

DESY 80/36  
May 1980



Klustermethoden zur Jet-Analyse\*

von

K. Lanius

*Institut für Hochenergiephysik  
der Akademie der Wissenschaften der DDR  
Berlin - Zeuthen*

**DESY behält sich alle Rechte für den Fall der Schutzrechtserteilung und für die wirtschaftliche Verwertung der in diesem Bericht enthaltenen Informationen vor.**

**DESY reserves all rights for commercial use of information included in this report, especially in case of apply for or grant of patents.**

**To be sure that your preprints are promptly included in the  
HIGH ENERGY PHYSICS INDEX ,  
send them to the following address ( if possible by air mail ) :**

**DESY  
Bibliothek  
Notkestrasse 85  
2 Hamburg 52  
Germany**

## 1. Einleitung

In vielen Situationen steht man vor dem Problem, eine Gruppe von Objekten (Lebewesen, Dokumente, Wechselwirkungsprozesse, ...) in Untergruppen - Kluster - zu unterteilen, wobei die Klassifikation auf der Ähnlichkeit der Objekte beruht. Objekte eines Klusters sollen untereinander "ähnlich" sein, während die Kluster voneinander "verschieden" sind. Eine wesentliche Schwierigkeit der Clusteranalyse besteht darin, den Begriffen "ähnlich" und "verschieden" einen dem jeweiligen Problem angepassten Sinn zu geben.

Klusteranalysen sind mathematische und statistische Methoden, deren Ziel der Nachweis von Clustern, d.h. Objektballungen in im allgemeinen hochdimensionalen Räumen ist.

Entsprechend dem unterschiedlichen Charakter der zu untersuchenden Probleme gibt es viele Algorithmen zur Auffindung von Clustern. Eine Klasse von Methoden beruht auf der Annahme, daß die Ähnlichkeit zweier Objekte (Objektgruppen) durch ein Ähnlichkeitsmaß oder einen Abstand, also durch einen numerischen Wert beschrieben werden kann. Das Ähnlichkeitsmaß ist hierbei die einzige Information, die zur Konstruktion der Cluster benutzt wird.

Anfang der siebziger Jahre wurden die ersten Versuche zur Anwendung von Klusteralgorithmen in der Hochenergiephysik unternommen /1/. Seitdem wurden umfangreiche Untersuchungen an verschiedenen Prozessen mit unterschiedlichen Algorithmen durchgeführt. Bei den verschiedenen Anwendungen ist zwischen Clustern von Ereignissen in Reaktionskanälen und Clustern von Spuren in Ereignissen zu unterscheiden. Einen aktuellen

## Klustermethoden zur Jet-Analyse\*

K. Lanus

Institut für Hochenergiephysik  
der Akademie der Wissenschaften der DDR  
Berlin - Zeuthen

\* Vortrag im DESY-Kolloquium am 15. April 1980

Überblick vermitteln die Proceedings des "3rd Topical Meeting on Multi-Dimensional Analysis of High Energy Data", Nijmegen, März 1978 /2/.

Phänomenologische Untersuchungen hadronischer Reaktionen suchen in der Regel zunächst eine Antwort auf die Frage zu finden, welche Unterkanäle zum betrachteten Reaktionskanal beitragen. Ausgehend von der Erfahrung, daß dynamische Mechanismen an bestimmte Bereiche des Phasenraumes gebunden sind, sollen durch entsprechende "Schnitte" die Unterkanäle identifiziert werden.

Mit steigendem Primärimpuls wächst die mittlere Multiplizität und damit auch die Informationsmenge pro Ereignis. Mit der Ausdehnung des Phasenraumes und der größeren kombinatorischen Vielfalt können mehr Unterkanäle zu einem Reaktionskanal beitragen. Die bei niedrigen Multiplizitäten noch brauchbaren ein- und zweidimensionalen Projektionen wie effektive Massenspektren, t-Verteilungen und Dalitzplots werden wegen der Reflexionen konkurrierender Unterkanäle immer weniger anwendbar.

Multidimensionale Analysen mit Klusteralgorithmen sind weitestgehend frei von speziellen Annahmen über die Dynamik der Prozesse. Eine Identifizierung der gefundenen Kluster mit der physikalischen Wirklichkeit liegt außerhalb der Methode.

Ein Ereignis einer Reaktion

$$a + b \rightarrow 1 + 2 + \dots + n$$

wird bei fester Energie und unpolarisierten Teilchen im Anfangszustand durch  $3n - 5$  unabhängige kinematische Variable  $\vec{x}$  beschrieben. Die Ergebnisse der Reaktion bilden im Raum der

kinematischen Variablen eine Punktverteilung, wobei die Grenzen des Phasenraumes durch die Erhaltung von Energie und Impuls gegeben sind.

Innerhalb des Phasenraumes läßt sich die Dichteverteilung der Punkte durch

$$\frac{d\sigma}{d\vec{x}} = |\mathcal{T}(\vec{x})|^2 \rho(\vec{x})$$

schreiben.  $\rho(\vec{x})$  ist ein kinematischer Faktor, der von der Wahl der Variablen  $\vec{x}$  abhängt und  $\mathcal{T}(\vec{x})$  die Streuamplitude. Da im allgemeinen der Endzustand durch  $m$  verschiedene Unterkanäle gebildet wird, ist

$$|\mathcal{T}(\vec{x})|^2 = \left| \sum_{i=1}^m \mathcal{T}_i(\vec{x}) \right|^2$$

Im Idealfall verschwindender Interferenzglieder wird daraus

$$|\mathcal{T}(x)|^2 = \sum_{i=1}^m |\mathcal{T}_i(x)|^2$$

Jeder der Unterkanäle führt zu einer oder mehreren in der Regel nicht überlappenden Punkthäufungen im Phasenraum.

Bei nicht sehr großen Energien tritt der Idealfall wegen der vorhandenen Interferenzeffekte nicht auf. Es zeigt sich jedoch, daß die Überlappung der Kluster in vielen Fällen nicht groß ist und konstruktive Interferenzen zu eigenen Klustern führen. Ferner zeigt die Erfahrung, daß Unterkanäle mit kleinem Wirkungsquerschnitt erst nach der Eliminierung der vorherrschenden Subkanäle sichtbar werden. In verschiedenen Fällen wurden die durch die multidimensionalen Analyse gewonnenen Kluster zum Ausgangspunkt weiterer Analysen wie etwa einer Partialwellenzerlegung.

Die Existenz von Klustern von Ereignissen (bzw. Spuren) wurde in einem großen Energiebereich mittels unterschiedlicher Methoden bei einer größeren Zahl von Reaktionen nachgewiesen.

Bei der Anwendung in der Hochenergiephysik haben vor allem zwei Klusteralgorithmen Eingang gefunden.

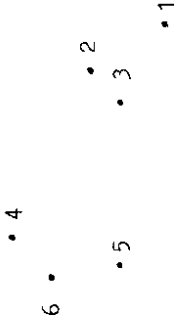
### 2. Hierarchische Methoden

Die Objekte liegen als Endpunkte - Knoten - in einer baumartigen Struktur. Knoten in verschiedenen Ebenen entsprechen jeweils einer möglichen Zuordnung, wobei eine höhere Ebene einer weiteren Zusammenfassung entspricht.

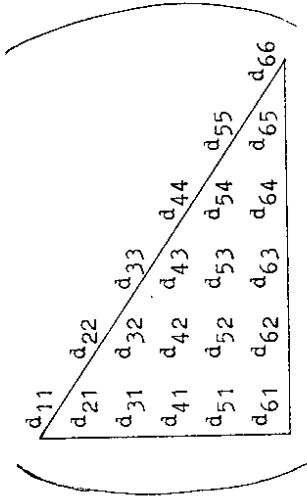
Die Klassifizierung geht von der Ähnlichkeits- oder Abstandsmatrix aus. Bei N Punkten hat diese Matrix  $N(N-1)/2$  Elemente. Wegen des rechen-technischen Aufwandes, der in der Regel mit  $N^2$  wächst, ist eine Anwendung auf große Mengen ( $\geq 10\ 000$  Objekte) nicht möglich.

Ein spezielles hierarchisches Verfahren ist die Verbindungsmethode der nächsten Nachbarn (Minimum Spanning Tree Method = MST-Methode). Die N Punkte werden als Knoten eines Verbindungsgraphen aufgefaßt. Der Graph hat  $N-1$  Kanten, so daß es für zwei Knoten nur einen Verbindungsweg gibt. Den Kanten wird ein Abstandsmaß zwischen den Knoten zugeordnet und der Verbindungsgraph konstruiert, der der kleinsten Abstandssumme entspricht (MST). Trennt man den Graph durch Entfernen der jeweils längsten Kante, so erhält man eine Hierarchie von Klustern.

Betrachten wir zur Illustration der hierarchischen Verfahren ein einfaches zweidimensionales Beispiel mit  $N = 6$  Objekten:

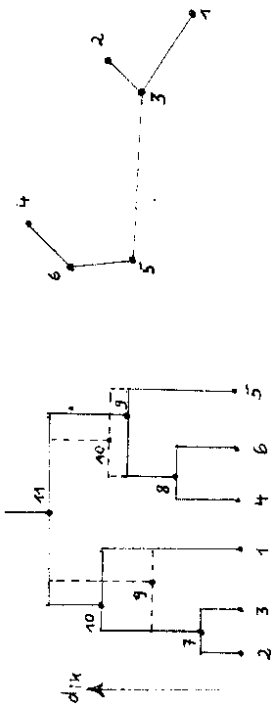


Die Abstandsmatrix bzw. die Ähnlichkeitsmatrix der 6 Objekte hat die Form



mit 15 Elementen, da  $d_{ij} = d_{ji}$  und  $d_{ii} = 0$  ist. (In der Ähnlichkeitsmatrix haben die Diagonalelemente den Wert  $d_{ii} = 1$ .)

Auf der rechten Seite der folgenden Skizze ist der MST mit den 5 Kanten gezeichnet, für den die Abstandssumme ein Minimum ist. Auf der linken Seite der Skizze ist die Zusammenfassung der Objekte auf verschiedenen Ebenen gezeigt.



Zunächst fügen wir die Objekte 2 und 3 mit dem kleinsten Abstand  $d_{32}$  in einer Gruppierung 7 zusammen, dann die nächstbenachbarten Objekte 4 und 6 zu 8 und diese Untergruppe mit 5 zu 9 usw.

Die Entscheidung zur Vereinigung einer Untergruppe mit einem Objekt bzw. zweier Untergruppen läßt sich auf unterschiedliche Weise vollziehen.

- Das Abstandsmaß ist die kleinste Distanz zwischen zwei Objekten der beiden Untergruppen (single-linkage) (ausgezogene Linien).

- Das Abstandsmaß ist die größte Distanz zwischen zwei Objekten der beiden Untergruppen (multiple-linkage) (gestrichelte Linien).

- Durch geeignete Mittelung zwischen den Objekten jeder der Untergruppen bildet man einen durchschnittlichen Abstand (average-linkage).

Das erste Kriterium des single-linkage führt zu einer Klassifikation, die der des MST entspricht. Charakteristisch für diese Methode ist, daß sie zu fadenförmigen Klustern führt.

Das gibt in vielen Anwendungsfällen, in denen andere Clusterformen vorliegen, schlechte Resultate. Nur wenn die Cluster näherungsweise linear in annehmbaren Raumbereichen sind, ist die MST-Methode gerechtfertigt.

T. Ludlam /3/ hat ein hierarchisches Verfahren bei der Analyse von 6000 Ereignissen der  $\pi^-p$ -Wechselwirkungen bei 150 GeV/c ( $n_{ch} \geq 3$ ) benutzt.

Als Abstandsmaß zweier Spuren in jedem Ereignis nahm er

$$d_{ij}^2 = M_{ij}^2 - (m_i + m_j)^2$$

wobei  $M_{ij}$  die effektive Masse des Zweikörpersystems und  $m_i, m_j$  die Ruhmassen der einzelnen Teilchen sind. Als Separationsverfahren wurde die MST-Methode benutzt. Alle Kanten mit  $d_{max} > 1.3 \text{ GeV}/c^2$  wurden geschnitten und damit für jedes Ereignis eine variiierende Zahl von Klustern ( $1 \leq 11$ ) erhalten. Im Durchschnitt liegt die Zahl der Spuren pro Cluster bei 3 und ändert sich nur wenig mit der Multiplizität. Bei der Analyse der Eigenschaften der Cluster ergeben sich einige Hinweise auf ein jetartiges Verhalten.

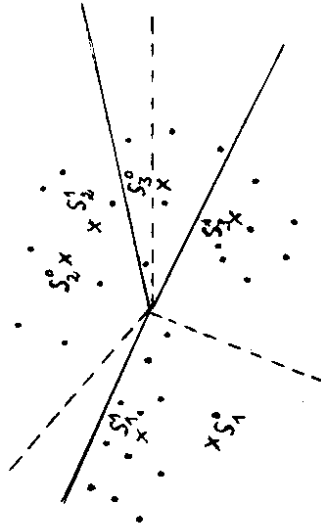
### 3. Nichthierarchische Methoden

Vorausgesetzt, daß Cluster existieren, führen diese Methoden zu einer Klustereinteilung, ohne daß der verwendete Algorithmus eine Hierarchie nacheinander zusammenfügbarer Cluster liefert.

Das Prinzip nichthierarchischer Verfahren sei an einer der in der Hochenergiephysik verwendeten Methoden - dem

Schwerpunktverfahren - charakterisiert. Die N Punkte im multidimensionalen Raum werden willkürlich in M Untergruppen aufgeteilt. Für jede der Untergruppen wird ein Schwerpunkt berechnet. Im zweiten Schritt wird jeder Punkt der Untergruppe zugeordnet, zu deren Schwerpunkt er den kleinsten Abstand hat. Für die damit erhaltenen Untergruppen werden die Schwerpunkte erneut berechnet. Das Verfahren wird bis zum Erreichen einer stabilen Clusterkonfiguration fortgesetzt. Die Methode ist schnell und der Rechenaufwand wächst mit N.

Betrachten wir zur Illustration wieder ein zweidimensionales Beispiel:



Im ersten Schritt wird der punkterfüllte Bereich willkürlich in drei Untergruppen geteilt (ausgezogene Linien). Die drei Punktgruppen führen auf die Schwerpunkte  $S_1^0$ ,  $S_2^0$  und  $S_3^0$ . Im zweiten Schritt werden die Abstände der Punkte zu den drei Schwerpunkten berechnet und eine neue Einteilung in drei Untergruppen erreicht, indem man jeden Punkt in die Gruppe einordnet, zu deren Schwerpunkt er am nächsten

liegt (gestrichelte Linien). Es werden die Schwerpunkte  $S_1^1$ ,  $S_2^1$ ,  $S_3^1$  dieser Untergruppen berechnet usw. Schwerpunktverfahren konvergieren in der Regel nach wenigen Schritten.

Die Erfahrungen in der Anwendung einfacher nichthierarchischer Algorithmen führten W.D. Nowak und H. Schiller /4/ zur Entwicklung eines Algorithmus, der folgenden Kriterien genügt:

- Er arbeitet mit einer variablen Zahl von Untergruppen, d.h. Untergruppen, die nicht mehr "verschieden" sind, werden zu einer vereint bzw. Untergruppen werden geteilt, wenn sich während der Iteration Dichtemaxima zeigen.
- Die Kenntnis typischer Projektionen (Dalitzplot, Prisma-plot) bei Dreiteilchenendzuständen legt die Forderung nahe, daß der Algorithmus zur Auffindung ellipsoidartig geformter Kluster in der Lage sein muß.
- Er soll die Bearbeitung von Reaktionskanälen mit  $10^4 - 10^5$  Ereignissen mit 15 - 18 Variablen gestatten.
- Bei der Messung des Abstandes jedes Punktes, d.h. Ereignisses vom Klustermittelpunkt geht die Kovarianzmatrix der Ereignisdichte als Gewicht ein.

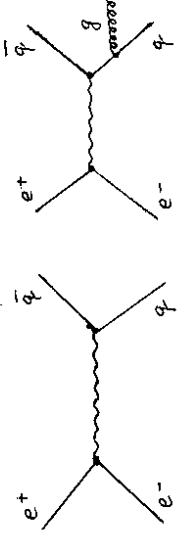
Dieser modellunabhängige Algorithmus wurde bei Verwendung eines lorentzinvarianten Variablensatzes zur Analyse von ca. 19 000 Ereignissen der Reaktion  $\pi^+ + p \rightarrow p + \pi^+ + \pi^-$  und ca. 5000 Ereignissen der Reaktion  $\pi^- + p \rightarrow p + \pi^- + \pi^+$  bei einem Primärimpuls der  $\pi^\pm$ -Mesonen von 16 GeV/c benutzt /5/.

Tabelle 1 vermittelt einen Eindruck, welcher hoher Grad an Übereinstimmung zwischen Klusterinhalt und durch Resonanzen charakterisierte Subkanäle erreichbar ist. Zum Vergleich sind in der Tabelle Resultate zweier unabhängiger Analysen gegeben /6/.

Abb.1 zeigt als Beispiel die effektiven Massenverteilungen und die Zerfallswinkelverteilungen für die Ereignisse in Kluster 9 der Tabelle 1 ( $\pi^+ p \rightarrow \Delta^{++}(1236) + \rho^0$ ). Die Massenverteilungen zeigen klare Resonanzsignale. Die Asymmetrie in  $\cos \theta$  ( $\pi^+ \pi^-$ ) des  $\rho$ -Mesons läßt sich als Interferenz des  $\rho^0$  mit einem verbleibenden ( $\pi^+ \pi^-$ ) - S-Wellen-Untergrund interpretieren.

4. Jet-Topologie in  $e^+e^-$ -Reaktionen

Bei der Untersuchung der topologischen Eigenschaften von Multihadronprozessen, die am Speicherring PETRA beim DESY mit Schwerpunktenenergien bis zu 31.6 GeV erzeugt wurden, beobachtete man neben den vorherrschenden 2-Jetstrukturen der hadronischen Ereignisse auch einige Prozesse, die eine 3-Jetstruktur zeigen (siehe etwa /7/). Die beiden Ereignisgruppen zuordenbaren Graphen sind



Zur Identifizierung von Jets wurden unterschiedliche Variable eingeführt. Häufig benutzte Größen sind Sphericity S und Thrust T

$$S = \frac{3}{2} \frac{\sum_i p_{Ti}^2}{\sum_i p_i^2} ; T = \frac{\sum_i |p_{Ti}|}{\sum_i |\vec{p}_i|}$$

wobei  $\vec{p}_i$  die Impulse der Teilchen und  $p_{Ti}$  ( $p_{Ti}$ ) ihre Transversalimpulse (Longitudinalimpulse) bezüglich der Jet-Achse sind. Die Jet-Achse ist durch die Forderung  $\sum p_{Ti}^2 = \min$ . bzw.  $\sum |p_{Ti}| = \max$ . definiert. Mit wachsender Schwerpunktsenergie wurde die kollineare 2-Jetstruktur der Hadronen sicher nachgewiesen, wobei die Winkelverteilung der Jet-Achsen bezüglich der Richtung der einlaufenden Teilchen einer Verteilung entspricht, wie man sie für Quarks mit dem Spin 1/2 erwartet.

Eine Gluonemission durch Bremsstrahlung der Quarks sollte zu einem Anwachsen der Transversalimpulse der Hadronen bezüglich der Jet-Achse eines der beiden Jets mit wachsender Energie führen. Ist die Energie des emittierten Gluons groß genug, so sollte man in einigen Fällen eine koplanare 3-Jetstruktur beobachten. Die Beobachtungen entsprechen den Vorhersagen der QCD.

Es ist naheliegend, Klustermethoden zur Suche nach Ereignissen mit einer Jetstruktur zu benutzen. Aus der Fülle der Algorithmen wählten H.E. Roloff und H. Schiller /8/ einen hierarchischen Algorithmus aus. Ausgangsinformation ist die Ähnlichkeitsmatrix mit den  $N(N-1)/2$  Elementen. Als Ähnlichkeitsmaß zwischen dem i-ten und dem k-ten Teilchen eines Ereignisses wird definiert:

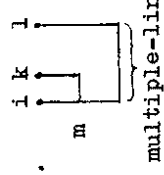
$$d_{ik} = \frac{1}{2} \left( \frac{|\vec{p}_i \cdot \vec{p}_k|}{|\vec{p}_i| |\vec{p}_k|} + 1 \right)$$



mit  $\vec{p}_j$  dem Teilchenimpuls. Diese Ähnlichkeit genügt den Forderungen  $0 \leq d_{ik} \leq 1$ , wobei  $d_{ik} = 0$  für maximal voneinander verschiedene bzw.  $d_{ik} = 1$  für identische Teilchen ist und  $d_{ik} = d_{ki}$ .

Die Wahl der Ähnlichkeit geht von der Vorstellung aus, daß Teilchen eines Jets untereinander kleine Winkel bilden. Es kann auch ein anderes Ähnlichkeitsmaß benutzt werden, das den obigen Forderungen genügt. Die gewählte Größe sollte immer die maximale Kenntnis über das, was als ähnlich anzusehen ist, enthalten.

Es werden jeweils die Teilchen bzw. Teilchengruppierungen miteinander kombiniert, die eine maximale Ähnlichkeit zeigen. Für jede neu entstehende Untergruppe muß anschließend die Ähnlichkeit zu den verbleibenden Teilchen bzw. Teilchengruppierungen neu definiert werden. Die Vereinigung einer Untergruppe mit einer Spur bzw. zweier Untergruppen erfolgt nach dem Kriterium des multiple-linkage.



Um die Ähnlichkeit der Untergruppe m der Teilchen i und k zum Teilchen l zu bestimmen, wird die kleinere der beiden Ähnlichkeiten  $d_{il}$  und  $d_{kl}$  gewählt:

$$d_{lm} = \min. (d_{il}, d_{kl}) ; l \neq i, k$$

Das Verfahren wurde an 90 Multihadron-Ereignissen getestet,

die mit der Anlage TASSO bei einer Schwerpunktsenergie von etwa 30,6 GeV erhalten wurden /9/. Die Zahl der geladenen Teilchen N in den Testereignissen variierte zwischen 5 und 19.

In den Abb.2a und 2b sind die Ähnlichkeitsbäume für zwei typische Ereignisse gezeichnet. Das Ereignis in Abb.2a läßt sich als eine 2-Jetstruktur und das Ereignis in Abb.2b als eine mögliche 3-Jetstruktur identifizieren.

Zur Klassifizierung der Ereignisse in die verschiedenen Untergruppen werden folgende Kriterien verwendet:

- a) Ereignisse mit  $d_{ik} < 0.8$  und  $\Delta(d_{ik}) = d_{i1} - d_{mk} \geq 0.28$
- b) Ereignisse mit  $d_{ik} < 0.68$  und  $\Delta(d_{ik}) \geq 0.36$ .

Das Resultat der Einteilung ist in Abb.3 gezeigt. In den Verteilungen sind Ereignisse schraffiert eingetragen, die nach einer Inspektion der räumlichen Verteilung der Einzelimpulse jedes Ereignisses auf einem interaktiven graphischen Display als 3-Jet-Kandidaten in Frage kommen.

Um ein weiteres Maß für die Gruppierungen auf jeder Stufe im Ähnlichkeitsbaum zu haben, müssen geeignete Größen, etwa ähnlich der Triplicity /10/, definiert werden, die nicht nur auf dem Cosinus des Winkels zwischen Teilchen bzw. Teilchengruppen beruhen.

Jedes Ereignis ist durch die N Impulse  $\vec{p}_1, \vec{p}_2, \dots, \vec{p}_N$  charakterisiert. Bei gegebener Gruppierung der N Teilchen in M disjunkte Gruppen  $K_1, \dots, K_g, \dots, K_M$  enthält jede der nicht überlappenden Gruppen  $m_j$  Teilchen mit

$$K_g = \left\{ \vec{p}_1^g, \vec{p}_2^g, \dots, \vec{p}_{m_j}^g \right\}.$$

Als Verallgemeinerung der Triplicity wird eine "M-Ticity" durch

$$T_M = \frac{\sum_{i=1}^m \left| \sum_{j=1}^m \vec{p}_i^g \right|}{\sum_{l=1}^N |\vec{p}_l|}$$

definiert.  $T_M$  variiert zwischen 0, wenn alle Teilchen in nur einer Gruppe sind und  $\sum_{i=1}^N \vec{p}_i = 0$  ist, und 1, wenn jedes Teilchen seine eigene Gruppe bildet.

In Abb. 4 ist die Variation von  $T_M$  mit der Zahl der Kluster  $M$  für die beiden in den Abb. 2a und 2b gezeigten Ereignisse skizziert.

Charakteristisch ist in beiden Fällen, daß mit Erreichen der "wahren" Klusterzahl die M-Ticity sich sprunghaft 1 nähert. Beide Beispiele zeigen auch, daß die Impulsbilanz, für den Grenzfall alle Teilchen in einer Gruppe, wegen des Fehlens der neutralen Komponente nicht aufgeht. In den beiden Beispielen ist der Grad der Nichterfüllung mit  $T_M$  (1 Kluster)  $< 0.2$  relativ gering.

In Abb. 5 ist der Ähnlichkeitsbaum für ein Ereignis gezeichnet, das nach den Kriterien a) und b) ein 3-Jet-Kandidat sein sollte. Die Inspektion am Display läßt es jedoch als einen 2-Jet-Kandidaten erscheinen. Abb. 6 zeigt die M-Ticity für dieses Ereignis. Die Impulsbilanz ist mit  $T_M$  (1 Kluster) = 0.32 relativ schlecht und  $T_M$  ist bereits für  $M = 2$  nahe 1.

Letzteres unterstützt die Interpretation als ein 2-Jet-Ereignis.

Zur Unterscheidung zwischen 2-Jet- und 3-Jet-Kandidaten führen wir als Maß für die Änderung der M-Ticity die Größe

$$D = T_3 - T_2$$

ein, wobei Ereignisse mit sehr schlechter Impulsbilanz der Geladenen Teilchen durch die Zusatzbedingung  $0 < T_1 < 0.49$  ausgeschlossen werden.

Die D-Verteilung der verbleibenden Ereignisse ist in Abb. 7 gezeigt.

Um die Zahl der visuell am Display ermittelten 3-Jet-Kandidaten weiter einzuschränken, wurde das von Wu und Zobernig /11/ vorgeschlagene Verfahren benutzt.

Zur Analyse der Form eines Ereignisses läßt sich aus den Impulsen  $\vec{p}_i$  ( $i = 1 \dots N$ ) der  $N$  geladenen Teilchen eine symmetrische  $3 \times 3$  Matrix bilden

$$M_{\alpha\beta} = \sum_{i=1}^N p_{i\alpha} p_{i\beta} ; (\alpha, \beta = x, y, z).$$

Die Einheitsvektoren  $\hat{n}_1, \hat{n}_2, \hat{n}_3$  des Impulsensors  $M$  sind mit dessen Eigenwerten  $\Lambda_1, \Lambda_2$  und  $\Lambda_3$  durch die Beziehung

$$\Lambda_j = \sum_{i=1}^3 (\vec{p}_i \cdot \hat{n}_j)^2 ; (j = 1, 2, 3)$$

verknüpft. Die Ereignisebene ist die  $\hat{n}_2, \hat{n}_3$  - Ebene und  $\hat{n}_1$  definiert die Richtung, in der die Summe der Quadrate der Impulskomponenten minimal ist.

Die normierten Eigenwerte sind durch

$$Q_j = \frac{\lambda_j}{\sum_{i=1}^N p_i^2} = \frac{\sum_{i=1}^N (p_i \cdot \hat{r}_i)^2}{2 \sum_{i=1}^N p_i}$$

definiert, wobei sie der Beziehung  $Q_1 + Q_2 + Q_3 = 1$  genügen.

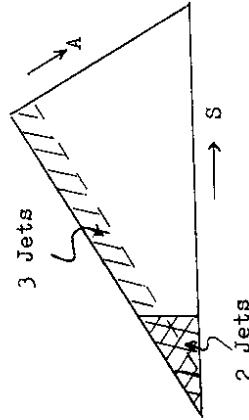
Die Form jedes Ereignisses läßt sich durch die beiden

Größen Aplanarity A und Sphericity S charakterisieren.

$$A = \frac{3}{2} Q_1$$

$$S = \frac{3}{2} (Q_1 + Q_2) = \frac{3}{2} \frac{\langle p_{\perp}^2 \rangle}{\langle p^2 \rangle}$$

Da  $0 < Q_1 < Q_2 < Q_3 < 1$  ist, liegen alle Ereignisse immerhalb eines Dreiecks. Kollineare Ereignisse liegen bei  $S \approx 0$ , koplanare Ereignisse bei  $A \approx 0$ ,  $S \neq 0$  und für sphärische Ereignisse haben S und A große Werte.



Zur Auswahl der 3-Jet-Kandidaten wählen wir die Bedingungen  $A \leq 0.06$  und  $S \geq 0.20$ . Diese Ereignisse sind in Abb. 7 schraffiert angedeutet. Acht der zehn 3-Jet-Kandidaten haben Werte von  $D \geq 0.12$  und die Gruppe dieser Ereignisse ist bis auf zwei Ereignisse mit den 3-Jet-Kandidaten identisch.

Durch einen einfachen hierarchischen Algorithmus mit einem sicher noch nicht ausreichenden Ähnlichkeitsmaß gelang bereits eine recht gute Klassifizierung der Jets. In der nächsten Zukunft wollen wir bessere Ähnlichkeitsmaße in ihrer Wirksamkeit zur Klassifizierung der Jets untersuchen.

Literaturverzeichnis

- /1/ H. Böttcher et al., Nucl. Phys. B 81, 365 (1974)
- /2/ Proceedings of the 3<sup>rd</sup> Topical Meeting on Multi-Dimensional Analysis of High Energy Data, Editor W. Kittel, Nijmegen 1978
- /3/ T. Ludlam in Proceedings of the 3<sup>rd</sup> Topical Meeting on Multi-Dimensional Analysis of High Energy Data, Editor W. Kittel, Nijmegen 1978
- /4/ W.D. Nowak and H. Schiller, Berlin Preprint PHE 75-12
- /5/ R. Honecker et al., Annalen der Phys. 36, 199 (1979)
- /6/ R. Honecker et al., Nucl. Phys. B 50, 157 (1972)  
M. Deuschmann et al., Nucl. Phys. B 99, 397 (1975)
- /7/ G. Wolf, Selected Topics on e<sup>+</sup>e<sup>-</sup>-Physics, DESY 80/13, February 1980
- /8/ H.E. Roloff and H. Schiller, Hierarchische Cluster Analyse, unpublished
- /9/ Wir danken Herrn Dr. G. Wolf für die Überlassung der TASSO-Ereignisse zum Testen der Methode
- /10/ S. Brandt and H.D. Dahmen, Z. Phys. C 1, 61 (1979)
- /11/ S.L. Wu and G. Zobernig, Z. Phys. C 2, 107 (1979)

Process	Cluster analysis		Maximum likelihood fit		Prism plot analysis	
	$\pi^+p$	$\pi^-p$	$\pi^+p$	$\pi^-p$	$\pi^+p$	$\pi^-p$
1. $A_1^+p$ $\leftarrow e^0\pi^+$	200 ± 22	220 ± 18	280 ± 67	204 ± 56		
2. $A_2^+p$ $\leftarrow e^0\pi^+$	200 ± 26	136 ± 23	167 ± 40	158 ± 33		
3. $A_3^+p$ $\leftarrow f\pi^+, e^0\pi^+$	114 ± 28	106 ± 16	115 ± 27	113 ± 22	590 ± 29	619 ± 40
4. $A_1^+p$ $\leftarrow e^0\pi^+, f\pi^+$	55 ± 15	40 ± 8	not found			
5. $N_{1470}^+p$ $\leftarrow A_{1238}^+\pi^+, \pi^-\pi^+$	68 ± 17	62 ± 13	146 ± 40	128 ± 22		
6. $N_{1760}^+p$ $\leftarrow A_{1238}^+\pi^+, \pi^-\pi^+$	115 ± 18	128 ± 15	106 ± 27	117 ± 22	189 ± 15	147 ± 18
7. $N_{2040}^+p$ $\leftarrow p\pi^+\pi^-$	48 ± 12	45 ± 8	not found			
8. $N_{2400}^+p$ $\leftarrow p\pi^+\pi^-$	28 ± 13	20 ± 6	not found		77 ± 11(e)	166 ± 19(a)
9. $A_{1260}^0$	160 ± 20	—	180 ± 47	—	194 ± 16	—
10. $A_{1236}^0$	108 ± 17	—	73 ± 13	—	88 ± 12	—
11. $A_{1236}^0$	30 ± 9	—	not found	—	32 ± 7	—
12. $A_{1300}^0$	40 ± 10	—	not found	—	67 ± 10	—
13. $A_{1296}^0$	—	48 ± 12	—	not found	—	not found
14. $A_{1296}^0$	—	11 ± 4	—	not found	—	not found
15. $N_{1520}^0$	—	48 ± 10	—	not found	—	not found
16. $N_{1520}^0$	—	19 ± 6	—	not found	—	not found
17. $A_{1236}^+\pi^-$	35 ± 15	80 ± 19	106 ± 28	127 ± 26	not found	44 ± 10
18. $A_{1296}^+\pi^+$	30 ± 8	not found	40 ± 7	45 ± 11	not found	not found
19. $N_{1520}^+\pi^+$	18 ± 6	not found	not found	not found	not found	not found
20. $p\pi^+e^0$	63 ± 34	29 ± 10	47 ± 15	68 ± 14	not found	90 ± 15
21. $p\pi^+f$	10 ± 8	not found	27 ± 11	40 ± 13	not found	30 ± 9
22. $p\pi^+g^0$	6 ± 5	15 ± 7	not found	not found	not found	not found
23. $p\pi^+$ (proton) $> 0$	6 ± 1	not found	not found	not found	not found	not found
24. $p\pi^+\pi^+\pi^-$	45 ± 32	72 ± 25	47 ± 17	35 ± 16	53 ± 20	32 ± 15

(a) These cross sections stand for  $\pi^+p \rightarrow (p\pi^+\pi^-\pi^-\pi^+)pp\pi^+$ .

Tabelle 1

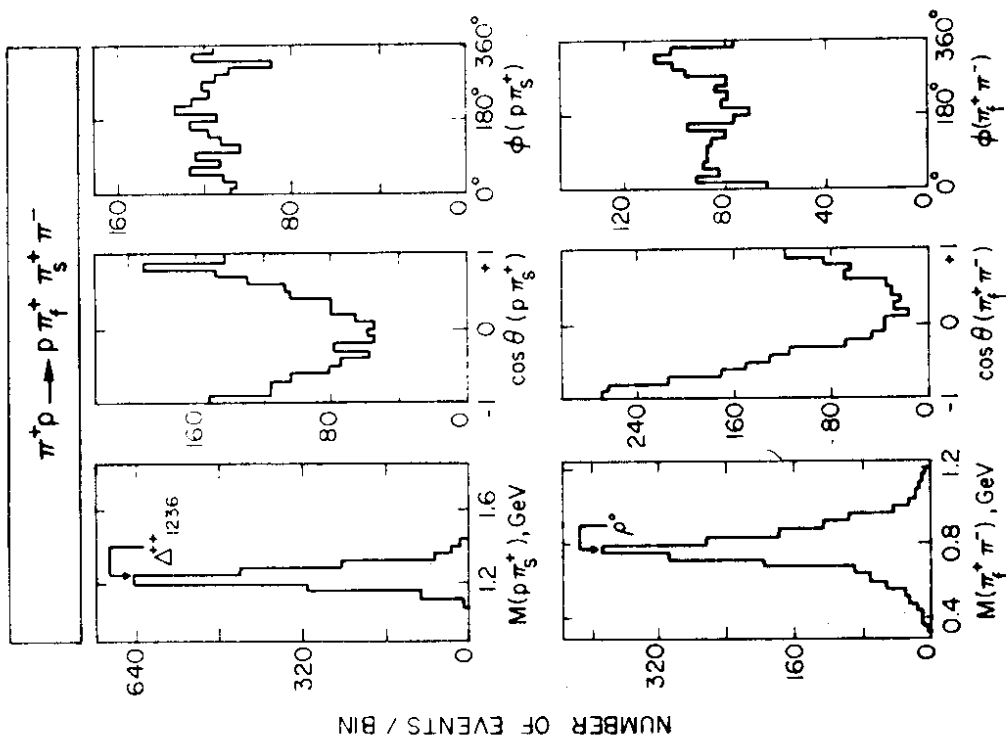
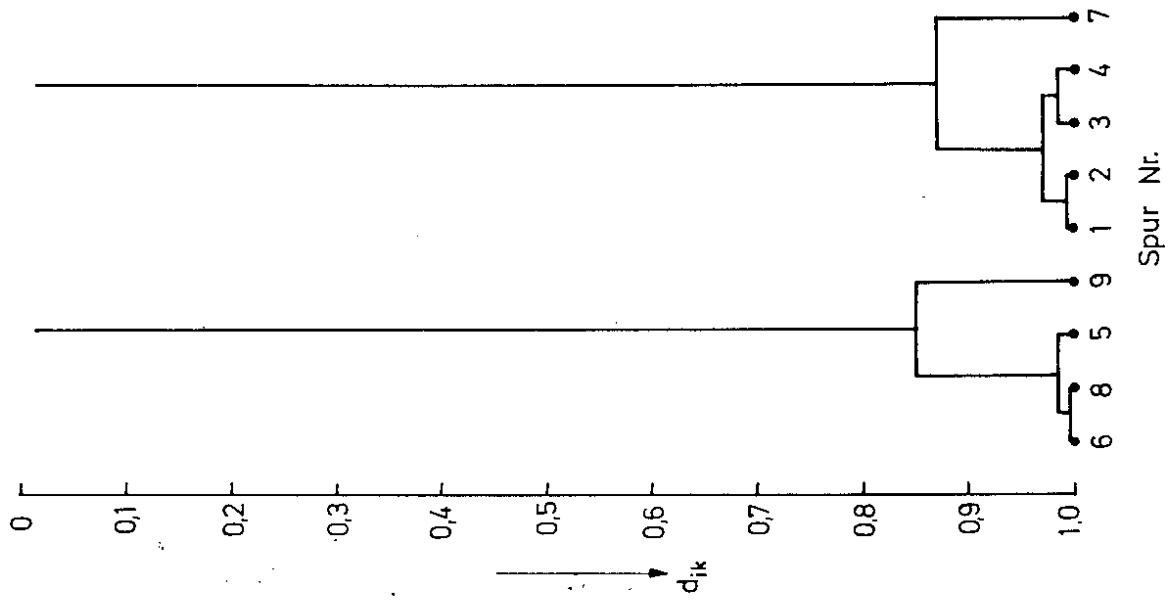


Abb. 1

# Einteilung der Ereignisse in Untergruppen

$d_{ik} < 0,8$   
 $\Delta(d_{ik}) \geq 0,28$

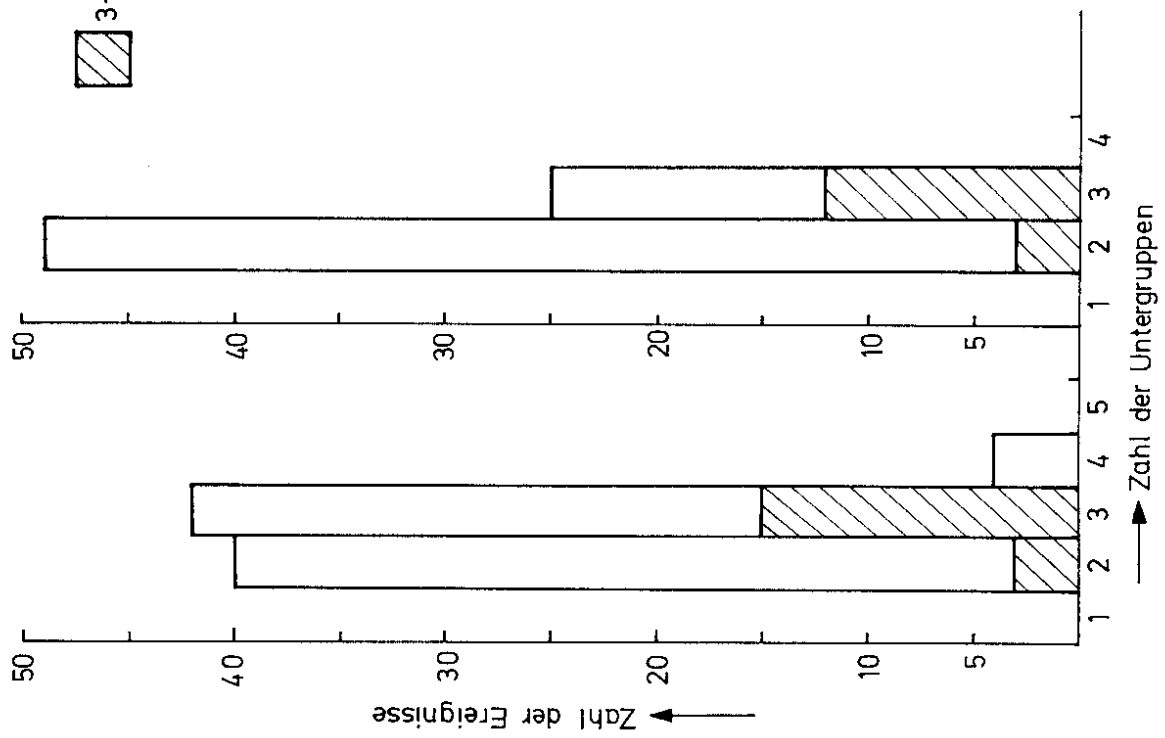


Abb. 3

# Ereignis 1056 / 2750

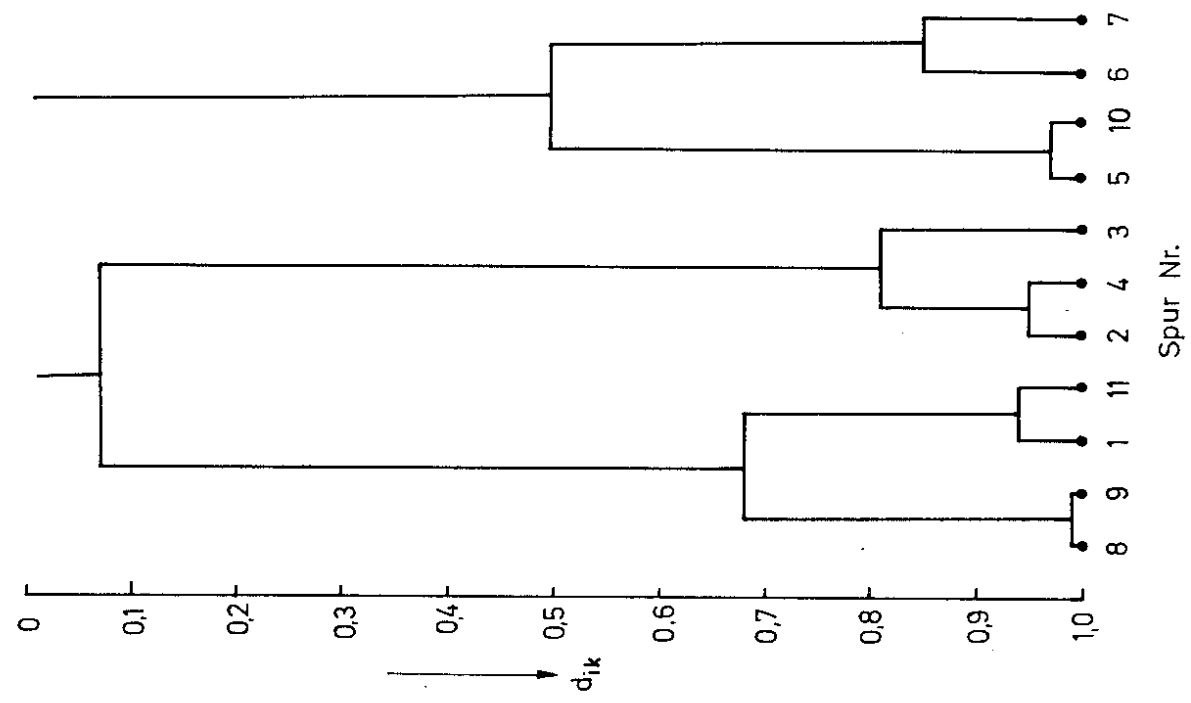


Abb. 2 b

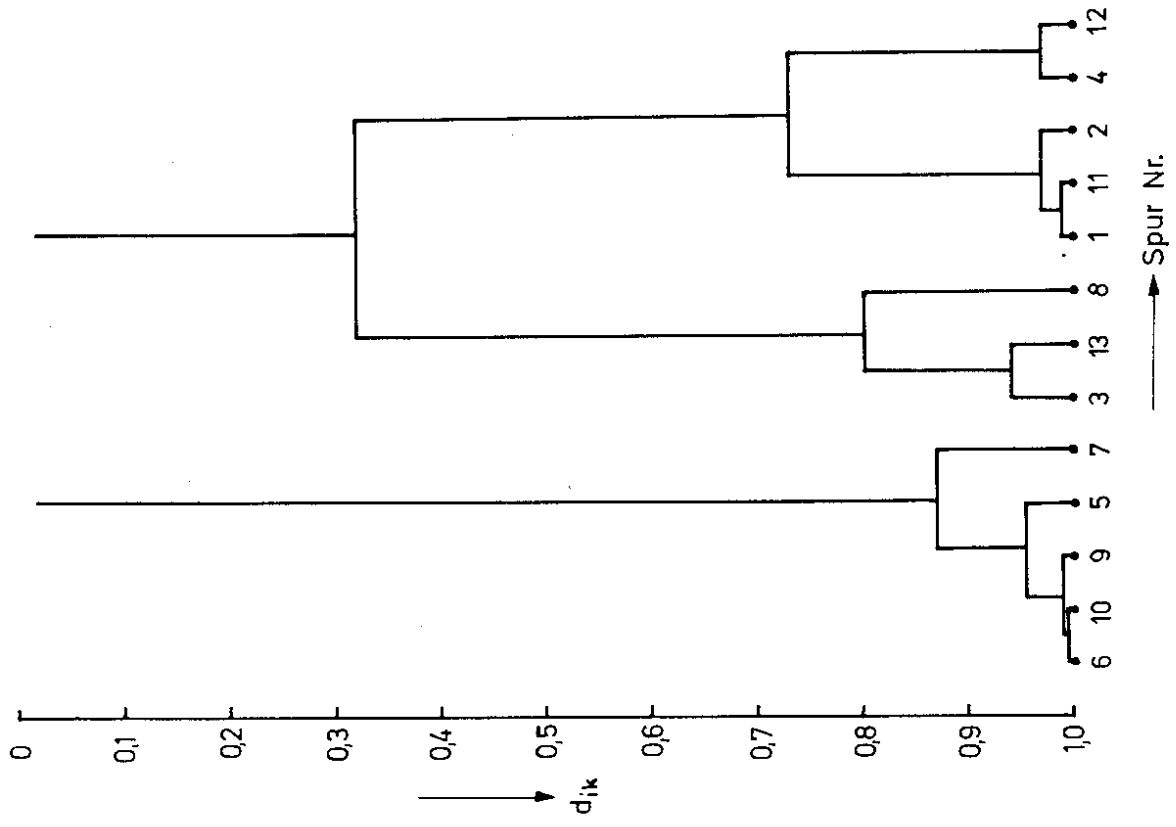


Abb. 5

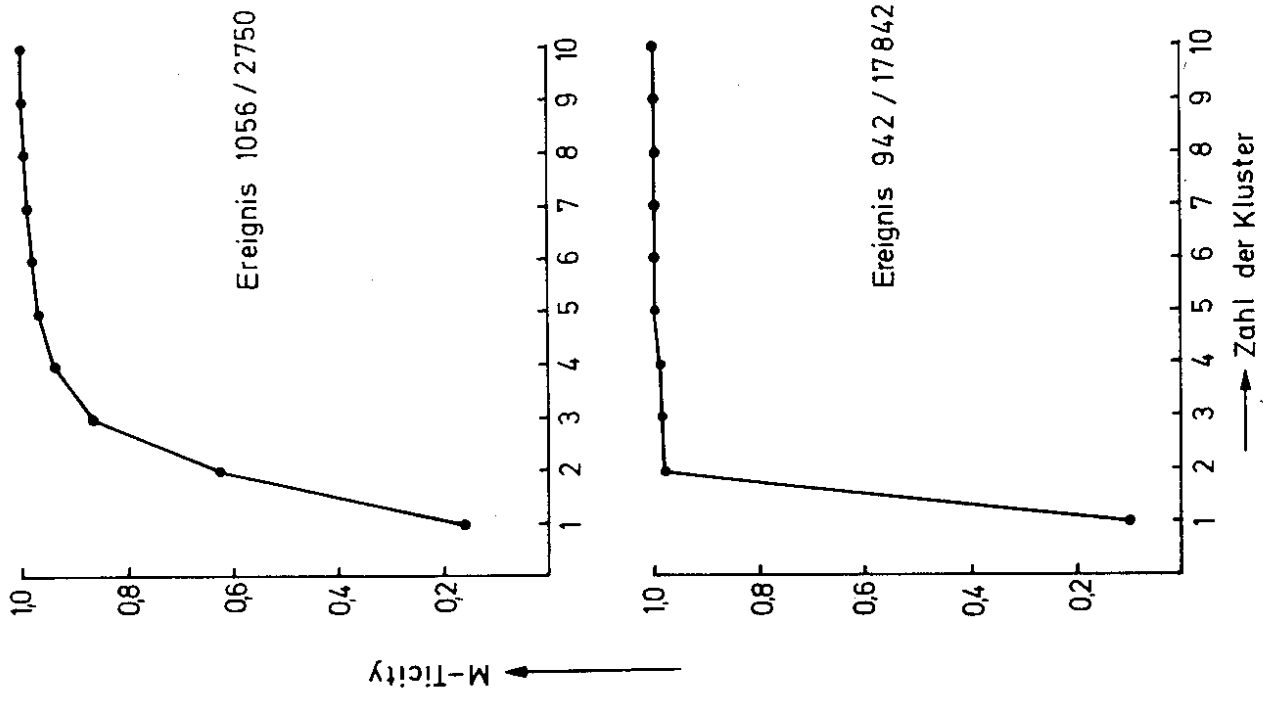


Abb. 4

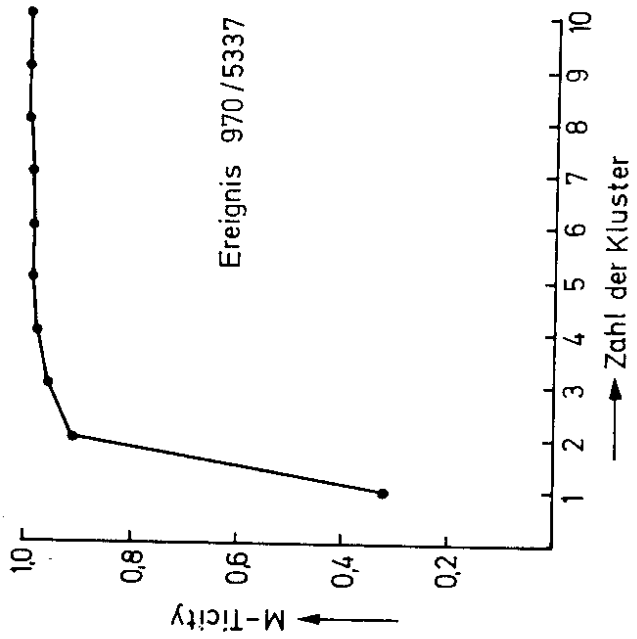


Abb. 6

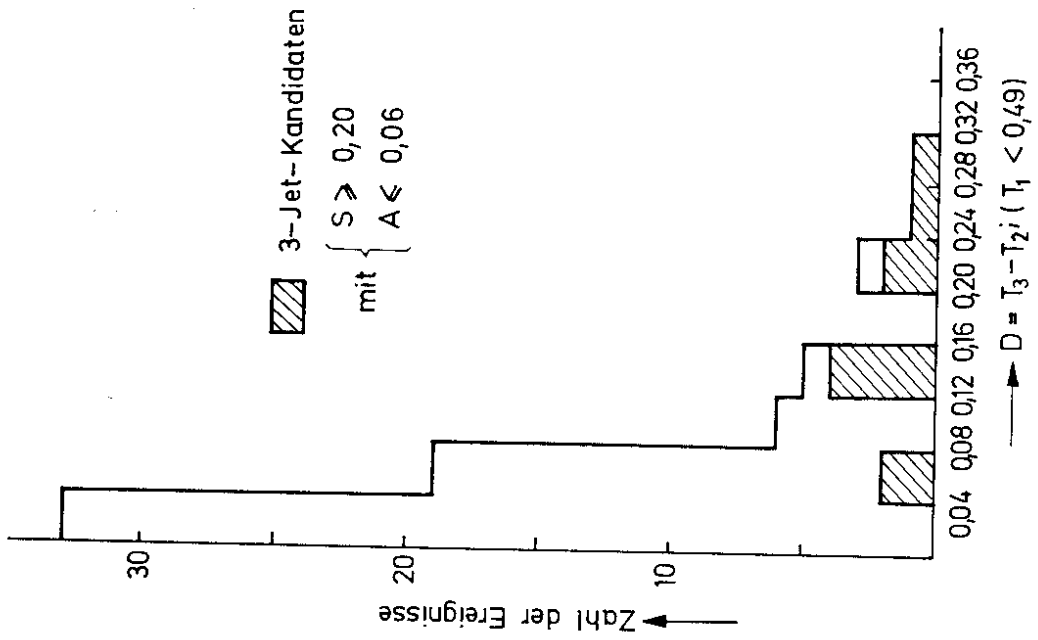


Abb. 7