ALEPH 89 - 150
PHYSIC 89 - 60
J.M. Scarr, I.ten Have
18 - 9 - 1989

PTCLUS: a Jet Finding Algorithm for High Energy Hadronic Final States

J. M. Scarr, I. ten Have

Department of Physics and Astronomy, Glasgow University,
Glasgow G12 8QQ, Great Britain

Due to confinement, quarks and gluons are not observed in a free state. The partons produced in e^+e^- collisions immediately hadronize and form a shower of physical particles, a jet. The particles formed by the hadronization process are collimated around the most energetic partons. To reconstruct the original parton direction and its energy, a jet finding algorithm is needed.

In this note we describe the PTCLUS jet finding algorithm[1]. PTCLUS will be tested on Monte Carlo events. Its performance will be compared to that of LUCLUS[2] and to that of the JADE mass algorithm[3].

1. The PTCLUS Jet Finding Algorithm

The PTCLUS jet finder consists of three steps.

Step I

The particles are sorted by energy. The most energetic particle is taken as the first jet initiator. The algorithm then loops through all the remaining particles. If the angle between the particle momentum vector and the cluster momentum vector is less than 90 degrees, $(\vec{p}.\vec{p}_{clus}) > 0$, a check is made whether the track should be added to the cluster. The p_T of the track with respect to the cluster is calculated using:

$$p_T^2 = \frac{(\vec{p}.\vec{p})(\vec{p}_{clus}.\vec{p}_{clus}) - (\vec{p}.\vec{p}_{clus})(\vec{p}.\vec{p}_{clus})}{(\vec{p}+\vec{p}_{clus}).(\vec{p}+\vec{p}_{clus})} \tag{1}$$

If the p_T^2 of the track with respect to the cluster is smaller than 0.25 $(GeV/c)^2$ then the track is added to the cluster:

$$\vec{p}_{clus} = \vec{p}_{clus} + \vec{p}_{track} \tag{2}$$

Where the p_T^2 exceeds the limit of 0.25 $(GeV/c)^2$ then the track is used as a new jet initiator.

The procedure is repeated until all tracks have been assigned to a cluster.

Step II

In step II, clusters separated by only a small distance are merged. First of all the angle between the two clusters considered is required to be less than 90 degrees. Subsequently the invariant mass of the two clusters, scaled by the visible energy squared, is used as a distance measure:

$$Y = \frac{M^2}{E_{vis}^2} \tag{3}$$

The mass is scaled by the visible energy to correct for particle losses. The invariant mass is calculated using:

$$M^2 = 2E_iE_j(1 - \frac{(\vec{p}_i \cdot \vec{p}_j)}{|\vec{p}_i||\vec{p}_j|}) \tag{4}$$

The cluster merging stops when Y < $Y_{\rm lim}$. $Y_{\rm lim}$ can be set by the user. A value of $Y_{\rm lim}$ = 0.01 is recommended. This corresponds to a mass of approximately 10 GeV/c^2. If the user requires a definite number of clusters the cluster merging can also be stopped when this number is reached.

Step III

After step II particles belonging to a certain merged cluster may have a smaller p_T with respect to another cluster than to the one they have been assigned to. If this is the case the particle is reassigned to the other cluster.

2. Performance of the PTCLUS Algorithm

The performance of the PTCLUS jet finding algorithm has been studied by comparing the number of clusters found to the number of partons present in Monte Carlo events. This study was done for different values of $Y_{\rm lim}$, so for different resolutions of the algorithm.

Firstly at parton level one has to determine how many parton clusters would be seen by the jet finder for a given resolution. If the angle between two partons is small their jets will probably be reconstructed as one cluster by the jet finder. The jets from low energy partons are likely to be merged into a nearby cluster. To determine the number of parton clusters the quarks and the gluons are used as input for PTCLUS. The same event is studied once more after fragmentation and weak decays. The final state, stable particles are now used as input for the jet finder. As the neutrinos remain undetected they are not included. In anticipation of future heavy flavour studies leptons are also excluded. The output now represents the number of jets that would be reconstructed in the event.

Figure 1 shows the average number of parton clusters (dashed line) found in a sample of 500 Lund JETSET 6.3 parton shower $b\bar{b}$ events. The full line represents the average number of jets found in the same sample of events. Figure 1 shows that for the range of $Y_{\rm lim}$ studied the agreement between the average number of parton clusters and the average number of jets is good. As one goes to lower $Y_{\rm lim}$, that is to higher resolution, the two lines start to diverge. At very high resolution some discrepancy may be expected. For low resolution (high $Y_{\rm lim}$ values) the agreement is very good. Figure 1 also shows that for low resolution a large fraction of the events become two jet events irrespective of the event structure.

Similar studies have been carried out for LUCLUS[4] and for the JADE mass algorithm[3]. The fractions of two, three and four jet events as a function of $Y_{\rm lim}$ is very similar for the three jet finders.

The performance of PTCLUS can also be tested by looking at how well the algorithm associates the cluster and the corresponding parton. For this purpose various sets of Monte Carlo events have been generated. In this paper we will present two cases. The first set is a set of $u\bar{u}$ events generated using Lund, JETSET 6.3. The $O(\alpha_s^4)$ matrix element processes were used in combination with the string fragmentation model. The second sample contains Lund JETSET 6.3 $b\bar{b}$ events generated using the parton shower model with again string fragmentation. The first set of events is somewhat simpler than the second set as u-jets are more collimated than b-jets. Also u-jets do not contain leptons with a high p_T relative to the jet as may be the case in a b-jet. Both samples, however, form quite a stringent test of the jet finding algorithm as the jet multiplicity is always four or more.

PTCLUS and the JADE mass algorithm were run with Y_{lim}=0.01 which is equivalent to a mass of 10 GeV/c^2 at $\sqrt{s} = 91$ GeV. In the LUCLUS algorithm the maximum distance was set to 3.5 GeV. This is roughly equivalent to $Y_{lim} < 0.01$.

For the Monte Carlo event samples one can compare the angular difference between the cluster and the original parton. Figure 2 (upper half) shows the angular difference as a function of the energy of the parent parton for the u-quark sample. Figures 2a, b, c show the distribution for PTCLUS, LUCLUS and the JADE mass algorithm respectively. In figure 3 (upper half) the same study is repeated for the b-sample. The figures clearly show that the reconstruction of low energy jets is far more difficult than the reconstruction of high energy jets. All three algorithms become unreliable at $E_{parton} < 5$ GeV. The lower halves of figures 2 and 3 show the energy difference between the reconstructed jet and the original parton as a function of the energy of the parton. Again figures a, b, c show the comparison for PTCLUS, LUCLUS and the JADE mass algorithm respectively. In general the energy difference is small, less than 2 GeV. From figures 2 and 3 one can see that globally the three algorithms show a similar behaviour.
A more detailed comparison is made in tables 1 and 2. The Monte Carlo events are now divided into three energy bins: $5 < E_{parton} < 10$ GeV, $10 < E_{parton} < 20$ GeV, $E_{parton} > 20$ GeV. The tables show that the performance of PTCLUS is comparable to that of LUCLUS. The association by the JADE mass algorithm on the other hand is less good.

Another important asset of the PTCLUS algorithm is its speed. The algorithm is at least a factor three faster than the LUCLUS algorithm, with a speed comparable to that of the JADE mass algorithm. Note, however, that PTCLUS can also become appreciably faster than the JADE mass algorithm when the number of tracks in the event is large ($N > 60$).

3. Conclusions

Over the range of Y_{lim} studied the number of clusters agrees well with the number of parton clusters found in the events. In the figures the angular difference and the energy difference between the parton and the associated jet looks quite similar for the three different jet finding algorithms. However, if one looks at the mean values and r.m.s. of the distributions one can see that the performance of PTCLUS and LUCLUS is clearly better than that of the JADE mass algorithm. The performance of PTCLUS is comparable to that of LUCLUS. However, PTCLUS has two advantages. It is at least three times faster than LUCLUS and it uses Y to define the jet resolution.

References

[1] R. Settles, Aleph 89-130, Physic 89-53

[2] T. Sjöstrand et al., The LUND Monte Carlo programs (1987)

[3] W. Bartel et al. (Jade collaboration), Z. Phys. C 33, (1986), 23

[4] D. Parker, Private Communication

Figure Captions

Figure 1: shows the percentage of 2-, 3- or 4-jet events (full line) found by PTCLUS as a function of Y_{lim}. This is to be compared to the percentage of parton clusters (dashed line). The input $b\bar{b}$ events were generated with the Lund, JETSET 6.3 parton shower Monte Carlo.

Figure 2: Comparison of the found jets with partons for $u\bar{u}$ events generated with the Lund JETSET 6.3 matrix element Monte Carlo. Above: $\Delta\theta$ (angle between the jet and the parton) versus E_{parton}, below: E_{parton}-E_{jet} versus E_{parton} for a) PTCLUS, b) LUCLUS, c) JADE mass algorithm

Figure 3 Comparison of the found jets with partons for $b\bar{b}$ events generated with the Lund JETSET 6.3 parton shower Monte Carlo. See figure 2.

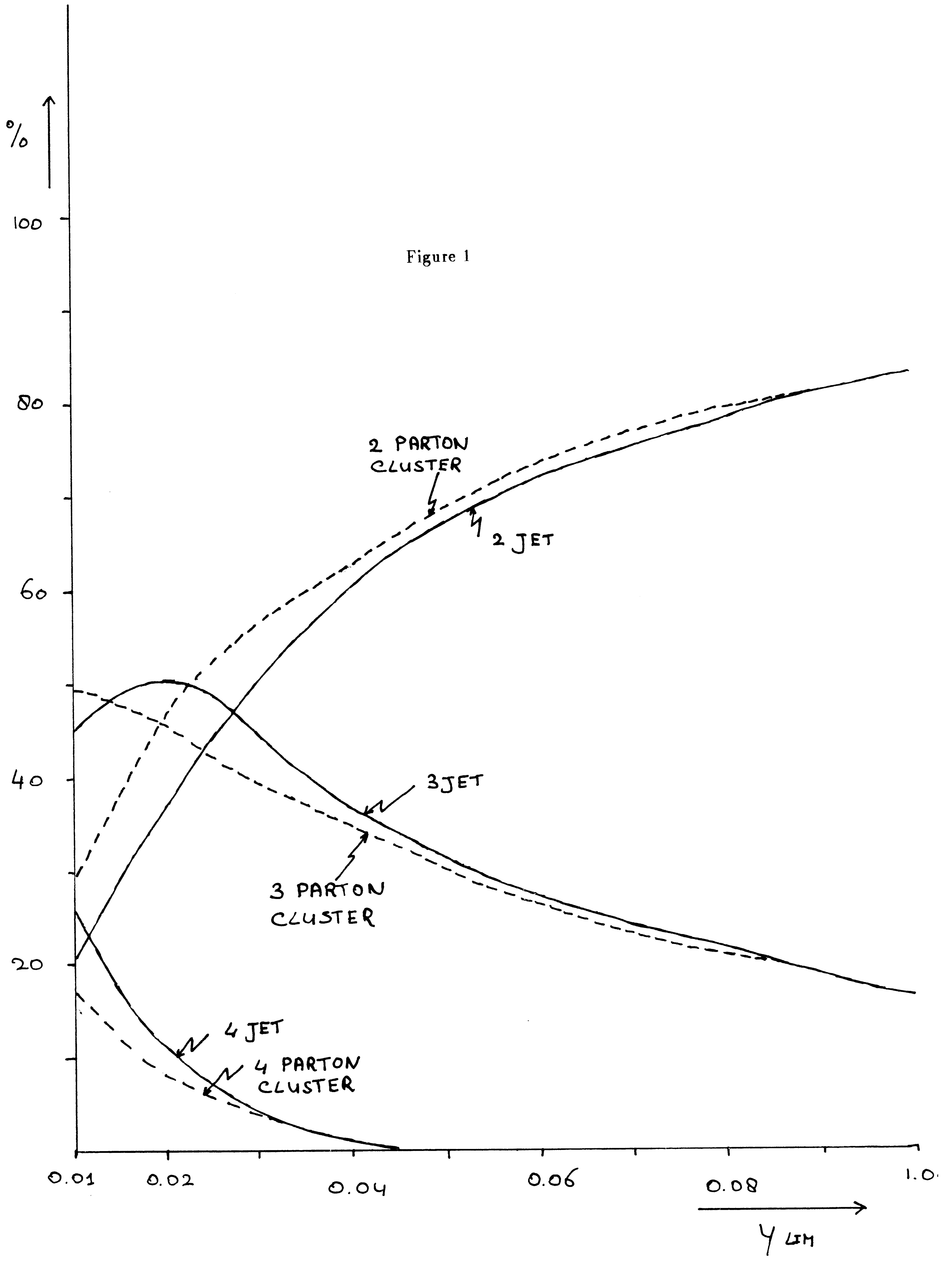
%
100
80
60
40
20
2 PARTON CLUSTER
2 JET
3JET
3 PARTON CLUSTER
4 JET
4 PARTON CLUSTER
0.01
0.02
0.04
0.06
0.08
y LIM

Figure 1

Figure 2

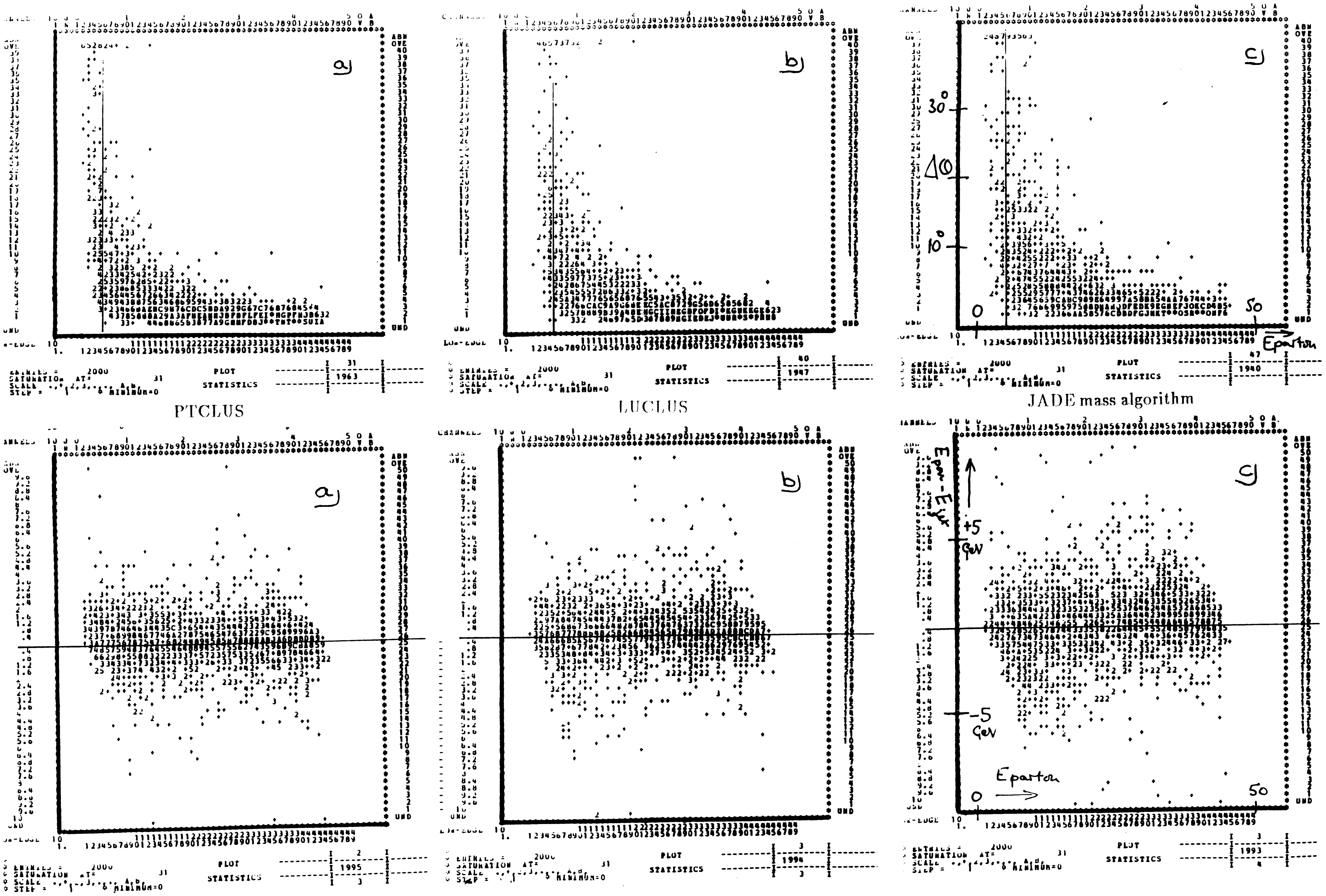
a)
b)
c)
30°
$\Delta\Theta$
10°
0
50
E_{parton}
PTCLUS
LUCLUS
JADE mass algorithm
$E_{par} - E_{jet}$
+5 GeV
-5 GeV
PLOT
STATISTICS

Figure 3

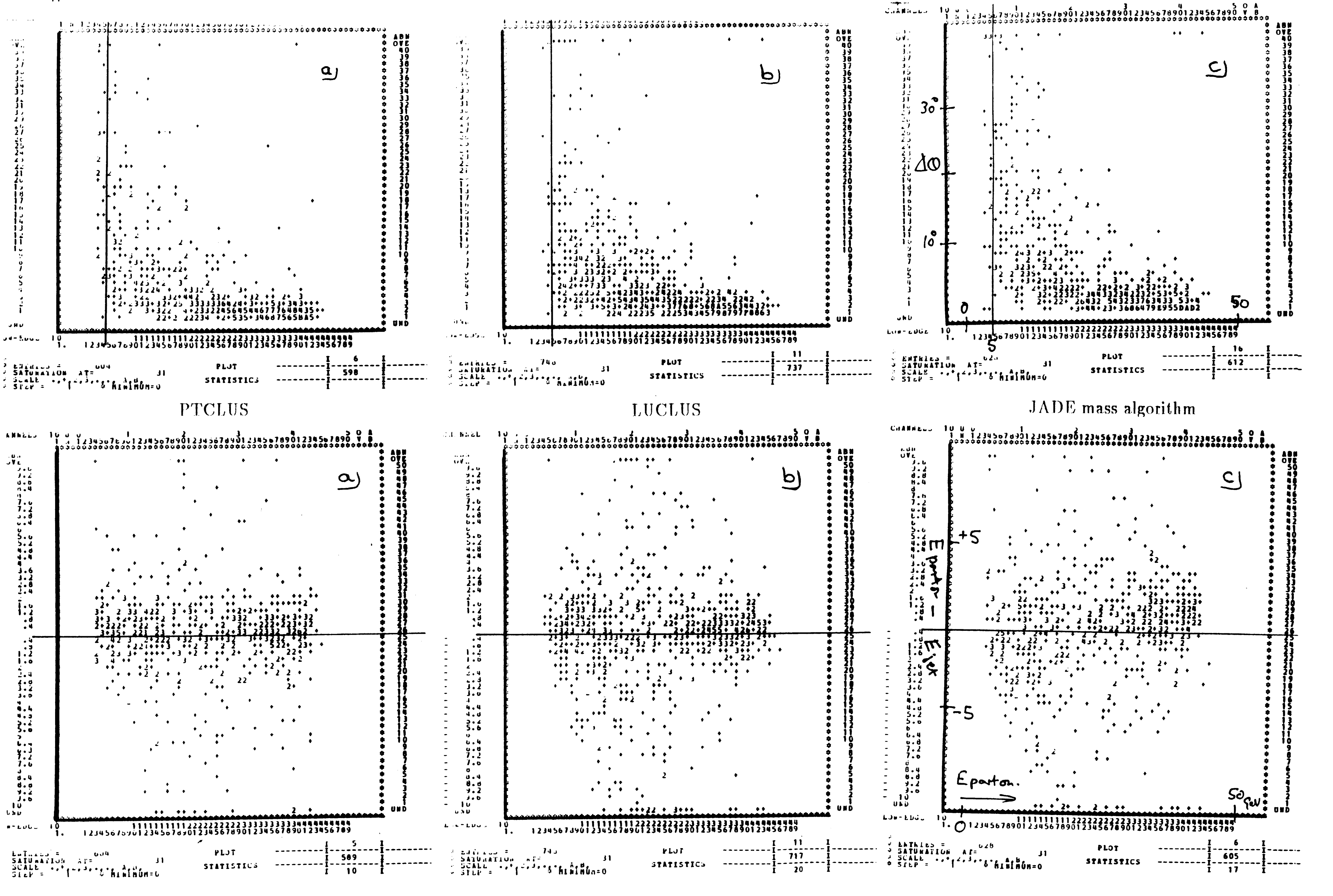
a)
b)
c)
30°
Δθ
10°
0
5
50
PTCLUS
LUCLUS
JADE mass algorithm
+5
-5
E parton - E jet
E parton
50 GeV
PLOT
STATISTICS

table 1a: $u\bar{u}$ Lund $O(\alpha_s^4)$ matrix element						
Algorithm	ΔE (GeV) $5<E_{part}<10$	r.m.s. $5<E_{part}<10$	ΔE (GeV) $10<E_{part}<20$	r.m.s. $10<E_{part}<20$	ΔE (GeV) $E_{part}>20$	r.m.s. $E_{part}>20$
PTCLUS	0.285	1.60	0.160	1.54	0.065	1.44
LUCLUS	-0.143	1.77	0.140	1.68	0.214	1.53
JADE Mass	-0.503	1.99	-0.103	2.36	0.414	2.02

table 1b: $b\bar{b}$ Lund parton shower model						
Algorithm	ΔE (GeV) $5<E_{part}<10$	r.m.s. $5<E_{part}<10$	ΔE (GeV) $10<E_{part}<20$	r.m.s. $10<E_{part}<20$	ΔE (GeV) $E_{part}>20$	r.m.s. $E_{part}>20$
PTCLUS	0.280	1.85	0.314	2.85	-0.370	2.58
LUCLUS	0.161	1.71	-0.043	3.47	-0.033	2.71
JADE Mass	-1.029	2.43	-0.493	3.39	0.449	2.75

table 2a: $u\bar{u}$ Lund $O(\alpha_s^4)$ matrix element						
Algorithm	$\Delta\theta$ (degr.) $5<E_{part}<10$	r.m.s. $5<E_{part}<10$	$\Delta\theta$ (degr.) $10<E_{part}<20$	r.m.s. $10<E_{part}<20$	$\Delta\theta$ (degr.) $E_{part}>20$	r.m.s. $E_{part}>20$
PTCLUS	7.8	5.3	4.1	3.3	1.5	1.2
LUCLUS	8.8	6.7	4.1	3.4	1.4	1.1
JADE Mass	10.6	7.8	5.3	4.9	1.6	1.7

table 2b: $b\bar{b}$ Lund parton shower model						
Algorithm	$\Delta\theta$ (degr.) $5<E_{part}<10$	r.m.s. $5<E_{part}<10$	$\Delta\theta$ (degr.) $10<E_{part}<20$	r.m.s. $10<E_{part}<20$	$\Delta\theta$ (degr.) $E_{part}>20$	r.m.s. $E_{part}>20$
PTCLUS	11.2	8.4	6.0	5.7	2.3	2.7
LUCLUS	10.1	7.5	6.9	6.2	2.6	3.4
JADE Mass	15.7	10.1	8.6	7.5	2.4	2.8