Higgs decays to third-generation fermions at CMS

Higgs Couplings 2019 - Oxford (UK) - 01.10.2019

Janek Bechtel on behalf of the CMS collaboration
Higgs Decays To Fermions

- Direct probe of the Yukawa coupling - does the Higgs boson couple to fermions as predicted in the Standard Model?
- Branching fraction of the Higgs boson to fermions proportional to particle masses
- Decays of the Higgs bosons into b-quarks and tau leptons are most sensitive to probe Higgs decays into fermions, however both have their challenges!
- This talk: SM measurements $H \rightarrow bb$ and $H \rightarrow \tau\tau$
  - ttH: Today at 14:20
Higgs Decays To Fermions: $H \rightarrow bb$ and $H \rightarrow \tau\tau$

- **We have a clear observation** of the Higgs decay into b quarks (*) and into tau leptons (**) and are now entering the **precision era** at both the **theory** and the **experimental level**

- So far no evidence for BSM physics at ~1 TeV energy scales
  - New phenomena might only appear at larger (out-of-reach) energy scales
  - Use of effective field theories to probe **deviations in the interactions of known SM particles** resulting from **possible BSM particles at an inaccessible energy scale**

- Precision measurements of the Higgs sector is an obvious tool to search for these deviations

- **We are now beginning to stress-test the Higgs couplings to fermions**

(*) *Phys. Rev. Lett. 121 (2018) 121801*

Precision Era Measurements: Simplified Template Cross Sections

- Measured signal strength will no longer be the ultimate figure of merit for analysts
- Simplified template cross sections (STXS) evolve future measurements towards measuring cross sections in mutually exclusive regions of the phase space (“STXS bins”)
- Possible BSM effects or deviations in effective field theories can be parameterized into STXS bins
- $H \rightarrow bb$ and $H \rightarrow \tau\tau$ provide access to all Higgs production modes, with high sensitivity to VH and VBF

STXS Stage 1.0 binning for gluon fusion production
MVA techniques are now widely used at various steps of the H→bb and H→ττ analyses, e.g. for
- identification of b jets or hadronic tau leptons,
- mass regression or
- signal vs background discrimination
  → Use of MVA score as final discriminator instead of $m_{jj}$ / $m_{ττ}$

→ Discovery of H→bb was possible by the use of modern MVA techniques!

CMS actively develops data-driven background estimation methods: In H→ττ around 90% of background events are modeled from data
- μ→τ embedding for genuine di-τ events
- $F_F$ method for estimation of jets misidentified as hadronic taus
VH(bb)
VH(bb): Overview

- Higgs decay into b quarks is the largest fermionic decay
- **Highest sensitivity of** $H \rightarrow bb$ **in VH production**, in which the Higgs boson is produced in association with a W or Z boson
- Leptonically decaying vector boson is helpful for online selection and reduces QCD multijet background
- Most sensitive channel for $H \rightarrow bb$, even though VH production cross section is only third-largest of all Higgs production processes
VH(bb): Analysis Strategy

- Selection of events with 0, 1 or 2 electrons or muons and 2 b-tagged jets

- Categorization based on 0, 1 or 2 lepton selection and on $p_T$ of W/Z boson

- Most important backgrounds are
  - production of W or Z bosons in association with jets (V+jets),
  - production of top-quark pairs (tt),
  - single top production,
  - diboson (WW, WZ or ZZ) and
  - QCD multijet events

- Deep neural network and kinematic fit in two-lepton channel improve mass resolution by $\sim 15\%$
VH(bb): Analysis Strategy

- Final discriminator derived by deep neural network classifier with 5 hidden layers
- 11-14 input variables depending on W/Z decay channel. Among the most relevant variables are the b-jet classifier (deepCSV), di-jet mass and $p_T(V)$
- Separate training of the classifier is performed for each lepton category
VH(bb) Results

- The analysis reported a measurement of the signal strength of $\langle \mu_{VH} \rangle = 1.06^{+0.26}_{-0.26}$ for 2016+2017.

- In combination with the less sensitive production modes (boosted ggH, ttH, VBF) and results from Run 1, the decay $H \to bb$ was observed beyond 5$\sigma$ last year.

- Result for full Run 2 with 137 fb$^{-1}$ is on the way.

---


<table>
<thead>
<tr>
<th>CMS</th>
<th>VH, $H \to b\bar{b}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run 2</td>
<td>Observed</td>
</tr>
<tr>
<td>2016</td>
<td>$1.06 \pm 0.20$ (stat) ± 0.17 (syst)</td>
</tr>
<tr>
<td>2017</td>
<td>$1.19 \pm 0.39$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Run 1</th>
<th>Observed</th>
<th>±1σ (syst)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$0.89 \pm 0.38$ (stat) ± 0.24 (syst)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Combined</th>
<th>Observed</th>
<th>±1σ (syst)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$1.01 \pm 0.17$ (stat) ± 0.14 (syst)</td>
<td></td>
</tr>
</tbody>
</table>
\( gg \rightarrow H(bb) \)
boosted $gg \rightarrow H(bb)$: Overview

- Both the Higgs production via gluon fusion and the Higgs decay into $b$ quarks have the highest production cross sections and branching fraction respectively.

- Still, the $gg \rightarrow H(bb)$ production is almost impossible to find due to the overwhelming QCD multijet background (at least for unboosted Higgs bosons).

- Analysis focuses on very high Higgs $p_T > 450$ GeV.
boosted $gg \rightarrow H(bb)$: Results

- Much lower significance than VH(bb) due to high QCD multijet background
- Highly boosted Higgs topology will be useful for measurements of high-$p_T(H)$ STXS bins
$gg \rightarrow H(\tau\tau)$ and $VBF(\tau\tau)$
Higgs decay into tau leptons has the second largest branching fraction of fermionic decays.

Among all Higgs decays: Highest experimental sensitivity to vector boson fusion production (VBF).

Latest public result CMS-PAS-HIG-18-032 using 2016+2017 data relies on several new analysis methods with respect to observation paper, most notably:

- Large-scale use of data-driven background estimation methods
- The use of a neural net classifier for the discrimination of signal and background events
- Measurement in bins of STXS
gg → H/VBF(ττ) Background Estimation

- Di-τ analysis uses four most sensitive final states: full-hadronic (τhτh), semi-leptonic (eτh, μτh), leptonic (eμ)
- ~90% of background events are estimated from data

<table>
<thead>
<tr>
<th>Process</th>
<th>Estimated by ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>genuine ττ events (Z→ττ, tt→ττ, VV→ττ, ...)</td>
<td>μ→τ embedded events (data-driven)</td>
</tr>
<tr>
<td>prompt leptons + genuine taus (tt→lτ, VV→lτ, ...)</td>
<td>Monte Carlo simulation</td>
</tr>
<tr>
<td>jets misidentified as τh (QCD, W+jets, tt+jets, ...)</td>
<td>$F_F$ method (data-driven)</td>
</tr>
<tr>
<td>e/μ misidentified as τh (Z→ee, Z→μμ, ...)</td>
<td>Monte Carlo simulation</td>
</tr>
</tbody>
</table>

All genuine di-τ (yellow) and jets misidentified as τh (green) processes estimated from data.
Background Estimation - Embedded Events

- \( Z \rightarrow \mu\mu \) and \( Z \rightarrow \tau\tau \) decays have the same rate and characteristics.
- This enables the embedding technique, which allows describing tau backgrounds almost completely from data.
- The two muons are removed from the reconstructed event record and replaced by two simulated tau lepton decays.
- Only the two tau decays are simulated - No simulation and tuning of pileup required.
Background Estimation - $F_F$ method

- Data-driven method of estimating jets misidentified as a $\tau_h$ (in $\tau_h, \mu \tau_h, e \tau_h$) from:
  - W+Jets
  - QCD multijet
  - $tt +$ jets

- 65% of tau leptons decay into hadrons
  $\rightarrow$ Challenging to discriminate from QCD background, even with advanced identification methods

- Extrapolation factors for regions given by two orthogonal tau identification requirements are determined for each background process

---

$JHEP$ 09 (2018) 007

**Tau ID working point:**
- Tight
- $!Tight \& \&$ Very Loose
Neural network is used to classify events into several signal and background categories. Each event will receive a score for each signal and background category. The highest score determines the category of the event.

Categorization performed not by analyst but by multiclass neural network - background processes are sorted into background control categories, and signal events into signal categories.

16-21 event variables depending on decay channel are used as input. Most relevant variables are di-tau mass, visible di-tau mass, di-jet mass and lepton and jet $p_T$.
HIG-18-032 results interpreted in merged Stage 1.0 STXS bins in 9 separate parameters-of-interest (5 for ggH, 4 for VBF as indicated by the boxes below)
$gg \rightarrow H/VBF(\tau\tau)$ Results

- The analysis reported the measurement of the inclusive $H \rightarrow \tau\tau$ signal strength of
  $<\mu> = 0.75^{+0.18}_{-0.17}$

- and the measurement of the $gg \rightarrow H(\tau\tau)$ and $VBF(\tau\tau)$ production of
  $<\mu_{gg \rightarrow H}> = 0.36^{+0.36}_{-0.37}$
  $<\mu_{VBF}> = 1.03^{+0.30}_{-0.29}$

- Result for full Run 2 with 137 fb$^{-1}$ is on the way
$VH(\tau\tau)$
VH(ττ): Analysis Strategy

- Lower event yield, however also greatly suppressed background due to additional W or Z boson decaying into leptons
- Events with additional ee or μμ from a Z, or additional electron and muon from a W decay are selected
- The latest publication using 2016 data reported a VH(ττ) signal strength of $<\mu_{VH}> = 2.5^{+1.4}_{-1.3}$ with the full Run 2 result on the way
Conclusion
Conclusion

- We are entering the era of precision measurements in the analysis of Higgs decays into third-generation fermions.

- Modern MVA techniques such as neural networks are now widely used in analysis of both $H\rightarrow bb$ and $H\rightarrow \tau\tau$ decays.

- Total significance no longer the only parameter of interest - Results from the full Run 2 data collected at the CMS experiments will be given in form of simplified template cross sections.
Backup
VH(bb): Mass Resolution

**Simulation Supplementary**

- **CMS**
  - Powheg PYTHIA Z(\Gamma) H(bb)
  - No recoil jets, p_T > 150 GeV

**Plots**

2017 (13 TeV)

<table>
<thead>
<tr>
<th>Plot</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Left Plot" /></td>
<td>Kinematic fit + b-jet regression&lt;br&gt;(\mu = 120.7) GeV, (\sigma = 9.9) GeV&lt;br&gt;b-jet regression&lt;br&gt;(\mu = 124.3) GeV, (\sigma = 14.9) GeV&lt;br&gt;PF-CHS jets&lt;br&gt;(\mu = 115.9) GeV, (\sigma = 17.4) GeV</td>
</tr>
<tr>
<td><img src="image2" alt="Right Plot" /></td>
<td>Kinematic fit + b-jet regression&lt;br&gt;(\mu = 120.0) GeV, (\sigma = 12.4) GeV&lt;br&gt;b-jet regression&lt;br&gt;(\mu = 125.2) GeV, (\sigma = 15.4) GeV&lt;br&gt;PF-CHS jets&lt;br&gt;(\mu = 116.2) GeV, (\sigma = 17.9) GeV</td>
</tr>
</tbody>
</table>
VH(bb): DNN output examples

1-muon category

High pT(V) - 2-muon category
## VH(bb): Dominant Uncertainties

<table>
<thead>
<tr>
<th>Uncertainty source</th>
<th>$\Delta \mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical</td>
<td>+0.26 -0.26</td>
</tr>
<tr>
<td>Normalization of backgrounds</td>
<td>+0.12 -0.12</td>
</tr>
<tr>
<td>Experimental</td>
<td></td>
</tr>
<tr>
<td>b-tagging efficiency and misid</td>
<td>+0.09 -0.08</td>
</tr>
<tr>
<td>V+jets modeling</td>
<td>+0.08 -0.07</td>
</tr>
<tr>
<td>Jet energy scale and resolution</td>
<td>+0.05 -0.05</td>
</tr>
<tr>
<td>Lepton identification</td>
<td>+0.02 -0.01</td>
</tr>
<tr>
<td>Luminosity</td>
<td>+0.03 -0.03</td>
</tr>
<tr>
<td>Other experimental uncertainties</td>
<td>+0.06 -0.05</td>
</tr>
<tr>
<td>MC sample size</td>
<td>+0.12 -0.12</td>
</tr>
<tr>
<td>Theory</td>
<td>+0.11 -0.09</td>
</tr>
<tr>
<td>Background modeling</td>
<td>+0.08 -0.08</td>
</tr>
<tr>
<td>Signal modeling</td>
<td>+0.07 -0.04</td>
</tr>
<tr>
<td>Total</td>
<td>+0.35 -0.33</td>
</tr>
</tbody>
</table>
H→ττ: Background Estimation - \( F_F \) method

- Data-driven method of estimating jets misidentified as a \( \tau_h \) (in \( \tau_h, \mu_{\tau_h}, e_{\tau_h} \))
- Used to estimate
  - W+Jets
  - QCD multijet
  - \( tt + \) jets

\[
F_F = \sum_i w_i F_F^i = \frac{N_{AR}^i}{\sum_j N_{AR}^j}
\]

\( i, j \in \{ \text{QCD, W+jets, tt} \} \)

New

Signal Region

Application Region

Determination Region

\( m_{W^\ell} \)
$H \to \tau\tau$: DNN output examples

$\mu_{\tau_h}$ channel: $Z \to \tau\tau$ background category

$\mu_{\tau_h}$ channel: $tt$ background category
$H \rightarrow \tau\tau$: DNN output examples

$\mu_\tau$ channel: VBF signal category - VBF Topology STXS bins
H→ττ: S/(S+B) binned

CMS Preliminary

77.4 fb⁻¹ (13 TeV)

log₁₀(S/(S+B))

Events / 0.15 units