MIP Reconstruction Techniques and Minimum Spanning Tree Clustering

Wolfgang F. Mader
The University of Iowa, 203 Van Allen Hall, 52242 Iowa City, IA

The development of a tracking algorithm for minimum ionizing particles in the calorimeter and of a clustering algorithm based on the Minimum Spanning Tree approach are described. They do not depend on information from the central tracking system. Both are important components of a particle flow algorithm currently under development.

1. INTRODUCTION

An e^+e^- linear collider is a precision instrument which would provide us with the possibility to test the Standard Model with unprecedented accuracy at the electroweak scale. It will have the potential to discover new physics processes in an energy regime up to 1 TeV or beyond [1]. In addition, the results obtained at the ILC will complement and supplement those obtained at LHC in many ways [2]. In order to achieve the maximum sensitivity to the physics anticipated from this machine, the detector design has to be optimized using experimental as well as reconstruction techniques never before implemented.

The studies described here are based on the Silicon Detector (SiD) [3]. The experimental technique currently investigated is a Particle Flow Algorithm (PFA). In this document two important components of a successful PFA are described: a tracking algorithm to identify minimum ionizing particles (MIPs) in the calorimeters, and a clustering algorithm based on the Minimum Spanning Tree (MST) approach.

2. THE SiD

The reconstruction of the events based on a PFA requires that all particles involved are correctly identified. In particular, the showers from photons and those from charged or neutral hadrons have to be separable. This implies a high granularity of the calorimeters, both transversely and longitudinally.

The SiD design currently used in the simulation consists of five layers of Silicon based pixel (CCD) vertex detectors ($1\,\mathrm{cm} < r < 7\,\mathrm{cm}$) and a five-layer Silicon microstrip tracking system ($20\,\mathrm{cm} < r < 125\,\mathrm{cm}$). The electromagnetic calorimeter (EMCal) consists of 30 layers of $(5 \times 5)\,\mathrm{mm}^2$ Silicon cells alternating with $0.25\,\mathrm{cm}$ ($0.5X_0$) Tungsten layers ($127\,\mathrm{cm} < r < 142\,\mathrm{cm}$), representing a total of 20 radiation lengths (X_0) or 0.8 hadronic interaction lengths (λ_I). The hadron calorimeter consist of 34 layers of $(1\times1)\,\mathrm{cm}^2$ Silicon cells with 1 cm scintillating material and 20 cm of stainless steel per layer, representing $40X_0$ and $4\lambda_I$. The inner part of the detector is surrounded by a solenoidal coil ($250\,\mathrm{cm} < r < 330\,\mathrm{cm}$), creating a five Tesla magnetic field. The muon tracking system ($340\,\mathrm{cm} < r < 545\,\mathrm{cm}$) is designed to also serve as a tail catcher for a possible punch-through of hadron showers.

3. BENEFITS OF A PARTICLE FLOW ALGORITHM

The implementation of a PFA implies the use of tracking to reconstruct charged particles, electromagnetic calorimetry to reconstruct photons, and electromagnetic and hadron calorimetry to identify and measure the energy deposit of neutral hadrons. In an average event at typical linear collider center-of-mass energies of $(0.5 - 1)$ TeV, $\approx 60\%$ of the energy deposited in the calorimeters is expected to come from charged particles, photons contribute $\approx 25\%$ and neutral hadrons $\approx 15\%$ [5].

Of particular importance for a PFA are the electromagnetic and the hadronic calorimeters. They play a key role in the reconstruction of jets, e.g. from decays of vector bosons, decays of top quarks or decays of possible Higgs bosons. To obtain the maximum sensitivity, it will be essential to be able to distinguish the presence of a Z^0 or $W^{\pm}$ boson by its hadronic decay modes into two jets. This implies a precision of the di-jet mass measurement which is comparable to the natural width of these bosons, requiring a jet energy resolution of $\sigma_E/E \approx 30\%/\sqrt{E}$. Methods used during the LEP era involving kinematic fits achieved a resolution of $\sigma_E/E \approx 60\% \cdot (1 + \cos\Theta)/\sqrt{E}$ [4]. Due to the huge amount of Beamstrahlung present at the ILC, these kinematic fits are not expected to substantially improve the measurements.

4. PFA RECONSTRUCTION IN THE EMCal

4.1. MIP Reconstruction in the EMCal

4.1.1. Benefits from MIP Tracking

An effective and efficient MIP reconstruction algorithm, without the assistance from the central tracking system is beneficial to a successful PFA in many ways. Three of them are discussed below.

1. Electromagnetic and hadron showers can be separated by analyzing their shower shapes. Having a longitudinally segmented calorimeter, the actual starting point of the shower and the possible MIP-track associated with it can be used as an additional distinctive feature.

2. In particular in the forward direction, the track reconstruction efficiency is reduced by the decreasing number of possible tracker-hits due to geometry. The extrapolation of identified MIP-tracks into the central tracking system can be used to assign unassociated hits and to improve the tracking efficiency.

3. Charged tracks produced in the decays of long-lived particles like K^0_S, Λ or possible SUSY particles only leave a limited number of hits in the central tracking system. The identification of possible MIP tracks from their decay products in the calorimeter can help to improve the detection efficiency in the same way as described above.

4.1.2. The MIP Reconstruction Algorithm

The seed for a MIP-candidate is initially defined as at least three isolated hits in four consecutive EMCal layers, where isolated means that no second hit is present in neighboring cells in the same layer. The energy deposit of each of the hits has to be consistent with coming from a MIP.

The MIP seeds are then extended as long as more isolated hits are found in consecutive layers. In order to account for possible energy deposits from δ electrons, in single layers a second energy deposit is allowed in one of the neighboring cells. In order to account for possible detection inefficiencies, a gap of the size of one layer at a time in radial direction are allowed.

The MIP reconstruction is a two-pass algorithm. The procedure described above is applied once, starting from the innermost layer, following the MIP-track to the outside part of the calorimeter, and once starting from the outside.

4.1.3. Association of MIP-Segments

The MIP-segments reconstructed with this algorithm are then combined to MIP objects. A five parameter helix fit is performed to the EMCal hits of each MIP-segment found. For every pair of MIP-segments, a two-step χ^2 criterion is calculated using the fit parameters and the corresponding error matrices.

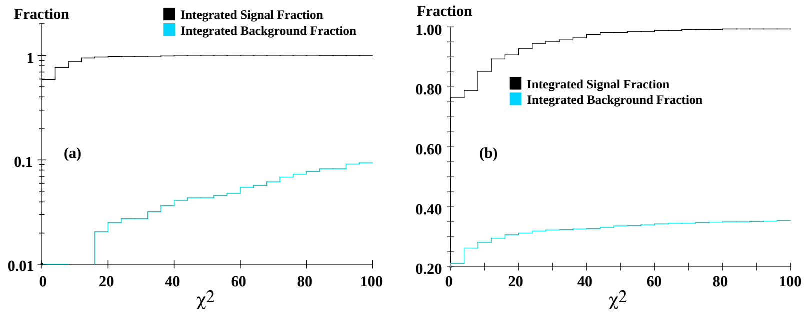

Figure 1: Integrated χ^2 distributions as calculated using the helix parameters (κ, ϕ_0, d_0, $\tan\lambda$) (a) and z (b). The upper curves indicate the efficiency for associating matching MIP-segments, while the lower curves indicate the background rate from random MIP segment combinations. The background fraction in (b) is relative to that remaining after requiring $\chi^2(\kappa, \phi_0, d_0, \tan\lambda) < 40$.

In the first step, only the track parameters (κ, ϕ_0, d_0, $\tan\lambda$) are used[1] and the efficiency (upper curve) for matching MIP-segments and the background rate from random combinations (lower curve) is displayed in Figure 1(a). By retaining all combinations with $\chi^2 < 40$, an efficiency of $\approx 100\%$ is achieved with a background fraction below 5%.

In the second step, a cut of $\chi^2_z < 80$ is used, where χ^2_z is calculated from the z component of the helix parameters only, on those MIP-combinations that pass the first selection step just described. This rejects 2/3 of the remaining background while maintaining high efficiency as shown in Figure 1(b). Pairs of MIP-segments satisfying these criteria are then combined to MIP objects.

4.1.4. The Use of Tracking System Layers in Calorimeter Tracking

The reconstruction of particles like K^0_S, Λ, or possible other long lived particles predicted from theories beyond the Standard Model is challenging in the current SiD design. In a typical Z^0 event, a K^0_S with an average energy of 5 GeV has a mean decay length of about 25 cm. Therefore, the charged pions from K^0_S decays would not leave any

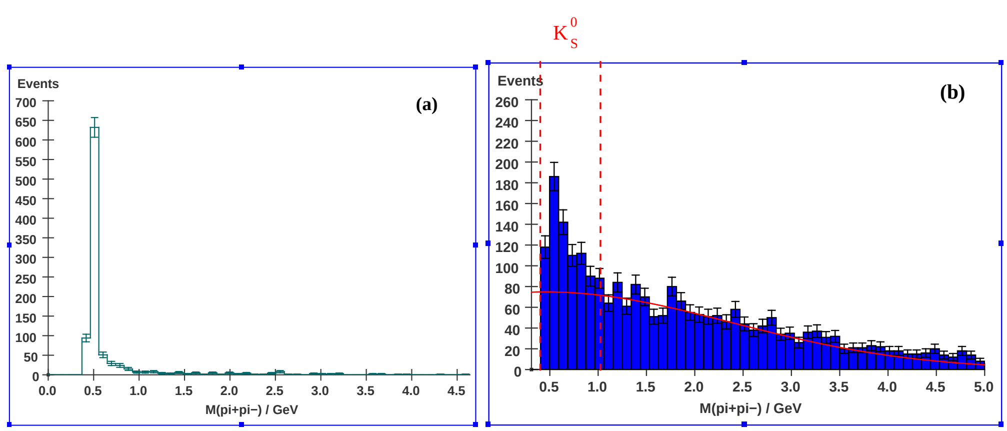

Figure 2: Invariant K^0_S mass distribution using MIP assisted tracking as described in the text. Plot (a) shows the invariant mass distribution as obtained from single K^0_S events, plot (b) shows the corresponding distribution as obtained from Z^0 events. In (b) the K^0_S is sitting on top of combinatorial background.

[1] The z component is not included in the first step, since its contribution exhibits artifacts due to the projective geometry of the EMCal configuration used in the simulation (for example, when a MIP passing through the EMCal in one tower of cells is scattered into a neighboring tower).

hits in the vertex detector, and would therefore not be reconstructed by the current tracking algorithm.

The tracking of MIP particles in the EMCal helps increasing the reconstruction efficiency of long-lived particles. The helix fit to the EMCal hits generated by MIP particles is extrapolated into the central tracking system. The radial position of the last tracking layer is $r = 125.06\,\mathrm{cm}$, only $\sim 2\,\mathrm{cm}$ away from the first layer in the EMCal. Using the extrapolated track, unassociated hits in the central tracking system (i.e. hits which have not yet been used for track reconstruction by the standard tracking algorithm) are picked up with high efficiency. The extrapolation algorithm has a resolution of better than 1 cm for the last tracking layer, whereas the average distance between unassociated hits in Z^0 events is of the order of 5 cm.

After picking up the hit from the last tracking layer, the track is refitted and the new track is then extrapolated to the second-to-last tracking layer. In case another unassociated hit is found, it is added to the MIP hits and the track is refitted again.

In order to examine the benefits from calorimeter tracking, pion tracks from neutral kaon decays were reconstructed as described above and fitted in pairs to a common vertex to form the K^0_S candidates. The results of the fits are illustrated in Figures 2 (a) and (b). Figure 2 (a) shows the K^0_S invariant mass for events from dedicated K^0_S event samples; a clear signal is obvious at the nominal K^0_S mass. The same algorithm is applied to Z^0 events and the result is shown in Figure 3(b). While the K^0_S peak is reconstructed at the nominal mass, it has a broader mass distribution on top of a large combinatorial background.

4.2. MST Clustering in the EMCal

An important component of a PFA is an efficient and effective clustering algorithm. An approach based on the Minimum Spanning Tree algorithm is described. Since the primary focus is on the reconstruction of hadronic showers in the EMCal and the HCal, identified energy deposits from the electromagnetic showers are removed prior to this study.

4.2.1. Description of the Algorithm

In the MST approach, energy deposits are assigned to a particular cluster on an individual hit-by-hit basis. To identify clusters from charged hadrons, the reconstructed shower points from the MIP-tracking algorithm and unassociated hits in the last tracking layer are used as cluster seeds. Starting from these seeds, the geometrical distance to all EMCal hits not yet associated to any cluster are calculated and the closest hit is added to the cluster, provided that the distance is smaller than the threshold. This procedure is repeated until all EMCal hits have been assigned to clusters.

An example of the performance of the algorithm for a typical K^0_S event is shown in Figures 3 (a) and (b). The section of the detector where the pions from a $K^0_S \rightarrow \pi^+\pi^-$ decay interact with the EMCal is shown in Figure 3(a).

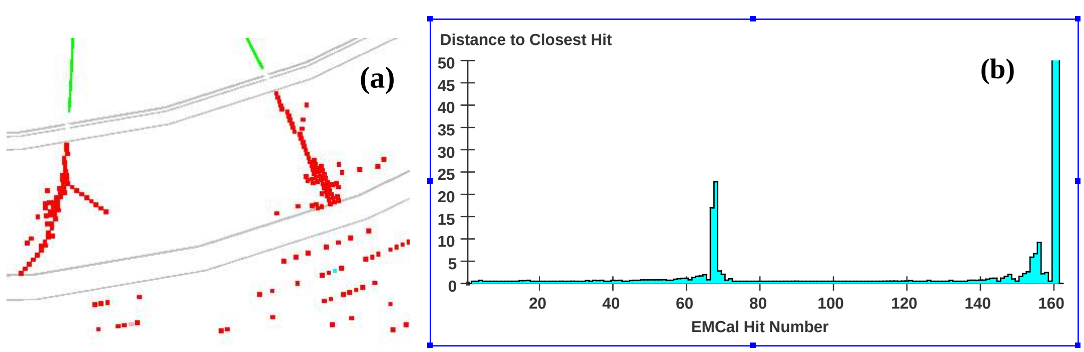

Figure 3: Plot(a) shows the signature of a typical $K^0_S \rightarrow \pi^+\pi^-$ decay from a single K^0_S event. Plot (b) displays the distance to the closest hit not yet assigned to the cluster as explained in the text.

Figure 3(b) shows the distance to the closest hit not yet assigned to the cluster. The spike around hit number 70, corresponding to a large geometrical distance, is used as separation criterion between the two pions showers.

The MST approach described here is based only on the coordinates of the energy deposits in the calorimeter. It is therefore independent of any particular calorimeter geometry.

4.2.2. Tuning of the MST Algorithm

The MST clustering algorithm has two free parameters. The threshold, which is the maximum distance allowed between a yet unassigned hit to the closest hit within a cluster, and the minimum number of hits required in a cluster (as opposed to a cluster fragment, which contains fewer hits).

The criteria these parameters are tuned on are:

- The number of reconstructed clusters should be close to the number of expected clusters in the event.
- The bulk of the total energy measured in the calorimeter is contained in the clusters (the rest is contained in the fragments).
- The clusters should be 'pure' objects, i.e. the fraction of the total energy within a given cluster comes from a single particle.
- The number of clusters an individual particle contributes energy to should be small.

Based on these criteria, the free parameters of the algorithm were tuned separately for the electromagnetic and the hadron calorimeter. Simulated Z^0 events as well as $t\bar{t}$ events at a center of mass energy of $\sqrt{s} = 500\,\mathrm{GeV}$ have been used. Figures 4 and 5 contain displays of the tuning criteria as a function of the free parameters. The optimal values for the tuning parameters in the current calorimeter design were found to be a 3 cm (10 cm) threshold for the EMCal (HCal) and minimum number of required hits in the clusters of five (eight).

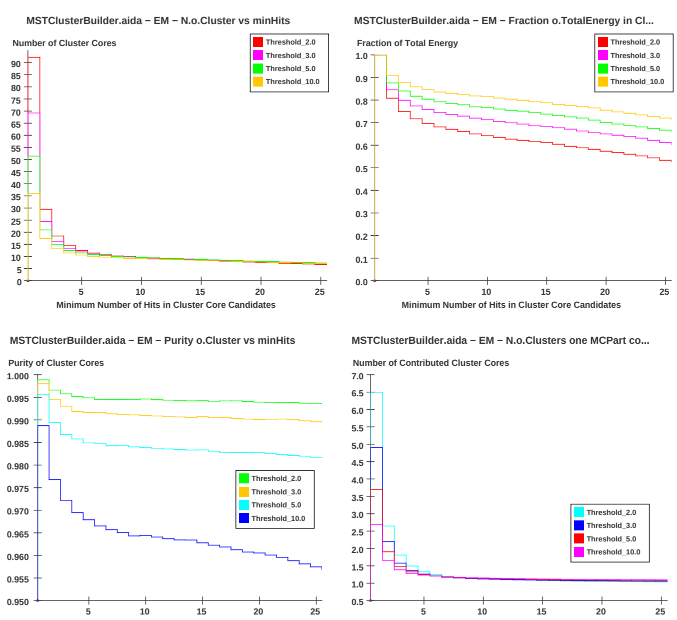

Figure 4: Performance of the MST clustering algorithm in the EMCal as a function of the tuning parameters. The distributions displayed are the number of reconstructed clusters and the fraction of the total energy contained in these clusters (upper row), and the purity of the clusters and the number of clusters a particular particle contributes to (lower row). The units for the threshold given in the plots is cm.

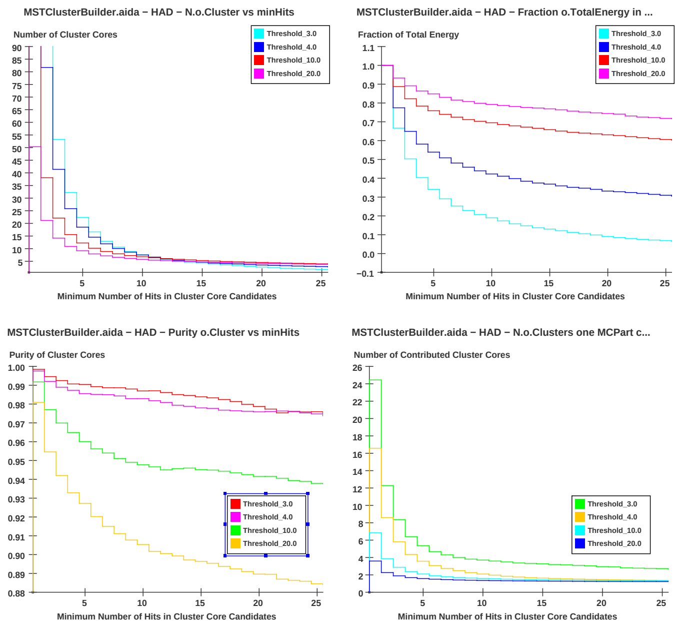

Figure 5: Performance of the MST clustering algorithm in the HCal as a function of the tuning parameters. The distributions displayed are the number of reconstructed clusters and the fraction of the total energy contained in these clusters (upper row), and the purity of the clusters and the number of clusters a particular particle contributes to (lower row). The units for the threshold given in the plots is cm.

5. SUMMARY

Two important components of a particle flow algorithm have been studied: a tracking algorithm for minimum ionizing particles in the electromagnetic calorimeter, and a clustering algorithm based on the minimum spanning tree approach. The implementation of these algorithms and their benefits to a particle flow algorithm have been discussed.

Acknowledgments

Work supported by Department of Energy contract DE-AC02-76SF00515.

References

[1] T. Csorgo, Phys. Lett. B **409**, 11 (1997)
[2] G. Weiglein *et al.* [LHC/LC Study Group], arXiv:hep-ph/0410364.
[3] www-sid.slac.stanford.edu/
[4] D. Buskulic *et al.* [ALEPH Collaboration], Nucl. Instrum. Meth. A **360**, 481 (1995).
[5] Presentation by S.Magill at the Victoria Linear Collider Workshop 2004: www.linearcollider.ca:8080/lc/vic04/abstracts/detector/calorimetry/steve_magill.ppt
[6] S. Agostinelli *et al.* [GEANT4 Collaboration], Nucl. Instrum. Meth. A **506**, 250 (2003).