Multivariative analysis of \( H \rightarrow b\bar{b} \) in \( t\bar{t}H \) associated production mode in ATLAS with fast simulation

Sergey Kotov

MPI für Physik

DPG meeting, Berlin, Mar 3-10, 2005
Motivation for $H \rightarrow b \bar{b}$ in $t\bar{t}H$ production mode

Channel description

Event reconstruction

Neural network training

Analysis results

Summary
Low mass SM Higgs boson overview

**LEP2 experimental bounds on Higgs mass**

- Precision measurements of EW observables: $m_H = 117^{+67}_{-45}$ GeV
- Direct searches: $m_H > 114$ GeV

**Signature channels for low mass SM Higgs**

- $H \rightarrow \tau^+\tau^-$ in VBF production mode
- $H \rightarrow WW^* \rightarrow l\nu l\nu$ in VBF production mode
- $H \rightarrow \gamma\gamma$ in gluon fusion production mode
- $H \rightarrow b\bar{b}$ in $t\bar{t}H$ associated production mode
Channel description

Features of $t\bar{t}H$, $H \rightarrow b\bar{b}$ channel

- Complex final state
  - 6 jets: 4 b-jets and 2 light jets
  - 1 high-$p_t$ lepton (trigger)
  - missing energy $E_{miss}^t$ from neutrino
  - additional jets from ISR/FSR

- Large backgrounds
  - combinatorial from mis-pairing of b-jets in signal events
  - irreducible from $t\bar{t}b\bar{b}$ events
  - reducible from $t\bar{t}jj$ events

full reconstruction of event and very good b-jet tagging are needed to suppress backgrounds

Signal and background cross sections

<table>
<thead>
<tr>
<th>Process</th>
<th>$t\bar{t}H$ $(m_H=120 \text{ GeV})$</th>
<th>$t\bar{t}b\bar{b}$ $(\geq 6 \text{ jets})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$, pb</td>
<td>0.5</td>
<td>9.2</td>
</tr>
</tbody>
</table>

$t\bar{t}H$ signal process diagram

$t\bar{t}b\bar{b}$ background

$t\bar{t}jj$ background
Event topology, preselection and data samples

**Preselection: event topology cuts**
- 1 isolated lepton with $p_t > 20(25)$ GeV for $\mu(e)$ and $\eta < 3$
- $\geq 6$ jets with $p_t > 20$ GeV and $\eta < 5$
- $\geq 4$ of jets identified as b-jets

**Data samples**

<table>
<thead>
<tr>
<th>Process</th>
<th>$t\bar{t}H$</th>
<th>$t\bar{t}b\bar{b}$</th>
<th>$t\bar{t}jj$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generated events, M</td>
<td>0.6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Preselection efficiency, %</td>
<td>2.1</td>
<td>0.8</td>
<td>0.02</td>
</tr>
</tbody>
</table>

**ATLFAST** underestimates the b-tagging efficiency in events with multiple b-jets ($\epsilon_b=0.42$ compared to $\epsilon_b=0.60-0.65$ in full simulation), so we use an efficiency scale factor of 3 in our final calculations.

**PYTHIA 6.2** and **AcerMC 2.3** programs are used for event generation. **ATLFAST** package is used for fast simulation of ATLAS detector response.
Event reconstruction

Main reconstruction challenge

How to reconstruct $t\bar{t}$-pair from all possible $b\ell\nu - bjj$ combinations properly with correct assignment of b-jets?

Various approaches to $t\bar{t}$-pair reconstruction

- ATLAS Technical Design Report: select combination with minimal $\Delta^2 = (m_{bl\nu} - m_t)^2 + (m_{bjj} - m_t)^2$
- a recent improved approach which uses likelihood techniques for reconstruction of $t\bar{t}$-pair
- this analysis uses neural network technique for $t\bar{t}$-pair reconstruction

Reconstruction strategy with Neural Network

- use events which pass preselection criteria (1 lepton, 4 b-jets, 2 light jets)
- determine $p_\nu$ from $p_l$ and $p_{miss}$ using $m_W$ constraint (if fails, use approximation $p_{\nu}^z = p_l^z$)
- select all possible reconstructed combinations of lepton, neutrino, 2 light jets and 2 b-jets for which the reconstructed invariant masses $m_{jj}$, $m_{bl\nu}$, and $m_{bjj}$ fit inside some mass windows of W boson and t-quark (30 GeV and 70 GeV respectively)
- feed parameters of these combinations through a neural network (which was trained beforehand on a sample of combinations matched and non-matched to MC generator truth table) and select combination with the highest NN output value
- assign the remaining 2 b-jets to the Higgs boson and plot their invariant mass $m_{bb}$
Multilayer Perceptron

Combination input variables

Input layer

Hidden layer

Output layer

Probability of being a signal combination

- 14 input variables: 5 invariant mass differences, 8 angular separations in \( \phi-\eta \) plane, 1 sum of transverse momenta
- **TMultiLayerPerceptron** ROOT built-in class is used as neural network
- 6000 of matched and 21000 of non-matched combinations were used (with proper weights) to train the neural network

\[
\begin{align*}
    \hat{h}_j &= f(\sum x_i w_{ij} + \theta_j) \\
    f &= \frac{1}{1+e^{-x}} \\
    O &= \sum h_j w_j + \theta_o \\
    E &= \frac{1}{N} \sum (O-T)^2
\end{align*}
\]
Neural network variables

Invariant mass of two light jets

$\phi$-$\eta$ grid between two light jets

Invariant mass of jjb system

Angle in $\phi$-$\eta$ grid between jj-system and b-jet

Mass difference of reconstructed $t$-quarks from nominal

Total transverse momenta of reconstructed $t$-quarks

Sergey Kotov (MPI für Physik)  Mutltivariative $t\bar{t}H$, $H \rightarrow b\bar{b}$ analysis

Berlin, Mar 4, 2005  8 / 11
Signal/background separation with neural network

**Effectiveness of the neural network for background suppression**

<table>
<thead>
<tr>
<th>Cut</th>
<th>$\epsilon_{ttH}$</th>
<th>$\epsilon_{cmbn}$</th>
<th>$\epsilon_{ttbb}$</th>
<th>$\epsilon_{ttjj}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.768</td>
<td>0.238</td>
<td>0.209</td>
<td>0.163</td>
</tr>
<tr>
<td>0.6</td>
<td>0.657</td>
<td>0.157</td>
<td>0.139</td>
<td>0.090</td>
</tr>
<tr>
<td>0.7</td>
<td>0.510</td>
<td>0.093</td>
<td>0.080</td>
<td>0.046</td>
</tr>
<tr>
<td>0.8</td>
<td>0.330</td>
<td>0.044</td>
<td>0.038</td>
<td>0.023</td>
</tr>
</tbody>
</table>
Expected number of events for $L = 30 fb^{-1}$

<table>
<thead>
<tr>
<th>$\epsilon$ scale factor</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon$, %</td>
<td>1.62</td>
<td>1.08</td>
<td>0.54</td>
<td>1.32</td>
</tr>
<tr>
<td>$N_{ttH}$</td>
<td>73.4</td>
<td>49.0</td>
<td>24.5</td>
<td>41.9</td>
</tr>
<tr>
<td>$N_{ttbb}$</td>
<td>421.5</td>
<td>281.0</td>
<td>140.5</td>
<td>164.2</td>
</tr>
<tr>
<td>$N_{ttjj}$</td>
<td>97.3</td>
<td>64.9</td>
<td>32.4</td>
<td>54.6</td>
</tr>
<tr>
<td>$S/\sqrt{B}$</td>
<td>3.2</td>
<td>2.6</td>
<td>1.9</td>
<td>2.8</td>
</tr>
<tr>
<td>Purity</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Purity of reconstructed $t\bar{t}H$ events (fraction of events with all 4 b-jets correctly assigned) is important for finding of the Higgs mass peak.
Conclusions

- A neural network approach was tried for reconstruction of $t\bar{t}$-pair in $t\bar{t}H$ events produced with fast simulation of the ATLAS detector ⇒ no considerable improvement with comparison to previously used likelihood approach was achieved.

- The obtained signal significance is still rather low, in large due to imperfections of the fast simulation algorithms.

Future plans

- ATLAS collaboration now is in the process of mass production of full simulation data for its summer physics workshop in Rome ⇒ first results show a HUGE improvement in b-tagging due to development of new combined algorithms.

- The neural network approach will be re-tried on this sample of fully simulated events as soon as it will be available.

- More elaborate consideration of the background sources is need to improve signal significance (ATLFAST program needs to be tuned for this using full simulation data, because it is impossible to produce the required amount of background events with full simulation of the ATLAS detector).