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\textsuperscript{th} Aug - 3\textsuperscript{rd} Sept 2021
Introduction

Higgs production modes

Higgs decay modes

Fermions/ Matter

<table>
<thead>
<tr>
<th>Gen</th>
<th>Quarks</th>
<th>Leptons</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Up (u), Down (d)</td>
<td>Electron (e), e neutrino</td>
</tr>
<tr>
<td>II</td>
<td>Charm (c), Strange (s)</td>
<td>Muon (μ), μ neutrino</td>
</tr>
<tr>
<td>III</td>
<td>Top (t), Bottom (b)</td>
<td>Tau (τ), τ neutrino</td>
</tr>
</tbody>
</table>

Bosons/Force carriers

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

Standard model of elementary particles

Higgs boson

Gen

Quarks

| Up (u), Down (d) | Charm (c), Strange (s) | Top (t), Bottom (b) |

Leptons

| Electron (e), e neutrino | Muon (μ), μ neutrino | Tau (τ), τ neutrino |

Fermions

- Quarks: Up (u), Down (d), Charm (c), Strange (s), Top (t), Bottom (b)
- Leptons: Electron (e), Muon (μ), Tau (τ)

Bosons

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

Decay

Signal

2-lepton channel

1-lepton channel

0-lepton channel
Background

Signal

- invariant mass has a peak at $m(V)$

- invariant mass has a peak at $m(H) \sim 125$ GeV

Main backgrounds

- falling mass distribution without peak, little softer $p_T$

- 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

2/09/2021

Krunal Gedia
Goal: Measurements in STXS

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

**Higgs boson:**
- Two jets with highest b-tagging score

**Vector boson:** Isolated leptons &/or MET

Improving jet mass resolution → dijet mass resolution

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

- Kinematic fit (2 lepton channel)
  - Constraints $m(\ell\ell) = m(Z)$ and $p_T(\text{total}) = 0$.  
  - Get constraints on jet resolution
• **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.

How to find control region and signal region

\[ m(\ell\ell) = m(Z) \]
• **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.
What do we fit?

Resolved topology in STXS bins

- SR
- CR
- DNN
- TT CR
- LF CR
- HF CR

Vpt observable observable Pure CRs

Multi-class DNN

What do we fit?

Boosted topology [link]

2/09/2021

Krunal Gedia
- 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**

**Optimized observable**

**Multi-class DNN**

**HF CR**
Expected VH results (resolved + boosted)

Uncertainties in STXS bins are statistically dominated!

Resolved+boosted topology
Summary of VHbb measurements in CMS

Inclusive (completed)

- **Goal**: Measurements done in inclusive phase space.
- Observed significance of VHbb with Run 1 + 2016 + 2017 data is $4.8\sigma$.

STXS (ongoing)

- **Goal**: Measurements in bins of $V_{pT}$ and $N_{\text{addjets}}$
- **Increased sensitivity reach**: Addition of boosted topology
- **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 ZHbb** & **WHbb**

- **l/v**
- **Z/W**

**gluon-induced ZHbb**

- **l/v**

**Dominant contribution**

- **3-channels**
  - **2-lepton**
    - b
    - e, μ
  - **1-lepton**
    - v
    - e, μ
  - **0-lepton**
    - v

Why VHbb to study H → bb coupling?

- Boost of the V-boson → QCD/V+Jets background
- Leptonic V decay → Trigger
- Large MET → Trigger
Identification of jet flavor (b-tagging)

Tracker

- Trajectory reconstruction

Calorimeter

- Measurement of Jet energy

Magnet

- Muon reconstruction

Muon chamber

Key:
- Blue: Muon
- Red: Electron
- Green: Charged Hadron (e.g., Pion)
- Gray: Neutral Hadron (e.g., Neutron)
- Dashed: Photon

CMS detector

Light jet

b-jet

1 centimeter
0.4 inches

collision point

"secondary" vertex where b hadron decayed

Z/W

1/ν

Iron return yoke interspersed with Muon chambers
2018

Observation of Higgs boson decay to bottom quarks

The CMS Collaboration

Inclusive analysis

STXS analysis

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

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

Signal strength

$1.24 \pm 0.38$  $0.88 \pm 0.29$

Resolved topology

2 AK4 jets

Boosted topology

1 AK8 Fatjet

2 AK4 jets

b

b

b

bb

$W H$  $Z H$

$p_T^V$

$0$

75

150

250

400

2/09/2021

Krunal Gedia
DNN-based b-jet regression [link]:

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

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 $m(\ell\ell) = m(Z)$ and $p_T(\text{total}) = 0$.
- Thus, get constraints on individual jets resolution.
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 $\alpha = pt(\text{sub})/pt(ll) \to 0$, 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.

2018
Scale: 1.0 +/- 0.019
Smear: 0.15 +/- 0.079
Event selection

Resolved topology

simplified strategy 0- & 1-lepton

DeepCSV

fail

LF CR

# add jets

pass

>=2

<2

m_{ll}

TT CR

∈ [90,150]

∉ [90,150]

SR

HF CR

simplified strategy 2-lepton

DeepCSV

fail

LF CR

m_{ll}

∈ [75, 105]

∉ [75, 105]

pass

TT CR

m_{ll}

∈ [90,150]

∉ [90,150]

SR

HF CR

Treatment of overlap events explained in Christina’s talk [link]

2/09/2021  Krunal Gedia
Resolved topology in STXS bins

Boosted topology in STXS bins

What we fit?

2 lepton, LF CR

DeepAK8bbVSLight

CMS Work in progress

60 fb⁻¹ (13 TeV)

Vpt shape

Vpt shape

HF DNN

(2 bin deepCSV in 2 lep)
Multivariate variables

**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.
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 continuity in pt(V) spectrum.

Resolved:
- Freefloat one rate parameter in each channel.

*Z+b is splitted in Z+1b, Z+2b for resolved topology.
Background process scale factors

<table>
<thead>
<tr>
<th>RECO $p_T(V)$</th>
<th>name</th>
<th>Znn</th>
<th>Wen+Wmn</th>
<th>Zee+Zmm</th>
</tr>
</thead>
<tbody>
<tr>
<td>75-150</td>
<td>low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150-250</td>
<td>med</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;250</td>
<td>high</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**free floating rate params:** 30 -> 15
**$p_T(V)$ shape migration uncertainty parameters:** +16

- **Z+light**
- **Z+c**
- **Z+b**
- **ttbar**

= SFs fully correlated
red = new since last Hbb meeting update

linear $p_T(V)$ shape migration uncertainty

250-400
> 250
400+
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
Combination with Run 1 and 2016/17 data

<table>
<thead>
<tr>
<th>Data set</th>
<th>Significance ($\sigma$)</th>
<th>Expected</th>
<th>Observed</th>
<th>Signal strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-lepton</td>
<td>1.9</td>
<td>1.3</td>
<td></td>
<td>0.73 ± 0.65</td>
</tr>
<tr>
<td>1-lepton</td>
<td>1.8</td>
<td>2.6</td>
<td></td>
<td>1.32 ± 0.55</td>
</tr>
<tr>
<td>2-lepton</td>
<td>1.9</td>
<td>1.9</td>
<td></td>
<td>1.05 ± 0.59</td>
</tr>
<tr>
<td>Combined</td>
<td>3.1</td>
<td>3.3</td>
<td></td>
<td>1.08 ± 0.34</td>
</tr>
<tr>
<td>Run 2</td>
<td>4.2</td>
<td>4.4</td>
<td></td>
<td>1.06 ± 0.26</td>
</tr>
<tr>
<td>Run 1 + Run 2</td>
<td>4.9</td>
<td>4.8</td>
<td></td>
<td>1.01 ± 0.23</td>
</tr>
</tbody>
</table>

Observation!

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)

2/09/2021

Krunal Gedia
Results of simultaneous fit

SF of background normalization

<table>
<thead>
<tr>
<th>Process</th>
<th>$Z(\nu\nu)H$</th>
<th>$W(\ell\nu)H$</th>
<th>$Z(\ell\ell)H \text{ low-}p_T$</th>
<th>$Z(\ell\ell)H \text{ high-}p_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>W + udscg</td>
<td>1.04 ± 0.07</td>
<td>1.04 ± 0.07</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>W + b</td>
<td>2.09 ± 0.16</td>
<td>2.09 ± 0.16</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>W + b̅b</td>
<td>1.74 ± 0.21</td>
<td>1.74 ± 0.21</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Z + udscg</td>
<td>0.95 ± 0.09</td>
<td>–</td>
<td>0.89 ± 0.06</td>
<td>0.81 ± 0.05</td>
</tr>
<tr>
<td>Z + b</td>
<td>1.02 ± 0.17</td>
<td>–</td>
<td>0.94 ± 0.12</td>
<td>1.17 ± 0.10</td>
</tr>
<tr>
<td>Z + b̅b</td>
<td>1.20 ± 0.11</td>
<td>–</td>
<td>0.81 ± 0.07</td>
<td>0.88 ± 0.08</td>
</tr>
<tr>
<td>tt</td>
<td>0.99 ± 0.07</td>
<td>0.93 ± 0.07</td>
<td>0.89 ± 0.07</td>
<td>0.91 ± 0.07</td>
</tr>
</tbody>
</table>

5.1 fb$^{-1}$ (7 TeV) + 19.8 fb$^{-1}$ (8 TeV) + 77.2 fb$^{-1}$ (13 TeV)

- Observed
- ±1σ (stat @ syst)
- ±1σ (syst)

Run 2
- 1.06 ± 0.20 (stat) ± 0.17 (syst)

Run 1
- 0.89 ± 0.38 (stat) ± 0.24 (syst)

Combined
- 1.01 ± 0.17 (stat) ± 0.14 (syst)
<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>Δμ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical</td>
<td>+0.26</td>
</tr>
<tr>
<td>Normalization of backgrounds</td>
<td>+0.12</td>
</tr>
<tr>
<td><strong>Experimental</strong></td>
<td>+0.16</td>
</tr>
<tr>
<td>b-tagging efficiency and misid</td>
<td>+0.09</td>
</tr>
<tr>
<td>V+jets modeling</td>
<td>+0.08</td>
</tr>
<tr>
<td>Jet energy scale and resolution</td>
<td>+0.05</td>
</tr>
<tr>
<td>Lepton identification</td>
<td>+0.02</td>
</tr>
<tr>
<td>Luminosity</td>
<td>+0.03</td>
</tr>
<tr>
<td>Other experimental uncertainties</td>
<td>+0.06</td>
</tr>
<tr>
<td><strong>MC sample size</strong></td>
<td>+0.12</td>
</tr>
<tr>
<td><strong>Theory</strong></td>
<td>+0.11</td>
</tr>
<tr>
<td>Background modeling</td>
<td>+0.08</td>
</tr>
<tr>
<td>Signal modeling</td>
<td>+0.07</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>+0.35</td>
</tr>
</tbody>
</table>
Combination with other Higgs production channels (where $H \rightarrow bb$)

Results (2018)

\[ \text{exp. (obs.) sig.} = 5.5 \sigma \ (5.6 \sigma) \]

\[ \mu = 1.04 \pm 0.14 \text{(stat)} \pm 0.14 \text{(sys)} \]
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.
VZ(bb) cross-check analysis

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

\[ \mu = 1.05 \pm 0.22 \text{ (obs.)} \]

\[ \text{sig.} : 5.2 \sigma \text{ (5.0 } \sigma) \]