Search for the Standard Model
Higgs Boson in $H \rightarrow bb$ @ CMS

S.Z. Shalhout [UC Davis]
on behalf of the CMS collaboration
April 5, 2012
University of Oregon
Outline

- Present results of CMS H→bb search in 4.7/fb @ 7 TeV.
- Touch on main analysis techniques
- Full details in arXiv:1202.4195v1
Introduction

• Direct searches have excluded much of the (lower) mass range for a SM Higgs:

- Small windows around 115 and 125 GeV remain un-excluded.
- In this region of $M_H$, SM Higgs mostly decays to $b\bar{b}$
SM Higgs @ The LHC

- Production Modes @ 7 TeV:

\[ \sigma(pp \rightarrow H+X) \text{ in pb} \]

\[ M_H \text{ [GeV]} \]

\[ \sqrt{s} = 7 \text{ TeV} \]

Focus of this Talk

@ 125 GeV:

- $g \rightarrow H$
- $q' \rightarrow q H$
- $\bar{q}' \rightarrow \bar{q} H$
- $q \rightarrow W^+ H$
- $\bar{q} \rightarrow W^- H$
- $g \rightarrow t H$
- $g \rightarrow \bar{t} H$

15 pb 1.2 pb 0.6 + 0.3 pb 0.09 pb
SM Higgs @ The LHC

• Branching Ratios:

![Branching Ratios Diagram]

- **below 135 GeV:**
  mostly $H \rightarrow bb$

- **above 135 GeV:**
  mostly $H \rightarrow W^+W^-$
CMS Detector & Data taking

- 5.2/fb recorded @ 7 TeV
- 90% data taking efficiency
H→bb Search Modes \((M_H < 135 \text{ GeV})\)

- \(gg\rightarrow H\rightarrow bb\) has largest rate, but comes with huge QCD dijet background!
Z/W Associated Production

- Next best is associated production (WH/ZH)
- Z and W decay decay products distinguish signal from QCD, provide ‘trigger-able’ topologies

Search in 5 final states

<table>
<thead>
<tr>
<th>Mode</th>
<th>Trigger Requirements</th>
<th>Signal Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W \rightarrow \mu^\pm\nu$</td>
<td>muon with $p_T &gt; 17$ to 40 GeV/c</td>
<td>90%</td>
</tr>
<tr>
<td>$W \rightarrow e^\pm\nu$</td>
<td>electron with $E_T &gt; 17$ to 30 GeV</td>
<td>95%</td>
</tr>
<tr>
<td>$Z \rightarrow \mu^+\mu^-$</td>
<td>muon with $p_T &gt; 17$ to 40 GeV/c</td>
<td>90%</td>
</tr>
<tr>
<td>$Z \rightarrow e^+e^-$</td>
<td>2 electrons with $E_T &gt; 17$ and $E_T &gt; 8$ GeV</td>
<td>99%</td>
</tr>
<tr>
<td>$Z \rightarrow \nu\nu$</td>
<td>momentum imbalance or missing transverse energy + jets</td>
<td>98%</td>
</tr>
</tbody>
</table>

Combination of Trigger Selection
Z/W Associated Production

- QCD background largely under control
- Main backgrounds are from W/Z+jets

W/Z + jets can also give

- $\mu^+ + \nu + bb$
- $e^+ + \nu + bb$
- $\nu + bb$
- $e^- + e^+ + bb$
- $\mu^- + \mu^+ + bb$

better, but still have large backgrounds from W/Z+jets
Boost away from W/Z+jets

- Perform search in “boosted” topology

- Require $p_T$ of dijet system and $p_T$ of $Z/W$ decay products to be large

arXiv:0802.2470v2
PhysRevLett.100.242001

Require $P_T(V)$ and $P_T(jj) > \sim 160$ GeV

better, but top backgrounds still large
**top backgrounds**

- $\text{WH} \rightarrow (e^\pm/\mu^\pm + \nu) \text{bb}$ has large top backgrounds even in the boosted regime

![Diagram showing normalized events vs. additional jet multiplicity](image)

Reduce top backgrounds by rejecting events with additional jet activity.
Putting it all together:

- Selections built around few main signal characteristics:
  
  - Dijet boost:
    
    WH ($p_T > 165$ GeV/c)
    Z($l^+l^-$)H ($p_T > 100$ GeV/c)
    Z($\nu\nu$)H ($p_T > 160$ GeV/c)

  - Significant missing transverse energy
    
    WH (MET > 35 GeV)
    Z($\nu\nu$)H (MET > 160 GeV)

  - Limited additional jet content
b-jet Identification:

- b-jet ID relies on secondary vertex info, displaced tracks with large impact parameters, &/OR semi-leptonic decays.

- H→bb search uses a “combined secondary vertex” tagger (CSV) which combines vertex and track info with a likelihood ratio:

  \[ \text{jet 1 CSV} > 0.898 \]
  \[ \text{jet 2 CSV} > 0.5 \]

- ~70% b-tag

- ≈1% usdg “mis” tag rate
S/B Optimized Cuts

- Cuts defining signal regions are optimized for highest S/B at each test mass
- As bkg. composition varies by Z/W decay produce a separate set of cuts for each of the 5 channels

Background Composition by SubChannel

- Background Fraction
- Cuts defining signal regions are optimized for highest S/B at each test mass
- As bkg. composition varies by Z/W decay produce a separate set of cuts for each of the 5 channels
Control Region Example #1

- \( W + bb \) control region (inverted boost)
• tt control region (additional jets, Z mass window sideband)
Control Region Example #3

- $Z \rightarrow \nu \nu$ control region (deleted muon $Z \rightarrow \mu \mu$ data)
Defining The Final Signal Regions

- Two approaches:

- cut & count $S$ & $B$ in $M_{jj}$ Window

- cut & count $S$ & $B$ above MVA minimum
Defining The Final Signal Regions

- Two approaches:

**Cut & count S & B in $M_{jj}$ Window**

**Cut & count S & B above MVA minimum**

“Cross-Check” Analysis

Main Result
BDT Search

- Main result uses a **Boosted Decision Tree** discriminant
  - ~100’s of trees returning S or B binary classification
  - hard-to-classify events receive higher weights in training of subsequent trees (boosting)
  - Final output is a weighted sum (reflecting classification purity) of tree decisions

- S/B optimized cut on BDT output + “counting experiment”

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M(jj)$</td>
<td>dijet invariant mass</td>
</tr>
<tr>
<td>$p_T(jj)$</td>
<td>dijet transverse momentum</td>
</tr>
<tr>
<td>$p_T(V)$</td>
<td>vector boson transverse momentum</td>
</tr>
<tr>
<td>CSV1</td>
<td>value of CSV for best b-tagged jet</td>
</tr>
<tr>
<td>CSV2</td>
<td>value of CSV for second-best b-tagged jet</td>
</tr>
<tr>
<td>$\Delta \phi(W,H)$</td>
<td>azimuthal angle between V and dijet</td>
</tr>
<tr>
<td>$\Delta \eta(J_1,J_2)$</td>
<td>different in $\eta$ between Higgs daughters</td>
</tr>
<tr>
<td>$N_{aj}$</td>
<td>number of additional central jets (in WH, Z(ee)Hand $Z(\nu\nu)H$)</td>
</tr>
</tbody>
</table>
BDT Search

• Main result uses a **Boosted Decision Tree** discriminant
  
  • ~100’s of trees returning S or B binary classification
  
  • hard-to-classify events receive higher weights in training of subsequent trees (boosting)
  
  • Final output is a weighted sum (reflecting classification purity) of tree decisions

• S/B optimized cut on BDT output + “counting experiment”
Mjj Search

- BDT result is cross-checked with a search using $M_{jj}$
- Dijet Mass resolution $\sim 10\%$
- Good separation between ZH/WH and backgrounds

- Counting experiment at each test-mass within an S/B optimized window

<table>
<thead>
<tr>
<th>search windows</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M(jj)(110)$</td>
</tr>
<tr>
<td>$M(jj)(115)$</td>
</tr>
<tr>
<td>$M(jj)(120)$</td>
</tr>
<tr>
<td>$M(jj)(125)$</td>
</tr>
<tr>
<td>$M(jj)(130)$</td>
</tr>
<tr>
<td>$M(jj)(135)$</td>
</tr>
</tbody>
</table>
Systematics

- Vary known systematics by ±σ and take resulting Δ efficiency as the uncertainty on S or B normalization

<table>
<thead>
<tr>
<th>Source</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>luminosity</td>
<td>4.5%</td>
</tr>
<tr>
<td>lepton eff.</td>
<td>2%</td>
</tr>
<tr>
<td>trigger eff.</td>
<td>2%</td>
</tr>
<tr>
<td>unclustered energy</td>
<td>3%</td>
</tr>
<tr>
<td>JES</td>
<td>2-3%</td>
</tr>
<tr>
<td>JER</td>
<td>3-6%</td>
</tr>
<tr>
<td>b-tag rate</td>
<td>6-15%</td>
</tr>
<tr>
<td>signal cross sections</td>
<td>4%</td>
</tr>
<tr>
<td>p_{T} shape</td>
<td>~10%</td>
</tr>
<tr>
<td>background normalization</td>
<td>up to 30%</td>
</tr>
<tr>
<td>MC stats</td>
<td>1-5%</td>
</tr>
</tbody>
</table>

- total signal uncertainty ~20%
- total background uncertainty ~30%
Search Summary:

I. Identify candidates selected by appropriate triggers.

II. Impose kinematic selections to reject major backgrounds (aim for maximum expected S/B)

III. Further reduce backgrounds by identifying b-jets ("b-tagging")

IV. Improve separation of signal from background with a multivariate discriminant (BDT)

V. Validate data model in several control samples.

VI. Using the number of observed candidates, SM expectation & systematics extract upper limits on SM Higgs production rate.
Results:

- No significant excess observed in any search channel
Results:

- Combined dijet mass:

- Sum of all VH modes
Results:

- 95% CL upper limits on $\sigma_H \times \text{BR}(H \rightarrow bb)$ using CLs method:

<table>
<thead>
<tr>
<th>$M_H$ (GeV)</th>
<th>BDT Analysis</th>
<th>$M(jj)$ Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CLs Exp Obs</td>
<td>CLs Exp Obs</td>
</tr>
<tr>
<td>110</td>
<td>2.74 3.14</td>
<td>3.05 3.44</td>
</tr>
<tr>
<td>115</td>
<td>3.12 5.18</td>
<td>3.21 5.56</td>
</tr>
<tr>
<td>120</td>
<td>3.56 4.38</td>
<td>4.14 6.07</td>
</tr>
<tr>
<td>125</td>
<td>4.27 5.72</td>
<td>4.74 6.31</td>
</tr>
<tr>
<td>130</td>
<td>5.28 9.00</td>
<td>6.42 10.5</td>
</tr>
<tr>
<td>135</td>
<td>6.74 7.53</td>
<td>7.67 8.92</td>
</tr>
</tbody>
</table>

- BDT analysis is up to 20% more sensitive than Mjj search
Results:

- 95% CL upper limits on $\sigma_{H} \times \text{BR}(H \rightarrow bb)$ using CLs method:

- BDT analysis is up to 20% more sensitive than Mjj search
Conclusions

• CMS has searched for $H \rightarrow bb$

• Challenging search mode due to small $ZH/WH$ rates and large backgrounds

• Achieved sensitivity of $4.3 \times SM$ at 125 GeV

Thanks!