A new jet reconstruction algorithm for lepton colliders

Based on preprint arXiv:1404.4294

M. Boronat, J. Fuster, I. García, M. Vos, IFIC (CSIC/UVEG) Valencia, Spain

With help from Gavin Salam (CERN)

America's Workshop on Linear Colliders

FermiLab, 14 May 2014
Introduction

Jet reconstruction is a crucial step in many physics analyses at the LC

Future high-energy lepton colliders present an environment that differs in several important respects from that encountered at the Z-pole

*Do we need to rethink jet reconstruction? Which algorithms are most suitable?*
A brief history of sequential recombination algorithms

JADE 1980s

\[ y_{ij} = \frac{E_i^2, E_j^2}{Q^2}(1 - \cos \theta_{ij}) \]

Experience on e^+e^- data at Z-pole

Generalised e^+e^- k_t algorithm

\[ d_{ij} = \min(E_i^2, E_j^2)(1 - \cos \theta_{ij})/(1 - \cos R) \]
\[ d_{iB} = E_i^2 \]

Feed back into e^+e^- algorithms

Durham or e^+e^- k_t algorithm (LEP and SLC)

\[ d_{ij} = 2\min(E_i^2, E_j^2)(1 - \cos \theta_{ij}) \]

Adapt to hadron colliders

\[ d_{ij} = \min(p_{Ti}^{2n}, p_{Tj}^{2n}) \Delta R_{ij}^{2n}/R^{2n} \]
\[ d_{iB} = p_{Ti}^{2n} \]

n=0: Cambridge-Aachen
n=1: Longitudinally invariant k_t
n=-1: Anti-k_t (LHC default)

Time to rethink e^+e^- algorithms!!

Moretti, Lonblad, Sjostrand, JHEP9808 (1998)
All algorithms available in FastJet

Marcel Vos, AWLC14
Boost invariance at hadron colliders

• At **hadron colliders** the **partons** that participate in the hard process generally **carry different fractions of the initial** hadron **energy**.

• The final state acquires a substantial **Lorentz boost** along the beam axis.
  
  • LHC di-jets: $\beta_z \sim 1$
  
  • LHC tt: $\beta_z \sim 0.5$

• Replace the **[energy, polar angle]** basis by

  **[transverse momentum, rapidity]**
Boost invariance at lepton colliders

- Photons emitted by the incoming beam particles (**Initial State Radiation**) can carry away a significant fractions of the nominal center-of-mass energy.
- For $e^+e^- \rightarrow Z/\gamma^* \rightarrow f\bar{f}$ process, with $m_f < M_Z/2$ a large fraction of events **returns** to the Z-pole.
- **In most interesting processes** at a future lepton collider ISR plays a much **less important role**.
- At lepton colliders we must deal with minor boosts and the basis $[E, \theta]$ remains the most natural choice.
• The **pile-up** at the **LHC** is a serious challenge that has led to a large body of work on mitigation and correction methods

• **LEP** or **SLC** presented effectively negligible background

• The $\gamma\gamma \rightarrow \text{hadrons}$ background at **CLIC** has **strong** impact on jet reconstruction performance [CLIC CDR, Marshall & Thomson, arXiv:1308.4537]

• Less pronounced, but **non-negligible** impact on ILC physics [many studies, arXiv:1307.8102]

• Using hadron collider algorithms can reduce these problems [CLIC CDR]
The Valencia jet algorithm

A new clustering jet reconstruction algorithm that combines the good features of lepton collider algorithms, in particular the Durham-like distance criterion;

\[ d_{ij} = \min \left( E_j^{2\beta}, E_j^{2\beta} \right) \left( 1 - \cos \theta_{ij} \right) / R^2 \]

with the robustness against background of the longitudinally invariant \( k_t \) algorithm

\[ d_{iB} = p_T^{2\beta} \]

The exponent \( \beta \) allows to tune the background rejection level

The algorithm has been implemented as a plugin for the FastJet package and is available in fjcontrib

https://fastjet.hepforge.org/trac/browser/contrib/contribs/ValenciaJetAlgorithm
Comparison of the distance criteria

Two test particles with constant energy ($E = 1$ GeV) and fixed polar angle separation (100 mrad)
Two test particles with constant energy ($E = 1$ GeV) and fixed polar angle separation (100 mrad)

Rotating from central to forward region
As the two-particle system rotates into the forward region, the distance $d_{ij}$ of longitudinally invariant $k_t$ decreases ($\Delta \eta$ increases, $p_T$ decreases)

Traditional $e^+e^-$ algorithms and Valencia have constant $d_{ij}$
The ratio of the inter-particle distance and the beam distance: \( \frac{d_{ij}}{d_{iB}} \) drives the robustness to (forward) background: the decision to assign the particle to final-state or beam jets depends on this ratio (and R)

Long. inv. \( k_t \)'s robustness is indeed due to its increasing \( \frac{d_{ij}}{d_{iB}} \) ratio

Valencia with \( \beta=1 \) is similar (by design) to long. inv. \( k_t \)
Jet reconstruction performance

CLIC di-boson (ZZ) production @ 500 GeV

Reconstruct Particle Flow objects using PANDORA

Reconstruct jets (exclusive, n=4)

Form Z boson candidates, selecting best jet pairs

Chosen to facilitate comparison with Marshall&Thomson, CLIC CDR

No background: it doesn't really matter which algorithm you pick
CLIC di-boson (ZZ) production @ 500 GeV
+ 300 BX of γγ → hadrons

Reconstruct Particle Flow objects using PANDORA
+ quality and timing cuts

Reconstruct jets
(exclusive, n=4)

Form Z boson candidates, selecting best jet pairs

Nominal background: Durham is severely affected, longitudinally invariant $k_t$ and Valencia OK
The choice of parameters corresponds to the optimal setting determined in a scan over a broad range of parameters.
Parameter optimisation: $\beta$ scan

![Graph showing $M_z$ (GeV) vs. $R = 1.0$, $\beta = 0.5$, $\beta = 1.0$, and $\beta = 1.5$.]

![Graph showing $m^2_{RMS}$ (GeV) vs. $\beta$.]
Jet reconstruction performance

The previous results in numbers: central value, width of the Z-boson mass peak and RMS$_{90}$

\[
\sqrt{s} = 500 \text{ GeV, no background overlay}
\]

<table>
<thead>
<tr>
<th></th>
<th>$m_Z$</th>
<th>$\sigma_Z$</th>
<th>RMS$_{90}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durham</td>
<td>90.6</td>
<td>5.4</td>
<td>13.8</td>
</tr>
<tr>
<td>long. inv. $k_t$</td>
<td>90.4</td>
<td>5.3</td>
<td>14.3</td>
</tr>
<tr>
<td>Valencia</td>
<td>90.3</td>
<td>5.2</td>
<td>12.5</td>
</tr>
</tbody>
</table>

\[
\sqrt{s} = 500 \text{ GeV, 0.3 $\gamma\gamma \rightarrow hadrons$ events/BX}
\]

<table>
<thead>
<tr>
<th></th>
<th>$m_Z$</th>
<th>$\sigma_Z$</th>
<th>RMS$_{90}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durham</td>
<td>101.1</td>
<td>13.6</td>
<td>28.8</td>
</tr>
<tr>
<td>long. inv. $k_t$</td>
<td>92.0</td>
<td>9.0</td>
<td>17.2</td>
</tr>
<tr>
<td>Valencia</td>
<td>92.5</td>
<td>9.2</td>
<td>16.2</td>
</tr>
</tbody>
</table>

e$^+$e$^-$ style algorithm can compete with hadron collider algorithm
Jet reconstruction performance

IFIC/LAL study of ILC lepton+jets tt @ 500 GeV, [arXiv:1307.8102]

Event Generation

Whizard 1.95

Reconstruct Particle Flow objects using PANDORA

Reconstruct jets (exclusive, n=4)

The signal is reconstructed by choosing the combination of b quark jet and W boson that minimises the following equation

\[ d^2 = \left( \frac{m_{\text{cand.}} - m_t}{\sigma_{m_t}} \right)^2 + \left( \frac{E_{\text{cand.}} - E_{\text{beam}}}{\sigma_{E_{\text{cand.}}}} \right)^2 + \left( \frac{p_b^* - 68}{\sigma_{p_b^*}} \right)^2 + \left( \frac{\cos \theta_{bW} - 0.23}{\sigma_{\cos \theta_{bW}}} \right)^2 \]
We consider four jet reconstruction algorithms

• **Durham** algorithm
• **Generic e+e− k_t** algorithm with beam jets with R = 1
• **Longitudinally invariant k_t** algorithm with R = 1.5
• **Valencia** algorithm with R = 1.2 and β = 0.8.

The choice of parameters corresponds to the optimal setting determined in a scan over a broad range of parameters.

Durham is affected by γγ -> hadrons, longitudinally invariant k_t and Valencia OK.
Degradation of all jet-related measurements due to $\gamma\gamma \rightarrow \text{hadrons}$ background

<table>
<thead>
<tr>
<th>RMS$_{90}$ [GeV]</th>
<th>$E_{4j}$</th>
<th>$E_W$</th>
<th>$m_W$</th>
<th>$E_t$</th>
<th>$m_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durham</td>
<td>23.2</td>
<td>19.6</td>
<td>20.3</td>
<td>19.5</td>
<td>21.4</td>
</tr>
<tr>
<td>$e^+e^- k_t$</td>
<td>25.6</td>
<td>20.8</td>
<td>21.6</td>
<td>20.5</td>
<td>22.8</td>
</tr>
<tr>
<td>long. inv. $k_t$</td>
<td>21.7</td>
<td>18.4</td>
<td>18.9</td>
<td>18.4</td>
<td>20.1</td>
</tr>
<tr>
<td>Valencia</td>
<td>21.4</td>
<td>18.0</td>
<td>18.8</td>
<td>18.2</td>
<td>20.0</td>
</tr>
</tbody>
</table>

Durham and e+e- $k_t$ significantly degraded.

Long. inv. $k_t$ algorithm and Valencia offer better reconstruction for all hadronic observables.

Four-jet system

Hadronic top candidate

Hadronic W candidate
Multi-TeV operation

CLIC at 3 TeV, pairs of highly boosted top quarks

Reconstruct two jets!

A very narrow top mass peak (FWHM = 24 GeV)

Long tail due to radiation

Better performance than long, invariant kt and much better than Durham.

Studies with realistic background are ongoing
A key requirement to jet algorithms

From Salam & Soyez, JHEP 0705 (2007)
An infinitely soft particle cannot lead to a new (hard) [jet] being found...
...it makes no sense for the structure of multi-hundred GeV jets to change radically just because hadronisation, the underlying event or pileup threw a 1 GeV particle in between them.

The sequential recombination structure underlying the Valencia algorithm is generally thought to be intrinsically safe

A large number of standard IR-safety tests were performed on the FastJet plugin. All succeeded.
Conclusions

- $\gamma\gamma \rightarrow$ hadrons bkg. forces us to rethink jet reconstruction algorithms. The performance of classical $e^+e^-$ algorithms is severally degraded.

- The Valencia jet algorithm retains the natural inter-particle distance criterion for $e^+e^-$ collisions and offers robust performance in the presence of the background levels expected at lepton colliders.

- Shown to work on several benchmark analyses. Pre-print out on the arXiv since a few weeks: Boronat, Garcia, Vos, A new jet reconstruction algorithm for lepton colliders, arXiv:1404.4294.

- Do try this at home! https://fastjet.hepforge.org/trac/browser/contrib/contribs/ValenciaJetAlgorithm
And, please, contact us if you do: Ignacio.Garcia at ific.uv.es.