Search for the Standard Model Higgs boson in its decays to all leptonic WW in CMS

Jake Anderson
Fermilab, on behalf of the CMS collaboration
Higgs production

$\sqrt{s} = 7$ TeV

$\sigma(pp \rightarrow H + X) \ [pb]$
Higgs decays

- $H \rightarrow WW \rightarrow \ell\nu\ell\nu$ does the heavy lifting of any Higgs search.
  - large branching fraction all of the mass range.
  - clean, triggerable experimental signature

- Principal drawback is the lack of the definite mass peak.

- Without a strong WW search a Higgs discovery will be severely delayed.
The CMS detector

Muon chambers
Drift tubes/RPC in barrel
Cathode strip/RPC in endcaps
covers $|\eta| < 2.4$

Inner tracker
Silicon pixels
Silicon strips

3.8T Solenoid

Electromagnetic Calorimeter
76k PbWO$_4$ crystals

Hadronic Calorimeter
Brass/scintillator
Iron/quartz fiber
The dataset

• We analyze the full dataset collected in 2011.
  ◦ LHC delivered: 6.1 fb\(^{-1}\)
  ◦ CMS recorded: 5.6 fb\(^{-1}\)

• Peak instantaneous luminosity: 4.02 nb\(^{-1}/s\)
Strategy

• Without a definite mass peak, understanding, modeling and controlling the backgrounds is key to sensitivity.

• Key backgrounds:
  ◦ V + jets
  ◦ Top
  ◦ EW diboson production

• Exploiting the scalar nature of the Higgs leads to best handle for irreducible non-resonant WW production.
Strategy (II)

- Given the various backgrounds and their relative contributions the analysis is split into 5 separate sub-channels.
  - same-flavor/different-flavor
  - 0-jet/1-jet
  - 2 jets (VBF selections)
- The most sensitive is the different-flavor, 0-jet channel.
The background (W+jets)

- Tight lepton identification and isolation requirements keep this background in check.
- Normalization of the residual contribution is taken from the data.
  - Relax the lepton selection to define the control sample.
  - Use jet faking lepton rates to extrapolate to the signal sample.
The background (Drell-Yan)

- Critical background in same flavor channels.
- Cut on relatively large amount of MET.
- Veto events with the same lepton flavor that fall near the Z peak.

- Residual background is determined from background subtracted data near the Z peak and extrapolated to the signal region using MC.
The background (top)

- Divide the data into jet bins (0, 1, and 2 extra jets).
  - 0 jet bin has much less top (i.e., more sensitivity)
- Veto events with b-tagged jets and soft muons (top-tagged)
- $\varepsilon_{\text{top-tag}} \sim 50\%$ as measured with data.
- Residual background estimated as

$$\frac{N_{\text{top-tag}}}{\varepsilon_{\text{top-tag}}} \times (1 - \varepsilon_{\text{top-tag}})$$

16-20 April 2012
J. Anderson - MCTP2012
The background (WW)

- The non-resonant WW production is an irreducible background.
- We can still exploit the kinematics of the event to separate this from Higgs decays.
- The variables of choice: $\Delta\phi_{\ell\ell}$ and $m_{\ell\ell}$. 
Kinematics

- Higgs decays have smaller opening angle for leptons.
- Higgs decays have smaller di-lepton mass.
The Di-boson background

- Normalization of the irreducible WW background is taken from data sidebands for low H mass and from simulation for larger masses.
- The remaining di-boson backgrounds are small and estimated from simulation.
- CMS has measured the electroweak WW production cross-section and it agrees very well with predictions.
Stacking up the background

- Extensive effort has been made to control and model the background components of this search.
- The agreement between the data and the MC is generally very good at every step of the analysis.
- Two complementary analysis techniques have been carried out
  - cut-based
  - multivariate based
Cut-based analysis

- There are mass dependent cuts on the lepton $p_T$, $m_{\ell\ell}$, $\Delta\phi_{\ell\ell}$ and the Higgs $m_T$.

- Events for the background and expected Higgs signal are counted and limits are set.

<table>
<thead>
<tr>
<th>$m_H$ [GeV]</th>
<th>$p_T^{\ell_{\text{max}}}$ [GeV]</th>
<th>$p_T^{\ell_{\text{min}}}$ [GeV]</th>
<th>$m_{\ell\ell}$ [GeV]</th>
<th>$\Delta\phi_{\ell\ell}$ [$^\circ$]</th>
<th>$m_T$ [GeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>$&gt;$</td>
<td>$&gt;$</td>
<td>$&lt;$</td>
<td>$&lt;$</td>
<td>[80,120]</td>
</tr>
<tr>
<td>130</td>
<td>25</td>
<td>10 (15)</td>
<td>45</td>
<td>90</td>
<td>[80,125]</td>
</tr>
<tr>
<td>160</td>
<td>30</td>
<td>25</td>
<td>50</td>
<td>60</td>
<td>[90,160]</td>
</tr>
<tr>
<td>200</td>
<td>40</td>
<td>25</td>
<td>90</td>
<td>100</td>
<td>[120,200]</td>
</tr>
<tr>
<td>250</td>
<td>55</td>
<td>25</td>
<td>150</td>
<td>140</td>
<td>[120,250]</td>
</tr>
<tr>
<td>300</td>
<td>70</td>
<td>25</td>
<td>200</td>
<td>175</td>
<td>[120,300]</td>
</tr>
<tr>
<td>400</td>
<td>90</td>
<td>25</td>
<td>300</td>
<td>175</td>
<td>[120,400]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$M_H$ [GeV]</th>
<th>data</th>
<th>background $\pm$</th>
<th>$H \rightarrow WW$ $\pm$</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td>136</td>
<td>136.7 $\pm$ 12.7</td>
<td>15.7 $\pm$ 0.8</td>
</tr>
<tr>
<td>130</td>
<td>193</td>
<td>191.5 $\pm$ 14.0</td>
<td>45.2 $\pm$ 2.1</td>
</tr>
<tr>
<td>160</td>
<td>111</td>
<td>101.7 $\pm$ 6.8</td>
<td>122.9 $\pm$ 5.6</td>
</tr>
<tr>
<td>200</td>
<td>159</td>
<td>140.8 $\pm$ 6.8</td>
<td>48.8 $\pm$ 2.2</td>
</tr>
<tr>
<td>400</td>
<td>109</td>
<td>110.8 $\pm$ 5.8</td>
<td>17.5 $\pm$ 0.8</td>
</tr>
</tbody>
</table>

0 jet bin
Cut-based limit

- expected exclusion: $129 \text{ GeV} < m_H < 236 \text{ GeV}$
- observed exclusion: $132 \text{ GeV} < m_H < 238 \text{ GeV}$
More advanced analysis

- To obtain more sensitivity, we employ a multivariate classifier based on boosted decision trees (BDT).
  - Uses kinematic observables from the cut based analysis and
    - $\Delta R_{\ell\ell} = [(\Delta \eta_{\ell\ell})^2 + (\Delta \phi_{\ell\ell})^2]^{\frac{1}{2}}$
    - $m_T$ for both $\ell$-MET pairs
    - lepton flavors
  - trained to discriminate Higgs from non-resonant $WW$ production.
  - Trained separately for each mass hypothesis and jet content bin.
  - Shapes of the BDT output for both signal and background are used to set the final limit.
• The BDT analysis was cross-checked using both an invariant mass shape analysis and a matrix element analysis.

• Shape variations in the background components are fully studied and taken as a systematic error.
Final limit (BDT)

- expected exclusion: $127 \text{ GeV} < m_H < 270 \text{ GeV}$
- observed exclusion: $129 \text{ GeV} < m_H < 270 \text{ GeV}$
Limit (low mass view)

95% CL limit on $\sigma/\sigma_{SM}$

- Median expected
- Expected $\pm 1\sigma$
- Expected $\pm 2\sigma$
- Observed

CMS preliminary

$H \rightarrow WW$ (BDT based)

$L = 4.6 \text{ fb}^{-1}$

Higgs mass [GeV]
Contribution to the limits

CMS, $\sqrt{s} = 7$ TeV
$L = 4.6-4.8$ fb$^{-1}$

95% CL limit on $\sigma/\sigma_{SM}$

Expected limits
- Combined
- $H \rightarrow bb$ (4.7 fb$^{-1}$)
- $H \rightarrow \tau\tau$ (4.6 fb$^{-1}$)
- $H \rightarrow \gamma\gamma$ (4.8 fb$^{-1}$)
- $H \rightarrow WW$ (4.6 fb$^{-1}$)
- $H \rightarrow ZZ \rightarrow 4l$ (4.7 fb$^{-1}$)
- $H \rightarrow ZZ \rightarrow 2l 2q$ (4.6 fb$^{-1}$)

Higgs boson mass (GeV)
Summary

• CMS has searched for the SM Higgs boson using the full dataset collected in 2011 in its leptonic WW decays, among other decay channels.

• We do not observe any significant excesses and are able to exclude the SM Higgs boson with masses from 127 GeV to 270 GeV.

• The data collected in 2012 will expand the reach of this decay mode.

• CMS also intends to include semi-leptonic $H \rightarrow WW$ to improve the sensitivity particularly at higher mass.
References

- https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig11024TWiki

**WH → WWW → 3ℓ3ν**

- https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig11034TWiki
Errors

• Errors on backgrounds
  ◦ Rate and shape systematic errors on background processes via changing generators, parameters and smearing.
  ◦ Normalization of most backgrounds taken from data control regions.
    – Errors based on both the control region and the extrapolation.
  ◦ Effect of pileup on distributions are also included.
  ◦ Total uncertainty on background is ~15% and dominated by statistics of the control regions.

• Errors on signal
  ◦ Signal efficiencies taken from MC.
  ◦ Errors on signal efficiencies fall into two categories
    – effects of signal migration between jet multiplicity bins – evaluated by comparing different generators.
    – effects of PDFs, $\alpha_s$, higher-order corrections, etc. – evaluated using the PDF4LHC prescriptions.
  ◦ Overall signal efficiency uncertainty ~20%.
Signal injection at 124 GeV

$H \rightarrow WW \rightarrow 2l2\nu + 0/1/2$-jets, mva based

- --- 95% CL exclusion: median
- Green: 95% CL exclusion: 68% band
- Yellow: 95% CL exclusion: 95% band
- Blue: pseudo-data (with Higgs(124))

95% CL Limit on $\sigma_{SM}$

Higgs mass, $m_H$ [GeV/$c^2$]

16-20 April 2012

J. Anderson - MCTP2012
2-jet VBF selection

- 2 jets $p_T > 30$ GeV
- no jets with $p_T > 30$ GeV in the $\eta$ region between the two jets.
- $|\Delta\eta_{jj}| > 3.5$
- $m_{jj} > 450$ GeV
Fermiophobic Higgs

H → WW → 2\ell2ν + 0/1/2-jets, fermiophobic

95% CL Limit on $σ/σ_{SM}$

Higgs mass, $m_H$ [GeV/c$^2$]