

Progress with Particle Flow Calorimetry

Mark Thomson

University of Cambridge, Cavendish Laboratory
J.J. Thomson Avenue, Cambridge, CB30HE, UK

One of the most important requirements for a detector at the ILC is good jet energy resolution. It is widely believed that the particle flow approach to calorimetry is the key to achieving the ILC goal of a di-jet invariant mass resolution $\sigma_m/m < \Gamma_Z/m_Z$. This paper describes the current performance of the PANDORAPFA particle flow algorithm. For simulated light quark jets in the Tesla TDR detector, the jet energy resolution achieved is better than $\sigma_E/E \approx 3.4\%$ for jet energies in the range 45 – 250 GeV. This represents the first demonstration that Particle Flow Calorimetry can reach the ILC jet energy resolution goals.

1 Introduction

Many of the interesting physics processes at the ILC will be characterised by multi-jet final states, often accompanied by charged leptons and/or missing transverse energy associated with neutrinos or the lightest super-symmetric particles. The reconstruction of the invariant masses of two or more jets will provide a powerful tool for event reconstruction and identification. Unlike at LEP, where kinematic fitting[1] enabled precise jet-jet invariant mass reconstruction almost independent of the jet energy resolution, at the ILC this mass reconstruction will rely on the detector having excellent jet energy resolution. The ILC goal is to achieve a mass resolution for $W \rightarrow q\bar{q}$ and $Z \rightarrow q\bar{q}$ decays which is comparable to their natural widths, i.e. $\sigma_m/m = 2.7\% \approx \Gamma_W/m_W \approx \Gamma_Z/m_Z$. For a traditional calorimetric approach, a jet energy resolution of $\sigma_E/E = \alpha/\sqrt{E(\text{GeV})}$ leads to a di-jet mass resolution of roughly $\sigma_m/m = \alpha/\sqrt{E_{jj}(\text{GeV})}$, where E_{jj} is the energy of the di-jet system. At the ILC typical di-jet energies will be in the range 150 – 350 GeV, suggesting the goal of $\sigma_E/E \sim 0.3/\sqrt{E(\text{GeV})}$. This is more than a factor two better than the best jet energy resolution achieved at LEP, $\sigma_E/E = 0.6(1 + |\cos\theta|)/\sqrt{E(\text{GeV})}$ [2]. Meeting the jet energy resolution goal is a major factor in the overall design of a detector for the ILC.

2 The Particle Flow Approach to Calorimetry

It is widely believed that the most promising strategy for achieving the ILC jet energy goal is the particle flow analysis (PFA) approach to calorimetry. In contrast to a purely calorimetric measurement, PFA requires the reconstruction of the four-vectors of all visible particles in an event. The reconstructed jet energy is the sum of the energies of the individual particles. The momenta of charged particles are measured in the tracking detectors, while the energy measurements for photons and neutral hadrons are obtained from the calorimeters. The crucial step in PFA is to assign the correct calorimeter hits to reconstructed particles, requiring efficient separation of nearby showers.

Measurements of jet fragmentation at LEP have provided detailed information on the particle composition of jets (e.g. [3, 4]). On average, after the decay of short-lived particles, roughly 62% of the energy of jets is carried by charged particles (mainly hadrons), around

27% by photons, about 10% by long-lived neutral hadrons (*e.g.* n/K_L^0), and around 1.5% by neutrinos. Assuming calorimeter resolutions of $\sigma_E/E = 0.15/\sqrt{E(\text{GeV})}$ for photons and $\sigma_E/E = 0.55\sqrt{E(\text{GeV})}$ for hadrons, a jet energy resolution of $0.19/\sqrt{E(\text{GeV})}$ is obtained with the contributions from tracks, photons and neutral hadrons shown in Tab. 1. In practise it is not possible to reach this level of performance for two main reasons. Firstly, particles travelling at small angles to the beam axis will not be detected. Secondly, and more importantly, it is not possible to perfectly associate all energy deposits with the correct particles. For example, if a photon is not resolved from a charged hadron shower, the photon energy is not counted. Similarly, if part of charged hadron shower is identified as a separate cluster the energy is effectively double-counted. This *confusion* degrades particle flow performance. Because confusion, rather than calorimetric performance, determines the overall performance, the jet energy resolution achieved will not, in general, be of the form $\sigma_E/E = \alpha/\sqrt{E(\text{GeV})}$.

The crucial aspect of particle flow is the ability to correctly assign calorimeter energy deposits to the correct reconstructed particles. This places stringent requirements on the granularity of electromagnetic and hadron calorimeters. Consequently, particle flow performance is one of the main factors driving the overall ILC detector design. It should be noted that the jet energy resolution obtained for a particular detector concept is the combination of the intrinsic detector performance and the performance of the PFA software.

Component	Detector	Energy Fraction	Energy Res.	Jet Energy Res.
Charged Particles (X^\pm)	Tracker	$\sim 0.6 E_{\text{jet}}$	$10^{-4} E_{X^\pm}^2$	$< 3.6 \times 10^{-5} E_{\text{jet}}^2$
Photons (γ)	ECAL	$\sim 0.3 E_{\text{jet}}$	$0.15 \sqrt{E_\gamma}$	$0.08 \sqrt{E_{\text{jet}}}$
Neutral Hadrons (h^0)	HCAL	$\sim 0.1 E_{\text{jet}}$	$0.55 \sqrt{E_{h^0}}$	$0.17 \sqrt{E_{\text{jet}}}$

Table 1: Contributions from the different particle components to the jet-energy resolution (all energies in GeV). The table lists the approximate fractions of charged particles, photons and neutral hadrons in a jet and the assumed single particle energy resolution.

3 The PandoraPFA Particle Flow Algorithm

PANDORAPFA is a C++ implementation of a PFA algorithm running in the MARLIN[5, 6] framework. It was designed to be sufficiently generic for ILC detector optimisation studies and was developed and optimised using events generated with the MOKKA[7] program, which provides a GEANT4[8] simulation of the Tesla TDR[9] detector concept. The PANDORAPFA algorithm performs both calorimeter clustering and particle flow in eight main stages:

i) Tracking: for the studies presented in this paper, the track *pattern recognition* is performed using Monte Carlo information[5]. The track parameters are extracted using a helical fit. The projections of tracks onto the front face of the electromagnetic calorimeter are calculated using helical fits (with no accounting for energy loss along the track). Neutral particle decays resulting in two charged particle tracks (V^0 s) are identified by searching from pairs of non-vertex tracks which are consistent with coming from a single point in the central tracking chamber. Kinked tracks from charged particle decays to a single charged particle and a number of neutrals are also identified, as are interactions in the tracking volume (prongs).

ii) Calorimeter Hit Selection and Ordering: isolated hits, defined on the basis of proximity to other hits, are removed from the initial clustering stage. The remaining hits are ordered into *pseudo-layers* which follow the detector geometry so that particles propagating outward from the interaction region will cross successive pseudo-layers. The assignment of hits to pseudo-layers removes the dependence of the algorithm on the explicit detector geometry whilst following the actual geometry as closely as possible. Within each pseudo-layer hits are ordered by decreasing energy.

iii) Clustering: the main clustering algorithm is a cone-based forward projective method working from innermost to outermost pseudo-layer. In this manner hits are added to clusters or are used to seed new clusters. Throughout the clustering algorithm clusters are assigned a direction (or directions) in which they are growing. The algorithm starts by *seeding* clusters using the projections of reconstructed tracks onto the front face of the calorimeter. The initial direction of a track-seeded cluster is obtained from the track direction. The hits in each subsequent pseudo-layer are then looped over. Each hit, i , is compared to each clustered hit, j , in the previous layer. The vector displacement, \mathbf{r}_{ij} , is used to calculate the parallel and perpendicular displacement of the hit with respect to the unit vector(s) $\hat{\mathbf{u}}$ describing the cluster propagation direction(s), $d_{\parallel} = \mathbf{r}_{ij} \cdot \hat{\mathbf{u}}$ and $d_{\perp} = |\mathbf{r}_{ij} \times \hat{\mathbf{u}}|$. Associations are made using a cone-cut, $d_{\perp} < d_{\parallel} \tan \alpha + \beta D_{\text{pad}}$, where α is the cone half-angle, D_{pad} is the size of a sensor pixel in the layer being considered, and β is the number of pixels added to the cone radius. Different values of α and β are used for the ECAL and HCAL with the default values set to $\{\tan \alpha_{\text{E}} = 0.3, \beta_{\text{E}} = 1.5\}$, and $\{\tan \alpha_{\text{H}} = 0.5, \beta_{\text{H}} = 2.5\}$ respectively. Associations may be made with hits in the previous 3 layers. If no association is made, the hit is used to seed a new cluster. This procedure is repeated sequentially for the hits in each pseudo-layer (working outward from ECAL front-face).

iv) Topological Cluster Merging: by design the initial clustering errs on the side of splitting up true clusters rather than clustering energy deposits from more than one particle. The next stage of the algorithm is to merge clusters from tracks and hadronic showers which show clear topological signatures of being associated. A number of track-like and shower-like topologies are searched for including looping minimum ionising tracks, back-scattered tracks and showers associated with a hadronic interaction. Before clusters are merged, a simple cut-based photon identification procedure is applied. The cluster merging algorithms are only applied to clusters which have not been identified as photons.

v) Statistical Re-clustering: The previous four stages of the algorithm were found to perform well for jets with energy less than ~ 50 GeV. However, at higher energies the performance degrades rapidly due to the increasing overlap between hadronic showers from different particles. To address this, temporary associations of tracks with reconstructed calorimeter clusters are made. If the track momentum is incompatible with the energy of the associated cluster re-clustering is performed. If $E_{\text{CAL}} - E_{\text{TRACK}} > 3.5\sigma_E$, where σ_E is the energy resolution of the cluster, the clustering algorithm, described in *iii)* and *iv)* above, is reapplied to the hits in that cluster. This is repeated, using successively smaller values of the α s and β s in the clustering finding algorithm (stage *iii)*) until the cluster splits to give an acceptable track-cluster energy match. Similarly, if $E_{\text{TRACK}} - E_{\text{CAL}} > 3.5\sigma_E$ the algorithm attempts to merge additional clusters with the cluster associated with the track. In doing so high energy clusters may be split as above.

vi) Photon Recovery and Identification: A more sophisticated photon identification algorithm is then applied to the clusters. The longitudinal profile of the energy deposition,

ΔE , as a function of number of radiation lengths from the shower start, t , is compared to that expected for an electromagnetic shower:

$$\Delta E \approx E_0 \frac{(t/2)^{a-1} e^{-t/2}}{\Gamma(a)} \Delta t \quad \text{where} \quad a = 1.25 + \frac{1}{2} \ln \frac{E_0}{E_c},$$

E_0 is the shower energy and E_c is the critical energy which is tuned give the appropriate average shower profile in the ECAL. The resulting level of agreement is used to improve the tagging of photons and to recover primary photons merged with hadronic showers.

vii) Fragment Removal: At this stage there is still a significant number of “neutral clusters” (not identified as photons) which are *fragments* of charged particle hadronic showers. An attempt is made to identify these clusters and merge them with the appropriate parent cluster. All non-photon neutral clusters, i , are compared to all clusters with associated tracks, j . For each combination a quantity, e_{ij} , is defined which encapsulates the evidence that cluster i is a fragment from cluster j . The requirement, R_{ij} , for the clusters to be merged, *i.e.* the cut on e_{ij} , depends on the location of the neutral cluster and the change in the χ^2 for the track–cluster energy consistency that would occur if the clusters were merged, $\Delta\chi^2 = (E_{\text{TRACK}} - E_j)^2/\sigma_E^2 - (E_{\text{TRACK}} - E_i - E_j)^2/\sigma_E^2$. If $e_{ij} > R_{ij}$ the clusters are merged. This *ad hoc* procedure gives extra weight to potential cluster matches where the consistency of the track momentum and associated cluster energy improves as a result of the match.

viii) Formation of Particle Flow Objects: The final stage of the algorithm is to create Particle Flow Objects (PFOs) from the results of the clustering. Tracks are matched to clusters on the basis of the distance closest approach of the track projection into the first 10 layers of the calorimeter. If a hit is found within 50 mm of the track extrapolation an association is made. If an identified kink is consistent with being from a $K^\pm \rightarrow \mu^\pm \nu$ or $K^\pm \rightarrow \mu^\pm \nu$ decay the parent track is used to form the PFO. The reconstructed PFOs are written out in LCIO[5] format.

4 Current Performance

Fig. 1a) shows an example of a PANDORAPFA reconstruction of a 100 GeV jet from a $Z \rightarrow u\bar{u}$ decay at $\sqrt{s} = 200$ GeV. The ability to track particles in the high granularity Tesla TDR calorimeter can be seen clearly. Fig. 1b) shows the total PFA reconstructed energy for $Z \rightarrow uds$ events with $|\cos\theta_{q\bar{q}}| < 0.7$, where $\theta_{q\bar{q}}$ is the polar angle of the generated $q\bar{q}$ system. These events were generated at $\sqrt{s} = 91.2$ GeV using the Tesla TDR detector model with a HCAL consisting of 63 layers and in total 6.9 interaction lengths. The root-mean-square deviation from the mean (rms) of the distribution is 2.8 GeV. However, quoting the rms as a measure of the performance over-emphasises the importance of the tails. It is conventional to quote the performance in terms of rms_{90} , which is defined as the rms in the smallest range of reconstructed energy which contains 90% of the events. For the data shown in Fig. 1b) the resolution achieved is $\text{rms}_{90}/E = 0.23/\sqrt{E(\text{GeV})}$, equivalent to a single jet energy resolution of 3.3%. The majority of interesting ILC physics will consist of final states with at least six fermions, setting a “typical” energy scale for ILC jets as approximately 85 GeV and 170 GeV at $\sqrt{s} = 500$ GeV and $\sqrt{s} = 1$ TeV respectively. Fig. 2 shows the jet energy resolution for $Z \rightarrow uds$ events plotted against $|\cos\theta_{q\bar{q}}|$ for four different values of \sqrt{s} . The current performance is summarised in Tab. 2. The observed jet energy

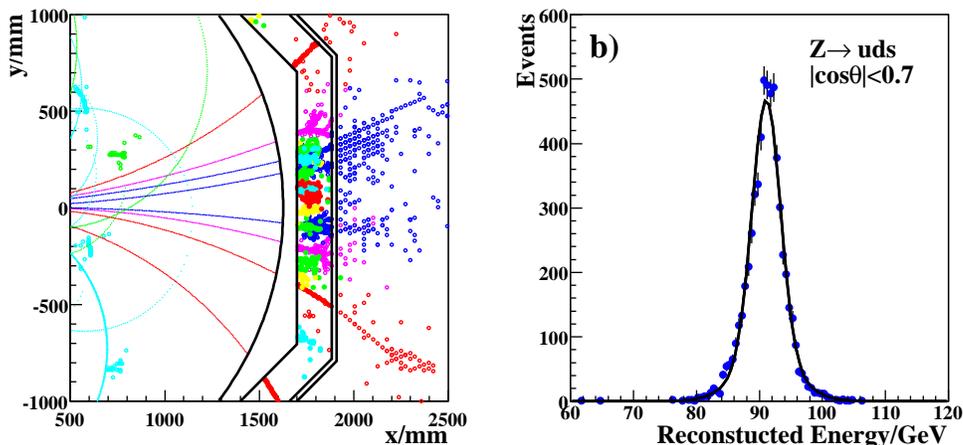


Figure 1: a) PANDORAPFA reconstruction of a 100 GeV jet in the MOKKA simulation of the Tesla TDR detector. b) The total reconstructed energy from reconstructed PFOs in $Z \rightarrow uds$ events for initial quark directions within the polar angle acceptance $|\cos\theta_{q\bar{q}}| < 0.7$. The solid line shows a fit to two Gaussians with a common mean; the broader Gaussian is constrained to contain 25% of the events. The narrow Gaussian has a width of 2.2 GeV.

resolution in simulated events is not described by the expression $\sigma_E/E = \alpha/\sqrt{E(\text{GeV})}$. This is not surprising, as the particle density increases it becomes harder to correctly associate the calorimetric energy deposits to the particles and the confusion term increases. The table also shows a measure of the single jet energy resolution, obtained by dividing rms_{90} by $\sqrt{2}$. For the jet energies considered (45 – 250 GeV) the fractional energy resolution is significantly better than the ILC requirement of 3.8% obtained from the consideration of gauge boson di-jet mass resolution. It should be noted that in a real physics analysis the performance is likely to be degraded by jet finding, jet-pairing and the presence of missing energy from semi-leptonic heavy quark decays. Nevertheless the results presented in this paper already provide a strong indication that Particle Flow Calorimetry will be able to deliver the ILC jet energy goals and it is expected that the performance of PANDORAPFA will improve with future refinements to the algorithm.

5 Conclusions

Particle flow calorimetry is widely believed to be the key to reaching the ILC jet energy resolution goal of a di-boson mass resolution of $\sigma_m/m < 2.7\%$. Consequently, the design and optimisation of detectors for the ILC depends both on hardware and on sophisticated software reconstruction. Based on the PANDORAPFA reconstruction of simulated events in Tesla TDR detector concept, it has now been demonstrated that particle flow calorimetry can meet this goal at the ILC. This was not true at the time of LCWS06 and, thus, represents a significant step forward in the design and future optimisation of the ILC detector(s).

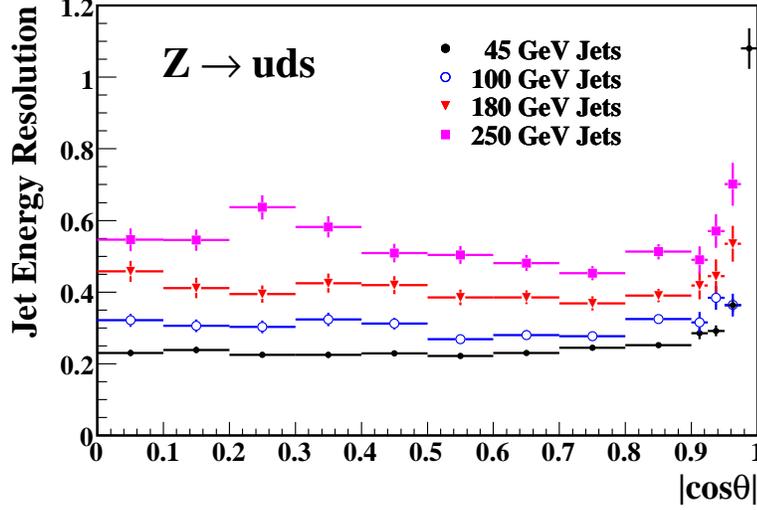


Figure 2: The jet energy resolution, defined as the α in $\sigma_E/E = \alpha\sqrt{E(\text{GeV})}$, plotted versus $\cos\theta_{q\bar{q}}$ for four different values of \sqrt{s} .

Jet Energy	rms ₉₀	rms ₉₀ / $\sqrt{E_{jj}(\text{GeV})}$	rms ₉₀ / $\sqrt{2}E_j$
45 GeV	2.2 GeV	23 %	3.3 %
100 GeV	4.1 GeV	29 %	2.9 %
180 GeV	7.4 GeV	39 %	2.9 %
250 GeV	12.0 GeV	54 %	3.4 %

Table 2: Jet energy resolution for $Z \rightarrow uds$ events with $|\cos\theta_{q\bar{q}}| < 0.7$, expressed as, rms₉₀ for the di-jet energy distribution, the effective constant α in $\text{rms}_{90}/E = \alpha(E_{jj})/\sqrt{E_{jj}(\text{GeV})}$, and the fractional jet energy resolution for a single jet.

References

- [1] M. A. Thomson, Proc. of EPS-HEP 2003, Aachen. Topical Vol. of Eur. Phys. J. C Direct (2004).
- [2] ALEPH Collaboration, D. Buskulic et al., Nucl. Inst. Meth. **A360** (1995) 481.
- [3] I.G. Knowles and G.D. Lafferty, J. Phys. **G23** (1997) 731.
- [4] M. G. Green, S. L. Lloyd, P. N. Ratoff and D. R. Ward, "Electron-Positron Physics at the Z", IoP Publishing (1998).
- [5] <http://www-flc.desy.de/ilcsoft/ilcsoftware/>.
- [6] F. Gaede, "Status of ILC-LDC Core Software", Proceedings of LCWS07, DESY, June 2007.
- [7] <http://polywww.in2p3.fr/activites/physique/geant4/tesla/www/mokka/>.
- [8] GEANT4 collaboration, S. Agostinelli et al., Nucl. Instr. and Meth. **A506** (2003) 3;
GEANT4 collaboration, J. Allison et al., IEEE Trans. Nucl. Sci. 53 (2006) 1.
- [9] TESLA Technical Design Report, DESY 2001-011, ECFA 2001-2009 (2001).