Submitted to: Physical Review D March 3, 1981

E-299

1 mg/n

FERMILAB-PUB-81-111-E

Multiple Correlations and High Transverse Momentum Jets in 147 GeV/c π⁻p Interactions

D. Brick, A. M. Shapiro, M. Widgoff Brown University, Providence, Rhode Island 021912

and

E. D. Alyea, Jr. Indiana University, Bloomington, Indiana 47401

and

E. S. Hafen, R. I. Hulsizer, V. Kistiakowsky, A. Levy^(a), P. Lutz^(b), S. H. Oh, I. A. Pless, J. P. Silverman, T. B. Stoughton, P. C. Trepagnier, R. K. Yamamoto, Department of Physics and Laboratory for Nuclear Science Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

and

H. O. Cohn Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830

and

R. J. Plano, P. Stamer^(C) Rutgers University, New Brunswick, New Jersey 08903

and

E. B. Brucker, E. L. Koller Stevens Institute of Technology, Hoboken, New Jersey 07030

and

W. M. Bugg, G. T. Condo University of Tennessee, Knoxville, Tennessee 37916

and

T. Ludlam^(d), H. D. Taft Yale University, New Haven, Connecticut 06520

ABSTRACT

We examine multiparticle correlations in a π p experiment at 147 GeV/c performed by the Proportional Hybrid System Consortium. The major aim of this paper is to demonstrate the existence of clusters in our data. We use different statistical algorithms to assemble into clusters the particles in each event which are associated by virtue of small relative angles. We find that these clusters are stable against different choices of metric and/or algorithm, and reproduce the effects previously observed in the data corresponding to clusters.

Some of these clusters have properties similar to high p_T jets. A detailed study of these jet-like clusters is described, and comparisons with some counter experiments are discussed.

I. Introduction

Ì

The observed short range rapidity correlations between particles in multiparticle events have often been interpreted in terms of production of these particles through an intermediate stage called a cluster¹. But aside from a rather rough comparison with the shape(s) of the correlation(s)--the magnitude of the correlation is fitted by adjusting the cluster size in a cluster model--this interpretation has been without a crucial test. If the ultimate goal is to be able to tell whether clustering is merely a convenient language, or whether clusters are actually produced, it becomes of interest to try to demonstrate the existence of clusters in the data of multiparticle events.

The concept of clusters is clearly related to that of the hadronic jets which have been observed in both $e^+e^$ collisions and hadronic interactions. Although there is a commonly accepted method for analyzing hadron jets in e^+e^- +hadrons, an operational definition of jets in hadronhadron+hadrons is still needed. The problem of a "trigger bias"² is a well known example of how difficult it is to define a jet. A review of the methods used, up to now, in order to define hadronic jets can be found in ref.3.

The aim of this paper is to develop methods of finding multiparticle correlations using all the kinematic variables of each particle, and to demonstrate that the particles are grouped in clusters. The main question is whether such clusters have properties which are independent of the methods used to form them and whether the clustered data reproduce known experimental results. In Section II, we discuss the definition of different distances between particles inside one event, while in Section III we describe two different algorithms used to aggregate particles into clusters. In Section IV we compare the results from clusters obtained by applying these algorithms to 147 GeV/c data with previously published results from the same data. In Section V we present our results, with a particular emphasis on their stability with respect to the choice of metric and algorithm. In Section VI we compare these results with two different Monte Carlo models. Finally, in Section VII, we discuss our results for clusters with large transverse momentum and compare them with the properties of jets observed in other experiments. Our conclusions are presented in Section VIII.

II. Distances Between Particles in Minkowsky-Space

The object of this study is to assemble each event into groups of particles which are somehow associated in full momentum-space, and not just in a one-dimensional space as in, for instance, the rapidity-gap method. We limit ourselves in this paper to methods based on nearest neighbor techniques. Thus, we have to define a distance between two particles of the same event, each particle being regarded as a point in the Minkowsky space of energy momentum.

-2-

Let us define p_i , \dot{p}_i , E_i , m_i to be respectively the 4-momentum vector, the 3-momentum vector, the total energy and the mass of particle i. The most general measure of the distance, using the known norm in a Minkowsky space, is:

Ż

$$d_0^2 (p_i, p_j) = -(p_i - p_j)^2 = -t_{ij}$$
 (1)
Unfortunately, This "distance" can be imaginary $(d_0^2 < 0)$

if $m_i \neq m_j$. So the expression actually used is:

$$M_1^2 (p_i, p_j) = M_{ij}^2 - (m_i + m_j)^2$$
 (2)

where M_{ij} is the two-body invariant mass. Expressions (1) and (2) are identical for the case where $m_i = m_j$.

The distance d_1 , despite some very good properties (see below), is not a mathematical distance because it does not obey the triangle inequality. Furthermore, we will see in section IV that we cannot reproduce old, classically obtained results with our data sample, by using d_1 . So we also used another Lorentz-invariant distance, the expression for which is:

$$d_2 = \cosh^{-1} \frac{p_i \cdot p_j}{m_i m_j}$$
(3)

This d₂ can be viewed either as:

- (1) the imaginary angle between the two four-vectors
- (2) the relative rapidity of the two particles in the(ij) rest frame.
- (3) the boost parameter of the Lorentz transformation that connects the particle i rest frame to the particle j rest frame.

-3-

These two distances (d₁ and d₂) have the following properties:

- 1) They are positive.
- 2) They are zero only if two conditions are fulfilled at the same time: 1) \vec{p}_i and \vec{p}_j parallel and in the same direction, and 2) $\beta_i = \beta_j$ where $\beta = |\vec{p}|/E$ is the velocity.
- 3) They are Lorentz invariants.
- 4) d₁ has dimension of a mass, d₂ is dimensionless,
- 5) d_2 obeys the triangle inequality, but d_1 does not.
- 6) d_1 and d_2 are related by the relation

$$d_{2} = \cosh^{-1} \left(1 + \frac{d_{1}^{2}}{2 m_{i} m_{j}} \right)$$
(4)

and are analytically related to the invariant mass M_{ij} of the two particles. Nevertheless, there is a difference between the behavior of these two distances, which can be explained by a simple example: take two pions $(\pi_1 \text{ and } \pi_2)$ and a proton (p) such that $d_1(\pi_1, \pi_2) = d_1(\pi_1, p)$. Then the corresponding d_2 distances are not equal: $d_2(\pi_1, \pi_2) > d_2(\pi_1, p)$. This difference is due to the product m_1m_2 which appears in equation (4).

Because we want to check the stability of the clusters against changes in the metric, we have to find (and use) another distance, with, if possible, a very different behavior from d_1 or d_2 . The picture of PETRA jets is well described by saying that the relative angle between two particles in a jet is small. With this in mind, we defined a third distance:

$$d_{3}(p_{i}, p_{j}) = \cos^{-1}(\vec{p}_{i}, \vec{p}_{j}) / |\vec{p}_{i}| |\vec{p}_{j}|$$
 (5)

-4-

This is a mathematically good distance, but not Lorentz invariant (we have chosen the center of mass frame). The interesting point is that it does not depend on the effective mass of the two particles, which gives rise to a very different behavior between d_2 and d_3 . Appendix I illustrates this point with a simple example.

III. Clustering Algorithms

From the rather large panoply of clustering algorithms, we have chosen two simple ones, hereafter referred to as MST and CAH.

A. MST (Minimal Spanning Tree) has been described by one of us (T.L.) in ref. 4 . In this algorithm the clusters are defined as the disjoined subtrees of the nearest neighbor graph. It is easily described by saying that:

Each particle is connected to its nearest neighbor.
 Thus, if a particle is in a cluster, so is its nearest neighbor.

2. All such connections which are larger than a distance DCUT are broken (this threshold is applied for picking out particles which are clearly isolated in momentum space).

In this way, there is no implicit scale of mass for determining the cluster sizes. The main disadvantage of this algorithm is that it is very unstable to slight changes in the positions of the particles.

B. CAH (our acronym for "Classification ascendante hierarchique selon la variance")⁵ can be described as follows:

1. Merge into a subset the two particles which are the closest to each other, and replace them by their 4-vector sum.

-5-

2. Recalculate distances between particles and/or subsets of particles, using as distance between subsets A and B:

$$d^{2}(A,B) = \frac{2n_{A}n_{B}}{n_{A}+n_{B}} d^{2}(p_{A},p_{B})$$
 (6)

where $p_A(p_B)$ are the 4-momenta of subsets A(B) $n_A(n_B)$ are the numbers of particles inside subsets A(B)

 Return to the first step, or stop if the new minimum distance is greater than a threshold DCUT.

Formula (6), which weights the distance between sub-sets, has a theoretical root; this chosen criterion is a measure of the increase of inertia involved in the merging, where

Inertia (A) = I(A) =
$$\sum_{i=1}^{n_{A}} d^{2}(p_{i}, p_{A})$$

and using the Huyghens theorem:

$$I(AUB) = I(A) + I(B) + \frac{n_A n_B}{n_A + n_B} d^2(p_A, p_B)$$

(The factor 2 in (6) is used simply to give a weight of 1 when $n_A=n_B=1$)

The results obtained by this algorithm are much less sensitive to slight changes in the positions of the particles; however, the size of the clusters is determined by the choice of DCUT. Figure 1 describes these two algorithms on a twodimensional example, with a DCUT = 2.5. IV. The Philosophy and Choice of the DCUT Thresholds

With the MST algorithm, a cut is clearly needed to obtain some one-particle clusters, some of which must correspond to the leading particles that are known to be in the data. With the CAH algorithm, the effect of the cut is different, since DCUT determines the size of the clusters. Thus, there is no reason, for a given distance, that DCUT(CAH)=DCUT(MST). We therefore have to find 6 values of DCUT (3 distances x 2 algorithms).

There is no theoretical best-value for the choice of DCUT, but two simple arguments allow us to at least reduce the possible range:

1) The $\pi\pi$ mass distribution, computed for pairs of pions both of which are in the same cluster, has to show a high-mass tail without sharp discontinuities, especially when the cluster contains only these two pions. In our data, this requirement implies an empirical lower bound for DCUT which is of the order of 1 GeV for d₁, and 4 for d₂. No minimum value is required by the d₃ distance.

2) Obviously, the two-prong elastic events have to be two-cluster events. This requires that the higher bound for DCUT is roughly:

> \sqrt{s} for d_1 $\cosh^{-1}(1+\frac{s}{2m_am_b})$ for d_2

and π for d_3

where m_a and m_b are the masses of respectively the beam and the target. It turns out that the diffractive events in the 4-prong 4C fit samples give the same upper limit for DCUT.

Since the possible range for DCUT is large, we have chosen

-7-

an empirical technique to determine the values of the DCUT. There are certain known "cluster"-type correlations produced when 147 GeV/c pions interact with protons. A general description of our π p experiment at that momentum can be found in ref. 6. We will require that any algorithm and DCUT reproduce these correlations, in particular those described in refs. 7, 8, and 9. We will denote hereafter a particular choice of algorithm/distance/DCUT in that order. For example, CAH/d₂/4.5 means CAH algorithm with distance d₂ and DCUT=4.5.

First, we can compare the results of our paper⁷ for the central fireball cross-section from a rapidity-gap analysis with our cross-section forevents having three or more clusters. This number, considered as a function of DCUT, is obviously a decreasing function. It turns out that the MST algorithm gives an unacceptably low value for this cross-section, for d_1 and d_2 , even for the minimum DCUT (7 mb for MST/ d_2 /4.0, and 8 mb for MST/ d_1 /1.0, compared to 13 mb in ref. 7).

Second, our clustering methods should result in leading particles appearing in single particle clusters with cross-sections agreeing with those given in ref 8. Thus, we applied the same cuts in Feynmann x as in ref. 8 to our l-particle clusters and compared the results. It turns out that $CAH/d_1/$ gives rise to a proton leading cross section which is systematically too high.

Third, using only the events $\pi^- p \rightarrow 2\pi^- 2\pi^+ x^0$, and applying the same cuts as in ref. 9 (off-mass-shell diffraction), we were able to reproduce the result of that analysis for the

-8-

three remaining combinations of algorithm/dist.

This discussion of DCUT, together with some numerical results, is summarized in Tables 1 and 2. In conclusion, clusters chosen by three combinations of algorithm and different metrics are in agreement with already published data. We must next address the question whether the clusters produced by these different techniques have the same properties.

V. General Results Obtained by These Clustering Algorithms

The number of clusters per event is displayed in Fig. 2 and the average values are shown on Table 3, as a function of final-state prong number. The average number of clusters per event is small, and increases with multiplicity by something on the order of .5 clusters per 2 prongs, starting from the approximate value of 2 clusters for the 4-prong events.

The next figure (Fig. 3) shows the internal charged multiplicity k inside the clusters. On the average, there are slightly less than 3 charged particles per cluster, and the distribution is always narrower than a Poisson distribution (σ^2 <2). The invariant mass distributions of the clusters are shown in Fig. 4. The agreement of the results obtained from the different algorithms or distances seems to indicate that these clusters are not an artifact of the analysis. This is supported by the observation that the mean mass of the clusters is of the same order for the two distances, which have a very different link to invariant mass.

To continue, a crucial point is whether or not the known resonance signals in the data remain intact when

-9-

events are broken up into clusters. The familiar resonances do not seem to appear as isolated clusters (no ρ^{0} signal is seen in the two pion invariant mass spectrum shown in Fig. 4). In order to study where the known resonances are, we selected events having a Δ^{++} resonance (Δ^{++} has been defined by $1.12 \leq M(p\pi^{+}) \leq 1.32$ and $|t_{p+\Delta}| < 1.$), and clustered these 389 events. The results are displayed in table 4 for both distances. Again, the main point to notice is the remarkable stability of the results against the choice of metric, together with a new measurement of the fraction, D, of directly produced Δ^{++} in the total number of Δ^{++} produced: D=(23.6+2.0)%.

The same kind of study cannot be done for the ρ resonance, because of the well known large combinatorial background. Nevertheless we know that some ρ 's must be found inside the beam diffraction dissociation component. So we present in Fig. 5 the $\pi^+\pi^-$ invariant mass spectrum for π 's belonging to the same cluster, this cluster having been chosen by the following procedure. In order to enrich the sample in beam diffraction dissociation, we took only two-cluster events, and computed the $\pi^+\pi^-$ invariant mass for the forward cluster, providing that its charge is -1 (charge of the beam) and its rapidity is greater than 1 (to remove the central region). A ρ signal is clearly visible in Fig. 5 for both distances and is even more enhanced if we select only the 3π clusters (dashed histograms).

Another check is to examine the charge of the clusters, since it has to reflect the well-known local compensation of charge. We found that 78% (respectively

-10-

75% and 76%) of the clusters arising from CAH/d2/4.7, CAH/d3/1.75, and MST/d3/1.15, respectively, have a charge of zero or + 1.

Summarizing the results for the 147 GeV/c data, the correlations in momentum space are such that events exhibit clusters of particles--with typically 3 charged particles per group--having masses and charges similar to those of ordinary particles or resonance states. It would be presumptuous to advocate that these clusters should be interpreted as "The Clusters" whose existence is universally inferred in recent literature on multiparticle correlation analysis. However, the very impressive stability of the results with respect to particular choices of algorithm and/or distance suggests that these structures may not be created by the methods. Whatever dynamics underlie this pattern, it provides a unique and apparently meaningful prescription for reducing multiparticle events to few-body structures. It then becomes of interest to examine the kinematic properties of the clusters.

Figure 6 shows the invariant cross section, integrated over the whole rapidity range, for both single pions and clusters, as a function of the transverse momentum of these objects. The single particle distribution shows the expected exponential behavior, with perhaps a higher tail than that given by a pure exponential. The clusters coming from all three algorithm/distance combinations show a unique behavior. When compared with the single particle distribution, a crossover appeared at p_T° .8 GeV/c. Above 2 GeV/c, the cross-section for cluster production is one order of magnitude higher than

-11-

for single particle production, which is just the behavior expected for what are now called high p_m jets. In Fig. 7, we present our result -- an average of the results of the three algorithms for production at 90°--and compare it both with a result from a calorimeter experiment 10 and with the theoretical prediction of ref. 11. Our cluster results lie systematically lower than both theoretical and experimental jet results by a factor of ~5 when compared with ref. This would be expected if our clusters 11. were in fact similar to the high \textbf{p}_{τ} jets because our clusters contain only charged particles and the jets also include neutrals. In fact, if we assume that in our clusters the neutral particles carry 15% of the total p_m of the jet, our cross-section curve shifts to remove the factor of ~ 5 difference¹².

VI. Nature of High p Clusters

In order to determine if our high p_T clusters can be identified with high p_T jets, we carried out three further comparisons.

1) The first one is a comparison on an event-byevent basis of the clusters chosen by different algorithms. This comparison allows us to give a quantitative statement on the stability with respect to algorithm/distance combinations in finding high p_T clusters. Table 5 is a summary of this event by event comparison, which indicates that one-third of the events have exactly the same clustering and 50% of all clusters are identified identically by the three combinations. Furthermore, the probability of finding the same cluster with \checkmark two different combinations is a rising function of the

-12-

energy of the cluster. In the central region and at low p_T , this probability is low (25%), reflecting mainly the differences already noted between d_2 and d_3 , while that same probability is high (above 70%) for clusters having an energy greater than 7 GeV, i.e. for clusters looking like jets (both low p_T or high p_T jets). Finally, if we restrict that comparison to clusters having a p_T greater than 1.5 GeV/c, and allowing the cluster to differ by only one slow particle, this probability is as high as 85%.

2) A comparison between the experimental clusters and those formed from Monte Carlo events has been performed. It was not possible to generate events which both satisfy energy-momentum conservation and reproduce the experimental single particle p_T and y distributions. Furthermore, since we are interested in differences and/or similarities in data lying in the tail of a distribution, where we have rather large uncertainties for the experimental data, it becomes difficult for that comparison to be a clear test. We therefore generated events according to three different Monte Carlo's, none of them perfect. In all three cases, we designated some of the particles which we generated as neutrals, and clustered only the "charged" particles.

a) One Monte Carlo, (1), required energy and momentum conservation, but resulted in single particle distributions which were different from the experimental results for the high p_T tail of the single particle p_T distribution (too steep fall-off) and the plateau in the central region of rapidity (no plateau). Cluster analysis of these Monte Carlo events resulted in a very small cross section for high p_T clusters.

-13-

b) The second Monte Carlo, (2), generated events according to the experimental single particle distributions, but did not require energy and momentum conservation. Because every particle was generated independently of the others, it was not necessary to generate "neutrals". This Monte Carlo gave a larger cross section for high p_T clusters than that for the experimental data, especially in the intermediate region $1.5 < p_T < 2.5$ GeV/c. This result is apparently attributable to the fact that this model gives too many events (compared to experiment) with a big p_T imbalance (p_T of the charged system).

The third Monte Carlo, (3), was similar to the second, c) but additionally required the angle in the transverse plane of the particle generated last to be opposite to that angle for the remaining system. Because of that correlation, we needed to designate some particles as neutrals. This resulted in a set of fake events with single particle distributions in exact agreement with those from the experimental data, and with a better but not perfect conservation of energy and There were still too many $p_{\boldsymbol{\pi}}$ imbalanced events momentum. (three times the expected amount) but less than in model #2. From this model, we obtained a cross section for high \textbf{p}_{m} fake clusters smaller than that for model #2, and comparable within the errors to the one from the experiment.

Fig. 8 illustrates and summarizes the results of these three independent Monte Carlo samples and their comparison with our experimental data. The conclusion of this study is the following. A "perfect" Monte Carlo would probably be somewhere between models #1 and #3, possibly near #3. Therefore we are not able at this point to conclude whether the high p_T clusters in our data are real jets or statistical fluctuations. However, we believe the existence of jets has been established, and we have to compare the properties of our high p_T clusters with both the theoretical and experimental properties of jets.

3) Using jets generated by the Feynman-Field (F.F.) Monte Carlo program described in ref. 13 with the same internal properties as those found for our high p_T clusters (charged multiplicity and energy), we have checked the

-14-

efficiency of the clustering algorithms to detect such jets. This efficiency can be measured by the quantity:

$\varepsilon = \frac{\text{number of jets found by the algorithm}}{\text{number of generated jets}}$

Because the program did not generate complete events with jets, but only a single jet, we had to perform this analysis in two steps.

a) We analyzed each generated jet as if it were a complete event, using each of the three algorithm/distance combinations. The value found for ϵ (Table 6) was of the order of 75% for CAH/d₂, and 55% for both of the others. More than 40% of the jets are correctly found by all three algorithms. If we add to the above sample those cases in which the particles missing from the generated jet all have a fractional longitudinal momentum $Z = \vec{p} \cdot \vec{p}_{jet} / |p_{jet}|^2$ less than .14, the efficiencies rise to 90% for any individual algorithm, and to 70% if we request the "jet" to be found by all three algorithms. Figure 9 shows both efficiencies as a function of the charged multiplicity inside the generated jets. The losses mainly come from the neutral particles generated by the F. F. program but removed before our clustering algorithms were applied to the data set. The neutral particles were removed from the data set in order to simulate our experimental situation.

b) We also used the cluster analysis on fake events consisting of a F.F. generated jet (correctly found when alone) and an experimental event formed by two low p_T clusters. The jet was always put in the central region of rapidity ($|y_{c.m.}| \le 1$). Table 7 presents the results of this analysis, and it can be seen that a cluster with a p_T greater than 1.5 GeV/c was found in more than 92% of these fake events. The

-15-

fake jet was found in only 40% of the events, but this percentage increases considerably if one includes cases in which the fake jet is "nearly" found. Three different definitions of "nearly" have been used (see table caption), which clearly overlap; in the case of the first definition, the efficiency of any of the three algorithms was greater than 75%.

These two complementary Monte Carlo studies allow us to claim that the clustering algorithms used are able to extract multiparticle correlations, especially those correlations which give rise to high transverse momentum jets. Nevertheless, this is not sufficient to permit identifying our high p_T clusters as jets. However, in order to pursue this identification, we can compare the properties of these high p_T clusters with those of experimental jets.

VII. Study of High p Clusters

In this section, we will call any cluster with a p_T greater than 1.5 GeV/c a "jet". This particular value of 1.5 GeV/c is a compromise between statistics and the unambiguous identification of a jet. We define the four-momentum of such a jet as the sum of the four-momenta of the (charged) particles belonging to that jet. For this jet study, we have chosen to use the results of CAH/d₂/4.7 for three reasons:

1) Simplicity

- 2) The distance is Lorentz invariant
- 3) The algorithm seems to be the most efficient for detecting jets (see § VI3)

Table 8 displays the general characteristics of our jet events. These results can be summarized as follows:

-16-

1) The inclusive cross section for $\pi^-p \rightarrow jet + X$ is, at our energy, of the order of 10% of the total π^-p cross-section.

2) Only 10% of jet-events have more than one high $p_{\rm T}$ jet.

3) The probability for finding a jet in one event increases with the internal multiplicity of that event.

4) The mean number of clusters found in jet events is slightly higher (3.56) than in all events (2.86).

Figure 10 gives the distribution in θ^{π} , polar angle (in the center-of-mass system) of the jet-axis, and it is interesting to note that this is the first analysis showing high p_{T} jets in the backward hemisphere from a π beam experiment. One may compare our data with predictions of the Feynman-Field quark model¹⁴, at our $x_{T} = 2p_{T}/\sqrt{s}$ of .22.

1) The ratio, R =
$$\frac{(\sigma \pi p \rightarrow jet + X)_{35^{\circ}}}{(\sigma \pi p \rightarrow jet + X)_{145^{\circ}}}, \text{ of}$$

cross sections for producing a jet at $0^*=35^\circ$ versus that at 145° is found to be R=1.47+.24, which agrees within three standard deviations with the value of ~ 2.2 predicted by ref. 14.

2) The experimental mean values of the charge of the jets allow us to deduce the charge ratios as a function of θ^* (see table 9). Our results are seen to be in good agreement with the predictions of ref. 14, also given in table 9.

We can also present some results for the momentum balance in our jet events, although this is necessarily tentative since we have no information on neutral particles. Let us first consider the planarity of these events. The conventional manner of studying this is to define a "trigger plane", which in our case is obviously the plane defined by

-17-

the beam axis and the axis defined by the charged particles in the jet (in the case of two jet events, the higher p_T jet axis), and to study the component out of that plane, p_{out} , the definition and the distribution of which are shown on Fig. 11. We obtain a mean value: $\langle p_{out} \rangle = (.490\pm.025)$ GeV/c, in very good agreement with other experiments (refs. 15, 16, 17), despite the fact that their "trigger plane" is somewhat differently defined.

Let us next consider the balance of \textbf{p}_{T} for the charged particles in jet events. We will call the net $p_{\mathbf{p}}$ component in the "trigger" plane p_{in} (see Fig. 11 for definitions). Considering only the 76 jet events for which Etot>14.GeV (in order to minimize the influence of the unknown neutral particles: $\sqrt{s}=16.6$ GeV in our experiment), and averaging the P_L (longitudinal momentum) and the p_{in} of each cluster of these events, we obtain the mean representation of these events given in fig. 12. It is easily seen that although the longitudinal component is well balanced, the pin component is not, despite the fact that the so-called spectator jets (beam and target jets) try to balance that component. It has been suggested that such a momentum imbalance would be the manifestation of an internal transverse momentum of the quarks inside hadrons (see for example ref. 15).

Turning to some internal properties of these jets, their charged multiplicity distribution is shown in Fig. 13(a). The mean charged multiplicity <k>=3.46±.17 is higher than the value of 2.74 obtained for all clusters, and the distribution is narrower than a Poisson distribution (σ^2 =2.05±.10).

Figure 13c shows the distribution in relative rapidity $\Delta y=|y-y^{jet}|$ of particles inside and outside jets. Because d₂ is already a rapidity difference, we also give this same distribution obtained by CAH/d₃ in fig. 13d. The remarkable similarity between these two distributions allows us to say that the very sharp spike near $\Delta y=0$ for particles belonging to the jet is not due to a particular choice of metric. It is in agreement with the expected correlation in rapidity between the first evidence for jets in hadron-hadron collisions)¹⁵.

In order to compare these jets with what would be expected if they resulted from the dressing of a single parton, it is of interest to study the momentum components of jet-particles along and perpendicular to the jet-axis. In fig. 14a, our experimental distribution of the fractional longitudinal momenta of jet-particles, $z=\vec{p} \cdot \vec{p}_{jet}/|p_{jet}|^2$, is compared to other experiments. The low values we find at small z are probably due to experimental or methodological loss of low-momentum particles (see our previous discussion on the ability of the algorithms to find fake jets), and the rest of the distribution is in good agreement with the results for jets produced either by lepton processes or in hadron-The slope is in excellent agreement although hadron collisions. our normalization suffers from the already noted bias due to our lack of neutral information.

The distribution of q_t , the transverse momentum component of jet particles, perpendicular to the jet axis, is shown in Fig.14(b). The exponential distribution has a mean value $\langle q_{m} \rangle = (.28 \pm .02)$ GeV/c, which is very good agreement with

-19-

the values obtained at PETRA (ref. 18).

VIII. Conclusions

Using purely statistical algorithms and two very different in order to find multiparticle correlations in a metrics $\pi^{-}p$ experiment performed with the Fermilab hybrid system at 147 GeV/c, clusters are found which exhibit a remarkable stability with respect to choice of distance and/or algorithm. This observation has led us to study the properties of these clusters, especially those with enough \textbf{p}_{m} to be equivalent to the high $\textbf{p}_{\boldsymbol{\psi}}$ jets observed in other experiments. Some of the clusters do, in fact, have properties which are very similar to those of high \textbf{p}_{π} jets seen in counter experiments and their observed structure is consistent with the assumption that they may originate from scattered hadronic constituents (quarks). However, our work with Monte-Carlo simulations indicates that the high $\textbf{p}_{\boldsymbol{\pi}}$ clusters we see in our data could also be explained as the result of statistical fluctuations. The observation of such clusters with a p_{π} above 5 GeV/c, which could be achieved by better statistics or higher energy and/or more information about neutral particles, should permit one to distinguish clearly between the above interpretations.

This work was supported in part by the U. S. Department of Energy and the National Science Foundation. We gratefully acknowledge the efforts of the 30-inch bubble chamber crew and the scanning and measuring personnel at the participating institutions.

-20-

Footnotes and References

a) On leave of absence from Tel-Aviv University, Ramat-Aviv, Israel.
b) On leave of absence from Lab. de Physique Corpusculaire, Collège de France, Paris, France.

- c) Permanent address: Seton Hall University, South Orange, N.J.
- d) Present address: Brookhaven National Laboratory, Upton, Long Island, New York.
 - 1. See, for example,

R. Diebold, Proceedings of 19th International Conference on High Energy Physics, Tokyo, 1978, and referencestherein.
2. M. Jacob and P. V. Landshoff, Nucl. Phys. <u>B13</u>, 395 (1976).
3. S. Nielsen, Ph. D. Thesis, Copenhagen, August 1978

(Niels Bohr Institute) unpublished.

See also: M. Deutschmann et al., Nucl. Phys. <u>B155</u>, 307 (1979) where a simple algorithm is used to produce multiparticle systems. The CCHK (ISR) and PLUTO (PETRA) collaborations are also using algorithms with a distance between particles for analyzing their data (private communications). Since July 1980, different papers have presented jet analyses using clustering algorithms. For example:

J. B. Babcock and R. E. Cutkosky, Nucl. Phys. <u>B176</u>, 113 (1980).
J. Dorfan, SLAC-PUB-2623 (1980.

K. Lanius, H. E. Roloff, H. Schiller, "Selection of Jets in Multihadron Final States Produced by e⁺e⁻ Annihilation", preprint from the Institute for High Energy Physics of the Akademie der Wissenschaften der DDR (Berline DDR).

Footnotes and References (cont'd)

- 4. T. Ludlam, Yale Report No. C00-3075-210 (March 1978).
- 5. J. P. Benzecri et al., "L'algorithme de Classification Ascendante Hierarchique." Internal report (unpublished) of the Laboratoire de Statistique Mathematique, Universite P. & M. Curie, Paris (1975).

The first idea of such an algorithm applied in high energy physics can be found in: T. Ludlam and R. Slansky, Phys. Rev. <u>D16</u>, 100 (1977).

- 6. D. Fong et al., Nucl. Phys. B102, 386 (1976)
- 7. I. A. Pless et al., Inclusive and Semi-Inclusive Charge Structure in π^-p Multiparticle Production at 147 GeV/c, contribution to Tokyo Conference (1978) (unpublished).
- 8. D. Fong et al., Phys. Lett. 53B, 290 (1974).
- 9. F. Barreiro et al., submitted to Phys. Rev. Lett.
- 10. C. Bromberg et al., Comparison of Hadron Jets Produced by π^- and p Beams on Hydrogen and Aluminum Targets (Fermilab preprint 7000.260).
- 11. R. P. Feynman, R. D. Field and G. C. Fox, Nucl. Phys. B128, 1(1977).
- 12. This percentage is very low, but this is because our experiment suffers from the following bias. If the neutral particles carry 50% or more of the total p_T of a jet, we are not able to say, using only the charged

Footnotes and References (cont'd)

information, if there is a high p_T jet or not. We can see high p_T clusters only if the fraction of p_T carried by the neutrals is small. This kind of bias is similar to the well known "trigger bias" in counter experiments (see ref. 2).

- R. D. Field and R. P. Feynman, Nucl. Phys. <u>Bl36</u>, 1 (1978).
 We take this opportunity to thank Professor Field for the permission to use his program.
- 14. R. D. Field and R. P. Feynman, Phys. Rev. D15, 2590 (1977).
- 15. M. Della Negra et al., Nucl. Phys. <u>B127</u>, 1 (1977).

<pout>=(.53 ± .02) GeV/c (CCHK).

- 16. Ref. 2. $\langle p_{out} \rangle = (.501 + .004) \text{ GeV/c}$ (BFS).
- 17. Ref. 10. =(.48+.01) GeV/c (E260).
- 18. M. Della Negra, Rapporteur's talk at Budapest Conference (1977).
- 19. B. Wiik, Rapporteur's talk at Madison Conference (July 1980).

Appendix 1

Comparison Between Distances d_2 and d_3 .

Let us take three pions, the first one with $p_1 = 10m_{\pi}$ and the two others with $p_2=p_3=m_{\pi}$, symmetric (angle α) with respect to the first one.



It is easy to compute the different distances between these particles and it turns out no matter what value of α is chosen, that:

$$d_{2}(1, 2) > d_{2}(2, 3)$$

and d₃ (1, 2) < d₃ (2, 3) (trivial)

This illustrates the different character of the two metrics.

- Table 1: Allowed values of DCUT for each algorithm/distance combination.
- Table 2: Comparison of the results obtained with different DCUT values with the experimental results in ref. 7, 8, and 9.
- Table 3: Average number of clusters per event. A good approximation to these results is <NCL>=1. + .25n_{ch}.
- Table 4: Study of events containing a Δ^{++} , defined by 1.12 < $M_{p\pi}$ + < 1.32 and $|t_{p \rightarrow p\pi}^{+}+| \leq 1$.
- Table 5: Comparison of the results obtained by pairs of algorithm/distance combinations, on an event by event basis: a) percentage of events with exactly the same clustering in the two algorithms; b) percentage of clusters found to be identical in the two algorithms; c) to f) same as b, but for different ranges in the energy of the clusters (E_{cl}) : c) $E_{cl} < 1 \text{ GeV}$, d) $1 < E_{cl} < 4 \text{ GeV}$, e) $4 < E_{cl} < 7 \text{ GeV}$, f) $E_{cl} > 7 \text{ GeV}$; g) percentage of clusters having a $p_{T} > 1.5 \text{ GeV/c}$, found to be either identical or different by only one slow particle (p < .3 GeV/c) in the two algorithms. Efficiency of the clustering algorithms for finding Table 6: jets generated according to Field-Feynman.13 ε_1 = number of jets correctly found by the clustering algorithm

number of jets generated

ε1

ε₂ = number of jets "nearly" found by the clustering algorithm + number of jets generated

TABLE CAPTIONS (cont'd)

"Nearly" found means that all particles (of the jet) missing from the cluster have a fractional longitudinal momentum z less than .14.

- Table 7: Efficiency of the clustering algorithms for finding a jet (see table 6) hidden in a real event. a) The fake jet was found correctly. b) The four vector difference between the jet and the high p_T cluster found has an absolue value less than .5 GeV ||Δ|| < .5 GeV. c) The angle between the jet axis and the axis of the cluster found is less than .12 rad (6.9°). d) The multiplicity of the cluster found is equal to or differs only by one unit from the multiplicity of the jet.</p>
- Table 8: General "jet" characteristics in π p interactions at 147 GeV/c.
- Table 9: The experimental results for the average charge <Q> and the charge ratio π^{-}/π^{+} for jets at three values of the center of mass production angle, θ^{*} . The Field-Feynman¹⁴ prediction for π^{-}/π^{+} is also given.

Algorithm/distance	DCUTmin	DCUThigh	Used value
1/ MST /d1/	does not fit	at all	excluded
2/ CAH /dl/	cannot fit d	ata of ref. 8	excluded
3/ MST /d2/	can only fit	data of ref. 8	excluded
4/ CAH /d2/	4.5	4.7	4.7
5/ MST /d3/	1.10(63°)	1.15(66°)	1.15(66°)
6/ CAH /d3/	1.55(89°)	1.8(103°)	1.75(100°)

Table l

		publ ished papers	MST/d3/1.0 (57°)	MST/d3/1.2 (69°)	CAH/d2/4.5	CAH/d2/4.7	CAH/d3/1.5 (86°)	CAH/d3/1.8 (103°)
^σ NCL≥3 (centra	3(mb) al fireball)	[Ref. 7] 12.9	13.6	11.2	13.0	11.9	13.9	12.2
Leadi (mł	ing part.	[Ref. 8]						
	P	.93 <u>+</u> .12	1.11	1.05	.95	.93	.99	.94
N=4	} π-	.91 <u>+</u> .16	.83	.79	1.08	1.05	.76	.69
	Sum	1.84± .20	1.95	1.84	2.03	1.98	1.75	1.63
	(P	.40 <u>+</u> .13	.62	.52	.41	.38	.45	.38
N=6	} π-	.18 <u>+</u> .09	.25	.22	.44	.41	.19	.16
	Sum	.58 <u>+</u> .16	.87	.74	.85	.79	.64	.54
2π ⁺ 2π	evts.	[Ref. 9)						
# evt	s. $(x_{\pi} - <.93)$	633	630	630	630	630	630	630
# evt 3 pa	s.NCL=2 art./l part.		267	312	284	300	279	317
#evts	."A" region	275	246	273	279	290	254	277
Percen mass diffra	ntage of off shell action	(43.4 <u>+</u> 2.6)%	39.1%	43.3%	44.3%	46.0%	40.3%	44.0%

(

Topology (number of charged part.)	CAH/d2/4.7	CAH/d3/1.75	MST/d3/1.15
4	2.07 <u>+</u> .06	2.03 <u>+</u> .06	2.17+.06
6	2.45+.07	2.52 <u>+</u> .07	2.58+.07
8	2.87 <u>+</u> .09	3.09 <u>+</u> .10	2.93 <u>+</u> .09
10	3.35+.13	3.65 <u>+</u> .14	3.35+.13
12	3.71 <u>+</u> .23	4.20 <u>+</u> .26	3.71+.23
14	4.1 9 <u>+</u> .43	4.76+.49	4.41 <u>+</u> .45
16	4.50 <u>+</u> .92	5.46 <u>+</u> 1.11	4.92 <u>+</u> 1.00
· · · · · · · · · · · · · · · · · · ·			
Inclusive	2.86 <u>+</u> .04	3.07 <u>+</u> .04	2.95+.04

	САН	/d2/4.7	CAH/d3/100°		
	Nber of evts	cross-section/ (µb)	Nber of evts	cross-section (µb)	
Direct A ⁺⁺	108	390. <u>+</u> 37.	88	315.+34.	
in a $(p2\pi)_{cl}$	180	638. <u>+</u> 47.	189	660.+48.	
in a (p3π) _{cl}	64	290. <u>+</u> 36.	70	294 <u>+</u> 35.	
in a (p4π) _{cl}	26	112. <u>+</u> 22.	29	143. <u>+</u> 27.	
in a (p>5π) _{cl}	8	45. <u>+</u> 16.	13	77. <u>+</u> 21.	
Total	386	1,475. <u>+</u> 75.	389	1,490. <u>+</u> 75.	
p and π^+ in two diff. cl.	3	15. <u>+</u> 9.	0	-	

	A (%)	B (%)	C (१)	D (%)	E (%)	F (%)	G (%)
CAH/d ₂ & MST/d ₃	34.0 <u>+</u> .5	48.5 <u>+</u> .6	25.2 <u>+</u> 1.1	45.6 <u>+</u> .9	51.9 <u>+</u> 1.3	72.5 <u>+</u> 1.9	84.6 <u>+</u> 2.1
CAH/d ₂ & CAH/d ₃	34.0 <u>+</u> .5	48.0 <u>+</u> .6	27.0 <u>+</u> 1.2	43.4 <u>+</u> .9	53.7 <u>+</u> 1.3	72.6 <u>+</u> 1.9	91.7 <u>+</u> 2.2
MST/d ₃ & CAH/d ₃	55.0 <u>+</u> .8	65.6 <u>+</u> .7	55.4 <u>+</u> 1.4	62.0 <u>+</u> 1.1	68.8 <u>+</u> 1.5	82.5 <u>+</u> 1.9	85.6 <u>+</u> 2.1

)

Efficiency	٤l	² ء
CAH/d2/4.7	.73 <u>+</u> .02	.90 <u>+</u> .02
CAH/d3/1.75	.54 <u>+</u> .01	.90 <u>+</u> .02
MST/d3/1.15	.56 <u>+</u> .01	.88 <u>+</u> .02
three algo. together	.41 <u>+</u> .01	.70 <u>+</u> .02

Table 6

		CAH/d2/4.7	CAH/d3/1.75	MST/d ₃ /1.15
A cluster with p _T >1.5 GeV/c is :	Eound	(93.2 <u>+</u> 4.1)%	(92.7 <u>+</u> 3.7)%	(99.8 <u>+</u> 3.9)%
A: jet exactly found		(42.8 <u>+</u> 2.8)%	(56.5 <u>+</u> 2.9)%	(39.8+2.5)%
the found	В	(34.2 <u>+</u> 2.5)%	(26.6 <u>+</u> 2.0)%	(37.0 <u>+</u> 2.4)%
cluster is near	{ c	(34.1 <u>+</u> 2.5)%	(23.3+1.9)%	(32.7 <u>+</u> 2.3)%
the jet	D	(34.2 <u>+</u> 2.5)%	(33.8 <u>+</u> 2.2)%	(51.3+2.8)%
fficiency (A+B)		(77.0 <u>+</u> 3.7)%	(83.1 <u>+</u> 3.5)%	(76.8 <u>+</u> 3.5)%

 $\sigma_{Tot} (\pi^- p \rightarrow jet + x) \qquad (2.24\pm.50) \text{ mb}$ $^{o}_{Tot} (\pi^- p \rightarrow jet + x) \qquad 1.81 \text{ GeV/c} (<x_T >=.218)$ Total number of jet-events 352 (7.5% of all events)
Number of events with one jet 313
Number of events with more than one jet 39

jet = cluster with $p_T^{>1.5} \text{ GeV/c}$

Topology of (n _{ch}) event	4	6	8	10	12	14	16
jet/event (%)	0.8	3.4	8.9	16.3	21.4	32.3	33.3
Total number of clusters per event (NCL)	2	3	4	<u>></u> 5			
number of jet events	29	181	107	31			
percentage of jet events among all events having that NCL	1.4%	9.9%	18.5%	30.4%			

Table 8: General jet characteristics in $\pi^-p \rightarrow jet + x$ at 147 GeV/c

Θ*	45°	90°	135°
<q>jet</q>	36 <u>+</u> .04	15 <u>+</u> .02	.26 <u>+</u> .03
$\frac{\pi}{\pi^+}$ (exp.)	2.10 <u>+</u> .25	1.35 <u>+</u> .15	.59 <u>+</u> .07
$-\frac{\pi}{\pi^+}$ (ref. [14])	2.25	1.18	.52

Table 9

Figure Captions

Fig. 1: Principle of the classification algorithms.

a) a two-dimensional event

b) The minimal spanning tree for that event: (the link between points 6 and 4 has been broken because neither of these two is the nearest neighbor of the other.)
As long as DCUT is greater than 1.7 (the length of the longest link) the MST algorithm gives two clusters (1-2-5-6 and 3-4-7).

c) CAH algorithm for the same event, the abscissa gives the labels of the points and the ordinate, the distance at which points or subsets are merged. For a DCUT between 2.2 and 4.5 two clusters (identical to the ones of MST) are obtained.

- Fig. 2: Distribution of the number of clusters per event (NCL) as a function of final-state prong number. (Solid line: CAH/d2/4.7, dashed line: CAH/d3/1.75, dot-dashed line: MST/d3/1.15). Lines connecting the data points are given rather than points for clarity and because the small uncertainties permit this representation.
- Fig. 3: Charged particle multiplicities of the groups formed by the three algorithm/distance combinations.

Solid line: $CAH/d_2/4.7 < k > = 2.74 \pm .04 \sigma_k^2 = 1.79$ Dashed line: $CAH/d_3/1.74 < k > = 2.67 \pm .04 \sigma_k^2 = 1.86$ Dot-dashed line: $MST/d_3/1.15 < k > = 2.56 \pm .04 \sigma_k^2 = 1.28$

Lines connecting the data points are given rather than the points for clarity and because the small uncertainties permit this representation.

Fig. 4: Invariant mass distributions of 2π and 3π groups in 147 GeV/c π -p data for three algorithm/distance combinations. It should be noticed that no p signal is seen in the 2π mass spectra.

> Solid line: $CAH/d_2/4.7$, $\langle M \rangle_{2\pi} = .55 \text{ GeV/c}^2$, $\langle M \rangle_{3\pi} = .98 \text{ GeV/c}^2$, Dashed line: $MST/d_3/1.15$, $\langle M \rangle_{2\pi} = .53 \text{ GeV/c}^2$, $\langle M \rangle_{3\pi} = .98 \text{ GeV/c}^2$, Dotted line: $CAH/d_3/1.75$, $\langle M \rangle_{2\pi} = .62 \text{ GeV/c}^2$, $\langle M \rangle_{3\pi} = 1.10 \text{ GeV/c}^2$ Invariant $\pi^+\pi^-$ mass distributions when both π 's belong to the

same forward (Y>1.) and negatively charged cluster, in twocluster events. a) CAH/d₂/4.7. b) CAH/d₃/1.75.

Fig. 5:

Dashed histograms: statistics restricted to such pairs in $(3\pi)^-$ clusters.

Fig. 6: Invariant cross section as a function of p_T , both for inclusive $\pi^{\pm}(+)$ and for clusters obtained by our algorithm/ distance combinations ($\Delta=MST/d_3/1.15$, $\bullet=CAH/d_2/4.70$, $x=CAH/d_3/1.75$) integrated over the whole rapidity range. Also shown are results of exponential fits:

> solid line: $747\exp(-(5.79\pm.05)p_{T})$ fitted between 0.and 1.5 Gev/c. dashed line: 136.6 $\exp(-(3.53\pm.08)p_{T})$ fitted between 1 and 3 GeV/c.

Fig. 7: Invariant cross section at 90° as a function of P_{T} , for clusters in our experiment (+) in the rapidity range $|Y| \leq .5$, and jets in a calorimeter experiment (•) (E260 ref. 10). The solid line is the prediction for jets produced with a π^{-} beam at our energy from Feynman, Field and Fox (ref. 11).

Fig. 8: Invariant cross section of clusters as a function of p_T in our experiment. The lines are the results of three different Monte Carlo calculations explained in the text (the numbers at the bottom of the lines refer to the paragraph where the corresponding model is described).

- Fig. 9: The ability of the clustering algorithms to find Field-Feynman jets¹³ as a function of the internal charged multiplicity of these jets. ε_1 have the same meaning as in Table 6. Solid line: CAH/d₂/4.7, dotted line: CAH/d₃/1.75 and dashed line: MST/d₃/1.15.
- Fig. 10: Cross section (top) and mean charge (bottom) of the jets as a function of the center of mass production angle of the jet.
- Fig. 11: a) The "trigger" plane is the plane defined by the c.m. momenta of the beam and the jet. The transverse momentum of any particle is decomposed into two components, one perpendicular to the "trigger plane"(pout) and one in the "trigger" plane (pin).

$$P_{out} = \frac{|(p_T^{jet} \times p_1) \cdot p|}{|p_T^{jet}| \cdot |p_1|} , \qquad p_{in} = \frac{p_T^{jet} \cdot p}{|p_T^{jet}|}$$

where p_1 is the momentum of the incoming beam, p_T^{jet} the transverse momentum of the jet and p is the momentum of the particle considered.

b) Distribution of the p_{out} component, for all particles outside the jet.

- Fig. 12: Momentum balance in "jet" events. These drawings represent an average of p_L , p_{in} and p_{out} computed for every cluster in each event for events with $E_{Tot}^{ch} > 14$ GeV. The figure represents these mean values, in the "trigger" plane: a) average over all 76 events satisfying the above criteria and b) same as a), but only for the 12 events containing two high p_{rr} "jets".
- Fig. 13: Correlations between "jet" particles. a) Charged multiplicity of the "jets" found by $CAH/d_2/4.7$: <k>=3.46; σ_K^2 =2.05, b) same for $CAH/d_3/1.75$ <k>=3.20; σ_k^2 =1.08, c) correlations in rapidity: $\Delta y = |y_{jet} - y|$, for particles belonging to the "jet" (•), and not belonging to the "jet"(x), for $CAH/d_2/4.7$, and d) same as c) but for $CAH/d_3/1.75$.
- Fig. 14: Quark fragmentation functions. Inclusive distributions of charged "jet" fragments as a function of a) z, the longitudinal fraction of "jet" momentum. (x) this experiment (our experimental results have been joined in order to guide the eye), (\bullet): pp \rightarrow jet + x (E260), (Δ) e⁺e⁻ \rightarrow jet + jet, and (+) $\nu \rho \rightarrow \mu^{-}$ +jet (data for other experiments have been taken from ref. 18; b) q_{T} , transverse momentum with respect to the jet axis. $\langle q_{T} \rangle = 1$ (.28 ± .02) GeV/c.





















Figure 4





Figure 6





Figure 8



Figure 9



. .









Figure 14

 \checkmark