Cone Algorithm For Jet Finding

N.J. Hadley
DØ Note 904

November 29, 1989

1 Introduction

This is a description of the DØ cone algorithm for jet finding. In the algorithm a jet is defined by summing up the Et in a cone in η, ϕ space of radius R = $\sqrt{\Delta\eta^2 + \Delta\phi^2}$ The algorithm is similar to that used by UA1 and CDF.

2 DØ Implementation

The algorithm starts from the ordered list of towers in the CATE bank. The towers are typically 0.1 in η and $2\pi/64$ in ϕ except in far forward and backward hemispheres where they are 0.2 in η and $2\pi/32$ in ϕ. The transition occurs at about $\eta = \pm 3$. Starting from the highest Et hadronic tower, preclusters are formed of contiguous cells out to a radius of about R = 0.3 . The preclusters are used to cut down on the number of towers considered as a possible starting point for jet formation. This is done in the routine CLUPRE. In a typical Monte Carlo two jet event there are more than 1000 hadronic towers in the CATE bank. These Preclusters are then used as the starting point for jets. The precluster center in η, ϕ space is used as the starting cone center. A new Et weighted center is then formed from all towers within a radius R of the center, and the process is repeated until the jet is stable. On Monte Carlo data, typically three or four times through the center finding loop are necessary. This is done in the subroutine CONCLU. If any jets share energy, these jets are then combined or split based on the fraction of energy shared relative to the Et of the lower ET jet. The routine that does this combining is SPLJET which is called from CONCLU. CONCLU and SPLJET fill the JETS, JTSH, and JPTS banks. Both CLUPRE and CONCLU are called from CLUFND which is the driving routine for the cone algorithm. (Most packages use additional driving routines which call CLUFND.)

All three routines, CLURPE, CONCLU, and SPLJET need to get parameters from the CAJETS_RCP bank at initialization. The initialization entry points are CLUINI, CONCLI, and SPLINI.

The parameters for the cone algorithm from the latest version of CAJETS.RCP are given below along with their default values.

```
!------------------------------------------------------------------------------
!              ETA,PHI CONE ALGORITHM:
!------------------------------------------------------------------------------
!   Constants for subroutine CONCLI
!------------------------------------------------------------------------------
RADIUS_OF_CONE          0.7       ! cone for finding jets
MIN_JET_ET              8.0       ! minimum ET for a jet
!------------------------------------------------------------------------------
!   Constants for subroutine SPLINI
!------------------------------------------------------------------------------
ET_SPLIT_FRAC           0.5       ! fraction of energy shared before COMBINE
MIN_CLUSTER_SEPARATION  0.01      ! minimum separation between clusters
!------------------------------------------------------------------------------
!   Constants for subroutine CLUPRI
!------------------------------------------------------------------------------
SEED_MIN_ET             1.0       ! SEED MINIMUM ET
CAND_MIN_ET             1.0       ! candidate minimum ET
PRECLU_MIN_ET           1.0       ! minimum ET for a PRECLUSTER
!------------------------------------------------------------------------------
```

The MIN_CLUSTER_SEPARATION parameter controls when to clusters are to be automatically combined because they are close in eta, phi space. The SEED_MIN_ET parameter is the minimum Et a tower is allowed to have if it is used initiate a precluster. The CAND_MIN_ET parameter is the minimum ET a tower can have and still be included in a Precluster.

3 Comparision with Monte Carlo Data

The algorithm has been tested on GEANT and FAKE events. We compared the generated jets with the jets found by the CONE algorithm by studying 143 "twojet" ISAJET/GEANT events. The GEAN jet quantities are those obtained from the Monte Carlo and the RECO jet quantitities those from reconstructing the events finding jets with the cone algorithm. The results were:

$$\langle \eta_{Gean} - \eta_{Reco} \rangle = 0.000 \pm 0.002. \text{ No offset within statistics.}$$

$$\sigma(\eta_{Gean} - \eta_{Reco}) = 0.026 \pm 0.002.$$

$$\langle \phi_{Gean} - \phi_{Reco} \rangle = -0.0016 \pm 0.0018. \text{ No offset within statistics.}$$

$$\sigma(\phi_{Gean} - \phi_{Reco}) = 0.025 \pm 0.002$$

$$\langle \text{Et Gean - Et Reco} \rangle = -0.26 \pm 0.22\text{GeV}.$$

$$\sigma(\text{Et Gean - Et Reco}) = 3.0 \pm 0.2\text{GeV}.$$

$$\text{Average jet Et} = 79 \text{ GeV}.$$

The distribution of jet Et's is nearly gaussian with a sigma of 16 GeV. Note that Resolution = $3/79 = 0.038\% \pm 0.003\%$ or Resolution = $34\%/\sqrt{Et}$. This "resolution" is not our resolution for jet Et, since the GEAN jets use the smeared calorimeter energies for those tracks that deposit energy in the calorimeter. Thus, they are *not* the parton energies.

The same studies done with the jets from the files RUN01_50.GEN1 through RUN01_63.GEN1 yield the results:

$$\langle \eta_{Gean} - \eta_{Reco} \rangle = 0.002 \pm 0.005.$$

$$\sigma(\eta_{Gean} - \eta_{Reco}) = 0.044 \pm 0.005.$$

$$\langle \phi_{Gean} - \phi_{Reco} \rangle = -0.006 \pm 0.003$$

$$\sigma(\phi_{Gean} - \phi_{Reco}) = 0.031 \pm 0.003$$

$$\langle \text{Et Gean - Et Reco} \rangle = -0.23 \pm 0.22\text{GeV}.$$

$$\sigma(\text{Et Gean - Et Reco}) = 2.3 \pm 0.2\text{GeV}.$$

$$\text{Average jet Et} = 41 \text{ GeV}.$$

The distribution of jet Et's is nearly gaussian with a sigma of 12 GeV. See Fig. 1. The resolution is worse for these events, most probably, because the average jet Et is lower.

In making the comparison between the GEAN jets and the RECO jets one must be careful. First, GEAN jet number 0 is the "beam" jet. The beam jet is all the other particles in the event that are not in the high Pt jets. Since these do not form a"JET" as one usually defined, this jet should be ignored in making the comparisons. Second, the number of GEAN jets (excluding the "beam" JET), is always 2 even if a high Pt gluon is emitted. An example of this is the first event in the file RUN01_50.GEN1 which was studied at Indiana. When displayed, this is clearly an event with three high PT jets. Nevertheless, the GEAN banks have only two high PT jets. This is because of the way the GEAN jets are defined using the ISAJ bank. We can still compare the RECO to the GEAN jets by examining the plot of $\phi_{Gean} - \phi_{Reco}$ or $\eta_{Gean} - \eta_{Reco}$ and noting that there is a large number of jets (about 70%) where this quantity is between -0.1 and +0.1. These jets are presumably those where no hard high Pt gluon was emitted. It is these distributions that were fitted with Gaussians to give the above results. See Fig. 2 and Fig. 3.

This problem can be avoided by using the PJET banks. The PJET bank will contain information about parton jets. This bank is needed to overcome the limitation that the ISAJ bank contains only two partons for"TWOJET" events even if the event contains many hard gluons or quark antiquark pairs coming either from the hard process or the initial radiation. These events with the extra partons what are usually called three or four or whatever jet events. The Parton jets are formed by ordering the partons in the ISAQ bank in Et and then forming jets starting with the highest Et parton using a cone algorithm. The second jet is then formed from the parton with the highest Et not included in the first jet. The process continues until there are no more partons. It is planned to have the PJET bank available soon.

Appendix A gives a list of the ISAJ jets, the PJET jets and the cone jets for the first 5 events in the file RUN01_50.GEN1. The cone jets were found using three different sets of jet parameters in the file CAJETS.RCP. Lowering the Et cuts used in CLUPRE increases the efficiency for finding low Et (less than 20 GeV) jets. At the same time, however, it increases sharply the number of jets found, particularly for Et less than 12 GeV. Many of these jets are presumably due to the minimum bias processes rather than hard scattering. See Fig. 4. Clearly more work is needed to optimize these parameters.

4 Conclusion

We conclude with a few general comments.

- The GEANT data has many towers with almost no energy. Here almost no means less than 20 MeV. In D0 we will almost certainly have to cut out towers with less than 50 MeV EM energy and less than 100 MeV hadronic energy to suppress noise hits. The program could be speeded up by putting such a cut into the filling of CATE or better yet CAEP.
- The JTSH banks contain much useful information.
- The number of JETS is stored in the CAPH bank at location Q(LCAPH +3).
- Other jet algorithms, such as nearest neighbor algorithm, are in progress.

Appendix A

hese are the first five events in the file RUN01_50.GEN1.
e compare jets found with the cone algorithm with jets from the artons. The "high thresholds" were SEED_MIN_ET = 2.0 GeV, AND_MIN_ET = 1.0 GeV, and PRECLU_MIN_ET = 4.0 GeV.
he "low thresholds" have the same values for the constants as given in page of this note: SEED_MIN_ET = 1.0 GeV, CAND_MIN_ET = 1.0 GeV and RECLU_MIN_ET = 1.0 GeV.

:vent 1

:ton Jets

Jet Number	Eta	Phi	Et (GeV)
1	0.29	5.13	33.0
2	0.07	3.12	43.2
3	-0.36	0.79	43.0
Cone R = 0.7 high thresholds			
1	0.04	3.14	34.2
2	0.28	5.16	24.5
3	-0.32	0.75	38.1
Cone R = 0.7 low thresholds			
1	0.04	3.14	34.2
2	0.28	5.16	24.5
3	-0.32	0.75	38.1
Cone R = 0.6 low thresholds			
1	0.03	3.14	33.6
2	0.28	5.16	24.3
3	-0.31	0.74	36.3

. ent 2

Parton Jets

Jet Number	Eta	Phi	Et (GeV)
1	0.65	1.13	43.9
2	0.83	4.06	38.0
3	1.69	5.06	9.7
Cone R = 0.7 high thresholds			
1	0.60	1.16	36.3
2	0.74	4.00	22.8
Cone R = 0.7 low thresholds			
1	0.60	1.16	36.3
2	0.74	4.00	22.8
3	1.65	5.15	9.2
Cone R = 0.6 low thresholds			
1	0.60	1.16	36.1
2	0.74	4.00	22.3
3	1.63	5.14	8.8

vent 3

arton Jets

et Number	Eta	Phi	Et (GeV)
1	-0.14	5.97	66.9
2	1.69	2.66	41.9
3	-0.66	3.77	16.4
4	-1.14	2.98	15.6

:one R = 0.7 high thresholds

1	1.68	2.70	36.4
2	-0.17	5.96	53.6

:one R = 0.7 low thresholds

1	1.68	2.70	36.4
2	-0.17	5.96	53.6
3	-0.42	4.52	12.1
4	-1.09	2.97	11.3

:one R = 0.6 low thresholds

1	1.67	2.70	35.9
2	-0.18	5.96	52.5
3	-0.43	4.66	9.0
4	-1.09	2.96	10.7

:vent 4

?arton Jets

Jet Number	Eta	Phi	Et (GeV)
1	-1.42	1.06	53.0
2	0.11	4.65	32.2
3	-0.56	3.78	24.7

Cone R = 0.7 high thresholds

1	-1.44	1.06	42.1
2	0.10	4.59	34.8

Cone R = 0.7 low thresholds

1	-1.44	1.06	42.1
2	0.10	4.59	34.8
3	0.75	3.44	10.2

Cone R = 0.6 low thresholds

1	-1.44	1.06	41.7
2	0.13	4.67	28.7
3	0.74	3.39	9.4
4	-0.44	3.97	20.0

Event 5

Parton Jets

Jet Number	Eta	Phi	Et (GeV)
1	1.52	6.08	124.2
2	0.20	2.82	122.5
3	-0.36	4.46	12.9

Cone R = 0.7 high thresholds

1	0.20	2.82	101.3
2	1.47	6.10	94.2
3	-0.37	4.50	16.8

Cone R = 0.7 low thresholds

1	0.20	2.82	101.3
2	1.47	6.10	94.2
3	-0.35	4.48	17.6

Cone R = 0.6 low thresholds

1	0.21	2.82	97.9
2	1.42	6.14	85.5
3	-0.38	4.50	15.8

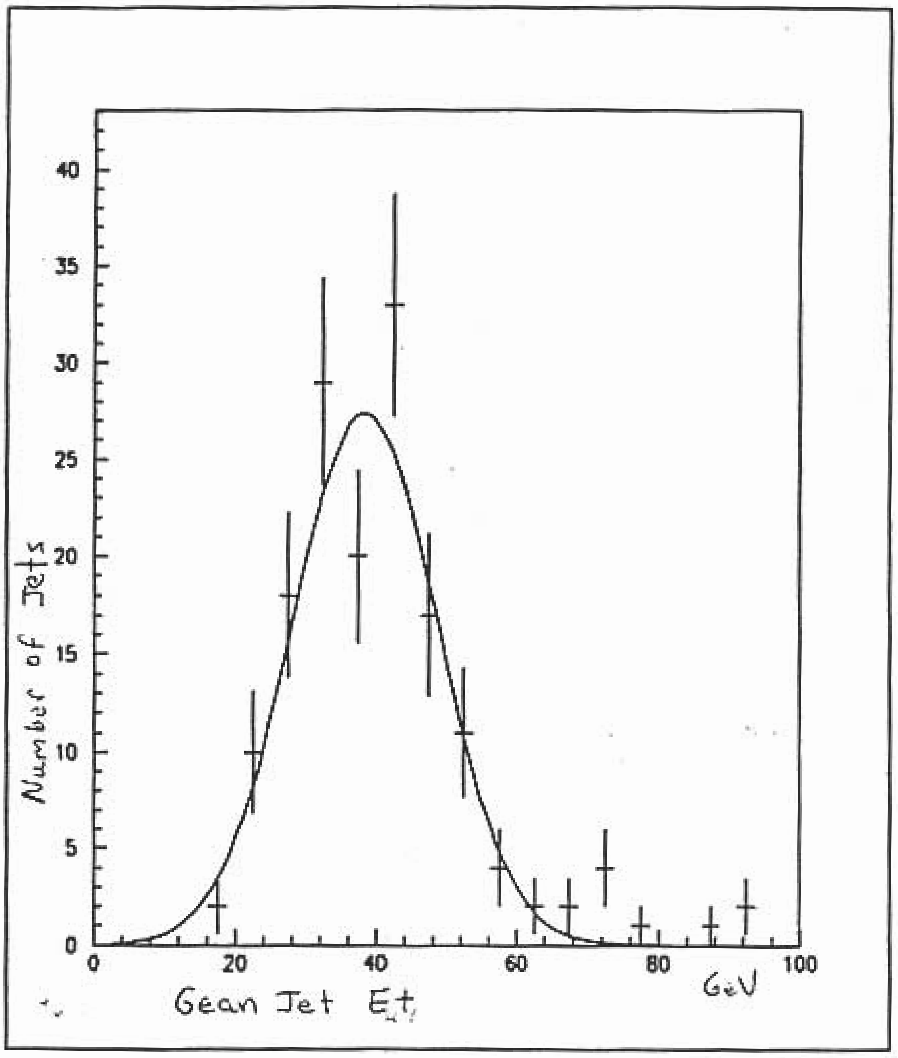

Figure 1 - Et Distribution of Jets from GEAN banks Run01-50.GEN1 through Run01-63.GEN1

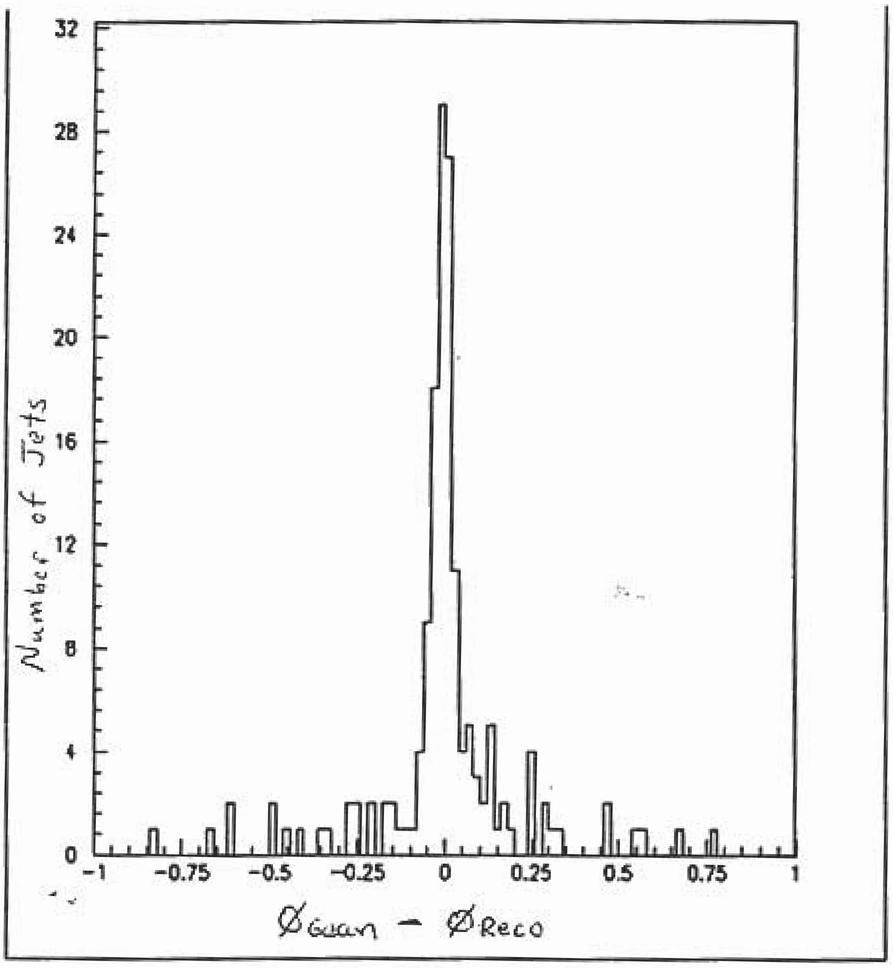

Figure 2- ØGean - ØReco

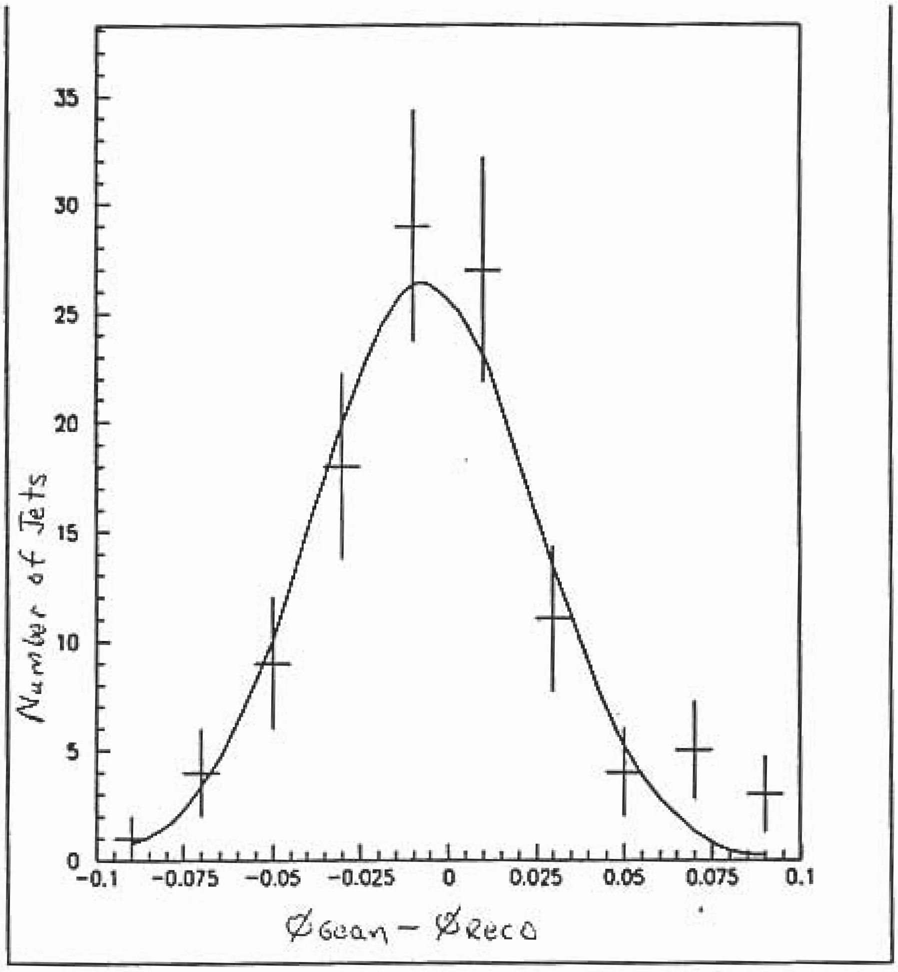

Figure 3 - ØGean - ØReco, same data as figure 2 with expanded scale. Gaussian Fit superposed

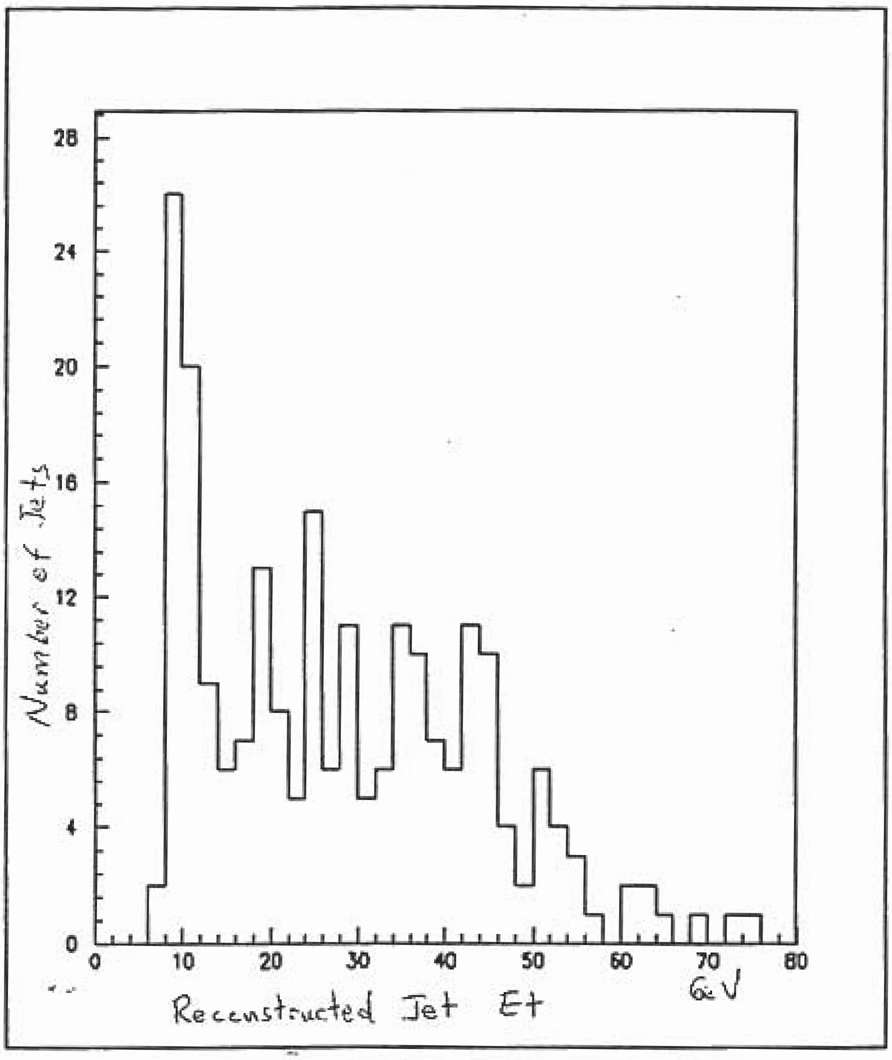

Figure 4 - Distribution of Reconstructed Jet Et
Same Data as Figure 1
Low Thresholds
The low thresholds are now the default,
See p. 2 of this note.