

1813  
1814  
CDF///CDFR/0000  
DOC/JET  
November 19, 1993

## A Guide to JETCLU: the CDF Jet Cluster Algorithm

Brenna Flaughner, Jim Mueller

The jet clustering algorithm JETCLU was chosen during the analysis of the 1987 jet data and is described in CDF note 605. Comparison of this algorithm to other algorithms are presented in that note and will not be repeated here. In this note we give a short summary of how the clustering algorithm works, applications to the 88-89 data, where to find the code and how to use it.

### 1 JETCLU

The CDF calorimeter is comprised of small cells (or towers) which form the basic unit of the clustering algorithm. A tower covers roughly 0.1 in pseudorapidity, For the azimuthal angle around the beam,  $\phi$ , the tower size is 15 degrees for  $|\eta| < 1.1$  and 5 degrees for  $1.1 < |\eta| < 4.2$ . Each tower has hadronic and electromagnetic sections. An energy vector for each tower is determined using the  $E_T$  weighted  $\eta - \phi$  centroid of the electromagnetic and hadronic components for the direction, and the sum of the energy deposited in the two components for the magnitude. The CDF clustering algorithm assigns the towers to clusters in three steps:

### 1. *Preclustering*

Calorimeter towers are merged when necessary into a common  $\phi$  segmentation of 15 degrees. A search is then made over all the merged towers for those above the seed tower threshold (e.g.  $E_{Tseed} \geq 1.0$  GeV). The seed towers are sorted in order of decreasing  $E_T$ , and placed in a list. A loop over the list is performed in which seed towers are grouped together to form *preclusters*. A seed tower is added to a precluster if it is within a distance  $R_0$  in  $\eta$  and  $\phi$  of the largest  $E_T$  tower in that precluster and it is adjacent to a seed tower already assigned to that precluster. If the seed tower can not be assigned to an existing precluster, it is used to start a new precluster. The list of the resulting preclusters is passed to the next step.

### 2. *Cone algorithm*

A loop over the preclusters from step 1 is performed. The centroid of each precluster is calculated by forming the  $E_T$  weighted  $\eta - \phi$  centroid of the towers assigned to it. For each cluster, a loop over towers with  $E_T$  above a low threshold (e.g.  $E_{Tmin} = 0.1$  GeV) is performed. Towers are added to a cluster if they are within  $R_0$  of the centroid. This parameter defines the so-called ‘CONE RADIUS’ of the algorithm. The radius used for most CDF analyses is 0.7. After a loop over towers, the centroid of the cluster is recalculated using all the towers now assigned to the cluster. If the list of towers for this cluster has changed, then the loop over towers is repeated using the new cluster centroid. For each cluster, this process is repeated until the list of towers is unchanged in two consecutive passes.

### 3. *Merge and/or Resolve Overlaps*

In the above steps, it is possible to have some towers which are assigned to more than one cluster. In this stage the overlapping towers are separated or the clusters are merged. First, a check is made on the amount of overlapping energy. If it is more than some fraction (typically 75% of the smallest cluster) then the clusters are merged. If the overlapping energy is less than the cut-off, then the common towers are assigned to the cluster with the nearest centroid

in  $\eta - \phi$  space. After the clusters have been separated or merged, the centroid of the each cluster is recalculated using the new list of towers.

## 2 Jet Energy and Momentum

From the list of towers assigned to each cluster, the jet energy and momentum components are defined as:

$$E_J = \sum_{i=1}^N E_i,$$

$$P_{x,J} = \sum_{i=1}^N E_i \sin\theta_i \cos\phi_i,$$

$$P_{y,J} = \sum_{i=1}^N E_i \sin\theta_i \sin\phi_i,$$

$$P_{z,J} = \sum_{i=1}^N E_i \cos\theta_i,$$

where  $i$  is the tower index and  $N$  is the number of towers in the cluster. The angles  $\theta_i$  and  $\phi_i$  are calculated with respect to event vertex along the beam axis. Note that the jets are not massless. From these quantities  $P_T$  and  $E_T$  are derived as:

$$P_{T,J} = \sqrt{P_{x,J}^2 + P_{y,J}^2},$$

$$P_J = \sqrt{P_{x,J}^2 + P_{y,J}^2 + P_{z,J}^2},$$

$$E_{T,J} = E_J \frac{P_{T,J}}{P_J}.$$

## 3 Applications

With the recent release of higher order QCD calculations, cone size dependence and “out-of-cone” corrections have become an issue. For comparisons to leading order QCD calculations it is reasonable to try to

correct the jet energies for the effect of bremsstrahlung and fragmentation which would cause particles to fall outside a cone radius. On the other hand, the next-to-leading order calculations are a first order approximation of this effect at the parton level. Since the next-to-leading order calculations can predict a dependence on cone size, three cone sizes,  $R_0 = 0.4, 0.7,$  and  $1.0$  have been used in analysis of the compressed jet data sample from the 88-89 run (see CDFNOTE 1380).

## 4 Using JETCLU

The dictionary file for JETCLU is QDJETC.UIC. The module name is JETCLU. To include JETCLU in your executable you need to DICT C\$DIC:QDJETC and LINK A,B,...,JETCLU in BUILD\_JOB. JETCLU requires that the TOWE and SETA banks exist, unless you request that JETCLU create the SETA bank itself. It will create the CLSL, CALL, CALS, and JETS banks by default. The CLSL bank created will by default have as algorithm name “ETHAT CLUSTERING”, algorithm version number 1, parameter bank name \$CLP, and parameter bank number 1. It is recommended that calorimetry cleanup code be run before JETCLU. To use JETCLU with default parameters, just put it in your job path after TOWE has been created and after any desired calorimetry cleanup.

There are only a few TALK\_TO verbs that an average user will need to know, these are listed at the top of the following list.

- ETMIN

This defines the minimum tower ‘ $E_T$ ’ necessary for a tower to be considered in cone clustering. The default value is 0.1 (GeV). This is a free parameter of the algorithm. Any physically reasonable value should yield physically reasonable results.

- ALPHA\_MAX

This defines the maximum eta-phi space distance from the cluster centroid to consider a tower for cone clustering. The default value

is 0.7, but it is known that the optimal value of this parameter depends strongly on the physics being analyzed. This is a free parameter of the algorithm. Any physically reasonable value should yield physically reasonable results.

- CALS\_MAKE

This verb will prompt the user for a decision on whether or not to make CALS banks for every final cone cluster. The default value is yes. DF uses CALS banks to display cluster parameters. If CALS banks are not being used in subsequent analysis, not making them will speed up the clustering.

- JETS\_MAKE

This verb will prompt the user for a decision on whether or not to make JETS banks for every final cone cluster. The default value is yes. If JETS banks are not being used in subsequent analysis, not making them will speed up the clustering.

- DETS\_MAKE

This verb will prompt the user for a decision on whether or not to make DETS and L2JT banks for each event. The default value is yes. If these banks are not being used in subsequent analysis, not making them will speed up the clustering.

- SETA\_MAKE

This verb will prompt the user for a decision on whether or not to make the SETA bank before performing jet clustering. The default value is no. Since CALORIMETRY makes the SETA bank, making it in JETCLU is not necessary.

- SETA\_PARAMETERS

This verb will prompt the user for the number of radiation lengths (default=10.0) and number of nuclear interaction lengths (default=3.0). These are passed to QDSETA if the user has requested that the SETA bank be made in JETCLU (see SETA\_MAKE above).

- ANAME

This verb will prompt the user for a new algorithm name. The algorithm name value in the CLSL bank will then be the input name.

- ETYPE

This verb allows one to change the type of ‘ENERGY’ the algorithm uses in clustering. The requested input is an integer which is interpreted in the following way;

1. Total (EM + HAD) Et (default value)
2. Total (EM + HAD) Energy
3. EM Et
4. EM Energy
5. HAD Et
6. HAD Energy

For jet clustering, Total  $E_T$  is probably the right variable. The other options are provided for specialty purposes (testbeam calibration, etc.). When the ETYPE has been changed, the cluster and tower “ $E_T$ ” cutoff parameters (verbs ESEED, EMIN, ETCUT and ETMIN) will be interpreted as values of the ‘ENERGY’ being clustered. The algorithm is equally efficient clustering in any energy type.

- CLUSTER

This verb invokes the clustering on the current event. If clusters (i.e. CALL, JETS, etc banks) already exist in IW with the same CLSL identifying parameters (algorithm name and number, parameter bank name and number) they will be overwritten.

- SEEDS

This decides whether to cluster only to seed level. If the input value is YES, only seed clusters will be formed. If not (default), clustering

will continue to cone clustering. Seed clustering is useful mainly as a diagnostic. No CALS or JETS banks will be produced for seed clusters, only CALL and CLSL.

- ADJACENCY

This sets the adjacency level for seed clustering. Two towers are defined as ADJACENT if the sum of the square of their distance (calculated in terms of number of towers) in  $\eta$  and  $\phi$  is less than or equal to the input number. Thus a value of 1 implies towers touching on an edge but not on corners are adjacent, a value of 2 that touching on edges or corners are adjacent, etc. The default adjacency is 2. Seed clusters are defined as groups of adjacent towers above the minimum tower ‘energy’ (verb EMIN) with at least one tower of ‘energy’ above the seed tower energy (verb ESEED), and all towers having monotonically decreasing energy moving out from the seed tower.

- EMIN

This defines the minimum tower ‘energy’ necessary for a tower to be considered in seed clustering. The default value is 1.0 (GeV).

- ESEED

This defines the minimum tower ‘energy’ necessary for a tower to be considered as a seed tower in seed clustering. The default value is 1.0 (GeV)

- ETCUT

This defines the minimum seed cluster ‘ $E_T$ ’ necessary for a seed cluster to be considered for growing into a cone cluster. The default value is 1.0 (GeV).

The following verbs were declared by Dave Brown to be ‘experts only’ parameters. They should only be changed by trained QCD professionals. Do not attempt this at home.

- MAX\_ITERATE

This defines the maximum number of iterations allowed for the cone clustering centroid definition, and the overlap cluster separation. If the number of iterations reaches this value, the iteration process stops and an error message is sent to LUTTYO. The default value is 100, although the typical number for convergence is around 4.

- RATCHET\_TOWERS

This sets a parameter which defines the level of ‘hysteresis’ in the centroid definition iteration. A value of 0 means no hysteresis (IE each iteration the centroid is computed with no reference to the previous iteration). The default value is 1, which means the towers in the original seed cluster are always retained in the JETCLU cluster.

- OVERLAP

This sets a parameter which defines the minimum fraction of Et overlap between two un-separated clusters at which to merge the two. If the fraction of Et overlap (defined as the Sum Et of towers held in common by two clusters divided by the minimum  $E_T$  of the two clusters) is greater than this fraction, the two are merged. If less, the towers in common are divided up between the two. The default value is .75.