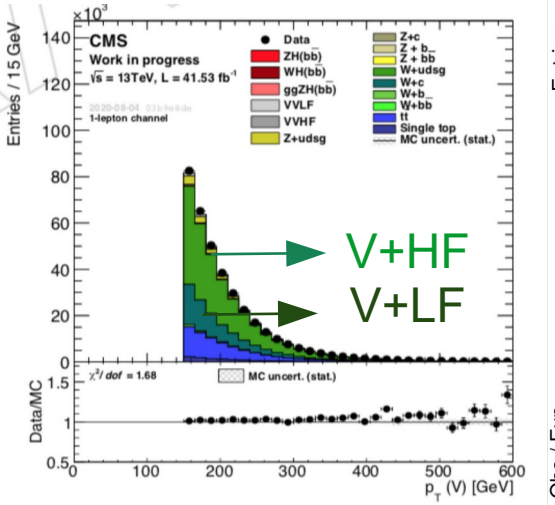
- **SR** (extract significance of S+B hypothesis)
- **CR** (accounts for data/MC mis-match for background, free-float these SF)

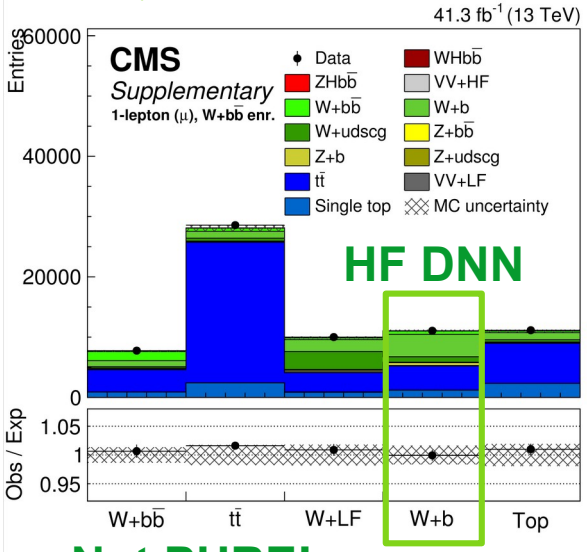
CR in fit:

- **TT** → Pure, float just the normalisation
- **V+LF** → Pure, float just the normalisation
- **V+HF** → Not pure in V+HF process

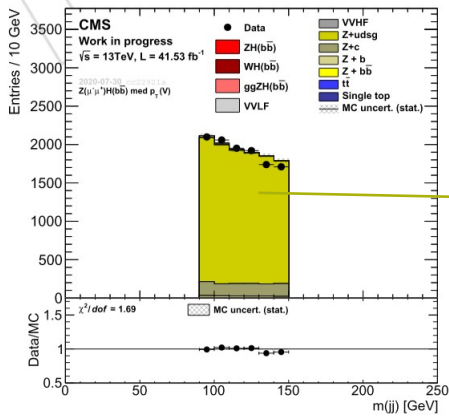


TT CR



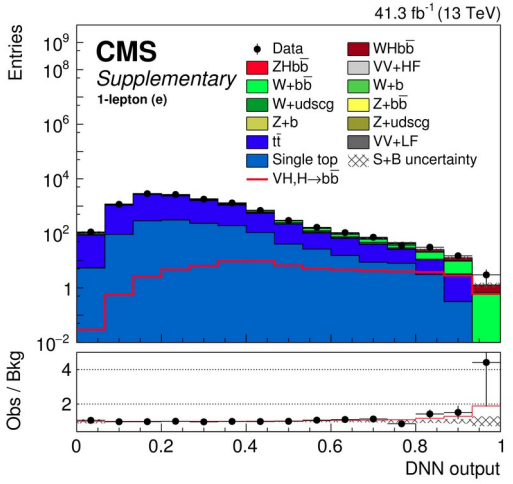


V+HF CR → Not PURE!

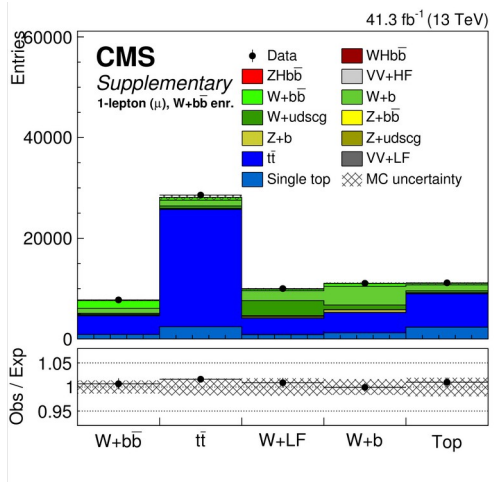


V+LF CR

Simultaneous fit of SR and CR to obtain signal strength/significance



SR: S/B DNN variable



HF CR :Multiclass DNN

V+LF & TT CR : Only yield

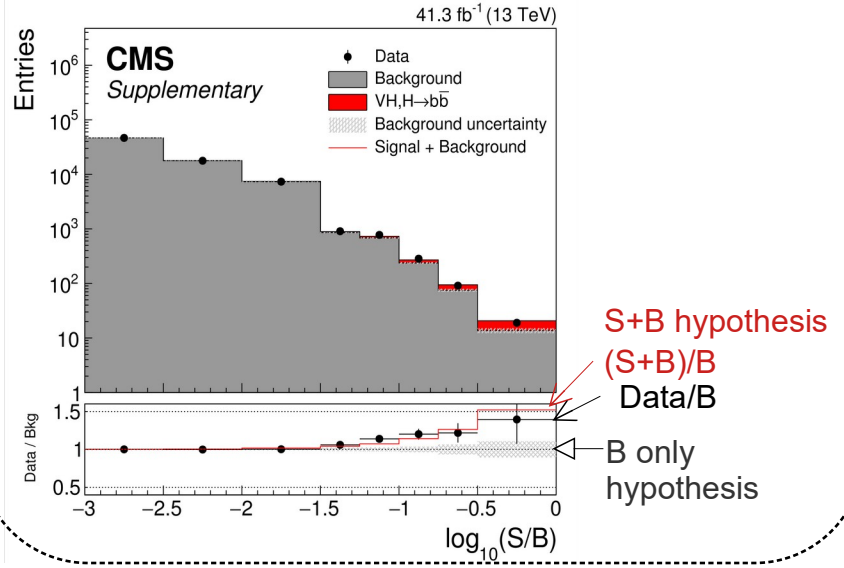
~300 sources of systematic uncertainties

Combination with Run 1 and 2016/17 data

Data set	Significance (σ) 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

Observation!

Excess of events behave as SM Higgs!



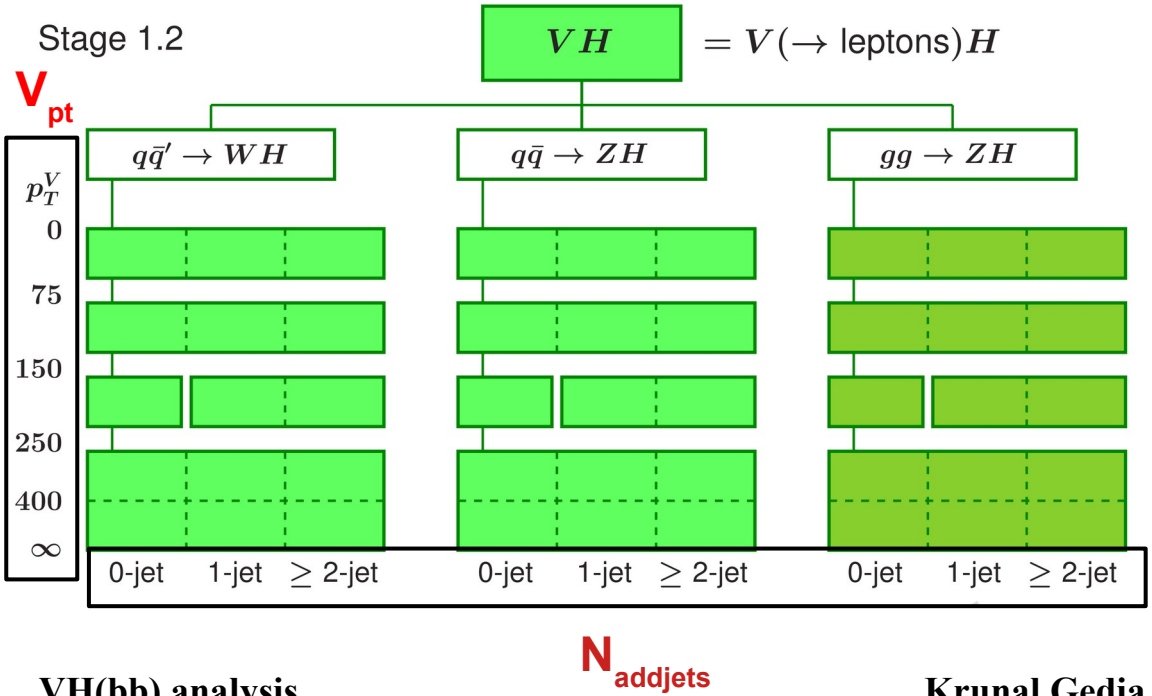
Measuerement of Higgs coupling to bottom quark

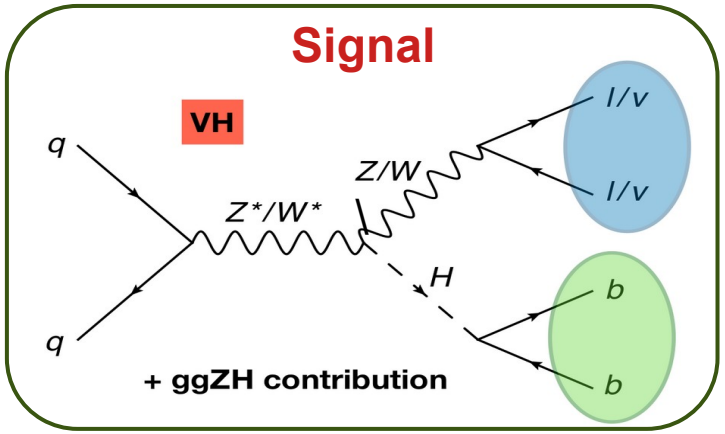
Inclusive (completed) → STXS (ongoing, pre-approval done) → Differential (future)

Real data: Collisions → Detector → Trigger → Reconstruction

MC: **Event generator** → Simulated detector → Trigger → Reconstruction

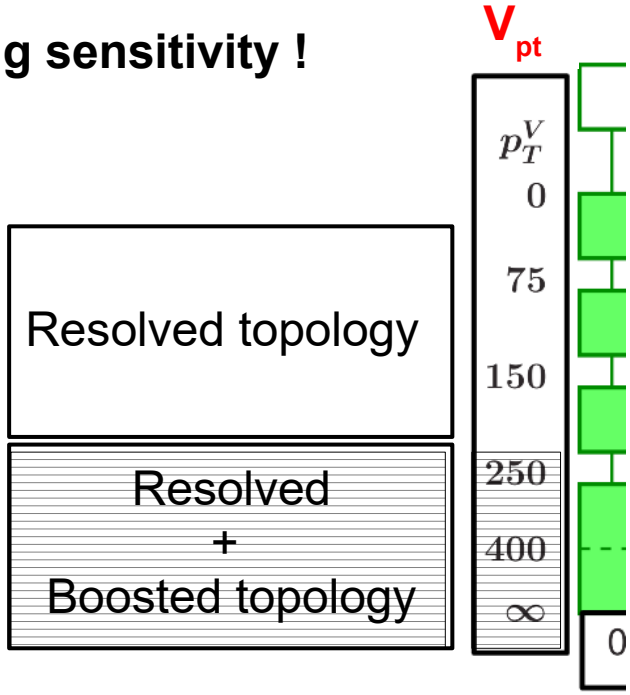
- Gen-level info/MC truth
- Partition signal in V_{pt} and $N_{addjets}$
- Reco-level info
- Analyse and measure H xsec, σ here

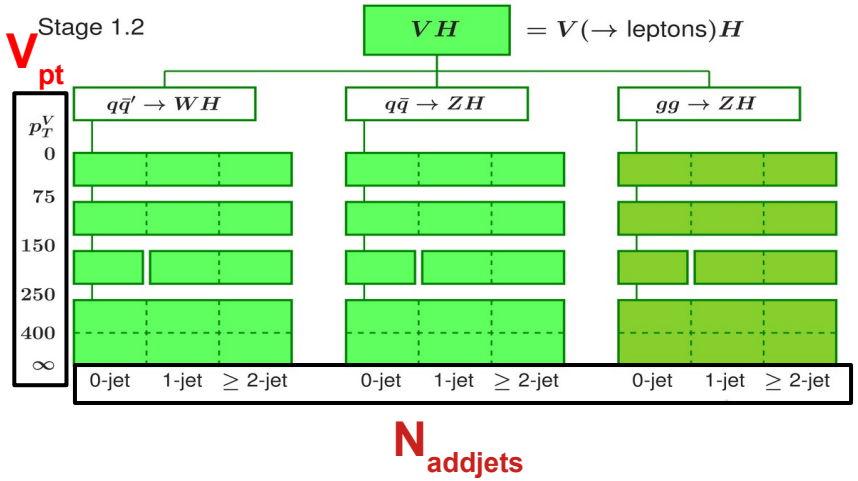


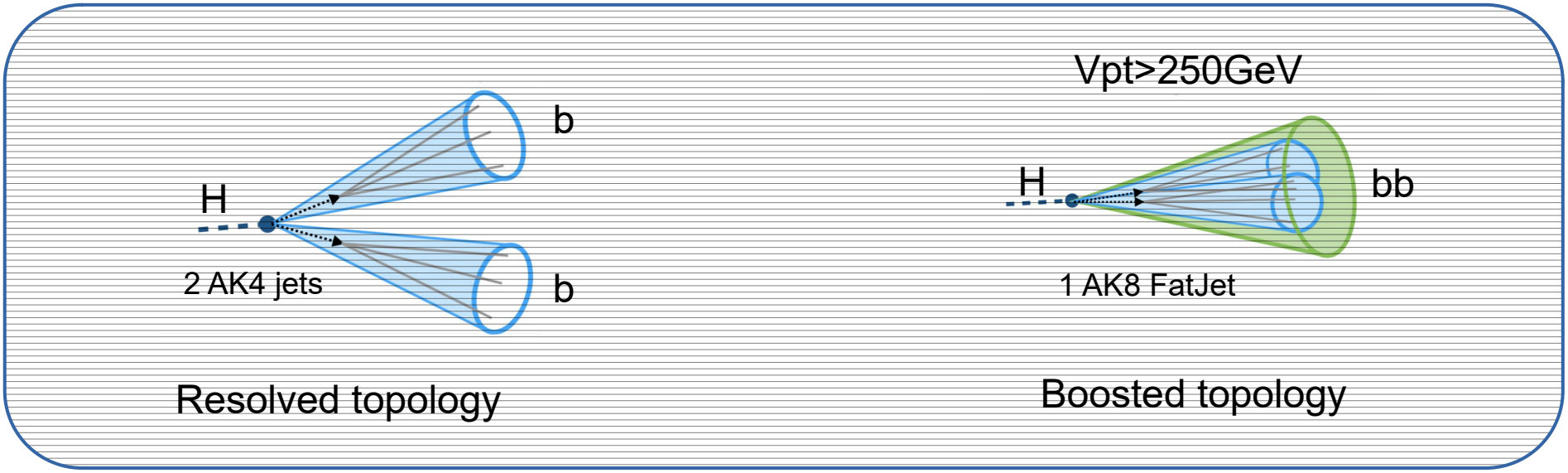


Advantages:
Reduce theoretical unc., NP models study, allows complexity

Increasing sensitivity !

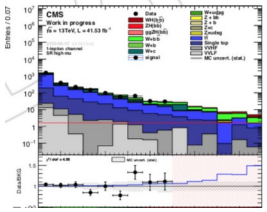




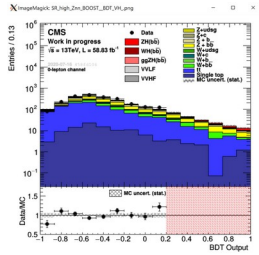


Simultaneous fit of SR and CR to obtain signal strength/significance

SR



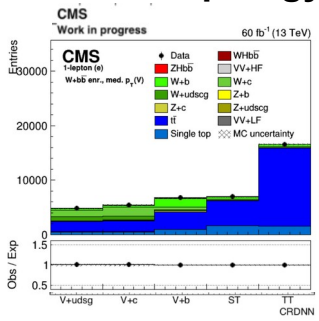
Resolved topology
S/B DNN classifier



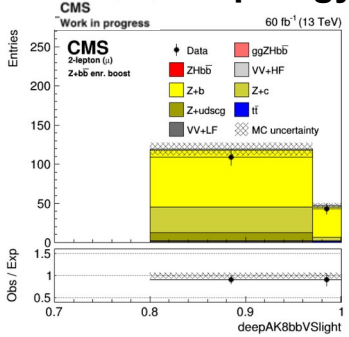
Boosted topology
S/B BDT classifier

CR

Resolved topology



Boosted topology



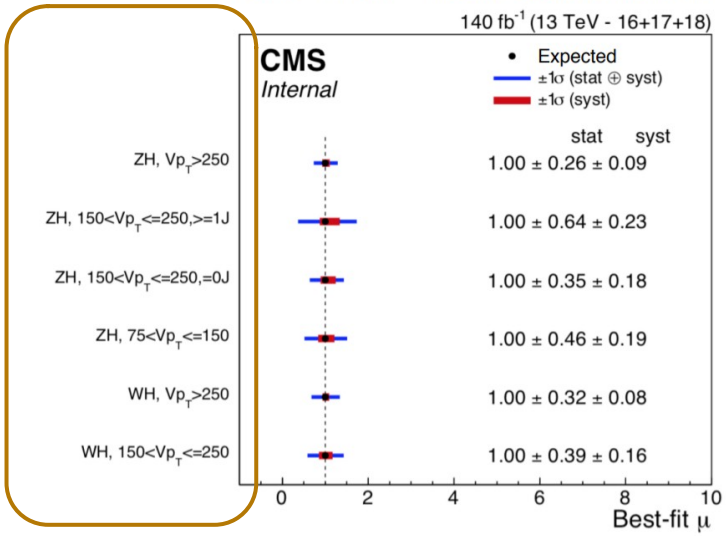
V+LF & TT CR: V_{pt}
V+HF : Multi-class DNN

FatJet tagger

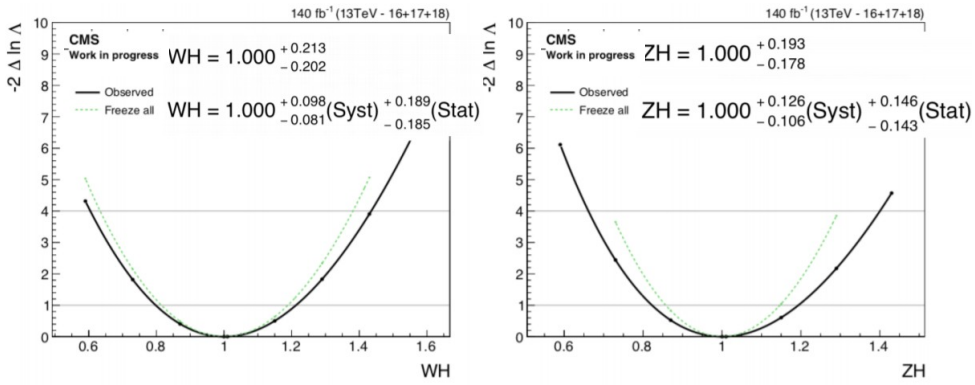
~300 sources of systematic uncertainties

New dataset: 2018

2016+2017+2018 combination

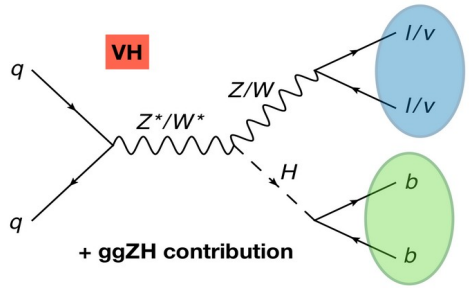


Expected STXS sensitivity

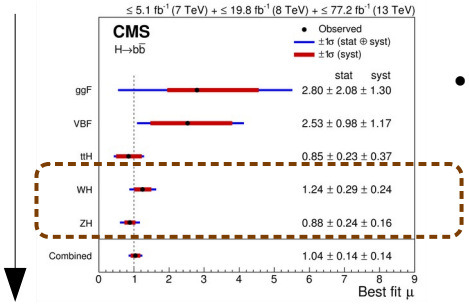


WHbb, ZHbb individually around 5 sigma level!

Summary of higgs boson coupling to b-quark measurements in associated vector boson production mode

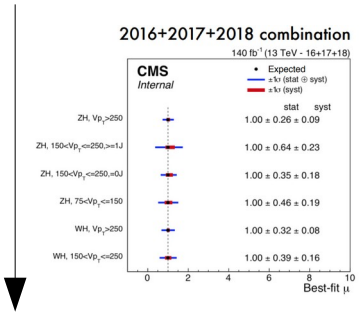


Inclusive (completed)



- Strategy: Measuements done in inclusive phase space.
- Observed significance of VHbb with Run 1 + 2016 + 2017 data is **4.8σ**.

STXS (ongoing)



- Strategy: Measurements done in bins of V_{pT} and $N_{addjets}$
- Increased sensitivity reach: Addition of boosted topology
- Expected significance of each of the WHbb and ZHbb using Run 2 dataset (2016 + 2017 + 2018) atleast **5.0σ** .

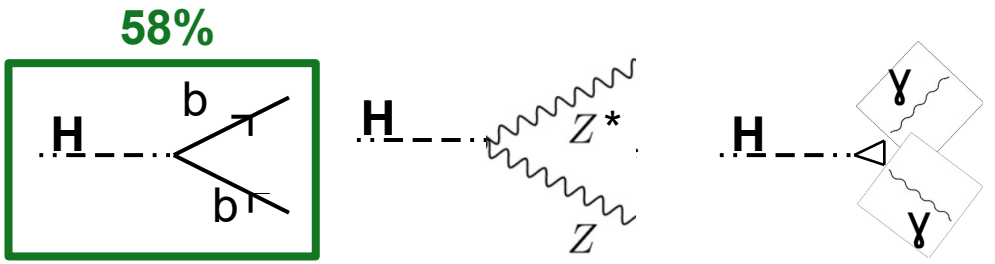
Differential (future)

BACK-UP

Introduction

- The Higgs boson discovered in H→ γγ , H→ZZ* →4l channels in 2012.
- m_H known with 0.12% precision level.
- Other observed properties & couplings compatible with SM but at precision level >10%

Higgs decay modes



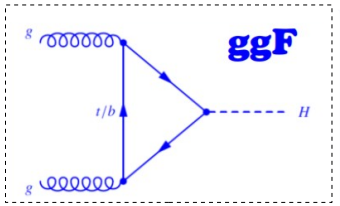
Why H→bb ?

- Measure Higgs boson to bottom quark coupling.
- Higgs boson has largest branching ratio to b-quarks.
- However, the H→bb coupling was not established in Run 1.

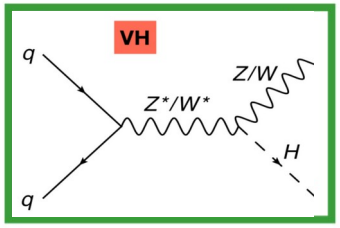
Challenges in Run 1:

- QCD multijet background
- Lack of SV discrimination in L1 trigger

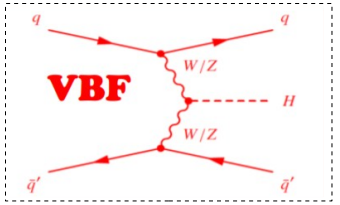
Higgs production modes



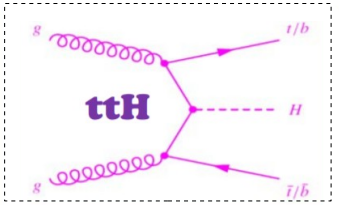
~87%



~4%

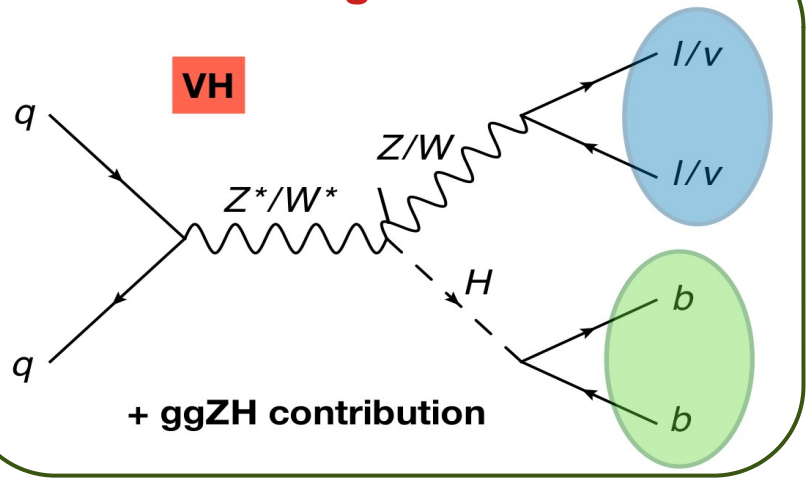


~7%



~1%

Signal



Why VHbb to study Hbb coupling?

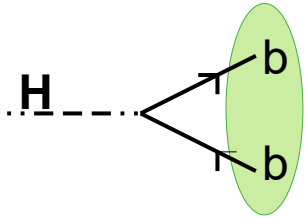
- QCD multi-jet background suppressed by boosted V and H (sensitivity mainly > 150 GeV)
- Leptonic decay modes of vector boson provides a trigger path for signal.
- Large MET is used to trigger event in case vector boson decays into neutrinos.
- Boosted V also reduces V+jets background.
- Better m_H resolution.

Story so far

- Evidence for H→bb established using 2016 dataset.
- Finally observed using Run 1 + 2016 + 2017 dataset. [link]

Object reconstruction

Higgs boson (improve mass resolution)



- **2 highest b-tag score jets.**

DNN based b-jet regression

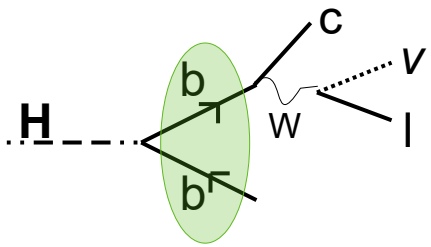
- Corrects for escaping neutrino, calibration mis-match, etc.

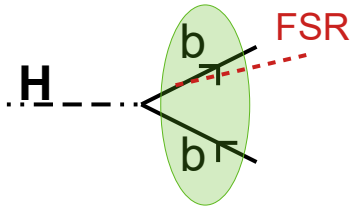
FSR recovery

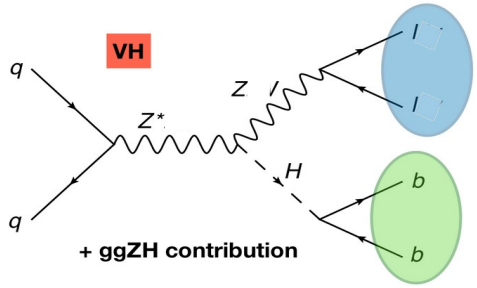
- Add 4-vectors of FSR jets to b-jets.

Kinematic fit

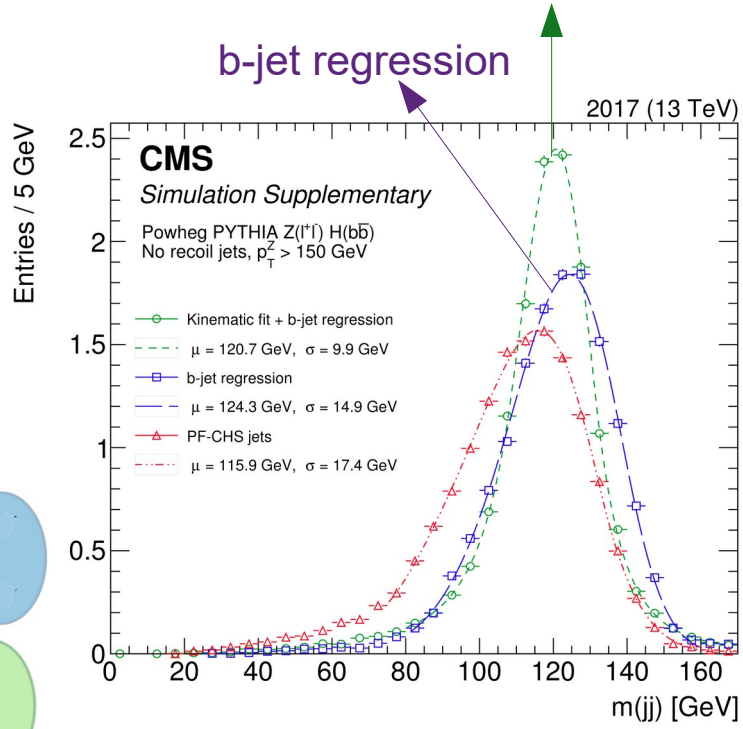
- Constraints: m(ll) = m(Z) and pT(total) = 0





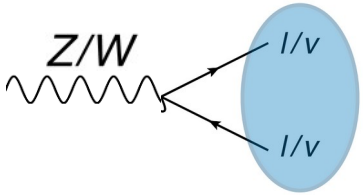


Kinematic fit+b-jet regression



Vector boson

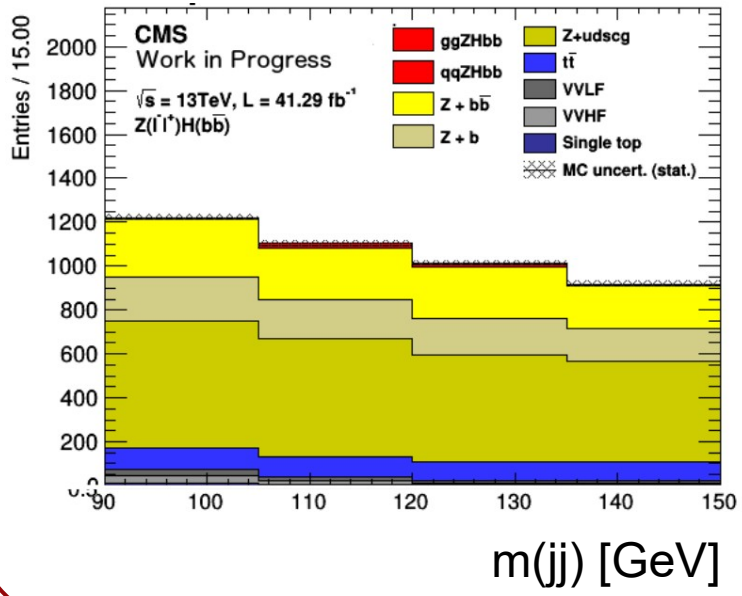
- 2 hardest isolated opposite sign leptons or isolated leptons + MET or just MET.



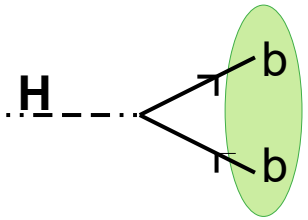
Asimov significance:

$$\sum_{bins} \frac{s}{\sqrt{b}}$$

- How well can we reject background only hypothesis
- Compute it on a S/B discriminating variable in signal region (SR)
- Use it also to optimize selection cuts to define the SR



Signal region



- Very low significance
- Signal region still dominated by background

- How to increase significance?
- Better selections? Already optimized!
- Better signal-background discriminating variable? Yes

Mjj → MVA variable (trained by all discriminating variables!)

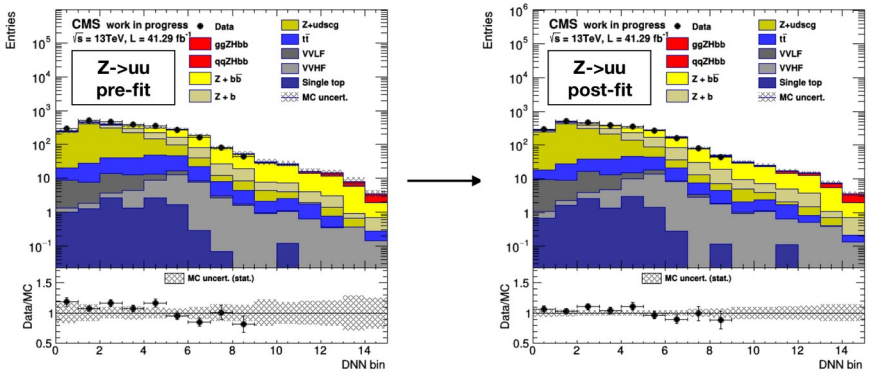
back-up

Results of simultaneous fit

SF of background normalization

Process	$Z(\nu\nu)H$	$W(\ell\nu)H$	$Z(\ell\ell)H$ low-p_T	$Z(\ell\ell)H$ high-p_T
$W+$udscg	1.04 ± 0.07	1.04 ± 0.07	–	–
$W+b$	2.09 ± 0.16	2.09 ± 0.16	–	–
$W+b\bar{b}$	1.74 ± 0.21	1.74 ± 0.21	–	–
$Z+$udscg	0.95 ± 0.09	–	0.89 ± 0.06	0.81 ± 0.05
$Z+b$	1.02 ± 0.17	–	0.94 ± 0.12	1.17 ± 0.10
$Z+b\bar{b}$	1.20 ± 0.11	–	0.81 ± 0.07	0.88 ± 0.08
$t\bar{t}$	0.99 ± 0.07	0.93 ± 0.07	0.89 ± 0.07	0.91 ± 0.07

Improvement in data/MC agreement & reduction in uncertainty



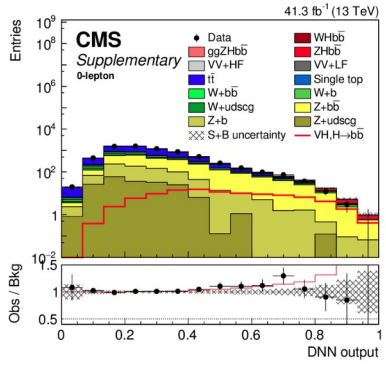
Back-up

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

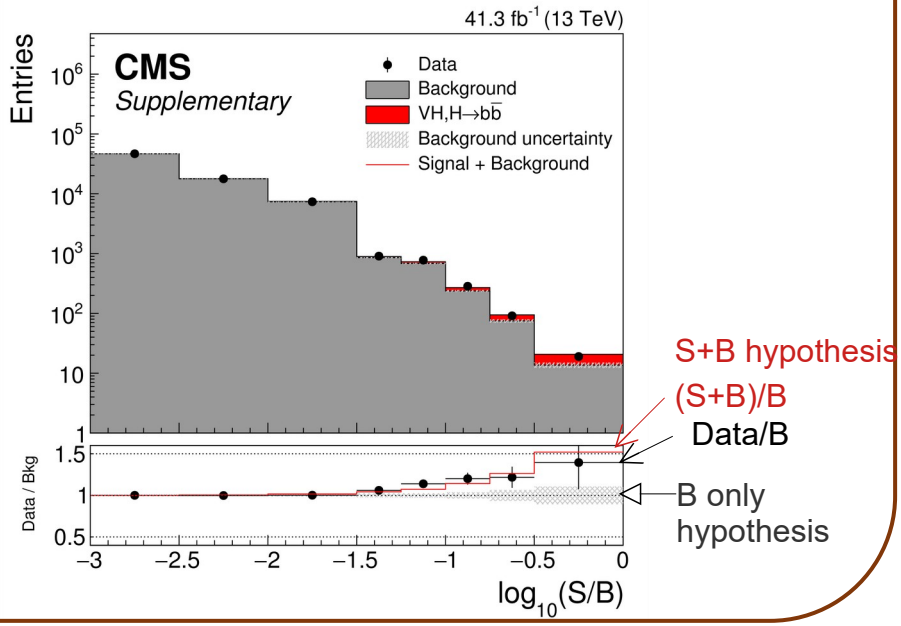
Combination of 0,1 and 2-lepton channels

Excess of events behave as SM Higgs!

Combine post-fit DNN scores of all SR



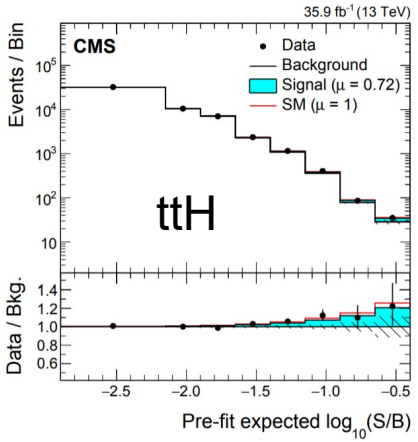
Combine bins with similar S/B

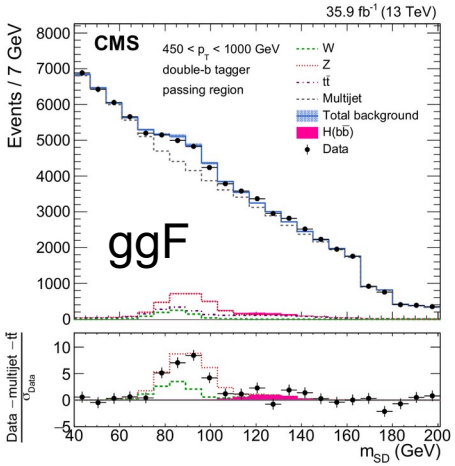


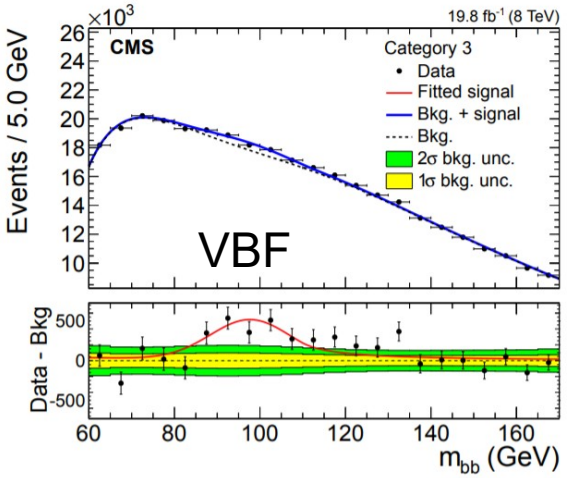
Combination with Run 1 and 2016/17 data

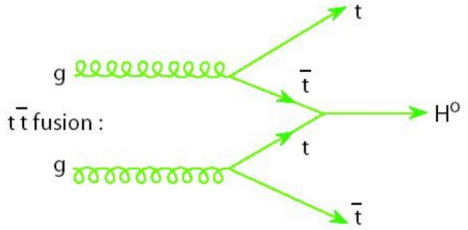
	Significance (σ)		
Data set	Expected	Observed	Signal strength
2017			
0-lepton	1.9	1.3	0.73 ± 0.65
1-lepton	1.8	2.6	1.32 ± 0.55
2-lepton	1.9	1.9	1.05 ± 0.59
Combined	3.1	3.3	1.08 ± 0.34
Run 2	4.2	4.4	1.06 ± 0.26
Run 1 + Run 2	4.9	4.8	1.01 ± 0.23

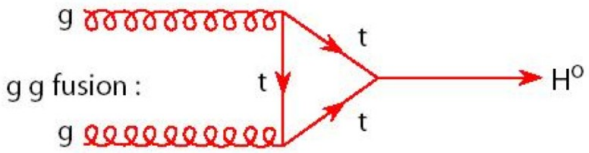
Combination with other Higgs production channels (where H→ bb)

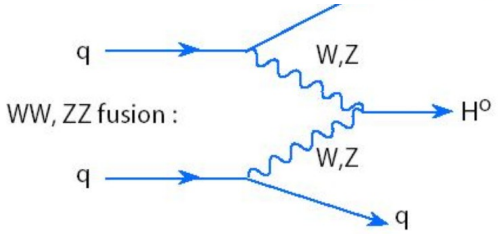


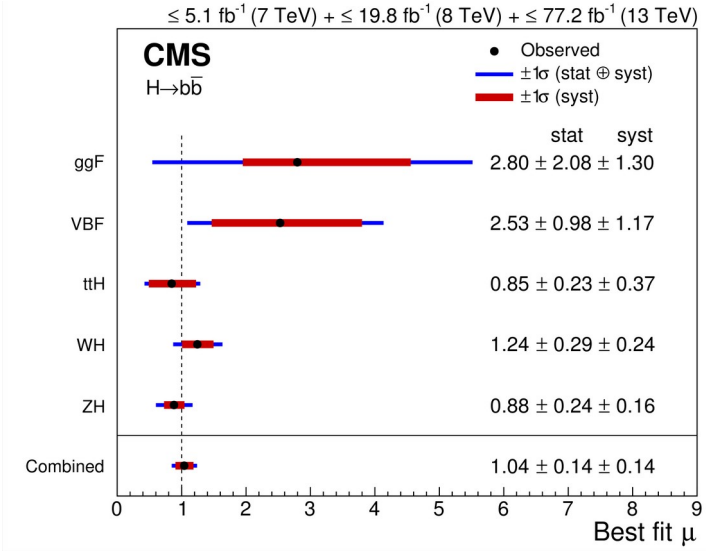










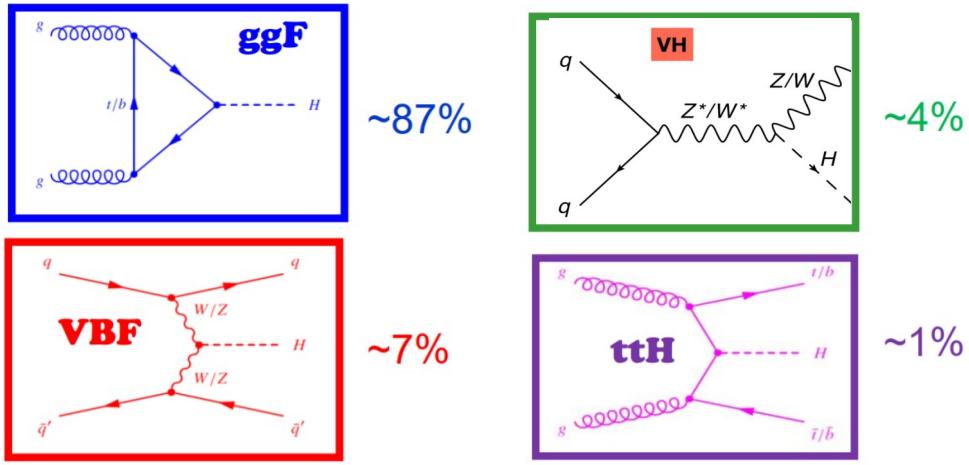


Results

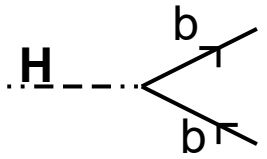
exp. (obs.) sig. = 5.5σ (5.6σ)
μ = 1.04 ± 0.14(stat) ± 0.14(sys)

Combination with other Higgs production channels (where H→ bb)

Higgs production modes



H→ bb decay mode

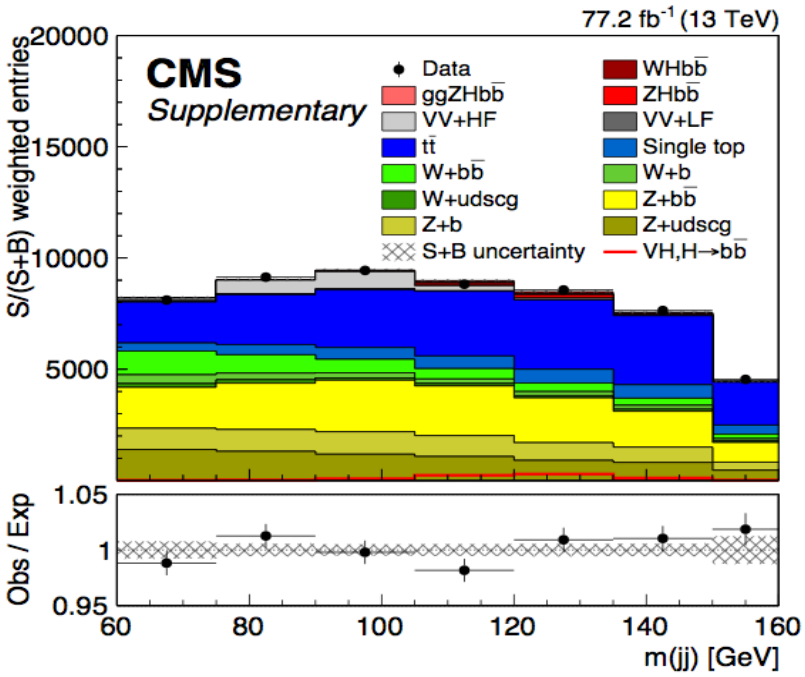


Results

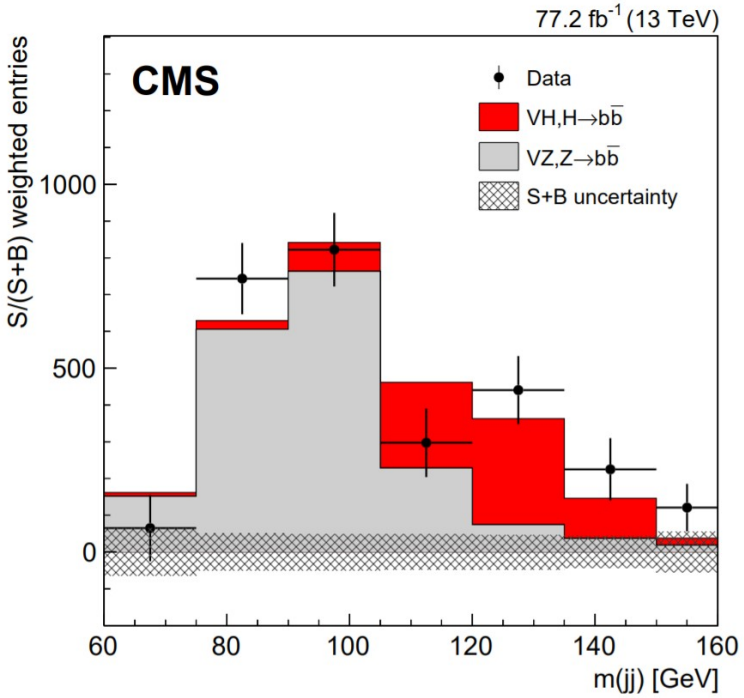
exp. (obs.) sig. = 5.5σ (5.6σ)
μ = 1.04 ± 0.14(stat) ± 0.14(sys)

mjj cross-check analysis

- Fit mjj distribution in 4 different bins of DNN score for SR.
- Same CR used in the fit.
- Combine SR post-fit mjj distribution of all channels by weighting events with S/(S+B).
- Sensitivity little lower than for fit with DNN score.

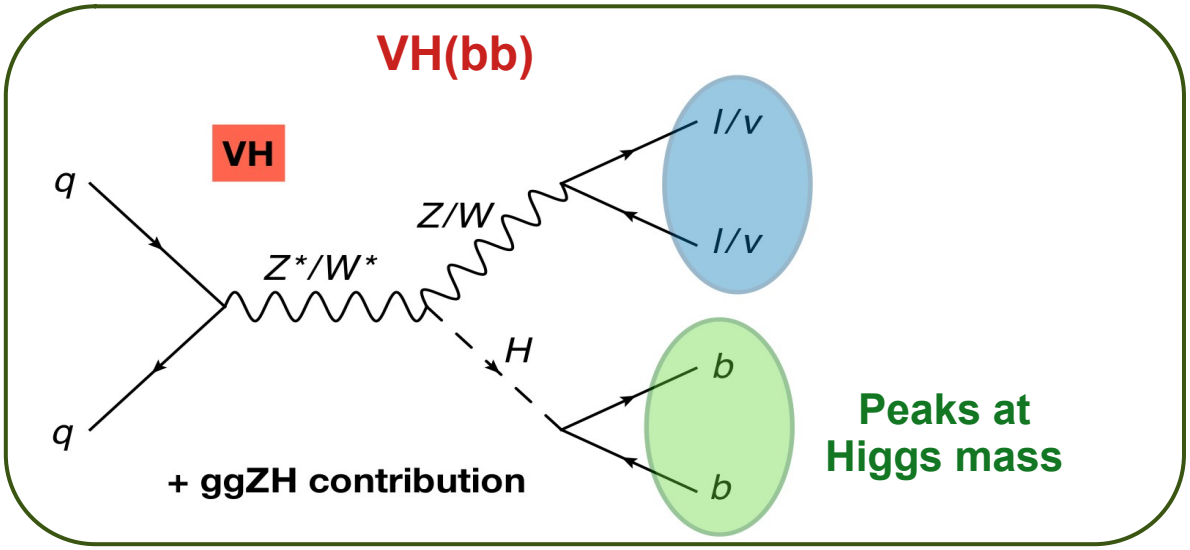


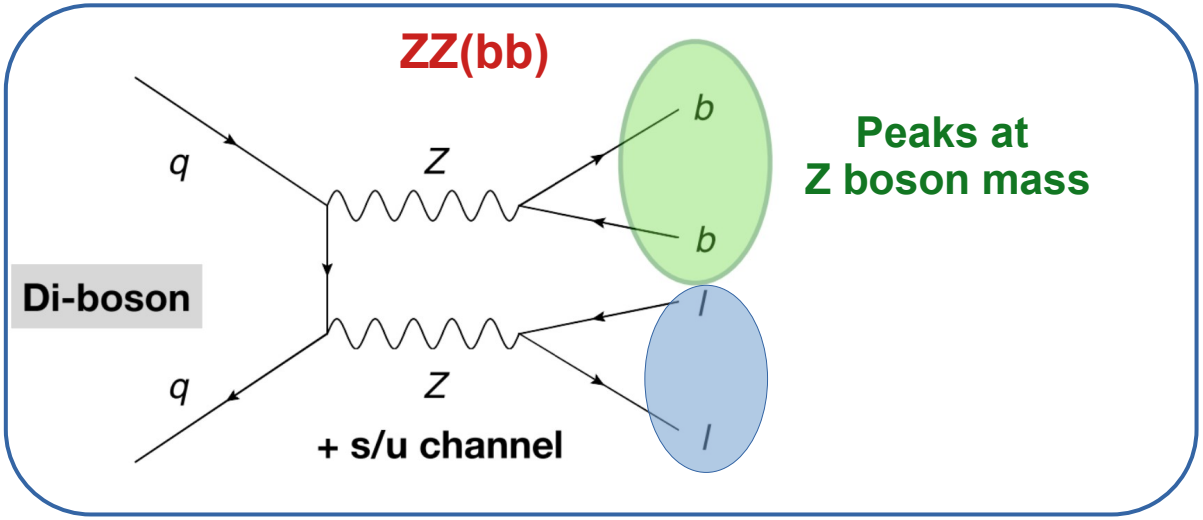
Bkg subtracted

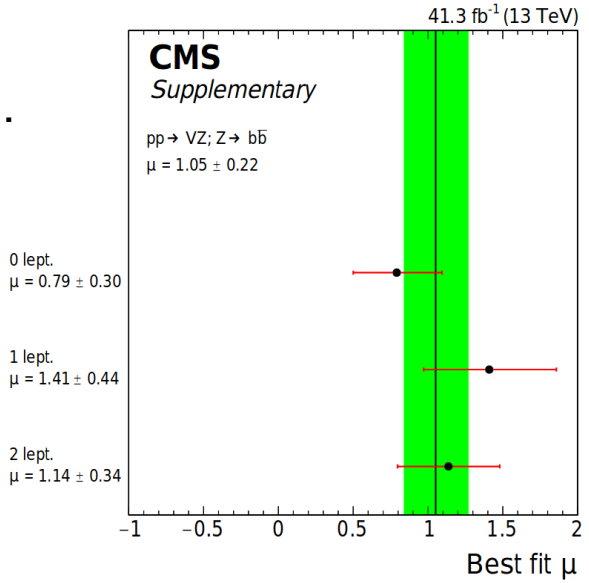


VZ(bb) cross-check analysis

- Take VZ(bb) as signal instead of VH(bb).
- Same final state, similar kinematics but different dijet invariant mass.







Result:
μ = 1.05 ± 0.22
obs. (exp.) **sig.** : **5.2 σ** (5.0 σ)

Overlap events

- Events passing selection for both resolved and boosted topology
- They can be assigned to only one of the topology.
- 4 different overlap strategies tested.

boosted \ resolved	SR	CR	-
SR	b	b	b
CR	b	b	b
-	r	r	

overlap to boosted

boosted \ resolved	SR	CR	-
SR	r	b	b
CR	r	r	b
-	r	r	

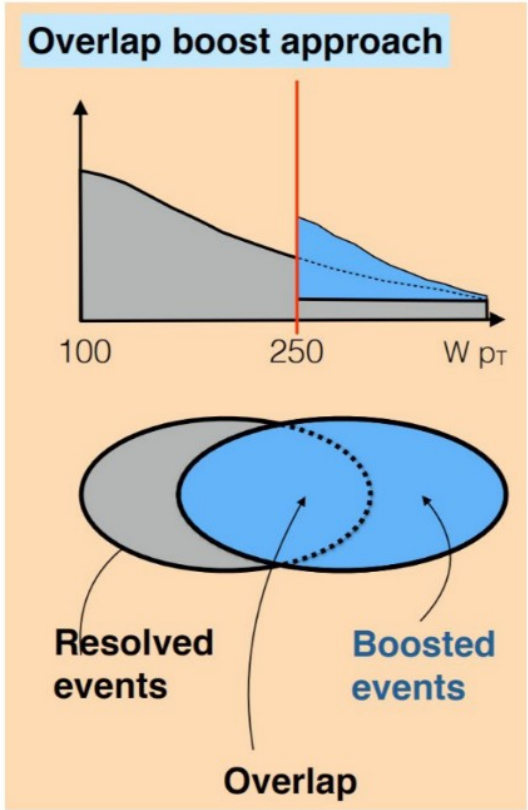
overlap to resolved, unless boosted SR+resolved CR

boosted \ resolved	SR	CR	-
SR	b	b	b
CR	r	b	b
-	r	r	

overlap to resolved, unless boosted SR+resolved CR

boosted \ resolved	SR	CR	-
SR	r	r	b
CR	r	r	b
-	r	r	

overlap to resolved



- best strategy for sensitivity:
  - resolved SR, boosted SR -> resolved
  - resolved CR, boosted CR -> resolved
  - resolved SR, boosted CR -> resolved
  - resolved CR, boosted SR -> boosted