Physics Letters B 750 (2015) 95–99

Contents lists available at ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

A robust jet reconstruction algorithm for high-energy lepton colliders

M. Boronat, J. Fuster, I. García, E. Ros, M. Vos *

IFIC (CSIC/UVEG), Valencia, Spain

ARTICLE INFO

Article history:
Received 9 February 2015
Received in revised form 30 July 2015
Accepted 25 August 2015
Available online 28 August 2015
Editor: H. Weerts

Keywords:
Jet reconstruction
Sequential recombination algorithm
Future lepton colliders
ILC
CLIC

ABSTRACT

We propose a new sequential jet reconstruction algorithm for future lepton colliders at the energy frontier. The Valencia algorithm combines the natural distance criterion for lepton colliders with the greater robustness against backgrounds of algorithms adapted to hadron colliders. Results on a detailed Monte Carlo simulation of $t\bar{t}$ and ZZ production at future linear e^+e^- colliders (ILC and CLIC) with a realistic level of background overlaid, show that it achieves better performance in the presence of background than the classical algorithms used at previous e^+e^- colliders.

1. Introduction

Experiments at lepton and hadron colliders use jet algorithms to cluster the collimated sprays of particles that form in processes with asymptotically free quarks and gluons in the final state. The first modern sequential recombination algorithms were developed for e^+e^- colliders operated at the Z-pole (a detailed historical account is found in Reference [1]). At the heart of the jet algorithm – and crucial to the definition of jets themselves – is a criterion to define the distance between two particles. In popular algorithms used at e^+e^- colliders the distance combines information on the angle between the particles and the energy of (the softest of the two) particles. Sequential recombination algorithms were adapted to the environment at hadron colliders in the early 1990s. At the Large Hadron Collider the large majority of analyses is based on inclusive jet reconstruction with the anti-k_t algorithm [2].

An intense R&D programme exists to develop the technology required for an e^+e^- collider with a center-of-mass energy well beyond that of previous lepton colliders. A linear e^+e^- collider can attain center-of-mass energies from several 100 GeV to several TeV [3,4]. The possibility of a large circular e^+e^- collider that can reach a center-of-mass energy of approximately 350 GeV [5] is also explored, as well as a muon collider [6]. Such machines present an environment that differs in several important respects from that encountered at the Z-pole. In this Letter we explore which jet reconstruction algorithms are most suitable for the e^+e^- colliders with a center-of-mass energy from 100 GeV to several TeV.

We start our discussion with a brief recapitulation of the properties of the most popular clustering algorithms in Section 2. We present a proposal for a new jet algorithm in Section 3. In Section 4 the key features of this algorithm are compared to popular algorithms. In Section 5 the Monte Carlo simulation setup that we used to benchmark the performance of the algorithms is introduced. Finally, in Sections 6 and 7 we present the results for top quark pair and di-boson (ZZ) production at the ILC and CLIC, in a realistic environment including the relevant background. In Section 8 we summarize the most important findings of this work.

2. Overview of jet reconstruction algorithms based on sequential recombination

The first modern clustering algorithm with a simple sequential recombination scheme algorithm is the JADE algorithm developed in the middle of the 1980s [7,8]. The distance y_{ij} assigned to any pair of particles i and j is given by:

$$y_{ij} = \frac{E_i^2 E_j^2}{Q^2}(1 - \cos\theta_{ij}) \qquad (1)$$

where E_i and E_j denote the energy of the two particles, Q is the total energy of the event, and θ_{ij} is the angle between the two particles. At each step the algorithm merges the pair of particles with

* Corresponding author.
E-mail address: marcel.vos@ific.uv.es (M. Vos).

http://dx.doi.org/10.1016/j.physletb.2015.08.055

the smallest distance y_{ij}. This process continues until the smallest distance exceeds a value y_{cut} (*inclusive* clustering) or a previously defined number of jets is obtained (*exclusive* clustering).

In the Durham or e^+e^- k_t algorithm [9] used extensively at LEP and SLC the distance between particles i and j is modified to depend on the minimum of the energies E_i and E_j, rather than the product E_iE_j:

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

For sufficiently small angles the distance reduces to the transverse momentum squared of the softer particle relative to the harder one. The distance measure is thus proportional to the squared inverse of the splitting probability for one parton k into partons i and j in the soft and collinear limit.

Jet reconstruction at hadron colliders presents a number of additional difficulties. The incoming beams radiate gluons that can form jets. Only a fraction of the energy of the composite projectiles is transferred in the hard parton–parton process and a hadron remnant continues to travel down the beam pipe. An important consequence is that the system formed by the reaction products is typically not at rest in the laboratory frame.[1] Clustering algorithms were adapted to meet these challenges in the 1990s.

The first important modification of the algorithms is the addition of so-called *beam jets*, introduced in Reference [10]. Any particle with a beam distance $d_{iB} = p_{Ti}^{2n}$ smaller than any d_{ij} is not merged with any other particle, but is associated to the beam jet. These are not considered part of the visible final state. Thus, the soft, collinear radiation emitted by the incoming hadrons and the hadron remnant travelling in the very forward and backward direction are discarded.

To cope with the boost along the beam direction, analyses at hadron colliders replace the particle energy E_i with its transverse momentum p_{Ti} and the angular distance between the particles $(1 - \cos\theta_{ij})$ with $\Delta R_{ij} = \sqrt{(\Delta\phi)^2 + (\Delta y)^2}$, where y denotes the rapidity. In the longitudinally invariant k_t algorithm [11,12] the distance criterion is based on the same observables "to improve the factorization properties [of the algorithm] and [achieve] closer correspondence to experimental practice [...]" [11]. We rewrite the generic inter-particle distance as follows:

$$d_{ij} = min(p_{Ti}^{2n}, p_{Tj}^{2n})\frac{\Delta R_{ij}^2}{R^2} \tag{3}$$

where R is the radius parameter that determines the maximum area of the jet. Setting n in the exponent to 1 yields the longitudinally invariant k_t algorithm. Alternative choices of the exponent yield the Cambridge–Aachen algorithm ($n = 0$), or the anti-k_t algorithm ($n = -1$), the default jet reconstruction algorithm at the LHC.

Finally, one can add beam jets to the k_t algorithm for e^+e^- experiments. This yields an algorithm we refer to as the generic e^+e^- k_t algorithm, with inter-particle distance:

$$d_{ij} = min(E_i^2, E_j^2)(1 - \cos\theta_{ij})/(1 - \cos R) \tag{4}$$

and beam distance given by $d_{iB} = E_i^2$.

[1] For di-jet production at the LHC $\beta_z = v_z/c$ of the di-jet system is very close to 1 and even a massive system such as a top quark pair acquires a typical $\beta_z = 0.5$. In contrast, for processes such as $e^+e^- \to ZH(\gamma)$ (Higgsstrahlung) at $\sqrt{s} = 250$ GeV and $e^+e^- \to t\bar{t}(\gamma)$ at 500 GeV β_z is smaller than 0.1 in 95% and 90% of the events, respectively. The exception to the rule is the $2 \to 2$ process $e^+e^- \to f\bar{f}(\gamma)$, with f any fermion lighter than the Z-boson, where ISR (return-to-the-Z) plays an important role.

3. The Valencia jet algorithm

Background levels at hadron colliders form an important consideration in the design of jet algorithms. The *pile-up* of several tens of minimum bias events on each bunch crossing at the LHC is a serious challenge that has led to a large body of work on mitigation and correction methods. In comparison, previous lepton colliders, such as LEP or SLD, presented an environment with essentially negligible background. Future lepton colliders are in between these two extremes. While very far from the background levels of the LHC, detailed studies of the $\gamma\gamma \to$ *hadrons* background at the ILC or CLIC have shown a non-negligible impact on the jet reconstruction performance [4,13]. Among several proposals to mitigate its effect, the use of the longitudinally invariant k_t algorithm, intended for hadron colliders, has led to the greatest improvement of the robustness [4].

We propose a new clustering jet reconstruction algorithm for future e^+e^- colliders that maintains a Durham-like distance criterion based on and can compete with the background resilience of the longitudinally invariant k_t algorithm. The algorithm has the following inter-particle distance:

$$d_{ij} = min(E_i^{2\beta}, E_j^{2\beta})(1 - \cos\theta_{ij})/R^2 \tag{5}$$

For $\beta = 1$ the distance is given by the transverse momentum squared of the softer of the two particles relative to the harder one, as in the Durham algorithm. We argue that a distance based on energy and angle, as opposed to the transverse momentum and ΔR distance of hadron collider algorithms, remains the most natural choice for the e^+e^- colliders of the foreseeable future. Equation (5) provides a uniform inter-particle distance over the central and forward detectors and is in line with the natural choice of basis for the analyses at such a machine. Note that we have redefined the meaning of the radius parameter R with respect to the generalized e^+e^- algorithm with beam jets. The R^2 in the numerator yields greater freedom than the $1 - \cos R$, that is limited to the interval $[0, 2]$.

The beam distance of the Valencia algorithm is:

$$d_{iB} = p_T^{2\beta} \tag{6}$$

For $\beta = 1$ this combination of inter-particle and beam distance metrics is similar to that of the $k_\perp$ algorithm proposed in Ref. [10], with the difference that $d_{iB} = p_{ti}^2 = E_i^2 \sin^2\theta_{iB}$, whereas in Ref. [10] it was given by $2E_i^2(1 - \cos\theta_{iB})$.

The Valencia algorithm is available as a plug-in for the FastJet [14,15] package.[2]

4. Comparison of the distance criteria of sequential recombination algorithms

The choice of distance criterion defines the essence of the jet algorithm and has profound implications on its performance in a given environment. The differences between the various algorithms are most easily visualized as follows. We calculate the distance between two test particles with an energy of 1 GeV emitted at a fixed relative angle of 100 mrad. The leftmost plot in Fig. 1 shows how the distance between the two particles evolves as the system is scanned from the central detector ($\cos\theta = 0$) to the forward region ($\cos\theta = 1$).

The distance d_{ij} of the generic e^+e^- k_t algorithm is independent of polar angle, as shown in Fig. 1. The same holds for the

[2] The code can be obtained from the "contrib" area under https://fastjet.hepforge.org/contrib/.