Jets Physics
Algorithms and Substructure Techniques

Matteo Cacciari
LPTHE Paris
Jet revolutions

Percentage of papers in INSPIRE containing the word "jet"

First ‘jet revolution’

Second ‘jet revolution’?

(1% ≈ 500 papers)

NB: Data not noise-subtracted
‘Jet substructure’ papers in INSPIRE

Number of papers containing the words ‘jet substructure’

More than 100 papers since 2008
(+ some background noise)

Pioneered by M. Seymour in the early ‘90s, rebooted by BDRS paper

15. Jet substructure as a new Higgs search channel at the LHC.
Published in Phys.Rev.Lett. 100 (2008) 242001
The pervasiveness of jets

- ATLAS and CMS have each published **300+** papers since 2010
  - More than a **third** of these papers make use of jets
  - **60%** of the **searches** papers makes use of jets

Why are jets so important?

(Source: INSPIRE. Results may vary when employing different search keywords)
One purpose of a ‘jet clustering’ algorithm is to reduce the complexity of the final state, simplifying many hadrons to simpler objects that one can hope to calculate.
Like Janus, jets can serve two purposes:

- **Observables**
  to be defined, calculated, measured

- **Tools**
  to be used to extract specific properties of the final state

---

From Wikipedia:

In ancient Roman religion and myth, Janus (Latin: Ianus, pronounced [ˈiaː.nuːs]) is the god of beginnings and transitions, and thereby of gates, doors, passages, endings and time. He is usually depicted as having two faces, since he looks to the future and to the past. The Romans named the month of January (Ianuarius) in his honor.
Requirements for jet algorithms

- In order to be most useful as an observable, a jet algorithm should be usable in perturbation theory and its jets not too affected by non-perturbative physics
  - Infrared and collinear safety
  - Controlled sensitivity to hadronisation and underlying event
  - Possibility to subtract effectively pile-up contamination

- Two main classes of jet algorithms:
  - cones
  - sequential recombination algorithms
    - Sequential recombination algorithms ($k_t$, Cambridge/Aachen, anti-$k_t$) provide a clustering history
      - When physically meaningful, the clustering history can be exploited to explore the substructure of a jet → jet as a tool
**IRC safe algorithms**

<table>
<thead>
<tr>
<th>$k_t$</th>
<th>$d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2}) \Delta R_{ij}/R^2$</th>
<th>Catani et al ‘91 Ellis, Soper ‘93</th>
<th>N\text{lnN}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cambridge/ Aachen</strong></td>
<td>$d_{ij} = \Delta R_{ij}/R^2$</td>
<td>Dokshitzer et al ‘97 Wengler, Wobish ‘98</td>
<td>N\text{lnN}</td>
</tr>
<tr>
<td><strong>anti-$k_t$</strong></td>
<td>$d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2}) \Delta R_{ij}/R^2$</td>
<td>MC, Salam, Soyez ’08 (Delsart, Loch)</td>
<td>$N^{3/2}$</td>
</tr>
<tr>
<td><strong>SISCones</strong></td>
<td>Seedless iterative cone with split-merge gives ‘economical’ jets</td>
<td>Salam, Soyez ‘07</td>
<td>$N^2\text{lnN}$</td>
</tr>
</tbody>
</table>

‘second-generation’ algorithms

All (as well as many IRC unsafe ones) are available in FastJet, [http://fastjet.fr](http://fastjet.fr)
Hierarchical substructure

Slide by
Gavin Salam

Matteo Cacciari - LPTHE

ICHEP - Valencia - July 2014
Why boosted objects

Heavy particle $X$ at **rest**

- Easy to resolve jets and calculate invariant mass,
- but signal very likely swamped by background (eg $H \rightarrow bb$ v. $tt \rightarrow WbWb$)

**Boosted** heavy particle $X$

- Cross section very much reduced, but acceptance better and some backgrounds smaller/reducible
The substructure of a jet can be exploited to:

- **tag** a particular structure inside the jet, i.e. a massive particle
  - First examples: Higgs (2-prongs decay), top (3-prongs decay)
- remove background contamination from the jet or its components, while keeping the bulk of the perturbative radiation
  - First examples: filtering, trimming, pruning
  - This operation is often generically denoted as **grooming**
Fat-jet finding
Often anti-kt, $R \approx 1$

Tagging step
large $p_t$, large mass fat-jet, signal or background

grooming step
signal jet candidate, still partly background-contaminated

final candidate, potentially with little background contamination

Note that in some taggers (e.g. pruning) the tagging and grooming steps are not explicitly factorised

Also, some tools may actually not follow rigidly this scheme
A two-prongs tagger/groomer for boosted Higgs, which

- Uses the **Cambridge/Aachen** algorithm
- Employs a **Mass-Drop** condition, as well as an **asymmetry cut** to find the relevant splitting (i.e. ‘tag’ the heavy particle)
- Includes a post-processing step, using ‘filtering’ (introduced in the same paper) to clean as much as possible the resulting jets of UE contamination (‘grooming’)
$pp \rightarrow ZH \rightarrow \nu\bar{\nu}b\bar{b}$

**Visualisation of BDRS**

Cluster with a large $R$

Undo the clustering into subjets, until a large mass drop is observed: tagging step

Re-cluster with smaller $R$, and keep only 3 hardest jets: grooming step

Butterworth, Davison, Rubin, Salam, 2008
The jet substructure maze

Some of the tools developed for boosted W/Z/H/top reconstruction

Jet Declustering
- Seymour93
- YSplitter
- Mass-Drop+Filter
- ATLASTopTagger

Jet Shapes
- Pruning
- Planar Flow
- Twist
- N-jettiness
- ACF
- CoM N-subjettiness (Kim)
- N-subjettiness (TvT)

Matrix-Element
- Templates
- JHTopTagger
- CMSTopTagger
- HEPTopTagger (+ dipolarity)
- Shower Deconstruction

Multi-variate tagger

Qjets
First taggers/groomers

- **Mass Drop + Filtering**
  - Declare with mass drop and asymmetry conditions
  - Recluster constituents into subjets at distance scale $R_{\text{filt}}$, retain $n_{\text{filt}}$ hardest subjets
  - Butterworth, Davison, Rubin, Salam, 2008

- **Jet ‘trimming’**
  - Recluster constituents into subjets at distance scale $R_{\text{trim}}$, retain subjets with $p_{t,\text{subjet}} > \varepsilon_{\text{trim}} p_{t,\text{jet}}$
  - Krohn, Thaler, Wang, 2009

- **Jet ‘pruning’**
  - While building up the jet, discard softer subjets when $\Delta R > R_{\text{prune}}$ and $\min(p_{t1}, p_{t2}) < \varepsilon_{\text{prune}} (p_{t1} + p_{t2})$
  - S. Ellis, Vermilion, Walsh, 2009

**Aim**: limit contamination from QCD background while retaining bulk of perturbative radiation
Robustness of substructure tools

Dasgupta, Fregoso, Marzani, Salam, 2008

Tools that are considered (or can be seen in Monte Carlo tests) to behave ‘similarly’ could cease to do so in different parameter regions.
Analytic calculations of jet substructure

Dasgupta, Fregoso, Marzani, Salam, 2013

Monte Carlo

Analytic
(resummed pQCD)

- Analytical understanding of ‘kinks’ in distributions
- Check of Monte Carlo predictions
- Other analytical investigations: Rubin 2010 (filtering), Walsh, Zuberi 2011 (jet substructure with SCET), Feige Schwartz, Stewart, Thaler 2012 (N-subjettiness), Dasgupta, Marzani, Powling 2013 (groomed jet mass), ...

\[
\frac{1}{\sigma} \frac{d\sigma}{dm^2}^{(\text{trim, LO})} = \frac{\alpha_s C_F}{\pi} \int_0^1 dz p_{gg}(z) \int \frac{d\theta^2}{\theta^2} \delta(m^2 - z(1-z)p_t^2\theta^2) \times \\
\times \left[ \Theta(z - z_{\text{cut}}) \Theta(1 - z - z_{\text{cut}}) \Theta(\theta^2 - R_{\text{sub}}^2) + \Theta(R_{\text{sub}}^2 - \theta^2) \right] \Theta(R^2 - \theta^2)
\]
Alternatives to (standard) substructure

- If what we are interested in is the structure of the constituents of a jet, the “jet” itself is not the most important feature.
- A different algorithm, or simply the study of the constituents in a certain patch will also do. Selected alternatives are:
  - Use of jet-shapes to characterise certain features
    - e.g. \textit{N-subjettiness}: how many subjets a jets appears to have
  - Alternative ways of clustering
    - e.g. \textit{Qjets}: the clustering history not deterministic, but controlled by random probabilities of merging. Can be combined with, e.g. pruning
  - Use information from matrix element
    - e.g. \textit{shower deconstruction}: use analytic shower calculations to estimate probability that a certain configuration comes from signal or from background
  - Use event shapes mimicking jet properties
    - e.g. \textit{JetsWithoutJets}, mimicking trimming

\textcite{Thaler, van Tilburg, 2011}
\textcite{Ellis, Hornig, Roy, Krohn, Schwartz, 2012}
\textcite{Soper, Spannowsky, 2011}
\textcite{Bertolini, Chen, Thaler, 2013}
Many of the substructure-based tools have been systematically compared, for instance in the context of the proceedings of the BOOST conferences. Efficiency-mistag curves for many top taggers. Improvements upon earlier tools have become hard to obtain.
Experimental validation

ATLAS and CMS have extensively validated the new grooming techniques, and also produced the first results using them.
Jet substructure in experiments

Extensive experimental work

Last 12 ATLAS & CMS preprints citing jet substructure work

Jet Cross-Section Measurements In CMS
CMS Collaboration

Performance of jet substructure techniques for large-$R$ jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector
ATLAS Collaboration
Measurement of jet shapes in top pair events at $\sqrt{s} = 7$ TeV using the ATLAS detector
ATLAS Collaboration

Searches for New Physics in Multijet Final States
for the CMS Collaboration

Search for Single and Pair-Production of Dijet Resonances with the CMS Detector
CMS Collaboration

Search for dark matter in events with a hadronically decaying W or Z boson and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector
ATLAS Collaboration

Searches for anomalous t\bar{t}bar production in pp collisions at $\sqrt{s} = 8$ TeV
CMS Collaboration

Search for heavy resonances decaying to top quarks
for the CMS Collaboration

Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks
CMS Collaboration

Search for the SM Higgs Boson Produced in Association with a Vector Boson and Decaying to Bottom Quarks
for the CMS Collaboration

Inclusive search for a vector-like T quark with charge 2/3 in pp collisions at $\sqrt{s} = 8$ TeV
CMS Collaboration

Search for top-quark partners with charge 5/3 in the same-sign dilepton final state
CMS Collaboration
Certainly incomplete list of phenomenology papers

**Discovering baryon-number violating neutralino decays at the LHC**

**Stop Reconstruction with Tagged Tops**
Tilman Plehn, Michael Spannowsky, Michihisa Takeuchi, Dirk Zerwas, arXiv:1006.2833

**Boosted Semileptonic Tops in Stop Decays**
Tilman Plehn, Michael Spannowsky, Michihisa Takeuchi, arXiv:1102.0557

**Scale-invariant resonance tagging in multijet events and new physics in Higgs pair production**
Maxime Gouzevitch, Alexandra Oliveira, Juan Rojo, Rogerio Rosenfeld, Gavin P. Salam, Veronica Sanz, arXiv:1303.6636

**Pulling Out All the Stops: Searching for RPV SUSY with Stop-Jets**

**Measuring boosted tops in semi-leptonic ttbar events for the standard model and beyond**
Mihailo Backovic, Ofir Gabizon, Jose Juknevich, Gilad Perez, Yotam Soreq, arXiv:1311.2962

**Reconstructing singly produced top partners in decays to Wb**
Nicolas Gutierrez Ortiz, James Ferrando, Deepak Kar, Michael Spannowsky, arXiv:1403.7490

**Hadronic b’ search at the LHC with top and W taggers**
Shuo Yang, Ji Jiang, Qi-Shu Yan, Xiaoran Zhao, arXiv:1405.2514
“Plain” jet clustering is mainstream
- All main algorithms implemented in a standardized, efficient and public way, used by all LHC collaborations and phenomenologists

Jet substructure techniques also largely mature
- tens of different approaches:
  - risk of fragmentation and poor readability
  - use of standard, public tools rather than ad hoc ones helps consolidation of the field
- are further improvements in performance becoming hard to achieve?
- many approaches only tested at MC level: robustness?
  - analytical control should become a must
There may still be room for further improvement in jet substructure techniques

To avoid fragmenting the field, and make progress efficient, we should

- Introduce techniques motivated by analytical arguments, not simply MC testing
- Ensure that they enjoy a **good analytical calculability**
  - very little reason to introduce today a novel substructure technique that does not enjoy a decent calculability, unless HUGE improvement can be shown (and still, it should be justifiable and robust)
- Provide a **public implementation** (e.g. in the FastJet contrib project, http://fastjet.hepforge.org/contrib, public repository for third-party contributions)
Soft Drop declustering

Larkoski, Marzani, Soyez, Thaler, van Tilburg, 2014

Decluster and drop softer constituent unless

\[
\text{Soft Drop Condition: } \frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left( \frac{\Delta R_{12}}{R_0} \right)^\beta
\]

i.e. remove wide-angle soft radiation from a jet

The paper contains

✓ analytical calculations and comparisons to Monte Carlos
✓ study of effect of non-perturbative corrections
✓ performance studies

Example of SoftDrop performance when used as a boosted W tagger
FastJet Contrib

The fastjet-contrib space is intended to provide a common location for access to 3rd party extensions of FastJet.

Download the current version: fjcontrib-1.011 (released 6 April 2014), which contains these contributions. Changes relative to earlier versions are briefly described in the NEWS file.

After downloading and unpacking, enter the fjcontrib-1.011/ directory and then run

```
./configure [--fastjet-config=FILE] [--prefix=...] [...] 
makes 
makes check   # optional
maked install
```

By default the package installs to the same directories as the FastJet installation.

A contribution named "SomeContrib" is usually accessed by including "fastjet/contrib/SomeContrib.hh" in your C++ file, and linking with -lSomeContrib.

Developers who wish to develop their own contribution or submit new ones should use svn to checkout the contrib framework,

```
svn checkout http://fastjet.hepforge.org/svn/contrib/trunk fjcontrib
```

then run a local script to get the current set of individual contribs

```
cd fjcontrib/
scripts/update-contribs.sh
```

and follow the instructions in the README and DEVEL-GUIDELINES files.
Version 1.011 of FastJet Contrib is distributed with the following packages

<table>
<thead>
<tr>
<th>Package</th>
<th>Version</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConstituentSubtractor</td>
<td>1.0.0</td>
<td>README NEWS</td>
</tr>
<tr>
<td>EnergyCorrelator</td>
<td>1.0.1</td>
<td>README NEWS</td>
</tr>
<tr>
<td>GenericSubtractor</td>
<td>1.2.0</td>
<td>README NEWS</td>
</tr>
<tr>
<td>JetCleanser</td>
<td>1.0.0</td>
<td>README NEWS</td>
</tr>
<tr>
<td>JetFFMoments</td>
<td>1.0.0</td>
<td>README NEWS</td>
</tr>
<tr>
<td>JetsWithoutJets</td>
<td>1.0.0</td>
<td>README NEWS</td>
</tr>
<tr>
<td>Nsubjettiness</td>
<td>1.0.3</td>
<td>README NEWS</td>
</tr>
<tr>
<td>ScJet</td>
<td>1.1.0</td>
<td>README NEWS</td>
</tr>
<tr>
<td>SubjetCounting</td>
<td>1.0.1</td>
<td>README NEWS</td>
</tr>
<tr>
<td>VariableR</td>
<td>1.0.1</td>
<td>README NEWS</td>
</tr>
</tbody>
</table>