Fermi National Accelerator Laboratory

FERMILAB-Conf-90/248-E
[E-741/CDF]

A Compilation of Jet Finding Algorithms *

Brenna Flaugher
Fermi National Accelerator Laboratory
P.O. Box 500
Batavia, Illinois 60510

Karlheinz Meier
Deutsches Elektronen Synchrotron (DESY)
Hamburg 52, Germany

December 1990

* To be published in the proceedings of the 1990 Summer Study on High Energy Physics, *Research Directions for the Decade*, Snowmass, Colorado, June 25 - July 13, 1990.

Operated by Universities Research Association Inc. under contract with the United States Department of Energy

A COMPILATION OF JET FINDING ALGORITHMS*

Brenna Flaugher

Fermi National Accelerator Laboratory, Batavia, IL 60510, USA

Karlheinz Meier

DESY, Notkestraße 85, D-2000 Hamburg 52, Germany

Abstract

Technical descriptions of jet finding algorithms currently in use in $p\bar{p}$ collider experiments (CDF, UA1, UA2), e^+e^- experiments and Monte-Carlo event generators (LUND programs, ISAJET) have been collected.

1 INTRODUCTION

This paper presents an overview of currently existing jet algorithms. Initially such algorithms were designed by experiments in order to interpret their data in terms of perturbative QCD calculations. With the advent of more sophisticated calculations like complex Monte-Carlo event generators and exact next-to-leading order QCD calculations[1,2,3], a constructive interplay between theory and experiment has started. The aim of this paper is to provide, in one place, descriptions of the different approaches and hopefully facilitate direct comparisons between experiments and between experimental data and theoretical calculations. An attempt for a standardized jet definition has recently been made for the case of hadron collisions [4].

---

*to appear in the Proceedings of the '1990 Summer Study on High Energy Physics - Research Directions for the Decade' Snowmass, Colorado, June 25 - July 13, 1990.

2 CDF JET ALGORITHM

A. Technical Description

The CDF calorimeter[5] is comprised of small cells (or towers) which form the basic unit of the clustering algorithm. A tower covers roughly 0.1 in pseudorapidity, η=-ln $tan(\theta/2)$, where θ is the angle with respect to the beam direction. For the azimuthal angle around the beam, ϕ, 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[6] assigns the towers to clusters in three steps:

1. *Preclustering*

   A search is made over all the calorimeter towers for towers above the seed tower threshold (e.g. $E_{T seed} \geq 1.0$ 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 within a radius of $R_0 = \sqrt{(\delta\eta^2 + \delta\phi^2)}$ of each other are grouped together to form ***preclusters***. The radius which is used for most CDF analyses is R_0=0.7. 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. 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.

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 predict a dependence of the cross section on cone size, CDF has used three cone sizes, $R_0 = 0.4$, 0.7, and 1.0, in their inclusive jet analysis[7]. For these measurements, no out-of-cone correction is applied.

At present, the dijet mass spectrum and dijet angular distribution have been calculated only to leading order, although there is some hope that they may be completed to higher order in the near future. As a result, the current analyses use the algorithm described above with a cone size of R_0=0.7 and the effects of making an out-of-cone correction are being investigated.

3 UA1 JET FINDING ALGORITHM

The UA1 jet finding algorithm[8] is also based on a cone type algorithm, but uses a different method for locating the clusters. The UA1 jet finding algorithm proceeds through the following steps :

1. Each calorimeter cell is assigned an energy vector whose magnitude is given by the energy deposited in the cell and the whose direction is given by the centroid of the electromagnetic and hadronic calorimeter energies.

2. All cells with $E_T > E_{T,seed}$ are ordered in decreasing transverse energy E_T.

3. The cell with the highest transverse energy initiates the first jet.

4. The next cell is added vectorially to the first if it is within a distance R_0 ($R_0 = 1$ in $\eta - \phi$ space). If the cell is outside this radius then a new jet is initiated. This procedure is repeated until all cells above the $E_{T,seed}$ threshold have been assigned to a jet.

5. Cells with $E_t < E_{seed}$ are then added to each jet if $R < R_0$.

A jet momentum vector is defined as the vectorial sum of all individual cells contained in a cone in $\eta - \phi$ space, centered on the highest transverse energy cell within the cone.

The original clustering algorithm used an $E_{T,seed}$ of 2.5 GeV. Later, it was found that low E_T jets were sensitive to this choice of the initiator threshold and a lower threshold, E_{Tseed}=1.5 GeV, was adopted. The effect of the change in initiator threshold on jets with $E_T > 35$ GeV was not significant[9].

4 UA2 JET FINDING ALGORITHMS

A. UA2 Cluster Algorithm

The UA2 cluster algorithm searches for structures of adjacent calorimeter cells having a (transverse) energy deposition (E_T)E in excess of a given threshold value. The algorithm consists of the following steps:

1. Order calorimeter cells in decreasing E (E_T).
2. Search for the calorimeter cell with highest E (E_T) not yet used in a cluster.
3. Join neighboring cells into the cluster if the deposited E (E_T) exceeds a given threshold in E (E_T) (e.g. 400 MeV).
4. Repeat 3 until no more neighbors are found.
5. Go to 2 if there are still cells unassigned.
6. Construct the jet momentum vectors from the scalar sum of transverse energy of all cells assigned to the cluster and the $\eta - \phi$ coordinate of the E(E_T) weighted center of the cluster. The jets are massless.

More details can be found in [10]. The cluster algorithm provides good spatial resolution for nearby jets. Roughly 50% of all jet pairs are resolved at an angular distance of 30 degrees for a specially tuned algorithm [11]. This feature is important for distinguishing gluon bremsstrahlung processes from alternative models which give large angular separation between jets (phase space,double parton scattering).

The clusters do not have a limited size in $\eta - \phi$ space. This can result in uncontrolled growth of clusters (e.g. close to the beam pipe). For the same reason the algorithm is sensitive to background (halo,cosmics) and multiple interactions. Another disadvantage is the sensitivity to potential calibration problems (pedestals) due to the low threshold for the definition of a neighbor.

The energy collection of this cluster algorithm is, on average, limited to the central core of the jet. For a jet with a transverse energy E_T= 40 GeV, typically 10% of the energy is not seen in the clustering procedure.

This algorithm is applied for multijet studies [12].

B. UA2 Cluster Merging Algorithm

To improve the energy collection, the basic cluster algorithm can be modified by merging clusters around a primary seed cluster within a well defined cone according to the following prescription:

1. Find clusters defined according to the UA2 cluster algorithm.
2. Select clusters with a transverse energy E_T> 3 GeV.
3. Order clusters in decreasing E_T.
4. Search for the cluster with highest E_T not yet used in a jet.
5. Collect all secondary clusters in a given radius $\sqrt{(\delta\eta^2 + \delta\phi^2)} < 1.3$ around the primary cluster axis to form a jet.
6. Go to 4 if there are still clusters unassigned.
7. Construct the jet momentum vectors from the scalar sum of transverse energy of all clusters assigned to the jet and the $\eta - \phi$ coordinate of the E_T weighted center of the cluster. The jets are massless.

Apart from reducing the amount of lost energy, this algorithm provides a sharply defined cone for the inclusion of hard, final-state radiation. This algorithm does not resolve close-by jets as expected from final state gluon radiation. The spatial resolution is instead sharply defined by the cone size. The algorithm is applied in the measurement of inclusive jet cross-sections.

C. UA2 Window/Cone Algorithm

The two UA2 algorithms previously described are based on interconnected structures of calorimeter cells (clusters). Low transverse momentum jets may deposit energy into cells not connected to such structures which is therefore not included in the total jet energy. Based on measurements of the transverse energy flow in two-jet events, a fixed jet aperture can be defined. The complete algorithm consists of the following steps:

1. Construct all possible rectangular windows (including overlaps) in the calorimeter map. The window size is 7x5 cells in the $\eta - \phi$ plane corresponding to 1.4 units in pseudorapidity and an azimuthal size of 75 degrees.

2. Search for the window with the highest measured transverse energy summed up over all cells contained in the window.

3. Remove all cells contained in the window from the cell map.

4. Go to 1 unless a pre-defined number of windows is found or all cells are used up.

5. Construct the jet momentum vectors from the scalar transverse energy sum of all cells assigned to the cluster and the $\eta - \phi$ coordinate of the E_T weighted center of the window.

6. Correct the transverse energy of individual windows by collecting cells in a cone of given size around the window axis.

7. Construct the jet momentum vectors from the scalar transverse energy sum of all cells in the cone and the $\eta - \phi$ coordinate of the E_T weighted center of the cone. The jets are massless.

The use of a window representing the typical size of a jet has proven to be a more stable initiator for a jet cone compared to a single cell. The standard cone size used in the UA2 analysis is $\sqrt{(\delta\eta^2 + \delta\phi^2)} < 0.8$. This is the result of an optimization procedure for the two jet mass resolution. This algorithm has been used in the analysis of hadronic decays of the W/Z bosons (jet spectroscopy) [13].

5 e^+e^- CLUSTER ALGORITHMS

This type of algorithm was originally introduced by the JADE collaboration [14]. All cluster algorithms proceed through the following steps :

1. Calculate invariant masses m_{ij} of all 'particle' pairs in the event. The definition of a particle is flexible. It can be a stable particle, a parton, a calorimeter cell or a 'pseudoparticle' as defined below. The calculated invariant mass is then scaled to provide a dimensionless quantity y_{ij} according to $y_{ij} = m_{ij}^2/W$. W is usually the center-of-mass energy of the e^+e^- collider or the total experimentally observed energy.

2. Call the pair with the smallest mass a 'pseudoparticle' and replace the two particles by the pseudoparticle.

3. Go to 1 and repeat the procedure until all y_{ij} values have passed a given cut-off value y_{cut}.

4. The number of jets is given by the number of clusters after passing the mass cut.

The prescriptions of how to combine two particles into one and the definition of m_{ij} are not unique. The following list summarizes five procedures currently in use [14,15,16].

A. The JADE - Algorithm

In the original JADE algorithm [14] two particles i and j are combined into one by adding their 4-vectors:

$$E_{ij} = E_i + E_j,$$

$$P_{ij} = P_i + P_j.$$

Here E is the energy of the particle and P its momentum vector. The pair mass used for the mass cut however, is calculated under the assumption that the particles i and j are massless:

$$m_{ij}^2 = 2E_iE_j(1 - cos\theta_{ij}).$$

B. The E - Algorithm

Here particles i and j are combined by adding their 4-vectors and the invariant mass of the system is also taken from these 4-vectors:

$$E_{ij} = E_i + E_j,$$

$$P_{ij} = P_i + P_j,$$

$$m_{ij}^2 = (E_i + E_j)^2 - (P_i + P_j)^2.$$

C. The E0 - Algorithm

Here the energies of particles i and j are added but the summed momentum vector is re-scaled to give zero invariant mass to the particle combination:

$$E_{ij} = E_i + E_j.$$

$$P_{ij} = (P_i + P_j)\frac{E_i + E_j}{|P_i + P_j|}.$$

The pair mass used for the mass cut is calculated as in the E - algorithm described above.

D. The P - Algorithm

Here the energy of the two particle combination is taken as the absolute value of the summed momentum vector. Also, the invariant mass of the two particle system is zero:

$$E_{ij} = |P_i + P_j|,$$

$$P_{ij} = P_i + P_j.$$

The pair mass used for the mass cut is calculated as in the E - algorithm described above.

E. The P0 - Algorithm

Here energy, momentum and mass of the two particle combination are calculated as in the P-algorithm:

$$E_{ij} = |P_i + P_j|,$$

$$P_{ij} = P_i + P_j,$$

$$m_{ij}^2 = (E_i + E_j)^2 - (P_i + P_j)^2.$$

The difference between this and the standard P-algorithm occurs in the calculation of y_{ij}. Instead of scaling m_{ij} with the center of mass energy of the collider, m_{ij} is scaled with the actual visible energy, which is calculated for each intermediate step from the sum of the energies of all (pseudo)particles in the event.

The implications of the different algorithms for experimentally observed multijet rates have recently been discussed by the OPAL collaboration [16].

6 THE LUCLUS AND LUCELL ALGORITHMS IN THE LUND JETSET PACKAGE

The jet algorithms provided by the JETSET [17] package are designed to represent the e^+e^- cluster algorithms (LUCLUS) and the cone algorithms used in $p\bar{p}$ physics (LUCELL). They are easy to use in the framework of the LUND Monte-Carlo packages. A detailed description of parameters can be found in the most up-to-date JETSET manual (e.g. [17]).

A. LUCLUS

Two different (free size) cluster algorithms with different 'distance measures' are available.

1. Relative Transverse Momentum

In this algorithm, particles with a relative transverse momentum below a given cut-off are joined together. The default value of this cut-off is 2.5 GeV. A detailed description can be found in [18].

2. Invariant Mass

The algorithm is implemented such that it corresponds to the $E0$-scheme described before[19]. The mass cut can be performed on the absolute cluster mass m^2 or on the scaled mass $y = m^2/W^2$. Here W is the total invariant mass of the system under study.

B. LUCELL

The JETSET package has an option for depositing particle transverse energies into a calorimeter grid (e.g. in $\eta - \phi$ space) defined by the user. Energies can be smeared to obtain a rough simulation of the calorimeter resolution. The cone algorithm is then executed as follows:

1. Order cells in decreasing transverse energy E_T.
2. Define a list of potential initiator cells exceeding a given threshold in E_T.
3. Collect the transverse energy of all cells in a cone of a given radius $\sqrt{(\delta\eta^2 + \delta\phi^2)}$ around the initiator cell. If the scalar sum of the transverse energy collected in this cone exceeds a given threshold value, this cone is defined as a jet.
4. Remove all cells used by the jet from the cell list.
5. Go to 2 if there are still cells unassigned.
6. Construct the jet momentum vectors either from the scalar transverse energy sum of all cells in the cone and the $\eta - \phi$ coordinate of the E_T weighted center of the cone (massless jets), or by adding vectorially the momenta of all cells assigned to the jet (jets with mass). Both variations are executed in the LUCELL program.

7 THE GETJET ALGORITHM IN THE ISAJET PACKAGE

The GETJET algorithm is a cone algorithm which is very similar to the UA1 algorithm. The GETJET program proceeds through the following steps [20] :

1. Order cells in decreasing E_T.
2. P_x, P_y, P_z and E of the cells above a threshold (e.g. 0.5 GeV) and within a radius, R_0, in $\eta - \phi$ space of this cell are added vectorially to the leading cell. This results in a four vector for the jet.
3. If the E_T of the jet is above the minimum cluster E_T cut, then the jet is kept. A loop over the remaining cells is performed to find the next jet. This process is repeated until a cluster with E_t less than the cluster E_T cut-off is found. This cut-off procedure assumes that the highest E_T cell will give the highest E_T cluster.

As in the UA1 algorithm, the highest E_T cell in a cluster defines the cluster centroid, it does not recalculate the cone centroid based on the cells it has assigned to the jet. The jet is defined as the vector sum of the cells in the cluster. These jets are not massless.

8 CONCLUSIONS

In this note, we have described various jet identification techniques which are in use in the $p\bar{p}$ collider experiments, e^+e^- collider experiments and in Monte Carlo packages (LUND and ISAJET). For the hadron collider experiments, the clustering methods fall into two catagories: cone algorithms (CDF, UA1), and nearest-neighbor algorithms (UA2). In addition, UA2 has employed a combination of both methods for some analyses. While there are clearly differences between the cone and nearest-neighbor algorithms, we have found that there are also differences among the cone algorithms in the details of how the centroid of a cone cluster is located and how the E_T and P_T of the jet are defined.

The most commonly used jet algorithm in e^+e^- experiments is the JADE-type cluster algorithm. Five of the various incarnations of this approach have been described.

References

[1] S. Ellis, Z. Kunszt, and D. Soper, Phys. Rev. D. **40** 2188 (1989); U. of Oregon Preprint OITS 436 (1990).

[2] F. Aversa, P. Chiapetta, M. Greco, J. Guillet, Phys. Lett. **210B** 225 (1988).

[3] R. Ellis and J. Sexton, Nucl. Phys. **B269** 445 (1986).

[4] J. Huth *et al.*, Toward a Standardization of Jet Definitions, contribution to these proceedings.

[5] CDF Collaboration, F. Abe *et al.*, Nucl. Inst. and Meth. **A271** 387 (1988).

[6] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **62** , 613 (1988).

[7] CDF Collaboration, presented by B. Flaugher, QCD at the Hadron Colliders, to appear in 'Proceedings of the Xth International Conf. on Physics in Collision', Duke University, June 1990; preprint FERMILAB-CONF-90/159-E.

[8] UA1 Collaboration, Arnison *et al.*, Phys. Lett. **132B** , 214 (1983).

[9] E. Buckley, Ph.D. Thesis, A Study of Two-Jet and Three-Jet Production at the CERN $p\bar{p}$ Collider, Rutherford Lab Preprint RAL T 029 (1986).

[10] A.Beer *et al.*, Nucl. Instr. Meth. **224** 360 (1984).

[11] UA2 Collaboration, J.A. Appel *et al.*, Z.Phys. **C30** 341 (1986).

[12] UA2 Collaboration, presented by P.Lubrano, Jet Physics with the UA2 Detector,to appear in the 'Proceedings of the 4th Recontre de Physique de la Vallee D'Aoste',La Thuile,Italy,March 1990.

[13] UA2 Collaboration, J.Alitti *et al.*, CERN-PPE/90-105.

[14] JADE Collaboration, W.Bartel *et al.*, Z.Phys. **C33** 23 (1986).

[15] Z.Kunszt and P.Nason, Physics at LEP I, eds. G.Altarelli *et al.*, CERN 89-08 (1989).

[16] OPAL Collaboration, M.Z. Akrawy *et al.*, CERN-PPE/90-143.

[17] T.Sjöstrand, The Lund Monte Carlo for Jet Fragmentation and e^+e^- Physics, LUTP85-10.

[18] T.Sjöstrand, Computer Phys. Comm. **28** 227 (1983).

[19] T.Sjöstrand, private communication.

[20] ISAJET Event Generation Package, written by F. Paige and S. Protopopescu is decribed in *Proceedings of the Summer Study on Physics of the Superconducting Super Collider*, Snowmass, CO 1986.