Improving Vector Boson Fusion (VBF) LHC Higgs Analyses with Fox-Wolfram Moments

Dr. Catherine Bernaciak - ITP Universität Heidelberg

High Energy Physics Seminar, Department of Physics University of Virginia, October 2, 2013
Outline

• Review of Standard Model Higgs Mechanism (4)
• Phenomenology of Standard Model Higgs (5)
• Higgs-like Boson Measurement at the LHC (2)
• Fox - Wolfram Moments (8)
• Results of Cut-Based and Boosted Decision Tree Analyses (16)
The Standard Model of Particle Physics

fermions = 6 quarks + 6 leptons

bosons = W, Z, photon, gluon, Higgs

impacts mass to W, Z, quarks, leptons
Electroweak Symmetry Breaking in the Standard Model - QED as a toy model

**QED**:

\[
L = -\frac{1}{4} (\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu)
\]

**U(1) gauge transformation**:

\[A_\mu(x) \rightarrow A_\mu(x) - \partial_\mu \eta(x)\]

adding mass to photon field:

\[
L = -\frac{1}{4} (\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) + \frac{1}{2} m^2 A_\mu A^\mu
\]

violates gauge invariance

a simple, realistic solution:

\[
L = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + |D_\mu \phi|^2 - V(\phi)
\]

a new complex, scalar field
Electroweak Symmetry Breaking in the Standard Model - QED as a toy model

\[ L = -\frac{1}{4} F_{\mu \nu} F^{\mu \nu} + |D_\mu \phi|^2 - V(\phi) \]

covariant derivative: \[ D_\mu = \partial_\mu - ieA_\mu \]

simplest, renormalizable, \( U(1) \) invariant potential

\[ V(\phi) = \mu^2 |\phi|^2 + \lambda(|\phi|^2)^2 \]

\[ \phi \rightarrow e^{-i\eta(x)} \phi(x) \]

\[ \langle \phi \rangle = \sqrt{-\frac{\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}} \]

\[ \mu^2 < 0 \quad \mu^2 > 0 \quad \langle \phi \rangle = 0 \]
\[ \langle \phi \rangle = \sqrt{-\frac{\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}} \]

\[ U(1) \text{ symmetry is broken with nonzero vev} \]

\[ \phi \equiv \frac{1}{\sqrt{2}} e^{i \frac{\chi}{v}} (v + h) \]

expand into of non-vev fields

and mass is acquired:

\[ L = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - evA_\mu \partial^\mu \chi + \frac{e^2 v^2}{2} A_\mu A^\mu + \cdots \]

photon field with \[ M_A = ev \]

same principle, when applied to electroweak theory causes weak bosons to acquire mass - Higgs field emerges as physical particle....
SM Higgs Interactions

- Once $m_H$ is known, couplings can be measured and compared to SM prediction.
SM Higgs Production at the LHC

the gluon fusion channel - main LHC production mechanism

"gluon fusion" ggf

\[ \sigma(gg \rightarrow H) \approx 15 \text{ pb at 7 TeV} \]

\[ \sigma(gg \rightarrow H) \approx 50 \text{ pb at 14 TeV} \]

for \( M_H = 125 \text{ GeV} \)

why?! more likely to find a gluon in the proton

bottom loop suppressed by \( \sim 0.1\% \) - lighter quark loops even less likely

H1 and ZEUS HERA I+II PDF Fit

\[ xg (\times 0.05) \]

\[ xS (\times 0.05) \]

HERAPDF1.5 NNLO (prel.)

exp. uncert.

model uncert.

parametrization uncert.

HERAPDF1.5 (prel.)

\[ Q^2 = 10000 \text{ GeV}^2 \]

HERAPDF Structure Function Working Group March 2011
SM Higgs Production at the LHC

Vector Boson Fusion
essential probe of EW higgs couplings - deviations from predicted rates could indicate BSM higgs physics

\[ \sigma(qqH) \approx 1.3 \text{ pb at } 7 \text{ TeV} \]
\[ \sigma(qqH) \approx 4 \text{ pb at } 14 \text{ TeV} \]
for \( M_H = 125 \text{ GeV} \)

distinctive “forward - backward” jet topology unlike any background processes

lack of central jet activity - handle for discerning from backgrounds
**SM Higgs Production at the LHC**

**solution:**
apply acceptance criteria on events to disfavor ggH + 2jet kinematics

\[
p_T^{j_1j_2} > 20 \text{ GeV} \\
\eta_{j_1} \cdot \eta_{j_2} < 0 \\
\Delta \eta_{j_1, j_2} > 4
\]

for after applying VBF selection, ggH events contribute only 4% - 5% to Higgs production

however, ggH + 2jet production could mimic VBF production
Recently the NLO QCD corrections to the NLO code for the signal process are not expected to change much. The inclusive partonic cross section for associated production depends on the final-state Higgs-boson decay.

The full next-to-leading-order (NLO) QCD corrections to SM Higgs production & decay are computed by convoluting it with the parton densities in the usual way. The remaining contributions. The hadronic cross section is obtained by more than 30% with momentum $k$.

For all search channels previously mentioned, a precise prediction of the signal cross section and its background might be extracted. These active-Higgs approximation which have not been re-optimized for the signal process are approximates Higgs radiation as a fragmentation process in D. Wackeroth.

Examples of LO Feynman diagrams for the partonic processes $\text{WH,ZH}$ for $\sqrt{s} = 90$ and 125 GeV at 7 TeV and 14 TeV.

The LO prediction for $\text{ttH}$ process has been calculated at 14 TeV $\sqrt{s}$ about $95\%$ and is fully analogous to the Drell–Yan expression. The resulting cross sections were computed a long time ago and are always positive and the factor varies between $611$ fb at 14 TeV and $\approx 0.6$ pb at 7 TeV for $M_H = 125$ GeV.

The LO prediction for the $\text{ttH}$ process has been calculated at 14 TeV $\sqrt{s}$ about $95\%$ and is fully analogous to the Drell–Yan expression. The resulting cross sections were computed a long time ago and are always positive and the factor varies between $611$ fb at 14 TeV and $\approx 0.6$ pb at 7 TeV for $M_H = 125$ GeV.

For all search channels previously mentioned, a precise prediction of the signal cross section and its background might be extracted. These active-Higgs approximation which have not been re-optimized for the signal process are approximates Higgs radiation as a fragmentation process in D. Wackeroth.

Examples of LO Feynman diagrams for the partonic processes $\text{WH,ZH}$ for $\sqrt{s} = 90$ and 125 GeV at 7 TeV and 14 TeV.

The LO prediction for $\text{ttH}$ process has been calculated at 14 TeV $\sqrt{s}$ about $95\%$ and is fully analogous to the Drell–Yan expression. The resulting cross sections were computed a long time ago and are always positive and the factor varies between $611$ fb at 14 TeV and $\approx 0.6$ pb at 7 TeV for $M_H = 125$ GeV.
SM Higgs Decay at the LHC

- $H \rightarrow b\bar{b}$: high decay rate, but $b$ tagging efficiency at 60%
- $WW \rightarrow l\nu l\nu$: clean and efficient, background under control
- $\tau \rightarrow l\nu l\nu$: clean and efficient, background under control

Branching ratios vs. $M_H$ [GeV]

- $\gamma \gamma$ (purple)
- $Z\gamma$ (black)
- $gg$ (blue)
- $cc$ (red)
- $\tau\tau$ (green)

LHC Higgs XS WG 2010
LHC Higgs-like Boson Discovery

combined mass measurement:

**ATLAS:**
\[ m_H = 125.5 \pm 0.2 \text{(stat)} \pm 0.5 \text{(sys)} \text{ GeV} \]

**CMS:**
\[ m_H = 125.3 \pm 0.4 \text{(stat)} \pm 0.5 \text{(sys) GeV} \]

combined signal strength measurement:

**ATLAS:**
\[ \mu = 1.33 \pm 0.14 \text{(stat)} \pm 0.15 \text{(sys)} \]

**CMS:**
\[ \hat{\mu} = \frac{\sigma}{\sigma_{SM}} = 0.87 \pm 0.23 \]

for \( M_H = 125 \text{ GeV} \)

consistent with SM Higgs hypothesis

\[ \mu = 0 \]
background only hypothesis

\[ \mu = 1 \]
SM Higgs boson hypothesis

\begin{tabular}{|c|c|}
\hline
\textbf{ATLAS} & \textbf{m}_H = 125.5 \text{ GeV} \\
\hline
\text{H} \rightarrow \gamma \gamma & \mu = 1.55^{+0.33}_{-0.31} \pm 0.21 \\
\text{Low } p_T & \mu = 1.6^{+0.4}_{-0.4} \\
\text{High } p_T & \mu = 1.7^{+0.7}_{-0.6} \\
2 \text{jet high mass (VBF)} & \mu = 1.9^{+0.8}_{-0.6} \\
VH \text{ categories} & \mu = 1.3^{+1.3}_{-1.1} \\
\hline
\text{H} \rightarrow Z Z^* \rightarrow 4 \ell & \mu = 1.43^{+0.46}_{-0.36} \pm 0.17 \\
\text{VBF--VW-like categories} & \mu = 1.2^{+0.9}_{-0.8} \\
\text{Other categories} & \mu = 1.45^{+0.49}_{-0.36} \pm 0.35 \\
\hline
\text{H} \rightarrow WW^* \rightarrow l
\bar{l}v
\bar{v} & \mu = 0.99^{+0.31}_{-0.28} \pm 0.21 \\
0+1 \text{ jet} & \mu = 0.82^{+0.38}_{-0.30} \pm 0.22 \\
2 \text{ jet VBF} & \mu = 1.4^{+0.7}_{-0.6} \pm 0.5 \\
\text{Comb. }H \rightarrow \gamma \gamma, ZZ^*, WW^* & \mu = 1.33^{+0.21}_{-0.18} \pm 0.15 \\
\hline
\end{tabular}
Theoretical Uncertainties in Higgs Measurement

large systematic uncertainty from higher order QCD calculations matched to parton shower - common to both ATLAS and CMS

### ATLAS

<table>
<thead>
<tr>
<th>Source (theory)</th>
<th>Uncertainty (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QCD scale</td>
<td>±8 (ggF), ±1 (VBF, VH), +4/ -9 (ttH)</td>
</tr>
<tr>
<td>PDFs + $\alpha_s$</td>
<td>±8 (ggF, ttH), ±4 (VBF, VH)</td>
</tr>
</tbody>
</table>

### CMS

<table>
<thead>
<tr>
<th>Source</th>
<th>Range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated luminosity</td>
<td>2.2-4.4</td>
</tr>
<tr>
<td>Lepton identification and trigger efficiency (per lepton)</td>
<td>3</td>
</tr>
<tr>
<td>Z($\nu\nu$)H triggers</td>
<td>2</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>2-3</td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td>3-6</td>
</tr>
<tr>
<td>Missing transverse energy</td>
<td>3</td>
</tr>
<tr>
<td>b-tagging efficiency</td>
<td>3-15</td>
</tr>
<tr>
<td>Signal cross section (scale and PDF)</td>
<td>4</td>
</tr>
<tr>
<td>Signal cross section ($p_T$ boost, EWK/QCD)</td>
<td>5-10/10</td>
</tr>
<tr>
<td>Statistical precision of signal simulation</td>
<td>1-5</td>
</tr>
<tr>
<td>Backgrounds estimated from data</td>
<td>10</td>
</tr>
<tr>
<td>Backgrounds estimated from simulation</td>
<td>30</td>
</tr>
</tbody>
</table>

include more EW and QCD higher-order corrections, resum EW Sudakov logs in VHbb ...

better match parton shower to existing NLO and NNLO and implement in simulation tool SHERPA, MC@NLO, MADGRAPH ...

---

1 PLB 726 (2013) 88-119, 2 JHEP 06 (2013) 081
The Fox-Wolfram Moments\textsuperscript{1}

a rotationally invariant set of observables constructed from Legendre polynomials

\[ H_\ell = \sum_{i,j} \frac{|\vec{p}_i| |\vec{p}_j|}{s} P_\ell(\cos \Omega_{ij}) \]

correlations between hadrons, jets, calorimeter entries...

weight factor

total angle between objects

\[ \cos \Omega_{ij} = \cos \theta_i \cos \theta_j + \sin \theta_i \sin \theta_j \cos(\phi_i - \phi_j) \]

\textsuperscript{1}Fox, Wolfram, PRL 1978
Legendre Polynomials

occur as series solution to Laplace’s equation in spherical coordinates

\[
\frac{d}{dx} \left[ (1 - x^2) \frac{d}{dx} P_n(x) \right] + n(n + 1) P_n(x) = 0
\]

\[
P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} \left[ (x^2 - 1)^n \right]
\]

\[
P_0(x) = 1 , P_1(x) = x
\]

\[
P_2(x) = \frac{1}{2} (3x^2 - 1)
\]

\[
P_3(x) = \frac{1}{2} (5x^3 - 3x)
\]

\[
\vdots
\]

\[
P_7(x) = \frac{1}{16} (429x^7 - 693x^5 + 315x^3 - 35x)
\]
The Fox-Wolfram Moments
an event shape observable describing correlations between four-momentum objects

✦ e+ e- to jets
Fox, Wolfram  Nucl. Phys. B 149 (1979) 413-496

✦ Top Quark signal at Tevatron
Field, Kanev, Tayebnejad PRD 55, 9 (1997)

✦ B meson decays at Belle:
Toru Iijima, hep-ex 0105005 (2001)

✦ Higgs physics at the LHC: VBF H tautau vs Z+2j and Top Pair
C.B., Buschmann, Butter, Plehn PRD 87, 073014 (2013)

✦ A Multivariate study of Fox-Wolfram Moments for Higgs Analyses at the LHC
C.B., Mellado, Plehn, Ruan, Schichtel, in preparation
The Fox-Wolfram Moments

\[ H_\ell = \sum_{i,j} \frac{|\vec{p}_i||\vec{p}_j|}{s} P_\ell(\cos \Omega_{ij}) \]

\[ 0 \leq H_\ell \leq 1 \]

\[ W^{T}_{ij} = \frac{p_T i p_T j}{p_T^2,\text{tot}} \]
\[ W^U_{ij} = 1 \]
\[ W^p_{ij} = \frac{|\vec{p}_i||\vec{p}_j|}{|\vec{p}|^2,\text{tot}} \]

- Transverse momentum weight
- Unit weight
- Magnitude momentum weight

"weight factor"
Fox-Wolfram Moments - 2 jet properties

\[ H_\ell = \sum_{i,j=1}^{2} \frac{W_i W_j}{W_{\text{tot}}^2} P_\ell(\cos \Omega_{ij}) \]

\[ = \frac{1}{(W_1 + W_2)^2} \left[ W_1^2 P_\ell(\cos 0) + W_2^2 P_\ell(\cos 0) + W_1 W_2 P_\ell(\cos \Omega_{12}) \right] \]

\[ = 1 + \frac{2W_1 W_2}{(W_1 + W_2)^2 P_\ell(\cos \Omega_{12})} \]

\[ = \frac{1 + 2r P_\ell(\cos \Omega_{12})}{1 + 2r + r^2} \]

\[ r = \frac{W_2}{W_1} \]

\[ 0 \leq r \leq 1 \]
Fox-Wolfram moments - 2 jet properties

odd moments - best for discriminating back-to-back jets, higher moments resolve larger angular $j_1 - j_2$ separation

$$r = \frac{W_2}{W_1}$$

multivalued function, no resolution to intermediate values of $\Omega_{12}$

$$H_\ell \to 0 \quad \text{for} \quad \Omega_{12} \to \pi$$
Fox-Wolfram moments - 2 jet properties

even moments - symmetry of even function reduces discriminatory power

\[ H_\ell \to 1 \quad \text{for} \quad \Omega_{12} \to 0 \quad \text{AND} \quad \Omega_{12} \to \pi \]

low, even moments may discern non forward-backward jets
Analysis for $H \rightarrow \tau \tau$

$(\text{process} + \text{hard jet}) \times \text{PS with CKKW using SHERPA}$

**Signal WBF**

**Background**
- QCD $ZJJ$
- Top Pair

**Fastjet anti-$k_T$ algorithm with $R = 0.4, 8\text{TeV}$**
## Cutflow Analysis

<table>
<thead>
<tr>
<th>acceptance</th>
<th>WBF + 1 jet</th>
<th>QCD ZJJ</th>
<th>Top Pair</th>
<th>S / B</th>
</tr>
</thead>
<tbody>
<tr>
<td>% fail</td>
<td>XS (fb)</td>
<td>% fail</td>
<td>XS (fb)</td>
<td></td>
</tr>
<tr>
<td>$pT_{j_1,j_2} &gt; 20$ GeV</td>
<td>29.4</td>
<td>18.7</td>
<td>115000</td>
<td>17200</td>
</tr>
<tr>
<td>$</td>
<td>y_{j_1,j_2}</td>
<td>&lt; 5.0$</td>
<td>1.49</td>
<td>13.0</td>
</tr>
<tr>
<td>$\Delta R_{j_1,j_2} &gt; 0.7$</td>
<td>2.73</td>
<td>12.6</td>
<td>0.97</td>
<td>7740</td>
</tr>
<tr>
<td>$m_{j_1,j_2} &gt; 600$ GeV</td>
<td>68.9</td>
<td>3.92</td>
<td>3.84</td>
<td>7440</td>
</tr>
<tr>
<td>$b$ - veto</td>
<td>NA</td>
<td>3.92</td>
<td>NA</td>
<td>253</td>
</tr>
<tr>
<td>$y_1 \cdot y_2 &lt; 0$</td>
<td>1.41</td>
<td>3.86</td>
<td>9.17</td>
<td>230</td>
</tr>
<tr>
<td>$</td>
<td>y_{j_1} - y_{j_2}</td>
<td>&gt; 4.4$</td>
<td>13.9</td>
<td>3.32</td>
</tr>
</tbody>
</table>

*can cuts on FWM replace or be added to current cuts used for VBF event selection?*

Wednesday, October 2, 13
Cuts on FWM Distributions

\[ \Omega_{12} \gtrsim 0.8\pi, \quad r \gtrsim 0.4 \]

\[ \Omega_{12} \sim 0.6\pi, \quad r \gtrsim 0.4 \]
Cuts on FWM Distributions

OR

\[ \Omega_{12} \sim 0, \pi \quad \text{any} \quad r \]
\[ r \leq 0.3 \quad \text{any} \quad \Omega_{12} \]
Cuts on FWM Distributions

\[ \Omega_{12} \sim 0, \pi \quad \text{any} \quad r \]
\[ r \leq 0.3 \quad \text{any} \quad \Omega_{12} \]

WBF + 1 jet
QCD ZJJ
Top Pair + 1 jet
## Cuts on FWM Distributions\(^1\)

<table>
<thead>
<tr>
<th>acceptance</th>
<th>WBF + 1 jet</th>
<th>QCD ZJJ</th>
<th>Top Pair</th>
<th>S / B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% fail</td>
<td>XS (fb)</td>
<td>% fail</td>
<td>XS (fb)</td>
</tr>
<tr>
<td>min cuts + b-veto</td>
<td>3.92</td>
<td>253</td>
<td>292</td>
<td>1/139</td>
</tr>
<tr>
<td>(H_T^3 &lt; 0.3)</td>
<td>38.4</td>
<td>2.41</td>
<td>44.4</td>
<td>141</td>
</tr>
<tr>
<td>(H_T^4 &gt; 0.8)</td>
<td>35.8</td>
<td>2.52</td>
<td>48.1</td>
<td>131</td>
</tr>
<tr>
<td>(H_T^8 &gt; 0.8)</td>
<td>50.1</td>
<td>1.96</td>
<td>60.5</td>
<td>100</td>
</tr>
<tr>
<td>(H_T^{12} &gt; 0.7)</td>
<td>64.5</td>
<td>1.39</td>
<td>73.0</td>
<td>68.3</td>
</tr>
<tr>
<td>rapidity gap</td>
<td>13.9</td>
<td>3.32</td>
<td>31.8</td>
<td>157</td>
</tr>
</tbody>
</table>

\(^1\)C.B. et.al, PRD 87, 073014 (2013)
### Analysis - Cutting on FWM

After typical WBF cuts are exhausted, can the moments help?

<table>
<thead>
<tr>
<th>acceptance</th>
<th>WBF + 1 jet</th>
<th>QCD ZJJ</th>
<th>Top Pair</th>
<th>S / B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% fail</td>
<td>XS (fb)</td>
<td>% fail</td>
<td>XS (fb)</td>
</tr>
<tr>
<td>minimal cuts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ b veto</td>
<td>NA</td>
<td>3.92</td>
<td>NA</td>
<td>253</td>
</tr>
<tr>
<td>central jet cuts</td>
<td>13.9</td>
<td>3.32</td>
<td>31.8</td>
<td>157</td>
</tr>
</tbody>
</table>

Top pair background can be further suppressed based on tagging jet correlations rephrased into FWM.

\[ H_{12}^T > 0.7 \]

1/57

Wednesday, October 2, 13
**Inclusive FWM:**

require *at least 2 tagging jets* satisfying minimal cuts

more power to discern WBF from ZJJ (3rd and higher jets have more drastically differing weights)

\[ H_\ell < 0.3 \text{ region populated} \]
**Classification Rule**

A “classifier” is a rule for determining which class an instance of a set belongs to.

- **sig/bkg**
- **event**
- **data or MC sample**

### Classification Table

<table>
<thead>
<tr>
<th>instance</th>
<th>$y_{j1}$</th>
<th>$y_{j2}$</th>
<th>$\Delta y_{12}$</th>
<th>$m_{12}$</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>event 1</td>
<td>2.79854</td>
<td>-1.33015</td>
<td>4.12869</td>
<td>264.056 GeV</td>
<td>S</td>
</tr>
<tr>
<td>event 2</td>
<td>1.5059</td>
<td>-1.09764</td>
<td>2.60354</td>
<td>156.285 GeV</td>
<td>B</td>
</tr>
<tr>
<td>event n</td>
<td>-1.10029</td>
<td>1.83929</td>
<td>2.93958</td>
<td>209.104 GeV</td>
<td>S</td>
</tr>
</tbody>
</table>

$y(\bar{x}) \in \mathbb{R}$ map all information of an event onto a real number - the “scalar output” of the classifier
Classification Response and ROC Curves

**rejection:**
- discard events

\[ r_s = 1 - \varepsilon_s \]
\[ r_b = 1 - \varepsilon_b \]

**efficiency:**
- keep events

\[ \varepsilon_s = \frac{S_1}{S_{\text{tot}}} \]
\[ \varepsilon_b = \frac{B_1}{B_{\text{tot}}} \]

\( \frac{dN}{dy(x)} \)

Each point on curve corresponds to a cut on classifier response

- if \( y_{\text{cut}} = 1 \)
  - \( (\varepsilon_s, r_b) = (0, 1) \)
  - ROC curve (Receiver Operating Characteristic)

- if \( y_{\text{cut}} = 0 \)
  - \( (\varepsilon_s, r_b) = (1, 0) \)

\( y(x) \)

\( y_{\text{cut}} \)
Boosted Decision Trees

each tree is a classification rule

boosting: combine trees into single rule

Adaptive Boost Algorithm:

\[ y(\vec{x}) = \frac{1}{N_{\text{boost}}} \sum_{i}^{N_{\text{boost}}} \ln(\alpha_i) h_i(\vec{x}) \]

events misclassified are reweighted, another tree is built, misclassification rate is updated, event is reweighted, etc...
Analysis for $H \rightarrow \text{diphoton}$

$(\text{process + hard jet}) \times \text{PS with CKKW using SHERPA}$

Signal VBF + 1 matrix element level jet

Background diphoton + 2 matrix element level jets

Fastjet anti-$k_T$ algorithm with $R = 0.4, 8\text{TeV}$
BDT Analysis with only Tagging Jet Correlations

use FWM after applying acceptance criteria for jets:

\[ p_{Tj} > 25 \text{ GeV for } |y_j| < 2.4 \]
\[ p_{Tj} > 30 \text{ GeV for } 2.4 \leq |y_j| < 4.5 \]
\[ |\Delta y_{j_1j_2}| \geq 2 \text{ and } m_{j_1j_2} > 150 \text{ GeV} \]

compare FWM with tagging jet correlations used by ATLAS

\[ \{ m_{j_1j_2}, y_{j_1}, y_{j_2}, \Delta y_{j_1j_2} \} \]

Decision Tree Settings\(^1\):

\[ N_{\text{train}}, N_{\text{test}} = 100K, 50K \]
\[ N_{\text{trees}}, N_{\text{layers}} = 400, 3 \]

\(^1\)Hoecker et.al., Toolkit for Multivariate Analysis, http://tmva.sourceforge.net
Results of BDT Analysis Including FWM

\[
\frac{S}{\sqrt{S + B}} = 198.7
\]

for cut at \( y = -0.14 \)

default ATLAS training variables

Background rejection versus Signal efficiency

88% bkg rejection at 40% sig efficiency

TMVA overtraining check for classifier: BDTA

Kolmogorov-Smirnov test: signal (background) probability = 1 (0.433)
Results of BDT Analysis Including FWM

in addition to default, train with:

\[ H_{\ell, \phi}^x = \sum_{i,j=1}^{N} W_{i,j}^x P_{\ell}(\cos \Delta \phi_{ij}) \]

\[ H_{1, \phi}^U = \frac{1}{2} + \frac{1}{4} \cos \Delta \phi_{12} \]

\[ H_{1, \phi}^T = \frac{p_{T1}^2 + p_{T2}^2}{p_{T_{tot}}^2} + \frac{p_{T1}p_{T2}}{p_{T_{tot}}^2} \cos \Delta \phi_{12} \]

<table>
<thead>
<tr>
<th></th>
<th>B rejection</th>
<th>( \frac{S}{\sqrt{S+B}} )</th>
<th>improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATLAS default</td>
<td>88.7%</td>
<td>198.7 (-0.14)</td>
<td></td>
</tr>
<tr>
<td>( H_{1, \phi}^T, H_{1, \phi}^U )</td>
<td>95.2%</td>
<td>209.166 (-0.07)</td>
<td>5.3%</td>
</tr>
<tr>
<td>( H_{1, \phi}^T )</td>
<td>94.9%</td>
<td>206.703 (-0.08)</td>
<td>4.0%</td>
</tr>
<tr>
<td>( H_{1, \phi}^U )</td>
<td>95.2%</td>
<td>208.821 (-0.08)</td>
<td>5.1%</td>
</tr>
<tr>
<td>( \cos \Delta \phi_{12} )</td>
<td>95.2%</td>
<td>208.821 (-0.08)</td>
<td>5.1%</td>
</tr>
</tbody>
</table>

1Bernaciak, Mellado, Plehn, Ruan, Schichtel, in preparation
## Results of BDT Analysis Including FWM

**improvement with redefinition of FWM:**

\[
H^x,\phi = \sum_{i,j=1}^{N} W_{ij} x P_\ell(\cos \Delta \phi_{ij})
\]

<table>
<thead>
<tr>
<th></th>
<th>$B$ rejection</th>
<th>$\frac{S}{\sqrt{S + B}}$</th>
<th>improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATLAS default</td>
<td>88.7%</td>
<td>198.7 (-0.14)</td>
<td></td>
</tr>
<tr>
<td>$H^T,\phi_1 \rightarrow H^T,\phi_2, H^U,\phi_1 \rightarrow H^U,\phi_2$</td>
<td>95.0%</td>
<td>208.901 (-0.07)</td>
<td>5.1%</td>
</tr>
<tr>
<td>$H^T,\phi_1, H^T,\phi_3, H^U,\phi_1, H^U,\phi_3$</td>
<td>95.3%</td>
<td>209.115 (-0.08)</td>
<td>5.3%</td>
</tr>
<tr>
<td>$H^T,\phi_1, H^T,\phi_2, H^U,\phi_1, H^U,\phi_2$</td>
<td>95.2%</td>
<td>209.132 (-0.08)</td>
<td>5.3%</td>
</tr>
<tr>
<td>$H^T,\phi_1, H^U,\phi_1$</td>
<td>95.2%</td>
<td>209.166 (-0.07)</td>
<td>5.3%</td>
</tr>
<tr>
<td>$H^T,\phi_1$</td>
<td>94.9%</td>
<td>206.703 (-0.08)</td>
<td>4.0%</td>
</tr>
<tr>
<td>$H^U,\phi_1$</td>
<td>95.2%</td>
<td>208.821 (-0.08)</td>
<td>5.1%</td>
</tr>
<tr>
<td>$\cos \Delta \phi_{12}, W^T_{12}$</td>
<td>95.3%</td>
<td>209.299 (-0.08)</td>
<td>5.3%</td>
</tr>
<tr>
<td>$\cos \Delta \phi_{12}$</td>
<td>95.2%</td>
<td>208.821 (-0.08)</td>
<td>5.1%</td>
</tr>
</tbody>
</table>

**redefinition of FWM offer modest improvement over ATLAS default variables**

---

Bernaciak, Mellado, Plehn, Ruan, Schichtel, in preparation
Conclusions - Future Work

✦ FWM suitable for both cut-based and decision tree analysis - offer consistent 5% improvement for azimuthal angle definition

✦ combinations of U and T weighted moments are better than T alone, U may be sufficient alone

✦ total angle moments - offer 1% improvement - need to understand why

✦ the FWM are an interesting addition to the variables currently used in Higgs analyses

Work Underway

✦ compare with Neural Network MVA

✦ incorporate 3rd jet and its scale uncertainty into this analysis

✦ can moments be used as a modified jet veto?
BACKUP SLIDES
Gauge Invariance, Briefly

A key aspect of any realistic field theory description of matter is **gauge invariance**.

The Lagrangian remains unchanged under a **local change of coordinate system**:

- Constructed from **gauge fields**.
- Spatial translation or rotation, internal field transformation:
  
  \[
  \psi \rightarrow e^{i\theta} \psi \\
  \psi(x^\mu) \rightarrow \psi(\Lambda^{\mu\nu} x_\nu + a^\mu)
  \]
Deciding Splitting Variables

avg. info needed to identify a class in T if it’s partitioned into 2 subsets:

\[ \text{info}_X(T) = \frac{|T_1|}{|T|} \text{info}(T_1) + \frac{|T_2|}{|T|} \text{info}(T_2) \]

information gain obtained by a particular test:

\[ \text{gain}(X) = \text{info}(T) - \text{info}_X(T) \]

variable 1

\( v_1 \)
\( v_2 \)
\( v_3 \)
\( v_4 \)
\( \ldots \)
\( v_m \)

- test 1: split variable 1
- test 2: split variable 1
- test \( m-1 \): split variable 1
- test \( m \): split variable 1

\( \text{var} > \text{threshold} \)

- \( T_1 \)
- \( T_2 \)

\( > v_2 \)
\( \leq v_2 \)
\( > v_3 \)
\( \leq v_3 \)

repeat for all variables: test with largest gain ratio becomes root node ...

repeat for subsequent nodes
Information Entropy

### Table

<table>
<thead>
<tr>
<th>instance</th>
<th>$y_{j1}$</th>
<th>$y_{j2}$</th>
<th>$\Delta y_{12}$</th>
<th>$m_{12}$</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>event 1</td>
<td>2.79854</td>
<td>-1.33015</td>
<td>4.12869</td>
<td>264.056 GeV</td>
<td>S</td>
</tr>
<tr>
<td>event 2</td>
<td>1.5059</td>
<td>-1.09764</td>
<td>2.60354</td>
<td>156.285 GeV</td>
<td>B</td>
</tr>
<tr>
<td>event n</td>
<td>-1.10029</td>
<td>1.83929</td>
<td>2.93958</td>
<td>209.104 GeV</td>
<td>S</td>
</tr>
</tbody>
</table>

### Probability of Finding an Event Belonging to S or B in the Entire Set T

$$P_{S,B} = \frac{N_{S,B}}{N_{\text{tot}}}$$

### Information Entropy (General)

$$I_E = \log_2(P_{S,B}) \text{ bits}$$

### Information Entropy of T

$$\text{info}(T) = -P_S \log_2(P_S) - P_B \log_2(P_B)$$

“avg. amount of info needed to identify the class of an event in T”
The Fox-Wolfram moments - brief history

an event shape observable describing correlations between four-momentum objects

e+ e- to jets

Fox, Wolfram Nucl. Phys. B 149 (1979) 413-496

Top Quark signal at Tevatron

Field, Kanev, Tayebnejad PRD 55, 9 (1997)

B meson decays at Belle:

Toru Iijima, hep-ex 0105005 (2001)

Higgs physics at the LHC: WBF H tautau vs Z+2j and Top Pair

C.B., Buschmann, Butter, Plehn PRD 87, 073014 (2013)
Classify Rules

\[
\begin{array}{cccc}
 y_{j_1} & y_{j_2} & \Delta y_{12} & m_{12} \\
 2.79854 & -1.33015 & 4.12869 & 264.056 \text{ GeV} \\
 1.5059 & -1.09764 & 2.60354 & 156.285 \text{ GeV} \\
 \vdots & \vdots & \vdots & \vdots \\
 -1.10029 & 1.83929 & 2.93958 & 209.104 \text{ GeV} \\
\end{array}
\]

- \( y(x) \in \mathbb{R} \)

Scalar Output Response Curve

- ROC curve (Receiver Operating Characteristic)

- A classification rule seeks to do better than random guessing, which is correct 50% of the time.
The Fox-Wolfram moments - 2 jet visualization

**odd moments** - best for discriminating back-to-back jets, higher moments resolve larger angular $j_1 \parallel j_2$ separation

$$H_\ell \rightarrow 0 \quad \text{for} \quad \Omega_{12} \rightarrow \pi$$

multivalued function, no resolution to intermediate values of $\Omega_{12}$

Wednesday, October 2, 13
The Fox-Wolfram moments - 2 jet visualization

**even moments** - symmetry of even function reduces discriminatory power

\[ H_\ell \rightarrow 1 \quad \text{for} \quad \Omega_{12} \rightarrow 0 \quad \text{AND} \quad \Omega_{12} \rightarrow \pi \]