High Energy Physics and Nuclear Physics

Volume 21, Number 1

On Jet Identification Algorithm in High Energy Hadronic Collisions

Liu Ming, Liu Feng (Y), Liu Feng and Liu Lianshou

(Institute of Particle Physics, Huazhong Normal University, Wuhan, China)

The "cone algorithm" for jet identification is studied in some detail. Two schemes are proposed to identify jets using the momenta of final state particles. Making use of the hadron-hadron collision data sample produced by LUND-PYTHIA Monte Carlo generator, the two schemes are tested and compared with each other. An effective scheme to identify jets is obtained.

Key words: high energy collision, multiparticle production, hard processes, jet, "cone algorithm."

1. INTRODUCTION

Jets have been observed in many experiments [1] and its identification is one of the most interesting fields in high energy experiments and theories. It is assumed that jets are related to hard processes in QCD, so the measurement of its production cross sections provides an effective tool to test QCD.

Jets give a simple way to approximately describe high energy experiment data. Its basic idea is to calculate the jet state Q out of the final states P for each event via the jet identification algorithm, i.e.,

$$P = [p_1, p_2, \cdots, p_{n_{ptc}}] \xrightarrow[\text{jet finder}]{} Q = [q_1, q_2, \cdots, q_{n_{jet}} + r_1, r_2, \cdots, r_{m_{ptc}}], \tag{1}$$

Received January 15, 1996. Supported by the National Natural Science Foundation of China and the Science Foundation of State Commission of Education.

where p_i is the 4-momenta of final particles, q_j and r_k are 4-momenta of jets and of the particles outside jets in the event, respectively. The multiplicity in high energy collisions can reach several hundreds, but the number of jets in a event is only a few, so the study on jets must be much simpler. At the same time, jets are closely related to the hadronization of quarks and gluons, and one can obtain more information about quarks gluons and their interactions (QCD) by studying jets.

However, the hadronic jet has an inaccurate angle width and an irregular shape. It is difficult to distinguish the particles inside the jet from those outside it. There exists ambiguity in determining whether a particle in an event belongs to a jet or not. Therefore, it is very important to identify a jet effectively in experiments.

In this paper, we obtain an effective scheme to identify the jet according to the "cone algorithm" which is a widely accepted jet definition in QCD. Its definition is briefly described in Section 2. In Section 3, a scheme to identify a jet based on this definition is given. Making use of the hadron-hadron collision data sample produced by the LUND-PYTHIA Monte Carlo generator, this scheme is tested. This scheme is improved in Section 4 by adding the local scanning within the cone radius. Finally, the discussion and summary are presented in Section 5.

2. JET DEFINITION-CONE ALGORITHM

The basic idea of the cone algorithm is to define the jet as a cone with radius R_0 in pseudorapidity and azimuth space (η, φ), if the total transverse energy of the particles in the cone exceeds a threshold E_{t0}. The cone center η_J and φ_J (i.e., jet axis) is the weighted average of all η_i and φ_i inside the jet cone (its transverse energy E_{ti}).

Practically, a cone is drawn around the cone center (η_J, φ_J) with radius R_0, if the distance

$$R = \sqrt{(\eta_i - \eta_J)^2 + (\varphi_i - \varphi_J)^2} \tag{2}$$

from the ith particle to the jet axis (η_J, φ_J) is smaller than R_0, we say that this particle is inside the cone. If the total transverse energy of these particles

$$E_{tJ} = \sum_{i \in \{R \le R_0\}} E_{ti} \tag{3}$$

is greater than the threshold E_{t0}, then the jet exists and its axis (η_J, φ_J) is the mean value of their η_i and φ_i weighted by their transverse energy E_{ti}, i.e.,

$$\left.\begin{aligned} \eta_J &= \frac{1}{E_{tJ}} \sum_{i \in \{R \le R_0\}} \eta_i E_{ti} \\ \varphi_J &= \frac{1}{E_{tJ}} \sum_{i \in \{R \le R_0\}} \varphi_i E_{ti} \end{aligned}\right\}, \tag{4}$$

The jet axis is (η_J, φ_J), and its transverse energy is E_{tJ}.

3. JET IDENTIFICATION SCHEME A

According to the cone algorithm, in order to identify a jet we need to know its axis η_J, φ_J. We obtain it in the following way:

Step 1: Select an axis A_0 to draw a cone C_0 with radius R_0 around it. If the total transverse energy of all particles in the cone exceeds 0.5 E_{t0}, it is the first candidate of the jet axis [2].

Step 2: After averaging over all particles η_i and φ_i weighted by E_{ti} inside the cone C_0, a new axis A_1 is obtained. If the total transverse energy E_1 of the new cone C_1 with radius R_0 around A_1 exceeds 0.5 E_{t0}, then the axis C_1 is the second candidate of the jet axis.

Generally, particles in the new cone C_1 do not always coincide with those in cone C_0 (therefore, $E_1 \neq E_0$), A_1 is obtained from the E_t weighted average over the particles in cone C_0, not over the particle in C_1. So it is necessary to do the next step.

Step 3: Sum over all particles η_i and φ_i weighted by their E_{ti} in cone C_1 to obtain a new axis A_2. A cone C_2 with radius R_0 is drawn around A_2. If the total transverse energy of all particles in cone C_2 is above 0.5 E_{t0}, then axis A_2 is the third candidate of the jet axis.

Such iteration repeats until the total transverse energy of all particles in the cone is below 0.5 E_{t0} during the iteration. This means that a jet cannot be found from axis A_0. If during the iteration the total transverse energy is always greater than 0.5 E_{t0}, and the total transverse energy E_i of all particles in cone C_i is equal to that in cone C_{i-1}, i.e., $E_i = E_{i-1}$, the iteration stops. In this case if E_i is above the threshold E_{t0}, we say the jet is found, and A_i is the jet axis.

The above iteration is the key procedure of the cone algorithm to find the jet. What remains is how to select the initial axis A_0 as our starting point. A simple way to do this is to scan in the whole (η, φ) region, namely:

$$-\eta_{\max} < \eta < \eta_{\max} \quad 0 < \varphi < 2\pi$$

(Only the central pseudorapidity region is discussed in this paper, $\eta_{\max}$ is the cut for pseudorapidity $|\eta|$). This region is divided into many segments and each lattice is selected as the initial axis A_0. This method wastes computing time and is unnecessary.

Considering that the jet axis often directs to the region where particles concentrate, as the first step we could select the momentum direction of a certain final state particle as the initial axis A_0. The scheme is found as follows:

Scheme A

(1) For any particle in the region $|\eta| \leq \eta_{\max}$, choose its (η_i, φ_i) as axis, and a cone with radius R_0 is drawn around this axis. Compare the total transverse energy of all cones; we will select the cone axis which has the maximum total transverse energy in that cone as the initial axis A_0.

(2) Starting from A_0, according to the iteration method described above, weighted average iteration repeats until the total transverse energy of any cone is below 0.5 E_{t0} during the iteration. Then this iteration stops. Now choose the axis with the total transverse energy next to maximum as the initial axis A_0 and repeat the above iteration. On the other hand, if a jet could be found after iteration, then proceed to the next step.

(3) After particles belonging to the jet are excluded from all final particles in the event repeat steps (1) and (2) for the left particles to find the next jet until no new jet can be found any more.

In order to test this scheme, we apply it to an actual data sample. Using the LUND hadron-hadron collision Monte Carlo generator PYTHIA, 200,000 nonsingle diffractive hadron-hadron collision events at $\sqrt{s} = 630$ GeV are generated. We consider only the central pseudorapidity region $|\eta| \leq 1.5$ and there are 188,552 events left. The parameters used to identify the jet are $R_0 = 1.0$ and $E_{tJ} = 4.0$ GeV, which correspond to low transverse momentum jets or minijets. Applying the above scheme A, the percentage of jet events is 8.09%. In Figs. 1(a) and 1(b), the multiplicity distributions of jet events and nonjet events are shown; Figs. 1(c) and 1(d) present the distributions of the average transverse momentum per event of the two samples. The multiplicity distribution of jet events is almost symmetric at the peak; however, that of nonjet events has a long tail. For jet events, the distribution of the average transverse momentum per event $\bar{p}_t$ exists a sharp curve at low $\bar{p}_t$ ($\bar{p}_t < 0.3$). These coincide with the properties of (mini-) jet events.

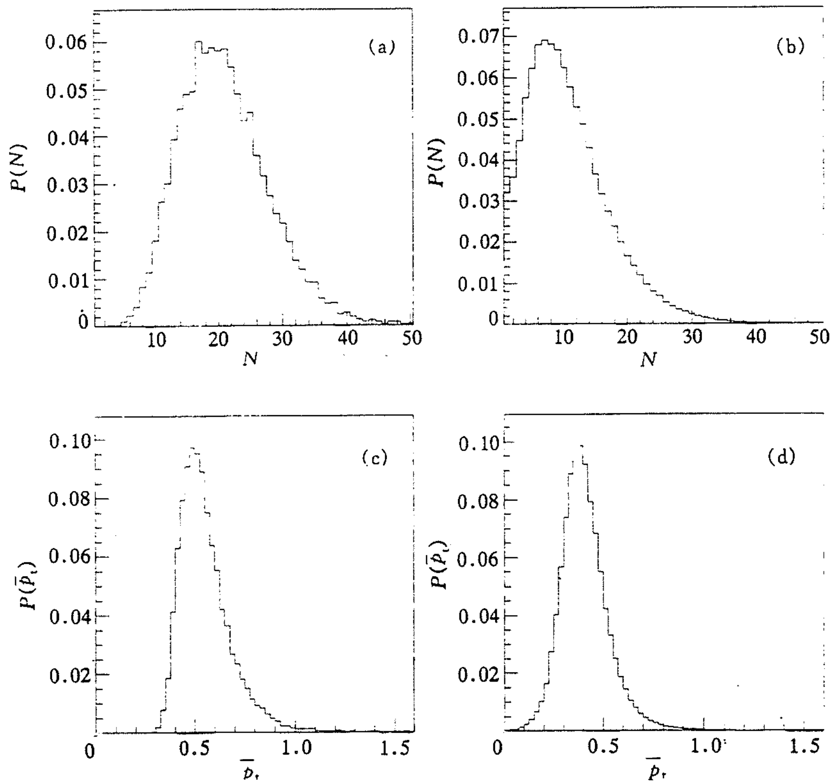

Fig. 1
The distributions obtained by Scheme A without scanning.

4. IMPROVEMENT OF JET IDENTIFICATION SCHEME

In the above scheme A, only those events where the jet axes are near the particles are taken into account. So the jets with their axes far away from particles might be lost. In order to overcome this defect, the local scanning is performed based on the scheme A to obtain a more complete scheme.

Scheme B

(1) The total transverse energy of all particles in a cone, with radius R_0 drawn around each particle's (η_i, φ_i), is calculated and compared with each other. The cone axis which corresponds to the maximum total transverse energy is singled out as the initial axis, and named A_{01}.

(2) Scanning around A_{01} in the region $\Delta\eta = \pm R_0$, and $\Delta\varphi = \pm R_0$ with the step 0.1 R_0. Each scanning point is regarded as the cone axis and the total transverse energy of the cone with radius R_0 is calculated. The axis corresponding to the maximum total transverse energy is singled out as A_{02}.

(3) Starting from A_{01} and A_{02}, the iteration is performed according to the above scheme, respectively. If the total transverse energy of a cone is below 0.5 E_{t0}, the iteration stops. If such case does not happen and we find two new cone axes, then compare the total transverse energy of these two cones, we select the one with larger total transverse energy as the jet candidate. If the total transverse energy of the this cone is above the threshold E_{t0}, a jet is found.

(4) After the particles inside the jet are excluded from all final state particles, steps (1)-(3) are repeated to find the next jet for the remaining particles. If jet is found again, the particles in it are excluded again. The steps (1)-(3) are repeated until one cannot find a new jet.

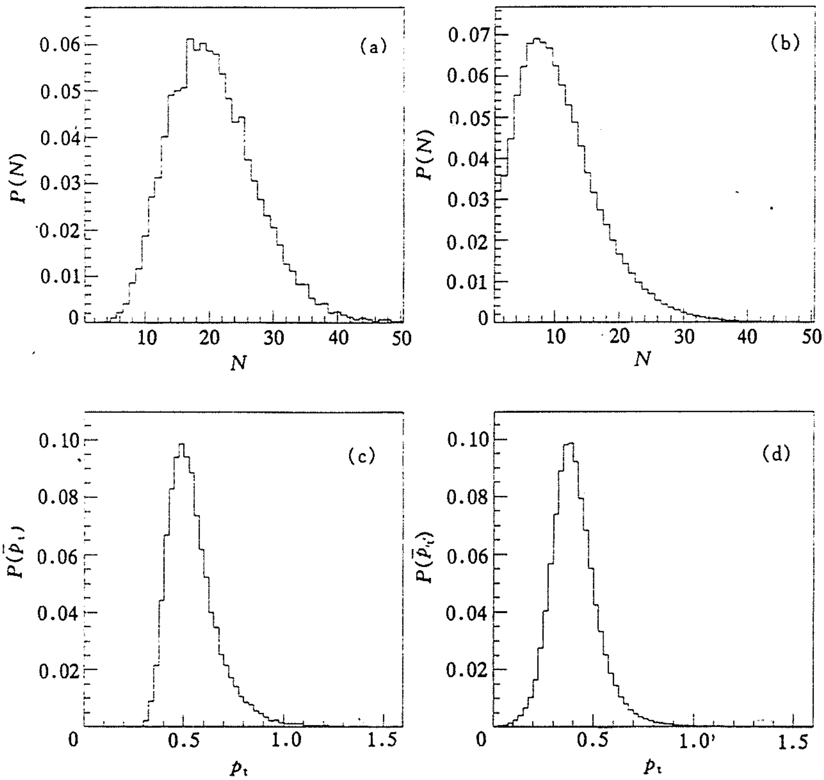

Fig. 2

The distributions obtained by Scheme B with scanning.

We apply scheme B to the same Monte Carlo data sample generated by PYTHIA, and the percentage of jet events is 9.19%. Figures 2(a) and 2(b) show the multiplicity distributions of jet events and nonjet events. The average transverse momentum per event $\bar{p}_t$ are presented in Figs. 2(c) and 2(d).

Comparing the results between the two schemes, there are about 12% jet events lost in the scheme without local scanning. In order to see the properties of the jet events that are not found in scheme A more clearly, two typical such events in (η_i, φ_i) space are shown in Fig. 3. [Figure 3(a) shows one jet found with scanning, but no jet without scanning; Fig. 3(b) shows two jets found with scanning, but only one without scanning.] In Fig. 3, the points mean the momentum directions of final particles, the circles mean jets. The letter S's nearest circles stand for jets found via scheme B (with local scanning), N's for jets found via scheme A (no scanning). The number is the transverse energy of the jets (GeV, accuracy in 0.5 GeV). It is shown in Fig. 3 that the jet can be found only through scheme B if its axis is far away from the momentum directions of final particles.

On the contrary, for some events, local scanning reduces the number of jets in the event. Figure 4 displays a typical event (the figure caption is the same as in Fig. 3). In Fig. 4, local scanning changes the two jets with lower transverse energy into one jet with much higher transverse energy. Such events are very few; there are only 65 such events within our data sample (about 190,000 events). Physically, it corresponds to hard gluon emission. If the emitted gluon is not too hard, it can be regarded as one jet, but it will form two jets if it is hard enough. Whether such events have one or two jets has some ambiguity.

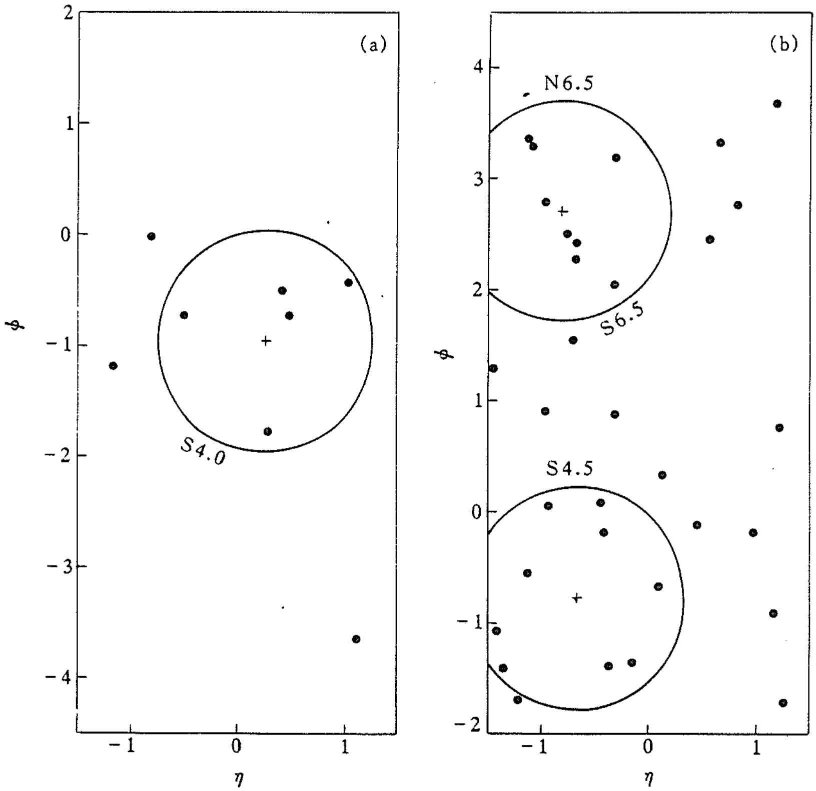

Fig. 3
Examples where scanning is necessary for finding jets.

5. SUMMARY AND DISCUSSION

According to the jet definition-cone algorithm, two schemes to identify jets are given. Scheme A, which is based on the momenta of final particles, is simple and easily carried out, but some (12%) of the jet events are lost. Such jet events can be found in the improved scheme B, because local scanning in the square region with the size of the diameter of the cone in (η_i, φ_i) plane is introduced. This is also easily carried out on the computer, so it is an ideal feasible scheme to identify jets.

From a dynamic point of view, jets are hadronic clusters fragmenting from hard partons. A jet has no sharp edge or accurate angle width in the phase space; there is no clear difference between one jet and two jets produced by hard gluon emission. The particles inside different jets and particles in the jet or produced by soft processes crisscross. This will make it difficult to define and identify jets in the event sample. All practical jet definition is an attempt to make the vague conditions precise, so in some sense it is an approximation.

According to the cone algorithm, we give a jet a distinct shape and size in our schemes without considering the intersection among the particles inside different jets and those in jets and from soft processes. This is not completely in accord with the actual dynamics; rather, it is an approximate description method.

There may exist an ambiguous case after a few jets are identified according to the cone algorithm and one must make a choice. In our schemes we assume that the higher transverse energy jet has privilege.

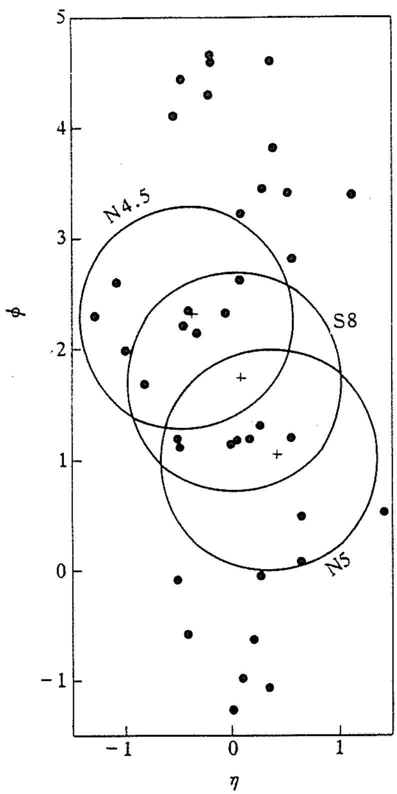

Fig. 4
Examples where scanning reduces the number of jets.

For example, if we start from particles a and b, after averaging weighted by E_t, we can find jets from both of them. However, if the particles in one jet are excluded and the remaining particles cannot be identified as a jet, then which jet that can be found depends on the order. In our schemes we choose the higher transverse energy jet.

As another example, we start from particles a and b, where we can find two jets. If starting from some points between them, named c, we can also find one jet (see Fig. 4); if the particles in the latter jet are excluded and the two original jets could not be formed from the remaining particles, then we will choose the latter as the final one if its transverse energy is larger than the two original jets.

In dynamics, when high-transverse momentum partons fragment into final hadronic jets, the "higher transverse energy has privilege" rule is reasonable. It also reasonable to use it to eliminate the uncertainty of the jet definition in the cone algorithm.

We apply our schemes in the Monte Carlo sample produced by the PYTHIA generator, and the results on the jet event cross sections and the distribution of multiplicity and the event average transverse momenta are accord with the UA1 experimental result. The difference between the jet cross sections of the two schemes is about 12% but the distributions of multiplicity and the average transverse momenta per event are almost the same. This indicates that both schemes can be used to identify jets though the scheme with local scanning is more complete.

REFERENCES

[1] G. Hanson, *et al.*, *Phys. Rev. Lett.*, **35** (1975): p. 1609; M. Banner, *et al.*, *Phys. Lett.*, **118B** (1982): p. 203; G. Arnison, *et al.*, *Phys. Lett.*, **123B** (1983): p. 115.
[2] S.M. Berman, J.D. Bjorken and J.B. Kogut, *Phys. Rev.*, **D4** (1971): p. 3388; G. Sterman and S. Weinberg, *Phys. Rev. Lett.*, **39** (1977): p. 1436.
[3] S.D. Ellis, Z. Kunszt and D.E. Soper, *Phys. Rev.*, **D40** (1989): p. 2188.
[4] H.U. Bengtsson and T. Sjöstand, *Comput. Phys. Commun.*, **46** (1987): p. 43.
[5] G. Giapetti (UA1), *Proc. of the 5th Topical Workshop on p-$\bar{p}$ Collider Phys., St Vincent Aosta Valley, Italy, 1985*, edited by M. Greco, Singapore: World Scientific, 1985.
[6] Y.F. Wu and L.S. Liu, *Phys. Lett.*, **B269** (1991): p. 28; Liu Feng, Wu Yuan Fang and Liu Lianshou, *Chinese J. Nucl. Phys.*, **15** (1993): p. 7.
[7] Y.F. Wu *et al.*, (UA1 mini bias), *Acta Phys. Slov.*, **44** (1994): p. 141.