Boosted top quarks & jet substructure

Marcel Vos (IFIC, CSIC/UV, Valencia, Spain)

MIAPP top physics day

Garching, 11 August 2014
The “boosted production” threshold

LHC: $\sqrt{s} \gg E_{\text{EW}}$

Even the heaviest SM particles often acquire $p_T \gg m$

$\rightarrow$ abundant production of “boosted objects”

LHC8: A top factory, the world's first sample of boosted top quarks

<table>
<thead>
<tr>
<th>Expected number of $t\bar{t}$ events in three different kinematical regimes</th>
<th>Tevatron run II</th>
<th>LHC 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusive $t\bar{t}$ production</td>
<td>57,000</td>
<td>2,600,000</td>
</tr>
<tr>
<td>Boosted production: $M_{t\bar{t}} &gt; 1$ TeV</td>
<td>25</td>
<td>30,000</td>
</tr>
<tr>
<td>Highly boosted: $M_{t\bar{t}} &gt; 2$ TeV</td>
<td>0</td>
<td>300</td>
</tr>
</tbody>
</table>

Tevatron: sufficient parton luminosity to discover the top quark, ~ no boosted production

*M.V.,* Boosting sensitivity to new physics, *CERN Courier, Oct 2012*

Results obtained with MCFM, arXiv:1204.1513 [hep-ph] & MSTW2008NLO PDFs

MIAPP, August 2014
The “boosted production” threshold

The next stage of LHC operation (LHC14 from ~April 2015) will produce millions of boosted top quarks and the first events with very high boost.

<table>
<thead>
<tr>
<th>Kinematical Regime</th>
<th>LHC 2012</th>
<th>LHC design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusive $tt$ production</td>
<td>2.600.000</td>
<td>155.000.000</td>
</tr>
<tr>
<td>Boosted production: $M_{tt} &gt; 1$ TeV</td>
<td>30.000</td>
<td>3.000.000</td>
</tr>
<tr>
<td>Highly boosted: $M_{tt} &gt; 2$ TeV</td>
<td>300</td>
<td>47.000</td>
</tr>
</tbody>
</table>

Millions of boosted top quarks, 50.000 extremely boosted events

M.V., Boosting sensitivity to new physics, CERN Courier, Oct 2012

Results obtained with MCFM, arXiv:1204.1513 [hep-ph] & MSTW2008NLO PDFs

Boosted top quarks

As the top quark $p_T$ increases, the opening angle between $W$ and $b$ decreases.

For very high $p_T$ the decay topology is so collimated that individual partons are not resolved with $R = 0.4$.

Rule of thumb for a 2-body decay:

$$R < \frac{2m_X}{p_T}$$

For top, assuming $p_T^W = p_T^t/2$

$$R < \frac{2m_t}{p_T^t} \quad \text{and} \quad R < \frac{4m_W}{p_T^t}$$

Classical algorithms are not adequate for boosted top quarks. See e.g. arXiv:1207.5644.
Boosted top quarks

For very high $p_T$ the decay topology is so collimated that individual partons are not resolved with $R = 0.4$

Reconstruct “boosted object” as a single $R \sim 1$ “fat” jet.


“resolved, at rest” $M_{tt} < 1$ TeV

“boosted, single-jet” $M_{tt} > 1$ TeV

Top tagging:
Kaplan, Rehermann, Schwartz, Tweedie, PRL101,
Boosted top and top-tagging literature

**Early phenomenology papers:**
Lillie, Randall, Wang, hep-ph/0701166,
Almeida, Lee, Perez, Sung, Virzi, arXiv:0810.0934,
New Physics at the LHC, Les Houches 2007

**New sources of boosted tops:**
Berger et al., arXiv:1111.6594 [hep-ph]

**More sophisticated techniques/taggers:**
Kaplan, Rehermann, Schwartz, Tweedie, PRL101
HepTopTagger arXiv:1006.2833
Shower deconstruction arXiv:1102.3480

**Early experimental work**
CMS-PAS-JME-09-001, CMS-PAS-EXO-09-002, CMS-PAS-EXO-09-08, CMS-PAS-TOP-09-009,
ATLAS-PHYS-CONF-2008-008/016, ATL-PHYS-PUB-2010-008

Jet substructure

Distinguish boosted objects from background through an analysis of the jet substructure

Hadronic top jet:

\[ m_j \sim m_t \]

(jet contains complete hadronic top decay, Images: first boosted top candidates, ATLAS-CONF-2011-073, \( m_t = 197 \text{ GeV} \))

Light-quark and gluon jets:

\[ m_j \mu \alpha_s p_T R \]

(peaks at approx. 70 GeV for \( p_T = 500 \text{ GeV} \))
Other jet substructure observables?

ATL-PHYS-PUB-2014-004

Splitting scale: \[ \sqrt{d_{ij}} = \min(p_T^i, p_T^j) \times R_{ij} \]
Butterworth, Cox, Forshaw, PRD65

Momentum balance: \[ \sqrt{y} = \frac{\sqrt{d_{ij}}}{m_{ij}} \]
BDRS, PRL100

Mass drop: \[ \mu_{ij} = \frac{\max(m_i, m_j)}{m_{ij}} \]
BDRS, PRL100

Jet width: \[ w = \frac{\sum_i \Delta R_{ij} p_T^i}{\sum_i p_T^i} \]

N-subjettiness: \[ \tau_N = \frac{\sum_k p_T^{Tk} (\min(\Delta R_{1k}, \Delta R_{2k}, \ldots, \Delta R_{Nk}))^\beta}{\sum_k p_T R_{0k}^\beta} \]
Thaler, van Tilburg, JHEP03

More sophisticated:
+ inject "physics", shower deconstruction (Soper & Spannowsky), templates (Perez et al.)
+ quantify uncertainty inherent in the decision making process, e.g. volatility (Krohn et al.)

Often correlated, combine with care
Jet substructure performance

ATLAS and CMS reach %-level jet energy scale uncertainty for some jets

Jet substructure is rather sensitive to the subtleties of the detector response

<table>
<thead>
<tr>
<th>position</th>
<th>threshold</th>
<th>energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta \phi \times \delta \eta = 0.1 \times 0.1 \text{ grid}$</td>
<td>$p_T &gt; 1 \text{ GeV}$</td>
<td>smear by 50% / $\sqrt{E}$</td>
</tr>
<tr>
<td>1.1%</td>
<td>0.5%</td>
<td>2.5%</td>
</tr>
<tr>
<td>3.5%</td>
<td>9%</td>
<td>4%</td>
</tr>
</tbody>
</table>

Impact on $p_T$ and mass resolution for QCD multi-jets with $400 < p_T < 500 \text{ GeV}$ of a discrete $\eta \times \phi$ grid, a $p_T$ threshold and energy smearing.

Ideally: bring jet constituents (close) to the particle-scale (CMS $\rightarrow$ particle flow objects, ATLAS $\rightarrow$ micro-jets)
Large-R jet and substructure performance

Use in-situ methods (track/calo, data/MC double ratio, W-mass constraint) to constrain jet mass response or large jets, JHEP09 (2013) 076

Recent progress: use $\gamma$+jets for energy response + topology systematic
Detector response

Can we measure jet substructure precisely and reliably?

Under the pile-up conditions of the HL-HLC?

2010: few events/crossing
2011: ~12 events/crossing
2012: up to ~25 events/crossing
HL-LHC: up to 200 events/crossing
Tools and Techniques: jet grooming

Jet substructure is often hidden by “soft stuff” (pile-up, underlying event and soft radiation from the jet itself).

Grooming reduces the effective jet area (substructure resolution, background rejection, and pile-up resilience)

Three flavours: mass-drop filtering (BDRS ’08), trimming (Krohn, Thaler, Wang ’09) pruning (Ellis, Vermillion, Walsh ’09)

SM tt

QCD incl. jets

 BOOST201 0 report EPJC73
Tools and Techniques: jet grooming (II)

- **Filtering**: break jet into subjets on angular scale $R_{\text{filt}}$, require mass drop and asymmetry, take $n_{\text{filt}}$ hardest subjets
  
  Butterworth, Davison, Rubin & Salam '08

- **Trimming**: break jet into subjets on angular scale $R_{\text{trim}}$, take all subjets with $p_{T,\text{sub}} > \epsilon_{\text{trim}} p_{T,\text{jet}}$
  
  Krohn, Thaler & Wang '09

- **Pruning**: building up the jet, if the two subjets about to be recombined have $R > R_{\text{prune}}$ and $\min(p_{T1}, p_{T2}) < \epsilon_{\text{prune}} (p_{T1} + p_{T2})$, discard the softer one.
  
  Ellis, Vermilion & Walsh '09

A detailed experimental comparison: *ATLAS large-R jet paper, JHEP1309*
Pile-up correction

Simulation (2008!) by Salam, Cacciari, Soyez

Jet area: shoot infinitely soft “ghost” particles at the jet and register area where they are clustered into the jet

Pile-up density: determine the level of pile-up activity on an event-by-event basis as the median of the

Subtract density x area

Jet area subtraction restores mass scale and the shape
Jet substructure and pile-up

First LHC data (2010): filtered jet mass flat up to 9 pile-up events for $R=1.2$

LHC 7 TeV: average trimmed jet mass flat up to 14 pile-up events for $R=1.0$

Simulation: jet mass scale for boosted top quarks with up to 200 pile-up vertices. Combination of grooming and pile-up subtraction restores the scale and mitigates impact on resolution

Jet substructure analysis will work for foreseeable future

BOOST2012 report, EPJC74 (2014)
CMS top-tagging

CMS Simulation, $\sqrt{s} = 8$ TeV

Matched parton $p_T > 400$ GeV/c

Matched parton $p_T > 600$ GeV/c

Matched parton $p_T > 800$ GeV/c

- CMS Top Tagger
- subj b-tag
- N-subjettiness ratio $\tau_3/\tau_2$
- CMS + subj b-tag
- CMS + $\tau_3/\tau_2 +$ subj b-tag
- HEP Top Tagger
- HEP + $\tau_3/\tau_2 +$ subj b-tag

CMS-JME-13-007
ATLAS Top-tagging

Performance on a $Z' \to tt$ sample ($m=1.75$ TeV)

ATLAS top-tagging note
ATLAS-CONF-2013-084 + shower deconstruction
ATLAS-CONF-2014-003
So, is everything done, then?

Where we stand:
Resolved regime has solutions since 1990s
Cope with the boosted regime ~now
Primitive combinations

The next challenge:
very high boost in the 13 TeV operation
- R=1/1.2 is too large
- Detector granularity is marginal

Solutions for a seamless connection of boosted and highly boosted regimes:
- combination of orthogonal signal regions
- scale-invariant tagging
  (Gouzevitch et al., JHEP1307)
- variable-R jet reconstruction
  (Krohn, Thaler, Wang, JHEP0906)
- inject more detailed knowledge of physics
  (template overlap and shower deconstruction, the subject of Dave Soper’s contribution)

Images: shower deconstruction, ATLAS-CONF-2014-003
tt resonance summary

First “boosted” analysis by ATLAS and CMS in JHEP 1209

Mass reach greatly enhanced

The “resolved” analysis offers complementary low-mass sensitivity

Recent analyses combine both (even if maybe not as well as they could)
**W' → tb**

Search for $W' \rightarrow tb \rightarrow bbqq$

Complete 2012 data set

arXiv:1408.0886 since last week

$\text{BR } (bbqq) = 3 \times \text{BR}(bb \text{ e/m n})$

Sensitive to right-handed $W'$ with only hadronic decays

Deal with large multi-jet background using b-tag and top-tag

Max. deviation 1.4 s, limits on new states with SM coupling:

$m(W') > 1.68 \text{ TeV }$ (left-handed)

$> 1.76 \text{ TeV }$ (right-handed)

Preliminary CMS result (B2G-12-009) has considerably tighter limits!

More on searches in Duc Bao Ta’s contribution this afternoon
Boosted W,Z

First envisaged application: vector boson scattering cross section


“A new method for identifying hadronically decaying W bosons is introduced, which we expect to be useful more generally [...]”

A lot of recent performance work in the experiments

CMS-PAS-JME-13-006

ATL-PHYS-PUB-2014-004

CMS X → VV and X → Vj search


ll,lv + j (arXiv:1405.3447)

ATLAS mono-V search
SM measurements of boosted production rate

Measurement of SM production of boosted objects


$\sigma = 8.5 +/- 1.7 \text{ pb} \ (p_T > 320, |\eta| < 1.9)$

Further scrutiny of inclusive top quark pair production rate limited by theory uncertainty

After first exploration, characterize differential distributions of top quark $p_T$, $t\bar{t}$ mass, charge asymmetry, as precisely as possible

Can theorists compute a precise differential cross section for pseudo-tops, if their definition involves jet grooming?
Boosted Higgs

Motivation to look for boosted objects:
- A new light particle (H, \chi^0, \ldots) is more easily isolated from background when boosted, for example because combinatorics are less of an issue
- Or, the measurement of a corner of phase space yields complementary sensitivity

Boosted Higgs (W/Z + light Higgs → b\overline{b})


"We conclude that subjet techniques have the potential to transform the high-pT WH,ZH (H → bb ) channel into one of the best channels for discovery of a low mass Standard Model Higgs at the LHC"

*ZH, Soper, Spannowksy, JHEP 1008:029 (2010)*

*gg → H, Grojean et al., JHEP1405*


"Combinatorial backgrounds are not a problem, and we find a multitude of distributions distinguishing between signal and continuum background."

A good shot at measuring the top and bottom Yukawa coupling
The longer-term future of boost

Future $e^+e^-$ machines:
Any machine: new ideas on jet reconstruction
Ultra-granular detectors: benchmark ultimate particle-flow performance
**ILC or CLIC at 1-3 TeV**: a small (few 100k events), but easy-to-control sample of (nearly) monochromatic boosted top quarks

Future hadron colliders:

<table>
<thead>
<tr>
<th>Expected number of boosted $tt$ events in three different kinematical regimes</th>
<th>LHC design (2015 – …) 300 fb$^{-1}$ @ 13 TeV</th>
<th>HE-LHC (&gt;2030) 3000 fb$^{-1}$ @ 33 TeV</th>
<th>VHE-LHC (&gt;2030) 3000 fb$^{-1}$ @ 100 TeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boosted production: $M_{tt} &gt; 1$ TeV</td>
<td>3.000.000</td>
<td>46.000.000</td>
<td>820.000.000</td>
</tr>
<tr>
<td>Highly boosted: $M_{tt} &gt; 2$ TeV</td>
<td>47.000</td>
<td>23.000.000</td>
<td>450.000.000</td>
</tr>
<tr>
<td>Extremely boosted: $M_{tt} &gt; 5$ TeV</td>
<td>30</td>
<td>150.000</td>
<td>9.500.000</td>
</tr>
</tbody>
</table>

So much boost we don't know how to deal with it!

MIAPP, August 2014
Future colliders & detectors

A 150 x 150 \( \mu \text{m}^2 \) pixel at \( r = 4.4 \text{ cm} \) → \( \Delta \phi/\Delta \eta \sim 0.003 \)
A 10 cm x 80 \( \mu \text{m} \) \( \mu \)-strip at \( r = 50 \text{ cm} \) → \( \Delta \phi/\Delta \eta \sim 0.006 \)
An ECAL crystal at \( r = 1.4 \text{ m} \) → \( \Delta \phi/\Delta \eta \sim 0.015 \)
A tile in HCAL at \( r = 1.5 \text{ m} \) → \( \Delta \phi/\Delta \eta \sim 0.1 \)

A jet may cover 5000 pixels/layer, 1000 ECAL crystals, but only HCAL 50 tiles

Preliminary: typical inter-particle distance vs. distance from IP

Superposed feature size, Moliere radius and interaction length of current experiments (and concepts)
BOOST: the workshop series

The workshop series:

SLAC 2009

Arizona, August 2013, report underway
UC London, August 2014

+ events at Boston, CERN LHCC, Manchester, Oregon, Seattle, SLAC

BOOST2014 is the sixth of a series of successful joint theory/experiment workshops which bring together the world leading experts from theory and the Tevatron and LHC experiments to discuss the latest progress and develop new approaches on the reconstruction and use of boosted decay topologies as search tools for new physics. This year, the workshop is hosted by the UCL HEP Group at the heart of London.
Conclusions

The LHC is a “boosted object” factory.

A lot of effort was invested to understand jet substructure: physics modeling, detector response and pile-up

Mature and complete tool-box now available

The first searches with these techniques indeed enhance the LHC physics potential

Measurements of boosted object production rates are starting to appear

Much more “boost” expected in the 2015 run of the LHC