Top-tagging at high jet multiplicity

Sebastian Schätzel
Universität Heidelberg, Germany

*Using Jet Substructure Workshop*
University of Oregon, Eugene, USA
23-26 April 2013

- HEPTopTagger
- fat jet quantities
  - mass
  - $p_T$
- multiplicity of substructure objects

Thanks to Tilman Plehn and Michael Spannowsky for suggestions
The past

http://de.arxiv.org/abs/1211.2202

clean back-to-back ttbar events
High jet multiplicity

4 tops from gluino pairs

benchmark model:

<table>
<thead>
<tr>
<th></th>
<th>mass (GeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{g}$</td>
<td>1300</td>
</tr>
<tr>
<td>$\tilde{t}$</td>
<td>2500</td>
</tr>
<tr>
<td>LSP</td>
<td>100</td>
</tr>
</tbody>
</table>
Top Taggers

<table>
<thead>
<tr>
<th>prong-based</th>
<th>fat jet shapes</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 3 subjets in fat jet substructure</td>
<td>• mass and $p_T$ of fat jet</td>
</tr>
<tr>
<td>• kinematics compatible with $t \rightarrow bW \rightarrow bq$</td>
<td>• N-subjettiness</td>
</tr>
<tr>
<td>example: HEPTopTagger (mass windows)</td>
<td>• ...</td>
</tr>
</tbody>
</table>

performance at high jet multiplicity?
The setup

- signal: Herwig++ 2.5.2
- background: QCD dijets, PYTHIA 8.175
- Delphes (0.1x0.1 cells)
- $E_{\text{CM}} = 14 \text{ TeV}$
- distributions scaled to 10 fb$^{-1}$
The jet environment – SUSY

akt04 jets

events / 10 GeV

p_T (GeV)

jet 1
jet 2
jet 3
jet 4
jet 5
jet 6
jet 7
jet 8
jet 9
jet 10

akt04 jet multiplicity

Mean 7.59

events
The jet environment – QCD

akt04 jets

akt04 jet multiplicity

Mean 0.02687
Interlude: Pythia

- practical reason for using it: 8M events at hand
- for multi-jets better use multi-leg generator
- take efficiencies in present study with grain of salt
- final numbers from data
N-Subjettiness of QCD jets without pile-up

- 2010 data, L1 calo trigger
- exactly 1 primary vertex (≥5 tracks w/ p_T > 150 MeV)
- corrected for detector effects

\[ \tau_{21} = \frac{\tau_2}{\tau_1} \]

\[ \tau_{32} = \frac{\tau_3}{\tau_2} \]

PYTHIA tune: AMBT1

- more indication for 2 subjets than for 3
- description within systematic errors, trend to underestimate substructure

±10% if no pile-up

S.S., BOOST2012
N-Subjettiness of inclusive QCD jets

\[ \tau_{21} = \frac{\tau_2}{\tau_1} \]

\[ \tau_{32} = \frac{\tau_3}{\tau_2} \]

with PU: up to 40% off, more in tails this is at low multiplicity!

POWHEG slightly better

- more indication for 2 subjets than for 3
- substructure underestimated by MC
  - POWHEG: 10% off
  - PYTHIA: up to 40% off

single jet trigger: anti-\( k_T \)
R=0.4, \( E_T > 350 \text{ GeV} \) (EF)
Master distributions

- ATLAS standard jets: $\text{anti-k}_T \ R=0.4$
- $n_{\text{tag}} \geq n$
- $n_{\text{jet}} \geq m$

experimentally:

\[ \begin{align*}
\text{SUSY} & \quad \text{QCD} \\
\text{jet multiplicity} & \quad \text{jet multiplicity} \\
0 & \quad 0 \\
2 & \quad 2 \\
4 & \quad 4 \\
6 & \quad 6 \\
8 & \quad 8 \\
10 & \quad 10 \\
12 & \quad 12 \\
14 & \quad 14 \\
3 & \quad 3
\end{align*} \]
I. identify hard subjets
- undo C/A clustering until all subjets have \( m < 50 \) GeV
- drop subjets that contribute less than 20% to parent mass

II. 3 subjets compatible with top decay?

a) **Filtering:** remove UE and PU
   - recluster with small \( R = \min (0.3, \frac{\Delta R_{ij}^{\text{min}}}{2}) \)
   - keep 5 hardest subjets
   - exclusively recluster constituents to 3 subjets

b) **apply kinematic cuts**
   - any \( m_{ij} \approx m_W \): \( \frac{m_{ij}}{m_{123}} \in \left[ 0.85 \frac{m_W}{m_t}, 1.15 \frac{m_W}{m_t} \right] \)
   - additional cuts (backup)
     - \( p_{T,t} > 200 \) GeV, \( 140 < m_t < 200 \) GeV

---

**HEPTopTagger (HTT)**

Plehn et al., arXiv:1006.2833

fat jet (C/A R=1.5)

top candidate 4-momentum \( p_t \)
HEPTopTagger efficiency

more fake tops at high multiplicity (accidental mass window hits)

significance improvement w.r.t. SUSY analysis without top quark identification
HEPTopTagger filtering

<table>
<thead>
<tr>
<th></th>
<th>normal</th>
<th>tight</th>
<th>loose</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{\text{cut}}$ (GeV)</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>$\max R_{\text{jet}}^{\text{filt}}$</td>
<td>0.3</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>$N_{\text{jet}}^{\text{filt}}$</td>
<td>5</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>$m_W$ window</td>
<td>±15%</td>
<td>±10%</td>
<td>±20%</td>
</tr>
</tbody>
</table>

Event tagging efficiency

Event fake tagging efficiency

Min jet multiplicity

Efficiency

Fake efficiency
HEPTopTagger filtering

- "tight > normal > loose"
- "2 tags > 1 tag"

<table>
<thead>
<tr>
<th></th>
<th>normal</th>
<th>tight</th>
<th>loose</th>
</tr>
</thead>
<tbody>
<tr>
<td>m_{cut} (GeV)</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>max R_{jet}^{filt}</td>
<td>0.3</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>N_{jet}^{filt}</td>
<td>5</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>m_W window</td>
<td>±15%</td>
<td>±10%</td>
<td>±20%</td>
</tr>
</tbody>
</table>
Unintegrated efficiencies

**SUSY**

**QCD**

Event tagging efficiency vs. jet multiplicity.

- **SUSY**
  - 0 tags
  - 1 tag
  - 2 tags

- **QCD**
  - 0 tags
  - 1 tag
  - 2 tags

Efficiency is shown on a linear scale for SUSY and a logarithmic scale for QCD.
Conclusions HEPTopTagger

- high multiplicity: larger chance to hit mass windows accidentally
- 2\textsuperscript{nd} tag: need 3 more jets
- tight filtering
  - reduces fake rate
  - improves S/B
Fat jet $p_T$ and top $p_T$

$P_T$ of leading C/A R=1.5 jet

$P_T$ of leading top candidate

QCD configurations passing HTT are more signal-like
Leading fat jet mass

mass of leading C/A R=1.5 jet

- SUSY, ≥4 jets
- SUSY, ≥7 jets
- QCD, ≥4 jets
- QCD, ≥7 jets

event tagging efficiency for (m_{fj1} > 350 GeV)

fraction of events / 20 GeV

mass (GeV)

keep

efficiency

min jet multiplicity

10^{-3}
Sum of fat jet masses

sum of masses of C/A R=1.5 jets

\[ \sum m (\text{GeV}) \]

- SUSY, \( \geq 4 \) jets
- SUSY, \( \geq 7 \) jets
- QCD, \( \geq 4 \) jets
- QCD, \( \geq 7 \) jets

keep

event tagging efficiency (\( \sum m(\text{fat jet}) > 800 \text{ GeV} \))

\( \text{efficiency} \)

SUSY

QCD

min jet multiplicity

0 1 2 3 4 5 6 7 8 9 10 11
Number of fat jets

Multiplicity of leading C/A R=1.5 jets

- ≥0 jets
- ≥4 jets
- ≥7 jets

**SUSY**

- Entries: 10000
- Mean: 2.594
- RMS: 0.7592

- Entries: 9909
- Mean: 2.603
- RMS: 0.7548

- Entries: 7055
- Mean: 2.744
- RMS: 0.7404

**QCD**

- Entries: 1145867
- Mean: 0.0003668
- RMS: 0.02351

- Entries: 10118
- Mean: 1.035
- RMS: 0.8863

- Entries: 454
- Mean: 2.488
- RMS: 0.7352
Improvement in significance w.r.t. SUSY analysis without top quark identification

- “$p_T > m$”
- “several fat jets $> 1$ fat jet”
Hard substructure in fat jet

HTT breakdown of fat jet into substructure objects (80% mass-drop, $m_{\text{cut}} = 50$ GeV)
Improvement in significance

counting does not help
Improvement in significance

for $n_{\text{jet}} \geq 7$:

- comparable performance
- $S/\sqrt{B}$ improvement: factor 2
## Uncertainties

<table>
<thead>
<tr>
<th></th>
<th>relative uncertainty on efficiency ($n_{\text{jet}} \geq 7$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEPTopTagger subjet calibration</td>
<td>4%</td>
</tr>
<tr>
<td>fat jet mass &gt;350 GeV</td>
<td>5%</td>
</tr>
<tr>
<td>fat jet $p_T &gt;500$ GeV</td>
<td>3%</td>
</tr>
<tr>
<td>sum fat jet mass &gt; 800 GeV</td>
<td>?</td>
</tr>
</tbody>
</table>
Summary

- Delphes study of high jet multiplicity events
- one SUSY model vs. QCD (PYTHIA)

- HEPTopTagger (prong-based):
  - jet combinatorics increase fake rate (accidental hits of mass windows)
  - tight filtering helps

- similar performance from cuts on $\sum m(\text{fat jet}), p_T(\text{fat jet 1})$
- substructure counting and cut on $m(\text{fat jet 1})$ perform worse

- ATLAS study underway (Maddalena Giulini)
Backup
Additional HEPTopTagger cuts

- for $m_i^2 \approx 0$: $m_{123}^2 \approx m_{12}^2 + m_{13}^2 + m_{23}^2$ sphere with radius $m_{123} \approx m_t$

- kinematics described by 2 angles:
  $\cos \theta = m_{23}/m_{123}$
  $\Phi = \arctan m_{13}/m_{12}$

- cut away background

Plehn et al., arXiv:1006.2833