***************INCOMPLETE PRELIMINARY DRAFT*****************

Cluster ID: an approach to reconstructing calorimeter data

Introduction

We present an approach to reconstructing calorimeter cell hits data for a modern highly segmented sampling calorimeter system used for measuring high energy particles. This may be thought of as a solution to the “energy flow” problem however it begins with a different set of assumptions than the ones that originally motivated the energy flow concept. The underlying assumption of the energy flow problem is that the hits from highly collimated hadronic jets of particles are such that it is impossible to sort out the hits from different particles into separate clusters. This is partially due to the belief that the cluster of hits from one particle will usually overlay the cluster of hits from other particles due to the close proximity of the particles. The other reason for skepticism about associating hits with particular particles is that even a single isolated particle is seen to produce a set of disconnected cluster of hits. For example, a pion knocks a neutron out of a nucleus of detector material and the neutron travels a large distance before it deposits its energy with a visible cluster of hits elsewhere in the detector. We use the term “fragment” to refer to these secondary clusters of hits.

The belief that these two problems, overlap and fragmentation of clusters, limit our ability to reconstruct calorimeter data is well motivated but incorrect. There are two sources for these beliefs. The calorimeters used at SLC and LEP although state of the art and highly segmented in their day were, in general, unable to isolate hits from different particles into separate sets of clusters in jets. This lead to the development of the energy flow concept that roughly stated is as follows. Tracks from a tracker lying inside the calorimeter system are extended into the calorimeter and hits along the projected trajectory in some cone-like volume are assumed to be due to the charged particle that created the track and they are tagged as such. The remaining hits are then assumed to be due to neutral particles. The “charged” particle calorimeter hits are ignored and the energy of the remaining “neutral” hits is combined with the tracking information about the charged particles to give an “energy flow” level of description of the hadronic jet from which the direction, energy and mass of the underlying jet producing particle is deduced with some considerable resolution error.

The second source of belief in the limitation of calorimeters to sort out particles in jets comes from a cursory perusal of graphic displays of hits from jets as seen in modern event detector event displays. What is seen is a bundle of close but clearly separate tracks in the tracker that lead into a single large cloud of hits in the calorimeter.

This evidence seems compelling, what is wrong with it? Obviously with a very coarsely segmented calorimeter, say projective cells subtending 300 milliradian angles from the IP, nearly all the particles in a tightly collimated jet would pass through the same cells.

However, considering the other extreme of say microradian sized cells and remembering that a calorimeter shower is just a large number of particles leaving an ionization trail one can see that there will be almost no overlap of hit energy. So as we move between the extremes of a calorimeter with one readout and a calorimeter with a semi-infinite number of readouts we move from complete overlap of all particles to essentially no overlaps of any particles. The claim we want to demonstrate here is that in the LEP/SLC era the detectors were still insufficiently segmented to get good separation but the next generation of LC detectors can achieve sufficient segmentation to render the overlap problem tractable.

Regarding event displays of next generation detector jet showers one must adjust for two things. First, in any display one is seeing a two dimensional projection of a three dimensional shower which makes matters appear much worse than they are and it is difficult even with zoom and rotate features to convince oneself visually that overlapping is not so common as it appears. One needs to color code hits from different particles and one needs to set hit thresholds to eliminate a pervasive haze of KeV level X-ray hits that permeate the detector. Once this is done it can be made more apparent visually that particle clusters are much more isolated than first appeared. However, analytic techniques are much better suited to demonstrating this claim as we will show in plots to follow.

The second problem, fragmentation, is also not as serious as it seems at first when properly approached. A simple way to demonstrate this is to do simulations with a finely segmented monolithic calorimeter. By monolithic we mean to replace the standard EM and HAD calorimeter separated by a large physical gap which is filled with passive support structures and sometimes a solenoid and instead use a single very deep calorimeter. If one simulates single pions interacting in the standard two calorimeter system with a physical gap and one clusters together hits in adjacent cells one will frequently find many separated clusters containing similar amounts of energy. This occurs in the typical design where the EM section is about one interaction length deep so that about 70% of the pions interact in the EM cal creating an initial fragment and then the secondary particles pass through the gap and each secondary creates a separate cluster in the had cal. The same pion interacting at the same point in the monolithic design typically creates a single cluster of contiguous hits including the primary and secondary particles. The next most energetic fragment that occurs tends to be about 20% the size of the primary. The origin of these fragments is two fold. First, knocked out neutrons as discussed before which because of their large mass compared to pions tend to pick up a relatively small portion of the pions energy due to basic collision kinematics. The second source is gammas from secondary pizero production which may travel a few cells before interacting thus although lying close to the primary cluster are not contiguous with it.

We have tried to suggest that there still may be hope for the idea of reconstructing calorimeter data from the point of view that separate clusters can be associated with individual particles. We call this concept Cluster ID as opposed to the concept of energy flow. In what follows we will demonstrate that this is indeed a promising approach to reconstructing calorimeter data. It is useful to compare to how we approach reconstructing tracking data from trackers that contain many hits, for example a TPC.

There we also group hits together and associate those hit sets with single particles. We will try to show that the same approach is viable with calorimeter data, albeit much more complex for reasons that will become clear as we proceed.

Cluster ID – The concept

The general idea of Cluster ID is to begin with a cluster builder of some kind. The most natural place to begin is by forming clusters of hits that are contiguous. In a projective tower geometry one can imagine the following algorithm. Start with a hit cell as a seed for a cluster. Look at its 26 neighboring cells (to the sides, below and on all the diagonals). Include in the cluster any such neighbor that is hit. Iterate this process with each of the hits as it is added to the cluster. This is called a contiguous-hits cluster. Now choose a new hit cell that does not belong to any previous cluster and use it as a seed until all hit cells belong to exactly one cluster. It is not essential for the Cluster ID approach to use this particular form of cluster builder as we will see more clearly later but we use it to elucidate the concepts involved.

Once the first step is completed and a set of clusters has been defined by a cluster builder, the goal is to understand the origin of each cluster. For example, was it created by a single photon interaction, a pion interaction, a fragment from another interaction, or does it contain hits from two particles, for example two overlapping photons or a photon shower which has been intersected by the minimum ionizing track of a pion which interacts much deeper in the calorimeter? There are many possibilities for the origin of each cluster. The key assumption of the Cluster ID approach is that each type of cluster has a set of unique measurable characteristics that can be used to identify its origin with some level of certainty.

To illustrate this idea consider the difference between how a photon showers and how a pion showers. The photon's shower is controlled by the radiation length of the calorimeter material and is typically about 10% of the interaction length of the material. The interaction length controls the showering of a hadronic particle. Thus, a photon in contrast to a pion will begin its shower early in the detector and again because of the comparatively short radiation length will interact repeatedly in a short distance producing a small compact cluster of hits. The pion will interact on average much deeper in the calorimeter and its charged secondaries will also travel interaction length distances before they interact. This results in a shower that is much less dense than that of a photon. The same is also true of a neutral hadron such as a K0-long. Thus, if we have a technique to measure shower properties that capture these ideas we can separate photons from hadrons. To separate charged and neutral hadrons we have the difference that the pion hadronic shower is preceded by a minimum ionizing track that always begins with hits in the first layer of the calorimeter, thus giving a method to also distinguish the pion from the K0L.

One can envision similar ideas to distinguish other types of clusters. A single photon shower has a cigar shape and overlapping photons that do not lie on exactly the same axis will have the shape of two cigars overlapping. Various techniques can be developed to identify this case. Similar ideas apply to pions piercing photons. In the case of overlapping pion showers there will be two minimum ionizing tracks beginning at the first layer as part of the cluster. If a pion overlaps with a neutral hadron one can consider other kinds of approaches. One can "scan" across the whole cluster and look for two concentrations of energy (or hits) with a low energy (or hit) density region in between. One can also associate the pion with a track with well measured momentum and see if the energy of the cluster which contains both the pion and the neutral hadron is too large to have been due solely to the pion.

Of course, for each of these techniques one can imagine special cases where they won't work. For example the two minimum ionizing tracks of a two pion overlap could also be a single pion with a backscattered particle or even a neutral hadron with two backscattered particles. All these special cases and many more need to be considered. But the efficacy of the approach is only to be found by quantitative work. Surely there is no technique that works correctly every time. The question is, quantitatively how well does each technique work?

There are at least two approaches to using these identifying properties. One way is with a set of cuts for each cluster type hypothesis and the other way is letting a neural net assign probabilities for each hypothesis. In the cuts approach, for example, consider using the property of density of energy in a volume that contains the cluster. A plot of that quantity for photons of various energies will be some kind of bell-shaped distribution with a mean and width. The same distribution for neutral hadrons will have a much lower mean and a width such that a cut at some point between the two mean values will serve to distinguish the two types at some level of efficiency. If a second property is measured which also has discriminating power such as first hit layer, then for the photons the mean and width of that number will only overlap the low end of the distribution for the neutral hadron. Now the two cuts can be adjusted to optimize efficiency of identification. Additional properties can be added to the mix until the problem of making the cuts has become quite multi-dimensional and a neural net is a better candidate than using cuts. Although this illustration has been the separation of photons and neutral hadrons in the real application one is interested in separating many types at the same time.

To further elucidate the concept of Cluster ID we emphasize it is critical that we develop a set of properties that are measured for all the clusters in the calorimeter. Each type of cluster is identified by its differences with all the other types. In fact, one develops a set of properties iteratively by beginning with a set of properties and testing its correctness at identification, then looking at cases where one type is misidentified as another type and trying to add to or modify the set of properties. Contrast this with the energy flow approach where a large set of hits (those found by track extension) are simply ignored. In other words, in ClusterID there is no separate pion or photon identification algorithm. They are all identified in relation to each other simultaneously.

As with any calorimeter reconstruction technique, the goal is to use the technique investigate how well various physics measurements of interest can be made. Then having chosen a specific algorithm and a specific physics measurement one can vary the physical calorimeter parameters such as segment size, material, etc to optimize the design for the particular algorithm used to make that particular physics measurement. It is reasonable to expect that a different algorithm applied to a different physics measurement will also lead to the same design optimization but we have not progressed that far in our studies to do more than speculate about it. But that is the kind of question that points the direction for further study.

Cluster ID – a specific implementation

We have described the Cluster ID concept in general terms above and we now turn to describing the specific implementation we have developed and tested. For the cluster builder we use the contiguous hits method described above. We will plot specific results derived using a detector with the following description… We will also show results that illustrate the difference between gaps and nogaps. For the particles we simulate we will use 500 GeV ttbar events to determine both single particle distributions and illustrate efficacy in a real event…

For the discriminating properties we have selected a set of 15 discriminators, each one has been chosen to contribute to solving particular problems of discrimination. We now described the discriminators in some detail.

The general question of describing the shape and size and orientation of a cluster is facilitated by defining a (symmetric) energy tensor for each cluster and calculating its three eigenvalues and the three principal axes. The tensor is defined exactly as one would define an inertia tensor except that the mass of each element in the object is replaced here with the energy of each hit. The location of each hit is taken to be the center of the hit cell. We label the eigenvalues for a cluster E1, E2 and E3 such that $E1 < E2 < E3$. It turns out to be advantageous to normalize the eigenvalues with the energy of the cluster calling the resulting normalized energy eigenvalues NE1, NE2 and NE3. This way clusters created by the same particle type but with varying energies have similar numerical NEn values.

Plots in figure 1 show the distribution of NE values for single photons, pions, and K0Ls [charged hadrons (pions, kaons,…) and neutral hadrons (K0L, K0S, neutron, lambda…)] that interact in the calorimeter. NE1 is the smallest normalized eigenvalue and corresponds to the…TO BE CONTINUED