Jet reconstruction algorithms in e^+e^- collisions

Javier Aparisi[1], Ignacio García[1], Martin Perelló[1], Ph. Roloff[2], Rosa Simoniello[2], **Marcel Vos**[1]

[1] *IFIC (UVEG/CSIC), València, Spain*
[2] *CERN, Geneva, Switzerland*

Abstract: Jet reconstruction is a key technique at future energy-frontier e^+e^- colliders. Classical e^+e^- algorithms are tested by several new challenges. In this contribution results are presented of studies into the jet reconstruction performance at high-energy e^+e^- colliders.

Introduction

A high-energy e^+e^- collider can provide precise measurements of the interactions of the Higgs boson and the top quark. The Higgs-strahlung process ($e^+e^- \to ZH$), with a maximum cross section at $\sqrt{s} \sim 250$ GeV), is accessible with a large circular machine [1][2]. The largest rings envisaged (with a circumference of 100 km) can reach the top quark pair production threshold at $\sqrt{s} \sim 350$ GeV [3]. Processes at still higher energies (vector-boson fusion Higgs production, associated production of a top quark pair and a Higgs boson, di-Higgs boson production) are accessible at a linear collider. The ILC [4][5] project envisages operation at 250 GeV and 500 GeV, with the possibility of an upgrade to 1 TeV. The CLIC [6][7] project aims for multi-TeV operation, with an initial stage at 380 GeV.

An accurate reconstruction of hadronic final states is a prerequisite for a precise measurements of Higgs boson and top quark couplings [8][9]. The linear collider detector concepts [10] achieve excellent single-particle reconstruction with highly granular calorimeters [11][12], and particle-flow algorithms [13]. Excellent jet clustering is required to take full advantage of the potential of the machine and detectors.

Challenges to jet clustering

Jet clustering at high-energy colliders differs in several respects from previous e^+e^- colliders, such as LEP or SLC. The most important effects are listed below:

- **Multi-jet final states:** processes with many jets in the final state become more important. Key measurements at the lowest energy require an accurate reconstruction of four final-state jets ($e^+e^- \to ZH$, with hadronic Higgs and Z boson decays, $e^+e^- \to t\bar{t}$ in the lepton+jets channel). Processes with six-jet, eight-jet and even ten-jet final states open up. A correct clustering of the reconstructed particles into jets turns out to be far from trivial even if the inputs to the algorithm are accurately reconstructed. In analyses such as the extraction of the Higgs self-coupling clustering has a dominant contribution to the mass resolution [14].
- **Hard emissions:** the phase space for the emission of hard gluons opens up. In some cases the distance or energy scale of the emitted gluon is no longer small compared to the typical distance between the decay products of gauge bosons. The n jets reconstructed with exclusive clustering using a sequential recombination algorithm (the standard procedure at previous lepton colliders and in benchmark studies of future e^+e^- colliders) may not correspond to

the n final-state quarks. This problem may be circumvented in the pair production of very energetic objects with hadronic decays (i.e. boosted gauge or Higgs bosons, or top quarks), by reconstructing two *fat* jets that capture the energy flow of the boosted object [15][16]. For final states with a strong hierarchy between energy scales this effect leads to failures in jet reconstruction in a small fraction of events (i.e. di-Higgs production at very high energy, where the radiated Higgs boson remains rather soft).

- **Forward processes:** t-channel processes become increasing important. At high energy the final-state products of processes such as vector-boson-fusion Higgs boson production are strongly peaked in forward and backward directions [17]. Special care is needed in the detector design and in the development of jet clustering algorithms to ensure robust jet reconstruction performance over the full polar angle coverage of the experiment.

- **Background processes:** energy flow superposed on the *signal* event can affect jet reconstruction. Where such backgrounds could safely be ignored at previous e^+e^- colliders, they may have a non-negligible effect at future installations. The $\gamma\gamma \rightarrow hadrons$ background renders the classical e^+e^- algorithms inadequate for high-energy operation of linear colliders [7][18][19][14]. At circular colliders the rate of $\gamma\gamma \rightarrow hadrons$ is several orders of magnitude smaller. The effect of synchrotron radiation, and appropriate shielding measures, remain to be evaluated.

Higher energy also has some benificial effects. An additional boost collimates the jets, so that confusion due to clustering is reduced. The relative size of non-perturbative corrections diminishes strongly with increasing center-of-mass energy.

Robust jet clustering in the presence of background

The impact of *pile-up* due to the energy flow of $\gamma\gamma \rightarrow hadrons$ was studied thoroughly as part of the CLIC conceptual design studies [7][18]. In multi-TeV operation a bunch train may deposit several TeV in the experiment. Timing cuts that select 1 ns around the *signal* bunch crossing reduce this contribution to the order of 100 GeV. In the classical approach (represented in most studies by exclusive clustering with the Durham algorithm [20]), all final-state particles are clustered into jets. The reconstructed jet properties are found to be strongly affected by the background energy flow [7][18][19][14]. At the highest energy classical e^+e^- algorithms are inadequate.

Several alternatives have been considered to achieve more robust performance in the presence of background. The generalization of the e^+e^- algorithms with a beam distance [21] yields jets with a limited area. Longitudinally invariant algorithms developed for hadron colliders [22][23] expose even less area in the forward and backward parts of the experiment, where the $\gamma\gamma \rightarrow hadrons$ background is most pronounced. In the VLC algorithm proposed in Ref. [19][14] this feature is combined with the traditional inter-particle distance of the Durham algorithm. The inter-particle and beam distance of these three classes of algorithms are given in Fig. 1, together with an indication of the jet area in the central and forward directions of the experiment.

The longitudinally invariant algorithms and VLC are found to be much more resilient than the classical e^+e^- algorithms. The VLC algorithm outperforms the hadron collider algorithms in the most demanding environment [19][14].

At the ILC, with a much smaller $\gamma\gamma \rightarrow hadrons$ rate and a much larger bunch spacing, the effect of this background is much less pronounced. Still, a modest but non-negligible improvement of the performance can be achieved by adopting the VLC or longitudinally invariant algorithm. At

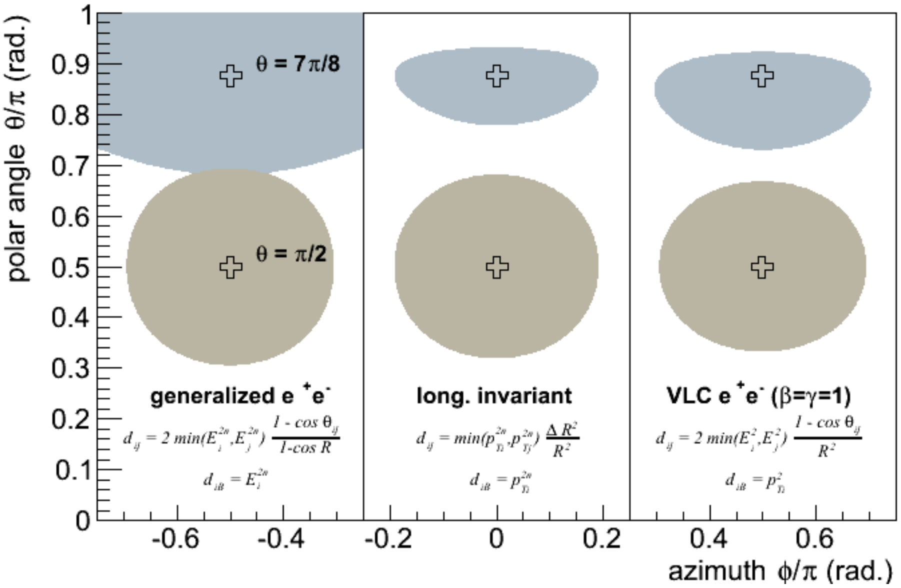

Figure 1: The area or *footprint* of jets reconstructed with a radius parameter $R = 0.5$, for the three major families of sequential recombination algorithms. The two shaded areas in each column correspond to a jet in the central detector ($\theta = \pi/2$) and to a forward jet ($\theta = 7\pi/8$). Reprinted from Ref. [14].

circular colliders the rate of $\gamma\gamma \to$ *hadrons* is so low that its effect is expected to be negligible. The impact of synchrotron radiation has not been evaluated in detail.

Perturbative and non-perturbative corrections

Even in the ideal case of a perfect detector response the jet energy differs from that of the final-state parton due to a number of effects. The largest correction for jets with a finite size can be addressed in a perturbative calculation. A smaller, non-perturbative correction remains. Ref. [14] (following Ref. [24]) estimates both on simulated $e^+e^- \to q\bar{q}$ and $e^+e^- \to t\bar{t}$ from a Monte Carlo event generator (the MadGraph5_aMC@NLO package [25] interfaced to Pythia 8.180 [26]).

As expected, the relative perturbative energy correction - evaluated as the difference between the energy of the jet reconstructed on stable particles and the quark produced in the hard scatter - is roughly independent of center-of-mass energy and decreases as the catchment area of the jet increases. The correction is smallest for the classical e^+e^- algorithms is smallest. For the longitudinally invariant and VLC algorithms with a radius parameter of 1.5, the average correction is less than 2%. A tail towards lower reconstructed jet energy remains, however, resulting in a median correction of approximately 5%.

Non-perturbative corrections are estimated as the difference between the energy of jets reconstructed on stable particles and at the parton-level. The results for $e^+e^- \to q\bar{q}$ production at $\sqrt{s} = 250$ GeV (left panel) and $e^+e^- \to t\bar{t}$ production at $\sqrt{s} = 3$ TeV (right panel) are shown in Fig. 2. The distributions are again asymmetric, with significant differences between the mean and median corrections (which can even have opposite signs). Non-perturbative corrections are reduced for large radius parameter, but do not vanish completely. The most striking effect is the strong

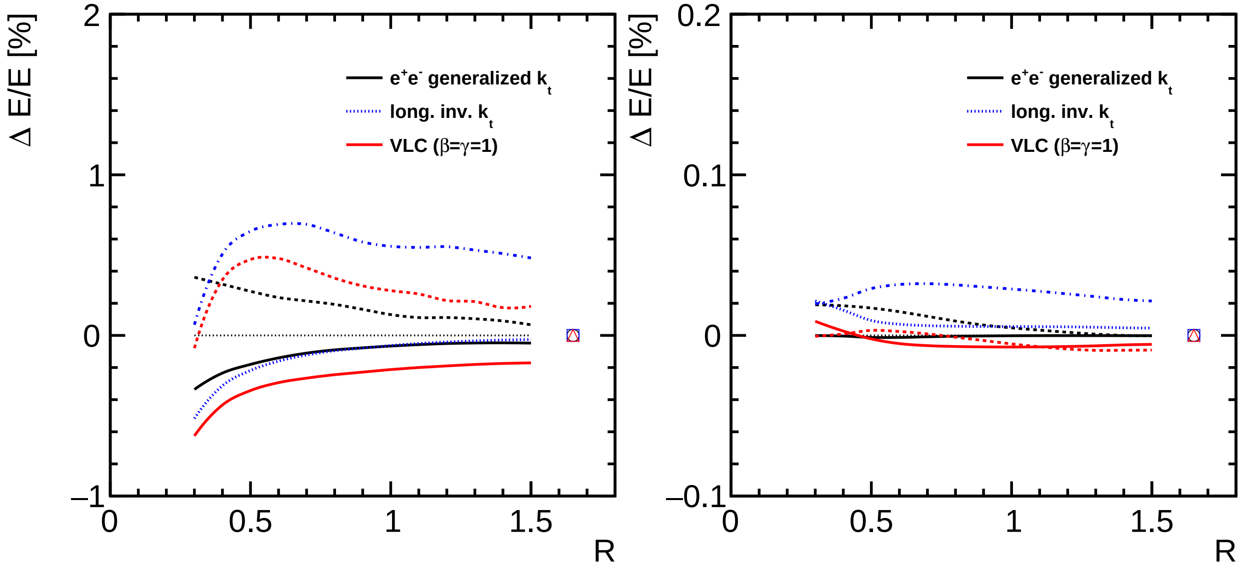

Figure 2: Non-perturbative jet energy corrections to the jet energy as a function of the jet radius parameter R in $e^+e^- \to q\overline{q}$ production at $\sqrt{s}$ = 250 GeV (left panel) and $e^+e^- \to t\overline{t}$ production at $\sqrt{s}$ = 3 TeV (right panel). The continuous line corresponds to the median relative correction, the dashed line to the median. Results are shown for three algorithms: the generalized e^+e^- algorithm, the longitudinally invariant k_t algorithm and the VLC algorithm with $\beta = 1$. Reprinted from Ref. [14].

reduction of the size of these corrections at high energy (the range of the Y-axis is reduced by a factor ten in the rightmost panel).

The non-perturbative contribution to the invariant mass of the jet is more important, with relative correction of several tens of % at low energy and a few % for $\sqrt{s} = 3$ TeV (for $R = 1$). Algorithms with the e^+e^- inter-particle distance (generalized Durham and VLC) converge slightly faster than the longitudinally invariant algorithms.

Conclusions

Jet clustering at future energy-frontier e^+e^- facilities faces several challenges that are new to e^+e^- colliders. Multi-jet final states and final states with very forward jets are much relevant than at LEP or SLC. Hard gluons emitted in events with relatively soft gauge bosons challenge exclusive jet reconstruction. Background such as $\gamma\gamma \to$ *hadrons* or synchrotron radiation may affect the jet reconstruction performance.

Detailed benchmark studies of the ILC and CLIC design study groups that jet clustering is the limiting factor in the analysis of complex multi-jet final states. Studies with realistic background levels show that for the most demanding environment in multi-TeV operation classical e^+e^- algorithms are inadequate. Longitudinally invariant algorithms and the VLC algorithm proposed in Ref. [19] prove to be much more resilient. Classical e^+e^- algorithms, on the other hand, show faster convergence of energy corrections with the radius parameter of the algorithm. The non-perturbative correction associated to hadronization decreases strongly with center-of-mass energy: from 1% at $\sqrt{s}$ = 250 GeV to less than a per mil at $\sqrt{s}$ = 3 TeV.

Given the relevance of jet clustering for the potential of high-energy e^+e^- colliders we encourage new ideas and more exhaustive performance study for existing algorithms (e.g. those of Refs. [27][28]).

Acknowledgement

The authors acknowledge the effort of the ILC and CLIC detector & physics groups in putting together the simulation infrastructure used to benchmark the algorithms. This work benefited from services provided by the ILC Virtual Organisation, supported by the national resource providers of the EGI Federation. The authors would like to thank Gavin Salam and Jesse Thaler for help with the VLC algorithm and the FastJet team for guidance creating the plugin code. The IFIC group is financially supported by the Spanish national programme for particle physics (FPA2015-65652-C4-3-R, MINECO/FEDER).

References

[1] M. Bicer, H. Duran Yildiz, I. Yildiz, G. Coignet, M. Delmastro, et al., *First Look at the Physics Case of TLEP*, [arXiv:1308.6176].

[2] CEPC-SPPC Study Group, *CEPC-SPPC Preliminary Conceptual Design Report. 1. Physics and Detector, IHEP-CEPC-DR-2015-01 (2015)*, .

[3] P. Janot, *Top-quark electroweak couplings at the FCC-ee*, JHEP **04** (2015) 182, [arXiv:1503.0132].

[4] K. Fujii et al., *Physics Case for the International Linear Collider*, [arXiv:1506.0599].

[5] H. Baer, T. Barklow, K. Fujii, Y. Gao, A. Hoang, et al., *The International Linear Collider Technical Design Report - Volume 2: Physics*, [arXiv:1306.6352].

[6] **CLICdp, CLIC** Collaboration, M. J. Boland et al., *Updated baseline for a staged Compact Linear Collider*, [arXiv:1608.0753].

[7] L. Linssen, A. Miyamoto, M. Stanitzki, and H. Weerts, *Physics and Detectors at CLIC: CLIC Conceptual Design Report*, [arXiv:1202.5940].

[8] M. Thomson, *Model-independent measurement of the $e^+ e^- \to HZ$ cross section at a future $e^+ e^-$ linear collider using hadronic Z decays*, Eur. Phys. J. **C76** (2016), no. 2 72, [arXiv:1509.0285].

[9] M. S. Amjad et al., *A precise characterisation of the top quark electro-weak vertices at the ILC*, Eur. Phys. J. **C75** (2015), no. 10 512, [arXiv:1505.0602].

[10] H. Abramowicz et al., *The International Linear Collider Technical Design Report - Volume 4: Detectors*, [arXiv:1306.6329].

[11] **CALICE** Collaboration, C. Adloff et al., *Electromagnetic response of a highly granular hadronic calorimeter*, JINST **6** (2011) P04003, [arXiv:1012.4343].

[12] **CALICE** Collaboration, C. Adloff et al., *Hadronic energy resolution of a highly granular scintillator-steel hadron calorimeter using software compensation techniques*, JINST **7** (2012) P09017, [arXiv:1207.4210].

[13] J. S. Marshall and M. A. Thomson, *The Pandora Software Development Kit for Pattern Recognition*, Eur. Phys. J. **C75** (2015), no. 9 439, [arXiv:1506.0534].

[14] M. Boronat, J. Fuster, I. Garcia, P. Roloff, R. Simoniello, and M. Vos, *Jet reconstruction at high-energy lepton colliders*, [arXiv:1607.0503].

[15] M. H. Seymour, *Searches for new particles using cone and cluster jet algorithms: A Comparative study*, Z. Phys. **C62** (1994) 127–138.

[16] A. Abdesselam et al., *Boosted objects: A Probe of beyond the Standard Model physics*, Eur. Phys. J. **C71** (2011) 1661, [arXiv:1012.5412].

[17] J. Fuster, S. Heinemeyer, C. Lacasta, C. Marinas, A. Ruiz Jimeno, and M. Vos, *Forward tracking at the next e+ e- collider. Part I. The Physics case*, JINST **4** (2009) P08002, [arXiv:0905.2038].

[18] J. Marshall, A. Muennich, and M. Thomson, *Performance of Particle Flow Calorimetry at CLIC*, Nucl.Instrum.Meth. **A700** (2013) 153–162, [arXiv:1209.4039].

[19] M. Boronat, J. Fuster, I. Garcia, E. Ros, and M. Vos, *A robust jet reconstruction algorithm for high-energy lepton colliders*, Phys. Lett. **B750** (2015) 95–99, [arXiv:1404.4294].

[20] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. Webber, *New clustering algorithm for multi - jet cross-sections in e+ e- annihilation*, Phys.Lett. **B269** (1991) 432–438.

[21] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, Eur.Phys.J. **C72** (2012) 1896, [arXiv:1111.6097].

[22] S. Catani, Y. L. Dokshitzer, M. Seymour, and B. Webber, *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, Nucl.Phys. **B406** (1993) 187–224.

[23] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys.Rev. **D48** (1993) 3160–3166, hep-ph/9305266].

[24] M. Dasgupta, L. Magnea, and G. P. Salam, *Non-perturbative QCD effects in jets at hadron colliders*, JHEP **0802** (2008) 055, [arXiv:0712.3014].

[25] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, [arXiv:1405.0301].

[26] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867, [arXiv:0710.3820].

[27] I. W. Stewart, F. J. Tackmann, J. Thaler, C. K. Vermilion, and T. F. Wilkason, *XCone: N-jettiness as an Exclusive Cone Jet Algorithm*, JHEP **11** (2015) 072, [arXiv:1508.0151].

[28] H. Georgi, *A Simple Alternative to Jet-Clustering Algorithms*, [arXiv:1408.1161].