PHYSICAL REVIEW D VOLUME 16, NUMBER 1 1 JULY 1977

# Search and classification of reaction mechanisms in multiparticle data

T. Ludlam*
*Physics Department, Yale University, New Haven, Connecticut 06520*

R. Slansky†
*Theoretical Division, Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87545*
(Received 21 September 1976)

The application of pattern-recognition techniques to the problem of separating reaction mechanisms in high-multiplicity collisions is explored. The "distance analysis," which we motivate and explain here, is an algorithm for finding the most densely populated regions of the multidimensional phase space. It provides a sensitive and model-independent means of identifying the individual events which make up each of the regions. It is then possible to use other techniques to interpret these. This procedure could also be useful outside particle physics. It is very simple and fast, and a detailed account of its implementation is given. As examples, we analyze the reactions $K^-p \to K^-\pi^+\pi^-p$ and $K^-p \to K^-\pi^+\pi^-p\pi^0$ at 12.6 GeV/$c$.

## I. INTRODUCTION

One goal of phenomenological analyses of high-energy hadron production is to partition the data into different reaction mechanisms, and to study each of these in isolation. It has proved possible to do this for low-multiplicity exclusive reactions with three or four particles in the final state. Here most of the information is carried by one or two dynamical variables, and the events can be classified from the projection onto these variables. For example, in the longitudinal-phase-space (LPS) analysis for four-body final states, two scaled longitudinal-momentum variables are measured for each event, and the resulting scatter plot shows a rich and readily interpretable structure.[1]

Unfortunately, higher-multiplicity final states are much harder both to analyze and to interpret. The five-body LPS is a three-dimensional projection, and very-high-statistics experiments are necessary to enable us to construct a three-dimensional density function from the data. Even where this has been done, it is difficult to distinguish clearly different reaction mechanisms.[2]

There are two possible explanations of this difficulty. The lack of structure in the three-dimensional projection could be a reflection of little structure in the full ten-dimensional five-body phase space. There might be so much overlap among the "different" mechanisms that there is only one fairly uniformly populated region in the phase space. The sudden failure of the five-body analysis to separate mechanisms, compared to the four-body results, might encourage this viewpoint. Of course, it is also possible that the one-, two-, and three-dimensional projections that have been studied are averages that obscure the interesting structure.

The evidence indicates that the latter is the case. The fluctuation analysis shows that models in which only one region of phase space is uniformly populated are inconsistent with the $n \geq 5$ data.[3] What is more significant is that this structure becomes even more marked as $n$ increases (at fixed energy), even though the usual plots of the data show little structure. (Perhaps we should emphasize that this evidence for separated mechanisms also occurs in reactions where leading particles are not manifestly present.) It appears necessary to be able to analyze these higher-multiplicity final states in higher-dimensional spaces, if the reaction mechanisms are to be identified. Specifically, programs to seek out the one or two variables that actually separate different reaction mechanisms begin to fail at $n \geq 5$. Thus, even if different reaction mechanisms are actually separated in the full phase space, their projections onto these variables overlap. (A nice review of such techniques is contained in Ref. 4.) On the other hand, higher-dimensional density plots require prohibitive quantities of data. We now briefly review several approaches for overcoming these problems.

The basic tactic of the prism-plot analysis[5] is to compare the data with a very flexible model. One needs a fairly good idea of the reaction mechanisms and their relative probabilities. The data are then compared with Monte Carlo–generated model events by some point-set–comparison method. Among the advantages of this method is its insensitivity to the choice of variables. (Jacobian effects in the data will also occur in the model.) Powerful point-set–comparison and Monte Carlo programs exist, and make the analysis quite feasible.[4,6] The disadvantages include its model dependence, and the fact that it is somewhat clumsy to vary the model parameters.

A complementary approach rests on the assumption that events belonging to different reaction mechanisms will be localized in different regions of phase space. No *a priori* model of the reaction mechanisms is needed. The idea is either to look for regions with a high density of events, or to look for boundaries or valleys between regions of high event density in phase space. The valley-seeking technique has been explored by Böttcher *et al.*,[7] in their analysis of $\pi^+p \to p\pi^+\pi^+\pi^-$ at 8 GeV/$c$. The method discussed here uses other techniques of pattern-recognition theory to search for regions of high event density in the multidimensional phase space.[8-10]

In order to implement these methods, it is necessary to define the distance between events in the multidimensional phase space, at least for events that are "near" to one another. Quantum mechanics requires only the existence of a probability density defined on the phase space; however, any practical definition of local density for a finite data sample involves the notion of distance between nearby points.

Distance measures ought to exist, if scattering amplitudes are analytic functions of the Lorentz invariants formed from the four-momenta. Phase space is then a Riemann space, which is also a metric space. This is probably not enough information to define uniquely a distance measure, since the symmetry properties of the space are also likely to be involved. In particular, we conjecture that the geometrical properties of non-Abelian gauge theories of elementary-particle interactions provide a unique measure of distance between events. We consider this to be a fundamental problem worthy of further scrutiny. From the more phenomenological viewpoint that we must necessarily take here, the "best" measure depends on the probability density and the algorithm to seek clusters.

Practically, the problem of defining a distance measure is not as serious as it might first appear. For parameterizing density, a reasonably good measure is needed only for nearby pairs. All that is needed for widely separated events is the fact that they are, indeed, far from one another, so only a rough estimate of these distances is needed. This reduces the problem to one of finding a local distance measure. Here the results become rather insensitive to the precise form of the metric, as long as certain general requirements are satisfied. The measure must carry information on many of the independent degrees of freedom of the final state. This will include multibody information. It must also respect, at least to some degree, certain obvious symmetry properties, such as rotational and Lorentz invariance, since events related by symmetry transformations should correspond to the same point in phase space.

Section II contains a brief introduction to pattern-recognition techniques for finding clusters of points in a multidimensional space. The basic notion is the minimal spanning tree.[11] (Although we do not directly compute these objects, they form the conceptual basis of our analysis.) We then describe our variant of a pattern-recognition method that has proved suitable for analyzing multibody exclusive data. It is closely related to the shared-neighbor techniques discussed in Refs. 8–10. A detailed outline of a fast and efficient algorithm for doing the "distance analysis" is described in the Appendix.

The problem of the distance measure is discussed in Sec. III, along with a report on our progress.

Section IV includes, as an example, a study of two final states observed in 12.6-Gev/$c$ $K^-p$ collisions: $K^-p \to K^-\pi^+\pi^-p$ and $K^-p \to K^-\pi^+\pi^-p\pi^0$. In the four-body final state, we reproduce well-known results: Beam excitation is about twice target excitation. We also resolve several mechanisms within each of these. The five-body final state is more complicated, although there is a vestige of the beam and target excitation.

## II. THE DISTANCE ANALYSIS

The main objective of this section is to give a brief conceptual account of the basic methods for recognizing and separating clusters of points in a multidimensional space.[8-10] One approach utilizes the minimal spanning tree.[11] Although the distance analysis is based on this idea, it does not require their explicit calculation. In this section we define the minimal spanning tree, describe a naive method for finding clusters, and then discuss improvements which render the method applicable to hadron data. This should provide a rationale for the distance analysis which we outline. A detailed description that is suitable for writing a computer program is given in the Appendix.

Minimal spanning trees (simply called "trees" below) are constructed as follows: Consider the set of $N$ points in a multidimensional space for which the distance between each pair of points is defined. We call the $\frac{1}{2}N(N-1)$ lines along which these distances are measured branches. The tree is the unique set of $N-1$ branches that connect all points together and for which the sum of the branch lengths is a minimum. It can be constructed by starting with any point $e_1$, and finding the point $e_2$ which is closest to $e_1$. Next find the point which is closest to $e_1$ or $e_2$, and add it to the tree

by the shorter branch. Continue the process until all $N$ points are added to the tree.

A simple (but naive) proposal for finding clusters of event points is to cut all branches of the tree that are longer than a threshold distance $r$. This will break up the tree into disconnected subtrees (clusters) whose branches are all shorter than $r$. For a proper choice of $r$, these subtrees might be identified with different reaction mechanisms. (See Ref. 11 for more details.)

In practice this procedure is quite unstable unless the clusters of points are widely separated. A small increase in $r$ triggers a transition from many small subtrees (some due to statistical fluctuations) to one large tree encompassing several clusters of points. (Of course a few small subtrees will be left over.) In hadron data, background events and small overlaps of different clusters easily provide series of branches by which different regions of high density "chained" together. One basic problem with this method is that is does not focus clearly enough on regions of high density.

Clearly, our goal is separate regions of high density. This can be done directly, or by seeking out the valleys (regions of low density) between the regions of high density. The latter proposal is explored in Ref. 7; we now discuss our technique for directly searching out regions of high density.

The distance analysis (and also Wishart's hierarchical mode analysis[8] which motivated us) parameterizes the density of points (or hadronic events) by the distance to the $k$th nearest neighbor, where $k$ is preassigned.[9,10] For an $N$ of several hundred to several thousand events, $k$ of five to ten provides a fairly stable definition of the clusters found.[12] The basic idea is then to build trees out of the events that are the most dense (events that have the shortest distances to the $k$th nearest neighbors). Background and overlap events in regions of low density are omitted from the trees, and the chaining problem is avoided, but at the cost of not necessarily classifying every event in the sample.

The distance analysis is a method of constructing trees, beginning with the most dense events and proceeding to the least dense. The main problem is one of organization: determining what fraction of the most dense events should be included, and what value of $r$ (the maximum branch length of the subtrees) should be taken. Let us first introduce some terminology. The "halo" of event $i$ is the set of its $k$ nearest neighbors. The "radius of the halo of $i$," $R_i$, is the distance to $i$'s $k$th nearest neighbor. The "$i$th event to become dense" is the event with the $i$th smallest value of $R_i$.

Roughly, the distance analysis proceeds in the following stepwise fashion: At the $i$th stage, the $i$th event becomes dense by taking $r = R_i$. Then construct the tree from the $i$ dense events, and cut the branches with $r > R_i$. The subtrees resulting from this division then give the cluster classification at the $i$th stage of the analysis.

Again, the question arises, at what stage (choice of $r$) should the analysis be stopped. We do not have a precise answer, and the decision depends to some degree on the judgment of the analyst. (Examples are given in Sec. IV.) For small values of $i$, many clusters will emerge, of which some are due to statistical fluctuations while larger clusters are often physically significant. For intermediate values of $i$, cluster formation will stabilize and increasing $i$ will simply add new events to subtrees which already exist. A long plateau of stability is a sign of significant clustering. Finally, the clusters will collapse due to chaining as $i$ is increased, although very widely separated clusters may not collapse for $r = R_N$.

In carrying out the $i$th stage, it is not necessary to construct trees explicitly, and some may find it confusing to contemplate trees while first studying our procedure. At the beginning of any given stage $i$, each event falls into one of three categories. (1) The event is unclassified. This means that it is not dense and has not appeared in the halo of any event more dense than $i$. (2) The event is a boundary point and has therefore appeared in the halo of at least one event more dense than $i$. (3) The event is dense. In practice it is easy to assign boundary points to the clusters. In the $i$th stage, the only changes of classification will be to the events in the halo of $i$. Unclassified events become boundary events; if there are dense events from different clusters, the clusters are merged together, and the boundary events are reassigned accordingly. This procedure has the advantage that at each stage, it focuses on the region of only one density. Thus, if two dense events in another region of phase space belong to different clusters and are within $R_i$, the clusters are not merged until they are both included in the halo of some event. That is, two sharply defined but close together clusters are not merged until the averaged density between them is large enough. A second obvious advantage is that, at any stage of the analysis, only $k$ events and not the whole event sample are examined.

In summary, the events in the most densely populated regions of phase space are classified first, and those events that will chain together the clusters are left until last. In studying the results, we have found it most informative to examine the classification just before clusters of some minimum size are about to be merged. It is as useful

to examine clusters formed early in the analysis as it is later on. By setting this minimum size large enough, or by looking for regions of stability, it is often clear where the classification is most significant.

A detailed outline of this algorithm can be found in the Appendix.

## III. DISTANCE MEASURE

The essential physics problem encountered in this kind of analysis is finding a distance measure between pairs of events in the multibody phase space. We do not know whether a unique solution can be found without a rather complete gauge theory of hadron production. Without a complete theory, we must take a phenomenological attitude toward finding a measure. The "best" measure depends on the probability density, the sample size, and the clustering algorithm. But it is also clear from Sec. II that it is not necessary to have a complete understanding of the measure. The pattern-recognition analysis is basically an ordering procedure in which events in densely populated regions are identified along with their nearby neighbors. Events that are far apart are not in each other's halos, and only a rough estimate is needed to establish this fact. The more important information is the distance between nearby points, since the order in which the events become dense and the content of the halos are the crucial inputs into the analysis.

The measure must carry information about most of the degrees of freedom, since we have found that one- and two-dimensional projections of the data do not separate reaction mechanisms in high-multiplicity final states. (For a given two-dimensional projection the pattern-recognition techniques will not be more effective than a visual examination for separating reaction mechanisms.) Each event is characterized both by the behavior of individual particles and by many-body coordinates.

In this paper we have taken a more phenomenological approach to the problem. We first discuss our choice of variables, and then give a solution to the problem of finding the distance between two four-vectors.

The flow of different sets of quantum numbers could be determined by the current structure of a hadronic theory. As a crude representation of such currents, we have used center-of-mass velocities of certain multiparticle sets for studying several 12.6-GeV/$c$ $K^-p$ reactions. Our choice was also guided by a study of the velocity plots. In keeping with the spirit of principal-components analysis techniques,[4] we have chosen velocities whose distributions are broad, and which seem to be fairly independent of one another. These variables presumably carry the most information. The velocities are taken to be coordinates in a Euclidean space.

For the reaction $K^-p \to K^-\pi^+\pi^-p$, we used four longitudinal velocities: $\beta(K^-\pi^-)$, $\beta(\pi^+\pi^-)$, $\beta(\pi^-p)$, and $\beta(\pi^+\pi^-p)$. We should emphasize that we have experimented with many other sets of variables and have found that, as long as the variables carry enough multiparticle information, the final results as reported in Sec. IV are qualitatively similar. (This comment also applies to the other reactions we have studied.) For the reaction $K^-p \to K^-\pi^+\pi^-p\pi^0$, we used five longitudinal velocities: $\beta(\pi^+\pi^-)$, $\beta(\pi^-p)$, $\beta(\pi^+\pi^-\pi^0)$, $\beta(\pi^+p\pi^0)$, and $\beta(K^-\pi^-\pi^0)$, all in the reaction center of mass. Thus, each event is a point in a five-dimensional space, and the distance is just the Euclidean distance.

Although satisfactory for phenomenological purposes, these measures lack rigorous theoretical motivation. (Covariance is not a problem since these quantities can be rewritten in manifestly covariant form.) Analyticity of scattering amplitudes does give some information about a distance measure. We cite an explicit example: The "three-" vectors for the two space, one time dimension Lorentz group can be projected onto the complex plane (Poincaré plane) where a natural distance measure can be defined.[13] The invariant distance between two three-vectors $p^{(i)}$ and $p^{(j)}$ is given by

$$d_{ij}^2 = p^{(i)} \cdot p^{(j)} - m^{(i)}m^{(j)}, \tag{1}$$

where

$$\begin{aligned} m^{(i)} &= (p^{(i)} \cdot p^{(i)})^{1/2}, \\ p^{(i)} \cdot p^{(j)} &= p_0^{(i)}p_0^{(j)} - p_1^{(i)}p_1^{(j)} - p_2^{(i)}p_2^{(j)}. \end{aligned} \tag{2}$$

This distance has an obvious generalization to one time, three space dimensions.

We have used measures based on this distance with success. The measure must satisfy several invariance properties, including invariance under the rotation of just one of the events about the beam axis. Of course, multibody momenta must be included in the component pieces of $d_{ij}^2$.

In summary, we have not derived from first principles a complete distance measure. This is not a serious problem in practice, since there are adequate phenomenological measures. In our pragmatic view, we have selected sets of coordinates that reflect multibody effects. The results are fairly independent of choice. However, in searching for detail it is important to realize that some measures may focus on certain regions of phase space better than others. Thus, it may be useful to analyse a set of data with several different measures.

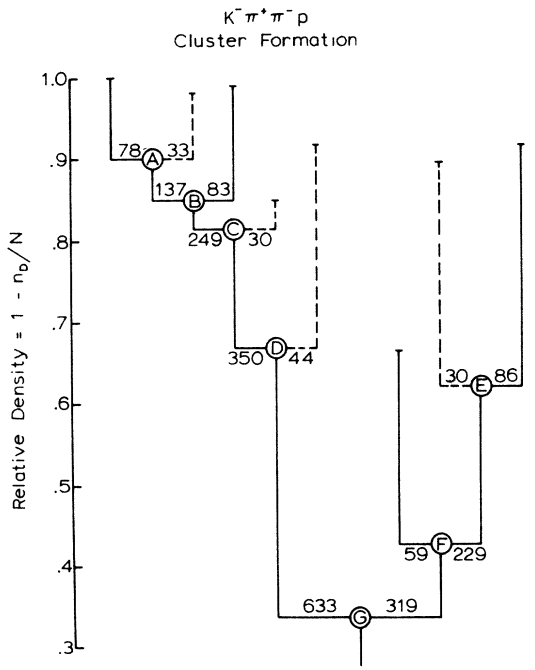


FIG. 1. Schematic of cluster formation for reaction 1. As described in the text, each branch of the treelike structure represents the development of a cluster. Classification proceeds from the top (most dense events) downward. The vertical scale measures the relative density of the event population being sampled at any point in the analysis: $n_D$ = number of dense events; $N$ = total number of events in the sample. For the sake of clarity, smaller clusters (less than 50 events) are represented by broken lines.

## IV. AN APPLICATION TO DATA

In this section we examine the effectiveness of the distance analysis on some real data by applying it to two exclusive channels from the Yale 12.6-GeV/$c$ $K^-p$ experiment. We discuss the results for the reactions

$$K^-p \to K^-\pi^+\pi^-p \quad \text{(reaction 1)},$$

$$K^-p \to K^-\pi^+\pi^-\pi^0p \quad \text{(reaction 2)}.$$

The first of these is a fairly well-understood channel in which the gross structure is dominated by strong-clustering effects (diffractive excitation of beam and target particles). Reaction 2 does not yield so simple an interpretation when confronted with conventional analysis techniques, although some diffractive component is clearly present,[14] and there is indirect evidence to indicate that similar clustering effects characterize a major portion of the cross section.[3]

The data were obtained from an exposure of the Brookhaven National Laboratory 80-in. bubble chamber to a beam of rf-separated $K^-$ mesons. We analyze 1291 examples of reaction 1, and 880 of reaction 2. The cross sections for reactions 1 and 2 are, respectively, $0.69 \pm 0.04$ mb and $0.40 \pm 0.04$ mb. All events were measured with manually operated Franckenstein engines, and the final samples isolated with the kinematic fitting program YACK. These yielded very good accuracy in the measurement of momenta and angles, and very few background events. A detailed description of the data is given in Ref. 14. High-quality data are needed for these kinds of precise analysis.

As discussed previously, an essential ingredient in the analysis is the construction of a multidimensional coordinate space in which to measure event-to-event distances. The theoretical problem of deriving a measure of distance from fundamental theory has not been solved. For the present analysis, we follow the more pragmatic view advocated in Sec. III. The multiparticle motion is characterized by the center-of-mass longitudinal velocities,

$$\beta_i = \frac{p_{L,i}}{E_i}, \tag{3}$$

where $p_{L,i}$ and $E_i$ are, respectively, the c.m. longitudinal momentum and energy for a particular combination of final-state particles labeled by subscript $i$. The distance $d_{ab}$ from event $a$ to event $b$ is assumed to be the Euclidean distance between the two event points in a space in which each coordinate axis measures the velocity of one of the chosen combinations:

$$d_{ab}{}^2 = \sum_i (\beta_i^a - \beta_i^b)^2 . \tag{4}$$

When analyzing a given sample of $n$-body data, we select approximately $n$ combinations of the final-state particles whose distributions in the chosen variable are well spread out, and also appear to carry significant dynamical information. For reasons discussed in Sec. III, this choice is adequate. Moreover, the method appears to yield more of the essential structure of these event samples than the conventional methods of analysis.

### Reaction 1: $K^-p \to K^-\pi^+\pi^-p$

The sample of 1291 events corresponding to reaction 1 was analyzed in a four-dimensional velocity space [Eqs. (3) and (4)] using the c.m. velocity variables $\beta_{K^-\pi^-}$, $\beta_{\pi^+\pi^-}$, $\beta_{\pi^-p}$, and $\beta_{\pi^+\pi^-p}$. The clustering pattern obtained is illustrated in Fig. 1. Here the vertical scale measures the relative local density, with the highest density at the top, and each branch of the treelike structure corresponds to the development of a significant cluster. Recall (Sec. II) that the analysis proceeds by clas-

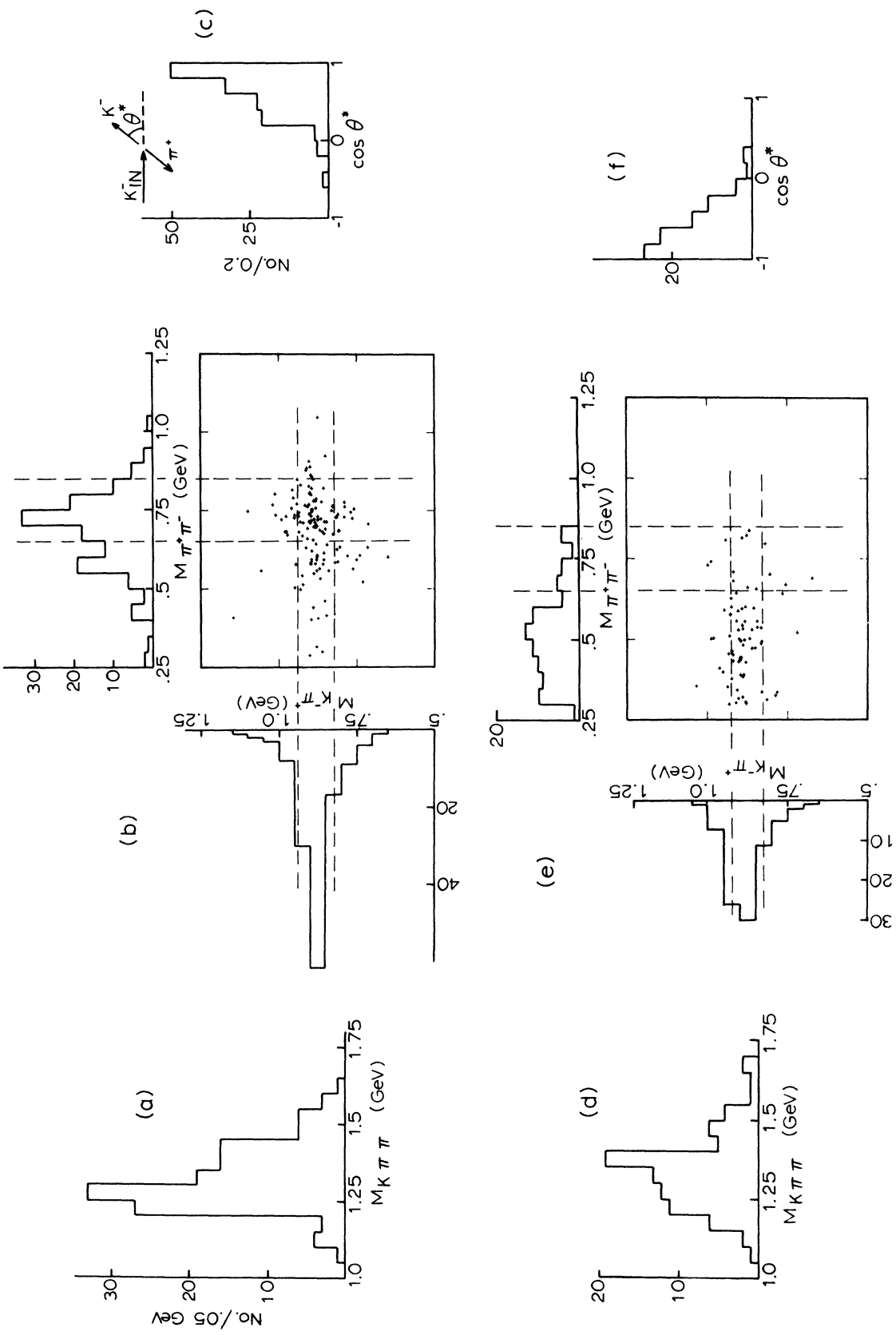


FIG. 2. Reaction 1. Structure of events in the two clusters which merge at point B in the diagram of Fig. 1. The 137-event cluster is shown in (a), (b), and (c); the 83-event cluster is shown in (d), (e), and (f). (a), (d): The $K^-\pi^+\pi^-$ invariant-mass distribution. (b), (e): The $K^-\pi^+$ invariant mass vs the $\pi^+\pi^-$ invariant mass. The dashed lines indicate the $K^*(890)$ and $\rho$ bands. (c), (f): The distribution in $\cos\theta^*$, where $\theta^*$ is the polar angle of the $K^-$ relative to the beam direction in the rest frame of the $K^-\pi^+$ system.

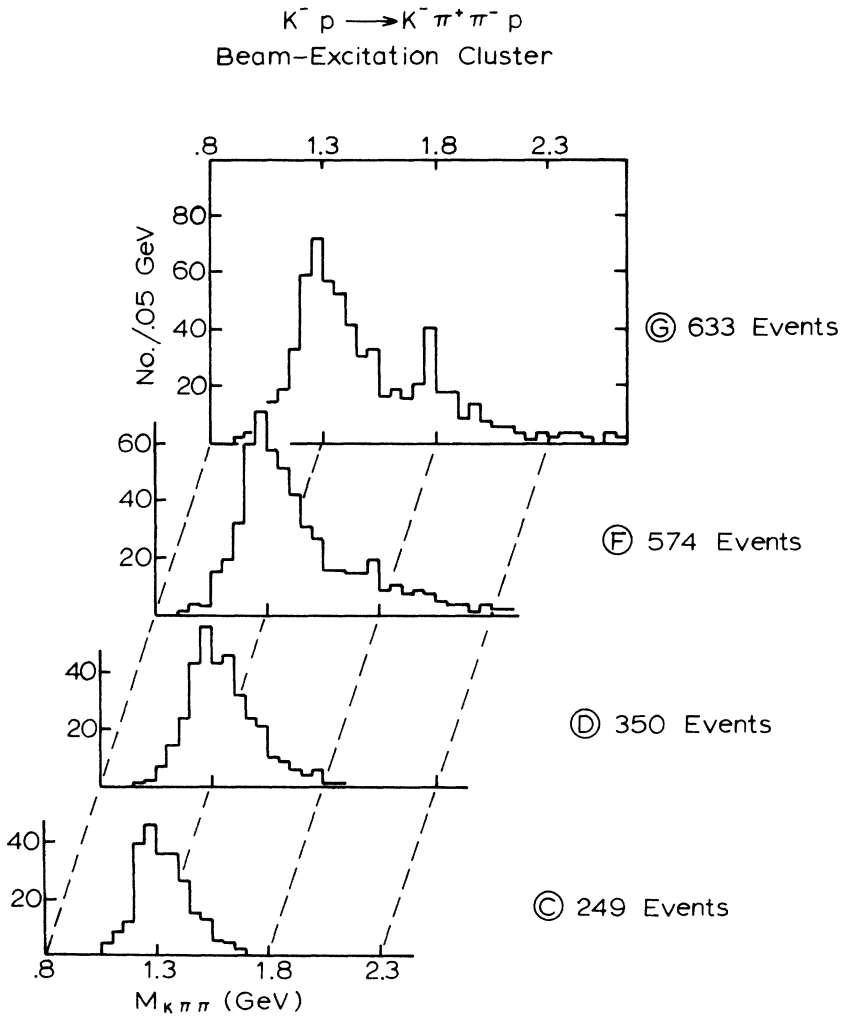


FIG. 3. The $K^-\pi^+\pi^-$ invariant-mass distribution at various points in the development of the beam-excitation cluster of reaction 1. The designations $C$, $D$, $F$, $G$ refer to the diagram of Fig. 1.

sifying successive events in the order in which they become dense. During this process clusters are formed, grow, and ultimately merge together, while new clusters may appear along the way. The signature for significant structure is the persistence of disjoint clusters of data points after many steps. In the diagram of Fig. 1, the branch corresponding to each cluster begins at the density level where the cluster originates, and extends downward to the level at which it merges with another cluster. These merger points are labeled $A$, $B$, $C, \ldots, G$. At each of these points the total number of events in each cluster (dense events plus boundary events) is indicated. The diagram has two distinct arms which result from the evolution of two large, well-separated classes of events. These two major clusters finally merge at point $G$ after $\frac{2}{3}$ of the events have become dense, and a total of 972 events (75% of the sample) have been classified. Upon inspection of the events in these clusters we shall find that the left-hand arm corresponds to diffractive excitation of the incident kaon, and the right-hand arm to diffractive excitation of the target proton.

We first examine the development of the arm $ABCDG$. Here we see that events in the most densely populated regions of our velocity space first form several relatively small clusters. These quickly merge together to form a larger cluster which then remains stable through many steps of the analysis.

Figure 2 shows the events in the two early clusters which merge together at point $B$. Both of these consist of the diffractively produced $Q$ meson, with $K\pi\pi$ masses confined to near-threshold values and $K^-\pi^+$ masses in the $K^*(890)$ band. The

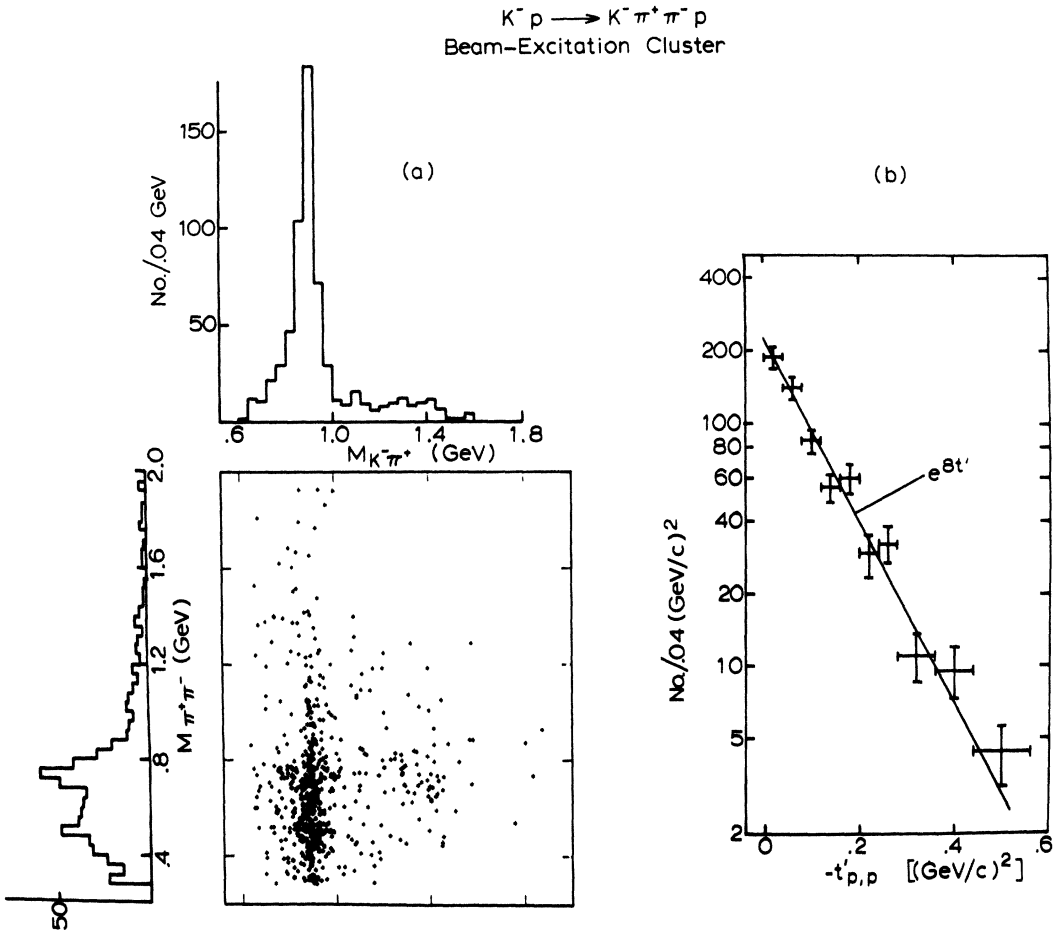


FIG. 4. Events in the beam-excitation cluster of reaction 1, as seen at point $G$ of Fig. 1. (a) Invariant mass of $\pi^+\pi^-$ vs invariant mass of $K^-\pi^+$. (b) Momentum-transfer distribution ($t'=t-t_{min}$) from the target to the outgoing proton. In this, as in the other momentum-transfer distributions, a curve of exponential slope is superimposed to guide the eye and facilitate comparisons.

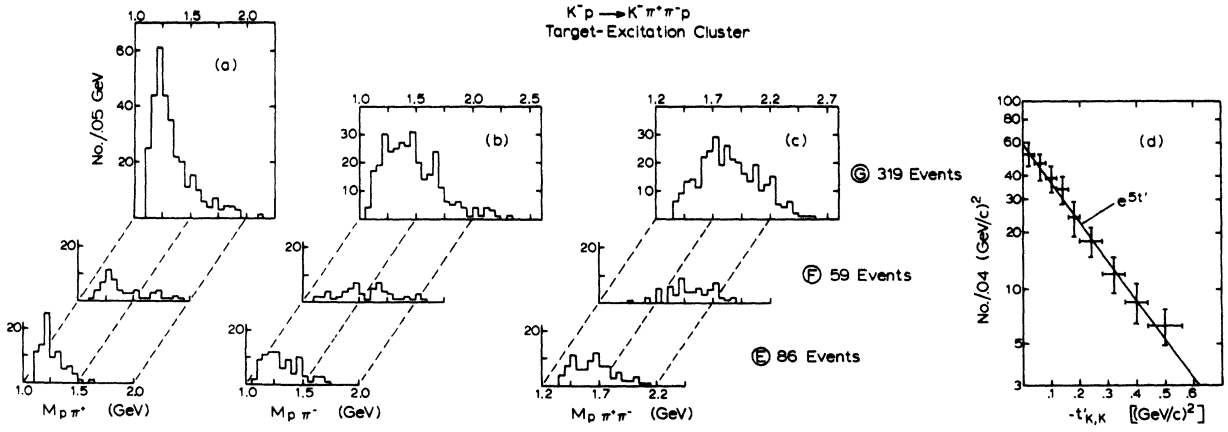


FIG. 5. The target-excitation cluster of reaction 1. (a), (b), (c): The $p\pi^+$, $p\pi^-$, and $p\pi^+\pi^-$ mass distributions for the 86-event cluster at point $E$, the 59-event cluster at point $F$, and the fully developed cluster at point $G$. The designations $E$, $F$, $G$ refer to the diagram of Fig. 1. (d): The distribution in the momentum transfer ($t'=t-t_{min}$) from the incident to the outgoing kaon for the full cluster just prior to the merger at point $G$.

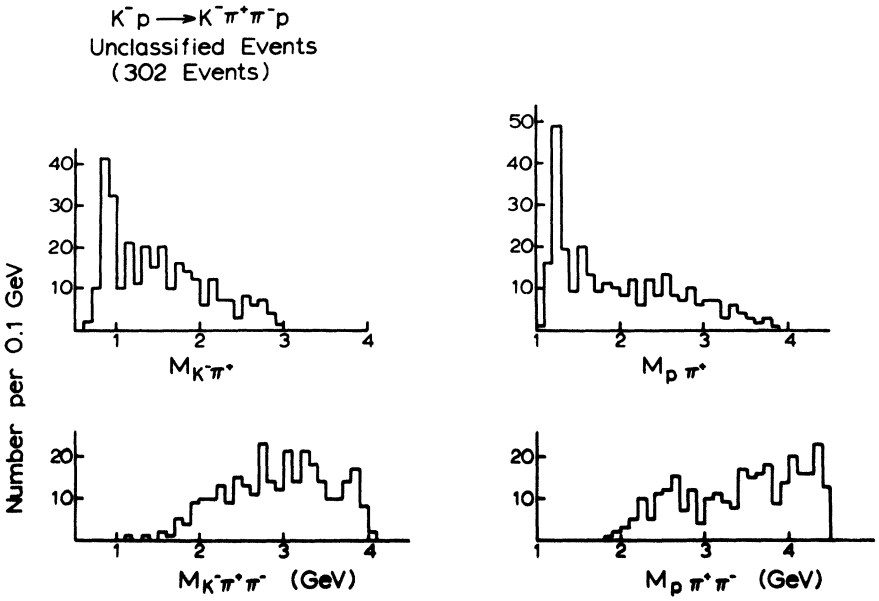


FIG. 6. Reaction 1. Invariant-mass distributions for events not yet classified at point $G$ of Fig. 1 (see text).

larger of these two clusters gives the appearance of a strong $\rho^0$ signal in the $\pi^+\pi^-$ mass distribution, with nearly all of the events lying in the $\bar{K}^{*0}/\rho^0$ overlap region [Fig. 2(b)]. There is no significant $\rho$ signal in the smaller cluster [Fig. 2(d)]. It is unlikely, however, that these two clusters result from two different production mechanisms, or from distinct decay modes of the $Q$. Rather, they reflect the alignment of the $K^*$ along the beam direction in the decay of the $Q$. One cluster picks out events in which the $K^-$ goes forward relative to the $\pi^+$ [Fig. 2(c)], and the other has selected the opposite configuration [Fig. 2(f)]. Thus, apparent clustering effects can result from the nonisotropic decay of resonances. The apparent $\rho^0$ in the larger cluster is just a kinematical reflection. In these data (and, we expect, in all cases of relatively high energy and final-state multiplicity) these effects give rise to relatively small clusters among the most dense events, which quickly merge together and are not to be confused with the more persistent structure characteristic of truly different production mechanisms.

Following the merger at point $B$ (referring again to the diagram of Fig. 1) we have a single large cluster which continues to grow as more events become dense, until the merger at point $G$. Figure 3 shows some "snapshots" of the $K\pi\pi$ mass spectrum as this process continues. As the cluster develops, the $Q$ peak first grows, and then the spectrum spreads out to higher values of the $K\pi\pi$

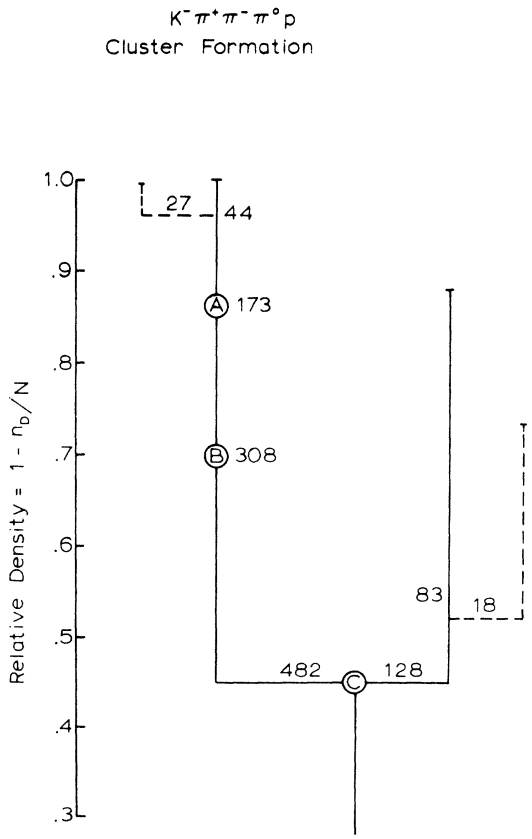


FIG. 7. Schematic of cluster formation for reaction 2 (see text).

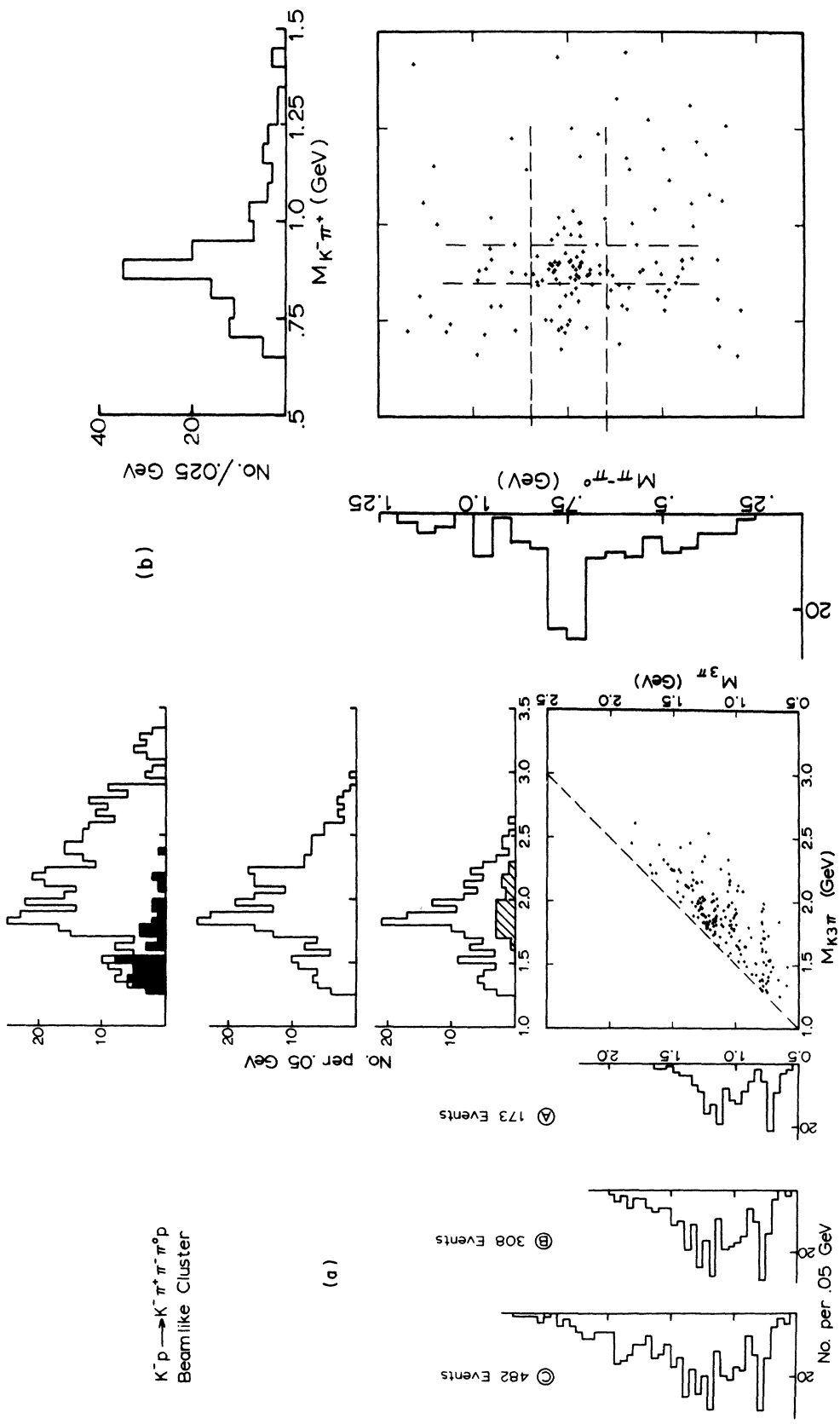


FIG. 8. Reaction 2. The beamlike cluster. (a) Invariant mass of $\pi^+\pi^-\pi^0$ vs $K3\pi$ invariant mass. The scatterplot shows the distribution at point $A$ (173 events) of the diagram in Fig. 7. The histograms show the projected mass spectra at points $A$, $B$, and $C$. The dashed line on the scatterplot indicates the kinematic boundary. On the $K3\pi$ mass distributions the shaded area is the $K\omega$ contributions, and the dashed area indicates the $K^*p$ contribution. (b): The $\pi^-\pi^0$ invariant mass vs $K^-\pi^+$ invariant mass for the cluster as it appears at point A of Fig. 7. Events in the $\omega$ peak $[0.07 \leqslant M(3\pi) \leqslant 0.85$ GeV)] have been removed.

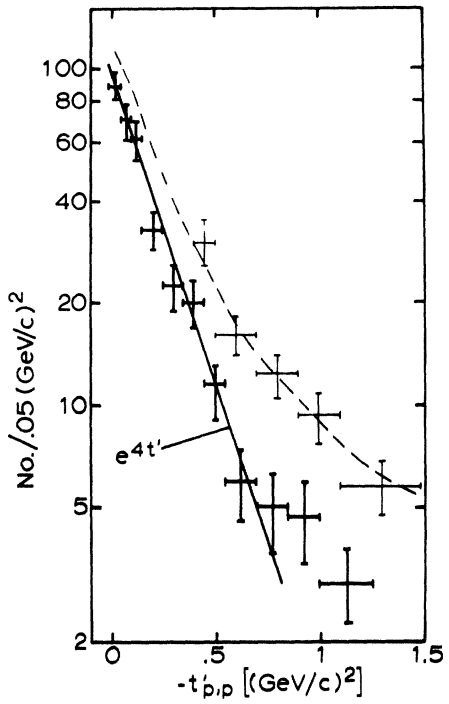


FIG. 9. Reaction 2. The beamlike cluster. Distribution in the momentum-transfer ($t' = t - t_{min}$) from the target to the outgoing proton. The dashed curve (hand drawn) shows the distribution for the total sample. The data points for the total sample are shown lightly drawn, but for the sake of clarity, are suppressed below $|t'| = 0.6$ (GeV/$c$)$^2$.

mass. Finally, a peak corresponding to the $L$ meson appears. Figure 4 shows in more detail the configuration of events in this cluster as it appears just prior to the merger at point $G$. The fully developed cluster has all of the characteristics of that component of reaction 1 which is due to diffractive excitation of the incident $K^-$.

Similarly, the right-hand arm ($EFG$) of the diagram in Fig. 1 consists of events corresponding to diffractive excitation of the target proton. As can be seen from the mass plots in Fig. 5, the clustering pattern indicates two contributions from this mechanism. The most dense events, which appear in the 86-event cluster at point $E$, exhibit a very sharp threshold rise in the $p\pi\pi$ mass spectrum, with $p\pi^+$ and $p\pi^-$ masses lying below 1.5 GeV. As the cluster develops further, very few additional events appear near threshold in the $p\pi\pi$ mass spectrum. The 59-event cluster that joins the larger cluster at point $F$, for instance, is much more spread out in its $p\pi^+$ and $p\pi^-$ mass spectra. The $p\pi\pi$ mass spectrum of the fully developed cluster at point $G$ appears to be a superposition of these higher-mass contributions on top of the initial, tightly confined (in velocity space) threshold enhancement. The momentum-transfer distribution for the full cluster [Fig. 5(d)] has the exponential shape typical of quasielastic processes.

After the merger at point $G$, no new clusters of significant size develop. We expect the events that remain unclassified at this point to consist partly of the tails of the diffractive clusters, partly of experimental background, and partly of events corresponding to various reaction mechanisms of small cross sections that have not been resolved as separate clusters (but which might be resolved with more statistics, or a better measure of distance). In fact, these unclassified events contain no diffractionlike signals. They do exhibit distinct $\bar{K}^{*0}(890)$ and $\Delta^{++}(1236)$ peaks (Fig. 6). The residual $K^*$ appears in association with low-mass $p\pi^-$, and is probably the result of simultaneous $K^*N^*$ production. The $\Delta^{++}$ is peripherally produced, with relatively low $K^-\pi^-$ mass recoiling, suggesting one-pion exchange.

Thus, the two large clusters at point $G$ appear to contain the whole of the diffractive component in reaction 1, and this selection has been made very cleanly. (Note that the ratio of beam-to-target excitation thus obtained is 2:1, in agreement with the wisdom of conventional analyses, factorization assumptions, and so forth.)

### Reaction 2: $K^-p \to K^-\pi^+\pi^-\pi^0 p$

The 880 events of this reaction were analyzed in a five-dimensional velocity space, constructed as before with the longitudinal velocities $\beta_{\pi^+\pi^-}$, $\beta_{\pi^- p}$, $\beta_{\pi^+\pi^-\pi^0}$, $\beta_{\pi^+\pi^0 p}$, and $\beta_{K^-\pi^-\pi^0}$. The hierarchy of clustering patterns is shown in Fig. 7, which is constructed in a fashion analogous to that of Fig. 1. Here we again observe the development of two major clusters, one significantly larger than the other, which remain separated until they merge at point $C$ after 70% of the events have been classified. Unlike the previous case, no large satellite clusters develop along the way. (Numerous small clusters of 15 events or less are not shown in the diagram. In this diagram, $A$ and $B$ are arbitrarily chosen points at which to illustrate the development of the larger cluster.)

Figure 8(a) shows the larger of these two clusters at successive stages of its development. It is a beamlike cluster with exponential falloff in the momentum transfer to the proton (Fig. 9). The most dense events have a prominent $\omega$ signal, with a corresponding threshold enhancement in the $K\omega$ mass spectrum [the shaded area in Fig. 8(a)]. These events also show a sharp, narrow enhancement at what appears to be a higher threshold near 1850 MeV in the $K3\pi$ mass. This second enhancement is associated with $3\pi$ masses be-

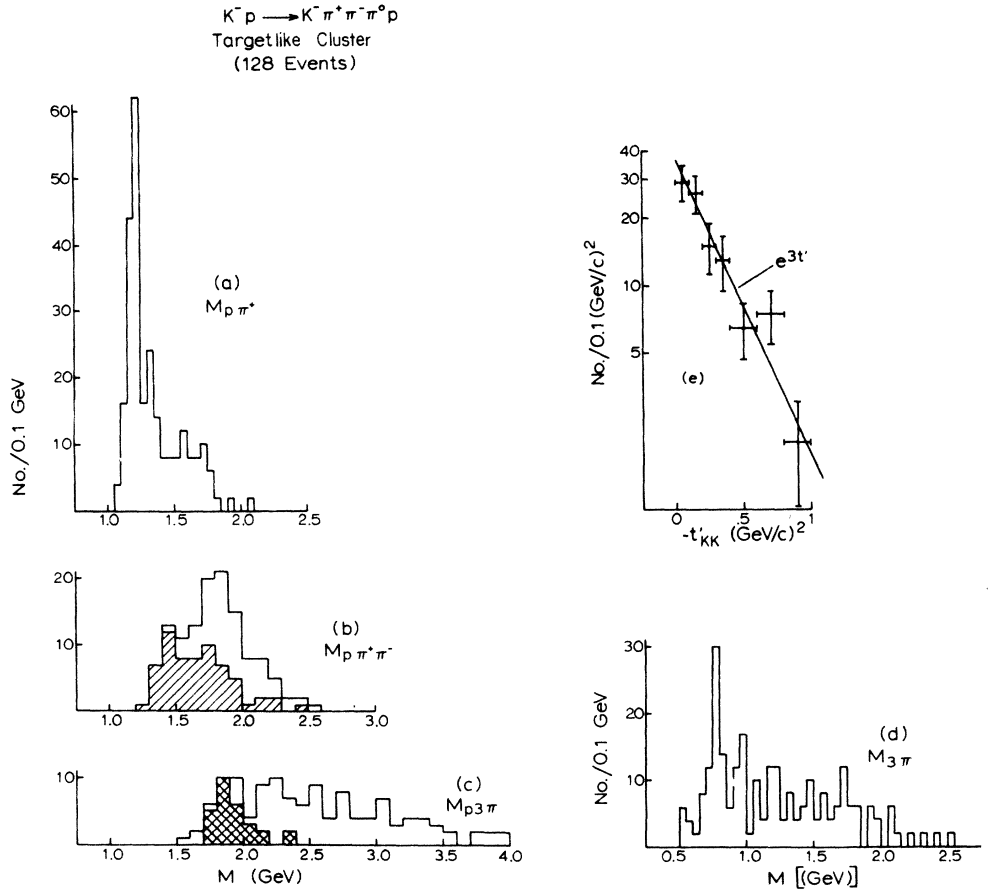


FIG. 10. Reaction 2. The targetlike cluster. (a), (b), (c), (d): The $p\pi^+$, $p\pi^+\pi^-$, $p3\pi$, and $3\pi$ invariant-mass distributions. (e) The distribution in momentum transfer ($t'=t-t_{min}$) from the incident to the outgoing $K^-$.

tween 1000 and 1100 MeV, and lies at the threshold for production of $KA_1$, $K^*(890)\rho$, or $K^*(1420)\pi$. We shall not attempt to make a case for $KA_1$, and there is no clear evidence for $K^*(1420)$ in the $K\pi\pi$ mass spectra. However, a $\bar{K}^{*0}(890)\rho^-$ signal is very much in evidence, as seen in Fig. 8(b). (For an $I=\frac{1}{2}$ $K^*\rho$ system decaying to $K^-\pi^+\pi^-\pi^0$ the $\rho^-$ dominates, which is observed to be the case; $\rho^+$ is not possible.) This sharp rise in the $K3\pi$ mass spectrum between 1750 and 1950 MeV appears to be a complicated amalgam of several different amplitudes. The $K\omega$ and $K^*(890)\rho^-$ contributions are shown shaded in Fig. 8(a). The pattern-recognition algorithm has not resolved the $\omega$ component as a separate cluster, although, upon inspection of the $3\pi$ mass spectrum, one expects that such a separation should be achieved with a more complete and appropriate set of variables.

The growth of the beamlike cluster as more events become dense follows a pattern similar to that seen in the previous reaction: The most dense events fall near thresholds in the $K3\pi$ mass spectrum, with subsequent events contributing at successively higher masses. There is a strong correlation between the $3\pi$ and $K3\pi$ masses, resulting in a striking similarity in the shapes of these mass distributions at each stage of the development shown in Fig. 8(a). The momentum-transfer distribution exhibited by events in this cluster (Fig. 9), while it is exponential in shape, is less steep than the elastic distribution or the quasielastic events found in reaction 1.

It is not obvious whether these events should be interpreted as "diffractive" or not. Indeed, the $K\omega$ and $K^*\rho$ components have steeper $t$ distributions (not shown) than that observed for the total sample. On the other hand, the kinematic structure of these components blends smoothly with that of the other events in the cluster, and the well-known mass dependence of the momentum-transfer slope could account for the observed $t$ distribution.

The smaller of the two clusters found in this reaction is targetlike in nature. The behavior of these events for the fully developed cluster as it appears at point $C$ in the diagram of Fig. 7 is shown in Fig. 10. The predominant two-body re-

sonance is $\Delta^{++}(1236)$. There is an $\omega$ signal in this cluster as well, smaller by a factor of ~2 than the $\omega$ contribution to the beamlike cluster. The $\Delta^{++}$ contributes to a threshold population of the $p\pi^+\pi$ mass spectrum [shaded area in Fig. 10(b)], which in turn is very similar in shape to that observed for $p\pi\pi$ masses in the targetlike cluster of reaction 1. The $p\pi^+\pi^0$ and $p\pi^-\pi^0$ distributions (the latter is shown in Fig. 11) are much broader, and exhibit no prominent threshold enhancements.

The $p3\pi$ spectrum is spread over a wide range of masses, with the $p\omega$ component [shown shaded in Fig. 10(c)] contributing near threshold. The momentum-transfer distribution falls significantly less steeply than does that of the beamlike cluster.

Figure 11 shows the contribution of the beamlike and targetlike clusters to a selection of the multiparticle mass distributions for this reaction. The events that remain unclassified at the point where the two clusters merge contain no additional

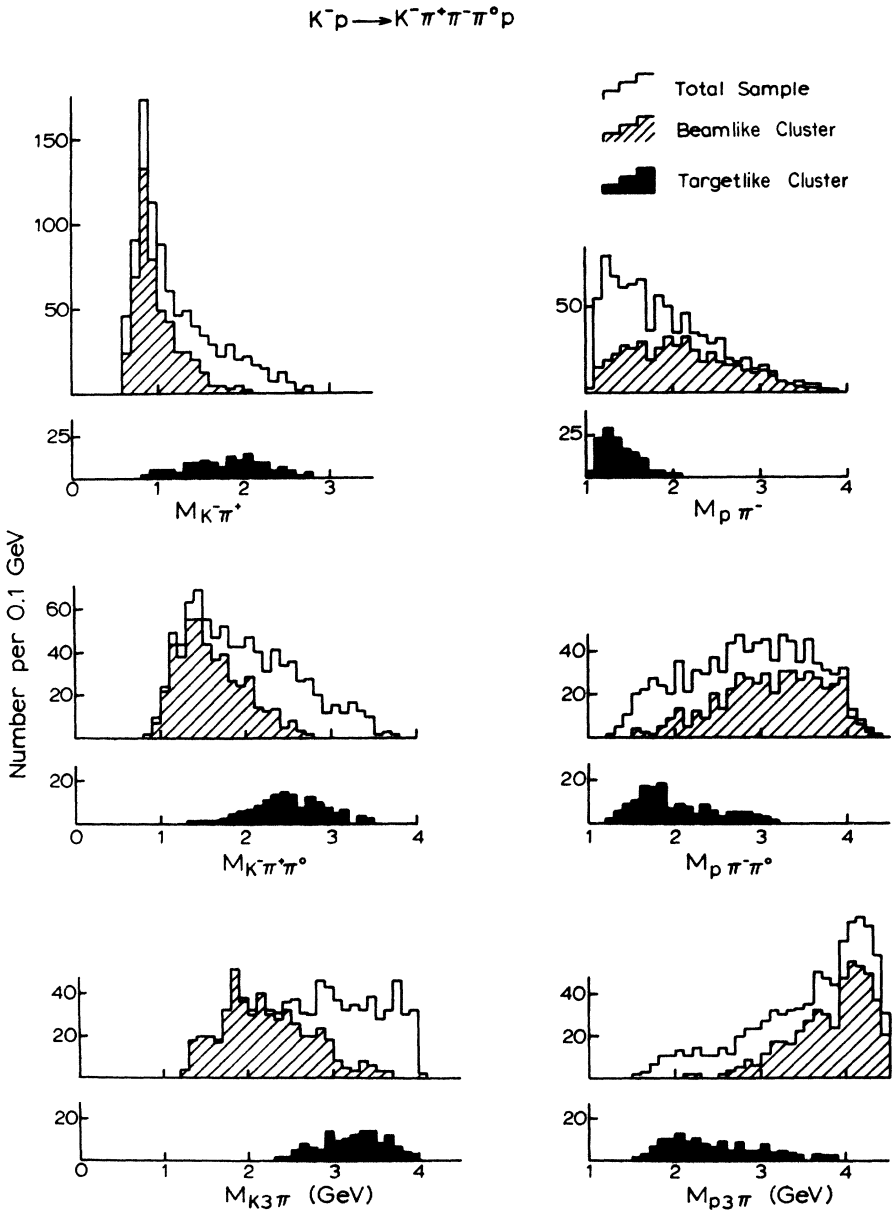


FIG. 11. Some two-, three-, and four-body invariant-mass distributions for reaction 2. In each case the open histogram shows the total sample (880 events), the dashed area contains the events in the beamlike cluster (482 events), and the heavily shaded histogram shows the events in the targetlike cluster (128 events).

threshold behavior in the $K3\pi$ or $p3\pi$ masses, and no additional $\omega$ signal.

For our present purposes we conclude that the clustering patterns that we have "discovered" in these two reactions do correspond to real dynamical structure, and appear to have revealed the essential features of the reaction mechanisms involved.

## APPENDIX: AN ALGORITHM FOR THE DISTANCE ANALYSIS

We present here a detailed outline of an algorithm which performs the distance analysis. In general, the approach to this problem will depend on whether computing time or storage space is more crucial. The following method is the fastest and conceptually the simplest we could devise.

Let $N$ be the total number of events to be analyzed and let $k$ be the threshold density. We first construct the halo array $h_{ij}$, where for each event $i$, $h_{ij}$ is its $j$th nearest neighbor: $1 \leq i \leq N$ and $1 \leq j \leq k$. Only one calculation of each distance $R_{lm}$ is necessary: If $R_{lm}$ is smaller than the distance between $l$ (or $m$) and its $k$th nearest neighbor so far computed, then $m$ (or $l$) is added to $l$'s (or $m$'s) halo and the previous $k$th nearest neighbor is removed from $l$'s (or $m$'s) halo. This requires storing $2kN$ numbers ($h_{ij}$ and $r_{ij}$) plus the event coordinates needed to calculate the distances. The $r_{ij}$ array of $k$ smallest distances can be eliminated if the distance between each pair is computed twice. However, a list of distances to the $k$th nearest neighbor, for example $r_{ik}$, is needed to compute the order in which the events become dense: $d_i$ is the event which has the $i$th shortest distance to its $k$th nearest neighbor, $1 \leq i \leq N$. Note that $h_{ij}$ and $d_i$ are integer arrays of event labels.

The distance analysis itself can be done entirely in terms of $N$, $k$, $h_{ij}$, and $d_i$; no further reference to the data is needed. The events and their halos are analyzed in the order $d_1, d_2, \ldots, d_i, \ldots, d_N$. At the $i$th stage, each event is either unclassified, or it is a boundary event assigned to some cluster, or it is a dense event assigned to some cluster. The classification of event $j$ (at stage $i$) is $\bar{c}_j$. The main information of the analysis is carried by this $N$-dimensional vector, and it is changed at each stage. At the $i$th stage ($d_i$ becomes dense), there are several possibilities: (1) The halo of $d_i$ has no dense events. Then start a new cluster with dense event $d_i$, and reclassify the unclassified events in the halo of $d_i$ as boundary events of this cluster. (2) The dense events in the halo of $d_i$ all belong to the same cluster. Then add $d_i$ as a dense event to this cluster and assign the unclassified events in the halo of $d_i$ as boundary events of the cluster. (3) The dense events in the halo of $d_i$ belong to two or more different clusters. Then all the dense and boundary events of these clusters (including events not in the halo of $d_i$) are merged into a single cluster, and the unclassified events in the halo of $d_i$ are classified as boundary of the merged cluster. We have found it convenient to write out the premerger classification if the second largest cluster has some minimum number of dense events. The classification vector $c_j$ at this stage then gives a partition of the data into clusters of events, but the interpretation is subject to the considerations discussed in Sec. II.

---

*Work supported by the U. S. Energy Research and Development Administration under Contract No. AT(11-1) 3075.

†Work supported by the U. S. Energy Research and Development Administration under Contract No. W-7405-ENG-36.

[1]See, for example, L. Van Hove, Phys. Rep. 1C, 347 (1971).

[2]W. Kittel, L. Van Hove, and S. Ratti, Nucl. Phys. B30, 333 (1971).

[3]J. Hanlon *et al.*, Phys. Lett. 46B, 415 (1973); T. Ludlam and R. Slansky, Phys. Rev. D 12, 59 (1975).

[4]For a review of this and other related techniques, see J. H. Friedman, Lectures presented at the CERN School of Computing, 1974 (unpublished).

[5]J. E. Brau *et al.*, Phys. Rev. Lett. 27, 1181 (1971).

[6]L. Van Hove, Nucl. Phys. B9, 331 (1971); M.-S. Chen and R. F. Peierls, J. Comput. Phys. 16, 195 (1974).

[7]H. Böttcher *et al.*, Nucl. Phys. B81, 365 (1974); W. D. Nowuk and H. Schiller, Comput. Phys. Commun. (to be published).

[8]D. Wishart, *Numerical Taxonomy*, edited by J. Cole (Academic, London, 1969), p. 282.

[9]R. A. Jarvis and E. A. Patrick, IEEE Trans. Comput. C-22, 1025 (1973).

[10]C. A. Goad, Los Alamos Report, 1975 (unpublished).

[11]C. T. Zahn, IEEE Trans. Comput. C-20, 68 (1971).

[12]This is an empirical prescription. The problem of defining local density by $k$ nearest-neighbor measures is discussed in, for example, K. Fukunaga and L. D. Hostetler, IEEE Trans. Inf. Theory IT-19, 320 (1973).

[13]C. L. Siegel, *Topics in Complex Function Theory* (Wiley-Interscience, New York, 1969), Vol. 2; F. Gürsey (private communication).

[14]A. J. Slaughter, Yale Ph.D. thesis, 1969 (unpublished).