Progress in defining jets for the LHC

Grégory Soyez
Brookhaven National Laboratory

in collaboration with G. Salam, M. Cacciari and J. Rojo
**Plan**

- **Foreword**: why jets? what are they?
  - introducing the basic terminology/concepts

- **Part 1**: building solid jet definitions
  - new algorithms to meet the fundamental requirements

- **Part 2**: optimizing jets in $pp$ collisions
  - which jet algorithm is best suited?
    - how to quantify the reconstruction efficiency
    - Results without pileup
    - Results with pileup (subtraction)
Foreword: why jets? what are they?
General (over)simplified picture

Hard scattering \((2 \rightarrow n)\)

computed exactly at \(\mathcal{O}(\alpha_s^p)\)

\[ gg \rightarrow gg, \quad gg \rightarrow ggg, \]
\[ gg \rightarrow gggg, \]
\[ gg \rightarrow H \rightarrow b\bar{b}, \]
\[ gg \rightarrow t\bar{t} \rightarrow \mu\nu_\mu b\bar{b}q\bar{q}, \]
\[ gg \rightarrow Z' \rightarrow q\bar{q}, \ldots \]
General (over)simplified picture

Hard scattering ($2 \rightarrow n$)

**Parton level**

\[ \approx \text{resummed collinear div.} \]

\[ \sum_i \alpha_s^i \log^i (p_t^2/\mu^2) \]

**Hadron level**: hadronisation

**Underlying event**

beam remnants interactions

\[ \Rightarrow \text{soft background} \]
General (over)simplified picture

Hard scattering ($2 \rightarrow n$)

**Parton level**

$\approx$ resummed collinear div.

$\sum_i \alpha_s^i \log^i \left( \frac{p_t^2}{\mu^2} \right)$

**Hadron level**: hadronisation

**Underlying event**

beam remnants interactions

$\Rightarrow$ soft background

**Pileup**

$\approx$ uniform soft background
General (over)simplified picture

**Hard scattering (2 → n)**

**Parton level**

\[ \approx \text{resummed collinear div.} \]

\[ \sum_i \alpha_s^i \log^i \left( \frac{p_t^2}{\mu^2} \right) \]

**Hadron level**: hadronisation

**Underlying event**

beam remnants interactions

⇒ soft background

**Pileup**

\[ \approx \text{uniform soft background} \]

“Jets” \( \equiv \) hard partons

Parton ambiguous

⇒ multiple jet definitions
Two classes of algorithms

<table>
<thead>
<tr>
<th>Class 1: recombination</th>
<th>Class 2: cone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Successive recombinations of the “closest”(^{(a)}) pair of particle</td>
<td>find directions of energy flow</td>
</tr>
<tr>
<td>Nice perturbative behaviour</td>
<td>(\equiv) stable cones(^{(b)})</td>
</tr>
<tr>
<td>Often used in (e^\pm e^\pm, e^\pm p)</td>
<td>Often used in (pp)</td>
</tr>
</tbody>
</table>

\(^{(a)}\) Distance: (stop when \(d_{\text{min}} > R\))

\[
k_t: \quad d_{i,j} = \min(k_{t,i}^2, k_{t,j}^2)(\Delta \phi_{i,j}^2 + \Delta y_{i,j}^2)
\]

Aachen/Cam.: \(d_{i,j} = \Delta \phi_{i,j}^2 + \Delta y_{i,j}^2\)

\(^{(b)}\) stable cones (radius \(R\)) such that:

the total momentum of its contents points in the direction of its centre
How the cone works...

- Seeded (iterative) approaches: iterate from an initial position until stable
  - seed = initial particle
  - seed = midpoint between stable cones found at first step
- One has to deal with overlapping stable cones: 2 subclasses
Seeded (iterative) approaches: iterate from an initial position until stable

- seed = initial particle
- seed = midpoint between stable cones found at first step

Class 2(a): cone with split-merge (ex.: JetClu, Atlas, MidPoint):

\[
\tilde{p}_{t,\text{shared}} > f \tilde{p}_{t,\text{min}} \\
\tilde{p}_{t,\text{shared}} \leq f \tilde{p}_{t,\text{min}}
\]
Seeded (iterative) approaches: iterate from an initial position until stable
- seed = initial particle
- seed = midpoint between stable cones found at first step

Class 2(a): cone with split-merge (ex.: JetClu, Atlas, MidPoint):
\[ \tilde{p}_{t,\text{shared}} > f \tilde{p}_{t,\text{min}} \]
\[ \tilde{p}_{t,\text{shared}} \leq f \tilde{p}_{t,\text{min}} \]

Class 2(b): cone with progressive removal (ex.: Iterative Cone)
- iterate from the hardest seed
- remove the stable cone as a jet and start again

Idea: “regular/circular” jets
**Recombination:**
- $k_t$ algorithm
- Cambridge/Aachen alg.

**Cone:**
- CDF JetClu
- CDF MidPoint
- D0 (run II) Cone
- PxCone
- ATLAS Cone
- CMS Iterative Cone
- PyCell/CellJet
- GetJet
Part 1

21st century: towards a solid toolkit
SNOWMASS accords, Tevatron 1990 (i.e. old!):

Several important properties that should be met by a jet definition are:

1. Simple to implement in an experimental analysis;
2. Simple to implement in the theoretical calculation;
3. Defined at any order of perturbation theory;
4. Yields finite cross section at any order of perturbation theory;
5. Yields a cross section that is relatively insensitive to hadronization.

i.e. usable by theoreticians (e.g. finite perturbative results)
and experimentalists (e.g. fast enough, not much UE sensitivity)
Speed improvement

Speeding up the $k_t$ and Cam/Aachen algorithms using computational-geometry techniques: $O(N^3) \rightarrow O(N \log N)$

C++ implementation in FastJet

http://www.fastjet.fr (M. Cacciari, G. Salam, G.S.)
More refined clustering ("2nd generation of algorithms")

Cambridge+Filtering algorithm:

- Cluster with Aachen/Cambridge and radius $R$
- For each jet, recluster it with Aachen/Cambridge and radius $R_{\text{sub}}$
  keep only $n_{\text{sub}}$ hardest sub-jets of the initial jet
More refined clustering ("2nd generation of algorithms")

Cambridge+Filtering algorithm:

- Cluster with Aachen/Cambridge and radius $R$
- For each jet, recluster it with Aachen/Cambridge and radius $R_{\text{sub}}$
  keep only $n_{\text{sub}}$ hardest sub-jets of the initial jet

**Aim:** remove the soft background

Properties:

- Proven to improve jet reconstruction, in $H \rightarrow b\bar{b}$
  
  [J.Butterworth, A.Davison, M.Rubin, G.Salam, 08]
- Additional parameters that deserve appropriate studies
- We will use the simplest choice: $R_{\text{sub}} = R/2$, $n_{\text{sub}} = 2$
QCD probability for gluon bremsstrahlung at angle $\theta$ and $\perp$-mom. $k_t$:

$$dP \propto \alpha_s \frac{d\theta}{\theta} \frac{dk_t}{k_t}$$

Two divergences:

- **Collinear**
  - $\theta \approx 0$

- **Soft**
  - $k_t \ll p_t$
QCD probability for gluon bremsstrahlung at angle $\theta$ and $\perp$-mom. $k_t$:

$$dP \propto \alpha_s \frac{d\theta}{\theta} \frac{dk_t}{k_t}$$

Two divergences:

- **Collinear**
  $$|\theta \approx 0$$

- **Soft**
  $$p_t \quad k_t \ll p_t$$

For pQCD to make sense, the (hard) jets should not change when

- one has a collinear splitting
  *i.e.* replaces one parton by two at the same place $(\eta, \phi)$

- one has a soft emission *i.e.* adds a very soft gluon
IR unsafety of the Midpoint alg

\[ p_t \]

\[ -1 \quad 0 \quad 1 \quad 2 \quad 3 \]

\[ \phi \]
IR unsafety of the Midpoint alg
IR unsafety of the Midpoint alg
IR unsafety of the Midpoint alg
IR unsafety of the Midpoint alg

\[ \begin{align*}
\phi & \quad 0 \quad 1 \quad 2 \quad 3 \\
pt & \quad 0 \quad 100 \quad 200 \quad 300 \quad 400
\end{align*} \]
IR unsafety of the Midpoint alg
IR unsafety of the Midpoint alg
IR unsafety of the Midpoint alg
IR unsafety of the Midpoint alg
IR unsafety of the Midpoint alg

Grégory Soyez

Stony Brook, NY, USA, Novembre 20th 2008

Jets at the LHC – p. 13/40
IR unsafety of the Midpoint alg

Stable cones:

Midpoint: \{1,2\} & \{3\}

Seedless: \{1,2\} & \{3\} & \{2,3\}

Jets: \(f = 0.5\)
**IR unsafety of the Midpoint alg**

Stable cones:
- **Midpoint:** \{1,2\} & \{3\}
- **Seedless:** \{1,2\} & \{3\} & \{2,3\}

Jets: \(f = 0.5\)
- **Midpoint:** \{1,2\} & \{3\}
- **Seedless:** \{1,2,3\}
IR unsafety of the Midpoint alg

Stable cones:
- Midpoint: \{1,2\} & \{3\} & \{2,3\}
- Seedless: \{1,2\} & \{3\} & \{2,3\}
- Jets: \(f = 0.5\)
  - Midpoint: \{1,2\} & \{3\}
  - Seedless: \{1,2,3\}

Stable cone missed \rightarrow IR unsafety of the midpoint algorithm
Solution: SISCone

- **Solution**: use a seedless approach, find **ALL** stable cones

- **Naive approach**: check stability of each subset of particle
Solution: SIS Cone

- **Solution**: use a seedless approach, find **ALL** stable cones

- **Naive approach**: check stability of each subset of particle
  Complexity is $O(N^{2^N})$
  $\Rightarrow$ definitely unrealistic: $10^{17}$ years for $N = 100$

- **Midpoint complexity**: $O(N^3)$
**Solution: SISCones**

- **Solution**: use a seedless approach, find **ALL** stable cones

- **Midpoint complexity**: $O(N^3)$

**Idea**: use geometric arguments

![Diagram](image)

- Each enclosure can be moved (in any direction) until it touches a point
- ... then rotated until it touches a second one

$\Rightarrow$ Enumerate all pairs of particles
- with 2 circle orientations and 4 possible inclusion/exclusion
$\Rightarrow$ find all enclosures
**Solution: SISCones**

- **Solution**: use a seedless approach, find **ALL** stable cones.

- **Midpoint complexity**: $O(N^3)$

**Idea**: use geometric arguments

⇒ Enumerate all pairs of particles
   with 2 circle orientations and 4 possible inclusion/exclusion
   → find all enclosures

- **Complexity**: $O(N^3)$, with improvements: $O(N^2 \log(N))$

→ C++ implementation: Seedless Infrared-Safe Cone algorithm (SISCones)

G. Salam, G.S., JHEP 04 (2007) 086; http://projects.hepforge.org/siscone

NB.: also available from FastJet
Execution timings:

- faster than midpoint without seed threshold
- at least as fast as midpoint with seed thresholds
Physical impact (2)

\[ \frac{(\text{Midpoint-SISCone})}{\text{SISCone}} \]

\[ \frac{d\sigma}{dp_t} \mid \text{Midpoint}(1) \]

\[ \frac{d\sigma}{dp_t} \mid \text{SISCone} \]

\[ \frac{d\sigma}{dp_t} \mid \text{Parton-level} \]

\[ pp \ \sqrt{s} = 14 \text{ TeV} \]

(b) 

Inclusive cross-section:

- Effect of a few %
- Less UE sensitivity
Inclusive cross-section:

- effect of a few %
- less UE sensitivity

Masses in 3-jet events:

- effects $\sim 45\%$
- Important for LHC!
Coll. unsafety of the iterative cone
Coll. unsafety of the iterative cone
Coll. unsafety of the iterative cone

Grégory Soyez
Stony Brook, NY, USA, Novembre 20th 2008
Jets at the LHC – p. 17/40
Coll. unsafety of the iterative cone
Before collinear splitting: 1 jet

After collinear splitting: 2 jets

→ collinear unsafety of the iterative cone algorithm
Come back to recombination-type algorithms:

\[ d_{ij} = \min(k_{t,i}^{2p}, k_{t,j}^{2p}) \left( \Delta \phi_{ij}^2 + \Delta \eta_{ij}^2 \right) \]

- \( p = 1 \): \( k_t \) algorithm
- \( p = 0 \): Aachen/Cambridge algorithm
Come back to recombination-type algorithms:

\[ d_{ij} = \min(k_{t,i}^{2p}, k_{t,j}^{2p}) \left( \Delta \phi_{ij}^2 + \Delta \eta_{ij}^2 \right) \]

- \( p = 1 \): \( k_t \) algorithm
- \( p = 0 \): Aachen/Cambridge algorithm
- \( p = -1 \): anti-\( k_t \) algorithm  \[\text{[M.Cacciari, G.Salam, G.S., JHEP 04 (08) 063]}\]
Come back to recombination-type algorithms:

\[ d_{ij} = \min(k_{t,i}^{2p}, k_{t,j}^{2p}) \left( \Delta \phi_{ij}^2 + \Delta \eta_{ij}^2 \right) \]

- \( p = 1 \): \( k_t \) algorithm
- \( p = 0 \): Aachen/Cambridge algorithm
- \( p = -1 \): anti-\( k_t \) algorithm [M.Cacciari, G.Salam, G.S., JHEP 04 (08) 063]

Why should that be related to the iterative cone ?!

- “large \( k_t \) \( \Rightarrow \) small distance”
  - \( i.e. \) hard partons “eat” everything up to a distance \( R \)
  - \( i.e. \) circular/regular jets, jet borders unmodified by soft radiation
- infrared and collinear safe
Hard event + homogeneous soft background

anti-\textit{k}_t \text{ is soft-resilient}

more later in this talk...
Execution timings:

As fast as the (fast) $k_t$ ([M. Cacciari, G. Salam, 06])
Recombination:
- \( k_t \) algorithm
- Cambridge/Aachen alg.
- anti-\( k_t \) algorithm

Cone:
- CDF JetClu
- CDF MidPoint
- D0 (run II) Cone
- PxCone
- ATLAS Cone
- CMS Iterative Cone
- PyCell/CellJet
- GetJet
- SISCone

4 available safe algorithms

All accessible from FastJet
Part 2

Jets in $pp$ collisions

(a) Choosing the adapted jet definition

Sample processes to study

We analyse 3 processes typical of kinematic reconstructions:

- \( Z' \to q\bar{q} \to 2 \text{ jets} \) and \( H \to gg \to 2 \text{ jets} \):
  
  simple environment: identify 2 jets and reconstruct \( M_{Z',H} \)
  
  source of monochromatic quark/gluon jets
  
  scale dependence: mass of the \( Z'/H \) varied between 100 GeV and 4 TeV
  
  ficticious narrow \( Z', H \)

- \( t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}bq\bar{q}b \to 6 \text{ jets} \):
  
  complex environment: identify 6 jets and reconstruct 2 top
  
  balance between reconstruction efficiency and identification

with

- the 5 IRC-safe algorithms: \( k_t \), Cambridge, anti-\( k_t \), SIS Cone, Cam+filtering

- jet radius varied between 0.1 and 1.5
We reconstruct histograms

How can we quantify the reconstruction efficiency?
Measure of the jet reconstruction efficiency

- Forget about measures related to parton-jet matching, → use the reconstructed mass peak

- Forget about fits depending on the shape of the peak

⇒ maximise the signal over background ratio \( \frac{S}{\sqrt{B}} \):
$Q_{f=z}^w(JA, R) = \text{minimal width of a window containing a fraction } f = z \text{ of the events}$

Fixed signal, minimal width (background)
figure of merit for quality measure

\[ Q_{w=\sqrt{M}}^w(JA, R) = \frac{1}{N} \text{maximal number of events in a window of width } x\sqrt{M} \]
it intuitively does what it should
it intuitively does what it should

relates to a signal significance (assuming constant background)

\[
\frac{\Sigma(JD_1)}{\Sigma(JD_2)} = \frac{N_{\text{signal}}}{\sqrt{N_{\text{bkg}}}} \]

\[
= \sqrt{\frac{Q^w_{f=z}(JD_2)}{Q^w_{f-z}(JD_1)}} = \frac{Q^w_{f=x\sqrt{M}}(JD_2)}{Q^w_{f=x\sqrt{M}}(JD_1)}
\]

\text{minimal } Q \equiv \text{better signal-to-background ratio}
it intuitively does what it should

relates to a signal significance (assuming constant background)

\[
\frac{\Sigma(JD_1)}{\Sigma(JD_2)} = \left[ \frac{N_{\text{signal}}}{\sqrt{N_{\text{bkg}}}} \right]_{\text{JD}} = \sqrt{\frac{Q_{f=z}^w(JD_2)}{Q_{f=z}^w(JD_1)}} = \frac{Q_{f=x}^w\sqrt{M}(JD_2)}{Q_{f=x}^w\sqrt{M}(JD_1)}
\]

minimal \( Q \equiv \) better signal-to-background ratio

we can associate an effective luminosity ratio

\[
\rho_{\mathcal{L}}(JD_2/JD_1) = \frac{\mathcal{L} \text{ needed with } JD_1}{\mathcal{L} \text{ needed with } JD_2} = \left[ \frac{\Sigma(JD_1)}{\Sigma(JD_2)} \right]^2
\]

e.g. \( \rho_{\mathcal{L}} = 2 \equiv JD_1 \) has \( \sqrt{2} \) the significance of \( JD_2 \)

\( \equiv JD_2 \) requires 2 times the integrated luminosity to achieve the same discriminative power.
Examples: best quality measures

Grégory Soyez
Stony Brook, NY, USA, Novembre 20th 2008
Jets at the LHC – p. 29/40

Allows to

- extract the best radius $R_{\text{best}}$
- compare the different algorithm
SISCon and Cam+filtering perform better

$R_{\text{best}}$ strongly depends on the mass
Same conclusions for gluon jets with slightly larger $R$
Luminosity ratios

Mandatory at the LHC:
Not choosing the best alg. AND $R$ can be very costly for new discoveries

Note: typical choice, $R \sim 0.5$
Part 2

Jets in $pp$ collisions

(b) pileup effect (jet areas & subtraction)
Pileup $\approx$ uniform soft background that shifts jets to higher $p_t$

... that needs to be subtracted!

$\Rightarrow$ Using jet areas!
**Basic idea:** [M. Cacciari, G. Salam, 08]

\[ p_{t, \text{subtracted}} = p_{t, \text{jet}} - \rho_{\text{pileup}} \times \text{Area}_{\text{jet}} \]
**Pileup subtraction**

**Basic idea:** [M.Cacciari, G.Salam, 08]

\[ p_{t, \text{subtracted}} = p_{t, \text{jet}} - \rho_{\text{pileup}} \times \text{Area}_{\text{jet}} \]

- **Jet area:** [M.Cacciari, G.Salam, G.S., 08]
  - region where the jet catches infinitely soft particles (active/passive)
  - tractable analytically in pQCD

*Example: area corrections from QCD radiation*

\[
\langle A(p_{t,1}, R) \rangle = \mathcal{A}_{\text{parton}}(R) + \frac{C_{F,A}}{\pi b_0} \log \left( \frac{\alpha_s(Q_0)}{\alpha_s(Rp_t)} \right) \pi R^2 d
\]

- area \(\neq \pi R^2\)
- area scaling violations

<table>
<thead>
<tr>
<th>(d)</th>
<th>passive</th>
<th>active</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k_t)</td>
<td>0.5638</td>
<td>0.519</td>
</tr>
<tr>
<td>Cam</td>
<td>0.07918</td>
<td>0.0865</td>
</tr>
<tr>
<td>SISCon</td>
<td>-0.06378</td>
<td>0.1246</td>
</tr>
<tr>
<td>anti-(k_t)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Grégory Soyez
Stony Brook, NY, USA, Novembre 20th 2008
Jets at the LHC – p. 35/40
Pileup subtraction

**Basic idea:** [M. Cacciari, G. Salam, 08]

\[ p_t,\text{subtracted} = p_t,\text{jet} - \rho_{\text{pileup}} \times \text{Area}_{\text{jet}} \]

- **Jet area:** [M. Cacciari, G. Salam, G.S., 08]
  - region where the jet catches infinitely soft particles (active/passive)
  - tractable analytically in pQCD

- **Pileup density per unit area:** \( \rho_{\text{pileup}} \)
  - e.g. estimated from the median of \( p_t,\text{jet} / \text{Area}_{\text{jet}} \)

implemented in FastJet on an event-by-event basis
**Subtraction at work**

- $M_{Z'} = 100$ GeV
- $k_t$, $R = R_{\text{best}}$

---

**Graph 1:**
- No PU
- Low-lumi PU
- High-lumi PU

---

**Graph 2:**
- $M_{Z'} = 100$ GeV
- $k_t$, $R = 0.5$

---

**Subtraction**
PU effects summary

Subtraction ⇒ (i) large improvement, (ii) $R_{\text{best}} \sim$ unchanged
Additional soft background has 2 effects:

- **Throw soft particles in the hard jet**: dealt with by subtraction
- **Modify the hard scattering (back-reaction)**
  - can be pointlike or diffuse
  - **gain**: $p_2$ gained when adding $p_m$
  - **loss**: $p_2$ lost when adding $p_m$
Additional soft background has 2 effects:

- Throw soft particles in the hard jet: dealt with by subtraction
- Modify the hard scattering (back-reaction)
  - can be pointlike or diffuse
  - tractable analytically (similar to areas)
- $k_t \gtrsim$ Cambridge > SISCon $\gg$ anti-$k_t$
Conclusions

Message 1: IRC safety is mandatory

Midpoint and the iterative cone IR or Collinear unsafe (at $\mathcal{O}(\alpha_s^4)$)

<table>
<thead>
<tr>
<th>Observable</th>
<th>1st miss cones at</th>
<th>Last meaningful order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inclusive jet cross section</td>
<td>NNLO</td>
<td>NLO</td>
</tr>
<tr>
<td>3 jet cross section</td>
<td>NLO</td>
<td>LO (NLO in NLOJet)</td>
</tr>
<tr>
<td>$W/Z/H + 2$ jet cross sect.</td>
<td>NLO</td>
<td>LO (NLO in MCFM)</td>
</tr>
<tr>
<td>jet masses in 3 jets</td>
<td>LO</td>
<td>none (LO in NLOJet)</td>
</tr>
</tbody>
</table>

+ We do not want the theoretical efforts to be wasted

- Note: 1 order worse for JetClu of the ATLAS Cone!
- All IRC-safe algorithms available from FastJet (http://www.fastjet.fr)
Message 2: flexibility in jet finding at the LHC

- Optimal jet definition (see also http://quality.fastjet.fr)
  - $R_{\text{best}} \sim 0.5$ at 100 GeV, $R_{\text{best}} \sim 1$ at 1 TeV
  - important to choose $R_{\text{best}}$, SISCones and Cam+filt. slightly better
  - same for quark and gluon jets, larger $R_{\text{best}}$ for gluons
  - TODO: understand this analytically/ improve clustering (e.g. filtering)

- Pileup subtraction using jet areas
  - Jet areas: clearly defined, analytic control
  - Simple systematic pileup subtraction
  - Same conclusions as without pileup
  - TODO: deal with fluctuating background (e.g. heavy ions)