Measurements of the Higgs boson to bottom quark coupling

(STXS measurements for VHbb resolved topology)

HIG-20-001

Krunal Gedia,
On behalf of the CMS Collaboration

Joint Annual Meeting of SPS and OPG

University of Innsbruck
30^{th} Aug - 3^{rd} Sept 2021

Standard model of elementary particles

Fermions/ Matter

mass ↓

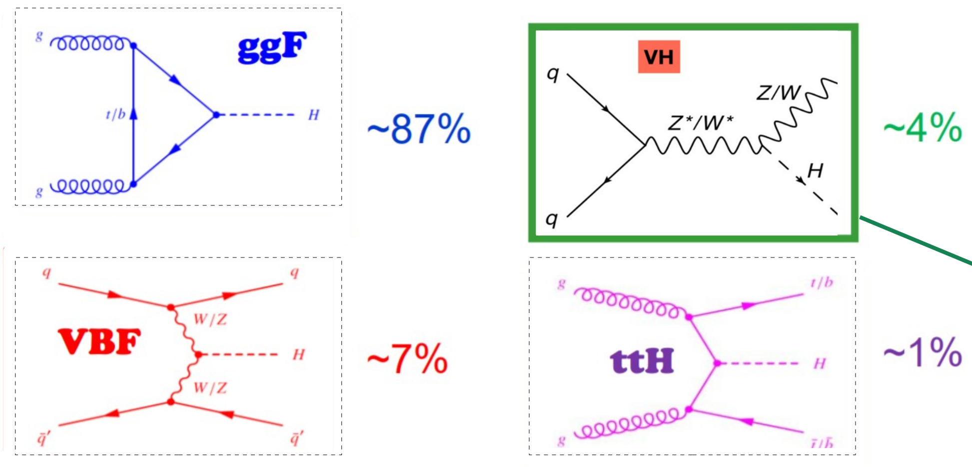
Gen	Quarks	Leptons
I	Up (u), Down (d)	Electron (e), e neutrino
II	Charm (c), Strange (s)	Muon (μ), μ neutrino
III	Top (t), **Bottom (b)**	Tau (τ), τ neutrino

Bosons/Force carriers

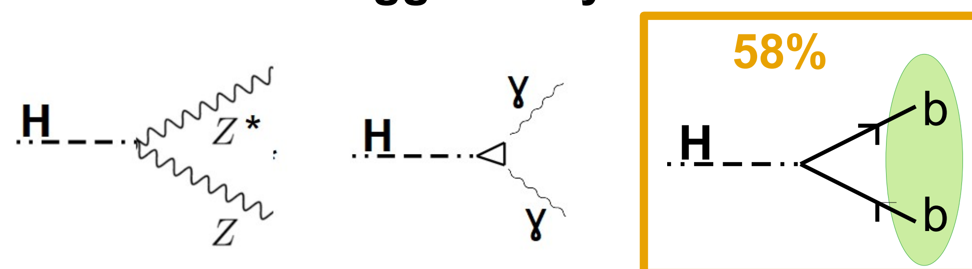
- Z boson
- Photon (γ)
- W boson
- Gluon (g)

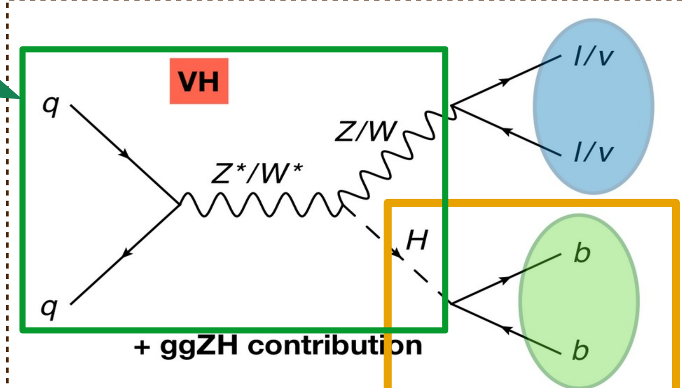
Higgs boson

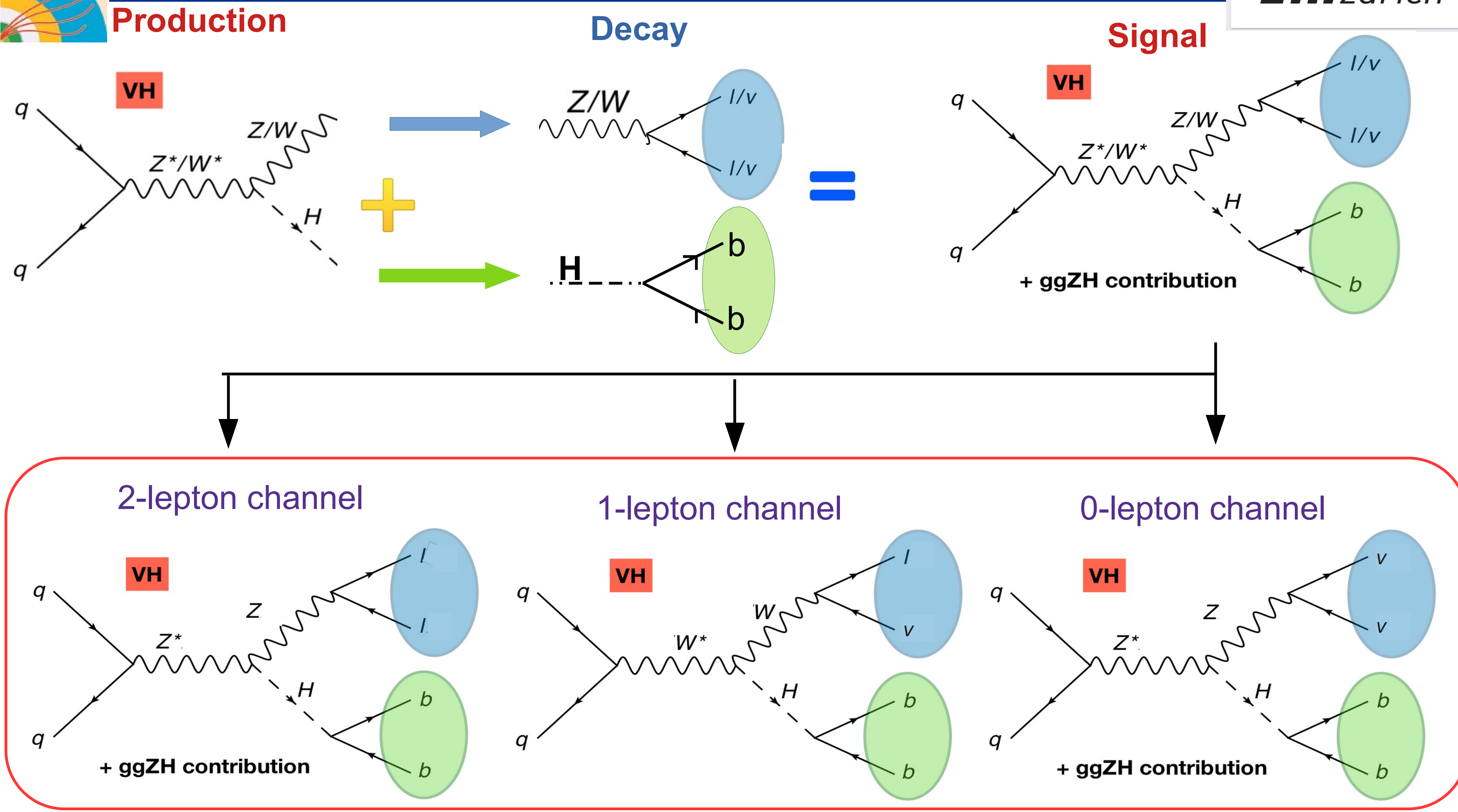
Higgs production modes



Higgs decay modes

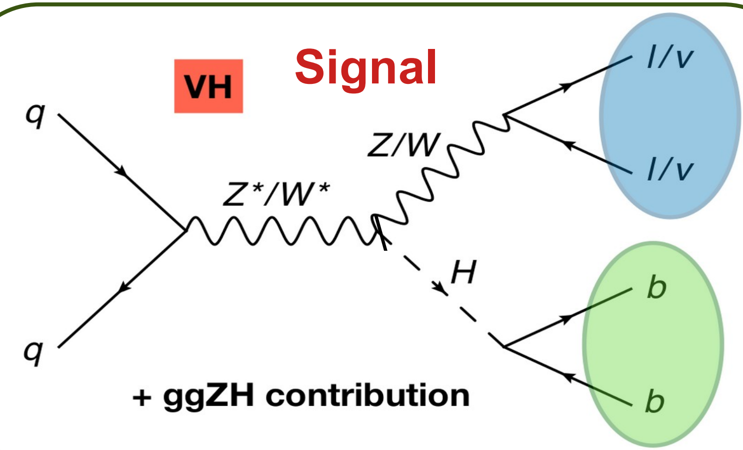






Production
Decay
Signal
VH
q
q
Z*/W*
Z/W
H
Z/W
l/ν
l/ν
H
b
b
VH
q
q
Z*/W*
Z/W
l/ν
l/ν
H
b
b
+ ggZH contribution
2-lepton channel
VH
q
q
Z*
Z
l
l
H
b
b
+ ggZH contribution
1-lepton channel
VH
q
q
W*
W
l
ν
H
b
b
0-lepton channel
VH
q
q
Z*
Z
ν
ν
H
b
b
+ ggZH contribution

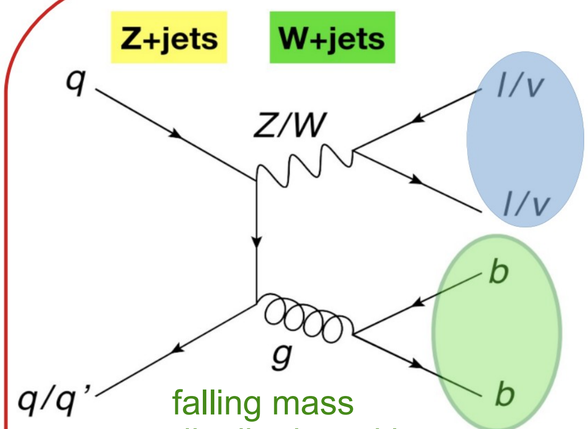
Signal



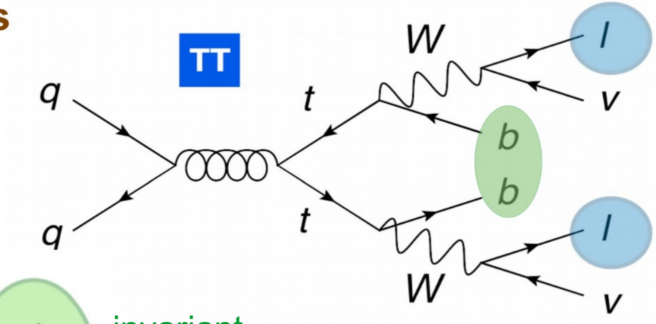
invariant mass has a peak at m(V)

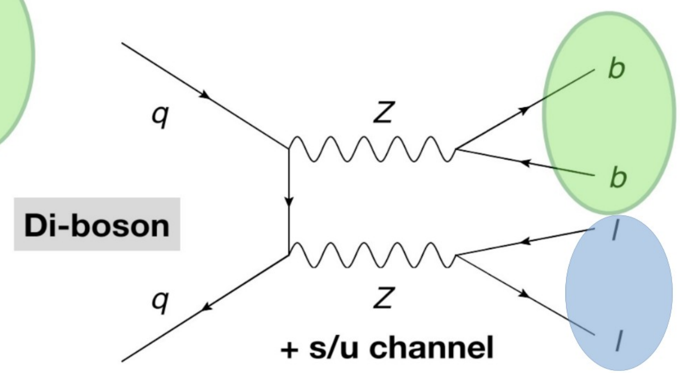
invariant mass has a peak at m(H) ~ 125 GeV

Main backgrounds



falling mass distribution without peak, little softer p_T

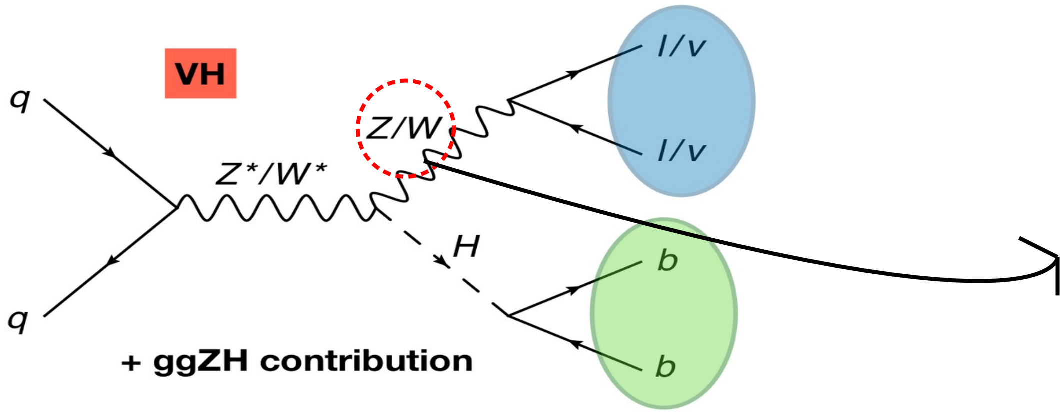


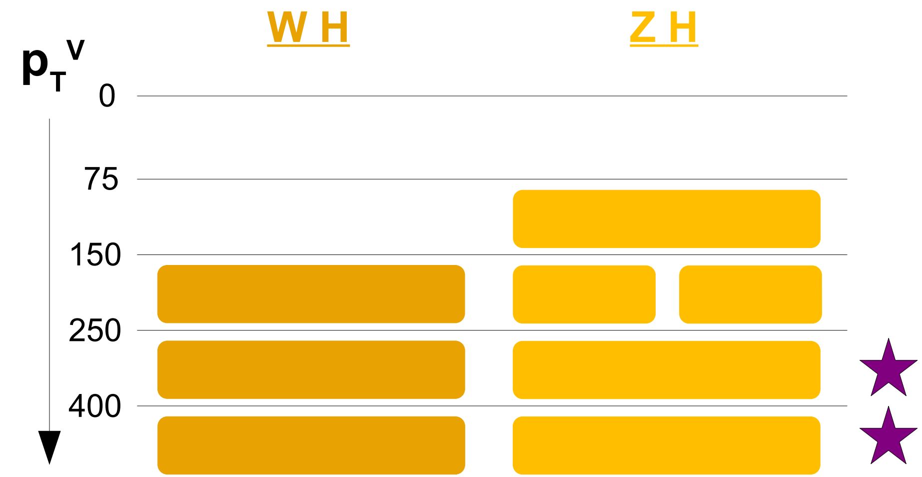


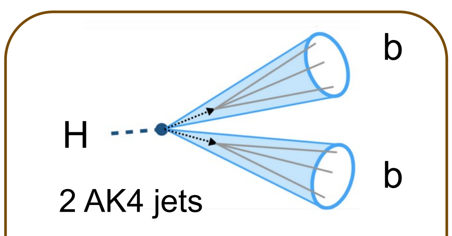
invariant mass has a peak at m(V)

invariant mass has flat distribution

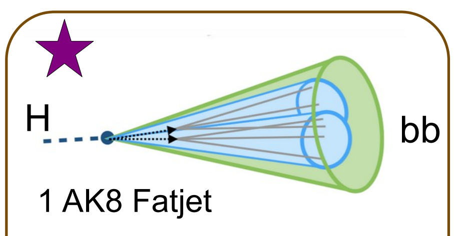
If W decays hadronically, more additional jets







Resolved topology



Boosted topology [link]

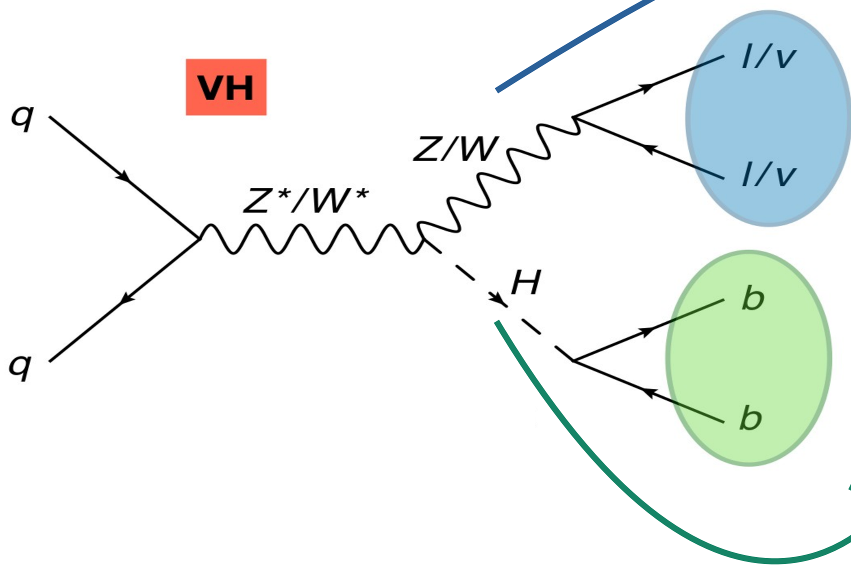
Advantages: Reduce theoretical unc., NP models study, allows complexity, combine ATLAS & CMS results,boosted

Analysis Strategy

- **Object reconstruction:** Reconstruct Higgs and vector boson.
- **Event Selection:** Derive signal enriched and background enriched regions by applying kinematic cuts.
- **Likelihood fit:** Choose some discriminatory variable in each of these regions and perform extended likelihood fit.

ETH zürich

Signal

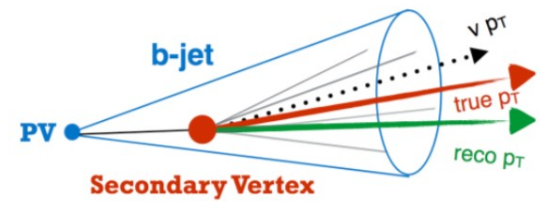


Vector boson: Isolated leptons &/or MET

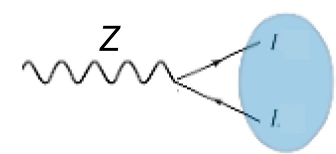
Higgs boson:
Two jets with highest b-tagging score

Improving jet mass resolution → dijet mass resolution

- DNN-based bjet regression
  - Energy correction due to escaping neutrino



- Kinematic fit (2 lepton channel)
  - Constraints m(ll) = m(Z) and pT(total) = 0.
  - Get constraints on jet resolution



Analysis Strategy

- **Object reconstruction:** Reconstruct Higgs and vector boson.
- **Event Selection:** Derive signal enriched and background enriched regions by applying kinematic cuts.
- **Likelihood fit:** Choose some discriminatory variable in each of these regions and perform extended likelihood fit.

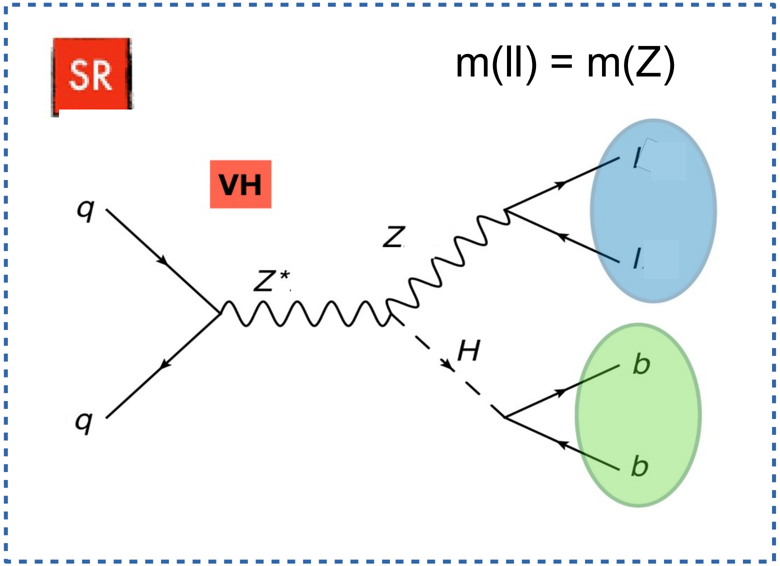
Event Selection

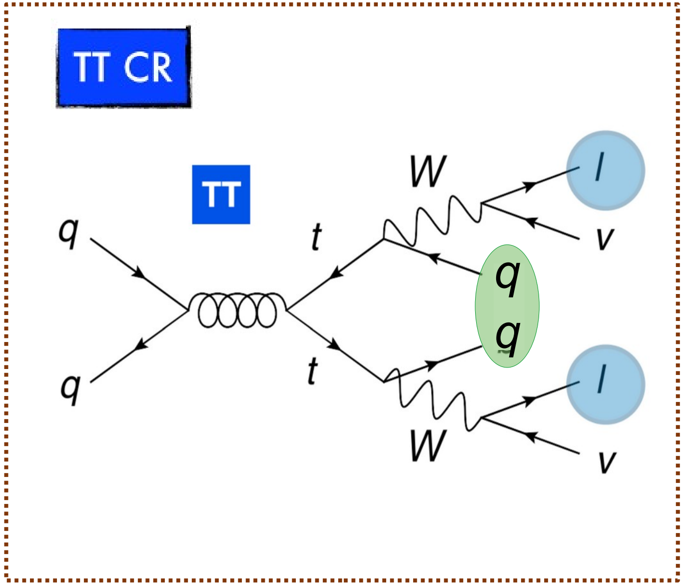
Signal region (SR)

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

Control region (CR)

- Enriched in one of the dominant background.
- Constrain normalisation of background processes in combined fit.
- HF CR , TT CR , LF CR





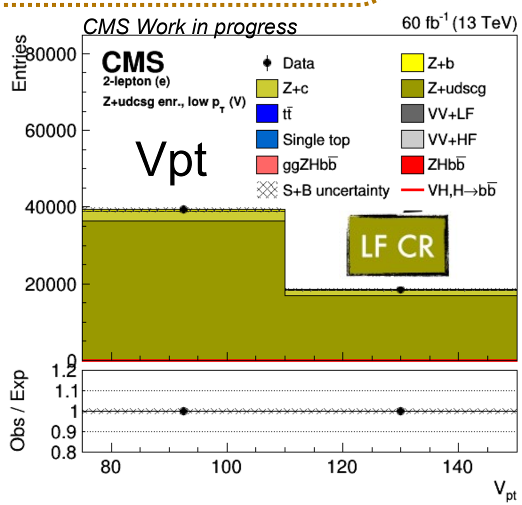
Analysis Strategy

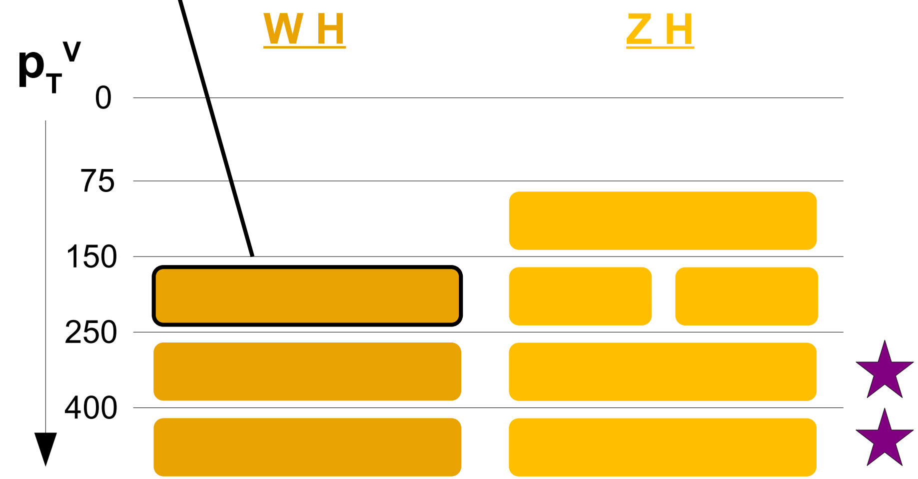
- **Object reconstruction:** Reconstruct Higgs and vector boson.
- **Event Selection:** Derive signal enriched and background enriched regions by applying kinematic cuts.
- **Likelihood fit:** Choose some discriminatory variable in each of these regions and perform extended likelihood fit.

ETH zürich

Resolved topology in STXS bins

- SR
  - DNN
- CR
  - TT CR — Vpt observable
  - LF CR — Vpt observable
  - (Pure CRs)
  - HF CR — Multi-class DNN

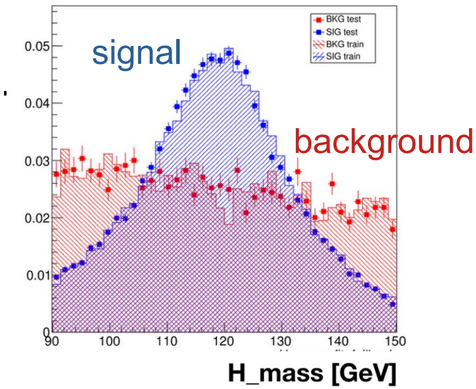




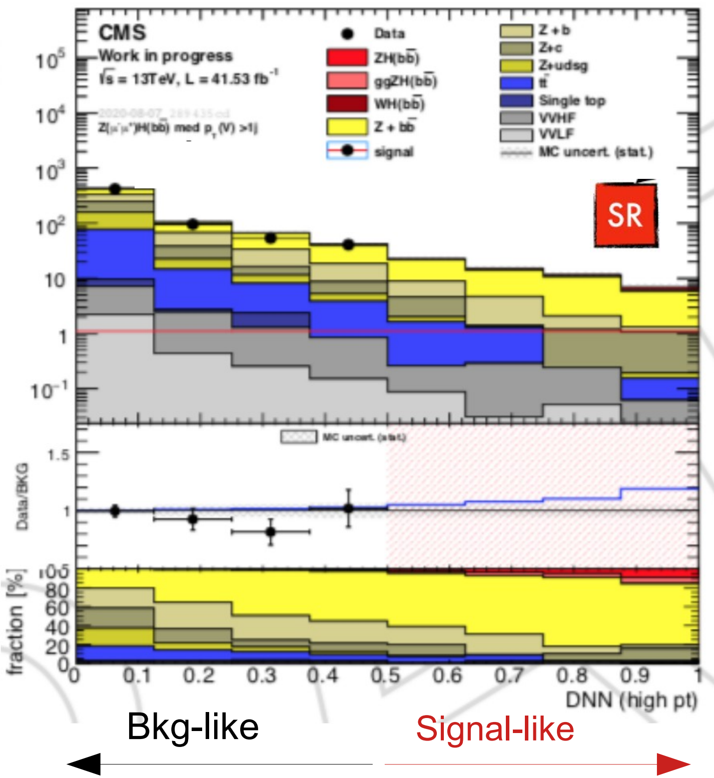
★ Boosted topology [link]

ETH zürich

- Channel dependent 15-27 high-level input features whose data/MC is verified in CR.
- Eg: Dijet invariant mass, btag score of jets



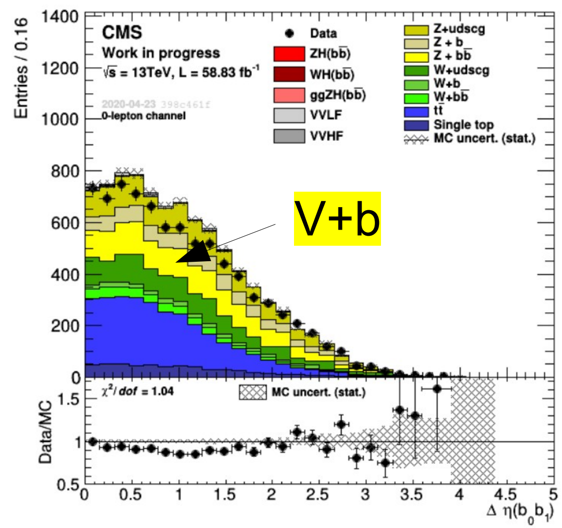
S/B DNN

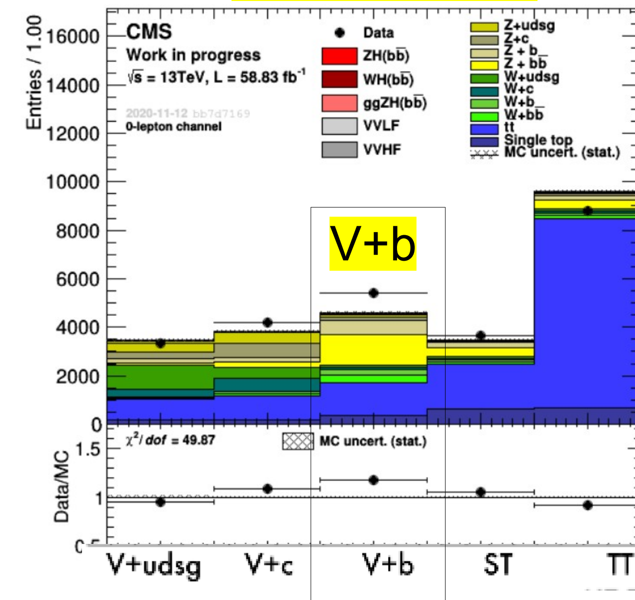


HF CR

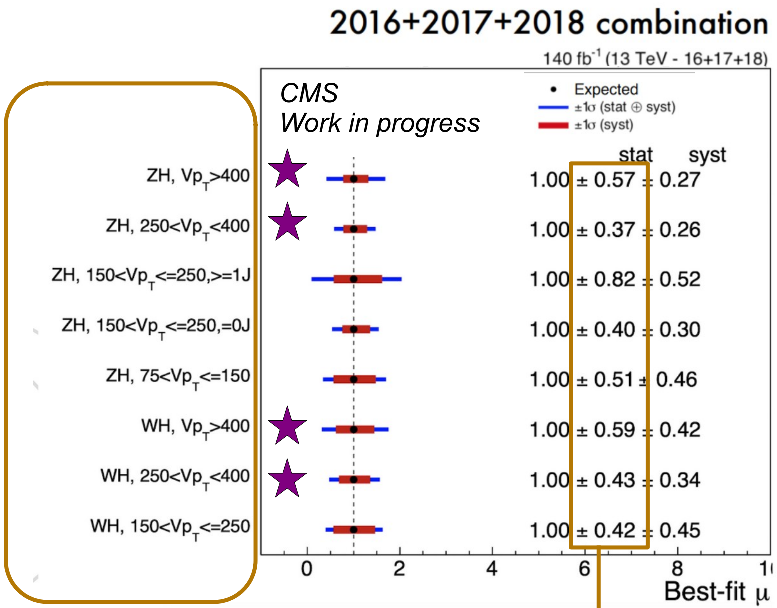
observable

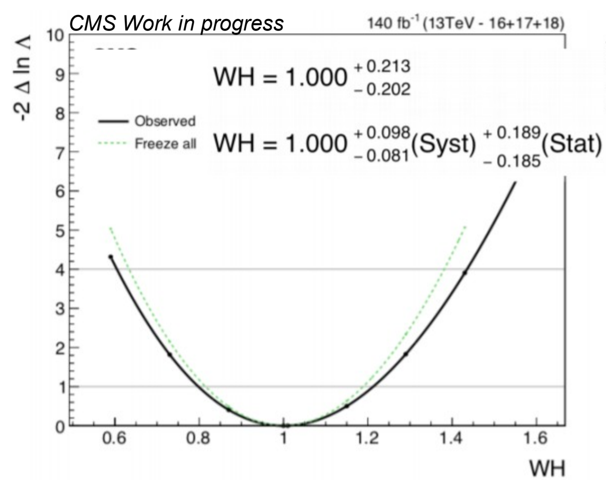
Optimized observable
Multi-class DNN

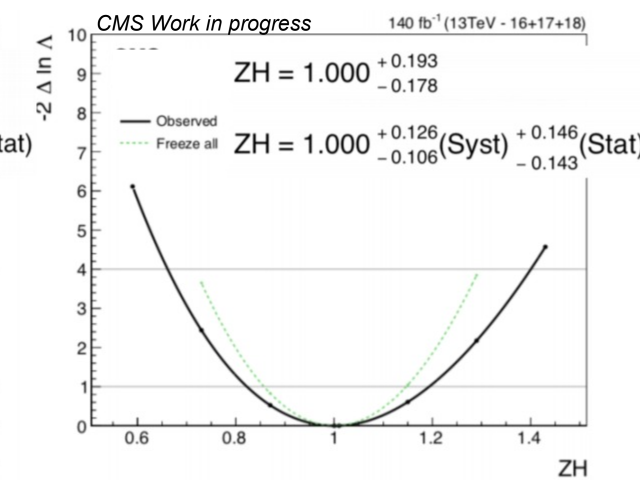




Expected VH results (resolved + boosted)







Uncertainties in STXS bins are statistically dominated!

★ Resolved+boosted topology

Summary of VHbb measurements in CMS

Inclusive (completed)

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

STXS (ongoing)

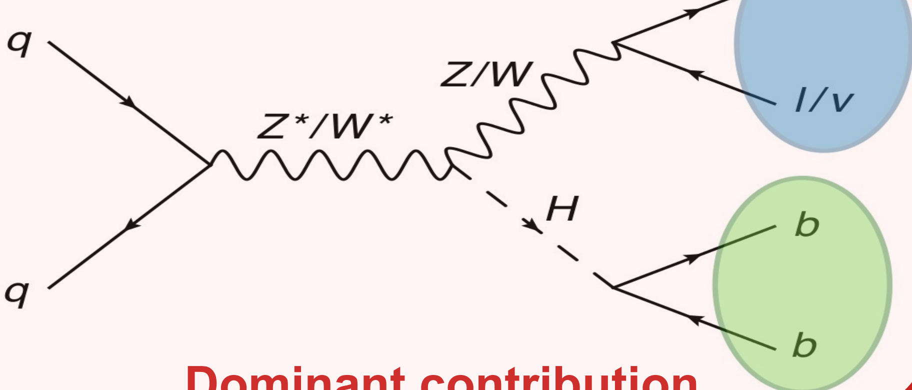
- Goal: Measurements in bins of $\mathbf{V_{pT}}$ and $\mathbf{N_{addjets}}$
- Increased sensitivity reach: Addition of boosted topology
- Full Run 2 data included (2016+2017+2018).
- Status: Analysis in the final Higgs group CMS internal review, expected to be released shortly

Anomalous coupling + Differential   Status: Just started internally in CMS

Back - Up

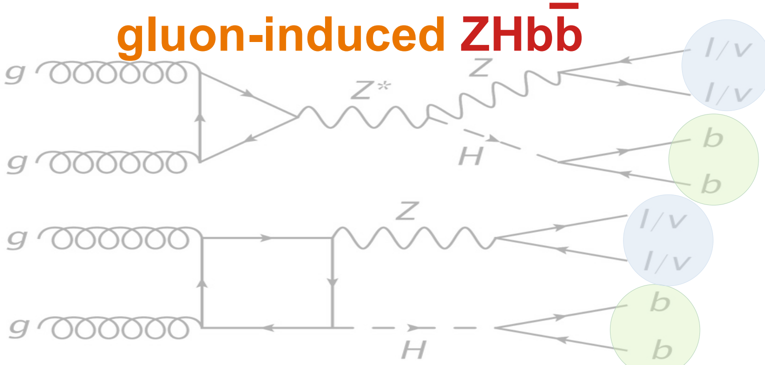
Signal: "Higgs-Strahlung"
Higgs produced with an associated vector boson

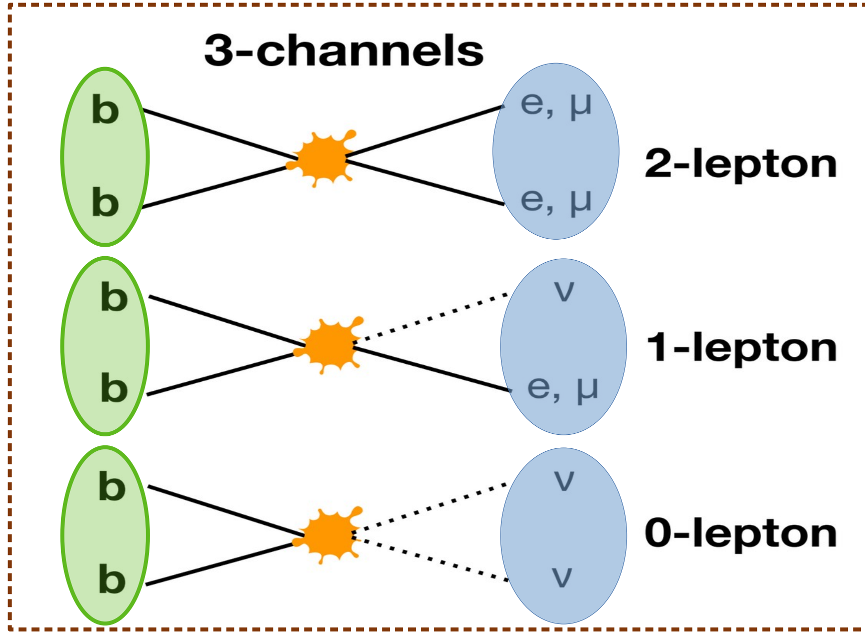
quark-induced ZHb$\bar{b}$ & WHb$\bar{b}$



Dominant contribution

gluon-induced ZHb$\bar{b}$

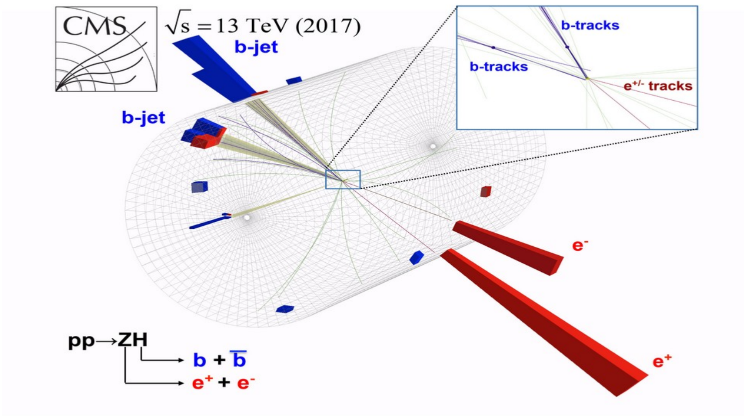


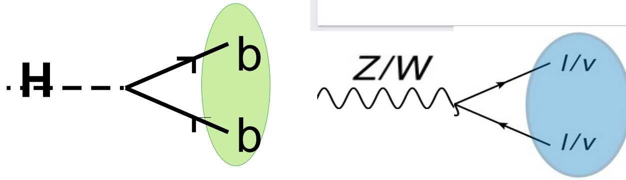


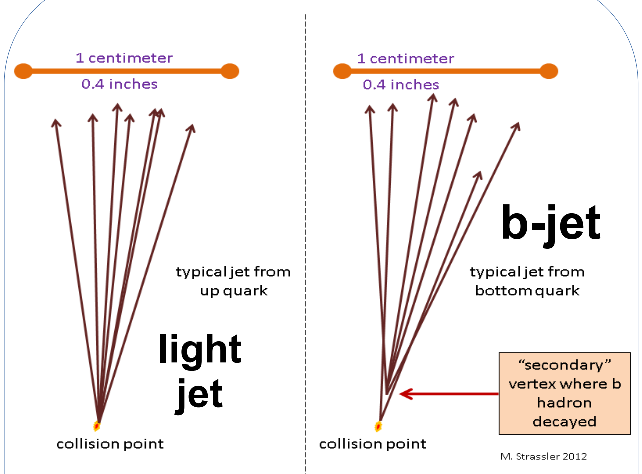
Why VHbb to study H→ b$\bar{b}$ coupling ?

- boost of the V-boson → QCD/V+Jets background
- Leptonic V decay → Trigger
- Large MET → Trigger

CMS detector

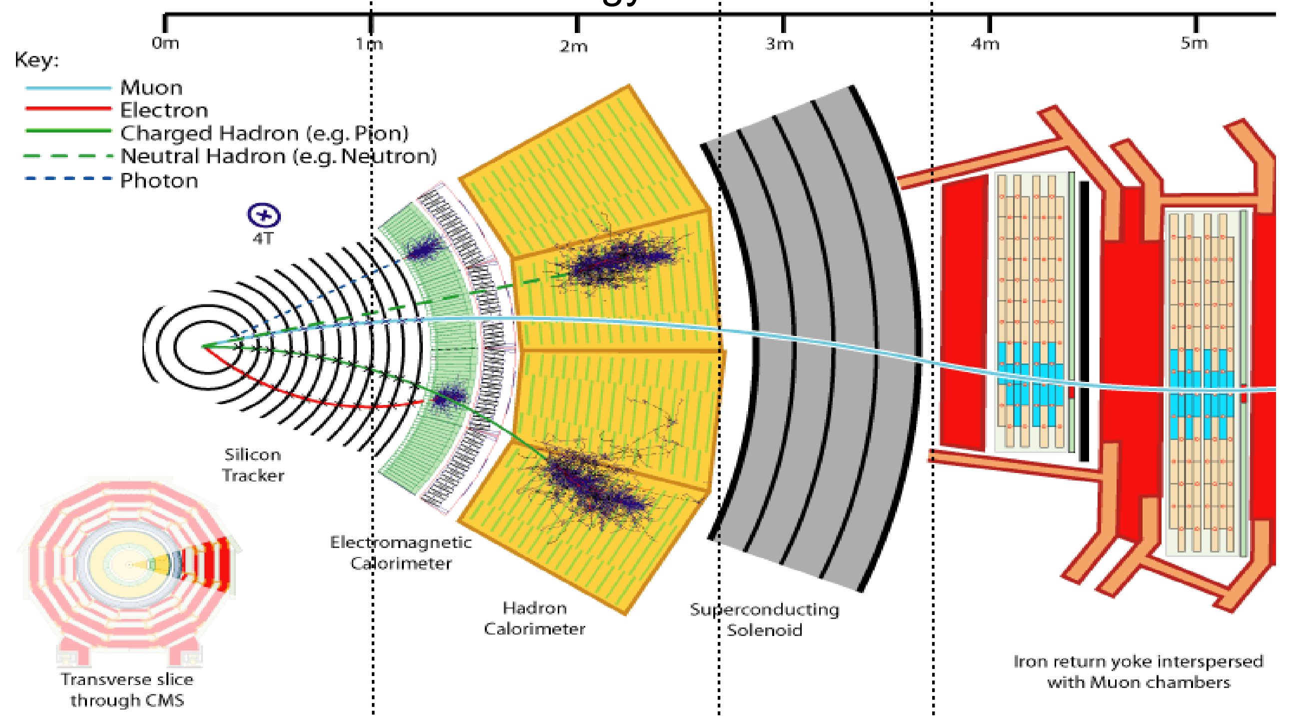






Identification of jet flavor (b-tagging)

Tracker	Calorimeter	Magnet	Muon chamber
Trajectory reconstruction	Measurement of Jet energy		Muon reconstruction



CMS-HIG-18-016

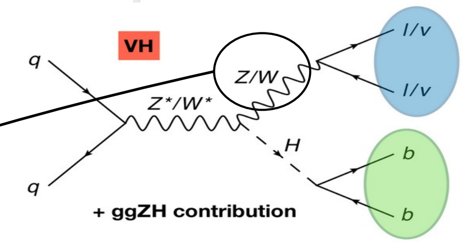
CERN-EP-2018-223
2018/09/20

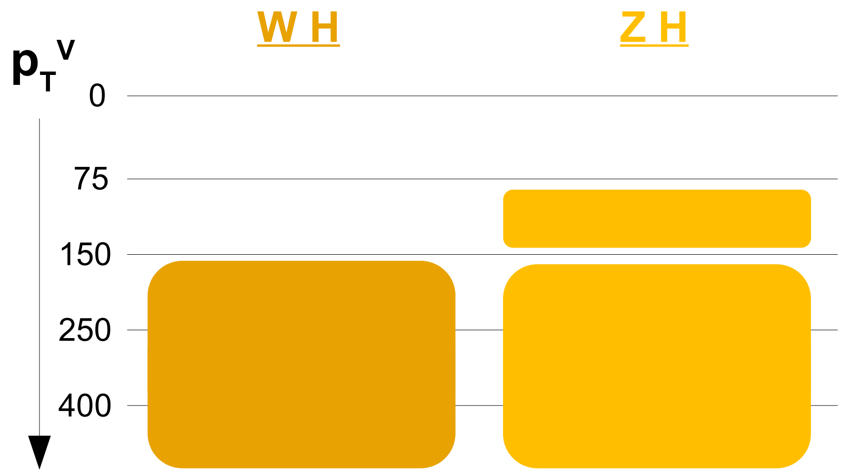
2018

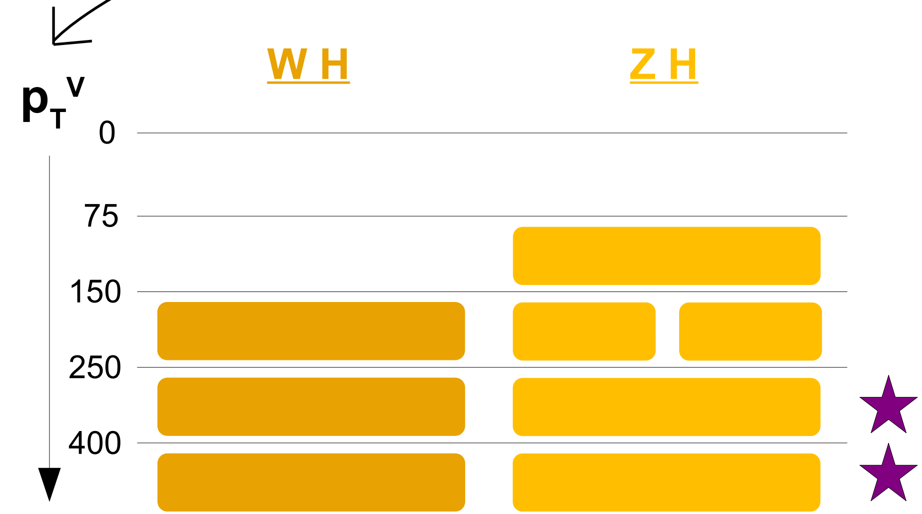
Observation of Higgs boson decay to bottom quarks

The CMS Collaboration*

[link]



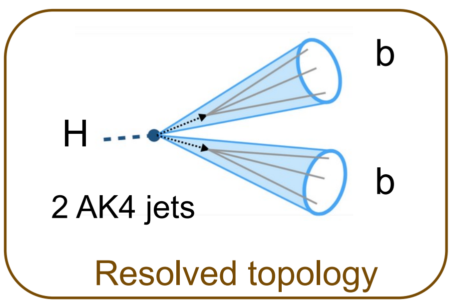


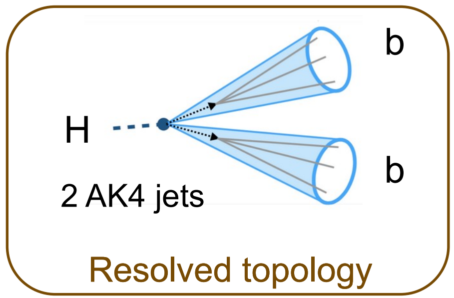


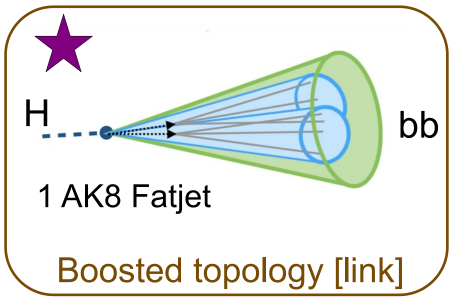
Signal strength $\mu = \sigma / \sigma_{SM}$

1.24 +/- 0.38

0.88 +/- 0.29



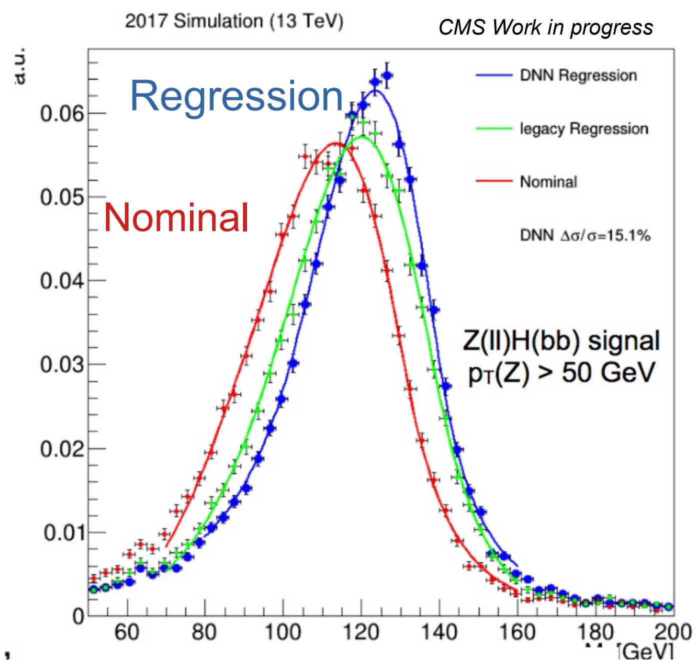




Advantages: Reduce theoretical unc., NP models study, allows complexity, combine ATLAS & CMS results,boosted

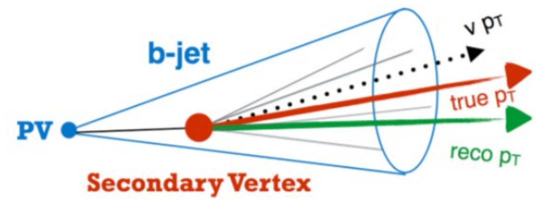
DNN-based b-jet regression [link]:

- Energy correction due to escaping neutrino from semi-leptonic decay, calibration mis-match.



Significant improvement in M(bb̄) resolution with respect to 2016 analysis.

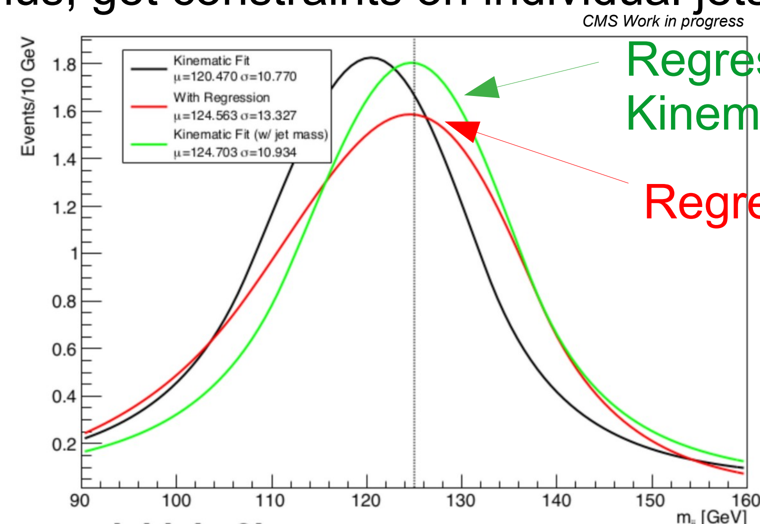
σ/μ (%)	16.4	15.2	12.3
improv.(%)		7.3	22.6



Kinematic fit: (only in two lepton channel)

- leptons have better momentum resolution than jets
- no intrinsic MET

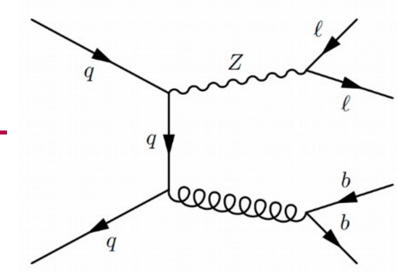
- Fit leptons and jets with uncertainties under constraints m(ll) = m(Z) and pT(total) = 0.
- Thus, get constraints on individual jets resolution.



med+high, 0j

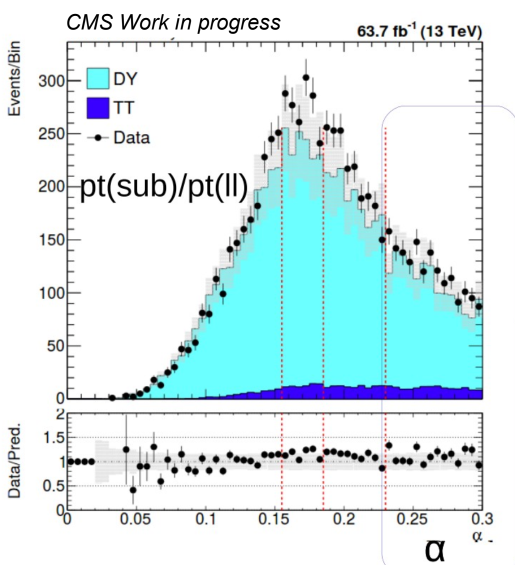
FSR recovery: (all channels)

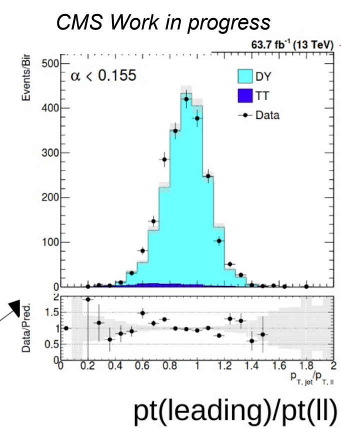
- Add 4-vectors of FSR jets in defined cone around b-jets.

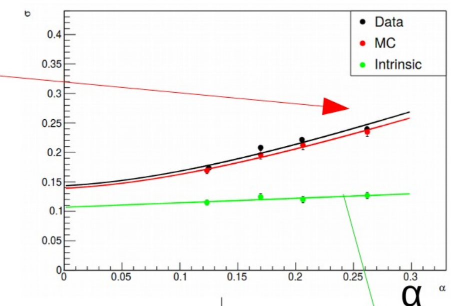


Dedicated smearing

- Good detector resolution of leptons allows us to use Z(ll)bb process
- The jet resolution can be measured by assuming it is balanced against the Z in the transverse plane.
- Apply tight b-tag cut on the leading jet and fix scale to 1.0
- So as α = pt(sub)/pt(ll) → 0 , we get one jet process through which we can extract resolution of data and MC and thus get scaling/smearing factors.
- Scale unc. are correlated while the smearing unc. decorrelated for signal & bkg.







$$f(\alpha) = (m \times \alpha) \oplus c \times (1 + c_k \times \alpha)$$

2018

Scale: 1.0 +/- 0.019

Smear: 0.15 +/- 0.079

Smear: $\sqrt{c_{data}^2 - c_{MC}^2}$

2

Event selection

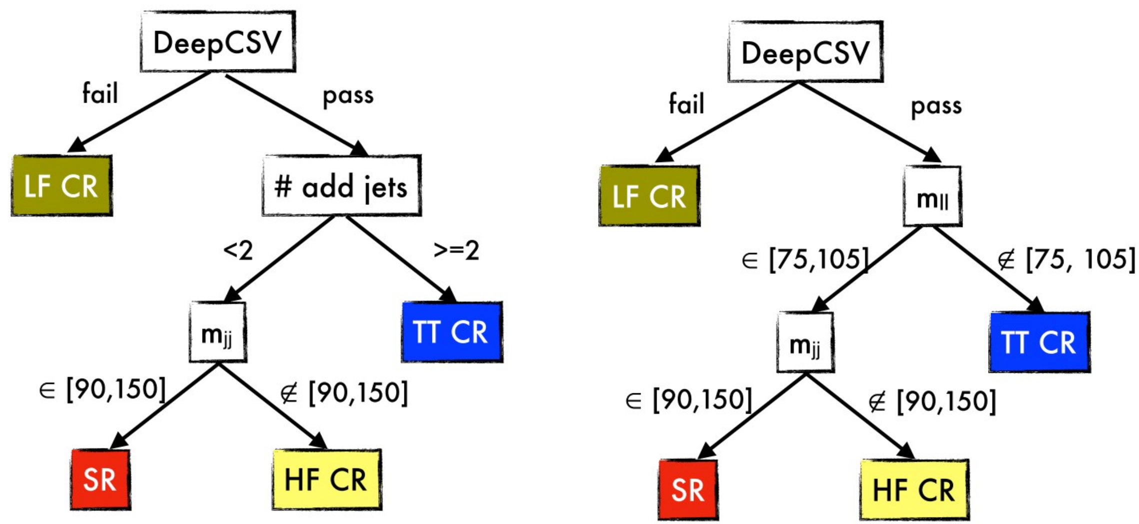
SR

HF CR TT CR LF CR

Resolved topology

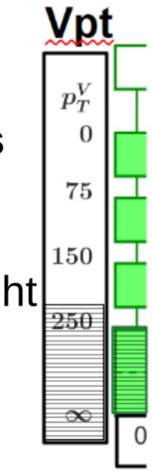
simplified strategy 0- & 1-lepton

simplified strategy 2-lepton



Treatement of overlap events explained in Christina's talk [link]

What we fit?



Resolved topology in STXS bins

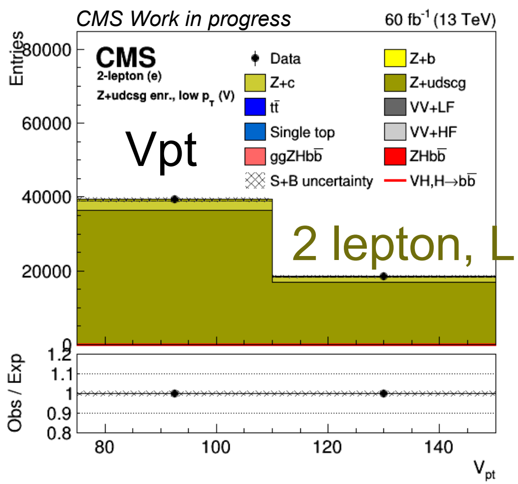
SR — DNN

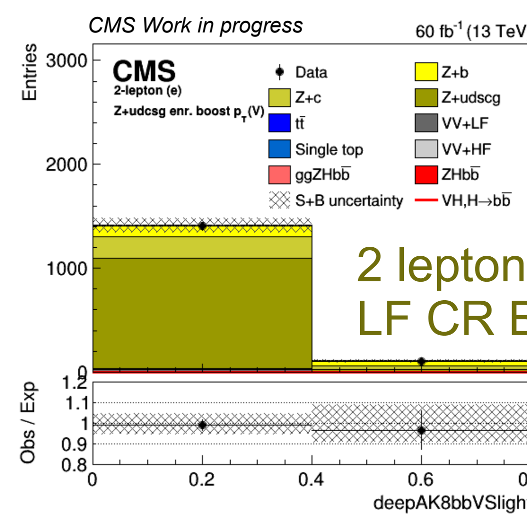
CR:
- TT CR — Vpt shape
- LF CR — Vpt shape
- HF CR — HF DNN (2 bin deepCSV in 2 lep)

Boosted topology in STXS bins

SR : BDT

LF CR, TT CR, HF CR : deepAK8bbVsLight

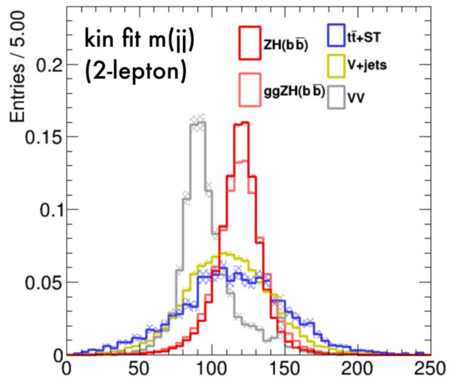




deepAK8bbVSlight

Multivariate variables

example input features (full list in AN)



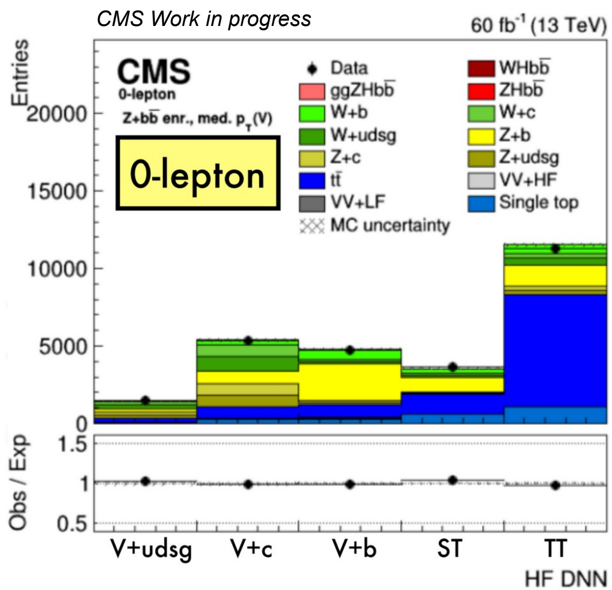
DNN/HF DNN:

- Channel dependent 15-27 high-level input features whose data/MC is verified in CR.

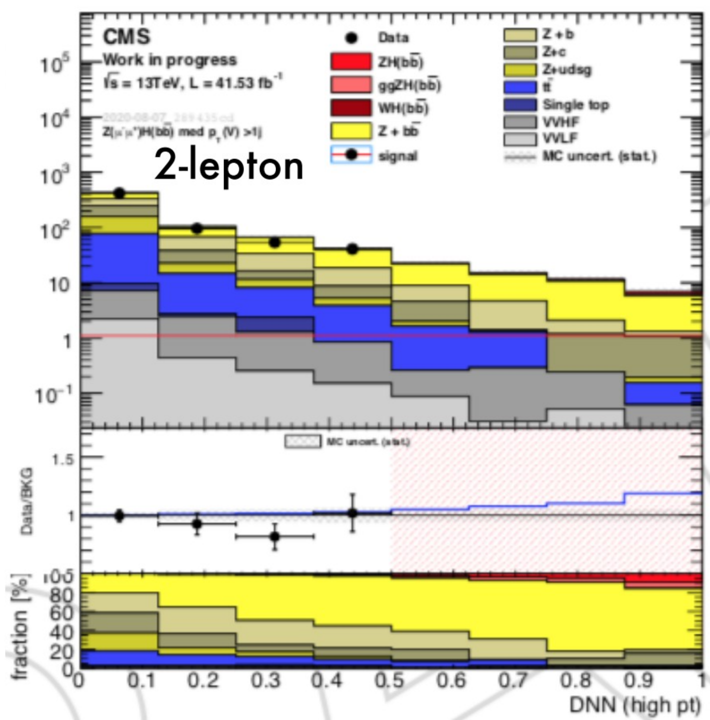
BDT:

- Uses FatJet kinematic variables + deepAK8.
- Overlap events have resolved features as well.

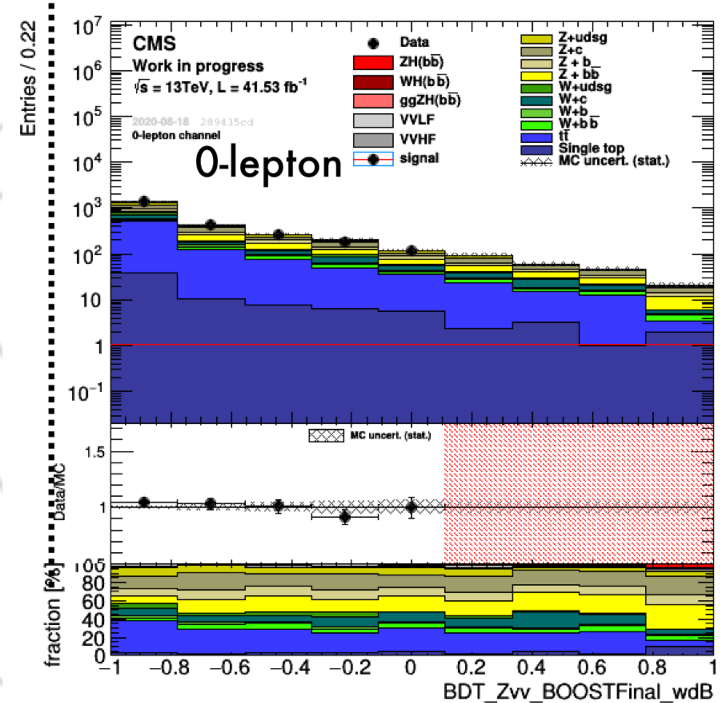
HF DNN



S/B DNN



S/B BDT



Background modelling

For ST, VV: 15% unc. on cross section.

For each of the TT, V+udsg, V+c, V+b processes in fit are obtained using

2 lepton channel (Zll)

1 lepton channel (Wlv)

0 lepton channel (Znn)

Boosted:
- Freefloat in-situ SF in each channel

Resolved + Boosted:
- Linear migration shape uncertainty at pt(V) bin boundaries at 150GeV and 250GeV.
- Constrained by fitting pt(V) templates in CR.
- Ensures continuty in pt(V) spectrum.

Zee Zmm

Wen Wmn

Znnn

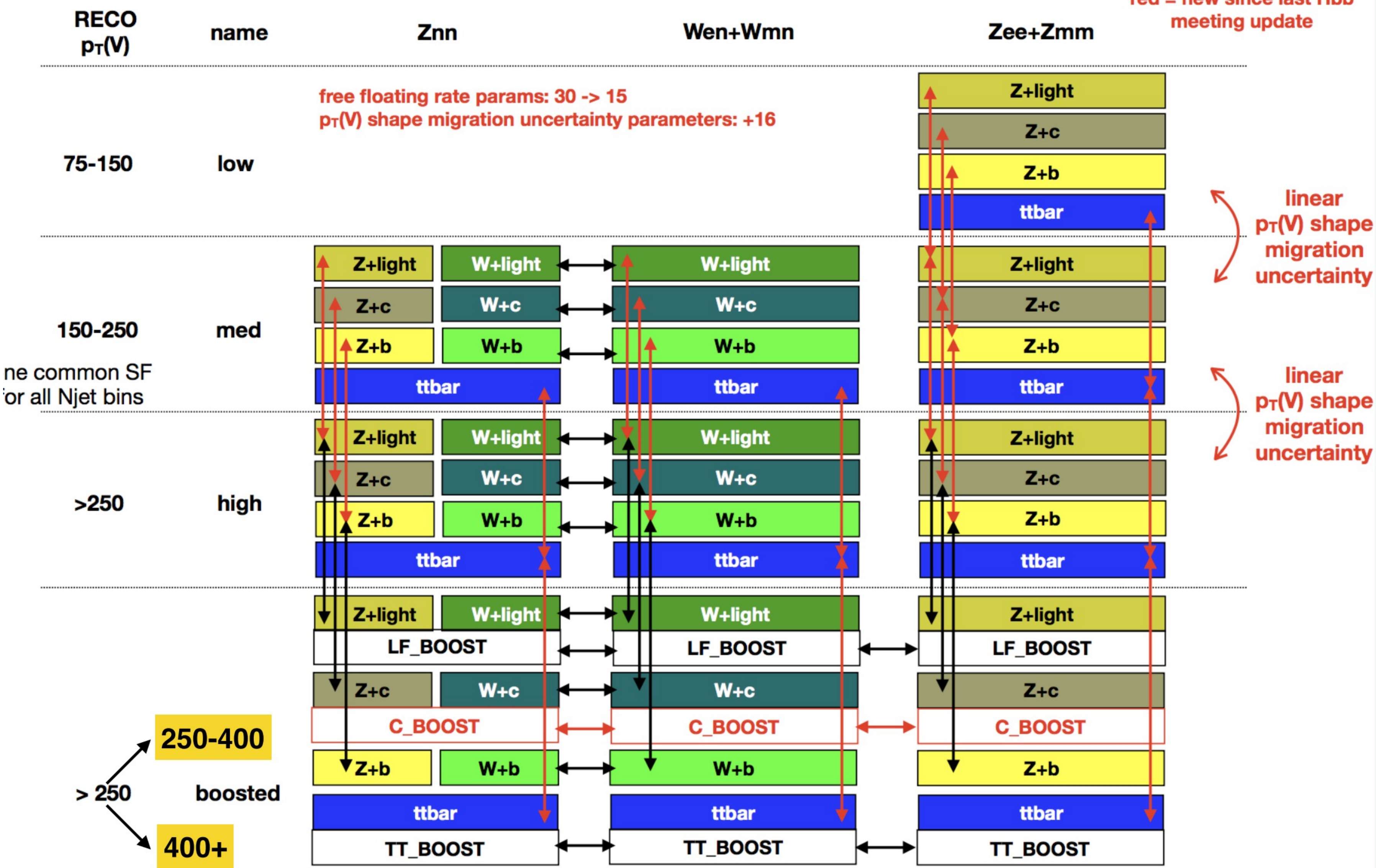
Resolved:
- Freefloat one rate parameter in each channel.

*Z+b is splitted in Z+1b, Z+2b for resolved topology.

Background process scale factors

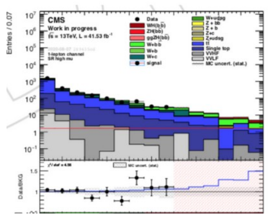
↕ = SFs fully correlated

red = new since last Hbb meeting update

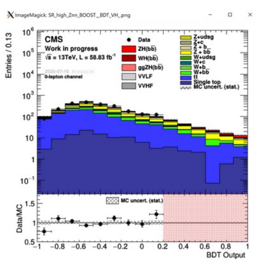


Simultaneous fit of SR and CR to obtain signal strength/significance

SR



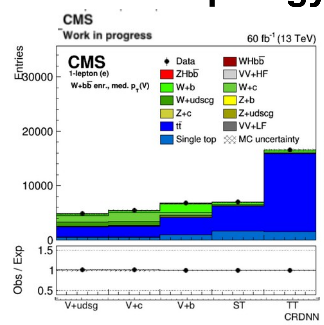
Resolved topology
S/B DNN classifier



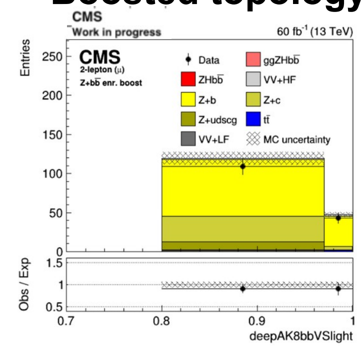
Boosted topology
S/B BDT classifier

CR

Resolved topology



Boosted topology



V+LF & TT CR: V_{pt}

V+HF : Multi-class DNN

FatJet tagger

+

~300 sources of systematic uncertainties

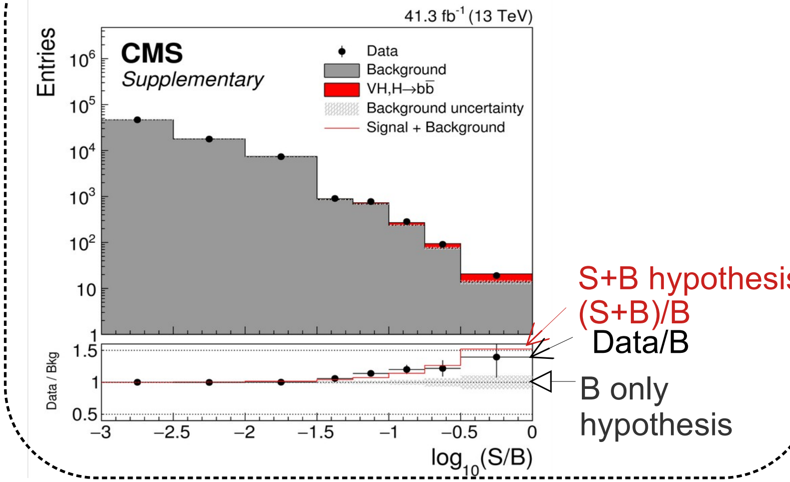
ETH zürich

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!

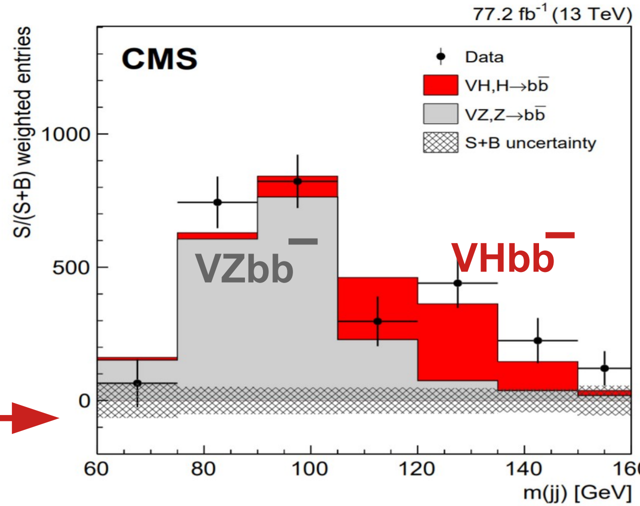


Cross check analysis:

m(jj) cross check analysis:
- Fit m(jj) distribution in SR instead of DNN score.

VZbb analysis:
- VZbb as signal instead of VHbb.

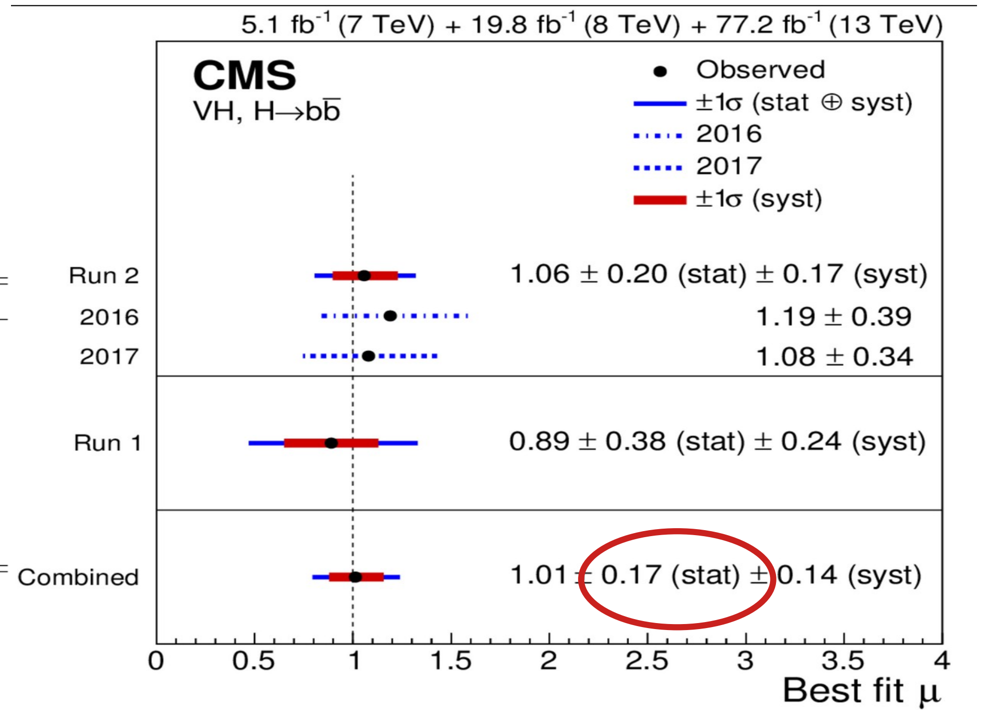
Dijet invariant mass in SR (bkg subtracted)



Results of simultaneous fit

SF of background normalization

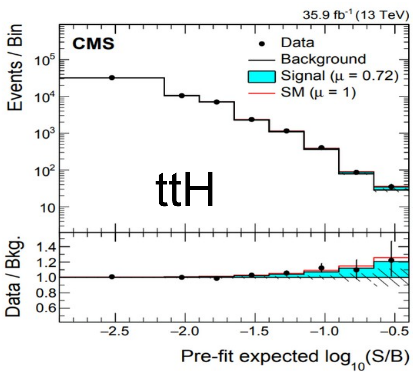
Process	Z(νν)H	W(ℓν)H	Z(ℓℓ)H low-p_T	Z(ℓℓ)H high-p_T
W + udscg	1.04 ± 0.07	1.04 ± 0.07	–	–
W + b	2.09 ± 0.16	2.09 ± 0.16	–	–
W + $b\bar{b}$	1.74 ± 0.21	1.74 ± 0.21	–	–
Z + udscg	0.95 ± 0.09	–	0.89 ± 0.06	0.81 ± 0.05
Z + b	1.02 ± 0.17	–	0.94 ± 0.12	1.17 ± 0.10
Z + $b\bar{b}$	1.20 ± 0.11	–	0.81 ± 0.07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

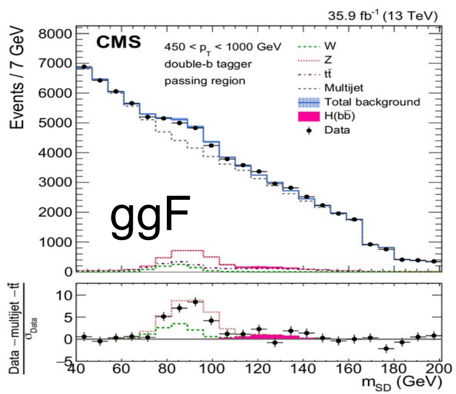


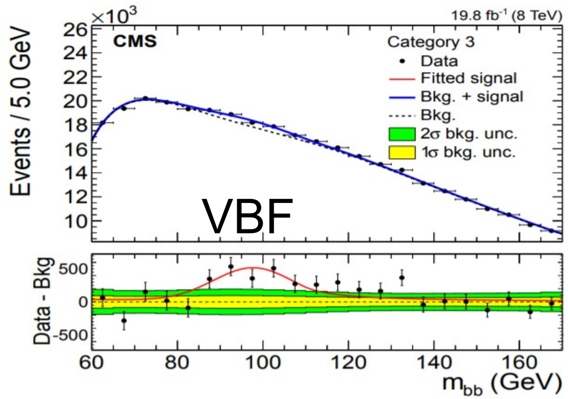
Results (2018)

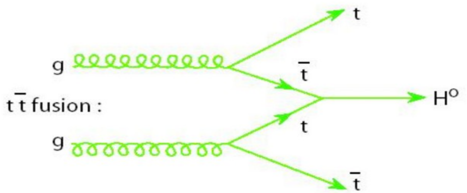
Uncertainty source	$\Delta\mu$	
Statistical	+0.26	−0.26
Normalization of backgrounds	+0.12	−0.12
Experimental	+0.16	−0.15
b-tagging efficiency and misid	+0.09	−0.08
V+jets modeling	+0.08	−0.07
Jet energy scale and resolution	+0.05	−0.05
Lepton identification	+0.02	−0.01
Luminosity	+0.03	−0.03
Other experimental 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

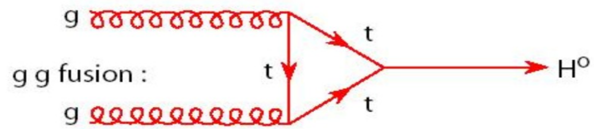
Combination with other Higgs production channels (where H→ bb)

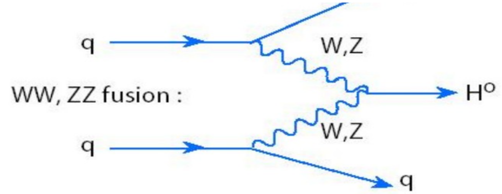


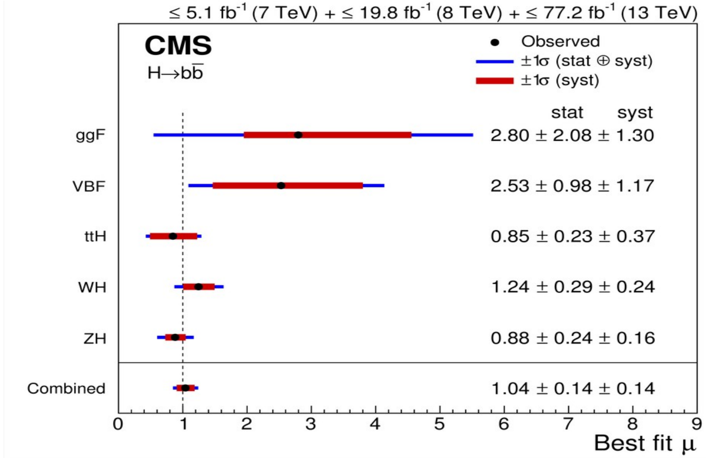












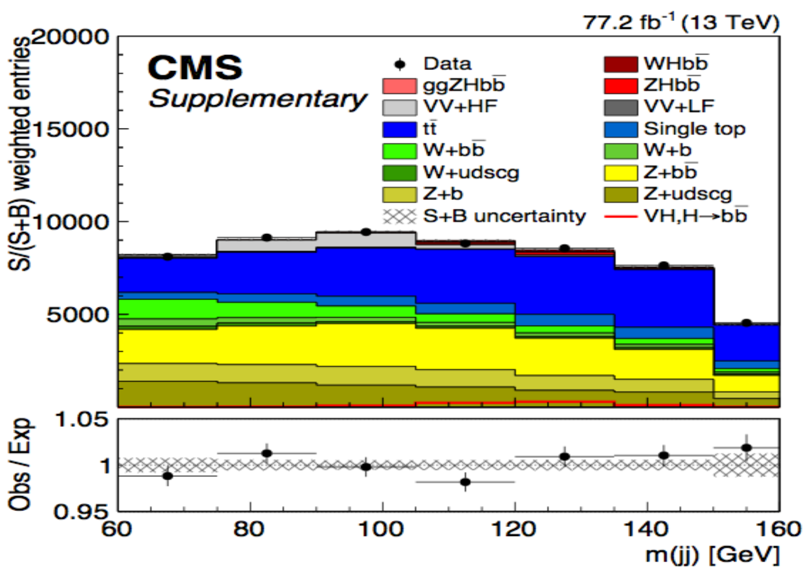
Results

exp. (obs.) sig. = 5.5σ (5.6σ)

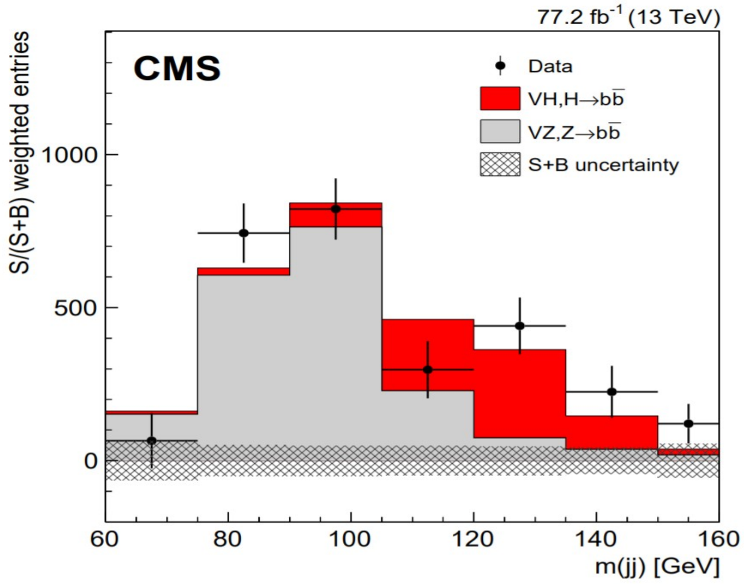
μ = 1.04 ± 0.14(stat) ± 0.14(sys)

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.

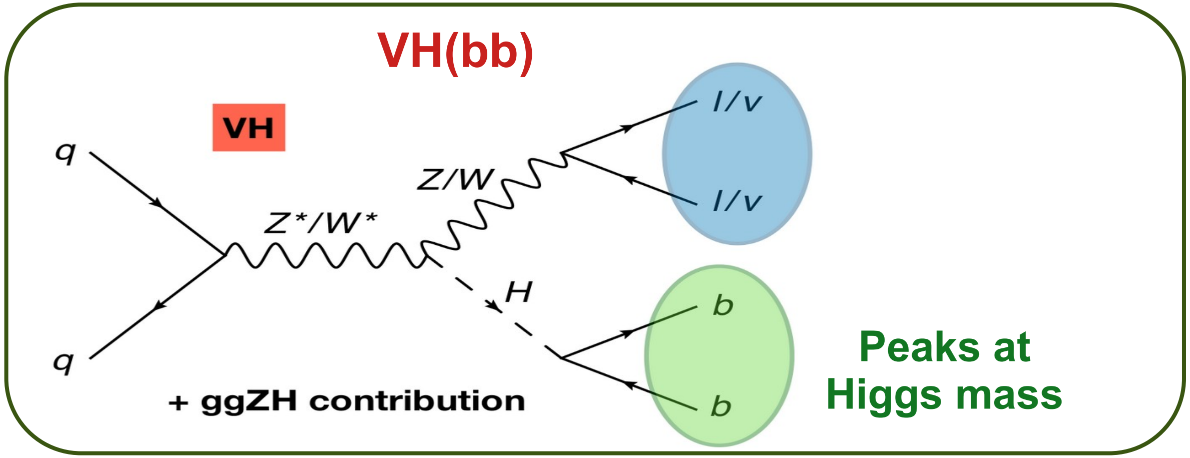


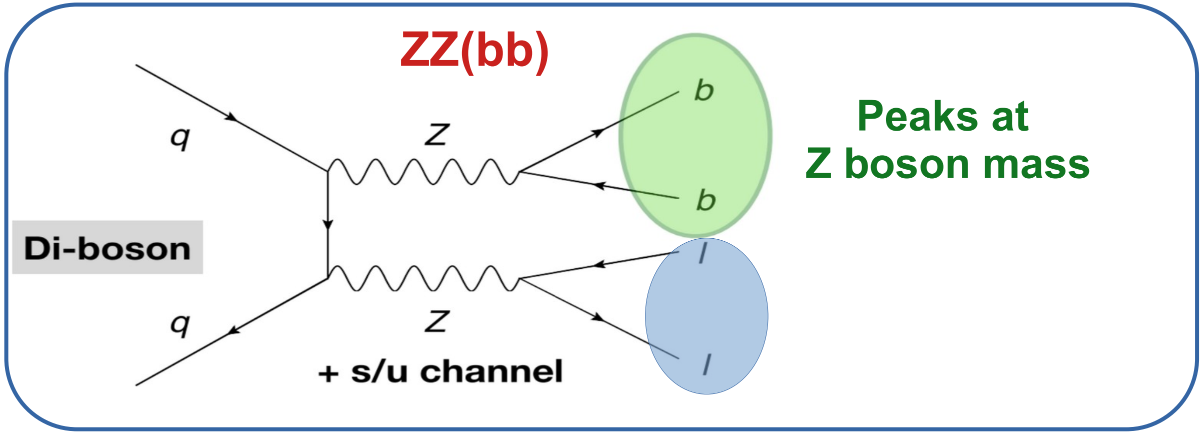
Bkg subtracted

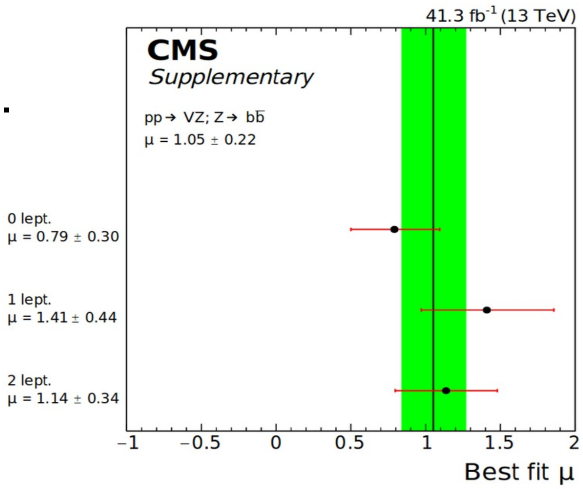


VZ(bb) cross-check analysis

- 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)