

J.-P. BENZÉCRI

**Actualité : analyse des données en physique
corpusculaire**

Les cahiers de l'analyse des données, tome 2, n° 4 (1977),
p. 451-466

http://www.numdam.org/item?id=CAD_1977__2_4_451_0

© Les cahiers de l'analyse des données, Dunod, 1977, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ACTUALITÉ :

ANALYSE DES DONNÉES EN PHYSIQUE CORPUSCULAIRE

II. — Ajustement des méthodes descriptives (à suivre)

[PHYS. COR.]

par J.-P. Benzécri ⁽¹⁾

4 L'analyse statistique des réactions à haute énergie :

4.0 Reportons-nous quarante ans en arrière : les principes généraux de la mécanique ondulatoire sont acquis ; l'édification, aujourd'hui encore inachevée, d'une théorie relativiste des champs quantifiés est déjà entreprise ; après la découverte de la radioactivité artificielle, l'étude de la structure du noyau a commencé. On ne connaît encore que peu de particules élémentaires : le photon ; l'électron (négatif d'abord ; puis positif) ; le proton, le neutron. Mais au nom des principes généraux Fermi a postulé l'existence du neutrino (particule de masse et charge nulles) ; le même cadre théorique de l'équation de Dirac, qui associe à l'électron négatif usuel un électron positif, requiert un anti-proton ; et aussi un antineutron ; Yukawa propose d'associer aux forces qui unissent les constituants du noyau, le champ d'une particule nouvelle, le méson auquel il assigne une masse de quelque 200 fois celle de l'électron... Aujourd'hui toutes ces particules postulées ou pressenties ont été préparées et font l'objet de multiples expériences qui les ont rendues très familières aux spécialistes ; bien plus : les particules et résonances déjà découvertes dans ces expériences, se comptent par dizaines ; et il apparaît nécessaire d'en expliquer le système. Les principes théoriques n'étant pas assurés, l'expérimentation à haute énergie apporte sans cesse des faits qui, plus ou moins attendus par certains, ne sont jamais exactement prévus par personne. Les découvertes d'événements vraiment nouveaux sont rares ; mais l'étude minutieuse des canaux déjà connus permet de découvrir des résonances et d'en préciser la forme. Cette spectroscopie est le travail quotidien des laboratoires de haute énergie ; fondée sur de vastes ensembles de données, elle requiert des opérations statistiques que nous devons présenter ici.

Le problème principal est, nous l'avons dit dans le précédent article (§ 3.5), la découverte et la description des sous-canaux d'un canal particulier. Nous partirons donc conceptions les plus éloignées de l'analyse des données sans *a priori* que nous pratiquons d'ordinaire : de celles où un modèle analytique joue le plus grand rôle (§ 4.1). Nous considérerons ensuite les techniques descriptives qui comme celles du diagramme de Dalitz ne dépassent pas les deux ou trois dimensions accessibles à nos sens, mais peuvent requérir une grande virtuosité dans le traitement informatique (§ 4.2). Nous réservons à un prochain article l'étude des généralisations multidimensionnelles où jusqu'à présent la classification automatique a joué le premier rôle (§ 4.3). Reste à citer d'autres problèmes qui, tel celui de la description d'un événement d'après quelques impacts de trajectoires, offrent matière à l'analyse multidimensionnelle (§ 4.4).

(*) Suite de l'article paru sous le même titre dans les Cahiers Vol II n° 3.

(1) Professeur de Statistique. Université Pierre et Marie Curie, Paris.

Dans ce § 4, outre des notes prises aux séances du colloque de Genève et les documents écrits distribués par certains conférenciers, nous avons utilisé diverses publications antérieures du CERN, notamment l'excellent rapport de synthèse de W. Kittel au symposium de Liblice, 1974 (*) ; et bien que notre objet premier soit ici de rendre compte et non d'innover ; nous avons tenter de suggérer, chemin faisant, quelques perfectionnements aux méthodes pratiquées par les physiciens (**). De plus afin de donner au lecteur une vue fidèle d'un colloque où l'on entendit plusieurs communications de statistique mathématique, nous commentons, en appendice, celle de P. Huber relative à l'estimation robuste en analyse multidimensionnelle (§ 4.5).

4.1 L'ajustement à des formules analytiques : Nous montrons d'abord à quelles sortes d'expressions analytiques on recourt ordinairement (§ 4.1.1). Puis nous exposons une étude de P. Gavillet & J. C. Morin (Genève 1976) qui ne tient pas compte des interférences entre sous-canaux (§ 4.1.2). Enfin nous dirons d'après W. Kittel (Prague 1974) comment d'autres auteurs, L. Van Hove notamment, tiennent compte de celles-ci. Pour un statisticien accoutumé aux méthodes paramétriques, l'ajustement requiert une épreuve de validité : il est en effet commun en physique d'apprécier par l'épreuve du χ^2 l'ajustement d'une loi unidimensionnelle à un histogramme ; mais dans le domaine multidimensionnel, apprécier adéquatement l'ajustement d'une loi de forme analytique compliquée à un nuage de points n'est pas facile : le physicien se borne d'ordinaire à des vérifications unidimensionnelles ; toutefois nous exposerons plus bas (cf § 4.4.3) les épreuves multidimensionnelles utilisées par J. Friedman.

4.1.1 Les formules semi-empiriques : Reprenons quelques notations introduites au § 3 :

V : sous-variété des éléments permis dans l'espace des phases (cf § 3.1) ;

$v = \{ \{p\}_{e_j} \}$: un point de V , décrivant un événement $v \in V$ (cf § 3.1) ; communément on introduira d'autres coordonnées que les composantes elles-mêmes des quadrimoments des particules émergentes ;

I : ensemble d'événements effectivement observés et soumis à l'élaboration statistique ;

$i \in I \subset V$: on identifiera l'étiquette i d'un événement avec le point correspondant de V : $i \in V$;

C : ensemble des sous-canaux considérés pour la réaction étudiée ; $c \in C$; (cf § 3.5) ;

$\psi_c(v)$; ou $A_c(v)$: fonction d'onde sur V (appelée encore amplitude A de probabilité) associée au sous-canal c (cf §§ 3.5 & 3.6) ;

$|\psi_c(v)|^2 dv$; densité de probabilité correspondant au cas modèle où le sous-canal c serait seul ouvert ; ici, dv désigne l'élément de volume invariant de V (cf § 3.2).

(*) W. Kittel : *Treating multiparticle processes* ; in Symposium on anti-nucleon nucleon interactions ; Liblice-Prague ; Juin 1974 ; ed. L. Montanet ; CERN 74-18. Cet article sera désormais cité : (Prague 1974) ; de même que le colloque objet de notre travail est cité (Genève 1976).

(**) Ces suggestions ont maintenant fait l'objet de la thèse de D. Matti (Paris 3° cycle Sept. 1977) ; thèse dont un extrait doit paraître dans les Cahiers.

Les expressions analytiques de $\psi_c(v)$ ou $|\psi_c(v)|^2$ utilisées pour l'analyse des grands ensembles d'événements à haute énergie, ne prétendent pas être des formules physiques exactes : ce sont des instruments mathématiques, conçus d'après des théories inachevées, pour déterminer suivant quels axes se distribuent les résultats d'expériences : c'est pourquoi l'on parle de formules semi-empiriques. Citons comme exemple, d'après W. Kittel (1974) deux formules utilisées par Mac Naughton & coll. pour la réaction $\pi^+ p \rightarrow \pi^+ \pi^0 p$.

Dans le sous-canal $\pi^+ p \rightarrow \rho p \rightarrow \pi^+ \pi^0 p$ (formation intermédiaire d'un méson ρ associé à la paire $\pi^+ \pi^0$; sur ces particules intermédiaires non directement observables, cf § 3.5), on postule l'amplitude :

$$A_\rho \approx ((m_\rho^2 - m^2) - i \Gamma m_\rho)^{-1} (a \cos \theta_{\pi\pi} + b \sin \theta_{\pi\pi} \cos \varphi) ;$$

dans cette formule, m , $\theta_{\pi\pi}$ et φ sont des coordonnées de l'événement ; et m_ρ , Γ , a , b sont des paramètres à ajuster aux données. De façon précise, $m_{\pi\pi}$ est la masse invariante (cf § 3.3) du système formé de deux mésons émergents π^+ et π^0 ; $\theta_{\pi\pi}$ est l'angle formé par les vitesses de ces deux mésons ; et φ est un angle qui repère la position du plan formé par les deux vitesses (on notera que le terme en $\cos \varphi$ s'annule avec $\theta_{\pi\pi}$; i.e. quand les deux trajectoires, étant collinéaires ne déterminent pas un plan) ; m_ρ est la masse (moyenne estimée) du méson ρ ; Γ est appelée largeur de l'état résonnant ρ (le méson ρ n'étant pas une particule stable, on doit lui attribuer une durée de vie finie Δt ; et corrélativement, selon la 4^e relation d'incertitude de Heisenberg sa masse n'est pas exactement définie, mais elle est distribuée autour de la valeur centrale m_ρ avec une largeur inversement proportionnelle à Δt). Le terme $((m_\rho^2 - m^2) - i \Gamma m_\rho)^{-1}$, facteur de résonance qui est maximum en module pour $m = m_\rho$ et dont l'argument varie de π à la traversée de ce maximum est appelé terme de Breit & Wigner : ces auteurs l'ont d'abord introduit en physique du noyau.

Pour la même réaction, on considère aussi le sous-canal $\pi^+ p \rightarrow \pi^+ (\pi^0 p)$, appelé diffraction : mécanisme qu'on peut décrire comme suit : le π^+ change de direction : le p dans le choc produit un π^0 ; Mac Naughton & coll. prennent une formule classique :

$$A_d \approx \exp((\gamma/2) t_\pi) e^{i\delta}$$

où γ et δ sont des paramètres à ajuster ; et t_π est le transfert de quadrimoment, i.e. la norme relativiste au carré de la différence entre le moment initial et le moment final du π^+ (plus exactement :

$\| \{p\}_{\pi^+} + \{p\}_{e\pi^+} \|^2 = - t_\pi$; et la variation du quadrimoment est un vecteur de type espace ; donc de carré de norme négatif, selon la signature la plus en usage) : en bref, la formule implique une décroissance exponentielle du phénomène avec l'angle de déviation du π^+ au carré.

Gavillet et Marin écrivent simplement pour la densité de probabilité (ils ne tiennent pas compte des interférences ; donc leur modèle ne comporte pas de phases ; cf § 4.1.2)

$$F(t) \times R \times W(\cos \theta) \times \dots ;$$

ce qu'ils commentent ainsi :

$F(t)$: facteur périphérique. Le statisticien doit savoir qu'on parle de réaction périphérique quand une particule poursuit sa trajectoire avec un faible transfert de quadrimoment, même s'il se produit des particules secondaires (cf $\pi^+ p \rightarrow \pi^+ (\pi^0 p)$ ci-dessus). Cette expression imagée se réfère au cas de deux particules sphériques (cf § 3.2 ; fig 3.1 & 4.1) dont les mouvements engendrent des cylindres qui ne se pénètrent qu'à leur périphérie.

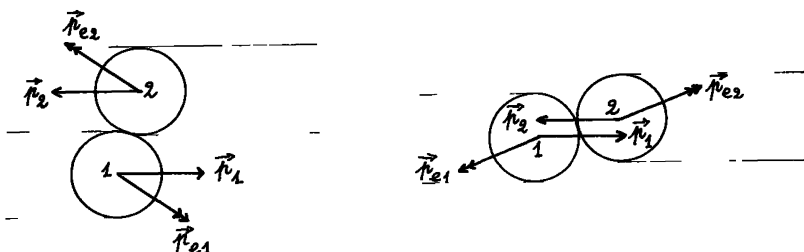


Figure 4-1: Deux cas de chocs entre deux sphères. À gauche, choc périphérique : faible transfert de quantité de mouvement. À droite, choc quasi-central : les deux sphères rebroussement chemin en échangeant presque toute leur quantité de mouvement.

4.1.2 La méthode de vraisemblance des canaux : Gavillet & Marin dans leur programme CHAFIT, (cf CERN/D. PHII/ PROG 75-2 ; Déc 1975) suivent Condon & Cowell (cf *Phys. Rev.* D 9 ; (1974) 2558) pour déterminer itérativement par la méthode du maximum de vraisemblance les parts afférentes à plusieurs sous-canaux (de forme fixée *a priori*) dans un échantillon d'événements d'une réaction donnée.

Une formule ayant été choisie pour l'amplitude d'un sous-canal ($\psi_c(v)$ selon nos notations), G & M normalisent la densité de probabilité correspondante en calculant l'intégrale :

$$F_c = \int_V |\psi_c(v)|^2 dv ;$$

(calcul effectué suivant la méthode de Monte-Carlo ; i.e. en tirant au sort un échantillon d'événements v dans V muni de la loi uniforme associée à son élément de volume naturel ; cf § 3.2). Ils posent :

$$\rho_c(v) = |\psi_c(v)|^2 / F_c ; \text{ d'où } : \int_V \rho_c(v) dv = 1 ;$$

$\rho_c(v) dv$ est une loi de probabilité sur V , c'est le profil du sous-canal c . On notera que cette normalisation faite, le taux de réaction entre deux faisceaux de densité déterminée (taux d'après lequel on calcule classiquement des sections efficaces, qui sont en bref l'analogue de la surface frontale πR^2 pour le cas modèle très simple des particules sphériques du § 3.2) n'apparaît plus dans la formule de ρ_c : c'est pourquoi nous parlons de profil.

Le système des profils $\{\rho_c | c \in C\}$ étant admis, (postulé pour être éventuellement retouché après exécution de leur programme) G. & M. cherchent leurs pondérations relatives d'après l'échantillon $I \subset V$ (cf 4.1.1). Sous l'hypothèse qu'il n'y a pas d'interférences (comme si les ondes

$\psi_c, \psi_c, \dots$ étaient incohérentes entre elles, ne provenaient pas d'une même source ; cf § 3.6), nous notons :

f_c : fréquence relative du sous-canal c ;

$f_c = \{f_c | c \in C\}$: loi sur C , décrivant le mélange des sous-canaux considérés ;

$P(v; f_c) = \Sigma \{f_c \rho_c(v) | c \in C\}$: densité de la loi sur V correspondant au système f_c des proportions relatives des sous-canaux.

$w_c^v = f_c \rho_c(v) / P(v; f_c)$: probabilité que (avec les fréquences relatives de la loi f_c), un événement v soit passé par le sous-canal c : c'est la formule usuelle de probabilité des causes, formule associée aux noms de Bayes et de Laplace.

Avec ces notations, le problème considéré par G. & M. après C. & C. s'énonce : trouver la loi $f_c \in \mathcal{P}_C$ (i.e. système de nombres positifs ou nuls f_c de somme 1) rendant maxima la fonction de vraisemblance $W(I; f_c)$ associée à l'échantillon I :

$$W(I; f_c) = \Pi \{P(i; f_c) | i \in I\} = \Pi \{\Sigma \{f_c \rho_c(i) | c \in C\} | i \in I\} ;$$

Et les auteurs donnent de ce problème la solution itérative suivante :

$$f_c^{(n+1)} = (1/\text{Card } I) \Sigma \{f_c^{(n)} \rho_c(i) / P(v; f_c^{(n)}) | i \in I\} ;$$

en bref, la part relative $f_c^{(n+1)}$ du sous-canal c est recalculée en répartissant la masse de chaque événement i sur l'ensemble C d'après les probabilités des causes w_c^i calculées en partant de la loi $f_c^{(n)}$ obtenue à l'itération précédente.

Il n'est pas précisé sous quelles conditions l'algorithme converge ; mais il est facile de montrer que la convergence (rapide en pratique selon G. & M.) ne peut conduire qu'à un *extremum* de la fonction de vraisemblance. Pour le voir, différencions logarithmiquement $W(I, f_c)$; il vient :

$$\partial W / (W \partial f_c) = \Sigma \{\rho_c(i) / P(i; f_c) | i \in I\} ;$$

différentiellement, la condition $f_c \in \mathcal{P}_C$ (f_c est une loi sur C , cf *supra*) implique : $\Sigma d f_c = 0$; cette condition équivaut à $dW = 0$ si et seulement si les dérivées partielles $\partial W / (W \partial f_c)$ sont toutes égales entre elles ; en ce cas leur valeur commune ne peut être que $\text{Card } I$, car :

$$\begin{aligned} \Sigma \{f_c \partial W / (W \partial f_c) | c \in C\} &= \Sigma \{f_c \rho_c(i) / P(i; f_c) | i \in I ; c \in C\} \\ &= \Sigma \{P(i; f_c) / P(i; f_c) | i \in I\} = \text{Card } I ; \end{aligned}$$

or la condition : $\forall c \in C : \partial W / (W \partial f_c) = \text{Card } I$, est justement celle de la convergence de l'algorithme de recalcul des fréquences f_c :

$$f_c = (1/\text{Card } I) \Sigma \{f_c \rho_c(i) / P(v; f_c) | i \in I\} = (f_c / \text{Card } I) \partial W / (W \partial f_c).$$

Certes, pour le physicien qui n'a que peu de confiance dans les

profils $\rho_c(v)$ attribués *a priori* aux sous-canaux c , les fréquences relatives f_c n'ont pas grand prix en elles-mêmes. Mais d'après les f_c , on peut calculer les w_c^1 i.e. partager entre les sous-canaux la masse 1 attribuée à chaque événement ; d'où pour chaque c un échantillon fictif *a posteriori*. On confronte alors, (notamment par divers calculs d'histogrammes ; cf § 4.2) cet échantillon *a posteriori* au profil ρ_c postulé *a priori* ; on modifie celui-ci, on en ajuste les paramètres. Généralement parmi les ρ_c postulés figure la densité uniforme sur V ; c'est ce qu'on appelle le *fond* : *a posteriori*, il reste principalement dans ce sous-canal les événements dont les sous-canaux déjà reconnus ne rendent pas compte ; c'est-à-dire en quelque sorte par soustraction, un échantillon (fictif ; *a posteriori*) susceptible de suggérer l'introduction d'un nouveau sous-canal. En somme, pour le physicien, l'algorithme G. & M., n'est qu'un élément répété au sein d'un processus itératif plus ample, visant à ajuster les paramètres de sous-canaux et à en découvrir de nouveaux.

4.1.3 *Ajustement semi-empirique et interférences* : Dans ce processus itératif manquent cependant les interférences entre sous-canaux. La formule additive : $P(v; f_c) = \sum_c f_c \rho_c(v)$ est inexacte ; on doit poser :

$$\psi(v, \alpha_c) = \sum \{ \alpha_c \psi_c(v) \mid c \in C \} ;$$

$$\Pr(v; \alpha_c) = |\psi(v; \alpha_c)|^2,$$

où les α_c sont des coefficients complexes, déterminés dans leur ensemble à la multiplication près par un nombre complexe de module 1 et assu-

jettis éventuellement à la condition de normalisation $\int_V \Pr(v; \alpha_c) dv = 1$;

condition qui s'exprime assez simplement par une forme hermitique(*) en

les α_c dont les coefficients sont les intégrales $\int_V \overline{\psi_c(v)} \overline{\psi_{c'}(v)} dv$, calculables par simulation comme les coefficients diagonaux $F_c = \int_V |\psi_c|^2 dv$, déjà rencontrés au § 4.1.2.

La fonction de vraisemblance associée à l'échantillon I s'écrit maintenant :

$$\text{Wr}(I; \alpha_c) = \prod \{ \Pr(i; \alpha_c) \mid i \in I \} ;$$

il n'y a plus, à notre connaissance, de processus itératif aussi simple que celui du § 4.1.2, pour trouver un optimum α_c . Si toutefois on a, par un calcul approché, fait choix de α_c , on peut ici encore, construire des échantillons fictifs *a posteriori*. Mais puisque dans $\Pr$ on a tenu compte des interférences, on devra de même partager le poids (unité) de chaque événement i entre termes principaux et termes d'interférences. Supposons pour fixer les notations, que les canaux sont désignés par des nombres $c = 1, 2, \dots$, $\text{Card } C = k$. On a :

$$\begin{aligned} \Pr(v; \alpha_c) = & \sum \{ |\alpha_c|^2 |\psi_c(v)|^2 \mid c = 1, \dots, k \} \\ & + 2 \sum \{ \text{Re}(\overline{\alpha_c} \alpha_{c'}) \overline{\psi_c(v)} \psi_{c'}(v) \mid 1 \leq c < c' \leq k \} ; \end{aligned}$$

(*) Il pourra être utile de faire sur les canaux un changement de base rendant cette forme diagonale ; i.e. introduire des combinaisons linéaires de canaux, orthogonales entre elles.

on notera donc :

$$w_c^v = |\alpha_c|^2 |\psi_c(v)|^2 / \text{Pr}(v, \alpha_c) ;$$

$$w_{cc'}^v = w_{c'c}^v = 2 \Re(\bar{\alpha}_c \alpha_{c'} \bar{\psi}_c(v) \psi_{c'}(v)) / \text{Pr}(v, \alpha_c).$$

L'échantillon *a posteriori* associé au sous-canal *c* comprend simplement les événements *i* de *I*, affectés chacun du poids (positif ou nul) w_c^i ; mais dans l'échantillon d'interférence entre les canaux *c* et *c'*, les événements *i* figurent avec un poids soit positif (interférence constructive), soit négatif (interférence destructive).

Selon W. Kittel (Prague 1974) L. Van Hove a eu recours avec avantage à ces échantillons *a posteriori* pour ajuster itérativement les paramètres de sous-canaux, et en découvrir de nouveaux en tenant compte des interférences. Au colloque (Genève 1976) des études d'interférences ont été présentées par J. Engelen, et J. Mac Naughton.

Reste à signaler une autre difficulté. Mac Naughton & coll. ainsi que Van Hove (cf Kittel 1974) utilisent des formules telles que celle donnée ci-dessous dans le cas de deux sous-canaux :

$$\text{Pr}(v) = |\alpha_1|^2 |\psi_1(v)|^2 + 2 d_{12} \Re(\bar{\alpha}_1 \alpha_2 \bar{\psi}_1(v) \psi_2(v)) + |\alpha_2|^2 |\psi_2(v)|^2 ;$$

Ici le coefficient d_{12} , nombre réel compris entre 0 et 1 permet de modérer l'influence du terme d'interférence : $d_{12} = 0$ élimine toute interférence, comme si les ondes ψ_1 et ψ_2 étaient incohérentes entre elles ; $d_{12} = 1$ donne la cohérence parfaite ; dans un cas intermédiaire (e.g. $d_{12} = 0,5$) on peut parler de cohérence partielle. A cette situation qu'on peut appeler *dépolarisation statistique*, nous croyons pouvoir attribuer une origine physique réelle, car le modèle ne tient pas compte du spin des nucléons ; de plus dans une formule phénoménologique imparfaite un paramètre supplémentaire tel que d_{12} peut aider à l'ajustement, même s'il n'est pas clairement interprétable.

Nous concluons ce § 4.1.3 en répétant au statisticien mathématicien que l'application de méthodes bien connues fondées sur le maximum de vraisemblance, pose en mécanique ondulatoire des problèmes nouveaux, du fait des interférences.

4.2 La statistique descriptive en faible dimension : Dans l'étude des réactions à haute énergie, les physiciens n'ont jamais suivi aveuglément des calculs de statistique paramétrique (cf § 4.1) sans en critiquer les résultats par des méthodes descriptives. L'histogramme et le diagramme triangulaire de Dalitz sont certes universellement employés (§ 4.2.1). Recourant parfois à des hypothèses physiques simplificatrices (§ 4.2.2), certains auteurs sont parvenus à une vision globale de l'espace des phases, sans pour autant s'aventurer au-delà de la dimension 3 (§ 4.2.3). A l'examen sous tous ses angles d'un nuage multidimensionnel, l'ordinateur couplé avec un écran cathodique peut servir comme une sorte de caméra mobile dans l'espace de dimension *n* (§ 4.2.4). Et parvenus à la limite de ces techniques dont les principes statistiques sinon la mise en oeuvre, sont simples, certains physiciens ont depuis plusieurs années abordé l'analyse multidimensionnelle (§§ 4.3 & 4.4).

4.2.1 Histogrammes et diagrammes triangulaires : Dans l'analyse d'un ensemble d'événements à haute énergie, l'histogramme est à la fois l'instrument élémentaire de la découverte d'un sous-canal et celui de la vérification des résultats les plus précis obtenus par d'autres méthodes. Montrons cela sur deux exemples.

Dans un texte manuscrit dont quelques copies furent distribuées à Genève, l'équipe du M.I.T. (R. Hulsinger & coll.) présente comme suit l'application à la réaction $\pi^+ p \rightarrow p \pi^+ \pi^0$ (à 3,9 GeV) de leur méthode des diagrammes tridimensionnels prismatiques (cf infra § 4.2.3).

"Point n'est besoin de beaucoup d'imagination pour faire les trois histogrammes des masses invariantes des couples $p\pi^+$, $p\pi^0$, $\pi^+\pi^0$ " (cf §§ 3.3 & 3.5). On voit immédiatement (cf fig 4.2) sur l'histogramme de la masse $m_{p;\pi^+}$ un maximum prononcé qui suggère le sous-canal $\pi^+ p \rightarrow \Delta^{++} \pi^0$ ($p\pi^+$) π^0 ; et de même sur l'histogramme de $m_{\pi^+;\pi^0}$ on découvre le sous-canal $\pi^+ p \rightarrow p \rho^+ \rightarrow p(\pi^+\pi^0)$. Mais l'importance relative et le profil exact

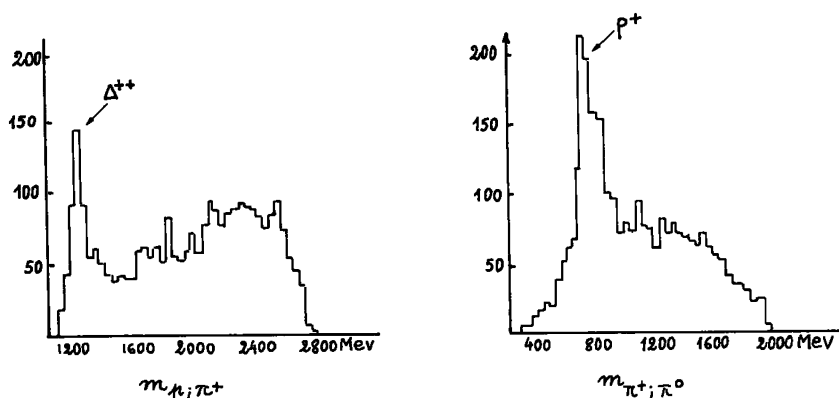


Figure 4-2 : histogrammes des masses invariantes d'après 2307 événements de la réaction $\pi^+ p \rightarrow p \pi^+ \pi^0$ à 3,9 GeV; chaque créneau des histogrammes a pour largeur 40 MeV; e.g. il y a ≈ 150 événements pour lesquels $m_{p;\pi^+}$ est compris entre 1200 et 1240 MeV/c² : c'est le pic Δ^{++} de l'histogramme de gauche, (reproduit d'après W. Kittel qui renvoie à J.F. Brau et coll. ; Phys. Rev. Letters 27 (1971) 1481).

de ces sous-canaux restent inconnus. Dans le diagramme triangulaire de Dalitz, Δ^{++} et ρ^+ se manifestent par des bandes denses (cf un cas analogue § 3.5 et fig 3.4) ; et d'autres mécanismes apparaissent que nous ne précisons pas ici. Les classes d'événements afférentes aux divers sous-canaux sont toutefois imparfaitement séparées : la séparation s'améliore grandement quand on considère le nuage des événements sur le diagramme prismatique tridimensionnel.

W. Kittel (1974) montre d'abord sur la même réaction $\pi^+ p \rightarrow p \pi^+ \pi^0$ (à 3,9 GeV), les mérites du diagramme prismatique. Il prend ensuite (après A. Ferrando & Coll) la réaction $K^+ p \rightarrow K^0 \pi^+ p$ (à 5 GeV) comme un exemple des résultats encore plus fins que peut fournir l'ajustement à des formules analytiques. Dans cette dernière réaction, se distinguent

deux sous-canaux : $K^+ p \rightarrow K^0 \Delta^{++} \rightarrow K^0 (\pi^+ p)$, et : $K^+ p \rightarrow K^* p \rightarrow (K^0 \pi^+) p$; le premier se révèle par un pic Δ^{++} sur l'histogramme de la masse invariante $m_{\pi^+, p}$; le second par un pic K^* dans l'histogramme de m_{K^0, π^+} . Pour définir un sous-échantillon d'événements relevant du sous-canal

$K^+ p \rightarrow K^* p$, on peut se borner à prendre tous les événements pour lesquels la masse invariante m_{K^0, π^+} est comprise dans l'intervalle de 1,37 à 1,47 GeV/c² (encadrant la masse moyenne 1,42 du K^*). Mais traçons pour ce sous-échantillon l'histogramme de la masse invariante $m_{\pi^+, p}$: celui-ci présente un pic net au voisinage de la masse Δ^{++} : d'où l'on peut conclure que 1/6 environ des événements du sous-échantillon construit pour le sous-canal $K^+ p \rightarrow K^* p$, relèvent en fait du sous-canal $K^+ p \rightarrow K^0 \Delta^{++}$. Si l'on recourt au diagramme prismatique pour définir un sous-échantillon relevant du sous-canal $K^+ p \rightarrow K^* p$, l'histogramme de $m_{\pi^+, p}$ associé à ce sous-échantillon présente encore un pic visible au niveau de la masse du Δ^{++} ; mais ce pic - les spécialistes disent "ce signal" - est bien moindre. Enfin par ajustement à des formules analytiques (ici il convient de prendre quatre termes ; un terme pour le sous-canal $K^0 \Delta^{++}$; et dans le sous-canal $K^* p$ trois termes, classiquement notés s, p, d, décrivant la distribution angulaire de la paire $K^0 \pi^+$ issue du K^*) on obtient un sous-échantillon du canal $K^+ p \rightarrow K^* p$, d'où le signal Δ^{++} a totalement disparu ; et il devient possible de déceler d'autres mécanismes à l'oeuvre dans cette même réaction.

Pour le statisticien, la morale de ces deux études sera donc qu'en physique des hautes énergies, tout commence et finit par des histogrammes.

4.2.2 Le choix des coordonnées : Nous l'avons dit (cf § 3.1), dans le cas de n particules émergentes, la sous-variété V des événements permis dans l'espace des phases a pour dimension $3n - 4$; mais compte-tenu de l'invariance par rotation autour de l'axe du faisceau incident, le nombre des paramètres invariants caractérisant un événement à n particules émergentes est $3n - 5$ (cf § 3.2). Si $n = 3$: $3n - 4 = 5$; $3n - 5 = 4$; un point du diagramme plan de Dalitz est projection d'une fibre de dimension 3 de V (cf § 3.4 *in fine*) ; cette fibre (compte-tenu de l'invariance par rotation) contient des événements qui doivent être différenciés par deux paramètres invariants ; il manque deux dimensions au diagramme de Dalitz. Dans le cas d'un événement à 4 particules émergentes, on a $3n - 5 = 7$: tel quel un nuage dans R^7 ne se prête pas à l'observation humaine. Avant de recourir à l'analyse statistique des données multidimensionnelles pour réduire celles-ci à un système de classes ou à des représentations de faible dimension (§ 4.3) on peut construire de telles représentations d'après la structure physique du problème : ceci revient à faire choix sur V de certaines variables importantes qui serviront de coordonnées pour une représentation à deux ou trois dimensions. Ces variables sont principalement d'une part les énergies des particules émergentes déjà utilisées pour le diagramme triangulaire (§ 3.3), d'autre part les composantes longitudinales des quadrimoments de ces particules.

On a vu (cf § 4.1.1, fig 4.1) que dans un choc périphérique, les quadrimoments des particules sont peu modifiés. Plus généralement on peut parler de réaction multipériphérique chaque fois que les quadrimoments $\{p\}_{ej}$ des particules émergentes sont à peu près situés dans le plan de R^4 défini par les quadrimoments $\{p\}_1$ et $\{p\}_2$ des particules initiales. Il est équivalent de dire que dans le système de référence

du laboratoire les particules émergentes ont des vitesses (ou des moments $\vec{p}_{ej}$) presque parallèles à l'axe du faisceau incident, soit de même sens soit de sens contraire ; les composantes transversales étant faibles. Ou encore, de dire, que dans un système où le centre de gravité des particules initiales 1 et 2 est immobile (système où l'axe des temps est orienté suivant $\{p\}_{\text{tot}}$, cf § 3.1), les particules émergentes ont des vitesses presque parallèles à l'axe défini par les deux vecteurs opposés $\vec{p}_1$, $\vec{p}_2$ soit dans le sens de l'un, soit dans celui de l'autre ; comme si chacune des particules incidentes 1 ou 2 avait dans le choc donné naissance à un groupe de particules émergentes J_1 ou J_2 de direction voisine de la sienne, ce que par un graphe on peut symboliser comme suit (cf § 3.5, fig 3-5) : dans le graphe de la figure 4-3, l'échange (ligne tiretée) entre les deux moments est d'autant plus important que e.g. le groupe J_1 peut ne contenir aucune particule identique à 1 (même à un changement de charge près).

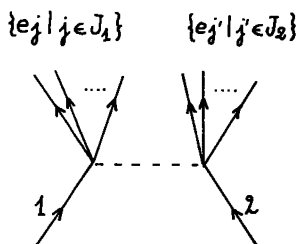


Figure 4-3. graphe symbolisant un événement pour lequel les moments transverses sont faibles.

Dans beaucoup des expériences réalisées jusqu'à présent, il est commun que les composantes transverses des $\{p\}_{ej}$ soient faibles ; et bien qu'à très haute énergie les événements à fortes composantes transverses aient une importance essentielle pour révéler la structure de la particule cible, (ces événements pouvant correspondre à des chocs avec une infraparticule composante de celle-ci), c'est par les composantes longitudinales que, conformément aux vues de Van Hove (*Nucl. Phys. B* 9 (1969) 331 ; cité d'après Kittel, 1974) débute l'analyse statistique.

Revenons donc au cas de trois particules émergentes. En projection sur le plan engendré par $\{p\}_1$ et $\{p\}_2$, le quadrimoment $\{p\}_{ej}$ a pour coordonnées invariantes naturelles E_{ej}/c et p_{Lej} (composantes longitudinales du moment, sur un axe orienté par $\vec{p}_2$, dans le système où le centre de gravité est immobile). De plus P.A. Miller (Genève 1976 ; cf *infra* § 4.2.3) a considéré les coordonnées isotropes $q_{ej}^+ = (E_{ej}/c) + p_{Lej}$ et $q_{ej}^- = (E_{ej}/c) - p_{Lej}$, dont l'intérêt est qu'à un facteur de proportionnalité près elles ont même expression si on se place dans un système d'axes où le centre de gravité n'est pas immobile, e.g. dans le système du laboratoire (*).

(*) Sur le choix des variables, cf P. Lutz, thèse, (Paris Mai 1977), 2^e Partie Ch II, § C.1 ; et notamment le § C.1.d, consacré aux variables isotropes.

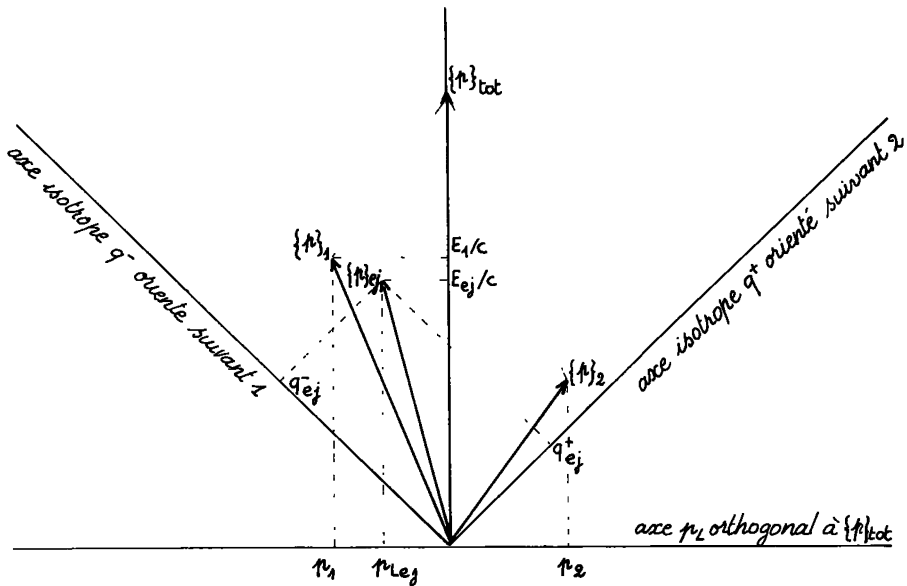


Figure 4-4 : Coordonnées longitudinales dans R^4 (espace-temps quadrudimensionnel) les deux quadrimoments $\{p\}_1$ et $\{p\}_2$ définissent un plan sur lequel il est avantageux de projeter le quadrimoment $\{p\}_{ej}$ de toute particule émergente. Dans ce plan on distingue un axe des temps (ou des énergies) défini par $\{p\}_{tot}$, un axe spatial (ou axe des moments longitudinaux) orthogonal au précédent, et deux axes isotropes (portant des vecteurs dont la norme hyperbolique est nulle) intersection du plan avec le cône de lumière.

Considérons un événement à trois particules émergentes. Les trois énergies $\{E_{e1}, E_{e2}, E_{e3}\}$ dont la somme est E_{tot} , sont bien représentées simultanément sur le diagramme de Dalitz (cf supra 3.3). De même pour un triplet de coordonnées isotropes de Miller $\{q_{e1}^+, q_{e2}^+, q_{e3}^+\}$ (ou pour les q^-) la représentation triangulaire s'impose, car il s'agit de trois nombres positifs dont la somme est q_{tot}^+ . Quant aux trois moments longitudinaux $\{p_{Le1}, p_{Le2}, p_{Le3}\}$, mesurés, répétons-le, dans le système où le centre de gravité est immobile, i.e. $p_{Ltot} = 0$, ils ont somme nulle ; on les représente généralement sur le diagramme hexagonal de Van Hove, que nous expliquerons comme suit (cf fig 4-5).

Dans l'espace tridimensionnel, le point $\{p_{Le1}, p_{Le2}, p_{Le3}\}$ appartient au plan H d'équation : $p_{Le1} + p_{Le2} + p_{Le3} = p_{Ltot} = 0$, plan perpendiculaire à la première diagonale du cube construit sur les trois vecteurs de base $\vec{e}_1, \vec{e}_2, \vec{e}_3$. Il revient au même de dire que dans le plan H , le triplet des moments longitudinaux est représenté par la somme vectorielle :

$$p_{Le1} \vec{v}_1 + p_{Le2} \vec{v}_2 + p_{Le3} \vec{v}_3 ,$$

où V_1, V_2, V_3 sont les projections sur H des vecteurs de base. Il est clair que le point $\{p_{Le1}, p_{Le2}, p_{Le3}\}$ ne décrit pas le plan H tout entier, car les moments longitudinaux sont bornés ; de façon précise rappe-
 pelons qu'on a l'égalité (cf § 3.1)

$$E_{\text{tot}}/c = \sum \{ (c^2 m_{ej}^2 + |\vec{p}_{ej}|^2)^{1/2} \mid j = 1, 2, 3 \} ;$$

si l'on substitue à $|\vec{p}_{ej}|^2$ la quantité p_{Lej}^2 qui lui est inférieure ou égale, on obtient une inégalité, l'égalité n'étant réalisée que si tous les moments transverses sont nuls, i.e. pour un événement multipériphé-
 rique (cf § 4.1.1 & fig 4-1) :

$$\sum \{ (c^2 m_{ej}^2 + p_{Lej}^2)^{1/2} \mid j = 1, 2, 3 \} \leq E_{\text{tot}}/c$$

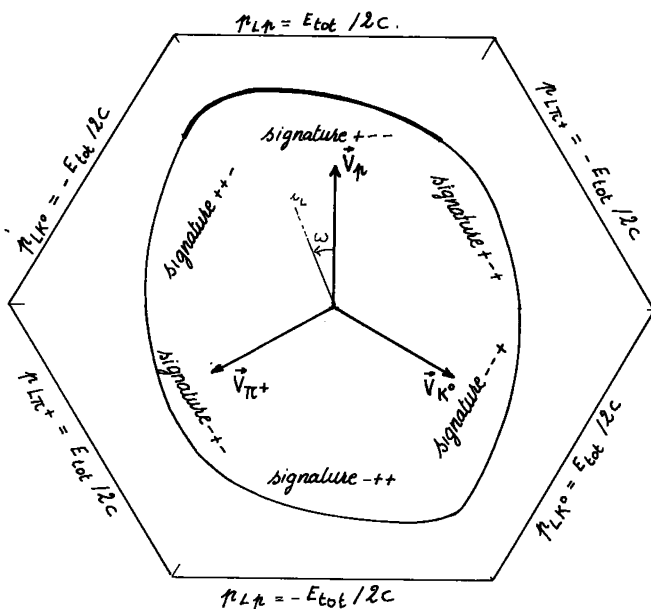


Figure 4-5 Diagramme hexagonal de Van Hove pour la réaction $K^+ p \rightarrow p^+ \pi^+ K^0$.
 Énergie du K^+ incident dans le système lié au laboratoire = 3 GeV ; énergie totale
 dans un système lié au centre de gravité : $E_{\text{tot}} = 2,56$ GeV ; dans ce système l'axe des
 p_L est orienté pour que le proton incident ait un moment positif $\approx 1,04$ GeV/c. Le plan
 est divisé en six angles égaux correspondant chacun à une signature : e.g. $++$: $p_{Lp} > 0$;
 $p_{L\pi^+} < 0$; $p_{LK^0} > 0$. On a représenté la frontière du domaine des triplets permis, ainsi que la
 frontière de l'hexagone : $\sum |p_{Lej}| = E_{\text{tot}}$; e.g. l'arête de signature $+-$ de cet hexagone a pour
 équation $p_{Lp} = E_{\text{tot}}/2c$; $p_{L\pi^+} + p_{LK^0} = -E_{\text{tot}}/2c$. Dans le système du laboratoire la vitesse de
 recul du proton a pour minimum 0,02c, sur la frontière du domaine des triplets permis
 on a renforcé l'arc sur lequel $v_{pL} \leq 0,2c$.

Dans le plan H , l'égalité $\Sigma\{(c^2 m_{ej}^2 + p_{Lej}^2)^{1/2}\} = E_{tot}/c$, définit la frontière du domaine accessible au point figuratif du triplet des moments longitudinaux. Cette frontière est intérieure à celle obtenue dans le cas limite où les masses m_{ej} sont toutes trois nulles (approximation qui peut toujours être faite si $c^2 m_{ej}^2 \ll E_{tot}^2$) : dans ce dernier cas on a simplement $\Sigma\{|p_{Lej}| \mid j = 1, 2, 3\} = E_{tot}/c$; équation qui définit dans H une frontière hexagonale : c'est proprement l'hexagone de Van Hove. Sur le diagramme de Van Hove de la figure 4-5, on a, comme sur le diagramme de Dalitz de la figure 3-4, considéré la réaction $K^+ + p \rightarrow K^0 + \pi^+ + p$ à 3 GeV (i.e. dans le cas où la quatrième composante du quadrimoment du K^+ incident est 3 GeV/c, dans un système lié au laboratoire)

Dans le cas de quatre particules émergentes, les énergies $\{E_{ej}\}$, nombres positifs de somme fixée E_{tot} sont bien représentées par un point intérieur à un tétraèdre : c'est la généralisation du diagramme triangulaire de Dalitz (la même représentation servirait pour les $\{q_{ej}^+\}$ ou les $\{q_{ej}^-\}$ de Miller). Quant aux $\{p_{Lej}\}$ dont la somme est nulle on a pour eux un diagramme de Van Hove généralisé, construit avec quatre vecteurs $\vec{v}_j$ de somme nulle (obtenus e.g. en joignant le centre d'un tétraèdre régulier aux quatre sommets) ; au lieu d'un hexagone, on obtient un cuboctaèdre. Le statisticien habitué aux représentations géométriques des ordres de préférence (cf e.g. J.P. Benzécri : Sur l'analyse des préférences ; in *Ordres totaux finis*, M. Barbut & coll. ; Gauthier-Villars, Mouton ; 1971) imaginera sans peine les constructions de diagrammes pour n particules : au fond ces représentations dérivent directement de l'usage des coordonnées rectangulaires dans R^n , leurs harmonieuses symétries sont celles de l'hypercube, en projection sur un hyperplan... Du point de vue physique, chaque face du diagramme des moments longitudinaux correspond à une partition de J en deux sous-ensembles J_1 et J_2 , (cf e.g. fig 4-5 : une signature pour chaque arête de l'hexagone ; signe positif J_2 , du côté de 2 ; signe négatif J_1), donc à un graphe du type de la figure 4-3 : il ne faudrait pas toutefois se hâter d'adopter cette décomposition géométrique comme une analyse en sous-canaux !

Selon-nous, la représentation longitudinale gagnerait à être modifiée comme suit (*). Pour chaque événement, le système des vecteurs $\vec{p}_{ej}$ (de somme nulle dans le système lié au centre de gravité) peut être ajusté (selon la méthode des moindres carrés ; i.e. l'analyse factorielle) à une droite Λ qui n'est pas exactement celle définie par $\vec{p}_1 = -\vec{p}_2$ (même si les composantes transversales sont faibles) ; sur Λ on peut compter les moments longitudinaux p_{Lej} , d'où une représentation de l'événement par un point $\{p_{Lej}\}$ du diagramme de Van Hove ; de plus on retiendra comme une variable intéressante, même s'il est faible, l'angle $(\vec{p}_1, \Lambda)$ qu'on peut appeler angle de diffraction. Ensuite, pour étudier les composantes transverses, on ajustera le système des $\vec{p}_{ej}$ à un plan P ,

(*) Les suggestions de cet alinéa ont fait l'objet de la thèse de D. Mañti, citée au début du présent article.

contenant Λ et dont on notera l'angle qu'il forme avec $\vec{p}_1$... Sous l'hypothèse d'uniforme distribution dans l'espace des phases, la direction de Λ est équirépartie sur la sphère, indépendamment de la direction de $\vec{p}_1$: on voit combien l'hypothèse de longitudinalité ($\Lambda // \vec{p}_1$) est éloignée de celle d'uniforme densité ; elle en est si éloignée que la notion même d'élément de volume naturel n'est ici qu'un terme de référence lointain, qui pourrait être corrigé par l'analyse statistique de la distribution de l'angle $(\vec{p}_1, \Lambda)$ en fonction de la position du système $\{p_{\Lambda e_j}\}$ dans le diagramme de Van Hove. (Il est classique de corriger les densités calculées sous l'hypothèse d'uniforme densité, en tenant compte de la distribution du transfert de quantité de mouvement).

4.2.3 Le diagramme prismatique et ses généralisations : Dans le cas de trois particules émergentes, chaque événement est caractérisé par 4 paramètres invariants. En prenant ces paramètres deux à deux, on dira encore que i est caractérisé par ses projections sur deux plans : par exemple le point i_E du diagramme triangulaire de Dalitz (diagramme des énergies) et le point i_L du diagramme hexagonal de Van Hove (diagramme des moments longitudinaux). On s'étonnera que ces diagrammes dont aucun ne tient compte des composantes transverses (souvent faibles, mais sur lesquelles en toute rigueur on n'a rien postulé) puissent fournir une représentation complète : c'est qu'il y a entre les 12 composantes des trois quadrimoments des particules émergentes un système de relations déjà étudié au § 3.1. Semblablement, on pourrait comme le propose Miller (Genève 1976, déjà cité) utiliser les deux diagrammes triangulaires des coordonnées isotropes q^+ et q^- , mais nous n'en dirons pas plus ici.

Cependant, comme l'affirme laconiquement l'équipe du M. I. T. dans le document déjà cité (Genève 1976) "Faire travailler un ordinateur en dimension 4 est trivial ; faire voir à un physicien des sous-nuages (*clusters*) en dimension 4 ne l'est pas. Il faut un compromis". Au diagramme triangulaire de Dalitz on ajoutera une troisième coordonnée : obtenant ainsi un diagramme prismatique. Voici comment :

Sur le diagramme hexagonal de Van Hove, (comme sur le diagramme triangulaire de Dalitz ; cf fig 3-4), la sous-variété V des événements permis se projette à l'intérieur d'un contour limite. De plus les événements périphériques, i.e. à faibles moments transverses, se trouvent projetés à proximité de ce contour (cf § 4.2.2) ; ils sont ainsi par une heureuse analogie, à la périphérie du nuage. Il s'impose donc de prendre pour troisième coordonnée l'angle polaire ω du point i_L (la valeur $\omega = 0$ correspondant au vecteur $\vec{V}_p$; e.g. à un événement périphérique où le proton cible donne un nucléon émergent e_1 qui dans le système du laboratoire recule dans la direction du faisceau incident, avec plus précisément, dans le système où le centre de gravité est immobile les relations suivantes : $p_{Le1} > 0$ (de même signe que p_{Le2} , moment du proton incident dans ce système) ; $p_{Le2} = p_{Le3} = -p_{Le1}/2$; cf fig 4-4). Simultanément, on peut sur des histogrammes contrôler la distribution du rayon polaire (plus exactement de son rapport à la valeur maximale sur le rayon polaire de i_L).

Nous avons déjà montré au § 4.1.1 l'efficacité de ces diagrammes prismatiques, utilisés depuis 1970 au M.I.T. . Au colloque de Genève, R. Hulsinger a pu avec une légitime satisfaction faire état de nombreuses applications d'un intérêt certain pour la physique. Bien qu'avec la dimension supplémentaire ω , un nuage uniformément dense dans l'espace

des phases ne se projette plus avec une densité uniforme (comme c'est le cas sur le diagramme de Dalitz cf § 3.4), il est possible par simulation de comparer les nuages d'événements réels à un nuage obtenu par projection d'un nuage uniforme : on évite ainsi des artefacts. Utilisée sur des échantillons fictifs engendrés d'après les paramètres des sous-canaux qu'elle a d'abord elle-même permis de déterminer, la méthode apparaît stable et fidèle. Cependant il faut rappeler après W. Kittel (1974) que ces calculs ne tiennent pas compte des interférences : la méthode du diagramme prismatique ne suffit pas à construire (e.g. d'après les histogrammes des sous-canaux qu'on a séparés) un modèle semi-empirique comportant des phases... De ce point de vue le diagramme prismatique suggère seulement des profils de sous-canaux, auxquels on pourra appliquer la méthode de vraisemblance (cf § 4.1.2) pour en calculer l'importance relative puis en préciser la forme et concevoir un nouveau modèle.

Enfin l'équipe du M.I.T. traite suivant les principes analogues les événements à quatre particules émergentes. Puisque $3 \times 4 - 5 = 7$, il faudrait 7 dimensions ! en se bornant aux 6 variables longitudinales (quatre énergies positives de somme E_{tot} ; quatre moments longitudinaux de somme nulle) il reste encore à confronter deux diagrammes tridimensionnels : le diagramme tétraédrique des énergies et le diagramme cuboctaédrique des moments longitudinaux (cf § 4.2.2). Comme le dit philosophiquement le document du M.I.T. déjà cité "Combiner ces deux diagrammes en un seul qu'un physicien à trois dimensions puisse comprendre, peut bien être impossible"... Mais il est permis d'aller de l'un à l'autre en projetant ici une sous-classe apparue là ; etc... Toutes opérations qu'un système informatique volubile peut rendre agréables et fécondes (cf § 4.2.4)

4.2.4 Vues d'un nuage en coupes et en projections : Découvrir les sous-canaux ; puis séparer des sous-échantillons d'événements relevant de chacun de ceux-ci afin d'en préciser le profil est l'objet de toute analyse de données issues d'une expérimentation à haute énergie. Un pic apparaît-il sur un histogramme, ou une bande dense sur un diagramme triangulaire ou prismatique ? Il faut isoler les événements correspondant à ce pic ou à cette bande ; en préciser la distribution angulaire ; rechercher sur l'histogramme de la masse invariante de toute paire (ou triplet) de particules émergentes, divers signaux parasites (cf § 4.2.1) ; comparer des profils à ce qu'ils seraient sous l'hypothèse de distribution uniforme dans l'espace des phases. Chacune de ces opérations, prise isolément semble simple ; mais les répéter aussi souvent que le voudrait la curiosité du physicien est plus qu'une tâche fastidieuse : c'est une entreprise au dessus des moyens naturels dont l'homme dispose pour traiter des informations aussi nombreuses que celles de certaines et de milliers d'événements ; il y faut un ordinateur couplé avec un écran cathodique. Fort heureusement les laboratoires de physique qui ont dû automatiser grandement l'acquisition des événements que recèlent les clichés de chambre à bulles et plus encore ceux saisis par d'autres détecteurs (cf § 4.4.1), comptent des virtuoses inégalables de l'informatique. Les statisticiens qui n'attendent que des listages pour l'analyse factorielle ou la classification automatique des données de biologie végétale et animale ou de sciences humaines, croient rêver quand ils voient avec quelle rapidité les très ingénieux systèmes conçus pour l'analyse interactive des données de physique de haute énergie semblent saisir au vol dans l'espace multidimensionnel les coupes et les projections qui se succèdent sur l'écran cathodique(*). Sans décrire aucun de ces systèmes, donnons-en un aperçu par quelques exemples.

(*) Voir renvoi page suivante.

Le système PAT (*Programmed Analysis Technique* ; objet à Genève d'une communication de l'équipe du M.I.T. qui l'a conçu) présente sous un angle quelconque une vue bidimensionnelle d'un nuage tridimensionnel (cf supra : diagramme prismatique) soit fixe, soit en rotation continue. L'utilisateur peut, par un crayon lumineux, mettre à part une classe d'événements ; et demander pour un tel sous-nuage des vues partielles et des histogrammes. L'ordinateur PDP-10 commandé principalement à distance, de l'écran cathodique ou d'un clavier spécial, lit les événements sur une bande magnétique usuelle, et peut garder en mémoire rapide un maximum de 2400 événements à 3 particules émergentes ou 2000 événements à 4 particules.

Le système PRIM6 (*"Projection, Rotation, Isolation, Masking in 6 dimensions"*), dont l'application à la réaction $K^-p \rightarrow K^-p \pi^+\pi^-$ (à 14 GeV) fut présentée à Genève par P. Lutz, offre d'un nuage de points de R^6 des vues bidimensionnelles, prises en quelque sorte par une caméra en mouvement continu, avec la possibilité de couper le nuage au vol, c'est-à-dire de séparer une classe d'événements encerclée d'un trait de crayon sur une vue plane (**).

Le NAMELESS (système innommé) réalisé au CERN par E. Pagiola, était en démonstration pour les participants au colloque de Genève. Ce programme interactif d'analyse multidimensionnelle, présente avec une flexibilité impressionnante les vues planes et les histogrammes demandés au clavier. De plus il effectue un classement des événements suivant le principe suivant : l'utilisateur principalement grâce à des bornes choisies au vu des pics d'histogrammes, définit des sous-classes d'événements $N_c, N'_c, N''_c, \dots$; chaque événement de l'échantillon traité, est rattaché par le système au noyau N_c dont il est le plus proche (proximité est à entendre ici en un sens particulier : pour chaque classe on calcule un critère de compacité : l'adjonction d'un point modifie la quantité critère ; un point étant dit d'autant plus proche d'une classe qu'est plus grande l'amélioration qu'il apporte au critère calculé par la classe) ; les sous-canaux $I_c, I'_c, I''_c, \dots$, ainsi obtenus sont appréciés d'après divers histogrammes (cf § 4.2.1) ; l'utilisateur peut alors proposer de nouveaux noyaux $N'_c, N''_c, N'''_c, \dots$; le processus interactif de définition des sous-canaux est arrêté quand les transferts d'individus entre classes, consécutifs aux modifications des noyaux (transferts recensés par le système) sont si peu nombreux que l'utilisateur n'espère plus d'amélioration.

On le voit, avec le NAMELESS project nous dépassons la statistique descriptive en faible dimension pour aborder l'analyse multidimensionnelle proprement dite, objet du § 4.3.

(*) M. Kerbaol qui, à Rennes, use d'une sortie sur un écran cathodique non seulement pour voir les cartes planes issues de l'analyse factorielle, mais aussi pour appeler les dossiers relatifs à des sous-ensembles d'individus définis d'un geste du crayon lumineux sur l'écran etc, nous a toutefois confié que le flot d'informations que peut ainsi débiter même un petit ordinateur, égare plus qu'il ne l'inspire le statisticien non averti. Et J.P. Fénelon dit volontiers que les étudiants qui apprennent l'analyse des données ne devraient recevoir que des listages de nombres et être ainsi contraints à tracer de leurs mains tous les graphiques.

(**) Cf P. Lutz, Thèse (Paris Mai 1977) ; 2° Partie Ch II § B.2.