SLAC-PUB-2623
October 1980
(T/E)

A CLUSTER ALGORITHM FOR THE STUDY OF JETS IN HIGH ENERGY PHYSICS

Jonathan Dorfan
Stanford Linear Accelerator Center
Stanford University, Stanford, California 94305

ABSTRACT. A cluster algorithm using angular correlations and leading particle effects is presented which is applicable to the study of jets produced in high energy storage ring collisions. The algorithm uses the concept of a minimal spanning tree and is computationally very efficient. Events are classified by their cluster number and the cluster number frequency distribution can be used for comparison with particle production models. Individual particles are assigned to the clusters and the vector sum of their momenta generate a cluster axis. These cluster properties permit the study of the dynamics of the jet production and fragmentation processes. The example of two and three jet production at PEP and PETRA energies is used to illustrate this technique.

Submitted to Zeitschrift für Physik C

---

*Work supported by the Department of Energy, contract DE-AC03-76SF00515.

Introduction

With the advent of high energy storage rings, the study of hadronic jets has become increasingly more important. Jets are the hadronic manifestation of high energy quarks and gluons which are produced in the storage ring collisions. The higher the quark and gluon energies the more faithfully the resulting hadrons follow the produced constituent direction, producing cone-like particle patterns (jets). Among the more interesting examples of jet production are: (a) high transverse momentum quark jets in pp collisions at the ISR; (b) the anticipated production of many ($\gtrsim 6$) jets at the high energy $\bar{p}p$ colliders; (c) the production of two collinear jets via one photon exchange in e^+e^- storage rings and the extension to three (and four) jet events at PEP and PETRA energies; (d) the production in e^+e^- collisions of two non-collinear hadron jets via the two photon process; and (e) the production of three gluon jets in the decay of resonances which are bounds states of heavy quark-antiquark pairs (Τ). The majority of the jet pattern recognition algorithms currently employed are not sufficiently general in their approach to handle all the different jet "geometries" which are manifested in examples (a) — (e). This paper presents a method which searches directly for the particle direction correlations (clusters) and classifies events by their cluster content. As the method is described it will become clear that it has broad applicability and is largely independent of the "geometry" of the produced jets.

The example of three jet events in e^+e^- storage ring collisions will be used to illustrate the power of the cluster method. At sufficiently high energies ($E_{CM} \gtrsim 8$ GeV) continuum hadronic events exhibit clear two jet structure; the jets arising from the hadronization of the quark and antiquark. Quantum Chromodynamics (QCD) predicts unambiguously that, at sufficiently high constituent energies, gluons are radiated by the quarks and antiquarks. These gluons can have large fractional energies and will themselves materialize as hadron jets. QCD also predicts that heavy quarkonium resonances (e.g., the Τ) should decay to hadrons via three gluons which will result in three hadronic jets. Experimental verification of three jet events thus becomes a crucial test of QCD.

One of the objectives of the cluster approach is to avoid biasing the pattern recognition procedure in favor of the predicted jet properties. In this way the algorithm will have general applicability and furthermore kinematic tests performed with the clusters will be more meaningful. The two most widely used algorithms for the study of gluon jets at the Υ resonance [1] and quark-antiquark-gluon jets at PETRA [2] are triplicity [3] and trijettiness [4]. Both of these methods have been successful at isolating three jet events at PETRA [2]. These two algorithms are designed specifically for the three jet topology in that they construct three and only three jet axes in the candidate hadronic events. Hence they do not have general applicability [5]. In the case of trijettiness the three axes are constructed to lie in a plane. This is not a necessary requirement for the triplicity method, but the use of a ghost particle to account for the missing momentum (as in Refs. [2,3]) guarantees momentum conservation and hence planarity of the three jet axes. Triplicity and trijettiness are therefore constructed to find planar three jet events. These two algorithms are protected against erroneous assignments by virtue of the quality test measures triplicity and trijettiness. This paper discusses an alternative algorithm which searches hadronic events for particle clusters and classifies the events according to the cluster number [6]. In this way 1, 2, 3, ... n jet events are handled by a single algorithm. Comparison with different production mechanisms is achieved using the cluster frequency distribution. The clusters are characterized by the particles which span them and an axis which is the vector sum of the momenta of these particles. Hence the dynamics of the cluster (jet) production and hadronization can be studied. Besides having broad applicability, the algorithm described below is computationally very fast.

The Algorithm

The algorithm described herein uses particle direction correlations to define clusters. The conciseness and success of this algorithm results from the use of a minimal spanning tree (MST) to form joining elements between particle direction vectors. The mathematical rigor and the application of MST's to pattern recognition problems is described in Ref. [7]. A short description of the construction of an MST is given here. Suppose you wish to connect the coordinate points shown in Fig. 1(a) with joining elements such that there is a connected path between any two points and the total length of the connecting elements is a minimum. There is a unique solution to this problem which is shown in Fig. 1(b) and the configuration of joining elements is called a minimal spanning tree. For convenience some further MST "jargon" is introduced. The coordinate points to be spanned by the MST are called nodes. The lines which connect nodes are referred to as edges and edges are either bridging [8] (the edge de) or nonbridging (the edge bc). If one removes a bridging edge from the MST one generates two subtrees (which incidentally are both MST's), but removal of a nonbridging edge from the tree merely isolates a single node. A bridging edge which is long compared to the typical edge distances in the MST is termed inconsistent. The procedure for cluster analysis using the MST is as follows. A distance measure (metric) is chosen which allows for the calculation of edge distances. Using this metric, the MST is obtained for the set of nodes under consideration. Clusters will result from groups of neighboring nodes which have small edges, where small is to be interpreted relative to the inconsistent edge(s) which separate different portions of the MST. Figure 1(c) shows how this procedure would "break" the MST in Fig. 1(b) at edges de and ek to yield three clusters. The discriminating power in any particular application arises from the presence of clearly defined inconsistent edges in the MST's

For this application the distance metric for particles i and j is given by

$$d_{ij} = \theta_{ij}^2 * W_{ij}\left(P_i, P_j\right) \tag{1}$$

where θ_{ij} is the angle between the two particles. The factor θ_{ij}, which is the geodesic distance between the two particles i and j as evaluated on the unit sphere, ensures that particles correlated in direction will have small interparticle distances. The function $W_{ij}(P_i,P_j)$ is a weight derived from the magnitudes of the particle momenta P_i and P_j. Several different weighting functions have been tried. The weights $W_{ij}=1$ and $W_{ij}=P_i*P_j$ have been rejected in favor of

$$W_{ij}(P_i,P_j) = w_i(P_i) * w_j(P_j) = P_i^{-1} * P_j^{-1} \quad . \tag{2}$$

This weighting function emphasizes leading particle effects in the clusters; the consequences of these effects are discussed later. It is important to realize that for the metric discussed above, a planar representation of the nodes which preserves the lengths of the edge distances (as in Fig. 1) is not in general possible. Figure 1 is used only to clarify the terminology and construction of an MST.

Monte Carlo generated events for the process $e^+e^- \rightarrow$ hadrons at $E_{CM} = 30$ GeV have been used to develop and optimize the pattern recognition code. These Monte Carlo models are described in more detail in the next section. The algorithm deals equally with neutral and charged particles although the software has the capability of ignoring the neutrals. The starting point is to construct an array of unit vectors $A_{ij} = P_j^{(i)}/|\vec{P}_i|$ $(i=1,N; j=1,3)$ for the N particles which are to be used in the cluster search. All stable particles (photons + charged particles) are used in the cluster search. The matrix A_{ij} along with the particle weights $w_j = P_j^{-1}$ are supplied to a routine PRIM [9] which constructs the MST using the metric defined in the previous section.

The next step is to search for inconsistent edges so that the MST can be subdivided. The problem of defining the appropriate distance scale, known in statistical parlance as the search for a "robust measure," is common in applications of this sort. The average (geometric mean) distance, for instance, is not a good measure. Instead, the scale is measured relative to the median distance which is obtained by ordering the $M=N-1$ edge distances on the real line and assigning the M/2th distance as the median [10]. The algorithm will subdivide the MST into two parts if the ratio of the longest inconsistent edge to the median distance is greater than the parameter R_1 ($R_1 \simeq 2$). If this condition is not met, the event is assigned the cluster number zero [11] and no further pattern recognition occurs. If an inconsistent edge is found, the algorithm proceeds to examine the two resulting subtrees for clusters. For each subtree the algorithm starts with the node at which the primal subdivision was made and investigates whether the neighboring nodes are bridging or not. If they are nonbridging they join with the primary node to form the beginnings of a cluster. If they are bridging and the ratio of the edge distance to the median is greater than the parameter R_2 ($R_2 \simeq R_1$) the neighbor is split off and becomes the possible generator of a new cluster. This procedure is continued until every node has been interrogated. To make this procedure more lucid let us consider the MST in Fig. 1(b). The bridging edges are cd, de, ef, ek, km and mn of which the longest is de. The distance de is much larger than the median (lm) so we "break" the tree by removing the edge de. Now according to our prescription, we start at node d and investigate its neighbors. Since cd is a bridging edge we must ask the question, "Do we want to 'break' the tree by removing the edge cd?" Since the distance cd is comparable to the median, the answer is "No," and c and d form the beginnings of a cluster. The nodes a and b are nonbridging neighbors of c and are therefore added to the cluster spanned by c and d. All the nodes in the first subtree have been interrogated, so the same procedure is applied to the subtree containing the node e. The nodes f, g, h and j are all associated with e and form the second cluster. The distance ek is much larger than the median (i.e., is inconsistent) and so the edge ek is removed and the search for a third cluster, generated by the node k, is begun. The edges km, and mn are comparable in length to the median and ml and np are nonbridging neighbors, so the nodes k, l, m, n and p form a cluster. The three clusters generated by the algorithm are shown in Fig. 1(c).

We now leave the example and return to the description of the algorithm. Each cluster is assigned a momentum which is the vector sum of the momenta of the particles which span the cluster. All clusters are tested against a set of minimal requirements and those which fail are deleted from the cluster list. Clusters must have a momentum greater than P_{min} and contain at least T_{min} particles. When a cluster(s) is deleted from the cluster list, the particles which spanned it are added to the cluster to which they are "closest". The criteria for "closeness" is simply the angle between the particle direction and the cluster axis as defined by its momentum. As each new particle is added to the cluster, the cluster momentum components are suitably augmented. For the results presented below the order in which these particles are added is of no consequence.

R_1, R_2, T_{min} and P_{min} are parameters which can be chosen by the user. The values used for the results presented in the next section are typically $R_1 = 2$, $R_2 = 1.5$, $T_{min} = 2$ and $P_{min} = 3.0$ GeV/c. It should be emphasized that in the cluster search no absolute distance cuts are made; instead ratios between distances, local to each event, are used. This has the advantage of avoiding an absolute distance scale, which corresponds roughly to a scale in angle.

Results

The MST cluster finding algorithm was developed with events generated by the SLAC/LBL Mark II Monte Carlo program. The results for three different models are presented in this section. The three models are (i) pure phase space (ii) the decay of a heavy quarkonium state into three gluons (ggg) and (iii) a model which generates quark-antiquark jets ($q\bar{q}$) as well as events with gluon bremsstrahlung ($q\bar{q}g$ and $q\bar{q}gg$). The center-of-mass energy for each set of events is chosen to be 30 GeV and in all cases radiative effects have been included. The phase space Monte Carlo events have full three dimensional symmetry. All generated particles are pions with a mean multiplicity (charged + neutral pions) of nineteen. The three gluon decay of heavy quarkonium is modeled using gluon energy and angle distributions as calculated by Koller et al. [12]. There are two reasons for including this example. Firstly, it demonstrates how the MST cluster method could be applied to the lowest lying $t\bar{t}$ bound state.

Secondly, it is the example used by Wu et al. [4] in their discussion of tri-jettiness and a direct comparison between the two methods is then possible.

The DESY QCD Monte Carlo of Ali et al., is used to generate the $q\bar{q}$, $q\bar{q}g$ and $q\bar{q}gg$ events. This Monte Carlo event generator is fully described in Ref. [13]. The parameters λ, α_s, σ_q and the number of flavors are chosen to be 350 MeV/c, 0.184, 300 MeV/c and 5 for the 30 GeV events. The criterion for generating a two jet event is that the thrust be greater than 0.95. The ratio of two, three and four jet events is 0.65:0.30:0.05. This QCD model has been chosen because it is widely used by the PETRA groups and, with these parameters, yields an excellent fit to the PETRA data in the 30 GeV center-of-mass energy region.

Figure 2 shows the cluster frequency distribution for the three models. The distribution for the $q\bar{q}$ events is shown separately from the $q\bar{q}g$ + $q\bar{q}gg$ events for reasons which will become apparent later. First let us consider using the MST cluster method to study the ggg events. The frequency distribution obtained by the MST cluster analysis for phase space (2(a)) is markedly different than that of the ggg events (2(b)) and, given real data, there would be no problem distinguishing between these two models. The algorithm assigns 52 % of the ggg events to the three jet topology and the remainder to the two jet topology. The events which are found to have two clusters are typically those in which the angle between two of the gluons is small ($\lesssim 40^{\circ}$). In this case the two jets overlap and reliable separation is difficult. The question arises as to how well the cluster parameters of the three cluster events agree with the generated gluon parameters. Each cluster is correlated with one of the produced gluons using the angle between the cluster momentum and the gluon momenta. After the assignment of each of the three clusters has been made to the three gluons, the difference between the generated gluon direction and the found cluster direction ($\delta\phi$) is calculated. This distribution is shown in Fig. 3. Similarly, Fig. 4 contrasts the reconstructed cluster energy and the generated gluon energy. Figs. 3 and 4 demonstrate that the algorithm is assigning particles to jets in good agreement with the hadronization process. One can thus conclude that the MST cluster algorithm would have good efficiency for studying the three gluon final state of a $q\bar{q}$ bound state in the energy region of 30 GeV. The gluon hadronic fragments are well reproduced by the algorithm, which would allow for a meaningful study of the fragmentation process.

As a check on the correctness of the n-jet assignment, Lanius [6] has generalized the notion of triplicity defining the n-ticity as:

$$\text{n-ticity} = \sum_{j=1}^{n} \left| \sum_{i\varepsilon j} \vec{P}_i \right| \Big/ \sum_{k=1}^{N} \left| \vec{P}_k \right| \quad . \tag{3}$$

If the cluster assignment, n, for an event is correct, the n-ticity should be large ($\gtrsim 0.8$). This measure provides information in addition to the cluster frequency distribution for distinguishing between different physics models. As an example, Fig. 5 contrasts the 3-plicity distributions for the three cluster phase space events and the three cluster ggg events. A cut at 3-plicity of 0.86 loses virtually no ggg events, but eliminates 92% of the phase space events. One obtains a similar phase space rejection for the two jet events using 2-plicity.

We now turn our attention to the more difficult problem of demonstrating that the three jet topology exists in 30 GeV hadronic events produced in e^+e^- collisions. The main issue pertaining to QCD is to distinguish between a model which has only $q\bar{q}$ events as opposed to one which has in addition gluon bremsstrahlung and hence $q\bar{q}g$ + $q\bar{q}gg$ events. It is for this reason that the components $q\bar{q}$ and $q\bar{q}g$ + $q\bar{q}gg$ have been separated in Fig. 2. (In order to get the total physics picture, one must sum the distributions in Fig. 2(c) and 2(d) weighted in the ratio of 0.65 to 0.35.) As in the previous example, we note that pure phase space cannot be confused with either the $q\bar{q}$ or the $q\bar{q}g$ + $q\bar{q}gg$ processes. From Fig. 2(c) we see that the MST cluster algorithm has an extremely high efficiency (95%) for classifying $q\bar{q}$ events as two jet events. It has been verified that the direction and energy of the clusters agree extremely well with the generated quantities. Only 1.5% of the $q\bar{q}$ events are classified as three cluster events. From Fig. 2(d), we find that 27% of the $q\bar{q}g$ + $q\bar{q}gg$ events are classified as three cluster events. The majority of the gluons are "soft" and it is very difficult to make a meaningful separation between the gluon and the quark which produced it. Hence most of the $q\bar{q}g$ + $q\bar{q}gg$ events are classified as two cluster events in which the one cluster is considerably broader than the other. However for gluon energies above 6 GeV, the algorithm has a gluon tagging efficiency of ~88%. As in the ggg example, we can test the accuracy of the algorithm for the three cluster events by comparing the properties of the found clusters and the generated jets. Figs. 6(a) − 6(c) show the agreement between the found cluster direction and the produced jet direction. (The correlation between the clusters and the generated jets is established in the same way as for the ggg example.) Here the distributions for the highest, intermediate and lowest energy jets are shown separately. The definition of the jet direction is clearly more difficult as the jet energy decreases. However, even for the lowest energy jet (usually the gluon) the jet direction is satisfactorily defined with 90% of the clusters being within 20^{o} of the produced jet [14]. Figure 7 shows the correlation between the generated jet energy and the found cluster energy. One may conclude that the MST cluster algorithm is a viable method for the study of both two and three jet events produced in high energy e^+e^- collisions. In particular the three jet events arising from gluon bremsstrahlung can be studied with little contamination from $q\bar{q}$ events. The good agreement between the cluster and generated jet energies and angles allow for the reliable study of the jet fragmentation properties. The 3-plicity distribution for the three cluster events is almost identical to that for the ggg events and 3-plicity is therefore a good discriminator for different physics models. The same conclusion is true for the 2-plicity. It should be noted that this algorithm also finds four jet events which arise from the $q\bar{q}gg$ events.

In the examples given above a perfect detector was assumed and all the generated stable particles were used in the cluster search. As a more realistic example, the Mark II is used as a model for the typical cylindrical storage ring detector. The MST method is applied to the problem of gluon bremsstrahlung using the detected particles. A complete description of the Mark II detector can be found in Ref. [15]. In the Mark II detector charged particle tracking is reliably achieved over 75 % of 4π. Photons are detected in the liquid argon shower counters which cover 67 % of 4π but have relatively poor low energy photon efficiency. For this example the endcap shower counters are not used. The result of these detector features is to reduce the average number of particles used by the algorithm from thirty (for a perfect detector) to seventeen. One finds that 70 % of the $q\bar{q}$ events are classified as two cluster events and only 1 % as three cluster events. For the $q\bar{q}g$ + $q\bar{q}gg$ events 57 % are classified as two cluster events and 13 % as three cluster events. (In both cases, the remaining events are classified as zero and one cluster events.) From the point of view of the total hadronic physics picture the algorithm would classify 66 % of the hadronic events as two cluster events and 6 % as three cluster events. The agreement in direction between the three produced jets and the three found clusters is shown in Fig. 8. Naturally they are not as good as for the ideal detector, but quite satisfactory for studying the properties of three jet events. Hence both two and three jet physics is possible, the three jet topology being relatively free ($\lesssim 10\%$) of $q\bar{q}$ contamination.

Discussion

One of the goals in designing this jet finding algorithm was to keep it as general as possible so that it could be applied to many different physics topologies. The idea was not to tailor the algorithm to the needs of one particular physics application, thereby incorporating the physics of the specific problem, but rather to use the most general physical properties common to jets as they are manifested in different physics regimes. In large part this has been achieved. The properties exploited by the algorithm are the direction correlation of particles in a jet and to a lesser extent the effects of leading particles. In essence these are the two properties which define the notion of a jet and it therefore seems appropriate to incorporate them. Physics models of jet fragmentation may differ greatly but they all embody leading particle effects because these are kinematic rather than dynamic. The leading particle bias enters via the weighting function $w_j = P_j^{-1}$ which has the effect of clustering low momentum particles around the leading particles. This weighting function has the advantage of optimising the algorithm's CPU time because the MST's have fewer bridging edges than those generated using, for example, $w_j = 1$. However the choice of weight $w_j = P_j^{-1}$ is not crucial to the success of the MST algorithm and "softer" momentum weighting, like the extreme weight $w_j = 1$, yields results which are similar (but somewhat worse) in quality and efficiency to those for the weight $w_j = P_j^{-1}$.

Another possible source of bias arises from the fact that the pattern recognition has been optimised using specific Monte Carlo physics models. The conclusions presented in the previous section are not sensitive to changes as large as 50% in the parameters R_1 and R_2. T_{min} has been set to its minimum value of two. A consequence of the MST algorithm is that it cannot find single particle clusters. The results are sensitive, however, to the choice of the minimum cluster momentum (P_{min}). P_{min} has been chosen to be large which gives stable and reliable results at the cost of three jet efficiency. However lowering this cut implies studying jets which are in general close to other jets. This makes it difficult to define with confidence the energy and direction of the two overlapping jets. In this paper three cluster efficiency has always been sacrificed so as to keep the quality of the cluster parameters high and the physics conclusions as unambiguous as possible. In the case of gluon bremsstrahlung it is easy to double the efficiency for classifying $q\bar{q}g + q\bar{q}gg$ events as three clusters by lowering P_{min} to ~ 2 GeV/c. The price paid is that many more ($\sim 25\%$) three cluster events result from the $q\bar{q}$ process. For some tests of QCD this might be acceptable. The more stringent and less efficient route has been chosen in this paper so as to demonstrate how the MST algorithm could be used to distinguish between a pure $q\bar{q}$ production process and one which included in addition gluon bremsstrahlung.

It should be emphasized that the cluster number acts as a single event measure, like sphericity or thrust, by which to generate a distribution for comparison with physics models. The cluster number frequency (cf., Fig. 2) can be used to distinguish between different physics models in the same sense that sphericity (thrust) was first used at SPEAR (PETRA) to infer the existence of quark (gluon) jets. The n-ticity provides a further test that the cluster number found by the algorithm is consistent with the kinematics for producing n such clusters. Random cluster assignments, such as those which arise in phase space events, have low n-ticity characteristic of the fact that the underlying particles are not produced in jets. The algorithm classifies clusters as a set of associated particles plus an axis which allows for the study of the individual jet dynamics as well as the dynamics of the n jet axes.

The MST algorithm is computationally very fast. The software package is written in Fortran IV and event analysis rates for an IBM 370/168 are 40 events per second for the ggg events and 50 events per second for the $q\bar{q} + q\bar{q}g + q\bar{q}gg$ events. The average number of particles used by the algorithm is thirty-six for the ggg events and thirty for the $q\bar{q} + q\bar{q}g + q\bar{q}gg$ events.

Conclusions

A cluster finding algorithm for use in high energy physics analysis has been developed which allows for the reconstruction and study of particle jets. The algorithm has broad applicability because it is not designed to handle specific jet geometries. It has been demonstrated that the cluster algorithm can be successfully applied to the problem of two and three jet events produced in e^+e^- collisions at 30 GeV. The hadronic jets are reliably reproduced by the found clusters which thereby allow for a meaningful study of the jet production and fragmentation processes. The method also has the advantage of being computationally very fast.

Acknowledgements

The author wishes to thank and acknowledge Jerry Friedman of the SLAC Computational Group for suggesting the use of the minimal spanning tree and for many useful discussions. Also thanks go to the members of the Mark II Physics Group for suggestions and criticisms. In particular the author wishes to thank Andreas Bäcker for sharing his insight and experience with the three jet topology and Vera Lüth for coding the three gluon Monte Carlo model especially for use in this paper.

References

1. C. Grupen, talk presented at the XXth International Conference on High Energy Physics, University of Wisconsin, Madison, Wisconsin, July 1980; SI-80-11 (1980).
2. R. Brandelik et al., Phys. Lett. 86B, 243 (1979); and Ch. Berger et al., Phys. Lett. 86B, 418 (1979).
3. S. Brandt and H. Dahmen, Particles and Fields (Z. Phys. C) 1, 61 (1979).
4. S. L. Wu and G. Zobernig, Particles and Fields (Z. Phys. C) 2, 207 (1979).
5. There are analysis techniques (e.g., G. Fox and S. Wolfram, Phys. Rev. Lett. 41, 1581 (1978)) which involve the use of "event shape" parameters and which have broad applicability. We will not discuss these methods here because we are interested in going beyond the "shape" parameter analysis to the full reconstruction of the hadronic jets.
6. There are at least two other cluster finding efforts in the process of development. They are K. Lanius, DESY 80/36 (1980), and V. Hepp, talk presented at the XXth International Conference on High Energy Physics, Madison, Wisconsin, July 1980. The MST method was conceived and developed independently and without knowledge of these two efforts.
7. C. T. Zahn, IEEE Trans. Computers, Vol. C-20, 1, 68 (1971)
8. The terms "bridging" and "nonbridging" are not used by statisticians, but were coined by the author for this paper.
9. The algorithm for finding the MST was originally presented by R. C. Prim, Bell Systems Tech. Jour. 36, 1389 (1957). The code needed to generate the MST is surprisingly simple and can be found per V. K. M. Whitney, Algorithm 422, COMM ACM 15, 273 (1972).
10. If N is even, the quantity M/2 is truncated for the purpose of finding the median.
11. It could be argued that such events should be assigned a cluster number of one. However since there is no clustering of particles in these events, they are defined to have cluster number zero.

12. K. Koller, H. Krasemann and T. F. Walsh, Particles and Fields (Z. Phys. C) 1, 71 (1979).

13. A. Ali et al., Phys. Lett. 93B, 155 (1980).

14. The typical half angle for a hadronic jet is $\sim 20^{\circ}$, so defining the energy flow to less than 20° almost certainly means that the correct assignment has been made.

15. The two most comprehensive descriptions of the Mark II detector are: V. Lüth, Proceedings of the 1979 International Symposium on Lepton and Photon Interactions at High Energy, Batavia, Illinois, 78 (1979); and J. Dorfan, AIP Conference Proceedings #59, Particles and Fields #19, 159 (1979).

Figure Captions

Fig. 1 a-c. A set of space points (nodes) are shown in 1(a) which are used to demonstrate the construction of a minimal spanning tree (1(b)) and the partitioning of the minimal spanning tree into three clusters as indicated by the circles in 1(c). The partitioning occurs because the edges de and ek are long compared to the rest of the edge distances.

Fig. 2. A plot of the cluster frequency distribution obtained by the MST cluster algorithm for the different production mechanisms described in the text. The distributions are normalized to the total number of events (N_{ev}).

Fig. 3. Comparison of the Monte Carlo generated jet axis direction with the found cluster axis direction. The difference in these two directions ($\delta\phi$) is plotted in degrees for those ggg events which are classified as three cluster events.

Fig. 4. A plot of the generated gluon energy and the found cluster energy for those ggg events which are classified as three cluster events. The fractional energy resolution (σ) averaged along the diagonal of Fig. 4 is 10%.

Fig. 5. The 3-plicity is plotted for the three cluster phase space events and ggg events. The overlap is very small which permits substantial discrimination against the phase space events with little loss of signal (ggg events).

Fig. 6. This figure plots $\delta\phi$ for those $q\bar{q}g + q\bar{q}gg$ events which are classified as three cluster events. The jets are ordered by energy. For interpretation it should be remembered that a typical jet spans ± 20 degrees.

Fig. 7. A comparison of the generated constituent (quark or gluon) energy and the found cluster energy for the $q\bar{q}g + q\bar{q}gg$ events for those events which are classified as three cluster events. The fractional energy resolution (σ) averaged along the diagonal of Fig. 7 is 11%.

Fig. 8. This figure plots $\delta\phi$ for those $q\bar{q}g + q\bar{q}gg$ events which are classified as three jet events. For this application the MARK II was used as a typical cylindrical storage ring detector.

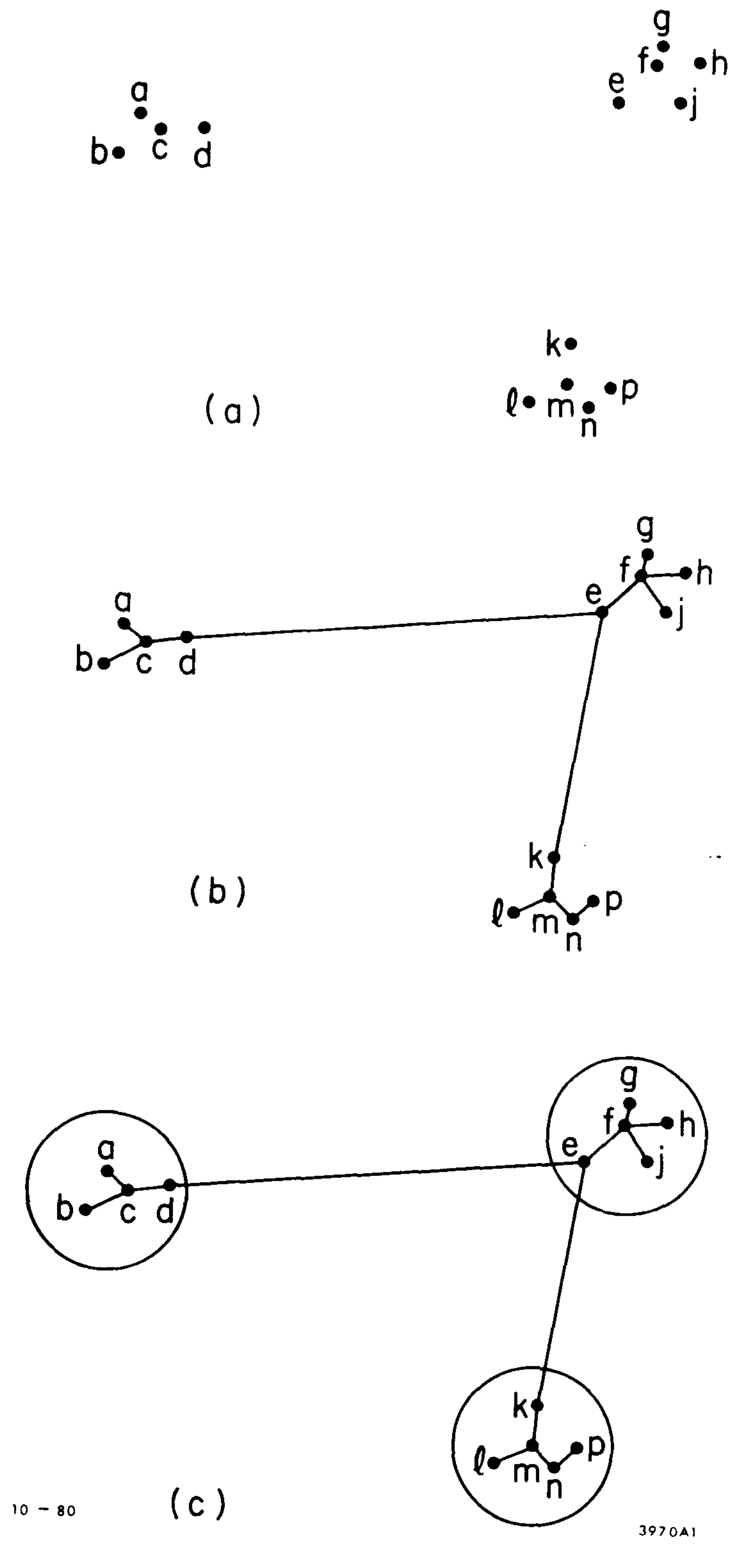


Fig. 1

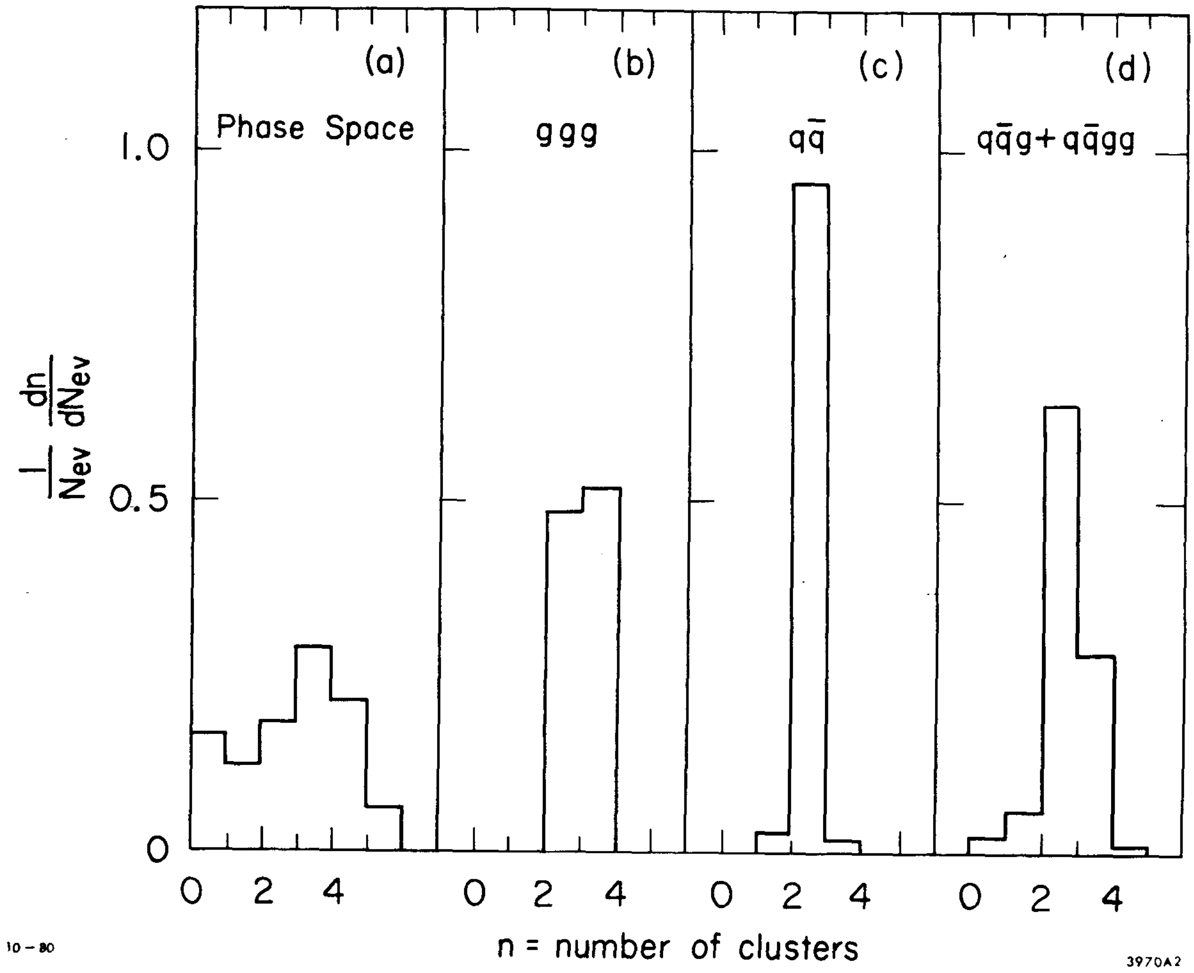

(a)
Phase Space
(b)
ggg
(c)
qq̄
(d)
qq̄g+qq̄gg
1.0
0.5
0
$\frac{1}{N_{ev}}\frac{dn}{dN_{ev}}$
0 2 4
0 2 4
0 2 4
0 2 4
n = number of clusters
10-80
3970A2

Fig. 2

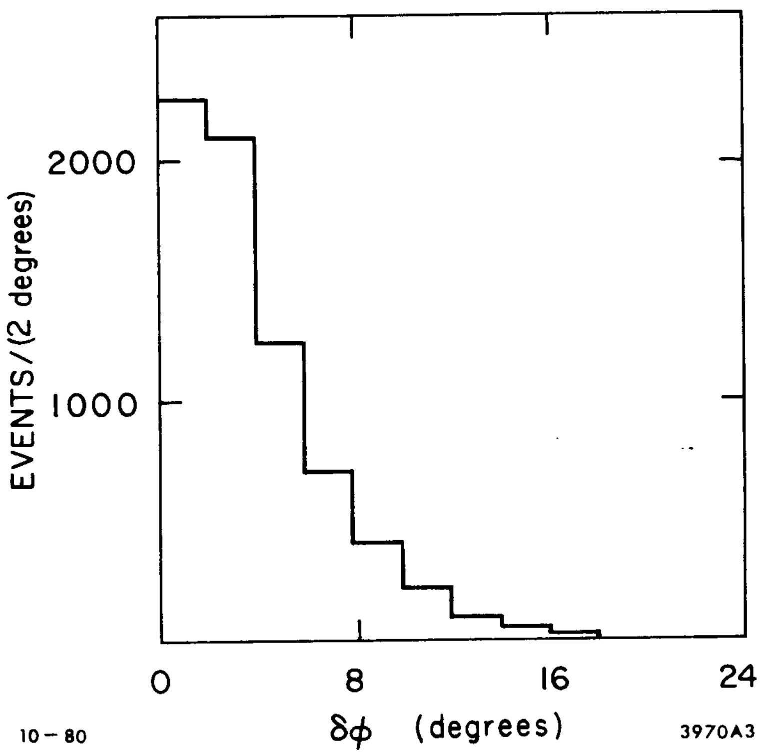


Fig. 3

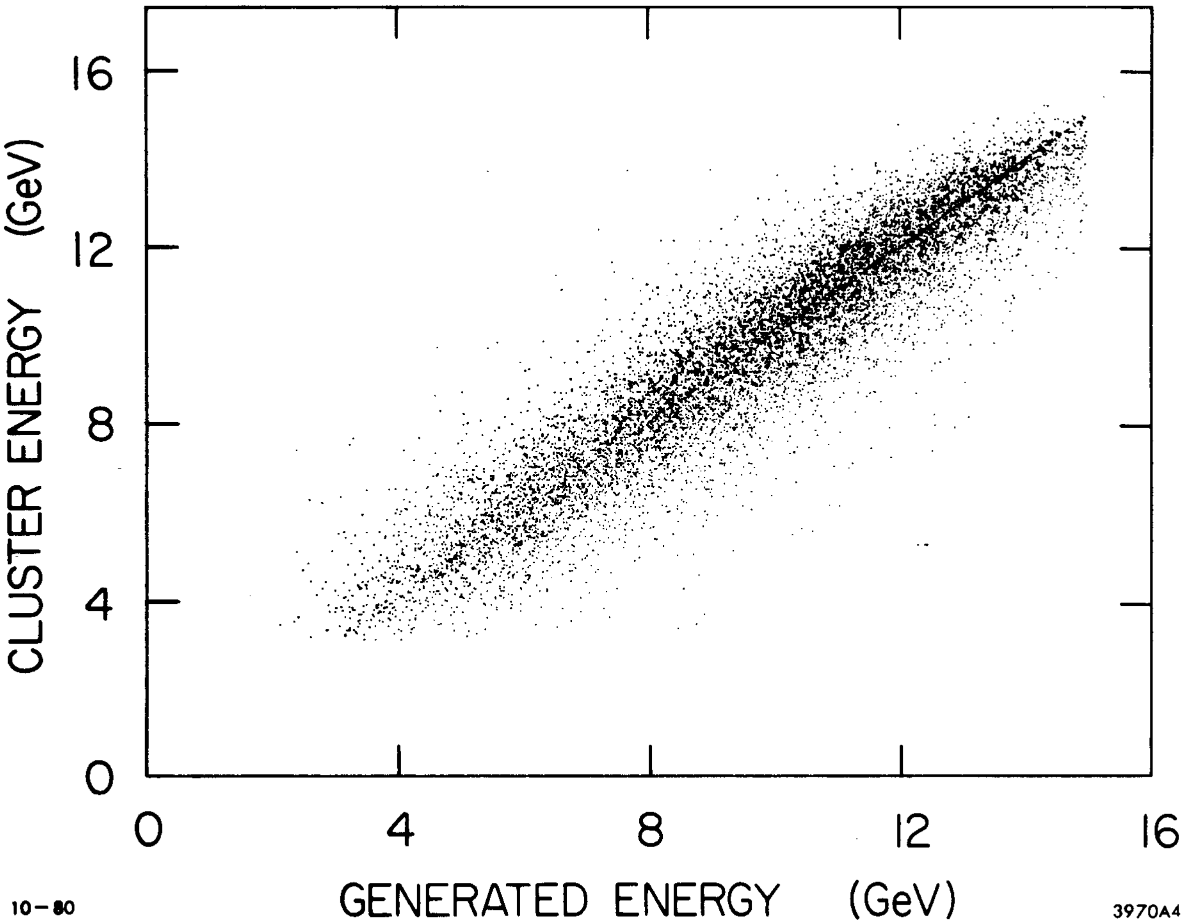
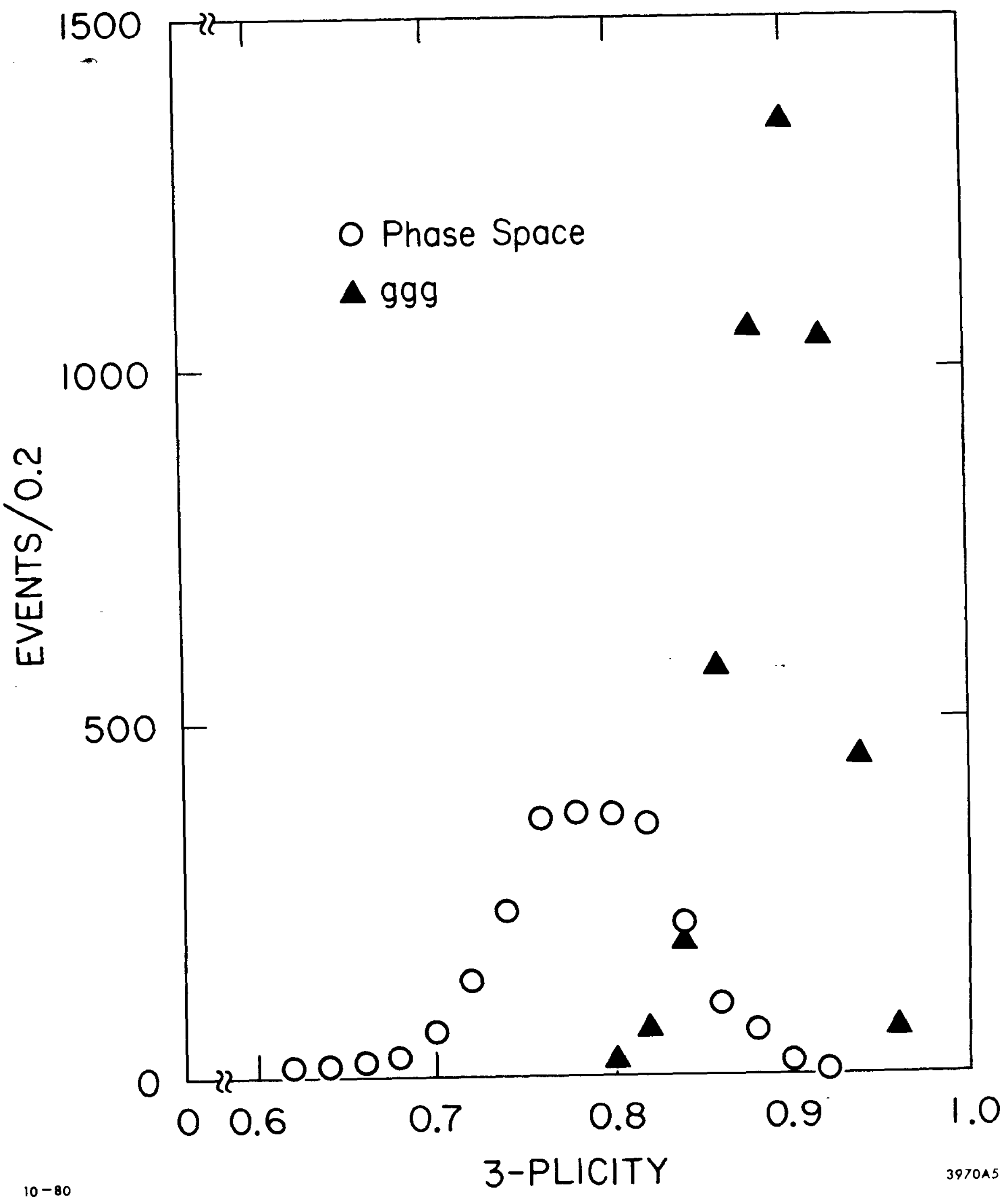
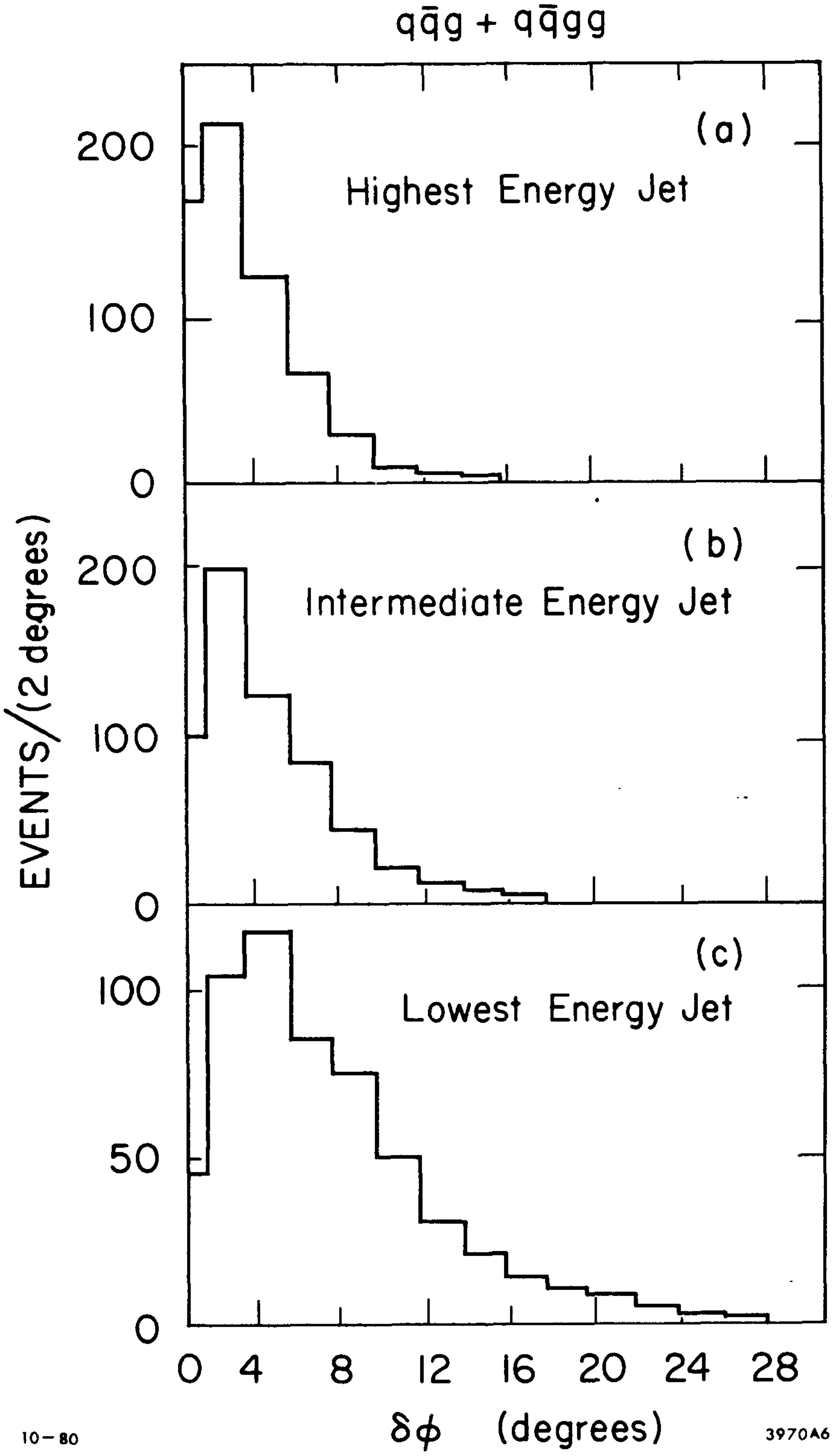

CLUSTER ENERGY (GeV)
16
12
8
4
0
0
4
8
12
16
GENERATED ENERGY (GeV)
10-80
3970A4

Fig. 4



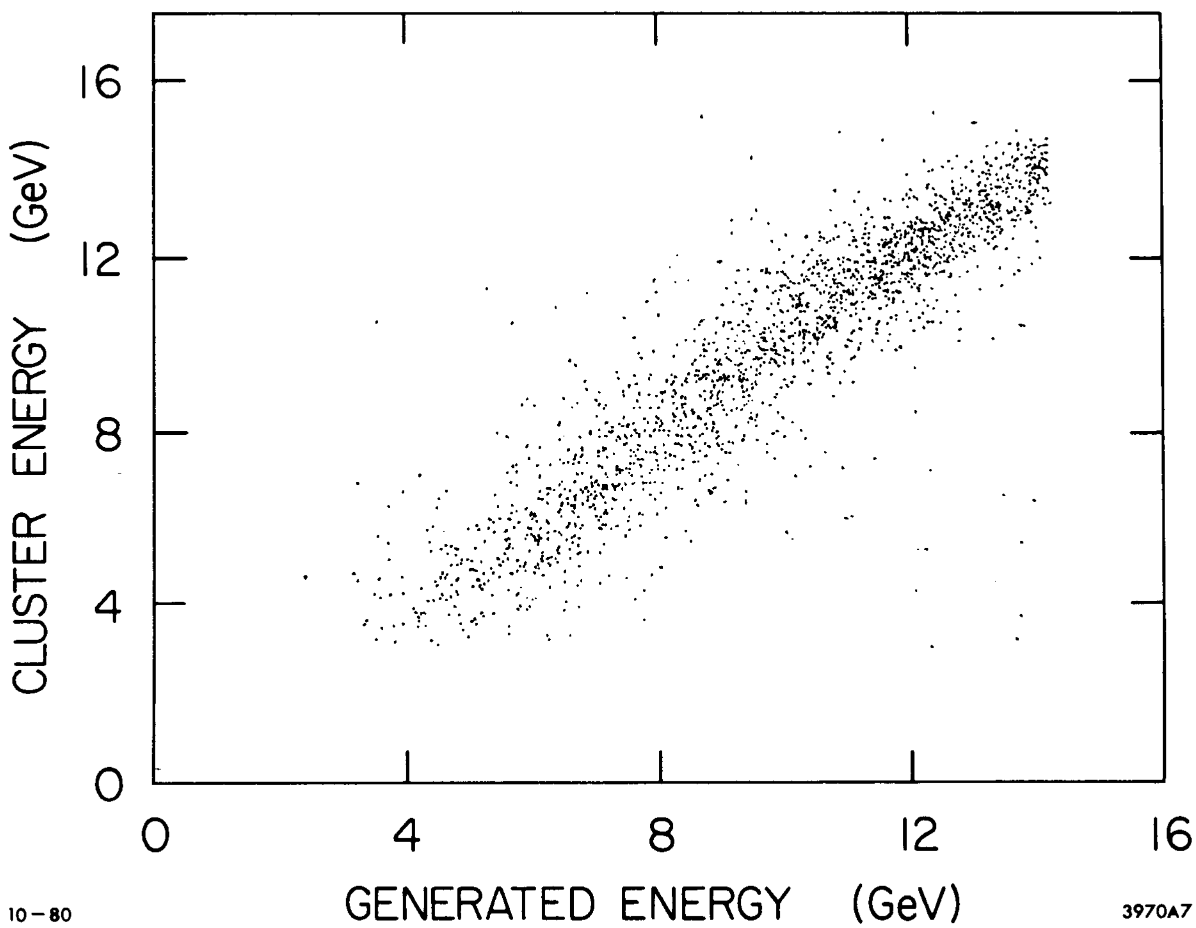
1500
1000
500
0
EVENTS/0.2
O Phase Space
▲ ggg
0 0.6
0.7
0.8
0.9
1.0
3-PLICITY
10-80
3970A5

Fig. 5



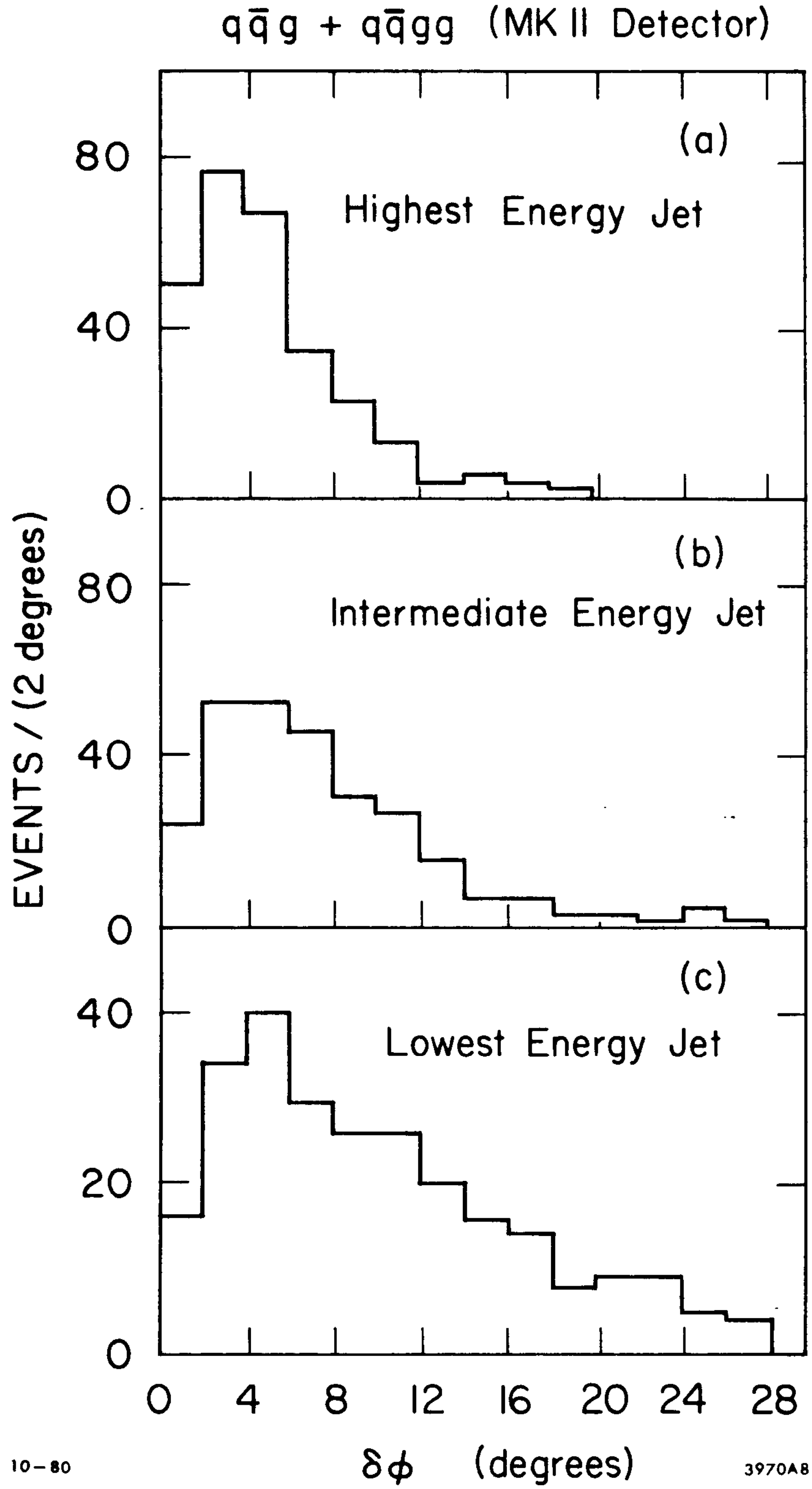
q$\bar{q}$g + q$\bar{q}$gg
(a)
Highest Energy Jet
(b)
Intermediate Energy Jet
(c)
Lowest Energy Jet
EVENTS/(2 degrees)
200
100
0
200
100
0
100
50
0
0 4 8 12 16 20 24 28
δφ (degrees)
10-80
3970A6

Fig. 6



CLUSTER ENERGY (GeV)
16
12
8
4
0
0
4
8
12
16
GENERATED ENERGY (GeV)
10-80
3970A7

Fig. 7



q$\bar{q}$g + q$\bar{q}$gg (MK II Detector)
(a)
Highest Energy Jet
(b)
Intermediate Energy Jet
(c)
Lowest Energy Jet
EVENTS / (2 degrees)
80
40
0
80
40
0
40
20
0
0 4 8 12 16 20 24 28
$\delta\phi$ (degrees)
10-80
3970A8

Fig. 8