Search for the Standard Model Higgs Boson in Missing Energy and b-jet Final States at the Tevatron

Tim Scanlon

On behalf of the CDF and DØ Collaborations
Outline

- Search overview
  - Signal signature
  - Background

- Analysis Technique
  - Event selection
  - b-jet identification
  - Multivariate selection

- Limits

- Conclusions

Delivered > 11.5 fb$^{-1}$
Recorded > 10 fb$^{-1}$
Using 8.4 fb$^{-1}$

[ Thanks to all my Tevatron colleagues ]

Tim Scanlon DPF 2011
Signal Signature

- Searching for:
  - Large missing energy
  - Two b-jets
  - No reconstructed isolated leptons

- Two dominant channels
  - $Z \rightarrow \nu \nu$
  - $W \rightarrow l \nu$ with missing lepton

- Two searches
  - D0 8.4 fb$^{-1}$
  - CDF 7.8 fb$^{-1}$
Define four samples:
- Signal sample
- Multijet model
- Multijet Control
- Physics Control

**Backgrounds**

- **Instrumental**
  - Multijet (MJ)
  - Large background
  - Difficult to model
    - Estimated from data

- **Physics**
  - V+jets (V=W/Z)
  - Diboson
  - Top

Tim Scanlon DPF 2011
Analysis Strategy

Kinematic event Selection
Analysis Strategy

Kinematic event Selection

b-jet Identification
Analysis Strategy

Kinematic event Selection

b-jet Identification

Multijet Removal
Analysis Strategy

Kinematic event Selection

b-jet Identification

Multijet Removal

Final Discriminant
Large missing energy (MET)

- CDF: MET > 35 GeV
- D0: MET > 40 GeV and MET Significance > 5

Two or more jets

- D0: 2 jets not back to back
- CDF: 2 or 3 jets

CDF relaxed requirements

- MET 50→35 GeV
- Lead jet pT 35→25 GeV
- Increase acceptance 30-40%

No isolated leptons
Multijet Model Sample

- **Multijet events**
  - MET tends to align with jets
  - Use as handle on MJ events

- **Define two regions**
  - **Signal**
    - MET and jets well separated in respect of azimuthal angle
  - **MJ model**
    - MET and jets closely aligned
    - Negligible signal

- **Events in MJ region used to model MJ in signal region**
  - Validated in control region
Multijet Control Sample

- Sample dominated by MJ
  - Kinematically similar to signal sample
  - D0
    - Relax MET cut
    - No MET significance cut
  - CDF
    - High MET
    - MET aligned with jet
- Validates modelling of MJ background
  - All variables well modelled
Physiology Conference Sample

- Form sample enriched in W+jets/top events

- Require:
  - D0: Isolated muon
  - CDF: Isolated electron/muon
  - Keep other cuts the same as signal region

- Validates modelling of EW/top backgrounds
  - All variables well modelled
b-Jet Identification

- Tools
  - Secondary vertex
  - Jet lifetime probability
  - Counting impact parameters
  - Combine all information using multivariate technique

Increase in $S/\sqrt{B} > x10$

Higgs x500

Higgs x10
b-Jet Identification

- **CDF - Secondary vertex and lifetime probability**
  - Three exclusive tagging channels
    - One single tag/two double tag

- **D0 - Multivariate (MVA) tagging**
  - Very loose single and double tag channels
    - ~80% b-jets, ~10% light-jets
  - Use MVA output as input to final discriminant
    - ~10% gain in sensitivity
Multijet Removal

- Train multivariate technique to remove large MJ background
  - D0 Decision Tree (DT)
  - CDF Neural Network (NN)
  - Kinematic variables used as input
    - D0 (30), CDF(14)

- Trained in pretag samples

Example variables:
- MET from jets, tracks and calorimeter
- Relative size of MET
- Relative position of jets/MET
- Event shape

---

Tim Scanlon DPF 2011
**Multijet Removal**

**CDF**

- Signal efficiency: ~90%
- MJ Rejection: ~87%

**D0**

- Signal efficiency: ~88%
- MJ Rejection: ~80%

MJ 4.5% of double tag signal sample

~ MJ 44% of double tag signal sample
Final Discriminant

- Train DT (D0) or NN (CDF) to separate signal from signal-like backgrounds
  - Trained separately in
    - Each b-tagging/jet channel

- Variables
  - D0
    - Inputs to MJ DT plus b-tagging and mass related variables
  - CDF
    - Output of MJ NN, mass, direction of jets and MET related variables
Final Discriminant

Single Tag

Double Tag
Systematic Uncertainties

- Uncertainties included on:
  - Shape
  - Normalisation

- Dominant uncertainties:
  - Jet energy
  - $b$-tagging
  - Background cross section
  - Luminosity

- Impact reduced by constraining uncertainties during limit setting
Limits

- Limits set at 95% CL relative to SM cross section

**Expected:** 4.0  
**Observed:** 3.2  
@115 GeV

**Expected:** 3.0  
**Observed:** 2.3
Conclusions

• SM Higgs search for missing energy and b-jets is one of the most sensitive at the Tevatron
  ➢ Vital channel in low mass Higgs search

• Large increase in sensitivity of searches
  ➢ Using up to 8.4 fb\(^{-1}\) of data
  ➢ More intelligent use of b-tagging
  ➢ Loosened kinematic cuts

• Many further improvements in pipeline
  ➢ More data
  ➢ Reduced systematic errors
  ➢ Increased acceptance

The SM Higgs is running out of places to hide!
Watch this space
Backup slides
Low Mass Higgs Search

- Low mass Higgs ($m_H < 135$ GeV)
  - Dominant decay to two b-quarks
  - Large heavy flavour jet background

- Associated production provides more distinctive decays
  - Search for various decay channels of $W/Z$

Tim Scanlon DPF 2011
B-tagging - (DØ) Certification

- Have MC / data differences - particularly at a hadron machine
  - Measure performance on data
    - Tag Rate Function (TRF)
      - Parameterized efficiency & fake-rate as function of $p_T$ and $\eta$
  - Use to correct MC b-tagging rate

- b and c-efficiencies
  - Measured using a b-enriched data sample

- Fake-rate
  - Measured using multi-jet data

Tim Scanlon DPF 2011
Region Definitions

CDF

<table>
<thead>
<tr>
<th>QCD CR1</th>
<th>EWK CR</th>
<th>QCD CR2</th>
<th>Signal region</th>
</tr>
</thead>
<tbody>
<tr>
<td>No lepton</td>
<td>At least one lepton</td>
<td>No lepton</td>
<td>No lepton</td>
</tr>
<tr>
<td>$E_T &gt; 70$ GeV/$c^2$</td>
<td>$E_T &gt; 35$ GeV/$c^2$</td>
<td>$E_T &gt; 35$ GeV/$c^2$</td>
<td>$E_T &gt; 35$ GeV/$c^2$</td>
</tr>
<tr>
<td>$\phi(j_2, E_T) \leq 0.4$</td>
<td>$\phi(j_2, E_T) &gt; 0.4$</td>
<td>$\phi(j_2, E_T) &gt; 0.4$</td>
<td>$\phi(j_2, E_T) &gt; 0.4$</td>
</tr>
<tr>
<td>$\phi(j_3, E_T) &gt; 0.4$</td>
<td>$\phi(j_3, E_T) &gt; 0.4$</td>
<td>$\phi(j_3, E_T) &gt; 0.4$</td>
<td>$\phi(j_3, E_T) &gt; 0.4$</td>
</tr>
<tr>
<td>$\phi(j_1, E_T) &gt; 1.5$</td>
<td>$\phi(j_1, E_T) &gt; 1.5$</td>
<td>$\phi(j_1, E_T) &gt; 1.5$</td>
<td>$\phi(j_1, E_T) &gt; 1.5$</td>
</tr>
<tr>
<td>$NN_{QCD} &lt; 0.45$</td>
<td></td>
<td>$NN_{QCD} &gt; 0.45$</td>
<td></td>
</tr>
</tbody>
</table>

TABLE IV: Main kinematic selection requirements for each of the control regions and the signal region.

DO

$$D = (\Delta \phi(p_T, Jet_L) + \Delta \phi(p_T, Jet_{NL}))/2$$

Signal region $D > \pi/2$
MJ model region $D < \pi/2$
### MJ Rejection Variables

#### CDF

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_T$</td>
<td>Magnitude of $E_T$</td>
</tr>
<tr>
<td>$p_T$</td>
<td>Magnitude of $p_T$</td>
</tr>
<tr>
<td>$\sqrt{\sum E_T}$</td>
<td>$E_T/H_T$</td>
</tr>
<tr>
<td>$E_T/H_T$</td>
<td>$\Delta \varphi$ between $E_T$ and $p_T$</td>
</tr>
<tr>
<td>$M(E_T, j_1, j_2)$</td>
<td>Maximum of $\Delta \varphi$ between any two jets</td>
</tr>
<tr>
<td>$\eta$ of $j_1$</td>
<td>Minimum of $\Delta \varphi$ between the $E_T$ and $j_i$</td>
</tr>
<tr>
<td>$\eta$ of $j_2$</td>
<td>Minimum of $\Delta \varphi$ between the $p_T$ and $j_i$</td>
</tr>
<tr>
<td>$\Delta \varphi(j_1, j_2)$</td>
<td>$\Delta \varphi(j_1, j_2)$ in the 2-jet rest frame</td>
</tr>
<tr>
<td>Sphericity</td>
<td></td>
</tr>
<tr>
<td>Centrality</td>
<td></td>
</tr>
</tbody>
</table>

#### D0

Variables used in the MJ DT and in the SM DT

- $\Delta \eta(j_1, j_2)$
- $\Delta \phi(j_1, j_2)$
- $\Delta R(j_1, j_2)$
- $\eta$ of $j_1$
- $\eta$ of $j_2$
- $p_T$ weighted $\Delta R(j_1, j_{all})$
- $p_T$ weighted $\Delta R(j_2, j_{all})$
- $E_T$
- $E_T$ significance
- $\Delta \phi(E_T, j_1)$
- $\Delta \phi(E_T, j_2)$
- $\Delta \phi(E_T, \text{dijet})$
- $\min \Delta \phi(E_T, j_{all})$
- $\max \Delta \phi(E_T, j_{all}) + \min \Delta \phi(E_T, j_{all})$
- $\max \Delta \phi(E_T, j_{all}) - \min \Delta \phi(E_T, j_{all})$
- $H_T$ (vectorial sum of $j_{all} p_T$)
- $H_T/H_T$ (with $H_T$ the scalar sum of $j_{all} p_T$)
- Asymmetry between $E_T$ and $H_T$
- $E_T$ component along the thrust axis
- $E_T$ component perpendicular to the thrust axis
- Sum of the signed components of the dijet and recoil momenta along the thrust axis
- Sum of the signed components of the dijet and recoil momenta perpendicular to the thrust axis
- Dijet $p_T$
- Scalar sum of $j_1$ and $j_2$ $p_T$
- Centrality (ratio of the scalar sum of $j_1$ and $j_2$ $p_T$ to the sum of $j_1$ and $j_2$ energy)
- Effective mass (sum of $E_T$ and of the scalar sum of $j_1$ and $j_2$ $p_T$)
- $\theta$ angle of $j_1$ boosted to the dijet rest frame
- $\theta$ angle of the dijet system
- Polar angle of $j_1$ boosted to the dijet rest frame with respect to the dijet direction in the laboratory
- Azimuthal angle of $j_1$ boosted to the dijet rest frame with respect to the dijet direction in the laboratory
- Color flow $j_1$
- Color flow $j_2$

---

Tim Scanlon DPF 2011
**Final MVA Variables**

**D0**

**MJ DT Variables plus**

<table>
<thead>
<tr>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dijet mass</td>
</tr>
<tr>
<td>Dijet transverse mass</td>
</tr>
<tr>
<td>$j_1$ ( p_T )</td>
</tr>
<tr>
<td>$j_2$ ( p_T )</td>
</tr>
<tr>
<td>$H_T$</td>
</tr>
<tr>
<td>$j_1$ b-tagging output</td>
</tr>
<tr>
<td>$j_2$ b-tagging output</td>
</tr>
</tbody>
</table>

**CDF**

<table>
<thead>
<tr>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invariant mass of the two leading jets in the event ( M_{jj} )</td>
</tr>
<tr>
<td>Invariant mass of ( \not{E}_T ), ( \not{j}_1 ) and ( \not{j}_2 )</td>
</tr>
<tr>
<td>Difference between the scalar sum of transverse energy of the jets ( H_T ) and ( \not{E}_T )</td>
</tr>
<tr>
<td>Difference between the vector sum of transverse energy of the jets ( H_T ) and ( \not{E}_T )</td>
</tr>
<tr>
<td>The output of the TRACKMET neural network</td>
</tr>
<tr>
<td>Maximum of the difference in the ( \eta - \phi ) space between the directions of two jets, taking two jets at the time</td>
</tr>
<tr>
<td>The output of ( NN_{QCD} )</td>
</tr>
</tbody>
</table>

**TABLE VI:** Input variables to the final discriminant neural network.