Boosting searches for new physics using jets and jet substructure at the energy frontier

Exploring the Physics Frontier with Circular Colliders

Aspen Winter 2015

David W. Miller

Enrico Fermi Institute

January 30, 2015
Roadmap

1. Introduction

2. Prospects for boosted object searches at the energy frontier

3. Pile-up mitigation and jet substructure

4. Substructure tools for boosted objects at the energy frontier

5. Summary and conclusions
Outline

1. Introduction
2. Prospects for boosted object searches at the energy frontier
3. Pile-up mitigation and jet substructure
4. Substructure tools for boosted objects at the energy frontier
5. Summary and conclusions
Run I of the LHC saw an unprecedented evolution in boosted objects
Run I of the LHC saw an unprecedented evolution in boosted objects

All of these analyses use boosted objects (top, $W/Z/h$, even $\tilde{g}$) in some way.
**We have already reached the TeV scale**

Limits on massive particles that decay to $W/Z/t/h$ are at or above the TeV-scale...and have been for some time.
### We have already reached the TeV scale

Limits on massive particles that decay to W/Z/t/h are at or above the TeV-scale...have been for some time.

<table>
<thead>
<tr>
<th>Extra dimensions</th>
<th>Model e,µ,τ,γ Jets Emiss T L dt[fb−1] Mass limit Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD G_{KK} + g/q</td>
<td>2e,µ,τ,γ Jets Emiss T L dt[fb−1] Mass limit Reference</td>
</tr>
<tr>
<td>ADD non-resonant ℓ ℓ</td>
<td>2e,µ,τ,γ Jets Emiss T L dt[fb−1] Mass limit Reference</td>
</tr>
<tr>
<td>ADD QM → ℓ q</td>
<td>2e,µ,τ,γ Jets Emiss T L dt[fb−1] Mass limit Reference</td>
</tr>
<tr>
<td>ADD BH high N_{het}</td>
<td>2e,µ,τ,γ Jets Emiss T L dt[fb−1] Mass limit Reference</td>
</tr>
<tr>
<td>ADD BH high Σ_{PR}</td>
<td>2e,µ,τ,γ Jets Emiss T L dt[fb−1] Mass limit Reference</td>
</tr>
<tr>
<td>RS1 G_{KK} → ℓ ℓ</td>
<td>2e,µ,τ,γ Jets Emiss T L dt[fb−1] Mass limit Reference</td>
</tr>
<tr>
<td>RS1 G_{KK} → WW → ℓ ℓ τv</td>
<td>2e,µ,τ,γ Jets Emiss T L dt[fb−1] Mass limit Reference</td>
</tr>
</tbody>
</table>

### Model ℓ,γ Jets Emiss T ∫ L dt[fb−1]

<table>
<thead>
<tr>
<th>Model ℓ,γ Jets Emiss T ∫ L dt[fb−1] Mass limit Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauge bosons</td>
</tr>
<tr>
<td>SSM Z' → ℓ ℓ</td>
</tr>
<tr>
<td>SSM W' → ℓτ</td>
</tr>
<tr>
<td>EGM W' → WZ → ℓ ℓ τv</td>
</tr>
<tr>
<td>EGM W' → WZ → q τψ</td>
</tr>
<tr>
<td>LRSM W_K → ℓ ℓ</td>
</tr>
<tr>
<td>LRSM W_K → ℓτ</td>
</tr>
</tbody>
</table>

### Heavy quarks

<table>
<thead>
<tr>
<th>Model ℓ,γ Jets Emiss T ∫ L dt[fb−1] Mass limit Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD G_{KK} + g/q</td>
</tr>
<tr>
<td>ADD non-resonant ℓ ℓ</td>
</tr>
<tr>
<td>ADD QM → ℓ q</td>
</tr>
<tr>
<td>ADD BH high N_{het}</td>
</tr>
<tr>
<td>ADD BH high Σ_{PR}</td>
</tr>
<tr>
<td>RS1 G_{KK} → ℓ ℓ</td>
</tr>
<tr>
<td>RS1 G_{KK} → WW → ℓ ℓ τv</td>
</tr>
</tbody>
</table>

### Model ℓ,γ Jets Emiss T ∫ L dt[fb−1]

<table>
<thead>
<tr>
<th>Model ℓ,γ Jets Emiss T ∫ L dt[fb−1] Mass limit Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauge bosons</td>
</tr>
<tr>
<td>SSM Z' → ℓ ℓ</td>
</tr>
<tr>
<td>SSM W' → ℓτ</td>
</tr>
<tr>
<td>EGM W' → WZ → ℓ ℓ τv</td>
</tr>
<tr>
<td>EGM W' → WZ → q τψ</td>
</tr>
<tr>
<td>LRSM W_K → ℓ ℓ</td>
</tr>
<tr>
<td>LRSM W_K → ℓτ</td>
</tr>
</tbody>
</table>

### Heavy quarks

<table>
<thead>
<tr>
<th>Model ℓ,γ Jets Emiss T ∫ L dt[fb−1] Mass limit Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD G_{KK} + g/q</td>
</tr>
<tr>
<td>ADD non-resonant ℓ ℓ</td>
</tr>
<tr>
<td>ADD QM → ℓ q</td>
</tr>
<tr>
<td>ADD BH high N_{het}</td>
</tr>
<tr>
<td>ADD BH high Σ_{PR}</td>
</tr>
<tr>
<td>RS1 G_{KK} → ℓ ℓ</td>
</tr>
<tr>
<td>RS1 G_{KK} → WW → ℓ ℓ τv</td>
</tr>
</tbody>
</table>

---

At 100 TeV, boosted objects are ubiquitous

Deployment of a new set of tools for SM and BSM physics

An enormous set of tools has been deployed by both collaborations to maximize our sensitivity to new physics, and to study old physics in new ways.

Jet $p_T > 350$ GeV and 3 $k_t$ subjets

arXiv:1306.4945

ATLAS-CONF-2013-084
“If you ain’t boostin’, you ain’t livin” – Nhan Tran, FNAL (Experimental Summary at BOOST 2014)
This trend will continue and we must be prepared

Table adapted from BOOST 2012 Report (arXiv:1311.2708): calc. w/ MCFM. Yields for boosted tops ($M_{tt} > 1$ TeV) and highly boosted tops ($M_{tt} > 2$ TeV) given in the 2nd and 3rd row.
The challenges we will face

- We have many tools to mitigate pile-up, but the increases in instantaneous luminosity are nonetheless very difficult to cope with.
- Good modeling of complex observables by Monte Carlos is very important and more unfolded measurements may be the key to improvements here.
Outline

1 Introduction

2 Prospects for boosted object searches at the energy frontier

3 Pile-up mitigation and jet substructure

4 Substructure tools for boosted objects at the energy frontier

5 Summary and conclusions
During the majority of 2015, we generally expect similar operating conditions as in Run I, except for bunch spacing.

Need to demonstrate performance for signal and background early, and likely use similar working points for top tagging as in Run I.
Early Run II extensions for Run I boosted analyses

Two advances made near the end of Run I are likely to be adopted early-on in Run II.

- LEFT: Optimized boson tagging using very large parameter space
- RIGHT: Track-jet-based $b$-tagging with very small radii. Now have the ability to define $b$-tagged objects independent of the calorimeter and then associated to a given calorimeter-based jet definition post facto
Prospects for boosted object searches at the energy frontier

New ideas in $b$-tagging

Using specialized $b$-tagging for boosted objects (II)

Subjet $b$-tagging an integral part of boosted top tagging algorithms in CMS

CMS $b$-tagging for boosted topologies commissioned at 8 TeV and analyses exploiting subjet $b$-tagging now public

CMS-JME-13-007 (also Ferencek at BOOST 2014)
Outline

1. Introduction

2. Prospects for boosted object searches at the energy frontier

3. Pile-up mitigation and jet substructure

4. Substructure tools for boosted objects at the energy frontier

5. Summary and conclusions
Both CMS and ATLAS implement the so-called ‘area-based’ pile-up subtraction

This is independent of the jet algorithm, and can thus also be applied to subjets if appropriate

Spoiler alert: it will be!
Area-based pile-up subtraction for jet kinematics

ATLAS-CONF-2013-083

- Reduces the dependence on pile-up significantly (it better!)
- Also reduces the impact on the jet resolution
- This is the default mode of pile-up mitigation on the jet 4-vector for Run II (being implemented in the trigger as well!)
CMS performed an excellent and comprehensive study of new constituent-level pile-up subtraction algorithms.

From “simple” (e.g. CHS) to “complex” (e.g. PUPPI), and maximum tracking information (e.g. PUPPI) to no tracking information (e.g. constituent subtraction)
PileUp Per Particle Id
Bertolini, Harris, Low, Tran (arXiv:1407.6013)

General Idea of the Algorithm

- Use the Jets without Jets paradigm
  - For each particle draw a cone around it

- In each particle cone
  - Compute metric $\alpha$
    - Distinguishes particle from hard scatter from PU
  - Calculate median $\alpha$ and $\alpha_{\text{RMS}}$ over an event for PU
    - Average over all particles associated to another vertex

- Compute a weight that a particle is from pileup

- Reweight particles and re-interpret the event

I view this more as a way to customize the inputs to the jet reconstruction
**PUPI in CMS**

Viola Sordini at BOOST 2014

Operates on the inputs to jet clustering (here PF candidate particles)  

$ \alpha_i = \log \sum_{j \in ch., PV} \left( \frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \Theta(R_0 - \Delta R_{ij})$ for $|\eta| < 2.5$

$ \begin{cases} 
\alpha_i = \log \sum \frac{p_{T,j}}{\Delta R_{ij}} \Theta(R_0 - \Delta R_{ij}) & \text{for } |\eta| \geq 2.5 \\
\alpha_i = \log \sum p_{T,j} \Theta(R_0 - \Delta R_{ij}) & \text{for } |\eta| < 2.5 
\end{cases}$

- The distribution for charged PU particles is used as template for the distribution for all PU particles

- For each neutral particle, a $\chi^2$ variable is constructed (for $|\eta| \geq 2.5$, sum the two $\chi^2$)

$$\chi_i^2 = \frac{|\alpha_i - \alpha_{PU}|^2}{RMS_{PU}^2}$$

I view this more as a way to customize the inputs to the jet reconstruction, which is extremely important
How do these approaches scale as we move into the future?

**ATLAS ECFA Studies on pile-up suppression**

- The mean number of jets with pile-up-corrected $p_T > 20$ GeV and $|\eta| < 1$ per event vs. $N_{PV}$

- For the curves with open markers, tracking information was used to suppress pileup jets, by imposing the same $R_{p_T} > 0.1$

$$R_{p_T} = \frac{\sum p_T^{track}(PV0)}{p_T^{jet}}$$
How do these approaches scale as we move into the future?

ATLAS ECFA Studies on pile-up suppression

- Hard-scatter jets are matched within $\Delta R < 0.3$ to a truth jet with $p_T > 10$ GeV, whereas pileup jets are required to have a minimal $\Delta R > 0.6$ from any truth jet with $p_T > 4$ GeV.

$$R_{p_T} = \frac{\sum p_T^{\text{track}}(PV0)}{p_T^{\text{jet}}}$$
How do these approaches scale as we move into the future?

ATLAS ECFA Studies on pile-up suppression

- Hard-scatter jets are matched within $\Delta R < 0.3$ to a truth jet with $p_T > 10$ GeV, whereas pileup jets are required to have a minimal $\Delta R > 0.6$ from any truth jet with $p_T > 4$ GeV.

$$R_{p_T} = \frac{\sum p_T^{\text{track}}(PV0)}{p_T^{\text{jet}}}$$
Outline

1. Introduction

2. Prospects for boosted object searches at the energy frontier

3. Pile-up mitigation and jet substructure

4. Substructure tools for boosted objects at the energy frontier

5. Summary and conclusions
Both collaborations are now doing very comprehensive scans of tagging observables for bosons
Incorporating multiple mass definitions simultaneously
Multivariante combinations for W tagging

- Now starting to incorporate many variables (some are already “composite” observables, e.g. QG-likelihood)
- Linear correlation coefficients (see ATLAS) are useful, but need to be honest about their limitations
- Similarly, need to be honest about the degree to which these complex combinations are necessary
**Boosted top tagging and pile-up at very high luminosity**

**High pile-up scenarios for top tagging**

- The anti-$k_t$ $R = 1.0$ jet mass distribution in $Z' \rightarrow t\bar{t}$ events where the mean number of interactions per bunch crossing $\langle \mu \rangle$ is 40, 80, 140, and 200.
- The pileup correction is made to the subjets only, before their pT fraction is calculated.
- When both trimming and jet 4-vector pileup subtraction are applied, the jet mass distribution is stable even at $\langle \mu \rangle = 200$.
Boosted top tagging and pile-up at very high luminosity

High pile-up scenarios for top tagging

- The anti-$k_t$ $R = 1.0$ jet mass distribution in $Z' \rightarrow t\bar{t}$ events where the mean number of interactions per bunch crossing ($\langle \mu \rangle$) is 40, 80, 140, and 200.
- The pileup correction is made to the subjets only, before their pT fraction is calculated.
- When both trimming and jet 4-vector pileup subtraction are applied, the jet mass distribution is stable even at $\langle \mu \rangle = 200$. 

![Graph showing jet mass distribution with different pileup corrections](image_url)
**Boosted top tagging and pile-up at very high luminosity**

**High pile-up scenarios for top tagging**

- The anti-\(k_t\) \(R=1.0\) jet mass distribution in \(Z' \rightarrow t\bar{t}\) events where the mean number of interactions per bunch crossing \(\langle \mu \rangle\) is 40, 80, 140, and 200.
- The pileup correction is made to the subjets only, before their pT fraction is calculated.
- When both trimming and jet 4-vector pileup subtraction are applied, the jet mass distribution is stable even at \(\langle \mu \rangle = 200\).
Boosted top tagging and pile-up at very high luminosity

High pile-up scenarios for top tagging

- The anti-$k_t$ $R=1.0$ jet mass distribution in $Z' \rightarrow t\bar{t}$ events where the mean number of interactions per bunch crossing ($\langle \mu \rangle$) is 40, 80, 140, and 200.
- The pileup correction is made to the subjets only, before their pT fraction is calculated.
- When both trimming and jet 4-vector pileup subtraction are applied, the jet mass distribution is stable even at $\langle \mu \rangle = 200$. 

![Graph showing jet mass distribution for different pileup scenarios](image-url)
Tagging tops with substructure

**ATLAS ECFA Studies on top tagging**

- The pile-up suppression uses the event-by-event median $p_T$ density ($\rho$) and the jet area.
- The pile-up correction is made to the subjets only before their $p_T$ is calculated for the trimming procedure.
Outline

1 Introduction

2 Prospects for boosted object searches at the energy frontier

3 Pile-up mitigation and jet substructure

4 Substructure tools for boosted objects at the energy frontier

5 Summary and conclusions
Conclusions and outlook

- **Pile-up mitigation:**
  - We are at the point where the experimental techniques to mitigate pile-up are extremely advanced
  - Collaborations will be testing these techniques at the LHC in the coming years
  - Adapting these techniques to the needs and constraints of new detectors and new technologies for future colliders will be the primary hurdle, not the methods themselves

- **Jet definitions:**
  - Wide variety of jet definitions available, and dynamically defined sizes and grooming parameters will become relatively commonplace
  - This will be relevant for wide dynamic range searches even for very early in Run II
  - Grooming is invaluable for $W/Z/h/t$ tagging and will need to be considered in any future collider detector study and optimization

- **Jet tagging:**
  - Very complex approaches are available, as well as very simple, yet effective, ones
  - Expect multivariate taggers to be commissioned by LHC experiments soon
  - Still need “simple” taggers for cross-checks, easy background/sideband estimates
  - $b$-tagging will be involved in almost all searches
Summary and conclusions

The last few years have seen a dramatic increase in the level to which we are able to exploit the hadronic final state to perform precision physics at the energy frontier by treating a jet as much more than just a fixed 4-vector surrogate for a parton.

Looking ahead

- These techniques will play a critical role in first Run II analyses
- SUSY, Exotics, Higgs (SM+Exotic) will make use of these taggers as we progress higher in the energy frontier
- At even higher instantaneous luminosities, the complex techniques now available for input-level pile-up mitigation will be come crucial

Thank you!

2015? 2045?
Outline

6 Backup slides and additional information
Additional Material
Outline

Backup slides and additional information

- Triggering on boosted objects in Run III
- Track-jet based $b$-tagging
- Caveats and pitfalls to watch out for
The entire calorimeter on a single board!
Triggering on boosted objects in Run III

Level-1 fat jet trigger: **global feature extraction (gFEX)**

- **Signal:** boosted top quarks
  - Build **gFEX jets (in red)** from seeds
  - Seed threshold: $E_T > 15$ GeV
  - Sum gTowers in $\Delta R < 1.0$
    (noise cut: $E_T > 3$ GeV)

- **Compare to Run I** (open markers)
  - Efficiency for Run I trigger **depends on the subjet multiplicity**
  - Worsening efficiencies for QCD jet $\rightarrow W$ jet $\rightarrow$ top jet
  - gFEX efficiency **does not depend on the jet structure**

**gFEX efficient despite significant jet substructure!**
Level 1 pile-up subtraction for boosted tops and bosons

- \( \rho \) calculation at Level 1 is independent of hard scattering process, as it should be
- Subtraction improves turn-on and allows for subjet tagging at Level 1
Track-jet based $b$-tagging for Run II

ATL-PHYS-PUB-2014-013

- Used heavily in the graviton $\rightarrow HH \rightarrow 4b$ search, but has significant implications for top & top-like final states.
- Now have the ability to define $b$-tagged objects independent of the calorimeter and then associated to a given calorimeter-based jet definition post facto.
Track-jet based $b$-tagging for Run II

ATL-PHYS-PUB-2014-013

- Used heavily in the graviton→$HH \rightarrow 4b$ search, but has significant implications for top & top-like final states.
- Now have the ability to define $b$-tagged objects independent of the calorimeter and then associated to a given calorimeter-based jet definition post facto.
Complex taggers compared to “simple” taggers (II)

- See identical performance with a cut on $\tau_{32}$ + cut on mass + cut on $b$-tagging as using full CMSTopTagger
"Best" performance vs. "realistic" performance

- The difference between "best" in boxes, and "second best" (e.g. rejection factor 41 vs. 52) is 0.4%.
- This is not an observable difference, and if we take such arguments literally, we may find ourselves in a corner.
- We should develop a threshold for improvement that requires that the complexity needed for that improvement be outweighed by the significance of the gain.
**Complex taggers compared to “simple” taggers (I)**

![Graph showing mistag rate versus top tag efficiency for different tagging scenarios.](image)

**Legend:**
- HEP Top Tagger
- HEP + $\tau_3/\tau_2$
- HEP + $\tau_3/\tau_2$ + sub. b-tag
- MultiR HEP Top Tagger
- CMS Top Tagger
- CMS Top Tagger + $\tau_3/\tau_2$
- CMS Top Tagger + $\tau_3/\tau_2$ + sub. b-tag
- Shower Deconstruction CA8
- Shower Deconstruction CA8 + sub. b-tag