CALORIMETER CLUSTERING WITH MINIMAL SPANNING TREES

GEORGIOS MAVROMANOLAKIS
Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, U.K.

We present a top-down approach to calorimeter clustering. An algorithm based on minimal spanning tree theory is described briefly.

1 Introduction

Clustering calorimeter hits is a complex pattern recognition problem with complexity depending on event type, energy and detector design. A successful clustering algorithm must be characterised by high efficiency and speed to cope with and to exploit the high granularity design forseen for both electromagnetic and hadronic calorimeters in a Future Linear Collider experiment. In the following we describe a top-down or divisive hierarchical clustering approach where the entire set of hits is first considered to be a single cluster, the minimal spanning tree, which is then broken down into smaller clusters.

2 Clustering With Minimal Spanning Trees

Given a set of nodes in a configuration space and a metric to assign distance cost or weight to each edge connecting a pair of nodes, we define the minimal spanning tree as the tree which contains all nodes with no circuits and of which the sum of weights of its edges is minimum (see Fig. 1). A minimal spanning tree is unique for the given set of nodes and the chosen metric, it is deterministic *i.e.* it has no dependency on random choices of nodes during construction, and it is invariant under similarity transformations that preserve the monotony of the metric [1]. First developed and applied to problems related to efficient design of networks [2], minimal spanning trees are well studied mathematical objects and there is a solid base of theorems which relate them to efficient clustering as well [1]. Applications to high energy physics can be found in [3].

A clustering algorithm based on minimal spanning trees has been developed. It can operate standalone or perform preclustering before a sophisticated energy-flow algorithm is applied [4]. Its operation is divided into three consecutive steps. First an appropriate metric, not necessarily euclidean, should be defined. Then the corresponding minimal spanning tree is constructed using Prim's algorithm [2]. The final step is to perform single linkage cluster analysis *i.e.* go through the tree and cut the branches with length above a proximity

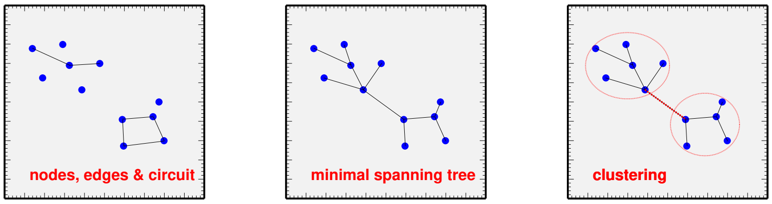

Figure 1: Illustration of terms and concepts discussed, nodes, edges and circuit, minimal spanning tree, single linkage cluster analysis.

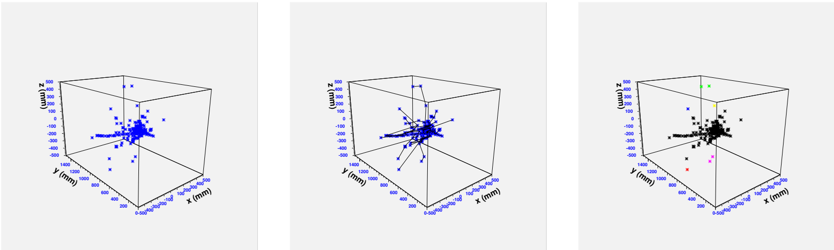

Figure 2: Example, clustering a single hadronic shower.

bound that the nodes belonging to the same cluster must obey. The algorithm is an $\mathcal{O}(N^2)$ loop, where N is the number of nodes. Also it should be emphasized that after defining an appropriate metric for the problem the rest of the algorithm has no dependency on detector geometry since only the metric deals with this. First tests of the algorithm with single and multiparticle events show satisfactory performance. A simple example is depicted in Fig 2.

3 Summary

We have discussed a top-down approach to calorimeter clustering based on minimal spanning trees, highlighting in brief their theoretical background and implementation in a clustering algorithm.

References

1. C.T. Zahn *IEEE Trans. Comput.* C **20** (1971) 68, J.C. Gower, G.J.S. Ross *Appl. Stat.* **18** (1969) 54, G.J.S. Ross *Appl. Stat.* **18** (1969) 103.
2. R.C. Prim *Bell System Techn. Journal* **36** (1957) 1389.
3. N. Saoulidou *Ph.D. Thesis* University of Athens 2003, K. DeWinter *et al.* (CHARM II Collaboration) *Nucl. Instrum. Methods* A**277** (1989) 170.
4. C. Ainsley *these proceedings.*