Clustering Algorithm Studies

Norman A. Graf

SLAC

LCWS2000

Calorimeter Clustering

- The basic concept of the "Energy Flow" algorithm for jet-finding is to use the tracking detector for the measurement of charged particle momenta and the calorimeter for neutrals.
- We therefore have to reconstruct and subtract neutral clusters before identifying the charged particle's energy deposition in the calorimeter.

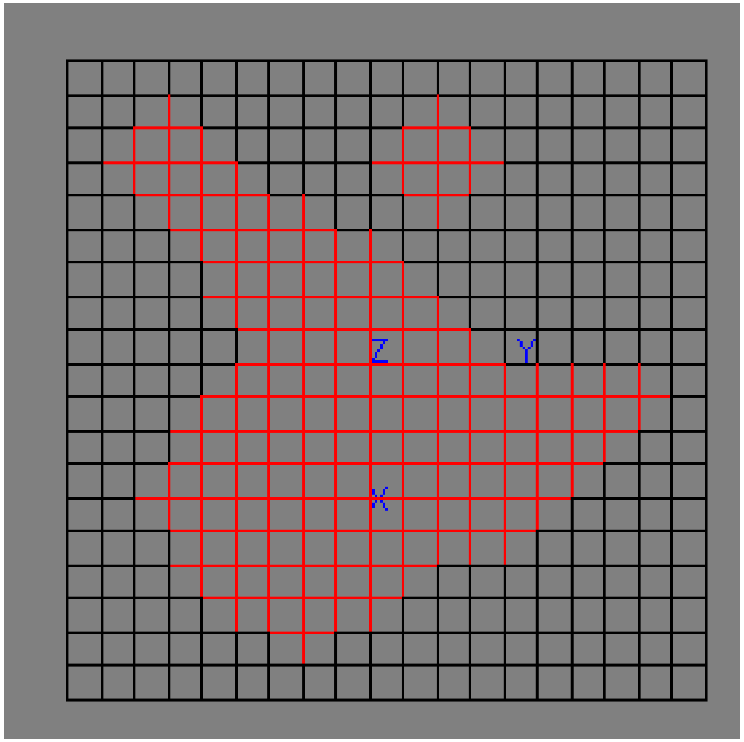
- **We also need pattern recognition algorithms to associate energy deposition in calorimeter cells with particles.**
- **EM showers are energetic, very localized and highly correlated.**
  - **Clustering works well.**
- **Muons deposit only minimum ionization, but do so along their trajectory**
  - **Tracking in calorimeter.**
  - **MIP deposition minimal in any case.**
- **Hadron showers are broad and unconnected.**
  - **More difficult to handle.**

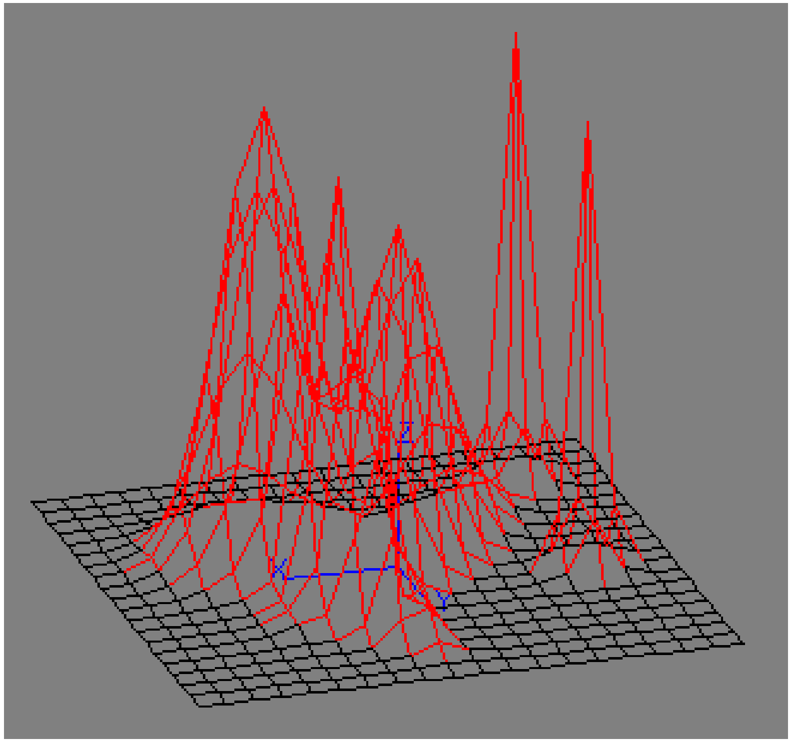
- **In complex events and within jets multiple particles will deposit energy in the same calorimeter cell, and showers will overlap**
  - **Good clustering is essential to resolve showers**
  - **A splitting/merging strategy is essential.**
- **Many cells are hit**
  - **An efficient algorithm is essential**

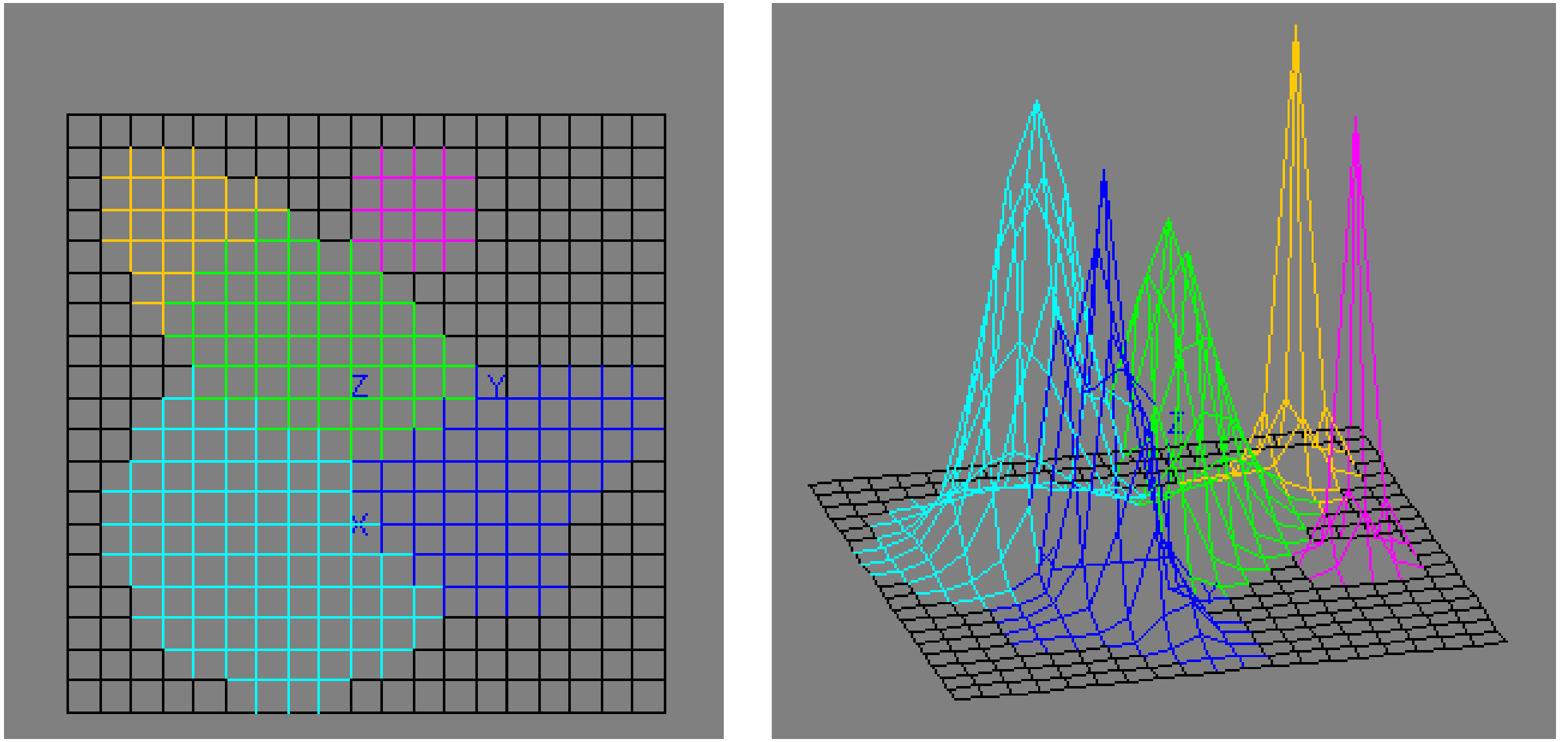
- **A fast, efficient, generic clustering algorithm has been developed to solve the problem.**
  - **Requires only one pass through the data to establish the clusters.**
  - **Based on a Nearest-Neighbor approach, but can be generalized to larger neighborhoods.**
  - **Works for arbitrary dimensions.**

- **Basic unit for clustering is a Cell.**
- **Cell contains an Index by which it is referenced.**
  - **E.g. Cell2D contains integer indices i,j.**
- **Cell also contains a value, for energy deposited.**
- **A Neighborhood encapsulates the topology of the detector. Given an Index, it is able to return a list of neighboring Indices.**
  - **Current implementation returns neighbors in user-defined region.**

- **One begins with each Cell pointing to itself as a proto-cluster.**
- **One then loops through all the neighboring Cells and establishes a pointer to the highest-valued neighbor.**
- **Linked lists of Cells comprise a Cluster.**


Z
Y
X


Z
X
Y



Z
Y
X

Cluster Splitting

- Nearby clusters, although identified as separate, will contaminate (leak) into each other.
  - Important to be able to remove background from nearby clusters.
    - Requires knowledge of shower shapes.
    - OK for EM showers and muons.
    - Not as obvious for hadronic showers.

Cluster Fitting

- **General Non-linear multidimensional fitter has been written to allow n-dimensional Gaussian, or exponential, to be fit to identified clusters.**
- **Fit each cluster separately to establish initial estimate, then perform global fit to all clusters.**
- **Can then subtract contributions from neighboring clusters.**

Cluster Merging

- **Clusters can also be incorrectly identified as separate, especially in hadronic showers.**
- **Will need to establish criteria for merging nearby clusters.**

Studies

- **Code has been incorporated into LCD framework.**
- **Single e, μ, π, γ generated to create catalogs of shower shapes.**
- **Preliminary results quite promising, γ γ resolved down to 20mr for 1GeV<E<10GeV.**
- **μ traces followed through calorimeters.**
- **Segmentation studies being undertaken to understand thresholds and neighborhoods.**

To Do

- Recognize and eliminate EM showers and μ traces.
- Extrapolate tracks into calorimeter to eliminate charged hadron showers.
  - May not need to explicitly reconstruct and recognize hadronic showers.
- Look at full MC events.