Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 1

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 1 (1960)

THE LOGARITHMIC POTENTIAL IN HIGHER DIMENSIONS

BY

BENT FUGLEDE

København 1960
i kommission hos Ejnar Munksgaard

Synopsis

Various classical potential theoretic properties of the logarithmic kernel in the plane are extended to the logarithmic kernel $-\log|x-y|$ in Euclidean n-space R^n. The key result is the following inequality for the energy of any (signed) mass distribution μ on a ball $B \subset R^n$ of radius ϱ:

$$-\int_B \int_B \log|x-y|\, d\mu(x)\, d\mu(y) \geq \log \frac{a_n}{\varrho} \cdot \left(\int_B d\mu\right)^2.$$

The best possible value of the constant a_n is determined explicitly in its dependence on the dimension n. In particular, the logarithmic kernel satisfies the energy principle on any ball of radius $\varrho < a_n$.

Printed in Denmark
Bianco Lunos Bogtrykkeri A-S

1. Introduction

In view of its role in the theory of analytic or harmonic functions, the logarithmic potential in the plane has been investigated thoroughly. Restricting the attention to the more recent literature on this subject and to the potential theoretic aspects thereof, we mention the works of O. Frostman [5], [6], M. Riesz [12], Ch. de la Vallée-Poussin [14], H. Cartan [1], and G. Choquet [3]. On the other hand, very little research seems to have been devoted to the logarithmic potential in Euclidean space R^n of dimension $n > 2$. The principle results on this topic are those of M. Riesz [12, § 4] and O. Frostman [6, § 1] concerning the logarithmic potential and energy of distributions of algebraic total mass zero, and further the calculation of the Fourier-Schwartz transform of the logarithmic kernel, cf. L. Schwartz [13, ch. VII, § 7] or J. Deny [4, note 3, p. 160 f.].

In the present paper we continue the study of the logarithmic kernel in R^n for arbitrary dimension n; that is, the kernel

$$-\log|x-y| \qquad (x \varepsilon R^n,\ y \varepsilon R^n),$$

interpreted as $+\infty$ for $x = y$. We shall use the terminology[1] and some of the results of a previous memoir [7]. Most of the results of the present paper are applied in a recent article [8]. Of independent interest is the main result asserting that the logarithmic kernel is *strictly* (positive) *definite* (that is, it satisfies the energy principle) when considered on a ball $A \subset R^n$ of sufficiently small radius a (cf. de la Vallée-Poussin [14, § 47] for the case $n = 2$). The least upper bound a_n of such radii is determined explicitly (§ 4, formula (5)). The proof is based on an explicit computation of the *equilibrium distribution* (in the sense of Deny [4, § 5]) on the unit ball in R^n. Combining this result with the known fact that the logarithmic kernel

[1] Observe, however, that the notations $k(x, \mu)$ and $k(\mu, \nu)$ for potential and mutual energy in [7] will be replaced by $U^{\mu}_{\alpha}(x)$ and $\langle \mu, \nu \rangle$, respectively, in the present paper (in which $k(x, y) = -\log|x-y|$).

is *regular* (that is, it satisfies the continuity principle), it follows, in essence from a theorem of M. Ohtsuka [10], that the logarithmic kernel is *perfect* in the sense of [7, § 3.3] when considered on a ball of radius $a \leqq (=) a_n$. This is Theorem 4.1 of the present paper. Using [7], one derives various corollaries from this theorem, in particular the existence of an interior or exterior capacitary distribution associated with any given bounded set, and further the capacitability of all bounded analytic subsets of R^n. This last result, which depends strongly on Choquet's theory of capacitability [2], was known previously for $n = 2$ (cf. Choquet [3], whose proof is based on special properties of the logarithmic potential in the plane). Further results involving the logarithmic potential or the logarithmic capacity are obtained in [8].

Since the logarithmic kernel is of variable sign, we shall consider the logarithmic potential and energy only of distributions of *compact support*. This limitation will not always be repeated. We shall mainly deal with distributions which are *measures* (not necessarily positive), but general distributions in the sense of Schwartz [13] will enter in the proof of the key result (Lemma 4.1).

2. Basic notions connected with the logarithmic kernel

The logarithmic *potential* of a measure μ on R^n (of compact support) is defined by

$$U^{\mu}(x) = -\int \log|x-y|\, d\mu(y) = U^{\mu^+}(x) - U^{\mu^-}(x)$$

at any point x for which the third member is defined (i. e., not of the form $(+\infty)-(+\infty)$). In particular, $U^{\mu}(x)$ is always defined and $\neq -\infty$ if $\mu \geq 0$.

The logarithmic *mutual energy* $\langle \mu, \nu \rangle$ of two measures μ and ν (both of compact support) is defined by

$$\begin{aligned}\langle \mu, \nu \rangle &= -\iint \log|x-y|\, d\mu(x)\, d\nu(y) \\ &= \langle \mu^+, \nu^+ \rangle + \langle \mu^-, \nu^- \rangle - \langle \mu^+, \nu^- \rangle - \langle \mu^-, \nu^+ \rangle,\end{aligned}$$

provided the third member is meaningful. In particular, $\langle \mu, \nu \rangle$ is always defined and $\neq -\infty$ if $\mu \geq 0$, $\nu \geq 0$. For $\mu = \nu$ we obtain the logarithmic *energy* $\langle \mu, \mu \rangle$ of a measure μ. An application of Fubini's theorem leads to the *formula of reciprocity*

$$\langle \mu, \nu \rangle = \int U^{\mu}(x)\, d\nu = \int U^{\nu}(x)\, d\mu,$$

valid whenever $\langle \mu, \nu \rangle$ is defined (cf. [7, § 2.1] for details).

In several respects the logarithmic kernel may be viewed as a limit case of the kernels of order α, $|x-y|^{\alpha-n}$, as $\alpha \to n$. This appears, e. g., from the identity

$$\log|x| = \left\{(\partial/\partial\alpha)\,|x|^{\alpha-n}\right\}_{\alpha=n}. \tag{1}$$

As observed by M. Riesz [12, § 4], the analogy is almost perfect when the measure μ in question has algebraic total mass 0, $\int d\mu = 0$. Note, in particular, the following formula due to M. Riesz (cf. also Frostman [5, § 33] and [6, § 1]):

$$\langle \mu, \mu \rangle = \frac{1}{\omega_n} \int [U^{\mu}_{n/2}]^2\, dx \,(\geqq 0) \quad \text{if} \quad \int d\mu = 0 \tag{2}$$

under the additional assumption that $\langle \mu, \mu \rangle$ exists and is finite.[1] Here

$$\omega_n = 2\pi^{n/2}/\Gamma(n/2)$$

denotes the surface of the unit sphere in R^n.

The *interior logarithmic capacity* $\gamma_*(E)$ of an arbitrary bounded set $E \subset R^n$ is defined by

$$-\log \gamma_*(E) = w(E);\ \text{i. e.,}\ \gamma_*(E) = \exp(-w(E)). \tag{3}$$

Here

$$w(E) = \inf \langle \mu, \mu \rangle \tag{4}$$

as μ ranges over the class of all positive measures of compact support contained in E and of total mass $\int d\mu = 1$. Cf. [7, § 2.3].

If E is compact, this infimum (4) is an actual minimum (cf. [7, Theorem 2.3]), attained by precisely one competing measure λ called the *capacitary distribution of unit mass* on E. (The uniqueness follows from (2) as explained in Remark 2 to Theorem 2.4 in [7], because the difference $\mu - \nu$ between any two competing measures is of zero total mass. Moreover, $U^{\mu}_{n/2} = 0$ almost everywhere implies $\mu = 0$ according to the uniqueness theorem of M. Riesz for the potentials of order α, cf. [12, § 10]). The logarithmic potential of this capacitary distribution λ has the following properties (cf. [7, Theorem 2.4]):

[1] By U^{μ}_{α} we denote the potential of order α of μ, that is, the potential of μ with respect to the kernel $|x-y|^{\alpha-n}$ of order α in R^n, $0 < \alpha < n$. If μ has a density f, that is, $d\mu = f(x)\,dx$, we may write U^{f}_{α} in place of U^{μ}_{α}.

(a) $U^{\lambda} \geq w(E)$ nearly everywhere[1] in E,

(b) $U^{\lambda} \leq w(E)$ everywhere in the support of λ.

For $n \leq 2$, the logarithmic kernel fulfills Frostman's maximum principle, and hence (a) and (b) may be replaced by: $U^{\lambda} = w(E)$ nearly everywhere in E, and $U^{\lambda} \leq w(E)$ everywhere in R^n, respectively. For $n \geq 3$ we have the following substitute for this latter inequality:

$$U^{\lambda} \leq w(E) + \log 2 \qquad \text{everywhere in } R^n \tag{5}$$

(cf. [8, § 2, formula (8)]). Likewise for arbitrary dimension n, the inequality $U^{\lambda} \geq w(E)$ holds everywhere in the interior of E. This may be proved in the manner devised by Frostman [5, p. 37] for the potentials of order α.

It is well known (cf. e. g., [7, § 2.3]) that, for any bounded set $E \subset R^n$,

$$\gamma_*(E) = \sup_K \gamma_*(K) \qquad (K \text{ compact}, \ K \subset E). \tag{6}$$

The *exterior* logarithmic capacity $\gamma^*(E)$ of an arbitrary bounded set E is defined by

$$\gamma^*(E) = \inf_G \gamma_*(G) \qquad (G \text{ open}, \ G \supset E). \tag{7}$$

A bounded set E is called *capacitable* (with respect to the logarithmic kernel) if $\gamma^*(E) = \gamma_*(E)$. If E is capacitable, we may write simply $\gamma(E)$ for the logarithmic capacity $\gamma^*(E) = \gamma_*(E)$ of E. This is the case, in particular, if E is open or compact (cf. e. g., [7, p. 153 f.]).

3. A substitute for M. Riesz' composition formula and its applications

The following lemma coincides in essence with a formula stated in Frostman [5, p. 61]. It serves as a substitute for the important composition formula of M. Riesz for the kernels of order α in R^n:

$$\int_{R^n} |x - z|^{\alpha - n} |z - y|^{\beta - n} dz = c_{\alpha, \beta} |x - y|^{\alpha + \beta - n},$$

valid for $\alpha + \beta < n$. The formula to be discussed here corresponds to the limit case $\alpha + \beta = n$.

[1] The expression "nearly everywhere in E" means "everywhere in E except possibly in some set $N \subset E$ for which $\gamma_*(N) = 0$". Replacing γ_* by γ^*, we arrive at an analogous concept called "quasi-everywhere".

LEMMA 3.1. *Let A and B denote two concentric closed balls in R^n of radii ϱ (fixed) and $R > 2\varrho$, respectively. The function $\psi(x, y; R)$ of $x \varepsilon A$ and $y \varepsilon A$ defined by*

$$\frac{1}{\omega_n}\int_B |x-z|^{\alpha-n}\,|z-y|^{-\alpha}\,dz = \log\frac{R}{|x-y|} + \psi(x, y; R)$$

is continuous on $A \times A$ and approaches a certain constant c uniformly in $A \times A$ as $R \to +\infty$.

Proof. We may of course suppose that the common centre of A and B is the origin 0, and so $B = \{z \varepsilon R^n : |z| \leq R\}$. We begin by studying the case $y = 0$. Introducing polar coordinates, whereby $dz = |z|^{n-1}\,d|z|\,dw$, we obtain for reasons of homogeneity, writing $t = |x|/|z|$,

$$\frac{1}{\omega_n}\int_B |x-z|^{\alpha-n}\,|z|^{-\alpha}\,dz = \int_{r/R}^{\infty} t^{-1}\,u_\alpha(t)\,dt,$$

where $r = |x|$, and where $u_\alpha(t)$ denotes the potential of order α of the uniform distribution of unit mass on the unit sphere in R^n, evaluated at a point of distance t from the origin. Clearly, $u_\alpha(t)$ is differentiable for $t > 1$ and for $0 \leq t < 1$, and integrable over a neighbourhood of $t = 1$. Moreover,

$$u_\alpha(0) = 1; \quad u'_\alpha(0) = 0; \quad u_\alpha(t) = O(t^{\alpha-n})$$

as $t \to +\infty$. Hence the function v_α defined by

$$v_\alpha(t) = \begin{cases} t^{-1}\,u_\alpha(t) & \text{for } t > 1 \\ t^{-1}\,(u_\alpha(t) - 1) & \text{for } 0 < t < 1 \end{cases}$$

is bounded near 0 and integrable over $(0, +\infty)$. We now obtain

$$\frac{1}{\omega_n}\int_B |x-z|^{\alpha-n}\,|z|^{-\alpha}\,dz = \log\frac{R}{r} + V\left(\frac{r}{R}\right),$$

where $V(t) = \int_t^{\infty} v_\alpha(s)\,ds$ is continuous and approaches the limit

$$c = \int_0^{\infty} v_\alpha(t)\,dt$$

as $t \to 0$.–In the general case, let $|x| \leq \varrho$, $|y| \leq \varrho$, and $R > 2\varrho$. We compare the integral of $f(z) = |x-z|^{\alpha-n}\,|z-y|^{-\alpha}$ over $B = \{z \varepsilon R^n : |z| \leq R\}$ with the integral of $f(z)$ over the ball $B' = \{z \varepsilon R^n : |z-y| \leq R\}$ of centre y. Since the poles x and y belong to both balls (because $|x-y| \leq 2\varrho < R$), the two integrals differ by a continuous function on $A \times A$, viz.

$$F(x, y; R) = \frac{1}{\omega_n}\int_{B-B'} f(z)\,dz - \frac{1}{\omega_n}\int_{B'-B} f(z)\,dz.$$

As the integrand f is $O(R^{-n})$, and the volumes of the sets $B-B'$ and $B'-B$ are $O(R^{n-1})$, we infer that $F(x, y; R)\to 0$ for $R\to\infty$, uniformly for $x\varepsilon A$, $y\varepsilon A$. Summing up, we obtain the representation

$$\frac{1}{\omega_n}\int_B |x-z|^{\alpha-n}\,|z-y|^{-\alpha}\,dz = \log\frac{R}{|x-y|} + V\left(\frac{|x-y|}{R}\right) + F(x, y; R),$$

from which the assertions of the lemma follow because

$$\psi(x, y; R) = V\left(\frac{|x-y|}{R}\right) + F(x, y; R).$$

It follows, in particular, from Lemma 3.1 that the logarithmic kernel $-\log|x-y|$ and the kernel

$$A_B(x, y) = \frac{1}{\omega_n}\int_B |x-z|^{\alpha-n}\,|z-y|^{-\alpha}\,dz$$

(considered on A) differ by the continuous function

$$\log R + \psi(x, y; R)$$

of $(x, y)\ \varepsilon A\times A$. We denote the supremum of the absolute value of this latter function over the compact set $A\times A$ by $M = M(R, \varrho)$. It depends on R, ϱ, α, and n. Taking, e. g., $R = 3\varrho$, we obtain a constant $M(3\varrho, \varrho)$ which we shall denote simply by $M(\varrho)$. Thus

$$A_B(x, y) - M(\varrho) \le \log\frac{1}{|x-y|} \le A_B(x, y) + M(\varrho) \tag{1}$$

for $x\varepsilon A$, $y\varepsilon A$. In particular, the class $\mathfrak{E}$ of measures μ, supported by A, whose energy is defined and finite, is the same for the two kernels $-\log|x-y|$ and $A_B(x, y)$. Since A_B is a definite kernel on A (cf. [7], § 3.5), we conclude that $\mathfrak{E}$ is a vector-space and that the class $\mathfrak{E}^+$ of positive measures in $\mathfrak{E}$ is a convex cone. Moreover, the logarithmic energy

$$\langle\mu, \mu\rangle = \iint \log\frac{1}{|x-y|}\,d\mu(x)\,d\mu(y)$$

is $+-\infty$. It is, in fact, $\ge -M(\varrho)\left\{\int|d\mu|\right\}^2$. This result will be improved considerably in § 4. Observe also that the logarithmic mutual energy $\langle\mu, \nu\rangle$ is defined and finite if $\mu, \nu\ \varepsilon\ \mathfrak{E}$, cf. [7, § 3.1].

The following two lemmas will not be used in the sequel. They are included on account of their role in [8, § 7]. The notations are those used in the preceding lemma (say, with $R = 3\varrho$). The characteristic function associated with a set E is denoted by φ_E. See also note 1, p. 5.

LEMMA 3.2. *Let μ denote a measure supported by the ball A, and put $f = \varphi_B \cdot U^{\mu}_{n-\alpha}$. Then the inequalities*

$$\frac{1}{\omega_n} U^f_\alpha - M(\varrho) \int |\, d\mu\,| \leq U^\mu \leq \frac{1}{\omega_n} U^f_\alpha + M(\varrho) \int |\, d\mu\,|$$

hold at any point of A at which the logarithmic potential U^μ is defined (hence everywhere in A if $\mu \geq 0$).

LEMMA 3.3. *Let μ denote a measure supported by A and of finite logarithmic energy $\langle \mu, \mu \rangle$. Then*

$$\left| \langle \mu, \mu \rangle - \frac{1}{\omega_n} \int_B [U^\mu_{n/2}]^2 \, dx \right| \leq M(\varrho) \left\{ \int |\, d\mu\,| \right\}^2 .$$

Each of these lemmas is derived from (1), or directly from Lemma 3.1, by integration with respect to $d\mu(y)$ (in Lemma 3.2) and $d\mu(x)d\mu(y)$ (in Lemma 3.3), followed by an application of Fubini's theorem. Thus we obtain

$$\frac{1}{\omega_n} \int_B [U^\mu_{n/2}(z)]^2 \, dz = \langle \mu, \mu \rangle + \iint [\log R + \psi(x, y; R)] \, d\mu(x) d\mu(y), \qquad (2)$$

from which Lemma 3.3 follows. Similarly in case of Lemma 3.2. In the special case where $\int d\mu = 0$ we arrive, following FROSTMAN [5, p. 61 f.], at the identity (2), § 2, when we let $R \to \infty$ in (2) under observation of the final assertion of Lemma 3.1.

4. The perfectness of the logarithmic kernel

LEMMA 4.1. *For any ball A of sufficiently small radius a, the restriction of the logarithmic kernel $-\log|x-y|$ to $A \times A$ is definite.*

Proof. Simple considerations of homogeneity will show that this property of the radius a is equivalent to the following inequality, valid for all measures (even of variable sign) of finite logarithmic energy, concentrated on some ball of arbitrary given radius ϱ:

$$\langle \mu, \mu \rangle \geq \log(a/\varrho) \cdot \left(\int d\mu \right)^2 . \qquad (1)$$

Moreover it suffices to prove this inequality in the case $\varrho = 1$ of the unit ball B_1. The idea of the proof is classical in the case $n = 2$ (cf. De la Vallée-Poussin [14, § 47] and Deny [4, p. 164]. It consists in producing a measure λ with $\int d\lambda = 1$ whose logarithmic potential U^λ is constant, say $= L$, everywhere in the unit ball B_1. If μ denotes any measure of finite energy concentrated on B_1, and if we write $m = \int d\mu$, then $\mu - m\lambda$ is likewise of finite energy, and since its algebraic total mass is 0, its logarithmic energy is ≥ 0 according to (2), § 2. Evaluating this energy, we get

$$\langle \mu, \mu \rangle - 2m \int U^\lambda d\mu + m^2 \langle \lambda, \lambda \rangle \geq 0,$$

and hence

$$\langle \mu, \mu \rangle \geq (2L - \langle \lambda, \lambda \rangle) m^2. \tag{2}$$

The existence of a measure λ (of compact support) with the stated properties: $U^\lambda =$ constant $(= L)$ on B_1, $\int d\lambda = 1$, can be proved as follows. For $n = 1$ or $n = 2$, the logarithmic kernel fulfills the maximum principle, and hence the capacitary distribution λ of unit mass on B_1 has the desired properties (cf. § 2), and we get $L = w(B_1) = \langle \lambda, \lambda \rangle$. This leads to the largest possible value a_n of a (in the case $n \leq 2$): $a_n = \exp(w(B_1))$. For $n = 2$, λ is simply the uniform distribution of unit mass on the unit circle, and hence $w(B_1) = 0$ (= the value of U^λ at the centre 0). This gives $a_2 = 1$. For $n = 1$, it can be shown that λ has the density τ given by $\tau(x) = 0$ for $|x| > 1$ and

$$\tau(x) = \pi^{-1}(1 - x^2)^{-1/2} \quad \text{for} \quad |x| < 1;$$

and this leads to $w(B_1) = \log 2$, $a_1 = 2$ (cf. below).

For $n \geq 3$, the capacitary distribution on the unit ball B_1 has no longer a constant logarithmic potential in B_1, and so the existence of a measure λ with constant U^λ in B_1 (and $\int d\lambda = 1$) must be verified in a different manner. Although it is possible to do this in an elementary way, we shall prefer to make use of the theory of distributions and at the same time determine explicitly the *best possible value* a_n of a. We propose to determine explicitly an *equilibrium distribution* T on the unit ball B_1 in R^n, that is, a distribution in the sense of Schwartz [13], supported by B_1, having the total integral $T(1) = 1$, and possessing a logarithmic potential which is constant on B_1. This equilibrium distribution T may then replace λ in the preceding argument in the case $n \leq 2$ (in which case, actually, $T = \lambda$).[1]

[1] If, nevertheless, we insist upon constructing a *measure* λ with the desired properties, we merely have to "regularize" T by subjecting it to a homothetic transformation of R^n with respect to the origin and of a ratio $1 + r > 1$, followed by a convolution with some infinitely differentiable function $\varphi \geq 0$, $\int \varphi(x)dx = 1$, supported by the ball of radius r about the origin. The logarithmic potential of the measure λ obtained in this manner has the constant value $\log [a_n/(1+r)]$ in the unit ball B_1.

We begin by solving the corresponding problem for the potentials of order α in R^n instead of the logarithmic potential. For $0 < \alpha < 2$, the equilibrium distribution on B_1 is the positive measure T_α whose density τ_α is given by $\tau_\alpha(x) = 0$ for $|x| > 1$ and

$$\tau_\alpha(x) = \frac{2}{\omega_n} \frac{\Gamma(1-\alpha/2+n/2)}{\Gamma(1-\alpha/2)\Gamma(n/2)} (1-|x|^2)^{-\alpha/2}$$

for $|x| < 1$. The constant value of the potential $U_\alpha^{T_\alpha}$ within B_1 coincides with the energy of order α of T_α. The common value is

$$u_\alpha = \frac{\Gamma(\alpha/2)\Gamma(1-\alpha/2+n/2)}{\Gamma(n/2)}. \tag{3}$$

These results may be verified in the manner described in PÓLYA and SZEGÖ [11] for the case $n \leq 3$ (cf. also M. RIESZ [12, § 16] for the general case).— For an arbitrary value of α, the equilibrium distribution of order α on the unit ball B_1 in R^n can be obtained by analytic continuation of the above distribution T_α, and the constant value u_α of the potential $U_\alpha^{T_\alpha}$ within B_1 is given again by (3). (The "spectral measure" of B_1 is, therefore, $1/u_\alpha$; cf. DENY [4, p. 127].). For $\alpha = 2$ we find T_2 = the uniform distribution of unit mass on the unit sphere. For $\alpha > 2$, T_α is no longer a measure, but can be expressed as a "finite part" in the sense of Hadamard and Schwartz. (For $\alpha = 2k$, $k = 1, 2, \ldots$, T_α is a "multilayer" of order k on the unit sphere, cf. DENY [4, p. 129].)

Next we pass to the logarithmic potential by a differentiation with respect to the order α at $\alpha = n$ (cf. (1), § 2). If we apply the operator $-(\partial/\partial\alpha)_{\alpha=n}$ to both sides of the equation

$$U_\alpha^{T_\alpha} = u_\alpha \text{ in } B_1,$$

we obtain on the left the logarithmic potential of T_n. (The additional term is the total integral of $-(\partial T_\alpha/\partial\alpha)_{\alpha=n}$, and this vanishes because $T_\alpha(1) = 1$ for every α). The resulting equation

$$U^{T_n} = -\{\partial u_\alpha/\partial\alpha\}_{\alpha=n} \text{ in } B_1 \tag{4}$$

shows that $T = T_n$ *is the equilibrium distribution on the unit ball* B_1 in R^n, corresponding to the logarithmic kernel. Similarly, the logarithmic energy of $T = T_n$ is $-\{\partial u_\alpha/\partial\alpha\}_{\alpha=n}$. The largest possible value a_n of the radius a in Lemma 4.1 is now determined by

$$-\log a_n = \{\partial u_\alpha/\partial\alpha\}_{\alpha=n} = \frac{1}{2}\Psi(n/2) - \frac{1}{2}\Psi(1),$$

where $\Psi(t) = \Gamma'(t)/\Gamma(t)$. Explicitly,

$$\log \frac{1}{a_n} = \begin{cases} (n-2)^{-1} + (n-4)^{-1} + \ldots + 2^{-1} & \text{for even } n, \\ (n-2)^{-1} + (n-4)^{-1} + \ldots + 1^{-1} - \log 2 & \text{for odd } n. \end{cases} \tag{5}$$

THEOREM 4.1. *The logarithmic kernel* $-\log|x-y|$ *is perfect when considered on a closed ball* $A \subset R^n$ *of radius* $\varrho < a_n$.

Proof. The restriction of the logarithmic kernel to such a ball is *strictly definite* according to the inequality (1) together with the fact that the logarithmic energy of a measure of compact support is either finite or $+\infty$ (if at all defined), cf. § 3. Moreover, the logarithmic kernel in R^n is *regular* (i. e., satisfies the principle of continuity) by virtue of Kametani's theorem (cf. KUNUGUI [9, p. 78]); and so is therefore the restriction of $-\log|x-y|$ to $A \times A$. It follows from these two properties that this restriction is *consistent*, and hence perfect, cf. [7, Theorems 3.4.1 and 3.3].—Actually, the assertion of the theorem remains valid in the case $\varrho = a_n$ provided $n \geq 3$, because the sign of equality in (1) never occurs for any *measure* μ, but only for the equilibrium distribution T_n which is not a measure when $n \geq 3$.

In view of this perfectness of the logarithmic kernel (considered on A), the logarithmic potential of measures supported by A has all the properties described in [7, Chapter II]. First of all, we may introduce the interior and exterior *Wiener capacity* of arbitrary sets $E \subset A$:

$$\operatorname{cap}_* E = 1/w(E) = -1/\log \gamma_*(E), \tag{6}$$

$$\operatorname{cap}^* E = -1/\log \gamma^*(E). \tag{7}$$

Next, we may consider the (unique) interior and exterior *capacitary distributions* associated with an arbitrary set $E \subset A$, cf. [7, § 4]; and finally we may apply CHOQUET's theory of capacitability [2]. We prefer to state the results thus obtained in terms of the *logarithmic* capacity (instead of the Wiener capacity) and the capacitary distributions of *unit mass*. In this way we avoid the limitation to subsets of A; the extension to arbitrary bounded sets is simply a matter of applying a homothetic transformation, and using the fact that, for any constant $k > 0$, the kernels $\log(k/|x-y|)$ and $\log(1/|x-y|)$ differ by the additive constant $\log k$.

THEOREM 4.2. *To any bounded set* $E \subset R^n$ *corresponds a unique measure* λ *with*

$$\int d\lambda = 1, \ \langle \lambda, \lambda \rangle = w(E) = -\log \gamma_*(E),$$

whose logarithmic potential has the following properties

(a) $U^{\lambda} \geqq w(E)$ nearly everywhere in E,

(b) $U^{\lambda} \leqq w(E)$ everywhere in the support of λ.

This measure λ is called the *interior capacitary distribution of unit mass* associated with E. There is a similar *exterior capacitary distribution* of unit mass, whereby $w(E)$ should be replaced by $w^*(E) = -\log \gamma^*(E)$, and the term "nearly everywhere" by "quasi-everywhere". If E is capacitable, these two capacitary distributions coincide. This is the case, in particular, if E is compact, in which case λ is supported by E and coincides with the capacitary distribution of unit mass on E discussed in § 2.—Returning to the two capacitary distributions associated with an arbitrary bounded set E, we finally observe that, as in § 2, properties (a) and (b) imply

$$U^{\lambda} \geqq w(E) \quad \text{everywhere in the interior of } E, \tag{8}$$

and

$$U^{\lambda} \leqq w(E) + \log 2 \quad \text{everywhere in } R^n. \tag{9}$$

THEOREM 4.3. *If a bounded set $E \subset R^n$ is the union of an increasing sequence of sets E_p, then*

$$\gamma^*(E) = \lim_p \gamma^*(E_p).$$

This follows from Theorem 4.1 in view of [7, Theorem 4.4] in the case where E is contained in a ball A of radius $\varrho < a_n$. In the general case we apply first a suitable homothetic transformation as described above.—In the terminology of CHOQUET [2, § 15.3], this result means that the logarithmic capacity $\gamma(K)$ is *alternating* of order 1, a (when considered as defined on the class of all compact subsets K of, say, a fixed ball in R^n). Applying CHOQUET [2, § 30.2], we therefore obtain the following conclusion:

THEOREM 4.4. *Every bounded analytic set* (in particular every bounded Borel set) *$E \subset R^n$ is capacitable with respect to the logarithmic kernel $-\log|x-y|$ in R^n:*

$$\gamma^*(E) = \gamma_*(E).$$

As mentioned in the introduction, the case $n = 2$ of Theorems 4.3 and 4.4 was settled by CHOQUET [3] even without the restrictions of boundedness. It is not known to the present author whether, for $n > 2$, these two theorems would subsist if the boundedness restrictions were dropped. (One could define $\gamma_*(E)$ and $\gamma^*(E)$ for arbitrary sets E by (6) and (7), § 2, respectively.)

References

[1]. H. Cartan, *Théorie du potentiel Newtonien.* Bull. Soc. Math. France, 73 (1945), 74—106.

[2]. G. Choquet, *Theory of capacities.* Ann. Inst. Fourier, 5 (1953—54), 131—295.

[3]. G. Choquet, *Capacitabilité en potentiel logarithmique.* Bull. Acad. Roy. Belgique, (5) 44 (1958), 321—326.

[4]. J. Deny, *Les potentiels d'énergie finie.* Acta Math., 82 (1950), 107—183.

[5]. O. Frostman, *Potentiel d'équilibre et capacité des ensembles.* Comm. Sém. Math. Lund, 3 (1935).

[6]. O. Frostman, *Potentiels de masses à somme algébrique nulle.* Kungl. Fysiogr. Sällsk. Lund Förh., 20 (1950), 1—21.

[7]. B. Fuglede, *On the theory of potentials in locally compact spaces.* Acta Math., 103 (1960), 139—215.

[8]. B. Fuglede, *On generalized potentials of functions in the Lebesgue classes.* To appear in Math. Scand., 8 (1960).

[9]. K. Kunugui, *Étude sur la théorie du potentiel généralisé.* Osaka Math. J., 2 (1950), 63—103.

[10]. M. Ohtsuka, *Sur un espace complet de mesures positives dans la théorie du potentiel.* Proc. Jap. Acad., 32 (1956), 311—313.

[11]. G. Pólya und G. Szegö, *Über den transfiniten Durchmesser* ... J. reine u. angew. Math., 165 (1931), 4—49.

[12]. M. Riesz, *Intégrales de Riemann-Liouville et potentiels.* Acta sci. math. Szeged, 9 (1938), 1—42.

[13]. L. Schwartz, *Théorie des distributions.* Paris, 1950—51.

[14]. Ch. de la Vallée-Poussin, *Le potentiel logarithmique.* Paris-Louvain, 1949.

Indleveret til Selskabet den 8. september 1960.
Færdig fra trykkeriet den 31. december 1960.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 2

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 2 (1961)

A STUDY OF NUCLEAR POTENTIAL ENERGY SURFACES AND GAMMA VIBRATIONS

BY

DANIEL R. BÈS

København 1961
i kommission hos Ejnar Munksgaard

CONTENTS

Synopsis

The theory of the collective properties of the nuclear shell model has progressed recently due to the introduction of the simple pairing force to simulate the residual nucleonic interaction. Working within the framework of the adiabatic approximation, the present paper studies the consequences of this model for the γ-dependent terms of the nuclear potential energy surface. The simplified case of nucleons in a harmonic oscillator potential is considered first. Then, the energies and transition probabilities are calculated for γ-vibrations of deformed nuclei of axial symmetric shape. In addition, numerical calculations, based on realistic wave functions for nucleons in deformed nuclei, have been performed in a few cases and are compared with empirical data.

Printed in Denmark
Bianco Lunos Bogtrykkeri A-S

I. Introduction

In recent years, promising progress has been made in deriving the nuclear collective properties, starting from a description in terms of independent-particle motion.

On the one hand, it has been recognized that the shell model binding field may be associated with the effect of the long range part of the nucleonic interaction. In particular, the deformations of ellipsoidal shape can be shown to be a consequence of the quadrupole component of the effective two-body force[1][2]. On the other hand, it was recognized that there are important effects of this force which cannot be incorporated into a smoothly varying binding field, such as the inertial properties of the collective motion[3] or the potential energy of the nuclear deformation.

To represent this "residual" force, an interaction of especially simple properties has been suggested[4]. This is the so-called "pairing force" which is analogous to that used in the recent theory of superconductivity[5], and which is a generalization of the force in terms of which seniority is defined[6]. Preliminary investigations have shown that such a nuclear model contains many of the qualitative features of the observed nuclear spectra[7][8][9][10]. A more quantitative test of this model has been performed for nuclei in the regions near closed shells[11].

The aim of the present investigation is to study in greater detail some of the features of the nuclear potential energy surface which follow from this model. In particular, we consider the dependence on the parameter γ, which describes the departure from axial symmetry of an ellipsoidal nuclear deformation. We also investigate the properties of vibrations in the γ-coordinate, which are expected for nuclei of spheroidal shape.

For a quantitative analysis of the collective nuclear properties it is necessary to start from a nuclear shell model with the appropriate single-particle level spacings and wave functions. However, in order to explore some of the qualitative features, we first consider the simplified case of a harmonic oscillator well. Subsequently, we present some calculations based on a realistic single-particle spectrum, and compare the results with experimental data.

II. Formulation of the model

The basic assumption of many studies of nuclear structure is that it is sufficient to consider the degrees of freedom associated with the particles outside closed shells, the particles within closed shells manifesting themselves only through the Pauli principle and through a renormalization of the effective interparticle force.

In this chapter, we give (a) a brief description of the solution of the problem of identical particles moving in a certain shell coupled by the pairing force; (b) a discussion of the deformed part of the single-particle field; and (c) the calculation of the potential energy surface and mass parameter.

a) *The independent quasi-particle approximation.*

We assume that the matrix of the single-particle Hamiltonian has been diagonalized and that ε_ν is the eigenvalue corresponding to the degenerate single-particle states labelled by $\nu+$ and $\nu-$. These states are related by the operation of time reversal. Using the formalism of second quantization, the total single-particle Hamiltonian can be written

$$H_{\mathrm{sp}} = \sum_{\nu} \varepsilon_\nu \, (c^{\dagger}_{\nu+} c_{\nu+} + c^{\dagger}_{\nu-} c_{\nu-}). \tag{1}$$

Here, $c^{\dagger}_{\nu}$ and c_ν are, respectively, the creation and annihilation operators for the single-particle state ν. They obey the usual anti-commutation relations.

In this formalism, the pairing force is given by

$$H_{\mathrm{pair}} = - G \sum_{\nu, \omega} c^{\dagger}_{\nu+} c^{\dagger}_{\nu-} c_{\omega-} c_{\omega+}. \tag{2}$$

The lowest eigenvalue of the total Hamiltonian $H = H_{\mathrm{sp}} + H_{\mathrm{pair}}$ can be approximated by means of a variational procedure. One uses a trial function[5]

$$| 0 \rangle = \prod_{\nu} [U_\nu + V_\nu c^{\dagger}_{\nu+} c^{\dagger}_{\nu-}] \, | \text{vacuum} \rangle. \tag{3}$$

The condition $U_\nu^2 + V_\nu^2 = 1$ ensures that the wave function (3) is normalized. From (3) it is seen that V_ν^2 is the probability that the states $\nu+$ and $\nu-$ are occupied. The V_ν are variational parameters to be determined by the condition that they minimize $\langle 0 \mid H \mid 0 \rangle$. This leads to the equation

$$2/G = \sum_{\nu} (\varepsilon_\nu^2 + \Delta^2)^{-1/2}, \tag{4}$$

where $\Delta = G \sum_{\nu} U_\nu V_\nu$. The excitation spectrum has energies $E_\nu + E_\omega$, where

$$E_\nu = (\varepsilon_\nu^2 + \Delta^2)^{1/2}. \tag{5}$$

In consequence, the energy of the first excited state is always greater than 2Δ.

An elegant formulation, equivalent to the above procedure, has been developed by BOGOLUBOV[12] and by VALATIN[13]. It has been applied to nuclei by BELYAEV[7]. We now summarize some of his results. One starts by introducing two new operators $(\alpha_\nu^\dagger, \beta_\nu^\dagger)$ defined by the following canonical transformation

$$\left.\begin{aligned} \alpha_\nu^\dagger &= U_\nu c_{\nu+}^\dagger - V_\nu c_{\nu-} \\ \beta_\nu^\dagger &= U_\nu c_{\nu-}^\dagger + V_\nu c_{\nu+}. \end{aligned}\right\} \tag{6}$$

Because this is a canonical transformation, the new operators obey the same anti-commutation relations as the old. Thus, they can be regarded as creation operators for "quasi-particles" obeying Fermi statistics.

By means of the transformation inverse to (6), we can express H in terms of $\alpha_\nu^\dagger$, $\beta_\nu^\dagger$, β_ν, and α_ν. Using the anti-commutation relations, H can be put into normal form, i. e., with the $\alpha_\nu^\dagger$, $\beta_\nu^\dagger$ to the left of the β_ν, α_ν. H has then the following structure:

$$H = U + H_{11} + H_{20} + H_{\text{int}}. \tag{7}$$

The term U is a constant. H_{11} contains terms proportional to $(\alpha_\nu^\dagger \alpha_\nu + \beta_\nu^\dagger \beta_\nu)$, H_{20} terms proportional to $(\alpha_\nu^\dagger \beta_\nu^\dagger + \beta_\nu \alpha_\nu)$. H_{int}, the remainder, is supposed to have a small influence on the properties of at least the lowest states. The requirement that the coefficient of $(\alpha_\nu^\dagger \beta_\nu^\dagger + \beta_\nu \alpha_\nu)$ vanishes leads to (4).

If H_{int} is neglected, the remaining

$$U + H_{11} = \sum_{\nu} \varepsilon_\nu 2V_\nu^2 - \Delta^2/G + \sum_{\nu} E_\nu (\alpha_\nu^\dagger \alpha_\nu + \beta_\nu^\dagger \beta_\nu) \tag{8}$$

describes a system of non-interacting quasi-particles. The single quasi-particle energies are given by (5). The wave functions can be characterized by the number of quasi-particles present. In particular, the ground state has no quasi-particles. Expressed in terms of the original particle-creation operators, it is just the state (3), so that

$$\alpha_\nu \mid 0 \rangle = \beta_\nu \mid 0 \rangle = 0. \tag{9}$$

The excited states all have even numbers of quasi-particles. Those with two quasi-particles are denoted by

$$\mid \nu\omega \rangle = \alpha_\nu^\dagger \beta_\omega^\dagger \mid 0 \rangle. \tag{10}$$

Unfortunately, the solutions (3) and (10) are not eigenstates of the operator representing the number of particles

$$n_{\mathrm{op}} = \sum_{\nu} (c^{\dagger}_{\nu+} c_{\nu+} + c^{\dagger}_{\nu-} c_{\nu-}). \tag{11}$$

However, we can at least ensure that the average particle number in the ground state has a prescribed value n, by using a Lagrange multiplier. That is, we replace H by $H - \lambda n_{\mathrm{op}}$. The Lagrange multiplier λ is to be determined by the condition

$$n = \langle 0 \mid n_{\mathrm{op}} \mid 0 \rangle = 2 \sum_{\nu} V_{\nu}^2. \tag{12}$$

Since the formal effect of the subtraction of λn_{op} is the replacement of the ε_{ν} by $\varepsilon_{\nu} - \lambda$, we see that λ can be interpreted as an effective Fermi energy.

b) *The deformation-dependent terms of the Hamiltonian.*

We have been using a representation in which the single-particle Hamiltonian is diagonal. However, the spherical part of this Hamiltonian is not necessarily diagonal. We denote its matrix elements by $\varepsilon^0_{\nu\omega}$. The non-spherical part, associated with ellipsoidal deformations, is represented by the scalar product of the single-particle and the total nuclear quadrupole moments. This lifts the degeneracies characteristic of the central field, and has been successfully used[14][15] in the explanation of many properties of deformed nuclei. Thus, the total single-particle matrix element is

$$\varepsilon_{\nu\omega} = \varepsilon^0_{\nu\omega} - \varkappa \sum_{\mu} Q^T_{\mu} (q_{\mu})_{\nu\omega} = \varepsilon_{\nu} \delta_{\nu\omega}, \tag{13}$$

where $\varkappa$ is a coupling constant ultimately determined by the quadrupole force, and

$$(q_{\mu})_{\nu\omega} = 4 \sqrt{\frac{\pi}{5}} \langle \nu \mid r^2 Y_{2\mu}(\theta) \mid \omega \rangle \tag{14a}$$

$$Q^T_{\mu} = 4 \sqrt{\frac{\pi}{5}} \sum_{k} \int \Psi^* \Psi r_k^2 Y^*_{2\mu}(\theta_k)\, dr_1 \ldots dr_k \ldots dr_n, \tag{14b}$$

Ψ being the total nuclear wave function. Therefore, (13) and (14b) imply the self-consistency condition that the quadrupole tensor of the field is the sum of the quadrupole tensors of the orbits determined by that field and the pairing force.

If we were to take $\langle \Psi \mid H \mid \Psi \rangle$ as the total energy, the contribution of

the quadrupole force would be $-\varkappa\sum_{\mu}(Q_{\mu}^{T})^{2}$. This is a factor of 2 too large, since it effectively counts each particle pair twice. Thus, the expectation value $\langle\Psi\left|\left(H+\frac{\varkappa}{2}\sum_{\mu}(Q_{\mu}^{T})^{2}\right)\right|\Psi\rangle$, regarded as a function of the Q_{μ}^{T}, gives the potential energy surface for quadrupole deformations. This picture is reasonable, if the frequencies associated with changes in the Q_{μ}^{T} are small compared to the single-particle frequencies (adiabatic hypothesis).

Since we prefer to treat only the degrees of freedom associated with particles outside closed shells, we should like to replace (13) by an expression involving the quadrupole tensor, Q_{μ}, of these particles alone. The ratio Q_{μ}^{T}/Q_{μ} has been studied[9][17][18] in several single-particle models for the equilibrium values of Q_{μ}^{T}. We make the additional assumption that this ratio is independent of Q_{μ}. Consequently, Q_{μ}^{T} in (13) can be replaced by Q_{μ}, and $\varkappa$ renormalized.

It is useful to perform a principal axis transformation so that the five degrees of freedom (14b) are replaced by three Eulerian angles specifying the orientation of an intrinsic system of axes, and two parameters describing the shape of the ellipsoid. In this intrinsic system $Q_{1}=Q_{-1}=0$ and $Q_{2}=Q_{-2}$. Following[16], we use the shape parameters β and γ defined by*

$$\left.\begin{aligned} Q_{0} &\equiv Q\equiv\beta\cos\gamma \\ Q_{2}=Q_{-2} &\equiv S/\sqrt{2}\equiv\beta\sin\gamma/\sqrt{2}. \end{aligned}\right\} \tag{15}$$

This definition of β differs from that given in[16] by a factor of dimension (length)^{2}. Consequently, the single-particle matrix elements can be written in the form

$$\varepsilon_{\nu\omega}=\varepsilon_{\nu\omega}^{0}-\varkappa\beta\left[\cos\gamma\,(q_{0})_{\nu\omega}+\sin\gamma\,s_{\nu\omega}\right]=\varepsilon_{\nu}\delta_{\nu\omega}, \tag{16}$$

where

$$s_{\nu\omega}=1/\sqrt{2}\left[(q_{2})_{\nu\omega}+(q_{-2})_{\nu\omega}\right]. \tag{17}$$

The self-consistency conditions can now be written

$$\left.\begin{aligned} Q &=\sum_{\nu}(q_{0})_{\nu\nu}\,2V_{\nu}^{2} \\ S &=\sum_{\nu}s_{\nu\nu}\,2V_{\nu}^{2}. \end{aligned}\right\} \tag{18}$$

These can be taken into account by means of two additional Lagrange multipliers $\hat{\mu}$ and $\hat{\sigma}$. One must thus replace the ε_{ν} in (1) by

* In the following, all the directed quantities refer to the intrinsic axes, unless otherwise specified.

$$\varepsilon_{\nu\omega} = \varepsilon^0_{\nu\omega} - \lambda\delta_{\nu\omega} - \mu\,(q_0)_{\nu\omega} - \sigma s_{\nu\omega} = \varepsilon_\nu \delta_{\nu\omega}, \tag{13^1}$$

where

$$\mu = \varkappa Q + \hat{\mu} \qquad \sigma = \varkappa S + \hat{\sigma}. \tag{19}$$

In order to clarify the role of these Lagrange multipliers it is convenient to return to the original description of the quasi-particle approximation in terms of a variational procedure. The expectation value of $\langle 0 \mid H \mid 0 \rangle$ has to be minimized with respect to the V_ν. The result is a set of V_ν which depend on λ, μ, and σ, which are in turn determined from (12) and (18). Then, if we allow small variations from this set of V_ν, but keep λ, μ, and σ fixed, we have

$$0 = \delta\langle 0 \mid H \mid 0 \rangle = \delta\langle 0 \mid H_{\text{spher}} \mid 0 \rangle - \mu\,\delta Q - \sigma\,\delta S. \tag{20}$$

Thus,

$$\frac{\delta\langle 0 \mid H_{\text{spher}} \mid 0 \rangle}{\delta Q} = \mu \qquad \frac{\delta\langle 0 \mid H_{\text{spher}} \mid 0 \rangle}{\delta S} = \sigma. \tag{21}$$

The energy for the optimum V_ν, which we take to define the potential energy surface, is

$$E = \langle 0 \mid H_{\text{spher}} \mid 0 \rangle + \lambda n - 1/2\,\varkappa Q^2 - 1/2\,\varkappa S^2 \tag{22}$$

(cf. p. 7), and so

$$\frac{\delta E}{\delta Q} = \mu - \varkappa Q = \hat{\mu}; \qquad \frac{\delta E}{\delta S} = \sigma - \varkappa S = \hat{\sigma}. \tag{23}$$

At equilibrium, the Lagrange multipliers $\hat{\mu}$ and $\hat{\sigma}$ therefore vanish, and the Hamiltonian used to generate the wave functions has the same deformation as the one used to calculate the energy.

c) *Calculation of the potential energy surface and mass parameters.*

We seek an expansion of the potential energy as a power series in Q and S. According to (23), it is sufficient to calculate the partial derivatives $\frac{\delta^m \mu}{\delta Q^{m-n}\delta S^n}$ and $\frac{\delta^m \sigma}{\delta Q^{m-n}\delta S^n}$. If we wish to calculate the restoring force we require second derivatives, and in most cases these are conveniently obtained as follows*. For simplicity, we treat only one independent variable which we call R. We write

$$H = H_{\text{spher}} - \varrho R_{\text{op}} \tag{24}$$

$$H_{\text{spher}} = \sum_\nu (\varepsilon^0_\nu - \lambda)\,(c^\dagger_{\nu+} c_{\nu+} + c^\dagger_{\nu-} c_{\nu-}) + H_{\text{pair}}. \tag{25}$$

* This method was suggested by A. Bohr (private communication).

We assume that the problem defined by (25) has been solved within the quasi-particle approximation. Thus we have values of λ, Δ, and the V_ν. R_{op} can then be written

$$\left.\begin{aligned} R_{\mathrm{op}} = & \sum_{\nu} r_{\nu\nu}\, 2V_\nu^2 \\ & + \sum_{\nu\omega} r_{\nu\omega}\,(U_\nu U_\omega - V_\nu V_\omega)\,(\alpha_\nu^\dagger \alpha_\omega + \beta_\nu^\dagger \beta_\omega) \\ & + \sum_{\nu\omega} r_{\nu\omega}\,(U_\nu V_\omega + U_\omega V_\nu)\,(\alpha_\nu^\dagger \beta_\omega^\dagger + \beta_\nu \alpha_\omega)\,. \end{aligned}\right\} \tag{26}$$

Treating $-\varrho R_{\mathrm{op}}$ as a perturbation, the new ground state $|\,0'\rangle$ is given in first-order perturbation theory by

$$|\,0'\rangle = |\,0\rangle + \varrho \sum_{\nu\omega} \frac{r_{\nu\omega}(U_\nu V_\omega + U_\omega V_\nu)}{E_\nu + E_\omega}\,|\,\nu\omega\,\rangle. \tag{27}$$

Here again the value of ϱ is determined from the assigned expectation value of R_{op}, i. e.,

$$R = \langle\, 0' \,|\, R_{\mathrm{op}} \,|\, 0'\,\rangle = 2\,\varrho \sum_{\nu,\,\omega} r_{\nu\omega}^2 \frac{(U_\nu V_\omega + U_\omega V_\nu)^2}{E_\nu + E_\omega}\,. \tag{28}$$

According to (23) and (28), the restoring force $C = \dfrac{\delta^2 E}{\delta R^2}$ is given by

$$C = \frac{\delta \varrho}{\delta R} - \varkappa = \frac{1}{2 \sum\limits_{\nu,\,\omega} r_{\nu\omega}^2 \dfrac{(U_\nu V_\omega + U_\omega V_\nu)^2}{E_\nu + E_\omega}} - \varkappa. \tag{29}$$

A simple physical interpretation can be given for the terms in (29). The first one, which tends to preserve the spherical shape, equals the increase in the expectation value of H_{spher} due to the deformation. To second order in R,

$$\langle\, 0' \,|\, H_{\mathrm{spher}} \,|\, 0'\,\rangle - \langle\, 0 \,|\, H_{\mathrm{spher}} \,|\, 0\,\rangle = \frac{R^2}{4 \sum\limits_{\nu,\,\omega} r_{\nu\omega}^2 \dfrac{(U_\nu V_\omega + U_\omega V_\nu)^2}{E_\nu + E_\omega}}. \tag{30}$$

The second term corresponds to the expectation value of the interaction which produces the deformation.

However, some precautions must be taken when using $|\,0'\rangle$ given by (27), since its average particle number differs from that of $|\,0\rangle$. In fact, to first order,

$$\langle 0' \mid n_{\mathrm{op}} \mid 0' \rangle - \langle 0 \mid n_{\mathrm{op}} \mid 0 \rangle = \varrho \Delta^2 \sum_{\nu} \frac{r_{\nu\nu}}{E_\nu^3}. \tag{31}$$

The linear term in (31) can produce a spurious contribution in (30). In order to eliminate this linear term and thus this spurious effect, we must require $\sum_{\nu} \frac{r_{\nu\nu}}{E_\nu^3} = 0$. (A second order variation in the number of particles does not affect (30), since the expectation value of H_{spher} is stationary in the number of particles).

In the Appendix, it is shown that the additional condition $\sum \frac{r_{\nu\nu}\varepsilon_\nu}{E_\nu^3} = 0$ must also be satisfied, due to the requirement that matrix elements of the form $\langle 0' \mid H_{\mathrm{int}} \mid (\nu\omega)' \rangle$ should contain no terms linear in ϱ.

The conditions $\sum_{\nu} \frac{r_{\nu\nu}}{E_\nu^3} = 0$ and $\sum_{\nu} \frac{r_{\nu\nu}\varepsilon_\nu}{E_\nu^3} = 0$ are satisfied for quadrupole vibrations about spherical equilibrium shape. The methods of this section can also be used to study the vibrations about non-zero equilibrium deformations. In this case, the conditions $\sum_{\nu} \frac{s_{\nu\nu}}{E_\nu^3} = 0$ and $\sum_{\nu} \frac{r_{\nu\nu}\varepsilon_\nu}{E_\nu^3} = 0$ are satisfied for γ-vibrations about $\gamma = 0$ or $\gamma = \pi$. However, for β-vibrations they are not satisfied, because the quadrupole operator connects the ground state to the spurious 2-quasi-particle state. Hence, one may not fix λ and Δ and then do the perturbation calculation; one must rather determine first the effect of the perturbation on the single-particle energies and wave functions, and then solve (4) and (12) for λ and Δ. Although we will not need the general expressions so obtained, we give them in the Appendix for completeness.

We calculate the mass parameter, using time-dependent adiabatic pertubation theory (the "cranking" model). (See also eq. (15) of ref.[7]).

$$\left.\begin{aligned} B &= 2\hbar^2 \sum_{\nu\omega} \frac{|\langle \nu\omega \mid \frac{\delta}{\delta R} \mid 0 \rangle|^2}{E_\nu + E_\omega} \\ &= 2\hbar^2 \left(\frac{d\varrho}{dR}\right)^2 \sum_{\nu\omega} \frac{|\langle \nu\omega \mid \left(\frac{\delta}{\delta\varrho}\right) \mid 0 \rangle|^2}{E_\nu + E_\omega}. \end{aligned}\right| \tag{32}$$

Using the relation $\frac{\delta H}{\delta \varrho} = \left[\frac{\delta}{\delta\varrho}, H\right]$, (32) can be shown to be equivalent to[19]

$$\left.\begin{aligned} B &= 2\hbar^2\left(\frac{d\varrho}{dR}\right)^2 \sum_{\nu,\omega} \frac{|\langle \nu, \omega \mid \dfrac{dH}{d\varrho} \mid 0\rangle|^2}{(E_\nu + E_\omega)^3} \\ &= 2\hbar^2\left(\frac{d\varrho}{dR}\right)^2 \sum_{\nu,\omega} \frac{|\langle \nu, \omega \mid R_{\mathrm{op}} \mid 0\rangle|^2}{(E\nu + E\omega)^3} \\ &= 2\hbar^2\left(\frac{d\varrho}{dR}\right)^2 \sum_{\nu,\omega} \frac{r_{\nu\omega}^2 (U_\nu V_\omega + U_\omega V_\nu)^2}{(E_\nu + E_\omega)^3}. \end{aligned}\right\} \quad (33)$$

The replacement of $\dfrac{\delta H}{\delta \varrho}$ by $-R_{\mathrm{op}}$ is valid only if λ and Δ have no first order terms in ϱ (cf. eq. (10 A)). The formulae for the restoring force and mass parameters of the vibrations simplify if the single-particle quadrupole moment has only diagonal elements (e. g., the harmonic oscillator)

$$C = \frac{1}{\Delta^2 \sum_\nu \dfrac{r_{\nu\nu}^2}{E_\nu^3}} - \varkappa \quad (34\,\mathrm{a})$$

$$B = \hbar^2 \frac{\sum_\nu \dfrac{r_{\nu\nu}^2}{E_\nu^3}}{\left(2\Delta \sum_\nu \dfrac{r_{\nu\nu}^2}{E_\nu^3}\right)^2}. \quad (34\,\mathrm{b})$$

Another simple case arises when the single-particle states are degenerate

$$C = \frac{G\Omega}{2\theta_n \sum_{\nu,\omega} r_{\nu\omega}^2} - \varkappa \quad (35\,\mathrm{a})$$

$$B = \frac{\hbar^2}{2G\Omega\theta_n \sum_{\nu,\omega} r_{\nu\omega}^2} \quad (35\,\mathrm{b})$$

$$\hbar\omega = G\Omega\left(1 - \frac{2\varkappa\theta_n}{G\Omega} \sum_{\nu,\omega} r_{\nu,\omega}^2\right)^{1/2}. \quad (35\,\mathrm{c})$$

Here, $\theta_n = 1 - x_n^2$ and $x_n = \dfrac{n}{\Omega}$, while Ω is the total number of pairs of states available.

The above adiabatic treatment of the quadrupole vibrations requires the energy of the first vibrational excitation $\hbar\omega$ to be small compared to twice the quasi-particle energy. A different approach to this problem has been given by B. MOTTELSON[10]. He considers particles moving in degenerate states, and coupled by pairing and quadrupole forces. The quadrupole force affects only one of the $I = 2$ two quasi-particle states, whose energy is given

by this model to be exactly the same as (35 c). Here the conditions of validity are complementary to ours, since they effectively imply* a small depression of the vibrational state compared to the 2-quasi-particle energy. Since (35 c) holds at both limits, we may expect it to be a reasonable approximation in between.

III. Expansion of the energy about the spherical equilibrium shape

BELYAEV[7] has already studied the dependence of the nuclear surface energy on an axially symmetric deformation, using a single-particle Hamiltonian with diagonal intrinsic quadrupole moments $(q_0)_{\nu\omega} = (q_0)_{\nu\nu}\,\delta_{\nu,\,\omega}$, and an assumed density of states. In the following, the simplified case of nucleons moving in a harmonic oscillator shell will be treated**, but the restriction to axial symmetry will be omitted.

In order to calculate the partial derivatives $\dfrac{\delta^n \mu}{\delta Q^{n-m} \delta S^m}$ and $\dfrac{\delta^n \sigma}{\delta Q^{n-m} \delta S^m}$, we can proceed as follows: λ and Δ are expanded as power series in the variables μ and σ. It is then possible to construct the power series for E_ν and V_ν^2 and, therefore, the right-hand side of the basic equations (4) and (12). We must put equal to zero the coefficients of the successive powers of μ and σ in the expressions for G and n, since these quantities are independent of the deformation. This provides us with a set of equations from which the coefficients in the expansions for λ and Δ can be derived. These coefficients are inserted in the power series for V_ν^2. The power series for Q and S can then be immediately obtained by using equation (18). After reversing these last two series and performing the necessary differentations, we obtain the following expression for the energy:

$$
\begin{aligned}
E = E_{(\beta=0)} &+ \frac{C}{2}\beta^2 \\
&+ \left(\frac{MW}{\hbar}\right)^2 \frac{G\beta^2}{2\,\theta_n N^2} \Bigg\{ - \frac{4\beta}{5\,(1+\alpha)\,Q_{\max}}\, x_n \cos 3\gamma + \left(\frac{4\beta}{5\,(1+\alpha)\,Q_{\max}}\right)^2 \left(\frac{1}{2} + \frac{3\,\theta_n}{64}\right) \\
&\quad - \left(\frac{4\beta}{5\,(1+\alpha)\,Q_{\max}}\right)^3 x_n \left(\frac{83}{56} - \frac{\theta_n}{2}\right) \cos 3\gamma \\
&\quad + \left(\frac{4\beta}{5\,(1+\alpha)\,Q_{\max}}\right)^4 \left[\frac{3635}{448} - \theta_n \frac{77955}{1792} + \theta_n^2 \frac{5535}{128}\right. \\
&\quad \left. + \left(\frac{223}{56} - \theta_n \frac{4979}{224} + \theta_n^2 \frac{1207}{64}\right) \cos^2 3\gamma \right] + \ldots \Bigg\}
\end{aligned}
\tag{36}
$$

* More precisely, the seniorities of the states mixed by the quadrupole force into the ground state should be small compared to Ω.

** A more general type of shell has been considered by J. M. ARAÚJO (private communication).

$$C = \left(\frac{MW}{\hbar}\right)^2 \frac{G}{\theta_n N^2} - \varkappa. \tag{36 a}$$

β and γ are defined in (15), M is the mass of the nucleon, and W the frequency of the oscillator field. N, which is assumed to be large compared to unity, is the principal quantum number of the oscillator shell. We also define n_z and $n_\perp$ to be the numbers of oscillator quanta along and perpendicular to the z-axis, respectively.

$Q_{\max}$ is the maximum value of Q which can be obtained with a given number of particles in the shell. One gets this value in the "aligned coupling scheme"[2].

α is defined so that αN is the maximum occupied value of $n_\perp$, for prolate deformation, with a given number of particles $n_\perp$ and no pairing force. In consequence, $0 \leq \alpha \leq 1/\sqrt{2}$. For values of $n > \Omega$ ($\alpha > 1/\sqrt{2}$), $Q_{\max}$ occurs for oblate deformation. In this case the previous expression also holds, holes playing the role of our previous particles.

A few comments can be made on equation (36).

1) The γ-dependence of the terms of a given order in β can be understood on the basis of general invariance arguments. The energy of the system must be invariant with respect to rotations. Therefore, it can be expressed as a linear superposition of the solutions of the five-dimensional quadrupole oscillator corresponding to zero total angular momentum. The γ-dependent part of these solutions can be expressed in terms of Legendre polynomials in the variable $\cos 3\gamma$[16]. The solutions for $I = 0$ can be characterized[21] by the quantum numbers (n_β, l), where n_β is the number of quanta for the β-motion and l is an integer that RAKAVY[22] has called the "seniority". It is related to $\mathfrak{N}$, the total number of phonons, by the equation $\mathfrak{N} = 2n_\beta + 3l$.

The γ-independence of the term proportional to β^2 simply reflects the fact that no function of $\cos 3\gamma$ can be formed from linear combinations of quadratic expressions in $\cos\gamma$ and $\sin\gamma$. The only invariant expression that can be made proportional to β^2 is the β-excitation built on the ground state. This wave function is characterized by the quantum numbers (1,0). For 3 phonons only one solution is possible, and is proportional to $\beta^3 \cos 3\gamma$ (0,1). The only allowed $I = 0$ state with 4 phonons is the second β-excitation of the ground state (2,0), which does not depend on γ. Also in the case of 5 phonons only the solution (1,1) appears. It corresponds to the β-excitation of the (0,1) state and, therefore, has the same γ-dependence, namely $\cos 3\gamma$. Two $\mathfrak{N} = 6$ states appear for $I = 0$. The triple β-excitation of the ground state with no γ-dependence (3,0) and the (0,2) state which is proportional to $\beta^6 \cos^2 3\gamma$.

In general, one can predict that terms which contain odd powers of $\cos 3\gamma$ are multiplied by odd powers of β; even powers of $\cos 3\gamma$ are multiplied by even powers of β.

2) The existence of a negative β^3 term* ensures that, for sufficiently small positive values of C, there is a maximum in the expression for the energy as a function of the axially symmetric deformation. It is situated at

$$\beta_{\max} = \frac{5\,(1+\alpha)}{3 \times \varkappa_n} Q_{\max} C . \tag{37}$$

The smallness of C (and therefore of $\beta_{\max}$) allows us to consider only the β^2 and β^3 terms in (36). One can then easily derive (37).

The existence of a maximum ensures the existence of a second minimum, provided the system does not collapse. Thus, the system has started to deform even before reaching the transition point $C = 0$.

It is interesting to note why there do not occur two minima in the curve which BELYAEV used to illustrate the energy of the system as a function of the axially-symmetric deformation. Let us consider a degenerate shell whose levels are split by a deformation in such a way that the final single-particle spectrum is symmetric with respect to the original energy. This system will have no preference for prolate rather than oblate deformations, or vice versa. Thus, no odd powers of β will appear in an expansion of the energy such as (36), because these terms are associated with odd powers of $\cos 3\gamma$, which can distinguish between $\gamma = 0$ and $\gamma = \pi$. In particular, no β^3 term can occur and therefore the sufficient condition for the existence of two minima no longer holds. BELYAEV has found the ground-state equilibrium deformation for a system of this kind (constant density of levels). One should remember, however, that this system has some kind of γ-unstability, because prolate and oblate deformations are equally favoured. Neither does the energy surface for the γ-deformation of an $n_\perp$-subshell in an axially symmetric harmonic oscillator field present two minima.

The density of states of an axially symmetric harmonic oscillator is proportional to the energy; the density in a deformed j-shell is inversely proportional to the magnitude of the magnetic quantum number. In both cases, the equilibrium deformation is such that the density increases with energy. If the shell is less than half filled, this favours prolate deformation

* A discussion of the β^3 terms, including their effect in the kinetic energy, has been made independently by A. KERMAN (to be published).

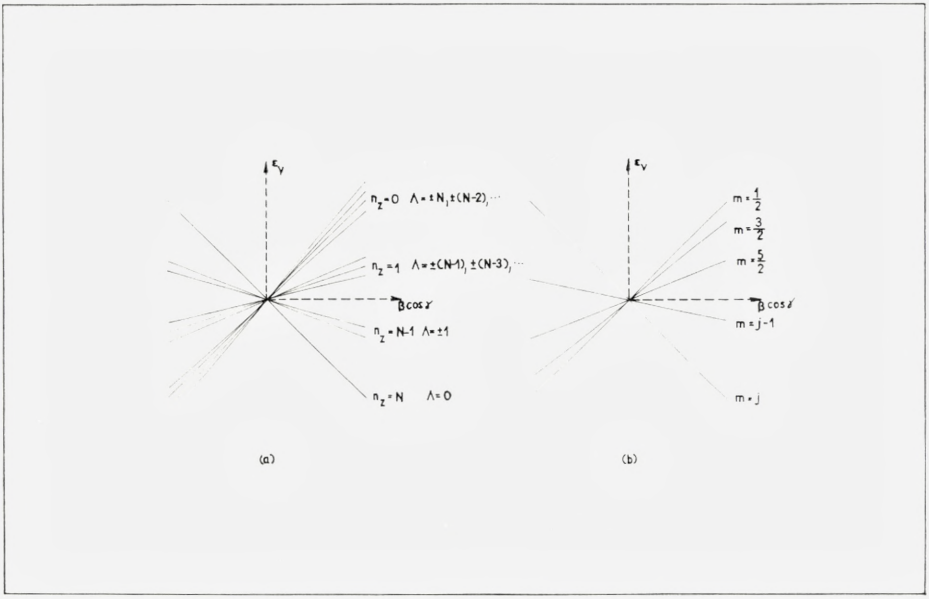

Fig. 1. Level spectra for (a) a harmonic oscillator field of cylindrical symmetry, (b) an axially symmetric field superposed on the spherical field giving rise to a single j-shell.

for the harmonic oscillator and oblate deformation for the j-shell (see Fig. 1). Past the middle of the shell the above arguments apply to the hole states, and thus the roles of prolate and oblate deformations are interchanged.

The most direct consequence of the existence of two minima would be the appearence of a sudden change in the deformation when the second minimum falls below the first. We have seen that the existence of two minima requires a β^3 term, which in turn implies γ-stability. This is consistent with the empirical fact that the transition to deformed nuclei is more abrupt at the beginning of the rare-earth region where the nuclei are γ-stable, than at the end where they approach γ-unstability. More accurate predictions cannot be given at present, because neither the harmonic oscillator nor the j-shell provides a realistic description of the actual single-particle spectra.

3) In the spherically symmetric harmonic oscillator, the consequences of the terms proportional to $\cos 3\gamma$ and $\cos^2 3\gamma$ have been studied by constructing their matrices and diagonalizing them in perturbation theory. The necessary γ-dependent part of the wave functions is given in reference[20]. The term $\cos 3\gamma$ shifts the first $2+$ and $4+$ states towards the positions that they would occupy in a rotational band. The second $2+$ state is pushed

rather high. On the contrary, the $\cos^2 3\gamma$ term tends to bring the second 2+ state below the first 4+ state.

The discussion of the influence of these terms on the transition rates is simplified by the existence of a "γ-parity"[20] which is equal to the parity of l. Any interaction which can be expanded in even powers of $\cos 3\gamma$ preserves a selection rule which forbids the transition from the second 2+ state to the ground state; the odd powers of $\cos 3\gamma$ violate this selection rule.

Most non-deformed, even mass nuclei[23] have their second 2+ level below their first 4+ level; in addition, the transition from the second 2+ level to the ground state is strongly retarded. The previous arguments suggest that both these features can be attributed to the effect of a term proportional to $\beta^6 \cos^2 3\gamma$. One can imagine situations in which the coefficient of the β^3 term would be reduced, for example if the single-particle spectrum is intermediate between those of the harmonic oscillator and the j-shell, or if protons and neutrons are filling opposite ends of similar shells (see 2). The main effect of a β^4 term would be on the position of the second 0+ state, about which very little is known experimentally.

We have considered only the β- and γ-dependence of the nuclear surface energy. Similar terms in the mass parameter should also be taken into account in a more detailed study of nuclear vibrations.

IV. Gamma vibrations in a deformed harmonic oscillator field

We assume that the system has a prolate axially symmetric equilibrium deformation ($\gamma = 0$), and we study the change in the potential energy for small changes in γ. In this chapter we consider the case of a harmonic oscillator field. Because of the very particular degeneracies associated with this field, we do not expect quantitative agreement with actual nuclei. However, the oscillator gives a first qualitative picture of a realistic nuclear shell, and has the advantage that closed expressions for the vibrational parameters can be obtained. In addition, we assume that $N(= n_z + n_\perp)$ is much greater than unity.

The operator corresponding to the γ-deformation has only diagonal matrix elements in a single-particle representation characterized by the quantum numbers N, $n_\perp$ and n_y.

$$s_{\gamma\gamma} = \frac{\hbar}{MW_\perp} \sqrt{3}\,(n_\perp - 2\,n_y). \tag{38}$$

Here, $W_\perp$ is the characteristic frequency for oscillations perpendicular to the z-axis.

We can therefore apply equations (34). One can easily evaluate the vibrational coefficients in two simple cases.

a) The deformation is so great compared to the pairing force that the problem reduces to coupled particles in the $n_\perp$-subshells. The necessary condition for the validity of this approximation is that the two quasi-particle energies are small compared to the distance between 2 subshells, i. e.,

$$G_\perp n_\perp \ll 3 \varkappa Q_{\text{eq.}} \frac{\hbar}{MW}, \tag{39}$$

where $G_\perp$ is the effective strength of the pairing force which acts between particles belonging to the $n_\perp$-subshell.

Due to renormalization effects of the other $n_\perp$-subshells, $G_\perp$ is greater than the G to be used if the entire N-shell is treated. We can calculate the renormalization by means of a procedure similar to those employed in [7] and [10] to account for the influence on a particular unfilled shell of the presence of other shells. Let us call $G_{\nu\omega}$ $(= G)$ the pairing force matrix element corresponding to a scattering of a pair of particles from the states $(\nu+,\ \nu-)$ to the states $(\omega+,\ \omega-)$. According to [10],

$$\left.\begin{aligned}(G_\perp)_{\nu\nu'} &= G_{\nu\nu'} + \sum_{\omega=n_\perp+1}^{\omega=N} \frac{G_{\nu\omega} G_{\omega\nu'}}{2(\varepsilon_\omega - \varepsilon_\nu)} + \sum_{\omega=n_\perp-1}^{\omega=0} \frac{G_{\omega\nu'} G_{\nu\omega}}{2(\varepsilon_\nu - \varepsilon_\omega)} + \cdots \\ &= G_{\nu\nu'} \left\{ \frac{1}{1 - \dfrac{G}{6\varkappa Q_{\text{eq}}} \dfrac{MW}{\hbar} \left[\sum_{\omega=1}^{\omega=N-n_\perp} \dfrac{\omega + n_\perp}{\omega} + \sum_{\omega=N_\perp-1}^{\omega=n_\perp} \dfrac{n_\perp - \omega}{\omega} \right]} \right\}.\end{aligned}\right\} \tag{40}$$

By performing the above summations, and using condition (39), we get

$$G_\perp = G \left\{ \frac{1}{1 - \dfrac{G n_\perp}{6\varkappa Q_{\text{eq}}} \dfrac{MW}{\hbar} ln\,[n_\perp (N - n_\perp)]} \right\}. \tag{41}$$

In this case, a) the simple expressions (35), corresponding to the "degenerate model", can be used for the vibrational parameters

$$\left.\begin{aligned} C &= \frac{G_\perp}{2\theta'_n n_\perp^2} \left(\frac{MW_\perp}{\hbar} \right)^2 - \varkappa \\ B &= \frac{\hbar^2}{2} \frac{1}{G_\perp n_\perp^4 \theta'_n} \left(\frac{MW_\perp}{\hbar} \right)^2, \end{aligned}\right\} \tag{42}$$

where $\theta'_n = \frac{n'}{n_\perp}\left(2 - \frac{n'}{n_\perp}\right)$, and n' is the number of particles in the $n_\perp$ subshell. If $\theta^{0'}_n$ is the value of θ'_n for which the axial shape is no longer stable, (42) and (36) imply that

$$\theta^0_{n'} = \frac{G_\perp}{2\varkappa n_\perp^2}\left(\frac{MW_\perp}{\hbar}\right)^2 = \frac{\theta^0_n}{2}\frac{G_\perp}{G}\frac{N^2}{n_\perp^2}\left(\frac{W_\perp}{W}\right)^2. \tag{43}$$

Neglecting the renormalization effect expressed by (41), we see that $\theta^0_{n'}$ and θ^0_n are roughly of the same order of magnitude. This implies that the fraction of nuclei with axially symmetric equilibrium deformation is of the same order of magnitude as the fraction of nuclei which are spherical.

Nuclei with $0 < \theta'_n < \theta^0_{n'}$ have an axially symmetric stable deformation. The ratio between the frequency of the γ-vibrations and the gap is

$$\frac{\hbar\omega_\gamma}{G_\perp n_\perp} = \left[1 - 2\varkappa\left(\frac{\hbar}{MW_\perp}\right)^2 n_\perp^2 \frac{\theta'_n}{G_\perp}\right]^{1/2}. \tag{44}$$

Here the adiabatic condition implies that $\frac{\hbar\omega_\gamma}{G_\perp n_\perp} \ll 1$.

For nuclei in the region of transition between axially symmetric and γ-deformed nuclei, the potential energy surface does not exhibit two minima (cf. p. 14).

b) We can also easily treat the deformed harmonic oscillator field if we replace summations over the variable $n_\perp$ by integrations, using a level density proportional to the single-particle energies (see p. 15). This is a particular case of the level density used by BELYAEV in his investigation of axial deformations. Equations (46)-(54) are a transcription of some of his results into our notation.

The single-particle energies ε_ν can be labelled by $n_\perp$. With a convenient choice of the zero-point energy, they are given by

$$\varepsilon_\nu = \varepsilon_{n_\perp} = 3\mu\frac{\hbar}{MW}n_\perp - \lambda. \tag{45}$$

W is the frequency of the harmonic oscillator. In neglecting the difference between $W_\perp$ and W_z we make an error of the order of the deformation, i. e., of order $A^{-1/3}$ or N^{-1} for the equilibrium deformation. This can be neglected in our limit $N \gg 1$.

A new parameter η characterizing the deformation is introduced:

$$\eta = \frac{3\,\hbar\mu}{MWG\bar{\varrho}}, \tag{46}$$

where $\bar{\varrho}$ is defined by the condition

$$\int_0^N \frac{\varrho\,(n_\perp)\,dn_\perp}{E_{n_\perp}} = \bar{\varrho}\int_0^N \frac{dn_\perp}{E_{n_\perp}}. \tag{47}$$

In the axially symmetric harmonic oscillator, the level density, $\varrho(n_\perp)$, equals $n_\perp$. Therefore, the parameters used by BELYAEV in order to characterize the level density are here

$$\varrho_0 = \frac{N}{2}; \qquad \xi = 1. \tag{48}$$

The parameters λ and Δ are always determined from (4) and (12):

$$\lambda = \frac{3\,\hbar\mu N}{2\,MW}(1 - x_n \coth\eta), \tag{49}$$

$$\Delta^2 = \left(\frac{3\,\hbar\mu N}{2\,MW}\right)^2 \frac{(1-x_n^2)}{\sin\hbar^2\eta}, \tag{50}$$

where x_n measures the number of particles in the shell and is also a slow-varying function of η

$$x_n = 2\left\{\frac{1 - \dfrac{n}{\Omega} - \dfrac{\gamma}{2}}{1 + \left[1 - 2\,\gamma\left(1 - \dfrac{n}{\Omega}\right) + \gamma^2\right]^{1/2}}\right\} \tag{51}$$

with

$$\gamma \equiv \coth\eta\left(1 - \frac{2\,\eta}{\sinh 2\,\eta}\right). \tag{52}$$

One can eliminate $\bar{\varrho}$ in (46) by its expression as a function of ϱ_0, ξ, x_n and η

$$\frac{\mu\hbar}{MW} = \frac{GN}{6}\left[\eta - x_n\,(\eta\coth\eta - 1)\right]. \tag{53}$$

The quadrupole moment Q is given by

$$\left.\begin{aligned} Q &= \frac{\hbar}{MW}\sum_{n_\perp}(2N - 3\,n_\perp)\,n_\perp\,2V_{n_\perp}^2 \\ &= \frac{\hbar}{MW}\sum_{n_\perp}(-2N + 3\,n_\perp)\,n_\perp\frac{\varepsilon_{n_\perp}}{E_{n_\perp}} \\ &\simeq \frac{\hbar}{MW}\frac{N^3}{4}(1-x_n^2)\left[\coth\eta - x_n\left(1 + \frac{3}{\sinh^2\eta}\right) + \frac{\eta}{\sinh^2\eta}(3\,x_n\coth\eta - 1)\right]. \end{aligned}\right\} \tag{54}$$

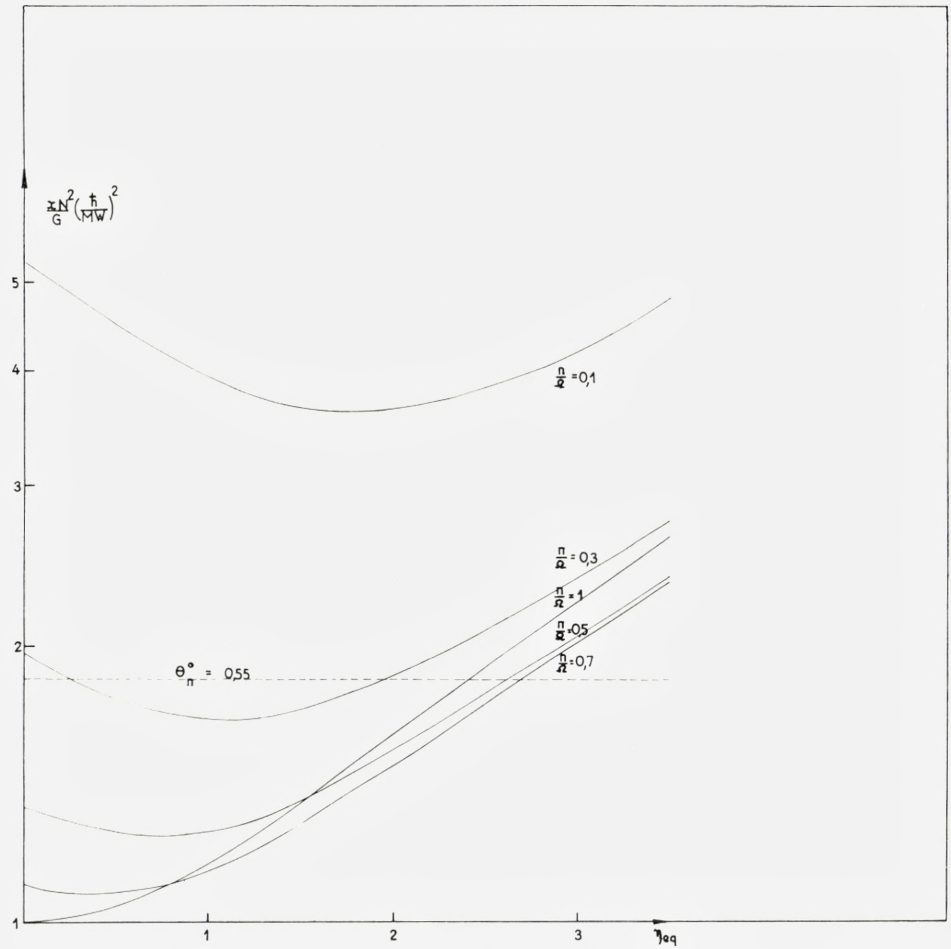

Fig. 2. The right-hand side of equation (55) represented as a function of $\eta = \eta_{\mathrm{eq}}$, for several values of $\frac{n}{\Omega}$. The intersection of these curves with the horizontal dashed line yields the values of η for which the energy is a minimum if the $\frac{\varkappa N^2}{G}\left(\frac{\hbar}{MW}\right)^2$ is such that $\theta_n^0 = 0.55$.

Equations (53) and (54), plus the condition of the vanishing of the Lagrange multiplier at equilibrium ($\mu = \varkappa Q$), define an implicit equation for the equilibrium value of η.

$$\frac{\varkappa N^2}{G}\left(\frac{\hbar}{MW}\right)^2 = \frac{4\sinh^2\eta\,[\eta - x_n(\eta\coth\eta - 1)]}{3\,(1-x_n^2)\,[\sinh 2\eta - 2x_n(3+\sinh^2\eta) + 2\eta\,(3x_n\coth\eta - 1)]}. \quad (55)$$

At $\eta = 0$ the right-hand side becomes indeterminate. However, (36) shows that in this case there is always an extremum in the potential energy surface. The right-hand side of (55) is plotted as a function of η in Fig. 2 for $\frac{n}{\Omega} =$ 0.1, 0.3, 0.5, 0.7, and 1. It is seen that, for a particular $\frac{n}{\Omega}$ and sufficiently small $\frac{\varkappa N^2}{G}\left(\frac{\hbar}{MW}\right)^2$, there are no solutions to (55) and thus the only extremum is the minimum at $\eta = 0$. For larger $\frac{\varkappa N^2}{G}\left(\frac{\hbar^2}{MW}\right)$, there are two values of η satisfying (55), the lower corresponding to a maximum and the upper to a second minimum (cf. discussions on p. 14). For still larger $\frac{\varkappa N^2}{G}\left(\frac{\hbar^2}{MW}\right)^2$, there is only one solution to (55), and thus only one minimum (the extremum at $\eta = 0$ is now a maximum). One can also see that for each $\frac{n}{\Omega} \neq 1$ there is a minimum value for stable deformation.

The expressions $\sum_\nu \frac{s_{\nu\nu}^2}{E_\nu^3}$ and $\sum_\nu \frac{s_{\nu\nu}^2}{E_\nu^5}$, needed in (34) for the evaluation of the restoring force and mass parameter, are to be calculated for the value of η which corresponds to the equilibrium situation. We find

$$\left.\begin{aligned}\sum_\nu \frac{s_{\nu\nu}^2}{E_\nu^a} &= \left(\frac{\hbar}{MW}\right)^2 \sum_{n_\perp=0}^{n_\perp=N} \sum_{n_y=0}^{n_y=n_\perp} \frac{(n_\perp - 2\,n_y)^2}{E_{n_\perp}^a} \\ &\simeq \left(\frac{\hbar}{MW}\right)^2 \sum_{n_\perp=0}^{n_\perp=N} \frac{n_\perp^3}{E_{n_\perp}^a} \\ &\simeq \left(\frac{\hbar}{MW}\right)^2 \int_0^N \frac{n_\perp^3\, dn_\perp}{E_{n_\perp}^a}\end{aligned}\right\} \quad (56)$$

$$\left.\begin{aligned}\sum_\nu \frac{s_{\nu\nu}^2}{E_\nu^3} = \left(\frac{\hbar}{MW}\right)^2 &\frac{8}{N^2 G^3 [\eta - x_n(\eta\coth\eta - 1)]^3 [1 + (1 - x_n^2)\sinh^2\eta]} \\ &\{ x_n [3 + (4 - 3x_n^2)\sinh^2\eta - (3 + x_n^2)\sinh^4\eta] \\ &+ \sinh\eta\cosh\eta\,(-3 + (1 + 3x_n^2)\sinh^2\eta] \\ &+ 3\eta(1 - x_n\coth\eta)[1 + (1 - x_n^2)\sinh^2\eta)]\}\end{aligned}\right\} \quad (57)$$

$$\left.\begin{aligned}\sum_\nu \frac{s_{\nu\nu}^2}{E_\nu^5} = \left(\frac{\hbar}{MW}\right)^2 &\frac{256\sinh^5\eta}{3N^6 G^5 [\eta - x_n(\eta\coth\eta - 1)]^5 [1 + (1 - x_n^2)\sinh^2\eta]^3} \\ &\{\cosh\eta [3 + (4 + 8x_n^2)\sinh^2\eta + (1 + 2x_n^2 - 3x_n^4)\sinh^4\eta] \\ &- x_n \sinh\eta [9 + 12\sinh^2\eta + (3 - 2x_n^2 - x_n^4)\sinh^4\eta]\}.\end{aligned}\right\} \quad (58)$$

In the limit of vanishing G, $\eta \to \infty$ and equations (50), (51), (55), (57), and (58) reduce to

$$\Delta^2 = \frac{N^4 G^2}{16} \frac{\eta^2}{\sinh^2 \eta} (1 - x_n)^3 (1 + x_n). \tag{50^1}$$

$$x_n = \frac{1 - \frac{2n}{\Omega}}{1 + \left[\frac{2n}{\Omega}\right]^{1/2}} \tag{51^1}$$

$$\lim_{\eta \to \infty} G[\eta - x_n(\eta - 1)] = \frac{3 \varkappa N^2}{2} \left(\frac{\hbar}{MW}\right)^2 (1 - x_n)^2 (1 + x_n) \tag{55^1}$$

$$\sum_\nu \frac{s_{\nu\nu}^2}{E_\nu^3} = \frac{8}{N^2 G^3} \left(\frac{\hbar}{MW}\right)^2 \frac{\sinh^2 \eta}{\eta^3 (1 - x_n^2)} \tag{57^1}$$

$$\sum_\nu \frac{s_{\nu\nu}^2}{E_\nu^5} = \frac{256}{3 N^6 G^5} \left(\frac{\hbar}{MW}\right)^2 \frac{\sinh^4 \eta}{\eta^5 (1 - x_n)^2 (1 - x_n^2)^2}. \tag{58^1}$$

According to (34), the restoring force for γ-vibrations is then given by

$$C = 2 \varkappa \frac{(1 + 2 x_n)}{1 - x_n}, \tag{59}$$

which implies that the axial symmetry is preserved until $x_n = -1/2$, which corresponds to $\frac{n}{\Omega} = \frac{9}{8}$. Thus the first half of the shell $\left(\frac{n}{\Omega} < 1\right)$ has stable prolate deformation.

The mass parameter for the γ-vibrations diverges as $\eta \to \infty$. However, the significant quantity is the ratio between the energy of the γ-vibration and 2Δ. This ratio remains finite, and is given by

$$\frac{\hbar w_\gamma}{2\Delta} = \left[\frac{1 + 2 x_n}{1 + x_n}\right]^{1/2}. \tag{60}$$

For non-zero G, we must solve (55) for η and then evaluate $\frac{\hbar w_\gamma}{2\Delta}$, using (50), (51), (57), and (58). The constant $\frac{\varkappa N^2}{G} \left(\frac{\hbar}{MW}\right)^2$ has been chosen so that $\theta_n^0 = 0.55$, corresponding to a situation in which the spherical shape becomes unstable when the $\frac{n}{\Omega} = \frac{1}{3}$. The results are shown in Fig. 3. The corresponding curve for oblate deformation is obtained by reflecting the curve for prolate deformation about the line $\frac{n}{\Omega} = 1$.

It is seen that the adiabatic condition for γ-vibrations is satisfied for $0.5 < \frac{n}{\Omega} < 1.5$. For prolate (oblate) deformation the frequency of the vibration decreases as the number of particles (holes) increases. Fig. 3 also

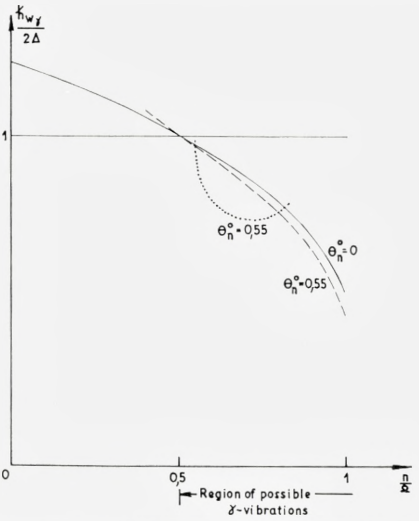

Fig. 3. The ratio between the energy of the γ-vibrations and twice the value of s plotted as a function of $\frac{n}{\Omega}$, for $\theta_n^0 = 0$ ($G = 0$) and $\theta_n^0 = 0.55$. The full and dashed lines represent the two cases in which the calculations were done by replacing the summations by integrations. The dotted line represents the result of the calculations done without this approximation, for $\theta_n^0 = 0.55$ and n such that $n_\perp = 4$ is at the Fermi surface for $G = 0$.

shows that the ratio $\frac{\hbar\omega_\gamma}{2\Delta}$ is not significantly affected by the presence of the pairing force.

If $G \to 0$, so that $\eta \to \infty$, we might expect to approach the situation dealt with in a) above. Nevertheless, the fact that Fig. 3 shows no subshell effects implies that the two methods do not lead to the same result. In fact, for fixed N, the validity of method a) places an upper limit on G (see (39)), whereas the validity of method b) places a lower limit on G. Evidently these regions of validity do not overlap. It is probable that the actual nuclear case is better represented by method b). On the one hand, the reduction in the observed moments of inertia compared to the rigid values implies a mixing by the pairing force of different $n_\perp$-subshells. On the other hand, performing the sums in (4), (12), (54), (56), and (57) exactly for $\theta_n^0 = 0.55$,

$N = 7$, and n such that $n_\perp = 4$ is at the Fermi surface for $G = 0$, leads to the dotted curve in Fig. 3. Although subshell effects do appear, the average ratio agrees well with the result given by method b).

V. Gamma vibrations in a realistic shell model

In the previous section we found that the occurrence of low-energy γ-vibrations requires states with high values of $n_\perp$. However, the oscillator model has very special features, in particular the degeneracy of the $n_\perp$-subshell. Before attempting a detailed comparison with experiment we must give up these special features and make the single-particle Hamiltonian more realistic.

The Nilsson model[24] has been very succesful in explaining the properties of odd-particle states in deformed nuclei[15]. The nucleons are supposed to be in states very similar to those of a deformed harmonic oscillator. The states are labelled by $(N, n_z, \Lambda, \Lambda+\Sigma)$. N and n_z have the same meaning as before, and Λ and Σ are the components, along the symmetry axis, of the orbital and spin angular momenta, respectively. However, this model differs essentially from the harmonic oscillator in that states with the same n_z are no longer degenerate.

We must now consider nuclei with both neutrons and protons outside closed shells. Apart from some very exceptional cases, there are no nuclei in the deformed region in which an external neutron and proton are occupying time-reversed states. Thus, the pairing force we have been using (2) will not couple the neutrons and protons. They will, however, be coupled by the deformed field. The matrix for the single-particle neutron Hamiltonian is, in the Nilsson representation,

$$(\varepsilon_n)_{\nu\omega} = (\varepsilon_n^N)_\nu \, \delta_{\nu\omega} - (\varkappa_n S_n + \varkappa_{np} S_p)\,(s_n)_{\nu\omega}. \tag{61}$$

A corresponding expression holds for protons. The coupling constants $\varkappa_n$, $\varkappa_p$, $\varkappa_{np}$ and $\varkappa_{pn}$ are to be determined, in principle, by the isotopic spin dependence of the nuclear force plus renormalization effects. It will be assumed in the following that $\varkappa_n = \varkappa_p$ and $\varkappa_{np} = \varkappa_{pn}$. The ε_ν^N are the single-particle energies calculated by Nilsson*. They already contain the terms depending on the axially symmetric part of the deformation. The subscripts n and p indicate neutrons and protons, respectively.

* Relatively small shifts will be made in the energy of some of Nilsson's levels in order to get closer agreement with empirical level ordering in odd-mass nuclei (cf. p. 30).

One has then to solve the equations (4) and (12) for neutrons and protons separately*. The neutron-proton coupling is expressed by the terms $-\varkappa_{pn} S_n (S_p)_{op}$ and $-\varkappa_{np} S_p (S_n)_{op}$. In the absence of these terms, we can calculate the vibrational parameters for separate neutron and proton γ-vibrations**. The problem is then equivalent to that of two coupled harmonic oscillators

$$H_\gamma = \frac{B_n}{2}\dot{S}_n^2 + \frac{B_p}{2}\dot{S}_p^2 + \frac{C_n}{2}S_n^2 + \frac{C_p}{2}S_p^2 - \varkappa_{np} S_p S_n . \tag{62}$$

The last term contains the usual factor of 1/2.

We can now decouple the oscillators by transforming to normal coordinates. The lower eigenfrequency is given by

$$\hbar w_\gamma = \left[\frac{1}{2}\left\{\frac{C_p}{B_p} + \frac{C_n}{B_n} - \left[\left(\frac{C_p}{B_p} - \frac{C_n}{B_n}\right)^2 + \frac{4\varkappa_{np}^2 \hbar^2}{B_p B_n}\right]^{1/2}\right\}\right]^{1/2}. \tag{63}$$

We can also calculate the probability of the electric quadrupole transition connecting the first γ-vibrational state with the ground state. For this purpose, it is convenient to regard the γ-vibration as a superposition of two travelling waves[22], Q_2 and Q_{-2}, with definite angular momentum projections along the symmetry axis, and with the same vibrational parameters. The operator $\mathfrak{M}(E2,\mu)$ responsible for the $E2$ transitions[14] is related to Q_μ by

$$\mathfrak{M}_{n,p}(E2,\mu) = \frac{1}{4}\sqrt{\frac{5}{\pi}}\, e_{np} Q_\mu^* , \tag{64}$$

where $e_{n,p}$ is the effective electrical charges carried by the neutron or the proton, respectively. Using eq. (V.34) of ref.[14] the square of the transition matrix element for a single oscillator is found to be

$$|\langle 2 | \mathfrak{M}(E2,2) | 0 \rangle|^2 = \frac{5\, e^2}{16\pi}\frac{\hbar}{2[BC]^{1/2}}. \tag{65}$$

For the coupled harmonic oscillators it is

$$|\langle 2 | \mathfrak{M}(E2,2) | 0 \rangle|^2 = \frac{5}{64\pi}\frac{\hbar}{w_\gamma}\left[\frac{e_p^2}{B_p}(1-\cos b) + \frac{e_n^2}{B_n}(1+\cos b) + \frac{e_p e_n}{\sqrt{B_p B_n}}\sin b\right], \tag{66}$$

where

$$\operatorname{tg} b = \frac{2\varkappa_{np}\sqrt{B_p B_n}}{(B_n C_p - B_p C_n)}. \tag{67}$$

* Some of the wave functions were kindly supplied by S. G. Nilsson; others were derived in collaboration with Z. Szymański.

** As the operator x^2-y^2 is not diagonal in the Nilsson representation, we have to use (29) and (33).

The reduced transition probability from the ground state to the first γ-vibrational level may be written (cf. eq. V.33 and note 175 of ref. 14)

$$B(E2; 0 \to 2) = 2 \, | \langle 2 \, | \mathfrak{M}(E2, 2) \, | \, 0 \rangle |^2. \tag{68}$$

We must now discuss the choice of the parameters G_n, G_p, $\varkappa_n$, $\varkappa_{np}$, e_n and e_p*.

In the deformed region, there is no clear distinction between filled and unfilled shells. Therefore, one has no definite prescription for the states into which the pairing force is allowed to scatter. However, states which lie far from the Fermi level contribute to the wave function only through a renormalization of G[7] [10]. Therefore, we have chosen to allow the pairing force to scatter only amongst the 24 states nearest to the Fermi level.

One has to choose an effective value of G_n and G_p such that $2\Delta_n$ and $2\Delta_p$ reproduce the average differences between the neutron and proton binding energies of even and odd-mass nuclei. Furthermore, the predicted quasi-particle excitations should be tested with experimental data. However, one should expect some shiftings due to quasi-particle interactions, to the blocking of some states near the Fermi surface, etc. Therefore, the empirical quasi-particle energies give only a lower limit on the value of G.

Finally we have chosen a value of $G_n = \dfrac{26.5}{A}$ Mev and of $G_p = \dfrac{32.1}{A}$ Mev.

The values of $\varkappa_n$ and $\varkappa_{np}$ enter into the calculation of the ground state quadrupole moment. The method used here is analogous to that used in the derivation of (55).

The part of Nilsson's potential responsible for the deformation is

$$-\frac{\delta M W^2}{3}(2z^2 - x^2 - y^2). \tag{69}$$

Comparing (69) with the corresponding term in our single-particle Hamiltonian (13[1]), we get

$$\frac{\delta M W^2}{3} = \mu = \varkappa Q + \hat{\mu}. \tag{70}$$

Using the Nilsson single-particle energies corresponding to a given value of δ, we solve (4) and (12) and thus obtain the electric Q_e and mass Q quadrupole moments as a function of δ. The inverse of the first function enables us to determine δ_{eq} from the observed equilibrium electric quadrupole

* The determination of these constants is only outlined here. It is given with more details in[25].

moment. At equilibrium, $\hat{\mu}$ vanishes and thus the value of $\varkappa$ which would yield these values of δ_{eq} and $Q(\delta_{\mathrm{eq}})$ is given by

$$\varkappa = \frac{MW^2}{3} \frac{\delta_{\mathrm{eq}}}{Q(\delta_{\mathrm{eq}})}. \tag{71}$$

In the rare-earth region, the observed electric quadrupole moments are reproduced by*

$$\varkappa = 122\, A^{-5/3} \left(\frac{MW}{\hbar}\right)^2 \mathrm{Mev}. \tag{72}$$

In the calculation of Q_e and Q, V_ν has been chosen to be unity for the states below the selected 24, and zero for those above. The single-particle mass quadrupole moment is given by**

$$\left.\begin{aligned} q_{\nu\omega} &= \frac{\hbar}{M} \langle \nu \mid \left(\frac{2z'^2}{W_z} - \frac{x'^2 + y'^2}{W_\perp}\right) \mid \omega \rangle \\ &= \frac{\hbar}{MW} \Big\{ \langle \nu \mid (2z'^2 - x'^2 - y'^2) \mid \omega \rangle \\ &+ \frac{\delta}{3} \left[\left(1 + \frac{3}{2}\delta + \ldots\right) \langle \nu \mid (2z'^2 - x'^2 - y') \mid \omega \rangle + (2 + \delta + \ldots) \left(N + \frac{3}{2}\right) \right] \Big\} \end{aligned}\right\} \tag{73}$$

where $X'^2 = \frac{MW_\perp}{\hbar} x^2$; $y'^2 = \frac{MW_\perp}{\hbar} y^2$; $z'^2 = \frac{MW_z}{\hbar} z^2$.

Use has been made of the relations[24]

$$W_z^2 = W^2 \left(1 - \frac{4}{3}\delta\right) \qquad W_\perp^2 = W^2 \left(1 + \frac{2}{3}\delta\right). \tag{74}$$

It has been verified that the contribution to the total quadrupole moment from the terms multiplied by δ in (73) is equal to the contribution due to the first term. In other words, the same results could be obtained by using a renormalized value of $\varkappa$ equal to twice[9] the value given in (72), and using for the single-particle quadrupole matrix elements the value given by the first term in (73). The coupling parameter $\varkappa$, calculated in this way, is to be considered here as an average value of $\varkappa_n$ and $\varkappa_{np}$.

We are going to calculate the energy and transition probability for the γ-vibrations in three cases, namely $\varkappa_n = \varkappa_{np} = x$; $6\varkappa_n = 2\varkappa_{np} = 3\varkappa$ and

* The $A^{-7/3}$ dependence of $\varkappa$ has been pointed out by Belyaev[7].

** The single-particle matrix elements have been obtained using the expression of the wave functions in terms of the asymptotic representation[26].

$\varkappa_n = \varkappa_{np} = 1.3\,\varkappa$. The value of x to be used is equal to twice that given in (72). The single-particle matrix element will be

$$s_{\nu\omega} = \frac{\hbar}{MW}\sqrt{3}\,\langle \nu \,|\, (x'^2 - y'^2) \,|\, w \rangle . \tag{75}$$

Determination of e_n and e_p. It was found above that renormalization effects doubled the mass quadrupole moment. We have used for the additional charge due to these effects a value of Z/A times the additional mass. Accordingly,

$$e_p = e\,(1 + Z/A) \qquad e_n = e\,Z/A\,, \tag{76}$$

where e is the charge of a free proton.

Results of the calculations. Table I a contains the value of the summations $\sum_1 = \sum_{\nu,\,\omega} \frac{s^2_{\nu\omega}\,(U_\nu V_\omega + U_\omega V_\nu)^2}{E_\nu + E_\omega}$ and $\sum_3 = \sum_{\nu,\,\omega} \frac{s^2_{\nu\omega}\,(U_\nu V_\omega + U_\omega V_\nu)^2}{(E_\nu + E_\omega)^3}$, calculated for some neutron numbers and for the deformation listed in column 1. Columns 4, 5 and 6 contain the restoring force for the neutron vibration, assuming $\varkappa_n = \varkappa_{np} = \varkappa$; $6\,\varkappa_n = 2\,\varkappa_{np} = 3\,\varkappa$ and $\varkappa_n = \varkappa_{np} = 1.3\,\varkappa$, respectively. Column 7 lists the mass parameter. Table I b is the analogous table corresponding to protons.

Table II contains the predicted energy of the first γ-vibrational level and the experimental value.

The two first calculations show that the value of the energy of the first vibrational level does not depend on the ratio $\varkappa_n/\varkappa_{np}$. At the beginning of the deformed region the predicted energies are about 80 per cent greater than the empirical ones. There is, however, a correlation between the empirical and theoretical trends (i. e., decrease in the energy for Er^{166}). This decrease is due mainly to the relative large values of $n_\perp$ for the states which come near to the Fermi energy.

At the end of the deformed region, the predicted trends and order of magnitude of the energy are in good agreement with the experimental values. However, a detailed comparison is hindered in the region of W and Os by the uncertainty in the parameters used. The predicted energies are rather sensitive to the position of the (5101/2) and (5123/2) neutron levels. In the calculations, these levels have been depressed by 250 kev in order to fit the spectrum of W^{183}. Calculations with the original Nilsson energies would decrease the energy of the γ-vibrations for 112 neutrons and increase it for 110 neutrons (keeping $\delta = 0.20$) and would thus give a somewhat better fit. Furthermore, the experimental evidence on the value of δ is not

TABLES I a and I b. The rare-earth region. The units of $\sum_1$ are $\left(\frac{\hbar}{MW}\right) \times \mathrm{Mev}^{-1}$. Those of $\sum_3$ are $\left(\frac{\hbar}{MW}\right)^2 \times \mathrm{Mev}^{-3}$. The units of C, C' and C'' are the inverse of those of $\sum_1$. Column C refers to the case $\varkappa_n = \varkappa_{np} = \varkappa$; column C' to the case $3\varkappa_n = \varkappa_{np} = \frac{3\varkappa}{2}$, and column C'' to the case $\varkappa_n = \varkappa_{np} = \varkappa\, 1.3$. The units of B are $(MW)^2 \times \mathrm{Mev}^{-1}$.

TABLE I a.

N	δ	Σ_1	Σ_3	C_n	C_n'	C_n''	B_n
90.....	0.25	39.8	3.38	0.0070	0.0098	0.0053	0.00107
	0.30	38.1	3.47	75	103	58	119
92.....	0.30	34.3	3.03	92	119	76	129
94.....	0.30	34.5	3.94	93	119	77	165
98.....	0.30	38.6	5.29	81	105	66	177
100.....	0.30	37.2	4.32	88	111	74	156
106.....	0.25	54.5	9.22	48	70	36	156
108.....	0.20	60.6	10.81	41	62	28	147
110.....	0.20	67.7	14.0	33	53	21	153
112.....	0.15	75.7	15.8	26	46	15	138
	0.20	68.2	14.2	34	53	22	152
114.....	0.15	74.1	14.4	28	48	17	131

TABLE I b.

Z	δ	Σ_1	Σ_3	C_p	C_p'	C_p''	B_p
62.....	0.25	24.8	2.01	0.0145	0.0173	0.0128	0.00163
	0.30	22.2	1.74	169	197	151	176
64.....	0.30	24.9	2.85	147	173	131	230
66.....	0.30	31.1	5.74	109	135	94	297
68.....	0.30	30.8	5.65	114	138	99	298
70.....	0.30	21.0	1.95	191	215	177	221
72.....	0.25	26.2	1.98	147	169	133	144
74.....	0.15	40.0	5.14	84	105	72	161
	0.20	34.3	3.82	105	126	93	162
76.....	0.15	44.2	6.95	74	93	62	178
	0.20	43.2	6.65	76	95	64	176

TABLE II. The rare-earth region. Columns 2, 3 and 4 are in Mev. They correspond to the case $\varkappa_n = \varkappa_{np} = \varkappa$; $3\,\varkappa_{np} = \frac{3\,\varkappa}{2}$ and $\varkappa_n = \varkappa_{np} = \varkappa\,1.3$, respectively. Column 6 lists the experimental energy.

Nucleus	δ	$(\hbar w_\gamma)_{\text{th}}$	$(\hbar w'_\gamma)_{\text{th}}$	$(\hbar w''_\gamma)_{\text{th}}$	$(\hbar w_\gamma)_{\text{exp.}}$ [27) 28)]
$_{62}Sm^{152}_{90}$	0.25	1.79	1.84	0.77	1.092
	0.30	1.91	1.97	1.18	
Sm^{154}_{92}	0.30	2.11	2.14	1.50	
$_{64}Gd^{156}_{92}$	0.30	1.90	1.91	1.30	1.152
Gd^{158}_{94}	0.30	1.81	1.82	1.26	1.182
$_{66}Dy^{160}_{94}$	0.30	1.45	1.46	0.89	0.964
Dy^{164}_{98}	0.30	1.38	1.38	0.78	
$_{68}Er^{166}_{98}$	0.30	1.43	1.43	0.88	0.787
Er^{168}_{100}	0.30	1.52	1.52	1.03	0.822
$_{70}Yb^{170}_{100}$	0.30	2.05	2.10	1.64	
$_{72}Hf^{178}_{106}$	0.25	1.44	1.56	0.71	
$_{74}W^{182}_{108}$	0.20	1.16	1.24	—	1.222
W^{184}_{110}	0.20	0.92	1.09	—	0.903
W^{186}_{112}	0.15	0.65	0.76	—	0.730
	0.20	1.02	1.10	—	
$_{76}Os^{186}_{110}$	0.20	0.75	0.81	—	0.768
Os^{188}_{112}	0.15	0.53	0.64	—	0.628
	0.20	0.80	0.85	—	
Os^{190}_{114}	0.15	0.66	0.73	—	0.558

so precise for the W and Os isotopes as in other rare-earth nuclei. Table II indicates good agreement for W^{184}, using $\delta = 0.20$, and for W^{186} using a value of δ intermediate between 0.15 and 0.20.

In addition, the restoring force becomes very small, so higher order terms in S could become more important.

We have also performed the calculations using a coupling constant $\varkappa$ which is 30 per cent greater than the one determined by considerations on the axially symmetric equilibrium deformations. These calculations give good agreement for the γ-energies at the beginning of the deformed region and they lead to γ-instability in W and Os. If such would be the case, the γ-vibrations in Hf should be especially low. However, this fact does not appear to be supported by experimental data.

One can estimate roughly the effect of the neglected Coulomb interaction by assuming an ellipsoid with constant density of charge[24]. The Coulomb energy is

$$E_c = \frac{3}{5}\frac{Z^2 e^2}{R}\left[1 - \frac{5}{36}\frac{(Q^2+S^2)}{A^2R^4}\right]. \tag{77}$$

The ratio $\frac{1}{12}\left(\frac{Z}{A}\right)^2 \frac{e^2}{R^5}\cdot\frac{2}{\varkappa} = 1.40\times 10^{-3}A^{2/3}$ is 0.04 if $A = 170$. $\left(R = 1.2\times 10^{-13} A^{1/3}\text{cm};\ \frac{Z}{A} = 0.4\right)$. Thus the effect of the Coulomb interaction amounts only to a 4 per cent change in $\varkappa$ in the middle of the first deformed region.

Table III contains the reduced transition probabilities calculated by means of (68) and parameters determined above. In the Gd, Dy, Er, and W-isotopes, the predicted values of the reduced transition probabilities are in agreement with experiment. In Os^{188} and Os^{190}, however, the predicted transition rates are about three times the experimental values. The disturbing aspect of the discrepancy is the fact that, experimentally, no increase in the transition rate occurs as the energy of the γ-vibrations decreases. We expect such an increase since the decrease in the γ-energy is principally due to a reduction in the restoring force, which should lead to oscillations of greater amplitude. On the other hand, if the restoring force goes to zero (γ-unstable oscillations[21]) the transition from the second 2+ state to the ground state is completely forbidden. This reveals an incompleteness in the present treatment, due to the fact that our wave functions do not have the required symmetry properties[16]. This symmetrization would give rise to interference terms which are responsible for the cancellation of the above-mentioned matrix element as the system approaches γ-instability. But these interference effects should be small if the root mean square value of γ is small compared with $\pi/3$. Estimated values for this quantity are also listed in Table III. They have been calculated by means of

$$\gamma_{\text{r. m. s}} = \sqrt{\frac{2Q_2^2}{Q_0^2}} = \sqrt{\frac{B(E2;\,00\to 22)_m}{B(E2;\,00\to 20)_m}}, \tag{78}$$

where $B(E2)_m$ is the usual reduced transition probability calculated, assuming the same charge for neutrons as for protons. However, the ratio (77) can be well approximated by the ratio between the reduced transition probabilities obtained with the effective charges (76).

It seems that for Os^{188} and for Os^{190} the above-mentioned interference effects could begin to be important. The inclusion of higher-order terms which may have "γ-parity" would increase these interference effects.

The present estimates of $\gamma_{\text{r. m. s}}$ may provide also a test about the validity

TABLE III. Reduced transition probabilities, in units af $10^{-48}\, e^2\,\mathrm{cm}^4$, for the transition from the ground state to the first γ-vibrational level. The last column lists estimated root-mean-square γ-values in units of $\pi/3$. The interference effects neglected here are unimportant if $\gamma_{\mathrm{r.\,m.\,s.}} \ll \frac{\pi}{3}$.

Nucleus	δ	$B(E2; 0\to\gamma)_{\mathrm{th}}$	$B(E2; 0\to\gamma)'_{\mathrm{th}}$	$B(E2; 0\to\gamma)''_{\mathrm{th}}$	$B(E2; 0\to\gamma)_{\mathrm{exp}}$ [28) 29)]	γ_{rms}
$_{62}Sm^{152}_{90}$	0.25	0.12	0.12	0.30	—	0.20
	0.30	0.10	0.10	0.17	—	0.20
Sm^{154}_{92}	0.30	0.09	0.10	0.13	—	0.17
$_{64}Gd^{156}_{92}$	0.30	0.10	0.11	0.16	~ 0.16	0.17
Gd^{158}_{94}	0.30	0.11	0.11	0.14	~ 0.16	0.16
$_{66}Dy^{160}_{94}$	0.30	0.13	0.13	0.24	—	0.18
Dy^{164}_{98}	0.30	0.12	0.12	0.22	—	0.16
$_{68}Er^{166}_{98}$	0.30	0.12	0.12	0.19	~ 0.22	0.16
Er^{168}_{100}	0.30	0.13	0.13	0.18	~ 0.22	0.16
$_{70}Yb^{170}_{100}$	0.30	0.07	0.08	0.09	—	0.13
$_{72}Hf^{178}_{106}$	0.25	0.10	0.10	0.21	—	0.17
$_{74}W^{182}_{108}$	0.20	0.16	0.18	—	~ 0.12	0.23
W^{184}_{110}	0.20	0.17	0.19	—	0.17 ± 0.05	0.26
W^{186}_{112}	0.15	0.37	0.34	—	0.17 ± 0.03	0.40
	0.20	0.17	0.18	—		0.28
$_{76}Os^{186}_{110}$	0.20	0.19	0.22	—	—	0.33
Os^{188}_{112}	0.15	0.58	0.47	—	—	—
	0.20	0.19	0.23	—	0.20 ± 0.06	0.36
Os^{190}_{114}	0.15	0.35	0.37	—	0.14 ± 0.03	0.48

All measurements of ref. 28 carry an experimental uncertainty of a factor of 2.

of the models which take into account only the degrees of freedom associated with an asymmetric rotor.

In order to summarize our results, we can say that, without any free parameter, we have been able to predict energies for the γ-vibrations which are in good agreement with experimental data at the end of the deformed region. At the beginning, the predicted γ-energies are too high, but the structure in the empirical curve is predicted theoretically. The experimental transition rates are also well accounted for, with the exception of the Os isotopes.

Some calculations have still to be performed in order to test the validity of some of our assumptions. For instance, we have to treat the closed shells explicitly in order to check the renormalization idea. In addition, the use of a central potential which is essentially an harmonic oscillator one, may

overestimate the dependence of the matrix elements $s_{\nu\omega}$ on the asymptotic quantum numbers.

The author wishes to express his gratitude to Professors NIELS BOHR and AAGE BOHR for their hospitality at the Institute for Theoretical Physics, Copenhagen, and for a grant. Thanks are due Professors A. BOHR and B. R. MOTTELSON for many suggestions offered during the course of this investigation. A special acknowledgement is due Dr. B. F. BAYMAN for criticisms and assistance in the preparation of the manuscript. Discussions with other members of the group working in Copenhagen on theoretical nuclear physics during 1958—1959 have been very useful and stimulating.

Institute for Theoretical Physics
University of Copenhagen.

Comisión Nacional de Energia Atómica
Buenos Aires, Argentina.

Appendix

General calculation of the restoring force and mass parameters.

We consider a system at its equilibrium deformation. We assume that we already know the values of λ, Δ and the V_ν. We allow a small change, R, in the previous single-particle field. The matrix (13[1]), with $\mu = \varkappa Q_{\text{equilibrium}}$, $\sigma = \varkappa S_{\text{equilibrium}}$, has to be modified by the addition of a term which is generally not diagonal

$$-(\varkappa R + \hat{\varrho})\, r_{\nu\omega} = -\varrho\, r_{\nu\omega}. \tag{1A}$$

$\hat{\varrho}$ is again a Lagrange multiplier.

The quantities denoted by a prime will refer to the new single-particle field. We shall expand in ϱ and keep only linear terms.

$$\left.\begin{aligned} \lambda' &= \lambda + \varrho\lambda_1 \\ \Delta'^2 &= \Delta^2 + \varrho\Delta_1^2 \end{aligned}\right\} \tag{2A}$$

We diagonalize the single-particle Hamiltonian by means of perturbation theory

$$\left.\begin{aligned} \varepsilon'_\nu &= \varepsilon_\nu - \varrho\,(r_{\nu\nu} + \lambda_1) \\ c'_\nu &= c_\nu + \varrho \sum_\omega \frac{r_{\nu\omega}}{\varepsilon_\omega - \varepsilon_\nu}\, c_\omega \end{aligned}\right\} \tag{3A}^*$$

and use a procedure similar to the one outlined in p. 12. The expressions for E_ν and V_ν^2 are

$$\left.\begin{aligned} E_\nu^1 &= E_\nu + \frac{\varrho}{E_\nu}\left[\frac{\Delta_1^2}{2} - (r_{\nu\nu} + \lambda_1)\,\varepsilon_\nu\right] \\ V_\nu'^2 &= V_\nu^2 + \frac{\varrho}{2E_\nu^3}\left[\frac{\Delta_1^2\,\varepsilon_\nu}{2} + (r_{\nu\nu} + \lambda_1)\,\Delta^2\right]. \end{aligned}\right\} \tag{4A}$$

* Since the quadrupole operator is even under time-reversal ($r_{\nu+\omega+} = r_{\nu-\omega-}$) it follows that the + sign in the second equation (3A) holds for both $c_{\nu+}$ and $c_{\nu-}$. In consequence, $H'_{\text{pair}} = H_{\text{pair}} + 0\,(\varrho^2)$. When this condition is not satisfied (i. e., for the Coriolis force) the following treatment may not be valid.

The two basic equations (4) and (12) must be satisfied independently of the value of ϱ. From the requirement that the terms proportional to ϱ vanish, it follows that

$$\left.\begin{aligned} &\lambda_1 \sum_\nu \frac{\varepsilon_\nu}{E_\nu^3} - \frac{\Delta_1^2}{2} \sum_\nu \frac{1}{E_\nu^3} + \sum_\nu \frac{r_{\nu\nu}\,\varepsilon_\nu}{E_\nu^3} = 0 \\ &\lambda_1 \sum_\nu \frac{1}{E_\nu^3} + \frac{\Delta_1^2}{2\Delta^2} \sum_\nu \frac{\varepsilon_\nu}{E_\nu^3} + \sum_\nu \frac{r_{\nu\nu}}{E_\nu^3} = 0 . \end{aligned}\right\} \quad (5\,\mathrm{A})$$

The solution of this system of equations is

$$\left.\begin{aligned} \lambda_1 &= \frac{1}{\left[\left(\sum_\nu \frac{\varepsilon_\nu}{E_\nu^3}\right)^2 + \Delta^2\left(\sum_\nu \frac{1}{E_\nu^3}\right)^2\right]} \left[\left(\sum_\nu \frac{\varepsilon_\nu}{E_\nu^3}\right)\left(\sum_\nu \frac{r_{\nu\nu}\,\varepsilon_\nu}{E_\nu^3}\right) + \Delta^2\left(\sum_\nu \frac{1}{E_\nu^3}\right)\left(\sum_\nu \frac{r_{\nu\nu}}{E_\nu^3}\right)\right] \\ \Delta_1^2 &= \frac{2\Delta^2}{\left[\left(\sum_\nu \frac{\varepsilon_\nu}{E_\nu^3}\right)^2 + \Delta^2\left(\sum_\nu \frac{1}{E_\nu^3}\right)^2\right]} \left[\left(\sum_\nu \frac{1}{E_\nu^3}\right)\left(\sum_\nu \frac{r_{\nu\nu}\,\varepsilon_\nu}{E_\nu^3}\right) - \left(\sum_\nu \frac{\varepsilon_\nu}{E_\nu^3}\right)\left(\sum_\nu \frac{r_{\nu\nu}}{E_\nu^3}\right)\right]. \end{aligned}\right\} \quad (6\,\mathrm{A})$$

Using equations (3A), (4A) and (6A), the new values of V_ν and U_ν can be calculated. Then the new ground state wave function $|\,0'\rangle$ can be expressed in the representation corresponding to the equilibrium deformation

$$\left.\begin{aligned} |\,0'\rangle &= \prod_\nu \left[U'_\nu + V'_\nu\, c'^{+}_{\nu+}\, c'^{+}_{\nu-}\right] |\,\text{vacuum}\,\rangle \\ &= |\,0\,\rangle \\ &+ \varrho \sum_\nu \frac{1}{4\,U_\nu V_\nu} \left[(r_{\nu\nu} + \lambda_1)\,\frac{\Delta^2}{E_\nu^3} + \frac{\Delta_1^2}{2}\,\frac{\varepsilon_\nu}{E_\nu^3}\right] \alpha_\nu^+\,\beta_\nu^+\,|\,0\,\rangle \\ &+ \varrho \sum_\nu \sum_{\omega \neq \nu} \frac{V_\nu\,U_\omega}{\varepsilon_\omega - \varepsilon_\nu}\, r_{\nu\omega}\,(\alpha_\nu^+\,\beta_\omega^+ + \alpha_\omega^+\,\beta_\nu^+)\,|\,0\,\rangle . \end{aligned}\right\} \quad (7\,\mathrm{A})$$

Because of the identity $(E_\nu + E_\omega)(U_\omega V_\nu - U_\nu V_\omega) = (\varepsilon_\omega - \varepsilon_\nu)(U_\omega V_\nu + U_\nu V_\omega)$ the third term in (7A), which contains non-diagonal single-particle matrix elements $r_{\nu\omega}$, can be cast into the form

$$\varrho \sum_{\nu,\,\omega} \frac{r_{\nu\omega}\,(U_\nu V_\omega + U_\omega V_\nu)}{(E_\nu + E_\omega)}\,\alpha_\nu^+\,\beta_\omega^+\,|\,0\,\rangle . \quad (8\,\mathrm{A})$$

Using the relations $\langle\,0\,|\,R_{o\rho}\,\alpha_\nu^+\,\beta_\omega^+\,|\,0\,\rangle = r_{\nu\omega}\,(U_\nu V_\omega + U_\omega V_\nu)$ and $2\,U_\nu V_\nu = \dfrac{\Delta}{E_\nu}$, and equations (7A) and (8A), we find

$$\left.\begin{aligned} R &= \langle 0' \mid R_{op} \mid 0' \rangle \\ &= \langle 0 \mid R_{op} \mid 0 \rangle \\ &+ \varrho \left[2 \sum_{\nu,\omega} \frac{r_{\nu\omega}^2 (U_\nu V_\omega + U_\omega V_\nu)^2}{E_\nu + E_\omega} + \lambda_1 \Delta^2 \sum_\nu \frac{r_{\nu\nu}}{E_\nu^3} + \frac{\Delta_1^2}{2} \sum_\nu \frac{r_{\nu\nu}\varepsilon_\nu}{E_\nu^3} \right] \end{aligned}\right\} \quad (9\,\mathrm{A})$$

from which the derivative $\frac{d\varrho}{dR}$, and thus the restoring force (eq. 29), can be immediately obtained.

The calculation of the mass parameter is done according to the prescriptions of the "crancking model":

$$\left.\begin{aligned} B &= 2\hbar \sum_{\nu,\omega} \frac{|\langle \nu\omega \mid \frac{\delta}{\delta R} \mid 0 \rangle|^2}{E_\nu + E_\omega} \\ &= 2\hbar^2 \left(\frac{d\varrho}{dR}\right)^2 \sum_{\nu,\omega} \frac{|\langle \nu\omega \mid \frac{d}{d\varrho} \mid 0 \rangle|^2}{E_\nu + E_\omega} \\ &= 2\hbar^2 \left(\frac{d\varrho}{dR}\right)^2 \left\{ 2 \sum_{\nu,\omega} \frac{r_{\nu\omega}^2 (U_\nu V_\omega + U_\omega V_\nu)^2}{(E_\nu + E_\omega)^3} \right. \\ &\quad \left. + \frac{1}{8} \sum_\nu \frac{1}{E_\nu^5} \left(\Delta^2 \lambda_1^2 - \frac{\Delta_1^4}{4} + \frac{\Delta_1^4 E_\nu^2}{4\Delta^2} + \lambda_1 \Delta_1^2 \varepsilon_\nu + 2\lambda_1 \Delta^2 r_{\nu\nu} + \Delta_1^2 r_{\nu\nu}\varepsilon_\nu \right) \right\}. \end{aligned}\right\} \quad (10\,\mathrm{A})$$

We see that if $\sum_\nu \frac{r_{\nu\nu}}{E_\nu^3} = \sum_\nu \frac{r_{\nu\nu}\varepsilon_\nu}{E_\nu^3} = 0$, it follows that $\lambda_1 = \Delta_1^2 = 0$. In this case, formulae (29) and (33) are correct. The first condition $\sum_\nu \frac{r_{\nu\nu}}{E_\nu^3} = 0$ is related to the non-conservation of the particle number (31). To interpret the second condition $\sum_\nu \frac{r_{\nu\nu}\varepsilon_\nu}{E_\nu^3} = 0$ we construct the two quasi-particle perturbed wave function in analogy to (27).

$$|(\nu\omega)'\rangle = |\nu\omega\rangle - \varrho \frac{\langle \nu\omega \mid R_{op} \mid 0 \rangle}{E_\nu + E_\omega} |0\rangle + \varrho \sum_{\xi\eta} \frac{\langle 0 \mid R_{op} \mid \xi\eta \rangle}{E_\xi + E_\eta} |\xi\eta\,\nu\omega\rangle, \quad (11\,\mathrm{A})$$

where $|\xi\eta\,\nu\omega\rangle = \alpha_\xi^+ \beta_\eta^+ \alpha_\nu^+ \beta_\omega^+ |0\rangle$ represents a four quasi-particle wave function. We require that the matrix element of H_{int} between $|\langle \nu\omega)'\rangle$ and $|0'\rangle$ contains no linear term in ϱ, and therefore can lead to no quadratic term in the expression for the total ground-state energy. From Appendix A of Belyaev[7] we have

$$\sum_{\xi\eta} \langle \xi\eta\, \nu\omega \mid H_{\text{int}} \mid 0 \rangle = \frac{G}{2} \delta_{\nu\omega} \sum_{\xi} (U_\nu^2 V_\xi^2 + U_\xi^2 V_\nu^2) \tag{12A}$$

$$\sum_{\xi\eta} \langle \xi\eta \mid H_{\text{int}} \mid \nu\omega \rangle = -\frac{G}{2} [\delta_{\nu\omega} \sum_{\xi} (U_\nu^2 U_\xi^2 + V_\nu^2 V_\xi^2) + 2\, U_\nu V_\nu U_\omega V_\omega]. \tag{13A}$$

The last term in (13A) is of order $\frac{1}{\Omega}$ compared to the first, as it contains no summation over all the single-particle states. Neglecting it and using (27), (11A), (12A), and (13A),

$$\left.\begin{aligned} &\langle 0' \mid H_{\text{int}} \mid (\nu\omega)' \rangle \\ &= \varrho \sum_{\xi\eta} \frac{\langle \xi\eta \mid R_{op} \mid 0 \rangle \langle 0 \mid H_{\text{int}} \mid \xi\eta\, \nu\omega \rangle}{E_\xi + E_\eta} + \varrho \sum_{\xi\eta} \frac{\langle 0 \mid R_{op} \mid \xi\eta \rangle \langle \xi\eta \mid H_{\text{int}} \mid \nu\omega \rangle}{E_\xi + E_\eta} \\ &= -\varrho \frac{G}{4} \delta_{\nu\omega} (U_\nu^2 - V_\nu^2) \sum_{\xi} \frac{\langle 0 \mid R_{op} \mid \xi\xi \rangle}{E_\xi} (U_\xi^2 - V_\xi^2). \end{aligned}\right\} \tag{14A}$$

Thus the validity of (27) implies the vanishing of (14A). This in turn requires

$$\sum_{\xi} \frac{\langle 0 \mid R_{op} \mid \xi\xi \rangle}{E_\xi} (U_\xi^2 - V_\xi^2) = \Delta \sum_{\xi} \frac{r_{\xi\xi}\, \varepsilon_\xi}{E_\xi^3} = 0. \tag{15A}$$

References

(1) J. P. Elliot, Proc. Royal Society A, 245, 128 (1958); 245, 562 (1958).
B. H. Flowers (and subsequent discussion), Proc. of the Rehovoth Conference on Nuclear Structure, Rehovoth, Israel, September 1957, p. 161. North-Holland Publishing Co., Amsterdam (1958).
B. R. Mottelson (to be published).

(2) B. F. Bayman, Comptes Rendus du Congrès International de Physique Nucléaire, Paris, July 1958, p. 740. Dunod, Paris (1959).

(3) A. Bohr and B. Mottelson, Mat. Fys. Medd. Dan. Vid. Selsk. **30**, no. 1 (1955).

(4) A. Bohr, B. Mottelson, and D. Pines, Phys. Rev., **110**, 936 (1958).

(5) J. Bardeen, L. N. Cooper, and J. R. Schriefer, Phys. Rev., **108**, 1175 (1957).

(6) G. Racah, Proc. of the Rehovoth Conference on Nuclear Structure, Rehovoth Israel, September 1957, p. 155. North-Holland Publishing Co., Amsterdam (1958).

(7) S. T. Belyaev, Mat. Fys. Medd. Dan. Vid. Selsk. **31**, no. 11 (1959).

(8) A. Bohr, Comptes Rendus du Congrès International de Physique Nucléaire Paris, July 1958, p. 203. Dunod, Paris (1959).

(9) B. Mottelson, The Many-Body Problem. Lectures given at "École d'Eté de Physique Théorique", Les Houches, France, 1958, p. 283. Dunod, Paris (1959).

(10) B. Mottelson, Lectures given at the Institute for Theoretical Physics, Copenhagen, autumn 1958 (Unprinted).

(11) L. Kisslinger and R. Sorensen, Mat. Fys. Medd. Dan. Vid. Selsk. **32**, no. 9 (1960).

(12) N. N. Bogolubov, JETP, USSR 34 (1958) 58 and 73, Nuovo Cimento, 7, 794 (1958).

(13) J. G. Valatin, Nuovo Cimento **7**, 843 (1958).

(14) K. Alder, A. Bohr, T. Huus, B. Mottelson, and A. Winther, Revs. Mod. Phys. **28**, 432 (1956).

(15) B. Mottelson and S. G. Nilsson, Mat. Fys. Skr. Dan. Vid. Selsk. **1**, no. 8 (1959).

(16) A. Bohr, Mat. Fys. Medd. Dan. Vid. Selsk. **26**, no. 14 (1952).

(17) Z. Szymański, Nuclear Physics **11**, 454 (1959).

(18) A. de Shalit, Phys. Rev. **113**, 373 (1959).

(19) D. Inglis, Phys. Rev. **96**, 1059 (1954); **97**, 701 (1955).

(20) D. R. Bès, Nuclear Physics **10**, 373 (1959).

(21) L. Wilets and M. Jean, Phys. Rev. **102**, 788 (1959).

(22) G. R. Rakavy, Nuclear Physics **4**, 289 (1957).

(23) C. A. MALLMANN, submitted to the Second International Conference on the Peaceful Uses of Atomic Energy, Geneva, September 1958.
(24) S. G. NILSSON, Mat. Fys. Medd. Dan. Vid. Selsk. **29**, no. 16 (1955).
(25) D. R. BÈS and Z. SZYMAŃSKI, to be published.
(26) A. J. RASSEY, Phys. Rev. **109**, 949 (1958).
(27) R. K. SHELINE, Revs. Mod. Phys. **32**, 1 (1960).
(28) O. NATHAN and V. I. POPOV, Nucl. Phys. **21**, 631 (1960).
(29) F. K. MCGOWAN, Comptes Rendus du Congrès International de Physique Nucléaire, Paris, July 1958, p. 225. Dunod, Paris (1959).

Indleveret til Selskabet den 18. november 1960.
Færdig fra trykkeriet den 21. juni 1961.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 3

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 3 (1961)

ON THE ANALYTIC PROPERTIES OF THE 4-POINT FUNCTION IN PERTURBATION THEORY

BY

ALFRED CHI-TAI WU

København 1961
i kommission hos Ejnar Munksgaard

Synopsis

The analytic properties of the 4-point function as a function of 6 complex invariants are studied in simplest perturbation theory examples. This is a generalization of the work by Källén and Wightman on the vertex function. The singularity manifolds are: one 4-point singularity manifold, 4 sets of the 3-point manifolds of the type discussed by KW, and 6 cuts. These are determined in three different ways, including an explicit evaluation of the 4-fold Feynman parameter integral which results in a sum of 192 Spence functions. It is shown from the existence of the non-trivial geometric envelopes that the regularity domain D_4^{pert} is in general not entirely bounded by the analytic hypersurfaces. The boundary of the domain is illustrated with the aid of the 1-mass surfaces in some typical configurations of the 6 complex variables, showing that the 4-point boundary will in general carve out bubble singularities from the 3-point boundary. It is hoped that the results here may give some insight into the problem of finding the envelope of holomorphy of the 4-point domain determined by the axioms of the local field theory alone.

Printed in Denmark
Bianco Lunos Bogtrykkeri A-S

I. Introduction*

In the study of the general structure of the local field theory on the basis of a few generally accepted postulates[1] (viz., field operators transforming according to the representations of the proper Lorentz group; positivity of energy of physical states; local commutativity for space-like separations; etc.), one is led to the investigation of the analytic properties of the vacuum expectation values of a product of field operators[2] and of related quantities such as the retarded commutators[3]. Several significant physical applications in this field have been made in recent years, e. g., the proofs of the dispersion relations[4], the CPT-theorem[5], and the connection between spin and statistics[6].

The significance of the vacuum expectation value of products of two fields (in short, the 2-point function) has been known for quite some time[7]. The complete 3-point analyticity domain $E(D_3)$ has been determined by KÄLLÉN and WIGHTMAN[8] as a consequence of the above axioms without mass spectrum, and more recently the integral representations of the Bergman-

* Preliminary results of Sec. IV were reported by J. S. TOLL at the Naples Conference (April, 1959) (see, ref. 13). I would like to thank Professor TOLL for this.

[1] See, e. g., A. S. WIGHTMAN, Phys. Rev. **101**, 860 (1956). See also, WIGHTMAN, in *Les Problèmes Mathématiques de la Théorie Quantique des Champs*, Lille (1957).

[2] For a comprehensive survey of the properties of such Wightman functions, see, e. g., R. JOST's Lecture Notes in the *International Spring School of Physics*, Naples (1959); and also JOST's article in „*Theoretical Physics in the Twentieth Century*", ed. FIERZ and WEISSKOPF, *Interscience Publishers*, New York (1960).

[3] See, e. g., H. LEHMANN, K. SYMANZIK, and W. ZIMMERMANN, Nuovo Cimento **1**, 205 (1955); and *ibid.* **6**, 319 (1957); V. GLASER, H. LEHMANN, and W. ZIMMERMANN, Nuovo Cimento **6**, 1122 (1957); O. STEINMANN, Helv. Phys. Acta **33**, 257 (1960); and *ibid.* **33**, 347 (1960).

[4] See, e. g., N. N. BOGOLIUBOV, B. V. MEDVEDEV, and M. K. POLIVANOV, Lecture Notes (translated at Institute for Advanced Study, Princeton, 1957), and FIZMATGIZ, Moscow (1958); H. J. BREMERMANN, R. OEHME, and J. G. TAYLOR, Phys. Rev. **109**, 2178 (1958); H. LEHMANN, Nuovo Cimento **10**, 579 (1958).

[5] R. JOST, Helv. Phys. Acta **30**, 409 (1957).

[6] N. BURGOYNE, Nuovo Cimento **8**, 607 (1958); cf. also G. LÜDERS and B. ZUMINO, Phys. Rev. **110**, 1450 (1958).

[7] In a 1951 paper by H. UMEZAWA and S. KAMEFUCHI, Prog. Theor. Phys. **6**, 543 (1951), one finds, e. g., the assumption about the positive definite energy of all physical states clearly stated. Furthermore, this paper also contains an explicit example of a reduction formula, viz., for the problem of vacuum polarization. See, further, G. KÄLLÉN, Helv. Phys. Acta. **25**, 417 (1952); H. LEHMANN, Nuovo Cimento **11**, 342 (1954).

[8] G. KÄLLÉN and A. S. WIGHTMAN, Mat. Fys. Skr. Dan. Vid. Selsk. **1**, No. 6 (1958). This paper will be referred to as KW.

Weil type have been given[9] as a most general representation for a function analytic in $E(D_3)$ and with arbitrary singularities outside.

The present investigation consists of a generalization to the 4-point case of a very special feature which was treated by KW in their discussion of the 3-point domain[10]. To make things perfectly clear as to how this might fit into the general framework in the 4-point case, it will perhaps be helpful to sketch briefly the necessary steps needed in the systematic exploitation of the analyticity domains of the n-point functions.

For an n-point function, one starts in the space of $(n-1)$ real 4-vectors ξ_i. The axiom of positivity of energy immediately allows an analytic continuation to the (complex) tube domain R_{n-1} with $\zeta_i = \xi_i - i\eta_i$ and all η_i lying inside the forward light-cone. Now there are three subsequent steps:

a) The Hall-Wightman theorem[11] maps this tube R_{n-1} into a domain $\boldsymbol{M}_{n-1}$ in the inner-product space of the $1/2\, n(n-1)$ complex variables[12]. The first problem is then to determine this primitive domain $\boldsymbol{M}_{n-1}$ (i. e., to characterize the boundary $\partial \boldsymbol{M}_{n-1}$). $\boldsymbol{M}_{n-1}$ is a natural domain of holomorphy[13].

b) By permuting the original vectors, one gets a permuted n-point function and thus a permuted domain $\mathfrak{P}\boldsymbol{M}_{n-1}$. Now by the axiom of strong locality, these permuted functions coincide on a certain space-like region $\boldsymbol{S}$. If $\boldsymbol{S} \cap \{\mathfrak{P}\boldsymbol{M}_{n-1}\} \neq 0$, then one gets a function analytic in the domain $D_n = \cup \{\mathfrak{P}\boldsymbol{M}_{n-1}\}$.

c) The domain D_n (because of the above union) is not a natural domain of holomorphy[14]. The final step is to find the envelope of holomorphy $E(D_n)$ of D_n[15].

We now briefly discuss separately the cases for $n \leqslant 4$.

Case 1) *2-point domain*: $\boldsymbol{M}_1$ is trivial; it is just the cut-plane (as is obvious from squaring a single (difference) vector ζ). The cut is along the positive real-axis. Steps (b) and (c) are unnecessary. $\boldsymbol{M}_1 = D_2 = E(D_2)$.

[9] G. KÄLLÉN and J. S. TOLL, in Pauli Memorial Volume, Helv. Phys. Acta. **33**, 753, (1960).

[10] See KW Appendix III and Section VII.

[11] D. HALL and A. S. WIGHTMAN, Mat. Fys. Medd. Dan. Vid. Selsk. **31**, No. 5 (1957).

[12] For $n \geqslant 5$, the number of independent inner products is reduced to $2(2n-5)$ by linear dependence of more than 4 vectors in 4-dimensional space-time.

[13] For $n \leqslant 3$, this is clear, since $\boldsymbol{M}_{1,\,2}$ are both bounded by analytic hypersurfaces, and one knows that one can go no further. For $n = 4$, one gets non-analytic hypersurfaces, however, this is still proved by KÄLLÉN and TOLL (private communication; and TOLL's Lecture Notes in International Spring School of Physics, Naples (1959)).

[14] Cf., for example, D. RUELLE, Helv. Phys. Acta **32**, 135 (1959) and thesis (1959), Bruxelles.

[15] For basic notions of the theory of functions of several complex variables, see, e. g., H. BEHNKE and P. THULLEN, *Theorie der Funktionen mehrerer komplexer Veränderlichen*, Ergebn. Math. **3** Nr. 3, Berlin (1934). For a physicist's summary, cf., e. g., KW Sec. VI ff.

Case 2) *3-point domain*:

a) Part of $\boldsymbol{M}_2$ was first treated by D. HALL[16]; it was simplified and exhausted by KW who show that $\boldsymbol{M}_2$ is bounded by the following pieces of analytic hypersurfaces:

$$F_{12}\colon\ z_3 = z_1 + z_2 + r + z_1 z_2/r,\ 0 < r < \infty, \qquad (\text{for } \boldsymbol{Im}\ z_1 \cdot \boldsymbol{Im}\ z_2 > 0);$$
$$S\colon\ z_3 = z_1(1-k) + z_2(1-1/k),\ 0 < k < \infty, \quad (\text{for } \boldsymbol{Im}\ z_1 \cdot \boldsymbol{Im}\ z_2 < 0),$$

and the cuts in z_1 and z_2.

b) Permutation is straightforward.

c) $E(D_3)$ turns out to be bounded also by analytic hypersurfaces:

$$\text{Cuts:}\ z_k = \varrho \geqslant 0,\ k = 1, 2, 3. \quad (0 < \varrho < \infty).$$
$$F'_{ij}\colon\ z_k = z_i + z_j - \varrho - z_i z_j/\varrho, \quad (\text{for } \boldsymbol{Im}\ z_k \cdot \boldsymbol{Im}\ z_i < 0,\ \boldsymbol{Im}\ z_k \cdot \boldsymbol{Im}\ z_j < 0);$$
$$\mathfrak{F}\colon\ z_1 z_2 + z_2 z_3 + z_3 z_1 - \varrho(z_1 + z_2 + z_3) + \varrho^2 = 0$$
$$(\text{for } \boldsymbol{Im}\ z_1 \cdot \boldsymbol{Im}\ z_2 > 0,\ \boldsymbol{Im}\ z_1 \cdot \boldsymbol{Im}\ z_3 > 0).$$

Case 3) *4-point domain*:

a) Part of the boundary of the primitive domain $\boldsymbol{M}_3$ has been very elegantly characterized by JOST[17] with a set of 3×3 matrices $M = \text{DANAD}$, where $M = ||(\zeta_i \cdot \zeta_i)||$, D is diagonal with positive diagonal elements, A is symmetric real except for diagonal elements which have positive imaginary parts, and N has zero diagonal elements and 1 everywhere else. That $\mathbf{M}_3$ is indeed a natural domain of holomorphy has been shown by KÄLLÉN and TOLL[18], who have also shown that $\boldsymbol{M}_3$ is *not* everywhere bounded by analytic hypersurfaces.

b) The permuted domain remains to be determined. This can be accomplished by the present technique if sufficient and careful work is carried through.

c) The real difficulty lies in the problem of finding the envelope of holomorphy $E(D_4)$, which is at the present moment completely unknown. It is therefore entirely an open question as to whether or not $E(D_4)$ will be bounded by analytic hypersurfaces.

At this point, we want to discuss the role of the domain D_n^{pert}, which one gets from simple yet non-trivial examples in perturbation theory. Let us recall the following facts:

[16] D. HALL, Ph. D. thesis, Princeton (1956).
[17] See Ref. cited in footnote 2.
[18] See Ref. cited in footnote 13.

Case 1a) $n = 2$: $D_2^{\text{pert}} = E(D_2)$.

Case 2a) $n = 3$: D_3^{pert} gives about three-fourths of the answer to $E(D_3)$, i. e., D_3^{pert} is bounded by cuts and F_{kl}' surfaces. *The only thing* D_3^{pert} *fails to tell is the* $\mathfrak{F}$*-surface* (which corresponds to the case when all $\boldsymbol{Im}\ z_k$ have the same sign). (In fact, it should perhaps be pointed out that it would be extremely difficult to discover the exact shape of $E(D_3)$ if one didn't know beforehand D_3^{pert}; a knowledge of which then enabled KW to actually prove the final results.)

It is in this spirit that the present study of the D_4^{pert} is undertaken. Namely, it is hoped that perhaps D_4^{pert} might again give some insight into the envelope of holomorphy $E(D_4)$ in the axiomatic approach.

The work divides itself into two parts. The first part (Sections II–V) is devoted to the explicit location of the singularities of the 4-point function in perturbation theory and their relevance criteria. The second part (Section VI) is to determine what constitutes the boundary of the domain; the study of this boundary is our primary interest.

The main result of this study is that D_4^{pert} is also *not* entirely bounded by analytic hypersurfaces. A lengthy analysis of the problem of the geometric envelopes for the 4-point singularity manifold is made (Section VI). The 4-mass envelopes and the 3-mass envelopes, although they can also exist, are shown to be trivial and cannot contribute to the boundary of the domain. On the other hand, the two-mass envelopes are quite nontrivial and have most natural relations with the 3-point boundary F_{kl}' surfaces. In principle, with the aid of an electronic computer, the boundary of D_4^{pert} can be explicitly plotted. However, we only give here the equations and illustrate instead the one-mass curves (which are analytic) for some typical configurations in the space of six complex variables to show the presence of the 2-mass envelopes.

It is evident that the fact that D_4^{pert} is not bounded by analytic hypersurfaces will make the problem for D_4^{pert} to provide some answer to $E(D_4)$ much less transparent than the previous 3-point case. Of course, it is trivial that $E(D_4) \subset D_4^{\text{pert}}$. However, as already mentioned above, it is still an open question whether or not $E(D_4)$ is bounded by analytic hypersurfaces. *If* D_4^{pert} does have anything to do with $E(D_4)$, then the present investigation gives a negative answer.

II. Simple Examples of the 4-Point Function

II. 1 The Vacuum Expectation Value of Products of Four Fields in Ward Theory (x-space)

We consider in perturbation theory an interaction via a Lagrangian $g\Phi_1\Phi_2\Phi_3\Phi_4$, where the Φ_i's are neutral scalar fields with field quanta m_i. Expanding in powers of g, we have

$$\Phi_j(x) = \Phi_j^{(0)}(x) + g\int dx'\, \Delta_R(x-x';\ m_j)\,\Phi_k^{(0)}(x')\,\Phi_l^{(0)}(x')\,\Phi_m^{(0)}(x') + \ldots\ldots \qquad (1)$$

where $(jklm)$ is a permutation of (1234).

To the first non-trivial order, the vacuum expectation value of the four fields reads:

$$\begin{aligned}
&<0\,|\,\Phi_1(x_1)\,\Phi_2(x_2)\,\Phi_3(x_3)\,\Phi_4(x_4)\,|\,0> \\
&= \frac{-g}{(2\pi)^9}\iiint dq_1\,dq_2\,dq_3\cdot\exp\left[i(q_1x_{14}+q_2x_{24}+q_3x_{34})\right] \\
&\times\Bigg[\Theta(-q_2)\,\Theta(-q_3)\,\Theta(q_1+q_2+q_3)\,\frac{\delta(q_2^2+m_2^2)\,\delta(q_3^2+m_3^2)\,\delta((q_1+q_2+q_3)^2+m_4^2)}{(q_1^2+m_1^2)_R} \\
&+\Theta(q_1)\,\Theta(-q_3)\,\Theta(q_1+q_2+q_3)\,\frac{\delta(q_1^2+m_1^2)\,\delta(q_3^2+m_3^2)\,\delta((q_1+q_2+q_3)^2+m_4^2)}{(q_2^2+m_2^2)_R} \\
&+\Theta(q_1)\,\Theta(q_2)\,\Theta(q_1+q_2+q_3)\,\frac{\delta(q_1^2+m_1^2)\,\delta(q_2^2+m_2^2)\,\delta((q_1+q_2+q_3)^2+m_4^2)}{(q_3^2+m_3^2)_R} \\
&+\Theta(q_1)\,\Theta(q_2)\,\Theta(q_3)\,\frac{\delta(q_1^2+m_1^2)\,\delta(q_2^2+m_2^2)\,\delta(q_3^2+m_3^2)}{((q_1+q_2+q_3)^2+m_4^2)_R}\Bigg],
\end{aligned} \qquad (2)$$

where $x_{ij} = x_i - x_j$. The Θ's are the usual step functions

$$\Theta(x_0) = \begin{cases} 0, & \text{for } x_0 < 0 \\ 1, & \text{for } x_0 > 0. \end{cases}$$

The scalar products for 4-vectors are here defined with the metric $(+++-)$.

For the choice of the 4-point function in x-space, a more convenient expression results if we multiply (2) with a suitable weight function $\mathfrak{G}(m_1^2, m_2^2, m_3^2, m_4^2)$ and integrate over all masses $m_k^2 \geqslant 0$. In particular, following KW, we choose

$$\mathfrak{G}(m_1^2, m_2^2, m_3^2, m_4^2) = \prod_{k=1}^{4} \Delta(-m_k^2;\ a_k), \quad a_k > 0, \qquad (3)$$

where

$$\bar{\Delta}(\lambda, \sigma) = \frac{1}{(2\pi)^4} \mathfrak{P}. \int dp \, \frac{e^{ip\alpha}}{p^2 + \sigma} \Big|_{\alpha^2 = \lambda} \tag{4}$$

in which $\mathfrak{P}.$ denotes the usual Cauchy principal part.
Using

$$\frac{2}{\pi} \mathfrak{P}. \int_0^\infty \frac{d\lambda \, \bar{\Delta}(-\lambda; \sigma)}{p^2 + \lambda} = \Delta^{(1)}(p^2; \sigma) = \frac{1}{(2\pi)^3} \int d\xi \, e^{ip\xi} \delta(\xi^2 + \sigma) \tag{5}$$

and

$$\Delta_R(x) = 2\Theta(x) \bar{\Delta}(x), \tag{6}$$

we have the expression

$$\left.\begin{aligned} I &= \int\int\int\int_0^\infty dm_1^2 \, dm_2^2 \, dm_3^2 \, dm_4^2 \, \varrho(m_1^2, m_2^2, m_3^2, m_4^2) < 0 \,|\, \Phi_1(x_1) \Phi_2(x_2) \Phi_3(x_3) \Phi_4(x_4) \,|\, 0 > \\ &= \frac{-g}{32 (2\pi)^8} \int\int\int dq_1 \, dq_2 \, dq_3 \cdot \exp\left[i(q_1 x_{14} + q_2 x_{24} + q_3 x_{34})\right] \\ &\quad \times \left[\Delta^{(1)}(q_1^2; a_1) \; \Delta_A(q_2^2; a_2) \; \Delta_A(q_3^2; a_3) \; \Delta_R((q_1 + q_2 + q_3)^2; a_4)\right. \\ &\quad + \Delta_R(q_1^2; a_1) \, \Delta^{(1)}(q_2^2; a_2) \; \Delta_A(q_3^2; a_3) \; \Delta_R((q_1 + q_2 + q_3)^2; a_4) \\ &\quad + \Delta_R(q_1^2; a_1) \; \Delta_R(q_2^2; a_2) \, \Delta^{(1)}(q_3^2; a_3) \; \Delta_R((q_1 + q_2 + q_3)^2; a_4) \\ &\quad \left. + \Delta_R(q_1^2; a_1) \; \Delta_R(q_2^2; a_2) \; \Delta_R(q_3^2; a_3) \, \Delta^{(1)}((q_1 + q_2 + q_3)^2; a_4)\right] \\ &= \frac{-g}{32 (2\pi)^{11}} \int\int\int\int\int d\xi_1 \, d\xi_2 \, d\xi_3 \, d\xi_4 \, \delta(\xi_1 + \xi_4 + x_{14}) \, \delta(\xi_2 + \xi_4 + x_{24}) \, \delta(\xi_3 + \xi_4 + x_{34}) \\ &\quad \times \left\{\frac{\delta(A_1)}{A_2 A_3 A_4} + \frac{\delta(A_2)}{A_1 A_3 A_4} + \frac{\delta(A_3)}{A_1 A_2 A_4} + \frac{\delta(A_4)}{A_1 A_2 A_3}\right\}, \end{aligned}\right\} \tag{7}$$

where

$$A_k = \xi_k^2 + a_k.$$

With the aid of the well-known identities

$$\sum_{\text{cyclic}} \frac{\delta(A_1)}{A_2 A_3 A_4} = -\int\int\int\int\int_0^1 d\alpha_1 \, d\alpha_2 \, d\alpha_3 \, d\alpha_4 \, \delta(1 - \Sigma\alpha_k) \, \delta^{(3)}(\Sigma\alpha_k a_k) \tag{8}$$

and

$$\frac{1}{\pi} \int d^4\xi \, \delta^{(3)}(\xi^2 + A) = -\frac{\partial}{\partial A} \left(\mathfrak{P}. \frac{1}{A}\right) \tag{9}$$

the integrals over all ξ_k can be easily carried out. The result is:

$$\left.\begin{aligned} I = \frac{g}{64 (2\pi)^{10}} \int\int\int\int\int_0^1 \frac{d\alpha_1 \, d\alpha_2 \, d\alpha_3 \, d\alpha_4 \, \delta(1 - \Sigma\alpha_k)}{(\alpha_1 \alpha_2 z_1 + \alpha_1 \alpha_3 z_2 + \alpha_1 \alpha_4 z_3 + \alpha_2 \alpha_3 z_4 + \alpha_3 \alpha_4 z_5 + \alpha_4 \alpha_2 z_6 - \Sigma\alpha_k a_k)^2} \\ a_k > 0, \end{aligned}\right\} \tag{10}$$

where the six z's are defined as follows:

$$\left.\begin{aligned} z_1 &= -x_{14}^2, \quad & z_4 &= -x_{12}^2 \\ z_2 &= -x_{24}^2, \quad & z_5 &= -x_{23}^2 \\ z_3 &= -x_{34}^2, \quad & z_6 &= -x_{13}^2. \end{aligned}\right\} \tag{11}$$

Equation (10) is the expression we shall take for the 4-point function $I(z; a)$ as a function of the six complex z's and four real a's.

II. 2 The Time-Ordered Product of Four Currents (p-space)

For completeness, we mention that the Fourier transform of the time-ordered product of four currents in perturbation theory gives rise in p-space to exactly the same integral expression (10). The expression for the square-loop Feynman graph is too well-known to warrant a derivation here[19]. Since, as we shall see later, the singularity manifold has a natural geometrical interpretation in terms of such graphs, we shall briefly sketch the necessary notations.

Consider also four scalar fields $\varphi_k^{(0)}(x)$, with characteristic masses m_k, $k = 1, \ldots, 4$. Write $a_k = m_k^2$, and

$$\begin{aligned} j_1(x) &= \varphi_4^{(0)}(x)\, \varphi_1^{(0)}(x) \\ j_2(x) &= \varphi_1^{(0)}(x)\, \varphi_2^{(0)}(x) \\ j_3(x) &= \varphi_2^{(0)}(x)\, \varphi_3^{(0)}(x) \\ j_4(x) &= \varphi_3^{(0)}(x)\, \varphi_4^{(0)}(x). \end{aligned}$$

Then

$$\left.\begin{aligned} F(z) &= <0 \mid T\{j_1(x_1)\, j_2(x_2)\, j_3(x_3)\, j_4(x_4)\} \mid 0> \\ &= \frac{1}{16}\, \Delta_F(x_{12}; a_1)\, \Delta_F(x_{23}; a_2)\, \Delta_F(x_{34}; a_3)\, \Delta_F(x_{41}; a_4) \\ &= \frac{i\pi^2}{(2\pi)^{16}} \iiint dp_{12}\, dp_{23}\, dp_{34} \cdot \exp.\,[i(p_{12}x_{12} + p_{23}x_{13} + p_{34}x_{14})] \\ &\qquad \times H(p_{12}, p_{23}, p_{34}) \end{aligned}\right\} \tag{12}$$

where double indices denote the differences $x_{ij} = x_i - x_j$.

[19] For general expressions of Feynman amplitudes, cf., e. g., J. S. R. Chisholm, Proc. Camb. Soc. **48**, 300 (1952); Y. Nambu, Nuovo Cimento **6**, No. 5, 1064 (1957).

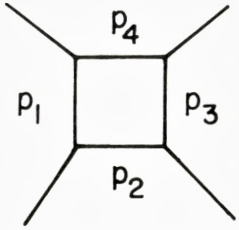

Figure 1. Square-loop Feynman graph

A standard computation then yields:

$$\left.\begin{aligned} &H(p_{12}, p_{23}, p_{34}) \\ &= \int_0^1 \int_0^1 \int_0^1 \int_0^1 \frac{d\alpha_1\, d\alpha_2\, d\alpha_3\, d\alpha_4\, \delta(1-\Sigma\alpha_k)}{(\alpha_1\alpha_2\zeta_1 + \alpha_1\alpha_3\zeta_2 + \alpha_1\alpha_4\zeta_3 + \alpha_2\alpha_3\zeta_4 + \alpha_3\alpha_4\zeta_5 + \alpha_4\alpha_2\zeta_6 - \Sigma\alpha_k a_k)^2};\ a_k > 0 \end{aligned}\right\} \tag{13}$$

where the six ζ's are defined by

$$\left.\begin{aligned} \zeta_1 &= -p_{12}^2, \quad & \zeta_4 &= -p_{23}^2 \\ \zeta_2 &= -p_{13}^2, \quad & \zeta_5 &= -p_{34}^2 \\ \zeta_3 &= -p_{14}^2, \quad & \zeta_6 &= -p_{42}^2. \end{aligned}\right\} \tag{14}$$

We see that expressions (13) and (10) are identical and the definitions for the z's and the ζ's are merely the same six invariants derived from a set of three independent four-vectors.[20]

III. Function of Six Complex Variables Represented by a 4-Fold Feynman Parameter Integral

III. 1 Definition of the Ψ-Manifold

Both the examples treated in Sec. II have led to the same integral expression, namely

$$I(z; a) = \int_0^1 \int_0^1 \int_0^1 \int_0^1 \frac{d\alpha_1\, d\alpha_2\, d\alpha_3\, d\alpha_4\, \delta(1-\Sigma\alpha_k)}{D^2}, \tag{15}$$

[20] Special cases of this square Feynman graph example have been treated independently for all six real variables by R. Karplus, C. M. Sommerfield, and E. H. Wichmann, Phys. Rev. **114**, 376 (1959). This was later extended to the case of two complex variables by J. Tarski, Jour. Math. Phys. **1**, 154 (1960). In both works, all the 3-point boundaries are restricted to the real domain, and all the masses (internal and external) are held fixed together with stability conditions. Subsequently, there appeared a number of papers on the methods of locating the singularities of the general Feynman amplitudes without the explicit completion of integrations. See, e. g., L. D. Landau, Proceedings of the International Conference on High Energy Nuclear

where

$$D \equiv \alpha_1\alpha_2 z_1 + \alpha_1\alpha_3 z_2 + \alpha_1\alpha_4 z_3 + \alpha_2\alpha_3 z_4 + \alpha_3\alpha_4 z_5 + \alpha_4\alpha_2 z_6 - \Sigma\alpha_k a_k . \tag{16}$$

This denominator D can be written in various manners for different purposes. For instance, using the identity $\Sigma\alpha_k = 1$ under the integral, we can write

$$D = \frac{1}{2}\sum_{i,j} \Psi_{ij}\,\alpha_i\alpha_j \tag{17}$$

where the 4×4 symmetric matrix (Ψ_{ij}) is defined as

$$(\Psi_{ij}) = \begin{pmatrix} -2a_1 & z_1 - a_2 - a_1 & z_2 - a_3 - a_1 & z_3 - a_4 - a_1 \\ z_1 - a_1 - a_2 & -2a_2 & z_4 - a_3 - a_2 & z_6 - a_4 - a_2 \\ z_2 - a_1 - a_3 & z_4 - a_2 - a_3 & -2a_3 & z_5 - a_4 - a_3 \\ z_3 - a_1 - a_4 & z_6 - a_2 - a_4 & z_5 - a_3 - a_4 & -2a_4 \end{pmatrix}. \tag{18}$$

The determinant $|\Psi_{ij}|$ will be simply denoted by Ψ throughout this paper, and the manifold $\Psi(z;\,a) = 0$ will be referred to as the Ψ-manifold. It will be shown that the 4-point type singularity of our function $I(z;\,a)$ comes just when this linear transformation (Ψ_{ij}) becomes a singular one (Section IV). The significance and the structure of this Ψ-manifold are given in Sec. III.4.

III. 2 Symmetry of the 4-Point Function

The symmetry of the problem is contained in that of Ψ. Equivalently, we shall define a 3×3 determinant $\Lambda(z)$ (a quantity which will repeatedly appear in our later discussion), as follows:

$$\Lambda(z) = \frac{1}{2}\begin{vmatrix} -2z_1 & z_4 - z_2 - z_1 & z_6 - z_3 - z_1 \\ z_4 - z_1 - z_2 & -2z_2 & z_5 - z_3 - z_2 \\ z_6 - z_1 - z_3 & z_5 - z_2 - z_3 & -2z_3 \end{vmatrix}. \tag{19}$$

$\Lambda(z)$ has the following interpretation[21]: Let $\zeta_1, \zeta_2, \zeta_3$, be a set of three independent 4-vectors, and let the z's and ζ's be related as

Physics, Kiev (1959); J. C. Polkinghorne and G. R. Screaton, Nuovo Cimento **15**, No. 2, 289 (1960); and ibid. **15**, No. 6, 925 (1960). An inherent disadvantage of such approaches is the lack of explicit knowledge of when and only when the cancellation of singularities will not occur.

[21] For real vectors in the Euclidean space, $\Lambda(x)$ has the significance of being proportional to the square of the volume of a tetrahedron. The principal minors of $\Lambda(x)$, which are exactly the type of function $\lambda(x)$ of KW, have the meaning of being proportional to the squares of areas of triangles (cf. remark following Eq. (82)).

$$\left.\begin{aligned} z_1 = -\zeta_1^2, \quad z_2 = -\zeta_2^2, \quad z_3 = -\zeta_3^2 \\ z_4 = -(\zeta_1 - \zeta_2)^2, \quad z_5 = -(\zeta_2 - \zeta_3)^2 \\ z_6 = -(\zeta_3 - \zeta_1)^2. \end{aligned}\right\} \tag{20}$$

Then

$$\Lambda(z) = 4 \times \text{Gram Determinant of } (\zeta_1, \zeta_2, \zeta_3) = 4\,|(\zeta_i \cdot \zeta_j)| \tag{21}$$

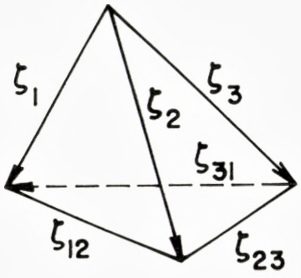

Figure 2. Tetrahedron representation for vectors

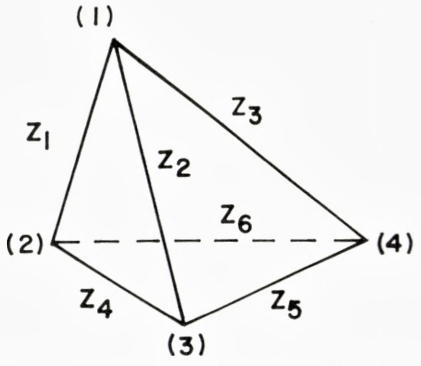

Figure 3. Tetrahedron representation for $\Lambda(z)$

The situation for ζ_i is depicted in Fig. 2 together with their difference vectors. The (Lorentz) squares of the vectors are just the z's given in (20). In Fig. 3, we labelled the six edges of the "tetrahedron" T by these z's. Each of the four faces of T picks out a triplet of z's at a time. Intuitively one would expect each triplet to obey the restriction of the three-point type of KW, and this is indeed the case, as will be shown explicitly later (Sec. IV).

It is clear then that our problem has the symmetry endowed in this tetrahedron, in particular, the permutation symmetry which leaves the set of four faces of T invariant. Let us first divide the edges of T into two classes: Two edges which meet at a vertex of T will be called *adjacent* edges, otherwise *conjugate* edges. Obviously for a tetrahedron, for each edge there are four adjacent edges and only *one* conjugate edge. Thus the six z's break into three pairs of conjugate indices[22]. In our present notation, they are: (1,5), (2,6), and (3,4). For convenience, the four faces of T will be denoted by F_k, $k = 1, \ldots, 4$ and labelled in the following order: (456), (235), (136),

[22] When properly identified, energy and momentum transfer variables are conjugate to each other in this sense. It is important to note that conjugate indices, *ipso facto*, do not appear simultaneously in any one of the 3-point quantities, e. g., $\Phi(z)$ or $\lambda(z)$.

and (124). Note that this is equivalent to labelling the 4 vertices of Fig. 3 in the counter-clockwise order.

Then the operations which transform the set of all F_k into themselves are obviously the permutations among any *two pairs* of the conjugate indices. For example, $(1,5) \leftrightarrow (2,6)$; by this we mean the following:

$$\text{either (i)} \begin{pmatrix} 1 \\ 5 \end{pmatrix} \begin{matrix} \leftrightarrow \\ \leftrightarrow \end{matrix} \begin{pmatrix} 2 \\ 6 \end{pmatrix}: \text{ thus } F_1, F_4 \text{ invariant; } F_2 \leftrightarrow F_3$$

$$\text{or (ii)} \begin{pmatrix} 1 \\ 5 \end{pmatrix} \times \begin{pmatrix} 2 \\ 6 \end{pmatrix}: \text{ thus } F_2, F_3 \text{ invariant; } F_1 \leftrightarrow F_4$$

$$\text{or (iii)} \begin{pmatrix} 1 \\ \updownarrow \\ 5 \end{pmatrix} + \begin{pmatrix} 2 \\ \updownarrow \\ 6 \end{pmatrix}: \text{ thus } F_1 \leftrightarrow F_4; \text{ and } F_2 \leftrightarrow F_3.$$

In other words, a permutation between adjacent edges is to be accompanied by the permutation between their respective conjugate indices (e. g. cases (i) and (ii) above); and a permutation within one pair of conjugate indices is to be accompanied by the permutation within *another* pair of conjugate indices (e. g. case (iii) above). This exhausts the symmetry of the problem.

We might remark that the above symmetry property, which is purely geometrical, is not confined to the perturbation theory. The quantity $\Delta(z)$ (or the Gram determinant of three 4-vectors) will undoubtedly play an important role in the case of the axiomatic approach. In the perturbation example Eq. (15), this symmetry is of course trivially implied by the permutation symmetry between any two $\alpha_i \leftrightarrow \alpha_j$ in the integrand, the net result there being the proper interchange of four z's and two a's, which (apart from the associated permutation among the mass parameters) agrees exactly with our above general prescription of the permutation among two pairs of conjugate indices.

III. 3 The Structure of the 3-Point Φ_k-Manifolds and The 2-Point R_μ-Manifolds

The 3-point Φ-manifold of KW has precisely the same structure as that of $\Delta(z)$ discussed above, except that a set of three z's emerging from one vertex in Fig. 3 is now replaced by a set of three mass parameters. Thus the Φ-determinant is (apart from a trivial factor of 4) just the Gram determinant of three 4-vectors ζ_i with the diagonal elements put on some mass-shells.

In the present 4-point problem, we have in all *four* sets of such Φ, one for each face of the tetrahedron T. Thus, for example, the structure of Φ_1 can be represented by the tetrahedron T_1 in Fig. 4.

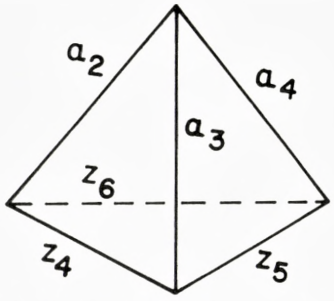

Figure 4. Tetrahedron representation for Φ_1

$$\Phi_1 = \frac{1}{2}\begin{vmatrix} -2a_2 & z_4-a_3-a_2 & z_6-a_4-a_2 \\ z_4-a_2-a_3 & -2a_3 & z_5-a_4-a_3 \\ z_6-a_2-a_4 & z_5-a_3-a_4 & -2a_4 \end{vmatrix}. \tag{22}$$

To every Φ_k-determinant, there are associated four 2×2 subdeterminants. One of them involves pure z's, i. e. the $\lambda(z)$ defined by KW, e. g.:

$$\lambda_1(456) = -\begin{vmatrix} -2z_4 & z_6-z_4-z_5 \\ z_6-z_4-z_5 & -2z_5 \end{vmatrix} \tag{23}$$

which is associated with the face with all z's in Fig. 4. To see how $\lambda(z)$ is related to $\Phi(z)$, we note that (22) can be written as

$$\Phi_1 = \frac{1}{2}\begin{vmatrix} -2a_3 & z_4+a_3-a_2 & z_5+a_3-a_4 \\ z_4+a_3-a_2 & -2z_4 & z_6-z_4-z_5 \\ z_5+a_3-a_4 & z_6-z_4-z_5 & -2z_5 \end{vmatrix} \tag{22a}$$

in which $-\lambda_1$ appeas as the first principal minor of Φ_1 when written in the form (22a). This feature will also appear in the 4-point case (cf. Sec. VI.2).

The other three quantities are the R_k-manifolds defined by KW, e. g.,

$$R_6 = -\begin{vmatrix} -2a_2 & z_6-a_2-a_4 \\ z_6-a_2-a_4 & -2a_4 \end{vmatrix} \tag{24}$$

which are associated with the faces of one z and 2 a's in Fig. 4.

It is well known that the manifold $R_k(z_k) = 0$ yields the cut in each variable z_k in the 3-point case. This feature is also carried over to the 4-point case where we have six such R-manifolds, giving rise to a cut along the positive real axis in each of the six complex variables. Note that each cut is actually an 11-dimensional manifold.

III. 4 The Structure of the 4-Point Ψ-Manifold

The generalization from the 2-point R-manifold to the 3-point Φ-manifold is strongly suggestive as to how the 4-point Ψ-manifold might be built up, and indeed the analogy turns out to be a valid one. As one can build up a tetrahedron T_k for Φ_k by adding three a's to the k-th face taken out from the tetrahedron T for $\Lambda(z)$, one may now build up a "pentahedron"[23] for Ψ by adding four legs of a's to the entire tetrahedron T as the base (Fig. 5). The remaining four hypersurfaces of this pentahedron, being tetrahedrons T_k with 3 a's and 3 z's, represent just the set of 4 Φ_k-manifolds in our problem[24].

The Ψ-determinant has the simple interpretation in the p-space as 16 times the Gram determinant $|(p_i \cdot p_j)|$ of four 4-vectors p_k such that the

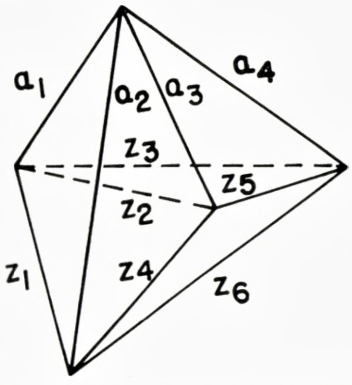

Figure 5. Pentahedron representation for $\Psi(z; a)$

diagonal elements are put on some mass-shells: $-p_k^2 = m_k^2 = a_k$, and the off-diagonal elements are re-expressed through the difference vectors, e. g.,

[23] The above intuitive terms such as "tetrahedron" and "pentahedron" should perhaps be properly changed into "n-simplex," $n = 4,5$ respectively.

[24] By our previous labelling of the four faces of T (Sec. III.2), the index k of Φ_k is such that a_k does *not* appear in Φ_k.

$$2\,p_i p_j = -(p_i - p_j)^2 - p_i^2 - p_j^2 = z_m - a_i - a_j$$

for some m. These four p_k's may just be identified with the four internal momenta of the square-loop Feynman graph (Fig. 1). On account of the momentum conservation at each vertex, these difference vectors $p_{i\,i+1} = p_i - p_{i+1}$ are just the four external momenta and the six z's are then the six invariants built up from these $p_{i\,i+1}$ (cf. Eq. (20)). From this, it is clear that the 4-point singularity manifold $\Psi = 0$ can be interpreted to arise just when the four p_k's are not linearly independent[25].

IV. Sources of Singularities of the 4-Point Function

IV. 1 General Discussion

From the integral representation (15) it is clear that, with a given set of parameters a_k, the singularities of $I(z; a)$ come from a certain manifold of z such that the denominator D vanishes somewhere within the range of integration. Of course, not all such points need be singular points of $I(z)$, as we can easily convince ourselves that the integration may very well smoothe out some of the singularities of the integrand. In fact, from the 3-point example treated in KW, we see that there are some delicate cancellations which made

a) only part of the Φ-manifold as relevant 3-point type singularities; and

b) the relevant portion of the cut (2-point singularity), in the case of non-vanishing masses, actually starts from $z_k = (\sqrt{a_m} + \sqrt{a_n})^2$, but not $(\sqrt{a_m} - \sqrt{a_n})^2$.

It will be shown in Sec. V that an inherent cancellation of this nature will again occur in the 4-point case.

In this section, we shall mainly locate the sources of all possible singularities of $I(z)$. It will be shown explicitly that these singularities arise only when the quadratic roots of D in α_i become double roots. The conditions for such double roots at each stage then yield the singularity manifolds for the 2-point, 3-point, and 4-point type, respectively.

We now briefly compare the methods we shall adopt in the 4-point case

[25] This was independently noted by LANDAU *loc. cit.*, and implicitly implied by KARPLUS, *et. al.*, *loc. cit.*, for the real case.

versus those available for the 3-point case. In the case treated by KW, the corresponding original expression is

$$H(z_1, z_2, z_3) = \int\int\int_0^1 \frac{d\alpha_1\, d\alpha_2\, d\alpha_3\, \delta(1 - \Sigma\alpha_j)}{D_1}, \tag{25}$$

where

$$D_1 = \alpha_1\alpha_2 z_3 + \alpha_2\alpha_3 z_1 + \alpha_3\alpha_1 z_2 - \Sigma\alpha_j a_j. \tag{26}$$

It is obvious that, when one of the 4 α's in the 4-point case becomes zero, D of (16) (apart from a trivial relabelling of the indices) goes over to D_1 of (26).

The 3-fold integration in (25) can be carried out in a straightforward manner, but the result contains a sum of 16 Spence functions[26] which are somewhat inconvenient. Instead, KW applies the differentiation $\Sigma\partial/\partial a_k$, which, on account of the identity $\Sigma\alpha_k = 1$, has the net effect of raising the power of D_1 by one for every such operation. Thus[27]

$$\sum_{k=1}^{3} \frac{\partial}{\partial a_k} H(z; a) = \int\int\int_0^1 \frac{d\alpha_1\, d\alpha_2\, d\alpha_3\, \delta(1 - \Sigma a_j)}{D_1^2}. \tag{27}$$

Now (27) when integrated out contains only logarithms (KW (A. (46))):

$$\sum_{k=1}^{3} \frac{\partial}{\partial a_k} H(z; a) = -\frac{1}{2}\frac{1}{\Phi} \sum_{k} \frac{P_k}{\sqrt{R_k}} \log \frac{z_k - a_m - a_n + \sqrt{R_k}}{z_k - a_m - a_n - \sqrt{R_k}}, \tag{28}$$

where Φ is of the structure of (22), R_k of (24) and $P_k = \partial\Phi/\partial a_k$.

The 4-fold integration (15) can of course be carried out by force, but at first sight one is rather inclined to feel uneasy about a sum of 192 Spence functions. In this respect, it resembles (25). Unfortunately, however, the above differentiation technique will no longer save the situation, and the Spence function terms always persist in any explicit expression for $I_\nu(z)$, where ν refers to the power of D in (15). Since the case $\nu = 2$ is the simplest of all, and there is no merit in going to higher ν, we shall just stay with (15).

At this point, it is instructive to learn the lesson from the 3-point case. A study of the 3-point function $H(z)$ of (25) *in the undifferentiated form*

[26] For a comprehensive treatment of Spence functions, see, e. g., L. LEWIN, *Dilogarithms and Associated Functions*, London (1958).

[27] One might note that, in a 2-dimensional (1-space, 1-time) space, the 3-point function without differentiation actually has the form (27). (This is a remark by Profs. KÄLLÉN and TOLL.)

led us to rederive the same singularity manifolds as those obtained from the differentiated form (28). There are two ways for this, which are essentially equivalent:

(a) The first method is to discuss the singularities under one remaining integral sign. We have, after the completion of a 2-fold integration, the following expression:

$$H(z;\, a) = \int_0^1 \frac{d\alpha}{\sqrt{N(\alpha)}} \log \chi(\alpha), \tag{29}$$

where

$$\left.\begin{aligned} N(\alpha) &= \lambda(z)\,\alpha^2 - 2P_3\,\alpha + R_3 = \lambda(z)\,(\alpha - \varrho_1)\,(\alpha - \varrho_2) \\ \varrho_{1,2} &= \frac{1}{\lambda(z)}\,[P_3 \pm 2\sqrt{z_3\,\Phi(z)}] \end{aligned}\right\} \tag{30}$$

and

$$\chi(\alpha) = \frac{\frac{1}{2}\left(\frac{\partial\lambda}{\partial z_1}\,\alpha - \frac{\partial R_3}{\partial a_2}\right) - \sqrt{N(\alpha)}}{\frac{1}{2}\left(\frac{\partial\lambda}{\partial z_1}\,\alpha - \frac{\partial R_3}{\partial a_2}\right) + \sqrt{N(\alpha)}} \cdot \frac{\frac{1}{2}\left(\frac{\partial\lambda}{\partial z_2}\,\alpha - \frac{\partial R_3}{\partial a_1}\right) - \sqrt{N(\alpha)}}{\frac{1}{2}\left(\frac{\partial\lambda}{\partial z_2}\,\alpha - \frac{\partial R_3}{\partial a_1}\right) + \sqrt{N(\alpha)}}.$$

Thus, as far as the integrand of (29) is concerned, when the 3-point roots $\varrho_{1,2}$ fall into the range (0,1) in the α-plane, $N(\alpha) = 0$ gives an apparent singularity. However, at this point $\log \chi(\alpha)$ becomes $\log 1 = n.\,2\pi i$ (n = integer), in which lies the inherent cancellation. As long as the two roots remain distinct, $H(z)$ can still be defined by analytic continuation into another sheet of the Riemann surface even when one (or both) ϱ_i has (have) actually passed through the open interval (0,1), since in this case one may very well deform the path of integration to avoid meeting with the roots. The upper end $\alpha = 1$ is perfectly harmless. At the lower end $\alpha = 0$, however, one gets the R_3-manifold (which gives the cut in the z_3-plane). On the other hand, when the roots tend to coincide after they have crossed over the range (0,1) an odd number of times, then the above deformation of the integration path is no longer possible, and $H(z)$ will have a singularity. The condition for such double roots gives precisely the manifold $\Phi(z) = 0$ (apart from the trivial alternative $z_3 = 0$ which we disregard). The only other singularities $H(z)$ can have is at the coincident zeros or poles of $\chi(\alpha)$ which can be easily seen to lead to the R_1 and R_2 manifolds (cf. (46) and the remark thereto).

In this way, one is able to relocate the singularities of the 3-point function $H(z)$ in the undifferentiated form, which agrees exactly with what one gets from the explicit differentiated form (28).

(b) The second method is to carry out the last integration of (29). As already mentioned before, one gets Spence function terms besides logarithms here. However, a careful examination of these terms shows that, with proper manipulation, they are still manageable. One first learns which combination of the variables go into each of the Spence functions by explicitly differentiating them with $\Sigma\partial/\partial a_k$. From this, one sees how the Spence functions unfold and all the inherent cancellations thereof. Once this is done, one can, taking into account the symmetry of the problem, again recover the singularities of the 3-point function $H(z)$ in the Spence function form. We did this only as an exercise to get an insight into properly handling the corresponding (and more complicated) Spence function terms in the 4-point case.

In the following, these two approaches are generalized to the 4-point case.

IV.2 The One-Fold Integral Representation

We now proceed to discuss the singularities of $I(z)$ after a straightforward completion of integrations over three of the four α's. We have, before a final integration, the following expression[28]:

$$I(z;\, a) = -\frac{1}{2}\int_0^1 \frac{d\alpha}{A(z)\alpha^2 + \frac{1}{2}Q_1\alpha + \Phi_1} \sum_{j=1}^{3} \frac{M_j(\alpha)}{\sqrt{N_j(\alpha)}} \log \chi_j(\alpha). \tag{31}$$

Here the denominator in front of the summation sign has singled out, in the language of Sec. III.4, the tetrahedron T_1, viz., the set of variables $(z_4, z_5, z_6;\ a_2, a_3, a_4)$. The summation is thus extended over the remaining three T_{j+1}, $j = 1, 2, 3$, of the pentahedron of Fig. 5. (Recall that T_k was defined by deleting a_k from the pentahedron).

Now the symbols in (31) stand for the following:

$$Q_k = \frac{\partial \Psi}{\partial a_k}, \qquad k = 1, \ldots, 4 \tag{32}$$

$$\Phi_k = \frac{1}{2}\Psi^{kk} \tag{33}$$

[28] We have performed the integrations over $\alpha_4, \alpha_3, \alpha_2$, The remaining integration is over α_1, where we drop the subscript. This singles out the triplet (*456*). Of course, by symmetry, the order of integration is entirely immaterial. Had one left the last integration over α_k undone for any k, the net effect would be a trivial permutation of $T_1 \leftrightarrow T_k$ from Eq. (31).

where Ψ^{kk} denotes the k-th principal minor of the Ψ-determinant. Note that Φ_1 is explicitly given in (22). $A(z)$ is given by (19). We have also defined λ_1 in (23). The three remaining similar expressions (one for each of the remaining three triplets) can be simply defined as

$$\lambda_{j+1} = (2A)^{jj}, \quad j = 1, 2, 3. \tag{34}$$

Furthermore, we have in (31)

$$M_j(\alpha) = \frac{\partial A(z)}{\partial z_j}\,\alpha - \frac{\partial \Phi_1}{\partial a_{j+1}}. \tag{35}$$

The quantities $N_j(\alpha)$ and $\chi_j(\alpha)$ are precisely of the same structure as those appearing in the undifferentiated form of the 3-point function $H(z)$ in (29). Here

$$N_j(\alpha) = \lambda_{j+1}\,\alpha^2 - 2\,P_{j'}\,\alpha + R_{j'}, \tag{36}$$

where

$$P_{j'} = \frac{\partial \Phi_{j+1}}{\partial a_1} \tag{37}$$

and the *primed* index j' denotes the conjugate of j in the sense of Sec. III.2, viz., $j = (1, 2, 3)$; $j' = (5, 6, 4)$, respectively.

The quantity R_6 is given explicitly in (24), and the remaining five $R_\mu(z_\mu)$ are obvious from symmetry, as they can readily be read off from the principal 2×2 minors of the Ψ-determinant.

Finally we have:

$$\chi_j(\alpha) = \frac{\dfrac{\partial M_j(\alpha)}{\partial z_{k'}} - \sqrt{N_j(\alpha)}}{\dfrac{\partial M_j(\alpha)}{\partial z_{k'}} + \sqrt{N_j(\alpha)}} \cdot \frac{\dfrac{\partial M_j(\alpha)}{\partial z_{l'}} - \sqrt{N_j(\alpha)}}{\dfrac{\partial M_j(\alpha)}{\partial z_{l'}} + \sqrt{N_j(\alpha)}}, \quad j = 1, 2, 3 \tag{38}$$

in which the indices $(j'kl)$ form a triplet. The identification of indices k and l is unique for each j.

We note in passing that the integrand in (31) evaluated at $\alpha = 0$ is precisely the final expression (28) for the 3-point function in the differentiated form, now for the variables (z_4, z_5, z_6). (This is certainly to be expected, and serves as a check for (31)).

Having thus identified all the quantities that appear in (31), we proceed to note a number of identities which will be important for our subsequent discussions. We have, for $j = 1, 2, 3$,

$$M_j^2(\alpha) = \lambda_1 N_j(\alpha) + 4\,z_{j'}\left[A(z)\,\alpha^2 + \frac{1}{2}\,Q_1\,\alpha + \Phi_1\right]. \tag{39}$$

$$\left.\begin{aligned}\left(\frac{\partial M_j(\alpha)}{\partial z_{k'}}\right)^2 &= \frac{1}{4}\left(\frac{\partial \lambda_{j+1}}{\partial z_l}\,\alpha - \frac{\partial R_{j'}}{\partial a_{l+1}}\right)^2 \\ &= N_j(\alpha) + 4\,z_{j'}\left[z_k\,\alpha^2 + \frac{1}{2}\frac{\partial R_k}{\partial a_1}\,\alpha + a_{k+1}\right].\end{aligned}\right\} \tag{40}$$

$$\left(\frac{\partial \Phi_{j+1}}{\partial a_{k+1}}\,\alpha + \frac{\partial \Phi_{j+1}}{\partial z_l}\right)^2 = R_k N_j(\alpha) + 4\,\Phi_{j+1}\left[z_k\,\alpha^2 + \frac{1}{2}\frac{\partial R_k}{\partial a_1}\,\alpha + a_{k+1}\right], \tag{41}$$

where $(j'\,kl)$ forms a triplet in (40) and (41).

Furthermore, we have

$$\sum_{j=1}^{3} M_j(\alpha) = -\lambda_1 \cdot (1-\alpha) \tag{42}$$

$$Q_k^2 = 16\,\Lambda(z)\,\Phi_k + 4\,\lambda_k \Psi(z;\,a), \qquad k = 1,\ \ldots,\ 4. \tag{43}$$

Note that, for $\alpha = 0$, (39) reads

$$\left(\frac{\partial \Phi_1}{\partial a_{j+1}}\right)^2 = \lambda_1 R_{j'} + 4\,z_{j'}\,\Phi_1 \tag{44}$$

which is just the 3-point relation (KW (A. 46d)), now for the variable (*456*). On the other hand, (41) reads for $\alpha = 0$

$$\left(\frac{\partial \Phi_{j+1}}{\partial z_l}\right)^2 = R_k R_{j'} + 4\,a_{k+1}\Phi_{j+1} \tag{45}$$

which is a variant of (44) in that the role of the corresponding a's and z's is now interchanged. The 4-point analogue of this will be noted in Eq. (110). Equations (43) which are the proper generalization of (44) to the 4-point case will also play a dominant role in our later discussion of the boundary (Sec. VI). It might be of some interest to point out that identities of the types (43) and (44) have a rather natural interpretation in terms of the determinant expansion by means of a theorem due to Jacobi[29]. An illustration of this is given in Appendix D.

For completeness, we might mention that the quadratic expression $(z_k\,\alpha^2 + 1/2\,\partial R_k/\partial a_m\,\alpha + a_n)$ appearing in (40) and (41) is the 2-point analogue of the 3-point quantity $N_j(\alpha)$ defined in (36), or (30). In fact, this is the expression used by KW to discuss the singularity on the cut, viz. (cf. KW (A.47)):

[29] See, Appendix D.

$$\frac{1}{\sqrt{R_k}} \log \frac{z_k - a_m - a_n + \sqrt{R_k}}{z_k - a_m - a_n - \sqrt{R_k}} = - \int_0^1 \frac{d\alpha}{z_k \alpha^2 + \frac{1}{2} \frac{\partial R_k}{\partial a_m} \alpha + a_n} \tag{46}$$

Note that this 2-point denominator is what one gets by multiplying the numerator and the denominator of the individual factor in (38) (cf. (40)). Therefore we see that the zeros or poles of $\chi_j(\alpha)$, which give apparent singularities to the logarithms in the integrand of (31), are really confined to the individual cuts in the z's.

We see from (46), (29), and (31) that in the passage from the 2-point to the 3-point and to the 4-point functions, there is a perfect pattern of generalization, especially in the respective denominators of the integrands before the final stages of integration, viz.:

	Quadratic Form:	*Discriminant*:
2-Point:	$z_k \alpha^2 + \frac{1}{2} \frac{\partial R_k}{\partial a_m} \alpha + a_n$;	R_k: 2×2 Determinant
3-Point:	$\lambda(z) \alpha^2 - 2 \frac{\partial \Phi}{\partial a_j} \alpha + R_j$;	Φ: 3×3 Determinant
4-Point:	$\Lambda(z) \alpha^2 + \frac{1}{2} \frac{\partial \Psi}{\partial a_k} \alpha + \Phi_k$;	Ψ: 4×4 Determinant

(47)

A word about the definition of the branches of $\log \chi_j(\alpha)$ in (31) is now in order. From the original integral representation (15), we note that, where all z's are negative real, $I(z)$ is not only analytic but also positive. Hence we may define the $\log \chi_j(\alpha)$ to lie on its principal sheet for such z's and the rest is done by analytic continuation from there. With this definition, for instance, we will always have on the physical sheet $\log \chi_j(1) = \log 1 = 0$ at the upper limit of integration. Note that $\chi_j(1) \equiv 1$, independent of the z's (cf. (52) below). It should perhaps also be pointed out that for $\alpha \varepsilon (0,1)$, $N_j(\alpha)$ are all positive for all negative real z's. For general z's, the sign of the square root $\sqrt{N_j}$ is rather unimportant since $\log \chi_j$ will just compensate for any change of sign in front of $\sqrt{N_j}$.

IV. 3 The 4-Point Roots

The 4-point roots $r_{1,2}$ are now defined as the zeros of the 4-point quadratic expression in (47), (which is the denominator in (31)). By virtue of (43), we have

$$r_{1,2}(z) = \frac{-Q \pm \sqrt{Q_1^2 - 16\,\Lambda\Phi_1}}{4\,\Lambda(z)} = \frac{-Q \pm 2\sqrt{\lambda_1\Psi}}{4\,\Lambda(z)}. \tag{48}$$

Here we see explicitly that the condition for r to be a double root corresponds to the Ψ-manifold. (The other alternative $\lambda_1(z_4, z_5, z_6) = 0$ is trivial).

Thus, from (39), it follows that

$$\frac{M_j(r_i)}{\sqrt{N_j(r_i)}} = \pm\sqrt{\lambda_1}; \quad \begin{matrix} i = 1, 2 \\ j = 1, 2, 3 \end{matrix} \tag{49}$$

and, together with (42), we have in particular

$$\sum_{j=1}^{3}{}' \sqrt{N_j(r_i)}\Big|_{r_i = 1} = 0. \tag{50}$$

Furthermore, it can be shown that

$$\prod_{j=1}^{3}{}' \chi_j(r_i) = 1, \tag{51}$$

where the summation sign Σ' and the product sign Π' are meant to take care of the sign condition of (49).

Note that

$$\chi_j(1) = 1, \quad j = 1, 2, 3 \tag{52}$$

holds automatically from (37), regardless of the manifold

$$\Lambda(z)\, r_i^2 + \frac{1}{2} Q_1 r_i + \Phi_1 = 0.$$

Finally the special case

$$\prod_{j=1}^{3}{}' \chi_j(0) = 1 \tag{53}$$

now holds on the Φ_1-manifold. This last identity was first established in KW and played an important role in their discussion of the 3-point function in the differentiated form[30].

The identities (49) and (51) are crucial for the 4-point case. Equation (49) says that at the vanishing of the 4-point denominator in (31), all the coefficients of the logarithms become identical, which allows the three log terms to be summed. Eq. (51) guarantees that they add up to log 1.

[30] Cf. KW (A. 50). There the factor $\frac{1}{\sqrt{\lambda(z)}}$ should read $\sqrt{\lambda(z)}$.

Therefore, just as in the 3-point case, the change of the branches of this final logarithm will determine the relevance of the 4-point singularity. We shall leave this problem to Sec. V and Appendix A.

With the above preliminary, the integral (31) can now be written as

$$\left.\begin{aligned} I(z;a) &= \frac{-1}{2A(z)\cdot(r_1-r_2)}\int_0^1 d\alpha \sum_{j=1}^{3}\left(\frac{M_j(r_1)}{\alpha-r_1}-\frac{M_j(r_2)}{\alpha-r_2}\right)\frac{1}{\sqrt{N_j(\alpha)}}\log\chi_j(\alpha) \\ &= -\frac{1}{2A(z)\cdot(r_1-r_2)}\sum_{j=1}^{3}[F_j(r_1)-F_j(r_2)]; \end{aligned}\right\} \quad (54)$$

where

$$F_j(r_i) = M_j(r_i)\cdot\int_0^1 \frac{d\alpha}{\alpha-r_i}\cdot\frac{1}{\sqrt{N_j(\alpha)}}\log\chi_j(\alpha). \qquad \begin{matrix} i=1,2 \\ j=1,2,3 \end{matrix} \quad (55)$$

The situation in the α-plane is quite clear. Namely, one has only to watch out for the three sets of roots (i. e. the 2-point, 3-point, and 4-point) of the expressions (47) versus the path of integration (0,1). Equation (54) explicitly shows that singularities of the 4-point type occur when the 4-point roots r_i become a double root *and* when there is no cancellation among the F's. We now discuss separately the two cases $r_1 \neq r_2$ and $r_1 = r_2$.

IV. 4 The 2-Point and 3-Point Singularities in the 4-Point Function

We first discuss the case when the 4-point roots are distinct: $r_1 \neq r_2$ in (54). Obviously any singularity must then come from each $F_j(r_i)$ and furthermore these singularities may still be subject to cancellation when the summation over j is carried out. The functions $F_j(r_i)$ defined in (55) are evidently multi-valued. When explicitly evaluated, they involve logarithms and a sum of 32 Spence functions for each $i = 1,2$ and $j = 1,2,3$. It is clear that the 4-point complication for each $F_j(r_i)$, as compared with the 3-point function $H(z)$ in the undifferentiated form (29), arises from the presence of the extra factor $(\alpha - r_i)^{-1}$ in (55), which at first sight may cause an apparent singularity for the integrand when r_i passes through the range (0,1). However, this is actually not a relevant source of singularity as long as $r_1 \neq r_2$, and $r_i \neq 0$, or 1, since in this case the path of integration can be easily deformed. Stated otherwise, on account of the identities (49) and (51), $\sum_{j=1}^{3} F_j(r_i)$ can still be defined by analytic continuation to a different

sheet of the Riemann surface whenever a single root r_i crosses over the open interval (0,1). As already remarked above (following (47) and (52)), the upper limit of integration is entirely harmless. On the other hand, $r_i = 0$ implies the Φ_1-manifold, which is exactly the 3-point singularity corresponding to the tetrahedron T_1. The other three manifolds are $\Phi_{j+1} = 0$ which arise from the set of 3-point denominators $N_j(\alpha)$ in (55). This is evidently clear from our discussion of the 3-point function in the undifferentiated form (29). The remaining singularities in $F_j(r_i)$ in (55) then come from

a) when the 3-point roots take on the lower limit 0: giving the manifolds $R_{j'} = 0$ for each j. This results in one cut each for (z_5, z_4, z_6); and

b) when the 2-point roots (i. e. the zeros and poles of $\chi_j(\alpha)$) become double roots within the open interval (0,1). These 2-point roots result in the manifolds $R_m(z_m) = 0$ for $m = 1,2,3$ and can take on the value 0 only when the appropriate masses are zero.

We thus conclude that, for the case $r_1 \neq r_2$, the singularities of our 4-point function $I(z)$ of (54) are the degenerate ones of the 3-point and the 2-point types.

The above statement can also be explicitly verified by completing the last integration of (55) and then discussing the resulting expression. This is done in Appendix A.

We might mention that, for the case $r_1 \neq r_2$, there exists yet another way of looking at the singularities of $\sum_j F_j(r_i)$. Consider now the expression

$$\left.\begin{aligned} J(z;\alpha) &= \sum_{k=1}^{4} \frac{\partial}{\partial a_k} [-2\Delta \cdot (r_1 - r_2) \cdot I(z;\alpha)] \\ &= \sum_{k=1}^{4} \frac{\partial}{\partial a_k} \sum_j [F_j(r_1) - F_j(r_2)] . \end{aligned}\right\} \tag{56}$$

As far as the singularities in z's are concerned, $J(z)$ will for all practical purposes yield as much information as $I(z)$, as long as we are away from the Ψ-manifold. Now the right-hand side of (56) is free from Spence functions; and one can readily see, after a straightforward computation, that one gets singularities of the 3-point and the 2-point type.[31]

[31] These details are contained in the Appendix B of the author's University of Maryland, Department of Physics Technical Report No. 186 (unpublished).

IV. 5 The 4-Point Singularity

Now we come to the case when the 4-point roots become coincident: $r_1 = r_2$, or we are on the Ψ-manifold. From (54), it is clear that one gets a 4-point singularity on the Ψ-manifold *unless* there is a cancellation among the $\sum_j F_j(r_i)$. For this we may divide the Ψ-manifold into Ψ^R and Ψ^{IR}, where the superscripts R and IR denote respectively the *relevant* (no cancellation) and the *irrelevant* (no jump) portions of the Ψ-manifold. It is easy to convince oneself that Ψ^{IR} is actually non-empty. Obvious examples are the cases when all $\boldsymbol{Im}\, z_\mu$ have the same sign, or when all z_μ are negative real, since in both cases we know from the original integral representation (15) that $I(z)$ is analytic there.

The relevance criteria for the Ψ-manifold are treated in Sec. V.

IV. 6 Summary of the Singularity Manifolds

In this section, we see that the 4-point function $I(z)$ admits the following types of singularities:

(a) 4-Point Singularity: on the manifold $\Psi(z; a) = 0$;

(b) 3-Point Singularity: on the manifolds $\Phi_k = 0$, $k = 1, \ldots, 4$;

(c) 2-Point Singularity: on the manifolds $R_\mu = 0$, $\mu = 1, \ldots, 6$.

In terms of the determinants, the Φ_k's and the R_μ's are just the appropriate principal minors of the Ψ-determinant (cf. Sec. III).

V. The Relevance Criteria for the 4-Point Singularity Manifold

We have seen in Sec. IV that the 4-point singularity arises when the roots r_i defined by (48) become coincident. Now we want to examine the behavior[32] of these merging roots more closely in connection with the question of distinguishing Ψ^R from Ψ^{IR}.

[32] In fact, the following technique was first applied to the 3-point case in the undifferentiated form (29) where one is able to re-derive the criteria for the change of relevance of the Φ-manifold. An explicit illustration of this is contained in the Appendix C of the reference cited in footnote 31.

To be specific, let us consider, for the sake of convenience, $z_1, \ldots, z_5$ as being fixed, the roots $r_{1,2}$ as functions of z_6 alone. Suppose we make an arbitrary path $\widehat{ab}$ in the z_6-plane, which connects a point a in the known analyticity region (such a point can always be chosen; e. g., at $-\infty$) to a point b lying on the Ψ-manifold (Fig. 6). Under the mappings $z_6 \to r_i(z_6)$, $i = 1, 2$, this path $\widehat{ab}$ is now mapped into, say $\widehat{A_1B}$ and $\widehat{A_2B}$, respectively, in the α-plane. Then there are the following possibilities:

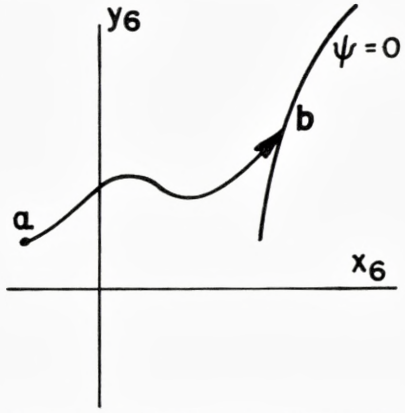

Figure 6. Path of continuation to Ψ-manifold.

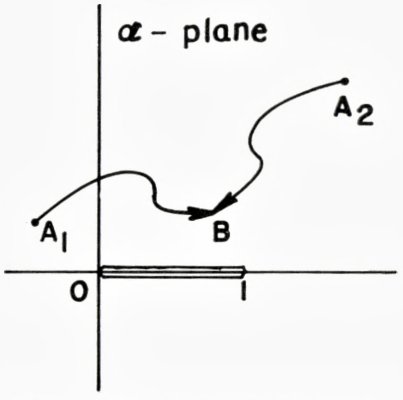

Figure 7. Behavior of the 4-point roots: Irrelevant merging.

(i) Neither of the paths $\widehat{A_iB}$ crosses the interval (0,1), e. g., Fig. 7;

(ii) One of the paths crosses over the interval (0,1) once, e. g., Fig. 8;

or if more than one crossing is made, then either

(i′) the net crossing is even and without encircling the endpoints; or

(ii′) the net crossing is odd, or with encircling of the end points.

Situation (i) or (i′) is obviously harmless. For such cases, (the path of integration can be easily deformed for the case (i′)), the function $\sum_j F_j(r)$ has no jump, hence there will be a cancellation in (54); the singularity at $r_1(b) = r_2(b)$ is thus removed, and one says that the portion of the Ψ-manifold, to which the point b belongs, must lie in Ψ^{IR}. On the other hand, for the situation (ii) or (ii′), the function does have a jump, and hence no cancellation. One gets then an actual singularity at the point b, and the portion of the Ψ-manifold to which b belongs will lie in Ψ^R.

The technique thus described, of plotting the explicit behaviors of the merging roots r_i in the α-plane versus the path of continuation in the

z-space from the known analyticity region to the part of the Ψ-manifold whose relevance is to be determined, although most primitive and tedious, is a rather useful and practical procedure to really pin down the relevance question. Except in some very special cases it is not necessary to plot these merging roots, as one can instead rely on more general criteria. Since we do not expect the whole Ψ-manifold to be relevant, the relevance of this must change when it intersects with some other manifolds. In the following,

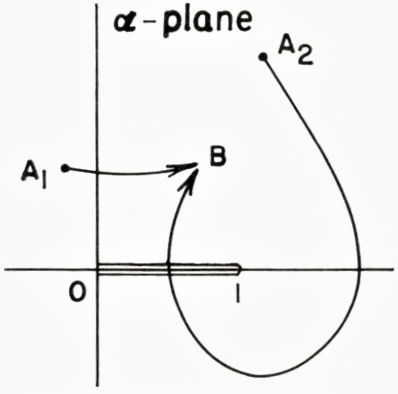

Figure 8. Behavior of the 4-point roots: Relevant merging.

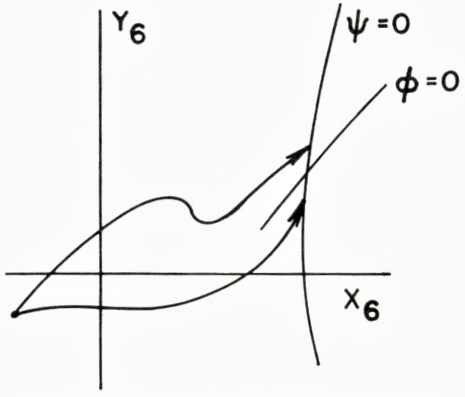

Figure 9. Path of continuation in the neighborhood of $(\Psi = 0) \cap (\Phi = 0)$.

we shall show that these other surfaces are just the relevant portions of the Φ_k-manifolds of the 3-point type.

We shall first state the relevance criteria for the 3-point singularity manifold $\Phi = 0$:

Lemma 1 (KW): The 3-point singularity manifold $\Phi = 0$ changes its relevance at its intersections with the relevant portions of the 2-point singularity manifolds $R_i = 0$, $i = 1, 2, 3$.

This statement is evident from the explicit form (28)[33].

We can now state in perfect analogy:

Lemma 2: The 4-point singularity manifold $\Psi = 0$ changes its relevance at its intersections with the relevant portions of the 3-point singularity manifolds $\Phi_k = 0$, $k = 1, \ldots, 4$.

[33] Actually in KW, the problem of choosing the relevant portion of the Φ-manifold is quite easy. Since one knows enough from the permuted domain D_3 where one must have analyticity, an explicit knowledge of the branches of the logarithms is not mandatory. A more transparent way of seeing this independently is by discussing the behavior of the 3-point roots. This is given in Appendix C of the reference cited in footnote 31.

It suffices to show this for $k = 1$, as the others will obviously follow from symmetry. There are two ways to see this:

(a) One observes that one of the two (4-point) roots, say r_1, goes through the end-point zero of the interval (0,1) in the α-plane when the z's cross the manifold $\Phi_1 = 0$ (cf. (48)), while the other root (r_2) does not and will essentially remain unchanged. Thus when the two roots tend to merge, in one case (i. e., corresponding to one side of the Φ_1-manifold), the paths of

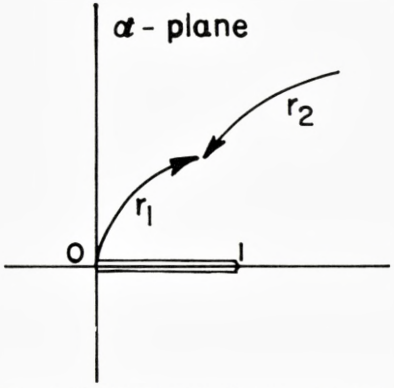

Figure 10. Merging of the 4-point roots: On one side of the Φ_1-manifold.

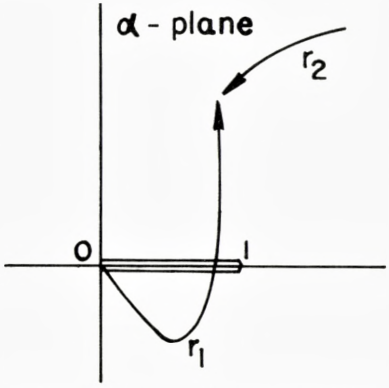

Figure 11. Merging of the 4-point roots: On the other side of the Φ_1-manifold.

the roots do not cross the cut (0,1) (Fig. 10); while for the other case (i. e. corresponding to the other side of the Φ_1-manifold), one of the roots (r_1) does cross over the cut (0,1) once (Fig. 11). Thus $\sum_j F_j(r_1)$ crosses over to a different sheet of the Riemann surface, while $\sum_j F_j(r_2)$ remains on the original sheet. Therefore, there is a cancellation on one side of the Φ_1-manifold, but not on the other side. Thus one concludes that the transition between Ψ^R and Ψ^{IR} takes place at the intersection with the Φ-manifolds.

(b) Another way to see this is by examining the explicit expression for $\sum_j F_j(r_i)$. The details are included in the Appendix A. We simply state that the results there confirm the above simple argument.

We conclude this section with a few remarks:

(1) It is clear that, since the whole Ψ-manifold cannot be all relevant, Ψ^R is non-empty only if the Ψ-manifold has an intersection with Φ_k^R (the relevant portion of the Φ_k-manifolds). Furthermore, as we shall see in Sec. VI that the singularity domain of the 4-point proper is actually compact, Ψ^R is non-trivial only if the Ψ-manifold intersects twice with the Φ_k^R. The con-

dition that such intersections occur is extremely complicated, and we shall only state later (cf. Sec. VI) some necessary conditions.

(2) The pattern of generalization from Lemma 1 to Lemma 2 strongly suggests that this feature may perhaps very well be valid for the general n-point domain in the perturbation theory. However, we do not attempt to prove (or disprove) this conjecture, since this lies outside the scope of the present investigation.

(3) Since $\partial D_3^{\text{pert}}$ is actually part of $\partial E(D_3)$, Lemma 1 is likewise valid in the axiomatic approach. In the 4-point case, from the preliminary results[34] for the $\partial D_4^{\text{prim}}$, the general spirit of Lemma 2 (i. e. deleting Ψ in leaving the question open as to whether this Ψ-manifold has any relation with the $\partial D_4^{\text{prim}}$ (cf. remark in Sec. VI.7)) seems also to be valid in the axiomatic approach.

VI. Determination of the Boundary of the 4-Point Domain

VI. 1 General Discussion

In the preceding two sections, we have shown that the 4-point singularities, subject to the relevance conditions, are confined to the manifold given by the vanishing of the following 4×4 determinant:

$$\Psi(z;a)=\left.\begin{vmatrix} -2a_1 & z_1-a_2-a_1 & z_2-a_3-a_1 & z_3-a_4-a_1 \\ z_1-a_1-a_2 & -2a_2 & z_4-a_3-a_2 & z_6-a_4-a_2 \\ z_2-a_1-a_3 & z_4-a_2-a_3 & -2a_3 & z_5-a_4-a_3 \\ z_3-a_1-a_4 & z_6-a_2-a_4 & z_5-a_3-a_4 & -2a_4 \end{vmatrix}\right\rangle ;\ a_k>0\,. \tag{57}$$

In this section, we wish to determine what constitutes the boundary surfaces for this singularity domain. As it stands, $\Psi(z_\mu;a_k)$ generates a 4-parameter family of surfaces in the space of six complex variables. In principle, the boundaries of such a family of surfaces could be made up from any of the following multitude of possibilities:

(1) The geometric envelope of this 4-parameter family of surfaces, which would correspond to a special path traversed by the a's in the *sedecimant* $a_k>0$. This will be called, for convenience, the 4-*mass envelope* and will be denoted by E_{1234}.

[34] Private communication from Prof. KÄLLÉN.

(2) Subcases of (1) when *one* of the 4 a's takes on the extreme value of 0, or ∞, and the other 3 a's, taking a path in the subspace of the octant $a_j > 0$, produce a 3-*mass envelope* E_{ijk}. In principle, there could be 8 such envelopes.

(3) Still further subcases of (1) are when two of the 4 a's take on the extreme values of 0, or ∞, and the remaining two a's, taking a path in the quadrant $a_i > 0$, produce a 2-*mass envelope* E_{ik}. There could be 24 such envelopes.

(4) Finally, we have the simplest of all cases when 3 of the 4 a's take on the extreme values of 0, or ∞, leaving the remaining one single a_s to vary along the semi-axis $a_s > 0$. In all, there could be 32 such 1-*mass surfaces* E_k.

Out of all these 65 possible candidates for the boundaries to the 4-point domain D_4^{pert}, our present task is to eliminate the ineligible ones. Fortunately, we can eliminate all cases in (2)–(4) which involve any a_k to be ∞. We recall that the a's have the physical meaning of the squares of the masses associated with the internal lines in the Feynman diagrams. Now if any a_k is arbitrary large, then the thresholds for virtual production processes which correspond to the onsets of the associated cut-planes will be proportionally high. Since the 3-point boundary F'_{kl} curves will not be relevant unless they have crosses over the cut beyond the threshold (Lemma 1 of Sec. V), and furthermore, since the relevance of the 4-point boundaries depends on whether or not they have intersected the relevant 3-point curves (Lemma 2), it is clear that the ∞-portion of any a_k would not give rise to any relevant singularity. This statement is also valid in the 3-point case, if we note that all the relevant portions of the F'_{kl} curves are actually confined to the lower ends of the a_m-ranges (from $a_m = 0$ up to a finite value).

This criterion has the further consequence that the singularity domain of the 4-point proper is actually compact. Unlike the 3-point case when the F'_{12} curve extends to ∞ in the z_3-plane at the $a_3 = 0$ end, the all $a_k = 0$ end is always finite in the 4-point case (apart from the trivial case when one of the z's stays zero) (cf. Eq. (120) below).

Thus we shall from now on consider in cases (2)–(4) those extreme values of a's to be zero only. In this way, the list of candidates for boundary is now radically trimmed from 65 down to 15, viz.,

(1) 1 E_{1234}
(2) 4 E_{ijk}
(3) 6 E_{ik}
(4) 4 E_k.

In the following, we shall first examine the questions of the various envelopes listed above. *A priori*, the question is two-fold:

(a) whether such envelopes can exist at all in the allowed all-positive ranges of the a's and

(b) if they do exist under certain circumstances, then it still remains to be seen whether they are really part of the boundary of our domain.

It would perhaps be helpful to recall the corresponding situations for the Φ-manifold in the 3-point case. KW have shown that the boundaries there are made of only the 1-mass curves (analogue of case (4) above). The envelope problem for the Φ-manifold is a much simpler one than we shall encounter below. We give a concise treatment for this in Appendix B. The result there can be simply summarized as follows: *Envelopes for the 3-point Φ-manifold can exist, but they do not lie off the R-manifolds*[35] (i. e. on the cut for each z). One concludes then that the boundaries are made up by the F'_{kl}, which are simple analytic hypersurfaces.

Our results in the following subsections will show that, unlike the 3-point case, the envelopes in the 4-point case are non-trivial[36] and in general the boundary of our domain will be made of pieces of (2-mass) envelopes. Thus we have here a fundamental difference between the 4-point domain and the 3-point domain, namely, the regularity domain of the 4-point function in perturbation theory, D_4^{pert}, is in general *not* everywhere bounded by analytic hypersurfaces.

Before we go into the details for each of the above cases, we shall formulate the envelope condition as follows:

The existence of the envelopes is purely a property that is related to the algebraic structure of the manifold. Consider in general the expression for an m-parameter family of surfaces, $f(z_i; a_k) = 0$, $i = 1, \ldots, n$; $k = 1, \ldots, m$, where the a's are the parameters under consideration, which are allowed to vary over a *real* domain $\boldsymbol{A}_m$.

Definition: A point on f is said to lie on the m-envelope of f if, together with $f = 0$, the set of $(m-1)$ independent equations

$$\boldsymbol{Im}\begin{pmatrix} \dfrac{\partial f}{\partial a_j} \\ \dfrac{\partial f}{\partial a_k} \end{pmatrix} = 0\,, \qquad j, k = 1\,, \ldots, m \tag{58}$$

admits a set of solutions $\{a_k^*\}$ such that $\{a_k^*\} \varepsilon \boldsymbol{A}_m$.

[35] In this connection, it is very tempting to conjecture that the envelopes for the Ψ-manifold would not lie off the Φ-manifolds, but this conjecture turns out to be not true.

[36] In the sense that in general they do not lie on the Φ-manifolds. However the 4-mass and the 3-mass envelopes do not contribute to the boundary (cf. Sec. VI.2 and Sec. VI.3).

Stated in another way, the $(m-1)$ independent equations (58) can be regarded as the $(m-1)$ constraints on the m-parameters, so that in principle one can always express all the other $(m-1)$ parameters a_s, $s>1$, as functions of the remaining parameter, say a_1. Let

$$\tilde{\boldsymbol{A}}_m = \{a\colon a_1 \varepsilon \boldsymbol{A}_m;\ a_s = a_s(a_1)\}$$

which shall be referred to as the "path" for the m-envelope. Note that in general $\tilde{\boldsymbol{A}}_m$ will not be completely contained in $\boldsymbol{A}_m$. If, regardless of the configuration of the z's,

$$\tilde{\boldsymbol{A}}_m \cap \boldsymbol{A}_m = 0,$$

then it is clear that the m-envelope in question does not exist at all. Otherwise, for $\tilde{\boldsymbol{A}}_m \cap \boldsymbol{A}_m \neq 0$, we will be able to find in the a-space (i. e. $\boldsymbol{A}_m$) an allowed path $\tilde{\boldsymbol{A}}_m \cap \boldsymbol{A}_m$ such that the image of this under the mapping

$$z_n = g(z_1, \ldots, z_{n-1};\ a_k)\colon\ f = 0$$

gives the desired m-envelope. Since the a's mix the real and imaginary parts of the z's, the envelopes will evidently in general *not* be analytic hypersurfaces. Equations (58) will be referred to as the envelope conditions.

VI. 2 The 4-Mass Envelope

We now proceed to apply the general equations (58) to our specific manifold $\det|\Psi_{ij}| \equiv \Psi(z;a) = 0$ of (57). Before we do this, we shall derive a number of identities which will be crucial for the subsequent discussion of the envelopes. First, we find it useful to rewrite the Ψ-determinant such that the a's shall appear only in one column and one row (cf. (22a)). For instance, we have from (57):

$$\Psi(z;a) = \begin{vmatrix} -2a_1 & z_1+a_1-a_2 & z_2+a_1-a_3 & z_3+a_1-a_4 \\ z_1+a_1-a_2 & -2z_1 & z_4-z_2-z_1 & z_6-z_3-z_1 \\ z_2+a_1-a_3 & z_4-z_1-z_2 & -2z_2 & z_5-z_3-z_2 \\ z_3+a_1-a_4 & z_6-z_1-z_3 & z_5-z_2-z_3 & -2z_3 \end{vmatrix} \equiv |\tilde{\Psi}_{ij}|. \quad (57\,a)$$

Here a_1 is singled out. Evidently there are 3 other such forms obtained by suitable permutations. For convenience, let us denote by $\tilde{\Psi}_{ij}$ the ij-th element and by $\tilde{\Psi}^{ij}$ its minor in (57a), while the corresponding uncurled quantities shall refer to those in the original from (57). Note that

$$\left.\begin{aligned} \tilde{\Psi}_{1k} &= \Psi_{1k} + 2\,a_1; \quad k \neq 1 \\ \tilde{\Psi}_{11} &= \Psi_{11}. \end{aligned}\right\} \tag{59}$$

We have already had occasion in Sec. IV to define such quantities as $Q_k = \dfrac{\partial \Psi}{\partial a_k}$ of (32). Now these have the most natural interpretation in terms of (57a), namely:
For $k = 2,3,4$

$$\tilde{\Psi}^{1k} = (-1)^k \cdot \frac{1}{2} \frac{\partial \Psi}{\partial a_k} = (-1)^k \cdot \frac{1}{2} Q_k; \quad k \neq 1, \tag{60}$$

$$\tilde{\Psi}^{kk} = 2\Phi_k = \Psi^{kk}; \quad k \neq 1, \tag{61}$$

$$\tilde{\Psi}^{11} = 2\,\Lambda(z). \tag{62}$$

Furthermore, from (57a), one immediately sees that

$$\sum_{i=1}^{4} Q_i = \sum_{i=1}^{4} \frac{\partial \Psi}{\partial a_i} = \left(\sum_{i=1}^{4} \frac{\partial}{\partial a_i} \tilde{\Psi}_{11} \right) \cdot \tilde{\Psi}_{11}$$

since

$$\sum_{i=1}^{4} \frac{\partial}{\partial a_i} \tilde{\Psi}^{1k} = 0, \quad \text{for} \quad k \neq 1.$$

Therefore

$$\sum_{i=1}^{4} Q_i = -4\,\Lambda(z), \tag{63}$$

in which the right-hand side is independent of the a's. Next, with the aid of (60), we have

$$\begin{aligned} \Psi &= \sum_{i=1}^{4} (-1)^{i+1} \tilde{\Psi}_{1i} \tilde{\Psi}^{1i} = \tilde{\Psi}_{11} \tilde{\Psi}^{11} + \sum_{k+1} (-1)^{k+1} \tilde{\Psi}_{1k} \tilde{\Psi}^{1k} \\ &= -\frac{1}{2} \Psi_{11} \sum_{i=1}^{4} Q_i + \frac{1}{2} \sum_{k+1} (-1)^{2k+1} \tilde{\Psi}_{1k} Q_k. \end{aligned}$$

Using (59), we get

$$\Psi = -\frac{1}{2} \sum_{j=1}^{4} \Psi_{ij} \cdot \frac{\partial \Psi}{\partial a_j} \tag{64}$$

for all $i = 1, \ldots, 4$.
Identities (63) and (64) will be of great importance to us in the following discussions. Another set of identities which we will need here already ap-

peared in (43)[37]. We are now ready to write down the envelope conditions for E_{1234} according to (58):

$$\boldsymbol{Im}\begin{pmatrix}\dfrac{\partial\Psi}{\partial a_j}\\[2ex]\dfrac{\partial\Psi}{\partial a_k}\end{pmatrix} = \boldsymbol{Im}\left(\frac{Q_j}{Q_k}\right) = 0\,; \qquad \begin{matrix} j,\ k = 1,\ \ldots,\ 4 \\ j \neq k \text{ (otherwise trivial).}\end{matrix} \tag{65}$$

In view of the identity (63), we can now define a set of four *real* numbers γ_k such that on E_{1234}

$$Q_k = -\,4\,\gamma_k \cdot \Lambda(z) \tag{66}$$

with

$$\sum_k \gamma_k = 1\,; \qquad \boldsymbol{Im}\,\gamma_k = 0\,. \tag{67}$$

From (43) we have on the manifold $\Psi = 0$:

$$Q_k = \pm\,4\sqrt{\Lambda\Phi_k}\,. \tag{68}$$

Therefore we have on E_{1234}

$$\frac{\Phi_j}{\Phi_k} = \frac{\gamma_j^2}{\gamma_k^2} > 0\,; \qquad j,\ k = 1,\ \ldots,\ 4. \tag{69}$$

In principle, the system of equations (69) together with $\Psi = 0$ contain all the information there is about the 4-mass envelope. (In fact, as we shall see later in Sec. VI.4 for a 2-mass envelope, one has only one such equation which actually exhausts the envelope condition). However, a frontal attack on (69) for both the 4-mass and the 3-mass envelopes could lead to tremendous algebraic complications. We find it much more convenient to go back to the system of equations (64). We have for $\Psi = 0$:

$$\sum_{k=1}^{4} \Psi_{ik} Q_k = 0\,, \qquad i = 1,\ \ldots,\ 4. \tag{70}$$

Now with (66), we get ($\Lambda(z) \neq 0$)

$$\sum_{k=1}^{4} \Psi_{ik}\,\gamma_k = 0\,, \qquad \sum_k \gamma_k = 1\,. \tag{71}$$

We emphasize that the γ_k's are real, so that the system of equations (71) is equivalent to the following set of 9 real linear algebraic equations[38]

[37] For a proof of such identities, see Appendix D.

[38] The fact that $\Psi = 0$ is automatically satisfied is obvious from (72).

$$\sum_{k=1}^{4} (\boldsymbol{Im}\, \Psi_{ik})\, \gamma_k = 0, \qquad i = 1, \ldots, 4, \tag{72a}$$

$$\sum_{k=1}^{4} (\boldsymbol{Re}\, \Psi_{ik})\, \gamma_k = 0, \qquad i = 1, \ldots, 4, \tag{72b}$$

$$\sum_{k=1}^{4} \gamma_k = 1. \tag{72c}$$

Since, according to (72c), the solution with all γ's being equal to zero is unacceptable, it follows that the determinants of the coefficients of any four equations taken at a time (out of the eight in (72a-b)) should vanish. We shall first discuss the consequences of (72a) which contain the most powerful restrictions on E_{1234}:

$$\det |\, \boldsymbol{Im}\, \Psi_{ij} \,| \equiv \begin{vmatrix} 0 & y_1 & y_2 & y_3 \\ y_1 & 0 & y_4 & y_6 \\ y_2 & y_4 & 0 & y_5 \\ y_3 & y_6 & y_5 & 0 \end{vmatrix} = 0. \tag{73}$$

Note that this determinant is equivalent to the λ-function of products of conjugate variables, viz:

$$\lambda(y_k\, y_{k'}) \equiv - \begin{vmatrix} -2\, y_1\, y_5 & y_2\, y_6 - y_1\, y_5 - y_3\, y_4 \\ y_2\, y_6 - y_1\, y_5 - y_3\, y_4 & -2\, y_3\, y_4 \end{vmatrix} = 0. \tag{73a}$$

From now on, we shall be more specific by keeping the other 5 z's fixed, and project everything into the z_6-plane. We see that the 4-mass envelope E_{1234} can only be satisfied on the two horizontal straight lines obtained by solving (73), viz:

$$y_6 = \frac{y_1\, y_5 + y_3\, y_4 \pm 2\sqrt{y_1\, y_3\, y_4\, y_5}}{y_2} \tag{74}$$

or

$$\pm \sqrt{y_2\, y_6} = \sqrt{y_1\, y_5} \pm \sqrt{y_3\, y_4}. \tag{74a}$$

From (74), it follows immediately that *there exists no 4-mass envelope whenever*

$$y_1\, y_3\, y_4\, y_5 < 0. \tag{75}$$

More generally, in view of (74a), we can state that *the necessary condition for the existence of the 4-mass envelope* E_{1234} *is that the three products of* $y_k\, y_{k'}$, $k = 1, 2, 3$ (k' = conjugate of k, cf. Sec. III.2) *must have the same sign*, or

$$\frac{y_1 y_5}{|\,y_1 y_5\,|} = \frac{y_2 y_6}{|\,y_2 y_6\,|} = \frac{y_3 y_4}{|\,y_3 y_4\,|} \tag{76}$$

which we shall refer to as the *sign convention* for the existence of the 4-mass envelope $E^{\pm}_{1234}$.

An obvious example which satisfies this sign condition but where $E^{\pm}_{1234}$ is entirely irrelevant is furnished by the configuration whenever 5 z's lie in the same half-plane. Then the $E^{\pm}_{1234}$ in the 6-th variable must also lie in this same half-plane. As we have already mentioned in Sec. V, the original function (15) has no singularity for all 6 z's having the same sign in the $\boldsymbol{Im}\, z$'s. Here we have a situation where the entire lines are irrelevant. For the other configurations[39], however, the situations are much more complicated, as we shall see below.

So far, we have only explored the existence condition of E_{1234} based on the consequence of the imaginary part equations (72a). A brief examination of the real part equations (72b) will convince oneself that there is no algebraic contradiction among the two sets of equations, so that in principle E_{1234}, satisfying (76), can exist provided that all the parameters a_k could be found to be positive at least for some configurations of the x's. This we now proceed to show.

To be specific, let the y's be given, satisfying (76); one can explicitly compute the γ_k's from (72a) and (72c)[40] (cf. Appendix C) in terms of the y's. Equations (72b) may now be regarded as those governing the a_k's. The solutions may be written as follows:

$$a_k = -\frac{1}{2}\sum_{i,j} X_{ij}\,\gamma_i\,\gamma_j + \sum_j X_{kj}\,\gamma_j\,; \qquad k, i, j = 1, \ldots, 4 \tag{77}$$

with

$$\sum_j \gamma_j = 1\,,$$

where the matrix X is given by

$$X = \begin{pmatrix} 0 & x_1 & x_2 & x_3 \\ x_1 & 0 & x_4 & x_6 \\ x_2 & x_4 & 0 & x_5 \\ x_3 & x_6 & x_5 & 0 \end{pmatrix}. \tag{78}$$

[39] The distinct configurations for which (76) is satisfied are given in Appendix C.

[40] Note that the following ratios hold on E_{1234}:
$\gamma_1^2 : \gamma_2^2 : \gamma_3^2 : \gamma_4^2 = y_4 y_5 y_6 : y_2 y_3 y_5 : y_1 y_3 y_6 : y_1 y_2 y_4$, the right-hand side can be regarded as Φ_k evaluated at all $x = 0$ and all $a = 0$ (cf. Appendix C).

Note that

$$X_{ij} = \mathbf{Re}\, \Psi_{ij}\big|_{a_k = 0}.$$

It is clear then that the configurations of the x's must be such that $\bigcap_k \{a_k > 0\} \neq 0$, or

$$-\frac{1}{2}\sum_{i,j} X_{ij}\gamma_i\gamma_j + \sum_j X_{kj}\gamma_j > 0, \quad \text{for} \quad k = 1, \ldots, 4. \tag{79}$$

The set of equations (79) which is linear and homogeneous in the six x's defines a region of the x's in the six-dimensional space R^6, which can be visualized as the intersection of the "positive sides" of the four linear manifolds defined by setting the left-hand side of (79) equal to zero for each k. Let Ω_x denote this intersection. The fact that Ω_x is non-empty is trivial (since the dimensionality of the variables (x's) exceeds the number of constraints by two). It may be of some interest to note the subset of Ω_x for which the a_k's are positive definite (i. e., regardless of the γ_k's). For this we may rewrite (77) in the following matrix notation:

$$a_k = -\frac{1}{2}\gamma^{(k)T} L^{(k)} \gamma^{(k)}, \quad k = 1, \ldots, 4 \tag{80}$$

in which $\gamma^{(k)}$ denotes a 3×1 column matrix of the γ_j's with the deletion of the γ_k, e. g., $\gamma^{(1)} = \begin{pmatrix}\gamma_2\\ \gamma_3\\ \gamma_4\end{pmatrix}$, etc.. $L^{(k)}$ is a set of 3×3 symmetric matrices in the x's:

$$L^{(1)} = \begin{pmatrix} -2x_1 & x_4 - x_2 - x_1 & x_6 - x_3 - x_1 \\ x_4 - x_1 - x_2 & -2x_2 & x_5 - x_3 - x_2 \\ x_6 - x_1 - x_3 & x_5 - x_2 - x_3 & -2x_3 \end{pmatrix},$$

$$L^{(2)} = \begin{pmatrix} -2x_1 & x_2 - x_4 - x_1 & x_3 - x_6 - x_1 \\ x_2 - x_1 - x_4 & -2x_4 & x_5 - x_6 - x_4 \\ x_3 - x_1 - x_6 & x_5 - x_4 - x_6 & -2x_6 \end{pmatrix},$$

$$L^{(3)} = \begin{pmatrix} -2x_2 & x_1 - x_4 - x_2 & x_3 - x_5 - x_2 \\ x_1 - x_2 - x_4 & -2x_4 & x_6 - x_5 - x_4 \\ x_3 - x_2 - x_5 & x_6 - x_4 - x_5 & -2x_5 \end{pmatrix},$$

$$L^{(4)} = \begin{pmatrix} -2x_3 & x_1 - x_6 - x_3 & x_2 - x_5 - x_3 \\ x_1 - x_3 - x_6 & -2x_6 & x_4 - x_5 - x_6 \\ x_2 - x_3 - x_5 & x_4 - x_6 - x_5 & -2x_5 \end{pmatrix}.$$

Note that[41]

$$\det L^{(k)} = 2\Lambda(x), \quad \text{for} \quad k = 1, \ldots, 4. \tag{81}$$

Now with the a_k's regarded as the quadratic forms in the γ_k's in (80), a standard procedure of diagonalization immediately shows that the subset ω_x of Ω_x for which the a_k's are positive definite is given by

$$\omega_x = \left\{ x\colon x_\mu > 0,\ \lambda_k(x) < 0,\ \Lambda(x) < 0, \quad \begin{matrix} k = 1, \ldots, 4 \\ \mu = 1, \ldots, 6 \end{matrix} \right\}. \tag{82}$$

It is trivial to check that ω_x is non-empty. Thus $0 \neq \omega_x \subset \Omega_x$. Geometrically, $-\lambda(z_i, x_j, x_k) = 16$ times the squares of the area of the triangle with the

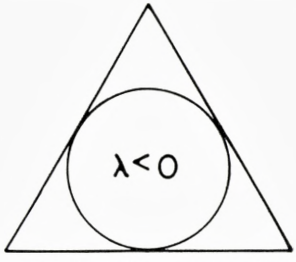

Figure 12. Projection of the $\lambda(x)$-cone.

sides $\sqrt{x_i}$, $\sqrt{x_j}$, $\sqrt{x_k}$; and $-\Lambda(x) = 144$ times the square of the volume of the tetrahedron formed by the six edges with lengths $\sqrt{x_\mu}$. In a 3-dimensional space, the region $\lambda(x) < 0$ is the interior of a cone[42] (tangent to all coordinate planes) within the octant $x_i, x_j, x_k > 0$ (cf. Fig. 12 above as projection). Now in the 6-dimensional space, one first goes to the *sexaginta-quadrant* $x_\mu > 0$, then takes the intersection of 4 sets of the λ-cones in the sub-3-spaces, and finally inscribes the surface of $\Lambda(x) = 0$ (which will be tangent to all four λ_k-cones). (No attempt is made to draw such a picture here, not even the projection).

This establishes that with suitably given z's (***Im*** z's satisfying (76), ***Re*** z's satisfying (79), and in particular (82)), the four parameters a_k can indeed be found simultaneously positive on the 4-mass envelope E_{1234}, and with this we conclude the existence of the 4-mass envelope.

[41] We note in passing that the structures of the $L^{(k)}$-matrices can be easily understood with the aid of the tetrahedron T of Sec. III.2. The diagonal elements in $L^{(k)}$ correspond to those edges emerging from the k-th vertex of Fig. 3, and the off-diagonal elements to the edges conjugate to this vertex (i. e. the k-th face).

[42] A beautiful picture of such λ-cone appeared in a recent paper of A. S. WIGHTMAN and H. EPSTEIN, Annals of Phys. **11**, 201 (1960), in an entirely different context.

We now proceed to discuss the relevance of E_{1234}. To be specific, consider $y_1, \ldots, y_5$ given according to (76), compute the y_6 from (74) (i. e. we get two horizontal lines $E^{\pm}_{1234}$ in the z_6-plane). From these 6 y's, compute the γ_k's from (72a). Now given more or less arbitrary $x_1, \ldots, x_5$, the linearity of (77) implies that a_k has one zero only on each of the $E^{\pm}_{1234}$. A

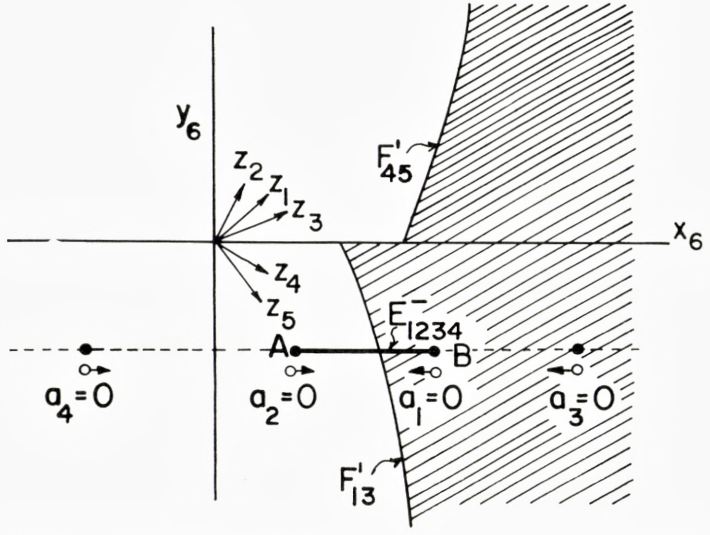

Figure 13. Straight-line segment in z_6-plane as the 4-mass envelope.

typical case is illustrated in Fig. 13. We use the symbol $0 \rightarrow$ to show the direction in which that particular a_j is positive. Hereafter, E_{1234} shall properly denote the allowed region of existence of the 4-mass envelope on which the intersection of all $a_k > 0$ has been taken (e. g., the segment between $a_2 = 0$ and $a_1 = 0$ in Fig. 13). By definition, E_{1234} is contained in Ω_x of (79); however, $E_{1234} \cap \omega_x$ may be empty. The case when $E_{1234} \cap \omega_x \neq 0$ has some pertinent features which we leave to the Appendix C. In general there are the following possibilities:

(a) E_{1234} is either empty for a particular configuration of the z's or is entirely contained inside the 3-point singularity domain: in such cases, the 4-mass envelopes are entirely irrelevant.
(b) E_{1234} is unbounded at one end which lies outside the 3-point singularity

domain: In this case (imagine all $\leftarrow 0$ pointing to the left in Fig. 13), it is also easy to dispose of by observing that the extreme far end of E_{1234} (which corresponds to all $a_k \to \infty$) is never relevant. Since the relevance of E_{1234} does not change unless it has an intersection; otherwise, the case is reduced to (c) below.

(c) E_{1234} is finite and partly lies outside the 3-point singularity domain (Figs. 13, 24, 25). This is the only outstanding situation of the 4-mass envelope which needs further discussion. It is clear that the path in the a-space corresponding to such a finite E_{1234} is a straight-line segment bounded by two 3-dimensional sub-spaces. As will be shown in Sec. VI.3, at the end $a_m = 0$ of E_{1234} comes the 3-mass envelope $E_{jkl}\,(j \neq k \neq l \neq m)$. Stated otherwise, the fact that E_{1234} suddenly comes to a stop must mean that there is another curve which would also pass through that point. For this, we must defer the remaining discussion of the role of the 4-mass envelope until we have treated the 3-mass envelopes in the next sub-section.

VI. 3 The 3-Mass Envelope

We have seen that the restrictions of the 4-mass envelope are so strong that one gets only rather trivial situations where E_{1234} is confined to a straight-line segment in the z-plane. The envelope condition (58) or (66) is relaxed when one goes from an m-envelope to an $(m-1)$-envelope; since, by definition, one of the parameters a_i now takes on the fixed extreme value 0, the corresponding restriction of $\partial \Psi / \partial a_i$ is then to be removed. We shall now sketch the necessary modification for the treatment of the 3-mass envelopes E_{jkl}. To be specific, let us consider E_{123}, the 3-mass envelope formed by a special path in the (a_1, a_2, a_3) 3-space. For this, we set once for all, $a_4 = 0$, in the expression for $\Psi(z; a)$. Strictly speaking, Q_4 which was defined as $\partial \Psi / \partial a_4$ is now meaningless, however, as a shorthand notation, we shall still use it as

$$Q_4 = \left(\frac{\partial \Psi}{\partial a_k} \right) \Bigg|_{a_4 = 0}$$

which, as stated above, is no longer restricted by the reality condition of (66). However, the identities (70) still hold with $a_4 = 0$. We may now define on E_{123} a set of 5 real γ_s, $s = 1, \ldots, 5$ such that

$$\left.\begin{aligned} Q_j &= \gamma_j\, G(z;\, a), \quad j = 1, \ldots, 3 \\ G(z;\, a) &= \sum_{j=1}^{3} Q_j = -4\,A(z) - Q_4\,; \\ \sum_{j=1}^{3} \gamma_j &= 1 \end{aligned}\right\} \tag{83}$$

and

$$Q_4 = (\gamma_4 + i\,\gamma_5)\cdot G(z;\, a).$$

Substituting (83) into (70) and dividing by $G(z;\, a)$, we get, after taking the imaginary and the real parts:

$$\sum_{k=1}^{4} (\boldsymbol{Im}\, \Psi_{ik})\,\gamma_k + (\boldsymbol{Re}\ \Psi_{i4})\,\gamma_5 = 0\,; \tag{84a}$$

$$\sum_{k=1}^{4} (\boldsymbol{Re}\ \Psi_{ik})\,\gamma_k - (\boldsymbol{Im}\, \Psi_{i4})\,\gamma_5 = 0\,; \quad \text{for} \quad i = 1, \ldots, 4. \tag{84b}$$

Now taking the fourth equation of (84b) together with (84a), we have

$$\sum_{t=1}^{5} U_{st}\,\gamma_t = 0, \quad s = 1, \ldots, 5 \tag{85}$$

where

$$U = \begin{pmatrix} 0 & y_1 & y_2 & y_3 & \xi_1^{(4)} \\ y_1 & 0 & y_4 & y_6 & \xi_2^{(4)} \\ y_2 & y_4 & 0 & y_5 & \xi_3^{(4)} \\ y_3 & y_6 & y_5 & 0 & 0 \\ \xi_1^{(4)} & \xi_2^{(4)} & \xi_3^{(4)} & 0 & 0 \end{pmatrix} \tag{86}$$

with

$$(\xi_j^{(4)}) = \begin{pmatrix} x_3 - a_1 \\ x_6 - a_2 \\ x_5 - a_3 \end{pmatrix} \tag{87}$$

where the superscript (4) is a reminder of $a_4 = 0$. Note that this column corresponds to the edges emerging from the 4-th vertex of the tetrahedron T of Fig. 3.

Since the $\det|U|$ must vanish for non-trivial solutions of the γ_s's, we have

$$\det|U| = -\,\xi^{(4)^T}\, V\, \xi^{(4)} = 0, \tag{88}$$

where V is a symmetric 3×3 *singular* matrix involving the y's alone:

$$V=\begin{pmatrix} 2y_4y_5y_6 & y_5(y_1y_5-y_3y_4-y_2y_6) & y_6(y_2y_6-y_1y_5-y_3y_4)\\ y_5(y_1y_5-y_3y_4-y_2y_6) & 2y_2y_3y_5 & y_3(y_3y_4-y_1y_5-y_2y_6)\\ y_6(y_2y_6-y_3y_4-y_1y_5) & y_3(y_3y_4-y_1y_5-y\ y\ ) & 2y_1y_3y_6 \end{pmatrix}. \quad (89)$$

Eq. (88) can be easily solved. The result is

$$\sum_j W_{ij}\,\xi_j^{(4)}=0, \quad \text{for} \quad i=1,2,3, \tag{90}$$

where W is also a 3×3 singular matrix (but in general unsymmetric):

$$W=\begin{pmatrix} V_{11} & V_{12}\mp\sqrt{\lambda(y_k y_{k'})} & V_{13}\pm\sqrt{\lambda(y_k y_{k'})}\\ V_{21}\pm\sqrt{\lambda(y_k y_{k'})} & V_{22} & V_{23}\mp\sqrt{\lambda(y_k y_{k'})}\\ V_{31}\mp\sqrt{\lambda(y_k y_{k'})} & V_{32}\pm\sqrt{\lambda(y_k y_{k'})} & V_{33} \end{pmatrix} \tag{91}$$

in which $\lambda(y_k y_{k'})$ is the determinant (73a) which vanishes on the 4-mass envelope.

From (91), it immediately follows that *the 3-mass envelopes cannot exist if*

$$\lambda(y_k y_{k'})<0. \tag{92}$$

This implies that

(a) If the y's satisfy the sign convention (76), then *the 3-mass envelopes can only lie outside the region bounded by the two lines of* (74). In particular, (92) implies that E_{jkl} *can never cross over* E_{1234} (cf. Fig. 15).
(b) On the other hand, if the y's do not obey the sign convention (76), then $\lambda(y_k y_{k'})>0$ always, and E_{jkl} may exist while E_{1234} cannot.

For case (a), i. e. when E_{1234} exists, we assert that E_{jkl} intersects with E_{1234} at the point which corresponds to the remaining parameter $a_m=0$ on E_{1234}. This is intuitively clear since, at the point $(a_j^*,\ a_k^*,\ a_l^*;\ a_m=0)$ on E_{1234}, we are in the 3-space of $(a_j,\ a_k,\ a_l)$ in which lies a path for E_{jkl}; now this point must actually lie on the path for E_{jkl}, since the condition for E_{1234} is sufficient for that of E_{jkl}. This statement can be explicitly verified by elementary computation. Considering the case $a_4=0$, we note that the following ratios hold for the γ_s's on E_{123}:

$$\left.\begin{aligned}
&\gamma_1 : \gamma_2 : \gamma_3 : \gamma_4 : \gamma_5 : = y_5 \left[(y_1 y_5 - y_2 y_6 - y_3 y_4) \pm \sqrt{\lambda (y_k y_{k'})} \right] \\
&: 2 y_2 y_3 y_5 : y_3 \left[(y_3 y_4 - y_1 y_5 - y_2 y_6) \mp \sqrt{\lambda (y_k y_{k'})} \right] : \\
&: y_2 \left[(y_2 y_6 - y_1 y_5 - y_3 y_4) \pm \frac{\xi_3^{(4)} y_3 + \xi_1^{(4)} y_5}{\xi_3^{(4)} y_3 - \xi_1^{(4)} y_5} \cdot \sqrt{\lambda (y_k y_{k'})} \right] \\
&: \frac{\mp 2 y_2 y_3 y_5}{\xi_3^{(4)} y_3 - \xi_1^{(4)} y_5} \cdot \sqrt{\lambda (y_k y_{k'})} .
\end{aligned}\right\} \quad (93)$$

It is clear then that, at $E_{1234} \cap E_{123}$, we have $\gamma_5 = 0$. Then the remaining four γ's will have exactly the same ratio as those in the case of the 4-mass envelope (cf. footnote 40, and Appendix C), and the solution to the 3-mass envelope will coincide with the solution to the 4-mass envelope E_{1234} at $a_4 = 0$ on the latter. This establishes our above statement that $E_{ijk} \cap E_{1234} \neq 0$.

We now return to the discussion of the situation (c) of E_{1234} in the last sub-section, in which E_{1234} has a finite strip lying outside the relevant 3-point

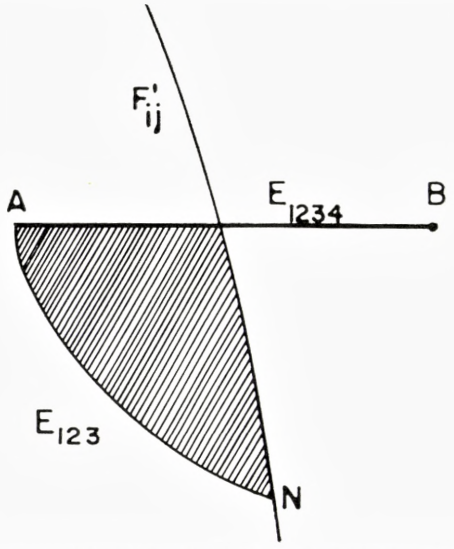

Figure 14. Inadmissible corners formed by the intersection between the 3-mass and the 4-mass envelopes.

singularity domain (cf. Fig. 13). The end-point A of E_{1234} corresponds to one particular $a_m = 0$ on E_{1234}, say $m = 4$. As we have just seen that E_{123} can only lie on one side of E_{1234} (e. g., below the segment AB in the z_6-plane,

cf. Fig. 14 above) and that E_{123} actually touches this end-point A. Let us imagine that E_{123} is depicted by some curve $\widehat{AN}$ in the z_6-plane (Fig. 14). The exact shape of E_{123} will not be important to us (cf. remark in connection with Fig. 15 below). Our discussion up to this point does not exclude the possibility that the shaded region in Fig. 14 might contain the 4-point singularity. But this we now proceed to show as inadmissible.

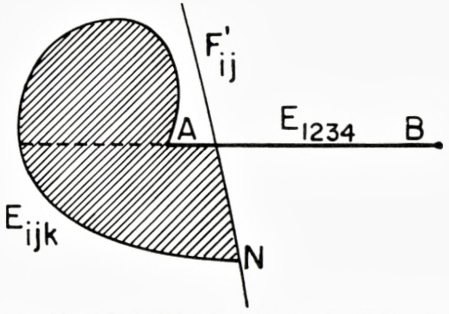

Figure 15. Admissible (but non-occurring) corners.

If this were really the case, the intersections of E_{ijk} with E_{1234} would be of such a kind that we had a corner in our domain. Since, as is characteristic of the theory of several complex variables, such corners are vulnerable to further analytic continuation[43], they cannot be part of the actual boundary of a natural domain of holomorphy. Note that if it were possible for E_{ijk} to cross over the 4-mass envelope like in the situation shown in Fig. 15, then this would in principle be admissible (since, in this case, the regularity domain would be the intersection of the two rather than the union as in Fig. 14). But our discussion of the 3-mass envelopes definitely excludes the possibility of such double intersections between E_{ijk} and E_{1234}. This leaves the only alternative of the corner as shown in Fig. 14, which one can reject as unacceptable for the boundary of our domain. Thus one concludes that the 4-mass envelope and the 3-mass envelopes do not contribute to

[43] A standard theorem is the well-known "Kantensatz". See, e. g., BEHNKE-THULLEN, *loc. cit.*, p. 52; KW's Sec. VI; and H. KNESER, Math. Ann. **106**, 656 (1932). Although, strictly speaking, this theorem has only been proved for corners formed by analytic surfaces, while in our present case we are presumably dealing with the non-analytic surfaces, one can in the neighborhood of such corners construct tangential analytic surfaces so that the shaving of the corner received from the "Kantensatz" on the enveloping analytic surfaces will automatically affect our present corner proper. I would like to thank both Professors JOST and KÄLLÉN for comments on this point.

the boundary[44]. In cases when the shaded region of Fig. 14 contains actual singularities, there must be another surface passing through and cover up this corner of Fig. 14. For this, we must go over to the treatment of the 2-mass envelopes.

VI. 4 The 2-Mass Envelope

As already mentioned in Sec. VI.3, the farther we go down to the envelopes of lower hierarchy, the less restrictions there are on the Q_k's. We shall first establish the intersection of the 2-mass envelope E_{ik} with the 3-mass envelope E_{ijk}. The method is quite analogous to the previous treatment of the 3-mass envelope.

We introduce a set of 6 real parameters γ_μ. For specificity, let us set $a_4 = a_2 = 0$ and consider E_{13} (i. e. the 2-mass envelope formed by a path in the quadrant $a_1 > 0$, $a_3 > 0$). As before, the quantities Q_4, Q_2 shall now be understood to stand for

$$Q_4 = \left(\frac{\partial \Psi}{\partial a_4}\right)\Bigg|_{(a_4 = a_2 = 0)}$$

$$Q_2 = \left(\frac{\partial \Psi}{\partial a_2}\right)\Bigg|_{(a_4 = a_2 = 0)}.$$

Since the envelope condition for E_{13} requires that

$$\boldsymbol{Im}\left(\frac{Q_1}{Q_3}\right) = 0,$$

we may set

$$\left.\begin{aligned} &Q_i = \gamma_i\, h(z;\, a), \quad i = 1, 3 \\ &\gamma_1 + \gamma_3 = 1 \\ &h(z;\, a) \equiv -4A(z) - Q_2 - Q_4 \\ &Q_2 = (\gamma_2 + i\gamma_6) \cdot h(z;\, a) \\ &Q_4 = (\gamma_4 + i\gamma_5) \cdot h(z;\, a). \end{aligned}\right\} \quad (95)$$

Substituting (95) into (70) and dividing by $h(z; a)$, we get, after taking the imaginary and the real parts:

[44] The role of the 3-mass envelopes in the case when the 4-mass envelope does not exist will not be discussed here. In view of the above feature for $m = 4$ that the $(m-1)$-envelope can only lie on one side of the m-envelope (i. e. meet at most tangentially), which will be seen later (Lemma 3) to be also valid for $m = 3$, one feels more confident that the 2-mass envelopes are actually more important even in this case.

$$\sum_{k=1}^{4} (\boldsymbol{Im}\, \Psi_{ik})\, \gamma_k + (\boldsymbol{Re}\, \Psi_{i2})\, \gamma_6 + (\boldsymbol{Re}\, \Psi_{i4})\, \gamma_5 = 0\,; \tag{96a}$$

$$\sum_{k=1}^{4} (\boldsymbol{Re}\, \Psi_{ik})\, \gamma_k - (\boldsymbol{Im}\, \Psi_{i2})\, \gamma_6 - (\boldsymbol{Im}\, \Psi_{i4})\, \gamma_5 = 0\,. \tag{96b}$$

Combining the second and the fourth equations of (96b) with (96a), we have:

$$\sum_{\nu=1}^{6} T_{\mu\nu}\, \gamma_\nu = 0\,, \quad \mu = 1, \ldots, 6 \tag{97}$$

where $(T_{\mu\nu})$ is a 6×6 symmetric matrix:

$$(T_{\mu\nu}) = \begin{pmatrix} 0 & y_1 & y_2 & y_3 & \xi_1 & \xi_3 \\ y_1 & 0 & y_4 & y_6 & 0 & \xi_6 \\ y_2 & y_4 & 0 & y_5 & \xi_4 & \xi_5 \\ y_3 & y_6 & y_5 & 0 & \xi_6 & \\ \xi_1 & 0 & \xi_4 & \xi_6 & 0 & -y_6 \\ \xi_3 & \xi_6 & \xi_5 & 0 & -y_6 & 0 \end{pmatrix} \tag{98}$$

where

$$\left.\begin{aligned} &\xi_i = x_i - a_1\,, \quad i = 1, 3 \\ &\xi_j = x_j - a_3\,, \quad j = 4, 5 \\ &\xi_6 = x_6\,. \end{aligned}\right\} \tag{99}$$

Let $T \equiv \det |\, T_{\mu\nu}\, |$. Now making use of the Jacobi theorem[45] on the expansion of the determinant in terms of the minors, we have

$$T = \frac{T^{55}\, T^{66} - (T^{56})^2}{\lambda (y_k\, y_{k'})}\,, \tag{100}$$

where $T^{\mu\nu}$ = minor of $T_{\mu\nu}$, being 5×5 determinants.

One immediately recognizes that T^{55} and T^{66} are precisely the determinants of the type (whose matrix is defined in (86)) for the 3-mass envelopes E_{123} (at $a_2 = 0$) and E_{134} (at $a_4 = 0$) respectively. From this, the intersection of the 2-mass envelope with the 3-mass envelope is quite obvious. Consider, e. g., $E_{13} \cap E_{123}$. Since T must vanish on E_{13}, and T^{55} vanishes on E_{123}, and consequently on $E_{13} \cap E_{123}$, we have

[45] See Appendix D.

$$\gamma_6 = 0$$

and

$$T^{55} = 0. \tag{101}$$

A straightforward computation with the aid of (90) will reveal that (101) reduces to the second equation of (96b) with $\gamma_6 = 0$ and $\gamma_1, \ldots, \gamma_5$ expressed

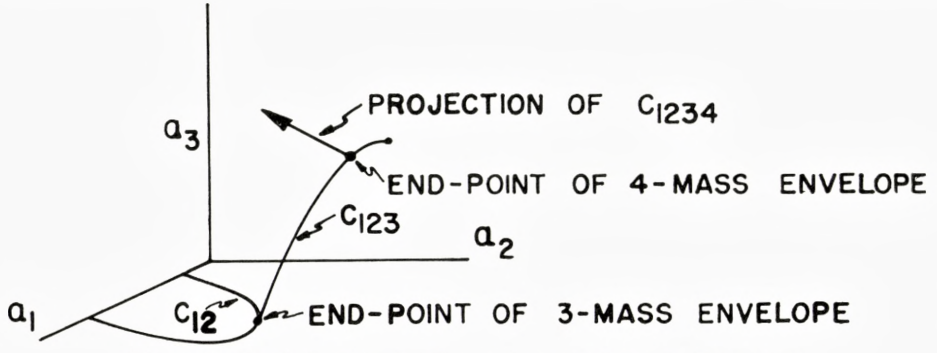

Figure 16. Typical paths for the various envelopes in the a-space.

by those on the 3-mass envelope E_{123} (cf. (93)). This means that (101) is automatically satisfied on $E_{13} \cap E_{123}$; hence there is no internal inconsistency. This shows that in general $E_{ik} \cap E_{ijk} \neq 0$.

This is also intuitively clear since the path C_{ijk} in the octant of all positive a's corresponding to the relevant portion of E_{ijk} is in general bounded by the coordinate 2-planes $a_m = 0$, $m = i, j,$ or k. Since the envelope condition for E_{ijk} is sufficient for E_{ik}, the end-points of C_{ijk} (in the finite case) must then necessarily lie on the path, say, C_{ik} for E_{ik}. This situation is depicted in Fig. 16, showing that the path of one of the $(m-1)$-envelopes passes through one of the end-points of the path for the m-envelope, $m = 2, 3, 4$. One further consequence for $E_{ik} \cap E_{ijk}$ is the following:

From (100), we have, since $T = 0$ on E_{13},

$$(T^{56})^2 = T^{55}\, T^{66}. \tag{102}$$

(102) can only be satisfied when T^{55} and T^{66} have the same sign. In the case when E_{134} and E_{123} are distinct, we have in the neighborhood of $E_{13} \cap E_{123}$, $T^{55} \approx 0$; while T^{66} (i. e. the determinant corresponding to E_{134}) will essentially remain unchanged in sign. Thus (102) immediately implies that T^{55} cannot change its sign in the neighborhood of $E_{13} \cap E_{123}$, i. e. E_{13} cannot cross over E_{123}. The same statement holds for E_{134}.

Collecting with this our previous result for $E_{ijk} \cap E_{1234}$, we have established the cases $m = 3, 4$ of the following:

Lemma 3: The intersection between the envelopes $E_f^{(m-1)}$ and $E_f^{(m)}$, for $m = 2, 3, 4$,

(1) is non-empty,
(2) occurs at the ends of $E_f^{(m)}$, and
(3) is "tangential".

Remark: (a) The subscript f is used to denote the case when the path $C^{(m)}$ for the m-envelope $E^{(m)}$ is finite (i. e. $C^{(m)}$ is bounded by the sub-$(m-1)$-spaces). Otherwise, in the case when $C^{(m)}$ is unbounded, one can always show that the corresponding envelopes are irrelevant.

(b) The term "tangential" is understood as saying that $E_f^{(m-1)}$ can only lie on one side of $E_f^{(m)}$ (i. e. cannot cross over $E_f^{(m)}$ at the intersection[46], in the z-space).

(c) Lemma 3 says *nothing* about the relationship between an $E^{(m-2)}$ and an $E^{(m)}$. Thus, for instance, a 2-mass envelope can cross over the 4-mass envelope to swallow the corner of Fig. 15 (cf. Fig. 18 below).

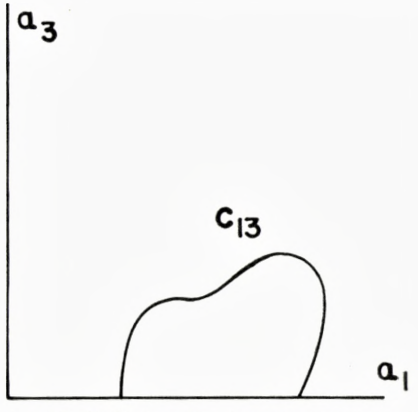

Figure 17. Path for the 2-mass envelope E_{13}.

(d) Whether $E_f^{(m)}$ will always contain the actual singularity is not fully settled here. This is true, however, in the 3-point case: while the $E_f^{(3)}$ and $E_f^{(2)}$, although not contributing to the boundary, do lie inside the sin-

[46] This feature seems to be also valid for the envelopes in the primitive domain of the 4-point function in the axiomatic approach. (Private communication from Professor G. KÄLLÉN).

gularity domain (on the cut)[47]. However, in the present case, we have one explicit example (cf. Fig. 24) where the corner formed by $E_f^{(4)}$ with $E_f^{(3)}$ is actually singular. (Of course, the case when $E_f^{(m)}$ is entirely contained inside the 3-point singularity domain is trivial).

It remains to say a few words about the case $m = 2$ in Lemma 3 which involves the 1-mass envelopes (strictly speaking, they are not envelopes).

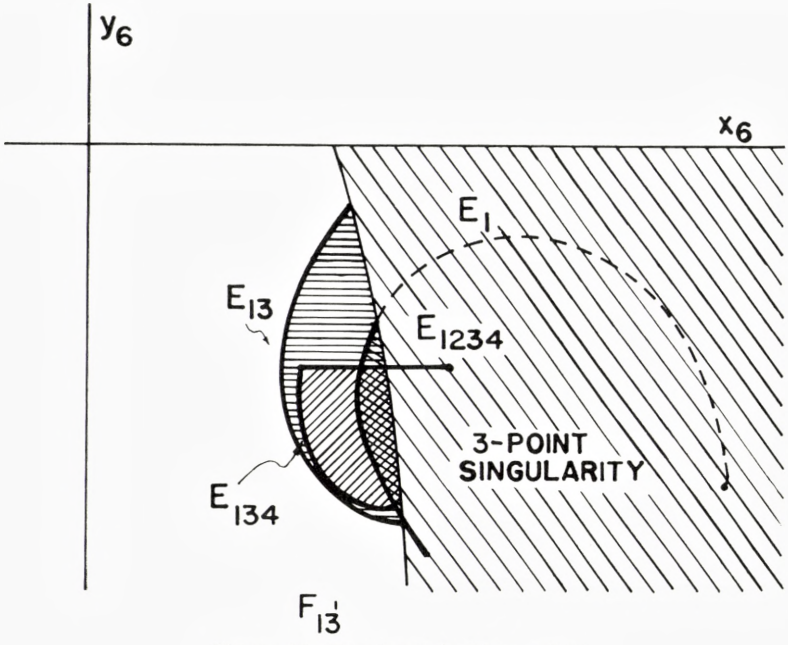

Figure 18. Envelopes in the z_6-plane.

One can, of course, explicitly show their intersections with the 2-mass envelopes in a perfectly analogous manner as was done above for $E_f^{(2)} \cap E_f^{(3)}$; we shall, however, omit this elementary computation here. Intuitively, it is clear in the quadrant $a_i > 0$, $a_k > 0$, since a finite path C_{ik} for E_{ik} must necessarily terminate on the semi-axes. A typical situation is shown in Fig. 17 in which C_{13} is bounded by the same axis. The image in the z_6-plane is shown in Fig. 18 where the 2-mass envelope E_{13} rides on top of the one-mass surface E_1, and the singularity domain is the union of the regions bounded by these two.

For completeness, we mention that, in the 3-point case, there occurs

[47] See Appendix B.

a peculiar situation where $E^{(1)} \cap E^{(2)} \cap E^{(3)} \neq 0$. This does not happen in general for the 4-point case. The only exception for $E^{(m-2)} \cap E^{(m-1)} \cap E^{(m)} \neq 0$ to occur would be when there are coincident zeros of the a's on $E^{(m)}$, e. g. Fig. 19. However, the situation in the 3-point case is actually of a slightly different nature than that of Fig. 19. There, the image of the 2-mass envelope in the z-plane happens to be a constant, so that E_{12} (the analogue of which in the 4-point case are the 3-mass envelopes) actually shrinks to a point which serves as the junction between the 1-mass F′-curves and the 3-mass envelope there[48].

The rest of this sub-section is devoted to the discussion of the connection

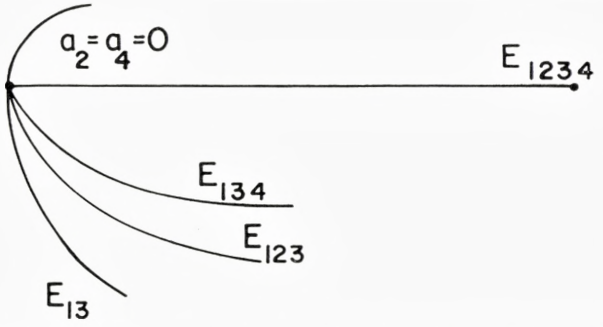

Figure 19. Multiple intersections among the envelopes in the 4-point case (Non-occurrence of).

of the 2-mass envelopes with the boundaries of the 3-point singularity domain, F'_{kl}- curves, and the equations for the former.

The conditions for the 2-mass envelopes are all contained in equations of type (96). However, for the 2-mass envelope, it is actually more convenient to take (94) together with the identities (68), (i. e. (69)). Thus, we have, for instance, on E_{13},

$$\left(\frac{\Phi_1}{\Phi_3}\right)_{a_2 = a_4 = 0} = \sigma^2, \quad \sigma \text{ real.} \tag{103}$$

Since, for $z_6 \neq 0$, we may write

$$\left(\frac{\Phi_1}{\Phi_3}\right)_{a_2 = a_4 = 0} = \frac{a_3 (z_6 - z_6^{(3)})}{a_1 (z_6 - z_6^{(1)})} = \sigma^2 > 0, \tag{104}$$

where

[48] See Appendix B.

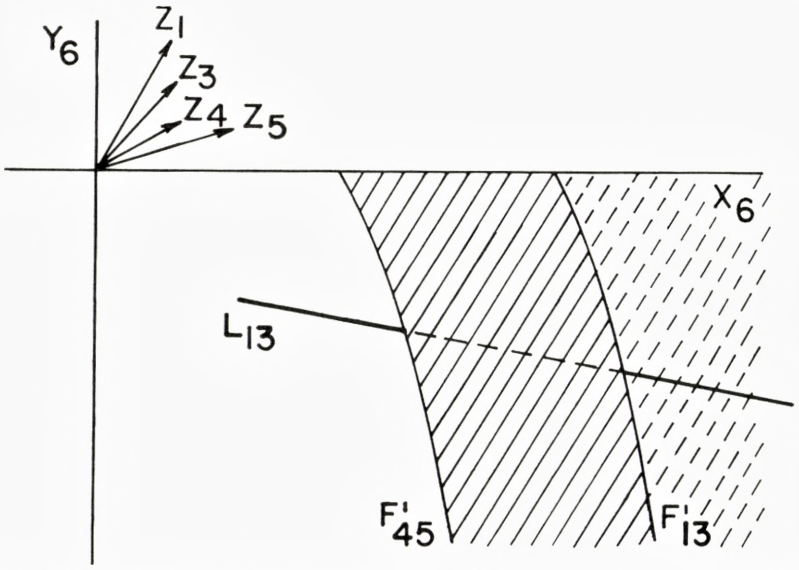

Figure 20. Allowed region for the 2-mass envelope: (outside the solid-line shaded region) when the two sets of F'-curves are in the same half-plane.

$$z_6^{(3)} = z_4 + z_5 - a_3 - \frac{z_4 z_5}{a_3} \tag{105a}$$

$$z_6^{(1)} = z_1 + z_3 - a_1 - \frac{z_1 z_3}{a_1} \tag{105b}$$

are the points on the F'_{45} and F'_{13}, respectively. Equation (104) allows a simple visualization of the location of the 2-mass envelope. Consider a point (a_1^*, a_3^*) on C_{13}, then in the z_6-plane one can locate two points $z_6(a_i^*)$, $i = 1, 3$, on F'_{13} and F'_{45}, respectively, according to (105). One sees then that the condition (104) for E_{13} at (a_i^*) can only be satisfied on the line L_{13} passing through $z_6(a_1^*)$, $z_6(a_3^*)$, excluding the segment between them. In other words, the 2-mass envelopes cannot exist in the region bounded by the two F'-curves, such as the shaded regions in Figs. 20 and 21. The exact image of the point (a_1^*, a_3^*) in the z_6-plane is given by the intersection of this line L_{13} with the Ψ-manifold, which now reads for $a_2 = a_4 = 0$:

$$z_6 = \frac{(z_1 - a_1)(z_5 - a_3) + (z_3 - a_1)(z_4 - a_3) \pm 2\sqrt{a_1 a_3 (z_6 - z_6^{(1)})(z_6 - z_6^{(3)})}}{(z_2 - a_1 - a_3)}. \tag{106}$$

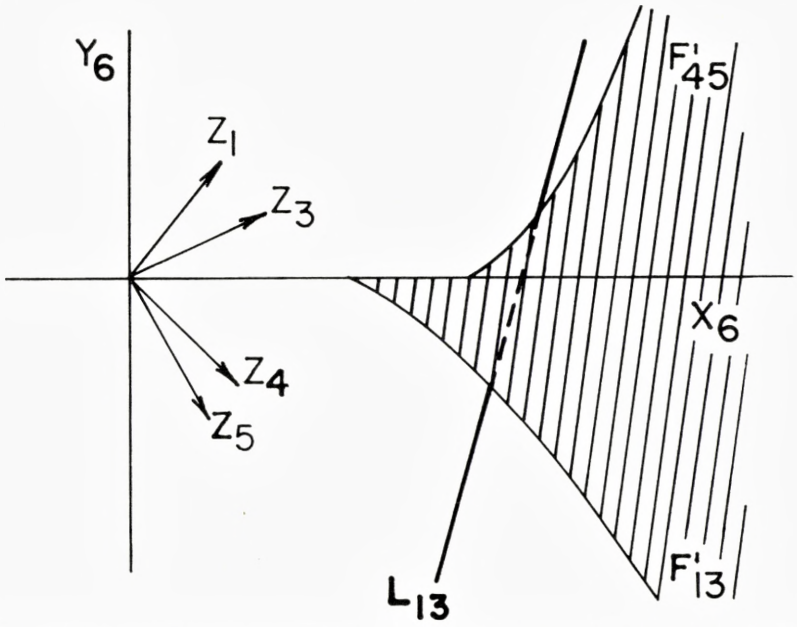

Figure 21. Allowed region for the 2-mass envelope: (outside the shaded region) where the two sets of F'-curves are in the opposite half-plane.

The elimination of z_6 from (104) and (106) is straightforward, and the resulting equation reads:

$$\left.\begin{aligned} 0 = \sigma^2 &\{a_1^3 - a_1^2 a_3 - a_1^2 (z_1 + z_3 - z_4 - z_5 + z_2) \\ &\quad - a_1 [z_1 z_5 + z_3 z_4 - z_1 z_3 - z_2 (z_1 + z_3)] + a_3 z_1 z_3 - z_2 z_1 z_3\} \\ + 2\sigma &\{a_1^2 a_3 - a_1 a_3^2 - a_1 a_3 (z_1 + z_3 - z_4 - z_5) - a_1 z_4 z_5 + a_3 z_1 z_3\} \\ - &\{a_3^3 - a_3^2 a_1 - a_3^2 (z_4 + z_5 - z_1 - z_3 + z_2) \\ &\quad - a_3 [z_1 z_5 + z_3 z_4 - z_4 z_5 - z_2 (z_4 + z_5)] + a_1 z_4 z_5 - z_2 z_4 z_5\}. \end{aligned}\right\} \quad (107)$$

Note that this equation is symmetric under the simultaneous permutation of

$$\left\{ a_1 \leftrightarrow a_3; \begin{pmatrix} z_1 \leftrightarrow z_4 \\ z_3 \leftrightarrow z_5 \end{pmatrix}; \quad \sigma \leftrightarrow \frac{1}{\sigma} \right\}$$

and the variation thereof (cf. Sec. III.2). The real and imaginary parts of (107) give two equations for the 3 parameters a_1, a_3, and σ. In principle, from these one is able to express two of the parameters in terms of the remaining one, say σ; thus, for fixed $z_1, \ldots, z_5$, one gets:

$$\begin{matrix} a_1 = a_1(\sigma) \\ a_3 = a_3(\sigma) \end{matrix}; \quad \sigma^2 > 0. \tag{108}$$

(108) then defines the path C_{13} when taken in the positive quadrant $a_1 > 0$, $a_3 > 0$. With this substituted into the equation resulting by solving for z_6 from (106), one gets the final equation for E_{13} in the z_6-lane, which for all other z's being fixed, reads

$$z_6 = z_6(\sigma). \tag{109}$$

In actual computation, however, the solutions of (108) from (107) involve great computational labor[49]. The other five E_{jk} envelopes are, fortunately, slightly less complicated. But we shall not go into all this.

Since, in the solutions (108) for the a's, the ***Re*** z's and the ***Im*** z's are well mixed, it is clear that (109) no longer gives an equation for an analytic hypersurface. Since, as we shall show in Sec. VI.5, the 1-mass surfaces (which are analytic) do not in general constitute the whole boundary to the 4-point domain, and since we have shown that in general the higher envelopes lead to the pathological situations shown in Fig. 14, the process of successive elimination forces the 2-mass envelopes to be the only remaining eligible candidates for our boundary. And indeed for one explicit configuration (cf. Fig. 24 in Sec. VI.5) we have shown that the 2-mass envelope does come in.

With this we conclude that non-analytic hypersurfaces do serve as part of the boundary to the 4-point domain in perturbation theory. In the final sub-section, we shall study those 1-mass surfaces[50] E_k and shall illustrate in some typical configurations the explicit behavior of E_k which indicates the presence of the envelopes.

VI. 5 The 1-Mass Surfaces

The 1-mass surfaces, as compared with the various envelopes we have discussed above, are much simpler objects, as they are simply the images of the four coordinate semi-axes in the a-space. Applying the technique of the determinant expansion of Appendix D, we have the following identity:

[49] With σ as a running parameter, one gets usually a 6th degree algebraic equation involving one final a_i.

[50] Chronologically, these 1-mass curves were investigated first. From these, we can easily convince ourselves that they do not give the whole boundary. One is then forced to undertake a lengthy treatment of the envelope problem which is summarized above.

$$\Psi = \frac{4\Phi_2\Phi_4 - (\Psi^{24})^2}{-R_2}. \tag{110}$$

Therefore, on the Ψ-manifold, we get

$$\Psi^{24} = \pm 2\sqrt{\Phi_2\Phi_4}. \tag{111}$$

Note that

$$\Psi^{24} = \frac{1}{2}\frac{\partial\Psi}{\partial z_6}. \tag{112}$$

Identity (110) is the proper generalization of (45) which holds in the 3-point case. In terms of z_6, (111) is equivalent to

$$z_6 = \frac{(z_4 - a_2 - a_3)\dfrac{\partial\Phi_2}{\partial z_5} + (z_1 - a_1 - a_2)\dfrac{\partial\Phi_2}{\partial z_3} \pm 2\sqrt{\Phi_2\Phi_4}}{R_2}. \tag{113}$$

For completeness, we mention that the analogue of KW (A. 48c) reads in the 4-point case as follows: On the Ψ-manifold

$$\pm\sqrt{\Phi_1} = \frac{\dfrac{\partial\Phi_4}{\partial z_1}\sqrt{\Phi_2} \pm \dfrac{\partial\Phi_2}{\partial z_3}\sqrt{\Phi_4}}{R_2} \tag{114a}$$

$$\pm\sqrt{\Phi_3} = \frac{\dfrac{\partial\Phi_4}{\partial z_4}\sqrt{\Phi_2} \pm \dfrac{\partial\Phi_2}{\partial z_5}\sqrt{\Phi_4}}{R_2} \tag{114b}$$

and the permutation thereof. (114) follows directly from (113), or equivalently also from (70) with the aid of (68).

The expressions for the 1-mass surfaces E_k (i. e. $a_k \neq 0$, for one k, all other a's being zero), which immediately follow from (113) by setting to zero 3 a's at a time, are summarized as follows:

For E_1: $a_1 > 0$:

$$z_6 = \frac{a_1}{(z_2 - a_1)^2}\left(\sqrt{z_4 w_4} \pm \sqrt{z_5 w_5}\right)^2, \tag{115}$$

For E_3: $a_3 > 0$:

$$z_6 = \frac{a_3}{(z_2 - a_3)^2}\cdot\left(\sqrt{z_1 w_1} \pm \sqrt{z_3 w_3}\right)^2, \tag{116}$$

where the w's are defined as

$$w_1 = z_1 - z_2 - z_4 + a_3 + \frac{z_2 z_4}{a_3} \tag{117a}$$

$$w_3 = z_3 - z_2 - z_5 + a_3 + \frac{z_2 z_5}{a_3} \tag{117b}$$

$$w_4 = z_4 - z_1 - z_2 + a_1 + \frac{z_1 z_2}{a_1} \tag{117c}$$

$$w_5 = z_5 - z_2 - z_3 + a_1 + \frac{z_2 z_3}{a_1} \tag{117d}$$

which *vanish* on the appropriate F'_{kl}-curves.

For E_2: $a_2 > 0$:

$$z_6 = \frac{1}{z_2}\left[z_1 z_5 + z_3 z_4 + a_2(z_2 - z_3 - z_5) \pm 2\sqrt{z_3 z_5 [a_2^2 + a_2(z_2 - z_1 - z_4) + z_1 z_4]}\right]. \tag{118}$$

For E_4: $a_4 > 0$:

$$z_6 = \frac{1}{z_2}\left[z_1 z_5 + z_3 z_4 + a_4(z_2 - z_1 - z_4) \pm 2\sqrt{z_1 z_4 [a_4^2 + a_4(z_2 - z_3 - z_5) + z_3 z_5]}\right]. \tag{119}$$

With $z_1, \ldots, z_5$ fixed, the above 4 curves E_k in the z_6-plane start from a common point G which corresponds to all $a_i = 0$ (for a given choice of the sign in front of the square root, cf. remark following (124) below)

$$z_6^{(0)} = \frac{\zeta}{z_2} \tag{120}$$

with

$$\zeta = z_1 z_5 + z_3 z_4 \pm 2\sqrt{z_1 z_3 z_4 z_5}.$$

The 4 curves E_k start from G with the following slopes:

$$\left(\frac{\partial z_6}{\partial a_k}\right)_G = \pm \frac{\sqrt{A(z)}}{z_2\sqrt{z_1 z_3 z_4 z_5}}\sqrt{\Phi_k^{(0)}}, \tag{121}$$

where $\Phi_k^{(0)}$ is Φ_k evaluated at all a's being zero, viz:

$$(\Phi_k^{(0)}) = \begin{pmatrix} z_4 z_5 z_6 \\ z_2 z_3 z_5 \\ z_1 z_3 z_6 \\ z_1 z_2 z_4 \end{pmatrix}. \tag{122}$$

On account of the identities (63) and (68), we have

$$\sum_{k=1}^{4}\left(\frac{\partial z_6}{\partial a_k}\right)_G = \mp\frac{\Lambda(z)}{z_2\sqrt{z_1 z_3 z_4 z_5}}. \tag{123}$$

One may note the analogy between the ratios among these slopes and those among the γ's on the 4-mass envelope (cf. footnote 40) if one replaces all the z's by ***Im*** z's.

Next we come to the asymptotic behavior of the E_k. For E_1 and E_3 in the z_6-plane, we have respectively (for other a's being zero or finite)

$$\lim_{a_1\to\infty} z_6 = \left(\sqrt{z_4}\pm\sqrt{z_5}\right)^2 \tag{124a}$$

$$\lim_{a_3\to\infty} z_6 = \left(\sqrt{z_1}\pm\sqrt{z_3}\right)^2. \tag{124b}$$

In other words, E_1 and E_3 terminate at finite points in the z_6-plane corresponding to $\lambda_1(z_4, z_5, z_6) = 0$ and $\lambda_3(z_1, z_3, z_6) = 0$, respectively. On the other hand, E_2 and E_4 extend to infinity in the z_6-plane as $a_2\to\infty$ and $a_4\to\infty$, with the following slopes:

$$z_6(a_2\neq 0)\xrightarrow[a_2\to\infty]{}\frac{1}{z_2}\left[z_2 - z_3 - z_5 \pm 2\sqrt{z_3 z_5}\right]a_2 \tag{124c}$$

$$z_6(a_4\neq 0)\xrightarrow[a_4\to\infty]{}\frac{1}{z_2}\left[z_2 - z_1 - z_4 \pm 2\sqrt{z_1 z_4}\right]a_4. \tag{124d}$$

We now proceed to investigate the relevance problem[51] of these 1-mass curves. First of all, the sign in front of the root in equations (115)–(119) should be chosen in such a way that one gets an enhancement rather than a cancellation among the terms. The latter is entirely irrelevant. This situation is also true for the lower order singularity manifolds. We recall that, in the 2-point case, the relevant cuts start from $z_k = (\sqrt{a_m} + \sqrt{a_n})^2$, but not from $(\sqrt{a_m} - \sqrt{a_n})^2$. In the 3-point case, the F'-curves are gotten by also choosing the sign which would add up terms (while the opposite sign gives exactly zero there). Of course, for complex quantities under the square roots, the sign is meaningful only with a suitable convention of the branches, which we shall take as the one with the positive imaginary part.

[51] To be precise, in view of the fact that part of the singular portion of E_k may be overriden by a 2-mass envelope (cf. Fig. 24), we are here seeking only the relevant portion of E_k in the following sense:

(i) it has actual singularities, *and*
(ii) it lies outside the 3-point singularity domain (but not necessarily as the actual 4-point boundary).

It is a consequence of Lemma 2 that E_k has a relevant portion if E_k intersects twice with the relevant portions of the dominating F'-curves *and* if the bubble formed by such double intersections lies outside the 3-point singularity domain. The condition for such double intersections between E_k and F' can in principle be stated algebraically as follows: Consider, for example, $E_1 \cap F'_{45}$. After rewriting (115) for E_k in the form

$$\left.\begin{aligned} &\lambda_1(z_4, z_5, z_6)\, a_1^2 - 2\, a_1 \{ z_6 [z_2 z_6 - z_1 z_5 - z_3 z_4 + 2\, z_4 z_5 - z_2 (z_4 + z_5)] \\ &\quad - (z_4 - z_5)(z_1 z_5 - z_3 z_4) \} + \lambda\, (z_k z_{k'}) = 0\,, \qquad 0 < a_1 < \infty, \end{aligned}\right\} \tag{125}$$

and with the relevant portion of F'_{45} given by

$$\left.\begin{aligned} z_6 &= z_4 + z_5 - \varrho - \frac{z_4 z_5}{\varrho} \\ 0 < \varrho &\leqslant \frac{\boldsymbol{Im}\, z_4 z_5}{\boldsymbol{Im}(z_4 + z_5)}, \end{aligned}\right\} \tag{126}$$

the problem is to find the condition on the configuration of the other 5 z's such that the system of equations (125), (126) admits at least two solutions for a_1 (or ϱ) in their respectively allowed ranges, as indicated above. This can be done by brute force, but the result is so complicated that we do not wish to display it here. The conditions are obviously dependent on the moduli (as well as the arguments) of the 5 z's, and we have not been able to deduce from it a concise statement about the desired configuration. (However, cf. (129)).

Instead, we shall in the following classify the configurations of the 5 z's by the location of the starting point G of E_k. There are three distinct cases:

Case (1); G lies *outside* the 3-point singularity domain;
Case (2): G lies *deep inside* the 3-point singularity domain;
Case (3): G lies *on* or *slightly inside* the 3-point singularity boundary.

From our studies of the E_k curves, we find that the first two cases do not yield anything of interest. They correspond to the situations where E_k has no intersection or non-relevant intersections with the F' curves. Therefore we shall concentrate on case (3) above, which also has an intuitively appealing feature for the desired intersections between the E_k and the dominant F'-curves.

The condition is then to require that at least one of the slopes for E_k

at G given by (121) has an intersection with the dominant F' curves. For the case when the latter are hyperbolas (i. e. $0 < \arg z_k + \arg z_l < \pi$), this implies that[52]

$$\arg z_k + \arg z_l < \arg \left(\frac{\partial z_6}{\partial a_i} \right)_G < \pi + \arg z_k + \arg z_l \qquad i = 1, \ldots, 4. \tag{127}$$

Condition (127), however, like the solutions to (125) and (126), is again dependent on the lengths of the z's in addition to their arguments.

It is clear that (127) is not sufficient to guarantee a double intersection even when G is chosen to lie on or slightly inside the F'-curves. However, only in such cases will the 1-mass curves E_k provide a useful hint as to how the 2-mass envelopes would come in. We illustrate this statement with the following 4 pictures: Fig. (22a) and Fig. (22b) show situations where the E_k's have the wrong slopes, and are irrelevant. In such cases, the envelopes are also irrelevant. Fig. (22c) shows a situation when one E_j comes out of the F'-region, while one other E_k stays inside. Although neither makes double intersections with F' (hence neither is relevant *per se*), the corresponding 2-mass envelope E_{jk} may very well form a bubble with F', which will serve as the 4-point boundary. Finally, in Fig. (22d), one sees a situation where one E_i does make a bubble with F' (the bubble can be shown to be relevant). On the other hand, another E_j also comes out of F', which by itself gives no contribution to the boundary; however, their 2-mass envelope E_{ij} may enlarge the bubble formed previously by E_i alone. This last phenomenon is what we have called the "overriding" of the relevant portion of 1-mass curves by a 2-mass envelope.

We shall now study some explicit examples. Let us first fix, for the sake of convenience, two (out of three in all) pairs of the conjugate variables (in the sense of Sec. II.2), say z_1, z_3, z_4, z_5. Ideally one would like to plot simultaneously in the product planes of the remaining pair of conjugate variables (i. e. z_2 and z_6), but for simplicity and practicality, we shall only plot in the z_6-plane (i. e. a 2-dimensional slice in the space of 12 dimensions) with suitable reference to the location of its conjugate variable z_2. The restriction on z_2 is as follows:

[52] The 3-point analogy of this condition is obvious: The relevance condition of the F'_{kl} curve itself, $z_m = z_k + z_l - r - z_k z_l / r$, can also be easily discussed by investigating the slope of the curve (actually the asymptote here for the hyperbola) at $r = 0$ (i. e. the analogue of the point G). Since one knows that the whole piece of F'_{kl} changes its relevance at its intersection with the cut along the positive x_m-axis, the relevance condition of F'_{kl} is to require that the slope at $r = 0$ should at least intersect with this x_m cut, i. e., $\pi < \arg(-z_k z_l) < 2\pi$, from which follows immediately the desired condition of the configuration: $0 < \arg z_k + \arg z_l < \pi$.

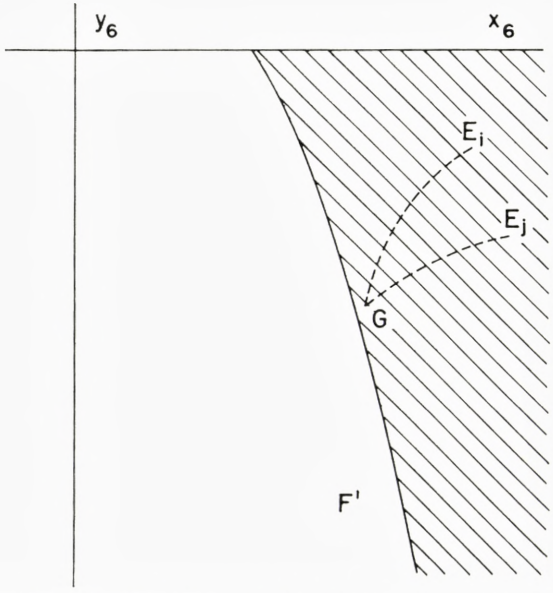

Figure 22a. Starting slopes of the 1-mass curves: (Irrelevant).

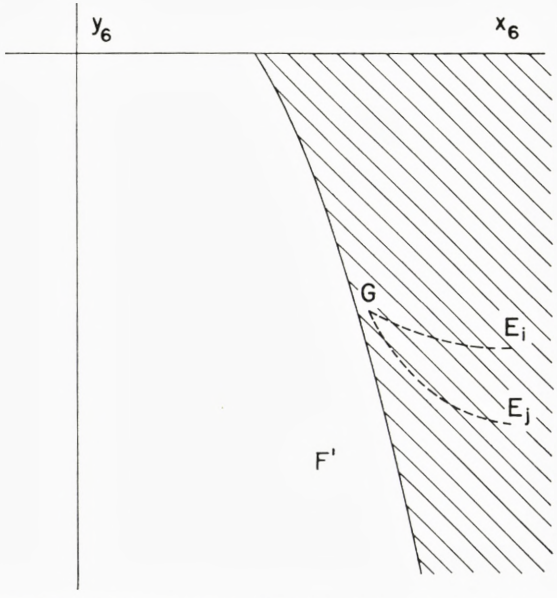

Figure 22b. Starting slopes of the 1-mass curves: (Irrelevant).

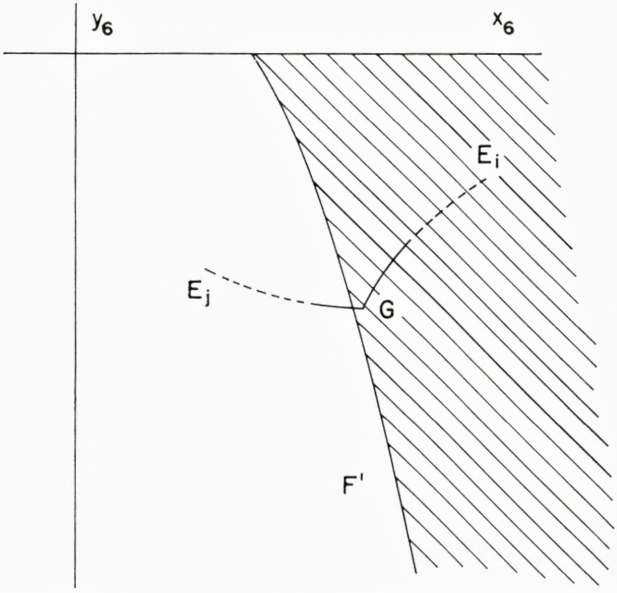

Figure 22c. Starting slopes of the 1-mass curves: (Irrelevant 1-mass curves, but relevant 2-mass envelopes).

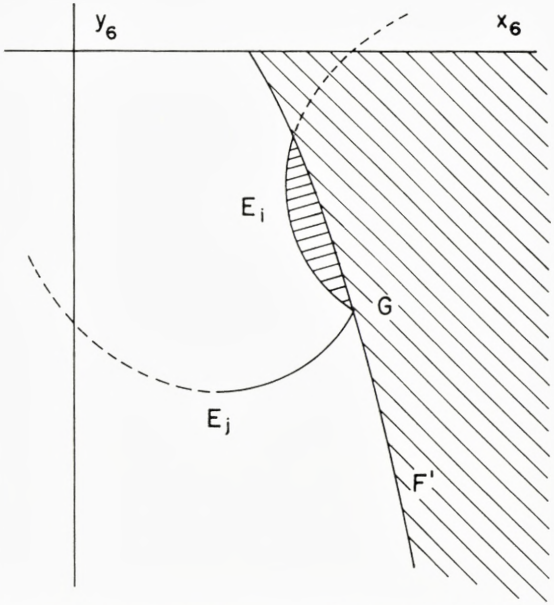

Figure 22d. Starting slopes of the 1-mass curves: (Relevant 1-mass curve and further 2-mass envelopes).

(1) z_2 shall not lie inside the relevant portions of the 3-point singularity manifolds $\Phi_2 = 0$ and $\Phi_4 = 0$, and
(2) z_2 shall be such that G of (120) lies on or slightly inside the dominating boundaries to the manifolds $\Phi_1 = 0$ and $\Phi_3 = 0$ in the z_6-plane. $\cdots$ (128)

Clearly there exists a limiting case of (120) when $\arg z_6^{(0)}$ and $\arg z_2$ approach respectively those of the asymptotes of the two dominating F' for z_6 and z_2. This implies, after a simple computation, the following necessary condition for the relevance of E_k for the case when $\boldsymbol{Im}\, z_i$, $i = 1, 3, 4, 5$, have the same sign[53]

$$2 \operatorname*{Max}_i \{\arg z_i\} < \sum_i \arg z_i, \qquad i = 1, 3, 4, 5. \tag{129}$$

In the following, we shall confine ourselves to the consideration of those configurations for which the four sets of the 3-point Φ_k-manifolds are simultaneously relevant. (A few remarks are, however, made near the end of the text, regarding the degenerate cases, cf. Lemma 4 of Sec. VI.6). This means that, if one is looking at the triplet (ijk) in the z_k-plane, one requires that the following 3-point conditions are to be satisfied:

$$\text{(a)} \quad 0 < \arg z_i + \arg z_j < \pi, \qquad \text{if } y_i y_j > 0,\ y_i > 0. \tag{130a}$$

$$\text{(b)} \quad 3\pi < \arg z_i + \arg z_j < 4\pi, \qquad \text{if } y_i y_j > 0,\ y_i < 0. \tag{130b}$$

$$\text{(c)} \quad \arg z_j > \pi + \arg z_i, \qquad \text{if } y_i y_j < 0,\ y_i > 0. \tag{130c}$$

One recalls that the configurations (a) and (b) yield hyperbolas and the configuration (c) gives a bubble in the z_k-plane.

There are *five* distinct configurations in the distribution of the 4 z_i's, $i = 1, \ldots, 4$. The first *four* cases correspond to $y_1 y_3 y_4 y_5 > 0$ (which imply the existence of the 4-mass envelope) and the remaining case is for $y_1 y_3 y_4 y_5 < 0$ (where E_{1234} does not exist).

(*A*) *All 4 Up*: (*Two sets of hyperbolas each for* z_6 *and* z_2).

In this case, we have:

$$y_1 y_3 y_4 y_5 > 0: \ y_i > 0. \tag{131}$$

[53] For other configurations with mixed signs of $Im\, z_i$, condition (129) can be easily modified by replacing some appropriate $\arg z_m$ by $2\pi - \arg z_m$ (Cf., e. g., Eq. (136)).

The 3-point boundaries are $F'_{13} \cup F'_{45}$ in the z_6-plane; and $F'_{14} \cup F'_{35}$ in the z_2-plane. (When there is no intersection among the two F' curves, one of the Φ-manifolds will be imbedded in the other, and the dominating F' curve is the one which corresponds to the *smaller* sum of the arguments. Otherwise, one has to take both of them into account.) The 3-point conditions are (130a) taken four times, or

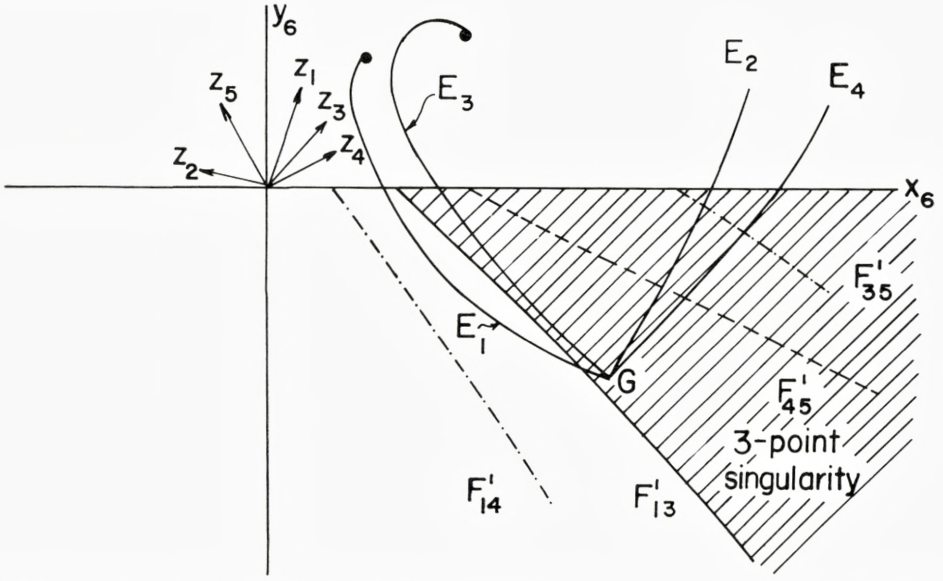

Figure 23. 1-mass curves in z_6-plane for the configuration (A): All 4 up (Two sets of hyperbolas each for z_6 and z_2).

$$\arg z_i + \arg z_j < \pi, \quad \text{for} \quad i,j = 1,3,4,5,: \quad \left.\begin{array}{l} i \neq j \text{ and} \\ i \neq \text{conjugate of } j. \end{array}\right\} \tag{132}$$

The 4-point condition (129) reads:

$$2 \operatorname*{Max}_i \{\arg z_i\} < \sum_i \arg z_i < 2\pi, \quad i = 1, 3, 4, 5. \tag{133}$$

In plotting in the z_6-plane, z_2 is to be chosen according to (128). A typical situation for this case is shown in Fig. 23.

(*B*) *Two Up and Two Down*: (*hyperbolas for* z_6: *bubbles for* z_2).

In this and the immediate next configurations, conjugate variables lie in the opposite half-planes. Here,

$$y_1 y_3 y_4 y_5 > 0; \; y_1 y_3 > 0; \; y_1 y_4 < 0. \tag{134}$$

The 3-point boundaries are $F'_{13} \cup F'_{45}$ in the z_6-plane; and $F'_{23} \cup F'_{25} \cup F'_{12} \cup F'_{24}$ in the z_2-plane. The 3-point conditions are explicitly (consider the case $y_1 > 0$)

$$\left.\begin{aligned} 0 < \arg z_1 + \arg z_3 < \pi \\ 3\pi < \arg z_4 + \arg z_5 < 4\pi \\ \arg z_4 > \pi + \arg z_1 \\ \arg z_5 > \pi + \arg z_3 . \end{aligned}\right\} \quad (135)$$

The 4-point condition (129) now takes the modified form

$$\text{(i)} \quad 2 \operatorname*{Max}_i \{\arg z_i\} < \sum_i \arg z_i < 5\pi \quad (136\text{a})$$

if

$$\operatorname{Max}\{\arg z_1, \arg z_3\} < 2\pi - \operatorname{Min}\{\arg z_4, \arg z_5\}$$

or

$$\text{(ii)} \quad 3\pi < \sum_i \arg z_i < 4\pi + 2 \operatorname{Min}\{\arg z_i\} < 5\pi \quad (136\text{b})$$

if

$$\operatorname{Max}\{\arg z_1, \arg z_3\} > 2\pi - \operatorname{Min}\{\arg z_4, \arg z_5\} .$$

A typical case is shown in Fig. 24.

Note: Figure 24 gives a very interesting example: E_1 makes a bubble with F'_{13} which can be shown to be singular. On the other hand, E_3 lies outside, and by itself is not relevant. Thus we have the situation shown in Fig. 22*d*. Now, if one takes the path $a_1 = a_3$ in the positive (a_1, a_3)-quadrant $(a_2 = a_4 = 0)$, one finds that its image in the z_6-plane makes another bubble with F'_{13}, which is also singular, but not contained by E_1. This shows definitely that

(a) *The 1-mass surfaces E_k do not in general give the whole boundary of D_4^{pert}*, and
(b) *Envelopes actually exist.*

Another curve, which corresponds to the path $a_1 = a_2 = a_3 = a_4$ in the positive sediciment, is also plotted in Fig. 24. However, it is not relevant in this case.

A plot of one of the simplest envelopes in the z_6-plane, namely E_{24}, is also made, but in this particular case, it is completely submerged inside the 3-point singularity domain.

Finally the 4-mass envelope E_{1234} is finite in this case, being bounded by $a_2 = 0$ and $a_1 = 0$. This is exactly the situation illustrated in Fig. 13.

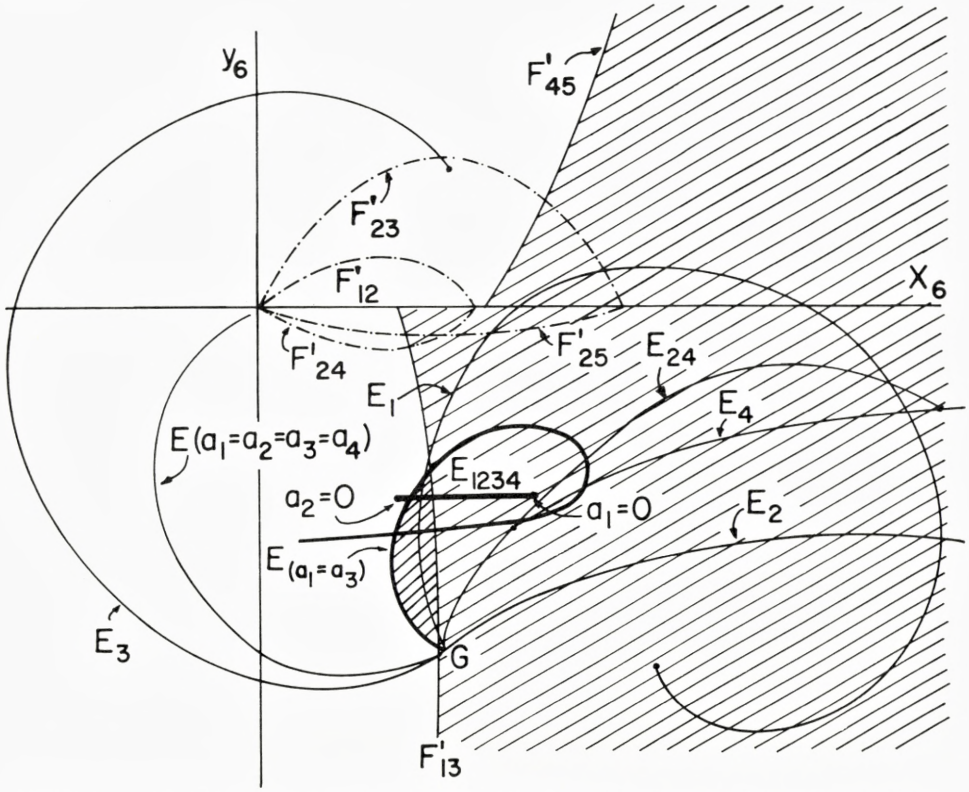

Figure 24. 1-mass curves in z_6-plane for the configuration (B): Two up and two down (hyperbolas for z_6; bubbles for z_2).

The end-point $a_2 = 0$ lies outside the F'_{13} and E_1 as well as $E_{a_1=a_3}$. The 3-mass envelope E_{134} can only come from below the E_{1234}. One will then have essentially a final situation similar to that shown in Fig. 18.

(*C*) *Two Up and Two Down*: (*bubbles for* z_6; *hyperbolas for* z_2).

This one gets from (*B*) by simply permuting within one pair of conjugate indices. The net result (cf. Sec. III. 2) is the interchange of the role of z_6 and z_2.
Thus, e. g., if one permutes z_3 and z_4 from (*B*),:

$$y_1 y_3 y_4 y_5 > 0: \; y_1 y_4 > 0, \; y_1 y_3 < 0. \tag{137}$$

The 3-point boundaries are $F'_{16} \cup F'_{36} \cup F'_{46} \cup F'_{56}$ in the z_6-plane, and $F'_{35} \cup F'_{14}$ in the z_2 plane. The 3-point and 4-point conditions are literally the same as (135) and (136) if one permutes z_3 and z_4. The E_k's are shown in Fig. 25. This suggests a 2-mass envelope.

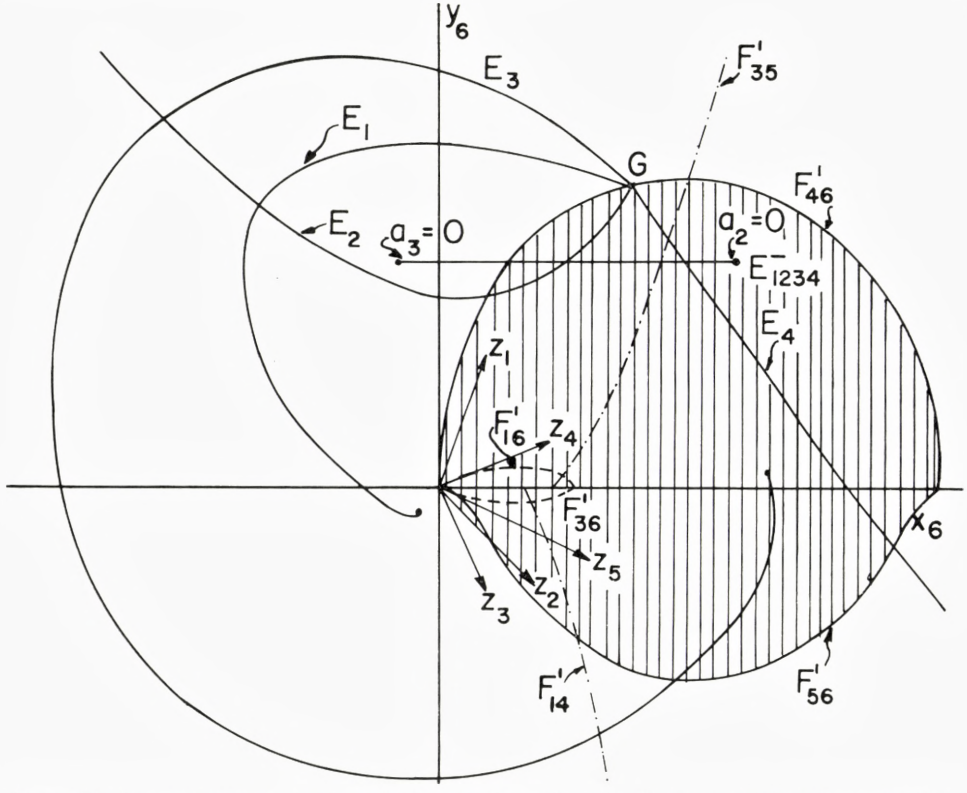

Figure 25. 1-mass curves in z_6-plane for the configuration (C): Two up and two down (bubbles for z_6; hyperbolas for z_2).

Note: In Fig. 25, one sees again the situation of Fig. 14. Here the 4-mass envelope E^-_{1234} is terminated at $a_3 = 0$. Now the 3-mass envelope E_{124} will intersect this point in the z_6-plane from above the line E^-_{1234} (since the other line E^+_{1234} in this case lies below E^-_{1234}, and from our analysis of (92), the 3-mass envelope must lie outside the region bounded by these two lines.) One gets again a corner in the intersection $E_{1234} \cap E_{124}$. A 2-mass envelope, say, E_{14}, is then expected to cover this corner. The situation is depicted in Fig. 25a.

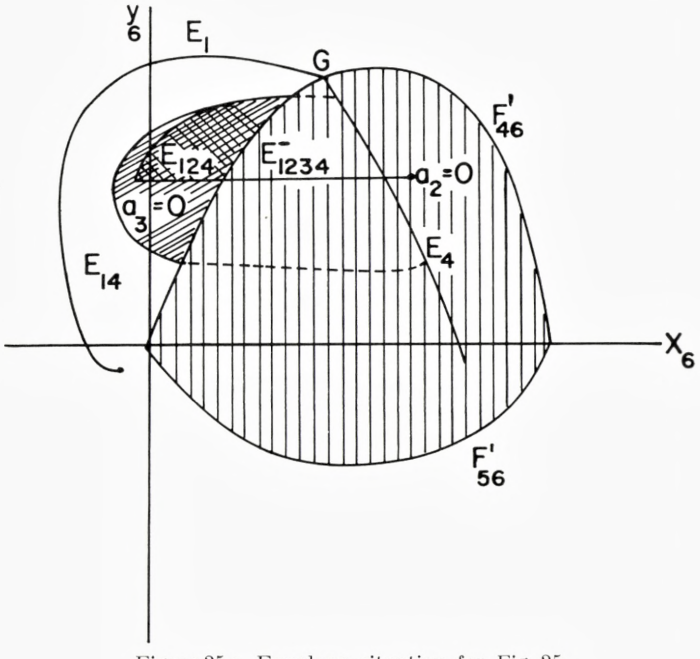

Figure 25a. Envelope situation for Fig. 25.

(*D*) *Two Up and Two Down*: (*bubbles for* z_6; *bubbles for* z_2).

One obtains this configuration from (A) when one shifts one pair of conjugate indices (*1,5*) or (*3,4*) to the opposite half-plane. Here we have, e. g.:

$$y_1 y_3 y_4 y_5 > 0: \quad y_1 y_5 > 0, \quad y_1 y_3 < 0. \tag{138}$$

The 3-point boundaries are $F'_{16} \cup F'_{36} \cup F'_{46} \cup F'_{56}$ in the z_6-plane, and $F'_{23} \cup F'_{25} \cup F'_{12} \cup F'_{24}$ in the z_2-plane. The 3-point condition is in this case (with (*3,4*) down)

$$\operatorname{Min}\{\arg z_3, \arg z_4\} > \pi + \operatorname{Max}\{\arg z_1, \arg z_5\} \tag{139}$$

and the 4-point condition reads:

$$\text{(i)} \quad \sum_i \arg z_i < 4\pi + 2 \operatorname{Min}\{\arg z_i\}, \quad i = 1, \ldots, 4 \tag{140a}$$

if

$$\operatorname{Max}\{\arg z_1, \arg z_5\} > 2\pi - \operatorname{Min}\{\arg z_3, \arg z_4\}$$

or

$$\text{(ii)} \quad \sum_i \arg z_i > 2\,\mathrm{Max}\{\arg z_i\}, \quad i = 1, \ldots, 4 \tag{140b}$$

if

$$\mathrm{Max}\{\arg z_1, \arg z_5\} < 2\pi - \mathrm{Min}\{\arg z_3, \arg z_4\}.$$

(*E*) *Three Up and One Down*: (*1 hyperbola and 1 bubble each for* z_6 *and* z_2). Here $y_1 y_3 y_4 y_5 < 0$. Consider, for example:

$$y_1, y_3, y_4 > 0, \text{ and } y_5 < 0. \tag{141}$$

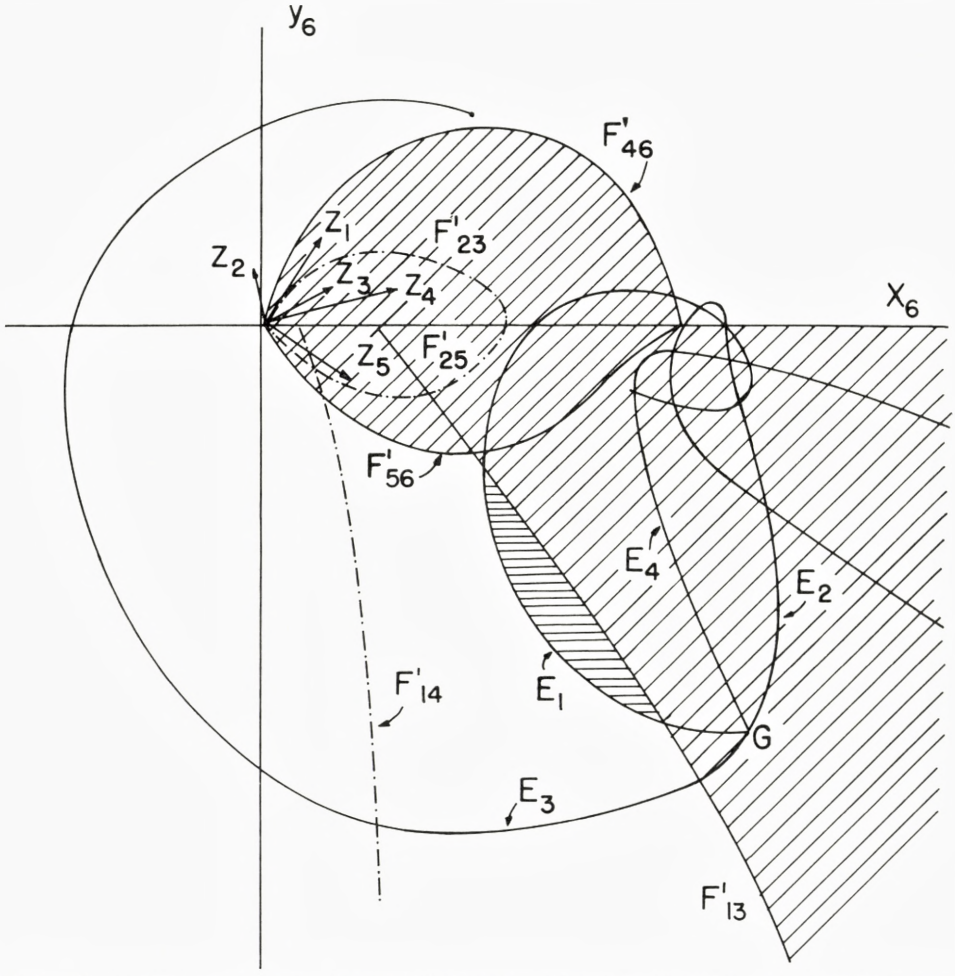

Figure 26. 1-mass curves in z_6-plane for the configuration (E): Three up and one down (one hyperbola and one bubble each for z_6 and z_2).

The 3-point boundaries are then $F'_{13} \cup F'_{46} \cup F'_{56}$ in the z_6-plane, and $F'_{14} \cup F'_{23} \cup F'_{25}$ in the z_2-plane. The 3-point conditions are:

$$\left.\begin{aligned} &\arg z_1 + \operatorname{Max}\{\arg z_3, \arg z_4\} < \pi \\ &\arg z_5 > \pi + \operatorname{Max}\{\arg z_3, \arg z_4\} \end{aligned}\right\} \quad (142)$$

and the 4-point condition reads for this case:

$$2 \arg z_5 + 2 \operatorname{Max}\{\arg z_1, \arg z_3, \arg z_4\} - 2\pi < \sum_i \arg z_i < 2 \arg z_5. \quad (143)$$

A typical case is shown in Fig. 26 (which suggests a 2-mass envelope).

VI. 6 Brief Remarks on the Degenerate Cases.

In our above description of the 1-mass curves, we have only considered the configurations where all four sets of the Φ_k-manifolds are simultaneously relevant. It would be of interest to see how the 4-point boundary changes its character when one or more Φ_k-manifolds become irrelevant. While we shall not attempt to enter into the discussion for this in detail, we offer two remarks on such degenerate cases:

(1) **Lemma 4:** Non-relevance of 2 sets of Φ-manifolds must imply the non-relevance of at least one more set.

Proof:

It suffices to show this for one particular configuration, say, in the case when 4 of the 6 z's are all in the upper half-plane, $0 < \arg z_i < \pi$, $i = 1, 3, 4, 5$ (cf. configuration (A) of Sec. VI.5), (since the proof for the other configurations can be easily carried through with only trivial modifications).

Suppose Φ_2 and Φ_4 manifolds are both irrelevant in the z_2-plane, then

$$\arg z_1 + \arg z_4 > \pi$$
$$\arg z_3 + \arg z_5 > \pi.$$

Assume Φ_3-manifold to be relevant in the z_6-plane (otherwise, nothing is to be proved), so

$$\arg z_1 + \arg z_3 < \pi.$$

Then Φ_1-manifold must be irrelevant, since

$$\arg z_4 + \arg z_5 > 2\pi - \arg z_1 - \arg z_3 > \pi.$$

(1a) An immediate consequence of Lemma 4 is the following. When Φ_2 and Φ_4 manifolds are both irrelevant (thus, e. g., one has the cut-plane in z_2), then in the z_6-plane, one has at most one set of relevant F' curves for the 3-point boundary. In this case, the 2-mass envelopes will not be expected to play a role, and the 4-point boundary will then be at most made up of the 1-mass surfaces which are analytic.
(2) The case when all 4 sets of Φ_k-manifolds are simultaneously irrelevant is, of course, trivial. Absence of any relevant 3-point boundary implies no change of relevance for the 4-point boundary. Since the latter cannot be entirely relevant, it must be entirely irrelevant. Thus, in this case, one gets the cut-planes.

VI. 7 Conclusion

It should be emphasized that we have by no means exhausted the boundary of the 4-point domain in perturbation theory. In fact, we have only explored it to the extent that we have shown how the 4-point correction to the already existing 3-point singularity might look. Our studies of the domain D_4^{pert} shows that the relevant 4-point singularities will carve out some bubbles from the dominating F' curves of D_3^{pert}. The singularity domain of the 4-point proper is seen to be compact. We have demonstrated that in general the 1-mass surfaces will not constitute the whole boundary of D_4^{pert} and that the presence of the envelopes implies that D_4^{pert} is not everywhere bounded by analytic hypersurfaces. Of the various envelopes we have discussed, the 2-mass envelopes are the most important ones.

It is hoped that, if the 3-point analogy is again valid in the 4-point case, the results derived here might be of some use to the problem of finding the holomorphy envelope $E(D_4)$ based on the axioms of local field theory alone.

We conclude by posing a question. One recalls again from the 3-point case that the domain D_3^{prim} is bounded by the F-curves (say, for the case when both $\boldsymbol{Im}\ z_j$, $\boldsymbol{Im}\ z_k$ have the same sign) of KW, which differ from the holomorphy envelope F'-curves only by the exactly opposite signs of the range of the parameters a_i (which, in the p-space, has the significance of being m_i^2). Intuitively, this can be understood as follows: If one starts from the original tube domain R_{n-1} of the vectors p_i where one requires $\boldsymbol{Im}\ p_i \varepsilon V$, this automatically forces one to go off the mass-shells and in particular one finds it convenient to go to negative values of the mass-squares $a_i = m_i^2 < 0$.

(This situation is clear, for example, in the proof of dispersion relations, with the technique of BOGOLIUBOV[54].) So $\partial D_3^{\text{prim}}$ essentially involves the manifold with the parameters still in the range $a_i < 0$. The problem of finding the holomorphy envelope then furnishes the necessary analytic continuation from $a_i < 0$ to $a_i > 0$, which is by no means trivial. This is exactly the relation between the F-curve of D_3^{prim} and the F'-curve of $E(D_3)$, or the F'-curve of D_3^{pert}, as shown by KW.

Therefore it will be of interest to see whether or not this analogy is a valid one in the 4-point case, viz., whether $\partial D_4^{\text{pert}}$ can be compared with $\partial D_4^{\text{prim}}$ with only a possible difference of the signs of the parameters*. Of course, the problem is much more complicated in the 4-point case, since one is dealing with the envelopes in both $\partial D_4^{\text{prim}}$ and $\partial D_4^{\text{pert}}$. An answer in the affirmative sense would further strengthen one's hope that $\partial D_4^{\text{pert}}$ may have something to do with $\partial E(D_4)$. But this we shall leave to a separate investigation.

Acknowledgement

I am most grateful to Professor JOHN S. TOLL for his continual interest and encouragement and for enabling me to accompany him to Lund, Sweden, where this work was started during my stay at *Institutionen för Teoretisk Fysik, Lunds Universitet*, in 1958–59. I am deeply indebted to Professor GUNNAR KÄLLÉN for suggesting this investigation and for his invaluable guidance and persistent help through all stages of this work. It is a great pleasure to thank Professor TORSTEN GUSTAFSON for his cordial hospitality in Lund. The financial support of the National Science Foundation and the U.S. Air Force during my year at Lund is gratefully acknowledged.

The work was completed during my stay at the *Institut für Theoretische Physik, Eidgenössische Technische Hochschule* in Zürich in 1959–60. I am very much obliged to Professor RES JOST for many enlightening discussions and for the hospitality of this Institute. Finally, the financial support of the *Schweizerische Nationalfond* is greatly appreciated.

[54] See, reference cited in footnote 4.

* *Note added in proof*: For the 3-point case, reversing the sign of the parameter a_i in the F_{jk}-curve is equivalent to changing the sign of $(z_i - z_j - z_k)$. The latter scheme is better suited in a generalization to the 4-point case. Preliminary investigation shows that the analogy of such a simple relationship between $\partial D_3^{\text{pert}}$ and $\partial D_3^{\text{prim}}$ may very well break down in the 4-point case.

Appendix A

Explicit Form of the 4-Point Function

Here we discuss the singularities of $I(z)$ of (54) after explicitly carrying out the final integration for $F_j(r_i)$ in (55). For the case $r_1 \neq r_2$, the singularities are found on the Φ_k-manifolds and the R_μ-manifolds (cf. Sec. IV. 4). Finally, for the case $r_1 = r_2$, the change of relevance of the Ψ-manifold is shown to occur at $(\Psi = 0) \cap (\Phi_k = 0)$ (cf. Sec. V).

By symmetry, it suffices to write down $F_j(r_i)$ only, say, for $j = 1$, $i = 1$, namely for the first half of the terms for the triplet (z_2, z_3, z_5). A straightforward computation from (55) yields:

$$\left.\begin{aligned} F_1(r_1) = \sqrt{\lambda_1}\Big\{ &\log\frac{1-r_1}{-r_1}\log\chi_1(r_1) + \log\frac{\eta_1(1)}{\eta_1(0)}\log\chi_1(r_1) \\ &+ S_1(\eta_{i1}(r_1)) - S_1(\eta_{i2}(r_1))\Big\} \end{aligned}\right\} \quad (A.1)$$

in which each $S_j(\eta_{ik}(r_i))$ is a sum of 16 Spence functions:

$$S_1(\eta_{i1}(r_1)) = \sum_{\mu=1}^{8} \varepsilon_\mu \left[\varphi\left(\frac{\eta_1(r_1)-\eta_1(1)}{\eta_1(r_1)-\omega_\mu}\right) - \varphi\left(\frac{\eta_1(r_1)-\eta_1(0)}{\eta_1(r_1)-\omega_\mu}\right)\right], \quad (A.2)$$

where

$$\left.\varepsilon_\mu = \begin{cases} 1, & \mu = 1, \ldots, 4 \\ -1, & \mu = 5, \ldots, 8 \end{cases}\right\} \quad (A.3)$$

$$\varphi(\zeta) = \int_1^\zeta \frac{dt}{t}\log(1+t) \quad (A.4)$$

and

$$\eta_{1,2}(\alpha) = \frac{1}{\lambda_2}\left[\frac{1}{2}\frac{\partial N_1(\alpha)}{\partial\alpha} \pm \sqrt{\lambda_2 N_1(\alpha)}\right] \quad (A.5)$$

$$\left.\begin{aligned} \omega_{1,2} &= \frac{2z_5}{\lambda_2}\cdot\frac{P_2 \pm \sqrt{\lambda_2 R_2}}{\frac{1}{2}\frac{\partial\lambda_2}{\partial z_3} - \sqrt{\lambda_2}} \\ \omega_{3,4} &= \frac{2z_5}{\lambda_2}\cdot\frac{P_3 \pm \sqrt{\lambda_2 R_3}}{\frac{1}{2}\frac{\partial\lambda_2}{\partial z_2} - \sqrt{\lambda_2}} \end{aligned}\right\} \quad (A.6)$$

$$\left.\begin{aligned} &\omega_{5,6} = A_3\omega_{1,2} \\ &\omega_{7,8} = A_2\omega_{3,4} \\ &\text{where} \\ &A_m = \frac{\frac{1}{2}\frac{\partial\lambda_2}{\partial z_m} - \sqrt{\lambda_2}}{\frac{1}{2}\frac{\partial\lambda_2}{\partial z_m} + \sqrt{\lambda_2}}, \qquad m = 2, 3. \end{aligned}\right\} \quad (A.6a)$$

The following identities can be easily verified:

$$\frac{4z_5}{\lambda_2^2}\Phi_2 = \eta_1(\alpha)\,\eta_2(\alpha) = A_3\omega_1\omega_2 = A_2\omega_3\omega_4 = A_3^{-1}\omega_5\omega_6 = A_2^{-1}\omega_7\omega_8 \quad (A.7)$$

and

$$\left.\log\chi_1(\alpha) = \log A_2A_3\frac{\prod_{\mu=1}^{8}(\eta_1(\alpha)-\omega_\mu)}{\prod_{\mu=5}^{8}(\eta_1(\alpha)-\omega_\mu)} = -\log A_2A_3\frac{\prod_{\mu=1}^{8}(\eta_2(\alpha)-\omega_\mu)}{\prod_{\mu=5}^{8}(\eta_2(\alpha)-\omega_\mu)}\right\} \quad (A.8)$$

We now briefly discuss the singularities of $\sum_j F_j(r_1)$, with $F_1(r_1)$ given by (A.1), for the case $r_1 \neq r_2$.

(1) The first term is $\log\frac{1-r_1}{-r_1}\cdot\log\chi_j(r_1)$. The point $r_1 = 0$ corresponds to the Φ_1-manifold (cf. (48)). It is clear that a cancellation of the 3-point type occurs here when the summation over j is carried out. Finally, for $r_1 \neq 0$, or 1, the zeros and poles of $\chi_j(r_1)$ can at most lead to the cuts in the z's. (cf. (51) and Sec. IV.4).

(2) The next term is $\log\frac{\eta_1(1)}{\eta_1(0)}\log\chi_j(r_1)$. Here the vanishing of $\eta_1(1)$ or $\eta_1(0)$ gives the Φ_{j+1}-manifold.

(3) Now we come to the Spence function terms. Each Spence function $\varphi(\zeta)$ is defined with a cut in the ζ-plane starting from its branch point at $\zeta = -1$ to infinity. Now the branch points in (A.2) occur at

$$\left.\begin{aligned} &\omega_\mu - \eta_1(1) = 0, \\ &\omega_\mu - \eta_1(0) = 0. \end{aligned}\right\} \quad \mu = 1, \ldots, 8 \quad (A.9)$$

and

With the aid of (A.8), we see that this happens at the two ends of the integration interval. Again the point $\alpha = 1$ is irrelevant. But the point $\alpha = 0$

leads to the Φ_1-manifold, which is to be expected. Note that the points $r_1 = 1$ or 0, which give $\varphi(0) = -\pi^2/12$, are entirely harmless for the Spence functions.

Another source of singularity for the Spence function $\varphi(\zeta)$ is at infinity. Now this happens when

$$\eta_1(r_1) - \omega_\mu = 0, \quad \mu = 1, \ldots, 8 \tag{A.10}$$

or, according to (A.8), this implies zeros or poles of $\chi_j(r_1)$. But these can at most correspond to the individual cut in each of the z's.

Thus we conclude that from the explicit expression (A.1) and its permuted form for the case $r_1 \neq r_2$, the singularities of the 4-point function $I(z)$ of (54) are confined to the 4 sets of Φ_k-manifolds and the 6 cuts, one for each z along the positive real axis. This agrees with our simple argument in Sec. IV.4.

Finally, from the representation (A.1) and its permuted forms of $\sum_j F_j(r_i)$, we now briefly discuss the change of relevance of the Ψ-manifold in the case $r_1 = r_2$. Here the expression (54) gets essentially a contribution from the first term in (A.1) (summed over j) in the neighborhood of the Φ_1-manifold:

$$\left.\begin{aligned} &\frac{1}{r_1 - r_2}\left[\log r_1 \log \prod_j \chi_j(r_1) - \log r_2 \log \prod_j \chi_j(r)\right] \\ &\quad \sim \frac{n \cdot 2\pi i}{r_1 - r_2} \cdot \log \frac{r_1}{r_2}\bigg|_{\Psi = 0} \sim \frac{nm(2\pi i)^2}{\sqrt{\Psi}}, \end{aligned}\right\} \tag{A.11}$$

where

$$\log \prod_j \chi_j(r_i) = \log 1 = n \cdot 2\pi i, \quad \text{by virtue of (51)},$$

and

$$\log \frac{r_1}{r_2}\bigg|_{\Psi = 0} \sim \log 1 = m \cdot 2\pi i; \quad n, m, \text{ integers}.$$

On the Φ_1-manifold, one of the r_i, say r_1, becomes zero, while the other is finite. Thus on one side of Φ_1-manifold, $m = 0$ (if we are on the principal sheet to start with, e. g., for all z's being negative real), but on the other side, $m \neq 0$. This shows a change of relevance of the Ψ-manifold at its intersection with the Φ_1-manifold. In a quite similar fashion, e. g., from the second term in (A.2), there develops a change of relevance across the Φ_{j+1}-manifold. To show this, it suffices to note that at $(\Psi = 0) \cap (\Phi_{j+1} = 0)$, one gets $N_j(r_i) = 0$, whence $\log \chi_j(r_i) = \log 1$ also, for each $j = 1,2,3$.

This confirms Lemma 2 in a more explicit way.

Appendix B

Envelope Problem for the Φ-manifold

In stating that the 3-point domain is bounded by analytic hypersurfaces F'_{kl} (obtained by setting two of the three mass-parameters equal to zero from the Φ-manifold), it is understood that the envelopes of the Φ-manifold are trivial, in the sense that they do not exist off the cuts and hence never actually contribute to the boundary (apart from what one has already on the cut). The purpose of this appendix is twofold:

(a) To give a proof of the above statement[55], and
(b) Since the Φ-manifold is of a much simpler structure, the analysis here actually serves as a prototype for the treatment of the Ψ-manifold (cf. Sec. IV), despite the fact that the final situations are quite different in two cases.

The notation here for the variables in the Φ-manifold follows that of KW.

I. 3-*Mass Envelope* E_{123}:

Let

$$\Phi(z;a)=\frac{1}{2}\begin{vmatrix} -2a_1 & z_3-a_2-a_1 & z_2-a_3-a_1 \\ z_3-a_1-a_2 & -2a_2 & z_1-a_3-a_2 \\ z_2-a_1-a_3 & z_1-a_2-a_3 & -2a_3 \end{vmatrix}. \tag{B.1}$$

The analogue of (63) is

$$\sum_{k=1}^{3} P_k=\lambda(z), \qquad P_k=\frac{\partial\Phi}{\partial a_k}. \tag{B.2}$$

The analogue of (64) is

$$\Phi=-\frac{1}{2}\sum_{k=1}^{3}\Phi_{ik}\frac{\partial\Phi}{\partial a_k}, \quad \text{for} \quad i=1,2,3, \tag{B.3}$$

where the Φ_{ik}'s denote the elements in the determinant (B.1) without, however, the factor 1/2.
The Analogue of (70) now reads on the Φ-manifold:

$$\sum_{k=1}^{3}\Phi_{ik}P_k=0, \quad i=1,2,3. \tag{B.4}$$

[55] This is previously known to KW, but remained unpublished. My sincere thanks are due Professor KÄLLÉN for his many enlightening discussions on this, and for his kind permission to include it here.

On the envelope E_{123}, we have

$$\left.\begin{array}{cc} \dfrac{\partial \Phi}{\partial a_i} & P_i \\ \dfrac{\partial \Phi}{\partial a_k} & P_k \end{array} = \begin{array}{c} \gamma_i \\ \gamma_k \end{array}\right\} \tag{B.5}$$

where the γ_k's are *real*, such that

$$\left.\begin{array}{l} P_k = \gamma_k \lambda(z) \\ \sum\limits_k \gamma_k = 1 . \end{array}\right\} \tag{B.6}$$

Thus the analogues of (72) are

$$\sum_k (\boldsymbol{Im}\, \Phi_{ik})\, \gamma_k = 0 \tag{B.7}$$

$$\sum_k (\boldsymbol{Re}\, \Phi_{ik})\, \gamma_k = 0 . \tag{B.8}$$

Now from (B.7) follows immediately the analogue of (73):

$$0 = \det |\, \boldsymbol{Im}\, \Phi_{ik} \,| = 2\, y_1 y_2 y_3 . \tag{B.9}$$

In general, for given y_1, $y_2, \neq 0$, (B.9) implies that y_3 must be zero on the 3-mass envelope. Or in other words:

No 3-mass envelope for the Φ-manifold can exist off the real axis.

This is also a horizontal line in the z_3-plane (cf. E_{1234} of (74) in the 4-point case). At this point, one can immediately see that E_{123} is irrelevant: It cannot be relevant on the negative real axis. Then *at most* E_{123} can lie on the positive real axis, which is already the cut.

The following, however, is devoted to an explicit solution to the real part equations (B.8), showing that E_{123} (as well as the 2-mass envelopes discussed below) is actually non-empty, and in one particular configuration (i. e. bubble) the 3-mass and the 2-mass envelopes are rather amusing (cf. Fig. 28).

With $y_3 = 0$, it follows further from (B.7) that

$$\left.\begin{array}{l} \gamma_3 = 0 \\ \dfrac{\gamma_1}{\gamma_2} = \dfrac{-y_1}{y_2} . \end{array}\right\} \tag{B.10}$$

Or, when normalized according to (B.6),

$$\gamma_1 = \frac{y_1}{y_1 - y_2}\,; \quad \gamma_2 = \frac{-y_2}{y_1 - y_2}. \tag{B.10a}$$

With these explicit values of the γ_k's, the real part equations (B.8) yield

$$y_1 a_1 + (-y_2)\, a_2 + (y_1 - y_2)\, a_3 = (x_2 y_1 - x_1 y_2) \tag{B.11}$$

$$y_1^2 a_1 - y_2^2 a_2 = 0 \tag{B.12}$$

$$x_3 = \left(1 - \frac{y_1}{y_2}\right) a_1 + \left(1 - \frac{y_2}{y_1}\right) a_2. \tag{B.13}$$

The path C_{123} in the a_k-space (which would give rise to E_{123}) is then the straight-line intersection of the two planes given by (B.11) and (B.12), within the octant $a_k > 0$. We now divide our discussion into two parts:

Case 1: $y_1 y_2 < 0$ (Bubble configuration).

Without loss of generality, we may take $y_1 > 0$. In this case, (B.11) is compact within the octant $a_k > 0$. Therefore its intersection with (B.12) gives

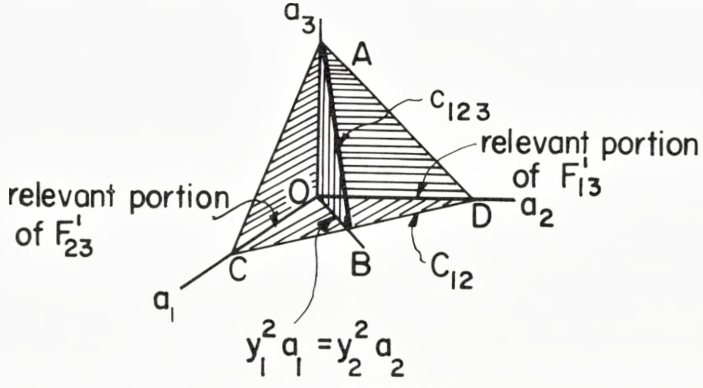

Figure 27. Paths in the a-space for the 3-mass and 2-mass envelopes and the 1-mass curves for the 3-point Φ-manifold: $y_1 y_2 < 0$.

a finite straight-line segment AB (Fig. 27). The image of AB in the z_3-plane is given by (B.13). More explicitly, we have from (B.12) and (B.13)

$$x_3 = \left(1 - \frac{y_1}{y_2}\right)^2 a_1 \tag{B.14}$$

which, together with (B.12), implies that a_1 and a_2 are positive if and only if x_3 is positive. Furthermore, one gets from (B.11)

$$a_3 = \frac{(-y_1 y_2)}{(y_1 - y_2)^2}(x_3^{(0)} - x_3), \tag{B.15}$$

where

$$x_3^{(0)} = \frac{(y_1 - y_2)}{(-y_1 y_2)}(x_2 y_1 - x_1 y_2) \tag{B.16}$$

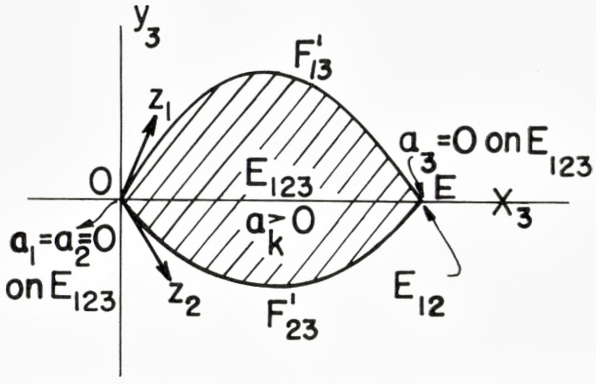

Figure 28. Φ-manifold envelopes for the bubble configuration in the z_3-plane.

is precisely the abscissa of the point E (Fig. 28) which is the common intersection of F'_{13} and F'_{23} with the x_3-cut.
Note that

$$x_3^{(0)} > 0, \quad \text{for} \quad \arg z_2 > \pi + \arg z_1,$$

which is the relevance criterion for the bubble of Fig. 28. From (B.14), (B.12), and (B.15), it is clear now that OE is the image of AB, since all $a_k > 0$ if and only if

$$0 < x_3 < x_3^{(0)}. \tag{B.17}$$

This shows that in the case *when the 3-point boundary is given by the bubble, the 3-mass envelope for the Φ-manifold is actually the segment of the cut on the real axis lying inside the bubble.* It will be shown later that the end point E (where $a_3 = 0$ on the 3-mass envelope) actually constitutes the 2-mass envelope E_{12} for the Φ-manifold in this case.

Case 2: $y_1 y_2 > 0$ (Hyperbola configuration).

In this case, the results become dependent on the ratios of the real and imaginary parts of z_1 and z_2.

(i) when $y_1 = y_2$, we must have also $x_1 = x_2$ as a consequence of (B.11) and (B.12). The allowed region in the a_k-space becomes unbounded, being the whole plane (B.12) within the octant $a_k > 0$ (i. e., $a_1 = a_2$, a_3 arbitrary). The image in the z_3-plane is a single point $x_3 = 0$, viz., E_{123} is at the origin.

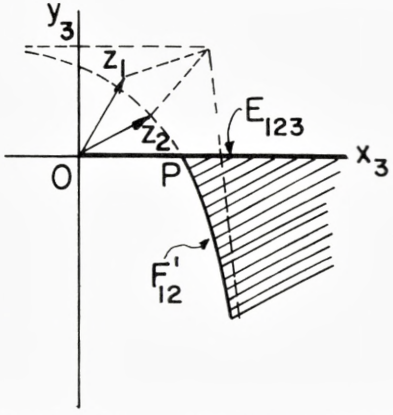

Figure 29. Φ-manifold envelopes in the z_3-plane for the hyperbola configuration: $x_3^{(0)} \leqslant 0$.

(ii) $y_1 \neq y_2$. (B.14) and (B.15) now imply that all $a_k > 0$ if and only if

$$x_3 > \mathrm{Max}\{0,\ x_3^{(0)}\}. \tag{B.18}$$

Thus

(iia) if $x_3^{(0)} \leqslant 0$, E_{123} is the whole cut $x_3 \geqslant 0$. (Fig. 29).

(iib) if $x_3^{(0)} > 0$, E_{123} starts from $x_3 = x_3^{(0)}$. However, this point has no significance for the case $y_1 y_2 > 0$, since the hyperbola F_{12}' (Fig. 30) intersects the real axis at P with

$$x_3^{(P)} = \frac{y_1 y_2 \left[(x_1 - x_2)^2 + (y_1 + y_2)^2\right]}{(y_1 + y_2)(x_1 y_2 + x_2 y_1)}. \tag{B.19}$$

In this case one has both

$$\left.\begin{aligned} &x_3^{(P)} > x_3^{(0)} \\ \text{and} \quad &x_3^{(P)} > 0 \end{aligned}\right\} \tag{B.20}$$

for $\arg z_1 + \arg z_2 < \pi$, which is the criterion for F_{12}' to be relevant.

II. *Two-Mass Envelopes.*

It can be easily seen that the 2-mass envelopes still lie on the cut along the real axis.

We shall only treat E_{12} here with a_3 set equal to zero; for the others the analysis can be easily adapted. The envelope condition reads:

$$\frac{P_1}{P_2} = \frac{\sqrt{R_1}}{\sqrt{R_2}} = \sigma, \quad \text{a real number.} \tag{B.21}$$

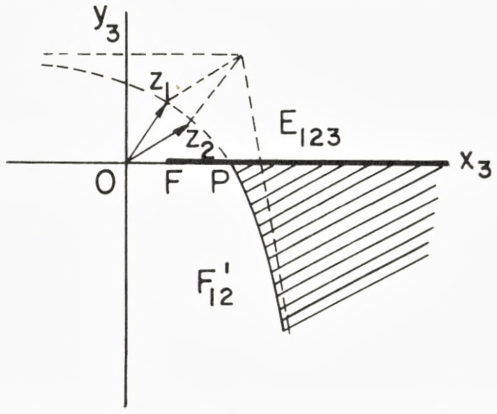

Figure 30. Φ-manifold envelopes in the z_3-plane for the hyperbola configuration: $x_3^{(0)} > 0$.

Case 1: $y_1 y_2 < 0$

(B.12) and (B.14) now no longer hold, however, (B.11) with $a_3 = 0$ is equivalent to (B.21). Furthermore, (B.13), which can be regarded as the equation for the Φ-manifold in this case, is still valid. From these, one gets rather unexpectedly that E_{12} is just a single point at $x_3 = x_3^{(0)}$, (viz., the point E of Fig. 28). Geometrically, in Fig. 27, CD is now the path for E_{12} in the positive quadrant. The entire segment CD is mapped into the point E, which is exactly the end-point $a_3 = 0$ of E_{123}.[56]

In this case, it is interesting to note that the path OC along the a_1-axis and the path OD along the a_2-axis in Fig. 27 map respectively into the rele-

[56] The fact that the path for E_{12} is simply the projection of the plane for E_{123} in 3-space onto the 2-plane must be regarded again as a peculiarity of the 3-point case. This is not true in the 4-point case (cf. Sec. VI), where we have shown that, although the path for E_{1234} is also a straight line in the 4-space, the paths for E_{ijk} and E_{ik} are both not projections, and are very far from being straight lines.

vant portion of F'_{23} and F'_{13} in Fig. 28 (with the common end-point E besides the origin). The occurrence of such multiple intersections of $E_{123} \cap E_{12} \cap E_1 \cap E_2$ must be regarded as a 3-point peculiarity (cf. Fig. 19 and the accompanying remark). In the 4-point case, we have seen, however, that in general we have only the intersection between an m-envelope and an $(m-1)$-envelope (cf. Lemma 3).

Case 2: $y_1 y_2 > 0$

Following E_{123} in this case, and the E_{12} for the above case, we see that E_{12} for this case also consists of a point at $x_3 = x_3^{(0)}$. Now

(i) If $x_3^{(0)} < 0$, E_{12} is irrelevant, and
(ii) If $x_3^{(0)} > 0$, E_{12} is the point F in Fig. 30, which is imbedded in the cut.

Appendix C

Some Algebraic Details for the 4-Mass Envelope

We give here the details for the values of the γ_k's on E_{1234}, and the dependence of their relative signs on the configuration of the y's.

Solving (72a), one gets

$$\frac{\gamma_1\gamma_2}{\gamma_3\gamma_4} = \frac{y_5}{y_1}; \quad \frac{\gamma_1\gamma_3}{\gamma_2\gamma_4} = \frac{y_6}{y_2}; \quad \frac{\gamma_1\gamma_4}{\gamma_2\gamma_3} = \frac{y_4}{y_3} \tag{C.1}$$

or equivalently:

$$\left.\begin{aligned} \frac{\gamma_1}{\gamma_2} &= \frac{y_3 y_4 \pm \sqrt{y_1 y_3 y_4 y_5}}{-y_2 y_3} \\ \frac{\gamma_3}{\gamma_2} &= \frac{y_1 y_5 \pm \sqrt{y_1 y_3 y_4 y_5}}{-y_2 y_5} \\ \frac{\gamma_4}{\gamma_2} &= \frac{\pm \sqrt{y_1 y_3 y_4 y_5}}{y_3 y_5}. \end{aligned}\right\} \tag{C.1a}$$

These may then be normalized according to (67). The $(\pm)$ signs correspond to the sign of $E^{\pm}_{1234}$ in (74). From these, one immediately notes that, for example,

on E^{+}_{1234},

$$\gamma_1\gamma_3 \gtrless 0 \text{ according as } y_1 y_3 \gtrless 0,$$

and

$$\gamma_2\gamma_4 \lessgtr 0 \text{ according as } y_3 y_5 \gtrless 0.$$

The exactly opposite statements hold on E^{-}_{1234}.

We summarize the results in Table 1:

TABLE 1: Relative Signs of γ_k on $E^{\pm}_{1234}$ Versus Configurations

Cases	Configuration of y		Relative Signs of γ_k's	
	Up	Down	On E^{+}_{1234}	On E^{-}_{1234}
I	1, 3, 4, 5	2, 6	all $\gamma_k > 0$	(i) $y_1, y_4 \,\|\|\, y_2, y_3$
				(ii) $y_1, y_2 \,\|\|\, y_3, \gamma_4$
II	1, 2, 5, 6	3, 4	$\gamma_1, \gamma_2 \,\|\|\, \gamma_3, \gamma_4$	(i) $\gamma_2, \gamma_4 \,\|\|\, \gamma_1, \gamma_3$
				(ii) all $\gamma_k > 0$
III	2, 3, 5	1, 4, 6	(i) $\gamma_1 \,\|\|\, \gamma_3, \gamma_2, \gamma_4$	$\gamma_4 \,\|\|\, \gamma_1, \gamma_2, \gamma_3$
			(ii) $\gamma_3 \,\|\|\, \gamma_1, \gamma_2, \gamma_4$	
IV	1, 2, 3	4, 5, 6	(i) $\gamma_2 \,\|\|\, \gamma_4, \gamma_1, \gamma_3$	$\gamma_3 \,\|\|\, \gamma_1, \gamma_2, \gamma_4$
			(ii) $\gamma_4 \,\|\|\, \gamma_2, \gamma_1, \gamma_3$	

Remark: (a) These are the only four distinct configurations of the y's for which $E^{\pm}_{1234}$ exists. The remaining case with all $y_\mu > 0$ is disregarded here, since the 4-mass envelope is entirely irrelevant in this case (cf. remark following (76)). The permutation of (*3, 4*) with (*1, 5*) in case II is trivial. So is the permutation of $(2 \leftrightarrow 6)$ in cases III and IV.

(b) The subdivision into (i) and (ii) is based on

(i) $|y_1 y_5| > |y_3 y_4|$

(ii) $|y_1 y_5| < |y_3 y_4|$, respectively. Note that, when $y_1 y_5 = y_3 y_4$, one of the lines $E^{\pm}_{1234}$ coincides with the cut.

(c) All signs except in the case when all $\gamma_k > 0$ are meant only in a relative sense. Thus we use the double bars to denote that the γ's lying on the same side of the double bar have the same sign, while any two γ's lying on the opposite sides of the double bar have opposite signs.

(d) The above results can be briefly stated as follows:

(1) When the signs of the 6 y's break into 4 || 2, the signs of the 4 γ's break into 4 || 0, or 2 || 2.

(2) When the signs of the y's break into 3 || 3, then those of the γ's break into 3 || 1.

With Table 1, one can readily infer from (77) or (80) the signs of the a_k's on the 4-mass envelope at $x_6 \to \pm\infty$. For a_1 and a_3, no other information is needed; however, for a_2 and a_4, there is a further dependence on the magnitude of γ_2 and γ_4 (when the latter are positive). Table 2 illustrates the situation for $z_6 \to -\infty$. Exactly opposite statements hold for the signs of the a_k at the other end $x_6 \to +\infty$.

TABLE 2: The Signs of a_k on $E^{\pm}_{1234}$ at $x_6 \to -\infty$.

Cases*		a_1 and a_3		a_2		a_4	
		On E^+_{1234}	On E^-_{1234}	On E^+_{1234}	On E^-_{1234}	E^+_{1234}	E^-_{1234}
I	(i)	+	−	−		−	
	(ii)	+	−	−		−	
II	(i)	−	+				
	(ii)	−	+		−		−
III	(i)	+	−	$\mp(\gamma_2 \gtrless 1)$	$\pm(\gamma_2 \gtrless 1)$	$\mp(\gamma_4 \lesseqgtr 1)$	−
	(ii)	+	−	$\mp(\gamma_2 \lesseqgtr 1)$	$\pm(\gamma_2 \gtrless 1)$	$\mp(\gamma_4 \lesseqgtr 1)$	−
IV	(i)	−	+	−	$\mp(\gamma_2 \gtrless 1)$	$\pm(\gamma_4 \lesseqgtr 1)$	$\mp(\gamma_4 \gtrless 1)$
	(ii)	−	+	$\pm(\gamma_2 \lesseqgtr 1)$	$\mp(\gamma_2 \gtrless 1)$	−	$\mp(\gamma_4 \lesseqgtr 1)$

* For the cases III and IV in Table 2, the signs of 4 γ's break into $3 \| 1$. Table 2 assumes that 3 γ's > 0 and one $\gamma < 0$.

The remainder of this appendix is devoted to the discussion of the case when the (all a_k positive) segment E_{1234} has an intersection with the set ω_x of (82). For this, it will be convenient to divide the discussion into the following two classes of configurations:

(1) *All γ_k positive.*

In this case, we have

$$0 < \gamma_k < 1\,, \quad \Sigma \gamma_k = 1\,. \tag{C.2}$$

From Table 1, we see that this happens only for the following two configurations (Figs. 31–32). We recall from Table 1 that all $\gamma_k > 0$ hold for the configuration (Fig. 32) only for $y_1 y_5 < y_3 y_4$ (otherwise 2 of the γ's become nega-

tive). When $y_1 y_5 - y_3 y_4 \to 0$, the line E^-_{1234} collapses into the cut on the real x_6-axis. On E_{1234}, we have, in general, by virtue of (72):

$$\left.\begin{aligned} \sum_{i,j} (\boldsymbol{Re}\, \Psi_{ij})\, \gamma_i \gamma_j &= 0 \\ \sum_{i,j} (\boldsymbol{Im}\, \Psi_{ij})\, \gamma_i \gamma_j &= 0. \end{aligned}\right\} \qquad \text{(C.3)}$$

Now the γ's of (C.2) may just be identified as playing the same role as our original integration variables α_k's. Therefore for this case, the denominator D

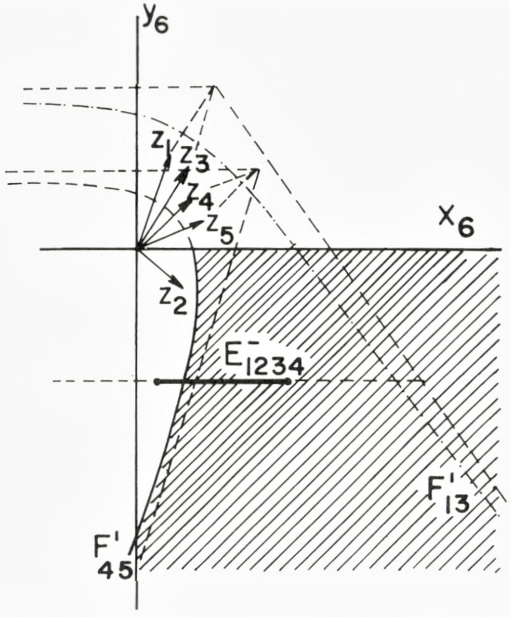

Figure 31. All $\gamma_k' > 0$ on the 4-mass envelope for the configuration (I) of Table 1.

of (15) will indeed vanish identically on E_{1234} (where all a_k's are positive). When this segment has an intersection with the Φ_k-manifolds, part of it will have actual singularities.

One observes from Table 2 that E_{1234} are finite for both of these configurations, since two of the a's (viz., a_2, a_4) are negative at $x_6 \to -\infty$, and the other two (viz. a_1, a_3) are negative at the other end ($x_6 \to +\infty$).

(2) *Not all γ_k's positive*:

In this case, identification of γ_k with α_k is not possible, thus (C.3) do not automatically imply that (16) will vanish on E_{1234}. In fact, it can be easily seen that $\boldsymbol{Re}\, D$ never vanishes for $x \varepsilon \omega_x$ of (82).

One notes from (16), after the substitution $\alpha_4 = 1 - \sum_{i=1}^{3} \alpha_i$,

$$\left.\begin{aligned} -\boldsymbol{Re}\, D) &= -\frac{1}{2}\sum_{i,j}(\boldsymbol{Re}\,\Psi_{ij})\,\alpha_i\alpha_j = \\ &= \left[x_6(\alpha_2+\alpha_2^0)^2 - \frac{\lambda_1(x)}{4x_6}(\alpha_3+\alpha_3^0)^2 + \frac{\Lambda(x)}{\lambda_1(x)}(\alpha_1+\alpha_1^0)^2 - \frac{\Psi(x;a)}{4\Lambda(x)}\right] \end{aligned}\right\} \quad (C.4)$$

where

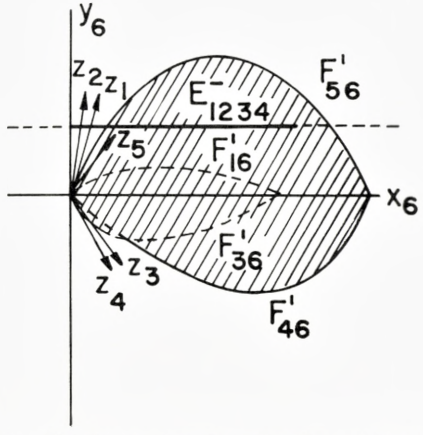

Figure 32. All γ_k's > 0 on the 4-mass envelope for the configuration (II) of Table 1.

$$\left.\begin{aligned} \alpha_2^0 &= -\frac{1}{2x_6}\left[\frac{\partial^2\Lambda(x)}{\partial x_1\partial x_2}\alpha_3 + \frac{\partial^2\Lambda(x)}{\partial x_4\partial x_2}\alpha_1 - \frac{\partial^2\Phi_1(x;a)}{\partial x_4\partial a_3}\right] \\ \alpha_3^0 &= \frac{1}{\lambda_1(x)}\left[\frac{\partial\Lambda(x)}{\partial x_2}\alpha_1 - \frac{\partial\Phi_1(x;a)}{\partial a_3}\right] \quad \text{(Cf. Eq. (35))}; \\ \alpha_1^0 &= \frac{Q_1(x;a)}{4\Lambda(x)} \end{aligned}\right\} \quad (C.5)$$

and $\Psi(x;a) = \det|\boldsymbol{Re}\,\Psi_{ji}|$
which vanishes identically on E_{1234}. Thus we see that (C.4) is positive definite for $x \varepsilon \omega_x$, unless simultaneously

and

$$\left.\begin{aligned} \alpha_i &= -\alpha_i^0, \quad \text{for} \quad i = 1, 2, 3 \\ \alpha_i^0 &< 0\,, \ \sum_i(-\alpha_i^0) < 1. \end{aligned}\right\} \quad (C.6)$$

It is now easy to see that (C.6) cannot happen when at least one of the γ_k's is negative. We have on E_{1234}, after treating the det $|\boldsymbol{Re}\,\Psi_{ij}|$ with exactly the same procedure which led to (64),

$$\gamma_k = -\frac{Q_k(x;\,a)}{4\,\Lambda(x)}. \tag{C.7}$$

Now, without loss of generality, we may take[57] $\gamma_1 < 0$. Then (C.5) implies that $\alpha_1^0 = -\gamma_1 > 0$, and (C.6) clearly cannot happen. Thus for all cases with γ_k not simultaneously positive, $-\boldsymbol{Re}\,D$ is positive definite on $\omega_x \cap E_{1234}$, and it follows that this portion of the 4-mass envelope can never be a relevant part of the boundary.

For completeness, we note the following identity *on the 4-mass envelope*:

$$2\,\boldsymbol{Re}\,D = \sum_{i,j} \boldsymbol{Re}\,\Psi_{ij}\,\alpha_i\,\alpha_j = \sum_{i,j} X_{ij}\,(\alpha_i - \gamma_i)\,(a_j - \gamma_j) \tag{C.8}$$

which can be easily verified with the aid of (77).

Appendix D

Note on the Determinant Expansion

We here observe that a great number of identities which have played an essential role in our preceding discussion, such as (43), (44), (45), (100), and (110), have a most natural interpretation in terms of their associated determinants. Take, for example, (43), which reads:

$$\left(\frac{1}{2}\frac{\partial \Psi}{\partial a_k}\right)^2 = 4\,\Lambda(z)\,\Phi_k + \lambda_k\,\Psi(z;\,a). \tag{D.1}$$

Recalling the quantities following (57a), we have, for $k = 2$

$$\left.\begin{aligned} \frac{1}{2}\frac{\partial \Psi}{\partial a_2} &= \tilde{\Psi}^{12} \\ 2\,\Lambda(z) &= \tilde{\Psi}^{11} \\ 2\,\Phi_2 &= \tilde{\Psi}^{22} \\ -\lambda_2 &= \tilde{\Psi}^{12,\,12} \end{aligned}\right\} \tag{D.2}$$

[57] Otherwise, a trivial permutation will bring (C.4) into the form where the last α_k^0 corresponds to the desired negative γ_k.

where $\tilde{\Psi}^{12,12}$ refers to the minor complementary to the 2×2 minor

$$\begin{vmatrix} \tilde{\Psi}_{11} & \tilde{\Psi}_{12} \\ \tilde{\Psi}_{21} & \tilde{\Psi}_{22} \end{vmatrix}$$

in the $\tilde{\Psi}$-determinant of (57a). Then (D.1) takes the form

$$\left.\begin{aligned} \Psi = \tilde{\Psi} = \frac{\begin{vmatrix} \tilde{\Psi}^{11} & \tilde{\Psi}^{12} \\ \tilde{\Psi}^{21} & \tilde{\Psi}^{22} \end{vmatrix}}{\begin{vmatrix} \tilde{\Psi}_{33} & \tilde{\Psi}_{34} \\ \tilde{\Psi}_{43} & \tilde{\Psi}_{44} \end{vmatrix}} \end{aligned}\right\} \qquad \text{(D.3)}$$

Now identity (D.3) can be easily verified to hold for a general 4×4 determinants Thus (D.1) is established for $k = 2$, and by symmetry the others follow. At this point, the corresponding identities for the 3-point case (KW (A46d)) are seen to be also derivable from such a determinant expansion.

Ir appears, however, that identities of the form (D.3) are actually very special cases of a general theorem, which, in various forms, has been dated back to Gauss (also for symmetric determinants) and others. We shall here quote a theorem due to Jacobi[58], which states that

Any minor of order k in A^{-1} *is equal to the complementary signed minor in* A' *(the adjoint of* A*), multiplied by* $|A|^{-1}$.

In other words, this technique of determinant expansion relates the block I in (D.4) with the block II in (D.5), their determinants being off by a factor of the original determinant:

$$A^{-1}: \qquad \begin{matrix} k \\ \left\{ \right. \end{matrix} \left(\begin{array}{c|c} \overbrace{\quad\quad}^{k} & \vdots \\ \text{I} & \vdots \\ \hline \cdots\cdots\cdots\cdots & \\ \vdots & \\ \vdots & \end{array} \right) \qquad \text{(D.4)}$$

[58] See, e. g., an elementary text by A. C. Aitken, *Determinants and Matrices*, 3rd ed., Edinburgh (1944).

$$A': \qquad \left(\begin{array}{c|c} & \\ \cdots\cdots\cdots\cdots\cdots & \\ & \\ \hline & \mathrm{II} \\ & \overbrace{\quad}^{(n-k)} \end{array} \right) \Big\} (n-k) \qquad (\mathrm{D}.5)$$

It is then a simple matter to derive all the identities we mentioned by simply writing down the desired $k \times k$ minors in this fashion.

Institute of Theoretical Physics, University of Lund, Lund, Sweden – and
Institute for Theoretical Physics, Swiss Federal Institute of Technology, Zurich, Switzerland.
Work supported in part by the U.S. National Science Foundation and U.S. Air Force Office of Scientific Research, Air Research and Development Command, under Contract AF 49 (638)–24 at the University of Maryland, College Park, Maryland.
Present address: Institute for Advanced Study, Princeton, New Jersey.

Indleveret til Selskabet den 21. oktober 1960.
Færdig fra trykkeriet den 25. august 1961.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 4

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 4 (1961)

THE STRUCTURE OF BARIUM BROMIDE DIHYDRATE

BY

EVA BANG

København 1961
i kommission hos Ejnar Munksgaard

Synopsis

$BaBr_2, 2H_2O$ belongs to the space group $C2/c$. The unit cell contains 4 molecules. $a = 10.44_9$ Å, $b = 7.20_4$ Å, $c = 8.38_5$ Å, $\beta = 113° 29\frac{1}{2}'$. X-ray analysis by 2-dimensional Fourier methods shows $[Ba(H_2O)_2]_\infty$ columns in the direction of the c-axis, placed in between buckled layers of bromine atoms. The structure shows features reminding of $SrCl_2, 6H_2O$. The shortest Ba–Ba distance is 4.44 Å, only a little larger than found for the element. The distances Ba–Br are 3.41 Å, and 3.49–3.52 Å. The distances Ba–H_2O are 2.82 Å and 3.00 Å and Br–H_2O are 3.25–3.33 Å and 3.60 Å.

Printed in Denmark
Bianco Lunos Bogtrykkeri A-S

Introduction

Very few detailed structure determinations on hydrated simple metal halides have been reported in the literature: three fluorides, eight chlorides, and one bromide. Five belong to the group of alkaline earth halogenides: $BaCl_2, H_2O$, $BaBr_2, H_2O$, $SrCl_2, 6\,H_2O$, $BaCl_2, 2\,H_2O$ and $SrCl_2, 2\,H_2O$.

The monohydrates, investigated by electron diffraction[1, 2] are isomorphous (Pmcn). They contain zig-zag chains of $\overset{H_2O}{\diagup}\underset{Ba}{\diagdown}\overset{H_2O}{\diagup}\underset{Ba}{\diagdown}\,Ba,$ where barium has 2 water molecules and 7 halogen atoms and the water molecule has 2 barium and 7 halogen atoms as nearest neighbours. The structure consists of a net of halogen atoms where barium and water alternate in the holes.

In $SrCl_2, 6\,H_2O$[3] (P321), we get a one-dimensional lattice complex of $[Sr(H_2O)_6]_\infty$, where strontium has 9 water molecules as nearest neighbours and no chlorine. There are 2 types of water molecules, one has 2 chlorine and 2 barium, the other has 3 chlorine and 1 barium as nearest neighbours.

The dihydrates $BaCl_2, 2\,H_2O$[4] ($P2_1/n$) and $SrCl_2, 2\,H_2O$[5] ($C2/c$) contain infinite two-dimensional lattice complexes, where the cation has 4 chlorine and 4 water molecules as nearest neighbours in slightly different arrangements.

To get more information about the behaviour of the water and the structural differences in the hydrates of the alkaline earth halogenides the structure of $BaBr_2, 2\,H_2O$ has been investigated.

The crystal class of $BaBr_2, 2\,H_2O$ was shown to be 2/m by O. Mügge[7, 8] and H. Dufet,[9] and the crystals to be optically positive. They can be obtained by slow evaporation at room temperature and very often appear from the same batch with quite different habits: plates, rods, or needles after [001] with parallel extinction, needles after $[1\bar{1}0]$ with oblique extinction. Twins occur very often and are difficult to recognize. The needles after $[1\bar{1}0]$ usually are not twinned. The morphology has been very carefully described by O. Mügge,[7] who used the deformation by pressure into twins to show

that the crystal class was monoclinic and not orthorhombic as suggested by C. RAMMELSBERG and others.[10, 11]

The powder diagrams of the 3 dihydrates mentioned above show that $BaBr_2, 2H_2O$ is neither isomorphous with $BaCl_2, 2H_2O$ nor with $SrCl_2, 2H_2O$.

X-Ray Examination, Unit Cell, and Space Group

Preliminary axes were taken from oscillation and Weissenberg diagrams (camera 57.3 mm) and refined from powder diagrams (Table 1). The powder diagrams have been obtained by means of a Bradley 19 cm camera and a Guinier type focusing camera with Cu radiation.

The unit cell dimensions are

$$a = 10.44_9 \pm .03 \text{ Å} \quad b = 7.20_4 \pm .02 \text{ Å} \quad c = 8.38_5 \pm .02 \text{ Å},$$
$$\beta = 113.49 \pm .3°.$$

The angle between [100] and [102] is 90.18°.

The axial ratio $a : b : c$ calculated 1.450 : 1 : 1.164.

The axial ratio $a : b : c$ given by O. MÜGGE[7] 1.44943 : 1 : 1.16559.

β angle (supplemental angle of β given above) from O. MÜGGE[7] 66° $30\frac{1}{2}'$.

The density $3.87_{2\ 18°}$[12] gives 4.05 molecules pr. unit cell.

The Weissenberg diagrams were taken by multiple film technique, the intensities estimated visually. As the crystals easily are deformed by mechanical pressure no attempts to cut a small fragment have been made. No corrections for absorption and temperature factor have been applied as only 0-layer lines have been used for the final calculations. The intensities from rotation round the b axis should be less accurate, as it is very difficult to find a crystal with a suitable cross-section in the a–c plane. The usual corrections for polarisation and Lorentz factor have been applied.

With the monoclinic setting given above, reflections were present only for $h + k = 2n$ and h0l for $l = 2n$. $|F(hkl)| = |F(h\bar{k}l)|$. The possible space groups are Cc or $C2/c$.

Investigations into piezoelectricity for the crystals gave a negative result.* It should, however, be mentioned that piezo- and pyroelectricity have been reported in a paper from 1897.[13] The possibility of a lack of symmetry centre will be discussed later in connection with the Patterson and electron density projections.

* Kindly performed by V. FRANK, dynamical method.

TABLE 1. Comparison of observed and calculated $\sin^2\theta$ values.

Indices	Estim. Int.	10^4 $\sin^2\theta$ obs.	10^4 $\sin^2\theta$ calc.
110	wm	0178	0179
$11\bar{1}$	wm	0216	0216
200		0259	0259
111	vvw	0345	0344
002	m	0403	0402
$20\bar{2}$			0404
$11\bar{2}$	vw br	0458	0453
020			0458
021	vvw	0559	0558
$31\bar{1}$			0605
310	vvw?	0688	0689
$22\bar{1}$	wm	0700	0697
112			0710
$31\bar{2}$	vs	0715	0714
220			0717
022	m	0861	0860
$22\bar{2}$			0862
$11\bar{3}$	m	0891	0891
202			0918
$40\bar{2}$			0923
221	wm	0948	0946
311	wm	0996	0990
$31\bar{3}$	vw	1024	1023
400	ms	1039	1036
130	ms	1093	1096
$13\bar{1}$	m	1133	1132
$22\bar{3}$	vvw	1235	1236
131	vvw	1263	1261
113			1277

Indices	Estim. Int.	10^4 $\sin^2\theta$ obs.	10^4 $\sin^2\theta$ calc.
$42\bar{1}$	dif	1341	1337
$42\bar{2}$			1353
222			1363
$13\bar{2}$	dif	1370	1369
023			1376
$20\bar{4}$			1382
420			1485
$51\bar{2}$	vw	1490	1492
312	w	1497	1494
$51\bar{1}$	w	1512	1512
$33\bar{1}$			1521
$31\bar{4}$	vw?	1529	1530
$11\bar{4}$	w	1536	1534
330			1608
004			1614
$40\bar{4}$			1615
$33\bar{2}$			1627
$42\bar{3}$	wm	1630	1627
$51\bar{3}$			1630
132	vw	1675	1673
510			1733
$13\bar{3}$			1808
224	wm	1812	1811
040			1833
$42\bar{1}$			1852
331	vw	1910	1907
041			1934
$33\bar{3}$			1940
402			1952
$60\bar{2}$	m	1960	1961
223	vw	2010	2007
114			2044
$51\bar{4}$	ms	2054	2055

The intensities applied in the calculations were taken from Weissenberg diagrams, where the crystals had the following dimensions:

Length along the rotation axis [001] .2 mm cross section .01 × .02 mm².
– – – – – [010] .15 mm – – .15 × .3 mm².
– – – – – [$1\bar{1}0$] .5 mm – – .1. × .1 mm².

Patterson and Electron Projections

The Patterson and electron projections were calculated on a Hägg-Laurent-Frank machine.[14] The results for the Patterson function projected on (001) and (010) by using observed $|F(hkl)|^2$ values are shown in figs. 1 and 2.

In case we have the space group $C2/c$ (No. 15 Int. Tab.) there are 8 atoms in the general position:

$$(0, 0, 0;\ \tfrac{1}{2}, \tfrac{1}{2}, 0) + x, y, z;\ \bar{x}, \bar{y}, \bar{z};\ \bar{x}, y, \tfrac{1}{2} - z;\ x, \bar{y}, \tfrac{1}{2} + z.$$

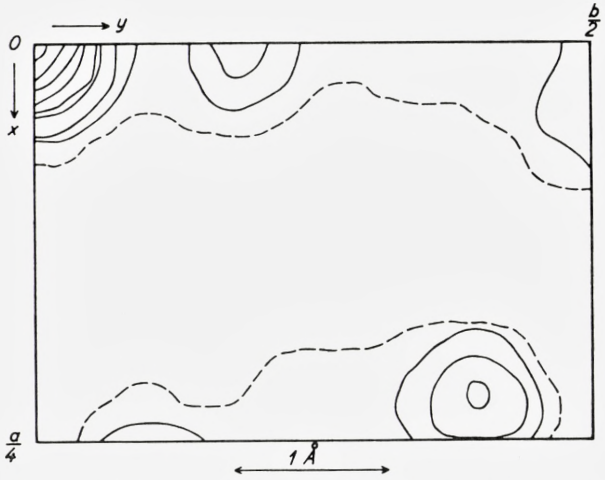

Fig. 1. Patterson projection on (001). Relative arbitrary scale.

The origin is at the symmetry centre of the glide plane.

The 4 Ba must lie in a special position. The positions denoted as c and d can be ruled out, as they both would give rise to a Ba–Ba vector peak in $0, \frac{1}{2}, \frac{1}{2}$, where the vector density is small for the [001] projection.

Assuming Ba at a symmetry centre a: 0, 0, 0 etc., or b: $0, \frac{1}{2}, 0$, etc., we should expect no separate Ba–Ba and 16 Ba–Br vector peaks if Br is in the general position (8 in the (001) projection and 8 in the (010) projection). In case the Ba position is on the twofold axis e: 0, y, 1/4, etc., we should get 4 separate Ba–Ba and 16 Ba–Br vector peaks (16 in the (001) projection and 8 in the (010) projection).

In the projection on (010) the Ba–Ba vector peaks are bound to fall in $0, 0;\ 0, \frac{1}{2}$, etc. The strong peaks outside the axes must be the Ba–Br vector

peaks. If Ba has the x, z coordinates 0, 0, the x, z coordinates of Br can be evaluated (.21, .10), and thus most of the signs of the structure factors. In case Ba is in e: 0, y, 1/4, etc., the resulting projection will be the same except for a translation 1/4 c.

The Patterson projection on (001) shows 8 strong peaks outside the axes

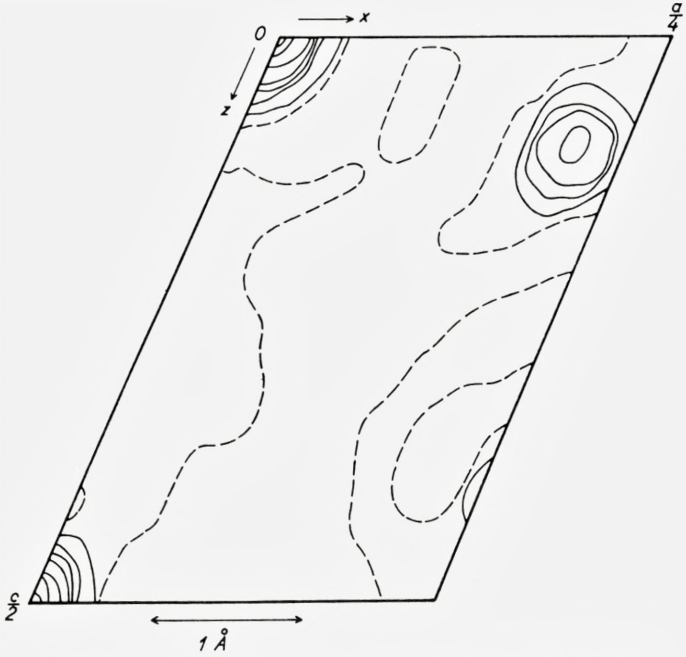

Fig. 2. Patterson projection on (010). Relative arbitrary scale.

which should be taken as Ba–Br vector peaks. In case we have Ba in 0, y, 1/4 the coordinates of the Ba and the Br atoms must have a correspondence that gives rise to superposed Ba–Br vector peaks in this projection.

The 4 Ba–Ba vector peaks in this case have the coordinates 0, $2y$, $\frac{1}{2}$, etc., and if the strongest peak outside the origin on the b axis is taken as a Ba–Ba vector peak ($y_{\mathrm{Ba}} = .1$ and $y_{\mathrm{Br}} = \frac{1}{2}$), we should only get 8 Ba–Br vector peaks and the coordinates of the Ba–Ba and the Br–Br vector peaks would correspond. On this assumption most of the signs could be evaluated and the electron density projection was calculated. This interpretation of the Patterson projection later turned out to lead to the correct structure.

TABLE 2. Atomic parameters in $BaBr_2$, $2\,H_2O$.

	x	y	z
Ba	0	.101	.250
Br	.218	.476	.350
O	.01	.748	.445

As another interpretation of the Patterson projection on (001) the coordinates for the Ba atoms could be 0, 0, 0; etc. or 0, 0, 1/4, etc. The x, y coordi-

Fig. 3. Patterson projection on $(1\bar{1}0)$. Relative arbitrary scale.

nates for the Br atom were evaluated (.22, .40) and thus most of the signs of the structure factors. The calculated electron density projection and the agreement between observed and calculated F values were not bad. (R = .3). It will be shown later that this structure was wrong.

In case there is no symmetry centre the space group is Cc, as mentioned above. The number of atoms in the general position is 4 and the coordinates are

$$(0, 0, 0; \tfrac{1}{2}, \tfrac{1}{2}, 0;) + x, y, z; \quad x, \bar{y}, \tfrac{1}{2} + z.$$

The origin is on the glide plane.

The 8 Br atoms must be coordinated in such a way that they give rise to only 8 Ba–Br vector peaks in the (010) and (001) Patterson projections. The combination of the 2 sets: x, y, z, etc., and $\bar{x}$, y, $\bar{z}$, etc., and the same relation between the Ba and Br y-coordinates as in the case C2/c fulfil this condition. The projections of this structure on (010) and (001) are the same as in the centrosymmetrical case and possible preliminary coordinates could be taken from the results mentioned above.

The projection of the structure along the $[1\bar{1}0]$ axis is noncentrosymmetrical. It turned out, however, that electron projections carried out for both space groups were of little help in distinguishing between the different possibilities at this stage of the structure investigations. The data were taken from a crystal needle showing oblique extinction. It was possible to coordinate the reciprocal net and the Patterson projection on $(1\bar{1}0)$ (fig. 3) throughout with the data from the crystals used before.

Only C2/c Ba in 0, 0, 0, etc., was cancelled out. The possible position of Ba is then 0, y, 1/4, etc., and of Br .22, y, .35, etc. The coordinates $y_{Ba} = 0.1$ and $y_{Br} = 0.5$ would be the most probable from a physical point of view as the Br–Br distance would be 3.8–4.2 Å, whereas $y_{Ba} = 0.0$ and $y_{Br} = 0.4$ would give a very short Br–Br distance 3.1–3.4 Å for both space groups. In the layer structure of $AlBr_3$ the shortest Br–Br distances are 3.59 Å[16]. r_{Br-} is 1.95–1.96 Å (PAULING–GOLDSCHMIDT).

A generalized projection on (010) made it possible to distinguish between the cases mentioned above.

Generalized Projection on (010)

In the generalized projection the electron density function $\varrho\,(x, y, z)$ is modified by a weighting function $e^{2\pi i L z}$, where L is a constant value of Index 1. Examples and theory are given in the monographs by H. LIPSON and W. COCHRAN[17] and by M. BUERGER.[18]

The generalized projection on (010) can be written

$$\varrho_K\,(x, z) = \int_0^1 \varrho\,(x, y, z)\, e^{2\pi i K y}\, b\,dz)$$

or

$$\varrho_K\,(x, z) = C_K\,(x, z) + iS\,(x, z),$$

where

$$C_K(x, z) = \frac{b}{V} \Sigma_h \Sigma_1 A(hKl) \cos 2\pi (hx + lz) + B(hKl) \sin 2\pi (hx + lz),$$

$$S_K(x, z) = \frac{b}{V} \Sigma_h \Sigma_1 B(hKl) \cos 2\pi (hx + lz) - A(hKl) \sin 2\pi (hx + lz).$$

The real part $C_K(x, z)$ should approximately give the projection of the structure on (010) with the electron density of the nth atom multiplied by

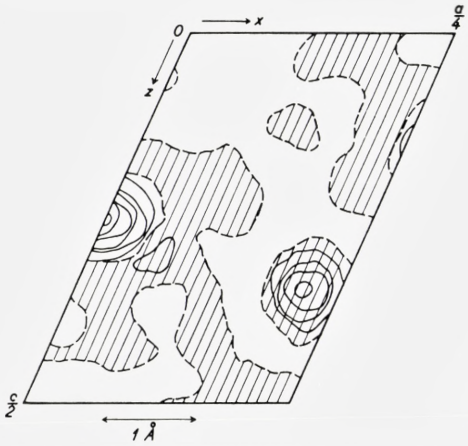

Fig. 4. $C(x, z)$

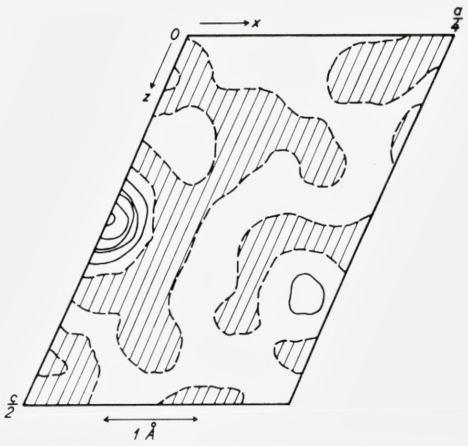

Fig. 5. $S(x, z)$

Fig. 4 and fig. 5. Generalized projections $C(x, z)$ and $S(x, z)$ on (010). $K = 1$. Negative areas are hatched. Relative arbitrary scale.

$\cos 2\pi$ Ky and the imaginary part $S_K(x, z)$ the electron density multiplied by $\sin 2\pi$ Ky.

As mentioned above, the possible position of Ba was $0, y, 1/4$ and of Br $.22, y, .35$. The possible y-coordinates for Ba and Br were $y_{Ba} = 0.0$, $y_{Br} = 0.4$ or $y_{Ba} = 0.1$, $y_{Br} = 0.5$.

The projections were carried out for $K = 1$ ($C\,2/c$). Owing to the symmetry of the space group the calculation work was small. The calculated $S_K(x, z)$ and $C_K(x, z)$ clearly showed that $y_{Ba} = 0.0$, $y_{Br} = 0.4$ was ruled out, as the $S_K(x, z)$ projection in this case was quite confused. Only the projections for $y_{Ba} = 0.1$, $y_{Br} = 0.5$ are given here (figs. 4 and 5; cf. fig. 7). The $S(x, z)$ projection shows a small bromine peak which indicates that y_{Br} is a little different from 0.5.

This information is in agreement with the Patterson projection along $(1\bar{1}0)$, but could not be obtained from it on account of overlapping peaks.

Position of the Water. Refinement of the Structure

The structure found was a layer structure with Ba in the ac-plane on the twofold axes and a bromine layer in between. As the Patterson projection on

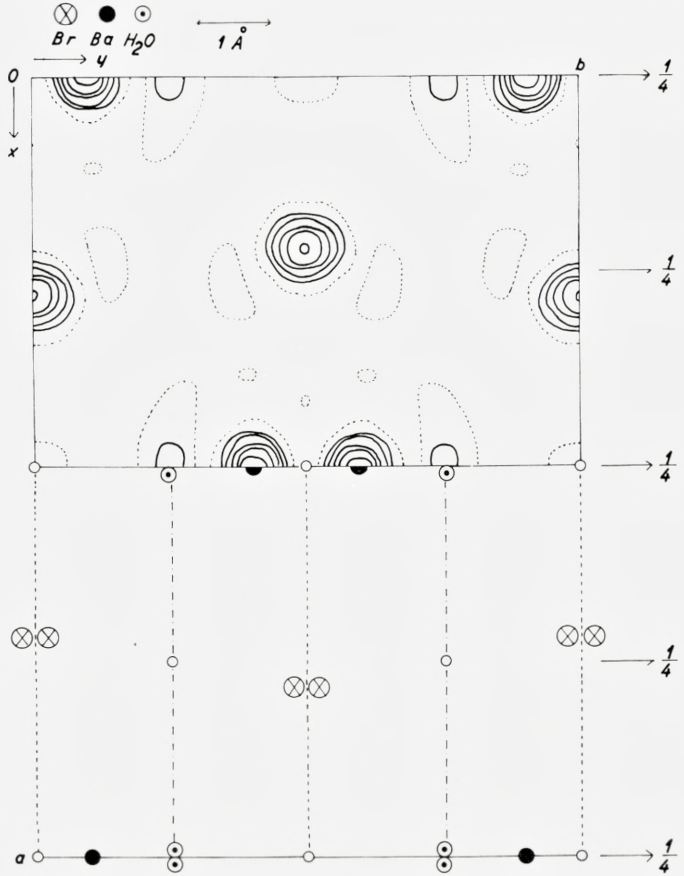

Fig. 6. Electron projection on (001). Relative arbitrary scale for electron density.

(001) showed vector density along the yz planes and as the water molecules from a packing point of view cannot be placed near the bromine layer, the next highest peak outside the origin on the y-axis was taken as a Ba-0 vector peak. The y_0 could easily be deduced. The x_0 should be very near to 0.0 and z_0 was taken to 0.5 from a geometrical consideration together with the electron density projection and the 2 generalized projections on (010).

The refining of the structure was carried out by difference syntheses using $F_{obs.} - F_{calc.}$ for 0-layer line projections along [001], [010] and [1$\bar{1}$0]. The calculated and observed structure factors were scaled for each projection by plotting $\log_{10} \frac{F_{obs.}}{F_{calc.}}$ against $\sin^2\theta$. The resulting coordinates are given in

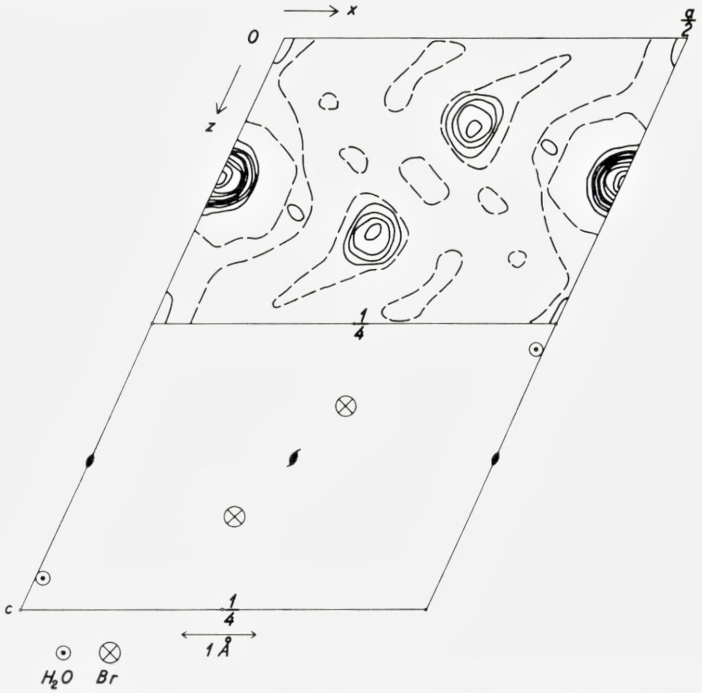

Fig. 7. Electron projection on (010). Ba on the twofold axis. Relative arbitrary scale for electron density.

Table 2 and the calculated and observed F-values in Table 3–7. The electron projections are given in figs. 6–8.

Neither further refining nor a calculation for the noncentrosymmetrical case $y_{Br} = 0.5$ should be made unless better experimental data have been obtained.

Description of the Structure

The interatomic distances between nearest neighbours are given in Table 7.

Each Ba atom is surrounded by 2 Ba, 6 Br, and 4 H_2O (fig. 9). The 2 Ba

atoms have a separation a little larger than found for the element. The 4 water molecules are nearly lying in the plane of the Ba atoms. 2 have the shortest distance, 2 are a little farther away, the difference being of the same magnitude as found for the 2 types of water molecules in $SrCl_2$, $6H_2O$[3]. Four Br are lying on one side of the Ba atom nearly in a plane perpendicular

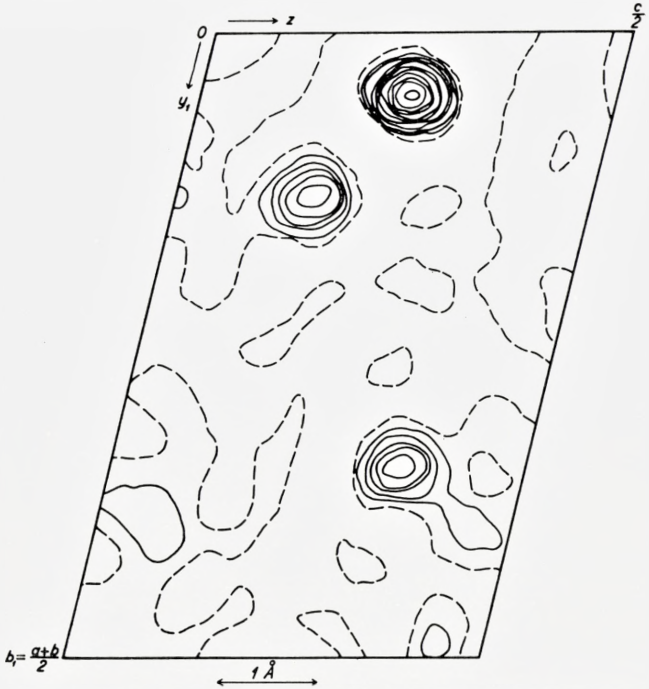

Fig. 8. Electron projection on $(1\bar{1}0)$. Relative arbitrary scale for electron density.

to the plane of the Ba atoms in a slightly longer distance than contact. 2 Br are on the other side of the Ba atom perpendicular to the plane of Ba atoms and the plane of the 4 Br atoms.

The water molecule has 2 Ba and 5 Br atoms as nearest neighbours (fig. 10). The size and the difference in distance to the Ba atoms are mentioned above. Two of the Br atoms are in distance of contact on one side of the Ba–H_2O–Ba group, 3 on the opposite side, 2 having distance of contact, 1 being a little farther away.

The Br atom is surrounded by 3 Ba, 5 H_2O, and 1 Br. One Ba has distance of contact, 2 are a little farther away. Four H_2O are tetrahedrally

TABLE 3. Comparison of calculated and observed structure factors. Rotation [001] 0-layer line.

h	k	F_{calc}	$\lvert F_{obs} \rvert$	h	k	F_{calc}	$\lvert F_{obs} \rvert$
2	0	8	13	9	3	$\overline{38}$	30
4	0	88	103	0	4	15	13
6	0	29	31	2	4	$\overline{58}$	58
8	0	36	36	4	4	1	
10	0	47	40	6	4	$\overline{32}$	32
1	1	29	31	8	4	$\overline{22}$	22
3	1	69	59	1	5	41	45
5	1	8	10	3	5	$\overline{18}$	22
7	1	66	52	5	5	54	47
9	1	$\overline{9}$		7	5	$\overline{6}$	
11	1	45	31	0	6	$\overline{8}$	15
0	2	57	42	2	6	$\overline{51}$	56
2	2	$\overline{47}$	51	4	6	$\overline{13}$	17
4	2	35	37	6	6	$\overline{34}$	30
6	2	$\overline{9}$	9	1	7	$\overline{11}$	10
8	2	3		3	7	2	
10	2	15	16	5	7	$\overline{20}$	14
1	3	$\overline{23}$	26	0	8	25	17
3	3	12	10	2	8	2	
5	3	$\overline{45}$	43	4	8	20	11
7	3	22	23	1	9	16	19

coordinated round the Br atom at a distance of contact, 1 H_2O at a slightly greater distance. The Br atom has a distance slightly greater than twice the ionic radius.

The structure could be described as consisting of flat $[Ba(H_2O)_2]_\infty$ columns parallel to the *c*-axis having the Ba atoms at a distance slightly greater than found for the element. The distance between the Sr atoms in the $[Sr(H_2O)_6]_\infty$ complex in $SrCl_2, 6H_2O$ is of the same size. Each $[Ba(H_2O)_2]_\infty$ column is surrounded by 6 parallel Br-columns (fig. 11). The Br in the columns form a row just a little buckled, with the Br–Br distance within the same size as the Cl–Cl distance in the Cl columns. Each Br contacts 3 $[Ba(H_2O)_2]_\infty$ columns through 1 Ba contact to one column, and $2\,H_2O$ contacts to each of the other 2 columns. There is a difference from $SrCl_2, 6H_2O$, where Cl only has water contacts to the $[Sr(H_2O)_6]_\infty$ complex.

Another one-dimensional complex $[Ba(H_2O)]_\infty$ occurs in the monohydrates $BaCl_2, H_2O$[1] and $BaBr_2, H_2O$[2] as zig-zag lines of Ba–H_2O–Ba–H_2O-

TABLE 4. Comparison of calculated and observed structure factors. Rotation [010] 0-layer line.

h	l	F_{calc}	$\lvert F_{obs} \rvert$	h	l	F_{calc}	$\lvert F_{obs} \rvert$
2	$\bar{10}$	1	8	2	$\bar{2}$	$\bar{46}$	35
4	$\bar{10}$	$\bar{50}$	37	4	$\bar{2}$	$\bar{13}$	12
6	$\bar{10}$	$\bar{18}$	18	6	$\bar{2}$	$\bar{70}$	53
8	$\bar{10}$	$\bar{28}$	20	8	$\bar{2}$	5	17
0	$\bar{8}$	35	34	10	$\bar{2}$	$\bar{63}$	78
2	$\bar{8}$	8	11	12	$\bar{2}$	1	8
4	$\bar{8}$	56	43	2	0	8	13
6	$\bar{8}$	$\bar{8}$		4	0	93	100
8	$\bar{8}$	54	44	6	0	29	44
10	$\bar{8}$	1		8	0	35	49
12	$\bar{8}$	35	36	10	0	46	60
0	$\bar{6}$	$\bar{5}$		12	0	6	5
2	$\bar{6}$	58	49	2	2	1	
4	$\bar{6}$	$\bar{34}$	36	4	2	$\bar{71}$	77
6	$\bar{6}$	$\bar{28}$	35	6	2	11	27
8	$\bar{6}$	$\bar{56}$	48	8	2	58	83
10	$\bar{6}$	$\bar{1}$		10	2	$\bar{1}$	7
12	$\bar{6}$	$\bar{60}$	52	2	4	63	56
0	$\bar{4}$	4		7	4	30	38
2	$\bar{4}$	90	58	6	4	23	32
4	$\bar{4}$	$\bar{7}$	9	8	4	46	69
6	$\bar{4}$	71	67	2	6	$\bar{68}$	64
8	$\bar{4}$	10		4	6	3	
10	$\bar{4}$	35	41	6	6	$\bar{52}$	55
12	$\bar{4}$	31	40	2	8	27	31
0	$\bar{2}$	$\bar{58}$	43	4	8	9	15

in layers of halogen atoms. Here the distance of contact Ba-halogen is found between the layers, and the shortest distance halogen-H_2O in the layers. The Ba–Ba distance is much greater than found for the element.

$[Sr(H_2O)_2]_\infty$ columns occur in the structure of $SrCl_2, 2H_2O$[5], but the arrangement in the complex is quite different from the arrangement in the $[Ba(H_2O)_2]_\infty$ complex in $BaBr_2, 2H_2O$, and the distance between the Sr atoms is much greater than for the element. $BaCl_2, 2H_2O$[4] shows infinite 2-dimensional Ba–H_2O complexes.

A description analogous to that given in (4) and (5) for $SrCl_2, 2H_2O$ and $BaCl_2, 2H_2O$, of $BaBr_2, 2H_2O$ as a layer structure, where the layers are neutral complexes of $[BaBr_2, 2H_2O]_{\infty\infty}$, does not fit so well here. The neutral

TABLE 5. Comparison of calculated and observed structure factors. Rotation [010] 1-layer line $k = 1$.

h	l	F_{calc}	$\lvert F_{obs} \rvert$	h	l	F_{calc}	$\lvert F_{obs} \rvert$
1	$\overline{10}$	$\overline{15}$	$\overline{17}$	5	$\overline{3}$	$\overline{34}$	25
3	$\overline{10}$	$\overline{36}$	26	7	$\overline{3}$	$\overline{28}$	26
1	$\overline{9}$	14		9	$\overline{3}$	$\overline{20}$	31
3	$\overline{9}$	17	16	11	$\overline{3}$	$\overline{24}$	35
5	$\overline{9}$	19	18	13	$\overline{3}$	$\overline{22}$	16
7	$\overline{9}$	14		1	$\overline{2}$	18	20
9	$\overline{9}$	21	11	3	$\overline{2}$	$\overline{88}$	50
1	$\overline{8}$	50	47	5	$\overline{2}$	3	
3	$\overline{8}$	3		7	$\overline{2}$	$\overline{47}$	70
5	$\overline{8}$	32	34	9	$\overline{2}$	32	42
7	$\overline{8}$	28	23	11	$\overline{2}$	$\overline{12}$	12
9	$\overline{8}$	8		13	$\overline{2}$	$\overline{38}$	20
11	$\overline{8}$	42	45	1	$\overline{1}$	31	18
1	$\overline{7}$	$\overline{16}$	17	3	$\overline{1}$	12	9
3	$\overline{7}$	$\overline{26}$	26	5	$\overline{1}$	26	36
5	$\overline{7}$	$\overline{16}$	14	7	$\overline{1}$	11	15
7	$\overline{7}$	$\overline{21}$	18	9	$\overline{1}$	18	32
9	$\overline{7}$	$\overline{13}$	17	11	$\overline{1}$	12	11
11	$\overline{7}$	$\overline{17}$	13	1	0	30	20
1	$\overline{6}$	$\overline{55}$	57	3	0	69	69
3	$\overline{6}$	9		5	0	$\overline{7}$	15
5	$\overline{6}$	$\overline{65}$	65	7	0	65	82
7	$\overline{6}$	14	10	9	0	$\overline{12}$	23
9	$\overline{6}$	$\overline{46}$	42	11	0	33	44
11	$\overline{6}$	$\overline{14}$	18	1	1	$\overline{22}$	25
1	$\overline{5}$	14	13	3	1	$\overline{26}$	26
3	$\overline{5}$	19	21	5	1	$\overline{25}$	46
5	$\overline{5}$	9	14	7	1	$\overline{16}$	25
7	$\overline{5}$	20	23	9	1	$\overline{23}$	41
9	$\overline{5}$	10		11	1	$\overline{13}$	10
11	$\overline{5}$	18	23	1	2	$\overline{76}$	70
13	$\overline{5}$	11	8	3	2	$\overline{4}$	
1	$\overline{4}$	13	12	5	2	$\overline{39}$	55
3	$\overline{4}$	34	30	7	2	$\overline{30}$	50
5	$\overline{4}$	47	39	9	2	$\overline{6}$	
7	$\overline{4}$	0		11	2	$\overline{43}$	22
9	$\overline{4}$	68	75	1	3	26	28
11	$\overline{4}$	$\overline{9}$	15	3	3	35	38
13	$\overline{4}$	45	26	5	3	19	30
1	$\overline{3}$	41	40	7	3	22	35
3	$\overline{8}$	$\overline{29}$	30	9	3	17	22

(to be continued)

TABLE 5 (continued).

h	l	F_{calc}	$\lvert F_{obs} \rvert$	h	l	F_{calc}	$\lvert F_{obs} \rvert$
1	4	63	58	1	6	$\overline{11}$	17
3	4	$\overline{9}$	14	3	6	$\overline{25}$	24
5	4	60	70	5	6	$\overline{33}$	38
7	4	$\overline{5}$	9	1	7	25	25
9	4	38	26	3	7	16	15
1	5	$\overline{13}$	14	5	7	19	18
3	5	$\overline{16}$	17	1	8	$\overline{7}$	
5	5	$\overline{9}$	13	3	8	$\overline{48}$	44
7	5	$\overline{16}$	22	1	9	$\overline{18}$	22

TABLE 6. Comparison of calculated and observed structure factors. Rotation $[1\bar{1}0]$ 0-layer line.

h	k	l	F_{calc}	$\lvert F_{obs} \rvert$	h	k	l	F_{calc}	$\lvert F_{obs} \rvert$
1	1	$\overline{10}$	$\overline{15}$	14	5	5	$\overline{5}$	$\overline{23}$	33
2	2	$\overline{10}$	20	10	6	6	$\overline{5}$	$\overline{25}$	34
1	1	$\overline{9}$	14	10	7	7	$\overline{5}$	3	5
2	2	$\overline{9}$	36	28	1	1	$\overline{4}$	14	
3	3	$\overline{9}$	24	20	2	2	$\overline{4}$	54	46
1	1	$\overline{8}$	52	45	3	3	$\overline{4}$	$\overline{10}$	13
2	2	$\overline{8}$	$\overline{7}$		4	4	$\overline{4}$	$\overline{60}$	58
3	3	$\overline{8}$	$\overline{27}$	24	5	5	$\overline{4}$	$\overline{21}$	34
4	4	$\overline{8}$	$\overline{4}$		6	6	$\overline{4}$	$\overline{4}$	14
5	5	$\overline{8}$	$\overline{20}$	22	7	7	$\overline{4}$	$\overline{19}$	14
1	1	$\overline{7}$	$\overline{16}$	17	1	1	$\overline{3}$	$\overline{44}$	37
2	2	$\overline{7}$	$\overline{30}$	30	2	2	$\overline{3}$	$\overline{28}$	30
3	3	$\overline{7}$	$\overline{36}$	48	3	3	$\overline{3}$	$\overline{17}$	18
4	4	$\overline{7}$	$\overline{8}$		4	4	$\overline{3}$	$\overline{42}$	46
5	5	$\overline{7}$	16	14	5	5	$\overline{3}$	$\overline{10}$	
1	1	$\overline{6}$	$\overline{55}$	58	6	6	$\overline{3}$	44	48
2	2	$\overline{6}$	$\overline{26}$	30	7	7	$\overline{3}$	25	25
3	3	$\overline{6}$	41	43	1	1	$\overline{2}$	18	19
4	4	$\overline{6}$	25	33	2	2	$\overline{2}$	$\overline{25}$	29
5	5	$\overline{7}$	4		3	3	$\overline{2}$	$\overline{28}$	28
6	6	$\overline{7}$	30	31	4	4	$\overline{2}$	47	56
7	7	$\overline{7}$	19	14	5	5	$\overline{2}$	54	61
1	1	$\overline{5}$	14	15	6	6	$\overline{2}$	7	13
2	2	$\overline{5}$	21	21	7	7	$\overline{2}$	2	
3	3	$\overline{5}$	48	56	1	1	$\overline{1}$	31	25
4	4	$\overline{5}$	33	37	2	2	$\overline{1}$	55	32

(to be continued)

TABLE 6. (Continued).

h	k	l	F_{calc}	$\lvert F_{obs} \rvert$	h	k	l	F_{calc}	$\lvert F_{obs} \rvert$
3	3	$\bar{1}$	24	27	4	4	3	9	
4	4	$\bar{1}$	20	32	5	5	3	$\overline{16}$	17
5	5	$\bar{1}$	22	15	6	6	3	2	
6	6	$\bar{1}$	$\overline{23}$	32	0	0	4	4	
7	7	$\bar{1}$	$\overline{44}$	40	1	1	4	65	56
1	1	0	29	31	2	2	4	28	29
2	2	0	$\overline{46}$	55	3	3	4	$\overline{42}$	37
3	3	0	17	28	4	4	4	$\overline{24}$	30
4	4	0	1		5	5	4	$\bar{5}$	
5	5	0	$\overline{55}$	57	6	6	4	$\overline{28}$	30
6	6	0	$\overline{33}$	36	1	1	5	$\overline{12}$	14
7	7	0	$\bar{7}$		2	2	5	$\overline{19}$	22
1	1	1	$\overline{23}$	16	3	3	5	$\overline{42}$	42
2	2	1	$\overline{61}$	41	4	4	5	$\overline{29}$	28
3	3	1	$\overline{37}$	44	5	5	5	20	20
4	4	1	2		0	0	$\bar{6}$	$\bar{5}$	
5	5	1	$\overline{10}$	12	1	1	6	$\overline{12}$	24
6	6	1	$\bar{4}$		2	2	6	$\overline{42}$	32
7	7	1	43	34	3	3	6	9	
0	0	2	$\overline{58}$	37	4	4	6	46	35
1	1	2	$\overline{89}$	44	1	1	7	26	27
2	2	2	10		2	2	7	19	20
3	3	2	36	30	3	3	7	14	16
4	4	2	4		4	4	7	30	24
5	5	2	23	32	0	0	8	35	29
6	6	2	40	31	1	1	8	$\bar{8}$	
1	1	3	26	27	2	2	8	14	11
2	2	3	38	37	3	3	8	18	8
3	3	3	54	44	1	1	9	$\overline{18}$	16

layers, which must be taken parallel to (110) or (100), where the $[Ba(H_2O)_2]_\infty$ columns have their shortest distance, are penetrated by bromine ions from adjacent layers. For a layer in the first case every second bromine atom in the columns parallel to [001] belongs to an adjacent layer, and for a layer in the second case every second bromine column in the buckled bromine layer parallel to (100) belongs to an adjacent layer (figs. 6 and 11).

In many of the structures of salt hydrates earlier investigated, especially the hydrated oxy-salts, the water molecule is surrounded by 3–4 neighbours: 1–2 metal ions and 2–3 anions. This simple rule can neither be applied for the hydrated alkaline earth halogenides nor for the hydrated halogenides of

TABLE 7. Interatomic distances between nearest neighbours (Frequency). The atoms are numbered by means of the sequence given in Int. Tab. for the general position of $C2/c$ (No. 15).

			Distance *in Å*	From PAULING Ionic Radii	Radius sum From GOLDSCHMIDT Ionic Radii	
Ba_1	Ba_2	(2)	4.44			4.347[19]
	Br_1	(2)	3.41	3.30	3.39	
	Br_7	(2)	3.49			
	Br_6	(2)	3.52			
	O_4	(2)	2.82	2.82	2.82	
	O_1	(2)	3.00			
Br_1	Ba_1	(1)	3.32			
	Ba_3	(1)	3.49			
	Ba_4	(1)	3.52			
	Br_6	(1)	4.00	3.90	3.92	
	Br_7	(1)	4.14			
	Br_3	(1)	4.18			
	Br_4	(1)	4.21			
	O_1	(1)	3.25	3.29	3.39	
	O_5	(1)	3.27			
	O_3	(1)	3.31			
	O_6	(1)	3.33			
	O_4	(1)	3.60			
	O_2	(1)	3.81			
O_1	Ba_2	(1)	2.82			
	Ba_1	(1)	3.00			
	Br_1	(1)	3.25			
	Br_5	(1)	3.27			
	Br_3	(1)	3.31			
	Br_6	(1)	3.33			
	Br_4	(1)	3.50			
	Br_2	(1)	3.81			

the transition elements.[15] In the first case it is only valid for $SrCl_2, 6H_2O$, and $BaCl_2, 2H_2O$. The number and positions of the atoms surrounding the water molecule given above for $BaBr_2, 2H_2O$ are not incompatible with the results for the group as a whole. It is not possible from the obtained data to say anything about the positions of the hydrogen atoms.

It should be mentioned that the direction of the $[Ba(H_2O)_2]_\infty$ columns

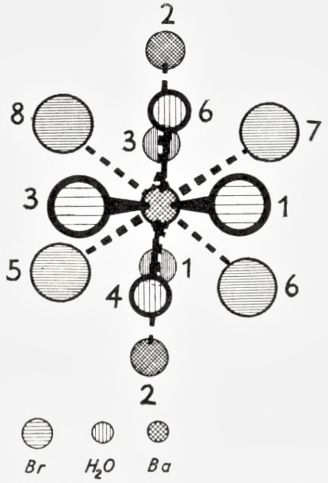

Fig. 9. Barium and surrounding atoms. Viewed in the direction along the *b*-axis. Shortest distances solid lines. Longer distances dashed lines. Lower lying atoms have a dark hatch-ing and thin circumference.

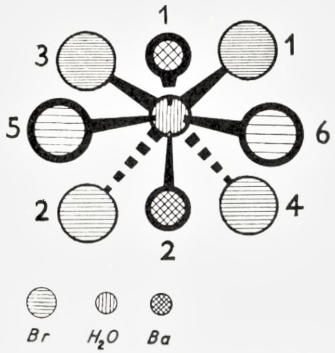

Fig. 10. Water molecule and surrounding atoms. Viewed in the direction along the *b*-axis. Signatures as in fig. 9.

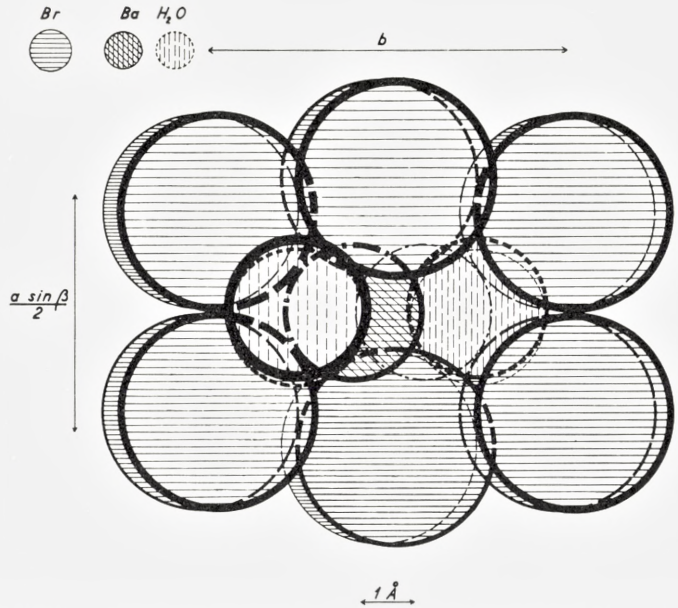

Fig. 11. $[Ba(H_2O)_2]_\infty$ column and surrounding Br columns viewed along tbe *c*-axis. Lower layers have dark hatching and a thin circumference.

[001] and the packing of the columns could be connected to the occurrence of the 2 different needle axes [001] and [110].

The deformation by pressure in the direction of the *c*-axis and the formation of twins are understandable if it is considered that the $[Ba(H_2O)_2]_\infty$ columns lie in this direction (fig. 12). $SrCl_2, 6\,H_2O$[20] shows cleavage per-

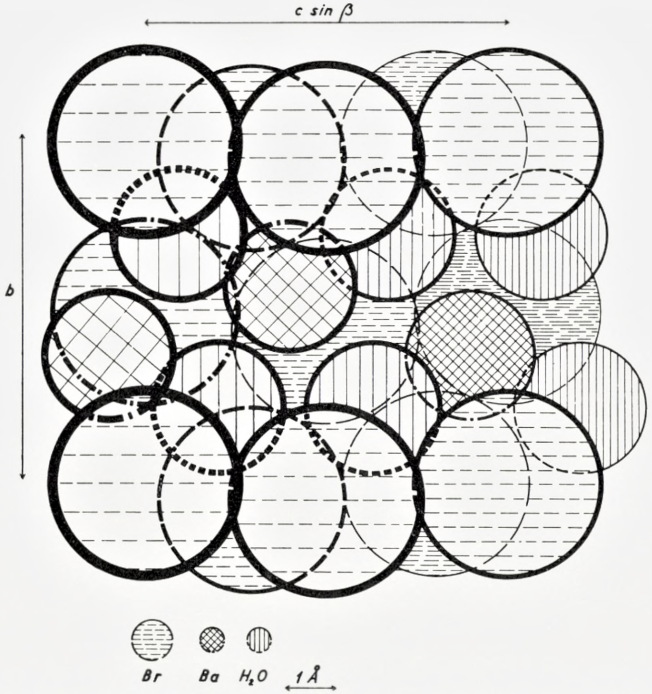

Fig. 12. $[Ba(H_2O)_2]_\infty$ column and surrounding bromine atoms viewed along the *a*-axis. Signatures as in fig. 11.

pendicular to the $[Sr(H_2O)_6]_\infty$ columns. It has not been possible to show cleavage for $BaBr_2, 2\,H_2O$.

Table 4 gives the shortest halogen-water distances for some halogen hydrates. The distances for $BaBr_2$, H_2O compare well with the other distances in the table. The mean value of the mean values of the distances for each crystal gives Br–H_2O = 3.29 Å and Cl–H_2O = 3.15 Å. This gives an empirical radius for H_2O against halogen ions $r^{an}_{H_2O}\,{}^{(P)}_{(G)}$ 1.34 Å for Cl and $r^{an}_{H_2O}\,{}^{(P)\ 1.34}_{(G)\ 1.33}$ Å for Br.

TABLE 8.

	Cl–H_2O Distance in A			
$BaCl_2$, H_2O[1]	3.14	3.24		
$BaCl_2$, $2H_2O$[4]	3.17	3.18	3.19	3.22
$SrCl_2$, $2H_2O$[3]	3.10			
$SrCl_2$, $6H_2O$[5]	3.10	3.17		
$CoCl_2$, $2H_2O$[21]	3.18	3.19		
$FeCl_2$, $4H_2O$[22]	3.07	3.18	3.22	
$AlCl_3$, $6H_2O$[6]	3.1			
HCl, H_2O[23]	2.95			
Adeninhydrochloridehemihydrate[24]	3.12			
	Br–H_2O			
$BaBr_2$, H_2O[2]	3.29	3.32		
$BaBr_2$, $2H_2O$	3.25	3.27	3.31	3.33
11–Aminoundecanoicacidhydrobromide-hemihydrate[25]	3.17	3.38		
Codeinhydrobromide Dihydrate[26]	3.27			
Strychninehydrobromide Dihydrate[27]	3.15	3.30	3.41	

Acknowledgements

I wish to thank Professor, dr. phil. A. TOVBORG JENSEN for his interest in this work and mag. scient. CHR. KN. MØLLER for help and discussions.

References

1. Z. G. Pinsker, B. K. Vajnstejn. Zur Fiz. Khim **23** p. 1058 (1949).
2. B. K. Vajnstejn, Z. G. Pinsker. Zur Fiz. Khim **24** p. 432 (1950).
3. A. Tovborg Jensen, Mat. Fys. Medd. Dan. Vid. Selsk. **17** No. 9 (1940).
4. — — — — — — — — — **20** No. 5 (1942).
5. — — — — — — — — — **22** No. 3 (1945).
6. — Krystallinske Salthydrater, Copenhagen (1948).
7. O. Mügge, Neues Jahrbuch f. Min. **1** p. 130 (1889).
8. — — — — — **1** p. 71 (1898).
9. H. Dufet, Bull. Soc. Fr. Min. XXVI p. 67 (1903)
10. C. Rammelsberg, Pogg. An. **55** p. 232 (1842).
11. Handl. Sitzb. Akad. Wien XXXII p. 244 (1858).
12. Heidweyller, Zeit. Phys. **3** p. 320 (1920).
13. W. G. Hankel, H. Lindenberg, Zeit. Krist. **27** p. 516 (1897).
14. V. Frank, J. Sci. Instr. **34** p. 310 (1957).
15. A. F. Wells, Quart. Rev. **8** p. 380 (1954).
16. P. A. Renes, C. A. Mac. Gillavry, Rec. Trav. Chim. **64**, p. 275 (1945).
17. H. Lipson, W. Cochran, The Crystalline State, vol. III.
    — Determination of Crystal Structures London (1953).
18. Martin J. Buerger, Crystal-Structure analysis, New York (1960).
19. W. Klemm, G. Mika, Z. Anorg. Chem. **248** p. 155 (1941).
20. A. Eppler, Zeit. Krist. **30** p. 129 (1899).
21. B. K. Vajnstejn, Doklady Akad. Nauk. S.S.S.R. **68** p. 301 (1949).
22. Bruce R. Penfold, J. A. Gregor, Acta Cryst. **12** p. 850 (1959).
23. Ku Yoon, Gene B. Carpenter, Acta Cryst. **12** p. 17 (1959).
24. June M. Broomhead, Acta Cryst. **1** p. 324 (1947).
25. G. A. Sim, Acta Cryst. **8** p. 833 (1955).
26. June M. Lindsey, U. U. Barnes, Acta Cryst. **8** p. 227 (1955).
27. J. H. Robertson, C. A. Beevers, Acta Cryst. **4** p. 270 (1951).

Indleveret til selskabet den 9. december 1960.
Færdig fra trykkeriet den 4. september 1961.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 5

Mat. Fys. Medd. Dan.Vid. Selsk. **33,** no. 5 (1961)

THE ELECTRONIC STRUCTURE OF FERROCENE

BY

J. P. DAHL AND C. J. BALLHAUSEN

København 1961
i kommission hos Ejnar Munksgaard

Synopsis

The electronic ground state and lowest excited states in ferrocene have been calculated, using the LCAO-MO-SCF method of ROOTHAAN. The various molecular integrals were evaluated by means of the approximations introduced by GOEPPERT-MAYER and SKLAR and by PARISER and PARR. Using the self-consistent field wave functions for iron, recently published by WATSON, we obtained an ionization potential for the molecule of 10.92 eV. Low excited states are further calculated to occur at 5.38 eV (A_{2g}), 5.44 eV (A_{1u}) 5.75 eV. (E_{1u}), 7.06 eV (A_{2u}), and 9.35 eV. (E_{1u}). A correlation is made between these numbers and the measured absorption spectra of the compound. Finally, the magnetic features of other related "sandwich compounds" are discussed in the light of the bonding scheme in ferrocene.

Printed in Denmark
Bianco Lunos Bogtrykkeri A-S

Introduction

Ferrocene is the trivial name for the compound $Fe(C_5H_5)_2$. Due to its high symmetry (D_{5d}) it should be well suited for a theoretical treatment in spite of the large number of electrons present. In this paper, we describe a calculation of the ground state and some excited states, using the simplified SCF–LCAO–MO theory given by ROOTHAAN[1]. A similar calculation has previously been performed by YAMAZAKI[2]; however, in view of this author's very short report, which makes it somewhat difficult to see the exact ordering of the levels, it was felt worthwhile to repeat the calculation.

Another incitement was the recent publication of the self-consistent field calculations by WATSON[3] of the electronic orbitals of the metals in the first transition group. These new orbitals should be vastly superior to those used by YAMAZAKI[2], since his are based on SLATER's rules. They are accordingly used in this work. Furthermore, the effect of overlap has been taken more explicitly into account than what appears to be the case in Yamazaki's paper.

In the course of our work, a paper appeared by SHUSTOROVICH and DYATKINA[4], treating ferrocene in a way similar to our procedure, but still using Slater orbitals for the metal ion. This allowed us to compare our results with that of the Russian authors. Due to the more contracted form of Watson's orbitals as compared to those of Slater, there are significant differences. A closer comparison is, however, given later.

Since an excellent review by WILKINSON and COTTON[5] deals very extensively with the history and subsequent theories of the chemical bonding present in "sandwich compounds", reference to previous works does not appear necessary. We restrict ourselves to mention that the qualitative aspects of the bonding present in these compounds have been given by MOFFITT[6] and by DUNITZ and ORGEL[7]. The more quantitative calculations reported here will again be seen to differ somewhat from the conclusions reached by these authors.

For discussions of the theoretical aspects of our approximation, Roothaan's fundamental paper[1] should be consulted.

The Orbitals

a) The Metal Orbitals.

Realizing that the iron orbitals $1s$, $2s$, $2p$, $3s$ and $3p$ are too contracted to participate in the chemical bonding, we utilize the $3d$, $4s$ and $4p$ orbitals for this purpose. They are of the general form of a radial function times a spherical harmonic

$$\psi(n, l, m) = \frac{1}{r} R_{n,l}(r)\, Y_l^m(\vartheta, \varphi).$$

They are assumed to be normalized to unity. Some of their transformation properties in the molecular point group symmetry D_{5d} (Fig. 1) are given in Table 1.

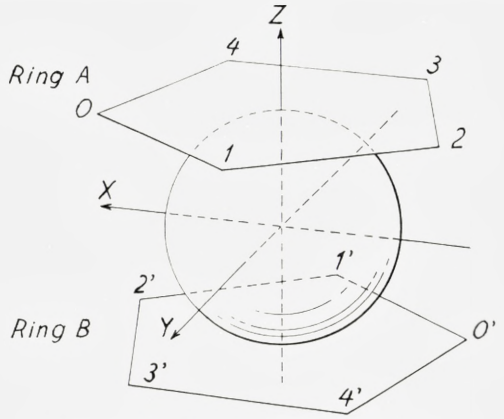

Fig. 1. Molecular structure of ferrocene.

TABLE 1. Some transformation properties af metal orbitals in the molecular symmetry D_{5d}. $\omega = \exp. 2\pi i/5$.

D_{5d}	E	$2C_5$	$2C_5^2$	$5C_2$	i	Designation
$3d_0$	$3d_0$	$3d_0$	$3d_0$	$3d_0$	$3d_0$	a_{1g}
$3d_1$	$3d_1$	$\omega\, 3d_1$	$\omega^2 3d_1$	$3d_{-1}$	$3d_1$	e_{1g}^+
$3d_{-1}$	$3d_{-1}$	$\omega^{-1} 3d_{-1}$	$\omega^{-2} 3d_{-1}$	$3d_1$	$3d_{-1}$	e_{1g}^-
$3d_2$	$3d_2$	$\omega^2 3d_2$	$\omega^{-1} 3d_2$	$3d_{-2}$	$3d_2$	e_{2g}^+
$3d_{-2}$	$3d_{-2}$	$\omega^{-2} 3d_{-2}$	$\omega\, 3d_{-2}$	$3d_2$	$3d_{-2}$	e_{2g}^-
$4s$	$4s$	$4s$	$4s$	$4s$	$4s$	a_{1g}
$4p_0$	$4p_0$	$4p_0$	$4p_0$	$-4p_0$	$-4p_0$	a_{2u}
$4p_1$	$4p_1$	$\omega\, 4p_1$	$\omega^2 4p_1$	$-4p_{-1}$	$-4p_1$	e_{1u}^+
$4p_{-1}$	$4p_{-1}$	$\omega^{-1} 4p_{-1}$	$\omega^{-2} 4p_{-1}$	$-4p_1$	$-4p_{-1}$	e_{1u}^-

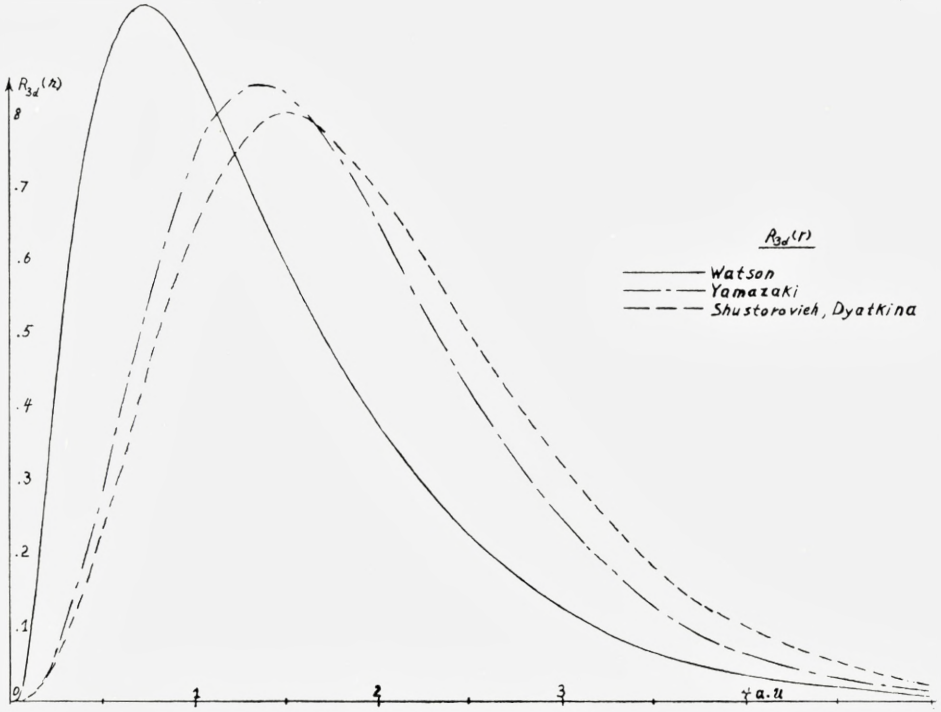

Fig. 2. Radial $3d$ orbital.

For the radial functions we choose Watson's self-consistent orbitals for the ground state of $Fe\,[A]\;(3d)^6(4s)^2$, where $[A]$ stands for a closed eighteen electronic shell. This choice instead of the Slater orbitals has the additional advantage that Watson's calculations contain most of the atomic interaction integrals which we have to use. Watson gives no $4p$ orbitals. We have therefore assumed that the radial part of the $4p$ orbital is nearly identical with that for the $4s$ orbital.

While Watson used a combination of four Slater orbitals for the $3d$-orbital with $n = 3$ and ten Slater orbitals with different values of n to describe the $4s$ orbital, we had to reduce these numbers for calculational reasons. If $R_n(\xi)$ stands for a Slater orbital with the quantum number n and the exponent equal to ξ, we use

$$R_{3d}(r) = 0.5978\;\; R_3(2.385) + 0.4982\; R_3(4.77)$$
$$R_{4s}(r) = R_{4p}(r) = R_3(1{,}093125).$$

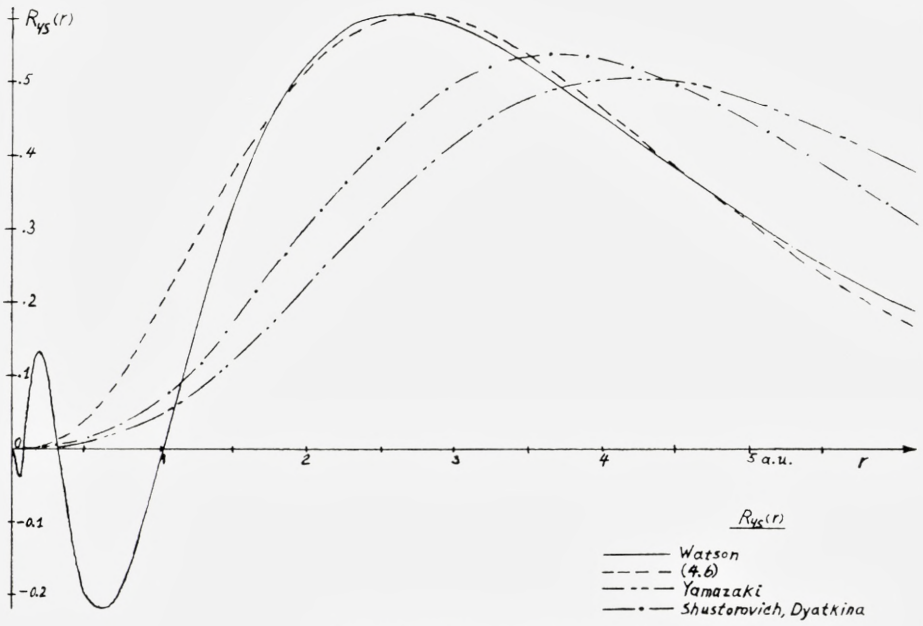

Fig. 3. Radial $4s$ orbital.

The values of ξ have been chosen in such a way that the orbitals are approximated in the best possible way in the region where the overlap takes place between the rings and the metal atom. The orbitals used are seen in Figs. 2 and 3. It appears that they are much more contracted than the Slater orbitals used by YAMAZAKI[2] and by SHUSTOROVICH and DYATKINA[4].

The orbitals used are then

$$\psi(a'_{1g}) = \frac{1}{r} R_{3d}(r) \sqrt{\frac{5}{16\pi}} (3\cos^2\vartheta - 1)$$

$$\psi(e^{\pm}_{1g}) = \frac{1}{r} R_{3d}(r) \sqrt{\frac{15}{8\pi}} \cos\vartheta \sin\vartheta \, e^{\pm i\varphi}$$

$$\psi(e^{\pm}_{2g}) = \frac{1}{r} R_{3d}(r) \sqrt{\frac{15}{32\pi}} \sin^2\vartheta \, e^{\pm i2\varphi}$$

$$\psi(a_{1g}) = \frac{1}{r} R_{4s}(r) \sqrt{\frac{1}{4\pi}}$$

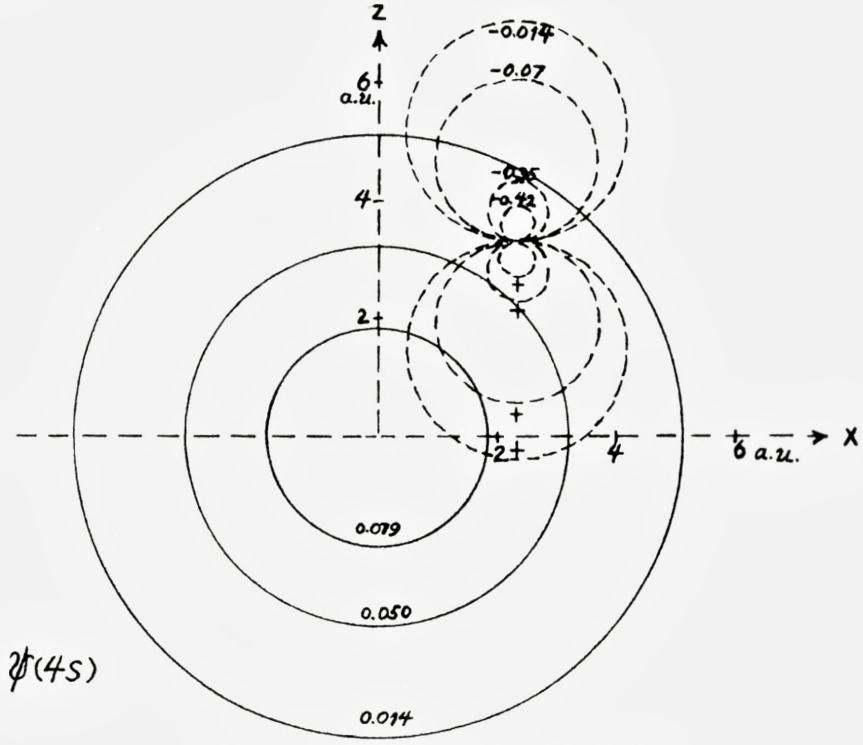

Fig. 4. $4s$ and $2p_z$ orbital, drawn to scale.

$$\psi(a_{2u}^{\pm}) = \frac{1}{r} R_{4p}(r) \sqrt{\frac{3}{4\pi}} \cos\vartheta$$

$$\psi(e_{2u}^{\pm}) = \frac{1}{r} R_{4p}(r) \sqrt{\frac{3}{8\pi}} \sin\vartheta \, e^{\pm i\varphi}$$

These orbitals are shown in Figs. 4, 5 and 6 together with a carbon $2p_z$ orbital. The pictures are drawn to scale.

b. The Ring Orbitals.

In order to construct symmetry orbitals transforming correctly in the molecular point group symmetry D_{5d}, we use the procedure outlined by MOFFITT[6]. The local coordinate systems upon the rings are always chosen

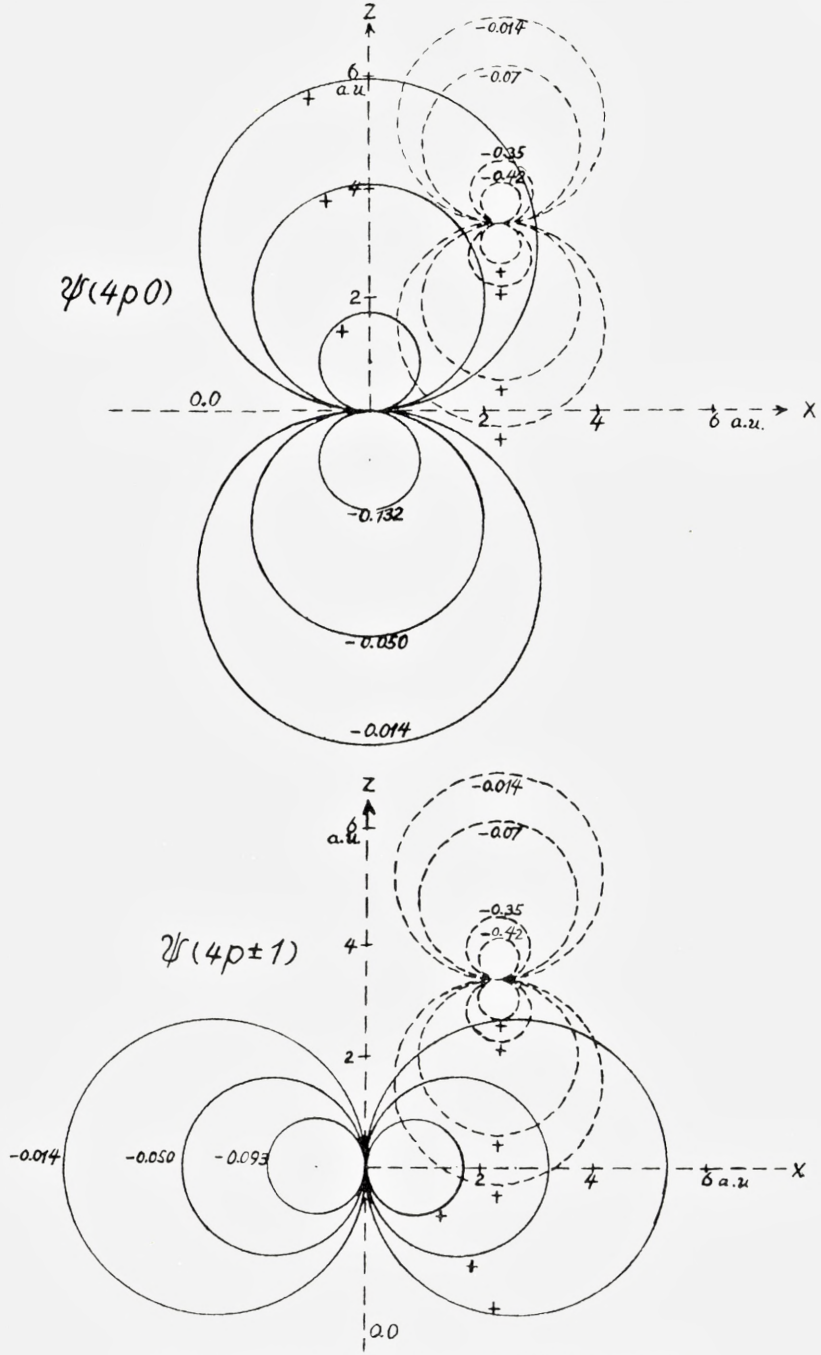

Fig. 5. $4p$ orbitals and a $2p_z$ orbital, drawn to scale.

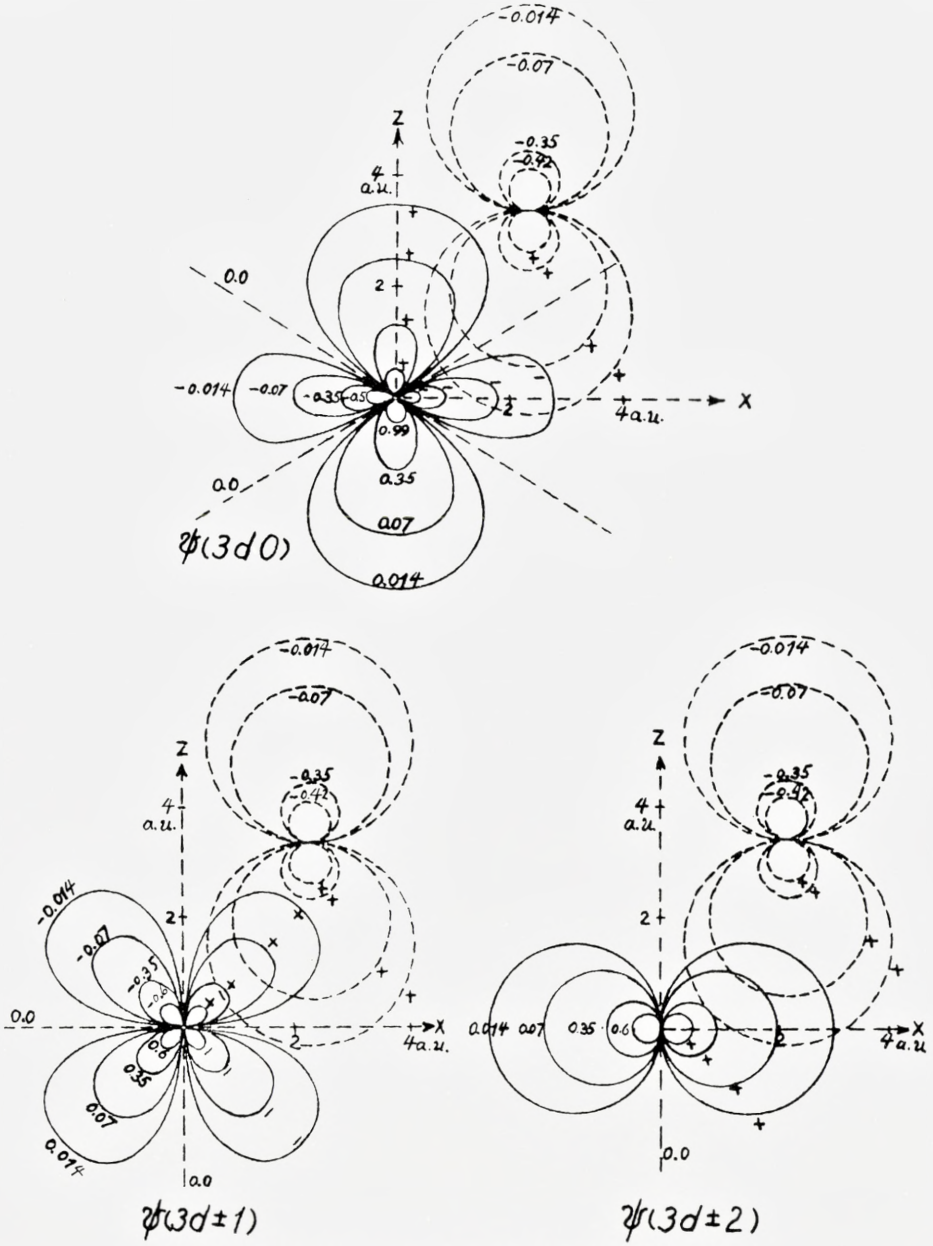

Fig. 6. $3d$ orbitals and a $2p_z$ orbital, drawn to scale.

in such a way that the z_i-axis is parallel to the z-axis of the molecule (Fig. 1) Furthermore, the z_i-axes on the two rings are pointing towards each other.

Note that in a *di*-π-cyclopentadienyl compound the plane of the rings is not a symmetry plane of the molecule. Consequently, the orbitals of the rings cannot be separated into σ- and π-orbitals. Nevertheless, these denotations will be used here in their conventional meaning.

A solution of Roothaan's equations will produce three sets of orbitals, all of which are linear combinations of the metal orbitals, the σ-bonding orbitals, the σ-antibonding orbitals, and the π-orbitals. The lowest placed set, comprising ten orbitals, is filled up. It is composed of nearly pure σ-orbitals, but contains in addition a small portion of the metal orbitals and the π-orbitals. The second set is that in which we are interested. It consists of metal- and π-orbitals, but may have a slight amount of σ-orbitals. This set is partly filled with electrons. The set with the highest energy is that built up mainly of antibonding σ-orbitals. All these orbitals are empty.

We assume here, as usual, that the first and third set are made up solely by bonding and antibonding σ-orbitals, respectively. It is furthermore assumed that the antibonding σ-levels do not interfere with the lowest excited states. Consequently, the eighteen valence electrons for ferrocene (eight metal electrons and ten π-electrons from the rings) are supposed to be in the molecular orbitals which are linear combinations of metal- and π-orbitals alone.

For the π-symmetry orbitals we have then, according to MOFFITT[6], calling the orbitals on one ring A, and those on the other ring B (see Fig. 1):

$$\varphi(a_{1g}) = \sqrt{\frac{1}{2}}\left[\varphi_A(a_2) + \varphi_B(a_2)\right] \quad \varphi(a_{2u}) = \sqrt{\frac{1}{2}}\left[\varphi_A(a_2) - \varphi_B(a_2)\right]$$

$$\varphi(e_{1g}^{+}) = \sqrt{\frac{1}{2}}\left[\varphi_A(e_1^{+}) + \varphi_B(e_1^{+})\right] \quad \varphi(e_{1u}^{+}) = \sqrt{\frac{1}{2}}\left[\varphi_A(e_1^{+}) - \varphi_B(e_1^{+})\right]$$

$$\varphi(e_{1g}^{-}) = \sqrt{\frac{1}{2}}\left[\varphi_A(e_1^{-}) + \varphi_B(e_1^{-}) \quad \varphi(e_{1u}^{-})\right] = \sqrt{\frac{1}{2}}\left[\varphi_A(e_1^{-}) - \varphi_B(e_1^{-})\right]$$

$$\varphi(e_{2g}^{+}) = \sqrt{\frac{1}{2}}\left[\varphi_A(e_2^{+}) + \varphi_B(e_2^{+})\right] \quad \varphi(e_{2u}^{+}) = \sqrt{\frac{1}{2}}\left[\varphi_A(e_2^{+}) - \varphi_B(e_2^{+})\right]$$

$$\varphi(e_{2g}^{-}) = \sqrt{\frac{1}{2}}\left[\varphi_A(e_2^{-}) + \varphi_B(e_2^{-})\right] \quad \varphi(e_{2u}^{-}) = \sqrt{\frac{1}{2}}\left[\varphi_A(e_2^{-}) - \varphi_B(e_2^{-})\right]$$

Here we have, for instance,

$$\varphi_A(a_2) = \sqrt{\frac{1}{5}}(\pi_0 + \pi_1 + \pi_2 + \pi_3 + \pi_4)$$

$$\varphi_A(e_1^+) = \sqrt{\frac{1}{5}}(\pi_0 + \omega\pi_1 + \omega^2\pi_2 + \omega^3\pi_3 + \omega^4\pi_4)$$

$$\varphi_A(e_1^-) = \sqrt{\frac{1}{5}}(\pi_0 + \omega^{-1}\pi_1 + \omega^{-2}\pi_2 + \omega^{-3}\pi_3 + \omega^{-4}\pi_4)$$

$$\varphi_A(e_2^+) = \sqrt{\frac{1}{5}}(\pi_0 + \omega^2\pi_1 + \omega^4\pi_2 + \omega^6\pi_3 + \omega^8\pi_4)$$

$$\varphi_A(e_2^-) = \sqrt{\frac{1}{5}}(\pi_0 + \omega^{-2}\pi_1 + \omega^{-4}\pi_2 + \omega^{-6}\pi_3 + \omega^{-8}\pi_4)$$

$$\omega = \exp\frac{2\pi i}{5}.$$

We have of course a similar combination for ring B.

The analytical form of the radial function of the carbon π orbital is given by

$$R_{2p}(r) = \frac{(2\xi)^{5/2}}{\sqrt{4!}} r^2 e^{-\xi r},$$

where, according to ZENER[8], $\xi = 1.59$.

c. Orthogonal Symmetry Orbitals.

The metal orbitals and the ring orbitals are within themselves orthogonal to each other. On the other hand, there are overlap integrals between those of the metal and ring orbitals which transform in the same way. By means of the given orbitals and the structural data of SEIBOLD and SUTTON[9], viz. for the distances Fe-C = 2.03 Å and C-C = 1.43 Å, we obtain by the usual methods[10] the following values of the overlap integrals:

$S(a_{1g})$	$S(a'_{1g})$	$S(e_{1g})$	$S(e_{2g})$	$S(a_{2u})$	$S(e_{1u})$
0.527	0.030	0.148	0.079	0.236	0.468

defining

$$S(a_{1g}) = \int \psi^*(4s)\varphi(a_{1g})\,d\tau$$

$$S(a'_{1g}) = \int \psi^*(3d\ 0)\varphi(a_{1g})\,d\tau$$

and similarly for the other overlap integrals. The way the overlap takes place is illustrated in Figures 4, 5, and 6. Since $S(a'_{1g})$ is very small, we take $\psi(3d0)$ and $\varphi(a_{1g})$ to be orthogonal to each other. (See, however, the following).

It is a great advantage, instead of the orbitals so far discussed, to use two sets of orbitals which are mutually orthogonal. Since the orbitals have overlap only pair by pair, we easily find for such a set of orbitals χ'_p:

$$\chi'_p = \sum_{r=1}^{2} \Delta_{rp}^{-\frac{1}{2}} \chi_r$$

with

$$\Delta_{rp}^{-\frac{1}{2}} = \frac{1}{2}\begin{bmatrix} (1+S)^{-1/2}+(1-S)^{-1/2} & (1+S)^{-1/2}-(1-S)^{-1/2} \\ (1+S)^{-1/2}-(1-S)^{-1/2} & (1+S)^{-1/2}+(1-S)^{-1/2} \end{bmatrix}.$$

We then obtain the orthogonal symmetry orbitals, denoting the essential metal orbital by μ and the essential ring orbitals by ϱ,

$$a_{1g}\begin{cases} \chi_1 = \mu(4s) \\ \chi_2 = \mu(3d0) = \psi(3d0) \\ \chi_3 = \varrho(a_{1g}) \end{cases} \qquad e^+_{1u}\begin{cases} \chi_{10} = \mu(4p1) \\ \chi_{11} = \varrho(e^+_{1u}) \end{cases}$$

$$a_{2u}\begin{cases} \chi_4 = \mu(4p0) \\ \chi_5 = \varrho(a_{2u}) \end{cases} \qquad e^-_{1u}\begin{cases} \chi_{12} = \mu(4p-1) \\ \chi_{13} = \varrho(e^-_{1u}) \end{cases}$$

$$e^+_{1g}\begin{cases} \chi_6 = \mu(3d1) \\ \chi_7 = \varrho(e^+_{1g}) \end{cases} \qquad e^+_{2g}\begin{cases} \chi_{14} = \mu(3d2) \\ \chi_{15} = \varrho(e^+_{2g}) \end{cases}$$

$$e^-_{1g}\begin{cases} \chi_8 = \mu(3d-1) \\ \chi_9 = \varrho(e^-_{1g}) \end{cases} \qquad e^-_{2g}\begin{cases} \chi_{16} = \mu(3d-2) \\ \chi_{17} = \varrho(e^-_{2g}) \end{cases}$$

$$e^+_{2u} = \chi_{18} = \varrho(e^+_{2u}) = \varphi(e^+_{2u})$$
$$e^-_{2u} = \chi_{19} = \varrho(e^-_{2u}) = \varphi(e^-_{2u}).$$

Energy Calculation

As shown by ROOTHAAN[1] and LÖWDIN[11], the minimization of the energy, using a molecular wave-function for the ground state which is an antisymmetrized product of LCAO molecular orbitals, leads to the matrix equation

$$\boldsymbol{Fc} = \varepsilon \Delta \boldsymbol{c},$$

where $\boldsymbol{F}$ is the matrix corresponding to the Hamiltonian $\hat{F}$ in the Hartree-Fock scheme, $\boldsymbol{c}$ the coefficients in the molecular orbitals, ε the energy of the molecular orbitals, and Δ the overlap matrix. Since, for our basis set defined in the previous section, $\Delta = \mathbf{1}$, we have

$$\boldsymbol{F}\boldsymbol{c} = \varepsilon \boldsymbol{c},$$

$\boldsymbol{F}$ is as written a 19 by 19 matrix, but due to the symmetry properties it is reduced to undermatrices: one three-dimensional, four two-dimensional, and one one-dimensional. As a whole, we get thus 19 orthogonal solutions, and the eighteen valence electrons will correspond to 9 of these.

The Hamiltonian operator $\hat{F}$ for an electron is given by

$$\hat{F} = \hat{H}^c + \hat{G},$$

where $\hat{H}^c$ (the core Hamiltonian) is made up of the kinetic and potential energy of the electron in the molecular skeleton, and $\hat{G}$ are the sums of the coulomb operators and exchange operators for the molecular orbital set under consideration.

Furthermore we have applied the GOEPPERT-MAYER-SKLAR[12] approximation and obtained for $\hat{H}^c$

$$\hat{H}^c = -\frac{\hbar^2}{2m}\nabla^2 + \sum_g U_g.$$

In other words, $\hat{H}^c$ is decomposed into the kinetic energy and a sum of contributions U_g from the various nuclear cores. Finally we have thus

$$\hat{F} = -\frac{\hbar^2}{2m}\nabla^2 + \sum_{j=0}^{9} U_{c_j} + U_{Fe^{+8}} + \sum_{j=0}^{9} (2\hat{J}_j - \hat{K}_j)$$

with

$$\hat{J}_j \varphi_i(1) = \left(e^2 \int \varphi_j^*(2)\ \varphi_j(2)\ \frac{1}{r_{12}}\, d\tau_2 \right) \varphi_i(1)$$

$$\hat{K}_j \varphi_i(1) = \left(e_2 \int \varphi_j^*(2)\ \varphi_i(2)\ \frac{1}{r_{12}}\, d\tau_2 \right) \varphi_j(1).$$

The calculation then starts with an assumed linear combination of the metal- and ring orbitals, for instance,

$$a_{1g} \begin{cases} \psi_1 = C_{1,1}\ \chi_1 + C_{2,1}\ \chi_2 + C_{3,1}\ \chi_3 \\ \psi_2 = C_{1,2}\ \chi_1 + C_{2,2}\ \chi_2 + C_{3,2}\ \chi_3 \\ \psi_3 = C_{1,3}\ \chi_1 + C_{2,3}\ \chi_2 + C_{3,3}\ \chi_3 \end{cases}$$

$$
\begin{aligned}
a_{2u} &\begin{cases} \psi_4 = C_{4,4}\ \chi_4 + C_{5,4}\ \chi_5 \\ \psi_5 = C_{4,5}\ \chi_4 + C_{5,5}\ \chi_5 \end{cases} \\
e^{+}_{1g} &\begin{cases} \psi_6 = C_{6,6}\ \chi_6 + C_{7,6}\ \chi_7 \\ \psi_7 = C_{6,7}\ \chi_6 + C_{7,7}\ \chi_7 \end{cases} \\
e^{-}_{1g} &\begin{cases} \psi_8 = C_{8,8}\ \chi_8 + C_{9,8}\ \chi_9 \\ \psi_9 = C_{8,9}\ \chi_8 + C_{9,9}\ \chi_9 \end{cases} \\
e^{+}_{1u} &\begin{cases} \psi_{10} = C_{10,10}\ \chi_{10} + C_{11,10}\ \chi_{11} \\ \psi_{11} = C_{10,11}\ \chi_{10} + C_{11,11}\ \chi_{11} \end{cases} \\
e^{-}_{1u} &\begin{cases} \psi_{12} = C_{12,12}\ \chi_{12} + C_{13,12}\ \chi_{13} \\ \psi_{13} = C_{12,12}\ \chi_{12} + C_{13,13}\ \chi_{13} \end{cases} \\
e^{+}_{1g} &\begin{cases} \psi_{14} = C_{14,14}\ \chi_{14} + C_{15,14}\ \chi_{15} \\ \psi_{15} = C_{14,15}\ \chi_{14} + C_{15,15}\ \chi_{15} \end{cases} \\
e^{-}_{2g} &\begin{cases} \psi_{16} = C_{16,16}\ \chi_{16} + C_{17,16}\ \chi_{17} \\ \psi_{17} = C_{16,17}\ \chi_{16} + C_{17,17}\ \chi_{17} \end{cases} \\
e^{+}_{2u}&: \ \psi_{18} = \chi_{18} \\
e^{-}_{2u}&: \ \psi_{19} = \chi_{19}\,.
\end{aligned}
$$

Since $\boldsymbol{F}$ depends upon the coefficients of the filled orbitals through $\hat{G}$, it is necessary to know which orbitals are used for the ground state. We follow most of the authors who have given qualitative and quantitative calculations in favouring the following lowest orbitals

$$\psi_1,\ \psi_2,\ \psi_4,\ \psi_6,\ \psi_8,\ \psi_{10},\ \psi_{12},\ \psi_{14},\ \text{and}\ \psi_{16}$$

as being the orbitals filled for the ground state. The remaining ten orbitals are then considered to be excited orbitals are used for the construction of the excited states.

In our calculation we have taken the a_{1g} orbitals to be of the form

$$
a_{1g} \begin{cases} C_{1,1}\ \chi_1 + C_{3,1}\ \chi_3 \\ \chi_2 \\ C_{1,3}\ \chi_1 + C_{3,3}\ \chi_3 \end{cases}
$$

This assumption, which is in accord with the low value of $S(a'_{1g})$, has also been made by Shustorovich and Dyatkina[4]. Furthermore, some calculations which do not involve this assumption indicate that the neglected coefficients really are small.

In the solution of our scheme we started with a set of coefficients nearly identical to Shustorovich and Dyatkina's, and after at most five iterations a set of self consistent solutions had been arrived at. Approximate convergence of the solutions was assumed to occur when agreement was obtained within 0.1 per cent of the proceeding set of coefficients.

TABLE 2. Molecular orbitals of ferrocene. This work.

Symmetry	Filled molecular orbitals		Energies of empty orbitals eV
	Form	Energies of orbitals eV	
a_{1g}	$0.633\ \mu(4s) + 0.774\ \varrho(a_{1g})$	-20.15	27.51
	$\mu(3do)$	-14.03	–
a_{2u}	$0.471\ \mu(4po) + 0.882\ \varrho(a_{2u})$	-17.77	12.39
e_{1g}	$0.454\ \mu(3d1) + 0.891\ \varrho(e^{+}_{1g})$	-12.48	8.96
e_{1u}	$0.591\ \mu(4p1) + 0.807\ \varrho(e^{+}_{1u})$	-14.74	24.64
e_{2g}	$0.898\ \mu(3d2) + 0.440\ \varrho(e^{+}_{2g})$	-10.92	3.98
e_{2u}		–	1.43

TABLE 3. Molecular orbitals for ferrocene according to SHUSTOROVICH and DYATKINA[4]. The energies in parenthesis are taken from YAMAZAKI[2].

Symmetry	Filled molecular orbitals		
	Form	Energies of orbitals eV	
a_{1g}	$0.49\ \mu(4s) + 0.87\ \varrho(a_{1g})$	-16.05	(-16.80)
	$\mu(3d0)$	-8.44	(-8.57)
a_{2u}	$0.10\ \mu(4p0) + 0.99\ \varrho(a_{2u})$	-13.74	(-15.45)
e_{1g}	$0.37\ \mu(3d1) + 0.93\ \varrho(e^{+}_{1g})$	-11.02	(-8.90)
e_{1u}	$0.59\ \mu(4p1) + 0.81\ \varrho(e^{+}_{1u})$	-12.62	(-7.68)
e_{2g}	$0.85\ \mu(3d2) + 0.52\ \varrho(e^{+}_{2g})$	-6.39	(-7.87)

Since, according to Koopman's[13] theorem the ionization potential is given as the energy of the highest occupied molecular orbital, we have $I_p = 10.92\ eV$. This is considerably higher than Yamazaki's value of 7.68 eV and the Russians' value of 6.39 eV.

The only—very unreliable—value for the ionization potential found in the literature[14] is based upon a mass spectrum measurement. It is 7.05 eV.

Excited States

The lowest electronic configuration for ferrocene is seen to be

$$(a_{1g})^2\ (a_{2u})^2\ (e_{1u})^4\ (a'_{1g})^2\ (e_{1g})^4\ (e_{2g})^4,$$

where the orbitals have been written down in order of increasing energy. From the equations of Roothaan, the excited orbitals are likewise obtained. They are, in order of increasing energy,

$$(e_{2u})\ (e^*_{2g})\ (e^*_{1g})\ (a^*_{2u})\ (e^*_{1u})\ (a^*_{1g}).$$

Following ROOTHAAN[1] we now construct the excited states of the molecule using these orbitals. Since an excitation of two or more electrons has very high energy, we only consider singly excited configurations. Brillouin's theorem then ensures that, when the *SCF* orbitals are used, no interaction takes place between the ground- and the excited-states.

It is possible to construct five singlet states possessing an energy less than 10 *eV* above the ground state. These states occur by excitation of the e_{1g} and e_{2g} electrons to the e_{2u} and e^*_{2g} orbitals. Using the nomenclature $(a \rightarrow b)$ to indicate that an electron has been excited from orbital a to orbital b, we get the excited states

$$A_{2g}(e_{2g} \rightarrow e^*_{2g}),\ A_{1u}(e_{2g} \rightarrow e_{2u}),\ E_{1u}(e_{2g} \rightarrow e_{2u})$$
$$A_{2u}(e_{2g} \rightarrow e_{2u}),\ E_{1u}(e_{1g} \rightarrow e_{2u}).$$

All of these states possess an energy less than the found ionization potential. The excitation energies are calculated using the standard methods[1,15]. We found with no configuration interaction between the two $^1E_{1u}$ states:

$A_{2g}(e_{2g} \rightarrow e^*_{2g})$: 5.38 *eV*
$A_{1u}(e_{2g} \rightarrow e_{2u})$: 5.44 *eV*
} orbitally not allowed transitions

$E_{1u}(e_{2g} \rightarrow e_{2u})$: 5.75 *eV*
$A_{2u}(e_{2g} \rightarrow e_{2u})$: 7.06 *eV*
$E_{1u}(e_{1g} \rightarrow e_{2u})$: 9.35 *eV*
} orbitally allowed transitions

Since the electric dipole vector transforms as $A_{2u}(\parallel)$ and $E_{1u}(\perp)$ with respect to the five-fold axis of the molecule, the first two excited states are orbitally not allowed transitions, whereas the last three states are orbitally allowed transitions.

The absorption spectrum of ferrocene shows one strong and two weaker band systems[16]. Denoting the molar extinction coefficient ε we find:

First band:	2.75 eV	($\varepsilon = 83.46$)
	3.1 eV	($\varepsilon = 50.0$)
Second band:	3.82 eV	($\varepsilon = 51.5$)
Third band:	4.77 eV	($\varepsilon = 5140$)
	6.17 eV	($\varepsilon = 53460$)

It is thus reasonable to make the assignments

	Experimental	Calculated
Band 1	2.75 eV	5.38 eV (A_{2g})
	3.1 eV	
Band 2	3.82 eV	5.44 eV (A_{1u})
Band 3	4.77 eV	5.75 eV (E_{1u})
	6.17 eV	7.06 eV (A_{2u})
		9.35 eV (E_{1u})

Discussion

The following picture results from our calculation. A system of 12 electrons occupies the strongly bonding orbitals: a_{1g}, a_{2u}, e_{1u} and e_{1g}. The corresponding antibonding orbitals have so high energies that they are of no importance for the lowest excited states. The six remaining valence electrons then occupy the non-bonding a'_{1g} orbital ($3d_{z^2}$) and the weakly bonding (e_{2g}) orbitals.

Our calculations show that no hybridization of the $4s$ and $3d_{z^2}$ orbitals occurs, as has been postulated by MOFFITT[6] and by DUNITZ and ORGEL[7]. A calculation of the charge distribution in the molecule, performed according to MULLIKEN[17], gave the result that the iron atom has a negative charge of -0.69 units in the ground state of the molecule. Consequently, each of the rings possesses a charge of $+0.35$. SHUSTOROVICH and DYATKINA[4] found exactly the same numbers, but with an *opposite* sign. The reversal of the sign in our calculation is due to the more contracted orbitals used in our work. The positive charge found upon the rings is supported by certain substitution reactions in the chemistry of ferrocene[5].

The absorption spectrum of ferrocene is due to transitions between the e_{2g} orbitals and the ring orbitals (charge transfer spectrum) and from the

e_{2g} orbital to the antibonding orbital e_{2g}^{*}. The twelve bonding electrons play a role similar to the σ-electrons in benzene; the a_{1g}' and e_{2g} electrons which are of minor importance as far as the bonding is concerned can thus be treated in a manner similar to the treatment of the π-electrons in benzene.

All neutral dicyclopentadienyl complexes possess this closed system of 12 electrons. These compounds only differ in the occupancy number of the a_{1g}' and e_{2g} orbitals. Further use of the e_{2u} (or e_{2g}^{*}) orbital for the Co and Ni compound allows us to write down the electronic structures for all of the neutral metal sandwiches. All these structures are consistent with the magnetic data[5]. The three unpaired spins found in $V(Cp)_2$ are specially noteworthy. Using the ferrocene orbitals of this work we have calculated that the ground state for $V(Cp)_2$ should be $^2E_{2g}$, but with an $^4A_{2g}$ state placed at 0.61 *eV*. A slight modification of the orbitals in going from iron to vanadium should thus be sufficient to explain the experimental findings.

$Ti(Cp)_2$:	(12) $(a_{1g}')^2$	$S=0$
$V(Cp)_2$:	(12) $(a_{1g}')^1(e_{2g})^2$	$S=3/2$
$Cr(Cp)_2$:	(12) $(a_{1g}')^2(e_{2g})^2$	$S=1$
$Fe(Cp)_2$:	(12) $(a_{1g}')^2(e_{2g})^4$	$S=0$
$Co(Cp)_2$:	(12) $(a_{1g}')^2(e_{2g})^4(e_{2u})^1$	$S=1/2$
$Ni(Cp)_2$:	(12) $(a_{1g}')^2(e_{2g})^4(e_{2u})^2$	$S=1$

The electronic states in Ti, V, Cr and Fe are in agreement with those proposed by LIEHR and BALLHAUSEN[18], but differ from their assignment in $Co(Cp)_2$ and $Ni(Cp)_2$.

It is interesting to not that the spectra of the other sandwich compounds have a close resemblance to that of ferrocene[16]. The view that the a_{1g}' and e_{2g} electrons are the "chemical electrons" makes this fairly understandable, and is at the same time in agreement with the basicity features of these compounds[23].

Even though the exact numbers in a semi-empirical theory as that outlined above must be considered with some caution, we believe that the essential features in the electronic structure of ferrocene have been given correctly here.

Appendix: Evaluation of Integrals

1) Evaluation of the "core" integrals.

We have $H^c = \frac{-\hbar^2}{2m} \nabla^2 + U_{Me} + \sum_r U_r$,

where U_{Me} is the potential from the metal nucleus and the closed argon shell, while $\sum_r U_r$ is the contribution from the 10 carbon nuclei.

Now, the diagonal matrix elements for the metal orbitals are called $\alpha(3d0)$, $\alpha(3d1)$ . . . and those for the ringorbitals $\alpha(a_{1g})$, $\alpha(e_{1g})$ . . . . The non-diagonal matrix elements are called $\beta(a_{1g})$, $\beta(e_{1g})$ and so forth. We have, for instance,

$$\alpha(3d0) = \int \psi^*(3d0)\, H^c\, \psi(3d0)\, d\tau$$
$$\beta(a_{1g}) = \int \psi^*(4s)\, H^c\, \psi(a_{1g})\, d\tau .$$

For the $\alpha(3d0)$ integral we get, by expanding,

$$\alpha(3d0) = \int \psi^*(3d0) \left\{ -\frac{\hbar^2}{2m} \nabla^2 + U_{Me} \right\} \psi(3d0) d\tau + \int \psi^*(3d0) \sum_r U_r \psi(3d0)\, d\tau$$

or

$$\alpha(3d0) = \alpha_{Me}(3d0) + \sum_r \alpha_r(3d0).$$

The integrals of the type $\alpha_{Me}(3d0)$ can be estimated from the spectrum of iron with the help of Watson's calculations[3]. We get

$$\alpha_{Me}(3d) = -5.554 \text{ a.u.}$$
$$\alpha_{Me}(4s) = -2.654 \text{ a.u.}$$
$$\alpha_{Me}(4p) = -2.474 \text{ a.u.}$$

The integrals of the type α_r are evaluated treating U_r as the potential coming from a π electron with opposite sign[12, 20]. The integral is then seen to be a usual Coulomb integral, the evaluation of which is given later. We get in atomic units a.u. = 27.210 eV

$\alpha(4s)$	$\alpha(3d0)$	$\alpha(3d1)$	$\alpha(3d2)$	$\alpha(4p0)$	$\alpha(4p1)$
−5.057	−8.260	−8.250	−8.223	−5.032	−4.798

In the evaluation of the integrals $\alpha(a_{1g})$ etc. we write for $\hat{H}^c$

$$\hat{H}^c = -\frac{\hbar^2}{2m}\nabla^2 + U_0 + \sum_{r=0}^{4} U_r + \sum_{r=0}^{4} U_r' + U_{Me},$$

where U_0, $U_1 \ldots U_4$ are the potentials from ring A, and $U_0' \ldots U_4'$ the potentials coming from ring B. Using a value of

$$\int \pi_0^* \left\{ -\frac{\hbar^2}{2m}\nabla^2 + U_0 \right\} \pi_0 \, d\tau = -0.4146 \text{ a.u.}$$

according to MULLIKEN, and the tables by PARR and CRAWFORD[19], and furthermore assuming a value of the "resonance integral"

$$\beta = \int \pi_0 \hat{H}^c \pi_1 \, d\tau = -0.0878 \text{ a.u., we get}$$

$$\alpha(a_{1g}) = \alpha(a_{2u}) = -4.623 \text{ a.u.}$$

$$\alpha(e_{1g}) = \alpha(e_{1u}) = -4.502 \text{ a.u.}$$

$$\alpha(e_{2g}) = \alpha(e_{2u}) = -4.305 \text{ a.u.}$$

Integrals of the type $\beta(l) = \int \psi^*(l) \hat{H}^c \varphi(l) \, d\tau$ are treated in the following way[20]. By symmetry, we must have

$$\beta(l) = \sqrt{10} \int \psi^*(l) \hat{H}^c \pi_0 \, d\tau .$$

With $\pi_0 = \bar{\pi}_0 - \frac{1}{2}\sum_r S_{r0}\bar{\pi}_r$, and since $\beta(l)$ is less than the corresponding $\alpha(l)$ values, we can put

$$\beta(l) = \sqrt{10} \int \psi^*(l) \hat{H}^c \bar{\pi}_0 \, d\tau .$$

We have further

$$\left\{ -\frac{\hbar^2}{2m}\nabla^2 + U_{Me} \right\} \psi(l) = \alpha_{Me}(l) \, \psi(l)$$

$$\left\{ -\frac{\hbar^2}{2m}\nabla^2 + U_0 \right\} \bar{\pi}_0 = W_{2p} \bar{\pi}_0 .$$

By expansion we then get

$$\beta(l) = \frac{1}{2} S(l) \left[W_{2p} + \alpha_{Me}(l) \right] + \frac{1}{2}\sqrt{10} \int \psi^*(l) \left[U_{Me} + U_0 \right] \bar{\pi}_0 \, d\tau$$

$$+ \sqrt{10} \int \psi^*(l) \sum_{r=1}^{9} U_r \bar{\pi}_0 \, d\tau .$$

The evaluation of the last integral follows from what has been said previously, but the calculation of the second integral requires a rotation of the coordinate

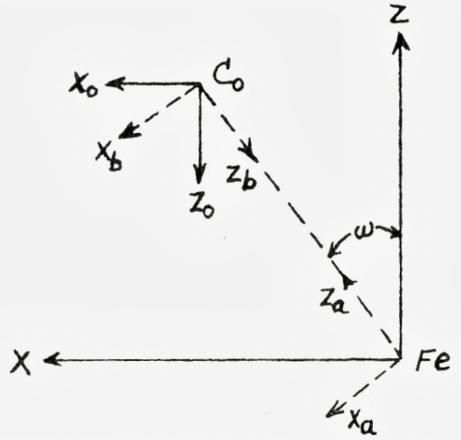

Fig. 7. Definition of local coordinate systems, x_a, y_a, z_a and x_b, y_b, z_b. $\omega = 36.^{\circ}8$.

systems. Transforming the coordinate systems as indicated in Fig. 7, we get, for instance, for the transformed metal orbitals:

$$\psi(3d_{z^2}) = (\cos^2\omega - \frac{1}{2}\sin^2\omega)\,[3d_{z^2}]_a - \frac{1}{2}\sqrt{3}\sin 2\omega\,[3d_{xz}]_a + \frac{1}{2}\sqrt{3}\sin^2\omega\,[3d_{x^2-y^2}]_a$$

$$\psi(3d_{xz}) = \frac{1}{2}\sqrt{3}\sin 2\omega\,[3d_{z^2}]_a + \cos 2\omega\,[3d_{xz}]_a - \frac{1}{2}\sin 2\omega\,[3d_{x^2-y^2}]_a$$

$$\psi(3d_{yz}) = \cos\omega\,[3d_{yz}]_a - \sin\omega\,[3d_{xy}]_a$$

$$\psi(3d_{x^2-y^2}) = \frac{1}{2}\sqrt{3}\sin^2\omega\,[3d_{z^2}]_a + \frac{1}{2}\sin 2\omega\,[3d_{xz}]_a + \frac{1}{2}(1+\cos^2\omega)\,[3d_{x^2-y^2}]_a$$

$$\psi(3d_{xy}) = \cos\omega\,[3d_{xy}]_a + \sin\omega\,[3d_{yz}]_a$$

$$\psi(4s) = [4s]_a$$

$$\psi(4p_z) = -\sin\omega\,[4p_x]_a + \cos\omega\,[4p_z]_a$$

$$\psi(4p_x) = \cos\omega\,[4p_x]_a + \sin\omega\,[4p_z]_a$$

$$\psi(4p_y) = [4p_y]_a$$

The transformation of the $2p_z$ orbitals in the $(x_b\ y_b\ z_b)$ coordinate system proceeds exactly in the same way.

The remaining integrals are now calculated with the help of these transformed orbitals, using a two-center potential; a charge of $+8$ upon the metal atom and a charge of $+1$ on the carbon atom. In this approximation, we get

$\beta(a_{1g})$	$\beta(e_{1g})$	$\beta(e_{2g})$	$\beta(a_{2u})$	$\beta(e_{1u})$
-3.086 au.	-1.143 au.	-0.603 au.	-1.475 au.	-2.629 au.

2) Evaluation of two-electron integrals.

The two-electron integrals occur in the Roothaan scheme in the evaluation of the matrix $\boldsymbol{G}$, whose elements are given by

$$G_{pq} = \sum_{r,s=1}^{m} p_{rs} \left\{ [pq\,|\,rs] - \frac{1}{2}[ps\,|\,rq] \right\},$$

where

$$[pq\,|\,rs] = \int\int \chi_p^*(1)\,\chi_r^*(2)\,\frac{e^2}{r_{12}}\,\chi_q(1)\,\chi_s(2)\,d\tau_1\,d\tau_2,$$

and

$$p_{rs} = 2\sum_{i=1}^{n} C_{ri}^* C_{si}$$

are the elements of the charge and bond order matrix.

Utilizing the symmetry properties of χ_r and χ_s and taking the "zero-differential overlap" as a criterium for the non-vanishing of the matrix elements[20], all the matrix-elements in G_{pq} can be reduced to "Coulomb" integrals. Using the method of ROOTHAAN[21] these were evaluated in a spheroidal coordinate system. In this way, all the integrals are reduced to linear combinations of so-called basis integrals. They were evaluated in the following way. First, one of the electronic coordinates was integrated analytically. The remaining function was then integrated numerically. Use was made of the molecular zeta function of COULSON and BARNETT[22]; we want to express our sincere thanks to Professor Coulson for having placed his tables of this function at our disposal.

Our final result for the nine Coulomb integrals is

$J(3d0,2p0) = 0.2706$ a.u.	$J(4s,2p0) = 0.2403$ a.u.
$J(3d\pm1,2p0) = 0.2696$ a.u.	$J(4p0,2p0) = 0.2559$ a.u.
$J(3d\pm2,2p0) = 0.2669$ a.u.	$J(4p\pm1,2p0) = 0.2325$ a.u.

1 a.u. $= 27.210\ eV$.

References

(1) C. C. J. Roothaan: Revs. Mod. Phys. **23**, 69 (1951).
(2) M. Yamazaki: J. Chem. Phys. **24**, 1260 (1956).
(3) R. E. Watson: Techn. Rep. No 12, Solid-State and Molecular Theory Group, M.I.T. (1959). Phys. Rev. **119**, 1934 (1960).
(4) E. M. Shustorovich and M. E. Dyatkina: Doklady Akademii Nauk. SSSR. **128**, 1234 (1959).
(5) G. Wilkinson and F. A. Cotton: Prog. Inorg. Chem. **1**, Interscience, New York (1959).
(6) W. Moffitt: J. Am. Chem. Soc. **76**, 3386 (1954).
(7) J. D. Dunitz and L. E. Orgel: J. Chem. Phys. **23**, 954 (1955).
(8) C. Zener: Phys. Rev. **36**, 51 (1930).
(9) E. A. Seibold and L. E. Sutton: J. Chem. Phys. **23**, 1967 (1955).
(10) R. S. Mulliken, C. A. Rieke, D. Orloff and H. Orloff: J. Chem. Phys. **17**, 1248 (1949).
(11) P. O. Löwdin: Phys. Rev. **97**, 1490 (1955).
(12) M. Goeppert-Mayer and A. L. Sklar: J. Chem. Phys. **6**, 645 (1938).
(13) T. Koopman: Physica **1**, 104 (1933).
(14) L. Friedman, A. P. Irsa and G. Wilkinson: J. Am. Chem. Soc. **77**, 3689 (1955).
(15) R. S. Mulliken: J. chim. phys. **46**, 497, 675 (1949).
(16) G. Wilkinson: Private communication.
(17) R. S. Mulliken: J. Chem. Phys. **23**, 1833 (1955).
(18) A. D. Liehr and C. J. Ballhausen: Acta Chem. Scand. **11**, 207 (1957).
(19) G. Parr and B. L. Crawford: J. Chem. Phys. **16**, 1049 (1948).
(20) R. Pariser and R. G. Parr: J. Chem. Phys. **21**, 466 and 767 (1953).
(21) C. C. J. Roothaan: J. Chem. Phys. **19**, 1445 (1951).
(22) M. P. Barnett and C. A. Coulson: Phil. Trans. Roy. Soc. **243**, 221 (1951).
(23) C. J. Ballhausen and J. P. Dahl: Acta Chem. Scand. To be published.

Indleveret til selskabet den 28. februar 1961.
Færdig fra trykkeriet den 4. september 1961.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 6

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 6 (1961)

ENERGY LOSS AND STRAGGLING OF PROTONS AND DEUTERONS

BY

LIS PERCH NIELSEN

København 1961
i kommission hos Ejnar Munksgaard

Synopsis

The energy loss and straggling of protons and deuterons have been measured in the energy range from 1.5 to 4.5 MeV. The elements investigated are Be, Al, Ni, Cu, Ag, and Au. The results are plotted in such units that all the points are expected to fall on a single curve. This is found to be the case and the shape of the curve is in good agreement with theoretical expectations.

Printed in Denmark
Bianco Lunos Bogtrykkeri A-S

I. Introduction

The present experimental investigation is concerned with the energy loss and straggling which protons and deuterons suffer when they penetrate foils of various elements.

The particles were accelerated in the 4.5 MV electrostatic accelerator at the Institute for Theoretical Physics in Copenhagen. A magnetic spectrograph was used as a precision instrument for the energy determinations.

In the energy range below 2 MeV numerous stopping power investigations have been carried out previously, but in the range from 2 to 10 MeV the experimental data are rather scarce. At higher energies ($\gtrsim$ 10 MeV) many investigations have been performed by means of cyclotrons. However, in general, the latter measurements are made only at the fixed energy which the cyclotron in question yields, and consist in a determination of the stopping power of various elements relative to a given standard. Air or aluminum are often used as such standards[6, 13]; this is unfortunate since, in particular for air, the results obtained from different experiments vary considerably. For this reason, it is difficult to combine the various results to an accurate description of the stopping phenomenon.

The present measurements are performed on metal foils of beryllium, aluminum, nickel, copper, silver, and gold, and the investigations cover the energy range from 1.5 to 4.5 MeV.

II. Summary of Theory

The penetration of charged particles through matter has been studied theoretically by many authors. The topic has been surveyed by NIELS BOHR[4] who, in particular, has discussed the conditions under which the various approximations to the problem can be applied.

The slowing down of a proton or a deuteron is caused by electronic collisions in which the energy is transferred to individual electrons in the atoms, resulting in atomic excitation and ionization processes. In a nuclear collision, the momentum is taken up by the target atom as a whole and,

because of the much larger mass involved, such collisions do not contribute essentially to the energy loss. They do, however, give rise to the multiple scattering which the particle suffers by the penetration, whereas the electronic collisions are of minor importance for this effect.

The energy loss. For the bombarding energies employed in the present investigation, the electronic collisions correspond to small values of the collision index $\varkappa$, i. e.,

$$\varkappa = 2\frac{Z_1 e^2}{\hbar v} < 1\,, \tag{1}$$

where $Z_1 e$ and v are the charge and velocity, respectively, of the bombarding particles. Under such circumstances the Born approximation can be employed, and on this basis BETHE[1] has shown that the average energy loss dE per range interval dR for non-relativistic particles is given by

$$\left.\begin{aligned} \frac{dE}{dR} &= \frac{4\pi e^4 {Z_1}^2}{mv^2} NZ_2 L\,, \text{ where} \\ L &= \log_e\left\{\frac{2mv^2}{I}\right\} \end{aligned}\right\} \tag{2}$$

In these expressions, e and m are the charge and mass of an electron, whereas N represents the density of the atoms in the target material which has the atomic number Z_2.

The energy I is an average over the excitation and ionization energies for the electrons in the target atoms. The average involves only those electrons which contribute to the stopping, i. e., electrons which have orbital velocities smaller than or comparable to the projectile velocity v. This implies that I will depend on the projectile energy E, unless

$$E \rangle\rangle \frac{A_1 M_0}{m} I_s \text{ for all } s, \tag{3}$$

where $A_1 M_0$ is the mass of the projectile, M_0 the nuclear mass unit, and I_s are the ionization energies of the various electrons in the atoms.

For very high (but non-relativistic) energies, where these conditions are all fulfilled, BLOCH[3] has shown that employment of the Thomas-Fermi statistical model of the atom leads to an average excitation potential

$$I = Z_2 I_0\,, \tag{4}$$

where I_0 is a constant of the order of the Rydberg energy $R_y = 13.6$ eV.

For the K-electrons the inequality (3) requires that even when bombarding an element as light as aluminum, the proton energy should be larger than 4 MeV. This means that, in the present investigation, the K-electrons of the heavier elements do not yield any significant contribution to the stopping, and that for aluminum one has to apply a correction to the simple expression (4) corresponding to a velocity dependence given by

$$I = Z_2 I_0 \exp\left\{\frac{C_K}{Z_2}\right\}, \tag{5}$$

where I_0 is the constant found at higher energies. BETHE and WALSKE[2,14] have calculated this C_K-correction and they find for low bombarding energies that C_K is negative, whereas for energies in the transition region (e. i., corresponding to an equality sign in (3) for the K-shell) it passes through a positive maximum before it goes to zero when (3) becomes valid.

For lower energies or heavier elements similar corrections would have to be applied also to other shells, but such calculations are only available for the L-shell[15]. However, it has been shown by LINDHARD and SCHARFF[11,12] that, to the extent the Thomas-Fermi model can be applied, one should expect the function L to be dependent on v and Z_2 in such a manner that

$$L = L\{x\}, \text{ where } x = \frac{v^2}{v_0^2 Z_2} \tag{6}$$

and

$$v_0 = e^2/\hbar .$$

It is evident that this is true for Bloch's solution for large values of x, where

$$L = \log_e\left\{\frac{2mv^2}{Z_2 I_0}\right\} = \log_e\left\{\frac{4R_y}{I_0}\right\} + \log_e\{x\}, \tag{7}$$

but (6) holds also in the general case where the function L has not been calculated explicitly.

The result that L to a first approximation depends only on a single parameter is of great value, as it makes it possible to present the experimental data in a comprehensive form which is well suited for interpolations to other elements than those actually measured.

The energy straggling. As a consequence of the statistical nature of the collision processes, not all the particles in a mono-energetic beam will lose the same amount of energy when they penetrate a foil of a definite thickness ΔR.

The standard deviation Ω of the energy distribution is called the energy straggling. The main contribution to Ω comes from the more violent collisions by which the electrons are given kinetic energies close to the maximum energy $4\,mv^2$ which a free electron can obtain in a collision with the projectile. This energy is large compared to I_s when the inequalities (3) are satisfied, and under these conditions one finds (ref. 4))

$$\Omega^2 = Z_1^{\,2} \cdot 4\pi e^4 Z_2 N \Delta R. \tag{8}$$

For lower energies where (3) is no longer valid, LINDHARD and SCHARFF[11] have shown that Ω can be expressed in terms of the function L which, according to equation (2), determines the average energy loss. They find that

$$\Omega^2 = Z_1^{\,2} 4\pi e^4 Z_2 N \Delta R \cdot \frac{1}{2} L\{x\} \tag{9}$$

$$\text{for } L\{x\} \lesssim 1.$$

A smooth transition between the two approximations is expected to take place for $L\{x\} \simeq 2$, but a more precise estimate is not available in this region.

III. Experimental Procedure

A thin layer of gold deposited on a carbon foil was bombarded with, e. g., the proton beam. The protons scattered at a backward angle of 145 degrees were passed through a broad-range magnetic spectrograph[8] and recorded by means of a photographic emulsion placed along the focal plane. The plate was exposed twice, first with a foil inserted in the primary beam and then, immediately afterwards, with the foil removed. As an example, Fig. 1 shows the corresponding groups of protons scattered from the gold layer. Protons scattered from the carbon backing have much lower energies at backward scattering angles, and do not disturb the measurements. The narrow peak obtained without the stopping foil shows that the energy spread of the primary beam is less than ± 0.1 per cent. The peak obtained with the foil inserted in the beam is shifted towards a lower energy and has an increased width. These effects are caused by energy loss and straggling in the stopping foil.

The two exposures corresponded to the same number of primary protons, as measured by collecting the beam in a Faraday chamber behind the carbon foil and recording the accumulated charge by means of a beam

integrator. The total number of tracks in each of the two peaks should thus be very nearly equal. It is determined by the thickness of gold scatterer and the bombarding conditions. In most of the exposures one aimed at

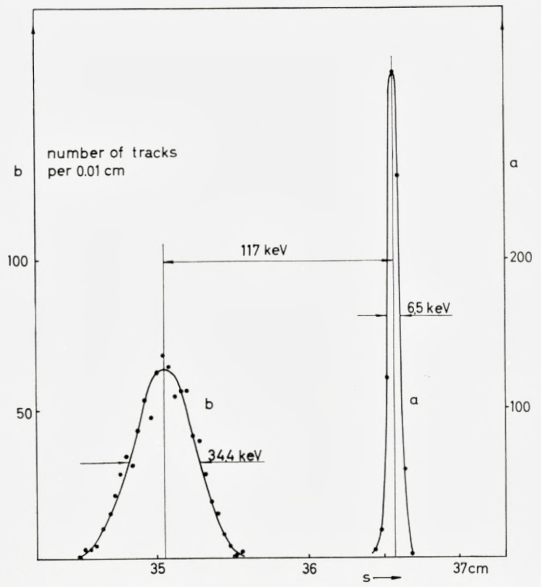

Fig. 1. Spectrum of protons scattered from a thin gold target, a) without foil and b) with a 3.3 mg/cm² Au-foil inserted in the 3 MeV proton beam. The spectrum was obtained by means of a magnetic spectrograph and the particles were recorded in a photographic emulsion.

about 2000 tracks, and this number could be obtained in a few minutes by collecting a charge of the order of 10 micro-coulomb in the Faraday chamber.

Determination of the energy loss and straggling. The spectrograph was calibrated by recording α-particles from $_{84}Po^{210}$ [5] at various field strengths, as measured by a nuclear induction Gauss-meter. By means of the calibration curve the average energy can be determined for each of the two peaks in Fig. 1, and from these energies the mean energy of the particles in the investigated foil can be computed as well as the average energy loss. This involves a correction for the recoil energy lost in the gold scatterer (1.8 per cent for protons scattered through 145 degrees) and, strictly speaking, one should also take into account that the energy scale is non-linear; however, the widths of the peaks are so small that this effect is quite negligible.

In a similar way, the energy straggling may be obtained from the standard deviations of the two distributions. The width of the broad curve is due mainly to the straggling phenomenon and the distribution may be expected to be approximately Gaussian. By plotting the integral distribution, i. e., the area $H\{s\}$ indicated in Fig. 2, one should therefore obtain an S-shaped curve with a steepest slope proportional to the reciprocal of the

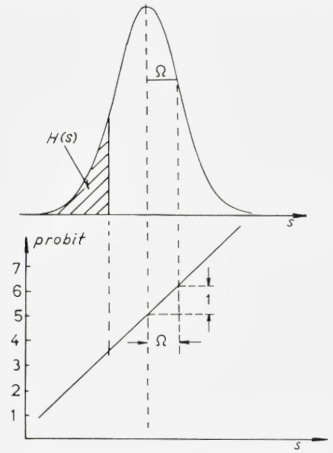

Fig. 2. Determination of the standard deviation of a Gaussian distribution by means of 'probits' (cf. section III).

standard deviation. In order to exploit all the points for the determination of this slope, the curve can be transformed to a straight line with the same slope by plotting the so-called probits (cf. Fig. 2) which represent a linear transform of the integrated Gaussian (cf., e. g., ref. 9).

From the standard deviation Ω_2 determined in this way for the energy distribution corresponding to the broad peak, one finds the energy straggling Ω itself by subtracting the contributions from other effects. The measured distribution results from a folding of the straggling curve with the curve which would be obtained if the straggling phenomenon was absent. Fortunately, the standard deviations add up geometrically, i. e., if we denote the standard deviation of the latter curve by Ω_0, then

$$\Omega^2 = \Omega_2^2 - \Omega_0^2 . \tag{10}$$

Hence, the corrections have a relatively small influence and for this reason the exact magnitude of Ω_0 is not important. One contribution to Ω_0 comes from the finite resolution of the spectrograph and the analyzing

magnet of the accelerator. It is given by the standard deviation Ω_1 of the reference curve ('a' in Fig. 1) and it can, for the present purpose, be estimated sufficiently well from the directly measured half-width.

Another contribution to Ω_0 comes from the multiple scattering in the stopping foil which smears out the beam spot on the target because of the spacing D needed between foil and target (D = 1.8 cm, cf. Fig. 3). The

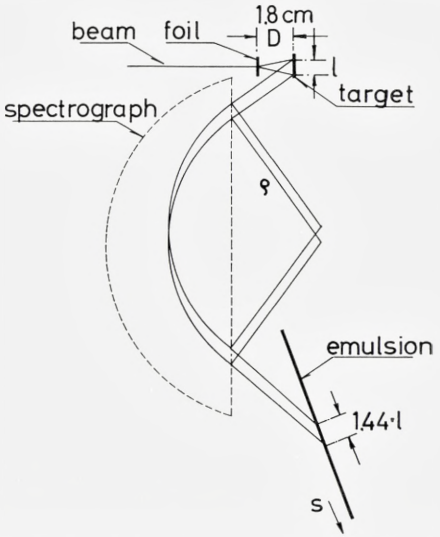

Fig. 3. Sketch of the spectrograph (not to scale), showing the broadening effect caused by multiple scattering in the stopping foil. The size of the effect is determined by the distance 'D' between foil and target and it is further magnified 1.44 times by the spectrograph. Two typical orbits are indicated.

magnitude Ω_s of the contribution from this effect was not measured directly, but it can be estimated from earlier measurements (cf., e. g., ref. 10) and from theory by the following considerations.

The standard deviation σ for the projected angular distribution is given by [4]

$$\sigma = \frac{Z_1 Z_2}{\sqrt{2}} \cdot \frac{[\pi e^4 \; N \; \Delta R \; \log_e \{\bar{n}\}]^{1/2}}{E}, \tag{11}$$

where $\bar{n}$ is the average number of nuclear collisions which a particle encounters by the passage of the foil. Hence Ω_s can be calculated from the expression

$$\Omega_s = \frac{dE}{ds} \cdot F \cdot D \cdot \sigma, \tag{12}$$

where $F = 1.44$ is the (constant) magnification and

$$\frac{dE}{ds} = 0.87 \frac{E}{\varrho} \tag{13}$$

is the reciprocal energy dispersion of the spectrograph.

Since for most of the exposures the magnetic field was adjusted to give a radius of curvature $\varrho \simeq 30$ cm, it is found from equations (11), (12), (13), and (8) that

$$\frac{\Omega_s}{\Omega} \simeq 2.7 \cdot 10^{-2} \sqrt{Z_2 \log_e \{\bar{n}\}}, \tag{14}$$

if $L\{x\} \gtrsim 1$. Hence, the correction factor is nearly independent of ΔR and E which enter only through the logarithmic term under the square root. In accordance with the theory it is found experimentally[10] that, for foil thicknesses of the order of mg/cm^2 and energies of the order of MeV, the square root decreases from a value of 15 for gold to 12 for copper, whereas for aluminum it is expected to be as low as 9. Because of the geometrical addition of the standard deviations, these values imply that even in the case of gold the correction to Ω_2 amounts only to approximately 8 per cent.

In order to ensure that the actual Ω_s correction was not underestimated, a few exposures were made with a target where the thin gold layer was confined to a narrow line, only 1 mm wide. A scatterer of this shape acts as a line source in the spectrograph even when a foil is introduced in the beam, and hence in this case no Ω_s correction is needed (cf. section IV). The energy straggling for the observed particles should be the same in the two geometries since the stopping and scattering in the foil is caused by two different processes (cf. section II) and therefore not correlated.

In addition to the above mentioned contributions to the widths of the observed peaks, one has to consider the effect of inhomogeneities in the investigated foils, and the degree of homogeneity of course depends on the technique by which the foils have been produced.

Preparation of the foils and determination of their thicknesses. The main source of error in stopping experiments comes from the difficulties involved in producing clean and very homogeneous foils. Great care is

needed in the preparation, and it is essential to check the homogeneity and cleanness of the foils sufficiently well to make the experimentally determined weight per unit area an accurate measure of ΔR.

The foils were weighed on a balance which could be read with an accuracy better than $\pm 10\ \mu g$. The absolute calibration of the balance was checked against a standard weight. The lightest foil weighed about 3 mg and the relative accuracy of the weighings was therefore better than 1 per cent.

The areas were computed from the linear dimensions of the foils which for the larger ones were measured by means of a movable microscope table calibrated to better than 10^{-2} mm. The measured areas were about 1 cm^2 or larger, and the relative uncertainty was less than 1 per cent.

The beryllium foils were not made with the present experiment in mind. They were rather thin and it was therefore necessary to use several layers together. The foils had a somewhat irregular shape, and for this reason the areas were determined by making blue prints on a homogeneous piece of paper. The figures were cut out and weighed relative to a piece of known area.

The aluminum foils were very uniform, rolled foils with a stated purity of 99.6 per cent. The impurities were mainly iron and silicon and the correction to the stopping power was therefore negligible. The nickel foils and some of the copper foils were rolled foils, produced commercially. The stated purity was better than needed for the present experiments.

The remaining copper foils, and all foils of silver or gold were made by evaporation in vacuum from a heated tungsten ribbon. The metal was evaporated onto a glass plate which had been prepared in the following way. First it was cleaned in sodium hydroxide and then a solution of polystyrene in chloroform was poured over it. After the chloroform had evaporated, the plate was left with a thin coating of polystyrene which was used as a basis for the evaporated metal. By means of a razor blade, the metal foil was then cut into rectangular pieces, each approximately 1 cm^2, which is the size of the standard frames used. Because of the polystyrene film, the foils came off the glass quite easily when a drop of water was added. Subsequently, the polystyrene was removed by dissolving it in chloroform. The areas of the rectangular pieces were measured both before they were taken off the glass and after the final preparation. They showed no tendency to shrink if the cuts were not made before the foil had reached room temperature after the evaporation.

The purity of the evaporated foils was checked by employing the same material for production of a very thin evaporated target from which protons could be scattered elastically and measured in the spectrograph. Each

contaminant gives rise to an elastic peak in the spectrum, and from the height of these peaks very small relative concentrations of the impurities can easily be determined. It was found that the gold did not contain any impurity large enough to justify a correction. Some of the silver foils contained 1.3 per cent of copper and this implies a correction to the energy loss of approximately 0.3 per cent, which is somewhat less than the estimated uncertainty of the measurements. The correction was included because the deviation was one-sided.

The foils were selected with respect to homogeneity by weighing neighbouring foils from the same evaporation, and if the deviations in thickness exceeded 2 per cent, the foils were not used. In order to reveal more localized inhomogeneities of the foils, investigations were made by means of a small range-meter, similar in construction to that described by CHILTON et al.[7]. A thin Po-source was placed on a movable table below a diaphragm with a small hole, comparable in size to the 1 mm^2 beam spot of the accelerator. The foil to be investigated was placed on the diaphragm. Some of the α-particles emitted from the source passed through the hole, penetrated the foil, and entered a Geiger counter. A plot of the counting rate versus the vertical position of the table yielded a well-defined half-intensity point corresponding to the range of the particles. In this position variations in the thickness of the foil were directly indicated by a change in counting rate when the foil was moved with respect to the diaphragm so that different parts of the foil were exposed.

The method was very sensitive and the foils were discarded if the measured thickness fluctuations exceeded 1.5 per cent. From the measurements one could extrapolate to the most probable value for the thickness at the center of the foil where the accelerator beam passed through in the actual experiment. The range-meter could not be used for some of the thicker foils because the range of the Po α-particles was too short. The homogeneity of these foils had to be checked in a more laborious way by bombarding them with the accelerator beam penetrating in several different positions. If the energy loss varied more than 2 per cent, the foils were discarded.

IV. Results and Discussion

The specific stopping power. In Figs. 4 and 5 the measured energy losses in KeV per mg/cm^2 are plotted as a function of the energy in MeV of the protons and deuterons, respectively. Measurements were made on 4 to 5 different foils of each of the metals indicated, the thicknesses ranging from

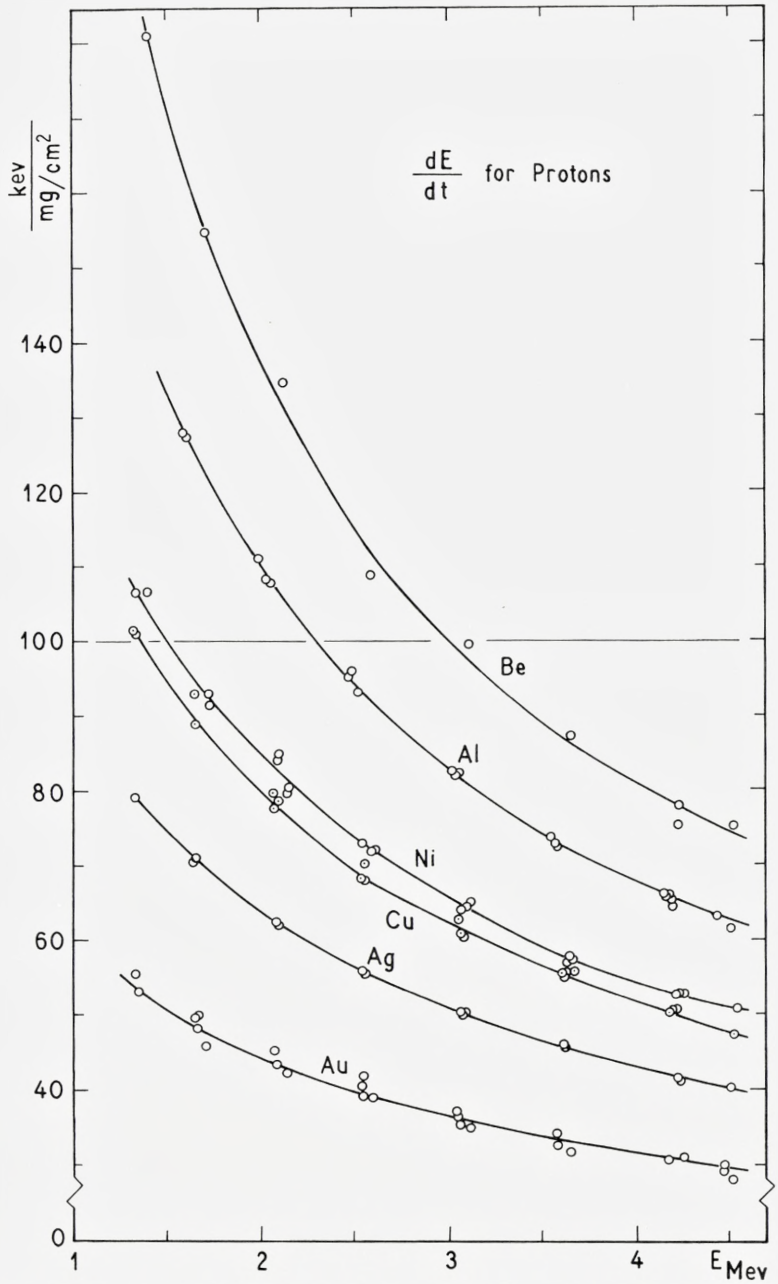

Fig. 4. Stopping powers for protons. The curves are experimental and drawn only in order to facilitate energy interpolations.

0.5 to 3.5 mg/cm^2. The uncertainty of the individual measurement is caused mainly by foil inhomogeneity, and is in the range from 1 to 3 per cent. The most reliable results were obtained for Al and Ag, the corresponding foils being particularly homogeneous. The measurements on Be were made only once, because the few foils available broke when they were taken out after the first exposure; the majority of the other measurements were repeated.

Figs. 4 and 5 are not well suited for interpolating the measured stopping powers to other elements. As mentioned in section II, it is more convenient to plot the data as a function of the parameter

$$x = \frac{v^2}{v_0^2 Z_2} = \frac{40}{A_1 Z_2} E_{\mathrm{MeV}}, \tag{15}$$

where E_{MeV} is the bombarding energy in MeV. As the measured thicknesses are given in units mg/cm^2, it is advantageous to introduce

$$dt = A_2 M_0 N \, dR \tag{16}$$

in the theoretical formulae. Equation (2) may then be rewritten in the form

$$\frac{dE}{dt} \sqrt{E Z_2} \left[\frac{A_2}{2 Z_2} \frac{1}{Z_1^2 \sqrt{A_1}} \right] = C \frac{L\{x\}}{\sqrt{x}}, \tag{17}$$

where

$$C = \frac{\pi \sqrt{2}\, e^2 \hbar}{m \sqrt{M_0}} = 14400 \frac{(\mathrm{KeV})^{3/2}}{\mathrm{mg/cm^2}}. \tag{18}$$

A plot of the experimental values of the quantity on the left-hand side of equation (17) as a function of x given by equation (15) should therefore give points falling on a single curve for all elements and projectiles. Since the measured dE/dt values are roughly proportional to $(E Z_2)^{-1/2}$, such a plot allows all the points to be presented with comparable precision, as shown in Fig. 6.

Within the experimental uncertainty, there are no differences between corresponding proton and deuteron points for a given element. This is not surprising since the mass is of negligible importance when the projectiles are very much heavier than the electrons. The points for different elements also fit in rather well with each other, although for the heavy elements there are deviations of a magnitude comparable to the experimental uncertainty. It is not clear whether the deviations are significant or not; on

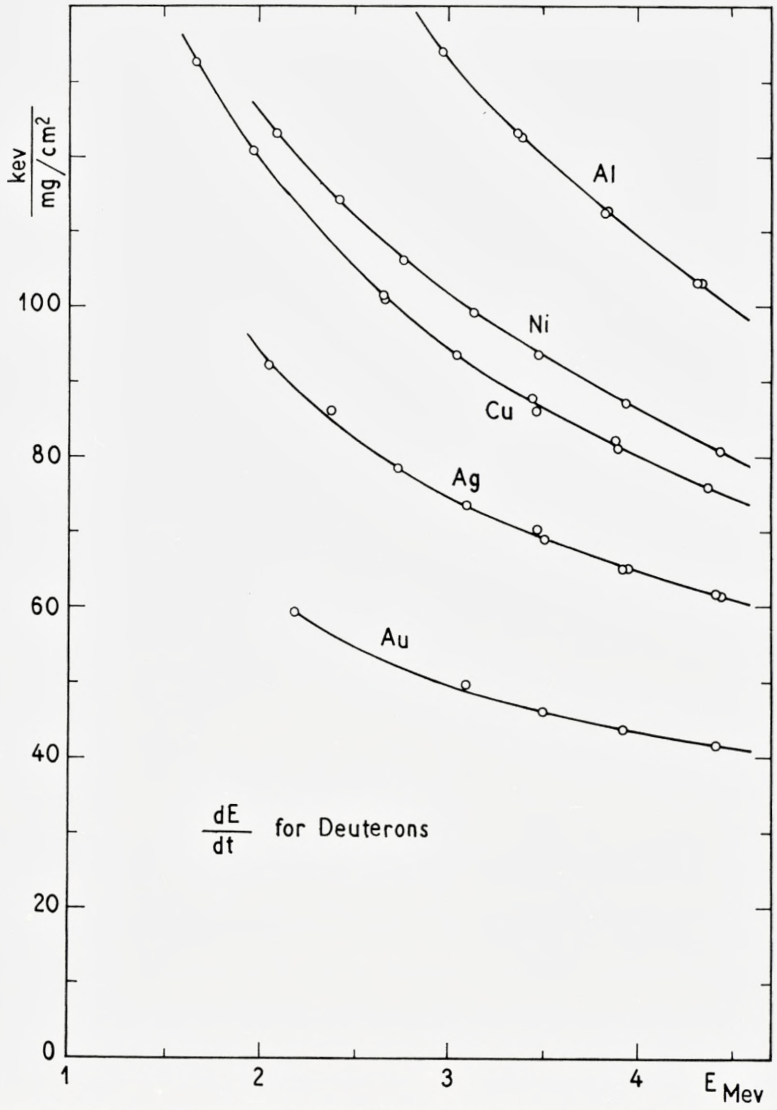

Fig. 5. Stopping powers for deuterons. The curves are experimental and drawn only in order to facilitate energy interpolations.

the other hand it would not be surprising if minor deviations occurred, since the Z_2 dependence is inferred from more specific assumptions. However, relatively safe interpolations to the stopping power of other elements

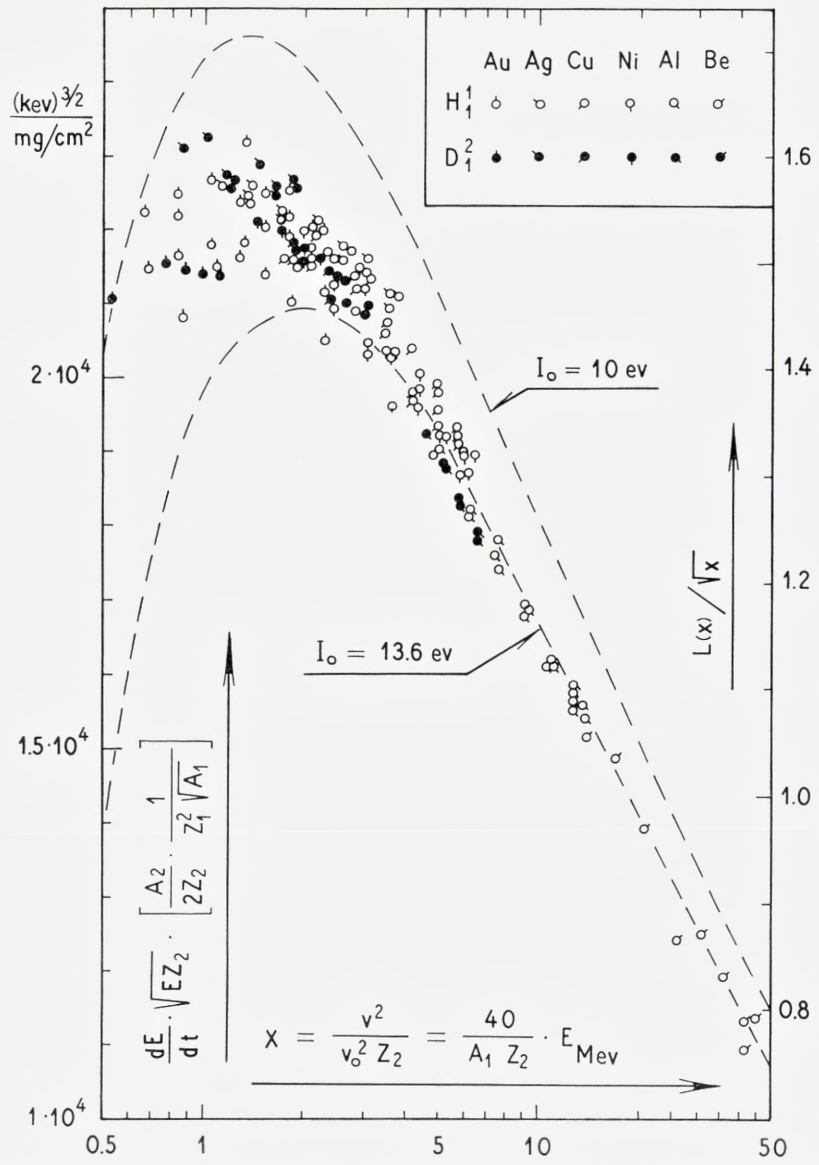

Fig. 6. Plot of the stopping power data contained in Figs. 4 and 5. According to theory, all points should fall on a single curve (cf. section IV). The factor in the square bracket does not depend much on the target material and is close to 1 for protons. The theoretical curves represent equations (2), (4), and (16), corresponding to the I_0-values indicated.

can be made from Fig. 6 in the investigated range of x-values. WARD WHALING[16] has compiled most of the existing data from stopping power measurements in the energy range below 2.5 MeV, as well as the data for gold investigated previously in the range from 1.5 MeV to 5 MeV; on this basis, he has extrapolated the results up to 10 MeV by means of the theoretical formulae. The present results agree well with the curves given by WHALING. In a few cases, the new points indicate slightly lower values, but the differences are less than 3 per cent.

In Fig. 6 the experimental data are compared with the theoretical expressions (4) and (2), corresponding to I_0-values of 10 eV and 13.6 eV. As explained in section II, the average excitation potential must at low bombarding energies be smaller than the constant value (4) which is approached at higher energies. From the figure it is evident that this is true, and the effect is shown more quantitatively in Table A, where the magnitudes of I have been evaluated at the various energies by means of equation (2) and the experimental curves in Fig. 4.

TABLE A. The average excitation potential I as derived from a comparison of formulae (2) and (16) with the experimental stopping power curves in Fig. 4. The proton energy is denoted by E_p, and ΔI represents the estimated uncertainty on I. The last row gives the ratio between the value of I at $E_p = 4$ MeV, and the atomic number Z_2 of the element in question (cf. eq. (4))

E_p	I_{Be}	I_{Al}	I_{Ni}	I_{Cu}	I_{Ag}	I_{Au}
MeV	eV	eV	eV	eV	eV	eV
1.5		185	365	377	556	866
2.0	56	184	368	371	572	937
2.5		180	371	379	576	974
3.0	56	180	371	382	588	995
3.5		175	378	378	587	1010
4.0	56	175	373	371	583	1000
ΔI	± 4	± 3	± 8	± 8	± 7	± 20
$I\{4\}/Z_2$	14.0	13.5	13.2	12.8	12.4	12.7

It is interesting to note that the variation of I for Al seems to go in the opposite direction. This is in agreement with the fact that, for high energies, various investigators[13] have found relatively low values, viz. $I_{Al} \simeq 163$ eV. The sign of the effect may be understood (cf. section II) by considering

that the C_K-correction in the present range of bombarding energies obtains its (positive) maximum value for elements in the neighbourhood of Al, as shown in Table B. From the table it appears, however, that the calculated C_K-correction does not suffice to account for the variation of the I_{Al}-values displayed in Table A.

TABLE B. Effect of the C_K-correction for Al. The correction has been estimated from the curve given by WALSKE[14].

E_p	C_K	I_{Al}^{uncorr}	I_{Al}^{corr}
MeV		eV	eV
1.5	0.3	185	181
2.0	0.6	184	174
3.0	0.9	180	168
4.0	1.0	175	162

The energy straggling. If the foil thickness t is introduced in accordance with equation (16), the relations (8) and (9) for the energy straggling Ω can be written in the form

$$\frac{\Omega}{\sqrt{t}}\sqrt{\frac{A_2}{2Z_2}\frac{1}{Z_1}} = \sqrt{\frac{2\pi e^4}{M_0}} \times \begin{cases} 1 \text{ for} & L \gtrsim 2 \\ \sqrt{\frac{1}{2} L\{x\}} \text{ for} & L \lesssim 2 \end{cases} \tag{19}$$

where

$$\sqrt{\frac{2\pi e^4}{M_0}} = 8.85 \frac{KeV}{(\text{mg/cm}^2)^{1/2}}. \tag{20}$$

In Fig. 7 the experimental values of the quantity on the left-hand side of equation (19) are plotted as a function of x. However, the plot does not include the Ω_s-correction. The magnitude of this correction can be estimated by means of the equations (11) to (14), and for each element the corresponding ordinate correction is indicated in the figure by the length of the arrows. The few exposures made with the 'line' target described in section III are consistent with the corrected values, but the points scatter too much to allow a quantitative determination of Ω_s. With the Ω_s-corrections included, the data conform quite well with theoretical expectations represented by the two curves. They are drawn in accordance with equations (19) and

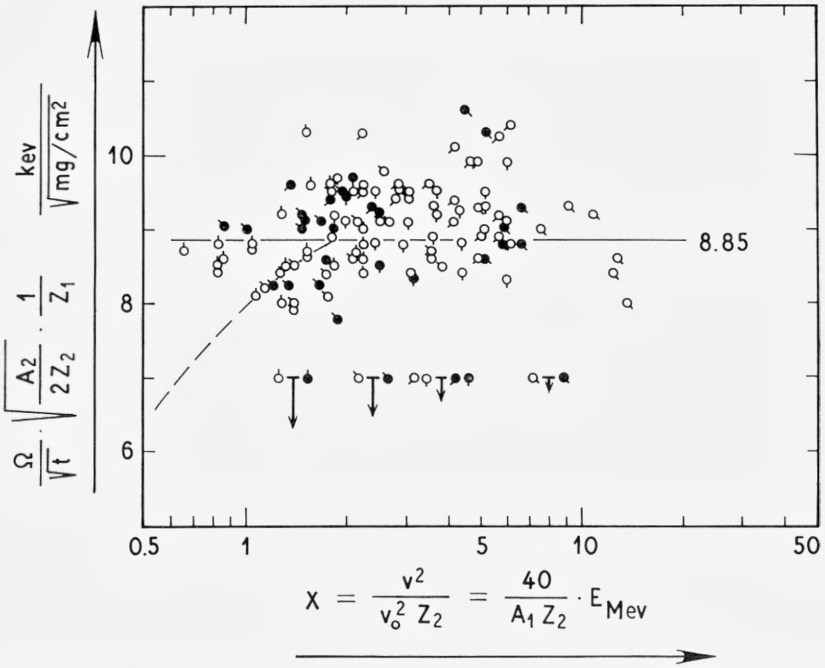

Fig. 7. Plot of the straggling data. The symbols are the same as in Fig. 6. The points have not been corrected for the scattering effect discussed in section III, but the magnitude of these corrections is indicated for each element by the length of the arrows. The curves represent the theoretical expressions (19), with the values of $L\{x\}$ derived from the experimental points in Fig. 6.

(20), and with $L\{x\}$ derived from the experimental points in Fig. 6, i. e., from the measured stopping powers.

Evidently, the points in Fig. 7 scatter too much to establish with certainty the expected decrease for small x-values, but it is difficult to obtain more reproducible data because of miscroscopic inhomogeneities in the foils. One variety of the commercial Al-foils, e. g., yielded rather large Ω-values, but these foils were only shiny on one of the sides and could therefore be discarded as far as the straggling measurements were concerned. Also the Be-foils were too poor to justify an evaluation of the straggling from the data obtained; Be is therefore not included in Fig. 7.

Acknowledgements

The author wants to thank Professor NIELS BOHR, as well as Professors JØRGEN BØGGILD and TORBEN HUUS for the excellent working conditions offered at the Institute for Theoretical Physics in Copenhagen. I am also grateful to Professor HUUS for his comments on the present paper. Furthermore, I want to thank Lektor BENT ELBEK for valuable advice and discussions. I am particularly indebted to Lektor MORTEN SCHARFF for his help and for the kind interest he showed the present investigation. Finally, thanks are due the members of the Van de Graaff staff of the Institute.

References

(1) H. A. BETHE, Ann. d. Phys., **5**, 325 (1930).
(2) H. A. BETHE and M. S. LIVINGSTONE, Revs. Mod. Phys. **9**, 245 (1937).
(3) F. BLOCH, Zs. f. Phys. **81**, 363 (1933).
(4) N. BOHR, Mat. Fys. Medd. Dan. Vid. Selsk. **18**, no. 8 (1948).
(5) G. H. BRIGGS, Revs. Mod. Phys. **26**, (1954).
(6) J. E. BROLLEY and F. L. RIBE, Phys. Rev. **98**, 1112 (1955).
(7) A. B. CHILTON, J. C. HARRIS, and J. N. COOPER, Phys. Rev. **93**, 413 (1954).
(8) B. ELBEK, K. O. NIELSEN, and M. C. OLESEN, Phys. Rev. **108**, 406 (1957).
(9) R. A. FISHER and F. YATES, Statistical Tables (Oliver and Boyd Ltd., London) (1938).
(10) T. HUUS, Mat. Fys. Medd. Dan. Vid. Selsk. **26**, no. 14 (1951).
(11) J. LINDHARD and M. SCHARFF, Mat. Fys. Medd. Dan. Vid. Selsk. **27**, no. 15 (1953).
(12) J. LINDHARD, Mat. Fys. Medd. Dan. Vid. Selsk. **28**, no. 8 (1954).
(13) C. P. SONETT and K. R. MACKENZIE, Phys. Rev. **100**, 734 (1955).
(14) M. C. WALSKE, Phys. Rev. **88**, 1283 (1952).
(15) M. C. WALSKE, Phys. Rev. **101**, 940 (1956).
(16) W. WHALING, Report from Kellogg Radiation Laboratory California Institute of Technology, Pasadena, California, (1957).

Indleveret til Selskabet den 10 juli 1961.
Færdig fra trykkeriet den 8. november 1961.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 7

Mat. Fys. Medd. Dan. Vid. Selsk. **33**, no. 7 (1961)

DISSOCIATION OF DIATOMIC MOLECULES CONSIDERED AS DIFFUSION IN PHASE SPACE

BY

THOR A. BAK AND KNUD ANDERSEN

København 1961
i kommission hos Ejnar Munksgaard

Synopsis

Dissociation of diatomic molecules is considered as the escape of a classical particle from a potential minimum due to its Brownian motion. The criterion for reaction is taken to be annihilation of the particle at a certain energy. For this model the Kramers equation is set up and solved exactly for potentials of the form $V = C|r|^n$, and the rate constant for escape from the potential minimum is found. It is also shown how the rate constant may be obtained from a variational principle, and as an example of this method the rate of escape from a Morse potential is found. The results obtained agree very well with machine calculations. Finally it is attempted to justify the Kramers equation in the limit of weak interaction by deriving it directly from the Liouville equation. It is shown that the equation obtained deviates significantly from the Kramers equation, except for the case of a harmonic oscillator molecule. It is remarkable, however, that rate constants obtained in this way for the rate of escape of a particle from very deep potentials of the above mentioned simple form are almost identical with those derived from the theory of Kramers.

Printed in Denmark
Bianco Lunos Bogtrykkeri A-S

Introduction

In recent years there has been a renewed interest in stepwise activation theories of chemical reactions[1–5] resembling the Brownian motion theory originally proposed by KRAMERS[6]. In such a theory the reacting molecule is considered as an effective mass point which performs a Brownian motion in an external potential due to the coupling to a thermostat. When the particle attains a sufficiently high energy a chemical reaction may occur, which in these theories is pictured as the crossing of a certain surface in phase-space, or simply as an annihilation of the particle as it reaches a certain energy.

Due to the Brownian motion of the particle in phase space the probability density function for the particle will satisfy a diffusion equation when the problem is treated classically, or a discrete analogue of a diffusion equation when it is treated quantum mechanically. It is generally assumed that the diffusion equation derived by Kramers using the semiphenomenological theory of Brownian motion is correct. So far it has only been possible, however, to compare it with more exact calculations in the case of a particle moving in a harmonic oscillator potential, since only in this case has it been possible to set up the equations. In the harmonic oscillator case, and with the assumption that the density in phase space does not depend on the angle variable of the particle, there is complete agreement between the classical equation of KRAMERS[6] and the equation derived by BAK, GOCHE and HENIN[7] for the case of the Brownian motion of an oscillator weakly coupled to a crystal lattice. Furthermore, for this case the quantum mechanical theory of MONTROLL and SHULER[1] also reduces to the Kramers equation in the limit $\hbar \to 0$.

In the case of a harmonic oscillator potential it is also comparatively simple to solve the diffusion equation, at least when the reaction is considered as an annihilation of particles. As is usual for such calculations, the rate constants obtained for the dissociation of diatomic molecules are far too small. One of the reasons for this is undoubtedly the use of the harmonic potentials, since one would expect that an anharmonic potential would speed

up the dissociation reaction. Although the neglect of anharmonicity in the potential is by no means the only reason for the discrepancy between theory and experiments, we shall here be concerned solely with this difficulty and disregard such questions as the relative importance of hard core collisions and weak interaction collisions, and the even more elusive question of whether it is permissible to use intermolecular potentials, derived by using the Born-Oppenheimer approximation, under the conditions which prevail in a molecular collision.

Throughout this paper we shall therefore assume that the reacting molecule, which we for simplicity shall think of as a diatomic molecule, is in weak interaction with a thermostat. The thermostat which is assumed to be in equilibrium may be a gas or a crystal (phonon gas). The criterion for reaction is that the molecule reach a certain energy, i. e., the reaction is pictured as an annihilation at a certain energy level.

For this model we derive the Kramers diffusion equation in phase space (or rather energy-time space) and solve it exactly for oscillator potentials $V = C|r|^n$. For arbitrary intermolecular potentials the equation can be solved approximately be a variational method. As an example of the use of this method, we find the rate constant for escape from a Morse potential.

Finally we attempt to justify the Kramers equation for the above model by deriving it directly from the Liouville equation, using the asymptotic time integration developed by Brout and Prigogine[12].

Due to mathematical difficulties we limit ourselves in this case to considering the oscillator potentials $V = C|r|$, $V = \frac{1}{2}\gamma r^2$ and the square-well potential, $V = 0$ for $|r| < l/2$, $V = \infty$ for $|r| > l/2$. These potentials have the common feature that when r is expressed in terms of the action-angle variables, J and r can be factorized, i. e. $r = r_0(J)\,\theta(\alpha)$.

In order to be able to compare the coefficients C, γ and l for these potentials one must make a convention about the different values of r at which dissociation occurs. We shall make the assumption that the value of r at which dissociation occurs is the same for the three potentials. We therefore have

$$C^2 = \frac{1}{2} D\gamma \qquad \text{and} \qquad l^2 = 2D/\gamma$$

where D is the activation energy.

The most remarkable feature of the exact theory is that it gives practically the same results for the rate of escape of a particle from a potential as does the theory of Kramers, in spite of the fact that the two equations for the

time behaviour of the density in phase space are completely different. We cannot, however, agree with the statement made by MAZUR[15] that Prigogine's theory confirms the phenomenological theory of Kramers. The agreement which one obtains with respect to rate constants indicates, however, that as long as only weak interactions are considered, the conceptually much simpler theory of Kramers which has later been elaborated by BRINKMAN[16] may be useful when considering the influence of anharmonicities, or when investigating the validity of annihilation as criterion for reaction.

Both the theory of Kramers and the more exact treatment based on Prigogine's integration of the Liouville equation show that for simple potentials one gets the result that the rate constant is

$$k \propto \frac{1}{\omega(\beta D)} \beta D \, e^{-\beta D}$$

which is precisely the result one would expect from a correspondence argument assuming the result

$$k = \beta D \, e^{-\beta D}$$

to be valid for the harmonic oscillator where the energy levels are equally spaced. Both theories therefore show that anharmonicities which decrease the frequency of vibration increase the rate of dissociation, in qualitative accord with the experimental findings. The aim of the present paper is, however, not to compare theory and experiments, but solely to study how the problem of anharmonic molecular potentials can be treated within the framework of a weak interaction theory.

Kramers' equation

We shall start by giving a derivation of the Kramers equation for diffusion in phase space for the case of small viscosity, i. e., weak interaction between the particle, the Brownian motion of which we consider, and the thermostat. The derivation takes its starting point in the Fokker-Planck-equation for the phase-density function $\Phi = \Phi(q, p, t)$, defined in the phase-space for the particle. In the case of a particle exposed to a force derived from an external field of force $V(q)$ in addition to the stochastically varying force due to the surrounding medium, the equation has the following form[6, 8]

$$\frac{\partial \Phi}{\partial t} = \eta \frac{\partial}{\partial p}\left(\Phi p + m\beta^{-1} \frac{\partial \Phi}{\partial p} \right) - \frac{p}{m} \frac{\partial \Phi}{\partial q} + \frac{dV}{dq} \frac{\partial \Phi}{\partial p}. \qquad (1)$$

For our purpose it turns out to be practical to change the variables specifying the state of motion of the particle from (q, p) to action-angle variables $J = \frac{1}{2\pi}\oint p dq$, which depends only on the energy, and α defined by $d\alpha = \omega dt$ where $\omega = \frac{2\pi}{T} = \frac{dE}{dJ}$. For the simple case where $dq = \frac{p}{m} dt$ we have therefore

$$J = \int_0^T \frac{p^2}{m} dt = \frac{2\pi}{\omega} \int_0^{2\pi} \frac{p^2}{m} d\alpha. \tag{2}$$

We now consider the case where the coupling-coefficient η can be taken to be so small that the particle will run through the region in phase space between the surfaces with constant energy E and $E + dE$ several times (α increasing each time by 2π) before it leaves this region and changes its energy. Expressed in another way, we suppose the particle density to be equally distributed over the region between E and $E + dE$, that is $\Phi = \Phi(J)$ or $\left(\frac{\partial \Phi}{\partial \alpha}\right)_J = 0$. Substituting

$$\frac{\partial}{\partial p} = \frac{\partial J}{\partial p}\frac{\partial}{\partial J} = \frac{p}{\omega m}\frac{\partial}{\partial J} \quad \text{and} \quad \frac{\partial}{\partial q} = \frac{\partial J}{\partial q}\frac{\partial}{\partial J} = \frac{1}{\omega}\frac{dV}{dq}\frac{\partial}{\partial J}$$

into eq. (1) and introducing the reduced energies $x = \beta E$ and reduced action variables $j = \beta J$, eq. (1) is transformed into

$$\frac{\partial \Phi}{\partial t} = \eta \left[\Phi + \beta \left(\frac{p^2}{\omega m} + \frac{1}{\omega} + \frac{p^2}{\omega m}\frac{d}{dj}\left(\frac{1}{\omega}\right)\right)\frac{\partial \Phi}{\partial j} + \beta \frac{p^2}{\omega^2 m}\frac{\partial^2 \Phi}{\partial j^2}\right].$$

Then integrating (averaging) over α from 0 to 2π using eq. (2), the following partial differential equation is obtained

$$\frac{\partial \Phi}{\partial t} = \eta \frac{\partial}{\partial j}\left(\Phi j + \frac{j}{\omega}\frac{\partial \Phi}{\partial j}\right) = \eta\omega \frac{\partial}{\partial x} j \left(1 + \frac{\partial}{\partial x}\right)\Phi \tag{3}$$

where in the last equation Φ is considered as a function of the reduced energy x and t instead of j and t. Introducing the reduced time $\tau = \eta t$ we get

$$\frac{\partial \Phi}{\partial \tau} = \omega \frac{\partial}{\partial x} j \left(1 + \frac{\partial}{\partial x}\right)\Phi$$

and separating the variables by setting

$$\Phi(x, \tau) = \Psi(x)\Theta(\tau)$$

we obtain

$$\frac{1}{\Theta}\frac{d\Theta}{d\tau} = \frac{1}{\Psi}\omega\frac{d}{dx}j\left(1+\frac{d}{dx}\right)\Psi = -k.$$

As Φ has to approach a stationary function for $\tau \to \infty$, the solution for Θ must be $\Theta \propto e^{-k\tau}$ with k positive and real. For Ψ we therefore obtain the differential equation

$$\omega\frac{d}{dx}j\left(1+\frac{d}{dx}\right)\Psi + k\Psi = 0.$$

Because we are primarily interested in the deviation from the Maxwell-Bolizmann distribution, we set $\Psi(x) = y(x)\, e^{-x}$, and obtain for y

$$\frac{d}{dx}\left(j\, e^{-x}\frac{dy}{dx}\right) + k\, e^{-x}\left(\frac{y}{\omega}\right) = 0 \tag{4}$$

or

$$j\omega\frac{d^2y}{dx^2} + (1-j\omega)\frac{dy}{dx} + ky = 0. \tag{5}$$

Furthermore, for the special potentials, which can be expressed by $V(q) = C\,|q^n|$ $\left(C \text{ being a constant, } 0 \leqslant q \leqslant \left(\frac{E}{C}\right)^{1/n}\right)$ as for example the square-well or the harmonic potential, $j\omega$ will be proportional to x as a consequence of the virial theorem. This can be shown simplest by introducing $p = \pm\sqrt{2m(E-V)}$ in the expression defining J when the potential is symmetrical about $q = 0$.

$$J = \frac{1}{2\pi}\oint p\,dq = 4\,\frac{1}{2\pi}\sqrt{2m}\int_0^{(E/C)^{1/n}}\sqrt{E - C\,q^n}\;dq.$$

When $q = z\left(\frac{E}{C}\right)^{1/n}$ is substituted we get

$$J = \frac{4}{2\pi}\sqrt{2\,m\,C^{-1/n}}\,E^{(1/2+1/n)}\int_0^1\sqrt{1-z^n}\;dz$$

$$\frac{1}{\omega} = \frac{dJ}{dE} = \left(\frac{1}{2}+\frac{1}{n}\right)E^{-1}J$$

$$J\omega = \frac{2n}{n+2}E = \frac{1}{c}E \qquad \text{and} \qquad j\omega = \frac{1}{c}x. \tag{6}$$

For a box-potential (i. e., "$n = \infty$") it is seen that c has to be put equal to $\frac{1}{2}$ in eq. (6). For these potentials the distribution function y thus obeys the differential equation

$$x\frac{d^2y}{dx^2} + (c-x)\frac{dy}{dx} + ay = 0 \qquad \text{with} \qquad a = ck. \tag{7}$$

Exact solutions of Kramers' equation

The exact solution of eq. (7) is a confluent hypergeometric series[9]

$$y = {}_1F_1(-a, c; x) = 1 - \frac{a}{c}x - \frac{a(-a+1)}{c(c+1)}\frac{x^2}{2!} + \ldots \tag{8}$$

and our result is therefore

$$\Phi = \sum_k b_k e^{-k\tau} e^{-x} {}_1F_1(-a, c; x). \tag{9}$$

The values of k which must be selected in the summation above are determined by the boundary condition: ${}_1F_1(-a, c; x^*) = 0$. The physical meaning of the condition: $\Phi = 0$ for $x \geqslant x^*$ is that the particle associated with the oscillating motion of the molecules considered is simply annihilated, when it—during its random motion—reaches the reduced energy x^*.

The eigenvalues $a = ck$ given by the equation

$${}_1F_1(-a, c; x) = 0 \tag{10}$$

lie very close to the integers 0, 1, 2, 3, . . . for large values of $x = \frac{D}{kT}$. Because of the factor $e^{-k\tau}$ in each term of eq. (9), the term corresponding to the smallest eigenvalue $a_0 \ll 1$ will be quite dominating, if only a certain time has elapsed since the system was "started" with some initial distribution, and the error made by setting Φ equal to this first term will therefore be completely negligible for reasonable values of x^* (say $x^* > 5$).

The reduced rate constant k_0 for the annihilation process—and for the chemical reaction—is defined by:

$$\frac{dN}{d\tau} = -k_0 N$$

where N is the total number of particles in the potential well

$$N = \int_0^{x^*} \Phi(x, \tau)\, dx.$$

It can therefore be calculated from the following expression in which all terms in the solution for Φ except the first have been neglected.

$$k_0 = \frac{-\frac{d}{d\tau}\int_0^{x^*} b_0 e^{\frac{a_0}{c}\tau} e^{-x} {}_1F_1(-a_0, c; x)\, dx}{\int_0^{x^*} b_0 e^{\frac{a_0\tau}{c}} e^{-x} {}_1F_1(-a_0, c; x)\, dx} = \frac{a_0}{c}. \tag{11}$$

The problem of calculating the rate constant is by eq. (11) reduced to the purely mathematical problem of finding the lowest value for a which satisfies (10).

If ${}_1F_1$ is Taylor-expanded in a_0, assuming that $a_0 \ll 1$, we obtain the following approximate solution of eq. (10) as ${}_1F_1(0, c; x) = 1$

$$\frac{1}{a_0} = \left(\frac{\partial\, {}_1F_1(a, c; x^*)}{\partial a}\right)_{a=0} = \frac{x^*}{c} + \frac{x^{*2}}{c(c+1)\,2} + \ldots \tag{12}$$

By means of eq. (12) it is possible to evaluate a_0 directly as a function of x^* (method 3), but a simpler, although somewhat more approximate, formula can be found by using

$$\int_0^{x^*} \frac{{}_1F_1(1, c; x) - 1}{x}\, dx = \frac{x^*}{c} + \frac{x^{*2}}{c(c+1)\,2} + \ldots = \left(\frac{\partial\, {}_1F_1(a, c; x^*)}{\partial a}\right)_{a=0}.$$

The integration above is carried out by setting

$$\int \frac{{}_1F_1(1, c; x)}{x}\, dx = {}_1F_1(1, c, x)\,(a_0 + a_1 x^{-1} + \ldots) + \text{const.} = {}_1F_1(1, c; x)\, P(x^{-1}) + \text{const.}$$

and determining the coefficients a_0, a_1, ... in the polynomium P by differentiation

$${}_1F_1(1, c; x) = x\,\frac{\partial\, {}_1F_1}{\partial x}\, P + x\, {}_1F_1\, \frac{dP}{dx}.$$

If at this place we substitute the asymptotic formula for ${}_1F_1(1, c; x)$[9], valid for $x \gg 1$

$${}_1F_1(1, c; x) = \frac{\Gamma(c)}{\Gamma(1)}\, e^x\, x^{1-c}$$

P can be determined to be

$$P = x^{-1} + cx^{-2} + c(c+1)\,x^{-3} + \ldots$$

and in this way we obtain

$$\int_0^{x^*} \frac{{}_1F_1(1, c; x) - 1}{x}\, dx = \int_0^1 \frac{{}_1F_1(1, c; x) - 1}{x}\, dx + \int_1^{x^*} \frac{{}_1F_1(1, c; x)}{x}\, dx - \ln x^*$$

$$= \int_{\bullet 0}^{\bullet 1} \frac{{}_1F_1(1, c; x) - 1}{x} dx - ln\, x^* + \left[{}_1F_1(1, c; x) \frac{1}{x} \right]_1^{x^*}$$

$$= \Gamma(c)\, e^{x^*}\, x^{*-c} + 0\,(ln\, x^*).$$

We have therefore obtained the following asymptotic formula

$$k = \frac{1}{\Gamma(c+1)}\, e^{-x^*} x^{*c} = \frac{1}{\Gamma(c+1)}\, e^{-\beta D} (\beta D)^c \tag{13}$$

which rapidly converges upon the exact result as the value of x increases. As mentioned in the introduction, this result is precisely what one would expect more intuitively, namely $k \propto \frac{1}{\omega(x^*)}\, x^*\, e^{-x^*}$.

The agreement between the k-values given by eq. (13) and those obtained by numerical solution of eq. (10) either by a machine method[10], which has been done for $c = 1$, or by other methods such as method 3, mentioned above, is fairly good for the higher values of x (see table 1).

The variational method

We now return to the general problem of solving the Kramers equation (3) without making assumptions about the form of the potential in which the particle moves. First of all we are interested in a method which allows a calculation of the smallest value of k for which Φ satisfies the boundary conditions, because this value is equal to the rate constant k. For that purpose a variational method is used. The method enables us to determine k-values only a little larger than the exact ones by approximating Φ with a trial function.

It is immediately observed that equation (4) is of the Sturm-Liouville type and that y satisfies the Sturm-Liouville boundary conditions:

$$y j\, e^{-x} \frac{dy}{dx} = 0 \text{ both for } x = 0 \ (j = 0) \text{ and for } x = x^* \ (y = 0).$$

Therefore

$$k_0 = \frac{-\int_{\bullet 0}^{\bullet x^*} y \frac{d}{dx}\left(j\, e^{-x} \frac{dy}{dx} \right) dx}{\int_{\bullet 0}^{\bullet x^*} y^2 \frac{1}{\omega}\, e^{-x}\, dx} = \frac{\int_{\bullet 0}^{\bullet x^*} \left(\frac{dy}{dx} \right)^2 j(x)\, e^{-x}\, dx}{\int_{\bullet 0}^{\bullet x^*} y^2 \frac{1}{\omega(x)}\, e^{-x}\, dx}.$$

TABLE 1.

Values for the rate constant k, in reduced units.

x^*	method	Potential $V = C \mid q^n \mid \left(n = \frac{2}{2c-1}\right)$			Morse potential
		$n = \infty$, $c = \frac{1}{2}$ box-potential	$n = 2$, $c = 1$ harmonic potential	$n = 1$, $c = \frac{3}{2}$ const. force potential	
2	1)	0,216	0,271	0,288	
	2)	0,249	0,348	0,463	0,414
	3)				
	4)		0,329		0,372
5	1)	0,0170	0,0337	0,0566	
	2)	0,0156	0,0288	0,0471	0,0388
	3)	0,0152			
	4)		0,0274		0,0345
10	1)	$1{,}62 \cdot 10^{-4}$	$4{,}54 \cdot 10^{-4}$	$10{,}8 \cdot 10^{-4}$	$6{,}21 \cdot 10^{-4}$
	2)	$1{,}54 \cdot 10^{-4}$	$4{,}09 \cdot 10^{-4}$	$9{,}28 \cdot 10^{-4}$	
	3)	$1{,}53 \cdot 10^{-4}$		$9{,}00 \cdot 10^{-4}$	
	4)		$4{,}02 \cdot 10^{-4}$		$5{,}86 \cdot 10^{-4}$
15	1)	$1{,}33 \cdot 10^{-6}$	$4{,}58 \cdot 10^{-6}$	$13.3 \cdot 10^{-6}$	
	2)				
	3)	$1{,}29 \cdot 10^{-6}$		$12{,}1 \cdot 10^{-6}$	
	4)		$4{,}26 \cdot 10^{-6}$		
20	1)	$1{,}04 \cdot 10^{-8}$	$4{,}12 \cdot 10^{-8}$	$13{,}9 \cdot 10^{-8}$	
	2)	$1{,}01 \cdot 10^{-8}$	$3{,}92 \cdot 10^{-8}$	$12{,}9 \cdot 10^{-8}$	$6{,}45 \cdot 10^{-8}$
	3)			$12{,}9 \cdot 10^{-8}$	
	4)		$3{,}90 \cdot 10^{-8}$		$6{,}45 \cdot 10^{-8}$

Method 1) is based on formula (13).

Method 2) is based on formula (17) using table 2.

Method 3) is based on formula (12).

The confluent hypergeometric series has been evaluated directly by means of a desk computer.

Method 4) are calculations on a digital computer.

The variational principle now guarantees that if—instead of the exact solution y—we substitute a "trial function" $Y(x)$ into this expression, a value k_0' comes out

$$k_0' = \frac{\int_0^{x^*} \left(\frac{dY}{dx}\right)^2 j(x)\, e^{-x}\, dx}{\int_0^{x^*} Y^2 \frac{1}{\omega(x)}\, e^{-x}\, dx} \tag{14}$$

which is larger than k_0. By varying the parameters in the trial function until k_0' attains it minimum value, we can therefore determine an approximate value for k_0.

If the denominator in eq. (14) is considered as a normalization constant for Y, this problem can be formulated as the problem of finding among all possible normalized functions for which $Y(x^*) = 0$ that function y which minimizes the integral

$$\int_0^{x^*} \left(\frac{dY}{dx}\right)^2 j(x)\, e^{-x}\, dx. \tag{15}$$

It is interesting to note that this integral is formally identical with the "generalized entropy production" discussed previously by one of us[11] in connection with ordinary diffusion. Therefore the variational principle can be considered formally as a case of the theorem of Prigogine stating that a stationary irreversible process is characterized by a minimal production of entropy.

If in eq. (14) we change to the new variable $X = x/x^*$, we obtain

$$k_0' = \frac{\int_0^{1} \left(\frac{dY}{dX}\right)^2 j(X)\, e^{-x^* X}\, dX}{x^{*2} \int_0^{1} Y^2 \frac{1}{\omega(X)}\, e^{-x^* X}\, dX}. \tag{16}$$

In the calculations of k_0'-values performed in this work we have used the trial function: $Y(X) = 1 - e^{x - x^*} = 1 - e^{x^*(X-1)}$, since a calculation for the harmonic potential with the use of a trial function including a parameter ξ, $Y = 1 - e^{\xi(x - x^*)}$ gave the lowest k_0'-value, when ξ was extremely close to the value 1. It should be noted that $y = 1 - e^{x - x^*}$ corresponds to a Ψ-function

$$\Psi = e^{-x} y = e^{-x} - e^{-x^*}.$$

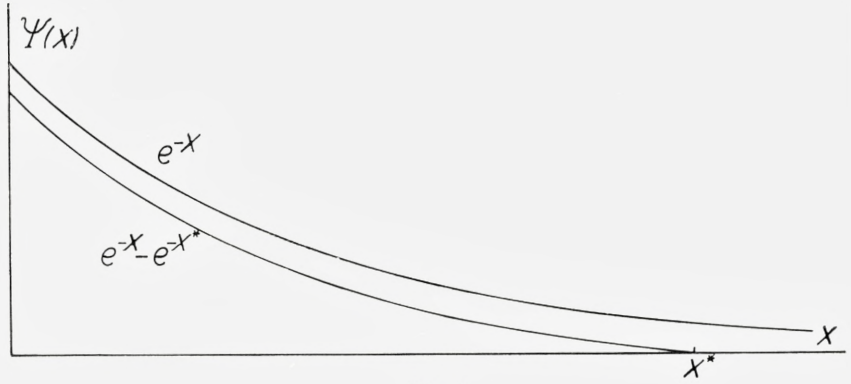

Fig. 1.

That is, except for a normalization factor it is simply an equilibrium distribution function from which a constant value e^{-x^*} has been substracted in order to make Ψ zero for $x = x^*$.

When this Y-function is substituted into eq. (16) we obtain

$$k_0' = e^{-2x^*} \frac{\int_0^1 e^{x^* X} j(X)\, dX}{\int_0^1 \frac{e^{-x^* X}}{\omega(X)} (1 + e^{2x^*(X-1)} - 2\, e^{x^*(X-1)})\, dX}. \tag{17}$$

From this expression the values for the rate constant k listed in table 1 were calculated (Method 2).

Escape of a particle from a Morse potential

In order to obtain a better approximation to the true intramolecular potential for a diatomic molecule than the simple potentials discussed above and to check the above variational principle, we have investigated the escape of a particle from a Morse potential given by

$$V(r) = D(1 - e^{-r/\delta})^2$$

$$D = \frac{m}{2} \omega_0{}^2 \delta^2$$

r is the length of the molecule minus its equilibrium length, and ω_0 is the frequency of oscillations for very small energies, that is, the harmonic frequency. The magnitude of $1/\delta$ measures the degree of anharmonicity.

From this expression it follows that

$$J = \frac{1}{2\pi}\oint p\,dr = \frac{2D}{\omega_0}\left(1 - \sqrt{1 - E/D}\right)$$

and

$$\omega = \omega_0\sqrt{1 - E/D} = \omega_0\sqrt{1 - x/x^*}$$

which when substituted in eq. (5) gives the equation

$$\frac{2}{x^*}\left[\sqrt{1-X} - (1-X)\right]\frac{d^2y}{dX^2} - 2\left[\sqrt{1-X} - (1-X) - \frac{1}{2x^*}\right]\frac{dy}{dX} + ky = 0.$$

Here we set $\sqrt{1-X} = 1-u\,(0 \leq u \leq 1)$ and obtain

$$\frac{1}{2x^*}u(1-u)\frac{d^2y}{du^2} - \left[u(1-u)^2 - \frac{1}{2x^*}\right]\frac{dy}{du} + k(1-u)^2 y = 0. \qquad (18)$$

This differential equation cannot be solved in terms of known functions, but by numerical integration the smallest value of k, for which a solution through the points $(u, y) = (0,1)$ and $(u, y) = (1,0)$ exists, can be found by a trial and error method. This was done on a digital computer and the results are listed in table 1.

In order to check the validity of the variational principle in the form developed above we have calculated the k_0' values given by eq. (17), where $j(X)$ and $\omega(X)$ for the Morse potential have been inserted. The expression used for these calculations is:

$$k_0' = \frac{1 - F(x^*)}{e^{x^*}\,G(x^*) + F(x^*) - 2} \approx 2x^* e^{-x^*} \text{ for large } x^*$$

in which

$$F(x^*) = \frac{1}{\sqrt{x^*}}\int_0^{\sqrt{x^*}} e^{-t^2}\,dt$$

and

$$G(x^*) = \frac{1}{\sqrt{x^*}}e^{-x^*}\int_0^{\sqrt{x^*}} e^{t^2}\,dt$$

$F(x^*)$ and $G(x^*)$ are listed in table 2. In the limit $x^* \gg 1$ F is asymptotically equal to $\frac{\sqrt{\pi}}{2x^*}$ and G asymptotically equal to $\frac{1}{2x^*}$. The expression for k_0' should be compared with that obtained for the harmonic potential:

$$k_0' = \frac{(x^*-1)\,e^{-x^*} + e^{-2x^*}}{1 - 2x^* e^{-x^*} - e^{-2x^*}} \approx x^* e^{-x^*} \text{ for large } x^*.$$

TABLE 2.
Values for the functions F and G

x^*	F	G
5	0,396	0,1157
10	0,280	0,0527
20	0,198	0,0256

Derivation of a diffusion equation in phase space

We have previously[7, 14] considered the Brownian motion of an oscillator coupled to a lattice which serves as a thermostat and shown that ϱ, the density in phase space of the system consisting of oscillator plus thermostat, satisfies the equation

$$\frac{\partial \varrho}{\partial t} = \pi \lambda^2 \frac{L}{2\pi} \sum_{\nu}^{\prime} \int d\omega_f \frac{df}{d\omega_f} \left\{ \left(\frac{\partial}{\partial J_f} + \nu \frac{\partial}{\partial J} \right) | V_{\nu, f} |^2 \delta (\nu\omega + \omega_f) \left(\frac{\partial}{\partial J_f} + \nu \frac{\partial}{\partial J} \right) \right. \\ \left. + \left(\frac{\partial}{\partial J_f} - \nu \frac{\partial}{\partial J} \right) | V_{\nu, f} |^2 \delta (\nu\omega - \omega_f) \left(\frac{\partial}{\partial J_f} - \nu \frac{\partial}{\partial J} \right) \right\} \varrho \qquad (19)$$

in which J_f and ω_f are the action variable and the frequency of a lattice oscillation with wave number f, J and ω are the corresponding variables for the oscillator. $V_{\nu, f}$ is a Fourier coefficient of the non-harmonic interaction energy which will be defined below. In deriving this result it was assumed that ω is independent of J so that the Hamiltonian has the form

$$H = \sum_f \omega_f J_f + J + V.$$

We shall now see that precisely the same equation arises when ω depends on J, except that $V_{\nu, f}$ has to be defined slightly differently.

For the perturbing potential V we take

$$V(r) = \sum_n W(r - a_n) u_n$$

where u_n is the displacement of the n'th particle in the lattice, and $W(r - a_n)$ is the force exerted on the oscillator by an infinitesimal displacement of the n'th particle. This force of course depends on the stretching of the oscillator which is given by r. Expressing u_n in normal coordinates q_f we obtain

$$V = \sum_f V_f q_f e^{ifr}$$

with

$$V_f = \sum_n W(r - a_n)^{-if(r - a_n)}.$$

As an approximation we assume V_f to be independent of r for all values of f.

The equation for ϱ given above is derived from the spectrally decomposed Liouville equation, using the integration technique developed by BROUT and PRIGOGINE[12]. In order to apply this method the interaction potential V must be Fourier expanded on the angle variable α. This can be done quite simply in the case of a harmonic oscillator by using

$$r = r_0 \sin \alpha$$

and

$$e^{ifr_0 \sin \alpha} = \sum_{m=-\infty}^{\infty} J_m(fr_0)\, e^{im\alpha} \tag{20}$$

where J_m is the Bessel function of m'th order.

Using $q_f = q_{f+}^0 e^{i\alpha_f} + q_{f-}^0 e^{-i\alpha_f}$ one then obtains

$$V = \sum V_f q_f e^{ifr} = \sum_m \sum_f \left\{ V_{m,f}^+ e^{i(\alpha_f' + m\alpha)} + V_{m,f}^- e^{i(-\alpha_f' + m\alpha)} \right\}$$

with

$$V_{m,f}^+ = V_f q_{f+}^0 J_m(fr_0)$$

$$V_{m,f}^- = V_f q_{f-}^0 J_m(fr_0).$$

The squares of the absolute values of these Fourier coefficients are independent of the index + or − and are the quantities $|V_{\nu,f}|^2$ used above.

When the oscillator is not harmonic we still have $r = r_0(J)\,\Theta(\alpha)$ for the simple potentials we are going to consider. Here $\Theta(\alpha)$ is a periodic function with the property $|\Theta(\alpha)| \leqslant 1$. We now define the functions Y_m as

$$e^{ifr_0\,\theta(\alpha)} = \sum_{m=-\infty}^{\infty} Y_m(fr_0)\, e^{im\alpha} \tag{21}$$

and replace $J_m(fr_0)$ in the equations above by $Y_m(fr_0)$. When $|V_{\nu,f}|^2$ is redefined in this way, and it is assumed that $\omega = \omega(J)$ is a function of J, equation (1) now describes the evolution of ϱ for the system of an anharmonic oscillator in a lattice of harmonic oscillators.

We then use that

$$|V_{\nu,f}|^2 = |V_f|^2 |q_f|^2 Y_\nu^2(fr_0)$$

and

$$|q_f|^2 = \frac{J_f}{M^{(N)} \omega_f}$$

where $M^{(N)}$ is the total mass of the lattice.

Also, as is usual in calculations of this sort, we use the Peierls assumption

$$\frac{|V_f|^2}{\omega_f^2} = \text{constant}$$

and the Debye spectrum for the phonons

$$\frac{d\omega_f}{df} = c$$

where c is the velocity of sound in the crystal. Finally we assume that the crystal is in equilibrium, i. e., we set

$$\omega_f J_f = kT.$$

Under these assumptions the integration over ω_f is elementary because of the δ function, and the sum over ν in the equation for ϱ amounts to evaluating

$$\sum_{\nu=-\infty}^{+\infty} \nu^2 Y_\nu^2\left(\nu \frac{\omega r_0}{c}\right).$$

For the case of a harmonic oscillator an analytical expression for the above sum has been found by SCHOTT[17], namely

$$\sum_{\nu=-\infty}^{+\infty} \nu^2 J_\nu{}^2(\nu z) = \frac{z^2(4+z^2)}{8(1-z^2)^{1/2}} \tag{22}$$

and since $z = r_0\,\omega/c$ is small compared to one, the right hand side is replaced by $\frac{1}{2} z^2$.

In general an analytical expression for the sum $\sum_\nu \nu^2 Y_\nu^2(\nu z)$ cannot be found, but we can find an approximate expression for the sum valid under the same conditions as above.

Expanding in powers of z we have

$$\sum_\nu \nu^2 Y_\nu^2(\nu z) = \sum_\nu \nu^2 Y_\nu^2(0) + 2z \sum_\nu \nu^3 Y_\nu(0) Y_\nu'(0)$$
$$+ z^2 \sum_\nu \nu^4 [Y_\nu'(0)^2 + Y_\nu(0) Y_\nu''(0)] + \ldots$$

From the definition of $Y_\nu(z)$ it is immediately seen that $Y_\nu(0) = \delta_{\nu,0}$ and therefore the above expansion reduces to

$$\sum_\nu \nu^2 Y_\nu^2(\nu z) = z^2 \sum \nu^4 Y_\nu'(0)^2 + 0(z^3).$$

Using the definition of $Y_\nu(z)$ and Parsevals theorem for the Fourier coefficients we have

$$\sum \nu^4 Y_\nu'(0)^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} (\vartheta''(\alpha))^2 \, d\alpha$$

and therefore

$$\sum \nu^2 Y_\nu^2\left(\nu \frac{\omega r_0}{c}\right) = \frac{\omega^2}{2\pi c^2}\int_{-\pi}^{\pi} (r''(\alpha))^2 \, d\alpha$$

$$= \frac{1}{2\pi c^2 m^2}\int_0^T (m\ddot{r}(t))^2 \, dt$$

$$= \frac{1}{c^2 m^2 \omega^2}\overline{\left(\frac{\partial V}{\partial r}\right)^2} \tag{23}$$

where the bar denotes time average.

With the same abbreviations as we have used before, i. e.,

$$\varkappa = \lambda^2 \frac{V_f^2}{\omega_f^2}$$

$$\mu = \frac{M^{(N)}}{L}$$

$$x = E\beta = \int \omega(J)\,\beta\, dJ$$

we have

$$\frac{\partial \Phi}{\partial t} = \frac{\varkappa\beta}{\mu c^3 m^2}\,\omega(x)\,\frac{\partial}{\partial x}\,\omega^{-1}(x)\overline{\left(\frac{\partial V}{\partial r}\right)^2}\left(1 + \frac{\partial}{\partial x}\right)\Phi \tag{24}$$

which deviates from the classical Kramers equation in that $J(x)$ has been replaced by (a constant times) $\omega^{-1}\overline{\left(\frac{\partial V}{\partial r}\right)^2}$. It is easily seen, for instance by using the virial theorem, that for a harmonic oscillator

$$\overline{\left(\frac{\partial V}{\partial r}\right)^2} = \gamma\beta x$$

where γ is the force constant, so that for this case we obtain the same result as previously, namely

$$\frac{\partial \Phi}{\partial t} = \eta \frac{\partial}{\partial x}\, x\left(1 + \frac{\partial}{\partial x}\right)\Phi$$

with

$$\eta = \frac{\varkappa\gamma}{\mu c^3 m^2}.$$

To be able to use this formalism on non-harmonic oscillators we must first estimate how large an error in the rate constant we commit by replacing

the right hand side of eq. (22) by $\frac{1}{2}z^2$. This is necessary because the validity of eq. (24) depends on a similar approximation.

Using the complete expression (22) in the master equation eq. (19) we find

$$\frac{\partial \Phi}{\partial t} = \eta \frac{\partial}{\partial x} x \frac{1+\varepsilon x}{\sqrt{1-\varepsilon x}} \left(1 + \frac{\partial}{\partial x}\right) \Phi$$

where ε is the ratio of kT to $\frac{1}{2}mc^2$, m being the mass of the particle and c the sound velocity.

Using the variational principle and the same trialfunction as above we find that for large values of x

$$k = x^* e^{-x^*} \left[1 + \frac{3}{2}\varepsilon x^* + \frac{7}{8}\varepsilon^2 x^{*2} + \ldots\right]$$

and since εx^* is smaller than one for all cases of interest this expression is permissible. Although the approximation obtained by replacing eq. (22) by $\frac{1}{2}z^2$ primarily is good when one only considers the Brownian motion of the molecule at low energies it is seen that the ratio between the correct rate constant and the approximate rate constant is only a factor 2–3 even when $\varepsilon x^* \approx 1/2$. The temperature dependence of the preexponential factor is of course somewhat different in the two expressions, but since this dependence is small anyway and not easily accessible experimentally this is of minor importance.

We can therefore presumably use eq. (24) derived above for estimating the influence of anharmonicities on the rate of escape of a particle from a potential minimum.

For the box-potential, which is zero for $|r| \leqslant l/2$, we interprete $\overline{\left(\frac{\partial V}{\partial r}\right)^2}$ as $\left(\overline{\left|\frac{\partial V}{\partial r}\right|}\right)^2 = 16\,E^2/l^2 = \frac{2\gamma}{D}E^2$, where γ is the spring constant for the equivalent harmonic potential i. e. the harmonic potential for which the stretching $l/2$ corresponds to the potential energy D.

We then get

$$\frac{\partial \Phi}{\partial t} = 2\,\eta x^{*-1} \sqrt{x} \frac{\partial}{\partial x} x^{3/2} \left(1 + \frac{\partial}{\partial x}\right) \Phi$$

and now η has the same meaning as above. Using the variational principle we get

$$k_0' = 2\, x^{*-1} \frac{x^* - \frac{3}{2}\left[1 - G(x^*)\right]}{2\, F(x^*)\, e^{x^*} + G(x^*) - 2}$$

or, in the asymptotic limit $x^* \to \infty$

$$k_0 = \frac{2}{\sqrt{\pi}}\, \eta \sqrt{x^*}\, e^{-x^*}.$$

This is precisely the same result as that obtained by Kramers' theory in the limit $x^* \to \infty$, in spite of the fact that the equations for Φ are completely different.

For the potential $C\,|r|$ we have $\left(\frac{\partial V}{\partial r}\right)^2 = C^2 = \frac{1}{2} D\gamma$ where γ is again the spring constant of the equivalent harmonic potential. We therefore have

$$\frac{\partial \Phi}{\partial t} = \frac{1}{2}\, \eta x^* \frac{1}{\sqrt{x}} \frac{\partial}{\partial x} \sqrt{x} \left(1 + \frac{\partial}{\partial x}\right) \Phi$$

and, using the variational principle, we get

$$k_0' = \frac{1}{2}\, \eta x^* \frac{e^{-x^*}\left[1 - G(x^*)\right]}{F(x^*) - e^{-x^*}\, G(x^*) - \frac{4}{3} x^*\, e^{-x^*}}$$

or, in the asymptotic limit

$$k_0 = \eta \frac{1}{\sqrt{\pi}}\, x^{*3/2}\, e^{-x^*}.$$

This result deviates only by a factor of $\frac{3}{4}$ from what one obtains from the theory of Kramers. We see therefore that for these simple anharmonic potentials the theory of Kramers and the more rigorous theory give practically the same results for the rate constants in the limit of large activation energy in spite of the fact that the equations for the density in phase space look quite different. This does not necessarily mean, however, that the result also would be almost identical for, for instance, the Morse potential, because in this case $r = r(\alpha, J)$ cannot be factorized, and therefore the short-cut used above cannot be applied.

Conclusion

The main result of the above calculations is that anharmonicities in the intramolecular potential changes the rate of a dissociation reaction by a factor which is approximately inversely proportional to the generalized frequency of the particle when its energy is equal to the dissociation-energy.

This only holds true for potentials of the form $C \mid r \mid^n$. For the Morse potential, for which the frequency goes to zero at the dissociation limit, the increase in rate over that for the harmonic potential is only a factor of two.

The result seems to be rather independent of whether one uses the theory of Kramers or a more refined theory. In the latter case however, owing to mathematical difficulties, only some of the simpler potentials could be treated, not the Morse potential.

The fact that the rate is increased when the frequency decreases with energy is well known. In the language of quantum mechanics it means that the rate is increased when the density of energy levels increases with the energy. It is remarkable, however, that the increase in rate obtained in this way for the Morse potential is far smaller than one should have expected. RICE[13], for instance, estimated that the anharmonicity in the Morse potential would speed up a dissociation reaction by a factor of twenty over that of a harmonic oscillator molecule. The most intuitive reason for this is probably to be sought in the fact that for potentials of the form $V = C \mid r \mid^n$ the anharmonicity is operative already at very low energies ($r \sim 0$) whereas the Morse potential is almost harmonic up to fairly high energies.

Although the results obtained here for the Morse potential using the Kramers theory conceivably could be changed somewhat by a more rigorous theory, we feel that the influence of anharmonicity in weak interaction theories has perhaps been somewhat overestimated in the past. Since hard core interactions play an important role in gasphase kinetics and no experimental result, to our knowledge, exists for dissociation of molecules interacting with phonons, it would be premature to try to compare with experiments at this stage.

Acknowledgement

This research has been supported in part by the Air Research and Development Command, United States Air Force and in part by Carlsbergfonden. K. A. acknowledges a grant from Teknisk Videnskabelig Forskningsråd, and T. A. B. acknowledges an illuminating discussion with Professor I. PRIGOGINE and Dr. J. HIGGINS.

References

1. E. W. Montroll and K. E. Shuler, Advances in Chemical Physics, *I*, 361 (1958).
2. K. E. Shuler, J. Chem. Phys. **31**, 1375 (1959).
3. E. E. Nikitin and N. D. Sokolov, J. Chem. Phys. **31**, 1371 (1959).
4. B. Widom, J. Chem. Phys. **31**, 1387 (1959).
5. I. Prigogine and T. A. Bak, J. Chem. Phys. **31**, 1368 (1959).
6. H. A. Kramers, Physica **7**, 284 (1940).
7. T. A. Bak, M. Goche and F. Henin, Mol. Phys. **2**, 181 (1959).
8. R. Becker, Theorie der Wärme, p. 274, Springer, Berlin 1955.
9. E. Jahnke und F. Emde, Funktionaltafeln, p. 275, Dover, New York, 1945. Bateman Manuscript Project. Higher Transcendental Functions. Vol. 1, p. 278, McGraw-Hill, New York, 1953.
10. T. A. Bak and I. W. Plesner, Acta Chem. Scand. **14**, 1310 (1960).
11. T. A. Bak, J. Phys. Chem. **60**, 1611 (1956).
12. R. Brout and I. Prigogine, Physica **22**, 35 (1956).
13. O. K. Rice, J. Chem. Phys. **9**, 258 (1941).
14. T. A. Bak, Contributions to the Theory of Chemical Kinetics, Munksgaard, Copenhagen, and Wiley, New York, 1959.
15. P. Mazur, Physica **25**, 29, 149 (1956).
16. H. C. Brinkman, Physica **22**, 29, 149 (1956).
17. G. N. Watson, A Treatise on the Theory of Bessel Functions. Second Edition, Cambridge University Press, Cambridge 1944, p. 573.

Indleveret til Selskabet den 24. maj 1961.
Færdig fra trykkeriet den 30. december 1961.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 8

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 8 (1962)

RANGES OF RECOIL IONS FROM α-REACTIONS

BY

LISE BRYDE, N. O. LASSEN AND N. O. ROY POULSEN

København 1962
i kommission hos Ejnar Munksgaard

CONTENTS

Synopsis

The recoil Ga^{66} ions, produced in (α, n) reactions when a thin copper layer is bombarded by α-particles from the cyclotron, are stopped in a pure gas. The thermalized ions are collected by means of an electric field, and from measurements of the activity distribution on the collector electrode the range distribution is obtained. In each gas, H_2, D_2, He, N_2 or A, the mean range is found to be nearly proportional to the energy E in the interval $0.6\ \text{MeV} < E < 1.2\ \text{MeV}$, in agreement with a theoretical formula given by Lindhard and Scharff. In this energy interval both electronic and nuclear stopping are of importance.

The reliability of the method is discussed. The shape of the range distribution in H_2 is compared with the calculated shape to be expected as a result of neutron emission from the compound nuclei, and from the half widths in various gases estimates of the straggling are obtained.

In a special experiment the range of Ga^{66} ions in copper is estimated. In other measurements the ranges of potassium ions in argon and F^{18} ions in nitrogen are obtained by the collector method. Gallium and potassium ions are found to be positive when thermalized, whereas F^{18} in nitrogen are found to be predominantly negative.

Printed in Denmark
Bianco Lunos Bogtrykkeri A-S

1. Introduction

The total charge z^* of a heavy ion moving through matter is determined by a balance between electron capture and loss processes[1,2]. A convenient, though not accurate, rule of thumb is the Bohr formula

$$z^* = z^{\frac{1}{3}} \cdot \frac{v}{v_0}, \tag{1}$$

where z is the nuclear charge, v the velocity of the ion, and $v_0 = 2.2 \times 10^8$ cm/sec is the orbital velocity of the hydrogen electron. For fast ions like fission fragments the mean charge is high at the beginning of the path, but low at the end. Accordingly, the energy loss caused by electronic encounters decreases along the range, and near the end it becomes smaller than the loss caused by nuclear collisions, which increases towards the end. The total charge depends on the stopping substance[3]. The variation of the charge with velocity and stopping substance makes range calculations rather difficult, and experimental data on range energy relations for heavy ions will always be of great value. This may be especially true for particles with an initial velocity $\sim v_0$, for which electronic and nuclear stopping may be of the same order of magnitude.

When a heavy particle is moving either through hydrogen or through deuterium, the average total charge corresponding to a given velocity must be expected to be the same in both gases and, consequently, the electronic stopping is the same[4]. The nuclear stopping, however, is smaller in D_2 than in H_2. Therefore the range of fission fragments is longer in D_2 than in H_2[5]. Since the difference stems from the part of the path where $v \lesssim v_0$, the relative difference should be greater for particles with an initial velocity of the order of v_0. Such particles may be obtained by bombarding medium heavy elements like copper with α-particles of 20 MeV, which is the energy of our cyclotron beam. If a thin copper foil is used as a target, the compound nucleus, produced when a copper nucleus is hit, will be expelled from the foil and move in the forward direction with the center of mass velocity. It was anticipated that the study of the range of such recoil par-

ticles in different gases might yield valuable information regarding the relative importance of the nuclear and electronic stopping.

Experiments of that kind were earlier made by HARVEY, DONOVAN, MORTON, and VALYOCSIK[6]. These authors measured ranges in various gases of recoil ions from the reaction Ra^{226} $(\alpha, 4n)$ Th^{226}, using 40 MeV α-particles. They found a slightly smaller range in D_2 than in H_2; this is opposite to the case of fission fragments, but the recoil Th-ions have velocities much smaller than v_0, and such low velocity particles may be assumed to behave in a different way[2, 8].

By a method very similar to that of HARVEY et al. we measured ranges in H_2, D_2, He, N_2, and A of Ga^{66} ions from the reaction

$$^{63}_{29}\mathrm{Cu} + ^{4}_{2}\mathrm{He} \rightarrow ^{67}_{31}\mathrm{Ga}^{*} \rightarrow ^{66}_{31}\mathrm{Ga} + ^{1}_{0}n, \qquad (2)$$

using α-particles of 10, 13, and 19.6 MeV, corresponding to average ion energies of 0.61, 0.79, and 1.19 MeV, respectively, or average ion velocities of 1.32, 1.50, and 1.84×10^8 cm/sec, respectively. Also, ranges of potassium ions in argon and F^{18} ions in nitrogen were measured. By another method the range of Ga^{66} ions in copper was estimated.

In the next section, the experimental arrangement will be described, and in section 3 the reliability of the method is discussed. In section 4, the results of the Ga^{66} measurements in gases are given and discussed, section 5 deals with the Ga range in copper, and section 6 with the ranges of K and F^{18} ions. Finally, in section 7, the widths of the Ga range distributions and the angular distribution will be discussed.

2. Experimental method and apparatus

Formula (1) is not valid for very small velocities, and the charge is not zero at the end of the path. It is well known from the standard way of producing, for instance, a ThB deposit, that recoil ions from some α-disintegration processes are positively charged when brought to rest in a gas. The present method is based on the fact that the Ga^{66} ions will also be positive when stopped, so that they can be collected on a negative electrode.

Since the α-beam from the Copenhagen cyclotron was used for these experiments only 1–2 hours per day, the experimental apparatus had to be made in a way which would allow the beam to be used for other pur-

poses the rest of the time. The recoil chamber was made so that it could be placed inside an existing scattering chamber and easily removed again after use. Apart from the fact that this arrangement was decisive for some of the dimensions, the special construction features implied by it are of no interest here, and Fig. 1 only shows the principal features.

The α-beam was stopped down to a diameter of 7 mm by a lead diaphragm 10 cm from the entrance window of the chamber. The window was 10 mm in diameter; it was made of a 1.2 mg/cm^2 plastic foil with a thin layer of copper on the inside surface, which served as the target. The

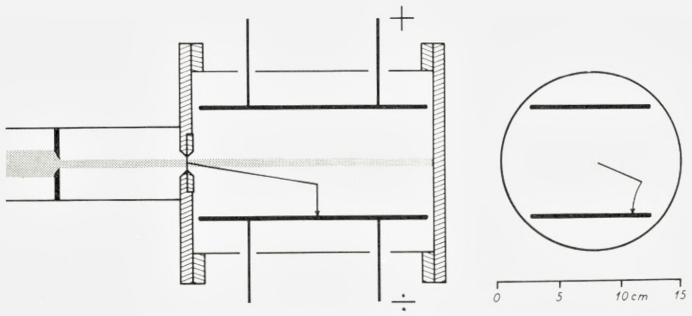

Fig. 1. Experimental apparatus.

layer, which was deposited on the plastic by evaporation in vacuo, was transparent and in some cases so thin that it was hardly conducting; the thickness was estimated from the amount of copper used and the geometry of the evaporation chamber. The uncertainty is about a factor of two. In the actual range measurements, a layer thickness of 5–10 μg/cm^2 was used, but occasionally a somewhat thicker layer served as the target in auxiliary experiments.

The chamber itself was a piece of a 6 inch steel tube. Inside it, there were placed two 3 mm brass plates, 10 × 19 cm^2, supported by Teflon insulators (not shown). One plate was held at +V volts, the other at −V volts. In some cases, the negative plate was replaced by a semicircular rod, 2 cm in diameter; the positive plate was then earthed like the rest of the chamber. V was chosen somewhat below breakdown potential, different for different gases and pressures. It ranged from 200 to 2000 volts. The ionization currents, of the order of 20–100 μA, were used by the cyclotron operator to maintain the machine at optimum conditions. The α-current itself which was not measured, was of the order of 0.05 μA.

On the inside surfaces of the plates grooves were cut lengthwise and crosswise; they were spaced 1 cm and formed a whole quadratic coordinate net. Before each experiment the plates were covered with aluminium foils 3 mg/cm^2 thick; the foils were bent round and fastened on the back sides by means of adhesive tape. By cautiously sliding a stick along the grooves the coordinate lines were transferred to the foils. After each bombardment the aluminium foils were cut along the lines, and the activities of the pieces were measured.

The chamber was filled with a pure gas. Before and after the bombardment the pressure was measured on a mercury gauge. The connection to the manometer was via a stopcock and, to avoid any possible influence of mercury vapour, the stopcock was opened only a few seconds and precautions were taken to have the main gas flow always going towards the manometer. When He was used, the chamber was connected to a liquid air charcoal trap. The other gases were continuously circulated through a side tube with hot calcium. This is a well working, standard procedure for the purification of A. For H_2, D_2, and N_2 special precautions had to be taken. When using these gases the temperature of the calcium was kept below a certain value (not known on an absolute scale), and before the actual experiments the calcium was saturated with the gas at the proper pressure and the temperature to be used. Separate purifiers were used for each gas.

The radioactivity of the aluminium pieces were measured by a 1 1/2 × 1 1/2 inch NaJ crystal. Each little piece of aluminium could be put in its own small specimen tube and pressed down against the flat bottom by a weight. During the counting the specimen tube was kept in a standard position right on top of the crystal by means of a holder. Small corrections had to be applied because the bottoms of the various specimen tubes were slightly different; corrections for decay were also applied. Often several aluminium pieces, for instance the 10 pieces from a whole row, were put in the same specimen tube. To speed up the counting four counter sets were used, each consisting of the crystal, the photomultiplier, the amplifier, and a single channel analyzer.

Reaction (2) was chosen, among other reasons, because Ga^{66} is a convenient nuclide, its half-life being 9^h, which leaves plenty of time for counting; its γ-spectrum contains rather strong high energy lines, and by simply using a bias of 1.7 MeV one can avoid counting almost any other possible activity. Na^{24} might be produced by high energy neutrons in the aluminium, but it was not found in significant amounts.

The other activities (K^{42}, K^{43}, and F^{18}) were measured with a properly chosen single-channel window, selecting a suitable γ-line. For the adjustment standard sources of Co^{60} (1.17 and 1.33 MeV), Cs^{137} (0.66 MeV), and Na^{22} (0.51 MeV) were used.

3. Discussion of the method

One might consider the following questions:

1. Will the Ga^{66} ions remain positive when stopped down to thermal velocities? Or will some be positive, some negative, and some neutral? Or will a particular ion have a fluctuating charge? It is clear that the collection along the electric field lines can only be good when the ions, after being thermalized, remain positive (or negative). If the ions are sometimes neutral, they will diffuse around, and the distribution will be smeared out.
2. If the ions are positive, will there still be some diffusion?
3. If the collection works well, what is the influence of the inhomogeneity of the field?
4. Will the ions, when collected on the aluminum foil, stick to the spot, or is it possible that they may again be liberated as neutral atoms?

The α-particles produce of the order of 10^{15} ion pairs per sec. If the electrons attach themselves to some impurity molecules to form negative molecular ions, some risk exists that they may collide with Ga^{66} ions and neutralize them. One reason for using very pure gases is to avoid attachment and to secure a fast removal of the negative ions. Other reasons are that, in pure gases, it is reasonable to expect[7] that clustering does not occur, that charge exchange reactions between thermal Ga^{66} ions and molecules can be neglected, and that the positive ion collection time is only a fraction of a milli-second, so that diffusion will be completely unimportant. Furthermore, an important reason is that possibly the fast Ga^{66} ions may have a mean charge and a mean range depending somewhat on even rather small impurity admixtures.

It was found that more than 90 per cent of the Ga^{66} activity was collected on the negative plate when the voltage was sufficiently high. Less than 5 per cent was found on the positive plate and less than 5 per cent on the walls of the chamber. As shown in Figs. 2 and 3, the activity on the negative plate was distributed in a rather broad peak, but this was to be expected, because the neutrons emitted from the compound Ga^{67} nuclei

will give the Ga^{66} nuclei recoil momenta varying in direction and magnitude. In fact, calculations which will be more closely discussed below, indicate that the width caused by neutron recoil is comparable to the experimental width found in the light gases. Experience thus seems to show that the method works for Ga^{66}.

In Fig. 2 are plotted the Ga^{66} activities of the aluminium pieces against

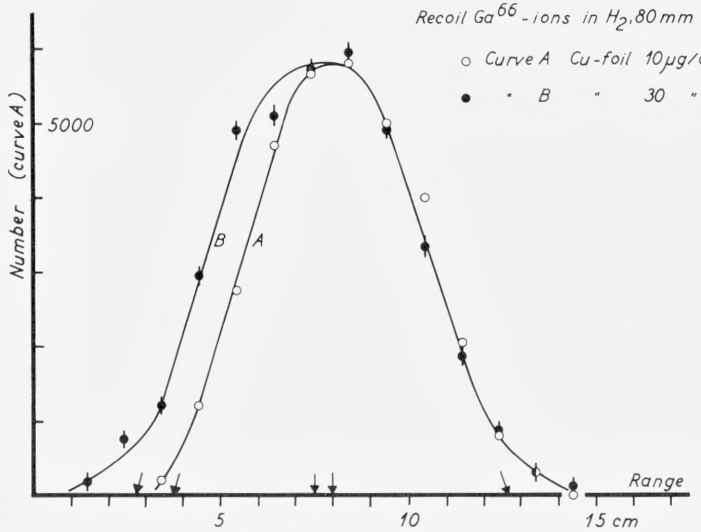

Fig. 2. Distribution of Ga^{66} activity along the negative collector plate for two thicknesses of the Cu layer and with the chamber filled with H_2 to a pressure of 80 mm Hg (23° C).

their positions along the collector plate. The ten pieces from each row are added. The abscissae are the distance from the window as measured in the beam direction. Since some particles diverge they will actually have travelled longer. The mean value as determined from the curve therefore is the mean of the projection of the ranges, and not the mean of the ranges themselves. The difference will be only a few per cent and can be neglected (cf. section 7). It may be emphasized that we are here talking about a purely geometrical effect, neglecting the influence of scattering in the gas. The latter phenomenon implies that the total path length, especially in the heavier gases, will be longer than the range, and this difference may be of much larger magnitude.

Fig. 2 gives further evidence for the reliability of the method. It should be

expected that curve B obtained with the thicker target would follow the thin target curve A on the right side, because the thick target may be considered to be made up of a stack of thin targets, but, on the left side, B would be displaced against smaller range values, in qualitative agreement with the figure. Since the mean range in copper is about 270 $\mu g/cm^2$ (see section 5), the displacement should be about $\frac{30-10}{270} \times 8 \sim 0.6$ cm, 8 being

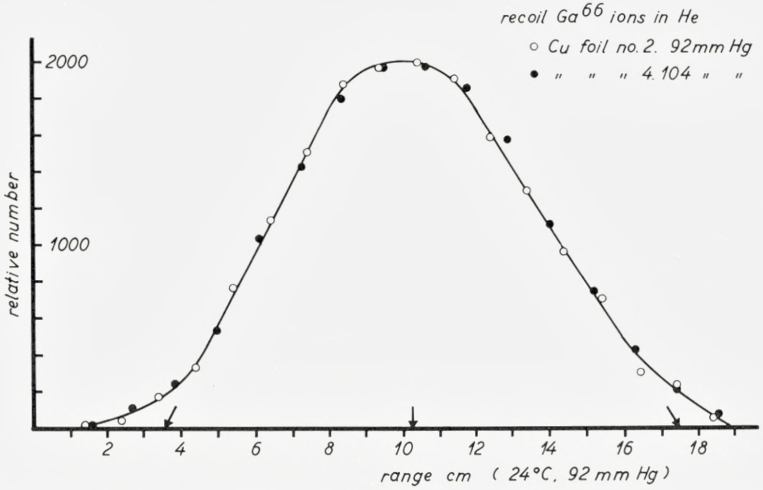

Fig. 3. Distribution of Ga^{66} activity along the negative collector plate when using a thin Cu layer (5–10 $\mu g/cm^2$) and He as stopping gas.

the mean range as obtained from curve A. The displacement is slightly larger, ~ 1 cm, but since neither the thickness nor the range in copper is accurately known, the quantitative agreement is not too bad. It is also inferred that, when a target thickness not exceeding 10 $\mu g/cm^2$ is used, the target contributes only little to the width of the distribution.

Fig. 3 shows the result of two measurements in He. Two different copper layers of about equal thickness (~ 10 $\mu g/cm^2$) and two He-pressures were used. Within a few millimeters the two sets of points show the same distribution. Here it might have been more convincing if the difference between the two pressures had been greater. However, in each experiment the pressure was purposely chosen in such a way that the peak fell not too far from the middle of the chamber where the electric field has no component in the

α-beam direction. Towards the end of the plates the field inhomogeneity will distort the results, and the lower parts of the curves — in Fig. 3 to the left of ~ 4 cm and to the right of ~ 16 cm — do not reflect accurately the actual range distributions.

The distribution curves were the same whether obtained with the collector plate or with the semi-circular rod. All evidence thus indicates that the longitudinal distributions may be regarded with some confidence.

On the contrary, lateral distributions measured by means of the activity on the plate are of no value. Even though the positive ions are rather quickly removed, the large number of them will create a space charge which will distort the electric field in a way as sketched in Fig. 4a. In Fig. 4b the black points and the full drawn curve show the activity of the 10 aluminium pieces in the row corresponding to the mean range. The appearance of the curve may be understood by help of Fig. 4 a. One consequence of the field distortion is the large broadening of the curve, demonstrated by comparison with the dotted curve and the white points which were obtained in the following way: 20 mm behind the window a circular lead disk, 10 mm in diameter, was placed; it stopped the beam as well as the recoil ions moving nearly forward. The dotted curve gives the activity distribution along the same row of aluminium pieces as before, but now there is no positive space charge. For the latter curve the central dip is due to the missing recoil ions in the forward directions, and the shape of the curve agrees with rough calculations. For the former curve the central dip is, at least mainly, a consequence of the field distortion.

The field distortion by space charge will have no influence on the longitudinal distribution. However, in order to measure the latter correctly, some knowledge of the lateral spread is necessary, because it has to be avoided that the recoil ions strike the plates before being thermalized in the gas. The dotted curve in Fig. 4b gives some information on the lateral spread and indicates the fulfilment of this requirement. Further indication was obtained in experiments where the plates were removed and the end flange of the chamber was covered with two aluminium foils. During bombardment the chamber was evacuated. Afterwards the foils were cut into circular rings by means of especially prepared punches, and the activities of the rings were measured. Fig. 5 shows the Ga^{66} activity on the catcher foil. The underlying foil was inactive (only γ-energies > 1.7 MeV were measured) with the exception of the innermost circle which was hit by the α-beam. For this circle the two foils were about equally active, but since it may not be justified

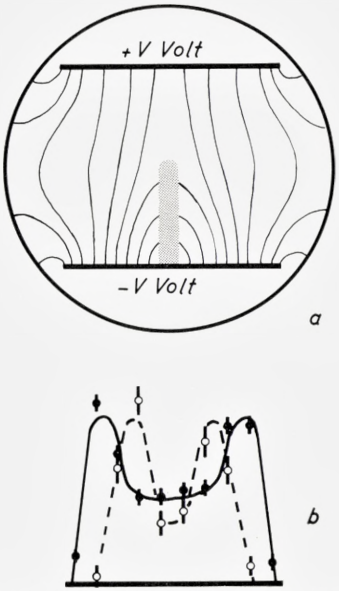

Fig. 4. a) Distortion of the electric field due to positive space charge (shaded area in the figure). Cross section perpendicular to the beam direction. Qualitative sketch. – b) Distribution of Ga^{66} activity across the negative plate. Full drawn curve under normal conditions, dotted curve when the beam and the recoil ions at small angles are stopped.

to use the difference between the activities of the two circles as a measure of the Ga^{66} nuclei from the copper layer*, the latter could not be determined for the innermost circle.

In Fig. 5 curve b gives a reasonably good fit to the experimental points. The integral curve e shows the percentage of the total number of particles within a cone of half angle θ equal to the abscissa. It may be seen that 85% of the recoils emerge from the target foil with $\theta < 12°$, and if the angular distribution were not changed by the stopping gas the full length of the chamber could be used without fear of distortion due to particles being lost by striking the plates. In all actual range measurements only the tail of the distribution curves were allowed to exceed a distance of some 13—14 cm from the foil. In argon, where the scattering is largest, the mean range was kept below 9 cm, and it is believed that a negligibly small amount of recoils was lost.

* Some of the active nuclei produced in the first foil will be thrown into the next foil, which, in the absence of Ga^{66} from the copper target, would have the higher activity.

Attempts were made to measure the angular distribution with gas in the chamber. An aluminium covered plate was placed perpendicular to the beam at a distance from the target foil corresponding to the mean range, and again the activity of 6 mm wide rings was measured. Curves obtained with and without H_2 were almost identical and in fairly good agreement with the distribution to be expected according to Fig. 5. In A the distribu-

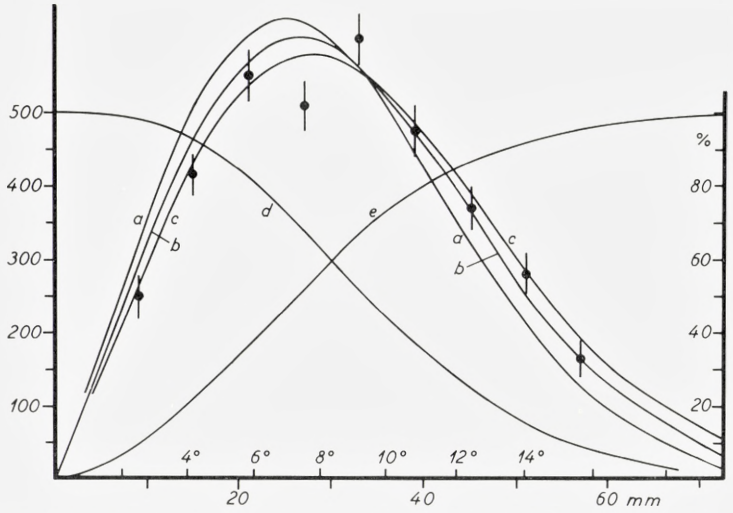

Fig. 5. Radial distribution of Ga^{66} activity on the end flange of the evacuated recoil chamber. The Ga^{66} was produced by bombarding Cu with 19.6 MeV α-particles. The abscissa is the radial distance from center. A scale showing the projection angle θ of the ions is also given. The points show the activity in relative units on circular rings, each 6 mm wide. a, b, and c, are calculated curves to be discussed in section 7, p. 26. They show the $I(\theta)\,d\theta$ distribution. Corresponding to b, curve d shows the $I(\theta)\,d\omega$ distribution, and curve e the integral $\int_0^\theta I(\theta)\,d\theta$.

tion was much broader; the measurements were not completely reproducible, perhaps because, since the actual collector plates were removed, no sufficiently good electric field was applied, and hence some Ga^{66} atoms stopped in the gas may have reached the end plate by diffusion. However, the measurements showed that less than 4% of the activity on the catcher foil was found at radii larger than 45 mm.

Before leaving the discussion of the method of collecting the recoil ions it may be mentioned that reproducible results were obtained only when the aluminium foils were handled with utmost care. By experiment it was found that 40—80% of the activity could be removed from the foil 1) by dipping

it in water or ethyl alcohol, 2) by rubbing it with a wet cloth or 3) by pressing a thumb against it. 20—40% was removed 1) by rubbing lightly with a clean, dry cotton wool cloth or 2) by touching gently with a clean, dry finger. Here is another reason for using a pure and dry gas. There may also be some reason for using as collector foil the aluminium which is chemically related to gallium.

4. Range of Ga^{66} ions in gases

Longitudinal distributions of Ga^{66} activity obtained in H_2, D_2, He, and A are shown in Fig. 6.

The difference in ranges in H_2 and D_2 demonstrates at once the importance of nuclear stopping, as discussed in the introduction. It also tells something about the electronic stopping.

For the velocity loss per cm due to nuclear encounters Bohr has given the formula (ref. 2, (5.1.2.))

$$-\frac{dv}{dx} = 2\pi N \frac{z_1^2 z_2^2 e^4}{m_1 m_2 v^3} L_v \tag{3}$$

with

$$L_v = \log\left\{ z_1 z_2 \sqrt{z_1^{2/3} + z_2^{2/3}} \frac{\mu (m_1 + m_2)}{m_1 + m_2} \cdot \left(\frac{v_0}{v}\right)^2 \right\}^{-\frac{1}{2}},$$

where N is the number of atoms per cm^3, m_1 and z_1 are the mass and nuclear charge numbers of the ion, m_2 and z_2 the corresponding values for the stopping substance, v is the ion velocity, and μ and e are the mass and charge of the electron. In a way described earlier (ref. 4, p. 31) the range energy relations in H_2 and D_2 may be calculated, assuming no electronic stopping. For Ga^{66} ions of velocity 1.84×10^8 cm/sec the range in D_2 would be 1.38 times the range in H_2. The experimental ratio is 1.17, thus indicating the importance of both electronic and nuclear stopping.

According to formula (3), the nuclear stopping power per cm will be about the same in D_2 and in He. The longer range in He shows again that the electronic stopping in D_2 is not negligibly small compared to the nuclear stopping. It shows furthermore that the electronic stopping is smaller in He than in D_2. In this connection it is interesting to remember that the total charge of fast fission fragments is smaller in He than in H_2 (or D_2)[3]. How-

ever, from the present measurements no conclusion regarding the ion charge in D_2 and in He can be drawn, because the range may be longer in He than in D_2 even if the charge values are equal. In fact, the ratio $R_{He}/R_D = 1.22$ between the ranges in He and in D_2 is closely the same as the ratio $\left(\frac{dE}{dx}\right)_H \Big/ \left(\frac{dE}{dx}\right)_{He} = 1.21$ between the stopping powers in H_2 and in He for

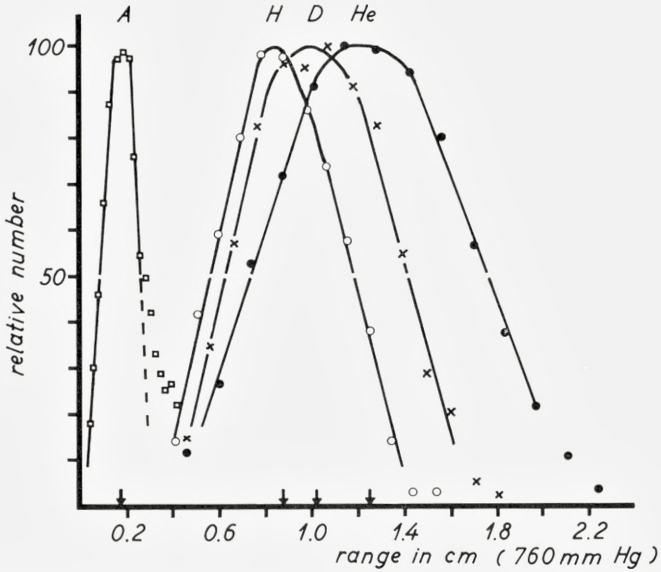

Fig. 6. Range distributions in H_2, D_2, He, and A of Ga^{66} ions produced by bombarding Cu with 19.6 MeV α-particles. The abscissa is the range in cm at 760 mm Hg and 23° C. – In the measurements the gas pressure was 71, 79, 104, and 17.3 mm Hg of H_2, D_2, He, and A, respectively.

5 MeV α-particles. This agreement between the figures is accidental; actually, the ratio between the electronic stopping powers for these slow ions may be expected to be higher than 1.2, but the ratio between the nuclear stopping powers in He and D_2 is about 1, and the range ratio depends on both.

For the heavier gases nitrogen and argon the experiments give $R_N/R_A = 1.02$, and one has for 5 MeV α-particles $\left(\frac{dE}{dx}\right)_A \Big/ \left(\frac{dE}{dx}\right)_N = 0.98$. This agreement may be understood in a somewhat similar way as for D_2—He.

When comparing the light and heavy gases one does not find such agreement between ratios of ion ranges and α-ranges. The experimental value for the ratio between the ranges in A and in He is $R_A/R_{He} = 0.135$,

whereas the ratio between the stopping powers for 5 MeV α-particles is $\left(\frac{dE}{dx}\right)_{\mathrm{He}} \Big/ \left(\frac{dE}{dx}\right)_{\mathrm{A}} = 0.183$, and for slower particles it may even be larger.

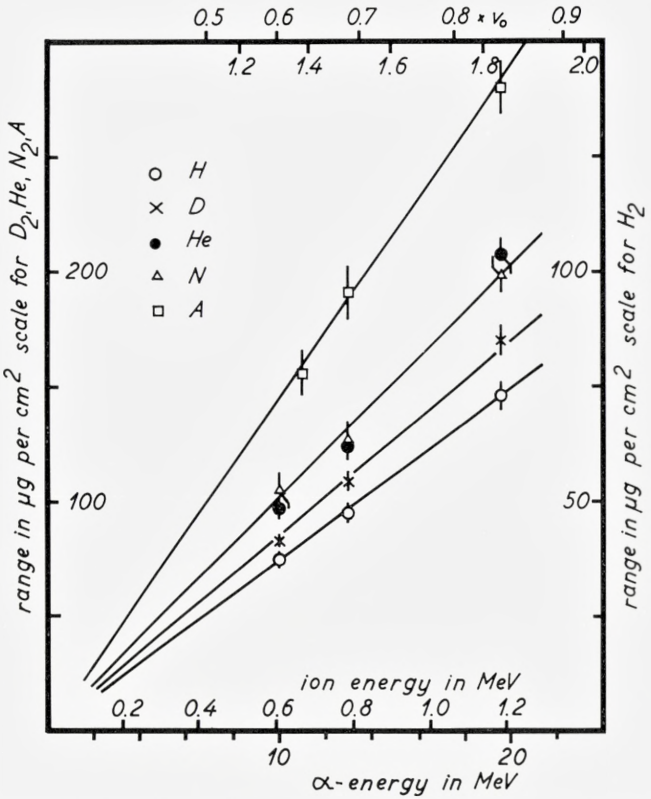

Fig. 7. Range in μg/cm^2 of Ga^{66} ions in gases. Note the different ordinate scales for H_2 and for the other gases. The ions are produced in the reaction Cu^{63} (α, n) Ga^{66} and the abscissa is the α-energy. A scale showing the mean ion energy is also given. On the top of the figure scales are given for the average ion velocity in units of 10^8 cm/sec and in units of v_0, the orbital velocity of the hydrogen electron.
The curves are straight lines through origo.

This again illustrates the influence of nuclear stopping which for the Ga ions, according to formula (3), is many times larger in A than in He.

Table 1 and Fig. 7 summarize the results obtained for various α-energies. A range correction of 2% for finite target thickness has been applied. To a rather good approximation the range in each gas is found to be proportional to the energy. The proportionality constants are given in Table 2.

TABLE 1. Range in gases of recoil Ga^{66} ions, given in mm (760 mm Hg, 23°C) and in $\mu g/cm^2$.

α-energy mean ion energy mean ion velocity	19.6 MeV 1.19 MeV 1.84×10^8 cm/sec				13.0 MeV 0.79 MeV 1.50×10^8 cm/sec				10.0 MeV 0.61 MeV 1.31×10^8 cm/sec			
	mm			$\mu g/cm^2$	mm			$\mu g/cm^2$	mm			$\mu g/cm^2$
	1.	2.	av.		1.	2.	av.		1.	2.	av.	
H_2	8.8	9.0	8.9	73	5.9	5.7	5.8	47.6	4.55	4.62	4.58	37.6
D_2	10.3	10.4	10.4	170	6.5	6.7	6.6	109	4.95	5.15	5.05	83
He	12.6	12.5	12.6	207	7.3	7.8	7.5	124	5.8	6.0	5.9	97
N_2	1.76	1.69	1.73	199	1.10			127	0.91			105
A	1.69	1.71	1.70	280	1.16			191	0.95*			156*

* α-energy 11.0 MeV.

Proportionality means, that the total stopping power $\frac{dE}{dx}$, equal to the sum, $\left(\frac{dE}{dx}\right)_e + \left(\frac{dE}{dx}\right)_v$, of the electronic and nuclear stopping powers, is constant in the energy range considered. Values for $\frac{dE}{dx}$ in various gases are also given in Table 2.

If the electronic stopping is neglected one should, for velocities $v \ll v_0$ just expect proportionality between range and energy (cf. formula (5.4.2) in ref. 2), while for $v \sim v_0$ one should expect a somewhat stronger energy variation (ref. 2, formula (5.3.2)). The present experiments show that both electronic and nuclear stopping play important roles, and none of them can be neglected. Now, for increasing velocities, the nuclear stopping decreases, but the electronic stopping increases and, in fact, it so happens that these two effects balance each other in such a way that, for a considerable interval of velocities, not only for $v \ll v_0$, the range is closely proportional to the energy. This is discussed by LINDHARD and SCHARFF[8] who give the formula

$$R' = k \times E \times \frac{m_2(m_1 + m_2)}{m_1} \cdot \frac{\sqrt{z_1^{2/3} + z_1^{2/3}}}{z_1 z_2}, \qquad (4)$$

where the units used for the range, the energy, and the masses are $\mu g/cm^2$, MeV, and mass units, respectively, and where theoretically $k = 600$.

It is borne out by the experiments that the formula gives a rather good approximation for Ga ions even for velocities comparable to v_0. Intro-

TABLE 2. R/E, range in $\mu g/cm^2$ divided by energy in MeV, and stopping power $\frac{dE}{dx}$ in keV per $\mu g/cm^2$ for Ga^{66} ions in gases.

	R/E for Ga^{66} ion energies			R/E Weighted mean	$\frac{dE}{dx}$
	1.19 MeV	0.79 MeV	0.61 MeV		
H_2	61.3	60.3	61.6	61.1	16.4
D_2	143	138	136	139	7.2
He	174	157	159	163	6.1
N_2	167	161	172	167	6.0
A	236	242	233*	237	4.2

* for energy 0.67 MeV.

ducing the experimental range values into the formula we find for k the values given in Table 3. It may be seen that the fit is very good for H_2 and for N_2. In D_2 the experimental values vary monotonically with velocity, which would indicate that the relative range difference between H_2 and D_2 decreases with decreasing velocity; however, the variations are hardly outside the experimental uncertainty. The rather large k-values in He show that, here, the electronic stopping plays a comparatively minor role than in the other gases. The small k-values in A reflect the influence of scattering in the stopping gas; actually, in (4) R′ stands for the average total path length, and the average (projected) range should be expected to be smaller than the former by a factor[8] $\frac{1}{1+\frac{1}{3}\frac{m_2}{m_1}} = 0.83$, thus leading to a k value of 500, in close agreement with the experiments.

TABLE 3. Experimental value of the constant k in formula (4).

Ga^{66} ion velocity in units of 10^8 cm/sec	1.84	1.50	1.38	1.31
H_2	580	570		580
D_2	660	640		620
He	750	680		690
N_2	590	560		600
A	510	520	550	
Cu	340			

5. Range of Ga^{66} ions in copper

A copper foil of thickness about 1.5 mg/cm^2 was bombarded with α-particles. A thin gold foil, thick enough to stop the recoiling Ga^{66} ions, was placed close behind it. If t is the thickness of the copper foil, a_1 its Ga^{66} activity, and a_2 the Ga^{66} activity of the gold foil, the quantity

$$R = t \frac{a_2}{a_1 + a_2}$$

may be taken as a measure of the mean range of Ga ions in copper.

Results of such measurements are shown in Fig. 9. It may be seen that the values for R are roughly proportional to the energy, and that they are not much different from the range in argon. If the R-values are multiplied by $\left(1 + \frac{1}{3}\frac{m_2}{m_1}\right)$ to give the total path lengths, the latter are found to be, within 2%, the same in A and in Cu.

In Table 3 a k-value is given. Assuming the range to be smaller than the path length by the factor $\frac{1}{1 + \frac{1}{3}\frac{m_2}{m_1}} \sim \frac{3}{4}$, the k-value to be expected would be ~ 450. However, when m_2 is about as large as m_1 just as for copper, the scattering gives rise to a very large smearing-out effect. Furthermore, the activity ratio $\frac{a_2}{a_1 + a_2}$ may depend on the collector foil, which in our case was gold, i. e. a substance with a rather high m_2 (back scattering).

6. Ranges of other recoil ions

In our measurements of the Ga^{66} activity we usually counted γ-rays with energies higher than 1.55 MeV. Using argon as a stopping gas it was found, however, that the range distribution had a foot on the right side. This may be seen in Fig. 6; it has only a negligible influence on the important part of the distribution curve. It is caused by a K^{42} activity, half life 12^h, produced in the gas by the reaction $A^{40}(\alpha, np)K^{42}$. K^{42} has a rather strong γ-line at 1.51 MeV. By counting, after the actual Ga^{66} measurements, γ-rays in the energy interval 1.45—1.60 MeV, it was found possible to obtain the K^{42} activity distribution as well as a corrected Ga^{66} distribution. From the corrected K^{42} distribution the range of K^{42} in A was obtained.

By the reaction $A^{40}(\alpha, p)K^{43}$ also K^{43} is produced. This nuclide has a half-life of 22^h; it has a strong γ-line at 0.615 MeV, which could be measured several days after bombardment.

In later measurements of the Ga^{66} range in A the counting limit was raised to 1.7 MeV and thus the Ga^{66} activity distribution was not disturbed

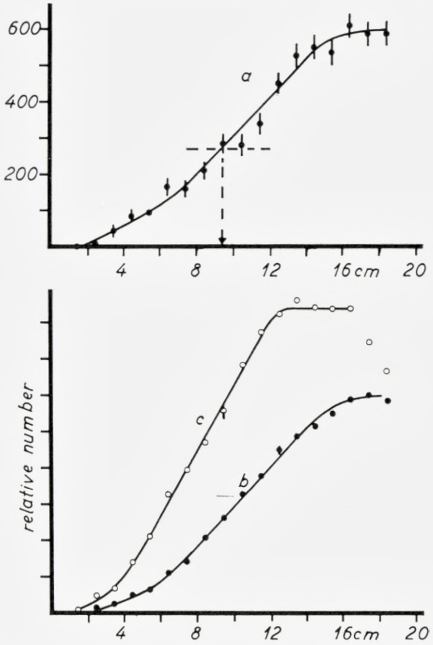

Fig. 8. Activity of K^{42} (curve a) and K^{43} (curves b and c) per cm of collector plate. The abscissae are the distances from the entrance window. The potassium is produced in the argon gas by (α, p) and (α, np) reactions. Curve a and b refer to the same argon pressure (17.3 mm Hg), curve c to a higher pressure (24.4 mm Hg).

by the K activity. The K^{43} distribution could still be obtained as a bi-product.

Fig. 8 shows some examples of distribution curves. Since the A-target is thick, the curves are of the integral type. If the range distribution were a sharp peak, the integral curve would reach half maximum height for an abscissa equal to the (mean) range R_m. Assuming a Gaussian distribution with full width at half maximum height equal to R_m*, one finds that the ordinate for R_m is not 0.5, but only 0.45 times the maximum height.

Clearly the determination of the mean range is less precise than for the

* This is a rough estimate; actually, the width may be larger (cf. section 7).

Ga ions, and the information about the width of the distribution is poor Also, the field inhomogeneity near the end of the plates may be more serious; in fact, the plateau must be reached before 15—16 cm, or the result will only be a lower limit for the mean range. For this reason the K^{42} and K^{43} ranges were measured in special experiments, in which no Cu-layer was

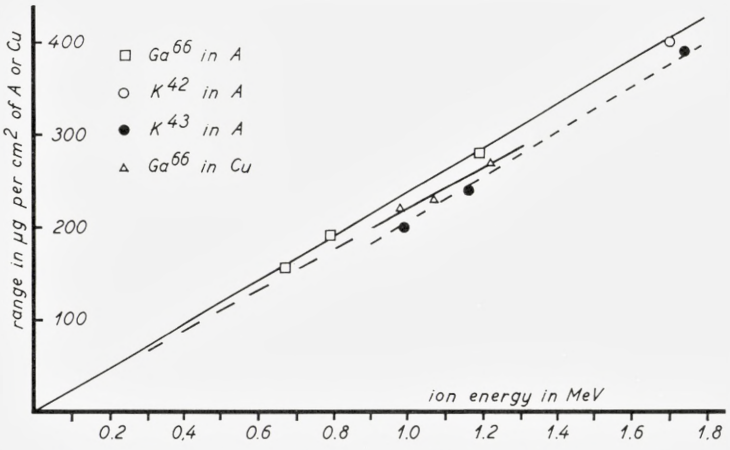

Fig. 9. Range of Ga and K-ions in A and of Ga ions in Cu.

used and the pressure was high enough to make the chamber length considerably larger than the ion range (see Fig. 8 c).

Similar experiments with no Cu-layer and with N_2 in the chamber yielded a value for the range of F^{18} ions in N_2. The 511 keV line was used; the half-life of 112^m was observed. Some shorter living activity produced in the plastic foil was allowed to die away, and only measurements made more than 3 hours after the bombardment were used.

The K ions in A were found to go predominantly to the negative plate, the positive plate having only about 10% of the activity. Both plates gave the same distribution.

The F^{18} ions in N_2 were found to behave in a different way, about 2/3 going to the positive plate and only 1/3 to the negative plate. The distribution on the positive plate was similar to the curve shown in Fig. 8 c. The distribution on the negative plate was probably identical, but it was measured on two other counters, and due to an accidental failure of the power supply, it was less reliable.

The different behaviour of the ions may give a hint concerning a possible

influence of the chemical nature of the ions on the ionic charge of thermal ions.

The values obtained for the ranges of K ions in A and F^{18} ions in N_2 are summarized in Table 4. The values printed in italics were determined in particularly designed experiments and they are considered to be the most reliable.

TABLE 4. Ranges in mm (760 mm Hg, 23°C).

α-energy in MeV	19.6		13.0	11.0
K^{42} ions in A	2.2	*2.38*	1	
K^{43} ions in A	2.2	*2.34*	1.4	1.1
F^{18} ions in N_2		*4.09*		

In Table 5 the ranges in $\mu g/cm^2$ and the k values to be inserted in (4) to fit the data are given. Here again $m_2 \sim m_1$, and if it were justified to use formula (4), one should expect $k \sim 450$. The small value of k for F^{18} ions in N_2 shows the non-validity of (4) for these rather fast ions ($v \sim 3.5\, v_0$). The nuclear stopping power computed from (3) would lead to a range more than 100 times larger than the experimental value, and it is thus found that for these ions the nuclear stopping is vanishingly small compared to the electronic stopping.

It may be noted that for K^{43} ions the range is found to be nearly proportional to the energy, and that for equal energy the K ions and the Ga ions have about the same range (see Fig. 9). The latter is contradictory to formula (4), as is also seen from the low k-values. In view of the close quantitative agreement between the formula and our Ga range values in gases, it seems strange that the formula should be in error by almost a factor

TABLE 5.

	Velocity* cm/sec × 10^8	Energy MeV	Range $\mu g/cm^2$	k
K^{42} in A	2.79	1.70	400	280
K^{43} in A	2.79	1.74	390	270
K^{43} in A	2.28	1.16	240	250
K^{43} in A	2.10	0.99	200	230
F^{18} in N_2	6.47	3.90	480	110

* Actually, the velocity of the compound nucleus which is assumed equal to the mean ion velocity.

of two for K ions which have velocities only slightly greater than the Ga ions. Neither can we imagine the experimental ranges to be so much wrong. A possible explanation for the discrepancy might be the following.

There is reason to believe that the Ga^{66} ions are produced in compound nuclear reactions and that the measured mean range corresponds to an ion velocity equal to the velocity of the compound nucleus (cf. next section). We have assumed that also the $A^{40}(\alpha, p)K^{43}$ reaction takes place via a compound nucleus, but if direct interaction processes are of importance, the residual nuclei may acquire smaller mean velocities, since the protons may be emitted predominantly in the forward direction.

This explanation does not seem too plausible. We should like to point out that the cases of disagreement are those in which $m_1 \sim m_2$ (Ga ions in Cu, K ions in A).

7. Discussion of range distributions and angular distributions of Ga^{66} ions

Remarks on straggling and nuclear temperature

As already mentioned, the spread in the range values is caused by 1) neutron emission from the compound Ga^{67} nuclei giving rise to a rather large energy spread of the ions, 2) straggling in the gas, 3) target thickness, 4) breadth of aluminium pieces, and 5) diffusion of the thermalized ions.

The contributions from the three last sources are small and will not be further discussed.

According to theory, the straggling increases with increasing mass number of the stopping gas, no matter whether the stopping is due predominantly to nuclear or to electronic collisions. This is also borne out by the experiments, as may be seen from Fig. 10.

Lindhard and Scharff give the formula

$$\frac{\sigma^2}{R'^2} = \frac{2}{3}\frac{m_1 m_2}{(m_1 + m_2)^2}$$

where σ is the standard deviation in range to be expected if the stopping were due entirely to nuclear collisions, and R′ is the average path length. The values A for the full width at half maximum height in per cent of the mean range, as obtained from this formula and by putting the average path length equal to $\left(1 + \frac{1}{3}\frac{m_2}{m_1}\right)$ times the mean range are given in Table 6.

The experimentally found total half widths B are also given and furthermore the values $C = \sqrt{B^2 - A^2}$. The relative uncertainty in the B-values may be estimated to about 4%. Then the uncertainties in the C-values are the figures given in the table.

If A were the correct scattering half-widths, C would be the partial half-widths resulting from other sources, i. e., essentially from neutron emission. Then, since the energy distribution resulting from this process does not

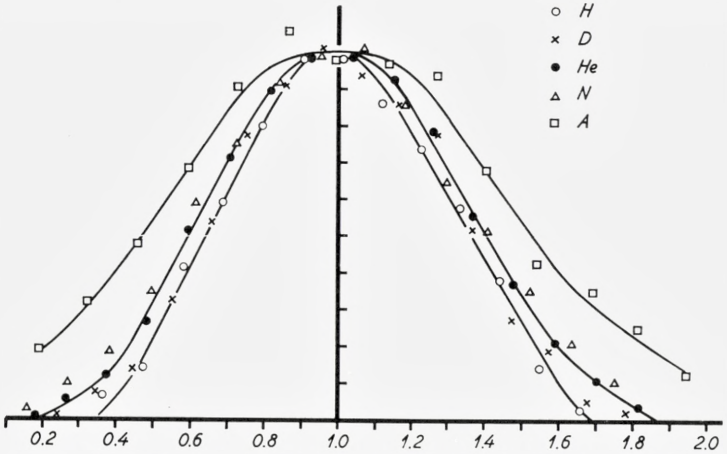

Fig. 10. Range distributions of Ga^{66} ions in gases. Ordinate: relative number of particles per unit interval of range. Abscissa: range in units of the mean range R_0. The Ga^{66} ions were produced by 19.6 MeV α-particles on Cu. Experimental points are given for H_2, D_2, He, N_2, and A. Curves are only drawn for H_2, He, and A.

depend on the gas, and since in each gas the range is proportional to the energy, the relative half widths C should be the same in all gases (not necessarily for all α-energies, see later). This is true within the experimental error for the light gases H_2, D_2, and He, whereas for N_2 the C-values come out too small, and for argon the experimental half widths are smaller than the A-values. This is not surprising; it merely shows once more that the electronic stopping cannot be neglected, and since the electronic collisions contribute less to the straggling than do the nuclear encounters, the real relative scattering half widths are smaller than the A-values.

For hydrogen the straggling is small compared to the range spread caused by neutron emission. As a first approximation, we may neglect the former and consider the value B_H as a measure for the latter. For the other gases the values $D = \sqrt{B^2 - B_H^2}$ will then represent the scattering half widths, the approximation being best for the heavy gases.

TABLE 6. Full width at half maximum height of range distribution in per cent of the mean range.
B are experimental values. For the meaning af A, C, and D, see text.

		$E_\alpha = 19.6$ MeV			$E_\alpha = 13$ MeV			$E_\alpha = 10$ MeV		
	A	B	C	D	B	C	D	B	C	D
H_2	23	68	64 ± 3	(0)	67	63 ± 3	(0)	63	59 ± 3	(0)
D_2	32	72	65 ± 3	(24)	72	65 ± 3	(26)	70	63 ± 3	(32)
He	45	81	68 ± 3	45 ± 8	84	59 ± 4	32 ± 12	76	62 ± 4	44 ± 8
N_2	78	85	(33)	51 ± 7	89	(43)	58 ± 6	88	(41)	62 ± 6
A	112	106		81 ± 6	104		79 ± 6	104		83 ± 5

We may ask, what should be the shape of the range spectrum if it is determined entirely by neutron emission? We shall make the two simplifying assumptions, (I) that the neutrons are emitted isotropically in the C.M. system and, (II) that the relative number of neutrons per unit interval of energy is given by

$$\frac{dn}{dE} = C_1 E e^{-\frac{E}{T}},$$

where C_1 is a constant and T, the nuclear temperature of the residual Ga^{66} nucleus, is also a constant [9) 10) 11)].

Introducing the momentum $P = \sqrt{2\,ME}$, where M is the neutron mass, gives

$$\frac{dn}{dP} = C_2 P^3 e^{-\frac{P^2}{2MT}},$$

where C_2 is a new constant. In this formula, $\frac{dP}{dn}$ may also stand for the number of recoil ions per unit interval of momentum. Denoting by Q the projection of P on the beam direction (see Fig. 11), the distribution in Q is given by

$$\frac{dn}{dQ}\,dQ = \iint \frac{\frac{dn}{dP}\,dP}{4\pi P^2\,dP}\,r\,d\varphi\,dr\,dQ = dQ \int \frac{\frac{dn}{dP}}{2P^2}\,r\,dr,$$

where φ is the azimuthal angle. From Fig. 11 we get $P^2 = Q^2 + r^2$, hence for a fixed Q: $rdr = PdP$, thus

$$\frac{dn}{dQ} = \frac{C_2}{2} \int_{P=|Q|}^{P=\infty} P^2 e^{-\frac{P^2}{2MT}}\,dP.$$

Let Q_0 denote the momentum in the laboratory system, due to center of mass motion. The projection of the lab. momentum is $Q + Q_0$. Introducing the assumption, which is justified from the previous results, that, (III), the projection R of the range*, is given by

$$R = C_3 (Q + Q_0)^2,$$

where C_3 is a constant, one finds

$$\frac{dn}{dR} = \frac{\frac{dn}{dQ}}{\frac{dR}{dQ}} = \frac{C_4}{Q + Q_0} \int_{P = |Q|}^{P = \infty} P^2 e^{-\frac{P^2}{2MT}} dP. \tag{6}$$

Fig. 12 shows curves corresponding to T = 1 MeV and T = 2 MeV, respectively. The experimental points show a thin target distribution in H_2; the arrows on the points on the left side of the peak indicate corrections

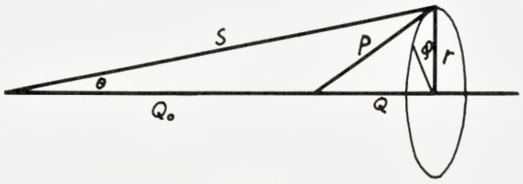

Fig. 11.

for the finite target thickness. As may be seen, the points are not inconsistent with a nuclear temperature between 1 and 2 MeV. In this region of the periodic system, and using α-particles of about 20 MeV, a nuclear temperature of about 1.2 MeV may be expected[9) 10) 11)]. Taking into consideration other contributions to the width (straggling in the target foil, finite breadth of collector foils) it is not surprising that the experimental points seem to indicate a somewhat higher temperature.

The calculated curves show a strong asymmetry. Of course, many effects will tend to remove this, but it is actually found that the experimental curves are also asymmetric, being steeper at the left than at the right side. It may be noted, however, that the points on the calculated curves in half maximum height lie closely symmetrical. The experimental value of R_0 was determined not from the position of the maximum activity, but as the mean of the two

* Since the projection angle θ is only small, it is not of much importance whether we talk about the range itself or its projection.

abscissae corresponding to half maximum intensity. R_0 thus determined is actually the range of Ga^{66} nuclei corresponding to the emission of neutrons with zero momentum in the forward direction, i. e. the range of Ga^{66} nuclei with a velocity equal to that of the compound nuclei. Thus, no correction should be applied for the difference between some of the ranges and their projections.

From Table 6 it may be seen that the width of the range distribution is

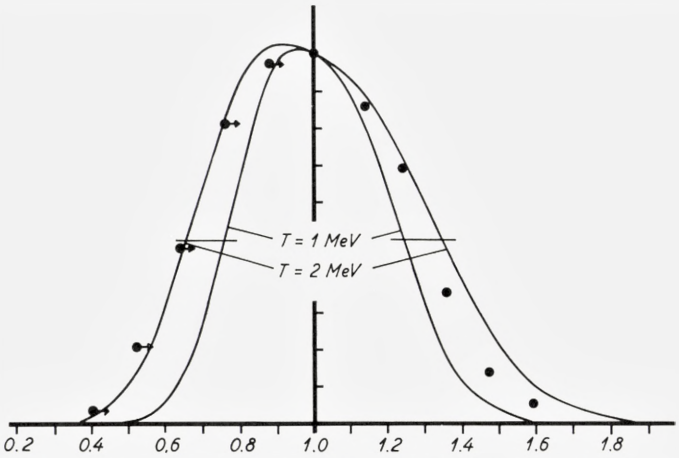

Fig. 12. Calculated curves showing the range distributions corresponding to temperatures of the Ga^{66} nucleus of 1 and 2 MeV, respectively. The points are an experimental distribution obtained in H_2 using α-particles with 19.6 MeV. Abscissa: range in units of R_0, the latter being the range corresponding to emission of a neutron with zero velocity component in the forward direction. Ordinate: relative number of particles per unit interval of range.

almost independent of the α-energy. This indicates that the nuclear temperature decreases with decreasing α-energy, a result which is in conformity with earlier observations[10, 11].

On the same assumptions (I) and (II), the angular distribution of the Ga^{66} ions is given by

$$\frac{dn}{d\theta} = C \sin\theta \cos\theta \int_{x_1}^{\infty} x^2 e^{-x^3} dx \qquad (7)$$

where $x_1 = 2\sqrt{\frac{E_\alpha}{T}}$ and where C is a constant.

In Fig. 5 (p. 12) curves a, b, and c correspond to T = 1.6, 1.8, and 2.0 MeV, respectively, and the experimental points agree fairly well. However, the compound scattering in the target foil may be responsible for a very

considerable part of the angular width. Due to the circular geometry contributions from α-beam divergence ($< \pm 0.^{\circ}5$) and finite target diameter are small.

These experiments were carried out at the Institute for Theoretical Physics, University of Copenhagen. It is a great pleasure to express our heartiest and most deepfelt gratitude to the Director of the Institute, Professor NIELS BOHR. For valuable discussions our thanks are due the late Dr. M. SCHARFF. We thank Mr. A. HEDEGAARD for preparing the target layers, Mr. PH. DAM for operating the cyclotron, Mr. FL. DALL and Mr. CLIVE LARSEN for help in the counting and the numerical calculations, and Mr. H. CHRISTENSEN for manifold and valuable help.

References

1. Niels Bohr and J. Lindhard: Mat. Fys. Medd. Dan. Vid. Selsk. **28**, no. 7 (1954).
2. Niels Bohr: Mat. Fys. Medd. Dan. Vid. Selsk. **18**, no. 8 (1948).
3. N. O. Lassen: Mat. Fys. Medd. Dan. Vid. Selsk. **26**, no. 5 (1950).
4. N. O. Lassen: Mat. Fys. Medd. Dan. Vid. Selsk. **25**, no. 11 (1949).
5. J. K. Bøggild, H. Arrøe, and Th. Sigurgeirsson: Phys. Rev. **71**, 281 (1947).
6. Bernard G. Harvey, Paul F. Donovan, John R. Morton, and Ernest W. Valyocsik: UCRL. 8618, p. 17 (1959).
7. Tyndall and Powell: Proc. Roy. Soc. London. A. **136**, 145 (1932).
8. J. Lindhard and M. Scharff: Phys. Rev. **124**, 128, (1961), and paper to be published.
9. H. W. Fulbright, N. O. Lassen, and N. O. Roy Poulsen: Mat. Fys. Medd. Dan. Vid. Selsk. **31**, no. 10 (1959).
10. N. O. Lassen and N. O. Roy Poulsen: Comptes Rendus du Congrès International, Paris 1958, p. 477 (Dunod, Paris 1959).
11. N. O. Lassen and V. A. Sidorov: Nucl. Phys. **19**, 579 (1960).

Indleveret til selskabet den 20. oktober 1961.
Færdig fra trykkeriet den 23. februar 1962.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 9

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 9 (1962)

ON THE FORMULATION OF THE DYNAMICAL LAWS IN THE QUANTUM THEORY OF FIELDS

BY

P. KRISTENSEN

København 1962
i kommission hos Ejnar Munksgaard

CONTENTS

Synopsis

A new formulation of the dynamical laws for a system of elementary particles is proposed. In addition to simple assumptions of a kinematical nature, the formulation rests on the principle that a classical field and classical sources which, according to the classical theory, describe the same physical situation, also do so when the classical system is coupled to a quantum field. For the simple example of the Hurst-Thirring field, it is shown that this principle may be formulated in finite mathematical terms and may serve as a substitute for the formal field equation of the renormalization theory. To the third order in the coupling constant—and presumably to all orders—the perturbation expansion gives the same result as the usual theory.

No infinities or similar mathematical ambiguities appear in the theory.

Printed in Denmark
Bianco Lunos Bogtrykkeri A-S

1. Introduction

The study of the fundamental assumptions of the relativistic quantum field theory has to a large extent been concerned with axioms of a general nature, such as f. inst. the axiom of microscopic causality, the asymptotic condition, and the requirement of unitarity. Basic assumptions of this type, valid in general, could be referred to as kinematical assumptions. It is well known how to express the kinematical assumptions either directly in terms of the field operators or in terms of various mathematical quantities closely connected with the field operators. In particular certain distributions, such as the τ-functions and the r-functions, have been studied. In terms of such distributions, one may express the kinematical assumptions in closed form[1a, 1b, 2, 3a, 3b].

Various suggestions for the incorporation into such formulations of the dynamical laws valid for a specific system of interacting elementary particles have been discussed. Thus LEHMANN, SYMANZIK and ZIMMERMANN[1a] have pointed out that the forces between elementary particles may be characterized by means of boundary conditions superimposed on the system of equations for the τ-functions*. Recently, NISHIJIMA[3b] and MURASKIN and NISHIJIMA[4] have proposed to use a postulated dispersion relation in terms of which the boundary conditions may be formulated in a simple manner. It might be true that the forces between elementary particles most conveniently are expressed in terms of boundary conditions imposed on equations of a purely kinematical character. Still, we hardly know the best way of characterizing the forces. It might therefore be of interest to investigate also formulations in which the basic assumptions are directly concerned with the dynamical properties of the system, and in which boundary conditions are used to exclude solutions of irregular behaviour only. So far, no direct formulation of the dynamical laws, such as, for example, an explicit construction in terms of the field operator of the source term in the field equation, could be given. More indirect approaches might therefore be acceptable.

* See also ref. 9.

The formulation investigated in the present work does not utilize field equations but is based, instead, on an assumption which concerns the behaviour of the quantum system when in interaction with a classical system of the same type as the quantum system. The assumption has a direct physical interpretation for the case of quantum electrodynamics. Consider the situation in which photons and electrons interact with an external electromagnetic system. The physical state of the classical electromagnetic system may be described in terms of a classical distribution of current and charge $j_\mu(x)$. In this case, the interaction between the quantum system and the classical system enters into the theory by addition of the classical source $j_\mu(x)$ to the operator source of the photon field. However, this is not the only possibility. According to the classical theory, we might also describe the physical situation of the classical electromagnetic system by the electromagnetic field which, according to the Maxwell equations of the classical theory, is produced by the classical distribution of four current $j_\mu(x)$. If this possibility is chosen, the interaction between the classical system and the quantum system is expressed by an additional term $ie\gamma_\mu A_\mu \psi$ in the field equation for the electron field, ψ being the electron field operator. In quantum electrodynamics, it has always been assumed that these two possibilities give the same physical result.*

We shall consider an assumption of this type as a basic principle of quantum physics. Admittedly, a direct physical interpretation of such a principle is possible only for the case of quantum electrodynamics. However, the principle may be generalized to other cases and formulated as a definite mathematical relation. To make the principle and some of its implications clear we study, in the present work, the simple example of the Hurst-Thirring field, i. e. the quantized version of the classical real field which satisfies the classical field equation

$$(-\Box + m^2)\, A(x) = gA^2(x). \tag{1.1}$$

Let us for the moment apply the formal version of the quantized form of (1.1). If the system interacts with an external field $A(x)$ and an external source $j(x)$, the real quantum field $\boldsymbol{A}(x)$ satisfies the field equation

$$(-\Box + m^2)\, \boldsymbol{A}(x) = \boldsymbol{j}(x) + 2g\, A(x)\, \boldsymbol{A}(x) + j(x), \tag{1.2}$$

where the formal expression for the source operator is

$$\boldsymbol{j}(x) = g\, \boldsymbol{A}^2(x). \tag{1.3}$$

* A proof was given by J. Schwinger, Phys. Rev. **76**, 790 (1949).

To emphasize that the field operator depends on both $A(x)$ and $j(x)$ we use the notation $\boldsymbol{A}[A, j; x]$.

Consider first the situation where $j = o$. If retarded boundary conditions are used we have the integral equation

$$\boldsymbol{A}(x) = \boldsymbol{A}_{in}(x) + \int \Delta_R(x-x') \{ \boldsymbol{j}(x') + 2\, gA(x')\, \boldsymbol{A}(x') \}\, d^4x'. \tag{1.4}$$

The notations used will be found in Appendix A. To give a definite meaning to the dependence of $\boldsymbol{A}$ on the physical state of the external system, we choose a representation of the free field operator $\boldsymbol{A}_{in}$ in which this operator is independent of the external field and source. If the formal expression (1.3) for the source operator is used, one finds that the operator

$$\hat{\boldsymbol{A}}(x) = A(x) + \boldsymbol{A}[A, o; x]$$

satisfies the integral equation

$$\hat{\boldsymbol{A}}(x) = \boldsymbol{A}_{in}(x) + \int \Delta_R(x-x')\, g\hat{\boldsymbol{A}}^2(x')\, d^4x' + A(x) - \int \Delta_R(x-x')\, gA^2(x')\, d^4x'.$$

Thus, if we define A_{in} and j by the classical field equation

$$A(x) = A_{in}(x) + \int \Delta_R(x-x') \{ gA^2(x') + j(x') \}\, d^4x', \tag{1.5}$$

where

$$(-\Box + m^2) A_{in}(x) = o,$$

we have

$$\hat{\boldsymbol{A}}(x) = \boldsymbol{A}_{in}(x) + A_{in}(x) + \int \Delta_R(x-x') \{ g\hat{\boldsymbol{A}}^2(x') + j(x') \}\, d^4x'.$$

In order to remove the classical radiation field A_{in}, we apply the time-independent unitary transformation*

$$\boldsymbol{U}[A_{in}] = exp\Big(-i \int \boldsymbol{A}_{in} \overleftrightarrow{\frac{\partial}{\partial y_o}} A_{in}(y)\, d^3\vec{y}\Big). \tag{1.6}$$

If one observes that

$$\boldsymbol{U}^{\dagger}[A_{in}]\, \boldsymbol{A}_{in}(x)\, \boldsymbol{U}[A_{in}] = \boldsymbol{A}_{in}(x) - A_{in}(x),$$

it is easily seen from the equation for $\hat{\boldsymbol{A}}$ that $\boldsymbol{U}^{\dagger}[A_{in}]\, \hat{\mathbf{A}}(x)\, \boldsymbol{U}[A_{in}]$ satisfies the integral equation which determines $\boldsymbol{A}[o, j; x]$. Thus, we have the relation

$$\boldsymbol{A}[o, j; x] = A(x) + \boldsymbol{U}^{\dagger}[A_{in}]\, \boldsymbol{A}[A, o; x]\, \boldsymbol{U}[A_{in}], \tag{1.7}$$

where the sources j and A_{in} are connected with the field A by the classical field equation (1.5) for the Hurst-Thirring field.

* We employ the notation: $f(x) \overleftrightarrow{\frac{\partial}{\partial x_o}} g(x) = f(x) \frac{\partial g(x)}{\partial x_o} - \frac{\partial f(x)}{\partial x_o} g(x)$.

The equation (1.7) is the mathematical expression for the assumption that sources j and A_{in} and a field A, which describe the same physical situation, according to the classical theory also do so when the classical system is coupled to a quantum system.

In a similar manner, one derives the more useful equation

$$\boldsymbol{A}[A, j+\delta j; x] = \delta A(x) + \boldsymbol{U}^{\dagger}[\delta A_{in}]\, \boldsymbol{A}[A+\delta A, j; x]\, \boldsymbol{U}[\delta A_{in}], \tag{1.8}$$

where the connection between the source variations δj, δA_{in} and the variation of the field is given by the varied form of the classical field equation, i. e.,

$$\left.\begin{aligned} \delta A(x) = \delta A_{in}(x) + \int \Delta_R(x-x') \{ & g(\delta A(x'))^2 \\ & + 2g A(x')\, \delta A(x') + \delta j(x') \}\, d^4x'. \end{aligned}\right\} \tag{1.9}$$

The proof given of (1.8), (1.9) is completely formal of course. However, these relations are in themselves meaningful mathematical expressions and we may assume that eqs. (1.8) and (1.9) are valid quite apart from the proof given. The formal derivation makes it plausible that, by such an assumption, essential characteristics of the dynamics are introduced in the theory. In fact, similar formal calculations with another expression for the operator source lead to a completely different result. This is also indicated by the fact that the characteristic non-linearity of the Hurst-Thirring field appears explicitly in equation (1.9).

We shall take (1.8), (1.9) as a basic assumption of the theory. It will be shown that such a postulate may be utilized in very much the same way as the formal field equation of the renormalization theory. The advantage gained is of course that we may maintain the attitude of ordinary mathematics that divergent quantities are allowed neither in the fundamental equations nor in any intermediate step of the calculations.

In paragraph 2, a list is given of the assumptions on which we propose to build a consistent formulation of the quantum theory of the Hurst-Thirring field. It will appear that (1.8) and (1.9), which we shall refer to as the variational equations for the field operator, are not totally of a dynamical nature. In fact, as a special result, we obtain from the variational equation for the field operator the reduction formulae of Lehmann, Symanzik and Zimmermann[1b] and of Nishijima[3a]. Thus, the asymptotic conditions are superfluous. The boundary conditions which are necessary to avoid solutions of irregular behaviour are discussed in paragraph 3. It is found that these boundary conditions may be formulated as a principle of maximum regularity of certain distributions, called π_R-functions, related to the source operator. The perturbation expansion is studied in paragraph 4,

where it is shown that the theory gives unambiguous answers to the third order (and presumably to all orders) in the coupling constant. The results found for the source operator agree with those of the renormalization theory.

The present investigation is of a rather preliminary nature. Several important problems have not been solved. Presumably, the validity of the variational equation for the field operator may be "proved" in the framework of the renormalization theory. From the mathematical point of view such a proof would be as formal as that given here. Still, a proof should be given in order to ensure that the correct results of the renormalization theory are reproduced to all orders in the coupling constant in the present formulation. This question has not yet been considered. The assumption of unitarity is not needed for the unique characterization of the theory. Ultimately we shall therefore be faced with the problem to prove the existence of a scattering matrix. This question has not been considered either. Only the Hurst-Thirring field has been studied and it is well known that this theory is not quite typical in several respects.

The mathematical techniques used are presented in the usual language of mathematical physics. Thus, the technical language of modern distribution theory is avoided, although a certain not too low standard of mathematical rigour should be maintained as regards questions of distribution theory. In other respects we benefit from the advantages of a purely formal approach, in particular with regard to topological questions in the underlying Hilbert space. For the purpose of the present investigation this is not dangerous. In fact it is easily seen that the situation may be remedied by a strict adherence to the weak topology, i. e. all definitions and calculations may be interpreted as relations between definite matrix elements in the Hilbert space. However, whether the weak topology is the appropriate one for a more thorough study of the theory is an open question.

2. The basic equations of the theory

To give a precise formulation of the variational equation for the field operator we need the connection between the quantum field $\boldsymbol{A}(x)$ and the incoming field $\boldsymbol{A}_{in}(x)$. We assume that $\boldsymbol{A}(x)$ is so regular that

$$\int \Delta_R(x-x')(-\Box'+m^2)\,\boldsymbol{A}(x')\,d^4x' \tag{2.1}$$

exists as a convolution integral, and that the operator

$$\boldsymbol{A}_{in}(x) = \boldsymbol{A}(x) - \int \Delta_R(x-x')(-\Box'+m^2)\,\boldsymbol{A}(x')\,d^4x' \tag{2.2}$$

is independent of the external fields and sources, and is quantized in the

usual way.* These assumptions involve two kinematical postulates, one concerning the distribution character of $\boldsymbol{A}(x)$, which serves to guarantee the existence of $\boldsymbol{A}_{in}(x)$, and a quantum rule which characterizes this operator.

For the special case of infinitesimal variations the fundamental dynamical assumption may now, in accordance with (1.8) and (1.9), be given in the following form:

$$\left.\begin{aligned}\int \frac{\delta \boldsymbol{A}(x)}{\delta j(y)}\,\delta j(y)\,d^4y = \delta A(x) + \int \frac{\delta \boldsymbol{A}(x)}{\delta A(y)}\,\delta A(y)\,d^4y \\ - i\int [\boldsymbol{A}(x), \boldsymbol{A}_{in}(y)]\,\frac{\overleftrightarrow{\partial}}{\partial y_0}\,\delta A_{in}(y)\,d^3\vec{y},\end{aligned}\right\} \tag{2.3}$$

where

$$\delta j(x) = (-\Box + m^2 - 2\,g\,A(x))\,\delta A(x), \tag{2.4}$$

and

$$\delta A_{in}(x) = \delta A(x) - \int \Delta_R(x-x')\,(-\Box' + m^2)\,\delta A(x')\,d^4x'. \tag{2.5}$$

A possible definition of the Volterra derivative is given in the Appendix B. As already mentioned in the Introduction, this postulate allows a derivation of the reduction formula.

As a final kinematical assumption we take the variational equation†, ††

$$\frac{\delta \boldsymbol{A}(x)}{\delta j(y)} = i\,\vartheta(x-y)\,[\boldsymbol{A}(x), \boldsymbol{A}(y)], \tag{2.6}$$

first proved by PEIERLS[5]. As is well known, this equation holds in the renormalized theory.

It will be seen in the next paragraph that the variational equations (2.3) and (2.6) have more than one solution. These equations should therefore be supplemented with subsidiary conditions, which excludes solutions of too irregular a behaviour. One such condition is the requirement of relativistic invariance. In the following it should be understood that only relativistically invariant solutions are admitted. The non-trivial question of the necessary boundary conditions will be discussed in the next paragraph.

Before taking up the discussion of the boundary conditions we derive

* Cf. Appendix A.

† In the present work only the vacuum expectation value of Peierls' variational equation is used.

†† For the Hurst-Thirring model it may be assumed that the commutator of two field operators is so regular that the right-hand side of (2.6) exists as a limit of $i\,\vartheta_\tau(x-y)\,[\boldsymbol{A}(x), \boldsymbol{A}(y)]$ where $\vartheta_\tau(x)$ is a sequence of testing functions which for $\tau \to 0$ converges (in the topology of the space of distributions) to the distribution $\vartheta(x)$. In practice this means that the retarded commutator may be treated as an ordinary product. Such a regularity assumption is not possible for the commutator of two source operators.

some direct consequences of the basic equations. First, we show that the reduction formula follows directly from the variational equation for the field operator.

Derivation of the reduction formula.

To derive the reduction formula we use twice the variational equation for the field operator. In the case of $\delta A_{in}(x) = o$, we find from (2.3)*

$$(K_y - 2gA(y))\frac{\delta \boldsymbol{A}(x)}{\delta j(y)} = \frac{\delta \boldsymbol{A}(x)}{\delta A(y)} + \delta(x-y), \tag{2.7}$$

while for $\delta j(x) = o$ the result is

$$o = \delta A(x) + \int \frac{\delta \boldsymbol{A}(x)}{\delta A(y)} \delta A(y)\, d^4y - i\int [\boldsymbol{A}(x),\ \boldsymbol{A}_{in}(y)] \frac{\overleftrightarrow{\partial}}{\partial y_0} \delta A_{in}(y)\, d^3\vec{y}, \tag{2.8}$$

where

$$\left.\begin{aligned} &\delta A(x) = \delta A_{in}(x) + \Delta_R\, 2gA(x)\,\delta A(x), \\ &K_x\, \delta A_{in}(x) = o. \end{aligned}\right\} \tag{2.9}$$

The conditions (2.9) should not be ignored in the derivation of the reduction formula, as these conditions severely limit the domain of the variations $\delta A(x)$. Thus, in the formula which results from (2.7) and (2.8),

$$\left.\begin{aligned} &\int \delta A(y)\,(K_y - 2gA(y))\frac{\delta \boldsymbol{A}(x)}{\delta j(y)}\, d^4y = \\ &+ i\int [\boldsymbol{A}(x),\ \boldsymbol{A}_{in}(y)] \frac{\overleftrightarrow{\partial}}{\partial y_o} \delta A_{in}(y)\, d^3\vec{y}, \end{aligned}\right\} \tag{2.10}$$

an integration by parts is not permitted. Instead, we use

$$K_y\, \delta A(y) = 2gA(y)\,\delta A(y), \tag{2.11}$$

and find

$$\left.\begin{aligned} &\int \left\{ \delta A(y)\, K_y \frac{\delta \boldsymbol{A}(x)}{\delta j(y)} - \frac{\delta \boldsymbol{A}(x)}{\delta j(y)} K_y\, \delta A(y) \right\} d^4y = \\ &+ i\int [\boldsymbol{A}(x),\ \boldsymbol{A}_{in}(y)] \frac{\overleftrightarrow{\partial}}{\partial y_o} \delta A_{in}(y)\, d^3\vec{y}. \end{aligned}\right\} \tag{2.12}$$

* In the following we use for convenience the notations

$$K_x f(x) = (-\Box + m^2) f(x), \text{ and } \Delta_R f(x) = \int \Delta_R(x-x') f(x')\, d^4x'.$$

Due to the retarded character of $\delta \boldsymbol{A}(x)/\delta j(y)$ we have

$$\int \Delta_R(y-z)\, K_y \frac{\delta \boldsymbol{A}(x)}{\delta j(y)}\, d^4 y = \frac{\delta \boldsymbol{A}(x)}{\delta j(z)}. \tag{2.13}$$

In fact, the difference between these two expressions vanishes for $x_o < z_o$ and satisfies as a function of z the homogeneous wave equation. If, further, the relation

$$\Delta_R\, K_x\, \delta A(x) = \delta A(x) - \delta A_{in}(x) \tag{2.14}$$

is taken into account, we find from (2.12)

$$\int \delta A_{in}(y)\, K_y \frac{\delta \boldsymbol{A}(x)}{\delta j(y)}\, d^4 y = +i \int [\boldsymbol{A}(x),\ \boldsymbol{A}_{in}(y)] \overleftrightarrow{\frac{\partial}{\partial y_0}} \delta A_{in}(y)\, d^3 \vec{y}. \tag{2.15}$$

With the aid of the well-known solution of the initial value problem of the wave equation

$$\delta A_{in}(y) = -\int \Delta(y-z) \overleftrightarrow{\frac{\partial}{\partial z_o}} \delta A_{in}(z)\, d^3 \vec{z}, \tag{2.16}$$

we find from (2.15) the reduction formula for the field operator

$$[\boldsymbol{A}(x),\ \boldsymbol{A}_{in}(y)] = -i \int \Delta(y-z)\, K_z \frac{\delta \boldsymbol{A}(x)}{\delta j(z)}\, d^4 z, \tag{2.17}$$

which alternatively, due to (2.6), may be written in the usual form

$$[\boldsymbol{A}(x),\ \boldsymbol{A}_{in}(y)] = \int \Delta(y-z)\, K_z\, \vartheta(x-z)\, [\boldsymbol{A}(x),\ \boldsymbol{A}(z)]\, d^4 z. \tag{2.18}$$

Thus, the reduction formula gets a heuristic motivation in the formulation studied here. Further, it might be remarked that in the derivation we have not made use of asymptotic formulae, which in fact do not form a part of the basic assumptions of the theory.

For the discussion of the contents of the formulation proposed here, we found it convenient to work with the source operator instead of the field operator itself. For the Hurst-Thirring field, in the presence of an external field and external sources, the source operator $\boldsymbol{j}(x)$ is most conveniently defined by the equation

$$(-\Box + m^2)\, \boldsymbol{A}(x) = \boldsymbol{j}(x) + 2g A(x)\, \boldsymbol{A}(x) + j(x), \tag{2.19}$$

whence by (2.2)

$$\boldsymbol{A}(x) = \boldsymbol{A}_{in}(x) + \int \Delta_R(x-x') \{ \boldsymbol{j}(x') + 2g A(x')\, \boldsymbol{A}(x') + j(x') \}\, d^4 x'. \tag{2.20}$$

We shall refer to the equation (2.20) as the field equation.

In terms of the source operator, the variational equation (2.7) takes the form of the variational equation for the source operator, viz.

$$[K_y - 2gA(y)]\frac{\delta \boldsymbol{j}(x)}{\delta j(y)} = \frac{\delta \boldsymbol{j}(x)}{\delta A(y)} + 2g\delta(x-y)\boldsymbol{A}(x), \tag{2.21}$$

while the reduction formula (2.17) yields the reduction formula for the source operator

$$[\boldsymbol{j}(x),\ \boldsymbol{A}_{in}(y)] = -i\int \Delta(y-z) K_z \frac{\delta \boldsymbol{j}(x)}{\delta j(z)}\, d^4 z. \tag{2.22}$$

Explicit expressions in terms of the field operator for the variational derivatives of $\boldsymbol{j}(x)$ and for $\delta \boldsymbol{A}(x)/\delta A(y)$ may be found from (2.6), (2.7) and the field equation (2.19). As an example, we quote the formula

$$\left.\begin{aligned}\frac{\delta \boldsymbol{j}(x)}{\delta A(y)} = {} & [K_x - 2gA(x)][K_y - 2gA(y)]\, i\vartheta(x-y)[\boldsymbol{A}(x),\ \boldsymbol{A}(y)] \\ & - [K_x - 2gA(x)]\delta(x-y) - 2g\delta(x-y)\boldsymbol{A}(x).\end{aligned}\right\} \tag{2.23}$$

The coupling constant appears explicitly here, where it plays a role in the characterization of the singularity at $x = y$. Outside the singularity the expression simplifies to

$$\frac{\delta \boldsymbol{j}(x)}{\delta A(y)} = i\vartheta(x-y)[\boldsymbol{j}(x),\ \boldsymbol{j}(y)], \quad \text{for } x_o \neq y_o. \tag{2.24}$$

This expression is well known from the formal canonical theory[6a, b], where the expression is assumed to cover the singularity for $x = y$ as well. It may easily be seen that the extrapolation of (2.24) and the corresponding expressions for $\delta \boldsymbol{j}/\delta j$ and $\delta \boldsymbol{A}/\delta A$ to all values of $x-y$ give the correct result if the commutation relations between $\boldsymbol{A}$, $\dot{\boldsymbol{A}}$ and $\boldsymbol{j}$ of the formal canonical theory are valid.

We see from (2.22) that, if the operator

$$\boldsymbol{j}(x;y) = K_y \frac{\delta \boldsymbol{j}(x)}{\delta j(y)} \tag{2.25}$$

is expanded in the series*

$$\left.\begin{aligned}\boldsymbol{j}(x;y) = {} & f(x;y) + \int f(x;y,\ 1)\boldsymbol{A}_{in}(1)\ d(1) \\ & + \frac{1}{2!}\int f(x;y,\ 1,\ 2) : \boldsymbol{A}_{in}(1)\boldsymbol{A}_{in}(2) : d(12) + \ldots,\end{aligned}\right\} \tag{2.26}$$

* Haag[7]. In the absence of bound states, the expansion functions f are c-numbers.

where :. . . : denotes the Wick product (se Appendix A), we have for the source operator

$$\left.\begin{aligned} \boldsymbol{j}(x) = f(x) &+ \int f(x;1)\,\boldsymbol{A}_{in}(1)\,d(1) \\ &+ \frac{1}{2!}\int f(x;1,\,2):\boldsymbol{A}_{in}(1)\,\boldsymbol{A}_{in}(2):\,d(12) \\ &+ \frac{1}{3!}\int f(x;1,\,2,\,3):\boldsymbol{A}_{in}(1)\,\boldsymbol{A}_{in}(2)\,\boldsymbol{A}_{in}(3):\,d(123) + \ldots . \end{aligned}\right\} \tag{2.27}$$

Here, $f(x) = \langle 0 \,|\, \boldsymbol{j}(x) \,|\, 0 \rangle$ is not yet determined, but may be assumed to be subject to the boundary condition

$$\langle 0 \,|\, \boldsymbol{j}(x) \,|\, 0 \rangle = f(x) = o, \quad \text{for } A = j = o. \tag{2.28}$$

We close this paragraph by a few comments on the formal theory which is obtained if (2.24) is extrapolated in a naive fashion to all values of x and y. In this formal theory the system of basic equations is easily seen to be complete. We take $j = A = o$ and find, by (2.21) and the extrapolated form of (2.24),

$$\boldsymbol{j}(x;y) = i\vartheta(x-y)\,[\boldsymbol{j}(x),\ \boldsymbol{j}(y)] + 2g\,\delta(x-y)\,\boldsymbol{A}(x), \quad \text{(wrong)}$$

where

$$\boldsymbol{A}(x) = \boldsymbol{A}_{in}(x) + \Delta_R\,\boldsymbol{j}(x).$$

These two equations determine the source operator anyhow if perturbation theory applies. To the lowest order in the coupling constant we find from (wrong)

$$\boldsymbol{j}(x;y) = 2g\,\delta(x-y)\,\boldsymbol{A}_{in}(x),$$

whence by (2.27) and (2.28) we find for the lowest order term in the source operator

$$\boldsymbol{j}(x) = g:\boldsymbol{A}_{in}^2(x):\,.$$

When this expression is inserted into the right-hand side of the equation (wrong) we find to the second order in g

$$\begin{aligned} \boldsymbol{j}(x;y) = i\vartheta(x-y)\,[&g:\boldsymbol{A}_{in}^2(x):,\ \ g:\boldsymbol{A}_{in}^2(y):] \\ &+ 2g\,\delta(x-y)\,\boldsymbol{A}_{in}(x) + 2g\,\delta(x-y)\,\Delta_R\,g:\boldsymbol{A}_{in}^2(x):\,. \end{aligned}$$

Proceeding in this manner we obtain the perturbation theory. However, already the second order expression demonstrates that the naive extrapolation of (2.24) is indeed not possible. The vacuum expectation value of the first term on the right-hand side of the above expression leads to the

well-known divergent expression for the self-mass of the meson and to a wave function renormalization, this being finite in the present model.

Still, the formal approach is not without interest. It shows that the axioms we have chosen are as complete as the usual axiomatic foundation of the formal canonical theory. It also demonstrates that, in order to have a complete dynamical theory, we must find the correct solution to (2.24), regarded as an equation for $\delta \boldsymbol{j}(x) / \delta A(y)$.

A direct approach to the multiplication problem (2.24) has not been found. The complete mathematical solution to the equation (2.24) involves an arbitrary distribution wich vanishes outside the subspace $x = y$, and what in particular complicates matters is, that this distribution is operator valued. We have instead chosen a more indirect method of investigation. In this method the arbitrary distribution is c-number valued and the correct solution may easily be characterized. The drawback of the method is that the external field has to be kept finite until the end of the calculations. This complicates somewhat the algebraic part of the calculations.

3. Discussion of the boundary conditions

In the following we take the external source equal to zero. Due to

$$\frac{\delta \boldsymbol{j}(x)}{\delta j(y)} = \int \Delta_R(z-y)\, K_z \frac{\delta \boldsymbol{j}(x)}{\delta j(z)}\, d^4 z, \tag{3.1}$$

we have, by (2.21) and (2.25),

$$[K_y - 2B(y)] \int \boldsymbol{j}(x;1)\, \Delta_R(1-y)\, d(1) = g \frac{\delta \boldsymbol{j}(x)}{\delta B(y)} + 2g\delta(x-y)\, \boldsymbol{A}(x), \tag{3.2}$$

where we have employed the notation

$$B(x) = gA(x).$$

Obviously only this combination is relevant for the problem. The discussion in the present paragraph as well as the explicit calculations in the next paragraph will be based on the equation (3.2), the reduction formula, which will be used in the form of the connection between the Haag series for $\boldsymbol{j}(x; y)$ and $\boldsymbol{j}(x)$ given by (2.26) and (2.27), and the equations

$$\left.\begin{aligned} g \frac{\delta \boldsymbol{j}(x)}{\delta B(y)} - g \frac{\delta \boldsymbol{j}(y)}{\delta B(x)} &= i[\boldsymbol{j}(x), \boldsymbol{j}(y)], \\ g \frac{\delta \boldsymbol{j}(x)}{\delta B(y)} &= o, \quad \text{for} \quad y_o > x_o, \end{aligned}\right\} \tag{3.3}$$

which are a direct consequence of the explicit expression (2.23). Actually, as will become clear, we need (3.3) only in the vacuum subspace of the Hilbert space.

For $j(x) = o$ the field equation (2.20) reads

$$\boldsymbol{A}(x) = \boldsymbol{A}_{in}(x) + \Delta_R(\boldsymbol{j}(x) + 2B(x)\boldsymbol{A}(x)). \tag{3.4}$$

Consider for the moment the perturbation solution. Here we regard $\boldsymbol{j}$ as of at least the order g. For $g = o$ we have by the field equation

$$\boldsymbol{A}^{(o)}(x) = \boldsymbol{A}_{in}(x) + \Delta_R 2B(x)\boldsymbol{A}^{(o)}(x).$$

By the solution of this equation we find $2g\delta(x-y)\boldsymbol{A}^{(o)}(x)$, i. e. the right-hand side of (3.2), to the first order in g. Hence we may calculate $\boldsymbol{j}(x;y)$ to the first order in g. To discuss in general terms how to proceed, assume that $\boldsymbol{j}(x)$ is known to the order g^n as a functional of B. The right-hand side of (3.2) and thus $\boldsymbol{j}(x;y)$ may then be calculated to the order g^{n+1}. Hence by the reduction formula, or more directly by (2.26), (2.27), we find $\boldsymbol{j}(x) - \langle 0|\boldsymbol{j}(x)|0\rangle$ to the order g^{n+1}. To be able to proceed in the iteration procedure we need $\langle 0|\boldsymbol{j}(x)|0\rangle$ to the order g^{n+1} or alternatively $g\delta\langle 0|\boldsymbol{j}(x)|0\rangle/\delta B(y)$ to the order g^{n+2}. This is the point where the vacuum expectation value of the system (3.3) comes into play. By the knowledge of $\boldsymbol{j}(x) - \langle 0|\boldsymbol{j}(x)|0\rangle$ to the order g^{n+1} we can calculate

$$C(x;y) = -C(y;x) = i\langle 0|[\boldsymbol{j}(x),\ \boldsymbol{j}(y)]|0\rangle \tag{3.5}$$

to the order g^{n+2}. Thus, what we need is to solve the system of equations

$$\left.\begin{aligned} g\frac{\delta\langle 0|\boldsymbol{j}(x)|0\rangle}{\delta B(y)} - g\frac{\delta\langle 0|\boldsymbol{j}(y)|0\rangle}{\delta B(x)} &= C(x;y), \\ g\frac{\delta\langle 0|\boldsymbol{j}(x)|0\rangle}{\delta B(y)} &= o, \quad \text{for } y_o > x_o, \end{aligned}\right\} \tag{3.6}$$

where the functional $C(x;y)$ is known. Observe that C is real valued.

We shall take the perturbation argument as an indication of the fact that, if we can characterize that solution of the system (3.6) which should be used in physics, we have a well-defined formalism for the Hurst-Thirring field. We therefore proceed to discuss the system (3.6) and from now on drop the assumption of the perturbation expansion. Thus we have converted the problem of the $i\vartheta(x-y)[\boldsymbol{j}(x), \boldsymbol{j}(y)]$ multiplication into a similar, but simpler, problem where only c-number valued distributions are involved.

The complete solution of the system (3.6) consists of a particular solution added to the complete solution of the corresponding homogeneous system, viz.

$$\left.\begin{aligned} &\frac{\delta l(x)}{\delta B(y)} - \frac{\delta l(y)}{\delta B(x)} = o, \\ &\frac{\delta l(x)}{\delta B(y)} = o, \quad \text{for} \quad x \neq y. \end{aligned}\right\} \tag{3.7}$$

Due to the requirement of relativistic invariance we need only discuss the relativistically invariant solutions of the homogeneous equations. Hence we have replaced the condition $x_0 > y_0$, which for the homogeneous system is extended to $x_0 \neq y_0$, by the condition $x \neq y$.

A functional of a function $B(x)$ determines an infinite set of distributions which we take as the expansion coefficients of the formal Volterra series. For $l(x)$ we denote these distributions by

$$l_{4n}(x-y_1,\ x-y_2, \ldots,\ x-y_n) = \delta^n l(x) / \delta B(y_1)\, \delta B(y_2) \ldots \delta B(y_n)\big|_{B=o}, \tag{3.8}$$

where we have used a notation which reflects the invariance of l under displacements in space-time. For the discussion of the equations (3.7) we found it necessary to restrict the domain of solutions to functionals analytical in the sense that they are determined uniquely by the set of expansion coefficients of the formal Volterra series. It need not be assumed that the formal Volterra series is convergent. To indicate the one-to-one correspondence between the set of distributions l_{4n} and the functional l, we write*

$$l(x) \approx \sum_{n=1}^{\infty} \frac{1}{n!} \int l_{4n}(x-y_1, \ldots, x-y_n)\, B(y_1) \ldots B(y_n)\, d^4y_1 \ldots d^4y_n. \tag{3.9}$$

The distributions $l_{4n}(z_1, z_2, \ldots z_n)$ are of course symmetric in $z_1, z_2, \ldots, z_n$ and are invariant under the homogeneous Lorentz group. By the second equation (3.7), $l_{4n}(x-y, x-y_2, \ldots, x-y_n)$ vanishes outside the subspace $x = y$, and hence, by the symmetry, $l_{4n}(z_1, z_2, \ldots, z_n)$ vanishes outside the intersection of the subspaces $z_\nu = o$, $\nu = 1, 2, \ldots n$, i. e. except at the single point $z_1 = z_2 = \ldots = z_n = o$. Thus, by a well-known theorem in the theory of distributions, $l_{4n}(z_1, z_2, \ldots, z_n)$ is a finite linear combination of $\delta(z_1)\delta(z_2)\ldots\delta(z_n)$ and its derivatives, viz.

$$l_{4n}(z_1, z_2, \ldots, z_n) = P\left(\frac{\partial}{\partial z_1}, \frac{\partial}{\partial z_2}, \ldots, \frac{\partial}{\partial z_n}\right) \delta(z_1)\, \delta(z_2) \ldots \delta(z_n). \tag{3.10}$$

Here, P is a symmetric relativistically invariant polynomial.

* The summation starts at $n = 1$ due to $\langle 0 \,|\, \boldsymbol{j}(x) \,|\, 0 \rangle\big|_{B=o} = o$.

We have not yet taken the first of the equations (3.7) into account. This equation resticts the distribution l_{4n} by the condition that $l_{4n}(x_1-x_2,\ x_1-x_3,\ \ldots,\ x_1-x_{n+1})$ is symmetric in $x_1, x_2, \ldots,\ x_{n+1}$ as well. Hence, $l(x)$ may be represented in the form

$$l(x) = \frac{\delta \Theta[B]}{\delta B(x)}, \tag{3.11}$$

where Θ is associated with the formal Volterra series

$$\left.\begin{aligned} \Theta \approx \sum_{n=2}^{\infty} \frac{1}{n!} \int l_{4(n-1)}(x_1-x_2, x_1-x_3, \ldots, x_1-x_n) B(x_1) B(x_2) \ldots B(x_n) \\ \cdot d^4x_1\, d^4x_2 \ldots d^4x_n. \end{aligned}\right\} \tag{3.12}$$

Thus, due to (3.10), we have

$$\Theta[B] = \int \mathfrak{L}(B(x),\ \partial_\mu B(x),\ \partial_\mu \partial_\nu B(x), \ldots)\, d^4x, \tag{3.13}$$

where the density function $\mathfrak{L}$ is an ordinary function of $B(x)$ and the derivatives of $B(x)$*. The function $\mathfrak{L}$ may involve derivatives of $B(x)$ of arbitrarily high order, but should be in accordance with (3.10) and (3.12).

The results may conveniently be expressed in terms of the Fourier transform of the distributions. We define

$$\left.\begin{aligned} l_{4n}(z_1, z_2, \ldots, z_n) = \frac{1}{(2\pi)^{4n}} \int \mathfrak{P}(q_1, q_1, \ldots, q_n) \\ \times \exp(-i q_1 z_1 - i q_2 z_2 - \ldots - i q_n z_n)\, d^4 q_1\, d^4 q_2 \ldots d^4 q_n, \end{aligned}\right\} \tag{3.14}$$

and have the result: $\mathfrak{P}(q_1, q_2, \ldots, q_n)$ is a symmetric and Lorentz invariant polynomium, i. e.

* An interpretation of $\Theta[B]$ may be given in the following manner. It may be shown that the first variational derivative of a scattering operator for the system is given by

$$-i \frac{\delta \boldsymbol{S}}{\delta A(x)} = \boldsymbol{S}\boldsymbol{j}(x).$$

Here the operator $\boldsymbol{S}$ is defined up to a phase factor by the equation

$$\boldsymbol{A}_{out}(x) = \boldsymbol{S}^\dagger \boldsymbol{A}_{in}(x)\, \boldsymbol{S},$$

and the conditions of unitarity and causality. It is easily seen that the condition of causality restricts the arbitrary phase factor to the form $\exp i\Theta$, with Θ given by an expression of the form (3.13). If the source operator is known to the order g^n in perturbation theory, we find $\boldsymbol{S}$, apart from such a phase factor, to the order g^n, and hence $\boldsymbol{j}$ to the order g^{n+1} is given by the formula

$$\boldsymbol{j}(x) = -i g \boldsymbol{S}^\dagger \frac{\delta \boldsymbol{S}}{\delta B(x)} + g \frac{\delta \Theta}{\delta B(x)},$$

where Θ is of the type described but otherwise unknown.

Thus the discussion in the following would be made superfluous if a characterization of the c-number phase factor of this scattering operator could be found by other means.

$$\mathfrak{P}(q_1, q_2, \ldots, q_n) = \mathfrak{P}(-q_1^2, -q_2^2, \ldots, -q_n^2, q_1 q_2, q_1 q_3, \ldots, q_{n-1} q_n),$$

and is symmetric in the indices. This is the result which follows from the second of the equations (3.7). The restrictions on $\mathfrak{P}$ which follow from the first of these equations will not be discussed further here.

The boundary conditions.

We now return to the system (3.6). Let $\langle 0 | \boldsymbol{j}(x) | 0 \rangle$ be represented by the formal Volterra series*

$$\left.\begin{aligned} \langle 0 | \boldsymbol{j}(x) | 0 \rangle \approx \sum_{n=1}^{\infty} \frac{1}{n! g^n} \int \pi_R(x-y_1, \ldots, x-y_n) B(y_1) \ldots B(y_n) \\ \cdot d^4 y_1 \ldots d^4 y_n . \end{aligned}\right\} \tag{3.15}$$

It follows from the discussion that, if distributions $\overset{\circ}{\pi}_R(x-y_1, \ldots, x-y_n)$ define a particular solution to the system (3.6), then

$$\pi_R(z_1, \ldots z_n) = \overset{\circ}{\pi}_R(z_1, \ldots, z_n) + P\left(\frac{\partial}{\partial z_1}, \ldots, \frac{\partial}{\partial z_n}\right) \delta(z_1) \ldots \delta(z_n), \tag{3.16}$$

where P is a polynomium of the type described.

Thus the boundary conditions we need are those which may serve to characterize coefficients of the polynomial $P(\partial/\partial z)$.

A simple boundary condition suggests itself at this place. It is tempting to require that the solutions should be regular at the origin, in other words, that there should be no δ-like singularity. Such a requirement gives the same (meaningsless) result as the naive approach to the $i\vartheta \cdot [\boldsymbol{j}, \boldsymbol{j}]$-multiplication problem. The point is that no such solution exists, and an enforcement of such an inconsistent requirement leads to mathematical inconsistencies which manifest themselves in terms of the well-known divergent integrals.

This is illustrated by a consideration of the simple example of the function $\pi_R(x-y)$. For this function we simply have the expression $i\langle 0 | [\boldsymbol{j}(x), \boldsymbol{j}(y)] | 0 \rangle|_{B=o}$, for the right-hand side of the first of the equations (3.6). It is easily seen that this expression may be represented in the form†

$$i\langle 0 | [\boldsymbol{j}(x), \boldsymbol{j}(o)] | 0 \rangle|_{B=o} = -\int \varepsilon(t)\, \delta(x^2+\tau)\, f(\tau)\, d\tau, \tag{3.17}$$

* In the formal theory the interpretation of the functions π_R is easily seen to be

$$\pi_R(x-y_1, \ldots, x-y_n) = \langle 0 | R(\boldsymbol{j}(x); \boldsymbol{j}(y_1) \ldots \boldsymbol{j}(y_n)) | 0 \rangle,$$

where R denotes the retarded product. This formal expression is ambiguous, and not of much use.

† See f. inst. the work of Gårding and Roos, reported in the lecture notes of Gårding and Lions[8]).

where $f(\tau)$ is a distribution with support on the positive part of the real τ-axes, $o \leqslant \tau < \infty$. If we split $\pi_R(x)$ in the symmetric part $\bar{\pi}(x) = \frac{1}{2}(\pi_R(x) + \pi_R(-x))$ and the antisymmetric part $\pi(x) = -\pi_R(x) + \pi_R(-x)$, we have according to (3.6) and (3.18)

$$\pi_R(x) = \bar{\pi}(x) - \frac{1}{2}\pi(x), \tag{3.18}$$

where

$$\pi(x) = \int \varepsilon(t)\,\delta(x^2+\tau)\,f(\tau)\,d\tau. \tag{3.19}$$

For $\bar{\pi}$ we now have the equation

$$\bar{\pi}(x) = -\frac{1}{2}\varepsilon(t)\,\pi(x), \quad \text{for } t \neq o. \tag{3.20}$$

If $f(\tau)$ is assumed to be sufficiently regular at the origin, we have the solution

$$\bar{\pi}(x) = -\frac{1}{2}\int \delta(x^2+\tau)\,f(\tau)\,d\tau. \tag{3.21}$$

However, for simple approximations, f. inst. in second order perturbation theory, we find that $f(\tau)$ is of the general type

$$f(\tau) \approx \delta'(\tau), \tag{3.22}$$

apart from regular terms, where δ' denotes the derivative of Dirac's δ-function. Assume, for the sake of the argument, that (3.22) is correct*. In this case, the expression (3.21) as it stands is without any mathematical meaning. To illustrate this point, we proceed in the calculations with complete disregard of the validity of the formal operations, and find the formal result

$$\bar{\pi}(x) = \frac{1}{2}\delta'(x^2) = \frac{1}{8\,rt}(\delta'(t-r) + \delta'(t+r)).$$

To see whether this expression makes sense we apply it to a testing function $\varphi(r, t)$, and find

$$\int \bar{\pi}(x)\,\varphi(r,t)\,d^4x = -\frac{\pi}{2}\int_0^\infty (\dot{\varphi}(r,r) - \dot{\varphi}(r,-r))\,dr + \frac{\pi}{2}\int_0^\infty \frac{\varphi(r,r) + \varphi(r,-r)}{r}\,dr.$$

* It may be seen that the assumption (3.22) essentially is equivalent to the assumption that the well-known spectral function $\Pi(\varkappa^2)$ in the Källén-Lehmann representation for the vacuum expectation value of the commutator of two source operators behaves like a constant for large values of $\varkappa^2$.

The last term diverges logarithmically at the origin. This of course is a manifistation of the well-known divergent self-energy. Thus the symbol $\delta'(x^2)$ does not represent a distribution.*

In the renormalization theory the situation is saved by an additional term in the solution of the type $\delta m^2 \delta(x)$, where δm^2 is a conveniently chosen logarithmically divergent constant. The renormalization theory gives of course the right result, but the detour over the mathematically undefined (divergent) expressions should be, and in fact is, superfluous.

As the above example shows, in our formulation the inconsistencies originate from mathematically inconsistent regularity assumptions. The mathematical form of (3.16) suggests a formulation of consistent regularity conditions in terms of the notion of the order of a distribution at a point (here the origin). Such a formulation may easily be given and might indeed be the most adequate one for the investigation of the fundamental problems of the theory, such as, for instance, the problem of the existence of rigorous solutions.

However, the concept of the order of a distribution at a point is difficult to work with in practical calculations, where concepts pertaining to the momentum space formulation are much more convenient. To avoid mathematical complications we assume that the π_R-functions are tempered distributions such that they possess a Fourier transform, and study the asymptotic behaviour of the Fourier transforms of the distributions instead of the singularity at the origin of the distributions themselves.

We here adopt a simple characterization of the asymptotic behaviour in momentum space of a tempered distribution given f. inst. by MEDVEDEV[9]. Let $T(z) = T(z_1, \ldots, z_n)$ be a tempered distribution with the Fourier transform $\tilde{T}(q_1, \ldots, q_n)$. We define the *rate of growth* at infinity in momentum space of $T(z)$ as the smallest integer $N = N(T)$ such that $\xi^{N+\alpha}$ increases faster than $\tilde{T}(\xi q_1, \ldots, \xi q_n)$ for $\xi \to \infty$ and any value of $\alpha > o$. Correspondingly, we call S more regular than T if $N(S) < N(T)$. If $N(T)$ is negative we simply say that T is regular. This ordering of distributions with respect to regularity is quite rough, as is illustrated by the remark that T and $T + P(\partial/\partial z)\delta(z)$ are equally regular whenever the degree of the polynomium P does not exceed $N(T)$. Still, the above characterization is sufficient for the discussion of the perturbation theory in the next paragraph. A more refined ordering of distributions with respect to regularity is proposed in the Appendix C.

* In contrast hereto, the distribution symbolized by $\varepsilon(t)\ \delta'(x^2)$ is perfectly well defined, as one sees by a similar calculation. A parallel, but simpler situation would arise in a two-dimensional theory (one x and one t), where $\delta(x^2 - t^2)$ is divergent, but $\varepsilon(t)\ \delta(x^2 - t^2)$ is convergent, if the usual way of treating δ-functions in mathematical physics is adopted.

Consider now the π_R-functions. If the system (3.6) does not admit regular solutions, we may instead look for the most regular solution. The rate of growth for the various π_R-functions of such a solution might be looked upon as constants characteristic of the singularity required by the interaction. Additional terms of the type $P(\partial/\partial z)\delta(z)$, when they make the π_R-functions less regular, introduce singularities of a complexity not required by the interaction. It is natural to assume that such singularities do not belong to the theory. This motivates the following formulation of the boundary conditions for the equations for the source operator: *The functions* $\pi_R(z_1,\ldots,z_n)$ *are as regular as compatible with the basic assumptions.* This boundary condition will be referred to as the principle of maximum regularity.

The principle of maximum regularity is not new, but has always been adhered to in the usual formulation of the theory. An explicit formulation of the principle may be found in the book by BOGOLIUBOV and SHIRKOV[10]. In the renormalization theory one simply introduces such renormalization constants, only, as are required to remove divergences according to the experience from the perturbation theory.

If the rate of growth at infinity in momentum space of a π_R-function is N, an additional singularity of the type $P(\partial/\partial z)\delta(z)$, where the degree of P does not exceed N, is left arbitrary. The coefficients of the various terms in the polynomial are thus arbitrary constants in the theory so far formulated. If this situation should occur, these constants should be determined by further conditions of the character of normalization conditions. Presumably no such arbitrary constants remain in the case of the Hurst-Thirring field. The situation in this respect might be different in, for instance, the π-meson theory, this theory being more singular. Alternatively such constants could be regarded as subject to physical interpretation, and thus as parameters belonging to the theory in the same way as the mass and the coupling constant.

4. Perturbation theory

As mentioned in the Introduction, we have not yet been able to prove that the theory proposed here gives the same results as the usual theory to all orders in the perturbation expansion. In the absence of a general proof we show in this paragraph that the two theories agree to the third order in the coupling constant. As the methods of calculation are somewhat different from the usual methods we present the third order calculation in some detail.

In perturbation theory we assume that the source operator is at least of the order g, and expand in powers of g. When only results for $B = o$ are

desired, it suffices to expand in powers of B as well. To indicate for a quantity F the sum of all terms of order $g^{\varrho} B^{\sigma}$, with $\varrho \leqslant r$ and $\sigma \leqslant s$, we employ the symbol $F^{(r;s)}$. It is easily seen that in the n'th order calculation we must calculate all terms $\boldsymbol{j}(x;y)^{(r;s)}$, with $r+s \leqslant n$ and $s < n$.

Second order perturbation theory.

To the first order in g and for $B = o$ we find from the variational equation for the source operator (3.2)

$$\boldsymbol{j}(x;y)^{(1;0)} = 2g\,\delta(x-y)\,\boldsymbol{A}_{in}(x); \tag{4.1}$$

whence, by (2.26) and (2.27),

$$\boldsymbol{j}(x)^{(1;0)} = g:\boldsymbol{A}_{in}^{2}(x):. \tag{4.2}$$

For $g = o$ and to the first order in B we have, by the field equation (3.4),

$$\boldsymbol{A}(x)^{(0;1)} = \boldsymbol{A}_{in}(x) + 2\int \Delta_R(x-1)\,B(1)\,\boldsymbol{A}_{in}(1)\,d(1). \tag{4.3}$$

These formulae allow us to calculate the operator $\boldsymbol{j}(x;y)^{(1;1)}$ from (3.2). From the resulting expression

$$\left.\begin{aligned}\boldsymbol{j}(x;y)^{(1;1)} = {} & 2g\,\delta(x-y)\,\boldsymbol{A}_{in}(x) + 4g\,\delta(x-y)\int \Delta_R(x-1)\,B(1)\,\boldsymbol{A}_{in}(1)\,d(1)\\ & + 4g\,\boldsymbol{A}_{in}(x)\,\Delta_R(x-y\,B(y)\end{aligned}\right\} \tag{4.4}$$

one finds, by the use of (2.26) and (2.27),

$$\left.\begin{aligned}\boldsymbol{j}(x)^{(1;1)} - \langle 0\,|\,\boldsymbol{j}(x)\,|\,0\rangle^{(1;1)} = {} & g:\boldsymbol{A}_{in}^{2}(x):\\ & + 4g\int \Delta_R(x-1)\,B(1):\boldsymbol{A}_{in}(x)\,\boldsymbol{A}_{in}(1):d(1).\end{aligned}\right\} \tag{4.5}$$

As explained in the beginning of paragraph 3, we determine the unknown vacuum expectation value $\langle 0\,|\,\boldsymbol{j}(x)\,|\,0\rangle^{(1;1)}$ by means of the system (3.6). A simple calculation gives for the right-hand side of the first of the equations (3.6) the result

$$\left.\begin{aligned}C(x)^{(2;0)} &= i\langle 0\,|\,[\boldsymbol{j}(x)^{(1;0)},\,\boldsymbol{j}(o)^{(1;0)}]\,|\,0\rangle\\ &= -\int_{4m^2}^{\infty} \Delta(x;\varkappa^2)\,\Pi^{(2)}(\varkappa^2)\,d\varkappa^2,\end{aligned}\right\} \tag{4.6}$$

with

$$\Pi^{(2)}(\varkappa^2) = \frac{g^2}{8\pi^2}\sqrt{\frac{\varkappa^2 - 4m^2}{\varkappa^2}}. \tag{4.7}$$

Thus, to this order, the system (3.6) becomes

$$\left.\begin{aligned} &\pi_R(x)^{(2;0)} - \pi_R(-x)^{(2;0)} = -\int_{4m^2}^{\infty} \Delta(x;\varkappa^2)\, \Pi^{(2)}(\varkappa^2)\, d\varkappa^2, \\ &\pi_R(x)^{(2;0)} = o, \quad \text{for } x_o < o. \end{aligned}\right\} \tag{4.8}$$

We have here used

$$\frac{g\delta \langle 0 \mid \boldsymbol{j}(x) \mid 0 \rangle^{(1;1)}}{\delta B(y)} = \pi_R(x-y)^{(2;0)}. \tag{4.9}$$

It is easily seen that, for $\pi_R(x)^{(2;0)}$, the naive ϑ-multiplication leads to a meaningless result. However, a particular solution to the system (4.8) is easily obtained by means of the identity $(-\Box + a^2)\Delta(x;\varkappa^2) = (a^2-\varkappa^2)\Delta(x;\varkappa^2)$. For the sake of convenience, we choose $a^2 < 4m^2$ and have the solution

$$\overset{\circ}{\pi}_R(x)^{(2;0)} = (-\Box + a^2)\int_{4m^2}^{\infty} \frac{\Delta_R(x;\varkappa^2)}{a^2-\varkappa^2}\, \Pi^{(2)}(\varkappa^2)\, d\varkappa^2. \tag{4.10}$$

Hence, the complete solution to (4.8) is given by

$$\pi_R(x)^{(2;0)} = \overset{\circ}{\pi}_R(x)^{(2;0)} + g^2 c_o\, \delta(x) + \ldots + g^2 c_N (-\Box)^N \delta(x), \tag{4.11}$$

where $c_o, c_1, \ldots, c_N$ are constants and N is an arbitrary positive integer. By the use of (4.5) and the above expression for the right-hand side of (4.9) we find, from (3.2),

$$\left.\begin{aligned} \boldsymbol{j}(x;y)^{(2;0)} = {} & \pi_R(x-y)^{(2;0)} + 2g\,\delta(x-y)\, \boldsymbol{A}_{in}(x) \\ & + 2g^2 \int \Delta_R(x-1) : \boldsymbol{A}_{in}^2(1) : d(1)\, \delta(x-y) \\ & + 4g^2\, \Delta_R(x-y) : \boldsymbol{A}_{in}(x)\, \boldsymbol{A}_{in}(y) : . \end{aligned}\right\} \tag{4.12}$$

Following the general pattern we next obtain $\boldsymbol{j}(x)$, to the same order, with the aid of (2.26) and (2.27). Due to $-\Box \boldsymbol{A}_{in}(x) = -m^2 \boldsymbol{A}_{in}(x)$ and the general formula

$$\int \overset{\circ}{\pi}_R(x-y)\, \boldsymbol{A}_{in}(y)\, d^4 y = c'\, \boldsymbol{A}_{in}(x), \tag{4.13}$$

where c' is a constant, we find

$$\left.\begin{aligned} \boldsymbol{j}(x)^{(2;0)} = {} & c''\boldsymbol{A}_{in}(x) + g : \boldsymbol{A}_{in}^2(x) : \\ & + 2g^2 \int \Delta_R(x-1) : \boldsymbol{A}_{in}(x)\, \boldsymbol{A}_{in}^2(1) : d(1). \end{aligned}\right\} \tag{4.14}$$

Here the constant c'' is given by $c'' = c' + g^2 c_o + \ldots + g^2 (-m^2)^N c_N$. To fulfill the regularity condition that (2.1) exists as a convolution integral, $c'' = o$

is required in order to avoid a $\delta(p^2+m^2)^2$—catastrophe in momentum space. Hence, by the field equation (3.4), we find the complete second-order expression for the field operator, viz.

$$\left.\begin{aligned} \boldsymbol{A}(x)^{(2;0)} = \boldsymbol{A}_{in}(x) + \int \Delta_R(x-1)\, g : \boldsymbol{A}_{in}^2(1) : d(1) \\ + 2g^2 \int \Delta_R(x-1)\, \Delta_R(1-2) : \boldsymbol{A}_{in}(1)\, \boldsymbol{A}_{in}^2(2) : d(12). \end{aligned}\right\} \tag{4.15}$$

The function $\pi_R(x)^{(2;0)}$ may now be calculated with the aid of the explicit expression (2.23). A simple calculation gives the result

$$\left.\begin{aligned} \pi_R(x-y)^{(2;0)} &= g \frac{\delta \langle 0 \,|\, \boldsymbol{j}(x) \,|\, 0 \rangle^{(1;1)}}{\delta B(y)} \\ &= K_x^2 \int_{4m^2}^{\infty} \frac{\Delta_R(x-y;\varkappa^2)}{(m^2-\varkappa^2)^2} \Pi^{(2)}(\varkappa^2)\, d\varkappa^2. \end{aligned}\right\} \tag{4.16}$$

From (4.16) we find $c_2 = \ldots = c_N = o$, and explicit expressions for c_0 and c_1 could be found. The results (4.16) and (4.15) are of course the well-known results of the renormalization theory.

Thus we see that, in the second order approximation, the function $\pi_R(x-y)$ is uniquely determined without the use of the principle of maximum regularity. Indeed, it may be shown by similar considerations as that above that this result is exactly true on the assumption that the spectral function $\Pi(\varkappa^2)$ in the Källén-Lehmann representation for $i\langle 0 \,|\, [\boldsymbol{j}(x), \boldsymbol{j}(o)] \,|\, 0 \rangle|_{B=o}$ is bounded for large values of $\varkappa^2$. If this is true one finds, as above, that the first of the π_R-functions is given by

$$\pi_R(x) = K_x^2 \int_{4m^2}^{\infty} \frac{\Delta_R(x;\varkappa^2)}{(m^2-\varkappa^2)^2} \Pi(\varkappa^2)\, d\varkappa^2. \tag{4.17}$$

This expression gives the well-known result for he polarization of the vacuum by a weak external field.

Third order perturbation theory.

For the higher-order calculations, the principle of maximum regularity is needed to determine the π_R-functions depending on two or more variables $z_1, z_2, \ldots$ It will be convenient to have a notation for the terms in a quantity $F^{(r;s)}$ which are proportional to $g^r B^s$. We shall denote these terms by the symbol $F^{(r;s)}$. Thus $\boldsymbol{j}(x)^{(3;0)} = \boldsymbol{j}(x)^{(3;0)} + \boldsymbol{j}(x)^{(2;0)}$. Hence, in the third order calculation, as we already know $\boldsymbol{j}(x)^{(2;0)}$, we need only calculate $\boldsymbol{j}(x)^{(3;0)}$.

To calculate this operator, we start from $\boldsymbol{j}(x;y)^{(1;2)}$ and follow the method outlined in the beginning of paragraph 3. By the variational equation for the source operator (3.2) we find

$$\boldsymbol{j}(x;y)^{(1;2)} = 2g\delta(x-y)\,\boldsymbol{A}(x)^{(0;2)} + \int \boldsymbol{j}(x;z)^{(1;1)}\,\Delta_R(z-y)\,d^4z\,2B(y), \tag{4.18}$$

where due to (4.4)

$$\left.\begin{aligned}\boldsymbol{j}(x;z)^{(1;1)} &= 4g\delta(x-z)\int \Delta_R(x-1)\,B(1)\,\boldsymbol{A}_{in}(1)\,d(1)\\ &\quad + 4g\,\boldsymbol{A}_{in}(x)\,\Delta_R(x-z)\,B(z),\end{aligned}\right\} \tag{4.19}$$

while $\boldsymbol{A}(x)^{(0;2)}$ is given by the field equation (3.4), viz.

$$\boldsymbol{A}(x)^{(0;2)} = \int \Delta_R(x-1)\,2B(1)\,\Delta_R(1-2)\,2B(2)\,\boldsymbol{A}_{in}(2)\,d(12). \tag{4.20}$$

Application of (2.26) and (2.27) — i. e. of the reduction formula for the source operator — yields the result

$$\left.\begin{aligned}&\boldsymbol{j}(x)^{(1;2)} - \langle 0\,|\,\boldsymbol{j}(x)^{(1;2)}\,|\,0\rangle\\ &\quad = 8g\int \Delta_R(x-1)\,B(1)\,\Delta_R(1-2)\,B(2):\boldsymbol{A}_{in}(x)\,\boldsymbol{A}_{in}(2):d(12)\\ &\quad + 4g\int \Delta_R(x-1)\,\Delta_R(x-2)\,B(1)\,B(2):\boldsymbol{A}_{in}(1)\,\boldsymbol{A}_{in}(2):d(12).\end{aligned}\right\} \tag{4.21}$$

According to the methods described in paragraph 3, the next step in the calculation consists in the evaluation of $g\langle 0\,|\,\delta\boldsymbol{j}(x)^{(1;2)}/\delta B(y)\,|\,0\rangle$ by the aid of the system (3.6). By (3.5)

$$C(x;y)^{(2;1)} = i\langle 0\,|\,[\boldsymbol{j}(x)^{(1;1)},\,\boldsymbol{j}(y)^{(1;0)}]\,|\,0\rangle - (x \leftrightarrow y), \tag{4.22}$$

where the relevant source operators are given by (4.5). One finds

$$C(x;y)^{(2;1)} = \int K(x,y;z)\,B(z)\,d^4z, \tag{4.23}$$

where the kernel $K(x,y;z)$ is given by

$$\left.\begin{aligned}K(x,y;z) &= 4g^2\,\Delta_R(x-z)\left\{\Delta^{(1)}(z-y)\,\Delta(y-x) - \Delta(z-y)\,\Delta^{(1)}(y-x)\right\}\\ &\quad - 4g^2\,\Delta_R(y-z)\left\{\Delta^{(1)}(z-x)\,\Delta(x-y) - \Delta(z-x)\,\Delta^{(1)}(x-y)\right\}.\end{aligned}\right\} \tag{4.24}$$

In this case, the formal solution of the system (3.6), i. e.

$$g\,\frac{\delta\langle 0\,|\,\boldsymbol{j}(x)^{(1;2)}\,|\,0\rangle}{\delta B(y)} = \frac{1}{g}\int \pi_R(x-y,\,x-z)^{(3;0)}\,B(z)\,d^4z, \tag{4.25}$$

with

$$\left.\begin{aligned}\pi_R(x-y,\,x-z)^{(3;0)} &= 4g^3\,\Delta_R(x-y)\left\{\Delta_R(x-z)\,\Delta^{(1)}(z-y)\right.\\ &\quad \left. + \Delta_R(y-z)\,\Delta^{(1)}(z-x)\right\}\\ &\quad + 4g^3\,\Delta_R(x-z)\,\Delta_R(z-y)\,\Delta^{(1)}(x-y)\end{aligned}\right\} \tag{4.26}$$

has a meaning, and defines a distribution. Simple considerations show that this distribution is regular in the sense defined in paragraph 3. Further, it is easily verified that all of the basic equations are fulfilled by the corresponding expression for $\boldsymbol{j}(x)^{(1;2)}$. The solution (4.26) is thus the one required by the principle of maximum regularity.

We may now calculate $\boldsymbol{j}(x;y)^{(2;1)}$. By (3.2)

$$\left.\begin{aligned}\boldsymbol{j}(x;y)^{(2;1)} &= 2g\delta(x-y)\,\boldsymbol{A}(x)^{(1;1)} + g\frac{\delta \boldsymbol{j}(x)^{(1;2)}}{\delta B(y)} \\ &\quad + \int \boldsymbol{j}(x;z)^{(2;0)}\, \Delta_R(z-y)\, d^4z\, 2B(y).\end{aligned}\right\} \tag{4.27}$$

Here $\boldsymbol{A}(x)^{(1;1)}$ may be obtained from the field equation (3.4) and the formula

$$\left.\begin{aligned}\boldsymbol{j}(x)^{(1;1)} &= \frac{1}{g}\int \pi_R(x-1)^{(2;0)}\, B(1)\, d(1) \\ &\quad + 4g \int \Delta_R(x-1)\, B(1) : \boldsymbol{A}_{in}(x)\, \boldsymbol{A}_{in}(1) : d(1),\end{aligned}\right\} \tag{4.28}$$

which is a consequence of (4.5) and (4.9). Thus,

$$\left.\begin{aligned}\boldsymbol{A}(x)^{(1;1)} &= \frac{1}{g}\int \Delta_R(x-1)\, \pi_R(1-2)^{(2;0)}\, B(2)\, d(12) \\ &\quad + 4g \int \Delta_R(x-1)\, \Delta_R(1-2)\, B(2) : \boldsymbol{A}_{in}(1)\, \boldsymbol{A}_{in}(2) : d(12) \\ &\quad + 2g \int \Delta_R(x-1)\, B(1)\, \Delta_R(1-2) : \boldsymbol{A}_{in}^2(2) : d(12).\end{aligned}\right\} \tag{4.29}$$

An expression for $g\delta \boldsymbol{j}(x)^{(1;2)}/\delta B(y)$ may be found from (4.21) and (4.25). Finally, $\boldsymbol{j}(x;y)^{(2;0)}$ is obtained from the expression (4.12). In this way $\boldsymbol{j}(x;y)^{(2;1)}$ may be calculated. The result is

$$\boldsymbol{j}(x;y)^{(2;1)} = \langle 0 \,|\, \boldsymbol{j}(x;y)^{(2;1)} \,|\, 0 \rangle + \boldsymbol{j}(x;y)^{(2;1)}_{\mathrm{II}}, \tag{4.30}$$

where $\boldsymbol{j}(x;y)^{(2;1)}_{\mathrm{II}}$ denotes a two-particle term. The vacuum expectation value is found to be

$$\left.\begin{aligned}\langle 0 \,|\, \boldsymbol{j}(x;y)^{(2;1)} \,|\, 0 \rangle &= 2\delta(x-y) \int \Delta_R(x-1)\, \pi_R(1-2)^{(2;0)}\, B(2)\, d(12) \\ &\quad + 2\int \pi_R(x-1)^{(2;0)}\, \Delta_R(1-y)\, d(1)\, B(y) \\ &\quad + (1/g)\int \pi_R(x-y, x-1)^{(3;0)}\, B(1)\, d(1),\end{aligned}\right\} \tag{4.31}$$

while the two-particle part of the operator $\boldsymbol{j}(x;y)^{(2;1)}$ becomes identical with the (meaningful) expression found by the application of the formal unrenormalized canonical theory. The first two terms on the right-hand side of (4.31) originate from the first and the last term on the right-hand side of (4.27), respectively. Thus both these "dangerous" terms are brought into

the third order calculation not by the second term on the right-hand side of (4.27), but by the two other terms and in the already properly normalized form obtained by the second order calculation.

By the expression for $\boldsymbol{j}(x;y)^{(2;1)}$ and by (2.26) and (2.27) we find $\boldsymbol{j}(x)^{(2;1)} - \langle 0 \,|\, \boldsymbol{j}(x)^{(2;1)} \,|\, 0 \rangle$. It is easily seen that, in fact, the vacuum expectation value is equal to zero. For by (3.6)

$$g \frac{\delta \langle 0 \,|\, \boldsymbol{j}(x)^{(2;1)} \,|\, 0 \rangle}{\delta B(y)} - (x \leftrightarrow y) = C^{(3;0)}(x;y), \tag{4.32}$$

where by (3.5)

$$C(x;y)^{(3;0)} = i \langle 0 \,|\, [\boldsymbol{j}(x)^{(2;0)}, \boldsymbol{j}(y)^{(1;0)}] \,|\, 0 \rangle - (x \leftrightarrow y), \tag{4.33}$$

i. e. by (4.14)

$$\left.\begin{aligned} C(x;y)^{(3;0)} = {} & 2ig^3 \int \Delta_R(x-1) \langle 0 \,|\, [:\boldsymbol{A}_{in}(x)\boldsymbol{A}_{in}^2(1):, :\boldsymbol{A}_{in}^2(y):] \,|\, 0 \rangle \\ & - (x \leftrightarrow y). \end{aligned}\right| \tag{4.34}$$

Hence

$$C(x;y)^{(3;0)} = o, \tag{4.35}$$

due to the fact that the vacuum expectation value is required for an odd number of incoming fields. It is now easily seen that the principle of maximum regularity requires

$$g \frac{\delta \langle 0 \,|\, \boldsymbol{j}(x)^{(2;1)} \,|\, 0 \rangle}{\delta B(y)} = o. \tag{4.36}$$

The equation (3.2) gives the expression

$$\boldsymbol{j}(x;y)^{(3;0)} = 2g\delta(x-y)\boldsymbol{A}(x)^{(2;0)} + g \frac{\delta \boldsymbol{j}(x)^{(2;1)}}{\delta B(y)}. \tag{4.37}$$

Here $\boldsymbol{A}(x)^{(2;0)}$ is given by (4.15) and $\boldsymbol{j}(x)^{(2;1)}$ is known already. From the resulting expression for $\boldsymbol{j}(x;y)^{(3;0)}$ and with the aid of equations (2.26) and (2.27) one finally finds the result, well known from the renormalization theory, that

$$\boldsymbol{j}(x)^{(3;0)} = \boldsymbol{j}(x)_{\mathrm{II}}^{(3;0)} + \boldsymbol{j}(x)_{\mathrm{IV}}^{(3;0)}, \tag{4.38}$$

where the two-particle part is given by

$$\left.\begin{aligned} \boldsymbol{j}(x)_{\mathrm{II}}^{(3;0)} = {} & g \int \pi_R(x-1)^{(2;0)} \Delta_R(1-2) :\boldsymbol{A}_{in}^2(2): d(12) \\ & + 2g^3 \int \Delta_R(x-1) \Delta_R(x-2) \Delta^{(1)}(1-2) :\boldsymbol{A}_{in}(1)\boldsymbol{A}_{in}(2): d(12) \\ & + 4g^3 \int \Delta_R(x-1) \Delta^{(1)}(x-2) \Delta_R(1-2) :\boldsymbol{A}_{in}(1)\boldsymbol{A}_{in}(2): d(12), \end{aligned}\right| \tag{4.39}$$

while the four-particle part* $\boldsymbol{j}(x)_{\mathrm{IV}}^{(3;\,0)}$ is identical with the four-particle part found by use of the formal theory.

The author has benefited from discussions with several members of the staffs at the institutes of physics and mathematics at Aarhus university. It is a pleasure here to thank Lars Mejlbo and Ebbe Thue Poulsen for helpful discussions as regards various questions in the theory of distributions. In particular my thanks are due Poul Werner Nielsen for much helpful criticism and valuable suggestions.

Institute of Physics
Aarhus University

APPENDIX A

The notations used for the theory of free mesons

We use the pseudo-Euclidean metric, $x = (x_1, x_2, x_3, x_4)$ where $x_4 = ix_0$. Further, $x^2 = x_\mu x_\mu = \vec{x}^2 - x_0^2$, and $-\Box = -\Delta + \partial^2/\partial x_0^2$. The free field operators $\boldsymbol{A}_{in}(x)$ are self-adjoint operator valued distributions which satisfy the commutation relations

$$[\boldsymbol{A}_{in}(x), \boldsymbol{A}_{in}(y)] = i\Delta(x-y),$$

where

$$\Delta(x-y) = -i(2\pi)^{-3} \int d^4p\, \varepsilon(p_0)\, \delta(p^2+m^2) \exp ipx.$$

Here, $\varepsilon(p_0) = p_0/|p_0|$. The retarded Greens function is $\Delta_R(x) = -\vartheta(x)\Delta(x)$, where $\vartheta(x)$ is the Heaviside function

$$\vartheta(x) = \begin{cases} 1, \text{ for } x_0 > o \\ o, \text{ for } x_0 < o. \end{cases}$$

The value of $\vartheta(x)$ for $x_0 = o$ is not important.

The Wick product is denoted by $:\ldots:$, and designates that the operators inside the double dots are ordered such that any positive frequency part stands to the right of any negative frequency part. Useful rules for the calculation with Wick products are found for instance in the book by Bogoliubov and Shirkov[10].

* I. e. an operator of the form

$$\frac{1}{4!} \int f(x; 1, 2, 3, 4) : \boldsymbol{A}_{in}(1)\, \boldsymbol{A}_{in}(2)\, \boldsymbol{A}_{in}(3)\, \boldsymbol{A}_{in}(4) : d(1234).$$

APPENDIX B

Volterra derivatives

Volterra derivatives seem to be an indispensable mathematical tool for the theory proposed here. It might therefore be useful to give a brief introduction to the theory of variational derivatives.

Consider a functional $\Phi[j]$, which maps from a certain space of functions j into the complex plane. Useful definitions of differentiability of such functionals are all of the following general type: The functional Φ is called differentiable of j if the variation of Φ is of the form

$$\Phi[j+\delta j]-\Phi[j]=\int \Psi[j;x]\,\delta j(x)\,d^4x+o[j,\delta j], \tag{B.1}$$

where $\int \Psi[j;x]\,\delta j(x)\,d^4x$ is a linear functional of δj and $o[j,\delta j]$ has certain properties. Roughly speaking, it is required that for j fixed and $\delta j\to o$, $o[j,\delta j]$ tends to zero "faster than" δj. Hence, to give a precise definition of differentiability, one has to specify

(i) the meaning of $\delta j\to o$, and
(ii) the meaning of the term "faster than".

For any such specification, we call $\Psi[j;x]$ the Volterra derivative of Φ, and use the symbol

$$\Psi[j;x]=\frac{\delta\Phi[j]}{\delta j(x)}. \tag{B.2}$$

A simple possibility is the following:

(i) $\delta j\to o$ means $\delta j=\xi j^{(1)}$, where $j^{(1)}$ is fixed and $\xi\to o$,
(ii) $o[j,\xi j^{(1)}]$ is required to be $o(\xi)$ for j and $j^{(1)}$ fixed, i. e.

$$\lim_{\xi\to o}\frac{o[j,\xi j^{(1)}]}{\xi}=o.$$

This immediately leads to the relation

$$\int \Psi[j;x]\,j^{(1)}(x)\,d^4x=\left.\frac{d\Phi[j+\xi j^{(1)}]}{d\xi}\right|_{\xi=o}. \tag{B.3}$$

We stress here that, to our knowledge, no argument is known which indicates that this particular definition is the most adequate one for use in the quantum field theory. However, other reasonable definitions seem to be more restrictive. Thus, in general, the existence of the right-hand side of (B. 3) will be a necessary condition for differentiability.

As mentioned in the Introduction, when operator valued functionals are considered, the weak topology in the Hilbert space may be used to give the relations in the text a well defined meaning. If further the definition (B. 3) is adopted, and if the field operator is regarded as an operator valued distribution, a symbol like $\delta \boldsymbol{A}(x)/\delta j(y)$ becomes endowed with the interpretation

$$\langle A \mid \int \frac{\delta \boldsymbol{A}(x)}{\delta j(y)} \psi(x)\, d^4x \mid B \rangle = \frac{\delta}{\delta j(y)} \langle A \mid \boldsymbol{A}(x)\, \psi(x)\, d^4x \mid B \rangle, \qquad \text{(B. 4)}$$

where $\psi(x)$ is a testing function for the operator valued distribution $\boldsymbol{A}(x)$, and the variational derivative on the right-hand side is the one defined above.

Finally, a remark about certain interchanges of limiting processes, frequently performed in the text, may be in its place. As an example consider the relation

$$(-\Box_x + m^2) \frac{\delta \boldsymbol{A}(x)}{\delta j(y)} = \frac{\delta}{\delta j(y)} (-\Box_x + m^2) \boldsymbol{A}(x). \qquad \text{(B. 5)}$$

If the possible interpretation mentioned above is employed, this equation means

$$\left.\begin{aligned} &\frac{d}{d\xi} (-\Box_x + m^2) \langle A \mid \boldsymbol{A}[j + \xi j^{(1)}; x] \mid B \rangle \\ &= (-\Box_x + m^2) \frac{d}{d\xi} \langle A \mid \boldsymbol{A}[j + \xi j^{(1)}; x] \mid B \rangle \end{aligned}\right\} \qquad \text{(B. 6)}$$

for $\xi = o$. This only requires an interchange of two ordinary differential operators. The validity of such relations is assumed in the text.

APPENDIX C

On the formulation of the principle of maximum regularity

In paragraph 3 the concept of the rate of growth at infinity in momentum space was used to formulate the principle of maximum regularity. Such a formulation is satisfactory due to its simplicity, but might not always work. It presupposes that the π_R-functions are tempered distributions for which the Fourier transforms are functions for large values of the momenta. Both these properties might be difficult to prove without recourse to an approx-

imation method. In this appendix we propose an ordering of distributions with respect to regularity which avoids these problems. Only standard notions in the theory of distributions are used, and for these we refer to the book of L. SCHWARTZ[11].

We consider distributions $T(\xi)$ defined on a f-dimensional Euclidean space, $\xi = (\xi_1, \xi_2, \ldots, \xi_f)$. Differential operators are denoted by $D^\alpha = \partial^{\alpha_1+\alpha_2+\ldots+\alpha_f}/\partial\xi_1^{\alpha_1}\,\partial\xi_2^{\alpha_2}\ldots\partial\xi_f^{\alpha_f}$, and $|\alpha| = \alpha_1+\alpha_2+\ldots+\alpha_f$ is the degree of the differential operator. As only local properties are considered, we need not specify the type of the distribution. Let $\mathfrak{O}$ be an open *bounded* set and $D(\mathfrak{O})$ the totality of testing functions which vanish outside $\mathfrak{O}$. To define the concept of the order of a distribution in $\mathfrak{O}$, we need the seminorms

$$p_k(\psi) = \max_{|\alpha|=k}\ \max_{\xi\varepsilon\mathfrak{O}} |D^\alpha\psi(\xi)|. \tag{C. 1}$$

The notion of the order of a distribution, as given in the book of Schwartz, is easily seen to be equivalent to the following definition (valid for bounded sets $\mathfrak{O}$ only): The order of the distribution T in $\mathfrak{O}$ is the smallest integer $m_T(\mathfrak{O})$ for which there exists a constant C, such that

$$|\int T(\xi)\,\psi(\xi)\,d\xi| \leqslant C\ p_{m_T(\mathfrak{O})}(\psi), \tag{C. 2}$$

for all $\psi\varepsilon\ D(\mathfrak{O})$. We define the amplitude $A_T(\mathfrak{O})$ of T in $\mathfrak{O}$ as the infimum of the possible constants C in (C. 2). The order and the amplitude of the distribution T at a point, say the origin o of ξ-space, may now be defined as follows: *The order of T at o is the smallest integer m_T for which there exists a neighbourhood $\mathfrak{O}$ of o in which the order of T is m_T. The amplitude of T at o, A_T, is the infimum of the amplitude over all neighbourhoods of o, i. e. $A_T = \inf_{\mathfrak{O}\ni o} A_T(\mathfrak{O})$.*

The two numbers, m_T and A_T, may now serve to order distributions with respect to their behaviour at $\xi = o$. We say that S is more regular than T at the origin if $m_S < m_T$ or if $m_S = m_T$ but $A_S < A_T$. In this manner all distributions may be compared to each other, and in particular to derivatives of δ. However, the comparison is quite rough. In general, if $m_S < m_T$, the distributions $T+S$ and T are equally regular at the origin.

The formulation of the principle of maximum regularity may now be taken over from paragraph 3.

It is obvious that the ϑ-multiplication thus defined (although not always uniquely defined) constitutes a generalization of the ordinary product of two functions. Indeed, for sufficiently regular distributions, i. e. for distributions

of the order zero and of the amplitude zero at the origin, the requirement of maximum regularity gives the same result as the "naive" ϑ-multiplication which succeeds in this case. It is easily seen that such distributions locally are measures continuous at the origin.

The concepts of the order and the amplitude of a distribution at a point are much more powerful than the concept of the rate of growth at infinity in momentum space to analyse the dominating singularity of the distribution. This fact is revealed by the following theorem, which we give without proof: *If m is the order at the origin in $(x_1, x_2, \ldots, x_n)$-space of the distribution $\pi_R(x_1, x_2, \ldots, x_n)$ there exists a unique relativistically invariant and symmetric polynomial $P_m(\partial/\partial x_1, \partial/\partial x_2, \ldots, \partial/\partial x_n)$, homogeneous of the m'th degree, such that $\pi_R - P_m\delta$ has the amplitude zero at the origin.*

The main point is that the polynomial exists and is *unique*. The relativistic invariance and the symmetry of the polynomial are then a trivial consequence of the fact that the statement that the amplitude at the origin is zero involves symmetric and invariant concepts only. In general one cannot reduce the order by a regularization process f this type.

References

1) H. Lehmann, K. Symanzik and W. Zimmermann, Nuovo Cimento **1**, 205 (1954); Nuovo Cimento **4**, 319 (1957).
2) V. Glaser, H. Lehmann and N. Zimmermann, Nuovo Cimento **4**, 1122 (1957).
3) K. Nishijima, Phys. Rev. **111**, 995 (1958); Phys. Rev. **119**, 485 (1960).
4) M. Muraskin and K. Nishijima, Phys. Rev. **122**, 331 (1961).
5) R. Peierls, Proc. Roy. Soc. London A, **214**, 143 (1952).
6) J. Schwinger, Proc. Nat. Acad. Sci. **37**, 452 (1951); Phys. Rev. **82**, 914 (1952).
7) R. Haag, Mat. Fys. Medd. Dan. Vid. Selsk. **29**, no. 12 (1955).
8) L. Gårding and I. L. Lions, Functional Analyses. Supplemento al vol. **14** del Nuovo Cimento, No. 1, p. 9 (1959).
9) B. V. Medvedev, Doklady Akademii Nauk SSSR. **135**, 1087 (1960).
10) N. N. Bogoliubov and D. V. Shirkov, Introduction to the Theory of Quantized Fields. London 1959. In particular p. 331.
11) L. Schwartz, Théorie des Distributions. Tome I. Paris (1957).

Indleveret til Selskabet den 26. september 1961.
Færdig fra trykkeriet den 6. april 1962.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 10

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 10 (1963)

INTEGRAL EQUATIONS GOVERNING RADIATION EFFECTS

(NOTES ON ATOMIC COLLISIONS, III)

BY

J. LINDHARD, V. NIELSEN, M. SCHARFF(†)
AND P. V. THOMSEN

København 1963
i kommission hos Ejnar Munksgaard

CONTENTS

Synopsis

A theoretical study is made of damage effects by particle radiations in matter, and their dependence on energy, mass and charge number of an incoming particle, as well as on the composition of the medium. Typical examples of damage effects are the number of ion pairs formed in a gas, or the number of vacancies created in a crystal. We are particularly concerned with the consequences of the competition between energy transfer to atomic electrons and to translatory motion of an atom as a whole. For these purposes, common integral equations are formulated and studied. We treat primarily average effects resulting from an atomic particle with given energy, but also their average fluctuation and probability distribution.

As an important example we study the division of the total energy dissipation, E, into energy given to recoiling atoms, ν, and energy given to electrons, $E - \nu$. Several radiation effects are accounted for from knowledge about E and ν.

The primary quantities in a study of radiation effects are the cross sections for all relevant collision processes. We use comprehensive estimates of cross sections, derived elsewhere in a Thomas-Fermi treatment. Various simple approximations are introduced; analytical and numerical estimates are made of solutions to the integral equations. For many purposes nuclear collisions and electronic collisions may be treated as if they were unconnected events, although this is not quite correct, especially at low energies. Considerable simplification is obtained by a suitable scaling of energy. A key to a common experimental and theoretical study is provided by an incoming particle identical with the atoms of the substance. Only few experiments can at present be compared quantitatively with theory.

Printed in Denmark
Bianco Lunos Bogtrykkeri A/S

§ 1. Introduction

When an atomic particle is slowed down in a substance, a wide variety of damage effects may be observed. Familiar phenomena of this kind are the number of ion pairs formed in a gas, the number of electron-hole pairs in a semiconductor, or the number of defects in a solid. Other damage effects have been studied less, or not at all, like the number of electrons ejected from atomic K-shells, or the number of dissociations of molecules. The observations of damage phenomena may be divided into two classes. The one is particle detection, where the effect of a single incoming particle is observed and possibly recorded in time, and the other is the total damage due to many particles, as in reactor materials.

All damage effects depend on a competition between the cross sections for a multitude of different processes. Theoretical studies have been made by many authors concerning some aspects of excitation and ejection of electrons. Other theoretical studies have been concerned with the average energy required to form defects in solids. Less attention has been paid to the question of the competition between, on the one hand, energy transfer to atomic electrons and, on the other hand, energy transfer to translatory motion of an atom as a whole. Our knowledge of collision processes for slow heavy particles has been scanty, and the mentioned competition does in fact occur primarily for slow heavy particles.

To a wide extent all above damage processes may be described by integral equations which are formally equivalent. The differences concern mostly the inhomogeneous parts or boundary conditions. But the competition between energy transfer to electrons and to atomic recoils can be described by equations which have even more in common. This is because there are extensive similarity properties, of Thomas-Fermi type, between the competing processes in this case. The homogeneous integral equations in different substances are actually quite closely connected. It can therefore be worthwhile to study them in some detail. When we have gained insight in the equations we can handle not only average damage effects, but also fluctuations and even the distribution in probability.

We shall be concerned mainly with one effect which corresponds to the simplest homogeneous equations. This effect is the division of the dissipated energy between electrons and recoiling atoms in the substance. More precisely, for an incoming particle of energy E we ask for that part η of the total energy loss, E, which is ultimately given to electrons, and that part ν, which is ultimately left in atomic motion. Since this division is a useful and simple concept, we comment on it in some detail as an example of the application of the general equations.

It might seem as if the division into η and ν were not quite well-defined, since we are not concerned with the final thermal equilibrium. However, on the one hand, the energy once given to electrons can be transferred back to atomic motion only extremely slowly and in exceedingly small bits.* On the other hand, sufficiently slow atoms no longer excite electrons and their energy may be frozen in or become thermalized. This may give a qualitative justification of the separation into ν and η.

For the present purpose the quantities η and ν may be specified as follows. We consider η as the sum total of the energy given to electrons, i. e. for ejected electrons it is the kinetic energy plus the original binding while for excited electrons it is the excitation energy. Correspondingly, ν is the total energy given to atoms, excluding internal excitation of atoms. Thus, η and ν are quite well-defined, and have the sum $\eta + \nu = E$.** It is clear that there must be a probability distribution $P(\nu, E)\,d\nu$ in the variable ν, such that

$$\int_0^\infty P(\nu, E)\,d\nu = 1\,, \quad \bar{\nu}(E) = \int_0^\infty \nu\, P(\nu, E)\,d\nu\,,$$

and similarly for the higher moments. For the present we may disregard fluctuations and consider only e.g. $\bar{\nu} = \bar{\nu}(E)$.

We shall attempt to show how $\bar{\eta}(E)$, $\bar{\nu}(E)$ and other cumulative effects may be derived for all kinds of particles in any medium. Since $\bar{\eta}$ and $\bar{\nu}$ are determined by the competition between energy transfer to electrons and to atomic recoils in all collisions during slowing-down, they are expected to depend on the medium, on the type of particle and on its energy. This enormous variability can be reduced somewhat by studying at first the more basic cases.

* An exception occurs if an electron by exciting atomic electrons gives rise to large vibrations or even disruption of bindings in molecules (through a Franck-Condon effect or an Auger effect). The energy transferred in this way from a moving electron into atomic motion can be appreciable. This effect must be studied separately, and is remarkable in that it does not occur in monatomic gases. – The role of the Auger effect is studied by Durup and Platzman (1961).

** If more subtle distinctions are necessary, we may divide E into components other than η and ν. Examples are the energy escaping as X-rays or as near-thermal excitations.

Let us start by considering the case where the medium consists of only one atomic species, of atomic number Z_2 and mass number A_2. Now, any incoming particle, irrespective of its type, gives rise to recoiling atoms of the medium, and we will have to make use of their value of $\overline{\nu}$ and $\overline{\eta}$. It follows that the simplest basic case occurs when the atomic number, Z_1, and the mass number, A_1, of the incoming particle are equal to those of the medium, i. e. $Z_1 = Z_2$, $A_1 = A_2$.

Suppose that a particle belonging to the medium ($Z_1 = Z_2$) initially has an energy E; we want to find $\overline{\eta}(E)$. Collisions with atoms result in recoiling atoms or ions which may have any energy E' within the interval $0 < E' < E$, and therefore the corresponding values of $\overline{\eta}(E')$ enter in the evaluation of $\overline{\eta}(E)$, as must also the differential cross section for energy transfer to recoiling atoms. Clearly, the procedure must be to build up $\overline{\eta}$ starting from zero energy, and the relative magnitude of the partial stopping cross sections $S_e(E)$ and $S_n(E)$ must be of direct importance. We shall therefore review briefly the behaviour of the relevant cross sections.

When deliberating the approach to these problems, one should first of all bear in mind that extreme accuracy and separate computation of each individual case cannot be the primary aim. Also, a discussion of quite low energies of heavy particles, less than 100 eV say, is either unnecessary or may be made separately. It is therefore desirable to use statistical methods, of type of the Thomas-Fermi treatment, as far as possible. Since at moderate energies the interpenetration of two atomic electron clouds can be considerable, many atomic electrons with moderate bindings play a part, and statistical methods seem promising. In the case $Z_1 = Z_2$, $A_1 = A_2$ the function $\overline{\eta}$ can depend on three variables, $\overline{\eta} = \overline{\eta}(Z_2, A_2, E)$. It turns out that the Thomas-Fermi treatment together with a suitable approximation to scattering reduces the number of variables. In fact, beside the energy measured in a suitable Thomas-Fermi scale there is only one further parameter, which even has approximately the same value in most cases. Such reductions in the number of variables lead to highly desirable simplifications in the theoretical treatment.

We shall already here give a brief summary of relevant stopping cross sections and differential cross sections. The cross sections are derived elsewhere (LINDHARD and SCHARFF (1961), and Notes on Atomic Collisions I and IV (unpublished)). We do not claim that the accuracy is very high, and in individual cases other authors may have obtained better estimates. The primary purpose for the present is to have available comprehensive formulas, applicable in as many cases as possible. Maybe the greatest uncertainty is the proportionality factor, k, in the electronic stopping. A considerable number of observations on range and on scattering have been made;

they appear to be in fair agreement with the formulas quoted here (LINDHARD, SCHARFF and SCHIØTT (1962)).

Stopping cross sections.

The nuclear stopping cross section $S_n = \int T_n d\sigma_n$ depends on the particle energy E, and on the parameters Z_1, Z_2, A_1 and A_2. An important region of low velocities corresponds to v less than $\sim 0.015\, v_0 Z^{2/3}$, where $Z^{2/3} = Z_1^{2/3} + Z_2^{2/3}$ and $v_0 = e^2/\hbar$. In this region S_n remains nearly constant, and we shall sometimes approximate S_n by the constant standard stopping cross section S_n^0 (similar to that quoted by BOHR (1948)),

$$S_n^0 = (\pi^2/2.7183) e^2 a_0 Z_1 Z_2 M_1 \cdot Z^{-1/3} (M_1 + M_2)^{-1}. \tag{1.1}$$

In a more accurate description S_n increases slowly towards a maximum (cf. Fig. 1), and (1.1) may be used in the neighbourhood of the maximum. Beyond it, S_n decreases corresponding to an increasing negative power of E, but always slower than E^{-1}. In fact, S_n approaches the classical stopping formula in a screened Coulomb potential.

It turns out that the nuclear stopping is most simply described by a suitable scaling of energy and cross section. Introduce the dimensionless quantities

$$\varepsilon = E \frac{a M_2}{Z_1 Z_2 e^2 (M_1 + M_2)} \text{ and } \varrho = RNM_2 \cdot 4\pi a^2 \frac{M_1}{(M_1 + M_2)^2}$$

as measures of energy and range, where $a = 0.8853 a_0 \cdot Z^{-1/3}$, while R is the usual range and N the number of atoms per unit volume. The derivative $(d\varepsilon/d\varrho) = S \cdot (M_1 + M_2)/(4\pi e^2 a Z_1 Z_2 M_1)$ is a dimensionless measure of the stopping cross section, S. To a good approximation all nuclear stopping cross sections are then described by one curve. This is shown in Fig. 1, where the solid curve was computed from the comprehensive scattering cross section in Fig. 2. The approximation $S_n = S_n^0$ is represented by the horizontal dotted line $(d\varepsilon/d\varrho)_n = 0.327$.

The electronic stopping cross section is nearly proportional to v in a considerable velocity interval, i.e. for $v < v_1 \cong v_0 \cdot Z_1^{2/3}$, and is of order of

$$S_e \cong 8\pi e^2 a_0 \frac{Z_1^{7/6} Z_2}{Z} \cdot \frac{v}{v_0}, \quad v < v_1. \tag{1.2}$$

This leads to an electronic contribution to stopping in an ε-plot $(d\varepsilon/d\varrho) = k \cdot \varepsilon^{1/2}$, where the quantity k as given by (1.2) depends somewhat on Z_1, Z_2, M_1 and M_2, but is often within the interval $0.10 < k < 0.20$. This holds in particular in the case of $Z_1 = Z_2$, $A_1 = A_2$, where $k = 0.133\, Z_2^{2/3} A_2^{-1/2}$, so that k varies only little with Z_2. Merely in the special case of $Z_2 \gg Z_1$, with Z_1 comparable to 1, does k appreciably exceed 0.20. The dashed straight line in Fig. 1 shows the electronic stopping for a representative value of k ($k = 0.15$). It cuts the horizontal line $S_n = S_n^0$ at an energy E_c corresponding to $\varepsilon_c = 4.75$.

In the neighbourhood of $v = v_1$ the electronic stopping has a maximum, upon which it decreases and gradually approaches the Bethe stopping formula.

Let us take the ratio $\xi(E) = S_e/S_n$ as a measure of the division of energy dissipation into electronic and atomic motion. The above summary of

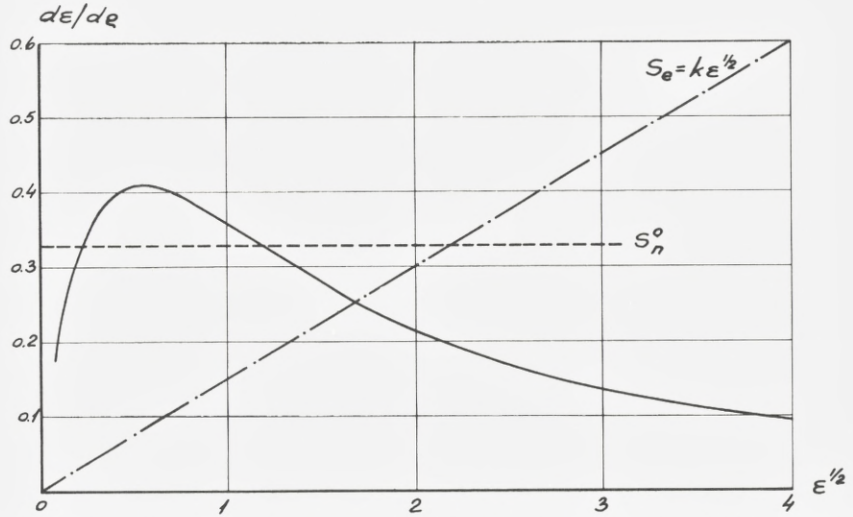

Fig. 1. Theoretical stopping cross sections in $\varrho - \varepsilon$ variables. The abscissa is $\varepsilon^{1/2}$, i.e. proportional to v. The solid curve is $(d\varepsilon/d\varrho)_n$ computed from the Thomas-Fermi cross section in Fig. 2. The horizontal dashed line indicates (1.1) and the dot-and-dash line is the electronic stopping cross section, $k\varepsilon^{1/2}$, for $k = 0.15$.

stopping cross sections then shows that there is a natural division into three regions of different behaviour. In the lowest energy region, region I, the nuclear stopping is dominating and relatively little energy goes into electronic motion. Region I is bounded upwards by an energy roughly equal to E_c. Above E_c the nuclear stopping falls off, while the electronic stopping goes on increasing as $E^{1/2}$. This is region II, with an upper bound given by v_1, i.e. ε_1 is of order of 10^3 or larger. In region II the ratio ξ increases rapidly, and the fraction of energy going into electronic motion must increase correspondingly. Finally, above ε_1 the electronic stopping starts decreasing, and the ratio ξ, though still increasing, approaches a maximum value of order of $2M_p/m \sim 4000$; this is region III. The division into three regions is convenient only when $Z_1 = Z_2$.

Differential cross sections.

Although the stopping cross sections are relevant and give a qualitative picture of the events, they contain only part of the necessary information. In fact, in the following the integral equations demand a detailed knowledge of the differential scattering cross section in nuclear collisions. As regards electronic collisions, we normally need no more than the stopping cross section itself.

We shall briefly recapitulate two different approximations to the differential cross section in nuclear collisions, assuming the scattering to be approximately

elastic (LINDHARD and SCHARFF (1961), and Notes on Atomic Collisions, I). First, in an s'th power potential, $V(r) = Z_1 Z_2 e^2 a_s^{s-1} s^{-1} r^{-s}$, with $a_s \approx a = 0.8853\, a_0 \cdot Z^{-1/3}$, the differential scattering cross section is approximately equal to

$$d\sigma_n = \frac{C_n}{T_m^{1-1/s}} \cdot \frac{dT}{T^{1+1/s}}, \quad s > 1, \tag{1.3}$$

where the incoming particle with energy E transfers an energy T to an atom originally at rest. Here, $T \leq T_m = \gamma E = 4\, M_1 M_2 (M_1 + M_2)^{-2} E$, T_m being the maximum energy transfer in the collision. Furthermore, the constant C_n is connected to the stopping cross section S_n, and is approximately given by

$$C_n = \frac{\pi}{s}\left(b^2 a_s^{2s-2} \frac{3s-1}{8s^2}\right)^{1/s} \cdot T_m = \left(1 - \frac{1}{s}\right) S_n, \quad b = \frac{2 Z_1 Z_2 e^2}{M_0 v^2},$$

M_0 being the reduced mass. In preliminary discussions these simple formulas are quite useful, especially for explorative purposes. The case of $s = 2$, where $S_n = S_n^0$ is independent of energy, appears to be a fair approximation at energies somewhat below E_c. At extremely low energies, $s = 3$ is preferable. At high energies s tends to 1.

A more accurate description is obtained from an interaction potential $V(r) = (Z_1 Z_2 e^2/r) \cdot \varphi_0(r/a)$, where $\varphi_0(x)$ is the Fermi function belonging to a single Thomas-Fermi atom. It turns out that the differential cross section is now to a good approximation, for all Z_1, Z_2, A_1, A_2 and all non-relativistic energies, equal to

$$d\sigma = \pi a^2 \frac{dt}{2 t^{3/2}} f(t^{1/2}), \tag{1.4}$$

where $t = \varepsilon^2 \cdot (T/T_m) = \varepsilon^2 \cdot \sin^2 \frac{\vartheta}{2}$. The variable t is proportional to the energy transfer T, and to the energy E through ε^2/T_m. Thus, one universal function of a single variable, $f(t^{1/2})$, describes the scattering at all energies and scattering angles, and for all atom-ion pairs. The function f was computed numerically from the Fermi function, and is shown in Fig. 2. At high energies and not too small angles the expression (1.4) becomes equal to the Rutherford cross section, where $f(x) = (1/2x)$. The equations (1.3) and (1.4) are used in the following in order to get first estimates of radiation effects.

Some reservations should be made in connection with the cross section (1.4) and the accompanying curve on Fig. 2. First, at high energies $\varepsilon > \varepsilon_1$, the curve on Fig. 2 is not very accurate at small angles, because the screening of the potential is reduced, the ion being stripped of most of its electrons. However, since at these energies most of the scattering is Rutherford scattering anyway, no major error is committed.

Second, a more interesting correction is due to the circumstance that for large angle ion-atom scattering a considerable energy is spent in electron excitation or ejection. This was observed by FEDORENKO and also by EVERHART and co-workers (cf. FEDORENKO (1959)). The result is that such collisions are not elastic, and that there is a correlation between nuclear collisions and electron excitation. Although ap-

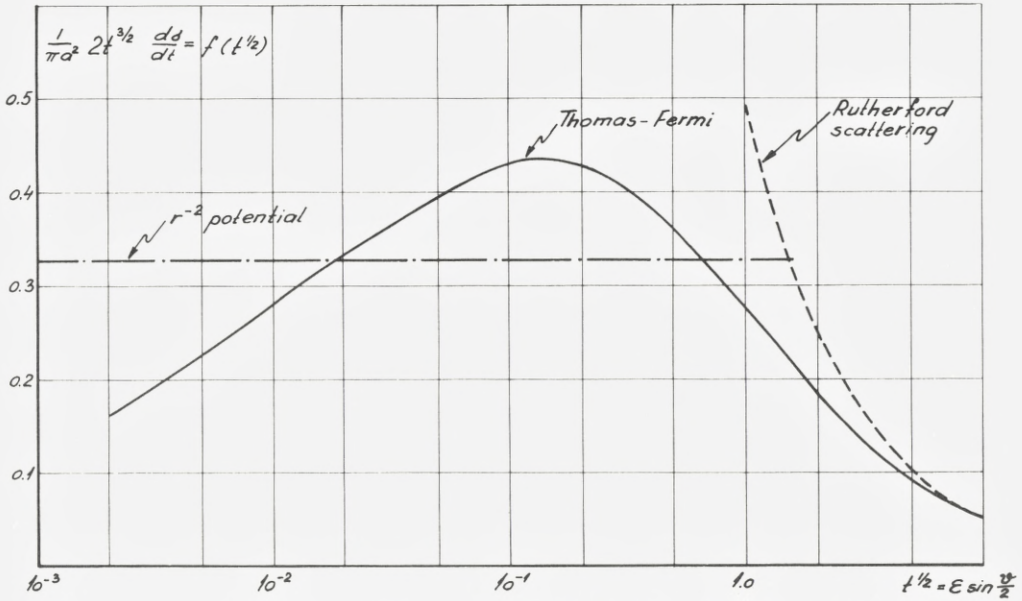

Fig. 2. Universal differential scattering cross section for elastic collisions, (1.4), based on a Thomas-Fermi type potential. At high values of $t^{1/2}$ it joins smoothly the Rutherford scattering. The cross section corresponding to power law scattering (1.3) with $s = 2$ is also shown.

proximate formulas may be quoted for the cross sections of such quasi-elastic collisions, the gain in generality hardly outweighs the complications due to the extra parameters in the treatment. Since the changes in our final results are presumably small (cf. p. 15), it seems preferable to verify at first the gross features of the simple formulas quoted above.

The general considerations in this introduction suggest a definite line of approach. It seems natural to develop first a formal theory of average damage effects, and to consider basic cases ($Z_1 = Z_2$) and possible simplifications, keeping in mind the main characteristics of the above cross sections. In this connection, the theory of fluctuations and of probability distributions should also be given. We therefore treat these general topics in § 2 and § 3. A direct application of the above cross sections to basic cases may then be made, first by analytical methods (§ 4) and next by numerical computations (§ 5). As an illustration of more complicated cases we consider a few examples, which also have bearing on experimental results (§ 6).

§ 2. The Basic Integral Equation

We shall now formulate and discuss the basic integral equation. The discussion, admittedly, is elaborate, but it seems profitable to make clear the contents of each assumption or approximation. We consider at first damage effects which are additive when due to independent events, so that e.g. saturation effects are excluded. The basic equation will be formulated in rather general terms, but immediate simplifications must be made when we treat solutions of actual cases. We study primarily the case where the particle belongs to the medium, and where the medium contains only one atomic species. When this case is solved, we may turn to equations for more complicated situations. For the present, we consider the simple case of average damage effects. Other averages, and the probability distribution in damage, will be discussed below.

We are concerned with a particle belonging to the medium, i.e. $Z_1 = Z_2$ (and $A_1 = A_2$). The particle has the energy E. We consider some unspecified physical quantity, φ, such as the number of ion pairs in a gas, the number of vacancies in a crystal, the energy given to electrons, etc. The quantity φ is arbitrarily taken to be zero before irradiation. The final average value of φ, after irradiation by a particle of energy E, we call $\overline{\varphi}(E)$. Although we use this simplified notation, the quantity $\overline{\varphi}$ depends not only on E, but also on Z_2 (and A_2), and to some extent on the physical state of the medium. Further, the physical quantity may be changed later by recombination processes, like in the case of ion pairs, but we shall disregard recombination effects and consider only the intermediate stage before recombination. In practice, recombination may be either avoided or accounted for separately. It is important that the physical quantity $\overline{\varphi}(E)$ in question is additive, i.e. for each separate slowing-down process all particles set in motion contribute additively to φ. This could hold for the three examples mentioned above.

The quantity $\overline{\varphi}(E)$ for the particle with energy E we may express in another way, if we suppose that the particle moves a path length dR in the medium with N atoms per unit volume. There is then a probability $NdRd\sigma_{n,e}$ for a collision specified by energy transfer T_n to the mass centre of the struck atom, together with energy transfer T_{ei} to electrons (electrons labelled by suffix i). The collision reduces the ion energy to the value $E - T_n - \sum_i T_{ei}$, i.e. the ion will now have a $\overline{\varphi}$-value equal to $\overline{\varphi}(E - T_n - \sum_i T_{ei})$. At the same time the struck atom gets the $\overline{\varphi}$-value $\overline{\varphi}\,(T_n - U)$, where U is the energy

wasted in disrupting the atomic binding. Finally, the electrons produced are described by another φ-function, which we denote as φ_e, and their contribution to $\overline{\varphi}$ after the collision in question is then $\sum_i \overline{\varphi}_e(T_{ei} - U_i)$, where U_i are the corresponding ionization energies. The above probability times the total $\overline{\varphi}$-value after the collision gives the contribution of this collision to $\overline{\varphi}(E)$. Afterwards we integrate over all collisions. There is left a probability $1 - NdR \int d\sigma_{n,e}$ that no collisions occur; in this event the $\overline{\varphi}$-value remains $\overline{\varphi}(E)$.

Collecting the above contributions we may write the original $\overline{\varphi}(E)$ as

$$\overline{\varphi}(E) = NdR \int d\sigma_{n,e} \left\{ \overline{\varphi}(E - T_n - \sum_i T_{ei}) + \overline{\varphi}(T_n - U) + \sum_i \overline{\varphi}_e(T_{ei} - U_i) \right\} + \\ + \left(1 - NdR \int d\sigma_{n,e}\right) \overline{\varphi}(E),$$

which leads to the basic integral equation

$$\int d\sigma_{n,e} \left\{ \overline{\varphi}\left(E - T_n - \sum_i T_{ei}\right) - \overline{\varphi}(E) + \overline{\varphi}(T_n - U) + \sum_i \overline{\varphi}_e(T_{ei} - U_i) \right\} = 0 \qquad (2.1)$$

This equation may be said to state simply that the $\overline{\varphi}$-value of the particle before the collision is equal to the sum of the $\overline{\varphi}$-values of, respectively, the particle, the struck atom and the ejected electrons after the collision, averaged over the probability of occurrence of the individual processes.

It may be noted that there is no necessity for the total cross section $\int d\sigma_{n,e}$ to be finite, and thus we do not attempt to normalize the probability of the various events. The actual physical quantities entering are integrals of $d\sigma_{n,e}$ times quantities tending to zero as e.g. T_n, or faster. The cross sections quoted in § 1 do in fact diverge. Of course, if classical cross sections larger than the atomic size become important in the final results, it may not be possible to separate into collisions with single atoms.

In equation (2.1) we have tried to avoid unnecessary details of notation. Thus, in specifying $\overline{\varphi}$ for the incoming particle or for the struck atom we might include a dependence on the degree of ionization of the particle in question. We shall assume such specifications to be included if necessary, but the interpretation of $\overline{\varphi}(E)$, if there can be doubt about the state of ionization, would normally be that in $\overline{\varphi}$ the number of electrons carried by the ion is considered to be a function of the ion velocity, and equal to the average number of electrons on the ion at the velocity in question.

The solution $\overline{\varphi}(E)$ of the equation (2.1) can be found if $\overline{\varphi}_e$ is a known function. This is the case if $\overline{\varphi}$ represents e.g. the number of vacancies pro-

duced in a crystal, since electrons with moderate energies may not be able to produce vacancies because of their low momentum, and thus $\overline{\varphi}_e = 0$. However, in general there is an additional integral equation describing $\overline{\varphi}_e$. For an electron of energy E the differential cross section is denoted by $d\sigma'_{n,e}$ and the collision results in a recoil atom of energy T_n and an energy transfer T_{ei} to atomic electrons. In analogy to (2.1) we immediately find

$$\int d\sigma'_{n,e} \left\{ \overline{\varphi}_e\left(E - T_n - \sum_i T_{ei}\right) - \overline{\varphi}_e(E) + \overline{\varphi}(T_n - U) + \sum_i \overline{\varphi}_e(T_{ei} - U_i) \right\} = 0. \quad (2.2)$$

Together with (2.1) the equation (2.2) leads to a solution for both $\overline{\varphi}$ and $\overline{\varphi}_e$. In equation (2.2) we may normally quite neglect the recoil of the nucleus; (2.2) then contains only $\overline{\varphi}_e$, and can be solved separately. An incoming electron usually gives only a small perturbation of the struck atom, and electron excitation may be separated into individual excitations. With a differential cross section $d\sigma'_e$ we thus find in all the simplified version of (2.2)

$$\sum_i \int d\sigma'_e \left\{ \overline{\varphi}_e(E - T_{ei}) - \overline{\varphi}_e(E) + \overline{\varphi}_e(T_{ei} - U_i) \right\} = 0. \quad (2.2')$$

Equation (2.1), supplemented by (2.2) if necessary, describes the simplest situation. It may be useful to comment on the set of integral equations belonging to other and more complicated cases. We give only a summary treatment, since the generalizations to be made are fairly obvious.

Firstly, if $Z_1 \neq Z_2$ we denote by $\overline{\varphi}_1(E)$ the average physical effect produced by particle 1. The equation for $\overline{\varphi}_1(E)$ is obtained in the same way as (2.1)

$$\int d\sigma_{1\,n,e} \left\{ \overline{\varphi}_1\left(E - T_n - \sum_i T_{ei}\right) - \overline{\varphi}_1(E) + \overline{\varphi}(T_n - U) + \sum_i \overline{\varphi}_e(T_{ei} - U_i) \right\} = 0, \quad (2.3)$$

where $d\sigma_{1\,n,e}$ is the differential cross section for collisions between particle 1 and the atom 2. Evidently, (2.3) requires that the solution of (2.1) is known. In this sense, equation (2.3) is secondary to (2.1); this applies also when we wish to compare experiments and basic collision theory. It is interesting to notice that (2.3), in contrast to (2.1), is not a typical integral equation; if $T_n + \sum_i T_{ei}$ is small, (2.3) becomes a differential equation.

Secondly, the substance may contain more than one atomic element. Then, primary cases are those where the incoming particle is one of the atoms in the substance, and the function $\overline{\varphi}^{(j)}(E)$ belongs to the case where the incoming particle is equal to the j'th atomic species of the substance. In place of (2.1) and (2.2) we now write generally

$$\sum_{j=1}^{m+1} \int dx\, S_{kj}(E,x)\, \overline{\varphi}^{(j)}(x) = 0, \qquad k = 1, 2, \ldots, m+1, \quad (2.4)$$

where m is the number of atomic elements in the substance, and $\overline{\varphi}^{(1)} \ldots \overline{\varphi}^{(m)}$ are the $\overline{\varphi}$-functions of these elements, while $\overline{\varphi}^{(m+1)}(E)$ represents $\overline{\varphi}_e(E)$. The integral

operator S_{kj} is associated with collisions between the k'th element, of energy E, and the j'th element at rest. As an example, we quote the basic case (2.1) and (2.2), where $m+1 = 2$, and e.g. $S_{21} = \int d\sigma'_{n,e}\, \delta(x - T_n + U)$, according to (2.2).

Let us now return to the basic integral equation (2.1) and discuss the approximations which might be made in solving the equation. It is useful to classify these approximations; roughly, they may be divided into five types.

Discussion of approximations.

The first approximation, (*A*), was introduced above. It consists in assuming that the electrons do not produce recoil atoms with appreciable energies. This is usually quite correct and implies that $\overline{\varphi}_e$ may be obtained separately, i.e. (2.2) simplifies into (2.2′). (*A*) is therefore normally fulfilled. An interesting exception occurs if the disruption of atomic bindings has significant influence on the measured effects (cf. footnote on page 4). A more straightforward exception is the case of incoming electrons of energies so high ($\gtrsim 1\,MeV$) that in violent collisions bound atoms can be directly dislodged. In the following, approximation (*A*) is always used.

The second approximation, (*B*), consists in neglecting the atomic binding term U in (2.1) so that $\overline{\varphi}(T_n - U)$ is replaced by $\overline{\varphi}(T_n)$. Since the bindings are of order of some eV, we are normally quite justified in neglecting U, for heavy particles at energies where the electronic stopping has any influence at all on the events. Approximation (*B*) is used everywhere in the following, if not directly otherwise stated.

At this stage it may be of interest to mention cases where (*B*) is invalid. In fact, if the binding energies contribute to (2.1) in a significant way, the particle energy E is not exceedingly large compared to the binding term U. This implies, on the other hand, that the electronic stopping is small and may be neglected. The approximation may be called (B^{-1}), and we then obtain the simplified equation

$$\int d\sigma_n \left\{\overline{\varphi}(E - T_n) - \overline{\varphi}(E) + \overline{\varphi}(T_n - U)\right\} = 0, \tag{2.5}$$

where $d\sigma_n$ is the differential cross section for elastic ion-atom collisions. This equation is essentially that used by Snyder and Neufeld (1955), and by other authors. It should be noted that the binding term U is introduced in a rather symbolic way. A thorough study demands a detailed description of the mechanism by which an atom in a lattice may be removed from its environment. Thus, beside the energy wasted irreversibly, U, when an atom is quickly removed, there is e.g. the threshold energy for adiabatic removal of the atom. The generalization of (2.5) to a substance containing several different atoms in various binding states should be obvious from (2.4). Note also that the approximation (*E*), introduced below, may be useful in studies of (2.5).

The third approximation, (*C*), is to assume that the energy transfers T_{ei} to electrons are small in a relative measure, or $T_{ei} \ll E - T_n$. Like (*B*) this approximation should hold quite well if the particle energy is not too low. In fact, we have approximately at high velocities $T_{ei} \sim E$ times electron mass divided by ion mass. In all, (*C*) applied to (2.1) leads to

$$\left.\begin{aligned} \int d\sigma_{n,e} \Big\{ \bar{\varphi}(E - T_n) - \bar{\varphi}(E) - \bar{\varphi}'(E - T_n) \sum_i T_{ei} + \\ + \bar{\varphi}(T_n) + \sum_i \bar{\varphi}_e(T_{ei} - U_i) \Big\} = 0 , \end{aligned} \right\} \tag{2.6}$$

where approximation (*B*) is also included. Like the two previous approximations, approximation (*C*) is used generally in the following, exceptions being clearly stated.

The fourth approximation, (*D*), is separation of nuclear and electronic collisions. The idea is that only a negligible part of the electronic excitation occurs at the small impact parameters where nuclear collisions play a role. In point of fact, most of the electronic excitations are associated with large impact parameters. It is then natural to disregard the slight overlap of the two types of collision effects, and (2.6) becomes

$$\bar{\varphi}'(E) \cdot S_e(E) = \int d\sigma_n \Big\{ \bar{\varphi}(E - T_n) - \bar{\varphi}(E) + \bar{\varphi}(T_n) \Big\} + \int d\sigma_e \sum_i \bar{\varphi}_e(T_{ei} - U_i), \tag{2.7}$$

where $d\sigma_n$ is the differential cross section for elastic nuclear collisions. $S_e(E) = \int d\sigma_e \sum_i T_{ei}$ is the electronic stopping cross section, $d\sigma_e$ being the differential cross section for energy transfers $T_{e1}, T_{e2}, \ldots, T_{ei}, \ldots$ to the individual electrons.

Approximation (*D*), as expressed by (2.7), is also used widely in the following. It contains a definite assumption, the justification of which is less apparent and less justified than the previous assumptions. In (2.7) we have disregarded the connection between electronic and nuclear collisions; they are even supposed to be separable. From a series development in (2.6) we find that the term neglected on the right hand side of (2.7) is approximately $\bar{\varphi}''(E) \int d\sigma_{n,e} T_n \sum_i T_{ei}$. It is of interest to investigate the justification of (2.7) using such correction terms.

In making approximation (*D*) we include approximation (*C*). This is reasonable since it implies only that $\bar{\varphi}(E) - \bar{\varphi}\left(E - \sum_i T_{ei}\right) = \bar{\varphi}'(E) \sum_i T_{ei}$. The

correction for this approximation is therefore $\frac{1}{2}\bar{\varphi}''(E)\int d\sigma_{n,e}\left(\sum_i T_{ei}\right)^2$ on the left hand side of (2.7), but is presumably not large.

Finally, the fifth approximation, (E), is to assume that also T_n is small compared to the energy E. Since the maximum energy transfer is normally quite large, and even equal to E if $A_1 = A_2$, it might seem that this approximation is poor. However, because the cross sections are strongly forward peaked, the approximation remains fairly good, as we shall see in § 4. Approximation (E), together with the previous simplifications, leads to

$$\bar{\varphi}'(E)\{S_e(E)+S_n(E)\} = \int d\sigma_n \bar{\varphi}(T_n) + \int d\sigma_e \sum_i \bar{\varphi}_e(T_{ei}-U_i), \qquad (2.8)$$

where $S_n(E) = \int d\sigma_n T_n$, and where the quantity neglected, as compared to (2.7), is approximately $(1/2)\,\bar{\varphi}''(E)\cdot\int d\sigma_n T_n^2$ on the right hand side of (2.8). The approximation (E) may be regarded as an expedient to get an approximate solution of (D), i.e. (2.7).

An interesting consequence of approximation (E) may be noticed. Thus, if we disregard (D), and use only (E), i.e. T_n and $\sum_i T_{ei}$ are small, we obtain again equ. (2.8), but now $S_n = \int d\sigma_{n,e} T_n$, $S_e = \int d\sigma_{n,e} \sum_i T_{ei}$. Further, the cross sections on the right of (2.8) should be $d\sigma_{n,e}$. The separation in (2.8) is therefore obtained independently of the separability of nuclear and electronic collisions assumed in (D). Conversely, it can be difficult to relate the integral equations for $\bar{\varphi}$ to the degree of correlation between electronic and nuclear collisions, as referred to in § 1, p. 9. In Fig. 6, the good agreement between approximations (D) and (E) indicates that correlation corrections to $\bar{\nu}(E)$ can not be large.

We shall sometimes use an approximation, (E'), which is much closer to (D) than (E) itself

$$\left.\begin{aligned} -\frac{1}{2}\bar{\varphi}''(E)\,\Gamma_n(E) + \bar{\varphi}'(E)\{S_e(E)+S_n(E)\} = \\ = \int d\sigma_n \bar{\varphi}(T_n) + \int d\sigma_e \sum_i \bar{\varphi}_e(T_{ei}-U_i), \end{aligned}\right\} \qquad (2.8')$$

where $\Gamma_n(E) = \int d\sigma_n T_n^2$.

When $\bar{\varphi}_e$ is determined by an equation like (2.2′) it only enters as a known source term in the basic integral equation (2.1). Clearly, the primary problem is then to find the complete solution of the homogeneous basic equation, i.e. omitting the $\bar{\varphi}_e$-term, in one of its formulations within the approximations (A) to (E).

It would be vain to ask for a detailed knowledge of $d\sigma_{n,e}$, let alone solve the equation (2.1) on this basis. However, from equations (2.7) and (2.8) it is seen that knowledge of the stopping cross sections S_e and S_n as functions of energy is essential to the solution of the basic integral equation. Apart from this, some knowledge of $d\sigma_n$ as a function of T_n is clearly required. This is seen in all versions of the basic integral equation, where the term $\int d\sigma_n \overline{\varphi}(T_n)$ always enters.

It need hardly be added that in the following we introduce approximations other than those listed above. Most of the approximations are connected with Thomas-Fermi-like properties or with the specific behaviour of the cross sections summarized in § 1. An example of general interest is the attempt to formulate asymptotic equations in the high energy limit, cf. (5.3) and (5.4).

§ 3. Fluctuations and Probability Distribution

Fluctuations.

So far, we have considered the average, $\overline{\varphi}(E)$, of an additive physical quantity, φ. However, it is of interest to discuss also other averages, for instance the average of the square of the physical quantity. In general, we might consider $<\varphi^m(E)>$, by which is meant the average over all events of the m'th power of φ, so that $<\varphi(E)> \equiv \overline{\varphi}(E)$. The equation governing $<\varphi^m(E)>$ is obtained in a similar way as (2.1), and we find in analogy to (2.1)

$$\int d\sigma_{n,e} \Big\{ - <\varphi^m(E)> + <\Big[\varphi\Big(E - T_n - \sum_i T_{ei}\Big) + \varphi(T_n) + \sum_i \varphi_e(T_{ei} - U_i)\Big]^m> \Big\} = 0 . \tag{3.1}$$

In principle, (3.1) may be used to construct the average of any function $f(\varphi)$, e.g. by means of a power series development in φ. In practice, however, it is preferable to study instead the equation for the probability distribution in φ, $P(\varphi, E)$. A brief discussion of the probability distribution is given below.

How ever this may be, it is always of considerable interest to treat the case of $m = 2$ in (3.1). This case indicates how equations of type of (3.1) may be solved, and gives at the same time the average square fluctuation in φ. We therefore put $m = 2$ in (3.1) and average over independent quantities like e.g. the product $<\varphi(E - T_n - \sum_i T_{ei})\varphi(T_n)> = \overline{\varphi}(E - T_n - \sum_i T_{ei})\overline{\varphi}(T_n)$,

where we average over the subsequent fate of two atoms of given energies, $E - T_n - \sum_i T_{ei}$ and T_n. We get thus

$$\left.\begin{aligned} &\int d\sigma_{n,e}\left\{\Omega_\varphi^2\left(E - T_n - \sum_i T_{ei}\right) - \Omega_\varphi^2(E) + \Omega_\varphi^2(T_n) + \sum_i \Omega_{\varphi e}^2(T_{ei} - U_i)\right\} = \\ &= -\int d\sigma_{n,e}\left\{\left[\bar{\varphi}\left(E - T_n - \sum_i T_{ei}\right) + \bar{\varphi}(T_n) + \sum_i \bar{\varphi}_e(T_{ei} - U_i)\right]^2 - \bar{\varphi}^2(E)\right\}, \end{aligned}\right\} \quad (3.2)$$

where we have introduced the average square straggling $\Omega_\varphi^2(E) = <\varphi^2(E)> - \bar{\varphi}^2(E)$, and $\Omega_{\varphi e}^2(E) = <\varphi_e^2(E)> - \bar{\varphi}_e^2(E)$.

The right hand side of (3.2) may be reformulated by means of (2.1), and we obtain

$$\left.\begin{aligned} &\int d\sigma_{n,e}\left\{\Omega_\varphi^2(E) - \Omega_\varphi^2(T_n) - \Omega_\varphi^2\left(E - T_n - \sum_i T_{ei}\right) - \sum_i \Omega_{\varphi e}^2(T_{ei} - U_i)\right\} = \\ &= \int d\sigma_{n,e}\left\{\bar{\varphi}\left(E - T_n - \sum_i T_{ei}\right) - \bar{\varphi}(E) + \bar{\varphi}(T_n) + \sum_i \bar{\varphi}_e(T_{ei} - U_i)\right\}^2. \end{aligned}\right\} \quad (3.3)$$

This is the integral equation which governs the straggling in φ, and it corresponds to the equation (2.1) describing $\bar{\varphi}$ itself. Also in a more formal respect (3.3) is similar to (2.1). In fact, if the right hand side of (3.3) could be neglected, the resulting equation for the quantity Ω_φ^2 would be exactly (2.1). Now, the right hand side of (3.3) is a positive source term completely determined by the known functions $\bar{\varphi}$ and $\bar{\varphi}_e$. It contains the square of a term whose average is zero, being the square of the change in $\bar{\varphi}$ in a collision, averaged over the different results of the first collision.

We shall not quote the separate equation for $\Omega_{\varphi e}^2(E)$, in analogy to (2.2) or (2.2′), since it would be of type of (3.3) and could be written down immediately. Moreover, simplifications in (3.3), corresponding to the approximations (*A*) to (*E*), are fairly straightforward. We consider explicitly only a few cases. Suppose that energy transfers to electrons are small, and that nuclear and electronic collisions are separable. This corresponds to approximation (*D*). In the cases where $\bar{\varphi}_e$ is zero we then get, in analogy to (2.7),

$$\left.\begin{aligned} S_e(E) \cdot \frac{d\Omega_\varphi^2(E)}{dE} &= \int d\sigma_n \left\{\Omega_\varphi^2(E - T_n) - \Omega_\varphi^2(E) + \Omega_\varphi^2(T_n)\right\} + \\ &+ \int d\sigma_n \left\{\bar{\varphi}(E - T_n) - \bar{\varphi}(E) + \bar{\varphi}(T_n)\right\}^2, \end{aligned}\right\} \quad (3.4)$$

where also the term $(\bar{\varphi}'(E))^2 \int d\sigma_e \left(\sum_i T_{ei}\right)^2$ is disregarded.

Assume here that T_n in (3.4) is small, i.e. approximation (E). From (3.4) we obtain, corresponding to the homogeneous part of (2.8),

$$\{S_n + S_e\} \frac{d\Omega_\varphi^2(E)}{dE} = \int d\sigma_n \Omega_\varphi^2(T_n) + \int d\sigma_n \{\bar{\varphi}(T_n) - T_n \bar{\varphi}'(E)\}^2. \qquad (3.5)$$

Although (3.5) appears to be simpler than (3.4), we shall find in § 5 that in a straightforward case equ. (3.4) has the advantage of simplicity.

Let us consider for a moment what kind of changes will result in (3.5), if approximation (D) is dropped and only (E) and (C) are kept. Then, T_n and $\sum_i T_{ei}$ are small, but a correlation between electronic and nuclear collisions remains. According to (3.3), all cross sections in (3.5) must be replaced by $d\sigma_{n,e}$, but moreover the term $(\bar{\varphi}(T_n) - T_n \bar{\varphi}'(E))^2$ on the right changes into $\left(\bar{\varphi}(T_n) - \bar{\varphi}'(E)\left[T_n + \sum_i T_{ei}\right]\right)^2$, and for this reason the effect of correlations can be distinguished. In this respect (3.5) differs from the corresponding equation (2.8), where we also discussed omission of approximation (D).

Corresponding to the equation (2.3) for $\bar{\varphi}_1(E)$, we shall also discuss the straggling in the case of $Z_1 \neq Z_2$. The average square straggling in φ_1 is denoted as $\Omega_{\varphi 1}^2(E)$. We consider again the case where φ_e does not contribute. Using approximation (D) an equation analogous to (3.4) is obtained

$$\left.\begin{aligned} S_{1e} \cdot \frac{d}{dE} \Omega_{\varphi 1}^2(E) = \int d\sigma_{1n} \{\Omega_{\varphi 1}^2(E - T_n) - \Omega_{\varphi 1}^2(E) + \Omega_\varphi^2(T_n)\} + \\ + \int d\sigma_{1n} \{\bar{\varphi}_1(E - T_n) - \bar{\varphi}_1(E) + \bar{\varphi}(T_n)\}^2, \end{aligned}\right\} \qquad (3.6)$$

where $\bar{\varphi}(E)$ is given by (2.7), $\Omega_\varphi^2(E)$ by (3.4) and $\bar{\varphi}_1(E)$ by (2.3) in approximation (D), while $d\sigma_{1n}$ is the differential nuclear cross section for collisions between the particle 1 and an atom 2. Further, S_{1e} is the electronic stopping cross section per atom for the particle 1 passing atoms 2. It is seen that (3.6) contains (3.4) as a special case. In (3.6), terms of type of $\left(\sum_i T_{ei}\right)^2$ are omitted.

Finally, we apply the approximation (E) to (3.6), i.e.

$$(S_{1e} + S_{1n}) \frac{d}{dE} \Omega_{\varphi 1}^2(E) = \int d\sigma_{1n} \Omega_\varphi^2(T_n) + \int d\sigma_{1n} \{\bar{\varphi}(T_n) - T_n \bar{\varphi}_1'(E)\}^2, \qquad (3.7)$$

where $\bar{\varphi}$, Ω_φ^2 and $\bar{\varphi}_1$ should be given in approximation (E) too. Note that (3.7) is a differential equation in the variable $\Omega_{\varphi 1}^2$, and may be integrated readily.

Probability distribution.

We have now studied average quantities, $\overline{\varphi}\,(E)$, described by rather simple equations, as well as fluctuations, $\Omega^2(E)$, which obey more elaborate equations. These are the first two steps in a series development, where successive moments $<\varphi^n>$ are calculated. The series development is convenient if the first moments give adequate information, since they may be calculated with comparative ease. Often, further information is needed. When the value of a series development becomes doubtful, a closed equation for the probability distribution itself is much to be preferred. Other approximation methods are then at our disposal.

It is thus of both theoretical and practical interest to study the probability distribution itself. We shall merely formulate the basic equations. Let us then ask for the equation analogous to (2.1), where one considers the effect of an incoming particle with energy E, and identical with the atoms in the medium. Introduce probability distributions $P(\varphi, E)$ and $P_e(\varphi, E)$ representing the probabilities that, respectively, the particle and an electron having energy E will produce the damage effect φ. Therefore, e.g. $\int_o^\infty \varphi P_e(\varphi, E)\, d\varphi = \overline{\varphi}_e(E)$ is the average effect produced by an electron of energy E. The equation governing $P(\varphi, E)$ is derived in the same way as (2.1), making the same assumptions. We find readily

$$\left.\begin{aligned} \int d\sigma_{n,e} P(\varphi, E) = \int d\sigma_{n,e} \int_o^\infty d\varphi' \int_o^\infty d\varphi'' \prod_j \int_o^\infty d\varphi_j P_e(\varphi_j, T_{ej} - U_j) \cdot \\ \cdot P\left(\varphi', E - T_n - \sum_i T_{ei}\right) \cdot P(\varphi'', T_n - U) \cdot \delta\left(\varphi - \varphi' - \varphi'' - \sum_i \varphi_i\right). \end{aligned}\right\} \tag{3.8}$$

The equation states that the probability for the value φ prior to the collision is equal to the product of the individual probabilities belonging to ejected particles, when averaged over the frequency of occurrence of the different events. There is an integration over all possible φ-values of the ejected particles, with the condition that their sum is equal to the original φ-value, as expressed by the δ-function. Thus, (3.8) assumes independent behaviour of the separate events, i.e. product of P's, and additivity of damage effect, i.e. $\varphi = \varphi' + \varphi'' + \sum_i \varphi_i$.

Equ. (3.8) determines $P(\varphi, E)$ and P_e is considered as a known function. If (3.8) is multiplied by φ and integrated over φ from 0 to ∞, equ. (2.1) results.

There is a similar equation for $P_e(\varphi, E)$. We write it down assuming for simplicity that electrons produce no atomic recoils (approximation (A) and equ. (2.2′))

$$\left.\begin{aligned} \int d\sigma'_e P_e(\varphi, E) = \int d\sigma'_e \int_o^\infty d\varphi' \prod_j \int_o^\infty d\varphi_j P_e(\varphi_j, T_{ej} - U_j) \cdot \\ \cdot P_e\left(\varphi', E - \sum_i T_{ei}\right) \cdot \delta\left(\varphi - \varphi' - \sum_i \varphi_i\right), \end{aligned}\right\} \tag{3.9}$$

where $d\sigma'_e$ is the differential cross section for transfer of energy T_{ei} to atomic electrons by an electron of energy E. There are further simplifications, if we take into account that an electron normally ejects at most one atomic electron in a collision.

In (3.8) let us assume that electrons do not contribute to the damage effect in question, i.e. $P_e(\varphi,E) = \delta(\varphi)$. In approximation (D) we then get, since $P(\varphi,0) = \delta(\varphi)$,

$$\left.\begin{aligned} S_e(E)\frac{\partial}{\partial E}P(\varphi,E) + \int d\sigma_n P(\varphi,E) = \\ = \int d\sigma_n \int_0^\infty d\varphi' \int_0^\infty d\varphi'' P(\varphi', E - T_n) P(\varphi'', T_n) \delta(\varphi - \varphi' - \varphi''). \end{aligned}\right\} \quad (3.10)$$

The bond expressed by the δ-function can be inconvenient. It is natural to introduce Laplace transforms of the probability distribution,

$$\tilde{P}(\lambda,E) = \int_0^\infty d\varphi P(\varphi,E) e^{-\lambda\varphi}.$$

The Laplace transforms are particularly useful because of the additivity of φ. From (3.8) we obtain the alternative version

$$\left.\begin{aligned} \int d\sigma_{n,e}\tilde{P}(\lambda,E) = \int d\sigma_{n,e}\tilde{P}\left(\lambda, E - T_n - \sum_i T_{ei}\right)\tilde{P}(\lambda, T_n - U)\cdot \\ \cdot \prod_j \tilde{P}_e(\lambda, T_{ej} - U_j), \end{aligned}\right\} \quad (3.11)$$

and if $\overline{\varphi}_e$ is zero we have $\tilde{P}_e(\lambda,x) \equiv 1$.

§ 4. Analytical Approximations in Homogeneous Equation

The first step towards a solution of (2.1), or its simplified versions, is to discuss its homogeneous part, i.e. put $\varphi_e = 0$. Now, it so happens that the quantity $\overline{\nu}(E)$, introduced in § 1 and described as the average energy transfer to atomic motion, is normally a solution of the homogeneous part of equation (2.1), because the energy transfer from electrons to atoms is negligible to nearly all purposes. By solving the homogeneous equation, we have therefore found one important physical property in slowing-down processes. In the following, the normal boundary condition on $\overline{\nu}(E)$ is $\overline{\nu}(E)/E \to 1$ for $E \to 0$, and thus $\overline{\eta}(E)/\overline{\nu}(E)$ vanishes in this limit.

It is necessary to gain some experience concerning solutions of the integral equation. To this end we consider at first analytical solutions using simplified approximations to cross sections; this can be of interest particularly at the lower energies. Secondly, in § 5 we solve the integral equations numerically with more accurate cross sections, using electronic computations. We are then led to new asymptotic or approximate solutions, which may be checked by the numerical and analytical results. The present

chapter may therefore be regarded as an exercise preliminary to the more precise treatment in § 5.

The simplest results obtain when we suppose that the differential cross section $d\sigma_n$ may be approximated by the power law scattering formula (1.3), corresponding to a potential proportional to r^{-s}. We can then arrive at analytical solutions of the various approximations to the integral equation. Let us start from approximation (E), i.e. (2.8). The homogeneous equation (2.8) for $\bar{\nu}$ becomes

$$(S_e + S_n) \cdot \bar{\nu}'(E) = \int_{T=o}^{E} \frac{d\sigma_n}{dT}\, dT\, \bar{\nu}(T), \tag{4.1}$$

where $\bar{\varphi}$ has been replaced by $\bar{\nu}$. We introduce (1.3) in (4.1), and multiply by $S_n^{-1}\, E^{1-1/s}$. Differentiating with respect to E we get a differential equation of second order in place of the integral equation (4.1),

$$(\xi(E)+1)\, E^2 \bar{\nu}'' + \left\{ E\xi'(E) + \left(1 - \frac{1}{s}\right)(1+\xi(E)) \right\} E\bar{\nu}' - \left(1 - \frac{1}{s}\right)\bar{\nu} = 0, \tag{4.2}$$

where $\xi(E) = S_e(E)/S_n(E)$. It is apparent that a differential equation was obtained from the original integral equation only because of the simple behaviour of the cross section (1.3), where the dependence of $d\sigma_n$ on E could be separated out as a factor.

Corresponding to (1.2) we shall assume that $S_e \propto E^{1/2}$, and since $S_n \propto E^{1-2/s}$ we get $\xi(E) \propto E^{2/s-1/2}$. It then turns out that the solutions of (4.2) are hypergeometric functions, of the kind $F(a, b;\ a+b;\ x)$, cf. Erdélyi et al. (1953). The complete solution of (4.2) is seen to be

$$\left.\begin{aligned} \bar{\nu}(E) = {} & C_1 E \cdot F\left(\frac{2s}{4-s}, \frac{s+2}{4-s}; \frac{3s+2}{4-s}; -\xi(E)\right) + \\ & + C_2 E^{\frac{1-s}{s}} \cdot F\left(\frac{2-2s}{4-s}, \frac{4-3s}{4-s}; \frac{6-5s}{4-s}; -\xi(E)\right), \end{aligned}\right\} \tag{4.3}$$

where C_1 and C_2 are arbitrary constants.

If we ask for the particular solution given by the normal boundary condition for $\bar{\nu}$ at $E = 0$, i.e. $\bar{\nu}(E)/E \to 1$ for $E \to 0$, we obtain $C_1 = 1$, $C_2 = 0$, if $s < 4$. Note that only for $s < 4$ does the present $\xi(E)$ tend to zero for $E \to 0$, and that this is the proper behaviour of $\xi(E)$.

If instead of (2.8) we start from the more correct equation (2.7), the cross section (1.3) is seen to lead to the equation

$$\frac{d}{dE}\left\{E^{1-1/s}\,\bar{\nu}'(E)\,\xi(E)\right\} = -\bar{\nu}''(E)\cdot E^{1-1/s} + \\ +\left(1-\frac{1}{s}\right)\int_0^E \frac{dT}{T^{1+1/s}}\cdot\left\{\bar{\nu}'(E-T)-\bar{\nu}'(E)+\bar{\nu}''(E)\,T\right\}.$$

The integrand on the right is large only for $T \approx E$. Making an underestimate of the integral (because $\bar{\nu}''(E)$ is always negative and increases with E) we then replace $T^{-1-1/s}$ by $E^{-1-1/s}$ in the integral. This gives the differential equation

$$E^2\cdot\bar{\nu}''\cdot\left(\xi+\frac{1}{2}+\frac{1}{2s}\right)+E\cdot\bar{\nu}'\cdot\left\{E\,\xi'+\left(1-\frac{1}{s}\right)(\xi+1)\right\}-\left(1-\frac{1}{s}\right)\bar{\nu}=0. \qquad (4.4)$$

This equation differs only little from (4.2), but is an underestimate of $\bar{\nu}$, as compared with the precise solution of (2.7) and (1.3). It is interesting that $\bar{\nu}$ from (4.2) is instead an overestimate of the solution of (2.7) and (1.3); this follows from $\bar{\nu}'(E)$ being a decreasing function of E. We have thus bracketed the solution of (2.7) between two approximate solutions. It turns out that (4.4) is generally a somewhat better approximation than (4.2). The solutions of (4.4) are seen to be hypergeometric functions, of the type

$$E\cdot F\left(\frac{2s}{(4-s)},\ \frac{s+2}{(4-s)};\ \frac{(4+s)\alpha+2s-2}{(4-s)\alpha};\ -\frac{\xi(E)}{\alpha}\right) \quad \text{and}$$

$$E^{\frac{1-s}{s\alpha}}\cdot F\left(\frac{2-2s}{(4-s)\alpha},\ \frac{(2-s)\alpha+2-2s}{(4-s)\alpha};\ \frac{(4-3s)\alpha+2-2s}{(4-s)\alpha};\ -\frac{\xi(E)}{\alpha}\right)$$

where $\alpha = \frac{1}{2}+\frac{1}{2s}$ is the coefficient of $\bar{\nu}''$ in the brackets in (4.4). The present solutions of (4.4) are similar to (4.3), and contain it as special case ($\alpha = 1$).

Region I. In region I, where $0 < E \lesssim E_c$, we may select a few suitable values of s, and study some of the approximate solutions. In doing this, we obtain not only a reasonable estimate of $\bar{\nu}(E)$, but also an insight in errors involved in some of the simplifications, (A) to (E), of the basic integral equation.

Let us consider the standard case, where $s = 2$ and $S_n = S_n^0$ is independent of energy, cf. (1.1) and (1.3). We put $\xi(E) = (E/E_c)^{1/2}$, and obtain from (4.2) and (4.3), with the boundary condition $\bar{\nu}(E)/E = 1$ at $E = 0$,

$$\bar{\nu}(E) = E_c\left\{-12+6\left[1+2\,(E_c/E)^{1/2}\right]\cdot\log\left(1+(E/E_c)^{1/2}\right)\right\}, \qquad (4.5)$$

representing the solution of (2.8) – i.e. approximation (E) – for power law scattering with $s = 2$. The solution (4.5) can be used only at energies where E/E_c is somewhat less than unity. This limitation must be made because a

decrease sets in in the actual function S_n at an energy somewhat lower than E_c, in most cases.

Let us consider in particular the limit of $(E/E_c) \ll 1$, where a more general approach is possible. In fact, in any one of the approximations (D) to (E') we get, when $s = 2$, a power series in $(E/E_c)^{1/2}$

$$\bar{\eta}(E) = E - \bar{\nu}(E) = \alpha_1 E^{3/2} E_c^{-1/2} - \ldots \quad , \quad E \ll E_c, \tag{4.6}$$

where α_1 is a constant, the value of which depends on the approximation used. We compare four solutions of the case $s = 2$. Firstly, approximation (E) given by (4.5) leads to $\alpha_1 = 1$. Secondly, a series development of the solutions of the approximation (4.4) leads to $\alpha_1 = 16/13 = 1.23$. Thirdly, the more correct integral equation (2.7), i.e. approximation (D), may be solved by a series development, leading to $\alpha_1 = 4/(3\pi - 6) = 1.17$. These three values for α_1 give an indication of the accuracy of the various approximations. As expected, (cf. the discussion of (4.4)) the solution (4.5) is an overestimate and (4.4) an underestimate of $\bar{\nu}(E)$; (4.4) is a somewhat better approximation. A fourth case may be mentioned, i.e. approximation (E') given by equation (2.8'). It consists in including the next term in the series development of $\bar{\nu}(E - T_n) - \bar{\nu}(E)$, i.e. subtract $(1/2)\,\bar{\nu}''(E) \int d\sigma_n T_n^2$ on the left hand side of (4.1). We find here $\alpha_1 = 8/7 = 1.14$, so that approximation (E') is superior to (E).

Region II. In this region the function S_e remains the same, increasing as $E^{1/2}$. However, S_n begins to decrease and the scattering approaches the Rutherford scattering, though with a screening at a distance $\sim a$. For a qualitative orientation we again base our description on (1.3), so that we assume that S_n is proportional to a power of E, i.e. $E^{1-2/s}$. This approach is qualitatively less justified than in region I, but we can learn about the possible approximation methods for solving the basic integral equation.

Let us suppose that S_n is proportional to $E^{-1/2}$ for $E > E_o$, so that $s = 4/3$ in (1.3), and $\xi(E) = (S_e/S_n) = (E/E_b)$. Then, $E_b = (E_o E_c)^{1/2}$ is the energy at which the two stopping cross sections become equal. Equation (4.2) for $\bar{\nu}$ now becomes

$$(4\,E^3 E_b^{-1} + 4\,E^2)\,\bar{\nu}'' + (5\,E^2 E_b^{-1} + E)\,\bar{\nu}' - \bar{\nu} = 0, \tag{4.7}$$

with the complete solution (cf. (4.3))

$$\left.\begin{aligned}\bar{\nu} = C_1 \cdot 5\, E_b \Big\{ 1 - \frac{1}{4\sqrt{2}}\, \xi^{-1/4} \log \frac{\xi^{1/2} + \sqrt{2}\, \xi^{1/4} + 1}{\xi^{1/2} - \sqrt{2}\, \xi^{1/4} + 1} - \\ - \frac{1}{2\sqrt{2}}\, \xi^{-1/4} \operatorname{arctg} \frac{\sqrt{2}\, \xi^{1/4}}{1 - \xi^{1/2}} \Big\} + C_2\, \xi^{-1/4} E_b^{-1/4}.\end{aligned}\right\} \quad (4.8)$$

The solution is determined by the boundary conditions at the energy E_o, where we find $\bar{\nu}$ and $\bar{\nu}'$ from (4.5). Thus, C_1 is given by

$$C_1 = \frac{12}{5} x^{-2} + \frac{6}{5} (x+1)(x-2) x^{-3} \log (1+x) \simeq 1 - \frac{2}{5} x + \ldots, \quad (4.9)$$

where $x = E_o / E_b$ is less than unity for all values of $Z_1 = Z_2$. The expression for C_2 is more involved, but $C_2 < 0$. This shows that although $\bar{\nu}(E)$ increases with E, it remains below the value C_1, and it increases only slowly towards this limit. In region III, however, it turns out that $\bar{\nu}$ will go on increasing without an upper limit, although still quite slowly. The behaviour of (4.5) in region I and (4.8) in region II is shown by the dashed curve in Fig.s 5 and 6 for the case of $\varepsilon_c = 4.75$ ($k = 0.15$) and $x = 0.56$, where it is compared with a numerical estimate based on the more accurate scattering formula (1.4). It may be noted that the value of C_1 is not far from unity, and that C_2 is small. If we were to put $C_1 = 1$ and $C_2 = 0$, we would instead have the solution where $s = 4/3$, i.e. (4.7), is used down to zero energy, and apparently this is satisfactory as a rough estimate.

In region I the solution (4.3) of the equation (4.2) was an application of approximation (E) using the cross section (1.3). It might therefore seem that also in region II the equations (4.7) and (4.8) are equivalent to approximation (E). However, we change from one cross section $d\sigma_n$ in region I to another in region II. Since (4.2) and (4.7) were obtained by differentiation of (4.1), they should be supplemented by inhomogeneous terms if the cross section changes at low values of TE. This circumstance is disregarded in (4.7), (4.8) and (4.9), giving some deviation from (E).

Straggling in region I. An evaluation of the straggling in ν or η, $\Omega_\nu^2(E) = \Omega_\eta^2(E)$, from (3.3) is more involved than the estimate of $\bar{\nu}$ itself. Still, at low energies in region I, a series development may be made and the first term is readily obtained. If (1.3) is applied, it turns out that the relative straggling in η becomes a constant, independent of energy (and atomic number and mass) at low energies,

$$\sigma^2 = \frac{\Omega_\eta^2(E)}{\bar{\eta}^2(E)} = \text{const., for } \xi(E) \ll 1. \quad (4.10)$$

If we consider the standard case, $s = 2$ in (1.3) and $\xi = (E/E_c)^{1/2}$, we find at low energies that $\overline{\eta}$ is proportional to $E^{3/2}$, cf. (4.6). We solve (3.5) and get $\sigma^2 = 1/14$. We may also solve directly the more basic integral equation (3.4) for Ω_ν^2, which corresponds to equation (2.7) for $\overline{\eta}$ itself. Then we obtain $\sigma^2 = (3\,\pi/4) - (23/10) = 0.0562$, which is somewhat less than the previous value of σ^2.

If, instead of the relative straggling, we consider the absolute straggling Ω_η, we find that (3.5) gives closely the same as (3.4), being only about 4 percent less than (3.4). The approximation (E) is therefore considerably better for the straggling than for the value of the function $\overline{\eta}$ itself.

Since Ω_η^2 is expected to be more accurate than σ^2, we quote the value of Ω_η^2 obtained in approximation (E), i.e. (3.5), using (1.3)

$$\Omega_\eta^2(E) = \frac{s\,(s-1)\,(s+2)^2\,(11\,s^2+23\,s+6)}{4\,(3\,s+2)^3\,(2\,s+3)\,(2\,s-1)}\,(E\cdot\xi(E))^2,\quad \xi(E)\ll 1, \qquad (4.11)$$

which shows that the coefficient of $(E\cdot\xi(E))^2$ only varies from 0.071 to 0.109 when s increases from 2 to 3. The corresponding variation of σ^2 may be found from (4.3) and (4.11).

§ 5. Numerical and Asymptotic Solutions for $Z_1 = Z_2$

Numerical results.

The analytical solutions in § 4 give merely some guidance in the problem, because they are based on the power law scattering, which has quite limited applicability. A fairly complete and reasonably accurate solution of the case $Z_1 = Z_2$ may be obtained from representative values of the electronic stopping constant, k, together with the universal cross section given by (1.4) and Fig. 2. It is convenient to use the $\varepsilon - t$ variables in (1.4). The electronic stopping is then assumed to be $(d\varepsilon/d\varrho)_e = k\cdot\varepsilon^{1/2}$ in regions I and II. The homogeneous integral equation for $\overline{\nu}$ is

$$\left(\frac{d\varepsilon}{d\varrho}\right)_e\cdot\overline{\nu}'(\varepsilon) = \int_0^{\varepsilon^2}\frac{dt}{2\,t^{3/2}}\cdot f(t^{1/2})\left\{\overline{\nu}\left(\varepsilon-\frac{t}{\varepsilon}\right)-\overline{\nu}(\varepsilon)+\overline{\nu}\left(\frac{t}{\varepsilon}\right)\right\}, \qquad (5.1)$$

where $f(t^{1/2})$ is shown in Fig. 2. Note that (5.1) is equivalent to approximation (D).

The integral equation (5.1), with $(d\varepsilon/d\varrho)_e = k\cdot\varepsilon^{1/2}$, was solved by numerical methods on the electronic computer DASK. Actually, a slight modification of (5.1) was advantageous in the numerical computations; it gives a

slight overestimate of $\bar{\nu}$ as compared to (5.1), and should be accurate within a few percent. When starting the solutions at small values of ε, the asymptotic behaviour of the cross section (1.4) was assumed to be $f(x) \propto x^{1/3}$, corresponding to power law scattering with $s = 3$. We could here use the ana-

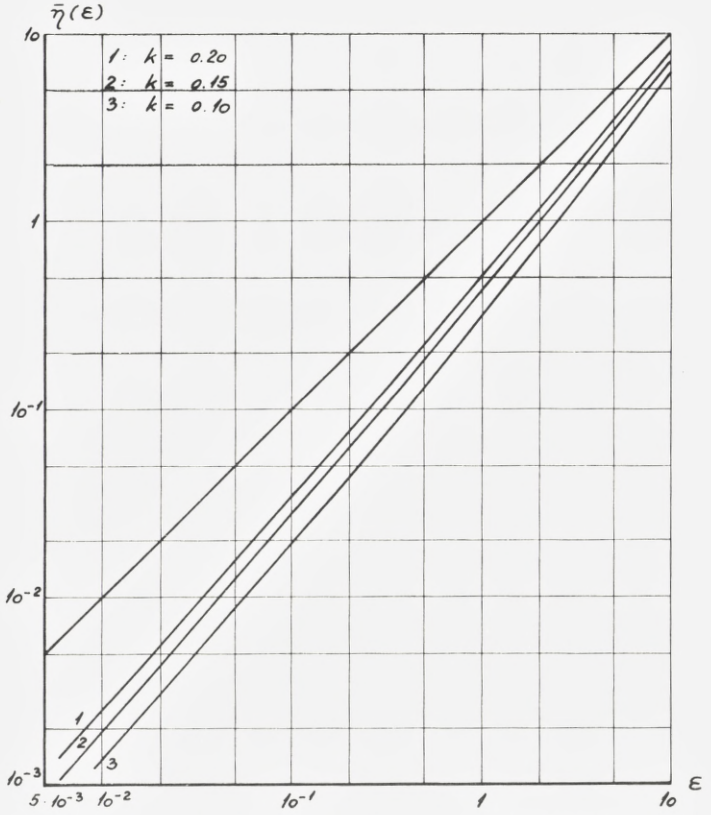

Fig. 3. The function $\bar{\eta}(\varepsilon)$ vs. ε at low values of ε, for $Z_1 = Z_2$ and in the three cases $k = 0.10, 0.15$ and 0.20. The curves were computed numerically from (5.1).

lytical estimates in § 4. In the following, solutions are presented for $k = 0.10$, 0.15 and 0.20, which covers the range of variation of k for $Z_1 = Z_2$.

The results of the coded computations of $\bar{\nu}(\varepsilon)$ from (5.1), i.e. approximation (*D*), are shown in Fig.s 3 and 4 for the above three values of k. Fig. 3 represents low values of the energy variable ε. In this region it is preferable to give the function $\bar{\eta}(\varepsilon) = \varepsilon - \bar{\nu}(\varepsilon)$, because $\bar{\nu}(\varepsilon)$ is nearly equal to ε. Fig. 4 is a continuation of the curves up to $\varepsilon = 100$. The function $\bar{\nu}(\varepsilon)$

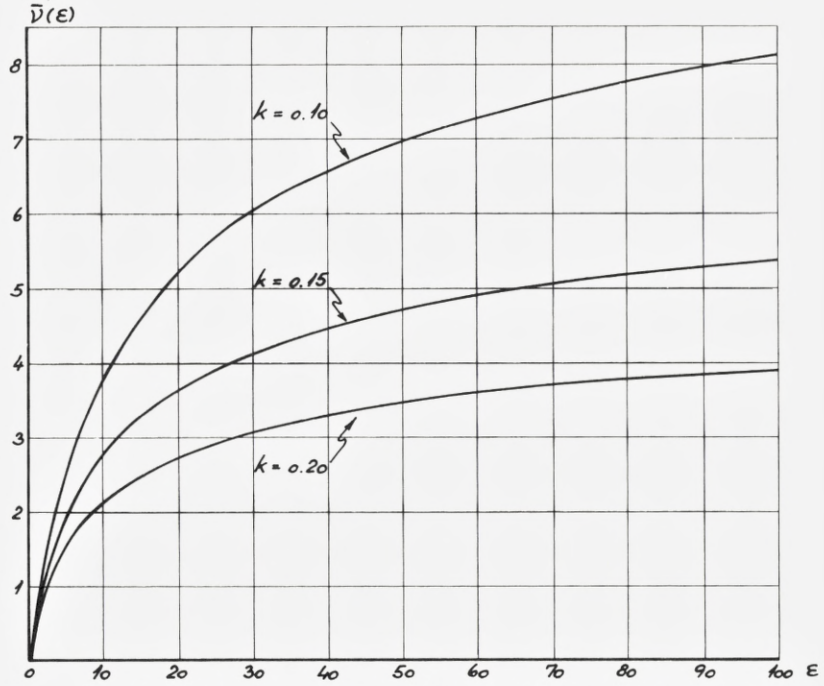

Fig. 4. The function $\bar{\nu}(\varepsilon)$ vs. ε for $0 < \varepsilon < 100$. The figure gives the continuation of the three solutions in Fig. 3.

increases initially as ε, but remains small compared to ε when ε is large. In region II, i.e. when $(d\varepsilon/d\varrho)_e = k \cdot \varepsilon^{1/2}$, $\bar{\nu}(\varepsilon)$ has an upper limit, as discussed below.

As a preliminary to the above calculations we made numerical calculations by hand in approximation (E), i.e. based on the homogeneous part of (2.8). It seems of interest to compare the two approximations. This is done in Fig.s 5 and 6, in the case of $k = 0.15$. The full-drawn curve in Fig. 5 is the accurate solution of (5.1). The dashed line is the analytical solution (4.5) for power law scattering, with $s = 2$. At $\varepsilon = 4.75$ this solution is continued by (4.8), corresponding to $s = 4/3$, cf. text in § 4. The accuracy of the power law solutions is seen to be moderate. Similarly, Fig. 6 shows $\bar{\nu}(\varepsilon)$ for $\varepsilon < 100$, in three approximations. The solid curve is the solution of (5.1). The analytical solution (4.8), for power law scattering with $s = 4/3$, is continued from Fig. 5, and shown by the dashed curve. This analytical solution is seen to become increasingly poor for large ε. The stipled curve represents the abovementioned computation by hand in approximation (E).

As expected, (E) is an overestimate of $\bar{\nu}$, by about 10 percent for high values of ε; this may be a tolerable accuracy for several purposes.

The average square fluctuation in ν, $\Omega^2(\varepsilon)$, may be computed from (3.4), i.e. approximation (D). In the coded computation we use $\varepsilon - t$ vari-

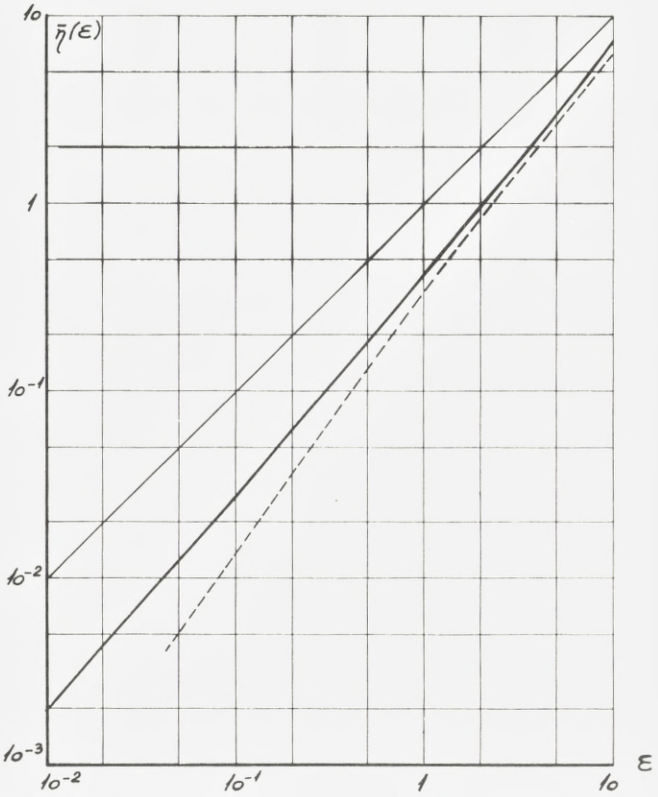

Fig. 5. Comparison of approximations for $Z_1 = Z_2$, $k = 0.15$ Curves show $\bar{\eta}(\varepsilon)$ vs. ε at low values of ε. Thick solid curve is solution of (5.1), like Fig. 3. Dashed line is power law formula (4.5), with $s = 2$. The curves approach the thin solid line $\bar{\eta} = \varepsilon$.

ables as in (5.1), and with $f(t^{1/2})$ given by Fig. 2. The equation contains inhomogeneous terms which may be computed from $\bar{\nu}(\varepsilon)$ in Fig.s 3 and 4. At $\varepsilon = 0$ the solutions were started from the analytical approximations in § 4, with $s = 3$. The results are shown in Fig. 7, for the three values of k used above, and relatively large values of ε. The figure gives $\Omega^2/\bar{\nu}^2$, the average square fluctuation divided by $\bar{\nu}^2$, and the resulting curves are seen to lie remarkably close to each other. It is instructive to compare various

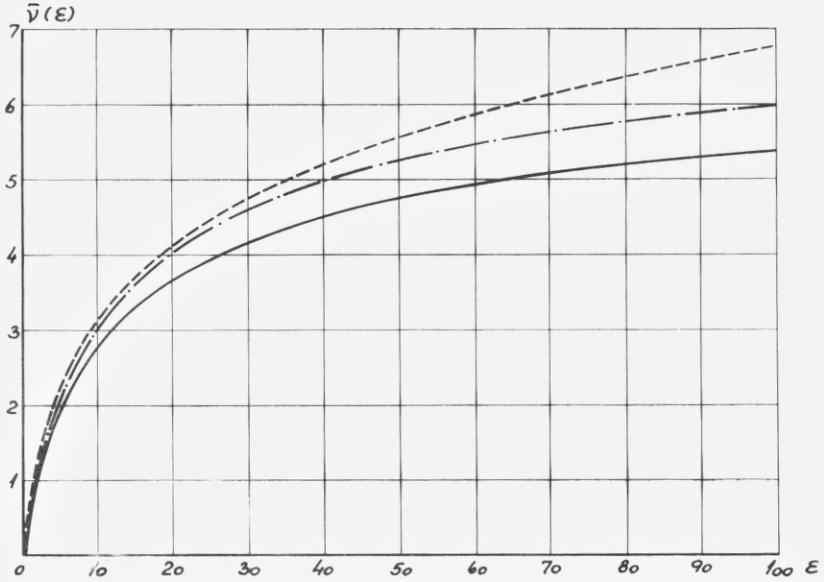

Fig. 6. Comparison of approximations for $Z_1 = Z_2$, $k = 0.15$. Curves show $\bar{\nu}(\varepsilon)$ vs. ε for $\varepsilon < 100$. The solid curve is solution of (5.1). Dashed curve is (4.8) continued from Fig. 5, corresponding to power law $s = 4/3$. Stipled curve was computed by hand in approximation (E).

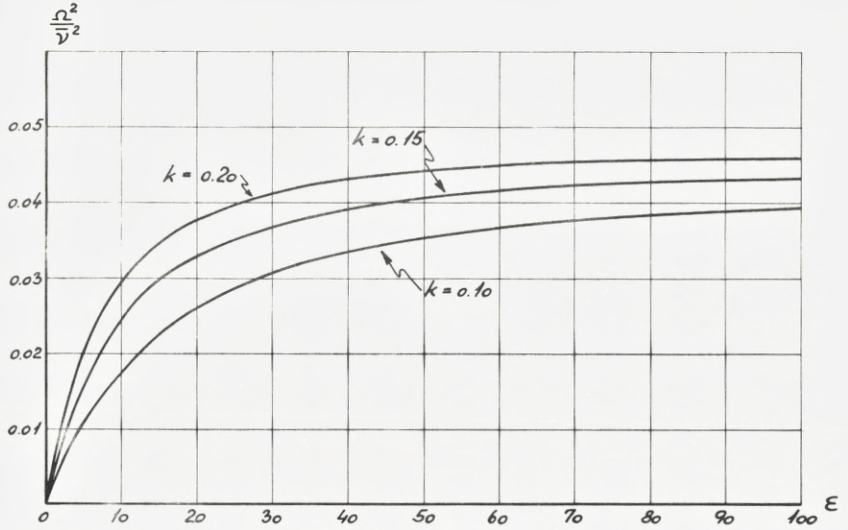

Fig. 7. Relative average square fluctuation in ν, $\Omega^2/\bar{\nu}^2$, for $k = 0.10$, 0.15 and 0.20. Coded computations in approximation (D).

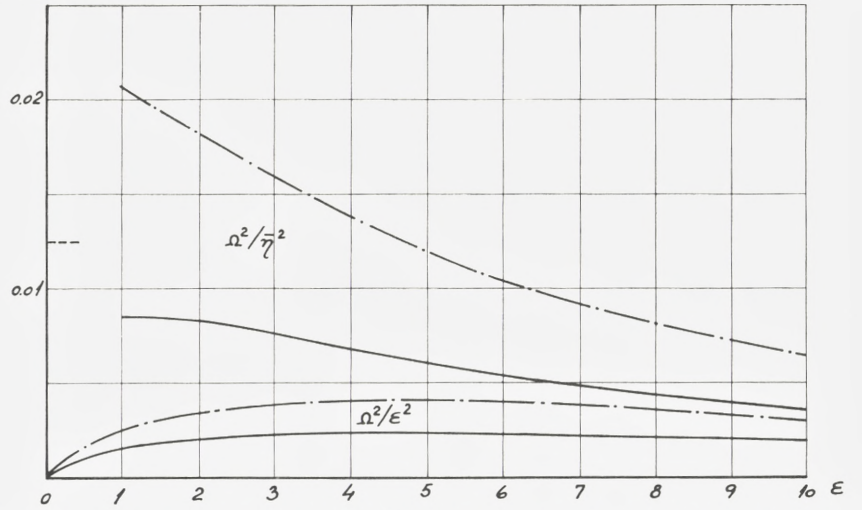

Fig. 8. $\Omega^2/\overline{\eta}^2$ (two upper curves) and Ω^2/ε^2 (two lower curves) for $k = 0.15$. Solid curves computed numerically in approximation (D), stipled curves in approximation (E).

approximations, as seen in Fig. 8 for $k = 0.15$. The solid curves represent approximation (D), as in Fig. 7. The stipled curves were computed by hand in approximation (E). The horizontal dashed line shows the point ($s = 3$), from which $\Omega^2/\overline{\eta}^2$ in approximation (E) was started at $\varepsilon = 0$. The difference between $\Omega^2/\overline{\eta}^2$ in approximations (D) and (E) is quite large, and here the errors in $\overline{\eta}$ and in Ω^2 seem to add, at low ε-values. We believe that the accuracy in Ω^2, at low values of ε, is not quite satisfactory in any of the approximations used.

One important reservation should be made as regards the above computations of $\overline{\nu}$ and $<(\nu-\overline{\nu})^2> = \Omega^2$. Apart from their definition as averages in the probability distribution $P(\nu)$, these two quantities acquire a simple meaning if $P(\nu)$ is approximately Gaussian, i. e. $P \sim C \cdot \exp\left\{-(\nu-\overline{\nu})^2/2\,\Omega^2\right\}$. However, sometimes the deviations from a Gaussian are noticeable. The probability distribution then has an asymmetric peak, with a most probable value ν^* slightly smaller than $\overline{\nu}$, and with a width at half maximum which may be considerably smaller than for the above Gaussian. There is also a tail towards high ν-values, decreasing with a power of ν of about -2 to -2.5, and having a cut-off at some high ν-value. Examples of this kind were studied in a recent paper (Lindhard and Nielsen (1962)). In any case, it depends on the experiment performed whether one may use the average values $\overline{\nu}$ and Ω^2, or take recourse to the probability distribution.

In a particle detector, where damage events due to single particles are recorded individually, one should normally consider the probability distribution. However, if many events are recorded together, like the damage by thousands of particles in a solid, the events collect into a Gaussian distribution, with average value $N \cdot \bar{\nu}$, and an average square fluctuation $N \cdot \Omega^2$, where N is the number of particles.

If the electronic stopping continued to rise as $k \cdot \varepsilon^{1/2}$, (region II), there would be an upper limit to $\bar{\nu}(\varepsilon)$. In the cases shown in Fig. 4 this upper bound may be obtained from (5.3); for $k = 0.15$ this leads to $\bar{\nu} < 7.8$. However, at an energy $\varepsilon_1 \sim 10^3$ the electronic stopping has a maximum and starts decreasing, so that approximately S_e/S_n tends to a constant $\sim 10^3$. Thus, in region III there is strictly no upper bound on $\bar{\nu}$, but its increase is extremely slow. We did not continue the coded computations into region III, partly because a new stopping parameter would be required, and partly because simple asymptotic equations take over, long before region III is reached.

Asymptotic equations.

Let us first consider a semi-empirical approximation to $\bar{\nu}$, which may be found from the numerical curves. In fact, from Fig. 4 it is seen that for large ε the function $\bar{\nu}$ is nearly reversely proportional to k, i.e. to the electronic stopping. This result cannot hold for $\varepsilon < 1$, where $\bar{\nu} \approx \varepsilon$. However, in this limit we found in § 4 that $\bar{\eta} = \varepsilon - \bar{\nu}$ is proportional to k, because the electronic stopping is small and a series development may be made of the function $\bar{\eta}$ in powers of k. A simple comprehensive formula joining the two results $\bar{\nu} \sim g_1(\varepsilon) k^{-1}$ and $\bar{\nu} \sim \varepsilon - k g_2(\varepsilon)$, is

$$\bar{\nu}(\varepsilon) = \frac{\varepsilon}{1 + k \cdot g(\varepsilon)}, \tag{5.2}$$

where $g(\varepsilon) \to 0$ for $\varepsilon \to 0$, and $g(\varepsilon) \to \varepsilon$ in region II. On the basis of the curves in Fig.s 3 and 4, we have estimated $g(\varepsilon)$ as shown in Fig. 9. It appears that (5.2) with Fig. 9 reproduces $\bar{\nu}(\varepsilon)$ within an accuracy of some percent, for all values of ε in regions I and II, and for the k-values of interest when $Z_1 = Z_2$.

A convenient approximation, valid for large ε, may be mentioned in connection with the numerical estimates. We note that for high energies E the differential cross section (1.4), as shown in Fig. 2, will be equal to the Rutherford cross section, $d\sigma_R$, except when $T \lesssim (E_0^2/E) \to 0$. If therefore we integrate a function of T, tending to zero as T, we may replace (1.4)

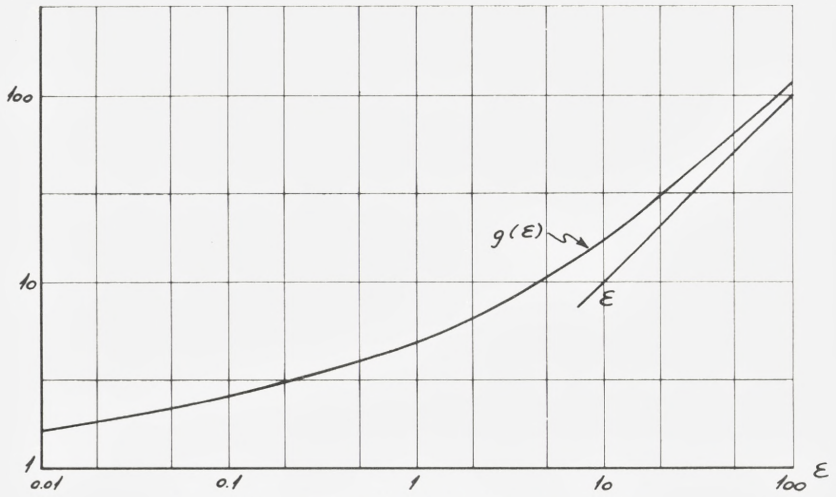

Fig. 9. The semi-empirical function $g(\varepsilon)$ in (5.2).

by the Rutherford cross section, integrated from the lower boundary $T = \lambda^2/E$, where the constant λ is determined by $\int_o^\infty T d\sigma = \int_{\lambda^2/E}^\infty T d\sigma_R$. We might make this replacement in the accurate equation (2.7), but for the present purpose (4.1) is accurate enough. Since $\bar{\nu}(T)$ increases slowly at high T-values, we can replace the upper limit E by ∞ in the integral in (4.1) and find, expressed in the $\varepsilon - t$ variables,

$$\varepsilon \frac{d\varepsilon}{d\varrho} \bar{\nu}'(\varepsilon) = \varepsilon \int_{t_o}^{\infty} \frac{dt}{4t^2} \bar{\nu}(t/\varepsilon) = \frac{1}{4} \log \varepsilon + C, \tag{5.3}$$

where $t_o = 0.60$, and C is a constant.

The fórmula (5.3) is a useful and rather accurate approximation, provided ε is larger than ~ 10. It may be readily integrated, without recourse to complicated coded computations. If we start using (5.3) at an energy ε_2, we may for instance fit $\bar{\nu}(\varepsilon_2)$ and $\bar{\nu}'(\varepsilon_2)$, the latter determining the constant C. We may normally disregard $(d\varepsilon/d\varrho)_n$ and write $d\varepsilon/d\varrho = (d\varepsilon/d\varrho)_e$. In region II we put $(d\varepsilon/d\varrho)_e = k \cdot \varepsilon^{1/2}$, and in this case (5.3) leads to an upper bound for $\bar{\nu}(\varepsilon)$, as mentioned on p. 31. We note furthermore that according to (5.3) the increase of $\bar{\nu}(\varepsilon)$ is proportional to k^{-1}, in agreement with (5.2).

An equation similar to (5.3) may be derived for the average straggling $\Omega^2(\varepsilon)$. For this purpose we consider equ. (3.4). Since the integrand on the

right tends to zero as T^2 or faster, we may directly put $d\sigma = d\sigma_R$ for high energies and integrate from 0 to E, because the integral converges rapidly at $T = 0$. We may also simplify the right hand side, since clearly $\bar{\nu}(T)$ is the dominating term for small or moderate T, where the differential cross section is large. Because $\bar{\nu}^2(T)$ saturates we then have the simple limiting approximation

$$\varepsilon\left(\frac{d\varepsilon}{d\varrho}\right)\cdot\frac{d}{d\varepsilon}\Omega^2(\varepsilon) = \varepsilon\int_0^{\varepsilon^2}\frac{dt}{4\,t^2}\,\bar{\nu}^2(t/\varepsilon) = \psi(\varepsilon). \tag{5.4}$$

We observe that the right hand side of (5.4), $\psi(\varepsilon)$, tends to a constant for large ε. The magnitude of $\psi(\varepsilon)$ may be estimated roughly by putting (cf. (5.2)) $\bar{\nu}(x) = x\cdot(1+kx)^{-1}$, leading to $\psi(\varepsilon) = \bar{\nu}(\varepsilon)\cdot 4^{-1} \to k^{-1}4^{-1}$. Now, in region II we then obtain $(d\Omega^2/d\varepsilon) \simeq \bar{\nu}(\varepsilon)4^{-1}\cdot k^{-1}\varepsilon^{-3/2}$, leading us to expect that for large ε the function Ω^2 is proportional to k^{-2}. Actually, this result fairly well corresponds to the curves in Fig. 8. In the opposite limit of low ε-values we have found that $\Omega^2 \propto k^2$.

§ 6. Outline of Treatment for $Z_1 \neq Z_2$

From the previous discussion it appears that the most direct connection between experiments and theory may be achieved in the case of $Z_1 = Z_2$. Unfortunately, there are as yet no measurements of this kind.

A brief treatment may now be given of more involved cases. We consider problems where the incoming particle does not belong to the medium, but the medium still contains only one atomic species; we write briefly $Z_1 \neq Z_2$. As we shall see, our previous division into three energy regions can no longer be upheld. At the lowest energies the description remains comparatively simple, and experiments are available for comparison with theory.

We shall not consider cases where the medium contains more than one element. The formulation of accurate general solutions can here become quite complicated, but solutions of special cases may be worked out numerically. Several measurements are available.

Consider then an incoming particle with atomic number Z_1 different from Z_2. We assume that the case of $Z_1 = Z_2$ is already solved, as described in the preceding paragraphs, and the corresponding solution for the energy given to atomic motion is $\bar{\nu}(E)$. The unknown function for the case $Z_1 \neq Z_2$

is denoted as $\bar{\nu}_1(E)$. The integral equation for $\bar{\nu}_1$ is obtained from (2.3), where we introduce approximation (*D*),

$$\bar{\nu}_1'(E)\cdot S_{1e} = \int d\sigma_1 \left\{ \bar{\nu}_1(E-T) - \bar{\nu}_1(E) + \bar{\nu}(T) \right\}. \tag{6.1}$$

Here, S_{1e} is the electronic stopping cross section for the ion Z_1 in the medium Z_2, and $d\sigma_1$ is the differential cross section for an elastic nuclear collision between an ion Z_1 and an atom Z_2, with corresponding stopping cross section S_{1n}.

In (6.1) enters $\bar{\nu}(T)$, where $T \leq T_m = \gamma E$; $\gamma = 4\, M_1 M_2/(M_1+M_2)^2$. Our previous division into three regions was characterized by the energies E_c and E_1, belonging to the atoms Z_2. Putting E_c and E_1 equal to the maximum recoil energy T_m we obtain for the particle Z_1 two characteristic energies $E_{2c} = \gamma^{-1} E_c$ and $E_{21} = \gamma^{-1} E_1$. However, the stopping cross sections S_{1e} and S_{1n} for the particle Z_1 give rise to a further subdivision. In fact, at energies lower than E_{1c} we may assume that S_{1e}/S_{1n} increases slowly, with a power of E between 1/2 and 1/6. At the energy E_{1c} the ratio S_{1e}/S_{1n} is comparable to 1. Next, above E_{1c} there is a decrease in S_{1n} while S_{1e} continues to rise as $E^{1/2}$ until the energy E_{11} is attained. For still higher energies S_{1e} decreases and the ratio S_{1e}/S_{1n} increases towards a constant $\sim 10^3$. Formally at least, we might then distinguish between five energy regions, separated by the energies E_{2c}, E_{1c}, E_{21} and E_{11}.

We limit the discussion to the lowest energy region. It is bounded upwards by either E_{1c} or E_{2c}. Approximate values of these energies are $E_{1c} \simeq A_1^3 (A_1+A_2)^{-2} Z^{4/3} Z_1^{-1/3} \cdot 500$ eV, and $E_{2c} \simeq (A_1+A_2)^2 \cdot A_1^{-1} Z_2 \cdot 125$ eV. When $Z_1 \gg Z_2$, E_{1c} will be larger than E_{2c}, while for $Z_2 \gg Z_1$ the energy E_{2c} becomes considerably larger than E_{1c}. For $Z_1 = Z_2$ the two energies are of course equally large.

Assume now that the energy is below E_{1c} and E_{2c}. We may then make the same approximation as in § 4 in region I. As an example we consider the standard case $s = 2$, leading to energy independent nuclear stopping cross sections, so that $S_{1e}/S_{1n} = (E/E_{1c})^{1/2}$ and $S_e/S_n = (E/E_c)^{1/2}$. For $\bar{\nu}(E)$ we can then apply approximation (4.6) with $\alpha_1 = 1$. The corresponding series development may be made in (6.1), i.e. in approximation (*E*). Using the expression (1.3) for $d\sigma$ we obtain

$$\bar{\eta}_1 = E - \bar{r}_1 = A E^{3/2}, \text{ for } E < E_{1c},\ E_{2c}, \tag{6.2}$$

where $A = \frac{2}{3}\left\{ E_{1c}^{-1/2} + \frac{1}{2}\gamma^{1/2} E_c^{-1/2} \right\}$.

Next, we determine the straggling Ω_1^2 in η_1. With the same low energy approximation as in (4.10), we apply (3.5). Like in (4.10) the relative straggling in η_1 becomes a constant,

$$\Omega_1^2(E)/\bar{\eta}_1^2(E) = \frac{1}{14}\gamma \left\{ \left(\frac{\gamma^{1/2}}{A E_c^{1/2}} - \frac{7}{4} \right)^2 + \frac{7}{16} \right\},\ E < E_{1c},\ E_{2c}. \tag{6.3}$$

The method actually used by us in solving equ. (6.1) is the following one. We introduce the $\varepsilon - \varrho - t$ variables described in § 1, and consider

those regions where electronic stopping cross sections are proportional to $\varepsilon^{1/2}$. The problem then contains two empirical constants, k and k_1, i.e. the proportionality factors in electronic stopping for particles Z_2 in Z_2 and Z_1 in Z_2, respectively. The values of k and k_1 are estimated in (1.2). Two further parameters enter, one being the mass factor, $\gamma = 4M_1M_2/(M_1+M_2)^2$, and the other the ratio, λ, between the ε-units for the particle pairs (Z_1, Z_2) and (Z_2, Z_2). The solutions are then of type of $\bar{\nu}_1 = \bar{\nu}_1(\varepsilon; k, k_1; \lambda, \gamma)$ and $\Omega_1^2 = \Omega_1^2(\varepsilon; k, k_1; \lambda, \gamma)$. A programme was coded for electronic computation on this basis, and solutions have been obtained in a number of cases. Three sets of solutions of this kind are quoted below. Other solutions were utilized in a recent paper on damage in Si (DENNEY et al. (1962)).

The numerical solutions should be regarded with some reservation, and they are of limited applicability. Firstly, they apply only at the low energies where electronic stopping cross sections are proportional to $\varepsilon^{1/2}$. This can be remedied by continuing the solutions by means of asymptotic equations similar to (5.3) and (5.4), cf. (6.6). Secondly, the connection to an actual measurement is rather longwinded and uncertain. The usefulness of the average quantities $\bar{\nu}_1$ and Ω_1^2 can differ much from one set of (Z_1, Z_2) to another. In any case, the three examples in the following may illustrate some of the difficulties.

Ionization efficiency.

One important experimental observation is the number of ion pairs N_i produced by a certain incoming particle; in a solid state detector we let N_i represent the number of electron-hole pairs. We shall not discuss the detailed mechanism by which electrons create ion pairs, but only note that the energy per ion pair, $W_\beta = E_{\text{electron}}/\bar{N}_i$, is approximately constant for swift electrons.* In the present case of an arbitrary incoming particle it is therefore natural to consider the total energy η given to electronic motion, and expect that the average number of ions is approximately given by the relation

$$\bar{N}_i = \frac{\bar{\eta}(E)}{W_\beta} . \tag{6.4}$$

Evidently, if η fluctuates, N_i should fluctuate proportionally. An average square fluctuation in η, $\Omega^2(E)$, must therefore contribute to the average square fluctuation, $\overline{(\Delta N_i)^2}$, in N_i by the amount

$$\overline{(\Delta N_i)_\eta^2} = \Omega^2(E)/W_\beta^2 , \tag{6.5}$$

* Experimental and theoretical discussions of W-values for electrons and α-particles are given in recent papers by JESSE (1961) and PLATZMAN (1961). The deviations of W_α/W_β from unity in polyatomic gases indicate one limitation in the accuracy of (6.4).

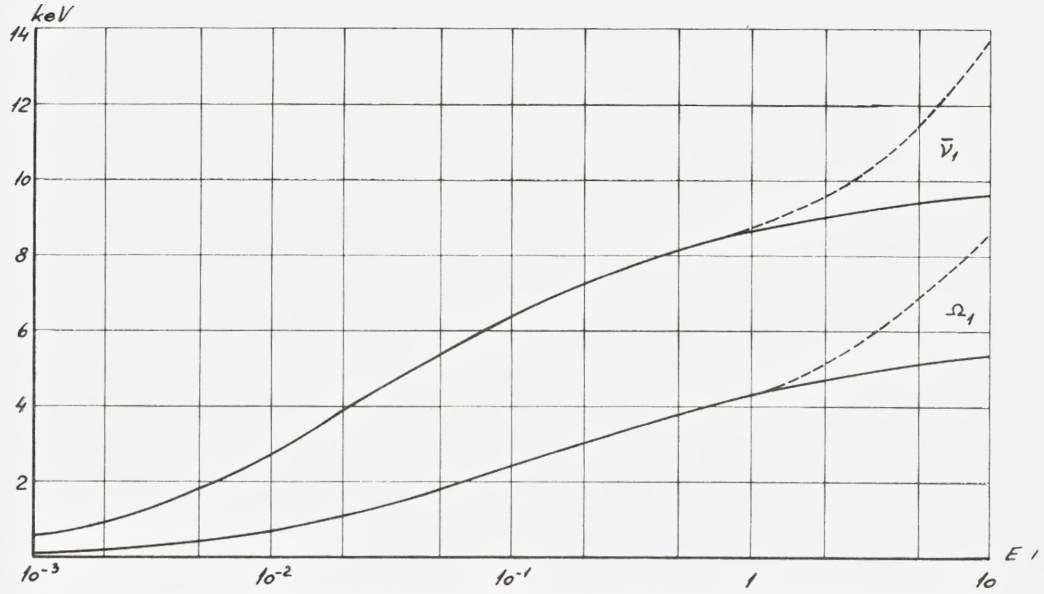

Fig. 10. Curves for $\bar{\nu}_1(E)$ and $\Omega_1(E)$ for α-particles in Si. Solid curves correspond to the coded computations. Dashed curves include corrections for decrease in electronic stopping, cf. (6.6).

but this is not the only cause of fluctuation of N_i. A direct statistical effect in ion pair production is that considered by FANO (1947), where the average square fluctuation was found to be $\overline{(\Delta N_i)^2}_F = F \cdot \bar{N}_i$, with $F \lesssim 0.5$, i.e. in some respects similar to a Poisson distribution. In many cases the fluctuation (6.5) dominates over the Fano effect.

In a treatment more precise than (6.4) and (6.5) one would introduce $\bar{N}_i$ directly as the variable $\bar{\varphi}$ in the basic integral equations. In fact, the basic case in production of ion pairs is an electron passing through a medium, and one must at first solve (2.2′) for $\bar{\varphi}_e(E) = \bar{N}_{ie}(E)$, i.e. the average number of ion pairs produced by an electron of energy E. Next, (2.1) is solved ($Z_1 = Z_2$) with respect to $\bar{N}_i(E)$, $\bar{N}_{ie}(E)$ being a source term. Thirdly, equ. (2.3) for $\bar{N}_{i1}(E)$ is solved. The Fano fluctuation is an estimate of the fluctuation in the first step only.

α-particles in Si.

Our first example of numerical computations illustrates the ionization by charged particles in a detector. We consider α-particles in Si, i.e. a solid state detector, but the results are quite similar to those for α-particles in A. In Fig. 10, the full-drawn curves show the behaviour of $\bar{\nu}_1(E)$ and $\Omega_1(E)$, as obtained from the coded computations mentioned above. Now, electronic stopping for α-particles in Si is proportional to velocity only up

to about 0.7 MeV, where a maximum obtains, upon which the stopping decreases as $\sim v^{-1}$. The full-drawn curves in Fig. 10 are therefore underestimates at energies above 1 MeV. A correction can be made rather easily, since $\bar{\nu}(E)$ for Si ions in Si at the energies in question is given in e.g. Fig. 6, or by (5.1), so that the asymptotic equation is

$$\varepsilon\left(\frac{d\varepsilon}{dQ}\right)\cdot\bar{\nu}_1'(\varepsilon) = \frac{\lambda\varepsilon}{\gamma}\int_{t_0}^{\varepsilon^2}\frac{dt}{4\,t^2}\,\bar{\nu}\left(\frac{\gamma t}{\lambda\varepsilon}\right), \tag{6.6}$$

where the right hand side is known, and $t_0 = 0.60$. A similar treatment may be made for $\Omega_1^2(\varepsilon)$. In this manner the two dashed curves were obtained for $\bar{\nu}_1(E)$ and $\Omega_1(E)$ in Fig. 10. By means of (6.4) and (6.5) may be found the resulting effects on signal size, $\bar{N}_i$, and on signal fluctuation, $\overline{\Delta N_i^2}$. However, the fluctuation Ω_1 is so large that the distribution in ν_1 must differ considerably from a Gaussian. The quantities $\bar{\nu}_1$ and Ω_1 are then less relevant than the most probable value of ν_1, the width at half peak height, and the shape of the tail in the probability distribution. In a recent note (LINDHARD and NIELSEN (1962)) the latter quantities are obtained by a method much simpler than the above one.

Ionization by α-recoils.

The recoil nucleus in α-decay is a very heavy particle with an energy of only 100—200 keV. In this case $\bar{\eta}_1(E) \ll E$, and a conspicuous effect should be observed in the number of ion pairs, according to (6.4). Detailed measurements have been made by B. MADSEN (1945), for Po, ThC and ThC′ α-recoils. In argon containing about 5 percent air, MADSEN observed the average number of ion pairs, $\bar{N}_i$, and also the width of the distributions.

The corresponding coded computations of $\bar{\nu}_1(E)$ and $\Omega_1^2(E)$ for a heavy recoil particle in pure argon have been performed. The three recoil nuclei have practically the same atomic number, and differ only in energy. The resulting behaviour of $\bar{\eta}_1(E)/W_\beta$ is shown by the full-drawn curve in Fig. 11. In the figure is also shown the result, if power law scattering with $s = 2$ is assumed, as indicated by the dashed line. The three experimental points of MADSEN are his values for $\bar{N}_i$, assuming $W_\alpha = W_\beta = 26.4$ eV, the energy per ion pair in pure argon. The points lie below the solid curve and, in view of the uncertainties, the agreement must be said to be satisfactory. From MADSEN's curves the mean square relative fluctuation, $\overline{\Delta N_i^2}/\bar{N}_i^2$, may be estimated roughly. It is of order of $\overline{\Delta N_i^2}/\bar{N}_i^2 \sim 0.02$. This is considerably

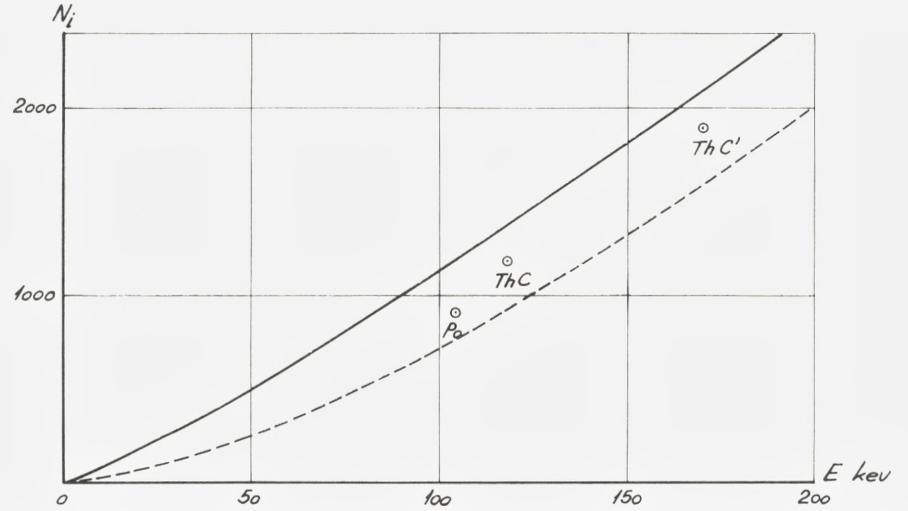

Fig. 11. Comparison with three α-recoil measurements by B. MADSEN. Solid curve is $\overline{\eta}_1(E)/W_\beta$ computed numerically. Dashed curve corresponds to the power law approximation (6.3). Comparison assumes $W_\alpha = W_\beta$, but magnitude of this constant is not important.

larger than the numerically computed average square straggling, $\Omega_1^2/\overline{\eta}_1^2 \sim 0.002$, but in approximate agreement with (6.3), i.e. $s = 2$. The latter is possibly fortuitous, and further measurements in the region of extremely low velocities are desirable.

Ionization by fission fragments.

As a third example we may consider the ionization by fission fragments in various gases. The question of the ionization efficiency of fission fragments was studied experimentally by SCHMITT and LEACHMAN (1956), cf. also UTTERBACK and MILLER (1959). SCHMITT and LEACHMAN observed the variation of the number of ions, $\overline{N}_i$, with fragment energy in several gases. It turned out that $\overline{N}_i$ was not quite proportional to the energy of the fragment. They therefore considered the difference between E and the energy $E_\alpha = W_\alpha \cdot \overline{N}_i(E)$, where $W_\alpha \approx W_\beta$ is the energy per ion pair for natural α-particles. This difference, $\Delta = E - W_\alpha \overline{N}_i$, was called the ionization defect. Now, if (6.4) holds very accurately, and if $W_\alpha = W_\beta$, it is apparent that Δ becomes equal to the present function $\overline{\nu}_1(E)$. However, since the observed Δ's are only some 5 percent of E, and since in some cases already W_α can deviate from W_β by several percent, it is abundantly clear that a comparison between Δ and $\overline{\nu}_1$ is only qualitative, as long as the excitation and ionization cross sections for fission fragments have not been studied in detail.

The coded computations of $\bar{\nu}_1$ and Ω_1^2 were carried through for fission fragments in a number of substances, taking one representative of the median light group ($Z_1 = 39$, $A_1 = 94.7$, $E_{\text{initial}} = 98.9$ MeV), and one representative of the median heavy group ($Z_1 = 53$, $A_1 = 138.8$, $E_{\text{initial}} = 66.9$ MeV). Several results of this kind are given in a recent paper (LINDHARD and THOMSEN (1962)). Results are quoted in Table 1 for the two groups of fission fragments with initial velocities in Ne and A, as compared with the observations of Δ by SCHMITT and LEACHMAN. There is quantitative agreement, and more could hardly be expected. It is seen that $\bar{\nu}_1$ is systematically smaller than Δ, which is not surprising since the value to be used for W may be greater than W_α.

TABLE 1

	Ne		*A*	
	Δ (MeV)	$\bar{\nu}_1$ (MeV)	Δ (MeV)	$\bar{\nu}_1$ (MeV)
Heavy group	4.8 ± 0.7	2.5	5.5 ± 0.5	3.1
Light group	4.3 ± 1.0	1.6	5.1 ± 0.8	2.0

Fluctuations have not been studied experimentally. As examples of the numerical computations it may be mentioned that for the heavy fission fragment group with initial velocities in Ne and A the values of $\Omega_1/\bar{\nu}_1$ are 0.066 and 0.097, respectively.

In an interesting theoretical treatment of the ionization yields of fission fragments KNIPP and LING (1951) have used a differential-integral equation for the average ionization of similar type as (E) in the present paper. Moreover, they introduced the description by ionization defect Δ employed by SCHMITT and LEACHMAN. The estimates of atomic collision cross sections by KNIPP and LING were necessarily somewhat uncertain. They considered the case of fission fragments in argon. For argon in argon their maximum ionization defect Δ was 780 keV, while our upper bound on $\bar{\nu}_1$ in region II (cf. p. 31) gives 600 keV for argon in argon. For the two fission groups in argon their estimates of Δ are also somewhat larger than our values of $\bar{\nu}_1$. KNIPP and LING made use of the connection to MADSEN's measurements.

Production of lattice defects.

In the present context mention should be made of the damage produced in a crystal lattice by irradiation. A general survey of radiation damage in solids is given by BILLINGTON and CRAWFORD (1961). Consider a solid composed of one element only. We may let φ represent e.g. the number

of vacancies N_v produced by a particle with $Z_1 = Z_2$. The discussion below applies just as well for the production of other lattice defects. In first approximation N_v should be proportional to the energy given to atomic motion, ν. The average value of N_v is therefore expected to obey an equation similar to (6.4)

$$\bar{N}_v = \frac{\bar{\nu}(E)}{U_v}, \tag{6.7}$$

where U_v may be regarded as an empirical constant. The relation (6.7) probably affords a more direct experimental check of the present results for $\bar{\nu}$ and $\bar{\eta}$ than does equ. (6.4). The reason is that in most cases $\bar{\nu} \ll E$ and $\bar{\eta} \approx E$, as in the ionization efficiency of fission fragments.

U_v can also be estimated theoretically from (2.5), i. e. approximation (B^{-1}), valid at low energies where no energy ends up in electronic motion. Having derived a constant U_v at such low energies, we have also justified the use of (6.7) at higher particle energies.

Several estimates have been made of the connection between U_v and atomic binding (Snyder and Neufeld (1955, 1956) and others, cf. Seitz and Koehler (1956), Billington and Crawford (1961)). It has become customary to use hard sphere ion-atom scattering, i.e. $d\sigma_n = \text{const.}\, dT$. Our present cross sections in § 1 are much more forward peaked and lead to a higher value of the ratio between U_v and atomic binding.

The fluctuation in N_v, $\overline{(\Delta N_v)^2}$, has a contribution from the fluctuation in ν. We find analogously to (6.5), $\overline{(\Delta N_v)^2_\nu} = \Omega^2(E)/U_v^2$. The magnitude of the relative fluctuation in N_v may be read off directly from the curves in Fig. 7, for $Z_1 = Z_2$.

In approximation (B^{-1}), and with hard sphere ion-atom scattering, Leibfried (1958) has derived a fluctuation in N_v, $\overline{(\Delta N_v)^2_L} = 0.15\, \bar{N}_v$, analogous to the Fano ionization fluctuations. Already at quite low energies the fluctuation of Leibfried is completely overshadowed by the present fluctuations.

The above relations, together with our previous computations of $\bar{\nu}(E)$ and $\Omega^2(E)$, cover the question of $\bar{N}_v$ and its fluctuation for $Z_1 = Z_2$. If $Z_1 \neq Z_2$ some cases are represented by the examples in this section, and others by Lindhard and Thomsen (1962). An interesting further example is the damage produced by neutrons, where the production spectrum of recoils by neutrons, times $\bar{\nu}(E)$ from § 5, may be integrated to give the production of lattice defects.

Finally, it should again be emphasized that (6.7) is an approximation. If necessary, more accurate treatments may be made. Thus, let us consider

the behaviour of $\overline{N}_v(E)$ at high particle energies. Here, an increasing fraction of the energy transfers to atoms are so small in magnitude that lattice bindings need not be disrupted. In fact, the logarithmic increase of the right hand side of (5.3) for increasing ε is due to such small energy transfers. In the evaluation of $\overline{N}_v$ we may therefore at a sufficiently high energy replace $\log \varepsilon$ by a constant, but this does not result in a large correction.

In conclusion we wish to express our deep gratitude to all who have encouraged us and assisted in this work. Miss SUSANN TOLDI has given untiring assistance in the preparation of the manuscript.

Institute of Physics,
University of Aarhus.

Note added in proof. In a recently published article by ABROYAN and ZBOROVSKII (Soviet Physics Doklady, **7**, 417 (1962)) the ionization pulse by potassium ions in a germanium detector is measured at ion energies ~ 1 keV. The authors find that the ratio β between the pulse for K ions and for electrons with the same energy is $\beta = 0.032$, 0.071, 0.114 and 0.135, for $E = 0.5$, 1, 3 and 8 keV, respectively. Now β should be equal to $\overline{\eta}/E$, and the simplified theoretical formula (6.2) gives $(\overline{\eta}/E) = 0.051\, E^{1/2}$, where E is measured in keV. This is in excellent agreement with the experimental values of β. However, numerical estimates corresponding to (5.2) are nearly a factor of 2 higher. In view of the smallness of β the results are promising in any case.

References

D. S. Billington & J. H. Crawford (1961). Radiation Damage in Solids. Princeton University Press.

N. Bohr (1948). The Penetration of Atomic Particles through Matter. Mat. Fys. Medd. Dan. Vid. Selsk. **18**, no. 8.

J. M. Denney, R. G. Downing & G. W. Simon (1962). Energy Dependence of Proton Damage in Silicon. 8653-6005-KU-000. To be published in Phys. Rev.

J. Durup & R. L. Platzman (1961). Role of the Auger Effect in the Displacement of Atoms in Solids by Ionizing Radiation. Disc. Far. Soc. No. 31.

A. Erdélyi, W. Magnus, F. Oberhettinger & F. G. Tricomi (1953). Higher Transcendental Functions, I. McGraw-Hill.

U. Fano (1947). Ionization Yields of Radiations. Phys. Rev. **72**, 26.

N. V. Fedorenko (1959). Ionization in Collisions between Ions and Atoms. Soviet Phys. Uspekhi **2**, 526.

W. P. Jesse (1961). Alpha-Particle Ionization in Polyatomic Gases and the Energy Dependence of W. Phys. Rev. **122**, 1195.

J. K. Knipp & R. C. Ling (1951). On the Ionization Yields of Heavy Particles. Phys. Rev. **82**, 30.

G. Leibfried (1958). Statistische Schwankungen der von einem Primärteilchen erzeugten Frenkel-Defekte. Nukleonik **1**, 57.

J. Lindhard & M. Scharff (1961). Energy Dissipation by Ions in the kev Region. Phys. Rev. **124**, 128.

J. Lindhard & P. V. Thomsen (1962). Sharing of Energy Dissipation between Electronic and Atomic Motion. 'Radiation Damage in Solids', vol. I, 65, IAEA, Vienna.

J. Lindhard, M. Scharff & H. E. Schiøtt (1962). Range Concepts and Heavy Ion Ranges (Notes on Atomic Collisions, II), preprint. To appear in Mat. Fys. Medd. Dan. Vid. Selsk.

J. Lindhard & V. Nielsen (1962). Nuclear Collisions and Ionization Fluctuations in Particle Detectors. Physics Letters **2**, 209.

B. S. Madsen (1945). Ionization Measurements on Single Recoil Particles from Po, ThC and ThC′. Mat. Fys. Medd. Dan. Vid. Selsk. **23**, no. 8.

R. L. Platzman (1961). Total Ionization in Gases by High-Energy Particles. Int. Journ. Appl. Rad. and Isot. **10**, 116.

H. W. Schmitt & R. B. Leachman (1956). Ionization vs. Energy Relation for Fission Fragments. Phys. Rev. **102**, 183.

F. Seitz & J. S. Koehler (1956). Displacements of Atoms During Irradiation. Solid State Phys. **2**, 305.

W. S. Snyder & J. Neufeld (1955). Disordering of Solids by Neutron Radiation. Phys. Rev. **97**, 1636.

W. S. Snyder & J. Neufeld (1956). Vacancies and Displacements in a Solid Resulting from Heavy Corpuscular Radiation. Phys. Rev. **103**, 862.

N. G. Utterback & G. H. Miller (1959). Ionization Yields for Fission Fragments. Phys. Rev. **116**, 976.

Indleveret til Selskabet den 15. november 1962.
Færdig fra trykkeriet den 6. maj 1963.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 11

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 11 (1963)

ON THE QUANTITATIVE EVALUATION OF AUTORADIOGRAMS

BY

HILDE LEVI AND A.W. ROGERS
IN COLLABORATION WITH
M. WEIS BENTZON AND ARNE NIELSEN

København 1963
i kommission hos Ejnar Munksgaard

Synopsis

Results of an investigation into the possibilities for a quantitative interpretation of autoradiograms on the basis of grain counting or track counting are presented. The relation between track length, number of grains, and initial energy of the beta particle is studied, and the blackening produced by a point source of a beta emitter, e. g. carbon-14, surrounded by nuclear emulsion is described in terms of the distribution of grains in space. It is shown how an analysis of this type can be carried out for any pure beta emitter. However, the present measurements are confined to beta particles of up to c. 400 keV.

Printed in Denmark
Bianco Lunos Bogtrykkeri A-S

Contents

Chapter I

I, 1. Introduction

When ionising particles, such as beta particles emitted by radioactive nuclides, pass through a photographic emulsion, some of their energy is transferred to the emulsion, and a number of silver halide grains is rendered developable. Due to this property, photographic emulsions have been used in order to demonstrate the presence of a source of beta particles in a specimen. In biological terms, this has often involved the recognition or identification of labelled cells in a tissue section or smear.

The autoradiographic techniques that have evolved fall broadly into two categories—those in which a relatively thick emulsion layer (of the order of 100 microns or more) is used and the passage of a beta particle is recognized as a continuous track of silver grains, and those in which a thin layer of emulsion (of the order of 10 microns or less) registers beta particles as a few blackened grains at most.

Clearly, the greater the number of ionising particles traversing a layer of photographic emulsion, the greater the number of tracks or grains one might expect to find. In principle, there is no reason why this method of recording the presence of ionising radiation should not be used in a strictly quantitative manner. In other words, it should be possible to calculate the number of beta particles entering a given volume of emulsion from the number of grains or tracks formed.

I, 2. Relative quantitation

Autoradiography has been used to provide estimates of the degree of labelling of one source relative to that of another by comparing the number of grains or tracks in comparable volumes of emulsion overlying or surrounding the sources studied. Also this type of quantitation, which should be called relative in contradistinction to the absolute measurements referred to in the previous section, requires strict control of technique at four stages, if it is to be reliable.

First, the source to be studied must be presented to the emulsion in such a way that loss or translocation of isotope does not occur, neither before nor during exposure. This prerequisite is difficult to ensure with biological material. The assumption is often made that loss of isotope during histological processing, if it does occur, will be the same in various parts of the tissue, so that a comparison of residual activities in different regions gives reliable information on the relative amounts of radioactive material present *in vivo*. LEVI, 1957, working with crystals of S-35 labelled barium sulphate, has drawn attention to the possible movement of even sparingly soluble radioactive material, particularly in ionic form, into and through the emulsion during exposure.

The *second* critical factor in technique is the geometrical relation between source and emulsion. Unless this is shown to be the same for the various structures whose autoradiograms are to be compared, the results cannot be related accurately. If, for example, one cell type is larger than another, or lies deeper in the tissue section (i. e., further away from the emulsion), if the emulsion layer is thicker in one place than in another, or the section varies in thickness, relative quantitation becomes an illusion, unless these factors have been recognised and corrected for (cf. LEVI, 1957; PERRY, 1961).

The *third* prerequisite for relative quantitation is the standardisation of the photographic processing. Conditions of exposure, viz. temperature, humidity, and the presence of oxidising agents, affect the stability of the latent image. Moreover, the type, the concentration, and the temperature of the developer have considerable influence on the image produced.

Finally, the grains or tracks in the relevant volumes of the emulsion must be counted, and the results compared. This comparison rests on the assumption that the emulsion response is linearly related to the number of incident particles. It is well known that at grain densities higher than those which can be evaluated visually, emulsion saturation begins to play a part, so that proportionality between incident radiation and emulsion response no longer prevails. It is, therefore, important that the exposure time be chosen so that all structures to be compared show blackening which does not exceed the linear part of the characteristic curve (*H*- and *D*-curve; MEES, 1952).

From the above it is clear that relative quantitation of autoradiograms is feasible only under carefully controlled conditions. However, it is worth keeping in mind that the sources of error become increasingly significant as the differences in labelling of different structures to be compared become

smaller. Aside from the purely technical problems, it is in the nature of radioactive decay that all observations are subject to statistical fluctuations, which brings an additional uncertainty to the evaluation.

I, 3. Absolute quantitation

The exact determination of the number of beta particles entering a given volume of emulsion, and hence the number of disintegrations taking place in the source per unit time, requires still more stringent control of technique. Obviously, the requirements discussed in the preceding paragraph must be fulfilled and, in addition, the emulsion response, i. e. the number of grains or tracks observed in the emulsion, must be converted into the number of disintegrations taking place in the source.

Grain counting.

Attempts at correlating the number of developed grains observed in the emulsion with the amount of label in the source have been made by several investigators (Howard and Pelc, 1951; Herz, 1957; Lajtha, 1952; Marinelli and Hill, 1948; Lammerton and Harris, 1954; Odeblad, 1950). They defined the term grain yield as the number of grains produced in an emulsion layer by a source which, in the majority of cases, was separated from the emulsion. Grain yield will then depend first of all on the geometrical relation between source and emulsion, and on a great many other variables.

Tables of grain yield are available in the literature (loc. cit.). Each figure represents the grain yield as determined under the conditions of one particular experiment. Some of these determinations have been made by covering a large, uniform source of beta particles with a thin layer of emulsion. The average optical density of the processed emulsion (after exposure) and the average number of incident beta particles were measured independently, and the number of grains produced per incident particle was calculated. Some investigators made use of a point source and determined the "grain yield" as above. However, since grains produced by a point source of beta particles may be found at a considerable distance from their origin, it is misleading to present a figure for grain yield without specifying the volume of emulsion within which the grains were observed.

It is therefore difficult to justify the use of such figures in converting grain counts into terms of disintegrations in the source under different

experimental conditions. Autoradiographic grain yield in this sense is comparable to the efficiency of a Geiger counter which is different in each arbitrary counting arrangement.

Track counting.

The determination of the number of disintegrations taking place in a source by means of track autoradiography requires the recognition of every track originating from the source.

A first attempt at correlating the number of beta tracks observed in the emulsion with the amount of label in the source was made by LEVI, 1954, who suspended C-14 and S-35 labelled algæ and yeast cells in Ilford G-5 emulsion. In these experiments, absolute quantitation was not even attempted in view of the variation in the degree of labelling of individual cells. When the same author used uniformly labelled barium sulphate crystals as beta emitters (1957), useful information on the possibilities of track autoradiography for absolute quantitation was gained. However, this system proved unsatisfactory mainly because of artefacts due to the solubility of the crystals.

LEVINTHAL, 1957, devised experiments based on absolute quantitation of beta track autoradiography. He suspended virus particles labelled with P-32 in thick layers of G-5 emulsion. The position of a virus was recognised by the star of beta tracks originating from a common centre. LEVINTHAL et al. (loc. cit.) were able to prove that the number of tracks in a star was a direct measure of the number of disintegrations that had occurred in the virus during exposure.

A possible approach.

An approach to absolute quantitation of autoradiograms can be made by establishing the possibility for quantitative *track* autoradiography and from there to proceed to an investigation into the quantitation of *grain* autoradiograms.

The starting point could be the study of a mathematically simple model, for example a point source emitting beta particles with a known energy spectrum, surrounded on all sides by an emulsion layer of infinite thickness as compared to the maximum range of the particles. If the source is small enough for self-absorption to be negligible (a prerequisite fulfilled in the present experiments, but seldom realized when the beta particles originate from biological material), the only adjustments that need be applied to the observed track count are 1) a correction for beta particles that have failed

to give rise to a recognisable track and 2) a correction for background tracks due to cosmic radiation.

An adequate description of this simple model must include an analysis of the relationships between particle energy, track length, number of grains per track, and spatial distribution of silver grains relative to the source. It then becomes possible to express the disintegration rate in the source both in terms of track counting and of grain counting.

The question remains, however, whether the description of the simple model will open the way for an analysis of the much more widely used, but mathematically complex systems where the source is covered by emulsion only over part of the space angle, and different media (air, gelatin, glass) surround the source over the remaining space angle. In this latter model, difficulties arise not only from the more complex geometry, but mainly from the fact that media with different densities will make range and yield evaluations extremely difficult. Practically all autoradiographic studies on biological material—by means of stripping film or thin coats with liquified emulsion—fall within the last mentioned category.

The work to be reported in the present paper was undertaken to follow up earlier investigations dealing with the above mentioned simpler case: source embedded in emulsion (Levi and Nielsen, 1959). By studying sources of molecular dimensions in Ilford G-5 emulsion and using beta emitters with a wider range of energies, the relationships between particle energy, number of grains per track, and spatial distribution of grains relative to the source were investigated. Some discrepancies between results reported earlier by Levi and Nielsen (loc. cit.) and by Zajac and Ross (1948) required clarification.

I, 4. Definition of terms

Grain yield.

The factor which relates the number of grains to the number of ionising particles producing them is known as the grain yield. It is defined as the ratio of the number of developed grains in a given volume of the emulsion to the number of ionising particles entering this volume. Several factors affect grain yield. They may be considered in three main groups.

1) The nature of the source: The energy spectrum of the isotope is the first consideration. Beta particles of tritium (E-max = 18.5 keV) will produce fewer grains in emulsion than those of radiophosphorus (E-max = 1.6 MeV).

2) The nature of the emulsion: If one emulsion is more highly sensitised than another with equal grain size, the grain yield will be higher in the former than in the latter. If, in two emulsions of equal sensitivity, the grain size of the first is smaller while the total concentration of silver halide is the same in both, a given loss of energy by an incident particle is likely to produce more grains in the first than in the second emulsion.

3) Conditions of exposure: Conditions of exposure will affect grain yield because, f. ex., oxidising agents or water promote latent image fading. It has also been demonstrated that, to a lesser extent, the time of development, the chemical composition of the developer as well as the temperature will influence the number of grains rendered visible in the emulsion (cf. DEMERS, 1959; AHMAD and DEMERS, 1959).

Beta tracks.

Tracks produced by beta particles in nuclear emulsion are characterised by the zigzag path of the particles and the irregular spacing of the silver grains in each track. The higher the energy of the beta particle, the longer its track and the wider its grain spacing in the beginning of the track. With decreasing energy, scattering and grain density (number of grains per unit path) increase.

A track is usually defined as 4 or more silver grains in a row. Therefore, a particle belonging to the low energy part of the continuous beta spectrum of any isotope may not produce sufficient grains to be recognised as a track. In the case of P-32, the proportion of unrecognisable tracks is very small and, at the same time, the majority of the tracks is fairly straight up to a considerable distance from the source (cf. LEVINTHAL's "stars", 1957).

In the case of C-14, the energy of a large fraction of the particles is so low that they will not produce recognisable tracks, because even 3 grains in a group cannot be distinguished from random arrangements of background grains. This problem has been studied in detail by LEVI and NIELSEN, 1959.

Delta tracks.

Beta particle tracks in nuclear emulsions sometimes branch. Branching occurs when the particle, on its passage through matter, imparts sufficient energy to an orbital electron to knock it out of orbit. Since the ejected electron has the same mass and charge as a beta particle, the characteristics of its track are the same as those of the primary particle. The orbital electron is known as a delta ray. The two branches of the resulting track configuration lie very nearly at right angles to each other.

Range-energy relation.

The term "range" has been applied to beta particles in two different senses. It is frequently used to describe the point-to-point distance travelled by a beta particle of a given energy in a medium of known composition. It is just as often used to denote the furthest distance from the point of origin that a beta particle of given energy will reach in a medium of known composition. Since beta particles undergo scattering on their passage through matter, the "range" by the second definition is always shorter than that given by the first. For low energy betas, f. ex. C-14, the difference between the two distances is appreciable.

To avoid confusion, the grain-to-grain distance travelled by a beta particle in the emulsion will be referred to in the present paper as the *track length*. The distance from the origin of the track to the point furthest away from the origin will be called the *radius* of the track. This is, obviously, the radius of the smallest sphere around the point of origin that contains the entire track.

The mean track lengths of beta particles of known energy can be predicted with great accuracy from theoretical considerations. Demers has constructed a table relating mean track length to energy for beta particles in Ilford G-5 emulsion, and this relationship is confirmed experimentally (cf. Zajac and Ross, loc. cit.).

The maximum probable radii for groups of beta particles have been investigated by interposing absorbers of known density between the source and the detector. This is the familiar "range" determination expressed in terms of mgm/cm^2 (cf. Glendenin, 1948).

The relationship between track length and radius has not been investigated except by Herz, 1949, who studied the depth of penetration of beta particles into a layer of nuclear emulsion.

Chapter II

Experimental

II, 1) Materials

The beta-emitting isotopes used in this work were

1) carbon-14 (E_{max} 155 keV) as labelled glucose in aqueous solution and as labelled yeast cells suspended in water;

2) calcium-45 (E_{max} 250 keV) in ionic form, carrier-free in water;

3) chlorine-36 (E_{max} 714 keV) in ionic form, carrier-free in water.

The emulsion was Ilford G-5 nuclear emulsion in gel form which was used within two weeks of dispatch. Prior to use, the emulsion was stored in a refrigerator at 4°C within a lead shield.

3″×1″ microscope slides were thoroughly cleaned in chromic acid, washed over night, and "subbed", i. e. coated with a thin gelatine layer, as prescribed for use with stripping film.

II, 2) Preparation of slides without cells (sandwich plates).

The stock solution containing the radioactive tracer was diluted with distilled water to a concentration which, by trial and error, was found to produce a reasonable track density in the emulsion in the course of 18–40 h of exposure.

In the dark room, the subbed slides were placed on a levelled glass surface at room temperature (20-25°C). In safelight, the required volume of G-5 emulsion was transferred to a 25 ml measuring cylinder placed in a thermostatically controlled water bath at 42°C. In the course of 10 min, the emulsion had liquified. Subsequently, the molten emulsion was filtered through gauze into another measuring cylinder, and diluted with half its volume of distilled water preheated to 42°C. After 2-4 min of gentle stirring at 42°C, to ensure complete mixing, the diluted emulsion was filtered a second time and kept in a measuring cylinder in the water bath ready for use. Using a Carlsberg pipette with a wide (broken) tip, 1 ml of the molten diluted emulsion was withdrawn and mixed, in a small glass jar, with 100 μl of the aqueous solution containing the tracer.

With another Carlsberg pipette, 1 ml of molten, diluted emulsion was placed on each slide, and spread gently with a fine paint brush to cover the whole surface of the slide. This layer was then left to gel for about 15 min. Next, one drop of the emulsion containing the radioactive isotope was placed in the centre of the slide and spread to cover an area of 2×3 cm^2. Finally, 1 ml of inactive emulsion was pipetted on top and spread to cover the whole slide.

The slides were left on the levelled glass plate in the dark in a gentle current of air for about 3 h, when they appeared dry. They were then placed in plastic slide boxes, and the open boxes were transferred into a desiccator containing dry silica gel. A current of dried CO_2, obtained from evaporating dry-ice in a flask, was passed over the plates for about 20 min and, finally, the slide boxes were stored in a refrigerator at 4°C inside an iron and lead shield.

Blanks were prepared in exactly the same way omitting, however, the central layer of emulsion containing the tracer.

The preparation of, say, 10 plates would require about 80 minutes, until drying began. Since the emulsion does not register tracks while the water content is high, the beginning of exposure cannot be timed exactly. It was estimated that the uncertainty of the exposure time of 20 h. was about 10 %.

Preparation of slides with labelled cells.

A series of plates were prepared containing yeast cells labelled with C-14[1]. The yeast, schizosaccharomyces pombe, was cultured for 24 h in a medium containing C-14 glucose. The cells were carefully washed 3 times in distilled water and resuspended in distilled water. In preparing the slides, a dilute suspension of cells was used to dilute the molten emulsion. 1 ml of emulsion containing the labelled cells was pipetted onto the slides and spread to cover the whole surface. Conditions of drying and exposure were the same as described above.

II, 3) Processing

The developer was made up as follows: 2.2 g of sodium sulphite (7 H_2O) were dissolved in 100 ml of dist. water. 0.46 ml of a sodium hydrogen sulphite solution (spec. gravity 1.34) were added to 210 ml of dist. water. These solutions were mixed and 1 g of Amidol was added. The developer was filtered and used immediately.

Development of the plates followed the temperature cycle method of Dilworth et al. (1948), with a modification suggested by Hauser (1959), who obtained more uniform development of thick plates, if the slides were soaked in full strength developer at 5°C and then immersed in dilute developer during the warm stage. Unless otherwise stated, the following routine was adopted.

The slides were taken from the exposure boxes in the refrigerator and placed in developer at 5°C in the refrigerator for 15 min. They were then transferred to a dish containing one part of developer diluted with two parts of distilled water. The dish was kept in a thermostatically controlled water bath at 20°C and the plates were developed in this warm stage for 25 min. They were then transferred to a stop bath of 1 % acetic acid at room

[1] The authors gratefully acknowledge the help of Dr. C. Chapman Andresen, Carlsberg Laboratory, who prepared the C-14 labelled yeast cells.

temperature for 15 min. During this period, the emulsion surface was gently swabbed a few times with moist cotton wool.

Fixation took place in 3 changes of 30 % hypo at room temperature with gentle mechanical stirring. The volume of each change of hypo was 1 liter and the fixation time 7 h. Finally, the slides were washed in running tap water for 3 h and in distilled water for additional 30 min. They were dried in air in a horizontal position and under a dust cover.

II, 4) Measurement of emulsion thickness

Several slides of both types were taken from the dark room after drying (unprocessed) for measurements of emulsion thickness. The slides were measured mechanically (micrometer), optically (microscope), and by weighing. The optical thickness estimates were obtained by cutting a thin channel through the emulsion layer approximately in the middle of the slide. Under the microscope, the upper surface of the emulsion layer and the upper surface of the glass slide exposed in the channel were brought in focus, and the distance between these surfaces was read on the fine adjustment screw of the microscope. The graduation of the fine adjustment provided a direct measure of emulsion thickness.

All slides were weighed before and after application of the emulsion layer. Since the density of the dried emulsion is given by Ilford to be 3.85, the thickness of the layer could be calculated.

The results obtained by the three methods agreed well, although the thickness determined by weighing was consistently less than that found optically or with the micrometer. The former gives a mean value for the whole surface of the slide, while the two others give the emulsion thickness in the centre of the slide. The results are listed in table I.

TABLE I

Emulsion thickness

Method	Sandwich slides	Single layer
Micrometer	130 μ (125–140)	55 μ (50–60)
Weighing	116 μ	50 μ
Optical	—	57 μ (48–68)
Accepted value	130 μ	55 μ

II, 5) Microscopy

In order to reduce the errors involved in measuring distances in the emulsion in a plane perpendicular to the surface of the slide, re-swelling of the emulsion was found advantageous. After photographic processing, the plates were soaked in 20% glycerol for 30 min, excess moisture was removed with a fan and, subsequently, the slides were mounted using a drop of Farrant's medium and a cover slip. The resulting preparation was found to be optically satisfactory. The refractive index of the emulsion was determined to be about 1.56.

The thickness of the emulsion after re-swelling was estimated optically. 21 measurements carried out on 7 different sandwich plates ranged from 100-149 microns with a mean of 125 microns, and this was regarded as sufficiently close to the estimated thickness during exposure of 130 microns.

All optical measurements were carried out by one observer (A.W.R.) using a Leitz ortholux microscope with a Ks×45 objective and ×10 Brillenträger oculars. One ocular contained a graduated scale. The length of this scale was calibrated by means of a reference micrometer slide.

Distances parallel to the surface of the slide were measured with this scale, and distances in the optical axis of the microscope were measured using the graduation of the fine adjustment screw of the microscope.

Track lengths were measured from the first grain of the track, which could easily be identified in most cases because beta tracks show a characteristically higher grain density and larger grain size at the termination of the track. For length measurements, each track was divided into a number of relatively straight sections. The projected length of each section in the plane of the slide as well as the "dip", i. e. the length in the optical axis of the microscope, were determined, and the true distance traversed was calculated as the hypotenuse of the right-angled triangle thus constructed. Where delta tracks were encountered, this fact was recorded and the length of the delta track measured.

As mentioned previously, the number of tracks in the emulsion was adjusted so that tracks seldom overlapped or came so close as to cause confusion. At this track density, it was very unlikely for two tracks to lie end-to-end. Wherever doubt arose whether a particular configuration of silver grains represented one or two tracks, the pattern was interpreted as one continuous track.

Chapter III

General outline of the experiments and their interpretation

III, 1) General

The purpose of preparing so-called sandwich slides was to confine the origins of the beta tracks to a narrow layer of emulsion, separated from the supporting glass slide and the air by at least 50 μ of emulsion. This aim was not fully achieved. Although the majority of the tracks started in the centre of the emulsion layer, the origin of some tracks was only within about 30 μ of the glass support and 25 μ of the upper surface. Presumably, the warm emulsion of the active drop melted the inactive emulsion layer immediately beneath it, and a certain amount of mixing occurred also later when the top layer was added. Diffusion of the labelled ions during exposure cannot be excluded, either.

The majority of the carbon-14 tracks was fully contained in the emulsion. When calcium-45 and chlorine-36 were used, increasing proportions (c. 25 % for Ca and a still higher, but not determined proportion for Cl) of the tracks left the upper and lower surfaces of the emulsion, and a similar increase was noted in the number of tracks re-entering the emulsion from the lower surface due to backscatter from the glass slide.

Most of the sandwich slides showed an artefact: shortly after processing of the emulsion, small brownish grains became visible throughout the emulsion. They were definitely smaller than the blackened grains of the tracks. In the course of a few days, these brownish grains were slowly replaced by large translucent crystals. Beta tracks could often be followed very close to such crystals without noticeable variation in grain size or -density. The nature of this artefact remained obscure. Since neither the visibility nor the grain yield of the tracks was affected, plates showing this artefact were used for counting.

III, 2) Track length and number of grains per track as functions of particle energy

One of the main problems of the present studies was to establish the relationship between beta particle energy and the number of grains produced in the emulsion. However, when using sources emitting a continuous beta spectrum, the initial energy of the particle giving rise to any particular track cannot be specified. Instead, the theoretical "range-energy relation" can be applied (cf. pp. 11 and 32). Zajac and Ross, 1949, found good

correlation between calculated and experimentally determined track lengths when using monoenergetic beta particles and a Kodak NT-4 emulsion. DEMERS, 1958, has calculated the corresponding values for Ilford G-5 emulsion. As might be expected, the two sets of values are practically identical.

An attempt was therefore made towards an indirect determination of the relation between particle energy and number of grains per track by establishing experimentally the ratio between track length and number of grains per track in Ilford G-5 emulsion. On the basis of the theory linking track length and particle energy, it becomes possible to arrive at the desired ratio between the energy of the particle and the number of grains it is likely to produce.

To this end, sandwich slides containing C-14, Ca-45, and Cl-36, respectively, were used and tracks were selected for counting, applying the following criteria (group A, cf. p. 31).

a) The complete track must be contained in the emulsion, i. e. both the beginning and the end of the track must be located at some distance from the surface of the emulsion layer.
b) Tracks were selected which ran predominantly parallel to the plane of the glass slide.
c) Tracks were given preference which did not pass other tracks at very close distance so that confusion could arise.
d) Very tortuous tracks were avoided.

Grain counts were carried out on a total of 101 tracks, namely 30 tracks of C-14, 30 tracks of Ca-45, and 41 tracks of Cl-36. Each track was divided into sections as described on p. 15, the length of each section and the number of grains in it were recorded. The presence of delta tracks was also noted down, and their length and grain count recorded.

The mathematical evaluation of the results is dealt with in a separate chapter (Chapter V.). Summarising the findings, it can be stated that

1) it is justifiable to regard the tracks studied in slides containing C-14, Ca-45, or Cl-36 as one uniform population of tracks;
2) the grain count in, say, the terminal 25 microns of a long track is the same as that of a track whose total length is 25 microns (cf. p. 38);
3) a simple relationship between the mean number of grains per track and the track length has been found, and the same relationship holds for ZAJAC and Ross' determinations;
4) the standard deviation of the present measurements is of the same order of magnitude as that found by ZAJAC and ROSS.

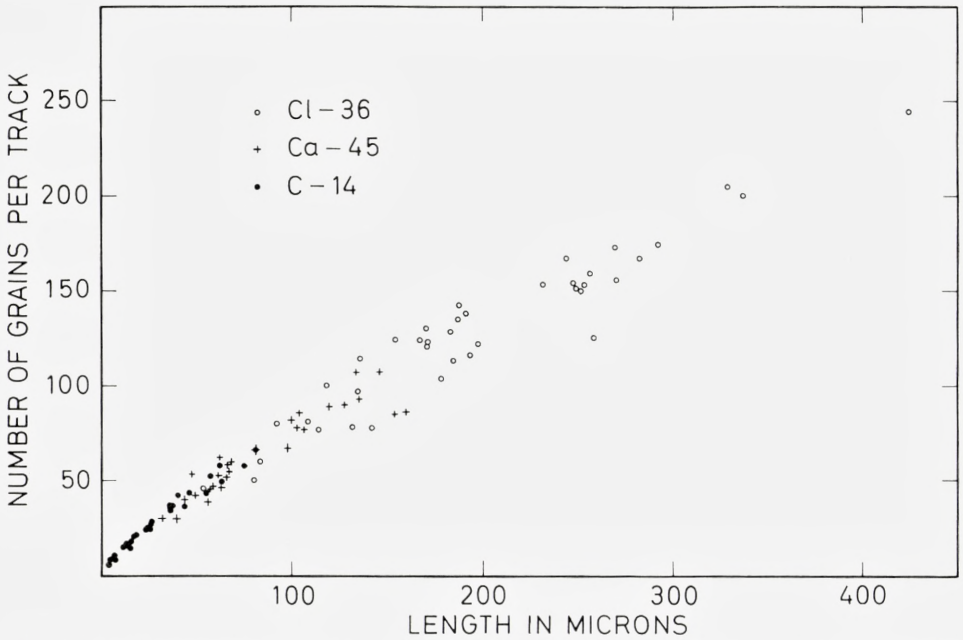

Fig. 1. Number of grains per track vs. track length measured on tracks of C-14, Ca-45, and Cl-36.

These observations are illustrated by Fig. 1 which shows a direct plot of the number of grains per track versus track length.

Using the figures for track length obtained in this series of measurements, the presumed initial energy of each of the beta particles studied was found from the tables given by DEMERS. The results obtained are plotted in Fig. 2 which shows the number of grains per track versus the calculated particle energy. Both the tracks analysed in the present study and those reported by ZAJAC and ROSS are plotted in the figure. The particles studied had energies between 20 and 390 keV. Where a delta track occurred, it was treated as a separate beta particle, in order to estimate its initial energy. The parent particle was assumed to have lost this energy and to have created the corresponding number of grains at the point of origin of the delta track.

The numerical treatment of these problems is likewise described in chapter V.

III, 3) The energy spectrum of C-14 betas in terms of grains per track

The energy spectrum of C-14 depicted in Fig. 3 rises to a low peak at about 25 keV and then falls to the maximum energy of 155 keV.

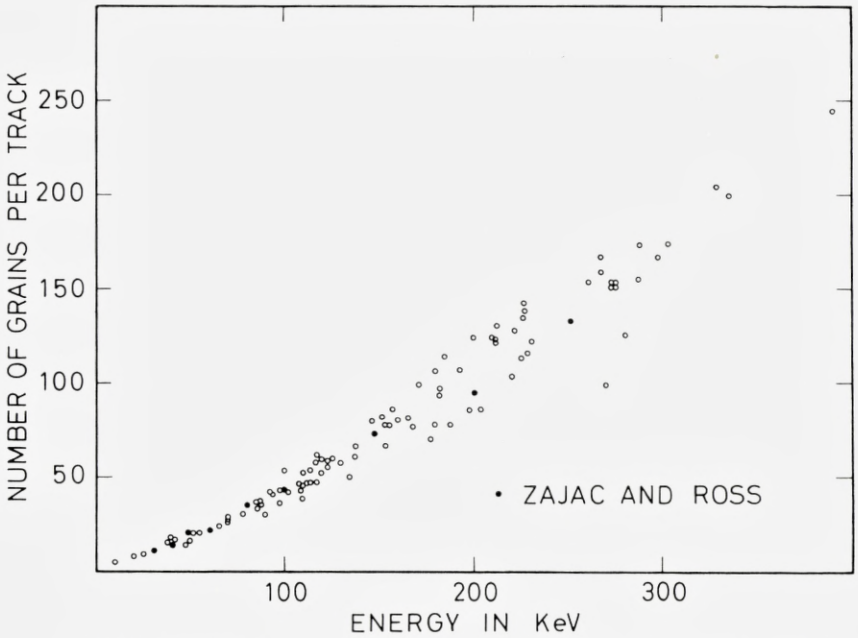

Fig. 2. Number of grains per track vs. presumed initial particle energy.

When calculating the distribution of the number of grains per track on the basis of the energy spectrum, it must be taken into account, however, that we are not dealing with a functional relationship but with a distribution of the number of grains per track for any given energy. On theoretical grounds (Bohr's equation) a considerable straggling in ranges and grain counts can be expected for a group of monoenergetic beta particles. ZAJAC and Ross (1949) found a standard deviation of about 20 % of the mean in their measurements of track length and their grain counts. If the distribution curve for the number of grains per track is corrected, assuming that each initial energy value will result in a population of grain counts, with a Poisson type distribution about the mean, a curve is obtained which is shown in Fig. 3 together with the energy spectrum. A detailed account of this calculation is given in chapter V.

In order to establish experimentally the distribution of the number of grains per track, a large number of randomly selected tracks was analysed on C-14 sandwich slides and on plates containing C-14 labelled yeast cells. The fields to be scanned under the microscope were chosen by predetermined settings of the mechanical stage (cf. LEVI and NIELSEN, 1959).

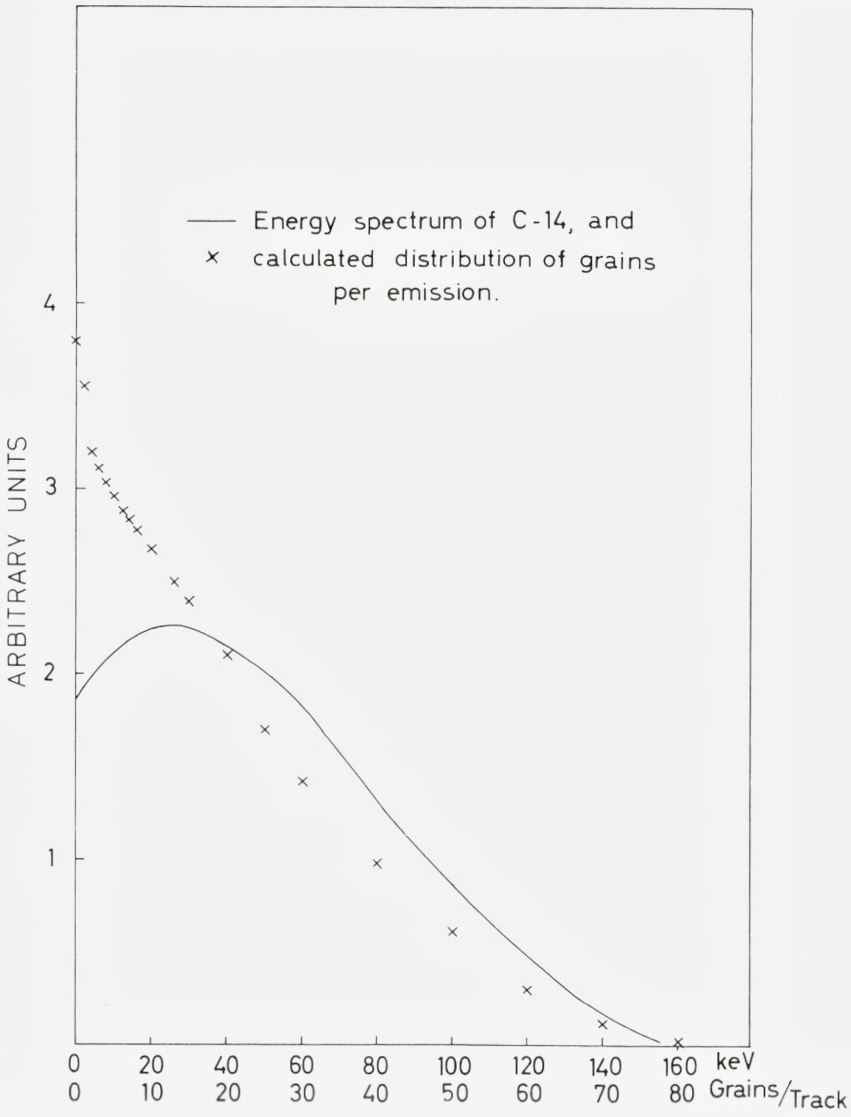

Fig. 3. Energy spectrum of C-14 and calculated distribution of grains per emission.

The following method of choosing tracks to be analysed in these fields of the sandwich plates was adopted (Group B p. 32). (1) All beta tracks lying in the chosen fields, whether complete or leaving the emulsion, were registered and the number of grains was counted. (2) Whenever tracks crossed the boundaries

of the chosen field, an attempt was made to decide whether the track origin or its termination lay within the field. Only the tracks whose origin was within the field were included in the counts.

21 fields were scanned, and 191 beta tracks analysed, of which 163 were completely contained within the emulsion while 28 tracks left the emulsion at either top or bottom.

In order to assess the contribution of background electron tracks to this total, and to estimate their energy distribution, 58 fields were scanned on a blank sandwich plate. 90 background tracks were observed, 53 of them completely contained within the emulsion.

The following criteria were adopted for selecting the tracks to be counted on the slides containing labelled yeast cells (Group E, p. 32).

(1) Only single yeast cells, or pairs of cells, in the emulsion were chosen; fragmented cells or cell clumps were omitted.
(2) Yeast cells lying within 20 μ of the top or the bottom surface of the emulsion were omitted.
(3) Yeast cells indicating the presence of a chemographic artefact were omitted[1].
(4) Yeast cells giving rise to 10 or more tracks were excluded because the difficulty of interpreting the track pattern became too great.

The slide was placed in a predetermined position under the microscope, and moved along a straight line from side to side "east-west". Every yeast cell satisfying the mentioned criteria was used for counting. Every beta track originating from the selected yeast cells was included.

A total of 304 beta tracks, originating from 88 yeast cells was analysed. Considering the frequency of background tracks previously determined on blanks, the probability of including a background track that might originate from the area covered by a selected cell was felt to be so slight as to be negligible.

Qualitatively, the results obtained on both types of plates are in good agreement. The observed distribution of the number of grains per track is depicted in Fig. 4.

[1] The slides containing suspended yeast cells showed a chemographic artefact: Where the cells clumped together, a solid mass of silver often surrounded them. However, this mass was completely different in appearance from the discrete grains of the beta tracks. It has been suggested that ruptured cells might be responsible both for the clumping and the chemography. On the same plates, numerous individual cells and pairs of cells were present which were free from silver deposits, and clearly recognisable beta tracks originated from them. These cells were therefore used for counting.

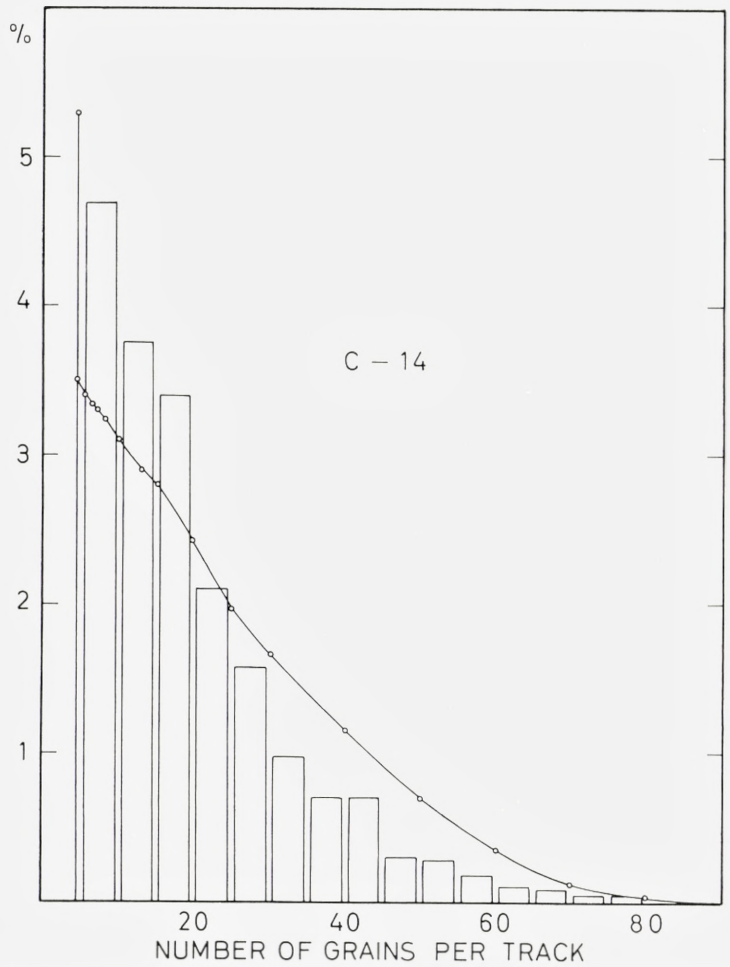

Fig. 4. Observed distribution of the number of grains per track for C-14 (histogram) and calculated distribution (cf. Fig. 3). The first bar of the histogram has been condensed into the line at 4 grains per track.

It appears from Fig. 4 that the shape of the calculated distribution curve agrees fairly well with the observed distribution, although the number of short tracks is somewhat higher than expected.

Since only 4 or more grains in a row form a recognisable track, 14 % of all C-14 beta particles (the sum of the first 4 ordinates) will not be recognised as tracks, and 4 % will not give rise to any grains at all under the conditions of this experiment, as calculated on p. 45.

III, 4) The energy distribution curve of Ca-45 betas in terms of grains per track

An attempt was also made to count the number of grains per track in a Ca-45 sandwich plate, and to compare the values observed with the predicted distribution. The criteria for selecting the tracks and for assessing the contribution of random background tracks to the observed total were the same as described on p. 20.

Thirteen fields were scanned and 195 beta tracks analysed, of which 136 were completely contained in the emulsion and 59 left the emulsion either at the upper or lower surface. As was to be expected, the statistical analysis of these figures was complicated by the higher proportion of tracks leaving the emulsion. Nevertheless, the agreement between the predicted and the observed distribution as illustrated by Figs. 5 and 6 is very close if the straggling effect (cf. above) is taken into consideration. 10 % of the Ca-45 beta particles must be expected to give rise to less than 4 grains under the conditions of this experiment.

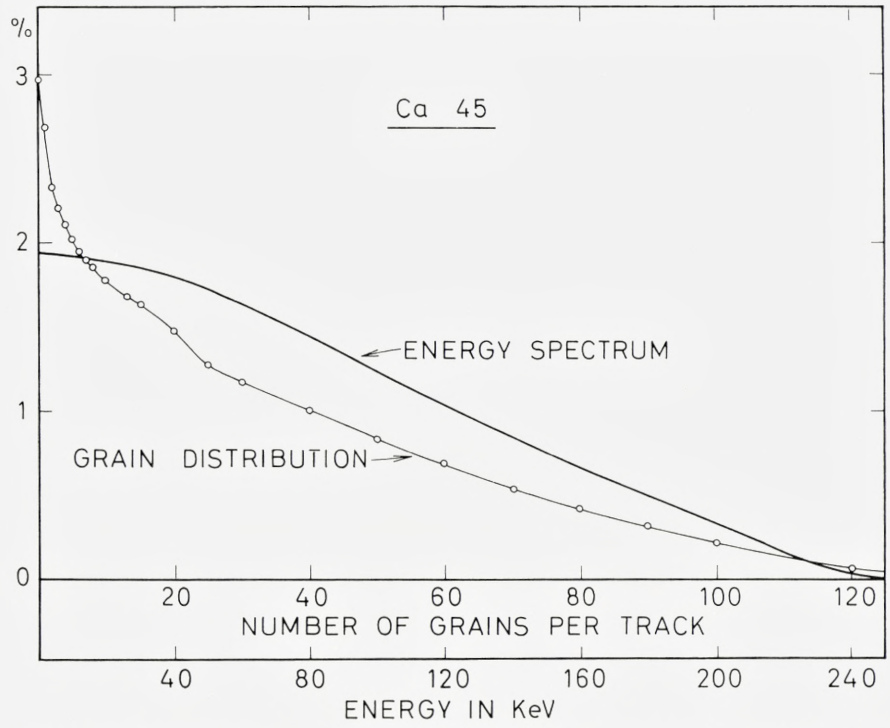

Fig. 5. Energy spectrum of Ca-45 and calculated distribution of grains per emission.

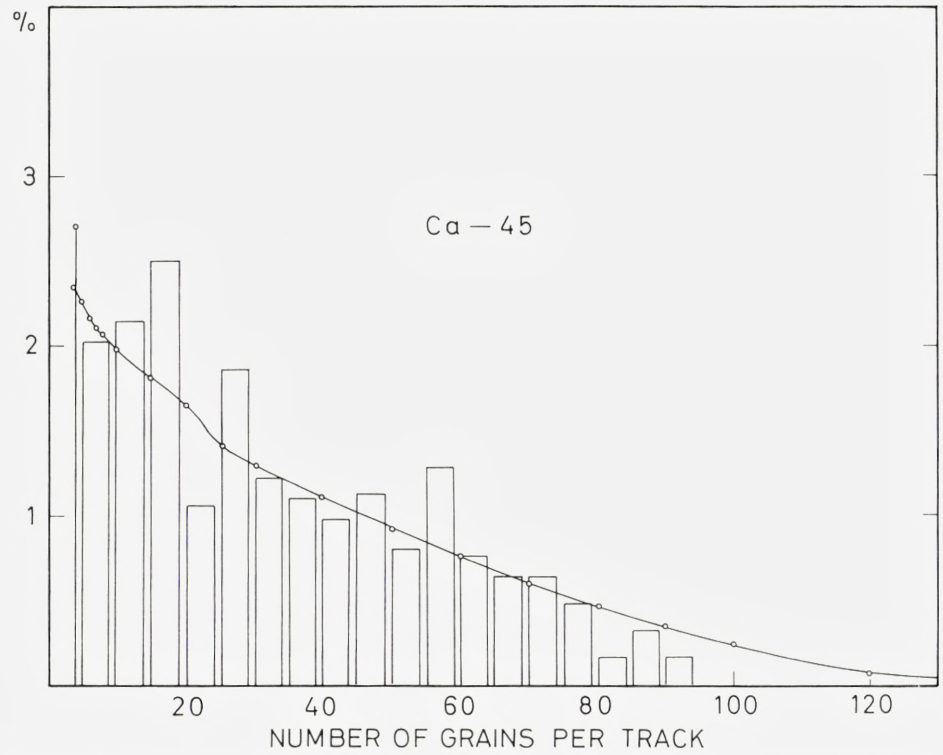

Fig. 6. Observed distribution of the number of grains per track for Ca-45 (histogram) and calculated distribution (cf. Fig. 5). The first bar of the histogram has been condensed into the line at 4 grains per track.

III, 5) The distribution of grains in space relative to the track origin

As discussed in the Introduction (p. 8) the simplest model to be studied with the aim of absolute quantitation must be a point source of a beta emitter surrounded by emulsion on all sides. The questions to be answered are 1) how many grains (or tracks) are being produced in the emulsion by a given number of disintegrations, and 2) what is the radius of the sphere around the source in which all grains (or tracks) will be contained? Therefore, the distance between the origin of a track and the point furthest away is of the greatest interest.

Little information could be found in the literature on the relationship between the track length and the radius of the sphere that will contain the entire track. In order to investigate this problem, two types of measurements were carried out on sandwich slides containing C-14 and Ca-45.

While the counts reported on pp. 20 and 23 were in progress, the distance between the first grain of the track and the grain lying furthest away was measured on every complete beta track encountered. Such measurements were made on 163 tracks of the C-14 slides and on 136 tracks of the Ca-45 slides.

The second type of measurement was performed on the C-14 plates, only (Group D p. 32). Tracks completely contained in the emulsion were divided into sections, each of which could be approximated by a straight line. The positions of the beginning and the end of each section relative to the first grain of the track were measured, and the number of grains in each section recorded. In this way,—and assuming fairly regular spacing of the grains within each section,—a three-dimensional picture of the position of each grain in each track relative to the track origin (the first grain) was obtained. 100 C-14 tracks were measured in this way. From this second type of measurement, the radius of each track can likewise be determined.

The figures obtained for C-14 and Ca-45 were found to constitute a homogeneous population. The mathematical derivation of the function which links the number of grains per track to its radius in microns is given on p. 40 and the distribution of radii is given in Table VI. Using the equations derived in section V, 3 for the interdependence of track length, particle energy, and number of grains per track, an equation is obtained that relates track length with radius, as well as initial particle energy with radius. It follows, for instance (cf. p. 42) that a sphere of G-5 emulsion with a radius of 20 microns around a point source of C-14 will contain 90 % of the beta tracks originating from this source. The corresponding radius of a sphere around a point source of Ca-45 is 45 microns (cf. p. 43).

For C-14, additional information can be obtained from the second type of measurement mentioned above. If the origins of the tracks examined are superimposed, a three-dimensional picture of the distribution in space of silver grains around this model point source can be constructed. Fig. 7 illustrates this distribution graphically. It follows that, for a point source of C-14 completely surrounded by G-5 emulsion, 29 % of the grains produced will lie within 5 microns of the source, 50 % within 9 microns, 75 % within 17 microns and 90 % within 25 microns. The second curve given on Fig. 7 is discussed on p. 45.

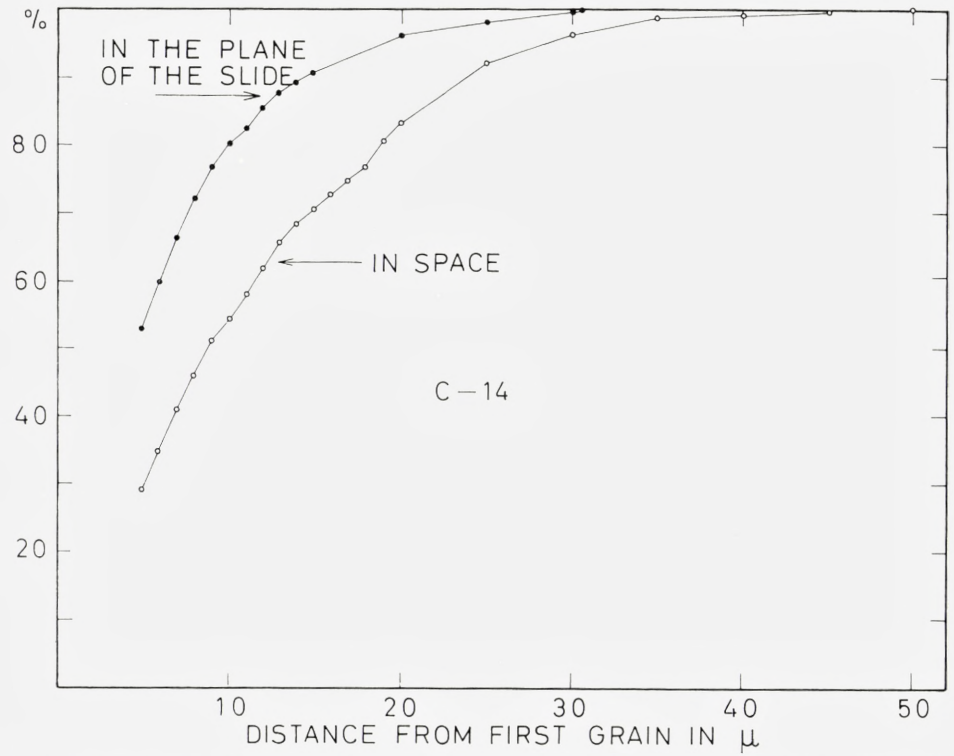

Fig. 7. Distribution in space, and in the plane of the slide, of silver grains around a model point source of C-14.

Chapter IV

Sources of error

The conclusions to be drawn from the experiments described in the preceding chapters rest entirely on the measurements of distances between points in the emulsion, and on grain counts. Some of the sources of error affecting these results are inherent in the processing and re-swelling of the emulsion and the resulting recognisability of detail. Others lie in the optical length and depth measurements, as well as in the criteria used for selecting the tracks to be studied. A critical discussion of these factors is therefore presented in the following sections.

IV. 1) Processing

As discussed earlier, (p. 10) processing conditions have a considerable influence on grain- and track yield. In particular, it was found desirable to investigate the effect of over- or underdevelopment, and to find the optimal development time for the plates prepared in the present study. A separate experiment was therefore carried out using a series of sandwich slides containing Ca-45. The plates were processed as described on p. 13, but the duration of the warm stage of development was varied between 5 min and 40 min. The criteria for the selection of the tracks (Group C p. 32) to be analysed were similar to those mentioned on p. 17. In order to obtain the desired information, only tracks longer than 75 microns were chosen. Using these criteria, the number of grains in the terminal 75 microns of the selected tracks was counted.

After only 5 min development, it proved extremely difficult to find tracks of the desired length. The terminal 20-30 microns could often be identified, but grain spacing was so irregular that the course of the earlier part of the track could only be guessed. This difficulty was not encountered after 10 min development. With increasing development time, the grain size as well as the number of grains increased. After 20 min development, the large grains in the terminal few microns of the tracks tended to fuse into a solid line of silver, which made grain counting difficult in this portion of the tracks. The random grain background was considerably higher after 40 min development as compared with the shorter processing times.

Twelve tracks were analysed at each of the following development times: 5, 10, 15, 20, 25, and 40 min.

The findings are illustrated in Fig. 8 where the grain counts in the terminal 75 microns of track are plotted against the development time. The mean grain counts increased considerably between 5 and 10 min, and more slowly from 10 to 15 and 25 min of developing, remaining practically constant up to 40 min. The grain densities expressed as number of grains per 25 micron sections, measured from the termination of the track, showed a similar pattern for all sections up to 125 microns (cf. p. 38).

From these observations it follows that significant fluctuations in grain yield did not occur as a result of slight variations in development time between 15 and 40 min in the particular conditions of this experiment. Clearly, the length of the plateau shown in Fig. 8 would have been shorter, and its slope steeper, if a more powerful developer had been used, or the temperature of development had been higher. Under different conditions,

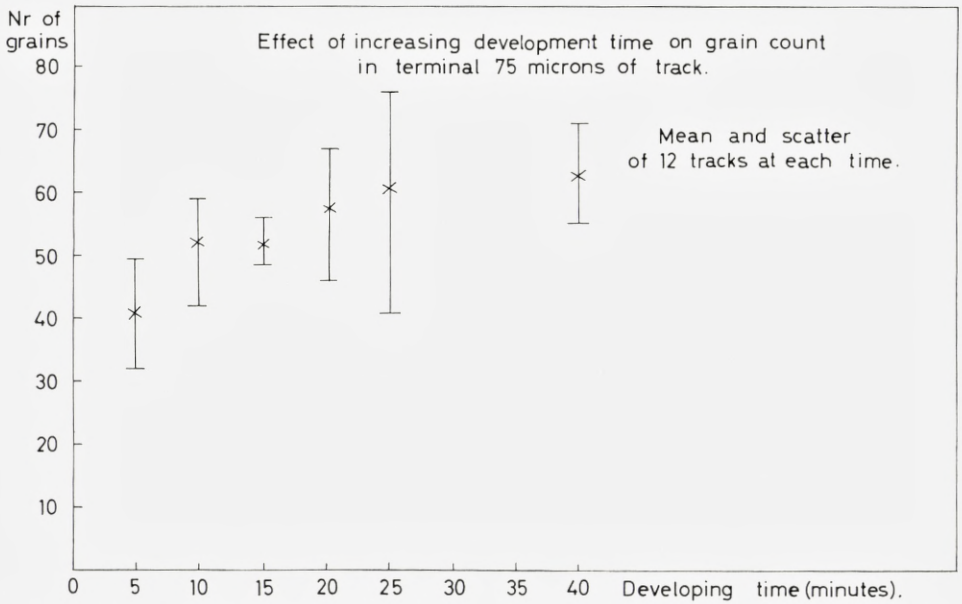

Fig. 8. Effect of increasing development time on grain count in terminal 75 microns of track.

the effect of development time might have been more critical. As mentioned in section II,3, the development time used in the present work was 25 minutes.

IV, 2) Reswelling of the processed emulsion

During fixing, the emulsion layer shrinks considerably, and the grains become superimposed very closely in the optical axis of the microscope. When a steep dip of the track occurs, individual grains cannot be identified. Similarly, delta tracks originating at right angle to the surface of the slide may be missed entirely. In sections of the track that lie perpendicular to the surface of the slide, grain counts are very difficult. ZAJAC and ROSS (1948, 1949) likewise encountered some difficulties in interpreting their track patterns due to the fact that the processed emulsion was viewed without reswelling. Recognition and counting of grains as well as measurements of distances are definitely much easier in the reswollen emulsion. In this connection, it should be emphasized that the discrepancy between the number of grains per track for carbon-14 reported by LEVI and NIELSEN (1959) and the results obtained in the present study is probably due to the different thicknesses of the processed plates. The material from which

the earlier results were derived was re-examined recently. Since the plates prepared by LEVI and NIELSEN had a higher track density, and the emulsion was viewed without reswelling, the track patterns were interpreted differently: whenever portions of a track lay at right angles to the plane of the slide, the heavily superimposed grains gave an appearance which was erroneously interpreted as the dense termination of a track; the continuation of a row of grains at a different focal level was assumed to represent the beginning of a new track. In this manner, long tracks were divided into a number of shorter ones. Undoubtedly, the figures presented in the present paper should be taken in preference to those of LEVI and NIELSEN.

IV, 3) Accuracy of track length measurements

a) When measuring track length, the irregular path of the beta particle was approximated by a series of straight lines. Even in the case of high energy beta particles, producing tracks which appear fairly straight over long distances, multiple scattering occurs between one grain and the next, and hence, the distance travelled by the particle is always greater than the distance measured between the end points of a fairly straight section. ZAJAC and Ross worked with the oil immersion objective of a Vickers projection microscope, and measured the distance from one grain to the next. This very time consuming method should lead to more accurate results; however, the results obtained by ZAJAC and Ross agree so closely with those reported in the present paper, that the simpler method of length measurement applied in the present study seems justified.

b) Since beta tracks originate from sources of molecular dimensions, the true point of origin of each track cannot be determined. It was therefore tentatively assumed that the tracks start from the first recognisable grain. In tracks of a total length of about 50 microns, the grains in the first section of the track were approximately 1 micron apart. In tracks 150 microns long, or more, the first grains lay about 2 microns apart. The total track lengths as stated in the present work have therefore been underestimated by not more than 1 or 2 microns, depending on the total length.

c) An estimate of the reproducibility of the measurements in the optical axis of the microscope was obtained in connection with the study of the spatial distribution of grains relative to the track origin. Since it must be assumed that beta particles travel randomly in all directions, the distribution of silver grains around a point source should be spherically symmetrical.

As described on p. 25, the spatial distribution of silver grains recorded for a large number of randomly selected C-14 tracks was evaluated. When

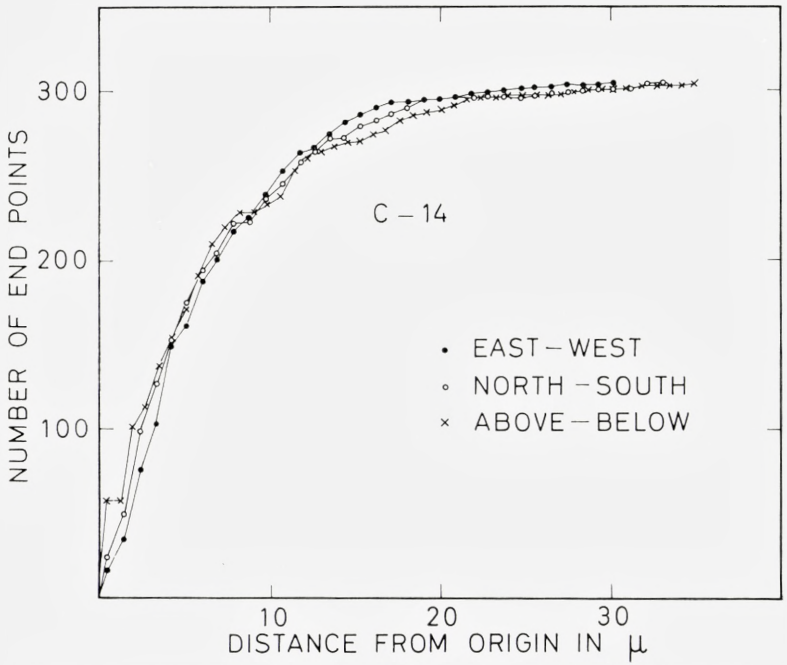

Fig. 9. Distribution of grains about the point of origin of the tracks.

the track origins of all tracks analysed were superimposed, the pattern of grains recorded was found to be symmetrical about the origin with respect to measurements taken in the plane of the slide. The distance measurements in the optical axis of the microscope were more variable, with a greater proportion of zero distances than were observed in the plane of the slide. Fig. 9 may serve to illustrate the distribution of grains about the point of origin of the tracks as being spherically symmetrical, if the very short distances of less than 1 micron which seem to have been recorded as zero, are excluded. This confirms the statement that reswelling restored the processed emulsion very closely to its original thickness.

IV, 4) The accuracy of grain counting

The only appreciable source of error in grain counting lies in the evaluation of the last few microns of the tracks, where the grains can be so large and closely spaced that the silver appears as a rather uneven blob, instead of a group of discrete grains. Since the chances for a short delta

track to occur near the termination of the track are likewise high, the number of grains may be underestimated. This error is likely to be consistent for each observer.

IV, 5) The selection of tracks to be analysed

The criteria for selecting tracks to be analysed, out of the practically infinite population of tracks available on each plate, have been stated in the previous chapters. However, some comments are needed if the possibility of a bias is to be evaluated. When the relationship between track length and number of grains per track was studied, very tortuous tracks, and tracks running predominantly in the optical axis of the microscope were avoided. Thereby, length measurements and grain counts were greatly simplified, but other errors may have been introduced.

A beta particle loses most of its energy in multiple scattering whereby a succession of small deviations from its course occur. In addition, abrupt loss of a considerable amount of energy will occur resulting in the formation of a delta track whose length can be measured, provided the delta track lies predominantly in the plane of the slide. If the abrupt loss of energy is due to a nuclear collision, however, the track will show a marked change of direction at the point of collision, but the energy loss cannot be estimated by examining the emulsion. In selecting tracks according to the criteria mentioned above, it is likely that tracks with relatively few abrupt changes in direction were chosen at the expense of those with more delta tracks and more frequent nuclear collisions.

The data thus obtained were used to derive the equation relating initial energy of the particle to the number of grains produced; the resulting equation may describe a situation in which a higher proportion of the particle energy is dissipated by multiple scattering than is normally the case.

Chapter V

The mathematical evaluation of results

V, 1) Groupings

Five series of measurements were carried out in the course of this investigation and, for convenience, they will be referred to in the manner given below.

Group A Selected beta tracks from C-14, Ca-45, and Cl-36 were divided into segments, and the length of each segment as well as the number of grains in it were recorded (cf. p. 17).

Group B Using C-14 plates, Ca-45 plates, and blank plates, randomly selected beta tracks were examined, and both the radius of each track and the number of grains in the track were recorded (p. 20). For each track, a note was made of whether it was completely contained in the emulsion or whether it passed through the top or bottom surface of the emulsion.

Group C A series of Ca-45 plates that had been developed for periods ranging from 5 minutes to 40 minutes was examined (p. 27). Selected beta tracks were divided into segments and the length of each segment, as well as the number of grains in it were recorded.

Group D Randomly selected C-14 tracks were studied (p. 25). Each track was divided into segments and the co-ordinates of the beginning and the end of each segment in three dimensions relative to the track origin (first grain) were recorded, as well as the number of grains in each segment.

Group E The C-14 beta tracks originating from selected yeast cells were examined (p. 21). The number of tracks per cell and the number of grains per track were recorded.

V, 2) The relation between track length, number of grains, and initial energy of the particles

Zajac and Ross (1949) determined the mean track length and the mean grain count for 9 groups of mono-energetic beta particles in Kodak NT 4 emulsion. Tables II and III summarize some of their findings.

Table II

Zajac and Ross' values for mean track length and mean grain count (loc. cit., 1949, table 1).

Particle energy keV	Mean track length (μ)	Mean grain count
E	$\bar{L}$	$\bar{G}$
30	7.0	11.0
40	10.8	13.8
50	15.8	20.4
60	21.4	22.4
80	32.7	35.5
100	46.7	43.3
147	95.4	74.2
200	141	95
250	201	133

TABLE III

Zajac and Ross' values of mean number of grains per micron ($\bar{g}$) at given distance (L) from termination of track (loc. cit., 1949, table 2).

L	$\bar{g}$	L	$\bar{g}$
2.5	1.49	70	0.584
7.5	1.11	90	0.593
12.5	0.93	110	0.545
17.5	0.90	130	0.540
25	0.761	150	0.501
35	0.766	170	0.515
45	0.680	190	0.495
55	0.700	..	..

In Fig. 10 the logarithm of the initial energy of the particles in keV ($\log E$) is plotted against the logarithm of the mean track length in microns ($\log \bar{L}$). The evidently linear relationship between $\log E$ and $\log \bar{L}$ can be described by the equation

$$\log \bar{L} = 1.59 \log E - 1.51 \tag{1}$$

In the same figure, $\log E$ is plotted versa the logarithm of the mean number of grains per track ($\log \bar{G}$) and this relation is described by the equation

$$\log \bar{G} = 1.19 \log E - 0.740 \, . \tag{2}$$

From equations (1) and (2), the functions linking track length and number of grains can be derived, viz.

$$\log \bar{G} = 0.747 \log \bar{L} + 0.385 \tag{3}$$

or

$$\bar{G} = \bar{L}^{0.747} \; 10^{0.385} = 2.43 \; \bar{L}^{0.747} \, . \tag{4}$$

The percentage standard deviation reported by ZAJAC and Ross is fairly constant; the coefficient of variation being approximately 20 % for the track lengths and slightly lower for the grain counts. However, the coefficients of variation for the grain counts are not incompatible with those predicted by

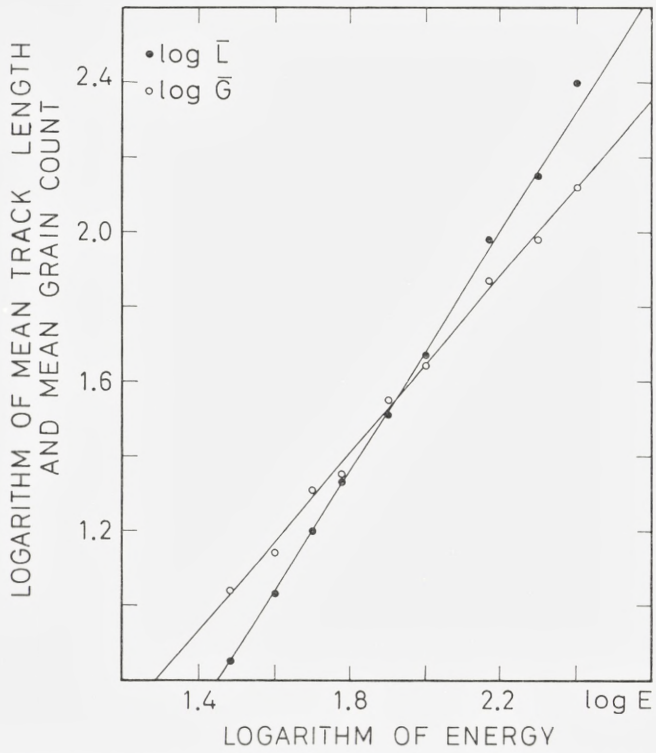

Fig. 10. Zajac and Ross' values for mean track length and mean grain count as a function of initial energy of the particles (log-log scale). These values are listed in Table II.

the Poisson distribution, except for the higher energy groups, where the experimentally determined coefficients are higher than those from the Poisson distribution.

In view of the similarity in composition between Ilford G 5 and Kodak NT 4 emulsion, the relation between particle energy and track length is expected to be the same in the two emulsions. In fact, ZAJAC and ROSS' figures for mean track length agree very well with those predicted on theoretical grounds for both emulsions.

From the group *A* measurements (cf. p. 31), track length and number of grains per track have been determined, but the initial energies of the particles cannot be stated. Fig. 11 shows the number of grains per track plotted against $L^{0.747}$, and, in addition, the energies corresponding to different values of L (equation 1) are plotted on the abscissæ. It can be seen that the

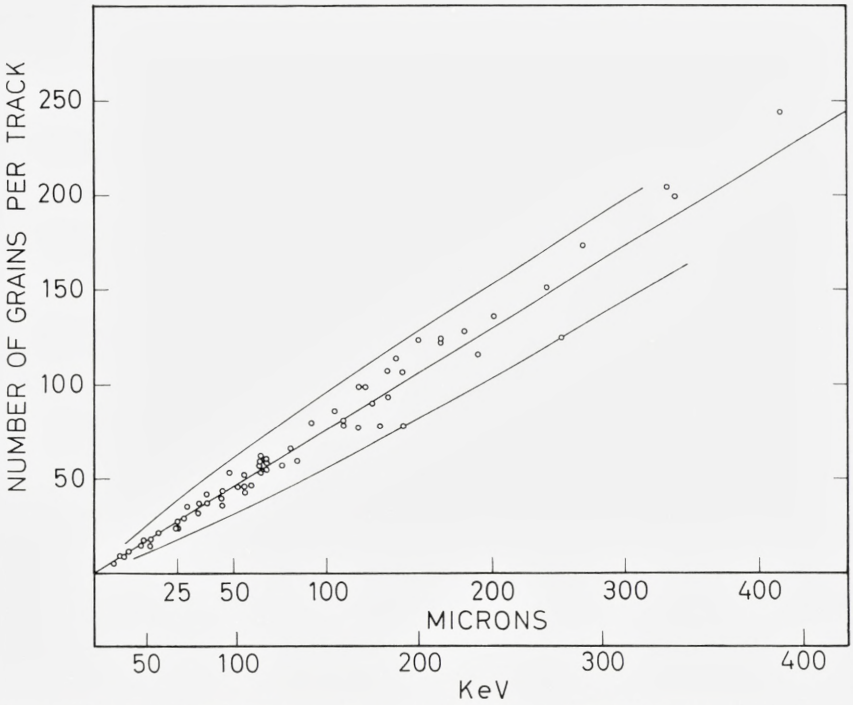

Fig. 11. Number of grains per track versus $L^{0.747}$ and versus initial energy of the particle (cf. equations (2) and (4), section V, 2; Group A).

results obtained for C-14, Ca-45, and Cl-36 tracks form a homogeneous group. The 95 % confidence limits calculated on the assumption of a Poisson distribution with a mean of $2.43\ L^{0.747}$ are shown on the graph. All points observed lie within these limits, which indicates a close agreement with equation (4). Furthermore, the experimental variance of the results is of the same order as that found by ZAJAC and ROSS, or even slightly lower.

In Fig. 12, the number of grains per track versus track length of 100 randomly selected C-14 tracks of group *D* are plotted in a log-log scale. The straight line corresponding to equation (4) is drawn up in full, and it can be seen that this line does not agree too well with the observed points, while the dashed line which has almost the same slope but is placed somewhat below the first mentioned line on the average fits the observed points much better. In addition, it will be noted that the scatter of the points around the straight line is clearly greater for group *D* tracks as compared with tracks of group *A* (cf. Fig. 11).

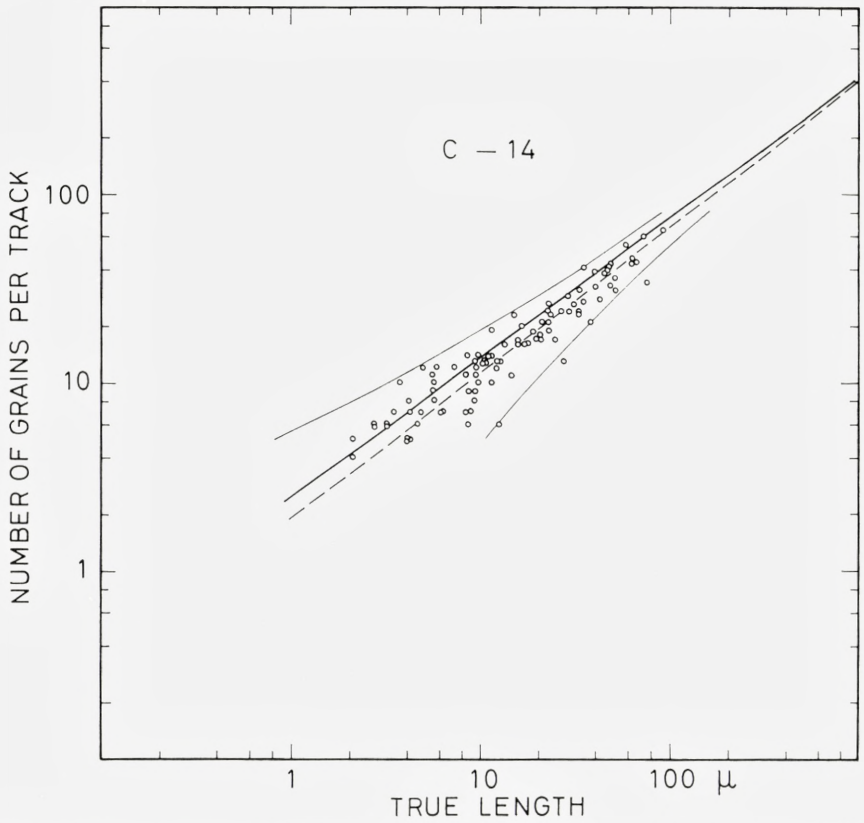

Fig. 12. Number of grains per track versus track length of 100 randomly selected C-14 tracks (Group D; log-log scale).

The difference in the results obtained from these two groups of tracks is undoubtedly due to the fact that the group *A* tracks have been selected so as to be easily recognisable and not very tortuous (cf. p. 17) whereby a lower scatter of the results was obtained. Furthermore, it is reasonable to assume that the difficulties encountered when evaluating the number of grains in sections of track running perpendicular to the plane of the glass slide led to an underestimate of the number of grains per track in group *D*. The scatter of the points of group *D* tracks agrees well with the expected Poisson distribution, while that of group *A* is lower than expected.

Since a beta particle with the initial energy E_0 loses energy on its passage through the emulsion, its energy at a given point in the track will be reduced to, say, E_1. It seems reasonable to assume that the length of the

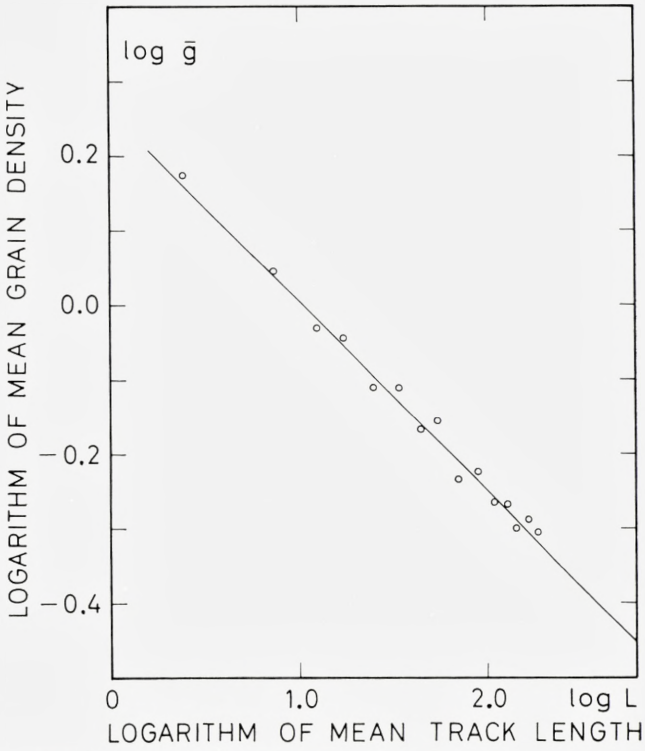

Fig. 13. Zajac and Ross' values of grain density versus track length (log-log scale). These values are listed in Table III.

remaining portion of the track, and the number of grains in that portion, will depend on E_1 in exactly the same manner as for another track produced by a particle with the initial energy E_1. In other words, the terminal section of any track will be independent of the initial energy of the particle and, on the average, independent of whether the complete track is long or short.

On this assumption, the average grain density $\bar{g}$ at a distance L from the termination of the track can be determined by differentiation of equation (4). This leads to

$$\bar{g} = \frac{d\bar{G}}{d\bar{L}} = 2.43 \times 0.747 \ \bar{L}^{-0.253} \quad \text{or} \tag{5a}$$

$$\log \bar{g} = 0.259 - 0.253 \ \log \bar{L} \, . \tag{5b}$$

In Fig. 13, ZAJAC and Ross' values are plotted with $\log \bar{g}$ as the ordinates and $\log L$ as abscissæ. There is very good agreement between these observed values and the straight line calculated on the basis of equation (5b).

TABLE IV

Mean number of grains per 25 microns of track (distance measured from termination).

Distance from end μ	Number of tracks			Mean number of grains.				Theory (Equ. 4)	Variance			
	C-14	Ca-45	Cl-36	C-14	Ca-45	Cl-36	total		C-14	Ca-45	Cl-36	total
0– 25....	17	19	33	26.0	26.2	25.5	25.8	26.9	7.2	11.5	35.8	21.6
25– 50....	6	16	30	17.7	20.1	18.4	18.7	18.3	9.0	10.9	14.5	13.4
50– 75....	2	8	28	16.6	18.9	16.9	17.2	16.0	2.0	4.2	13.3	10.8
75–100....	..	8	24	..	15.4	15.2	15.3	14.7	..	7.1	9.2	8.4
100–125....	..	4	18	..	14.7	14.5	14.5	13.7	..	3.9	7.7	6.8
125–150....	..	..	14	..	..	14.5	14.5	13.1	..	..	8.0	8.0
150–175....	..	..	9	..	..	12.2	12.2	12.5	..	..	7.1	7.1
175–200....	..	..	7	..	..	12.5	12.5	12.1	..	..	4.9	4.9
200–225....	..	..	7	..	..	12.7	12.7	11.7	..	..	2.9	2.9
225–250....	..	..	7	..	..	12.6	12.6	11.4	..	..	4.0	4.0

Tables IV and V give the results obtained from group *A* tracks. The straight segments into which these tracks originally were divided for the purpose of measurements, were of variable length. For convenience of comparison, and in agreement with the considerations outlined above, the tracks listed in tables IV and V have been divided into sections 25 microns long, starting at the termination of the tracks. The number of grains in each such section has been calculated. Table IV records the mean number

TABLE V

Mean number of grains per 25 microns acc. to total length of track and distance from end.

Total length of track μ	Distance from end of track					
	0–25 μ		25–50 μ		50–75 μ	
	number of tracks	mean no. of grains	number of tracks	mean no. of grains	number of tracks	mean no. of grains
25– 50.....	13	25.8	..	..	..	..
50– 75.....	14	26.4	14	18.3	..	..
75–150.....	19	24.1	18	20.1	18	16.2
150–250.....	14	28.0	11	17.5	11	17.8
250–	9	25.3	9	18.2	9	16.0

of grains found in 25 μ sections of track for beta particles from each of the three isotopes studied. In table V, the mean grain numbers in the three 25 micron sections nearest the termination are listed for tracks of different total length.

It appears from these tables that the grain density, viz. the mean number of grains per 25 micron section of track, is independent of the isotope and of the total track length. Moreover, it can be seen that these values are in close agreement with equation (4) (column 9 of Table IV). The grain density decreases from the termination of the track towards its origin, reaching a reasonably constant figure above 150 microns from the termination. The variance of grain density decreases according to a similar pattern. The high variance found in the terminal 25 microns of track probably reflects in part the great variability in track pattern found in this region, and the increasing difficulty of grain counting when the grains lie very close together.

When delta tracks occur, the relation between total number of grains and total track length cannot be described by equation (4), since both branches represent track terminations. If delta tracks were to be compared with unbranched tracks, one of the branches should be interpreted as a middle section of an unbranched track, and this would involve a different relation between length, number of grains, and energy. A correction could be applied in such cases, but since this correction would be based on an assumption we have set out to prove, it was considered preferable to omit branching tracks from Fig. 11. However, the longest branch of each of these tracks is included in the calculation of grain density listed in tables IV and V.

V, 3) The relation between track length, radius, and distribution of grains in space

Direct measurements of the radius and the number of grains per track are available from group *B* for C-14 and Ca-45, and from group *D* for C-14. These figures are useful, only, if the tracks are fully contained in the emulsion.

For group *D*, it was possible to calculate track length as well as radius. These results are plotted in Fig. 14. In the narrow range from c. 20 keV to 150 keV, the relationship between track length and radius can be expressed as

$$\log R = 0.816 \log L + 0.042. \qquad (6)$$

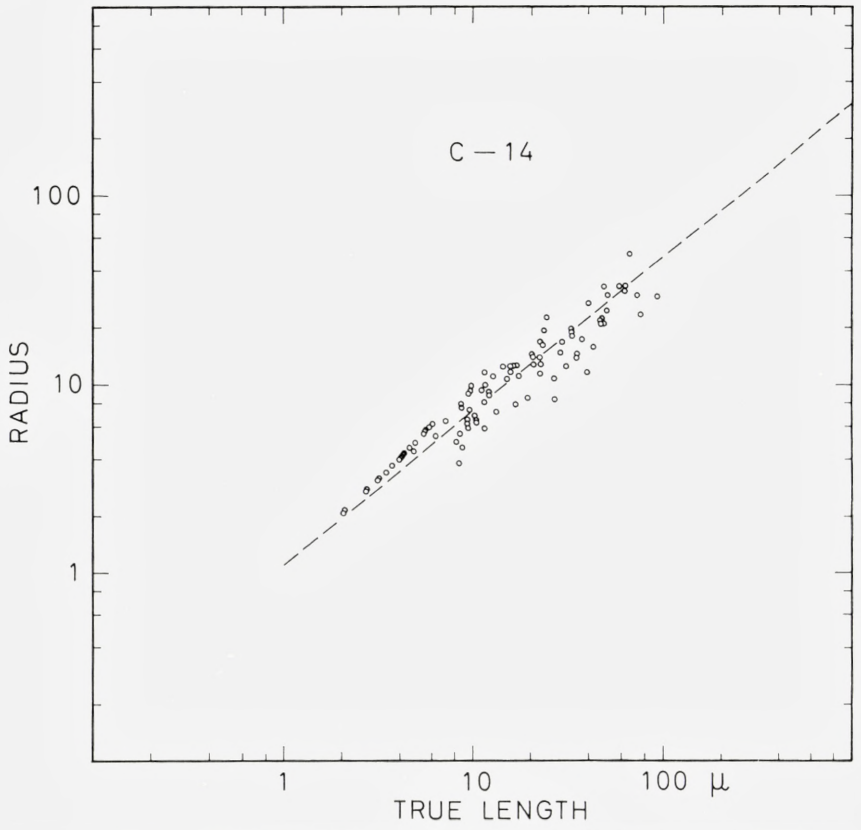

Fig. 14. Track length versus radius for C-14 (Group D; log-log scale).

It can be seen that the points scatter considerably about the mean. Using the relation between $\log G$ and $\log L$ derived from the dashed line in Fig. 12, viz.

$$\log G = 0.279 + 0.774 \log L , \tag{7}$$

the following relation between $\log G$ and $\log R$ is found

$$\log G = 0.229 + 0.949 \log R . \tag{8}$$

Figs. 15, 16, and 17 illustrate that this relation fits the C-14 tracks of groups B and D and also the Ca-45 tracks of group B. Also here, the scatter of the points about the mean is fairly wide.

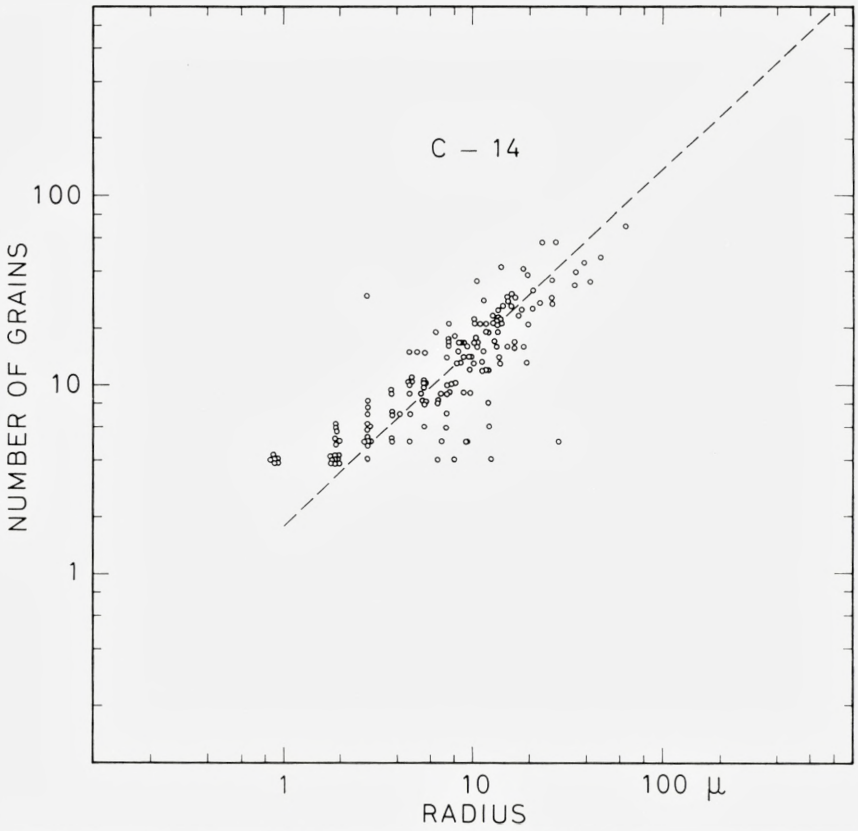

Fig. 15. Number of grains per track versus radius for C-14, (Group *B*, log-log scale).

Since the equations linking the presumed initial energy of a beta particle with its mean track length and mean number of grains have been calculated already, it is possible to apply these equations in order to convert equations (6) and (7) into a function describing the relationship between initial energy and mean radius

$$\log E = 0.784 \log R + 0.967. \tag{9}$$

It must be emphasized, however, that this relationship holds in a very limited range of energies, only (about 20–200 keV). Moreover, the variance of this particular function has not been estimated, and it is likely to be very large. In the absence of any information concerning the shape and width of the distribution curve about the mean radius for a group of mono-

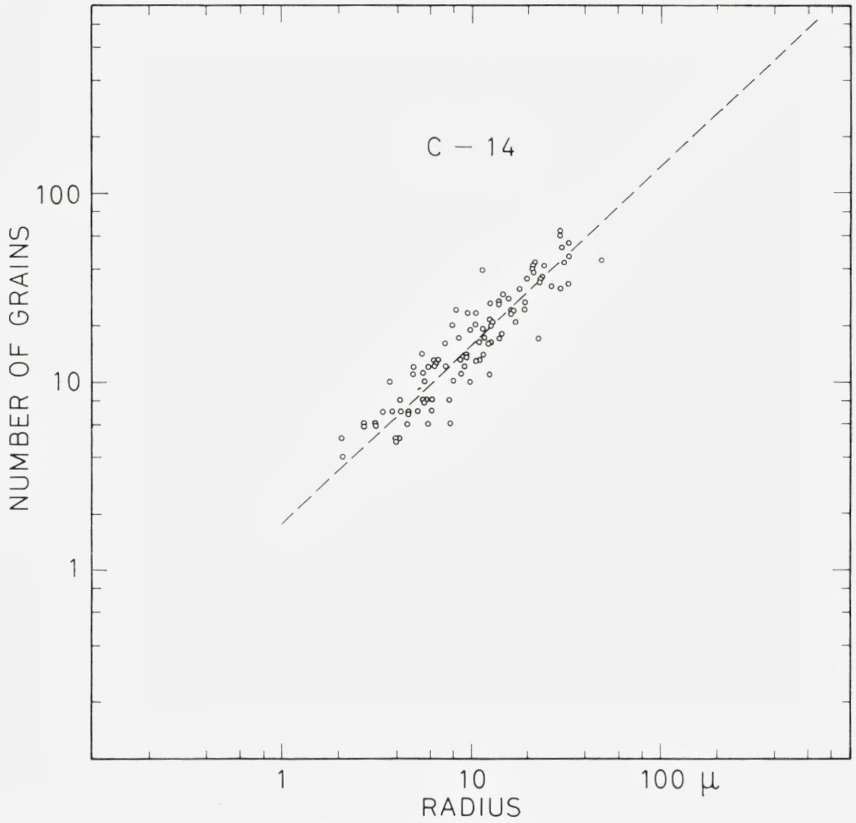

Fig. 16. Number of grains per track versus radius for C-14, (Group *D*, log-log scale).

energetic electrons, it is impossible to predict the probability of finding a C-14 track with a radius greater than a given value.

The data available for C-14 in the measurements of groups *B* and *D*, however, comprise 263 randomly selected tracks. Corrected for particles that have given rise to less than 4 grains (cf. p. 22), this represents 306 disintegrations.

It seems reasonable to assume that all disintegrations giving rise to less than 4 grains have radii shorter than 5 microns. Table VI gives the frequency distribution found for C-14 tracks of different radii. It follows that a sphere of emulsion of a radius of 20 microns around a point source of C-14 will contain about 90 % of the disintegrations.

The data available for Ca-45 are based on fewer tracks and the computation is complicated by the unknown fraction of tracks of group *B* that

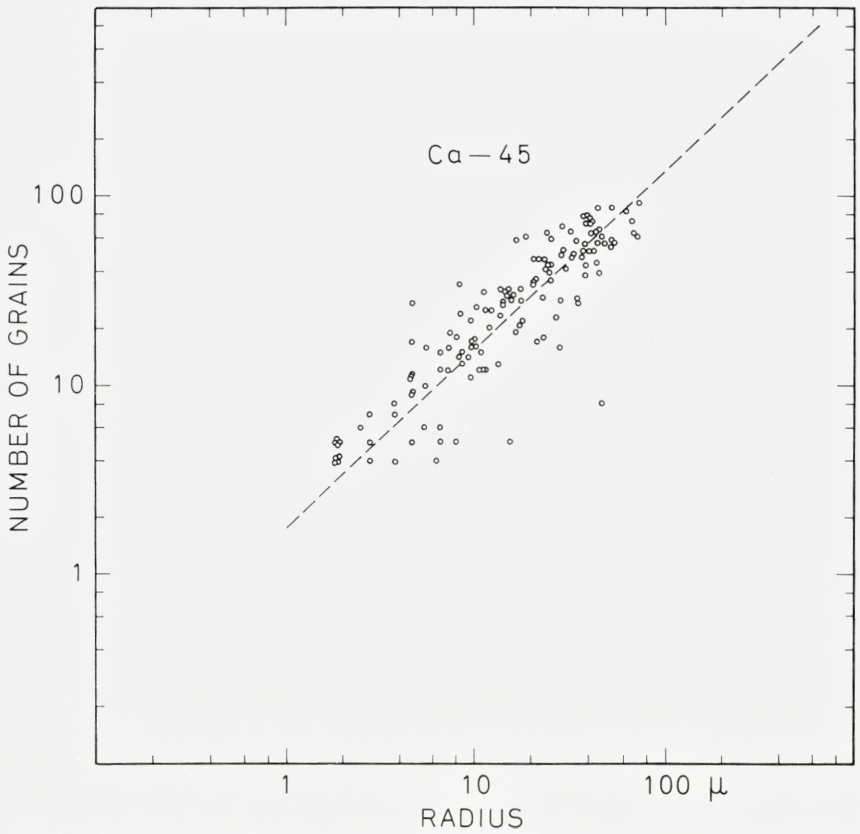

Fig. 17. Number of grains per track versus radius for Ca-45 (Group B; log-log scale).

left the emulsion. As a rough estimate, the sphere of emulsion that should contain 90 % of the disintegrations originating from a point source of Ca-45 would have a radius of 45 microns.

The basic data for calculating the spatial distribution of grains around a point source within the emulsion are available for 100 tracks of C-14 only (group *D*). These measurements were carried out in the following way: The cross in the eyepiece of the microscope was brought to coincide with the first grain of the track to be measured, thus forming the origin of a three-dimensional co-ordinate system. In order to ascertain that no systematic error was made in this system, the symmetry of tracks within the co-ordinate systems was checked using only the end points of all track segments.

TABLE VI

Distribution of radii of randomly selected C-14 tracks.
(Groups B and D, completely contained in the emulsion).

Radius in microns	Nr. of tracks with 4 grains or more	% of all tracks	% of expected nr. of disintegrations
0 – 5.0	71	27.0	37.1
5.1–10.0	76	28.9	24.8
10.1–15.0	55	20.9	18.0
15.1–20.0	25	9.5	8.2
20.1–25.0	14	5.3	4.6
25.1–30.0	10	3.8	3.3
30.1–35.0	6	2.3	2.0
35.1–40.0	2	0.8	0.7
40.1–45.0	1	0.4	0.3
45.1–50.0	2	0.8	0.7
50.1–55.0	0	..	..
55.1–60.0	0	..	..
60.1–65.0	1	0.4	0.3
Total...	263	100.1	100.0

Of 304 end points registered, 104 were situated west of the origin, 173 east of the origin, and 17 just on the north-south axis. Although the difference between east and west seems rather large, it must be assumed that this is a chance occurrence. 148 end points were situated south and 131 north of the origin, while 25 were on the east-west axis. Finally, 138 were above and 108 below in the direction of the optical axis of the microscope, and 58 were in the plane of the origin. The distribution of distances in the three dimensions—however omitting the sign—is depicted in Fig. 9. Apart from the slightly higher uncertainty in the optical axis of the microscope, the differences are insignificant.

From the co-ordinates of the end points of the segments, the co-ordinates of each of the 1968 grains in these tracks have been calculated in approximation. It has been assumed that the distances between the grains in a segment are equal. Moreover, the distance in space of each grain from the origin, i. e. the first grain of the track, as well as the corresponding projected distance in the plane of the glass slide have been calculated.

If the origins of these tracks are superimposed, a model point source of C-14 will result, and the observed distribution of grains around it may be studied.

Since only tracks with 4 or more grains have been registered, a correction must be applied for the tracks with less than 4 grains. According to the calculations described on pp. 22 and 46, 14 % of all C-14 tracks will have 3 grains or less, and the calculated distribution of tracks according to their number of grains has been taken into consideration. On this basis, it can be estimated that 22 grains have been missed in the above total, and these grains are assumed to lie within a distance of 5 microns or less from the first grain.

The distribution of grains in space according to the distance from the first grain is shown in Fig. 7. 50 % of all grains lie less than 9 microns, and 90 % less than 24 microns from the first grain. The figure also illustrates the distribution of grains according to the projected distance from the first grain in the plane of the glass slide. Here, however, only those grains are included which lie less than 5 microns from the first grain in the optical axis of the microscope. When viewed in this manner, 50 % of the grains lie less than 5 microns and 90 % less than 15 microns from the first grain.

The distribution of the distances from the first grain will be very similar to the distribution from the source. For C-14, the mean grain density does not fall below 15 grains per 25 microns at the origin of the longest tracks. It is unlikely, therefore, that the true track origin lay further than 1.5 microns from the first grain, and in many cases it is probably only 1 micron away.

V, 4) The calculation of the distribution of the number of grains per track from the energy spectrum of a beta emitter

In order to deduce a distribution curve for the number of grains per track from the energy spectrum of a pure beta emitter, two assumptions have been made, viz. 1) that the mean number of grains per track for a group of monoenergetic beta particles is given by equation (2) (cf. p. 33)

$$\log \bar{G} = 1.19 \log E - 0.74$$

and 2) that the scatter of the number of grains per beta particle about the mean follows a Poisson distribution.

The energy spectra used in the present study were calculated from the theoretical nomograms of Marshall (1955).

The calculation of the distribution curve proceeded as follows: A large number of equally spaced energy levels was chosen for which the desired distribution of the number of grains per track was calculated (Table VII column 1). These energy levels were 5 keV apart from 2.5 keV and up to 152.5 keV in the case of C-14, and up to 247.5 keV in the case of Ca-45. From the known energy spectrum, the probability of emission of a beta

col. 1 kev	col. 2 $k \cdot p(e)$	col. 3 $m(g)$	col. 4 $p(g\|e)$								Tai
			0	1	2	3	4	5	6	7	8
2.5	3.3	0.5	.6065	.3033	.0758	.0126	.0016	.0002	..	..	.
7.5	3.5	2.0	.1353	.2707	.2707	.1804	.0902	.0361	.0120	.0034	.00
12.5	3.7	3.6	.0273	.0984	.1771	.2125	.1912	.1377	.0826	.0425	.01
17.5	3.8	5.4	.0045	.0244	.0659	.1185	.1600	.1728	.1555	.1200	.08
22.5	3.9	7.5	.0006	.0041	.0156	.0389	.0729	.1094	.1367	.1465	.13
27.5	3.8	9.3	.0001	.0009	.0040	.0123	.0285	.0530	.0822	.1091	.12
32.5	3.8	11.4	..	.0001	.0007	.0028	.0079	.0180	.0341	.0556	.07
37.5	3.7	13.6	..	..	.0001	.0005	.0018	.0048	.0109	.0212	.03
42.5	3.6	16	..	..	..	.0001	.0003	.0010	.0026	.0060	.01
47.5	3.5	18	..	..	..	..	.0001	.0002	.0007	.0019	.00
52.5	3.4	20	..	..	..	..	..	.0001	.0002	.0005	.00
57.5	3.2	22	..	..	..	..	..	..	..	.0001	.00
62.5	3.0	25	..	..	..	..	..	..	..	..	.00
67.5	2.8	28	..	..	..	..	..	..	..	..	.
72.5	2.5	30	..	..	..	..	..	..	..	..	.
77.5	2.3	33	..	..	..	..	..	..	..	..	.
82.5	2.1	35	..	..	..	..	..	..	..	..	.
87.5	1.9	38	..	..	..	..	..	..	..	..	.
92.5	1.7	40	..	..	..	..	..	..	..	..	.
97.5	1.6	43	..	..	..	..	..	..	..	..	.
102.5	1.4	45	..	..	..	..	..	..	..	..	.
107.5	1.2	48	..	..	..	..	..	..	..	..	.
112.5	1.1	51	..	..	..	..	..	..	..	..	.
117.5	0.9	54	..	..	..	..	..	..	..	..	.
122.5	0.8	56	..	..	..	..	..	..	..	..	.
127.5	0.6	59	..	..	..	..	..	..	..	..	.
132.5	0.5	62	..	..	..	..	..	..	..	..	.
137.5	0.3	65	..	..	..	..	..	..	..	..	.
142.5	0.2	68	..	..	..	..	..	..	..	..	.
147.5	0.1	71	..	..	..	..	..	..	..	..	.
152.5	0.0	73	..	..	..	..	..	..	..	..	.
$p(g) \cdot k = \Sigma p(e) \cdot p(g\|e) \cdot k$			2.596	2.425	2.182	2.121	2.067	2.012	1.966	1.931	1.89
$p(g)$ in %			3.81	3.56	3.20	3.11	3.03	2.95	2.88	2.83	2.78
			13.68								
$\frac{p(g)}{0.1368}$ in %			..	..	..	..	3.51	3.42	3.34	3.28	3.22

particle at each of the chosen energy levels ($p(e)$) (column 2) was calculated (cf. MARSHALL, loc. cit.). From equation (2), the mean grain number ($m(g)$) corresponding to each energy level was found (column 3). In the case of C-14, these mean grain numbers lay between 0.5 and 73 grains per track, in the case of Ca-45, between 0.5 and 128 grains per track.

0	13	15	20	25	30	40	50	60	70	80	90	100
											grain levels	
.	..	..	..	..	..	..	..	..	..	..	..	..
.	..	..	..	..	..	..	..	..	..	..	..	..
028	.0001	..	..	..	..	..	..	..	..	..	..	..
262	.0024	.0003	..	..	..	..	..	..	..	..	..	..
858	.0211	.0057	.0001	..	..	..	..	..	..	..	..	..
219	.0572	.0235	.0009	..	..	..	..	..	..	..	..	..
144	.0987	.0611	.0063	.0002	..	..	..	..	..	..	..	..
740	.1085	.0955	.0239	.0017	.0001	..	..	..	..	..	..	..
341	.0814	.0992	.0559	.0092	.0006	..	..	..	..	..	..	..
150	.0509	.0786	.0798	.0237	.0026	..	..	..	..	..	..	..
058	.0271	.0516	.0888	.0446	.0083	..	..	..	..	..	..	..
020	.0127	.0292	.0809	.0654	.0197	.0002	..	..	..	..	..	..
004	.0033	.0099	.0519	.0795	.0454	.0014	..	..	..	..	..	..
001	.0007	.0027	.0249	.0673	.0677	.0065	.0001	..	..	..	..	..
.	.0002	.0010	.0134	.0511	.0726	.0139	.0002	..	..	..	..	..
.	..	.0002	.0045	.0276	.0631	.0314	.0013	..	..	..	..	..
.	..	.0001	.0020	.0162	.0499	.0447	.0033	..	..	..	..	..
.	..	..	.0005	.0063	.0293	.0598	.0101	.0002	..	..	..	..
.	..	..	.0002	.0031	.0185	.0629	.0177	.0007	..	..	..	..
.	..	..	..	.0009	.0080	.0566	.0328	.0026	..	..	..	..
.	..	..	..	.0004	.0043	.0472	.0431	.0054	.0001	..	..	..
.	..	..	..	.0001	.0015	.0310	.0541	.0128	.0006	..	..	..
.	..	..	..	..	.0005	.0175	.0558	.0243	.0020	..	..	..
.	..	..	..	..	.0001	.0086	.0484	.0373	.0055	.0002	..	..
.	..	..	..	..	.0001	.0050	.0403	.0447	.0094	.0005	..	..
.	..	..	..	..	..	.0020	.0273	.0510	.0181	.0015	..	..
.	..	..	..	..	..	.0007	.0162	.0498	.0290	.0041	.0002	..
.	..	..	..	..	..	.0002	.0086	.0422	.0395	.0089	.0006	..
.	..	..	..	..	..	.0001	.0041	.0315	.0463	.0164	.0017	.0001
.	..	..	..	..	..	..	.0018	.0209	.0473	.0258	.0040	.0002
.	..	..	..	..	..	..	.0010	.0150	.0447	.0322	.0067	.0005
19	1.702	1.640	1.432	1.156	.969	.671	.403	.201	.067	.012	.001	..
7	2.50	2.40	2.10	1.70	1.42	.98	.59	.29	.10	.018	.001	..
9	2.89	2.79	2.43	1.96	1.65	1.14	.68	.34	.11	.020	.002	..

Subsequently, a series of grain levels was chosen (horizontal row 2) and for each value of $(m(g))$, the probabilities of occurrence of a track with the number of grains listed in row 2 were calculated from the standard tables of Poisson distributions. The figures thus obtained were multiplied by the corresponding values of $(p(e))$. The products $p(g|e) \times p(e)$ at each

grain level were summed (row 3 from bottom) to give the overall probability of a track containing that number of grains to occur.

This overall probability ($p(g)$) was then normalised to percentage values (row 2 from bottom), and the distribution curve constructed.

Since, according to definition, a track contains 4 grains or more, the percentage values of $p(g)$ for 0,1,2, or 3 grains per track were subtracted from the total, and the remaining values of $p(g)$ were adjusted to give 100 % (last row).

Fig. 4 shows the calculated distribution curve and the observed distribution of the number of grains per track for C-14. The observed distribution was obtained by pooling the results of groups *B*, *D*, and *E*, and it agrees fairly well with the predicted curve. However, the observed distribution shows a preponderance of short tracks, a discrepancy which cannot be explained at present. If assumption (2) mentioned above is not correct, this might account for the difference between observed and calculated values in the beginning of the curve; it is conceivable, for instance, that the scatter of the number of grains per beta particle about the mean has a skewed distribution. It is also possible that, for low energies, the distribution about the mean is wider than that given by a Poisson distribution. Both these factors would tend to increase the probability of occurrence of tracks with few grains.

The predicted and observed distributions for Ca-45 are depicted in Fig. 6. Here, the situation is complicated due to the fact that a larger fraction of the tracks escaped from the emulsion. Still, over a major part of the energy spectrum, agreement between the calculated and the observed grain distribution is good.

Chapter VI

Summary

Beta tracks produced by carbon-14, calcium-45, and chlorine-36 in Ilford G-5 emulsion were studied in order to establish the possibility for accurate evaluation of track- or grain autoradiograms.

From measurements of track length and of the number of grains in the track, equations relating the initial energy of the beta particle to the mean number of grains produced have been derived. On the basis of these relationships, it is possible to predict the pattern of distribution of tracks in terms of the number of grains they contain for any isotope with a known

beta spectrum. In this manner, the percentage of particles that will not produce recognisable tracks can be calculated in each case. Clearly, this fraction will be relatively insignificant for isotopes with a high maximum energy, such as P-32, but it becomes much greater for low energy beta emitters such as C-14 and S-35.

It was found that 14 % of all beta particles emitted by a point source of C-14 will give rise to less than 4 grains, and thus not produce a recognisable track. 4 % will not give rise to any grains at all under the conditions of the present experiments.

Hence, the absolute disintegration rate in a small beta source completely surrounded by Ilford G-5 emulsion can be determined with a considerable degree of accuracy, either by track counting or, if the track density is too high, by grain counting in a well-defined volume of emulsion.

Moreover, the position of silver grains relative to the track origin was measured. From these data, the distribution of grains in space around a point source can be constructed, and the radius of the sphere that will contain a given percentage of the grains can be calculated for any beta-emitter. In the case of C-14, 50 % of all grains lie less than 9 microns, and 90 % less than 25 microns from the source. Viewing the projected distance from the first grain in the plane of the glass slide, and including only those grains which lie less than 5 μ from the first grain in the optical axis of the microscope, 50 % of the grains lie less than 5 microns and 90 % less than 15 microns from the first grain. This resembles the conditions of a stripping film autoradiogram.

Discrepancies in the measurements of track length and grain number as found by LEVI and NIELSEN versus those found by ZAJAC and ROSS, could be explained on technical grounds, and the values obtained by ZAJAC and ROSS were confirmed in the present study.

The influence of development time on grain yield was investigated and the advantage of re-swelling of the processed emulsion prior to microscopic inspection was demonstrated.

On the basis of the results so far obtained, absolute quantitation in other geometrical arrangements of source and emulsion—such as the familiar tissue section covered by stripping film, or a cell smear coated with a thin layer of nuclear emulsion—is still not directly possible. It is feasible, however, to calibrate small reference sources by track autoradiography, and subsequently to expose them in the conditions of the desired experiment, thus determining the grain yield and, from this, the disintegration rate in the source.

In connection with the calculation of the spatial distribution of silver grains around a point source, the definition of autoradiographic resolution was found to be ambiguous and in need of revision. This problem will be dealt with in a separate communication in the near future.

Acknowledgements

Most of the present work was carried out at the Zoophysiological Laboratory A, University of Copenhagen, while one of the authors (A.W.R.) was on leave of absence from the Department of Anatomy, Medical School, University of Birmingham, England, and in receipt of a research grant from the Ministry of Education, Denmark, which is gratefully acknowledged. In addition, financial assistance towards this project was received from the Research Fund of the Faculty of Medicine, Birmingham.

The work was supported in part by the Damon Runyan Memorial Fund for Cancer Research, New York, who defrayed the expenses in connection with the mathematical evaluation of the data.

Our thanks are due to Dr. E. Dahl-Jensen and Dr. Knud Hansen, Institute for Theoretical Physics, Copenhagen, for several helpful discussions on emulsion technique and the theory of energy loss by beta particles, and to Miss E. Frederiksen for skillful assistance.

Hilde Levi: Zoophysiological Laboratory A, University of Copenhagen.

A.W. Rogers: On leave from Department of Anatomy, Medical School, University of Birmingham, Birmingham, England.

M. Weis Bentzon: Statens Seruminstitut, Dept. of Bio-Statistics, Copenhagen.

Arne Nielsen: Institute of Human Genetics, University of Copenhagen.

References

Ahmad, I. and Demers, J., Can. J. Phys. **37**, 1548 (1959).
Demers, P., Ionographie. Presses Universitaires de Montreal (1958).
Dilworth, C. C., Occhialini, C. P. S., and Payne, R. M., Nature **162**, 102 (1948).
Glendenin, L. E., Nucleonics **2**, 12 (1948).
Hauser, J., Photographie Corpusculaire **2**, 207 (1959).
Herz, R. H., Phys. Rev. **75**, 478 (1949).
Howard, A. and Pelc, S. R., Exptl. Cell Res. **2**, 178 (1951).
Lajtha, L. G., Exptl. Cell Res. **3**, 696 (1952).
Lamerton, L. F. and Harris, E. B., J. Photogr. Sci. **2**, 135 (1954).
Levi, H., Exptl. Cell. Res. **7**, 44 (1954).
— Exptl. Cell Res., Suppl. **4**, 207 (1957).
— and Nielsen, A., Lab. Invest. **8**, 82 (1959).
Levinthal, C. and Thomas, C. A., Biochim. Biophys. Acta **23**, 453 (1957).
Marinelli, L. D. and Hill, R. F., Am. J. Roent. **59**, 396 (1948).
Marshall, J. H., Nucleonics **13**, 34 (1955).
Mees, C. E. K., The Theory of the Photographic Process. McMillan, New York 1952.
Norris, W. P. and Woodruff, L. A., Ann. Rev. Nucl. Sci. **5**, 297 (1955).
Odeblad, E., Nordisk Medicin **43**, 1056 (1950).
Pelc, S. R., Intern. J. Appl. Rad. and Isotopes **1**, 172 (1956).
Perry, R., First European Symposium on Autoradiography in Medical Sciences, Rome 1961.
Ross, M. A. S. and Zajac, B., Nature **162**, 923 (1948).
Zajac, B., Thesis (1949).
— and Ross, M. A. S., Nature **164**, 311 (1949).

Indleveret til Selskabet den 12. oktober 1962.
Færdig fra trykkeriet den 2. april 1963.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 12

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 12 (1963)

THE INFRARED SPECTRUM OF CH_3D

BY

FLEMMING ALLAN ANDERSEN

København 1963
i kommission hos Ejnar Munksgaard

CONTENTS

Synopsis

The infrared spectrum of CH_3D has been measured in the region 400–6000 cm^{-1} by means of a prism instrument (Beckman IR3) of medium resolving power. The band-centre frequencies have been derived by rotational analysis of the fundamental bands. The results are compared with previous prism and grating data, and a slightly revised set of normal vibration frequencies is given.

A value of the rotational constant $A''(A_0)$ has been obtained, which agrees within the limits of error with a recent Raman value.

Coriolis coupling constants have been derived for the three doubly degenerate fundamentals.

Possible assignments of the observed combination bands are given.

Printed in Denmark
Bianco Lunos Bogtrykkeri A-S

I. Introduction

The infrared and Raman spectra of methane and its deuterated species, CH_3D, CH_2D_2, CHD_3, and CD_4, have been the object of several investigations.[1] However, when one examines the literature, it is obvious that several of the fundamental frequencies of the partly deuterated methanes are uncertain. As the investigations cited above[1] of these molecules are now more than 20 years old and the experimental technique since then has improved considerably, a re-investigation of all the fundamental bands of the partly deuterated methanes appeared to be desirable.

In 1953 BOYD and THOMPSON[2] and later ALLEN and PLYLER[3] (1959) have measured the band near 2200 cm^{-1} connected with the C–D stretching in CH_3D with high resolution. At the time when this investigation had been started, in 1956, REA and THOMPSON[4] published normal vibration frequencies of CHD_3 obtained from infrared measurements. The assignments of two of the fundamental bands as well as the frequency value of one more fundamental are, however, in disagreement with infrared results of WILMSHURST and BERNSTEIN[5] (1957). The latter authors have published fundamental frequencies of all the deuterated methanes obtained with low resolution, the frequency values being the position of maximum intensity in the bands rather than the true band centre frequencies.

Recently JONES[6] (1960) has published the results of an infrared study of the degenerate C–H stretching fundamental of CH_3D at 3016 cm^{-1} using medium resolution ($\approx$ 1 cm^{-1}). Only the central part and the high frequency side of the band were measured. Some of his band constants deviate somewhat from the results of a recent Raman investigation of this band by RICHARDSON et al.[7]

The present paper will deal only with CH_3D. Similar results obtained for CH_2D_2 and CHD_3 will be given in separate papers.

II. Experimental

1. Preparations

CH_3D was prepared from CH_3I by allowing the halogen compound to react with a mixture of Zn-dust, acetic anhydride, and heavy water (99.83 % D_2O) at about 30° C. The methyl iodide was carefully distilled before use. The Zn-dust and acetic anhydride were of the highest purity commercially available. The acetic anhydride was carefully distilled and kept for a few days over metallic sodium in order to remove a small content of acetic acid.

The reaction was carried out by adding the methyl iodide dropwise to the mixture of Zn-dust, acetic anhydride, and D_2O (vigorous reaction!). The CH_3D evolved was collected in a gasometer and from there passed through a dry ice trap and then condensed in a liquid air trap. A volume of the gas first collected, equal to the volume of the reaction vessel, was rejected, as it would mainly consist of air. The CH_3D was then distilled *in vacuo* to another liquid air trap. This procedure was repeated once more. In this way the deutero-methane was separated from less volatile impurities (e.g. H_2O from the gasometer).

In a similar way CH_2D_2 and CHD_3 can be prepared from respectively CH_2I_2 and $CHCl_3$.

It follows from the method of preparation that the deutero-methanes will be contaminated by some D_2 and air. These contaminants were removed by successively pumping off small fractions of deutero-methane kept in a liquid air trap until an equilibrium pressure of 20–25 mm Hg was obtained.

In each sample the only spectroscopically detectable impurity was the nearest lower deuterated methane, the amount of this, about 5 per cent., being spectrally of little significance. Table 1 gives the results from the preparations.

Table 1. Results of preparations.

Deuterated compound	Mol halogen compound	Gram atom Zn	Mol $(CH_3CO)_2O$	Mol D_2O	Yield Litre and %
CH_3D	0.29 (CH_3I)	1.26	0.54	3.75	3.5 (50)
CH_2D_2	0.21 (CH_2I_2)	1.33	1.00	2.50	3.7 (73)
CHD_3	0.05 (CHI_3)	0.21	0.13	3.10	0.36 (30)

2. Spectroscopic Procedure

The spectra were taken on a slightly modified Beckman IR3 infrared spectrometer equipped with KBr, NaCl, and LiF prisms.[8] The path length could be varied from 10 to 2000 cm. The effective slit width, s_{eff},[9] is indicated at each spectrum reproduced below together with values of the gas pressure in mm Hg (p) and the path length in cm (l).

The relative positions of the fine structure components within a single band are believed to be accurate within $\pm 0.1-0.2$ cm^{-1}. The absolute accuracy of the frequencies given is estimated at about ± 0.5 cm^{-1}. All the frequencies are in cm^{-1}.

III. Fundamental Bands

The molecule CH_3D is a symmetric top belonging to the point group C_{3v} and therefore has 6 normal vibrations: 3 totally symmetric of species A_1 and 3 doubly degenerate of species E. They are all infrared active. The A_1 vibrations give rise to parallel bands while the E vibrations appear in the spectrum as perpendicular bands.

1. Parallel Bands (A_1 Fundamentals)

The parallel bands corresponding to the vibrations ν_{4a}, ν_{3a}, and ν_1 lie in the regions 1200–1400 cm^{-1}, 2050–2300 cm^{-1}, and 2900–3000 cm^{-1} with observed band centres at 1306, 2200, and 2970 cm^{-1}, respectively. The J fine structure of the P and R branches was easily resolved because of the high value of the rotational constant B'' (≈ 3.9 cm^{-1}), the spacing in the branches being roughly $2B''$.

On the low frequency side, the ν_{4a} band at 1306 cm^{-1} is overlapped by the strong perpendicular band, ν_{4bc}, at 1157 cm^{-1}, and on the high frequency side by the weak perpendicular band, ν_{2ab}, at 1471 cm^{-1}. Furthermore, fine structure components from the strong band $\nu_4(F_2)$ of CH_4, the band centre of which nearly coincides with ν_{4a} of CH_3D, were present in the spectrum indicating a small content of CH_4 in the CH_3D sample. However, because of the high intensity of the ν_{4a} band, assignment of P and R lines up to $J = 11$ could be made quite easily.

The 2200 cm^{-1} band (ν_{3a}) has been analyzed by Boyd and Thompson,[2] and recently by Allen and Plyler.[3] The author's measurements of the

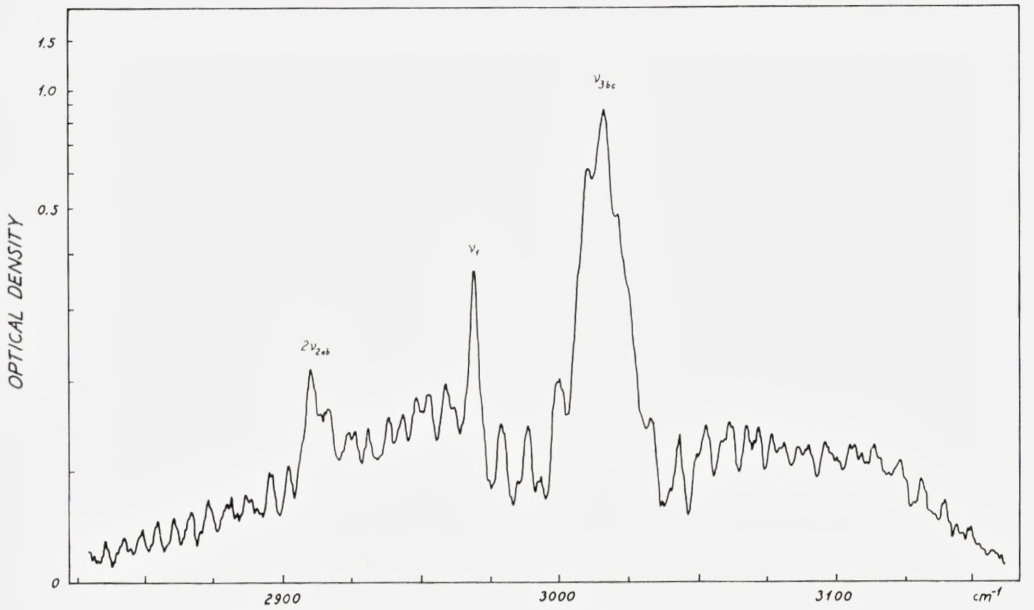

Fig. 1. $\nu_1(A_1)$ and $\nu_{3\,bc}(E)$. 2830–3170 cm^{-1}. p = 93 mm. l = 10 cm. s_{eff} = 2.1–2.3 cm^{-1}.

same band are in good agreement with their results in spite of the lower resolving power of the IR 3 instrument. In this way a satisfactory check on the spectroscopic procedure used in the present investigation was obtained.

The high intensity of the nearby overtone $2\nu_{4\,bc}$ ($2\times1157 = 2314$ cm^{-1}) observed at 2316 cm^{-1} indicates a Fermi resonance between the A_1 component of this overtone and $\nu_{3\,a}$. I have estimated the unperturbed frequency of $\nu_{3\,a}$ at 2210 ± 5 cm^{-1}.

Also ν_1 (Figs. 1 and 4) appears in the spectrum as one of the components of a doublet caused by Fermi resonance with the A_1 part of the overtone $2\nu_{2\,ab}$ ($2\times1471 = 2942$ cm^{-1}). The components are observed at 2970 and 2910 cm^{-1}, WILMSHURST and BERNSTEIN found 2973 and 2914 cm^{-1}. If we assume that the anharmonicity of the overtone $2\nu_{2\,ab}$ is -10 cm^{-1}—as seems reasonable—then the unperturbed level ν_1 has been raised 22 cm^{-1}, which means that the unperturbed frequency of ν_1 is close to $2970 - 22 = 2948$ cm^{-1}. As the most probable value, $\nu_1 = 2948 \pm 5$ cm^{-1} has been adopted (see note added in proof, p. 34).

Since the observed band at 2970 cm^{-1} is of low intensity and badly overlapped by the strong perpendicular band, $\nu_{3\,bc}$, at 3016 cm^{-1}, the assignment of the P and R lines is rather uncertain and must be regarded as tentative only.

The analysis of the rotational structure has been carried out by means of the well-known "method of combination differences". Neglecting the effect of centrifugal distortion, the P and R branches can be represented by

$$P(J) = \nu_0 - (B' + B'')\,J + (B' - B'')\,J^2 \tag{1}$$

and

$$R(J) = \nu_0 + 2\,B' + (3\,B' - B'')\,J + (B' - B'')\,J^2. \tag{2}$$

Double-primed quantities refer to the vibrational ground state, single-primed constants to vibrationally excited levels. The combination relations are:

$$R(J-1) + P(J) = 2\,\nu_0 + 2\,(B' - B'')\,J^2 \tag{3}$$

$$\Delta_2 F''(J) = R(J-1) - P(J+1) = 4\,B''\,(J + 1/2) \tag{4}$$

and

$$\Delta_2 F'(J) = R(J) - P(J) = 4\,B'\,(J + 1/2). \tag{5}$$

If we take centrifugal distortion into account, D_J'' and D_J' being the centrifugal distortion coefficients, the following relations hold:

$$R(J-1) + P(J) = 2\,\nu_0 + 2\,(B' - B'')\,J^2 - 2\,(D_J' - D_J'')\,J^2\,(J^2 + 1) \tag{6}$$

$$\Delta_2 F''(J) = 4\,B''\,(J + 1/2) - 8\,D_J''\,(J + 1/2)^3 \tag{7}$$

and

$$\Delta_2 F'(J) = 4\,B'\,(J + 1/2) - 8\,D_J'\,(J + 1/2)^3. \tag{8}$$

The difference between D_J' and D_J'' can be ignored. Eq. (6) then becomes identical with Eq. (3). Eqs. (7) and (8) can be rewritten

$$\frac{\Delta_2 F''(J)}{J + 1/2} = 4\,B'' - 8\,D_J''\,(J + 1/2)^2 \tag{9}$$

and

$$\frac{\Delta_2 F'(J)}{J + 1/2} = 4\,B' - 8\,D_J'\,(J + 1/2)^2. \tag{10}$$

Graphical representation of Eqs. (3)–(5) and (9) and (10) give the band-centre frequencies ν_0, the rotational constants B'' and B', and the difference B'–B''. The results are summarized in Table 2. Only for the ν_{3a} band, values (approximate) of D_J'' and D_J' could be obtained. For this band the results agree well with the results of the high resolution study by Allen and Plyler. The observed frequencies of the fine structure components have therefore been omitted. Observed frequencies and their interpretation for two of the bands (ν_1 and ν_{4a}), together with calculated frequency values, are given in Tables 3 and 4. The agreement appears to be satisfactory.

Table 2. Band constants obtained for parallel bands.

	ν_1* cm^{-1}	ν_{3a}** cm^{-1}	ν_{4a} cm^{-1}
ν_0***	2970.1 ± 0.5	2200.0 ± 0.5 (2200.03)	1306.5 ± 0.5
$B' - B''$	− 0.02 ± 0.02	− 0.040 ± 0.002 (− 0.0422)	− 0.088 ± 0.002
B''	3.90 ± 0.02	3.875 ± 0.005 (3.880_0)	3.89 ± 0.01
B'	3.92 ± 0.04	3.835 ± 0.005 (3.837_8)	3.80 ± 0.01
D_J''	—	$4._4 \times 10^{-5}$ (5×10^{-5})	—
D_J'	—	$4._4 \times 10^{-5}$ ($5._5 \times 10^{-5}$)	—

* High frequency component of a Fermi doublet. The unperturbed value of ν_1 is close to 2948 cm^{-1} (see text).

** The values in brackets are those obtained by ALLEN and PLYLER[3] (see text).

*** In Table 14 of this paper ν_0 values obtained by other investigators are compared to the author's.

Table 3. Observed and calculated fine structure lines of $\nu_1 (A_1)$ band.

$P(J)$				$R(J)$			
J	Obs. cm^{-1}	Calc. cm^{-1}	Δ Calc.-obs.	J	Obs. cm^{-1}	Calc. cm^{-1}	Δ Calc.-obs.
				0	2978.3	2977.9	− 0.4
1	2962.1	2962.3	0.2	1	2985.7	2985.6	− 0.1
2	2954,0	2954.5	0.5	2	2993.5	2993.3	− 0.2
3	2946,4	2946,6	0.2	3	3001.0	3000.9	− 0.1
4	2938.5	2938.7	0.2	4	3009.0	3008.5	− 0.5
5	2931.1	2930.7	− 0.4				
6	2923.3	2922.7	− 0.6				
7	2914.0	2914.7	0.7				
8	2906.7	2906.6	− 0.1				

Effective slit width s_{eff} = 1.3–1.4 cm^{-1}.

2. Perpendicular Bands (*E* Fundamentals)

The perpendicular bands arising from the normal vibrations ν_{2ab}, ν_{3bc}, and ν_{4bc} are, as already mentioned, observed at 1471, 3016, and 1157 cm^{-1} (Figs. 1, 2, 3, 4, and 5). The K fine structure was resolved for all three bands. The average Q line spacings $\Delta\nu_{2ab}$ and $\Delta\nu_{3bc}$ were found to have approximately the numerical values 5.8 cm^{-1}, respectively 1.9 cm^{-1}. Each of the spacings can be either 'positive' or 'negative'. If the Q line spacing of a band is 'negative', it implies that the ${}^RQ_K$ lines occur on the *low* frequency

Table 4. Observed and calculated fine structure lines of $\nu_{4a}(A_1)$ band.

$P(J)$				$R(J)$			
J	Obs. cm^{-1}	Calc. cm^{-1}	Δ Calc.-obs.	J	Obs. cm^{-1}	Calc. cm^{-1}	Δ Calc.-obs.
				0	1313.7	1314.1	0.4
1	—	1298.7	—	1	1320.9	1321.5	0.6
2	1290.2	1290.8	0.6	2	1328.0	1328.8	0.8
3	1282.3	1282.6	0.3	3	1335.8	1335.9	0.1
4	1274.4	1274.3	−0.1	4	1342.3	1342.8	0.5
5	1265.8	1265.9	0.1	5	1349.5	1349.5	0.0
6	1257.6	1257.2	−0.4	6	1356.0	1356.1	0.1
7	1248.3	1248.4	0.1	7	1362.4	1362.4	0.0
8	1239.5	1239.4	−0.1	8	1368.6	1368.7	0.1
9	1230.0	1230.2	0.2	9	1374.6	1374.7	0.1
10	1220.8	1220.8	0.0	10	1380.7	1380.5	−0.2
11	1211.1	1211.3	0.2	11	1386.5	1386.2	−0.3

Effective slit width $s_{\text{eff}} = 1.8$–$2.4\ cm^{-1}$.

side of the band centre. For ν_{3bc} the sign of the spacings can be established in the following way. Using the approximate formula for the Coriolis coupling coefficient given by MEAL and POLO[10]

$$\zeta_{3bc} = \frac{(1-\cos\alpha)\, m_H}{m_C + (1-\cos\alpha)\, m_H},$$

we find $\zeta_{3bc} = 0.10$. From the value of ζ_{3bc} we can calculate an approximate value for $\Delta\nu_{3bc}$. The average spacing, $\Delta\nu_i$, in a perpendicular band corresponding to the vibration ν_i is equal to $2\,[A_i'(1-\zeta_i) - B_i']$. Setting $A_{3bc}' \simeq A'' = 5.25\ cm^{-1}$ (see page 30) and $B_{3bc}' \simeq B'' = 3.88\ cm^{-1}$ (see Table 2), we obtain $\Delta\nu_{3bc} = 1.7\ cm^{-1}$, which is close to the observed value. This shows that the sign of $\Delta\nu_{3bc}$ is positive.

The question whether $\Delta\nu_{2ab}$ is positive or negative is connected with the magnitude of the spacing in the third perpendicular band ν_{4bc}. To a good approximation the average Q line spacings of the perpendicular bands will obey the sum rule[11]

$$\Delta\nu_{2ab} + \Delta\nu_{3bc} + \Delta\nu_{4bc} \simeq 6\,A'' - 7\,B'' = 4.34\ cm^{-1},$$

where $A'' = 5.25\ cm^{-1}$ and $B'' = 3.88\ cm^{-1}$. From the relation it follows that at least one of the bands must have a 'negative' spacing, which means that $\zeta_i > 1 - B_i'/A_i' \simeq 1 - B''/A'' = 0.26$. Setting $\Delta\nu_{2ab} = +5.8\ cm^{-1}$ and $\Delta\nu_{3bc} =$

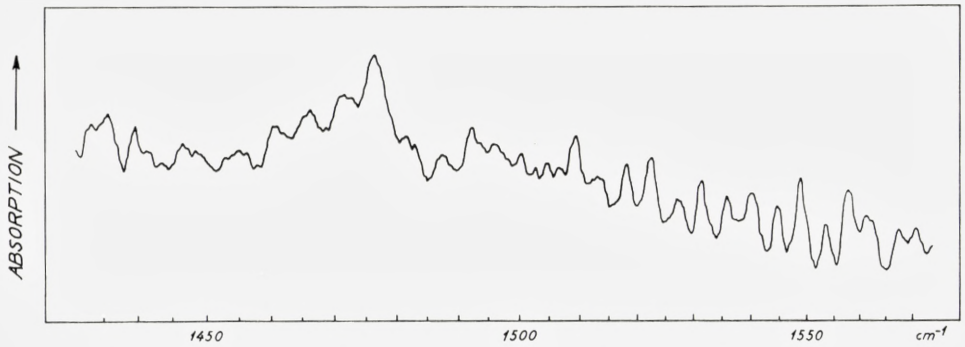

Fig. 2. $\nu_{2ab}(E)$. 1430–1570 cm^{-1}. $p = 760$ mm. $l = 10$ cm. $s_{eff} = 2.0$–2.5 cm^{-1}.

$+1.9\ cm^{-1}$ in the sum rule, one obtains $\Delta\nu_{4bc} = -3.4\ cm^{-1}$, while $\Delta\nu_{2ab} = -5.8\ cm^{-1}$ and $\Delta\nu_{3bc} = +1.9\ cm^{-1}$ gives $\Delta\nu_{4bc} = +8.2\ cm^{-1}$. For $|\Delta\nu_{4bc}|$ the value $3.4\ cm^{-1}$ is found experimentally. Consequently, the spacings of the ν_{2ab} and ν_{4bc} bands must be respectively 'positive' and 'negative'.

This result is also supported by the observed relative intensities of the Q branches near the two band centres. As ${}^{R}Q_0$ is the strongest of the Q lines, and ${}^{P}Q_1$ is stronger than ${}^{R}Q_1$, the spacing in the ν_{2ab} band (see Fig. 2) should be positive. In the ν_{4bc} band (see Fig. 5) it follows from the same kind of argument that the spacing is negative.

The rotational analysis of the bands is based on the preceding discussion.

A characteristic feature of all three perpendicular bands is the strong central part caused by an agglomeration of the Q lines near the band centre, and the complicated fine structure, consisting of ${}^{P}Q_K$, ${}^{R}Q_K$, ${}^{P}P_K(J)$, and ${}^{R}R_K(J)$ lines and the much weaker ${}^{P}R_K(J)$ and ${}^{R}P_K(J)$ lines. The K numbering of the Q lines was carried out in the usual way, and the observed frequencies are shown in Tables 8, 9, and 10. The assignments were to some extent complicated by the limited resolving power of the instrument and the overlapping by other bands.

On the high frequency side of the band ν_{4bc} some of the fine structure lines are masked by the strong band at $1306\ cm^{-1}$. The low frequency side of the weak band ν_{2ab} is overlapped by the absorption at $1306\ cm^{-1}$. For this reason, only a few ${}^{P}Q_K$ lines and one ${}^{P}P_K(J)$ line could be picked out of this band. On the other hand, the high frequency side of the band seems to be free from overlapping by other bands and has therefore been used for the evaluation of the band constants as discussed below. In the band ν_{3bc} there is an overlapping on the long wave side by the weak parallel band

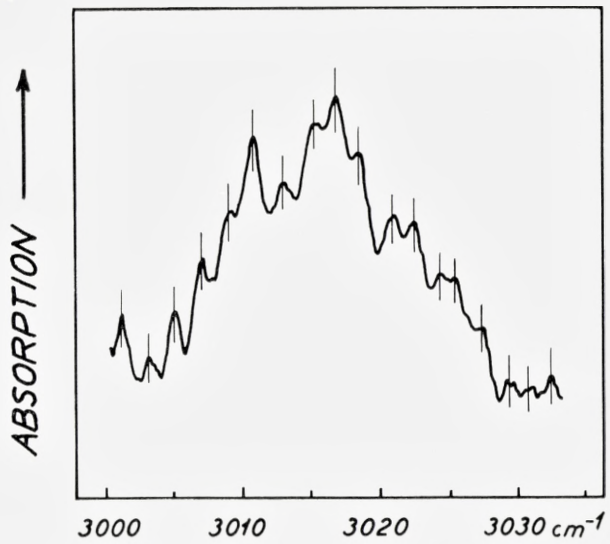

Fig. 3. $\nu_{3\,bc}(E)$. 3000–3030 cm^{-1}. $p = 65$ mm. $l = 10$ cm. $s_{eff} = 1.4$ cm^{-1}.

at 2970 cm^{-1}. However, from the fine structure analysis band constants could be obtained which gave satisfactory agreement between calculated and observed frequencies.

The positions of the Q branches of a perpendicular band, neglecting centrifugal distortion, are given by

$$\left.\begin{aligned} {}^{P}Q_K(J) = {} & [\nu_0 + A_i'(1-\zeta_i)^2 - B_i'] + (B_i' - B'')\,J(J+1) \\ & - 2\,[A_i'(1-\zeta_i) - B_i']\,K + [(A_i' - A'') - (B_i' - B'')]\,K^2 \end{aligned}\right\} \tag{11}$$

and

$$\left.\begin{aligned} {}^{R}Q_K(J) = {} & [\nu_0 + A_i'(1-\zeta_i)^2 - B_i'] + (B_i' - B'')\,J(J+1) \\ & + 2\,[A_i'(1-\zeta_i) - B_i']\,K + [(A_i' - A'') - (B_i' - B'')]\,K^2. \end{aligned}\right\} \tag{12}$$

Disregarding the difference between B_i' and B'' the following combination relations can be obtained:

$$\Delta_2 F''(J,K) = {}^{R}Q_{K-1} - {}^{P}Q_{K+1} = 4\,[A'' - A_i'\zeta_i - B'']\,K \tag{13}$$

$$\Delta_2 F'(J,K) = {}^{R}Q_K - {}^{P}Q_K = 4\,[A'(1-\zeta_i) - B_i']\,K \tag{14}$$

$${}^{R}Q_K + {}^{P}Q_K = 2\,[\nu_0 + A_i'(1-\zeta_i)^2 - B_i'] + 2\,[(A_i' - A'') - (B_i' - B'')]\,K^2. \tag{15}$$

Plotting these expressions for the bands $\nu_{3\,bc}$ and $\nu_{4\,bc}$, we obtain the values of the band constants given in Table 5.

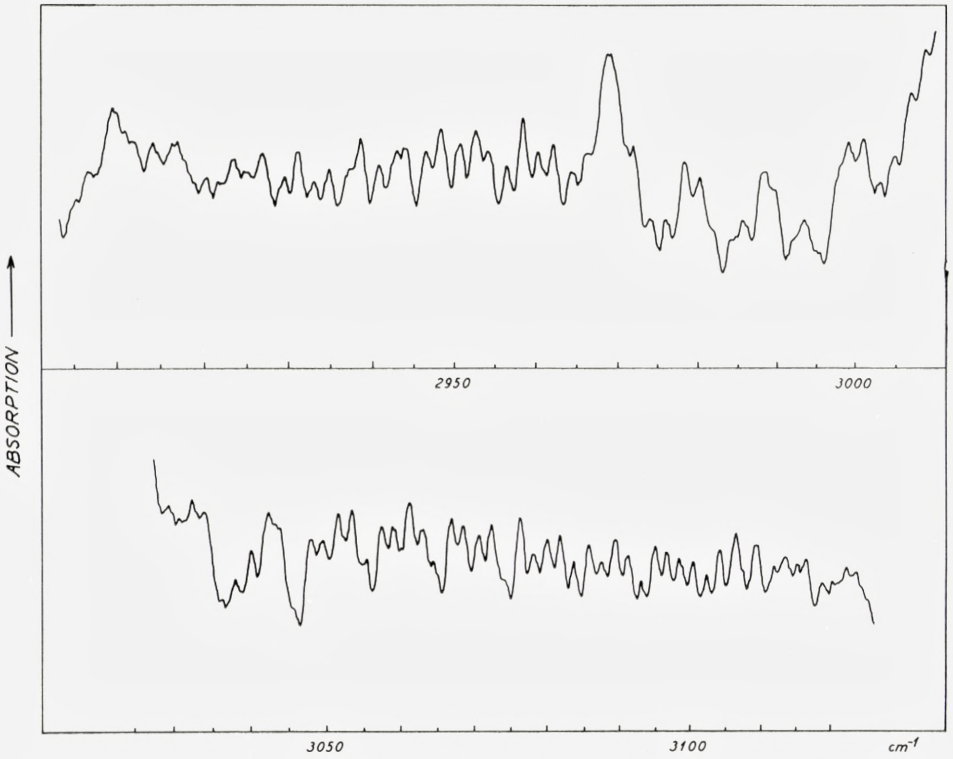

Fig. 4. $\nu_1(A_1)$ and $\nu_{3\,bc}(E)$. Upper curve: 2905–3010 cm^{-1}. Lower curve: 3030–3125 cm^{-1}. p = 152 mm. l = 10 cm. s_{eff} = 1.3–1.5 cm^{-1}.

Table 5. Preliminary values of band constants for perpendicular bands.

	$\nu_{2\,ab}$ cm^{-1}	$\nu_{3\,bc}$ cm^{-1}	$\nu_{4\,bc}$ cm^{-1}
$A'_i(1-\zeta_i)-B'_i$	2.91 ±0.03	0.944 ± 0.005	−1.69 ±0.02
$A''-A'_i\zeta_i-B''$	(2.95)	0.954 ± 0.005	−1.68 ±0.01
$(A'_i-A'')-(B'_i-B'')$	−0.040 ± 0.01	−0.011 ± 0.003	0.013 ± 0.003
$\nu_0+A'_i(1-\zeta_i)^2-B'_i$	1476.1 ±0.5	3016.8 ±0.5	1154.6 ±0.5

As only two ${}^PQ_K$ lines of the band $\nu_{2\,ab}$ could be utilized, another procedure was attempted. From the observed values of ${}^PQ_5$, ${}^PQ_6$, ${}^RQ_5$, and ${}^RQ_6$ we calculate for $A'_{2\,ab}(1-\zeta_{2\,ab})-B'_{2\,ab}$ from Eq. (14) the values 2.905 cm^{-1} and 2.913 cm^{-1}, the average value being 2.91 cm^{-1}.

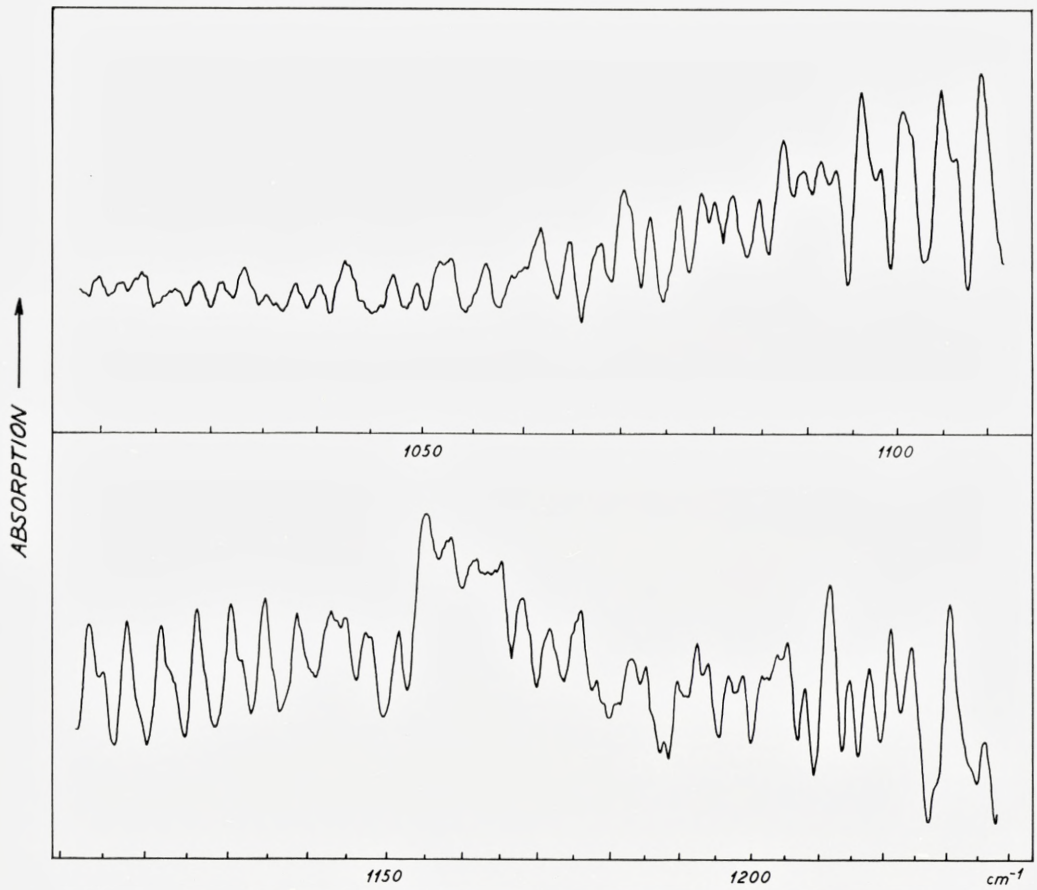

Fig. 5. $\nu_{4\,bc}(E)$. Upper curve: 1020–1110 cm^{-1}. p = 760 mm. Lower curve: 1110–1235 cm^{-1}. p = 323 mm. l = 10 cm. s_{eff} = 1.3–1.9 cm^{-1}.

In order to get a rough estimate of $(A'_{2\,ab} - A'') - (B'_{2\,ab} - B'')$, Eq. (15) was used. Taking the observed ${}^{R}Q_0 = 1476.3 = \nu_0 + A'_{2\,ab}(1 - \zeta_{2\,ab})^2 - B'_{2\,ab}$, and using the observed frequencies of ${}^{P}Q_5$, ${}^{P}Q_6$, ${}^{R}Q_5$, and ${}^{R}Q_6$, we obtain the values -0.03 and -0.05 cm^{-1}, the average being -0.04 cm^{-1}.

In order to make use of the observed ${}^{R}Q_K$ lines, Eq. (12) was written in the form

$$ {}^{R}Q_K - 2\,[A'_{2\,ab}(1 - \zeta_{2\,ab}) - B'_{2\,ab}]\,K = [\nu_0 + A'_{2\,ab}(1 - \zeta_{2\,ab})^2 - B'_{2\,ab}] + [(A'_{2\,ab} - A'') - (B'_{2\,ab} - B'')]\,K^2, $$

neglecting the term $(B'_{2\,ab} - B'')\,J\,(J+1)$. From a plot of this expression, where $A'_{2\,ab}(1-\zeta_{2\,ab}) - B'_{2\,ab} = 2.91\ \mathrm{cm}^{-1}$, we obtain

$$\nu_0 + A'_{2\,ab}\,(1-\zeta_{2\,ab})^2 - B'_{2\,ab} = 1476.1\ \mathrm{cm}^{-1}$$

and

$$(A'_{2\,ab} - A'') - (B'_{2\,ab} - B'') = -0.040\ \mathrm{cm}^{-1}$$

in good agreement with the values above.

The assignment of the ${}^{P}P_K(J)$ and ${}^{R}R_K(J)$ lines of the three bands was carried out in the following way.

The positions of the ${}^{P}P_K(J)$ and ${}^{R}R_K(J)$ lines are given by the equations

$${}^{P}P_K(J) = {}^{P}Q_K(J) - 2\,B'_i\,J + 4\,D_J\,J^3 \tag{16}$$

and

$${}^{R}R_K(J) = {}^{R}Q_K(J) + 2\,B'\,(J+1) - 4\,D_J\,(J+1)^3, \tag{17}$$

where the following approximations have been made: $D''_J = D'_J = D_J$, $D''_{JK} = D'_{JK} = 0$ and $D''_K = D'_K = 0$. It should, however, be emphasized that for higher J and K values, i.e. $J \geq 10$ and $K \geq 6$, approximately, it may not be permissible to ignore the D_{JK} and D_K terms. The D_{JK} and D_K values found by Richardson *et al.* for the $\nu_{3\,bc}$ band show this.

In order to calculate approximate values for the ${}^{P}P_K(J)$ and ${}^{R}R_K(J)$ lines of the three bands, Eqs. (16) and (17) were used together with Eqs. (11) and (12), the band constants given in Table 5 being inserted, and $B'_i = B'' = 3.88\ \mathrm{cm}^{-1}$. D_J was taken to $5.5\times10^{-5}\ \mathrm{cm}^{-1}$, which is close to the value of D'_J and D''_J for $\nu_{3\,a}$ obtained by Boyd and Thompson[2] and Allen and Plyler.[3] Then, calculating the relative intensities of the transitions from the formulae quoted by Herzberg,[12] taking $B'' = 3.880\ \mathrm{cm}^{-1}$ and $A'' = 5.245\ \mathrm{cm}^{-1}$ (see page 30), it was possible to pick out a number of ${}^{P}P_3(J)$ and ${}^{R}R_3(J)$ lines in the $\nu_{3\,bc}$ and $\nu_{4\,bc}$ bands. For the analysis, the following combination relations were used:

$$\left.\begin{aligned} {}^{R}R_K(J) - {}^{P}P_K(J) &= 4\,[A'_i(1-\zeta_i) - B'_i]\,K + 4'_i\,B\,(J+1/2) \\ &\quad - 4\,D_J[(J+1)^3 + J^3] \end{aligned}\right\} \tag{18}$$

$$\left.\begin{aligned} {}^{R}R_K(J-1) - {}^{P}P_K(J+1) &= 4\,[A'_i(1-\zeta_i) - B'_i]\,K + 4\,B''\,(J+1/2) \\ &\quad - 4\,D_J\,[J^3 + (J+1)^3] \end{aligned}\right\} \tag{19}$$

$$\left.\begin{aligned} {}^{R}R_K(J-1) + {}^{P}P_K(J) &= 2\,[\nu_0 + A'_i(1-\zeta_i)^2 - B'_i] \\ &\quad + 2\,[(A'_i - A'') - (B'_i - B'')]\,K^2 + 2\,(B'_i - B'')\,J^2, \end{aligned}\right\} \tag{20}$$

where $K = 3$. The small term $4\,D_J\,[J^3 - (J-1)^3]$ has been omitted. Using the values of $A'_i(1-\zeta_i) - B'_i$ and $(A'_i - A'') - (B'_i - B'')$ from Table 5, and $D_J = 5.5\times10^{-5}\ \mathrm{cm}^{-1}$, graphical representations of Eqs. (18), (19), and (20) gave

Table 6. Preliminary values of band constants for $\nu_{3\,bc}$ and $\nu_{4\,bc}$ bands.

	$\nu_{3\,bc}$ cm^{-1}	$\nu_{4\,bc}$ cm^{-1}
B'	3.873 ± 0.005	3.86 ± 0.01
B''	3.882 ± 0.005	3.88 ± 0.01
$B' - B''$	-0.011 ± 0.002	-0.013 ± 0.005
$\nu_0 + A_i'(1-\zeta_i)^2 - B_i'$	3016.9 ± 0.5	1154.7 ± 0.5
D_J	$5_{.5} \times 10^{-5}$	—

the values of B_i', B'', $B_i' - B''$, and $\nu_0 + A_i'(1-\zeta_i)^2 - B_i'$, listed in Table 6. For the $\nu_{3\,bc}$ band, it was possible to obtain a value of D_J. Using the approximation $J^3 + (J+1)^3 \simeq 2(J+1/2)^3$, Eqs. (18) and (19) were written:

$$\frac{{}^R R_3(J) - {}^P P_3(J) - 12\,[A_i'(1-\zeta_i) - B_i']}{J+1/2} = 4\,B_i' - 8\,D_J\,(J+1/2)^2 \qquad (21)$$

and

$$\frac{{}^R R_3(J-1) - {}^P P_3(J+1) - 12\,[A_i'(1-\zeta_i) - B_i']}{J+1/2} = 4\,B'' - 8\,D_J\,(J+1/2)^2. \qquad (22)$$

Graphical representations gave B_i', B'', and D_J (5×10^{-5} and 6×10^{-5} cm^{-1}, see Table 6).

For the $\nu_{2\,ab}$ band, it was assumed that $B_i' - B'' = -0.015$ cm^{-1}, which then gives $B_i' = 3.865$ cm^{-1}.

Table 7 shows the final values of the band constants used for calculating the frequencies in Tables 8, 9, and 10. As will be seen, some of the band constants have been slightly adjusted in order to improve the agreement between observed and calculated frequencies.

It has been possible to explain nearly all the observed lines as ${}^P Q_K$, ${}^R Q_K$, ${}^P P_K(J)$, and ${}^R R_K(J)$ lines. Only a few lines had to be interpreted as ${}^P R_K(J)$ and ${}^R P_K(J)$ lines. Although ${}^P R_K(J)$ and ${}^R P_K(J)$ lines generally contribute to the intensities of the observed lines, it was thought permissible to omit them in most cases because of their low intensity and the limited resolving power of the instrument. ${}^R P_K(J)$ lines have, however, had to be included in the *low* frequency region of the $\nu_{4\,bc}$ band in order to get reasonable agreement between observed and calculated intensities.

The frequencies of the ${}^R P_K(J)$ and ${}^P R_K(J)$ lines were calculated from the expressions

$${}^R P_K(J) = {}^R Q_K(J) - 2\,B_i' J + 4\,D_J J^3 \qquad (23)$$

$${}^P R_K(J) = {}^P Q_K(J) + 2\,B_i'(J+1) - 4\,D_J (J+1)^3. \qquad (24)$$

Table 7. Finally adopted values of band constants for perpendicular bands.

	$\nu_{2\,ab}$ cm^{-1}	$\nu_{3\,bc}$ cm^{-1}	$\nu_{4\,bc}$ cm^{-1}
$\nu_0 + A_i'(1-\zeta_i)^2 - B_i'$	1476.1	3016.8	1154.7
$A_i'(1-\zeta_i) - B_i'$	2.91	0.944	−1.70
$(A_i' - A'') - (B_i' - B'')$	−0.040	−0.011	0.013
B''	3.880	3.880	3.880
B_i'	3.865	3.869	3.865
$B_i' - B''$	−0.015	−0.011	−0.015
D_J	5.5×10^{-5}	5.5×10^{-5}	5.5×10^{-5}

For higher K and J values, deviations between observed and calculated frequencies can be expected due to the neglect of the D_{JK} and D_K terms in Eqs. (16) and (17). In those cases the assignments must be regarded as tentative, although they are often supported by the observed relative intensities of the lines (see Table 10). The relative intensities have been calculated by the author for the ${}^{P}P_K(J)$, ${}^{R}R_K(J)$, ${}^{R}P_K(J)$, and ${}^{P}R_K(J)$ lines. For the $\nu_{2\,ab}$ and $\nu_{4\,bc}$ bands, the relative intensities of the Q lines will not deviate much from the corresponding Q line intensities of the $\nu_{3\,bc}$ band, calculated by JONES.[6]

Table 8. Observed and calculated fine structure lines of $\nu_{2\,ab}(E)$ band.

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
1432.8[a]	${}^{P}P_1(5)$	1431.1	−1.7	3.1
	${}^{P}P_2(4)$	1433.1	0.3	4.6
1435.4	${}^{P}P_3(3)$	1434.9	−0.5	14.0
1439.5	${}^{P}Q_6$	1439.7	0.2	
	${}^{P}P_1(4)$	1439.0	−0.5	3.1
1441.5[a]	${}^{P}P_2(3)$	1440.9	−0.6	4.8
1446.4	${}^{P}Q_5$	1446.0	−0.4	
	${}^{P}P_1(3)$	1446.8	0.4	2.9
1448.3[a]	${}^{P}P_2(2)$	1448.8	0.5	4.8
1453.0[a]	${}^{P}Q_4$	1452.2	−0.8	
1455.0[a]	${}^{P}P_1(2)$	1454.7	−0.3	2.5
1460.7	${}^{P}Q_3$	1458.3	−2.4	
	${}^{P}P_1(1)$	1462.4	1.7	1.8
1465.8	${}^{P}Q_2$	1464.3	−1.5	

(To be continued)

Table 8 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
1471.3	$^{P}Q_1$	1470.2	−1.1	
1476.3	$^{R}Q_0$	1476.1	−0.2	
1481.8	$^{R}Q_1$	1481.9	0.1	
1482.9[a]	$^{R}R_0(0)$	1483.8	0.9	3.8
1487.4	$^{R}Q_2$	1487.6	0.2	
1492.1	$^{R}R_0(1)$	1491.5	−0.6	5.5
1493.0[a]	$^{R}Q_3$	1493.2	0.2	
1496.0	$^{R}R_1(1)$	1497.3	1.3	5.5
1498.0[a]	$^{R}Q_4$	1498.8	0.8	
1500.2	$^{R}R_0(2)$	1499.2	−1.0	6.9
1504.5	$^{R}Q_5$	1504.2	−0.3	
	$^{R}R_1(2)$	1505.0	0.5	5.7
1506.2	$^{R}R_0(2)$	1506.8	0.6	7.7
1509.4	$^{R}Q_6$	1509.6	0.2	
	$^{R}R_2(2)$	1510.7	1.3	8.3
1513.1	$^{R}R_0(4)$	1514.3	1.2	8.0
	$^{R}R_1(3)$	1512.6	−0.5	5.7
	$^{R}Q_7$	1514.8	1.7	
1518.0	$^{R}R_1(4)$	1520.1	2.1	5.6
	$^{R}R_2(3)$	1518.3	0.3	7.9
	$^{R}Q_8$	1520.0	2.0	
1522.7	$^{R}R_0(5)$	1522.0	−0.7	7.8
	$^{R}R_3(3)$	1523.9	1.2	20.2
1527.3	$^{R}R_1(5)$	1527.8	0.5	5.2
	$^{R}R_2(4)$	1525.9	−1.4	7.3
1531.3	$^{R}R_0(6)$	1529.5	−1.8	7.2
	$^{R}R_2(5)$	1533.5	2.2	6.5
	$^{R}R_3(4)$	1531.5	0.2	18.1
1535.8	$^{R}R_0(7)$	1537.0	1.2	6.3
	$^{R}R_1(6)$	1535.3	−0.5	4.6
	$^{R}R_4(4)$	1537.0	1.2	10.7
1540.2	$^{R}R_2(6)$	1541.0	0.8	5.6
	$^{R}R_3(5)$	1539.1	−1.1	15.7
1544.5	$^{R}R_0(8)$	1544.4	−0.1	5.2
	$^{R}R_1(7)$	1542.8	−1.7	3.9
	$^{R}R_4(5)$	1544.6	0.1	9.1
1548.8	$^{R}R_1(8)$	1550.2	1.4	3.2
	$^{R}R_2(7)$	1548.5	−0.3	4.7
	$^{R}R_5(5)$	1550.1	1.3	10.3
	$^{R}R_3(6)$	1546.6	−2.2	13.2

(*To be continued*)

Table 8 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
1553.5	$^{R}R_0(9)$	1551.8	−1.7	4.1
	$^{R}R_3(7)$	1554.1	0.6	10.8
	$^{R}R_4(6)$	1552.1	−1.4	7.6
1558.0	$^{R}R_0(10)$	1559.2	1.2	3.1
	$^{R}R_1(9)$	1557.6	−0.4	2.5
	$^{R}R_2(8)$	1555.9	−2.1	3.7
	$^{R}R_4(7)$	1559.6	1.6	6.1
	$^{R}R_5(6)$	1557.6	−0.4	8.4
1561.6	$^{R}R_2(9)$	1563.3	1.7	2.8
	$^{R}R_3(8)$	1561.5	−0.1	8.5
	$^{R}R_6(6)$	1563.0	1.4	18.2
1567.6	$^{R}R_0(11)$	1566.5	−1.1	2.2
	$^{R}R_3(9)$	1568.9	1.3	6.4
	$^{R}R_4(8)$	1567.0	−0.6	4.7
1570.9	$^{R}R_2(10)$	1570.8	−0.1	2.1
	$^{R}R_5(8)$	1572.5	1.6	5.1
	$^{R}R_6(7)$	1570.5	−0.4	14.3

s_{eff} = 2.0–2.5 cm^{-1}.

[a] Not resolved.

Table 9. Observed and calculated fine structure lines of $\nu_{3\,bc}(E)$ band.

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
2914.0	$^{P}P_4(12)$	2914.8	0.8	0.9
	$P(7)(\nu_1)$	2914.7	0.7	
2916.6	$^{P}P_3(12)$	2916.8	0.2	1.7
	$^{P}P_7(11)$	2916.7	0.1	1.6
	$^{P}P_8(11)$	2914.7	−1.9	1.6
2920.1	$^{P}P_1(12)$	2920.6	0.5	0.6
	$^{P}P_{10}(10)$	2918.4	−1.7	1.2
2921.5[a]	$^{P}P_5(11)$	2920.8	−0.7	1.5
	$^{P}P_9(10)$	2920.5	−1.0	4.7
2923.3	$^{P}P_4(11)$	2922.8	−0.5	1.3
	$^{P}P_8(10)$	2922.6	−0.7	2.4
	$P(6)(\nu_1)$	2922.7	−0.6	

(To be continued)

Table 9 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
2924.8	$^{P}P_3(11)$	2924.7	−0.1	2.4
	$^{P}P_7(10)$	2924.6	−0.2	2.3
2927.0	$^{P}P_2(11)$	2926.7	−0.3	1.1
	$^{P}P_6(10)$	2926.7	−0.3	4.5
2929.4	$^{P}P_1(11)$	2928.6	−0.8	0.9
	$^{P}P_5(10)$	2928.7	−0.7	2.1
	$^{P}P_9(9)$	2928.4	−1.0	6.9
2931.1	$^{P}P_4(10)$	2930.7	−0.4	1.9
	$^{P}P_8(9)$	2930.5	−0.6	3.4
	$P(5)(\nu_1)$	2930.7	−0.4	
2932.8	$^{P}P_3(10)$	2932.6	−0.2	3.5
	$^{P}P_7(9)$	2932.5	−0.3	3.3
2934.8	$^{P}P_2(10)$	2934.6	−0.2	1.5
	$^{P}P_6(9)$	2934.6	−0.2	6.3
2937.1[a]	$^{P}P_1(10)$	2936,5	−0.6	1.3
	$^{P}P_5(9)$	2936.6	−0.5	2.9
2938.5	$^{P}P_4(9)$	2938.6	0.1	2.7
	$^{P}P_8(8)$	2938.4	−0.1	4.7
	$P(4)(\nu_1)$	2938.6	0.2	
2940.6	$^{P}P_3(9)$	2940.5	−0.1	4.8
	$^{P}P_7(8)$	2940.4	−0.2	4.6
2942.8[a]	$^{P}P_2(9)$	2942.5	−0.3	2.1
	$^{P}P_6(8)$	2942.5	−0.3	8.6
2944.2[a]	$^{P}P_1(9)$	2944.4	0.2	1.7
	$^{P}P_5(8)$	2944.5	0.3	4.0
2946.4	$^{P}P_4(8)$	2946.5	0.1	3.6
	$P(3)(\nu_1)$	2946.6	0.2	
2948.2	$^{P}P_3(8)$	2948.4	0.2	6.2
	$^{P}P_7(7)$	2948.3	0.1	6.1
2950.6	$^{P}P_2(8)$	2950.4	−0.2	2.6
	$^{P}P_6(7)$	2950.4	−0.2	11.4
2952.6	$^{P}P_1(8)$	2952.3	−0.3	2.2
	$^{P}P_5(7)$	2952.4	−0.2	5.2
2954.0	$^{P}P_4(7)$	2954.4	0.4	4.6
	$P(2)(\nu_1)$	2954.5	0.5	
2956.4	$^{P}P_3(7)$	2956.3	−0.1	7.9
2958.2	$^{P}P_2(7)$	2958.3	0.1	3.3
	$^{P}P_6(6)$	2958.2	0.0	14.7

(*To be continued*)

Table 9 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity	
2960.2	$^{P}P_1(7)$	2960.2	0.0	2.6	
	$^{P}P_5(6)$	2960.2	0.0	6.6	
2962.1	$^{P}P_4(6)$	2962.2	0.1	5.7	
	$P(1)(v_1)$	2962.3	0.2		
2964.4	$^{P}P_3(6)$	2964.2	−0.2	9.6	
2966.0[a]	$^{P}P_2(6)$	2966.1	0.1	3.9	
2971.9	$^{P}P_3(5)$	2972.0	0.1	11.2	
2974.1	$^{P}P_2(5)$	2974.0	−0.1	4.3	
2976.0	$^{P}P_1(5)$	2975.9	−0.1	3.2	
2978.3	$^{P}P_4(4)$	2977.9	−0.4	8.1	
	$R(0)(v_1)$	2977.9	−0.4		
2980.2	$^{P}P_3(4)$	2979.8	−0.4	12.7	
2981.7[a]	$^{P}P_2(4)$	2981.8	0.1	4.7	
2984.0[a]	$^{P}P_1(4)$	2983.7	−0.3	3.2	
2985.7	$^{R}P_0(4)$	2985.6	−0.1	3.8	
	$R(1)(v_1)$	2985.6	−0.1		
					(Ref. 6)
2988.2	$^{P}P_3(3)$	2987.7	−0.5	14.1	15.1
2989.8[a]	$^{P}P_2(3)$	2989.6	−0.2	4.8	5.2
2991.9[a]	$^{P}P_1(3)$	2991.5	−0.4	3.0	3.2
2993.5	$^{R}P_0(3)$	2993.4	−0.1	3.0	3.2
	$R(2)(v_1)$	2993.3	−0.2		
2995.2[a]	$^{P}Q_{11}$	2994.7	−0.5		0.8
	$^{R}P_1(3)$	2995.3	0.1	0.5	0.5
2999.2	$^{P}Q_9$	2998.9	−0.3		5.2
	$^{P}P_1(2)$	2999.4	0.2	2.5	2.7
3001.0	$^{P}Q_8$	3001.0	0.0		4.7
	$R(3)(v_1)$	3000.9	−0.1		
3003.1	$^{P}Q_7$	3003.0	−0.1		7.7
3005.0	$^{P}Q_6$	3005.1	0.1		24.5
3007.0	$^{P}Q_5$	3007.1	0.1		18.3
	$^{P}P_1(1)$	3007.1	0.1	1.8	1.9
3009.0[a]	$^{P}Q_4$	3009.1	0.1		26.0
	$R(4)(v_1)$	3008.5	−0.5		
3010.6	$^{P}Q_3$	3011.0	0.4		70.0
3012.9	$^{P}Q_2$	3013.0	0.1		44.3
3015.3	$^{P}Q_1$	3014.9	−0.4		52.4
3016.7	$^{R}Q_0$	3016.8	0.1	100	107.5
3018.3	$^{R}Q_1$	3018.7	0.4		45.4
3020.8	$^{R}Q_2$	3020.5	−0.3		36.0

(*To be continued*)

Table 9 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity	
					(Ref. 6)
3022.3	$^{R}Q_3$	3022.4	0.1		54.0
3024.2[a]	$^{R}Q_4$	3024.2	0.0		19.6
	$^{R}R_0(0)$	3024.5	0.3	3.8	4.0
3025.5[a]	$^{R}Q_5$	3026.0	0.5		13.2
3027.5[a]	$^{R}Q_6$	3027.7	0.2		17.1
3029.5	$^{R}Q_7$	3029.5	0.0		5.2
3030.9	$^{R}Q_8$	3031.2	0.3		3.0
3032.5	$^{R}Q_9$	3032.9	0.4	3.4	3.4
	$^{R}R_0(1)$	3032.3	−0.2	5.5	5.8
3034.3	$^{R}Q_{10}$	3034.6	0.3		1.0
	$^{R}R_1(1)$	3034.2	−0.1	5.4	5.8
3036.1[a]	$^{R}Q_{11}$	3036.2	0.1		0.6
	$^{P}R_2(2)$	3036.1	0.0	0.6	0.6
3037.9	$^{P}R_1(2)$	3038.1	0.2	1.7	1.8
3039.9	$^{R}R_0(2)$	3040.0	0.1	6.8	7.2
3042.2	$^{R}R_1(2)$	3041.9	−0.3	5.6	5.9
3043.7[a]	$^{R}R_2(2)$	3043.7	0.0	8.3	8.7
3045.7[a]	$^{P}R_1(3)$	3045.7	0.0	2.3	2.4
3047.8	$^{R}R_0(3)$	3047.6	−0.2	7.6	8.0
3049.4	$^{R}R_1(3)$	3049.5	0.1	5.7	6.0
3051.4	$^{R}R_2(3)$	3051.3	−0.1	7.8	8.2
3053.3	$^{R}R_3(3)$	3053.2	−0.1	20.1	21.1
3055.4	$^{R}R_0(4)$	3055.3	−0.1	7.9	8.3
3057.4	$^{R}R_1(4)$	3057.2	−0.2	5.5	5.8
3059.0	$^{R}R_2(4)$	3059.0	0.0	7.2	7.5
3061.2	$^{R}R_3(4)$	3060.8	−0.4	17.9	18.7
3062.9	$^{R}R_0(5)$	3062.9	0.0	7.7	8.0
	$^{R}R_4(4)$	3062.7	−0.2	10.7	11.1
3064.4	$^{R}R_1(5)$	3064.8	0.4	5.1	5.3
3066.9	$^{R}R_2(5)$	3066.6	−0.3	6.4	6.7
3068.4	$^{R}R_3(5)$	3068.4	0.0	15.5	16.1
3070.7	$^{R}R_0(6)$	3070.5	−0.2	7.0	7.3
	$^{R}R_4(5)$	3070.3	−0.4	9.1	9.4
3072.3[a]	$^{R}R_1(6)$	3072.4	0.1	4.5	4.7
	$^{R}R_5(5)$	3072.1	−0.2	10.2	10.7
3074.1[a]	$^{R}R_2(6)$	3074.2	0.1	5.5	5.7
3076.3	$^{R}R_3(6)$	3076.0	−0.3	13.0	13.5
3078.2	$^{R}R_0(7)$	3078.0	−0.2	6.1	6.3
	$^{R}R_4(6)$	3077.9	−0.3	7.5	7.7

(*To be continued*)

Table 9 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity	
					(Ref. 6)
3080.1	$^{R}R_1(7)$	3079.9	−0.2	3.8	3.9
	$^{R}R_5(6)$	3079.7	−0.4	8.4	10.1
3081.8	$^{R}R_2(7)$	3081.7	−0.1	4.5	4.7
	$^{R}R_6(6)$	3081.4	−0.4	18.1	21.8
3083.6	$^{R}R_3(7)$	3083.6	0.0	10.6	10.9
3085.6	$^{R}R_0(8)$	3085.5	−0.1	5.1	5.2
	$^{R}R_4(7)$	3085.4	−0.2	6.0	6.2
3087.4	$^{R}R_1(8)$	3087.4	0.0	3.1	3.2
	$^{R}R_5(7)$	3087.2	−0.2	6.6	6.8
3089.4	$^{R}R_2(8)$	3089.2	−0.2	3.6	3.7
	$^{R}R_6(7)$	3088.9	−0.5	14.1	14.5
3091.2	$^{R}R_3(8)$	3091.1	−0.1	8.3	8.5
	$^{R}R_7(7)$	3090.7	−0.5	7.4	7.6
3093.0	$^{R}R_0(9)$	3093.0	0.0	4.0	4.1
	$^{R}R_4(8)$	3092.9	−0.1	4.6	4.7
3095.0	$^{R}R_1(9)$	3094.9	−0.1	2.4	2.4
	$^{R}R_5(8)$	3094.7	−0.3	5.0	5.2
3096.8	$^{R}R_2(9)$	3096.7	−0.1	2.8	2.8
	$^{R}R_6(8)$	3096.4	−0.4	10.7	10.9
3098.5	$^{R}R_3(9)$	3098.6	0.1	6.2	6.4
	$^{R}R_7(8)$	3098.2	−0.3	5.5	5.7
3100.6	$^{R}R_0(10)$	3100.5	−0.1	3.0	3.1
	$^{R}R_4(9)$	3100.4	−0.2	3.4	3.5
	$^{R}R_8(8)$	3099.9	−0.7	5.7	5.8
3102.4	$^{R}R_1(10)$	3102.4	0.0	1.8	1.8
	$^{R}R_5(9)$	3102.2	−0.2	3.7	3.8
3104.2	$^{R}R_2(10)$	3104.2	0.0	2.0	2.1
	$^{R}R_6(9)$	3103.9	−0.3	7.8	8.0
3106.4	$^{R}R_3(10)$	3106.0	−0.4	4.5	4.5
	$^{R}R_7(9)$	3105.7	−0.7	4.0	4.1
3107.4[a]	$^{R}R_0(11)$	3107.9	0.5	2.2	2.2
	$^{R}R_4(10)$	3107.9	0.5	2.5	2.5
	$^{R}R_8(9)$	3107.4	0.0	4.1	4.2
3109.4	$^{R}R_1(11)$	3109.8	0.4	1.3	1.3
	$^{R}R_5(10)$	3109.7	0.3	2.6	2.7
	$^{R}R_9(9)$	3109.1	−0.3	8.1	8.3
3111.9	$^{R}R_2(11)$	3111.6	−0.3	1.4	1.4
	$^{R}R_6(10)$	3111.4	−0.5	5.5	5.6

(To be continued)

Table 9 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity	(Ref. 6)
3113.5	$^{R}R_3(11)$	3113.4	−0.1	3.1	3.2
	$^{R}R_7(10)$	3113.2	−0.3	2.8	2.9
3115.3	$^{R}R_0(12)$	3115.3	0.0	1.6	1.5
	$^{R}R_4(11)$	3115.3	0.0	1.7	1.7
	$^{R}R_8(10)$	3114.9	−0.4	2.9	2.9
3116.5	$^{R}R_1(12)$	3117.2	0.7	0.9	0.7
	$^{R}R_5(11)$	3117.1	0.6	1.8	1.9
	$^{R}R_9(10)$	3116.6	0.1	5.6	5.7
3118.9	$^{R}R_2(12)$	3119.0	0.1	1.0	1.0
	$^{R}R_6(11)$	3118.8	−0.1	3.8	4.5

$s_{eff} = 1.3–1.5\ cm^{-1}$.
[a] Not resolved.

Table 10. Observed and calculated fine structure lines of $\nu_{4\,bc}(E)$ band.

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
1028.6	$^{R}P_2(15)$	1029.2	0.6	0.1
	$^{R}P_0(16)$	1027.8	−0.8	0.2
1030.7	$^{P}P_1(16)$	1031.2	0.5	0.1
	$^{P}P_3(17)$	1030.1	−0.6	0.1
	$^{R}P_4(14)$	1030.5	−0.2	0.1
1033.0	$^{P}P_4(17)$	1033.6	0.6	0.1
	$^{P}P_6(18)$	1031.3	−1.7	0.1
	$^{R}P_1(15)$	1032.5	−0.5	0.1
	$^{R}P_3(14)$	1033.8	0.8	0.3
	$^{R}P_6(13)$	1032.1	−0.9	0.1
1035.0	$^{P}P_2(16)$	1034.7	−0.3	0.1
	$^{R}P_0(15)$	1035.9	0.9	0.3
	$^{R}P_5(13)$	1035.3	0.3	0.1
1037.9	$^{P}P_3(16)$	1038.1	0.2	0.2
	$^{P}P_5(17)$	1037.1	−0.8	0.1
	$^{R}P_2(14)$	1037.2	−0.7	0.2
	$^{R}P_4(13)$	1038.6	0.7	0.1
1040.0	$^{P}P_1(15)$	1039.3	−0.7	0.2
	$^{R}P_1(14)$	1040.5	0.5	0.2
	$^{R}P_6(12)$	1040.1	0.1	0.2

(*To be continued*)

Table 10 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
	$^{P}P_2(15)$	1042.8	0.2	0.2
	$^{P}P_4(16)$	1041.6	− 1.0	0.1
1042.6	$^{P}P_6(17)$	1040.7	− 1.9	0.1
	$^{R}P_0(14)$	1043.9	1.3	0.4
	$^{R}P_3(13)$	1041.9	− 0.7	0.4
	$^{R}P_5(12)$	1043.3	0.7	0.1
	$^{P}P_1(14)$	1047.3	0.2	0.3
	$^{P}P_3(15)$	1046.2	− 0.9	0.4
1047.1	$^{P}P_5(16)$	1045.1	− 2.0	0.1
	$^{R}P_4(12)$	1046.6	− 0.5	0.2
1049.4	$^{P}P_4(15)$	1049.7	0.3	0.2
	$^{R}P_1(13)$	1048.6	− 0.8	0.3
	$^{P}P_2(14)$	1050.8	− 0.6	0.3
	$^{P}P_6(16)$	1048.7	− 2.4	0.3
1051.4[a]	$^{P}P_8(17)$	1047.8	− 3.6	0.1
	$^{R}P_0(13)$	1052.0	0.6	0.7
	$^{R}P_3(12)$	1049.9	− 1.5	0.5
	$^{R}P_5(11)$	1051.3	− 0.1	0.2
	$^{P}P_3(14)$	1054.2	1.2	0.6
	$^{P}P_5(15)$	1053.2	0.2	0.2
1053.0[a]	$^{P}P_7(16)$	1052.2	− 0.8	0.1
	$^{R}P_2(12)$	1053.3	0.3	0.3
	$^{R}P_4(11)$	1054.6	1.6	0.3
	$^{P}P_1(13)$	1055.4	− 0.6	0.4
1056.0	$^{P}P_9(17)$	1051.5	− 4.5	0.1
	$^{R}P_1(12)$	1056.6	0.6	0.4
	$^{P}P_2(13)$	1058.9	− 0.1	0.5
	$^{P}P_4(14)$	1057.7	− 1.3	0.3
	$^{P}P_6(15)$	1056.8	− 2.2	0.4
1059.0[a]	$^{P}P_8(16)$	1055.8	− 3.2	0.1
	$^{R}P_3(11)$	1057.9	− 1.1	0.7
	$^{R}P_5(10)$	1059.3	0.3	0.2
	$^{P}P_3(13)$	1062.3	0.8	1.0
	$^{P}P_5(14)$	1061.2	− 0.3	0.4
	$^{P}P_7(15)$	1060.3	− 1.2	0.2
1061.5	$^{P}P_9(16)$	1058.6	− 2.9	0.2
	$^{R}P_0(12)$	1060.0	− 1.5	1.0
	$^{R}P_2(11)$	1061.3	− 0.2	0.5
	$^{R}P_4(10)$	1062.6	1.1	0.3

(*To be continued*)

Table 10 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
1064.5	$^{P}P_1(12)$	1063.4	−1.1	0.6
	$^{P}P_6(14)$	1064.8	0.3	0.8
	$^{R}P_1(11)$	1064.6	0.1	0.6
1067.6	$^{P}P_2(12)$	1066.5	−1.1	0.7
	$^{P}P_4(13)$	1065.8	−1.8	0.6
	$^{P}P_8(15)$	1063.9	−3.7	0.2
	$^{R}P_0(11)$	1068.0	0.4	1.5
1070.2[a]	$^{P}P_3(12)$	1070.3	0.1	1.6
	$^{P}P_5(13)$	1069.3	−0.9	0.6
	$^{P}P_7(14)$	1068.3	−1.9	0.4
	$^{P}P_9(15)$	1067.6	−2.6	0.4
	$^{R}P_2(10)$	1069.3	−0.9	0.6
	$^{R}P_4(9)$	1070.5	0.3	0.3
1071.0[a]	$^{P}P_1(11)$	1071.4	0.4	0.9
1073.1	$^{P}P_4(12)$	1073.9	0.8	0.9
	$^{P}P_6(13)$	1072.9	−0.2	1.2
	$^{P}P_8(14)$	1071.9	−1.2	0.4
	$^{R}P_1(10)$	1072.6	−0.5	0.8
	$^{R}P_3(9)$	1073.8	0.7	1.0
1076.3	$^{P}P_2(11)$	1074.9	−1.4	1.0
	$^{P}P_5(12)$	1076.9	0.6	0.9
	$^{R}P_0(10)$	1076.0	−0.3	2.0
	$^{R}P_2(9)$	1077.2	0.9	0.8
1078.5	$^{P}P_3(11)$	1078.3	−0.2	2.4
	$^{P}P_7(13)$	1076.4	−2.1	0.6
	$^{P}P_9(14)$	1075.6	−2.9	0.8
	$^{R}P_4(8)$	1078.5	0.0	0.3
1079.9	$^{P}P_1(10)$	1079.4	−0.5	1.2
	$^{R}P_1(9)$	1080.5	0.6	1.0
1081.9	$^{P}P_2(10)$	1082.9	1.0	1.5
	$^{P}P_4(11)$	1081.8	−0.1	1.3
	$^{P}P_6(12)$	1081.2	−0.7	2.0
	$^{P}P_8(13)$	1080.0	−1.9	0.6
	$^{R}P_3(8)$	1081.8	−0.1	1.1
1084.6	$^{P}P_5(11)$	1085.3	0.7	1.4
	$^{R}P_0(9)$	1083.9	−0.7	2.7
	$^{R}P_2(8)$	1085.2	0.6	0.9
1087.1	$^{P}P_1(9)$	1087.3	0.2	1.7
	$^{P}P_3(10)$	1086.3	−0.8	3.4

(*To be continued*)

Table 10 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
1087.1	$^{P}P_7(12)$	1084.4	−2.7	1.0
	$^{P}P_9(13)$	1083.7	−3.4	1.3
1089.3	$^{P}P_4(10)$	1089.8	0.5	1.9
	$^{P}P_6(11)$	1088.9	−0.4	3.0
	$^{P}P_8(12)$	1088.0	−1.3	1.0
	$^{R}P_1(8)$	1088.5	−0.8	1.2
	$^{R}P_3(7)$	1089.7	0.4	1.0
1091.2	$^{P}P_2(9)$	1090.8	−0.4	2.0
	$^{R}P_0(8)$	1091.9	0.7	0.9
1092.8	$^{P}P_5(10)$	1093.3	0.5	2.1
	$^{P}P_7(11)$	1092.4	−0.4	1.5
	$^{P}P_9(12)$	1091.7	−1.1	2.0
	$^{R}P_2(7)$	1093.1	0.3	0.9
1095.8	$^{P}P_1(8)$	1095.3	−0.5	2.1
	$^{P}P_3(9)$	1094.2	−1.6	4.6
	$^{P}P_6(10)$	1096.9	1.1	4.4
	$^{R}P_1(7)$	1096.4	0.6	1.3
1098.0	$^{P}P_2(8)$	1098.8	0.8	2.6
	$^{P}P_4(9)$	1097.7	−0.3	2.6
	$^{P}P_8(11)$	1096.0	−2.0	1.6
1100.5[a]	$^{P}P_5(9)$	1101.2	0.7	2.9
	$^{P}P_7(10)$	1100.4	−0.1	2.3
	$^{P}P_9(11)$	1099.7	−0.8	3.1
	$^{R}P_0(7)$	1099.8	−0.7	3.8
1101.5[a]	$^{P}P_3(8)$	1102.2	0.7	6.1
1104.6	$^{P}P_1(7)$	1103.2	−1.4	2.5
	$^{P}P_6(9)$	1104.8	0.2	6.2
	$^{P}P_8(10)$	1104.0	−0.6	2.3
1106.3[a]	$^{P}P_2(7)$	1106.7	0.4	3.2
	$^{P}P_4(8)$	1105.7	−0.6	3.5
1109.1	$^{P}P_3(7)$	1110.1	1.0	7.7
	$^{P}P_5(8)$	1109.2	0.1	3.9
	$^{P}P_7(9)$	1108.3	−0.8	3.3
	$^{P}P_9(10)$	1107.7	−1.4	4.7
1113.2	$^{P}P_4(7)$	1113.6	0.4	4.5
	$^{P}P_6(8)$	1112.8	−0.4	8.5
	$^{P}P_8(9)$	1111.9	−1.3	3.4
1114.8	$^{P}P_2(6)$	1114.6	−0.2	3.8
	$^{P}P_9(9)$	1115.6	0.8	6.8
	$^{R}P_0(5)$	1115.6	0.8	4.2

(*To be continued*)

Table 10 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
	$^{P}P_3(6)$	1118.0	0.4	9.4
1117.6	$^{P}P_5(7)$	1117.1	−0.5	5.1
	$^{P}P_7(8)$	1116.3	−1.3	4.5
1119.1[a]	$^{P}P_1(5)$	1119.0	−0.1	3.1
	$^{P}P_2(5)$	1122.5	0.7	4.3
	$^{P}P_4(6)$	1121.5	−0.3	5.6
1121.8	$^{P}P_6(7)$	1120.7	−1.1	11.4
	$^{P}P_8(8)$	1119.9	−1.9	4.7
	$^{R}Q_{10}$	1122.0	0.2	
1122.8[a]	$^{R}P_0(4)$	1123.5	0.7	3.8
1123.4[a]	$^{P}P_7(7)$	1124.2	0.8	6.1
	$^{P}P_3(5)$	1125.9	−0.2	11.1
1126.1	$^{P}P_5(6)$	1125.0	−1.1	6.5
	$^{R}Q_9$	1125.2	−0.9	3.3
1127.0[a]	$^{P}P_1(4)$	1126.9	−0.1	3.1
	$^{R}Q_8$	1128.3	1.3	
	$^{P}P_2(4)$	1130.3	0.3	4.6
1130.1	$^{P}P_4(5)$	1129.4	−0.7	6.8
	$^{P}P_6(6)$	1128.6	−1.5	14.6
	$^{R}Q_7$	1131.5	0.0	
1131.5[a]	$^{P}P_5(5)$	1132.9	1.4	8.1
	$^{R}P_0(3)$	1131.3	−0.2	3.0
	$^{R}Q_6$	1134.8	0.3	
1134.5	$^{P}P_1(3)$	1134.7	0.2	2.9
	$^{P}P_3(4)$	1133.8	−0.7	12.6
	$^{R}Q_5$	1138.0	−0.5	
1138.5	$^{P}P_2(3)$	1138.2	−0.3	4.8
	$^{P}P_4(4)$	1137.3	−1.2	8.1
	$^{R}Q_4$	1141.3	−1.5	
1142.8	$^{P}P_1(2)$	1142.6	−0.2	2.5
	$^{P}P_3(3)$	1141.6	−1.2	14.0
1144.4	$^{R}Q_3$	1144.6	0.2	
1147.0[a]	$^{P}P_2(2)$	1146.1	−0.9	4.9
1148.0[a]	$^{R}Q_2$	1148.0	0.0	
	$^{R}Q_1$	1151.3	0.0	
1151.3	$^{P}P_1(1)$	1150.3	−1.0	1.8
1154.8	$^{R}Q_0$	1154.7	−0.1	100
1158.0	$^{P}Q_1$	1158.1	0.1	
	$^{P}Q_2$	1161.6	0.3	
1161.3	$^{R}R_0(0)$	1162.4	1.1	3.8

(To be continued)

Table 10 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
1164.9	$^{P}Q_3$	1165.0	0.1	
1167.7	$^{P}Q_4$	1168.5	0.8	
	$^{R}R_1(1)$	1166.7	−1.0	5.5
1171.4	$^{P}Q_5$	1172.0	0.6	
	$^{R}R_2(2)$	1171.1	−0.3	8.3
	$^{R}R_0(1)$	1170.1	−1.3	5.5
1174.7[a]	$^{R}R_1(2)$	1174.4	−0.3	5.7
1175.2[a]	$^{R}R_3(3)$	1175.3	0.1	20.2
1175.7	$^{P}Q_6$	1175.6	−0.1	
1177.8	$^{R}R_0(2)$	1177.8	0.0	6.9
	$^{R}R_2(3)$	1178.7	0.9	7.9
1179.0[a]	$^{P}Q_7$	1179.1	0.1	
1180.7[a]	$^{R}R_4(4)$	1179.6	−1.1	10.7
1182.8	$^{R}R_1(3)$	1182.0	−0.8	5.7
	$^{R}R_3(4)$	1182.9	0.1	18.1
	$^{P}Q_8$	1182.7	−0.1	
1185.0	$^{R}R_0(3)$	1185.4	0.4	7.7
	$^{R}R_2(4)$	1186.3	1.3	7.3
	$^{R}R_5(5)$	1183.9	−1.1	10.3
1186.1[a]	$^{P}Q_9$	1186.4	0.3	
1187.6	$^{R}R_4(5)$	1187.2	−0.4	9.1
1189.4	$^{R}R_1(4)$	1189.6	0.2	5.6
	$^{R}R_3(5)$	1190.5	1.1	15.7
	$^{R}R_6(6)$	1188.2	−1.2	18.2
	$^{P}Q_{10}$	1190.0	0.6	
1192.2	$^{R}R_5(6)$	1191.3	−0.9	8.4
	$^{R}R_7(7)$	1192.4	0.2	7.5
	$P(13)$ (ν_{3a})	1191.7	−0.5	
1193.7	$^{R}R_0(4)$	1193.0	−0.7	8.0
	$^{R}R_2(5)$	1193.9	0.2	6.5
	$^{R}R_4(6)$	1194.6	0.9	7.6
1196.6	$^{R}R_1(5)$	1197.2	0.6	5.2
	$^{R}R_6(7)$	1195.7	−0.9	14.3
	$^{R}R_8(8)$	1196.6	0.0	5.7
1198.6	$^{R}R_3(6)$	1197.9	−0.7	13.2
	$^{R}R_5(7)$	1198.9	0.3	6.7
1201.3[a]	$^{R}R_0(5)$	1200.6	−0.7	7.8
	$^{R}R_2(6)$	1201.5	0.2	5.6
	$^{R}R_4(7)$	1202.2	0.9	6.1
	$^{R}R_7(8)$	1199.8	−1.5	5.6
	$P(12)$ (ν_{3a})	1201.6	0.3	
1203.5	$^{R}R_6(8)$	1203.1	−0.4	10.8

(*To be continued*)

Table 10 (continued).

Obs. cm^{-1}	Assignment	Calc. cm^{-1}	Δ (Calc.-obs.)	Calc. relative intensity
	$^{R}R_1(6)$	1204.7	−0.3	4.6
1205.0	$^{R}R_3(7)$	1205.5	0.5	10.8
	$^{R}R_8(9)$	1204.0	−1.0	4.1
	$^{R}R_0(6)$	1208.1	0.4	7.2
	$^{R}R_2(7)$	1208.9	1.2	4.7
1207.7	$^{R}R_5(8)$	1206.3	−1.4	5.1
	$^{R}R_7(9)$	1207.2	−0.5	4.1
	$^{R}R_1(7)$	1212.2	0.8	3.9
	$^{R}R_3(8)$	1212.9	1.5	8.5
	$^{R}R_4(8)$	1209.6	−1.8	4.7
1211.4	$^{R}R_6(9)$	1210.5	−0.9	8.0
	$^{R}R_8(10)$	1211.4	0.0	2.9
	$^{R}R_9(10)$	1208.3	−3.1	5.7
	$P(11)$ (ν_{3a})	1211.3	−0.1	
	$^{R}R_0(7)$	1215.6	1.1	6.3
1214.5	$^{R}R_5(9)$	1213.7	−0.8	3.8
	$^{R}R_7(10)$	1214.6	0.1	2.9
	$^{R}R_2(8)$	1216.3	−1.4	3.7
	$^{R}R_4(9)$	1217.0	−0.7	3.5
1217.7	$^{R}R_6(10)$	1217.9	0.2	5.6
	$^{R}R_9(11)$	1215.6	−2.1	3.9
	$^{R}R_1(8)$	1219.6	−1.2	3.2
	$^{R}R_3(9)$	1220.3	−0.5	6.4
1220.8	$^{R}R_5(10)$	1221.1	0.3	2.7
	$^{R}R_8(11)$	1218.7	−2.1	2.0
	$P(10)$ (ν_{3a})	1220.8	0.0	
	$^{R}R_0(8)$	1223.1	−1.0	5.2
	$^{R}R_2(9)$	1223.7	−0.4	2.8
	$^{R}R_4(10)$	1224.4	0.3	2.5
1224.1	$^{R}R_6(11)$	1225.2	1.1	3.9
	$^{R}R_7(11)$	1221.9	−2.2	2.0
	$^{R}R_9(12)$	1222.9	−1.2	2.5
	$^{R}R_1(9)$	1227.0	−1.0	2.5
1228.0[a]	$^{R}R_3(10)$	1227.7	−0.3	4.7
	$^{R}R_8(12)$	1226.0	−2.0	1.3
	$P(9)$ (ν_{3a})	1230.2	0.1	
	$^{R}R_0(9)$	1230.4	0.3	4.1
1230.1	$^{R}R_2(10)$	1231.1	1.0	2.1
	$^{R}R_5(11)$	1228.4	−1.7	1.9
	$^{R}R_7(12)$	1229.2	−0.9	1.3
1235.9	$^{R}R_3(11)$	1235.0	−0.9	3.3
	$^{R}R_5(12)$	1235.7	−0.2	1.2

s_{eff} = 1.3–1.9 cm^{-1}. [a] Not resolved.

3. Results

From the rotational analysis values have been derived for the rotational constants A'' and A_i', the Coriolis coupling factors ζ_i, and the band-centre frequencies ν_0. It can be shown[13] that to a good approximation the following expression is valid:

$$A'' = \frac{1}{3}\sum [A_i'(1-\zeta_i) - B_i'] - \frac{1}{3}\sum [(A_i' - A'') - (B_i' - B'')] + \frac{7}{6}B''.$$

As the magnitude of B'' is known (3.880_0 cm^{-1}), A'' can be calculated. From the known values of $B_i' - B''$, $(A_i' - A'') - (B_i' - B'')$, and A'', A_i' can be evaluated for the three perpendicular bands. The values of A_i', B_i', and $A_i'(1-\zeta_i) - B_i'$ have then been used for the calculation of the ζ values. Finally, the band-centre frequencies were obtained from $\nu_0 + A_i'(1-\zeta_i)^2 - B_i'$. The results are given in Table 11. The value $A'' = 5.25_7 \pm 0.02$ cm^{-1} derived here from the analysis of the three perpendicular bands is close to the one (5.245 cm^{-1}) calculated from B'', assuming $r_{CH} = r_{CD}$ and regular tetrahedral structure, and it also compares well with the value 5.243 cm^{-1} obtained from the Raman study of the $\nu_{3\,bc}$ band.[7]

Table 11. Rotational constants A', Coriolis coupling factors ζ, and band-centre frequencies ν_0 for perpendicular bands.

	$\nu_{2\,ab}$	$\nu_{3\,bc}$	$\nu_{4\,bc}$
A_i'(cm^{-1})	5.20_2	5.23_5	5.25_5
ζ_i	−0.302	0.081	0.588
ν_0(cm^{-1})	1471.2	3016.2	1157.7

In Table 12 the results for the $\nu_{3\,bc}$ band are compared with the values obtained from the Raman investigation by Richardson *et al.* (E.H.R.) and the infrared study by Jones (L.H.J.). It is seen that the author's results are in somewhat better agreement with the Raman investigation than with the infrared study.

In Table 13, experimental ζ values are compared with theoretical values calculated recently from force constants by Jones and McDowell,[14] and by Mills.[15] The agreement between the values obtained by the present investigations and the theoretical work seems satisfactory, except in the case of $\zeta_{3\,bc}$.

The sum of the zetas must satisfy the ζ sum rule:[16] $\Sigma\zeta_i = B_e/2A_e$, as far as anharmonicity can be neglected, and no resonances occur. As the "equilibrium" rotational constants, A_e and B_e, are not known, one has to use the

Table 12. Band constant values for the $\nu_{3\,bc}$ band in cm^{-1}.

	Raman[7] E.H.R. et al.	Infrared[14] L.H.J.	This invest.
$A'(1-\zeta)-B'$	0.955	0.924	0.944
$A''-A'\zeta-B''$	0.961	0.920	0.954
$(A'-A'')-(B'-B'')$	−0.0058	0.004	−0.011
$\nu_0+A'(1-\zeta)^2-B'$	3017.17	3017.4	3016.8
A'	5.223	5.24	5.23_5
B'	3.863	3.865	3.87_3
B''	3.877	—	3.88_2
$B'-B''$	−0.0138	−0.013	−0.011
ζ	0.0775	0.08_6	0.081
ν_0	3016.59	3016.9	3016.2
D_J'	5×10^{-5}	5.5×10^{-5} a	$5._5\times10^{-5}$
D_J''	4.7×10^{-5}		

a Assumed value.

Table 13. Experimental and theoretical ζ values for doubly degenerate normal vibrations.

	$\zeta_{2\,ab}$	$\zeta_{3\,bc}$	$\zeta_{4\,bc}$
Experimental values:			
N. Ginsburg and E. F. Barker	−0.218 to −0.312	0.261 to 0.213	0.625 to 0.692
L. H. Jones		0.08_6	
E. H. Richardson et al.		0.0775	
This investigation	−0.302	0.081	0.588
Theoretical values:			
L. H. Jones and R. S. McDowell	−0.260	0.040	0.589
I. M. Mills	−0.263	0.044	0.588

values for the vibrational ground state. This must, however, be considered a rather good approximation. Thus, the ζ sum rule may be written: $\Sigma\zeta_i = B''/2A'' = 0.370$. The sum of the zetas is 0.367, which should be compared with 0.668 to 0.593 found earlier by Ginsburg and Barker.[17]

A comparison between the present fundamental frequencies and values from previous investigations is given in Table 14. The values adopted here mean a revision of the results of Wilmshurst and Bernstein regarding the fundamentals $\nu_{3\,a}$, $\nu_{4\,a}$, $\nu_{3\,bc}$, and $\nu_{4\,bc}$.

Table 14. Comparison of the present results with previous grating or prism data (cm^{-1}).

Symmetry	Normal vibration	Present investigation	Other investigations	Adopted values
A_1	ν_1	2948*	2945[a]	2948
	$\nu_{3\,a}$	2210*	2200[a]	2210
			2200.03[b]	
	$\nu_{4\,a}$	1306.5	1306.8[c]	1306.5
			1300[a]	
E	$\nu_{2\,ab}$	1471.2	1477.1[c]	1471.2
			1471[a]	
	$\nu_{3\,bc}$	3016.2	3021[a]	3016.6
			3016.9[d]	
			3016.59[e]	
	$\nu_{4\,bc}$	1157.7	1156.3[c]	1157.7
			1155[a]	

[a] See Ref. 5. * Estimated unperturbed frequency (see page 6).
[b] See Ref. 3.
[c] See Ref. 17.
[d] See Ref. 6.
[e] See Ref. 7.

IV. Overtones and Combinations

Overtone and combination bands have been observed in the region 2000–6000 cm^{-1}. Thirty-six pronounced absorption bands have been measured and interpreted as summation bands. It was possible to explain sixteen bands as binary combinations, whereas the rest of the bands have been interpreted as ternary and quaternary summation bands. Only in three cases, however, it has been necessary to make use of quaternary combinations.

The results are given in Table 15. In the column with observed frequencies square brackets indicate the possible presence of Fermi resonance, which makes definite assignments uncertain.

Only below 5000 cm^{-1} the bands have been characterized as parallel (∥) or perpendicular (⊥) bands. The reason for this is the limited resolving power of the spectrograph at high frequencies and the increased possibility of interactions between bands, which may change the shape of the bands considerably.

For the calculation of the combination frequencies, observed frequencies have been used rather than calculated values. In some cases this gives more

Table 15. Possible assignments of overtones and combinations.

Assignment	Symmetry	Band structure obs.	Intensity	Frequency (cm^{-1}) Obs.	Calc.	Δ (Calc.-obs.)
$4bc+4bc$	A_1+E	∥	s	2316[a]	2315	−1
$4a+4bc$	E		vw	ca. 2460	2464	ca. +4
$4a+4a$	A_1	∥	w	2597	2613	+16
$2ab+2ab$	A_1+E		s	2910[a]	2942	+32
$3a+4bc$	E	∥	w	3337	3358	+21
$3a+4a$	A_1		m	3498	3506	+8
$4a+4bc+4bc$	A_1+E		vw	3617	3622	+5
$2ab+3a$	E		vw	3670	3671	+1
$4a+4a+4bc$	E	⊥	vw	ca. 3750	3771	ca. +21
$2ab+4bc+4bc$	A_1+A_2+2E	⊥	vw	ca. 3795	3787	ca. −8
$4a+4a+4a$	A_1		vw	3874	3919	+45
$2ab+4a+4a$	E		w	4056	4084	+28
$2ab+2ab+4bc$	A_1+A_2+2E		w	4072	4100	+28
$1+4bc$	E	⊥	w	ca. 4126	4128	+2
$2ab+2ab+4a$	A_1+E	∥	vw	4216[b]	4249	+33
$3bc+4a$	E	⊥	w	4313	4323	+10
$3a+3a$	A_1	∥	vw	4342	4400	+58
$2ab+3bc$	A_1+A_2+E		w	4474	4488	+14
$3a+4a+4a$	A_1		vvw	4783	4813	+30
$4a+4a+4bc+4bc$	A_1+E		vvw	4902	4928	+26
$2ab+4bc+4bc+4bc$	A_1+A_2+3E				4944	+42
$2ab+3a+4a$	E	⊥	vvw	4962	4978	+16
$2ab+2ab+3a$	A_1+E		vw	5105	5142	—
$1+3a$	A_1	∥	vw	5164	5170	—
$3a+3bc$	E		vw	5223	5217	—
$1+4bc+4bc$	A_1+E		vw	5257	5285	—
$3bc+4bc+4bc$	A_1+A_2+2E		vw	5311	5332	—
$2ab+4a+4a+4a$	E		vw	5367	5391	+26
$3a+3a+4bc$	E		vvw	5494	5558	—
$2ab+2ab+4a+4a$	A_1+E		vvw	5558	5555	—
$2ab+2ab+2ab+4bc$	A_1+A_2+3E				5571	—
$1+4a+4a$	A_1		vw	5585	5583	—
$1+2ab+4bc$	A_1+A_2+E				5599	—
$3bc+4a+4a$	E		vw	5626	5630	—
$3a+3a+4a$	A_1		vvw	5692	5706	+14
$1+1$	A_1		w	5762	5940	—
$2ab+3a+3a$	E		w	5860	5871	—
$1+3bc$	E		w	5983	5987	—
$3bc+3bc$	A_1+E		w	6024	6033	—

[a] Displaced by Fermi resonance with fundamental.
[b] CH_4?

reasonable agreement between calculated and observed frequencies. Also from a theoretical point of view this procedure is the more correct.

In general quite large negative anharmonicities are observed, e.g. the difference $\nu_{calc.}-\nu_{obs.}$ is positive. However, a few of the observed combination frequencies show small positive anharmonicities.

The prominent absorption at 6024 cm^{-1} has been interpreted as the first overtone of the carbon-hydrogen stretching frequency, $3bc+3bc$, the anharmonicity being -7 cm^{-1}. This anharmonicity is, however, much smaller than one should expect for the first overtone of a C–H stretching frequency, in which case an anharmonicity of the magnitude of -100 to -200 cm^{-1} would seem reasonable. A probable explanation is that the band, because of Fermi resonance, has been displaced towards higher wave-numbers. Another possibility would be that the band is the quaternary combination $3a+4a+4a+4a$, the calculated frequency being 6120 cm^{-1}. The observed intensity of the band, however, seems too high for a quaternary combination, although Fermi resonance may have increased it.

Generally, it must be emphasized that on account of the great anharmonicities and the possible effects of Fermi resonance and other interactions, the assignment of the bands to specific combinations, especially in the region 5000–6000 cm^{-1}, is only tentative.

Acknowledgments

I am greatly indebted to the late Professor A. Langseth for suggesting the present investigation, for many helpful discussions on the subject and for much good advice. My thanks are also due to Dr. Svend Brodersen for having recorded the spectra on the infrared spectrometer.

Note added in proof: A calculation of the unperturbed frequencies using first-order perturbation theory[18] and assuming that the intensity ratio $I(2\nu_{2\,ab})/I(\nu_1) = 0.5$ (see Fig. 1) gives $2\nu_{2\,ab} = 2930$ cm^{-1} and $\nu_1 = 2950$ cm^{-1}.

References

1. See G. Herzberg, Infrared and Raman Spectra of Polyatomic Molecules (D. van Nostrand Co., New York, 1945), p. 306 and 309.
2. D. R. J. Boyd and H. W. Thompson, Proc. Roy. Soc. **A 216**, 143 (1953).
3. Harry C. Allen, Jr., and Earle K. Plyler, J. of Research of the Nat. Bur. of Standards **63 A**, 145 (1959).
4. D. G. Rea and H. W. Thompson, Trans. Faraday Soc. **52**, 1304 (1956).
5. J. K. Wilmshurst and H. J. Bernstein, Canad. Journ. Chem. **35**, 226 (1957).
6. Llewellyn H. Jones, J. Mol. Spectroscopy **4**, 84 (1960).
7. E. H. Richardson, S. Brodersen, L. Krause, and H. L. Welsh, to be published (private communication from Dr. Brodersen).
8. Svend Brodersen and A. Langseth, Mat. Fys. Skr. Dan. Vid. Selsk. **1**, no. 1 (1956).
9. Svend Brodersen, J. Opt. Soc. Am. **43**, 877 (1953).
10. Janet Hawkins Meal and S. R. Polo, J. Chem. Phys. **24**, 1126 (1956).
11. G. Herzberg, *loc. cit.* p. 436.
12. G. Herzberg, *loc. cit.* p. 426.
13. F. Allan Andersen, Børge Bak, and Svend Brodersen, J. Chem. Phys. **24**, 989 (1956).
14. Llewellyn H. Jones and Robin S. McDowell, J. Mol. Spectroscopy **3**, 632 (1959).
15. I. M. Mills, Spectrochim. Acta **16**, 35 (1960).
16. W. H. Schaffer, J. Chem. Phys. **10**, 1 (1942).
17. Nathan Ginsburg and E. F. Barker, J. Chem. Phys. **3**, 668 (1935).
18. G. Herzberg, *loc. cit.* p. 216.

Indleveret til Selskabet den 31. marts 1962.
Færdig fra trykkeriet den 20. maj 1963.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 13

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 13 (1963)

THE REACTION $Ne^{20}(\alpha,C^{12})C^{12}$

BY

N. O. LASSEN AND JANUS STAUN OLSEN

København 1963
i kommission hos Ejnar Munksgaard

Synopsis

Natural neon was bombarded by α-particles with laboratory energies ranging from 12 to 20 MeV. The neon gas serving as target was used at the same time as filling gas for a gridded, electron collection type ionization chamber used to measure the emitted carbon ions.

The differential cross section at C.M. angle 90° was measured as a function of energy; the excitation curve shows several peaks (Fig. 10). The angular distributions were measured for each of these peaks (Fig. 12). One distribution follows a $[P_8(\cos\theta)]^2$ curve; this peak may correspond to a single $8+$ state in the compound Mg^{24} nucleus at 25.2 MeV excitation energy, having $\Gamma < 150$ keV. The other angular distributions also show strong maxima and minima, but at least two even angular momentum states must contribute to each peak. It is believed that a somewhat larger number of states participate in the reaction, statistical variations in this number being responsible for the appearance of the excitation curve.

Printed in Denmark
Bianco Lunos Bogtrykkeri A/S

1. Introduction

At the International Conference on Nuclear Structure, Kingston, Canada, in the autumn of 1960, results obtained by ALMQVIST, BROMLEY and KÜEHNER[1] of the elastic scattering of carbon ions and of the reaction $^{12}_{6}C\,(^{12}_{6}C,\,\alpha)\,^{20}_{10}Ne$ were reported. These authors found peculiar resonances in the elastic C–C process for ion energies above the Coulomb barrier and in some reaction processes for energies just below the barrier. Therefore, it might be expected that also the reaction

$$^{20}_{10}Ne + {}^{4}_{2}He \rightarrow {}^{12}_{6}C + {}^{12}_{6}C \quad (-4.6\ \text{MeV}) \tag{1}$$

would show interesting features, and it was decided to make a study of this process by means of the α-beam from the Copenhagen cyclotron.

Since the energy of our α-particles is 20 MeV, carbon ions with an energy of about 6 MeV in the C.M. system are produced in reaction (1). An apparatus was constructed by means of which these fission carbon ions could be detected[2]. The principle of the method was to use an ionization chamber filled with neon to such a pressure that the short range heavily ionizing fission ions were stopped inside the chamber, giving pulses corresponding to their full energy, whereas α-particles and other lighter ions coming from the target spent only a small part of their range in the chamber and thus gave only small pulses. At the same time the neon gas acted as the target, the α-beam being passed across the ionization chamber inside a tube with small side holes.

In three separate experiments with different tubes the number of fission ions at the lab. angles 90°, 62° and 58° were counted. Search for other modes of fission than reaction (1) was made, but none were seen. A search for the reaction $^{28}_{14}Si\,(\alpha,\,0^{16})\,0^{16}$ was also made, using SiH_4 in the chamber[3], but neither was this process seen.

The results for reaction (1) indicated a strong angular variation of the cross section. To measure the angular distribution in more detail an apparatus based on the same principle, but having a different geometry, was

built; it will be described in the present paper. Preliminary results have been briefly reported[4].

It was soon found that the cross section was also varying strongly with the energy of the α-particles. Therefore it was necessary to construct a device for measuring the α-energy continuously during the experiments in order to ensure that it remained constant. With this energy monitor in use,

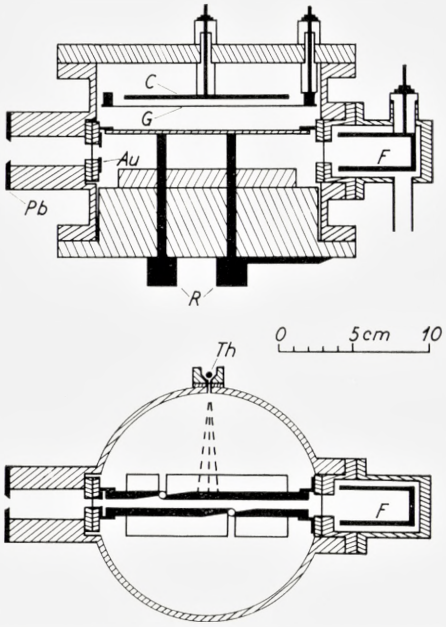

Fig. 1. The neon chamber. The α-beam is limited by the lead stop Pb and the (defining) gold stop Au, and it is measured by the evacuated Faraday cup F. G is a Frisch grid, C collector electrodes, and Th a source of natural α-particles used for calibration. R are two rotatable brass rods carrying the channels which define the emission angles of the carbon ions. Vertical (upper) and horizontal (lower) sections are shown.

an excitation curve was obtained with a somewhat different geometry for C^{12} detection, in which particles emitted in a wide angular range, approximately 20°–160°, were registered. The excitation curve showed marked resonances[5]. However, the wide angular range involved a rather thick target; furthermore, by the geometry the different angles had different detection efficiencies. Therefore, it was realized that a better way to find an excitation function might be to measure the differential cross section for $\theta = 90°$ as a function of energy.

The present paper deals with such measurements as well as with measurements of the angular distribution for a number of resonances.

2. Experimental apparatus

The carbon ions were detected by means of a gridded ionization chamber. Part of the apparatus is shown in Fig. 1. The ionization chamber is housed in a steel tube with 150 mm inside diameter. The α-beam traverses the tube along a diameter; it enters through a side tube carrying a lead plate with a hole, 6×11 mm^2. A nickel foil, 1 mg/cm^2 thick, separates the chamber

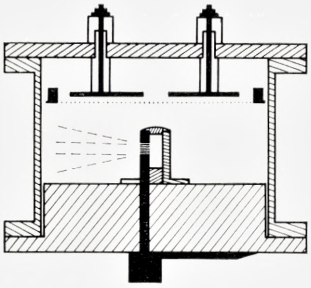

Fig. 2. Section through the neon chamber perpendicular to the beam.

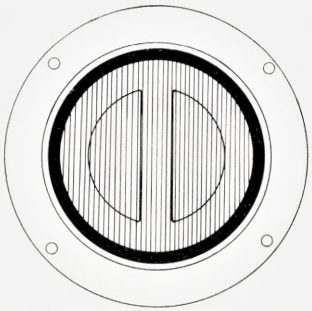

Fig. 3. View of the electrode system.

from the cyclotron vacuum; the window is 6×11 mm^2, but just behind it is placed a gold diaphragm which reduces the beam cross section to 5×10 mm^2. Right opposite the entrance tube another side tube housing a Faraday cup is placed; it is connected to the cyclotron vacuum through a rubber tube, and is separated from the ionization chamber by means of a tantalum foil of thickness ~ 20 mg/cm^2; this window is circular and 15 mm in diameter. Inside the chamber the α-beam passes through a duct so that, from the active volume of the ionization chamber outside this duct, the beam can be seen only through some narrow channels in the side walls.

Fig. 2 shows a vertical cross section perpendicular to the beam direction, and Fig. 3 shows the electrodes of the ionization chamber viewed from below. The grid consists of parallel 0.15 mm platinum wires, spaced 3 mm, and held by a brass ring supported by three teflon insulators (one of them shown in Fig. 1). There are two collector electrodes, each connected to its own preamplifier; thus, there are actually two ionization chambers.

The duct is made of four pieces of dural screwed to the bottom plate and placed very accurately; on top is a lid. At the two ends special shields

are placed, and the only openings left are the actual emission channels in two circular brass rods placed in the side walls of the duct in the way shown in Fig. 1 and, in more detail, in Fig. 4. The rods can be rotated from outside and their position can be read by means of a pointer and a scale; the vacuum seal is provided by o-rings. The rods are 5.0 mm in diameter; the upper end of each rod is made of two half circular pieces screwed to-

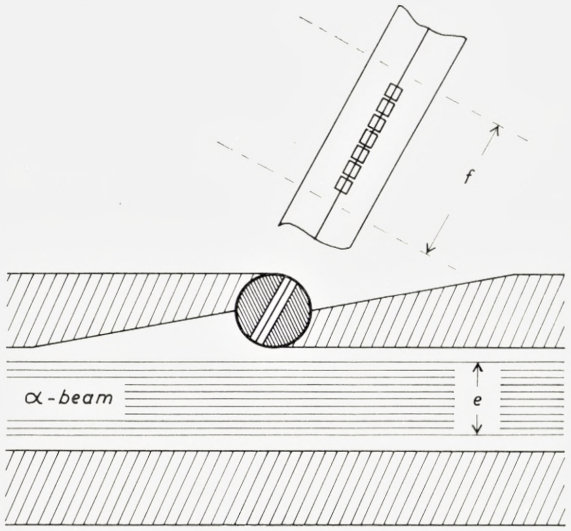

Fig. 4. Details of a rod with channels. e and f denote the horizontal and vertical extensions of the α-beam. $e = 0.5$ cm. $f = 1.0$ cm. A cross section and a side view of a rod are shown.

gether, and between them is a 0.1 mm bronze foil separating the two rows of channels cut in the pieces. There are 14 channels in each rod, each channel 0.5×1.0 mm^2 in cross section. To obtain a higher angular resolution some measurements of the angular distribution were made by using rods with narrower channels; each of these rods had 28 channels (in four rows) of cross section 0.2×1.0 mm^2.

The chamber was filled with pure neon. In our earlier experiments, a few percent of methane had been added, but to avoid trouble from recoil carbon ions, pure neon was used in the present experiment. The chamber was found to work equally well without CH_4 and with much lower voltages. Mostly, neon pressures of the order of 200 mm of Hg were used, and the voltage on the collector electrode was 100 V; at very low neon pressures

(< 50 mm) the voltage had to be lower. The gas was continuously purified by circulation over hot calcium in a side tube.

Via a standard amplifier the chamber was connected to a 100-channel pulse height analyzer. Calibration was made by ThC + C' α-particles crossing the chamber in a well collimated beam. The particles were stopped by hitting the duct; they dissipated an energy of the order of 1 MeV in the chamber. The energy resolution (full width at half maximum height) was about 8 %, approximately what should be expected from straggling. For

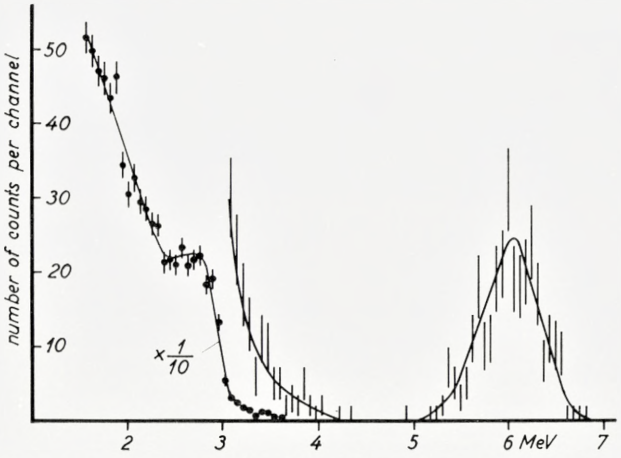

Fig. 5. Pulse height distribution showing a peak at 6 MeV corresponding to C^{12} ions. The spectrum was obtained with 0.5 μ Coulomb of 19.5 MeV α-particles and $\vartheta_{Lab} = 62°$. The neon pressure was 300 mm of Hg.

emission angles in the range $20° < \vartheta < 80°$ the fission ions gave a peak in the energy distribution curve well separated from the background; an example is seen in Fig. 5.

The energy of the α-particles was changed by means of absorbers, and it was measured by another gridded ionization chamber connected to another 100-channel pulse height analyzer. The analyzed[6] beam from the cyclotron was passed through a lead diaphragm with a hole, 22 mm in diameter (Fig. 6). Behind it three foil holders were placed in a slide arrangement; each could be set in four different positions and, in this way, different absorber thicknesses could be introduced in the beam. The absorber foils were 25 mm in diameter. Some were of nickel, the thicknesses being multiples of 0.5 mg/cm²; others were of beryllium, in multiples of 5 mg/cm².

Also Al foils were used. To cover the energy range 12–20 MeV several sets of foil holders were used. Behind the foils a second, vertically adjustable, lead diaphragm reduced the beam cross section to 5×10 mm^2, and behind it the beam passed through a 0.16 mg/cm^2 gold foil, placed at an angle of 45°. The elastically scattered α-particles from this foil were registered by the α-ionization chamber; it was similar in construction to a chamber earlier described[8]. It was filled with argon to about three atmospheres with

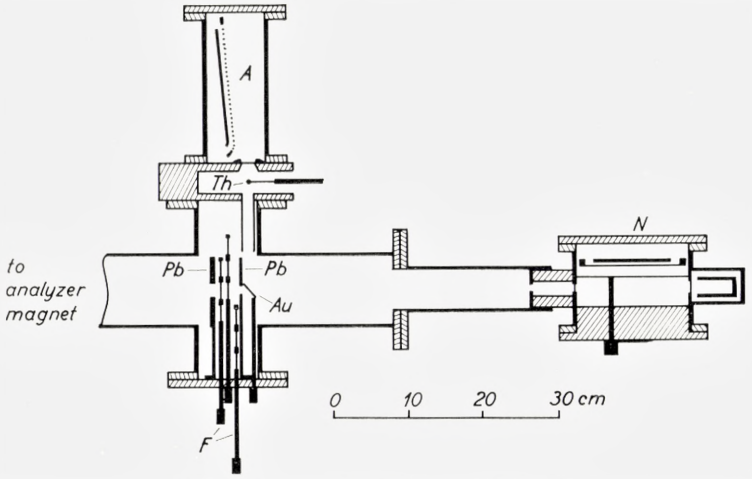

Fig. 6. The experimental arrangement.
N: Neon chamber. A: Ionization chamber for measurement of energy of scattered α-particles. Pb: Lead stops. F: Movable holders with absorber foils. Au: Gold scattering foil. Th: Removable source of natural α-particles.

an admixture of a few percent methane; the gas was continuously purified by circulation over hot calcium. The resolution for ThC α-particles was 2 %, or 160 keV. For some unknown reason, maybe because the width of the chamber was insufficient, the resolution was slightly inferior for the scattered α-particles, about 250–300 keV. Other measurements by solid-state counters have indicated that the energy spread of the undegraded beam was about 70 keV.

The thinnest Ni foil reduced the α-energy by 70–100 keV. Smaller changes were sometimes obtained by deliberately changing the Dee positions or other cyclotron parameters. The energy of the α-particles at the target could also be varied by changing the pressure in the neon chamber and thus the absorption in the neon gas in front of the target volume.

3. Evaluation of the differential cross sections from the measured numbers

Figs. 7 and 8 show the geometry. We first want to find the number of carbon ions escaping through one channel. Suppose the channels are set at an angle ϑ_0. We take the horizontal plane through the channel as xy-plane (Fig. 7). Consider such particles in this plane which move in directions having angles in the interval $\vartheta_o + \Delta\vartheta < \vartheta < \vartheta_o + \Delta\vartheta + d\vartheta$ (here, for the

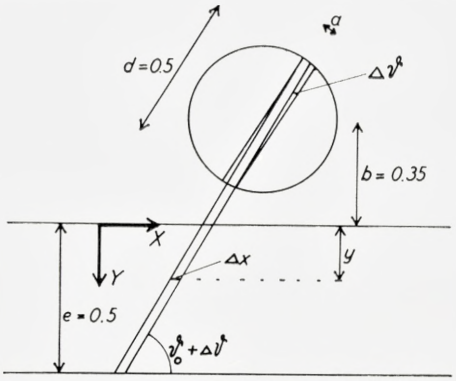

Fig. 7. Horizontal section through a channel.

present considerations $\Delta\vartheta$ is a fixed angle). If they are to escape through the channel, they must start within a region of the target having the thickness

$$\Delta x = \frac{a - |\Delta\vartheta \cdot d|}{\sin(\vartheta_o + \Delta\vartheta)}, \tag{2}$$

where a and d are the width and the length of the channel.

If they start from the depth y they have to be emitted within a solid angle

$$d\omega = \frac{r d\vartheta \cdot h}{r^2} \tag{3}$$

where h is the height of the channel and where

$$r = \frac{d}{2} + \frac{b + y}{\sin(\vartheta_o + \Delta\vartheta)}, \tag{4}$$

b having the meaning shown in Fig. 7.

However, it is necessary to take into account also the z-dimension. Consider a line element of target parallel to the x-axis and having coordinates y and z. If $|z|$ is not too high, particles may escape with a scattering angle $\vartheta_o + \Delta\vartheta$; to a first approximation the target thickness corresponding to such particles is again given by (2), the scattering angle and its projection on the xy-plane not differing much from each other.

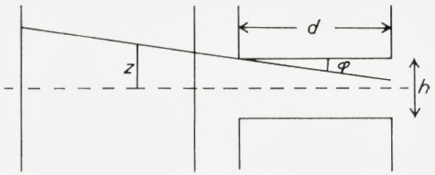

Fig. 8. Vertical section through a channel.

For $|z| > \frac{1}{2} h$, the solid angle is (Fig. 8)

$$d\omega = \frac{r d\vartheta (h - |\varphi d|)}{r^2}, \tag{5}$$

where

$$r = \left[\frac{d}{2} + \frac{b+y}{\sin(\vartheta_o + \Delta\vartheta)}\right] \cdot \frac{1}{\cos\varphi} \approx \frac{d}{2} + \frac{b+y}{\sin(\vartheta_o + \Delta\vartheta)} \tag{6}$$

$$\varphi \approx \operatorname{tg}\varphi = \frac{|z| - \frac{1}{2} h}{\frac{b+y}{\sin(\vartheta_o + \Delta\vartheta)} - \frac{d}{2}} < \frac{h}{d} = \frac{1}{5}. \tag{7}$$

For $|z| < \frac{1}{2} h$ the solid angle is given by (3).

For the angle $\vartheta_o + \Delta\vartheta$ the yield from the line element is

$$d^3 Y = \left(\frac{d\sigma}{d\omega}\right)_{\vartheta = \vartheta_o + \Delta\vartheta} \cdot 2 \cdot \frac{n\, dy\, dz}{ef} \cdot N \Delta x \cdot d\omega, \tag{8}$$

where N is the number of Ne^{20} atoms per cm^3 of target gas, n is the number of α-particles having passed the total beam cross section of breadth e (= 0.5 cm) and height f (= 1 cm); Δx is given by (2), $d\omega$ by (3) or (5), and the factor 2 enters because two identical particles are created in the fission process.

Inserting $a = 0.05$, $h = 0.1$ and $d = 0.5$, one finds

$$d^2Y = 2\int_{z=0}^{z=\frac{1}{2}h} d^3Y + 2\int_{z=\frac{1}{2}h}^{z=z_{\max}} d^3Y = 0.001 \cdot k_1 \cdot \frac{1 - 10 \cdot |\Delta\vartheta|}{\sin(\vartheta_o + \Delta\vartheta)}\, d\vartheta\, dy\,, \tag{9}$$

where $z_{\max} = \varphi_{\max} \dfrac{b+y}{\sin(\vartheta_o + \Delta\vartheta)} = \dfrac{1}{5} \dfrac{b+y}{\sin(\vartheta_o + \Delta\vartheta)}$, and $k_1 = \dfrac{2\,nN}{ef} \dfrac{d\sigma}{d\omega}$,

the cross section $\dfrac{d\sigma}{d\omega}$ being assumed to be constant for all possible values of $\Delta\vartheta$ corresponding to a fixed ϑ_o. By further integrating between the limits $y = 0$ and $y = 0.5$, one gets

$$dY = 0.0005\ k_1 \frac{1 - 10 \cdot |\Delta\vartheta|}{\sin(\vartheta_o + \Delta\vartheta)}\, d\vartheta\,. \tag{10}$$

Fig. 9 shows, for $\vartheta_0 = 30°$, this value dY as a function of $\vartheta = \vartheta_0 + \Delta\vartheta$. The distribution is almost a triangle. If we are interested merely in the total area, we may just as well replace the denominator $\sin(\vartheta_0 + \Delta\vartheta)$ by simply the constant value $\sin\vartheta_0$. Integration of (10) then gives

$$Y(\vartheta_o) = 2\int_o^{\Delta\vartheta_{\max}} dY = 0.00005 \cdot \frac{k_1}{\sin\vartheta_o} = \frac{1}{\sin\vartheta_o} \cdot 0.0002\, nN \cdot \frac{d\sigma}{d\omega}\,.$$

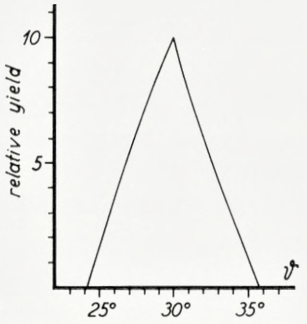

Fig. 9. Approximate response curve for channels set at $\vartheta_0 = 30°$. A correction to the curve is given in the text.

Taking into account the number of channels one finds for the number of carbon ions

$$n_c = 28 \cdot 10^{-4} \cdot \frac{1}{\sin\vartheta_o} \cdot nN \frac{d\sigma}{d\omega}\,.$$

From this expression the differential cross section in the lab. system is found. The C.M. cross section is found in a well-known way[7].

The distribution given by (10) and shown in Fig. 9 is not correct. If we want to consider the actual distribution we cannot use the approximation in which the angle and its projection on the xy-plane are put equal. If ϑ is the angle in the xy-plane, φ the inclination and ϑ' the actual angle, then $\cos \vartheta' = \cos \vartheta \cos \varphi$. For $\vartheta = 30^\circ$ and $|\varphi| = \varphi_{\max} = \frac{1}{5}$ this gives $\vartheta' = 32^\circ$. In this example, the center of gravity of the distribution is displaced $0^\circ.7$, so instead of being $29^\circ.8$ as in Fig. 9, it is about $30^\circ.5$. Thus for small angles, and especially when the higher angular resolution is used, a correction to the angle is needed (note: $\vartheta = 30^\circ$ corresponds to $\Theta_{\mathrm{Cm}} \sim 45^\circ$).

4. Results and discussion

By setting the channels at the proper angle, slightly dependent on α-energy, and counting the number of carbon ions with various foils interposed in the beam, the excitation curve giving $\frac{d\sigma}{d\omega}$ for $\Theta_{\mathrm{Cm}} = 90^\circ$ was obtained (Fig. 10). Like the earlier curve[5] it shows a number of peaks, the positions of which are, as a rule but not always, the same in the two curves; the peaks at 12.6, 13.6, 16.0 and 16.3 from the earlier curve occur in Fig. 10 at 12.7, 13.4, 15.9 and 16.2. However, the two curves are not identical and they should not be so, because they deal with different quantities. The earlier curve was obtained by counting ions in a wide angular range, and it thus gives the variation of the "total" cross section, whereas the present curve refers to the differential cross section at a certain angle. Since the angular distribution is not isotropic, and since it is not the same for different energies, the two quantities must vary with energy in different ways. In fact, it is surprising that the two curves are so much alike, and it may indicate that, although the angular distribution varies with energy, this variation may be rather smooth.

One demonstration of the difference is the fact that we were unable to find for $\Theta = 90^\circ$ any peak corresponding to the pronounced maximum in the "total" yield curve at 14.3 MeV. The observation that the excitation function for $\Theta \sim 45^\circ$ has a sharp peak at this energy is noteworthy, and it is in accordance with the angular distribution (Fig. 12).

Another demonstration may be the smaller widths of the peaks in Fig. 10 compared to the earlier curve; however, the earlier geometry involved a

somewhat larger effective target thickness, which is responsible for part of the increase in width. The difference between the two curves with respect to relative heights of the various peaks is more significant.

As regards the reliability of the curve in Fig. 10, we think that it is substantially correct for energies above 12 MeV. For lower energies the measurements become difficult, because the thick degrader foils and the ensuing rather large compound scattering results in a considerable reduction of the

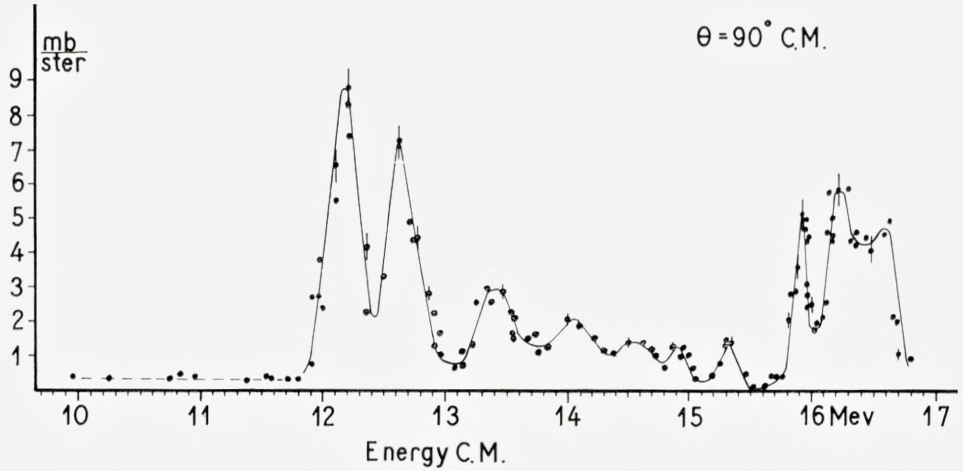

Fig. 10. Differential cross section for the Ne^{20} (α, C^{12})C^{12} reaction at C.M. angle 90°.

part of the beam which enters the neon chamber through the various stops. Since furthermore the fission cross section is rather small below 12 MeV, only few pulses were recorded. Additional troubles arise because, at these low energies of the incoming α-particles, the pulse height spectrum from the neon ionization chamber does not show a carbon peak which is well separated from the background of other pulses. This background has a tail extending to energies higher than the carbon energies, probably caused partly by processes in the chamber initiated by neutrons from the beryllium absorber foils. For these reasons, some ambiguity is involved in the estimation of the fission cross section below 12 MeV. For the inverse process the Chalk River group[8] has, for energies corresponding in our scale to 10–11 MeV, found an almost constant cross section of about 0.1 mb/ster. The present estimations give 0.3 mb/ster.

A peculiar feature of the curve is the non-existence of resonances below 12 MeV and the rise in the fission cross section at this energy. Part of the

explanation may be that at lower energies the low penetrability through the Coulomb barrier between the two carbon nuclei prevents fission from competing successfully with other modes of decay of the compound nucleus. In this connection, a comparison with the C–C elastic scattering curve measured at Chalk River[1) 8)] is interesting (see Fig. 11). The prominent peak that we find at 12.1 MeV is just barely seen in the elastic scattering, but, apart from this, there is a very close agreement between the resonance values found in the two experiments.

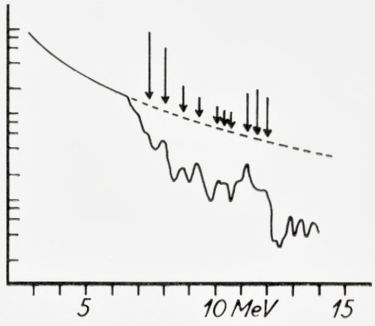

Fig. 11 The curve shows the cross section for elastic C^{12}–C^{12} scattering at C.M. angle 90° as a function of C.M. carbon energy, as obtained by the Chalk River group[1]). Arrows indicate positions and amplitudes of our peaks for the $Ne^{20}(\alpha, C^{12})C^{12}$ reaction.

In Fig. 10 the narrowest peak is the one at 15.9 MeV; it has a half width of 145 keV or, when the α-energy is measured in the lab. system, 175 keV. This is presumably somewhat more than the energy spread of the beam having passed the degrader foils, the window and the neon gas in front of the target volume. Also the two peaks between 12 and 13 MeV show half widths ($\sim$ 300 keV C.M.) somewhat larger than the estimated straggling in the beam. Comparing the cross sections from Fig. 10 with the elastic C–C cross sections, one can obtain rough estimates for the ratio Γ_c/Γ_α of the partial widths for emission of carbon nuclei and of α-particles; for the 15.9 MeV resonance one finds $\Gamma_c \sim 6\,\Gamma_\alpha$.

Fig. 12 shows the angular distributions for the various peaks in Fig. 10 and for one energy (14.3 MeV) at which the "total" cross section (the earlier curve) has a peak, but Fig. 10 has none. When obtaining these distributions, the energy of the incoming α-particles was measured simultaneously with the measurement of each point, and the mean energy was constant to better than $\pm$ 30 keV. When the angle is changed by rotating the rod, the position

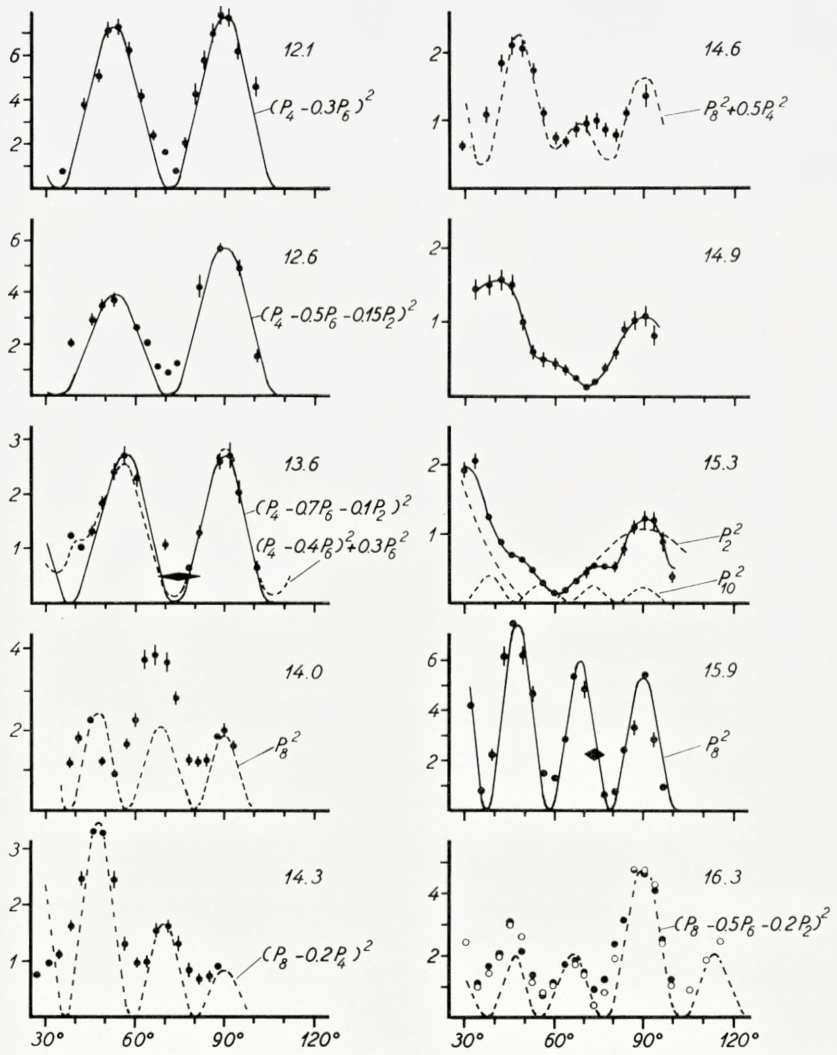

Fig. 12. Angular distributions of carbon ions from the reaction $Ne^{20}(\alpha, C^{12})C^{12}$ at various C.M. energies (entrance channel) written in the figure. For one point in the 13.6 MeV distribution the approximate angular response function for the channels is indicated. One distribution, for 15.9 MeV, was measured with a higher angular resolution, again indicated for one point.

of the target volume is shifted, and therefore the mean energy of the reacting α-particles is slightly altered. However, this effect was compensated by varying the pressure in the neon chamber.

Since the reaction particles, in both the entrance and exit channels, are spinless and all have positive parity, and since two identical particles are emitted, the compound states which may be involved must have even parity and even angular momentum. If the reaction goes through a single level in the compound nucleus, the angular distribution will follow a $[P_l(\cos\Theta)]^2$ curve. This is considered to be true for the α-energy 15.9 MeV, and it is inferred that this peak corresponds to a $8+$ level in the compound Mg^{24} nucleus at excitation energy 25.2 MeV and with a total width smaller than 150 keV.

The situation is not so clear for any of the other peaks, and it may be concluded that each peak covers more than one level. If two levels contribute to a peak, the angular distribution will be described by a combination of two even Legendre polynomials. If the levels do not overlap, but lie so close together that the spread in the α-energy prevents their resolution, the angular distribution will be given by

$$\frac{d\sigma}{d\Omega} \sim (P_{l_1})^2 + a_2(P_{l_2})^2, \tag{11}$$

a_2 being real and positive, l_1 and l_2 even. If the levels overlap, the distribution will be

$$\frac{d\sigma}{d\Omega} \sim (\alpha_1 P_{l_1} + \alpha_2 P_{l_2})^2. \tag{12}$$

If in formula (12) the coefficients are real, the distribution will have zero points. In other cases there are generally no zero points. Of course, the limited angular resolution tends to fill out the valleys; however, in some cases the experimental points approach the axis of abscissae, whereas for other α-energies the valleys are not so deep.

The angular distributions in Fig. 12 may be divided into four groups. The three first distributions for 12.1, 12.6 and 13.6 MeV have two maxima between 30° and 100°. The second group, for energies 14.0, 14.3 and 14.6 MeV, have three maxima and not very deep valleys. The third group, 14.9 and 15.3 MeV, have three or four maxima, one deep minimum, the other minima not being so deep. The fourth group, 15.9 and 16.3 MeV, resembles the second by having three maxima, but deviates by having two deep minima in between.

For 12.1 MeV the cross section is very low for 35° and 75°. We conclude that the distribution is of type (12) with real coefficients, corresponding to (at least) two levels so broad that they overlap. In fact,

the curve $(P_4 - 0.3\,P_6)^2$ fits nicely the experimental points, indicating levels with $l = 4$ and $l = 6$.

The 12.6 and 13.6 MeV peaks have angular distributions resembling that of the 12.1 MeV peak. This points to levels with the same angular momenta, $l = 4$ and $l = 6$. For the 13.6 MeV peak the curve corresponding to (12) with $\alpha_4 = 1$ and $\alpha_6 = -0.4 + 0.55\,i$ fits the experimental points, if the 70° point is disregarded. However, the distribution may be equally well fitted with $(P_4 - 0.7\ P_6 - 0.1\ P_2)^2$. Therefore, it is not possible to conclude that only two levels contribute, there might also be a level with $l = 2$. For similar reasons, although the curve $(P_4 - 0.5\ P_6 - 0.15\ P_2)^2$ fits the 12.6 distribution, it cannot be concluded that three levels contribute to the peak.

The second group of angular distributions, for the energies 14.0, 14.3 and 14.6 MeV, cannot be fitted with combinations of P_4 and P_6, but a P_8 must be involved. Since the distributions have no zero points, attempts were made to fit them with curves of type (11), but with little success. For the 14.6 MeV distribution a curve $P_8^2 + 0.5\,P_4^2$ reproduces roughly the three maxima between 30° and 100°, but it is not considered to give a satisfactory fit to the points.

Attempts to fit the 14.9 and 15.3 MeV distributions by curves of type (11) or type (12) by using real coefficients were also unsuccessful. The 15.3 MeV distribution seems to indicate a contribution from angular momentum of at least 10, but due to the small cross section the uncertainties are rather large, and a too detailed analysis would not seem justified.

The last group for energies 15.9 and 16.3 MeV has distributions dominated by $l = 8$. It may be mentioned that, although the 15.9 MeV peak shows a pure P_8^2 distribution, it may nevertheless contain more than one level, if only all significantly contributing levels have $l = 8$. The 16.3 MeV peak must contain a level with $l = 8$, and besides that at least one more level with lower l, most probably $l = 6$. The asymmetry of the central (90°) peak for this energy is a most peculiar effect. If the reaction we observe is described by expression (1), no other particles or photons being involved, the distribution must be symmetric around 90°. For all other energies (except maybe 12.1 MeV, see Fig. 12) the distributions never showed any significant departure from symmetry, but for 16.3 MeV the asymmetry was observed in four different measurements using three different rods and pointers.

One may try to understand the excitation curve in different ways:

1°. One may regard the peaks of the curve as being caused by individual levels in the compound nucleus. With the exception of the 15.9 MeV peak, each peak must then cover at least two levels, as discussed above.

2°. One may assume a somewhat larger number of participating levels and regard the peaks as being caused by statistical fluctuations in the level density. In some cases a peak may then correspond to one or a few strongly excited levels (for instance 15.9 MeV), but in other cases to many relatively weakly excited levels. Even in the latter cases the angular distributions may be expected to be of the type found; in expression (12) α_1 may stand for the sum of the coefficients of any number of levels with angular momentum l_1. Considering the influence of the centrifugal barrier in the C–C system one would, if many levels are involved in each peak, expect a gradual change of the angular distributions; when the energy is increased, more levels with higher l values will contribute to the distribution. This is just what has been observed. The 12.1 MeV peak is nearly described by P_4^2, but a small admixture of P_6 is needed to give an exact fit to the positions of the maxima. For 12.6 and 13.6 MeV gradually larger admixtures of P_6 are appropriate. For energies above 14 MeV levels with $l = 8$ begin to play a role.* This result, especially for the energy region 12–14 MeV, seems to be strongly in favour of the many-level hypothesis. Other features of the curve in Fig. 10 support this way of looking at things. If each peak corresponded to only a few levels one would expect much stronger variations in the relative peak heights. Also, if the levels were so scant that each peak covered only two or three, one would expect more cases of single levels, i. e. pure P_l^2 distributions. However, the rather deep minima in the excitation curve show that the density of participating levels cannot be very high. In this connection, it may be remarked that repeated careful investigations have shown that the cross section at 15.6 MeV is less than 2 % of the cross section at 15.9 MeV. From considerations of the kind intimated above, the average spacing of participating levels may be roughly estimated to about 40 keV.

3°. A third point of view may be mentioned. One might assume a very large number of weakly excited levels giving a general, low "background";

* The Chalk River group has measured[8]) some angular distributions for the ground state α-particles from the inverse reaction in an energy range corresponding to 10–11 MeV in our scale. The distributions indicate rather low angular momenta and seem to vary rather smoothly with energy.

the peaks on top of it correspond to individual, strongly excited levels. The very small cross section at 15.6 MeV and the small variation in peak heights are in disfavour of this point of view.

These experiments were carried out at the Institute for Theoretical Physics, University of Copenhagen, and the authors wish to express their deep gratitude to the Director of the Institute, the late Professor NIELS BOHR. In conclusion we thank mag. sc. N. O. ROY POULSEN for valuable discussions, Ing. PH. DAM and Ing. A. HEDEGAARD for technical assistance.

Institute for Theoretical Physics,
University of Copenhagen, Denmark

References

1. D. A. Bromley, J. A. Kuehner and E. Almquist: Proc. Int. Conf. Nucl. Structure, Kingston, Canada. (Univ. Toronto Press and North-Holland Publishing Co. 1960), p. 255.
   E. Almquist, D. A. Bromley and J. A. Kuehner: ibid. p. 258.
   J. A. Kuehner, B. Whalen, E. Almquist and D. A. Bromley: ibid. p. 261. See also Phys. Rev. Lett. **4**, 365 and 515, 1960, and Proc. Sec. Conf. Reactions between Complex Nuclei, Gatlinburgh, 1961, paper C-4, p. 151 and paper E-5, p. 282.
2. N. O. Lassen: Nucl. Phys. **38**, 442, 1962.
3. N. O. Lassen and Gunnar Sørensen: Nucl. Phys. **38**, 450, 1962.
4. N. O. Lassen: Phys. Lett. **1**, 65, 1962.
5. N. O. Lassen: Phys. Lett. **1**, 161, 1962.
6. H. W. Fulbright, N. O. Lassen and N. O. Roy Poulsen: Mat. Fys. Medd. Dan. Vid. Selsk. **31**, 10, 1959.
7. R. D. Evans: The Atomic Nucleus, McGraw-Hill, N. Y. 1955, p. 421.
8. H. E. Gove: Nucl. Instr. and Meth. **11**, 63, 1961.

Indleveret til Selskabet den 14. januar 1963.
Færdig fra trykkeriet den 6. maj 1963.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 14

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 14 (1963)

RANGE CONCEPTS AND HEAVY ION RANGES

(NOTES ON ATOMIC COLLISIONS, II)

BY

J. LINDHARD, M. SCHARFF(†) AND H. E. SCHIØTT

København 1963
i kommission hos Ejnar Munksgaard

CONTENTS

Synopsis

A theoretical discussion is given of the range of heavy ions with moderate velocity. The treatment is based on the theory of quasi-elastic collisions given elsewhere. The region where electronic and nuclear stopping compete is of particular interest. Use is made of a simple velocity proportional Thomas-Fermi type formula for electronic stopping, and a universal approximate differential cross section for scattering. Simplified models of nuclear scattering assuming power law scattering are also included. They turn out to be useful for exploratory computations of various range quantities.

The straightforward theory of ranges is studied in § 2. Range curves are computed for any atomic numbers of particle Z_1, and substance Z_2. It is found that when nuclear stopping is dominating, a $\varrho - \varepsilon$ plot gives a universal range energy description.

Probability distribution in total range and various averages are studied (§ 3), in order to assess corrections to measurements when necessary. Similarly, corrections to measurements of projected ranges are obtained (§ 4). The range correction due to nuclear stopping is obtained for ions of high initial energy.

In § 5 a survey is given of numerous recent measurements of range. They are found to be in fair accord with theoretical results, for energies between 100 MeV (fission fragments) and ~ 1 keV.

Printed in Denmark
Bianco Lunos Bogtrykkeri A/S

§ 1. Introduction

The present paper is a theoretical study of ranges of heavy ions of low velocity, and their connection to the basic problem of quasi-elastic collisions between ions and atoms. Three characteristic features give rise to complications. First, both electronic and nuclear stopping must be studied thoroughly, because they are similar in magnitude. Second, because of the frequent large deflections of the ions one must distinguish carefully between various range concepts. Third, the variety of choice of atomic number of both ion and substance gives an additional difficulty. We shall try to show that our present knowledge of quasi-elastic collisions, in spite of the above complications, can give us a simple and fairly accurate range theory. In point of fact, in the following we use a much simplified description of quasi-elastic collisions, which could be improved upon without difficulty. Aspects of quasi-elastic collisions are studied also in three associated papers: Notes on Atomic Collisions I, III, and IV. The aim is to exploit similarity properties of Thomas-Fermi type in collisions between heavy ions and atoms. In fact, similarity enables us to treat in a comprehensive way both slowing-down and damage effects by heavy ions.

The total range of a swift particle may be observed in track detectors like photographic emulsions. The observation of many tracks can then give the probability distribution in total range. In measurements of this kind the observed range depends on energy losses only, and not on scattering of the particle. For energetic heavy particles this separation of energy loss from scattering is especially valuable, since the two are due to unconnected processes, i. e. respectively electron excitation and Coulomb scattering by the atomic nuclei.

However, in nearly all other cases one observes somewhat different and less well-defined types of ranges. It is then customary to make corrections for multiple scattering in order to obtain the total range, but since these corrections are not insignificant—even in cases like high energy protons where deflections are small—it would seem appropriate to introduce explicitly these other types of ranges.

The scattering of a particle—in contrast to its energy loss—is always dominated by nuclear collisions, i. e. deflections in the screened electric field of the atom. In the case of electrons, large scattering angles are quite common during slowing-down. For heavy particles of high energy (e. g. protons with MeV-energies), scattering effects are relatively small, but since a high precision is desirable here, the distinction between different types of ranges again becomes important. Although the description in the following could be applied to electrons and to fast heavy particles, we shall aim at the case mentioned in the beginning of the introduction. In fact, for heavy ions of low velocity, e. g. $v \sim v_0 = e^2/\hbar$, scattering effects are large and the scattering can not be completely separated from energy loss, simply because the nuclear collisions here begin to dominate the energy loss too. This somewhat complicated case will be used as a basic example in our general discussion of range concepts.

The following discussion does not at all pretend to give an exhaustive treatment of range concepts. Thus, we are throughout concerned with stopping by a random system of atoms, i. e. uncorrelated atoms and separated collisions. This might never seem to include stopping of a relatively slow heavy ion in a solid, where the interatomic distance is short and atoms are arranged in a periodic lattice. Still, the effects are only sometimes large; they are not well understood and appear to be dependent on the structure of the lattice (cf. § 5).

Before turning to the various—and often complicated—range concepts and range distributions, we may take a more straightforward point of view. In § 2 we proceed as if the energy loss along the path was a nearly continuous process. This is not at all a poor first approximation. It both enables us to get a clearer picture of the essential points and permits comparisons with experiments (cf. § 5).

§ 2. Simple Unified Range Theory

Suppose that the range along the path is a well-defined quantity, so that we need not distinguish between e. g. average range, most probable range, and median range. We may introduce first the simple concept of specific energy loss, (dE/dR),—or average energy loss per unit path length—defined by

$$\frac{dE}{dR} = N \cdot S = N \int d\sigma \, T, \tag{2.1}$$

where N is the number of scattering centres (e. g. atoms) per unit volume and S the stopping cross section per scattering centre. Further, $d\sigma$ is the differential cross section for an energy transfer T to atoms and atomic electrons.

The basic range concept is then obtained simply by integration of (dE/dR),

$$R(E) = \int_0^E \frac{dE'}{(dE'/dR)} = \frac{1}{N}\int_0^E \frac{dE'}{S(E')}. \tag{2.2}$$

The formulations (2.1) and (2.2) give a simple connection between range, specific energy loss, and differential cross section. We do not at present distinguish between different types of ranges. A better understanding of the connection between (2.2) and e. g. the average range is obtained in the detailed discussions in § 3.

In an analogous way we may introduce the range straggling (cf. Bohr (1948)). Similarly to (2.1) the average square fluctuation in energy loss becomes

$$\overline{(\Delta E)^2} = N\Omega^2\, dR = NdR\int d\sigma\, T^2, \tag{2.3}$$

if the individual events have average occurrence $NdRd\sigma$, and are uncorrelated. We may next derive the average square fluctuation in range, $(\Delta R)^2$, using the present assumption that fluctuations are small,

$$(\Delta R)^2 = \int_0^E \frac{dE' N\Omega^2(E')}{(dE'/dR)^3} = \frac{1}{N^2}\int_0^E \frac{dE'\cdot\Omega^2(E')}{S^3(E')}. \tag{2.4}$$

If we were precise, we would say that the interpretation of (2.4) as the average square fluctuation in range is not quite correct. For the present purposes, however, we have by means of (2.2) and (2.4) defined the range, R, and its fluctuation, ΔR, and the results are sufficiently accurate for most purposes. We now use (2.2) and (2.4) in a first study of the ranges of slow heavy ions.

Quite apart from using at first simple expressions like (2.2) and (2.4), it seems important—at the present stage of accuracy of theory and experiments—to be able to give a comprehensive description of slowing-down. It would for instance be futile to aim at an individual stopping curve for every one out of $\sim 10^4$ possibilities for the set of atomic numbers (Z_1, Z_2), where the suffixes 1 and 2 denote the penetrating particle and the atoms of the medium, respectively. If we are concerned with very high velocities, where the Bethe-Bloch stopping formula applies, the question of

dependence on Z_1 drops out because the stopping is simply proportional to Z_1^2. In that case the dependence on Z_2 is not far from being given by a Thomas-Fermi description, i. e. Bloch's relation $I = Z_2 \cdot I_0$, and only when high accuracy is demanded need we introduce deviations from the Thomas-Fermi results. Considering again the present case of comparatively low velocities, where the stopping is not proportional to Z_1^2, it is very important that descriptions of a Thomas-Fermi-like character are introduced, even though the resulting accuracy might not be high.

In point of fact, we hope to show in this section, and in § 5, that a Thomas-Fermi-like treatment of the dependence on both Z_1 and Z_2 has a quite satisfactory accuracy at the present stage of experimental precision. Our treatment should be based on a self-contained theory of the quasi-elastic collisions between ions and atoms. This theory will not be derived here; it is studied in two associated papers (Notes on Atomic Collisions, I and IV, unpublished). We shall merely summarize a few results of interest to us in the present connection (cf. also LINDHARD and SCHARFF, (1961)).

Electronic stopping

It is well known that for penetrating charged particles of high velocity, the energy loss to atomic electrons is completely dominating. The corresponding stopping cross section per atom is denoted by S_e, so that the specific energy loss is $N \cdot S_e$, where N is the number of atoms per unit volume. At high velocities S_e increases with decreasing particle velocity and has a maximum for a velocity of order of $v_1 = v_0 \cdot Z_1^{2/3}$. However, we shall consider low velocities only and in fact assume that $0 < v < v_1$. In the whole of this velocity region simple theoretical considerations lead to velocity proportional stopping, and a Thomas-Fermi picture shows that (Notes on Atomic Collisions, IV; see also LINDHARD and SCHARFF (1961))

$$S_e = \xi_e \cdot 8 \pi e^2 a_0 \cdot \frac{Z_1 Z_2}{Z} \cdot \frac{v}{v_0}, \quad v < v_1 = v_0 \cdot Z_1^{2/3}, \tag{2.5}$$

where the constant ξ_e is of order of $Z_1^{1/6}$, and $Z^{2/3} = Z_1^{2/3} + Z_2^{2/3}$. It is interesting that the approximate formula (2.5) holds down to extremely low velocities, i. e. also for $v << v_0$, in contrast to previous theoretical descriptions, where S_e was assumed to vanish for $v \lesssim v_0$ (cf. BOHR (1948), SEITZ (1949)).

It should be emphasized that (2.5) is approximate in more than one sense. The constant in (2.5) is based on Thomas-Fermi arguments, and it is to be expected that fluctuations around this constant can occur, especially for $Z_1 \lesssim 10$*. Moreover, a precise proportionality to v will not be correct over the whole of the velocity region $v < v_1$. However, in the present context we shall not analyse electronic stopping in detail. As to stopping near the maximum $v \sim v_1$, cf. NORTHCLIFFE (1963).

* The presence of such ionic shell effects is confirmed in the systematic measurements by ORMROD and DUCKWORTH (1963), WIJNGAARDEN and DUCKWORTH (1962).

Another important circumstance may be mentioned. The energy loss to electrons is actually correlated to the nuclear collisions, and in close collisions considerable ionization will take place. Although the correlations are fairly well known, we disregard them in first approximation and consider electronic stopping as a continuous process. The correlation may be of some importance especially in straggling or higher order moments of the range.

Nuclear stopping and scattering cross section

A basic quantity is the nuclear stopping cross section, S_n. However, since the energy transfer in individual collisions can be quite large, the slowing-down by nuclear collisions cannot always be considered as a nearly continuous process. It is therefore important to know the differential cross section too. We shall here consider various approximations, of which the first one lends itself to a particularly simple mathematical treatment.

Suppose that there is a potential $V(r)$ between the ion and the atom, such that $V(r) = (Z_1 Z_2 e^2 a_s^{s-1}/s\, r^s)$, with $a_s \approx a = 0.8853\ a_0\, Z^{-1/3}$ (the number $0.8853 = (9\pi^2)^{1/3} 2^{-7/3}$ is a familiar Thomas-Fermi constant). It is interesting that then the classical differential scattering cross section may be obtained approximately from an extrapolated perturbation procedure (Notes on Atomic Collisions I), leading to the simple result

$$d\sigma_n = \frac{C_n}{T_m^{1-1/s}} \frac{dT}{T^{1+1/s}}, \qquad s \geqq 1, \tag{2.6}$$

for an energy transfer T from the ion of energy E to an atom at rest. Here $T \leqq T_m = \gamma E = 4 M_1 M_2 (M_1 + M_2)^{-2} E$, T_m being the maximum energy transfer in the collisions. Furthermore, the constant C_n is connected to the stopping cross section S_n, and is approximately given by

$$C_n = \frac{\pi}{s} \left(b^2 \cdot a_s^{2s-2} \cdot \frac{3s-1}{8 s^2} \right)^{1/s} \cdot T_m = \left(1 - \frac{1}{s}\right) S_n, \tag{2.7}$$

where the collision diameter b is equal to $2 Z_1 Z_2 e^2 / M_0 v^2$, $M_0 = M_1 M_2/(M_1 + M_2)$. In the particular case of $s = 1$, i. e. simple Coulomb interaction, equation (2.6) also gives the correct Rutherford scattering, but in this case S_n in (2.7) does not represent the stopping cross section, the convergence of which is a result of adiabaticity in distant collisions.

As we shall demonstrate below, formulas of type of (2.6) are valuable for explorative purposes, interesting values of s being 1, 3/2, 2, 3 and 4. The cross sections (2.6) are furthermore in accord with the Thomas-Fermi scaling of units. Corresponding to the case of $s = 2$, we shall sometimes approximate S_n by constant standard stopping cross section S_n^0 (similar to that quoted by Bohr (1948)),

$$S_n^0 = (\pi^2 / 2.7183)\, e^2\, a_0\, Z_1\, Z_2\, M_1 \cdot Z^{-1/3} (M_1 + M_2)^{-1}. \tag{2.7'}$$

Beside the simple power potential we study the case provided by a screened potential, $U(r) = (Z_1 Z_2 e^2/r) \cdot \varphi_0 (r/a)$, where φ_0 is the Fermi function, and further

$a = a_0 \cdot 0.8853\,(Z_1^{2/3} + Z_2^{2/3})^{-1/2}$, which is a fair approximation to the ion-atom force. BOHR (1948) has employed a similar potential, with $\exp(-r/a_B)$ in place of $\varphi_0(r/a)$; however, an exponential function falls off too rapidly at large distances.

A screened Coulomb potential, involving only one screening parameter, a, leads for dimensional reasons to a natural measure of range and energy, for an ion colliding with atoms at rest. In fact, we may introduce, respectively,

$$\varrho = RNM_2 \cdot 4\pi a^2 \frac{M_1}{(M_1 + M_2)^2} \quad \text{and} \quad \varepsilon = E \frac{aM_2}{Z_1 Z_2 e^2 (M_1 + M_2)} \tag{2.8}$$

as dimensionless measures of range and energy. Note that ε^{-1} is essentially the parameter ζ used by BOHR (1948). The scattering in the screened potential, $U(r)$, is obtained by means of the extrapolated perturbation method for classical scattering used in deriving (2.6), and one obtains a universal differential cross section

$$d\sigma = \pi a^2 \frac{dt}{2\, t^{3/2}} f(t^{1/2}), \tag{2.9}$$

where $t^{1/2} = \varepsilon \cdot \sin(\vartheta/2)$ and ϑ is the deflection in centre of gravity system. When elastic collisions are assumed, we find $\sin^2(\vartheta/2) = (T/T_m)$, where T and T_m are the energy transfer and its maximum value, respectively, in a collision with an atom at rest. The function $f(t^{1/2})$ is shown in Fig. 1. At high values of t it approaches the Rutherford scattering. In Fig. 1 is also shown (2.6) for the case of $s = 2$.

It may be noted that the power law (2.6) leads to $f = f_s$, where

$$f_s(t^{1/2}) = \lambda_s \cdot t^{\frac{1}{2} - \frac{1}{s}}, \quad 0.3 \lesssim \lambda_s \lesssim 1. \tag{2.6'}$$

In the above, we have at first considered approximate potentials representing the ion-atom interaction and next, in an approximative way, derived the scattering from the potentials. However, we shall in the following take a simpler and more direct point of view. We consider (2.6) and (2.9) directly as approximations to the true scattering cross section and disregard the connection to a corresponding potential. This is the more justified, since the scattering is only quasi-elastic and cannot in detail be described by a potential between two heavy centres.

From (2.9) and Fig. 1 may be derived the nuclear stopping cross section, by means of the formula $(d\varepsilon/d\varrho)_n = \int_0^{\varepsilon} dx\, f(x)\, \varepsilon^{-1}$. The result is shown in Fig. 2, together with the stopping from (2.6) for $s = 2$. Also the electronic stopping may be expressed in $\varrho - \varepsilon$ units, and is then $(d\varepsilon/d\varrho)_e = k \cdot \varepsilon^{1/2}$, where the constant k varies only slowly with Z_1 and Z_2, and according to (2.5) is given by

$$k = \xi_e \cdot \frac{0.0793\, Z_1^{1/2} Z_2^{1/2} (A_1 + A_2)^{3/2}}{(Z_1^{2/3} + Z_2^{2/3})^{3/4} A_1^{3/2} A_2^{1/2}}, \quad \xi_e \simeq Z_1^{1/6}. \tag{2.10}$$

Thus, k is normally of order of 0.1 to 0.2, and only in the exceptional case of $Z_1 << Z_2$ can k become larger than unity. If $Z_1 = Z_2$, $A_1 = A_2$, the constant k is given by the simple expression $k = 0.133\, Z_2^{2/3} A_2^{-1/2}$. A representative case of electronic stop-

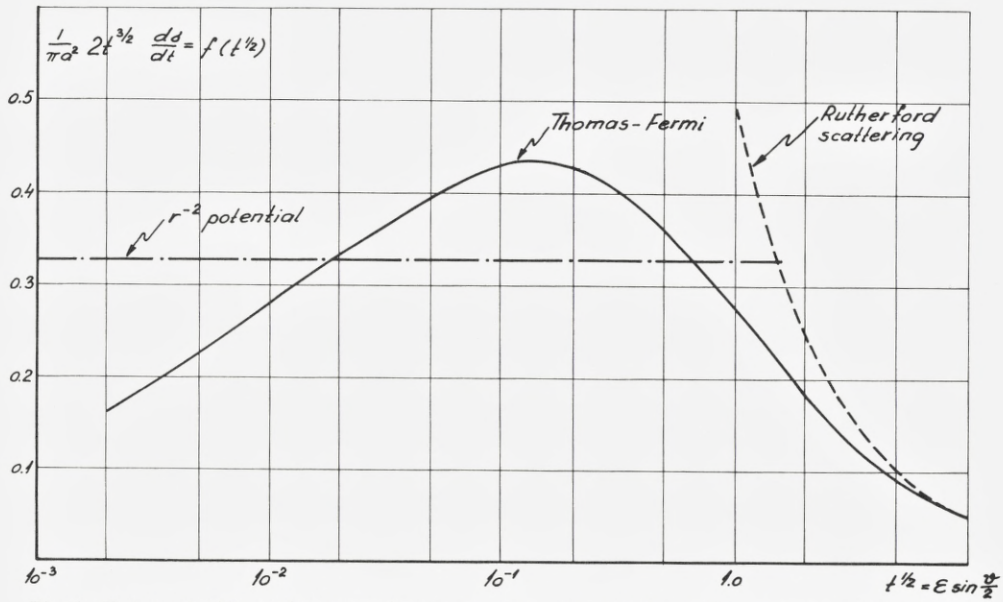

Fig. 1. Universal differential scattering cross section for elastic nuclear collisions, (2.9), based on a Thomas-Fermi type potential. At high values of $t^{1/2}$ it joins smoothly the Rutherford scattering. The cross section corresponding to power law scattering (2.6), or (2.6′), with $s = 2$ is also shown.

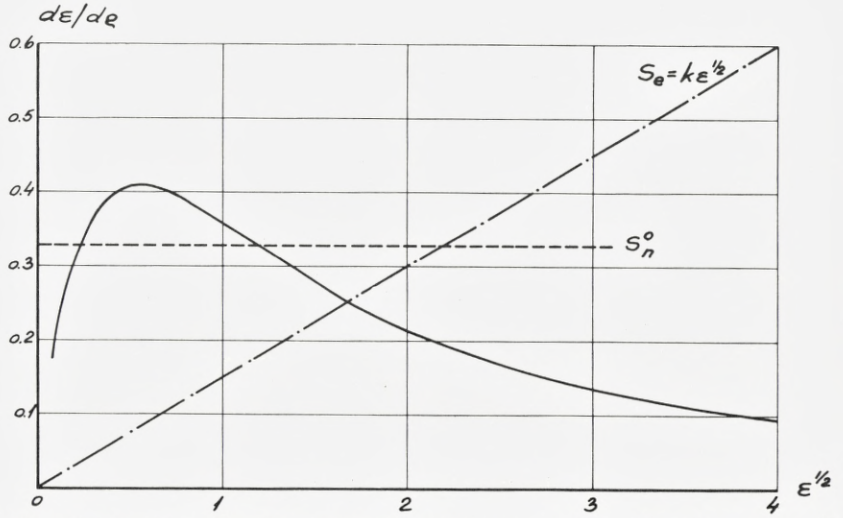

Fig. 2. Theoretical nuclear stopping cross section in $\varrho - \varepsilon$ variables. The abscissa is $\varepsilon^{1/2}$, i. e. proportional to v. The full-drawn curve is $(d\varepsilon/d\varrho)_n$, computed from Fig. 1. The horizontal dashed line indicates (2.7′). The dot-and-dash line is the electronic stopping cross section, $k\varepsilon^{1/2}$, for $k = 0.15$.

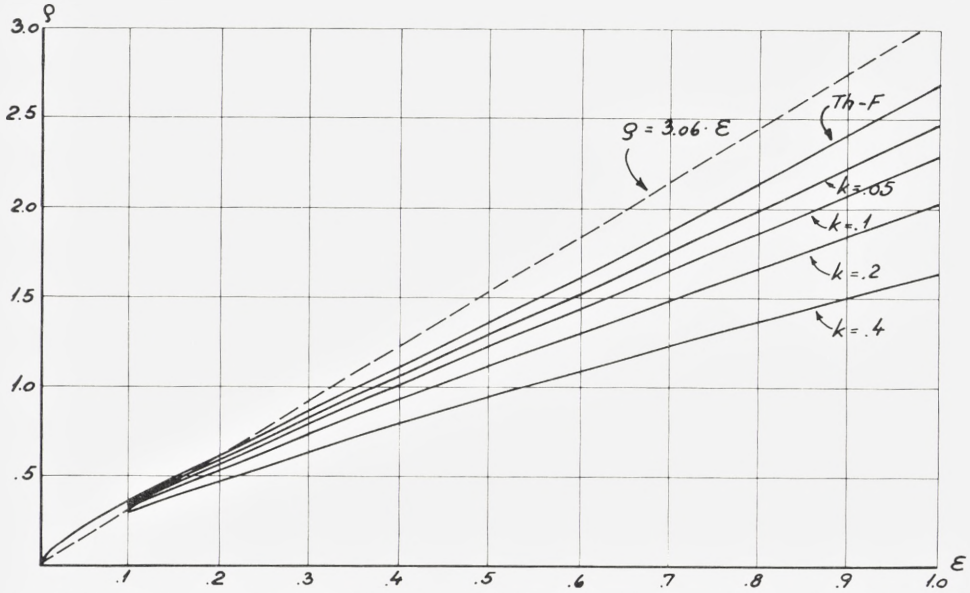

Fig. 3. Universal range-energy plot for $\varepsilon < 1$, cf. § 2 and § 3. The curve Th.–F. gives $\overline{\varrho}_1(\varepsilon)$, i. e. (2.2), as a function of ε with neglect of electronic stopping. Curves for various values of the constant k in electronic stopping are also shown. Dotted straight line is the standard range, $\varrho = 3.06\ \varepsilon$.

ping, $k = 0.15$, is shown in Fig. 2. Formula (2.10) applies for $v < v_1$, or approximately $\varepsilon < 10^3$. In the above we have for simplicity distinguished between electronic excitation and elastic nuclear collisions. This is not quite justified, since in close collisions there is a strong coupling between the two, i. e. the nuclear collisions are not elastic. In first approximation this need hardly be taken into account; the reader is referred to Notes on Atomic Collisions IV for a more detailed treatment of quasi-elastic collisions.

The nuclear scattering cross section is expected to be fairly accurate, but while shell effects should be of little importance, a systematic overestimate may occur, due to neglect of inelastic effects. A more thorough discussion is given in Notes on Atomic Collisions I. At low energies nuclear stopping dominates over electronic stopping (2.5). It must be emphasized though, that at extremely low ε-values, $\varepsilon \lesssim 10^{-2}$, the nuclear scattering and stopping becomes somewhat uncertain, because the Thomas-Fermi treatment is a crude approximation when the ion and the atom do not come close to each other.

Range-energy relations

By means of the simple formula (2.2), and the above stopping cross sections, we are now able to estimate total ion ranges. Now, if we consider nuclear stopping only, and one screening length a in the scattering, the

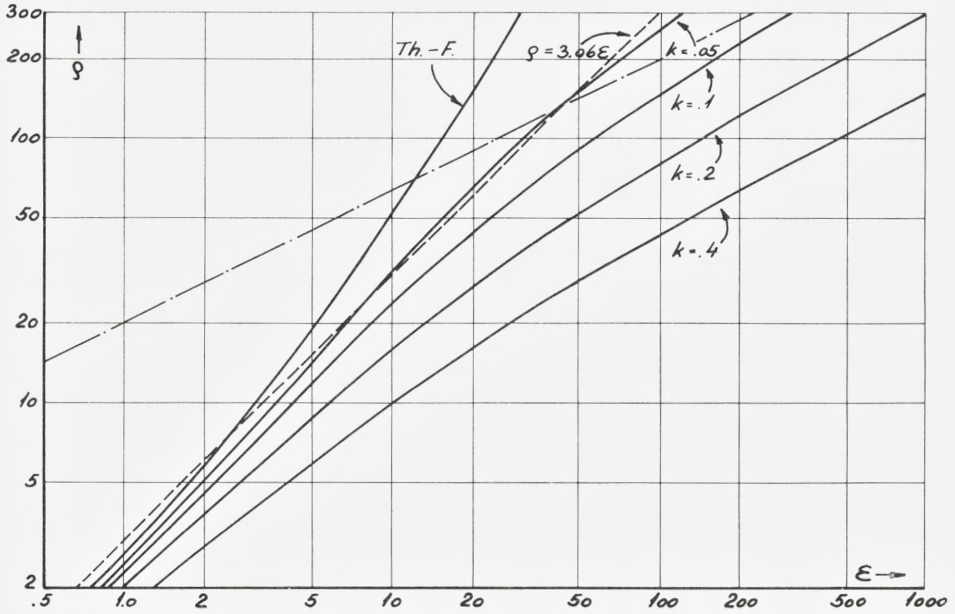

Fig. 4. The continuation at higher ε-values of the ranges $\overline{\varrho}_1(\varepsilon)$ in Fig. 3, for various values of constant k in electronic stopping. Straight dot-and-dash line is hypothetical range without nuclear stopping and $k = 0.1$.

dimensional arguments leading to (2.8) apply, and in these units the range in (2.2), ϱ, must be a function of ε only, i. e.

$$\varrho = \varrho(\varepsilon)$$

for all ions and atoms. This formula holds both when (2.7) and when (2.9) is introduced in (2.2). The resulting range, based on (2.9) and $f(t^{1/2})$ from Fig. 1 is shown by the solid curve in Fig. 3, for relatively small values of ε. The particular approximation of $s = 2$, i. e. the constant standard stopping cross section in (2.7′) and Fig. 2 leads to the straight line $\varrho = 3.06\ \varepsilon$ in Fig. 3. This standard range is closely similar to the range formula used by Bohr (1948) and also by Nielsen (1956). For small ε-values the numerical curve remains above the straight line and has a downward curvature, corresponding to the effective power of the potential being higher than 2, in fact of order of 3. The detailed behaviour of the range curve can be easily understood from the stopping curves in Fig. 2. If we use the straight line as a standard in Fig. 3, i. e. the horizontal line as a standard in Fig. 2, the range must at first be higher than the standard straight line in Fig. 3. Next, since the actual stopping rises above the horizontal line, the range

must drop considerably relative to the straight line, and actually fall below it. Finally, since the nuclear stopping becomes small in the high energy region with Rutherford scattering, the range must again increase above the straight line as may be seen in Fig. 4.

In this description we have so far omitted electronic stopping. This omission is justified at low energies because S_e/S_n tends to zero for small velocities, but at higher energies it becomes less and less adequate until the range finally is dominated by the electronic stopping, as may be judged from the stopping cross section in Fig. 2. Let us therefore take electronic stopping into account and write

$$\frac{d\varepsilon}{d\varrho} = \left(\frac{d\varepsilon}{d\varrho}\right)_n + k \cdot \varepsilon^{1/2}, \tag{2.11}$$

where $(d\varepsilon/d\varrho)_n$ is shown in Fig. 2, and the electronic stopping is assumed to be proportional to $\varepsilon^{1/2}$, i. e. we are concerned with moderate velocities, $v < v_1$. We choose a number of representative values of the constant k, $k = 0.05$, 0.1, 0.2 and 0.4. Values of k between 0.1 and 0.2 are quite common, according to (2.5). In Figs. 3 and 4 are shown the range curves for the above four values of k. The most conspicuous effects of electronic stopping are, first, that it leads to appreciable range corrections even at quite low ε-values. Second, for ε large compared to unity, the reduction in range always dominates, so that the range never increases above the straight line $\varrho = 3.06\,\varepsilon$, in contrast to the range with neglect of electronic stopping. In Fig. 4 is also shown the hypothetical range $\varrho = (2/k)\varepsilon^{1/2}$, which would result if there were no nuclear stopping, in the case of $k = 0.1$.

By means of curves like those in Figs. 3 and 4 we are able to compare or estimate ranges for all ions in all substances. But only for ε-values below, say, $\varepsilon = 10$ are curves for the various k-values fairly close together and easy to compare. For light ions in heavy substances deviations start at even smaller ε-values, because k becomes quite large. Moreover, only for these low values are we able to check in a direct manner the nuclear stopping, which here remains dominating.

Although we may well use Fig. 4 for estimates of ranges when $\varepsilon >> 10$, we can in this case introduce a more critical comparison between theory and experiments. In fact, it is apparent from Fig. 2 that for high values of ε the range is mainly determined by the electronic stopping, and only a minor range correction is due to nuclear stopping which dominates at low values of ε. Since nuclear stopping drops off quickly while electronic stopping increases, the nuclear stopping correction to the range remains fairly con-

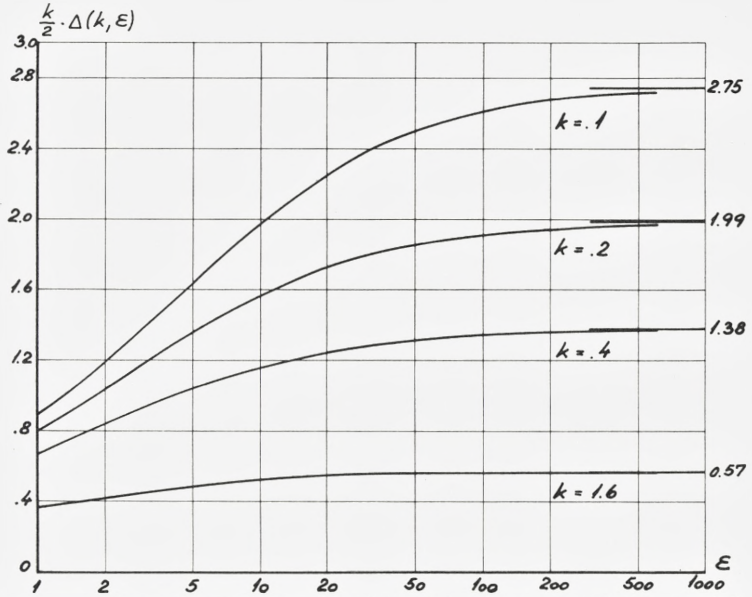

Fig. 5. Range corrections for nuclear stopping, $(k/2)\,\Delta(k, \varepsilon)$, from equ. (2.12). Curves are shown for $k = 0.1$, 0.2, 0.4 and 1.6. Asymptotic values are roughly $\Delta \to 1.76 \cdot k^{-3/2}$.

stant above a certain value of ε. We then introduce an extrapolated electronic range

$$\varrho_e(\varepsilon) = \int_0^{\varepsilon} \frac{d\varepsilon'}{(d\varepsilon'/d\varrho)_e} = \int_0^{\varepsilon} \frac{d\varepsilon'}{(d\varepsilon'/d\varrho)} + \int_0^{\varepsilon} \frac{(d\varepsilon'/d\varrho)_n \cdot d\varepsilon'}{(d\varepsilon'/d\varrho) \cdot (d\varepsilon'/d\varrho)_e} \qquad (2.12)$$

$$= \varrho(\varepsilon) + \Delta(k, \varepsilon).$$

The quantity $\Delta(k, \varepsilon)$ can be computed from the above formulas, and adding Δ to an observed $\varrho(\varepsilon)$, we obtain the extrapolated electronic range, which in our case of $v < v_1$ should be equal to $\varrho_e = (2/k)\,\varepsilon^{1/2}$ (cf. dot-and-dash line in Fig. 4).

The function $\Delta(k, \varepsilon)$ is shown in Fig. 5 for k-values between 0.1 and 1.6. This procedure is probably the most direct way of comparing theoretical predictions of electronic stopping like (2.5) with range observations. The point is here that Δ often is a relatively small correction, and in estimating the range correction Δ we may use (2.5), even if this formula be not too accurate. Examples of the application of (2.12) and Fig. 5 are shown in § 5, cf. Figs. 14 and 15.

Another circumstance may be noted in this connection. Since Δ tends to a constant at high ε-values, we may moreover use (2.12), together with Fig. 5, for comparisons with measurements at high ε-values, i. e. $v >> v_1$, where electronic stopping no longer increases proportionally to v, but instead decreases approximately as v to a power between -1 and -2.

In the present paragraph we do not make comparisons with actual range measurements, one of the reasons being that measured ranges require corrections of the kind discussed in § 4. Instead, we have presented these comparisons in § 5, where recent measurements are compiled. We do not discuss critically the accuracy of the measurements; this is perhaps unsatisfactory, because several new experimental methods have been applied. We merely make approximate and obvious range corrections, corresponding to the results in § 4. One result emerging from § 5 is that the theoretical nuclear stopping, as leading to the range curves in Figs. 3 and 4, for moderate ε-values appears to be in good agreement with observations, perhaps within $\sim$20 percent. It should be noted that the theory is somewhat uncertain at quite low ε-values, i. e. $\varepsilon \lesssim 10^{-2}$.

Beside the general experimental checking of the present range-energy relations there are several other ways of comparison. An immediate possibility is to measure directly stopping powers, which has been done in a few cases, but mostly when electronic stopping dominates. We shall not enter more critically into these questions, since the theory of electronic stopping is not the topic of the present paper. Nor will we attempt a detailed discussion of individual inelastic collisions between energetic ions and atoms at rest. But it may be mentioned that more subtle comparisons of ranges may be made. For instance, isotope effects are quite informative, and can elucidate both electronic and nuclear stopping, cf. § 5.

Range straggling

The simple description used here, with a range along the particle path based on (2.2), may now be extended to include an average square fluctuation in range, given by (2.4). This description contains the assumption that range fluctuations are relatively small. We may suppose that the fluctuations around the average correspond nearly to a Gaussian. In fact, if this were not so, the distribution in range would have a sizable skewness. Then we would have to distinguish between e. g. the most probable and the average range, and the simple relation (2.2) would have to be revised. Still, even in such cases the results in the present paragraph may be useful. We can in fact consider the present ranges, i. e. (2.2) as an approximation to the

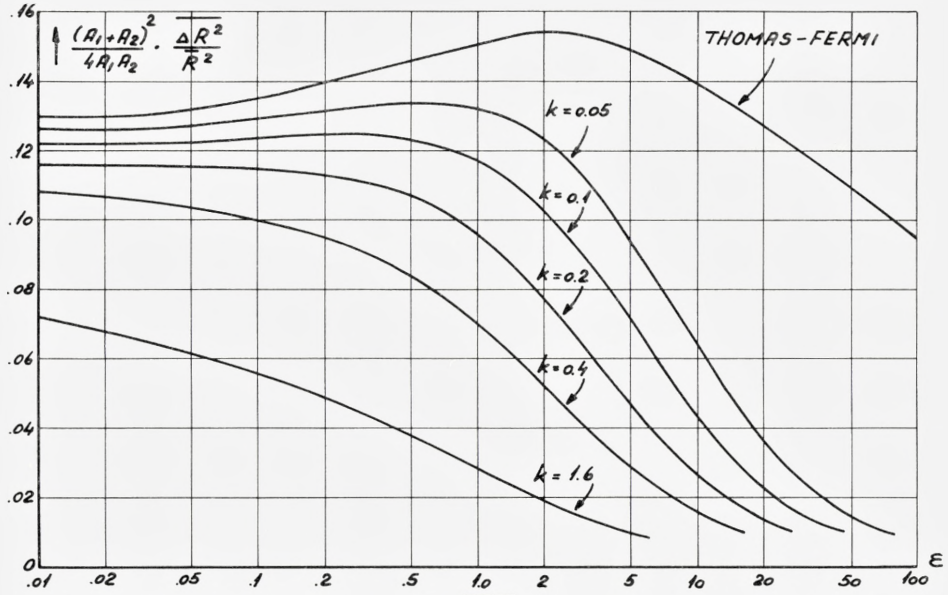

Fig. 6. Relative square straggling in range $(\Delta R^2)_1/\overline{R}_1^2$, divided by $\gamma = 4M_1M_2/(M_1+M_2)^2$. Curves are shown for several values of constant k in electronic stopping.

average range, and similarly consider the present range fluctuation, i. e. (2.4), as an approximation to the average square fluctuation in range. These averages are defined irrespective of the skewness of the distribution; they are studied in more detail in § 3, where also the accuracy of the present treatment is discussed more closely.

It is convenient to consider the relative square straggling in range, $(\Delta\varrho/\varrho)^2 = (\Delta R)^2/R^2$. Consider first nuclear stopping only, and in particular the power potentials represented by (2.6). Then we easily find

$$\left(\frac{\Delta\varrho}{\varrho}\right)^2 = \frac{s-1}{s(2s-1)}\gamma, \tag{2.13}$$

where $\gamma = 4M_1M_2/(M_1+M_2)^2$. We thus obtain the extremely simple result that the relative straggling is independent of the range itself. It is moreover interesting to note that the result (2.13) is rather insensitive to s in the neighbourhood of $s = 2$. When s increases from 2 to 3 the relative square straggling decreases by only 20 percent. Thus the simple model predicts that at low energies $(\Delta\varrho/\varrho)^2$ should be of order of $\gamma/6$ (cf. also LINDHARD and SCHARFF (1961), LEACHMAN and ATTERLING (1957), HARVEY (1960)). We have here

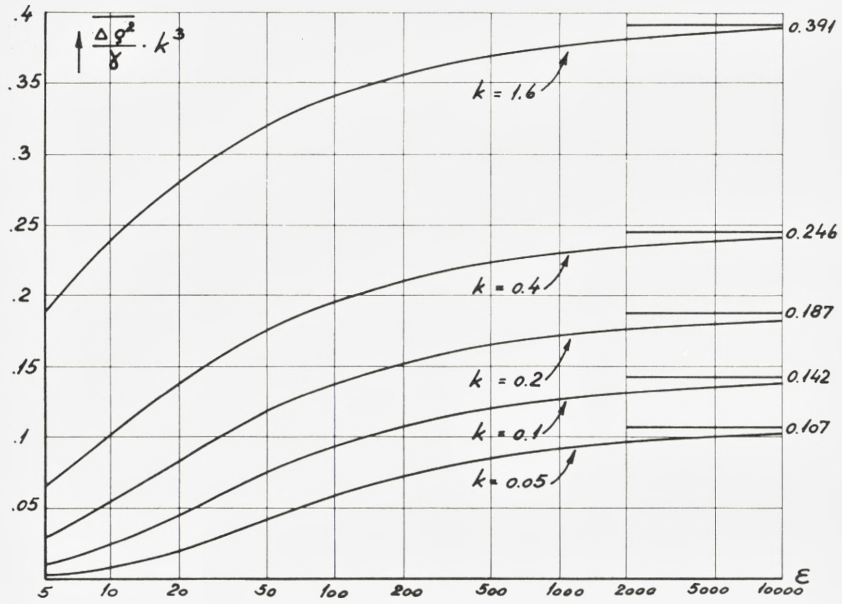

Fig. 7. Absolute straggling in range $(\overline{\Delta\varrho^2})_1$, at high values of ε. Saturation values of the straggling are indicated.

considered the application of (2.6) to the simple formulas (2.2) and (2.4). A more detailed study of the probability distribution in range is made in § 3, on the basis of the power law scattering (2.6). It is shown there that the right hand side of (2.13) is only the first term in a power series expansion in γ.

We next apply the numerical Thomas-Fermi-type cross section (2.9) for scattering in nuclear collisions. We do this at first with neglect of electronic stopping, and by means of (2.4) we compute $(\Delta\varrho/\varrho)^2 \cdot \gamma^{-1}$ against ε, as shown by the upper curve in Fig. 6. The relative straggling is seen to behave as expected from the simple power potential. Next, we include electronic stopping, using (2.5) and assuming that the contribution to straggling from electronic stopping is negligible*. Clearly, it must lead to a reduced relative straggling. The results are shown in Fig. 6, for $k = 0.05$, 0.1, 0.2, 0.4 and 1.6. At ε-values around 1 to 10 a considerable reduction in the relative straggling sets in. The reduction corresponds to the circumstance that in this

* This assumption can be questioned, since quasi-elastic collisions imply a correlation between the two types of energy loss, i. e. nuclear and electronic stopping. The assumption requires that a considerable part of electronic stopping occurs at impact parameters where recoil of the atom is small or moderate.

region electronic stopping has become quite dominating, and the absolute value of the square straggling, $(\Delta\varrho)^2$, does not increase much beyond this point. For high ε-values it is then convenient to consider the absolute value of the range straggling. The corresponding curves are given in Fig. 7, for various values of k. We therefore conclude that accurate measurements of straggling in range at high energies, where the electronic stopping does not at all correspond to (2.5), may give information about the predicted values of k, as given by (2.10).

The above treatment of simple ranges and range straggling is intended to be fairly comprehensive, and from the accompanying curves it is easy to obtain reasonable estimates of these quantities for any value of Z_1, A_1, Z_2, A_2 and v. However, we have disregarded completely those cases where the substance contains several atomic elements, $Z_2^{(1)}$, $Z_2^{(2)}$, etc., in given ratios. In all such cases, the nuclear stopping contribution from each element may be derived from the solid curve in Fig. 2, with a rescaling of units. The electronic stopping contributions are obtained from (2.5) or (2.10). The resulting ranges can be derived by numerical integration. However, considerable simplification occurs in an energy region where, e. g. the stopping cross section $S^{(i)}$, due to any atomic component i, is proportional to the same power of E, because in this case straightforward computations of averages may be made. For two components, a and b, we have $R = R_a R_b (R_b x_a + R_a (1 - x_a))^{-1}$, where R_a and R_b are the ranges in a and b, and x_a and $1 - x_a$ are the relative abundances of a and b. Similar procedures may be used in the case of straggling in range.

§ 3. Distribution in Range Measured Along the Path

In the present chapter we shall try to go one step beyond the treatment in § 2, where only a simple range straggling was considered, and where it was tacitly assumed that straggling effects were small. We wish to check the validity of this picture and also to extend it. A basic reason for the extended treatment are the large fluctuations, known to result from encounters between slow heavy ions and atoms. We therefore attempt to study the probability distribution in range measured along the path. Although this distribution is much simpler than the distribution in space of the endpoint of the path, it is not easily obtained. One might perhaps employ Monte Carlo methods* for the solution of representative cases, but we shall limit the treatment to typical and simple approximations, and in particular consider the power law scattering cross sections given by (2.6).

Consider again a particle (Z_1, A_1) with energy E, in a medium (Z_2, A_2).

* Monte Carlo methods were applied by e. g. Robinson, Holmes and Oen (1962) to various models of nuclear scattering, but with neglect of electronic stopping, cf. also Holmes (1962).

We denote by R the range measured along the particle path, i. e. the total distance traversed by the particle. Let $p(R, E)\,dR$ represent the probability that the particle has a range between R and $R+dR$, so that

$$\int_0^\infty p(E, R)\,dR = 1 \quad \text{and} \quad <R^m> = \int_0^\infty p(E, R)\,R^m\,dR\,.$$

An integral equation for $p(E, R)$ may be derived as follows. Suppose that the particle with energy E moves a path length δR in a medium containing N atoms per unit volume. There is then a probability $N\delta R d\sigma_{n,\,e}$ for a collision specified by energy transfer $\sum_i T_{ei}$ to electrons (electrons labelled by suffix i) and by an energy transfer T_n to translational motion of the struck atom. The particle will thus have an energy $E - T_n - \sum_i T_{ei}$. If the collision takes place, the particle has a probability $p(R-\delta R,\ E - T_n - \sum_i T_{ei})$ of obtaining the total range R. Multiplying by the probability of collision, $N\delta R d\sigma_{n,e}$, we get the contribution from this specified collision to the total probability for range R. We next sum over all collisions. There is left a probability $1 - N\delta R \int d\sigma_{n,\,e}$ that no collision occurs. In this event we clearly get a contribution $(1 - N\delta R \int d\sigma_{n,\,e}) \cdot p(R-\delta R, E)$ to the total probability for the range R.

Collecting the above contributions we have an alternative expression for $p(R, E)$,

$$p(R, E) = N\delta R \int d\sigma_{n,e}\, p(R-\delta R, E - T_n - \sum_i T_{ei})$$
$$+ (1 - N\delta R \int d\sigma_{n,e}) \cdot p(R-\delta R, E),$$

and in the limit of $\delta R \to 0$,

$$\frac{\partial p(R, E)}{\partial R} = N \int d\sigma_{n,e} \left\{ p(R, E - T_n - \sum_i T_{ei}) - p(R, E) \right\}, \tag{3.1}$$

which expression constitutes the basic integral equation governing the probability distribution in range along the path. In the remainder of this chapter we study the integral equation (3.1) and its consequences, using a number of approximations. We shall not further elaborate on the derivation of (3.1), but it may be noted that the formal limit of $\delta R \to 0$ corresponds to separability between consecutive collisions. If there is no separability, the equation still holds, or may be easily amended, as long as collisions with moderate or large T-values remain separable.

Besides separability we have assumed that successive collisions are not correlated. This holds if the atoms in the substance are in fact randomly distributed, or if e. g. impact parameters corresponding to sizable deflections are extremely small compared to interatomic distances, giving effectively uncorrelated events. A system where collisions are separated and uncorrelated may be termed a random system of atoms. The derivation of (3.1) is based on a random system, and we limit our treatment to this case. A solid with periodic lattice is for many purposes a random system, but at low ion energies deviations from (3.1) can occur. These deviations contain directional effects and are sensitive to lattice structure, cf. p. 32.

On the assumption that energy losses to electrons are small and separated from nuclear collisions, we obtain

$$\left.\begin{aligned}\frac{\partial p(R,E)}{\partial R} &= N\int d\sigma_n \left\{p(R, E-T_n) - p(R,E)\right\} \\ &\quad - NS_e(E)\frac{\partial}{\partial E}p(R,E),\end{aligned}\right\} \tag{3.2}$$

which formula is somewhat less general, but applicable to our previous cross sections for scattering.

We may rewrite (3.2) on the assumption that the Thomas-Fermi-like scattering formula (2.9) applies (note that this also includes (2.6) and (2.6′)), and then introduce the variables ϱ and ε. We readily obtain

$$\left.\begin{aligned}\frac{\partial}{\partial\varrho}\Pi(\varrho,\varepsilon) &= \frac{1}{\gamma}\int_0^{\varepsilon^2}\frac{dt}{2\,t^{3/2}} f(t^{1/2})\left\{\Pi\left(\varrho, \varepsilon - \frac{\gamma t}{\varepsilon}\right) - \Pi(\varrho,\varepsilon)\right\} \\ &\quad - \left(\frac{d\varepsilon}{d\varrho}\right)_e \frac{\partial}{\partial\varepsilon}\Pi(\varrho,\varepsilon),\end{aligned}\right\} \tag{3.3}$$

where $\Pi(\varrho, \varepsilon)\,d\varrho$ is the probability that a particle with energy parameter ε has a range between ϱ and $\varrho + d\varrho$, and where $\gamma = 4M_1M_2/(M_1+M_2)^2$. We have seen that in a wide region ($v < v_1$, i. e. roughly $\varepsilon < 10^3$), one may write $(d\varepsilon/d\varrho)_e = k\cdot\varepsilon^{1/2}$. In equation (3.3) we then have two parameters, k and γ.

A simple approach to the study of the integral equations (3.1), (3.2) or (3.3) is to obtain from these equations the moments $<R^m>$, whereby—at least in principle—the probability distribution itself may be determined too.

From (3.1) we obtain directly, when multiplying by R^m and integrating by parts

$$\left.\begin{aligned} m<R^{m-1}(E)> &= \\ N\int d\sigma_{n,e}\left\{<R^m(E)> - <R^m(E - T_n - \sum_i T_{ei})>\right\}&.\end{aligned}\right\} \tag{3.4}$$

Similarly, if (3.3) holds we arrive at a somewhat simpler relation

$$m<\varrho^{m-1}(\varepsilon)>=\frac{1}{\gamma}\int_0^{\varepsilon^a}\frac{dt}{2\,t^{3/2}}f(t^{1/2})\left\{<\varrho^m(\varepsilon)>-<\varrho^m\left(\varepsilon-\frac{\gamma\,t}{\varepsilon}\right)>\right\}\Bigg] \\ +\left(\frac{d\varepsilon}{d\varrho}\right)_e\cdot\frac{d}{d\varepsilon}<\varrho^m(\varepsilon)>\,.\Bigg] \tag{3.5}$$

By means of equations (3.4) or (3.5) we may successively derive the first, second, etc., moments of the range. In the resulting formulas the equations (3.4) are applied, because they have a wider applicability. In actual evaluations, however, we turn to (3.5), and to the analogous reformulations of (3.6) to (3.13) in $\varrho-\varepsilon$ variables, although the reformulations are not explicitly stated. Let us ask for the average range $\bar{R}(E)=<R(E)>$. According to (3.4)

$$1=N\int d\sigma_{n,e}\left\{\bar{R}(E)-\bar{R}(E-T_n-\sum_i T_{ei})\right\}. \tag{3.6}$$

An obvious procedure in solving (3.6) is to make a series development in powers of $T=T_n+\sum_i T_{ei}$. This approximation might seem poor when $M_1\sim M_2$, because $E-T$ can then take on any value between E and 0. However, we can profit from the circumstance that the energy transfer to electrons, $\sum_i T_{ei}$, is normally quite small, and that the nuclear scattering cross sections (2.9) are strongly forward peaked, since $f(t^{1/2})t^{-3/2}$ decreases approximately as t to a power between -1 and -2. We shall presently look into the accuracy of the various approximations.

Take at first only the first order terms in the brackets and denote the corresponding approximation to average range by $\bar{R}_1(E)$. We obtain from (3.6)

$$\frac{dR_1(E)}{dE}=\frac{1}{NS(E)},\quad \bar{R}_1(E)=\int_0^E\frac{dE'}{NS(E')}, \tag{3.7}$$

where $S(E)=S_n(E)+S_e(E)$ is the total stopping cross section. The formula (3.7) is exactly the straightforward equation (2.2) used in § 2.

Similarly, we can include higher order terms from (3.6),

$$1=NS(E)\frac{d}{dE}\bar{R}(E)-\frac{1}{2}N\Omega^2(E)\frac{d^2}{dE^2}\bar{R}(E)+\ldots\,, \tag{3.8}$$

where the quantity $\Omega^2(E)=\int d\sigma_{n,e}T^2$ is related to the straggling. If we include only the second order term we obtain a second order differential

equation which may be solved directly. Still, since the second order term may be considered small, we may express the second derivative by means of $\bar{R}_1$. This leads to $\bar{R}_2(E)$, the second approximation to average range

$$\bar{R}_2(E) = \int_0^E \frac{dE'}{NS(E')} \left\{ 1 + \frac{\Omega^2(E')}{2} \frac{d}{dE'} \left(\frac{1}{S(E')} \right) \right\}. \tag{3.9}$$

The average square fluctuation in range, $\overline{\Delta R^2}(E) = \overline{R^2}(E) - \bar{R}^2(E)$, is obtained from the second moment in (3.4), if we multiply (3.6) by $2\bar{R}(E)$ and subtract

$$\left.\begin{aligned} \int d\sigma_{n,e} \left\{ \overline{\Delta R^2}(E) - \overline{\Delta R^2}\left(E - T_n - \sum_i T_{ei}\right) \right\} = \\ \int d\sigma_{n,e} \left\{ \bar{R}(E) - \bar{R}\left(E - T_n - \sum_i T_{ei}\right) \right\}^2. \end{aligned}\right\} \tag{3.10}$$

In this equation the right hand side is a known source term. If we take the same successive steps as in the computation of $\bar{\varrho}$, we make a series development in (3.10), in powers of T. The first terms on both sides of the equation lead to the approximation $(\overline{\Delta R^2})_1$,

$$S(E) \frac{d}{dE} (\overline{\Delta R^2})_1 = \Omega^2(E) \left(\frac{d}{dE} \bar{R}(E) \right)^2, \tag{3.11}$$

where for $\bar{R}(E)$ we should use the first approximation, $\bar{R}_1(E)$. Therefore, also (3.11) brings us back exactly to our previous assumptions in § 2, in this case to (2.4).

Including terms in (3.10) up to second order, we get

$$\left.\begin{aligned} S(E) \frac{d}{dE} (\overline{\Delta R^2}) - \frac{\Omega^2(E)}{2} \frac{d^2}{dE^2} (\overline{\Delta R^2}) = \\ \left(\frac{d}{dE} \bar{R} \right)^2 \Omega^2 - \frac{K(E)}{2} \frac{d}{dE} \left(\frac{d}{dE} \bar{R} \right)^2, \end{aligned}\right\} \tag{3.12}$$

where $K(E) = \int d\sigma_{n,e} T^3$. When assuming the new terms in (3.12) to be small, we obtain the second approximation to $(\overline{\Delta R^2})$,

$$\frac{d}{dE} (\overline{\Delta R^2})_2 = \frac{\Omega^2(E)}{S^3(E) N^2} \left\{ 1 + \left(\frac{K}{\Omega^2 S} - \frac{5\Omega^2}{2S^2} \right) \frac{dS}{dE} + \frac{1}{2S} \frac{d\Omega^2}{dE} \right\}. \tag{3.13}$$

By means of the expression (3.13) we are able to estimate the accuracy of the straightforward formulas (3.11) and (2.4). It is important to notice that

TABLE 1

Comparison of first and second approximation of expansion in γ, for power law scattering. Results for average range and range straggling.

s	$\bar{R}_2/\bar{R}_1$	$\overline{(\Delta R^2)}_2/\overline{(\Delta R^2)}_1$
3/2	$1+\gamma/24$	$1+\gamma\cdot 0.10$
2	1	$1+\gamma/6$
3	$1-\gamma/15$	$1+\gamma\cdot 0.14$

the successive approximations made above are simply series expansions of average range and straggling to successive powers of $\gamma = T_m/E$.

It is of interest to compare the above approximations. For simplicity let us consider low energies and disregard electronic stopping. Since electronic stopping here tends to diminish fluctuation effects, we obtain in this way slightly exaggerated differences between successive range approximations. Moreover, we use power law scattering cross sections (2.6) or (2.6′). This permits exact computation of $\bar{R}(E)$. Note that according to (2.6) the ranges are proportional to $E^{2/s}$, while the square straggling in range behaves as $E^{4/s}$. We may compare $\bar{R}_1$, $\bar{R}_2$ and $\bar{R}$, and similarly $(\overline{\Delta R^2})_1$, $(\overline{\Delta R^2})_2$ and $\overline{\Delta R^2}$. The results depend on γ, i. e. on the mass ratio. For small values of γ, a series development in powers of γ is accurate. Since γ is often close to its maximum value, $\gamma = 1$, we also compare the approximations in this case. The results are listed in Table 1 ($\gamma << 1$) and Table 2 ($\gamma = 1$), in the cases $s = 3/2$, 2 and 3. Notice that at low energies values of s between 2 and 3 are of particular interest.

In the approximation used in Table 1 the range $\bar{R}_2$ and its fluctuation $(\overline{\Delta R^2})_2$ are equal to the exact average values $\bar{R}$ and $\overline{\Delta R^2}$, respectively. From Tables 1 and 2 it is apparent that $\bar{R}_2(E)$ is always a very good approximation to $\bar{R}(E)$, and one need not distinguish between the two. The range $\bar{R}_1(E)$ is somewhat less accurate, but deviates from $\bar{R}(E)$ by no more than 10 percent in the least favourable case ($\gamma = 1$). In actual range observations the deviation is reduced by electronic stopping and by the change in effective s with particle energy. There remains a difference between $\bar{R}_1$ and $\bar{R}$ only at the lowest values of ε. For our present purposes where all range curves (e. g. Figs. 3 and 4) are stated in terms of $\bar{R}_1(E)$ we need hardly distinguish between $\bar{R}_1(E)$ and $\bar{R}(E)$, because of obvious uncertainties in theory and experiment. Still, one might ask why the range curves are computed for $\bar{R}_1$ in place of $\bar{R}_2$. This is simply because a universal range curve would not result when $\bar{R}_2$ is used.

TABLE 2

Comparison of first and second approximation with exact formula when $\gamma = 1$. Average ranges and range straggling for power law scattering.

s	$\bar{R}/\bar{R}_1$	$\bar{R}/\bar{R}_2$	$(\overline{\Delta R^2})/(\overline{\Delta R^2})_1$	$(\overline{\Delta R^2})/(\overline{\Delta R^2})_2$
3/2	1.053	1.01	1.03	0.94
2	1	1	1.20	1.03
3	0.904	0.97	1.26	1.10

The straggling approximations $(\overline{\Delta R^2})_1$ and $(\overline{\Delta R^2})_2$ are, as a rule, a little smaller than $\overline{\Delta R^2}$ when $\gamma = 1$. This deviation becomes quite pronounced if instead we consider the relative straggling in range. Thus, in the extreme cases of $s = 3$ and $\gamma = 1$ we have $(\overline{\Delta R^2})_1/\bar{R}_1^2 = 0.133$ according to (2.13), while $\overline{\Delta R^2}/\bar{R}^2 \cong 0.20$ for $\gamma = 1$ and $2 < s < 3$. At quite low values of ε, and $\gamma = 1$, the straggling in Fig. 6 is therefore somewhat lower than the straggling in average range; still it is noteworthy that the electronic stopping has a considerable influence on straggling also for quite low values of ε. We infer moreover that the absolute values of range straggling in Fig. 7 are expected to represent $\overline{\Delta R^2}$ quite accurately, i. e. they are superior to the relative straggling values in Fig. 6. Note that the deviations are only important when $\gamma \cong 1$. The outcome of the discussion in the present chapter is therefore that the simple quantities $\bar{R}_1$ and $(\overline{\Delta R^2})_1$, introduced already in § 2, are satisfactory estimates of average range and average square fluctuation in range.

Results for power law scattering

In the interesting case of power law scattering, (2.6′), the formula (3.3) takes a particularly simple form if electronic stopping is neglected. In fact, we then obtain

$$\frac{\partial}{\partial r} P(r,\varepsilon) = \int_0^1 \frac{dy}{y^{1+1/s}} \left\{ (1-\gamma y)^{-2/s} P\left(r \cdot [1-\gamma y]^{-2/s},\ \varepsilon \cdot [1-\gamma y]\right) - P(r,\varepsilon) \right\}, \quad (3.13)$$

where $r = \lambda_s \varrho \cdot (2\gamma\varepsilon^{2/s})^{-1}$ and $\int_0^\infty P(r,\varepsilon)\,dr = 1$. If the power law holds down to zero energy, equation (3.13) permits us to choose $P(r,\varepsilon)$ independent of ε, and an extremely simple recursion formula is obtained for the moments of the distribution,

$$m < r^{m-1} > = < r^m > \cdot I(\gamma, m, s), \qquad I(\gamma, m, s) = \int_0^1 \left\{ 1 - (1-\gamma y)^{2m/s} \right\} \frac{dy}{y^{1+1/s}}. \quad (3.14)$$

The moments therefore only depend on one parameter, γ, for any given power law scattering.

This result, where virtually the whole range distribution is determined immediately for any energy when merely the power s is stated (and γ is known), is

clearly a direct consequence of universal cross sections, $f(t^{1/2})$. In a more qualitative sense, it is apparent that if at one particle energy a cross section is given as a function of $T/T_m = \sin^2 \vartheta/2$, this cross section leads to a certain ion-atom potential from which the scattering at all lower energies may be derived. This circumstance is expressed in an approximate way by the unified cross section, (2.9), and the results happen to be analytically simple for a power law cross section.

The integral $I(\gamma, m, s)$ may be expressed by means of the incomplete beta function (cf. ERDÉLYI et al. (1953)),

$$I(\gamma, m, s) = -s\left\{1-(1-\gamma)^{2m/s}\right\} + 2m\gamma^{1/s} B_\gamma\left(1-\frac{1}{s}, \frac{2m}{s}\right), \tag{3.15}$$

and is particularly simple when $\gamma << 1$, in which case a power series in γ converges rapidly,

$$\left.\begin{aligned} I(\gamma, m, s) = \frac{2m\gamma}{s-1}\Bigg\{1 - \frac{\gamma}{2s}\,\frac{s-1}{2s-1}(2m-s) \\ + \frac{\gamma^2}{3s^2}\,\frac{s-1}{3s-1}(2m-s)(m-s) + \ldots\ldots\Bigg\},\ \gamma << 1. \end{aligned}\right\} \tag{3.16}$$

An interesting case is also $\gamma = 1$, where the incomplete beta function in (3.15) becomes the usual beta function $B_1(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$.

The results in (3.14), (3.15) and (3.16) were used in Tables 1 and 2 for the computation of the first and second moments in various approximations. It is easy to derive also higher moments.

§ 4. Projected Ranges and Associated Quantities

Average projected range

An interesting quantity appears to be the projection of the range on the initial direction of the particle path. This quantity is often observed directly. Thus, one might be concerned with a collimated beam of particles passing through a number of foils perpendicular to the direction of the beam; the number of particles collected in each foil gives just the distribution in range projected on the initial direction of the beam. We may, in fact, define the concept of projected range as follows. A particle starts inside an infinite homogeneous medium from the origin in the direction of the x-axis; the value of x for the end point of the path is the projected range, R_p. The distribution in x is the distribution in projected range. Quantities of particular interest here are the average projected range, $\bar{R}_p = \bar{R}_p(E)$, and the average straggling in projected range, $\Delta R_p^2 = \overline{R_p^2} - \bar{R}_p^2$.

An integral equation for the average projected range may be obtained in analogy to the derivation of (3.1). We find readily

$$1 = N\int d\sigma_{n,e}\{\bar{R}_p(E) - \bar{R}_p(E-T)\cos\varphi\}, \tag{4.1}$$

where $T = T_n + \sum_i T_{ei}$, and φ is the deflection of the ion in the laboratory system. There is a close similarity to the integral equation (3.6) for the average range, the only difference being the factor $\cos\varphi$ in (4.1).

Let us consider some approximations which can be useful in solving (4.1). If always $T << E$, i. e. $\gamma << 1$, or if $\bar{R}_p$ is nearly proportional to E, we may write

$$1 = \bar{R}_{p1}(E)\, N\int d\sigma_{n,e}(1-\cos\varphi) + \frac{d\bar{R}_{p1}(E)}{dE} N\int d\sigma_{n,e}\cdot T\cdot\cos\varphi\,. \tag{4.2}$$

This approximation is similar to the one for $\bar{R}_1$ in (3.7) and (2.2), and we therefore use the notation $\bar{R}_{p1}$ for the projected range in (4.2). Actually, if the deflection φ may be neglected, we obtain $(d\bar{R}_{p1}/dE) = N\cdot S$, i. e. $\bar{R}_{p1}$ becomes equal to $\bar{R}_1$.

When solving (4.2) we can introduce the familiar transport mean free path, λ_{tr}, and a transport stopping cross section, S_{tr},

$$\frac{1}{\lambda_{tr}} = N\int d\sigma_{n,e}(1-\cos\varphi),\ S_{tr} = \int d\sigma_{n,e}\, T\cos\varphi\,. \tag{4.3}$$

With this notation, equ. (4.2) becomes

$$1 = \frac{\bar{R}_{p1}(E)}{\lambda_{tr}(E)} + \frac{d\bar{R}_{p1}(E)}{dE}\cdot NS_{tr}(E), \tag{4.4}$$

which equation (4.4) has the solution

$$\bar{R}_{p1}(E) = \int_0^E \frac{dE'}{NS_{tr}(E')} \exp\left\{\int_E^{E'} \frac{dE''}{\lambda_{tr}(E'')\, N\cdot S_{tr}(E'')}\right\}, \tag{4.5}$$

and this result should be a good approximation to $\bar{R}_p(E)$ if γ is small, or if R_p is nearly proportional to energy. We may solve the equation for $\bar{R}_p$ in the lowest approximation. This corresponds to taking the leading term in a series development in $\mu = M_2/M_1$, assuming μ to be small. The approximation is similar to that in § 3, for $\gamma << 1$. In the limit of small μ, the angle φ is always small and we need only include φ^2-terms in (4.3). Using

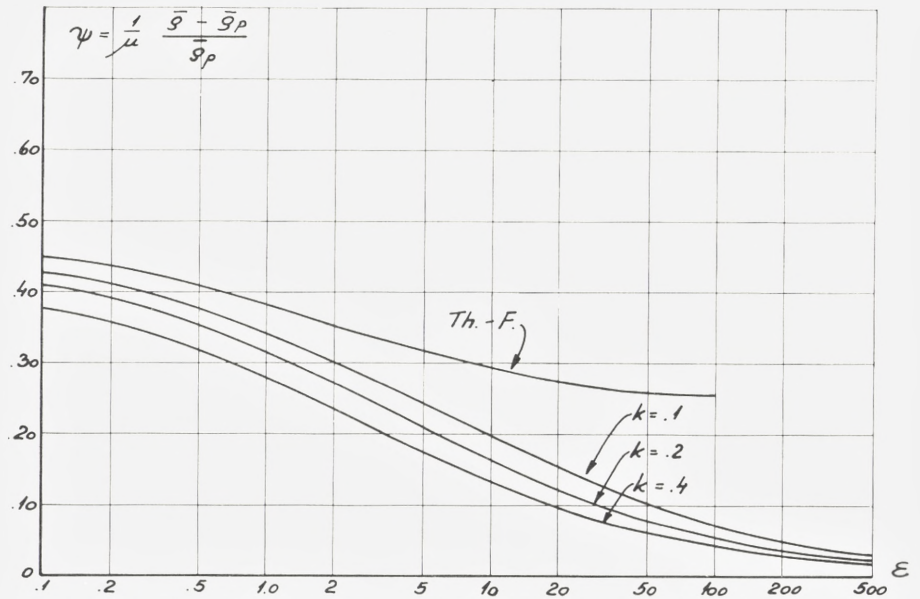

Fig. 8. Correction for projected ranges $(\bar{\varrho}_1 - \bar{\varrho}_{p1})/\bar{\varrho}_{p1} = \mu\psi$, to first order in the mass ratio $\mu = M_2/M_1$. Curves are shown for pure nuclear stopping and for three values of electronic stopping parameter k.

the nuclear scattering cross section (2.9) and electronic stopping $(d\varepsilon/d\varrho)_e = k \cdot \varepsilon^{1/2}$, we have computed the first order correction from average projected range to average range along the path. The resulting curves are shown in Fig. 8, for various values of k, and also for pure nuclear stopping.

It is more difficult to obtain accurate approximations to $\bar{R}_p$ when μ is large, corresponding to large angles of scattering, φ. We use the approximate equation (4.5) and profit from the circumstance that $\bar{R}_p$ is not far from being proportional to energy. By means of (2.9), solutions were obtained for $\mu = 1$ and $\mu = 2$, and a few representative values of the electronic stopping parameter k. The results are shown in Fig. 9.

The power law approximation of nuclear scattering, (2.6), with neglect of electronic stopping, permits accurate solutions for $\bar{R}_p$. We utilize the circumstance that $\bar{R}_p \propto E^{2/s}$. As an example, we consider the useful case of $s = 2$. The exact solution of (4.1) and (3.6) leads to (LINDHARD and SCHARFF (1961))

$$\bar{R}/\bar{R}_p = \frac{1}{4}\left\{-1-3\mu+(5+\mu)\frac{1+\mu}{2\,\mu^{1/2}}\arccos\frac{1-\mu}{1+\mu}\right\} \cong 1+\frac{1}{3}\mu, \quad s=2. \qquad (4.6)$$

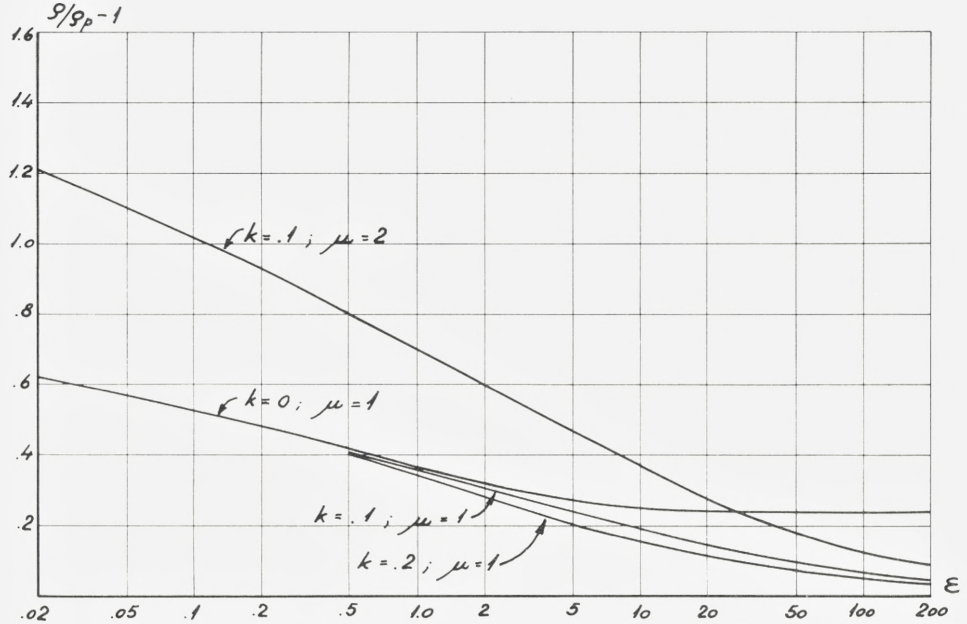

Fig. 9. Approximate curves for $\bar{\varrho}_1/\bar{\varrho}_{p1}$ for large values of the ratio $\mu = M_2/M_1$ and a few values of k.

As may be seen from Fig. 8, the rule-of-thumb $\bar{R}/\bar{R}_p = 1 + \mu/3$ is a fair approximation at low energies.

As a further example we may quote the value of $\bar{R}/\bar{R}_p$ for small μ, and any value of s,

$$\bar{R}/\bar{R}_p \simeq 1 + \mu \frac{s^2}{4\,(2\,s-1)}, \tag{4.7}$$

which approximation is quite accurate up to $\mu \sim 1$.

Associated range concepts

The average projected range is determined by one closed equation. However, the equations governing the higher moments of the projected range are far more complicated. If we treat the average square of the projected range, $\overline{R_p^2}$, we must also introduce the average square of the range projected on the plane perpendicular to the initial direction, $\overline{R_\perp^2}$. The average square of the distance between the starting-point and the end point of the path is then $\overline{R_c^2} = \overline{R_p^2} + \overline{R_\perp^2}$. We may describe R_c as the chord range (also referred to as vector range). These range concepts are illustrated in Fig. 10.

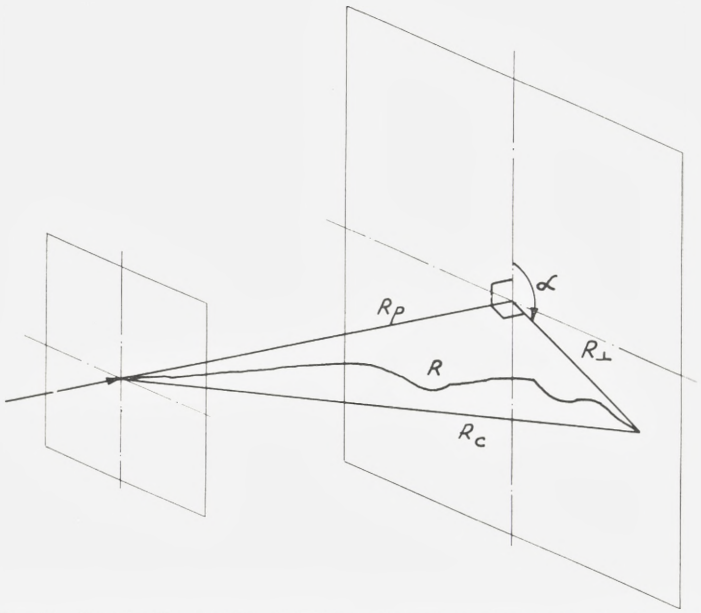

Fig. 10. Sketch illustrating definition of range concepts R, R_p, R_c and $R_\perp$.

The integral equations for $\overline{R_p^2}$ are derived in a similar way as (3.1). The following two equations are obtained, after rearrangement of terms,

$$2\,\bar{R}_p(E) = N\int d\sigma_{n,e}\left\{\overline{R_c^2}(E) - \overline{R_c^2}(E-T)\right\}, \tag{4.8}$$

$$2\,\bar{R}_p(E) = N\int d\sigma_{n,e}\left\{\overline{R_r^2}(E) - \left(1-\frac{3}{2}\sin^2\varphi\right)\overline{R_r^2}(E-T)\right\}, \tag{4.9}$$

where

$$\overline{R_c^2} = \overline{R_p^2} + \overline{R_\perp^2} \text{ and } \overline{R_r^2} = \overline{R_p^2} - \frac{1}{2}\overline{R_\perp^2}\,. \tag{4.10}$$

The two equations (4.8) and (4.9) may be solved separately, and then $\overline{R_p^2}$ is found from (4.10).

First order solutions of (4.8) and (4.9), for $\mu << 1$, can be obtained in a direct manner. However, we shall merely consider the case of power law scattering, with neglect of electronic stopping. The exact solutions may then be expressed as beta functions. In Table 3 we quote the results for $\mu = 1$ and various values of s. It is seen that in these cases $\overline{\Delta R_p^2}$ is of order of $\overline{\Delta R^2}$.

TABLE 3
Straggling in projected range for power law scattering and $\mu = 1$.

s	3/2	2	3
$\overline{\Delta R^2}/\overline{\Delta R_p^2}$	1.25	1.33	1.38
$\overline{\Delta R_p^2}/\overline{R_p^2}$	0.204	0.275	0.341

§ 5. Comparison with Experiments

As an illustration of the connection to experiments, we present a brief survey of recent experimental results, interpreted on the lines of the theory of this paper. Before that, it may be worth-while to summarize briefly and comment on the salient features of this theory.

A primary result is that a simple-minded theory of ranges and their fluctuations, as described in § 2, is quite accurate and that corrections of various kinds for projected ranges, etc., may be made without much difficulty, if necessary. A second result, somewhat independently of the details of the theory of collisions, is that a $\varrho - \varepsilon$ plot is useful for a study of ranges of particles with $\varepsilon < 1000$, and particularly for $\varepsilon \lesssim 10$. A third result is that for any ion of high energy a range correction, Δ, for the effect of nuclear stopping has been obtained, which permits a more accurate study of electronic stopping. Fourth, e. g. various isotope effects can serve to check several details of the theory, as may also observations of range straggling.

A theoretical result of special interest is that for $Z_1 = Z_2$ the electronic stopping constant is $k \sim 0.15$, except when $Z_1 = 1$. Therefore, the range energy curve for $Z_1 = Z_2$ should be closely a single curve in a $\varrho - \varepsilon$ plot. However, the corrections for e. g. projected ranges are not negligible in this case.

The numerical results computed here are based on a much simplified model of collisions. It is certainly possible to introduce a more detailed description of the collisions (cf. Notes on Atomic Collisions I and IV), and thereby improve on the present theoretical results. However, it may be more important to remove uncertainties and to correct misconceptions in the theory by measurements of range and stopping.

Another important circumstance is that direct comparisons with measured ranges may be made preferably in gases, where successive collisions are uncorrelated. In several respects stopping in solids may also answer the purpose, but experiments at low ion energies clearly seem to indicate the

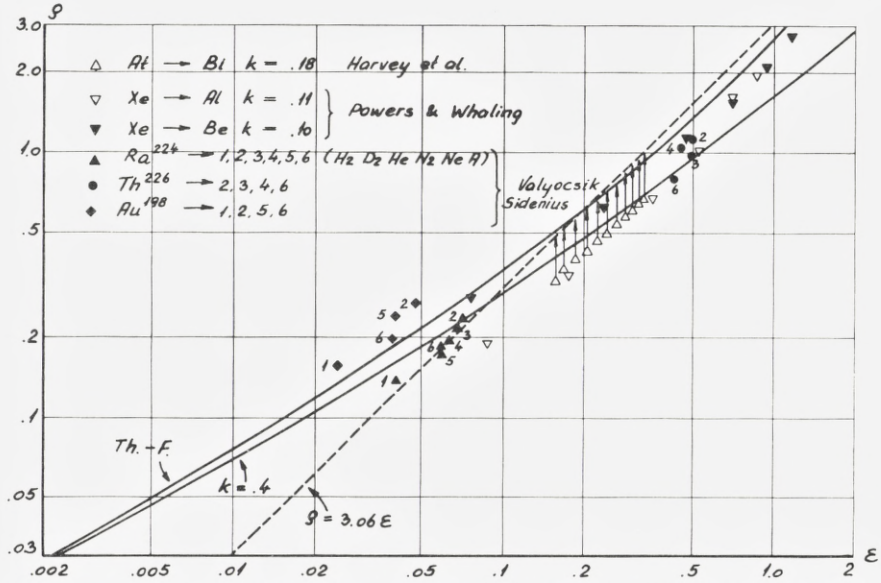

Fig. 11. Comparison between theoretical curves for $\bar{\varrho}_1(\varepsilon)$ given by (2.2), (2.9), (2.10) and (2.11), and measurements for $\varepsilon < 2$. As indicated on the figure, numbers 1, 2, 3, 4, 5 and 6 refer to stopping gases H_2, D_2, He, N_2, Ne and A, respectively. For further comments cf. text.

kind of correlation of collisions described as tunnelling (cf. PIERCY et al. (1963)), with strong directional effects and range lengthening in certain crystal structures. Although these range effects are in themselves highly interesting, their special character make them less suited in a general first comparison between range theory and experiments. In e. g. amorphous solids the effect appears to be absent, as was to be expected.

It should be appreciated that in the following we have merely made a compilation of measurements; not all of them are plotted in the figures. We are not in a position to make any critical examination of the experiments, some of which are in mutual disagreement or obviously inaccurate. We have included primarily the more recent measurements. A review of previous observations is given by HARVEY (1960). We are mainly interested in experiments where nuclear stopping is dominating, and do not discuss electronic stopping. NORTHCLIFFE (1963) has given a valuable survey of measurements on stopping in the energy region just above the one considered here, i. e. when electronic stopping dominates and goes through a maximum.

In plotting the results we have made approximate corrections for projected ranges, etc. Normally, the range measurements are plotted directly

on the figures, and range corrections are indicated by arrows. In some cases our knowledge of the measurements was too scanty to permit a range correction. As a general rule, we have corrected for projected ranges, etc., only if the correction exceeds ~10 percent.

Fig. 11 shows the theoretical range curve for values of ε smaller than 2, where nuclear stopping is quite dominating. The ranges for pure nuclear stopping are given by the upper solid curve, denoted as Th.-F. on the figure. A curve for exceptionally large electronic stopping, i. e. $k = 0.4$, is also shown. The actual k-values are quite small, and thus the expected ranges should be close to the Th.-F. curve. Further, note the dashed straight line corresponding to range proportional to energy, $\varrho = 3.06\varepsilon$. It should be emphasized that for extremely low energies, $\varepsilon \lesssim 10^{-2}$, the theoretical curve is not too well-defined.

Harvey, Wade and Donovan (1960) observed projected ranges for At^{205} and At^{207} ions in bismuth. The At recoil ions were produced by α-bombardment of a bismuth foil, leading to an (α, xn) process. This resulted in At ions with various energies between 400 and 900 keV; the energies were not sharply defined. Approximate corrections for projected range are shown by arrows in Fig. 11. The observations of Harvey, Wade and Donovan are in satisfactory accord with the predicted ranges.

Powers and Whaling (1962) studied projected ranges of monoenergetic ions of nitrogen and inert gases in several solids. The depth of penetration of the ions was obtained from a subsequent analysis of the distribution in angle and energy loss of protons scattered from the ions imbedded in the target. The ranges of Powers and Whaling are generally in good agreement with the theoretical curves. In the figure, we have included only their range measurements for Xe in Be and in Al. The corrections for projected ranges are quite small and are omitted. The ranges in Al may be compared with those of Davies et al. in Fig. 12. These two range observations for Xe in Al give quite different results and are placed on either side of the theoretical curve.

Valyocsik (1959) made accurate observations of ranges of Ra^{224} and Th^{226} recoil atoms with, respectively, 97 and 725 keV energies. Ranges are measured in gases using the electrostatic collection technique of Ghiorso and Sikkeland. Ranges and range stragglings were observed in deuterium, helium, nitrogen and argon, and in hydrogen and neon (only for Ra ions). The observations are shown in Fig. 11. They are in good agreement with theory (between 0 and 20 percent below theoretical ranges), and correspond to $k = 0.12$, except in hydrogen where $k = 0.16$.

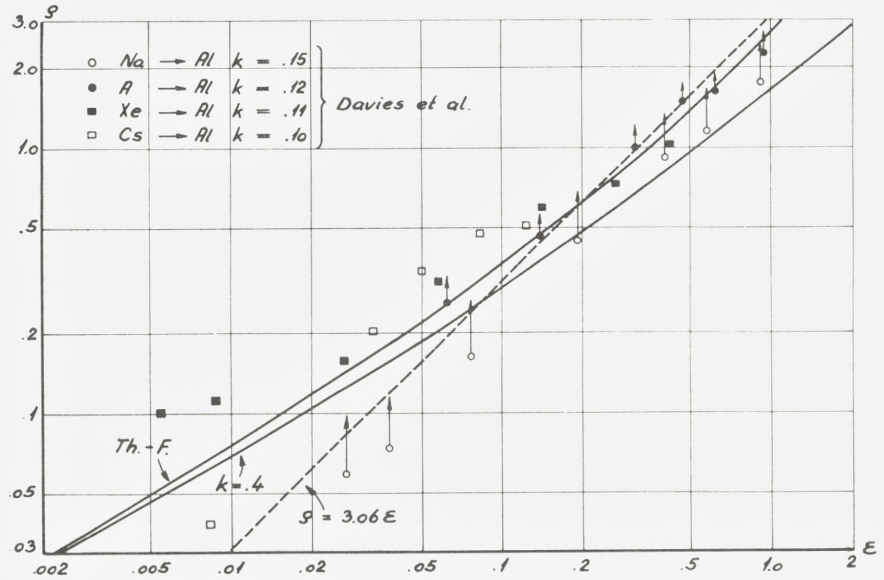

Fig. 12. As Fig. 11; measurements of median ranges by DAVIES et al. in Al. Ranges at low energies exceed theoretical curves, probably as an effect of tunnelling in crystal lattice.

A few measurements by the Copenhagen group (SIDENIUS, private communication) are also included in Fig. 11. The projected range of Au^{198} ions of energy 50 keV is measured by electrostatic collection. The correction for projected range is negligible. The ranges are slightly above theoretical curves. The k-values are as in VALYOCSIK's measurements.

DAVIES et al. (1960, 1961 and private communication) have observed projected ranges in Al, for the following ions: Na^{24}, A^{41}, K^{42}, Rb^{86}, Xe^{133} and Cs^{137}. Monoenergetic radioactive ions of energies between 1 keV and 2 MeV enter a polished Al surface. Thin layers of Al are removed successively by electro-chemical means and the residual activity is measured. In this way the distribution in projected range is obtained. The range values of DAVIES et al. in Fig. 12 are median ranges. At the higher energies there is good agreement with theoretical curves.

The measurements by DAVIES et al. were made with polycrystalline Al. It has turned out that the structure of Al is such that tunnelling of the ions may occur, whereby the average range becomes considerably larger than for a random system, and the range distribution has an exponential tail (PIERCY et al. (1963)). The results of PIERCY et al. for Kr^{85} in Al and Al_2O_3 at 40 keV are compared with theoretical estimates in Table 4. There is

TABLE 4

Ranges (in $\mu g/cm^2$) of 40 keV Kr^{85} in Al and Al_2O_3, and average square straggling in range. Experimental results by PIERCY et al. Computed results (columns 3 and 5) are for random system, as indicated.

	R_{exp}^{med}	$\bar{R}_{exp}$	$\bar{R}_{rand}$	$\overline{(\Delta R)^2_{exp}}$	$\overline{(\Delta R)^2_{rand}}$
Al.....	9.0	11.5	7.1	91	4.6
Al_2O_3.....	7.7	7.7	6.5	7.8	3.5

satisfactory agreement in the amorphous substance Al_2O_3, both as regards ranges and straggling. It appears also from Table 4 that the experimental median range in Fig. 12 is probably somewhat larger than the average ranges of a random system of Al atoms. We therefore infer that the results of DAVIES et al. in Fig. 12 are not in contradiction to the theoretical ranges of a random system. Note the very large experimental range straggling in Table 4 for Al, characteristic of an exponential distribution, where $\overline{\Delta R^2} = \bar{R}^2$.

There are several other measurements in the regions of energy corresponding to Figs. 11 and 12. Thus, BAULCH and DUNCAN (1957) obtain ranges of α-recoils ($\varepsilon \lesssim 0.1$) from 0 to 10 percent below theoretical curves. The results of VAN LINT et al. (1961) are at the higher energies at least about a factor of 2 above theoretical expectations, while at lower energies ($\varepsilon \sim 0.04$) agreement is fair. However, these measurements show a very considerable scatter. GUSEVA, INOPIN and TSYTKO (1959) measured ranges of monoenergetic Si^{30} ions in Ta and Cu backings, at energies between 10 and 25 keV. The depth of penetration was estimated from proton energies necessary for a (p, γ) process, together with knowledge of proton stopping. Their results are about a factor of 2 above the theoretical curves.

Fig. 13 shows some observations for $1 < \varepsilon < 100$, and corresponds to Fig. 4 in § 2. We are here in a region where the electronic stopping begins to take over. It is then important to know the value of the constant k. Some of the projected ranges observed by POWERS and WHALING (1962) are shown in Fig. 13, including one where the ratio $\mu = (M_2/M_1) \sim 2$, i. e. the corrections for projected range are large. The agreement with theoretical curves is good.

WINSBERG and ALEXANDER (1961) and ALEXANDER and SISSON (1962) measured projected ranges for Tb^{149} ions in aluminium, at energies between 4 and 30 MeV, and for At and Po ions in aluminium and gold, at energies between 3.5 and 13 MeV. The projected ranges and the range stragglings were obtained from the activities in stacks of catcher foils. In Fig. 13 we have included results for At and Po in gold and for Tb^{149} in

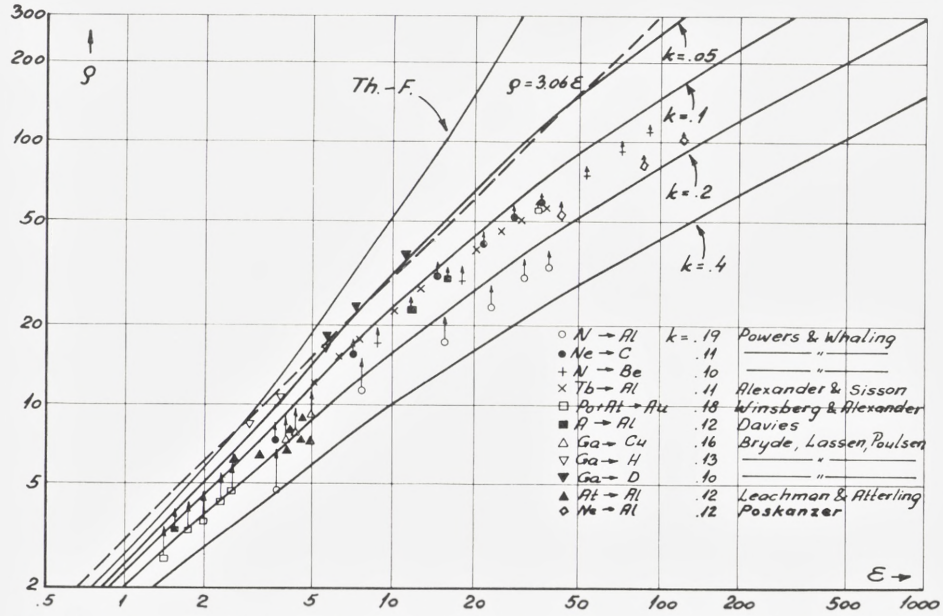

Fig. 13. Comparison with range measurements in the region $1 < \varepsilon < 100$, where electronic stopping becomes important. Theoretical k-values are given, indicating the theoretical curve with which to compare the observations.

aluminium. There is good agreement with the theoretical curves. It may be noted that the ions were formed in a nuclear reaction with subsequent neutron evaporation.

In the case of A^{41} in aluminium, DAVIES et al. (private communication) performed measurements at energies so high that electronic stopping is important. The ranges are in good agreement with the theoretical curves in Fig. 13.

BRYDE, LASSEN and POULSEN (1962) measured projected ranges for radioactive Ga^{66} recoil ions in gases using electrostatic collection. As typical representatives of their observations we have in Fig. 13 included ranges in hydrogen and deuterium. These ranges are about 40 percent above theoretical ranges. BRYDE, LASSEN and POULSEN also observed projected ranges for Ga^{66} in copper; the latter ranges are in good agreement with the theoretical curve. Also included in Fig. 13 are three measurements by POSKANZER (1963) of 1–3 MeV Ne^{22} ions in aluminium; these ranges are smaller than theoretical ranges. Finally, in Fig. 13 is shown the early measurements of ranges by LEACHMAN and ATTERLING (1957), where recoil ions of At^{203}

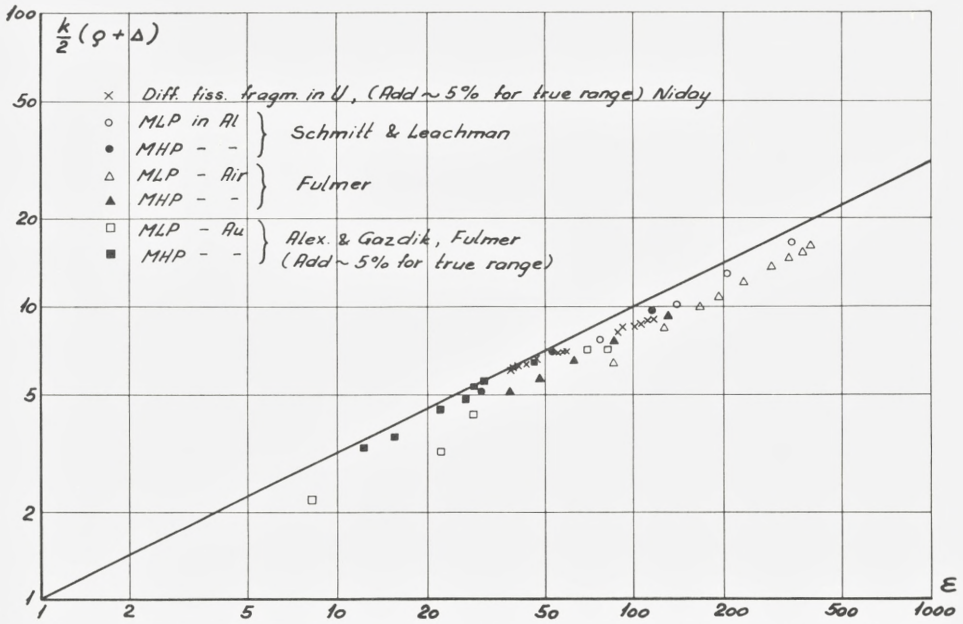

Fig. 14. Comparison between theoretical curve and range measurements for fission fragments, nuclear stopping being eliminated. For large values of ε the representation shown here is superior to that in Fig. 13.

and At^{205} penetrated a stack of aluminium foils, and projected ranges were measured. There is fair agreement, but apparently some fluctuations between individual measurements.

As mentioned previously, in the present paper we do not attempt a systematic study of electronic stopping as obtained from measurements at high values of ε. We may merely show two sets of representative measurements, where the nuclear stopping is eliminated, so that the extrapolated electronic range is obtained. For $v < v_1$ the theoretical extrapolated electronic range is $\varrho_e = 2\varepsilon^{1/2}/k$. Using theoretical range corrections for nuclear stopping, $\Delta(k, \varepsilon)$, as indicated in Fig. 5, we have plotted in Figs. 14 and 15 values of $(k/2)\{\varrho + \Delta(k, \varepsilon)\}$ obtained from measurements of ϱ. The theoretical curve is the straight line $k\varrho_e/2 = \varepsilon^{1/2}$. Fig. 14 contains only measurements of ranges of fission fragments. In Fig. 14 is shown measurements by Niday (1961) of fission fragment ranges in uranium. Niday used a thick uranium foil packed in aluminium catcher foils. Fission fragments resulted from thermal neutrons. The fragments ending up in aluminium were separated by radiochemical means. In this way an estimate of the ranges along the

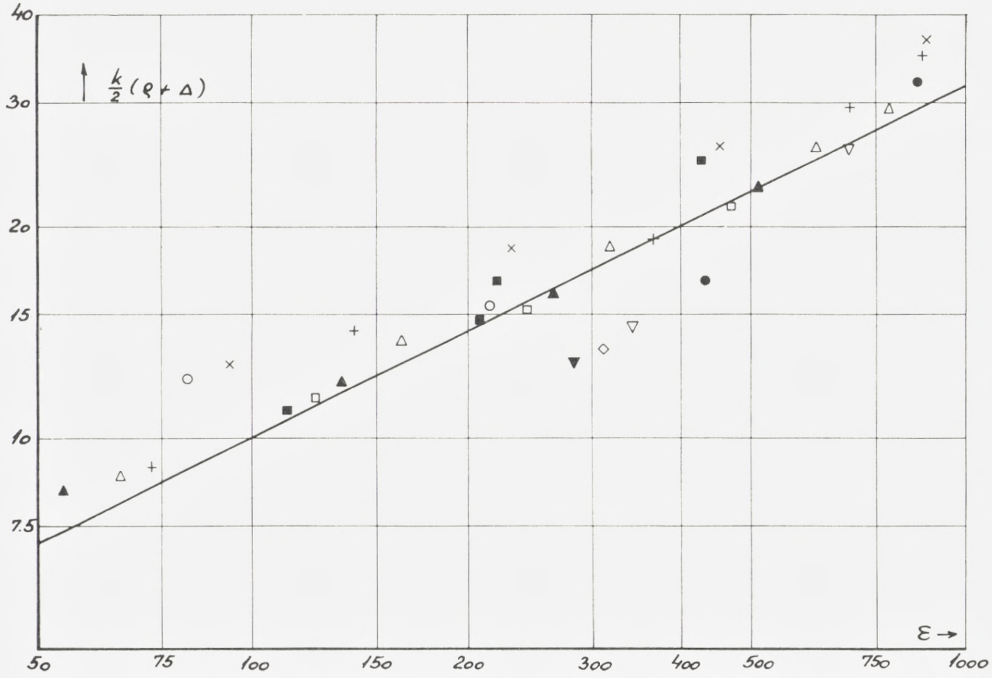

Fig. 15. Some recent measurements of projected ranges for light atoms in gases, corrected for nuclear stopping only, like in Fig. 14. Full-drawn curve is theoretical range $\varepsilon^{1/2}$. Points stand for following ions in air: × Li, + B, △ C, ▲ O, □ F, ■ Ne, ○ Na, and following ions in argon: ● Li, ▽ B, ▼ N (measurements by TEPLOVA et al.). Further, ◇ indicates F in nitrogen, measured by BRYDE, LASSEN and POULSEN.

chord was obtained. The ranges of NIDAY should be corrected by approximately +5 percent in order to obtain true ranges. The agreement with the theoretical range is good.

In Fig. 14 is also included observations on fission fragment ranges by ALEXANDER and GAZDIK (1960), FULMER (1957) and LEACHMAN and SCHMITT (1954). In the case of gold, about 5 percent should be added in order to obtain true ranges. There is agreement within ~10 percent.

A number of other authors have measured ranges of fission fragments (SMITH and FRANK (1959), KATCOFF, MISKEL and STANLEY (1948), GOOD and WOLLAN (1956), BØGGILD, ARRØE and SIGURGEIRSSON (1947), DOUTHETT and TEMPLETON (1954), SUZOR (1949), PORILE and SUGARMAN (1957), cf. also the review article by HARVEY (1960)). Some of the earlier measurements may be less accurate than those shown in Fig. 14, but generally there is approximate agreement with theory.

As an example of light ions with substantial energies we have taken measurements of projected ranges by TEPLOVA et al. (1962). A number of ions, from Li to Na, with energies in the interval 1–10 MeV, were slowed down in air, argon and hydrogen. Many of these measurements are shown in Fig. 15. On the figure is also shown a range value for F^{18} in nitrogen gas, measured by BRYDE, LASSEN and POULSEN (1962). We have not indicated corrections for projected ranges on Fig. 15, since the largest correction would be $\sim +8$ percent (for Li in argon gas).

In connection with electronic stopping it should be noted that at low atomic numbers, and particularly at low values of Z_1, there may be deviations from the theoretical k-value based on a Thomas-Fermi treatment. At low atomic numbers one may expect variations in the measured k-values due to shell effects. As an extreme example from a Thomas-Fermi point of view, in the case of Li ions in hydrogen, deuterium and helium, it appears from measurements of stopping (ALLISON and LITTLEJOHN (1957)) and of ranges (CLERC, WÄFFLER and BERTHOLD (1961)) that the electronic stopping may be as much as 2–3 times less than given by (2.5). Measurements by ORMROD and DUCKWORTH (1963) of electronic stopping in carbon for all ions with $Z_1 \lesssim 11$ indicate minor shell variations around the value in (2.5).

Range straggling

As to straggling in range (cf. p. 14) we have not attempted any closer analysis. High accuracy is difficult to obtain in range straggling, and at low ε-values ($\varepsilon < 0.5$) the rule-of-thumb $(\Delta\varrho/\varrho)^2 = \gamma/6 = M_1 M_2 (M_1 + M_2)^{-2} \cdot (2/3)$ is often sufficient. In many experiments a considerable fluctuation was present in the initial ion beam, e. g. because the ion resulted from a compound nucleus after neutron evaporation. The experimental range stragglings are often considerably above the curves. The measurements by VALYOCSIK on 97 keV α-recoils (cf. HARVEY (1960)) correspond to rather well-defined conditions. For 97 keV Ra the straggling in nitrogen, neon and argon is comparable with the theoretical one (cf. Fig. 6), but in the light gases, hydrogen, deuterium and helium, the straggling is much in excess of theoretical estimates. When subtracting a common constant of order of 0.016 from the experimental straggling $(\Delta\varrho)^2_{exp}$, one obtains a relative straggling $\gamma^{-1}(\Delta\varrho/\varrho)^2 \cong 0.14$–$0.18$, in excellent agreement with theory (since $\varepsilon \approx 0.03$–0.07, and $k \cong 0.12$). For 725 keV Th ions, where $\varepsilon \approx 0.4$–0.5, the experimental relative straggling is much too large in deuterium and helium. A reduction of $(\Delta\varrho)^2_{exp}$ by ≈ 0.04 in all gases would give a reasonable order of magnitude of the straggling. As a further example, many measurements by the Copen-

hagen group show rather large straggling effects, but some results (e. g. ranges of 50 keV Ga^{66} in hydrogen, helium, nitrogen and argon, shown in Fig. 11) with $\varepsilon \approx 0.3$–0.5, have a straggling $(\Delta\varrho/\varrho)^2\,\gamma^{-1} \approx 0.15$–$0.25$. Even in the difficult case of the lightest gases, where the theoretical straggling is extremely small, there is reasonable accord with theory.

Isotope effects

It is of interest to study isotope effects in range measurements. We shall treat the question of different isotopes used as stopping medium*. Although electronic stopping may dominate in the value of the range itself, isotope effects can still give direct information about the nuclear stopping. An instructive example is provided by the measurements of Bryde, Lassen and Poulsen (1962, and private communication). They observed ranges of Ga^{66} in hydrogen and deuterium; at high energies R_D is slightly larger than R_H, while at low energies R_H exceeds R_D. Now, if there was only electronic stopping, the two ranges would be equal, so that differences are due to nuclear stopping. It is seen from (2.7) that the nuclear stopping behaves as $S_n \propto M_2^{1-2/s}$, when $M_1 >> M_2$. At quite low energies, where the ion cannot penetrate deeply into the atom, the effective power of the potential is of order of $s = 3$, and thus $S_{nD} > S_{nH}$. At high energies, where the screening is weak, the effective power approaches $s = 1$, and therefore $S_{nH} > S_{nD}$ (Lindhard and Scharff (1961)). According to Fig. 2, the change-over in stopping occurs at an ε-value smaller than 0.5. Correspondingly, in Fig. 4 the change-over in slope—from lower to higher than that of the straight dashed line—occurs at $\varepsilon \sim 1$ for the Th.-F. curve.

Instead of this qualitative explanation of experimental results we may directly compare experimental range differences with theoretical ones deduced from Figs. 3 and 4. The results are shown in Table 5. Agreement between theoretical and experimental range differences is quite good,

Table 5

Differences between ranges in D_2 and H_2 for Ga^{66} ions. Ranges are in mm at 300° K, 760 mm Hg.

Energy (keV)	1190	790	610	50
$(R_D - R_H)_{th}$	0.9	0.7	0.6	−0.05
$(R_D - R_H)_{exp}$	1.5	0.8	0.5	−0.05

* A measurement, where different isotopes are chosen for the incoming particle, is discussed by Lindhard and Scharff (1961).

especially at the lower energies. This result is obtained in spite of the fact that at the three higher energies the absolute ranges of BRYDE, LASSEN and POULSEN are as much as ~40 percent higher than theoretical ranges (Fig. 13).

In further measurements by the Copenhagen group (SIDENIUS, private communication), other examples of isotope effects were obtained for 50 keV ions. Thus, for Na^{24} in hydrogen and deuterium ($\varepsilon = 2.4$ and 4.65) one found $(R_D - R_H)_{exp} = +0.157$ mm, while $(R_D - R_H)_{th} = +0.104$ mm, the ranges themselves being of order of 0.9–1.0 mm, and ~50 percent larger than theoretical ranges. For Au^{198} ions in hydrogen and deuterium, ε is so small ($\varepsilon = 0.024$ and 0.047) that the effective power has shifted to $s > 2$, and $(R_D - R_H)_{exp} = -0.061$ mm, while $(R_D - R_H)_{th} = -0.087$ mm; experimental ranges are ~ 0.4 mm, i. e. about 30 percent larger than theoretical ranges. Finally, for Ga^{66} in helium isotope gases ($\varepsilon \sim 0.4$) one found $(R_{He^4} - R_{He^3})_{exp} = -0.016$ mm, to be compared with $(R_{He^4} - R_{He^3})_{th} = -0.006$ mm; experimental ranges are ~0.4 mm, or 20 percent above theoretical ranges. All ranges quoted here are in mm at 300° K, 760 mm Hg. The agreement with theoretical isotope shifts of ranges is thus fairly good, and it is interesting that normally the change from larger to shorter range in the heavier isotope occurs at $\varepsilon \sim 1$.

Acknowledgments

A few of the above results were obtained seven years ago, following discussions with Dr. R. B. LEACHMAN on his observations of range distributions. They have been referred to various times in the literature. A brief summary of the present work (LINDHARD and SCHARFF (1961)) was published at the time of MORTEN SCHARFF's death.

We are much indebted to Drs. R. B. LEACHMANN, J. M. ALEXANDER, B. G. HARVEY, N. O. LASSEN, N. O. ROY POULSEN, W. WHALING, J. A. DAVIES, H. E. DUCKWORTH, Mr. G. SIDENIUS and many others for discussions and communication of experimental results prior to publication.

We wish to express our gratitude to all who have encouraged us and assisted in this work, in particular to P. V. THOMSEN, M. Sc.

We are much indebted to Miss S. TOLDI and Mrs. A. GRANDJEAN for assistance in the preparation of the paper.

Institute of Physics,
University of Aarhus.

References

J. M. Alexander & M. F. Gazdik (1960). Recoil Properties of Fission Products. Phys. Rev. **120**, 874.

J. M. Alexander & D. H. Sisson (1962). Recoil Range Evidence for the Compound-Nucleus Mechanism in Reaction between Complex Nuclei. UCRL-10098.

S. K. Allison & C. S. Littlejohn (1957). Stopping Power of Various Gases for Li Ions of 100–450 keV Kinetic Energy. Phys. Rev. **104**, 959.

D. L. Baulch & J. F. Duncan (1957). The Range-Energy-Relation for α-Recoil Atoms. Austral. J. Chem. **10**, 112.

N. Bohr (1948). Penetration of Atomic Particles through Matter. Mat. Fys. Medd. Dan. Vid. Selsk. **18**, no. 8.

L. Bryde, N. O. Lassen & N. O. Roy Poulsen (1962). Ranges of Recoil Ions from α-Reactions. Mat. Fys. Medd. Dan. Vid. Selsk. **33**, no. 8.

J. K. Bøggild, O. H. Arrøe & T. Sigurgeirsson (1947). Cloud-Chamber Studies of Electronic and Nuclear Stopping of Fission Fragments in Different Gases. Phys. Rev. **71**, 281.

H. G. Clerc, H. Wäffler & F. Berthold (1961). Reichweite von Li^8-Ionen der Energie 40–450 keV in H_2, D_2 und He. Zeitsch. f. Naturf. **16a**, 149.

J. A. Davies, J. D. McIntyre, R. L. Cushing & M. Lounsbury (1960). The Range of Alkali Metal Ions of Kiloelectron Volt Energies in Al. Can. J. Chem. **38**, 1535.

J. A. Davies & G. A. Sims (1961). The Range of Na^{24} Ions of Kiloelectron Volt Energies in Al. Can. J. Chem. **39**, 601.

J. A. Davies, J. D. McIntyre & G. A. Sims (1961). Isotope Effects in Heavy Ion Range Studies. Can. J. Chem. **39**, 611.

B. Domeij, I. Bergström, J. A. Davies & J. Uhler (1963). A Method of Determining Heavy Ion Ranges by Analysis of α-Line Shapes. To appear in Arkiv f. Fysik.

E. M. Douthett & D. H. Templeton (1954). The Ranges of Fragments from High Energy Fission of Uranium. Phys. Rev. **94**, 128.

A. Erdélyi, W. Magnus, F. Oberhettinger & F. G. Tricomi (1953). Higher Transcendental Functions, I. McGraw-Hill.

U. Fano (1953). Degradation and Range Straggling of High Energy Radiations. Phys. Rev. **92**, 328.

C. B. Fulmer (1957). Scintillation Response of CsI(Tl) Crystals to Fission Fragments and Energy vs. Range in Various Materials for Light and Heavy Fission Fragments. Phys. Rev. **108**, 1113.

W. M. Good & E. O. Wollan (1956). Range and Range Dispersion of Specific Fission Fragments. Phys. Rev. **101**, 249.

M. I. Guseva, E. V. Inopin & S. P. Tsytko (1959). Depth of Penetration and Character of Distribution of Atoms Injected into Si^{30} Isotope Targets. Sovj. Phys. JETP **9**, 1.

B. G. Harvey (1960). Recoil Techniques in Nuclear Reaction and Fission Studies. Ann. Rev. of Nucl. Sci. **10**, 235.

B. G. Harvey, P. F. Donovan, J. R. Morton & E. W. Valyocsik (1959). Range Energy Relation for Heavy Atoms. UCRL-8618.

B. G. Harvey, W. H. Wade & P. F. Donovan (1960). Recoil Studies of Heavy Element Nuclear Reactions, II. Phys. Rev. **119**, 225.

D. K. Holmes & G. Leibfried (1960). Radiation Induced Primary Knock-Ons in the Hard Core Approximation. J. Appl. Phys. **31**, 1046.

D. K. Holmes (1962). The Range of Energetic Atoms in Solids. "Radiation Damage in Solids", vol. I, IAEA, Vienna.

F. Joliot (1934). Ranges of α-Recoils in Cloud Chamber. J. Phys. Rad. (7), **5**, 219.

S. Katcoff, J. A. Miskel & C. W. Stanley (1948). Ranges in Air and Mass Identification of Plutonium Fission Fragments. Phys. Rev. **74**, 631.

R. B. Leachman & H. W. Schmitt (1954). Fine Structure in the Velocity Distribution of Slowed Fission Fragments. Phys. Rev. **96**, 1366.

R. B. Leachman & H. Atterling (1957). Nuclear Collision Stopping of Astatine Atoms. Arkiv f. Fysik **13**, 101.

J. Lindhard & M. Scharff (1961). Energy Dissipation by Ions in the keV Region. Phys. Rev. **124**, 128.

J. Lindhard, V. Nielsen, M. Scharff & P. V. Thomsen (1963). Integral Equations Governing Radiation Effects. Notes on Atomic Collisions, III. Mat. Fys. Medd. Dan. Vid. Selsk. **33**, no. 10.

V. A. J. van Lint, R. A. Schmitt & C. S. Suffredini (1961). Range of 2–60 keV Recoil Atoms in Cu, Ag, Au. Phys. Rev. **21**, 14517.

J. B. Niday (1961). Radiochemical Study of the Ranges in Metallic Uranium of the Fragments from Thermal Neutron Fission. Phys. Rev. **121**, 1471.

K. O. Nielsen (1956). The Range of Atomic Particles with Energies about 50 keV. "Electro-Magnetically Enriched Isotopes and Mass Spectrometry". Proc. 1955. Harwell Isotope Conf.

L. C. Northcliffe (1963). Passage of Heavy Ions through Matter. To appear in Ann. Rev. Nucl. Sci.

J. H. Ormrod & H. E. Duckworth (1963). Stopping Cross-Sections in Carbon for Low Energy Atoms with $Z \leq 12$. To appear in Canad. J. Phys.

G. R. Piercy, F. Brown, J. A. Davies & M. McCargo (1963). An Experimental Study of a Crystalline Structure on the Ranges of Heavy Ions. Phys. Rev. Letters **10**, 399.

N. T. Porile & N. Sugarman (1957). Recoil Studies of High-Energy Fission of Bi and Ta. Phys. Rev. **107**, 1410.

A. M. Poskanzer (1963). Range of 1–3 MeV Ne^{22} Ions in Al and the Analysis of Some Na^{24} Recoil Data. Phys. Rev. **129**, 385.

D. Powers & W. Whaling (1962). Range of Heavy Ions in Solids. Phys. Rev. **126**, 61.

M. T. Robinson, D. K. Holmes & O. S. Oen (1961). Monte Carlo Calculation of the Ranges of Energetic Atoms in Solids. ORNL-3212.

R. A. Schmitt & R. A. Sharp (1958). Measurement of the Range of Recoil Atoms. Phys. Rev. Letters **1**, 12.

F. Seitz (1949). On the Disordering of Solids by Action of Fast Massive Particles. Disc. Far. Soc. **5**, 271.

E. R. Smith & P. W. Frank (1959). Recoil Range of Fission Fragments in Zirconium. WAPD-TM-198.

F. Suzor (1949). Tracks in Various Materials of Uranium Fission Fragments. Ann. Phys. **4**, 269.

Ya. A. Teplova, V. S. Nikolaev, I. S. Dmitriev and L. N. Fateeva (1962). Slowing Down of Multicharged Ions in Solids and Gases. Sovj. Phys. JETP **15**, 31.

E. W. Valyocsik (1959). Range and Range Straggling of Heavy Recoil Atoms. UCRL-8855.

W. Whaling (1958). The Energy Loss of Charged Particles in Matter. Hdb. d. Phys. vol. **34**, 193.

A. van Wijngaarden & H. E. Duckworth (1962). Energy Loss in Condensed Matter of H^1 and He^4 in the Energy Range 4 to 30 keV. Can. J. Phys. **40**, 1749.

L. Winsberg & J. M. Alexander (1961). Ranges and Range Straggling of Tb^{149}, At and Po. Phys. Rev. **121**, 518.

Indleveret til Selskabet den 24. april 1963.
Færdig fra trykkeriet den 17. oktober 1963.

Matematisk-fysiske Meddelelser
udgivet af
Det Kongelige Danske Videnskabernes Selskab
Bind **33,** nr. 15

Mat. Fys. Medd. Dan. Vid. Selsk. **33,** no. 15 (1964)

MESON PRODUCTION AT HIGH ENERGIES AND THE PROPAGATION OF COSMIC RAYS THROUGH THE ATMOSPHERE

BY

YASH PAL AND B. PETERS

København 1964
Kommissionær: Ejnar Munksgaard

Synopsis

An analysis of the intensity, charge composition and energy spectra of secondary cosmic ray components suggests that the average incident extraterrestrial nucleon, in its passage through the atmosphere, is repeatedly excited to one of the low lying levels of the pion-nucleon system and decays between successive excitations by the emission of several mesons. The particle distribution in the atmosphere in the energy range between a few GeV to ~ 1000 GeV can be understood and calculated accurately in terms of such decay products and their progeny; other processes of meson production play only a minor role.

The experimental data available on secondary cosmic radiation determine approximately the properties of the excited "average" baryon state, rather similar to those known to exist at accelerator energies

Probability of excitation	$s = 0.7 \pm 0.07$
Isobar mass	$M_B \geqslant 2300$ MeV
Average number of pions emitted per decay	$n_B = 3.5 \pm 0.5$
Average charge excess among decay pions	$\lvert\pi^+ - \pi^-\rvert = 0.35 \pm 0.15$
Ratio of hyperons to nucleons among the decay products	$Y/N = 7 \pm 7\ ^0/_0$

PRINTED IN DENMARK
BIANCO LUNOS BOGTRYKKERI A-S

CONTENTS

I. Introduction

In this paper an attempt is made to describe the propagation of cosmic radiation in terms of a phenomenological model for high energy nuclear interactions, which is in harmony with ideas on the process of particle creation as evolved in accelerator laboratories and at the same time simple enough for accurate calculations of intensity, charge, and energy distribution of secondary particles in the atmosphere.

A new investigation of this problem seemed to be required in view of the fact that, in recent years, a number of structural features corresponding to definite quantum states have been observed in the pion-nucleon system, which are known to play an important role at low energies and may do so also at high energies. Upto 3 GeV, single and double pion production in nucleon-nucleon collisions can be understood in terms of the excitation and subsequent deexcitation of such pion-nucleon isobaric states.[1] This mode of description, which has been developed largely by LINDENBAUM and STERNHEIMER[2], is known as the isobar model of meson production. Recently, DAMGAARD and HANSEN[3] have presented evidence that the same process may account for the majority of particles created by 22 GeV protons.

At very high energies ($\geqslant$ 100 GeV), on the other hand, a study of the energy and angular distribution of the great majority of particles requires a different mechanism for particle creation; it suggests a kind of fireball model[28], e. g., it can be described in terms of nearly isotropic emission of low energy particles from a cloud which is approximately at rest in the centre of mass system of the colliding nucleons (see Appendix A). Cosmic ray evidence indicates that the colliding nucleons themselves do not form part of this fireball; their energy is high in the C-systems, even after collision. This can be deduced from the propagation of nucleons through the atmosphere; in the majority of encounters, a nucleon emerges which retains a large fraction of the original energy.

However, it appears that nucleons are not the only particles which move with high velocity in the centre-of-mass system after collision; there is strong evidence that, at least in a considerable fraction of collisions, a

small number of pions is generated with energies high in the rest system of the fireball and low in the rest system of one of the baryons.[4]. It seems natural to expect strong final-state interaction between these mesons and the nucleon so that one may describe their creation as being the result of the deexcitation of a baryon isobar. The term baryon isobar will be used only in this sense. One can then divide the particles created in high energy nucleon-nucleon encounters into two phenomenologically distinct groups:

the fireball (which contains the majority of particles, all relatively slow in the centre-of-mass system), and

two other sets, each containing a nucleon as well as a small number of mesons, which are slow in the baryon system and therefore dynamically related to the nucleon.

As the primary energy increases the multiplicity, and hence the size of the fireball, is found to increase. On the other hand, the number of mesons from deexcitation of isobaric states appears to remain constant, suggesting an essentially energy independent mass distribution of the excited baryon states.

This general picture, which appears to be in accord with the energy distribution among particles emerging from high energy collisions[3, 4, 5], receives strong support from other cosmic ray data; as shown below, the positive to negative ratio among sea level muons and among kaons observed in balloon exposed emulsion stacks provides evidence in favour of the frequent excitation of baryons in high energy collisions.

a) The Ratio of Positive to Negative Muons.

The ratio μ^+/μ^- measured on the surface of the earth is nearly five to four at low energies and remains constant (or possibly increases) for muon energies above 100 GeV (see Fig. 8). Since the available target nuclei in the atmosphere contain protons and neutrons in equal numbers, charge symmetry requires that the observed positive excess among muons be due to the excess of protons over neutrons in the primary cosmic radiation, corresponding to an average charge excess of 0.37* positive charges per collision. The relations $\frac{\mu^+}{\mu^-} \approx \frac{5}{4}$ and $\mu^+ - \mu^- \leqslant 0.37$ show that the muons measured on the ground are descendants of a subgroup (consisting of about 3 mesons) whose members satisfy each of two conditions:

* At a given energy per nucleon, the extra-galactic nucleons consist of 87 % protons and 13 % neutrons[6], corresponding to an average charge excess of 0.37 per primary interaction.

they have a preferential share in the charge excess of the incident nucleon, and

they receive an abnormally high energy so as to remain distinct among the bulk of particles created in the same interaction.

Since the observed ratio μ^+/μ^- remains essentially constant over a considerable energy range, the character of the subgroup should not depend on primary energy.

The muons in this subgroup come from parent particles which receive an energy proportional to that of the incident nucleon; this can be deduced from the fact that their energy spectrum (when corrected for their decay and for the interaction of parent particles) follows the same power law as the primary and secondary nucleon components (Figs. 2 and 5).

All these conditions are satisfied automatically if one assumes that the incident baryon, emerging from a nuclear collision in the atmosphere, finds itself some of the time in an excited state from which it returns to the nucleon ground state by meson emission.

b) The Ratio of Positive to Negative Kaons.

Cosmic ray produced kaons observed in emulsion stacks furnish quite independently another indication of the existence of baryon excitation in high energy collisions. Stopping kaons show an uncommonly high positive to negative ratio $K^+/K^- \approx 20$ [7]. The observations correspond, in the mirror system of nucleon-nucleon collisions, to kaons which receive more than $25\,\%$ of the energy of the incident primary. Such a high fractional energy is normal if a kaon arises from the deexcitation of the forward isobar. The large positive excess follows directly from the assumption that most of the excited isobars have strangeness number zero:

> Positive kaons should then be produced in processes of the type $N^* \to K + Y$, $(M > 1610$ MeV$)^\dagger$; kaons of negative strangeness should be much rarer because in non-strange baryon decay they can occur only as members of kaon-antikaon pairs and have much narrower production channels:
>
> $N^* \to (K\overline{K}) + N$ $\quad (M > 1960$ MeV) or
> $N^* \to K + Y^{**} \to K + \overline{K} + N$ $\quad (M > 2010$ MeV).

In view of all these considerations, it seems useful to adopt a dual picture of particle generation in the high energy range relevant to most cosmic ray

† There is now strong evidence that this process contributes appreciably to positive kaon production at accelerator energies.[3]

phenomena and investigate its consequences. We consider a very simple model which incorporates the two distinct processes of meson production. The fact that it represents existing data adequately seems to point to an underlying simplicity in the high energy collision process itself. (The model is more fully discussed in Appendix A, where it will also be shown that it is compatible with all well established experimental data on high energy interactions.) The main features of the meson production processes may be summarized as follows:

1) A fairly isotropic emission of mesons from a fireball, moving with small, i. e. non-relativistic velocity in the C-system of a nucleon-nucleon collision. The number of these "pionization" mesons increases with energy, but their C-energy does not (in conformity with the well-known energy independence of transverse momenta); as a result, their energy in the L-systems is *proportional to the square root of the primary energy.*
2) Emission of mesons from various excited baryon states whose nature is independent of collision energy above $\sim$ 10–15 GeV. In the L-system the energy of decay mesons associated with the forward moving baryon is therefore *proportional to primary energy* (that of the backward moving baryon is non-relativistic and essentially independent of primary energy).

The relative importance of the two processes depends on the phenomenon to be studied. In this paper we confine the investigation to the intensity and energy distribution of secondary particles in the atmosphere, i. e. to the combined effect of collisions produced by primaries whose energy distribution follows the well-known and rather steep power law. In this particular case a simplification arises for purely kinematical reasons, namely:

In the case of nucleon-nucleon collisions the two processes of particle generation cannot both contribute comparable numbers to the flux of the secondary particles in the cosmic radiation; this is a consequence of the fact that the spectra of mesons from the two processes have very different dependence on primary energy and that the steepness of the primary cosmic ray spectrum emphasizes those processes in which a large fraction of the primary energy is transferred to individual secondary particles. The calculation of the relative importance in the cosmic radiation of mesons generated by nucleons in pionization and mesons generated in the deexcitation of baryon states is straightforward on the basis of this model; it is carried out in Appendix A and the result is shown in Fig. 1. One sees that, for

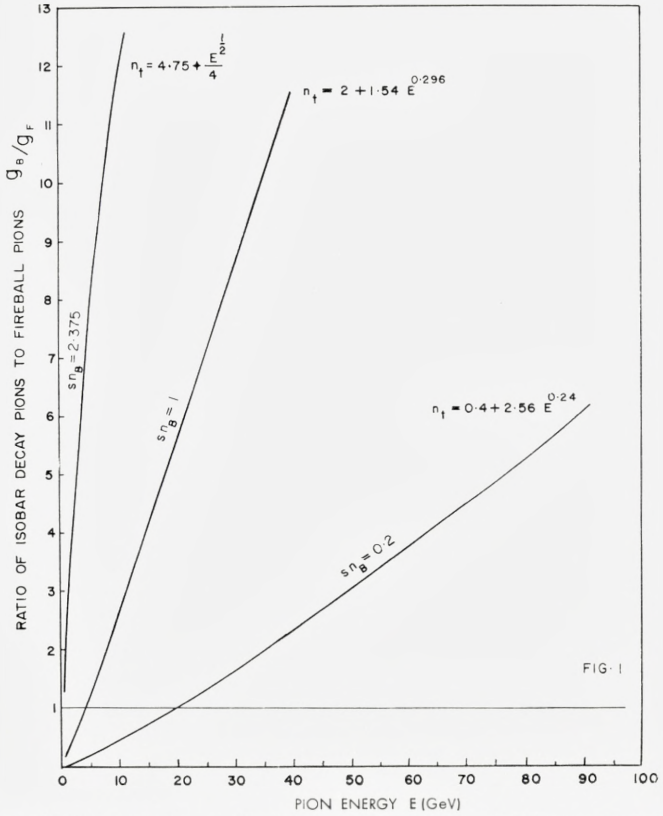

Fig. 1. Pion production from the decay of baryon isobars is compared with pion production from evaporation of a fireball (pionization). The ratio g_B/g_F (eqs. A 7, A 6) has been plotted for various values of the number of isobar decay pions, sn_B, which are produced in an average collision with target nuclei of low atomic weight. The total number of created particles per collision, n_t, is assumed to be related to that of charged particles, n_c, by $n_t = 1.6\,(n_c - 1) = 2\,sn_B + n_0\,E^{\varrho}$. The constants n_0 and ϱ have been chosen such that

$$n_t = \begin{cases} 6 \text{ for } E = 25 \text{ GeV} \\ 18 \text{ for } E = 2700 \text{ GeV} \end{cases}^{(24)}.$$

purely kinematical reasons, the mesons from the deexcitation of baryon isobars account for almost all particles above a few GeV even if the probability of excitation is fairly small.

We have, therefore, a situation which at first sight seems paradoxical: inspite of the fact that pions from the pionization process are more numerous than decay mesons in individual high energy collisions, *the secondary cosmic rays* observed in the atmosphere, at the surface of the earth or below

ground, *represent a reasonably pure sample of the decay products of nucleon isobaric states and their progeny.*

In view of this general consequence of the assumption of isobar excitation at high energy, namely the preponderance of its deexcitation products in the atmosphere, we calculate in Section II the flux of different components of secondary cosmic radiation arising from the production and decay of isobars and neglect in the case of nucleon collisions the contribution from the pionization process, except in so far as it represents a source of energy loss for the nucleon component in the atmosphere.

The creation of particles by the collision of pions in air is included. Since the nature of these collisions is not well known, a parameter is introduced which describes essentially the degree of elasticity in pion-nucleon interactions. Except at large depths in the atmosphere, the generation of pions by pions plays only a minor role in the secondary cosmic radiation.

The charge composition p/n, $\pi^{\pm}/p$ and the ratio μ^{+}/μ^{-} as a function of the properties of the excited isobaric states are discussed in Section III.

In both Sections II and III the calculations have been carried out for a single hypothetical isobar of "average" properties. The generalization to an arbitrary set of isobars decaying in an arbitrary manner is carried out in Appendix B. With the appropriate interpretation of symbols, the formulae derived in Sections II and III remain valid for the general case.

In Section IV it is shown that one obtains a very close agreement between measured and calculated spectra of nucleons and muons over the entire energy range above a GeV by a straightforward application of this simple model, and that one obtains the experimental ratio of pions to nucleons in the lower atmosphere provided one assumes that the collisions of pions with air nuclei are essentially inelastic.

In Section V we discuss the conditions which the excitation and deexcitation of baryon states have to satisfy in order to reproduce measurements on the secondary component of cosmic radiation with regard to absolute intensity, energy distribution, and charge composition.

The results of this investigation are summarized in Section VI. The applicability of the model to collisions at ultra-high energies (i. e. air showers) remains to be investigated.

II. The Propagation of Cosmic Rays Through the Atmosphere on a General Isobar Model

The calculations are based on the following assumptions:

1) Incident and target nucleons can be treated as free. (Since nucleon-nucleon collisions exhibit a high degree of elasticity at high energy, a nucleon which enters the atmosphere as constituent of a heavy primary is not shielded effectively and will have the same interaction mean free path in the atmosphere as a primary proton).
2) Upon emerging from the collision the incident baryon has lost some energy and may find itself either in the nucleon ground state or in an excited state from which it returns to the ground state by a succession of two-body decays leading to a total emission of n_B pions, or else by the emission of heavier bosons which subsequently disintegrate into n_B pions.
3) A fixed fraction of the incident energy is used up in creating particles through the pionization process, but in the presence of mesons from the decay of baryon isobars, the energy of the particles created in the pionization process is too low to contribute significantly to the flux of secondary particles in the atmosphere, as shown in Appendix A and Fig. 1. (Also mesons from the decay of the baryon which is emitted backwards in the C-system do not contribute since their energy in the laboratory system is still lower than that of the mesons from the pionization process).

Thus, according to this model, a nucleon after collision and deexcitation will have retained a substantial fraction η of its original energy, where η does not depend on energy but does depend on the type of isobar created in the collision, on its mode of decay, and the angle at which the mesons are emitted.

In the text we shall treat the production of secondaries as if they were decay products of a single type of isobar which returns to the nucleon ground state by the emission of n_B pions. In Appendix B we treat the more general case of deexcitation of a mixture of baryon states, each according to its own decay scheme, characterized by its mass and a set of decay branching ratios. The resulting formulae for the flux of secondary particles are identical with those given in this section, provided the quantities enclosed in brackets $\langle\ \rangle$ are replaced by the appropriate average values derived in Appendix B.

II.1 *The Energy Spectrum of Nucleons in the Atmosphere*

Let the differential energy spectrum of primary cosmic ray nucleons be represented by

$$N(0, E)\, dE = S_0 \frac{dE}{E^{\gamma+1}} \tag{II.1}$$

over the entire energy range of interest.*

The number of nucleons of energy E which have suffered j collisions is given by

$$N(E)_j = N(0, E)\ \eta^{\gamma j}.$$

The probability that a nucleon has suffered j collisions by the time it has reached an atmospheric depth of x g/cm^2 is

$$e^{-x/\lambda}\left(\frac{x}{\lambda}\right)^j \frac{1}{j!}.$$

Therefore, the flux of nucleons of energy E at a depth x g/cm^2 is given by

$$N(x, E) = N(0, E)\, e^{-x/\lambda} \sum_{j=0}^{\infty} \left(\frac{x}{\lambda}\right)^j \frac{\eta^{\gamma j}}{j!} = N(0, \mathrm{E})\, e^{-x/\Lambda}, \tag{II.2}$$

where $\Lambda = \dfrac{\lambda}{1 - \langle \eta^{\gamma} \rangle}$ is the attenuation length of nucleons in air and λ their interaction mean free path.

The bracket around $\langle \eta^{\gamma} \rangle$ indicates, as explained before, that it will have to be replaced by an appropriate average (eq. B.9).

Various corrections to eq. II.2 are required in the low energy region; they will be discussed in Appendix D (eq. D 2, 3).

II.2 *The Production Spectrum and Flux of Charged Pions*

The production spectrum of charged pions from baryon decay can be calculated in a straightforward manner on the basis of the model (see Appendix B) and is given by

$$P_{\pi}(x, E)\, dx = \langle B \rangle\, N(0, \mathrm{E})\, e^{-x/\Lambda} \frac{dx}{\lambda}. \tag{II.3}$$

$\langle B \rangle$ is defined by relation eqs. B.15, 17 in terms of the relative production rates of different isobars and their decay properties.

* Effects due to the apparent steepening of the primary spectrum above $\sim 10^{14}$ eV can be observed at present only in extensive air shower frequencies and perhaps near the upper end of the γ-ray spectrum; this effect will not be considered here.

(For a single isobar with transition directly to the ground state by isotropic emission of a charged pion $\langle B \rangle = B$ and is equal to

$$B = (\eta' \varepsilon)^{\gamma} \frac{(1+\beta)^{\gamma+1} - (1-\beta)^{\gamma+1}}{2\beta(\gamma+1)}, \tag{II.4}$$

where η' is the fraction of the incident nucleon energy retained by the isobar after pionization, ε is the fractional energy in the isobar rest system which is carried away by the decay pion, β its velocity, and γ is the exponent of the primary spectrum).

The interactions of pions in the atmosphere represent another source of particle creation; this increases the pion flux significantly in the lower parts of the atmosphere. Since the nature of pion-nucleon collisions is not well known, one must introduce a parameter which describes essentially the degree of elasticity which characterizes such collisions.

Complete elasticity implies that practically the entire energy is carried away by a single pion. Complete inelasticity implies that the available energy in the C-system is shared in a non-preferential manner by all the created mesons. (Unless the C-system energy of created pions is assumed to decrease with increasing collision energy, maximal inelasticity means that pion multiplicity is proportional to the square root of the incident energy). One can describe these extreme as well as intermediate conditions by assuming that (apart from a possible excitation of the target baryon) collisions of pions with nucleons lead on the average to the creation of

$$\nu = \nu_0 E_0^{\left(\frac{t-1}{t}\right)} \tag{II.5}$$

mesons and that the incident energy in the L-system is shared by half of them, so that their energy is

$$E = \frac{2E_0}{\nu}. \tag{II.6}$$

The production of charged pions by pions is then given by

$$P'_{\pi}(x, E) = (q_+ + q_-) \frac{t}{\lambda_{\pi}} (KE)^{2(t-1)} F_{\pi}\left[x, \frac{(KE)^t}{K}\right], \tag{II.7}$$

where

$$K = \left(\frac{\nu_0}{2}\right)^{\frac{t}{t-1}},$$

$q_{\pm 0}$ is the fraction of created pions of different charge which share the available energy, and $F_{\pi}(x, E)$ is the flux of charged pions at depth x with energy E.

This flux of charged pions is given by the solution of the differential equation

$$\frac{dF_\pi}{dx} + F_\pi\left(\frac{1}{\lambda_\pi} + \frac{u}{x}\right) = P_\pi + P'_\pi, \tag{II.8}$$

where $u = \dfrac{h_0 m_\pi}{c\,\tau_\pi E} = \dfrac{\varepsilon_\pi}{E}$ ($\varepsilon_\pi = 128$ GeV if $h_0 = 7$ km is the scale height of the atmosphere); λ_π is the interaction mean free path of pions in air.

The extreme cases are represented by $t = 1$ (elastic) in which case the equation reduces to

$$\frac{dF_\pi}{dx} + F_\pi\left(\frac{q_0}{\lambda_\pi} + \frac{u}{x}\right) = P_\pi, \tag{II.9}$$

and $t = 2$ (complete inelasticity).

The exact solution of equation II.8 may be written in the form

$$F_\pi(x, E) = \frac{S_0 \langle B \rangle}{E^{\gamma+1}} \frac{x}{\lambda} e^{-x/\lambda_\pi} \sum_{i=0}^{\infty} a_i(E) \left(\frac{x}{\lambda_\pi}\right)^i, \tag{II.10}$$

where

$$a_i = \sum_{n=0}^{i} \frac{[(q_+ + q_-)\,t]^n}{(KE)^{(\gamma-1)(t^n-1)}} \frac{\left(1 - \dfrac{\lambda_\pi}{\Lambda}\right)^{i-n}}{(i-n)!} \left[\prod_{j=0}^{n}\left(1 + i - j + \frac{u}{(KE)^{t^j-1}}\right)\right]^{-1}. \tag{II.11}$$

This can be verified by substitution. The pions which come directly from isobar decay are represented by the term $n = 0$; higher terms are important only in the lower atmosphere and for pions of intermediate energy (10–100 GeV).

II.3 *The Production of Neutral Pions*

The production spectrum of neutral pions is obtained from eqs. II.3 and II.7:

$$P_{\pi^0}(x, E) = \frac{1}{2} P_\pi(x, E) + \frac{q_0 t}{\lambda_\pi} (KE)^{2(t-1)} F_\pi\left[x, \frac{(KE)^t}{K}\right]. \tag{II.12}$$

From this equation the production spectrum and flux of γ-rays can be calculated in a straightforward manner if one assumes that γ-rays arise primarily from the decay of neutral pions.

II.4 *The Flux of Muons*

The production spectrum of muons from pion decay is

$$P_\mu(x, E) = \frac{\varepsilon_\pi}{xE} F_\pi[x, (rE)]. \tag{II.13}$$

Here, in analogy to eq. II.4,

$$r = \frac{2}{1 + \frac{m_\mu^2}{m_\pi^2}} \left[\frac{2\beta(\sigma+1)}{(1+\beta)^{\sigma+1} - (1-\beta)^{\sigma+1}} \right]^{\frac{1}{\sigma}} \approx 1.27, \tag{II.14}$$

where $\beta = 0.28$ is the muon velocity in the pion rest system

and
$$-[\sigma(x, E) + 1] = \frac{d[\log F_\pi(x, E)]}{d \log E} \tag{II.15}$$

is the exponent for the "best fitting" power law describing the pion spectrum. (Since σ varies slowly between γ and $\gamma+1$, the expression in brackets in II.14 is very close to unity and can be neglected; this is equivalent to the assumption that each muon receives 79 % of the pion energy, irrespective of the angle of emission).

The probability that a muon, produced at atmospheric depth z with energy E_z survives until it reaches depth x while losing energy by ionization at the rate b, so that it arrives with energy $E = E_z - b(x - z)$ is given by

$$\omega(z, E_z; x) = \left[\frac{z}{x} \left(1 - \frac{b(x-z)}{E_z} \right) \right]^{\frac{\varepsilon_\mu}{E_z + bz}} \tag{II.16}$$

$$\left(\varepsilon_\mu = \frac{h_0 m_\mu}{c\tau_\mu} \approx 1.12 \text{ GeV} \right).$$

Thus the muon flux is

$$\left.\begin{aligned} F_\mu(x, E) &= \int_0^x dz\, \omega\, P_\mu(z, E_z) \\ &= \frac{S_0 \langle B \rangle \varepsilon_\pi}{\lambda r^{\gamma+1}} \left(\frac{\lambda_\pi E}{x} \right)^v \frac{\sum_{i=0}^{\infty} a_i (rE')}{E'^{\gamma+2+v}} \int_0^{x/\lambda_\pi} dy\, e^{-y} y^{i+v}, \end{aligned}\right\} \tag{II.17}$$

where the integral can be replaced by the gamma function $\Gamma(i+v+1)$ for $x \gg \lambda_\pi$,

$$v = \frac{\varepsilon_\mu}{E + bx},$$

$$E' = E + b(x - \bar{x})$$

and the "mean height of production"

$$\bar{x} = \frac{\int_0^x dz\, z\, \omega\, P_\mu(z, Ez)}{\int_0^x dz\, \omega\, P_\mu(z, E_z)} = \lambda_\pi (1 + v) + \frac{\lambda_\pi \sum_{i=0}^{\infty} i a_i (rE')\, \Gamma(i + v + 1)}{\sum_{i=0}^{\infty} a_i (rE')\, \Gamma(i + v + 1)} \qquad \text{(II.18)}$$

for $x \gg \lambda_\pi$.

When $\lambda_\pi \approx \Lambda$, the first term in the summation, i. e. $a_0 = \dfrac{1}{1 + \varepsilon_{\pi/rE'}}$, accounts for more than 80 % of the muons in the lower atmosphere at all energies.

II.5 *The Flux of Neutrinos from $\pi - \mu$ Decay*

The neutrino flux has the same form as the muon flux, but without the terms due to ionization and decay; it is, therefore, obtained from the muon flux (eq. II.17) by setting $v = 0$ and $b = 0$ and replacing r by

$$r' = \frac{2}{1 - \left(\dfrac{m_\mu}{m_\pi}\right)^2} \frac{(1 + \sigma)^{\frac{1}{\sigma}}}{2} \simeq 4.0 \pm 0.2 \qquad \text{(II.19)}$$

for $\gamma < \sigma < \gamma + 1$.

$$F_\nu(x, E) = \frac{S_0 \langle B \rangle \varepsilon_\pi}{(\lambda/\lambda_\pi)\, r'^{\gamma+1}} \frac{\sum_{i=0}^{\infty} a_i (r'E)\, i!}{E^{\gamma+2}}. \qquad \text{(II.20)}$$

III. The Charge Composition of Secondary Cosmic Rays

The target nuclei in the atmosphere contain equal numbers of protons and neutrons. If the incident cosmic ray beam also were charge symmetric, then all secondary components of cosmic radiation would have to exhibit charge symmetry on any model of particle creation. Actually, the primary cosmic ray particles bring a known amount of excess positive charge. The manner in which this excess is shared by various secondary components of the cosmic radiation provides clues to the nature of high energy inter-

actions and, in particular, to the excitation and decay modes of isobars. In this section we investigate the charge composition of different components of cosmic radiation on a general isobar model.

III.1. *The Neutron to Proton Ratio in the Atmosphere*

Let w represent the probability that a nucleon after colliding with a charge symmetric target and after possible excitation and decay emerges in a different charge state. After j collisions the original composition of the nucleon beam† δ_0 will be changed into

$$\delta_j = \left(\frac{p-n}{p+n}\right)_j = \delta_0 (1-2w)^j. \tag{III.1}$$

Proceeding as in the derivation of eq. (II.2) one finds

$$\delta_x = \frac{N_p - N_n}{N} = \delta_0 e^{-x\left(\frac{1}{\lambda}-\frac{1}{\Lambda}\right)} \sum_{j=0}^{\infty} \left(\frac{x\eta'}{\lambda}\right)^j \frac{(1-2w)^j}{j!} = \delta_0 e^{-\frac{2x}{\lambda}\langle \eta' w\rangle}. \tag{III.2}$$

Thus

$$N_p(x, E) = N(0, E)\, e^{-\frac{x}{\Lambda}} \left(\frac{1+\delta_x}{2}\right), \tag{III.3}$$

$$N_n(x, E) = N(0, E)\, e^{-\frac{x}{\Lambda}} \left(\frac{1-\delta_x}{2}\right). \tag{III.4}$$

An explicit expression for $\langle \eta' w \rangle$ in terms of isobar properties is given in Appendix B (eq. B.10).
At energies $E \leqslant 10$ GeV eqs. (III.3, 4) have to be corrected as discussed in Appendix D.

III.2. *The Ratio of Positive to Negative Pions and Muons*

Let $\delta_\pi = \dfrac{n_+ - n_-}{n_+ + n_-}$ be the composition of the charged pions emitted in the deexcitation of a baryon which entered the collision as a proton. (Because

† The composition of the primary beam $\delta_0 = 0.74$ is known[6] to be constant within experimental error upto nucleon energies of order 10^{13} eV. At still higher energies there is evidence for a steepening of the primary spectrum and, if real, this will be accompanied presumably by changes in the chemical composition and the proton to neutron ratio.[8]

of charge symmetry, this changes sign if the incident particle is a neutron). The composition of the decay pions produced at depth x is then the product

$$\frac{P_{\pi^+} - P_{\pi^-}}{P_{\pi^+} + P_{\pi^-}} = \delta_x \langle \delta_\pi \rangle , \tag{III.5}$$

where $\langle \delta_\pi \rangle$ is the average positive excess mong the decay pions; it is defined by eqs. (B.15, 16) for an arbitrary mixture of isobaric states.

In order to determine the charge composition of the pion flux in the atmosphere, one may rewrite (eq. II.8.)

Setting

$$\left.\begin{aligned} F_\pi &= F_{\pi^+} + F_{\pi^-} \\ \text{and} \qquad G_\pi &= F_{\pi^+} - F_{\pi^-} \end{aligned}\right\} \tag{III.6}$$

one finds

$$\left.\begin{aligned} &\frac{dF_\pi(x,E)}{dx} + F_\pi(x,E)\left(\frac{1}{\lambda_\pi} + \frac{u}{x}\right) \\ &\quad = \frac{S_0 \langle B \rangle e^{-x/\Lambda}}{\lambda E^{\gamma+1}} + (q_+ + q_-)\frac{t}{\lambda_\pi}(KE)^{2(t-1)} F_\pi\left[x, \frac{(KE)^t}{K}\right] \end{aligned}\right\} \tag{III.7}$$

and

$$\left.\begin{aligned} &\frac{dG_\pi(x,E)}{dx} + G_\pi(x,E)\left(\frac{1}{\lambda_\pi} + \frac{u}{x}\right) \\ &\quad = \delta_x \langle \delta_\pi \rangle \frac{S_0 \langle B \rangle}{\lambda E^{\gamma+1}} e^{-x/\Lambda} + (q_+ - q_-)\frac{t}{\lambda_\pi}(KE)^{2(t-1)} G_\pi\left[x, \frac{(KE)^t}{K}\right]. \end{aligned}\right\} \tag{III.8}$$

q_+, q_- is the fraction of positives or negatives among the pions which share the energy available in a pion induced interaction for the case that the incident pion had *positive* charge. Charge symmetry requires that these quantities change sign if the incident pion is negative; on the other hand, charge conservation in collisions of pions with charge symmetric targets requires that

$$\frac{1}{2} \leqslant (q_+ - q_-)\frac{\nu}{2} = (q_+ - q_-)\left(\frac{KE}{2}\right)^{t-1} \leqslant 1 , \tag{III.9}$$

depending on whether the excess charge brought in by the incident pion is uniformly distributed among all secondaries or appears preferentially among the more energetic ones. Making use of eqs. (III.2, 3, 4 and 7), eq. (III.8) can now be rewritten in a form similar to the differential equation (III.7:

$$\left.\begin{aligned} &\frac{dG_\pi(x,E)}{dx}+G_\pi(x,E)\left(\frac{1}{\lambda_\pi}+\frac{u}{x}\right)\\ &=\delta_0\langle\delta_\pi\rangle\frac{S_0\langle B\rangle}{\lambda E^{\gamma+1}}e^{-x/\Lambda'}+\frac{\alpha t}{4\lambda_\pi}(KE)^{t-1}G_\pi\left[x,\frac{(KE)^t}{K}\right], \end{aligned}\right\} \text{(III.8 a)}$$

where

$$\frac{1}{\Lambda'}=\frac{1}{\Lambda}+2\frac{\langle\eta^\gamma w\rangle}{\lambda} \tag{III.10}$$

and

$$1\leqslant\alpha\leqslant 2\,.$$

In analogy to eq. (II.10) the solution can be written in the form

$$G_\pi(x,E)=\frac{\delta_0\langle\delta_\pi\rangle S_0\langle B\rangle}{E^{\gamma+1}}\frac{x}{\lambda}e^{-x/\lambda_\pi}\sum_{i=0}^{\infty}a_i'(E)\left(\frac{x}{\lambda_\pi}\right)^i, \tag{III.11}$$

where a_i' can be obtained from a_i by making the substitutions

$$\Lambda\to\Lambda'$$

$$\gamma\to\gamma+1$$

$$(q_+ + q_-)\,t\to\frac{\alpha}{4}.$$

Thus, the charge composition of the pion flux is

$$\frac{F_{\pi^+}-F_{\pi^-}}{F_{\pi^+}+F_{\pi^-}}=\delta_0\langle\delta_\pi\rangle\frac{\sum_{i=0}^{\infty}a_i'(E)\left(\frac{x}{\lambda_\pi}\right)^i}{\sum_{i=0}^{\infty}a_i(E)\left(\frac{x}{\lambda_\pi}\right)^i} \tag{III.12}$$

and that of the muons from pion decay (see II.17)

$$\delta_\mu=\frac{F_{\mu^+}-F_{\mu^-}}{F_{\mu^+}+F_{\mu^-}}=\delta_0\langle\delta_\pi\rangle\frac{\sum_{i=0}^{\infty}a_i'(rE')\int_0^{x/\lambda_\pi}dy\,e^{-y}\,y^{i+v}}{\sum_{i=0}^{\infty}a_i(rE')\int_0^{x/\lambda_\pi}dy\,e^{-y}\,y^{i+v}}. \tag{III.13}$$

The Effect of Kaons on the Muon Charge Ratio at Sea Level.

In this and the preceding section it was tacitly assumed that pions are the only particles which, by their decay, contribute to the muon flux. However, non-strange isobars of sufficiently large mass can also decay into a kaon and a hyperon, or a kaon-antikaon pair and a non-strange baryon; by subsequent decay these particles contribute to the observed flux of pions, γ-rays, and muons.

In general, one expects that such decay modes make only a moderate contribution to the flux of secondaries in the atmosphere. In the case of high energy muons, however, the part of the contribution which does not involve an intermediate pion (viz. the process $K \to \mu + \nu$) is amplified because the mean life of charged kaons is significantly shorter than that of charged pions and their mass is greater. Therefore, at a given energy, the probability of decay before interacting in the atmosphere is larger for a kaon than for a pion and, in the energy range where pion interaction becomes more probable than decay, an increasing fraction of muons arises from the decay of charged kaons.

Also hyperons and neutral K-particles can decay into muons without an intermediate pion, but they will contribute little, because in the case of hyperons and K_1^0 the branching ratio is small, and in the case of K_2^0 the lifetime is long. Therefore, it is the presence of charged kaons decaying directly into muons which produces the largest effect.

Among the decay modes which contribute kaons, the mode

$$N^* \to K + Y$$

is no doubt dominant;† it can produce positive but not negative kaons.

(Negative kaons can arise in two ways; either through $N^* \to (K\overline{K}) + N$ or through $N^* \to K + Y^*$, where the hyperon state is highly excited and therefore has a certain probability for fast decay via $Y^* \to \overline{K} + N$).

The ratio of positive to negative muons becomes therefore

$$\frac{\mu^+}{\mu^-} = \frac{1+\delta_\mu}{1-\delta_\mu} + \frac{2}{1-\delta_\mu} b_K \frac{F_{\mu K}}{F_\mu}, \qquad \text{(III.14)}$$

† As pointed out in Section I, the strong excess of positive over negative kaons among the very slow and the very fast particles produced in nuclear interactions indicates that modes which produce only kaons of strangeness +1 dominate in the decay of non-strange isobars.

where b_K is the branching ratio for the decay mode $N^* \to K + Y$ and F_μ, δ_μ, and F_{μ_K} are given by eqs. (II.17, III.13 and C.6), respectively. $\frac{F_{\mu_K}}{F_\mu}$ depends on energy approximately as

$$\frac{1+\frac{E}{\varepsilon_\pi/r}}{1+\frac{E}{\varepsilon_K/r_K}} = \left(\frac{1+\frac{E}{100}}{1+\frac{E}{570}}\right)_{E \text{ in GeV.}} . \qquad \text{(III.15)}$$

IV. Comparison between Measured and Calculated Properties of the Secondary Cosmic Radiation

In the previous section explicit expressions have been obtained for the altitude and energy dependence of protons, neutrons, pions, muons, and neutrinos.

It remains to assign numerical values to the various *energy independent* parameters which characterize the primary cosmic radiation and the interactions of nucleons and pions in the atmosphere. In principle, it should be possible to get the information on interaction cross sections from the asymptotic behaviour of these particles in the upper energy range of present-day accelerators. However, some of these data, especially on the formation of isobars, are not yet known adequately and it is necessary to use a few cosmic ray measurements to assign numerical values to some of these constants. The following values have been used for drawing Figs. 2–6.

a) The interaction mean free path of nucleons in air

$$\lambda = 75 \pm 5 \text{ g/cm}^2 \dagger$$

has been obtained from the absorption in graphite of neutrons capable of producing charged penetrating particles [10].

b) The interaction mean free path for pions in air

$$\lambda_\pi = 120 \text{ g/cm}^2$$

† Bad geometry absorption measurements at accelerator energies[9]) yield a mean free path which is at least 20 % longer. It is difficult to account for this discrepancy, unless it is due to interactions in which the incident proton suffers only a small energy loss and receives a transverse momentum less than 150 MeV/c. Such collisions could be missed in the accelerator experiment because the emerging proton falls within the diffraction peak.

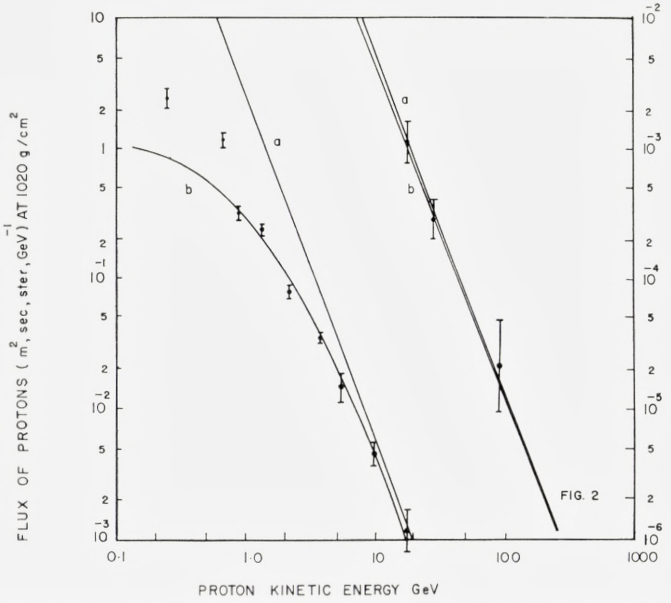

Fig. 2. The energy spectrum of protons at sea level. Curve 'a' is a line of constant slope $\gamma + 1 = 2.67$. Curve 'b' represents the calculated spectrum (eq. D.3) based on an attenuation mean free path $\Lambda = 120$ g/cm² and an ionization loss of 2 MeV (g/cm²). The experimental points are taken from ref. 12. The excess of observed protons at low energy is of the right order of magnitude to be attributed to terrestrial protons from the target nuclei in the atmosphere (see Appendix Db).

has been estimated by multiplying λ with the ratio of the cross section $\frac{\sigma_{pp}}{\sigma_{\pi p}} \approx 1.6$ obtained in high energy laboratories [11].

c) The exponent of the primary spectrum

$$\gamma = 1.67$$

is consistent with direct measurements at the top of the atmosphere; the exact value has been chosen so as to give the best fit to the sea level proton spectrum at high energy [12].

d) The absorption length for nucleons in the atmosphere

$$\Lambda = 120 \pm 5 \text{ g/cm}^2$$

has been obtained from the absolute intensity of protons at sea level as measured by the DURHAM group [12], and the absolute value of the primary cosmic ray flux as given by MCDONALD and WEBBER [13]. The expected deviations of the proton spectrum from a simple power law at low energies are discussed in Appendix D.

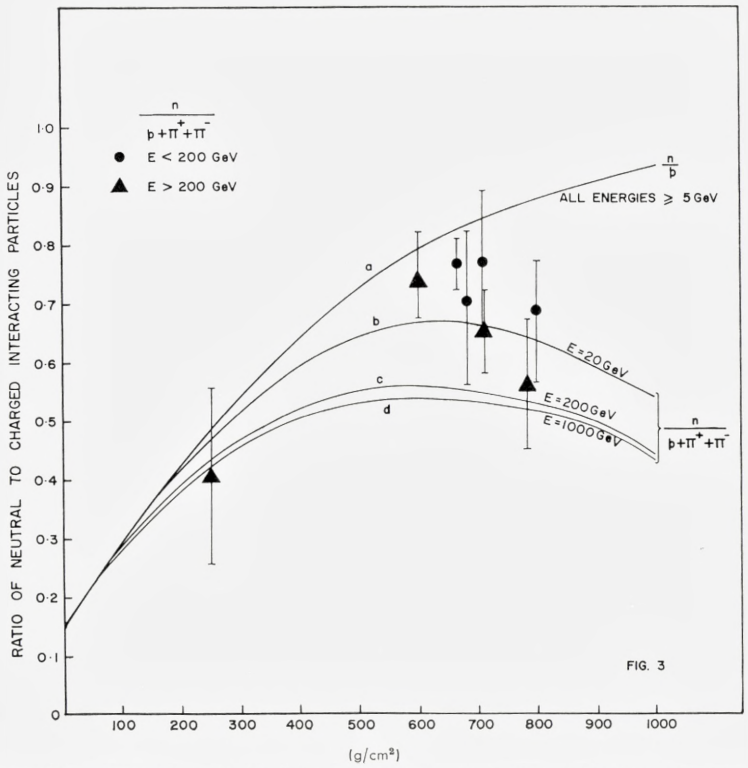

Fig. 3. Curve 'a' represents, as a function of altitude, the ratio of neutrons to protons at all energies for which ionization loss and contribution of recoil nucleons (Appendix D) can be neglected. Curves 'b', 'c', and 'd' represent the ratio of neutrons to the sum of protons and charged pions for various energies. The experimental data are those of ref. 26.

e) The composition of the primary radiation

$$\delta_0 = \frac{p_0 - n_0}{p_0 + n_0} = 0.74 \pm 0.01$$

has been obtained from the measurements of the primary chemical composition as reviewed by WADDINGTON [6].

f) The charge exchange probability for nucleons colliding in the atmosphere

$$w = 0.3 \begin{pmatrix} +0.2 \\ -0.08 \end{pmatrix}$$

has been chosen to reproduce the neutron to proton ratio as measured at mountain altitude [14]. (It is assumed that the average $\langle \eta^\gamma w \rangle$ can be replaced by $\langle \eta^\gamma \rangle w$).

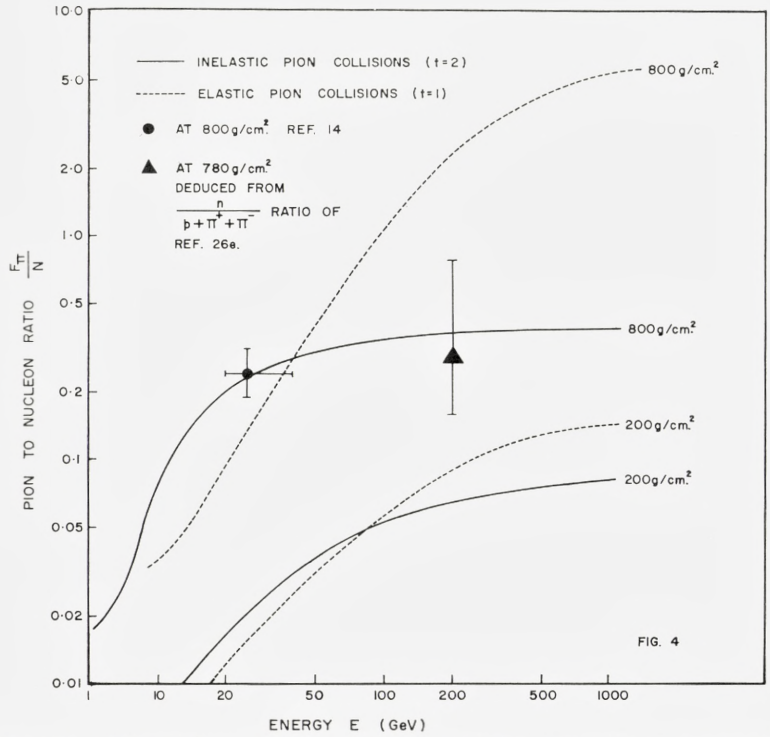

Fig. 4. The ratio of charged pions to nucleons as a function of energy at atmospheric depths of 200 g/cm² and 800 g/cm². The solid curves are calculated on the assumption of complete inelasticity in pion-nucleon collisions, the dotted curves for complete elasticity.

g) The constants which determine the number of pions which share the energy in pion nucleon collisions, eq. (II.5), have been chosen to represent the extreme cases

α) Complete elasticity $t = 1$ (for $t = 1$ the equations are independent of K)

β) Complete inelasticity $t = 2 \qquad \nu = 0.7\, E^{\frac{1}{2}}_{\mathrm{GeV}}$ (IV.1)

$$\left[\text{i. e. } K = \left(\frac{\nu_0}{2}\right)^2 = \frac{1}{8}\right].$$

K is chosen to fit the measured $\pi^{\pm}/p$ ratio at 800 g/cm² in the energy region 20–40 GeV [14].

h) The constant $\langle B \rangle$ which characterizes the fractional energy given to pions in the decay of baryon isobars is obtained by comparing calculated and measured sea level flux of muons at 40 GeV [15]

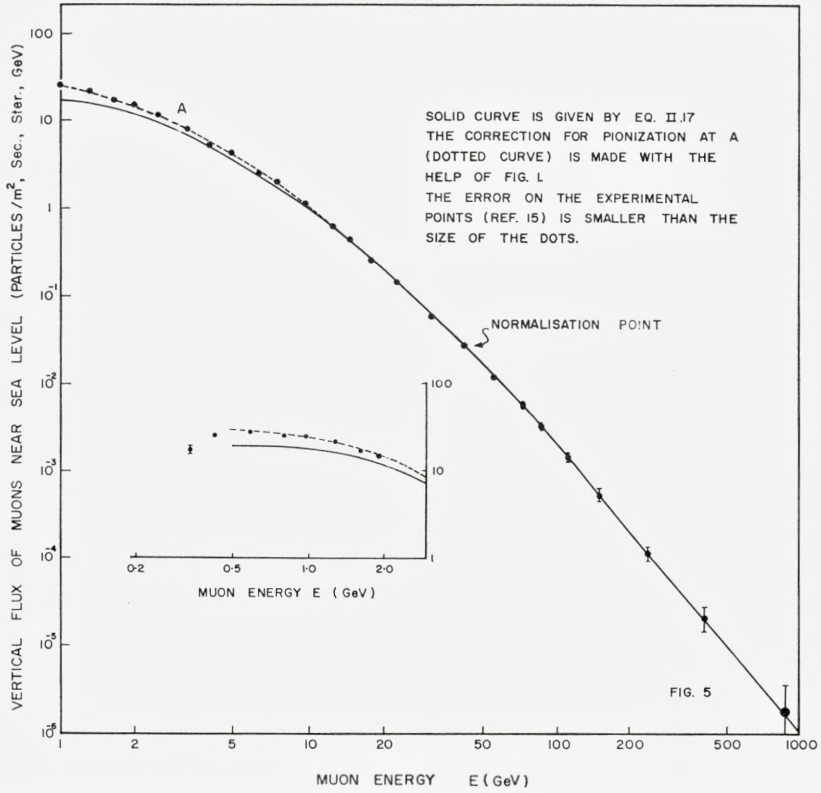

Fig. 5. The energy spectrum of muons at sea level. The solid curve represents the calculated spectrum, eq. II.17, normalized at $E \approx 40$ GeV. The dotted curve represents the spectrum after a rough correction for contribution from pionization has been made with the help of Fig. 1. the closeness of its fit to the experimental data is therefore somewhat fortuitous.

$\langle B \rangle = (3.35 \pm 0.3) \cdot 10^{-2}$ (both for the elastic case and for the inelastic case).

Using these constants the following curves have been calculated and reproduced together with the available experimental points:

α) The energy spectrum of protons Fig. 2 (eq. D3 (Appendix D)).

β) The ratio of neutrons to protons as a function of atmospheric pressure Fig. 3 a (eqs. III.3, 4)

γ) The ratio of neutral to charged interacting particles $\frac{n}{p + \pi^+ + \pi^-}$ as

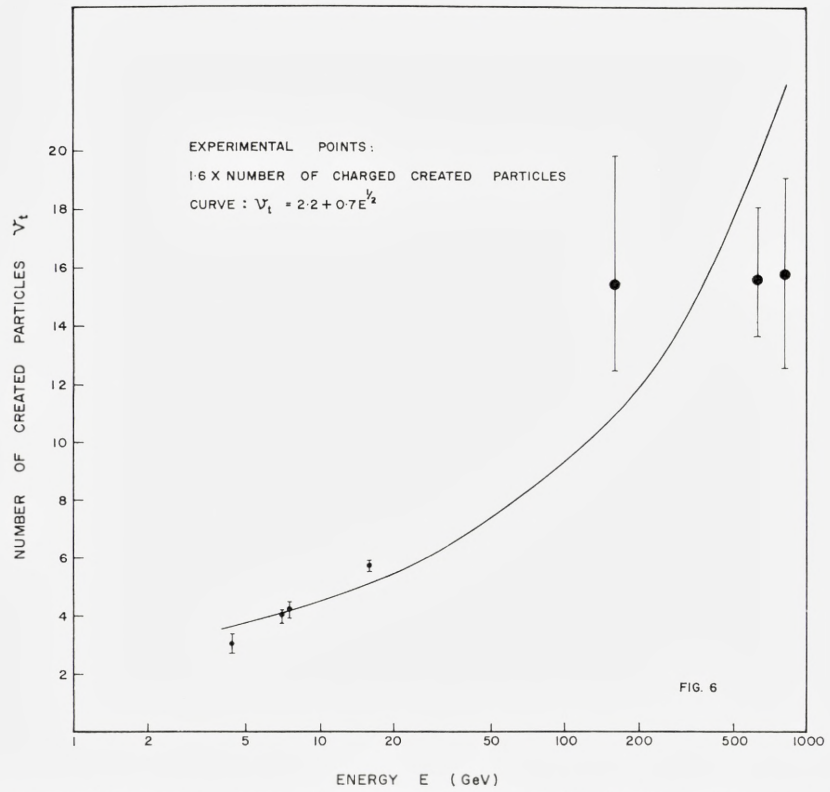

Fig. 6. The total number of created particles, charged and neutral, emitted in collisions of charged pions with target nuclei of low atomic weight, is plotted as a function of energy. The curve represents the relation IV.1a, with $sn_B \approx 2.4$. Experimental points are from the data compiled in ref. 27.

a function of atmospheric pressure for various energies, on the assumption that pion-nucleon collisions are completely inelastic. Figs. 3 b, c, d (eqs. II.10 and III.3, 4).

δ) The ratio of pions to nucleons at an atmospheric depth of 200 g/cm^2 and 800 g/cm^2 for the two extreme cases: that pion nucleon interactions are completely elastic and that they are completely inelastic. Fig. 4 (eqs. II.10 and III.3).

ε) The energy spectrum of muons at sea level. Fig. 5 (eq. II.17).

φ) The multiplicity of created particle in pion nucleon collisions. Fig. 6.

$$\nu_t = sn_B + 0.7\ E^{\frac{1}{2}}_{(\mathrm{Gev})}. \qquad \text{(IV.1a)}^{\dagger}$$

† The first term has been added to eq. (IV.1) to represent low energy mesons from the decay of an excited target baryon.

Formula (IV. 1 a), which corresponds to completely inelastic pion-nucleon interactions at high energy, cannot be expected to reproduce the multiplicity accurately at low energy; however, it describes the available data reasonably well provided one choses the mean number of mesons contributed by the target baryon, $sn_B \approx 2$ (Fig. 6). [The experimental data for high energy pion collisions are extremely poor; data below 20 GeV seem to indicate that the multiplicity of created particles increases not slower than $E^{\frac{1}{2}}$, independent of the choice of sn_B].

As shown in Fig. 4, the difference in the pion flux for the assumption of complete elasticity and complete inelasticity is small at 200 g/cm^2 and, therefore, the muon spectra at sea level are insensitive to the nature of pion-nucleon interactions. On the other hand, the pion flux near ground is very sensitive; compared to the elastic case the assumption of strong multiplication of pions leads to a large excess between 10 GeV and $\sim$30 GeV and a very large deficit above $\sim$50 GeV. The inelastic case is in better agreement with the existing determination of the pion-nucleon ratio than the elastic case.

Assuming then a high degree of inelasticity in pion interactions, the comparison between calculated and measured quantities (Figs. 2–6) shows that the very simple, energy independent model of high energy collisions, which has been adopted, is adequate for describing the distribution of secondary cosmic radiation within the accuracy of existing measurements.

It is now of interest to discuss the restrictions which are imposed on the masses, the excitation probabilities, and the decay modes of isobars by the numerical values of $\langle B \rangle$ and Λ and by the observed energy dependence of multiplicity of created particles in nucleon-nucleon collisions.

V. Average Parameters Characterizing the Production and Properties of the Dominant Baryon Isobars

In order to find the simplest isobar model capable of accounting for all existing observations on secondary cosmic radiation, one may assume tentatively that the incident nucleon, after having lost a fixed fraction $(1-\eta')$ of its energy in the pionization process, has a probability of emerging as an excited baryon of mass M_B and a probability $(1-s)$ of emerging in the nucleon ground state. If excited, the baryon is assumed to decay to the ground state by the isotropic emission of n_B mesons, all of which have the same energy in the baryon rest system.†

† These n_B mesons may of course be themselves decay products of a boson isobar.

V.1. *The Excitation Probability of Isobars and the Average Number of Decay Mesons*

The average number of mesons emitted in the decay of nuclear isobaric states can be estimated from the shape of the multiplicity-energy relation for particles created in nucleon-nucleon collisions.

It is shown in Appendix A.2 that the assumptions underlying the model discussed in this paper determine the form of the relation between multiplicity of created particles and energy (eq. A.4)

$$n_t = 2\, s n_B + n_0\, E^{\frac{1}{2}},$$

where the first term represents the average number of decay mesons from isobaric states of the baryon and the second the contribution from pionization. The actual relation may be expected to show some structure if the size of the heavy bosons constituting the fireball were quantized in units $M_F \gg m_\pi$.

When comparing relation A.4 with existing experimental values shown in Fig. 9, one sees that it is consistent with existing data

$$\text{for} \quad 2 < s n_B < 3$$

$$\text{and} \quad n_0 \approx \frac{1}{4}, \text{ for } E \text{ in GeV}.$$

The difference between this relation and the frequently employed empirical form $n_t \sim E^{\frac{1}{4}}$ becomes significant only at energies above $\sim 10^{15}$ eV, i. e. in the study of extensive air showers.

Separate values for the excitation probability s and the mean number of isobar decay pions n_B can be obtained by using the numerical values for $\langle \eta^\gamma \rangle$ and $\langle B \rangle$, obtained in the preceding section.

$\langle \eta^\gamma \rangle$ is related to the mean elasticity of collisions and can be expressed in terms of a ratio of nucleon interaction mean free path and attenuation length (II.2) which have been given in the previous section:

$$\langle \eta^\gamma \rangle = 1 - \frac{\lambda}{\Lambda} = 0.37 \pm 0.06. \tag{V.1}$$

With the help of eq. (B.9) one finds (for the particular case of a single type of baryon isobar) the relation between this parameter and the isobar properties:

$$\langle \eta^\gamma \rangle = \eta'^\gamma \left[1 + s\,(A-1)\right], \tag{V.2}$$

where A (defined by B.5) is given by

$$A = \mathfrak{E}_p^{\gamma} \frac{(1+\beta_p)^{\gamma+1} - (1-\beta_p)^{\gamma+1}}{2\beta_p(\gamma+1)}.$$

$\mathfrak{E}_p = \dfrac{M_B^2 + M_p^2 - n_B^2 m_\pi^2}{2 M_B^2}$ is the average fractional energy in the rest system of the isobar carried away by the deexcited nucleon and β_p is its velocity.

A second relation involving the same parameters is obtained from eqs. (B.15, 17), which determine the $\langle B \rangle$ in terms if isobar properties. ($\langle B \rangle$ is the scale factor which relates pion production to nucleon intensity).

For the case under discussion eqs. (B.15), 17 yield

$$\langle B \rangle = \tfrac{2}{3} s n_B \alpha \eta'^{\gamma} = (3.35 \pm 0.3) \times 10^{-2}; \qquad \text{(V.3)}$$

here α (defined by B.13) is given by

$$\alpha = \mathfrak{e}^{\gamma} \frac{(1+\beta)^{\gamma+1} - (1-\beta)^{\gamma+1}}{2\beta(\gamma+1)},$$

where $\mathfrak{e} = \dfrac{M_B^2 - M_p^2 + n_B^2 m_\pi^2}{2 M_B^2 n_B}$ is the average fractional energy in the rest system of the isobar, carried away by a pion, and β is its velocity.

After eliminating the unknown factor η'^{γ} between eq. (V.2) and (V.3) one obtains a relation between

the isobar mass,	M_B,
the probability of its excitation,	s,
and the number of decay mesons,	n_B.

This relation is rather insensitive to the value of M_B, which cannot therefore be determined accurately from the available data. Therefore, we have plotted the relation in Figs. 7 a, b for two extreme values of M_B:

$$M_B = \infty \text{ and}$$
$$M_B = (M_B)_{\min}.$$

$(M_B)_{\min}$ is defined by

$$\frac{(M_B)_{\min}}{M_p} = 0.288\, n_B + \sqrt{1 + 0.06\, n_B^2}; \qquad \text{(V.4)}$$

it is the lowest possible mass for an isobar which decays into n_B pions and gives to each just the energy which a pion would receive in the deexcitation

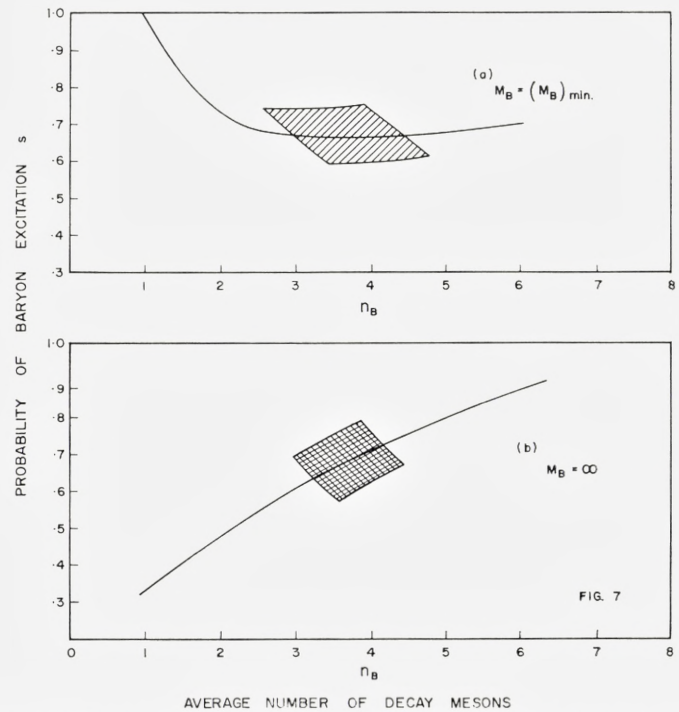

Fig. 7. The probability of baryon excitation, s, is plotted against the average number of decay mesons n_B. The shaded area indicates the values of s and n_B which are compatible with the observed ratios between the fluxes of primary nucleons, sea level protons, and sea level muons and with the observed energy dependence of multiplicity of meson production in high energy collisions (see Section V). Fig. 7a refers to a baryon with minimum mass as defined by eq. V.4; Fig. 7b refers to a baryon of infinite mass.

of the (3,3) resonance, i. e. the lowest excited state of the pion nucleon system.

The shaded area in Fig. 7 a indicates the range of values of s and n which satisfy the cosmic ray data

$$s = 0.70 \pm .07, \tag{V.5}$$

$$n_B = 3.5 \pm 0.5. \tag{V.6}$$

The vertical spread of the area of uncertainty in Fig. 7, i. e. the error in excitation probability, reflects mainly the uncertainty in the value of the nuclear interaction mean free path λ; the horizontal spread (the error in n_B) reflects mainly the present uncertainty in the mean multiplicity of created particles in the 1000 GeV region.

Fig. 7b ($M_B = \infty$) yields values for s and n_B which are not very different from those given in Fig. 7a, but it corresponds of course to an unrealistic assumption.

With $n_B = 3.5$, eq. (V.4) gives a lower limit for the average excited baryon mass $(M_B)_{\min} = 2300$ MeV. An indication that the actual value lies close to this limit is provided by the transverse momentum distribution of pions; the highest transverse momentum from an isobar of mass 2300 MeV occurs when it decays directly to the nucleon ground state by emission of a single pion; this value is 900 MeV/c and is very close to the upper limit of the observed transverse momentum distribution of neutral pions [16].

The fractional energy $(1-\eta')$ given to pionization can now be determined from eq. (V.2) or (V.3); it is (25 ± 7) %. This is not inconsistent with the observed multiplicity and the C-system energy of particles evaporated from the fireball if one also takes into account that ca. 20 % of the particles are heavier than pions (i. e. kaons or nucleon-antinucleon pairs).†

It must be remembered, however, that η' as determined here, and also the multiplicity relation which we have used, refer to targets consisting of light nuclei, not of single nucleons. Similarly, the value deduced for the excitation probability, s, refers to collisions of nucleons with light nuclei. The fractional energy loss to pionization in air $(1-\eta') \approx 25$ % should correspond to an energy loss of about 18 % for nucleon-nucleon collisions.

The fraction of the incident energy which goes into the pionization process is seen to be about half of the total energy loss of nucleons; because this energy is distributed among many particles, while in isobar deexcitation the energy is concentrated on a small number of pions, it is the latter which completely dominate the secondary cosmic radiation (see Fig. 1).

V.2. *The Composition of Isobar Decay Products*

a) The Ratio of Positive to Negative Pions and the Branching Ratio for K-Decay.

The positive excess $\langle \delta_\pi \rangle$ among decay pions of the forward isobars created by proton collisions with a charge symmetric target, as well as the

† If pions were the only particles evaporated from the fireball, the model predicts: $(1-\eta') = \dfrac{\varepsilon_F \gamma_C n_0 E^{\frac{1}{2}}}{E}$ (i. e. ≈ 11 %, if ε_F, which is the average pion energy in the C-system, is taken as $\varepsilon_F = 460$ MeV). But if only 80 % of the created particles are pions and 20 % are nucleons and antinucleons created with energy $\varepsilon \sim 3 M_p$, the corresponding value of $(1-\eta')$ lies above 20 %.

branching ratio b_k for the decay mode $N^* \to K + Y$, can be obtained from the observed muon charge ratio as a function of energy. Inserting the experimental values of quantities entering into eq. (III.15), the ratio μ^+/μ^- has been plotted as a function of energy in Fig. 8 for various values of b_k. The experimental points shown are those of reference 17; above 100 GeV, this includes all measurements in the vertical direction. All other experimental data (ref. 18) agree within errors, but cover only the lower part of the energy spectrum. (One experimental point[18c] which is in disagreement with the others has been indicated in the figure).

Each of the curves in Fig. 8 corresponds to a definite value of the branching ratio b_k for the decay $N^* \to K + Y$ and to a definite value for the charge excess $\langle \delta_\pi \rangle$ amongst pions. On each curve there is also shown the value of δ'_π, which represents the charge excess for $(1 - b_k)$ decays which do not involve strange particles. It is related to δ_π by

$$\delta'_\pi = \delta_\pi + c\, b_k ,$$

where $c \approx 0.8$ depends slightly on the particular hyperon states involved.

While the data seem to indicate an energy dependence of the form given by eq. (III.15), they do not rule out the value zero for b_k.

We take

$$b_k = (10 \pm 10)\ \%$$

and, correspondingly,

$$\frac{1}{\delta'_\pi} = 6.4\ {}^{+4.6}_{-1.3}.$$

(The highest value of the charge excess, $\frac{1}{\delta'_\pi} = 5.1$, corresponds to $b_k = 0$). A value of $b_k \approx 10\ \%$ corresponds to a kaon to pion production ratio

$$\left(\frac{K}{\pi}\right)_{\text{total}} = \frac{2\, b_k}{\frac{2}{3}(1 - b_k)} \frac{B_k}{B} \approx 16\ \%$$

(an upper limit on the production ratio $K_{\mu_4}/(\pi^+ + \pi^-) \approx 40\ \%$ at about 70 GeV, corresponding approximately to $\left(\frac{K}{\pi}\right)_{\text{total}} \approx 80\ \%$ was obtained by Ashton and Wolfendale[19] using the variation of muon flux with zenith angle).

b) The Ratio of Hyperons to Nucleons.

From the branching ratio $b_K = 0.10 \pm 0.10$ for the isobar decay mode $N^* \to K + Y$ one obtains the ratio of hyperons to nucleons among the forward emitted baryons in high energy collisions:

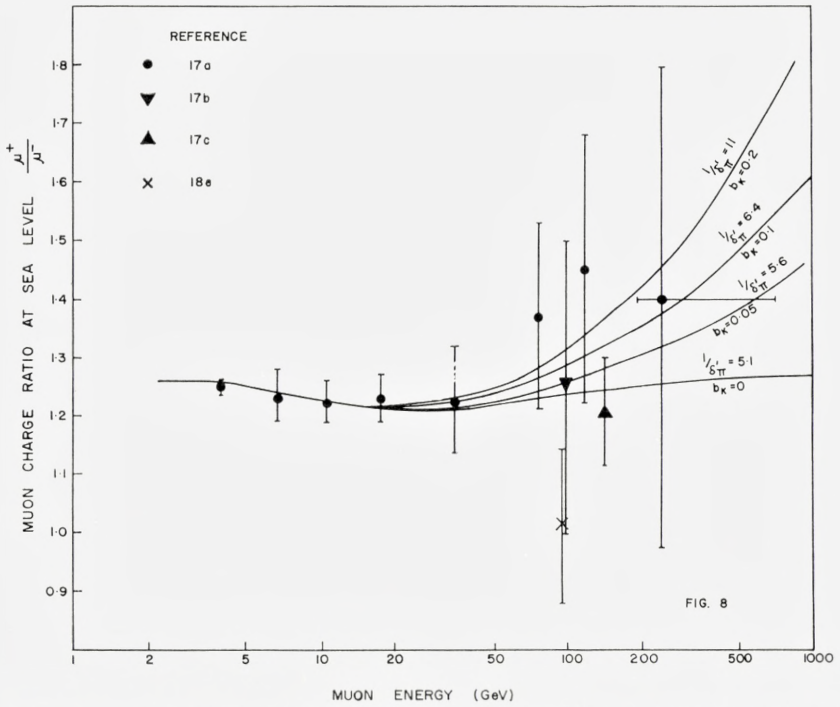

Fig. 8. The ratio of positive to negative muons at sea level[17] as a function of their energy. The relation given by eq. (III.15) has been drawn for various pairs of values $\langle \delta'_\pi \rangle$ and b_k consistent with the experimental data. [$\langle \delta'_\pi \rangle$ is the charge excess among pions from the isobar decays which do not involve strange particles, and b_k is the branching ratio for the decay mode $N^* \to K + Y$.]

$$\frac{Y}{N} = sb_K \approx 7\ {}^0/_0$$

and an estimate for the hyperon production cross section in high energy collisions:

$$\sigma_Y = 2\, sb_K\, \sigma_{\text{inelastic}} \approx 4.5\ \text{mb}.$$

However, because of the large experimental error which is still attached to the charge ratio of muons at high energy, one can place at present only an upper limit of 9 mb. on this cross section. (An earlier suggestion by one of the authors[4a], that this ratio may be high, is not borne out by this analysis). Fig. 8 shows that it is necessary to measure the ratio μ^+/μ^- at energy ~ 250 GeV to better than 5 $^0/_0$ in order to determine this cross section with an accuracy of ~ 30 $^0/_0$.

VI. Discussion

It has been shown that the model of meson production which has been adopted permits a rigorous calculation of the intensity, composition, and energy distribution of secondary cosmic ray particles in terms of those of the primary radiation. The model contains only one fundamental (and plausible) assumption, namely, that the excitation of baryon isobaric states, which is a prominent feature of nuclear collisions at laboratory energies, remains of some importance for collisions at higher energies. The comparative unimportance in the secondary cosmic radiation of mesons produced with low energies in the center-of-mass system of nucleon-nucleon collisions follows directly from this basic assumption (see Appendix A and Fig. 1). The contribution of mesons produced in pion-nucleon collisions is unimportant when calculating the muon flux, but affects the ratio of pions to nucleons in the lower atmosphere. The derivation of expressions for the spectra of various secondary components is then straightforward and can be carried out for an arbitrary set of baryon states and an arbitrary combination of excitation probabilities and decay modes.

Making use of existing experimental data on secondary components one obtains then fairly definite values for the excitation probability of some "average" baryon state (~ 0.7) and for its mean mass (~ 2200 MeV) (although arguments against a higher value cannot be considered entirely conclusive). One obtains also an average value for the number of mesons emitted in the decay of the isobar ($n_B \approx 3.5$) and for their net charge ($|\pi^+ - \pi^-| \approx 0.35$). The branching ratio for the decay of the "average isobar" in the mode $N^* \to K + Y$ can be determined only roughly $b_k = 10\ (\pm 10)$ Finally, one can break down the average fractional energy loss suffered by nucleons in high energy collisions into a part given to the decay mesons ($\sim 25\ \%$) and a part spent in creating particles in the C-system of nucleon-nucleon collisions (15–20 %).

On the other hand, it does not seem possible at present to deduce in a unique way, solely from the average properties of the secondary cosmic radiation, the excitation probabilities and decay schemes of the individual baryon states involved. This seems feasible only, either by extrapolation from the lower energy region accessible to accelerators, or by studying individual high energy events, jets or airshowers, with the particular aim of obtaining data on the very fast or the very slow particles created in the interaction.

The assumption that, in nuclear collisions at all energies above ~ 10 GeV, isobars are produced with masses and decay properties very similar to those needed to reproduce the pion production spectra at 22 GeV [3], is sufficient to account for the existing observations on secondary cosmic rays in the atmosphere. It is consistent with existing data on particle production in high energy laboratories and with observations on jets. More complicated models to account for secondary cosmic ray particles are clearly possible; the very simple hypothesis which has been explored in this paper is not unique, but it seems to be adequate at this stage.

An application of the model discussed in this paper to the structure of airshowers will be the subject of a later study. It seems worth mentioning, however, that, if the mechanism described here remains essentially valid also in the energy region responsible for airshowers, the probability that an incident nucleon loses all but 25 $^0/_0$ of its energy to neutral pions in a single collision will not be small. Such events will give rise to a few percent of air showers with abnormally high electron to muon ratios, i. e. with properties not unlike those of γ-ray induced airshowers whose possible occurrence and frequency is now under active investigation in various parts of the world. [20]

In view of the fact that particles created with low energy in the C-system of nucleon-nucleon collisions contribute little to the secondary cosmic ray flux (Appendix A) and that the recoil nucleons of terrestrial origin contribute only to the non-relativistic region of the energy spectrum (Appendix D), the high energy cosmic ray nucleons at any depth in the atmosphere represent essentially a sample of the extraterrestrial matter brought in by the incident cosmic radiation. Therefore, it is possible to study the fundamental question of whether the very high energy primary cosmic ray particles contain an appreciable fraction of anti-matter. If the primaries contained anti-nucleons, the nucleon spectrum on the ground above a few GeV should contain a corresponding fraction. By analysing the charge composition of the nucleon component at sea level upto $\sim$60 GeV, it seems possible to investigate a possible fraction of antimatter in the primary radiation upto energies of order 10,000 GeV/nucleon.

Acknowledgements

The authors are indebted to Mr. S. N. Tandon for assistance with the numerical computations. One of us (Y. P.) wishes to express his gratitude for the hospitality received at the Institute for Theoretical Physics, University of Copenhagen, where most of this work was carried out.

Appendix A

Experimental Justification for the Choice of the Model

It will be shown in this Appendix that each of the assumptions, on which the calculations in this paper are based, either follows from, or is at least consistent with, all experimentally well established facts regarding high energy collisions and also that once the not too infrequent occurrence of excited baryon states is admitted, one is led necessarily to the conclusion that the secondary cosmic radiation is dominated by the decay products of these baryon isobars.

The basic features of the model are:

1) A fireball is created at rest in the C-system of a nucleon-nucleon collision.*

2) The fireball evaporates, giving rise to n_F mesons, with an isotropic or moderately anisotropic angular distribution in its rest system. The ratio of pions to non-pions among the created particles does not change with energy.

3) In the evaporation process each of these mesons receives a momentum whose average value

$$\overline{P}_F = \frac{4}{\pi} \overline{P}_\perp \sim 450 \text{ MeV/c},$$

and whose maximum value

$$(P_F)_{\max} \approx 800 \text{ MeV/c}.$$

Since the average transverse momentum, $\overline{P}_\perp$, is known to be independent of energy, $\overline{P}_F$ also is independent of energy.

4) The number n_F of mesons evaporated from the fireball increases in proportion to the energy available in the C-system of the nucleon-nucleon collision.

5) The incident baryon emerges in some excited state with a probability, s, and decays by emitting on the average n_B pions, whose momentum in the baryon rest system is P_B.

Taken together, conditions 1) to 4) imply also that the incident baryon transfers on the average a constant fraction $(1 - \eta')$ of its energy to the fireball.

* The terms "fireball" and "pionization" were first introduced by G. Cocconi.

A.1 *The Pionization Process*

a) Energy spectra of mesons.

The energy distribution of mesons in the C-system of a nucleon-nucleon collision can be measured reliably provided the following conditions are satisfied:

α) The event occurs within or in the vicinity of the detector.
β) The energy of the incident particle is known with adequate accuracy.
γ) Observability and measurability are not functions of particle momentum, at least within a broad and well defined momentum range.

The only experiments carried out so far which satisfy all these conditions are those of the Moscow group [5a]. Here, the incident energy is measured by a total absorption calorimeter and the momenta of secondary charged particles by the curvature of their tracks in a magnet cloud chamber. They obtain $\overline{P}_F = 450$ MeV/c. Unfortunately, the experiments are limited to incident energies below 500 GeV.

If one relaxes the second condition and accepts also those experiments in which the incident energy is not measured but deduced from symmetry arguments and from the angular distribution of shower particles, one admits three other classes of experiments:

i) Emulsion experiments in which geometrically favourable events are selected and the momenta of all shower particles of a given collision are determined by scattering measurements. Experiments of this type, carried out by JAIN[5c], are discussed in ref. 1; the average momentum of particles belonging to the pionization process is $\overline{P}_F = 430$ MeV/c. Measurements of SCHEIN et al.[5b], analysed in an analogous manner, yield a somewhat higher value $\overline{P}_F = 600$ MeV/c. All the measurements refer to incident energies of order 2500 GeV.

ii) Cloud chamber experiments with magnetic field but without calorimeter. The experiment of HANSEN and FRETTER[5d] yields $\overline{P}_F = 470$ MeV at 100 GeV and $\overline{P}_F = 370$ MeV at 1000 GeV primary energy; that of MONTANET et al.[5e] yields $\overline{P}_F = 410$ MeV/c at 100 GeV.

iii) Cloud chambers containing enough absorbers both of low and of high atomic weight, so that all γ-rays are converted and observable and the energy of neutral pions produced in the interaction can be estimated from the ensuing showers. Such an experiment is that of LAL et al.[5f]

which yielded $\overline{P}_F = 430$ MeV/c. This experiment refers to a range of incident energies 20 GeV $\leqslant E \leqslant$ 150 GeV.

One sees that the experimental results on C-system momenta agree fairly well with each other and suggest the adopted value $\overline{P}_F \approx 450$ MeV/c.

The average transverse momentum of shower particles has been measured by many experimenters. Most particles have transverse momenta $P_\perp > m_\pi c$, i. e. they are relativistic in the C-system; the generally accepted value of $\overline{P}_\perp$ is independent of the incident energy and lies between 350 and 400 MeV/c. Thus, assumption 3), that $\overline{P}_f$ is energy independent and about 20–30 % larger than the average transverse momentum of shower particles, is in conformity with existing measurements.

b) Angular Distribution of Mesons.

The ratio of average total to average transverse momentum of shower particles in the C-system suggests a fairly isotropic angular distribution for the particles created in the pionization process. Direct measurements of the angular distribution of shower particles in the L-system support this conclusion.

Since the transverse momentum distribution shows that most particles are relativistic in the C-system, one may use, when transforming to the L-system, the approximation

$$\beta_c/\beta^* \approx 1.$$

With this approximation a fireball which emits particles with an angular distribution proportional to $\cos\theta^n$ produces an angular distribution in the L-system

$$N(x)\,dx = \tfrac{1}{2}\,d\left(\tanh^{n+1}(x + \log\gamma_c)\right), \tag{A.1}$$

where γ_c is the C-system energy of the incident nucleon in rest mass units and $x = \log\tan\theta$.

This distribution, when plotted against x, has two maxima:

$$x_m = -\log\left\{\gamma_c\left[\gamma_F \pm \sqrt{\gamma_F^2 - 1}\right]\right\}, \tag{A.2}$$

where $\gamma_F = \sqrt{1 + \frac{n}{2}}$. In the case of complete isotropy, ($n = 0$), the maxima coalesce into a single maximum at $x_m = -\log\gamma_c$, and one obtains a quasi-gaussian distribution with a root mean square deviation $\sigma = 0.39$. Normally,

however, fireballs must be expected to have intrinsic angular momentum, so that a certain measure of anisotropy in the evaporation and therefore double maxima in the log tan θ distribution should be common; they should become somewhat more pronounced with increasing energy. In many showers only a single hump can be seen, but when two maxima can be resolved, separations x_m are typically in the range $\frac{1}{2} < x_m < 2$, corresponding to C-system angular distributions between $\cos \theta^{0.125}$ and $\cos \theta^{2.5}$. The observed L-system angular distribution can therefore be interpreted as due to moderately anisotropic emission from a spinning fireball at rest; alternatively, (as one can see from eq. A.2) it can also be interpreted as being due to emission from two spinless fireballs moving parallel to the colliding nucleons with velocities in the C-system $\beta_F = \sqrt{\frac{n}{n+2}}$ and a fairly symmetric distribution for each hypothetical emission centre

$$\frac{\text{forward} - \text{backward}}{\text{forward} + \text{backward}} = 1 - 2\left(\frac{n}{n+2}\right)^{\frac{n+1}{2}} \approx 25 - 30\ \%.$$

The latter interpretation has been stressed especially by the Krakow group[21].

An upper limit for $(P_F)_{\max} \approx 800$ MeV/c is dictated by the observed absence or rarity of transverse momenta higher than this value[16].

Thus the assumptions 1), 2), and 3), i.e. creation of a fireball at rest emitting relativistic particles fairly isotropically, are in accord with experiments.

It is possible that fireballs have discrete mass values of the order of 2 GeV, as suggested by HASEGAWA[22], which may account for their absence or rarity in $p-p$ collisions at accelerator energies and for the large fluctuations observed in the angular distributions in the energy range of 100–300 GeV[5a].

A.2 *Relation between Multiplicity and Energy.*

Only few measurements of multiplicity exist where the energy of the incident nucleon is high and measurable and where at the same time interactions have been collected without strong bias against low multiplicity events. Below 30 GeV one has quite accurate results based on accelerator data. At 70,100, 300, and 1000 GeV one has cloud chamber data[5a, d, f]. The measured mean multiplicities may be slightly too high, because of some residual bias against very low multiplicities[5d].

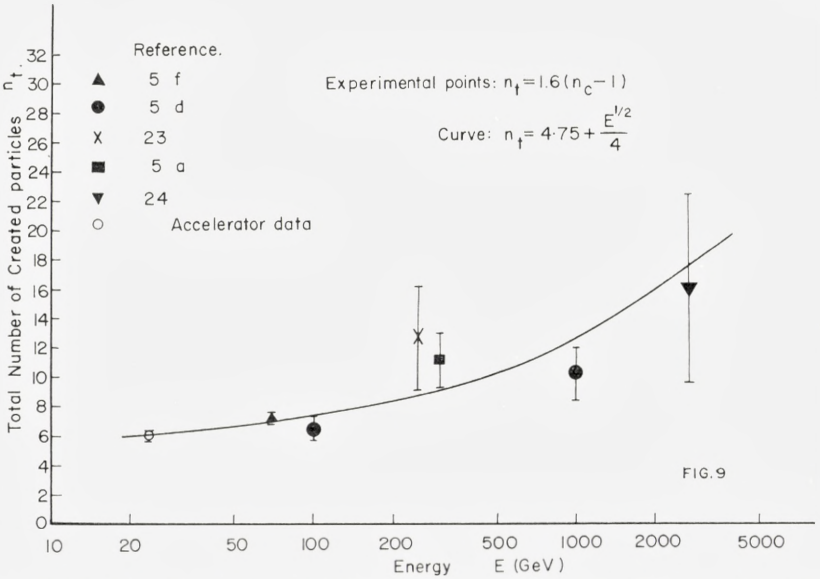

Fig. 9. The total number of created particles (charged and neutral), which on the average are emitted in a collision between a nucleon and a target nucleus of low atomic weight, is plotted as a function of the energy of the incident nucleon. The curve represents relation A.4.

At 250 GeV and at 2700 GeV, fairly reliable though statistically poor measurements of the multiplicity of charged particles were obtained by LOHRMANN et al.[23] and by ABRAHAM et al.[24] These authors used heavy primary nuclei whose energy per nucleon can be determined by various well-known methods; they then searched for interactions by scanning along the tracks of singly charged break-up products, i.e. fragments of the incident nucleus. One source of error could be the presence of deuterons or tritons among the break-up products, and another, a possible inclusion of interactions with heavy nuclei, silver or bromine, even when only events with less than five heavy ionizing particles are accepted. Both types of error may lead to some overestimate of multiplicity.

In Fig. 9 the total number of created particles, n_t, at various energies is shown; n_t has been calculated from the measured number of charged particles, n_c, using the relation

$$n_t = \left[0.8\left(\frac{3}{2}\right) + 0.2\,(2)\right] \cdot (n_c - 1) = 1.6\,(n_c - 1). \qquad (A.3)$$

The first term in eq. (A.3) refers to pions and the second to particles of isospin $\frac{1}{2}$. The relation (A.3) is energy independent, because the ratio of non-pions to pions among created particles is known to be independent of energy.[25]

The model requires a multiplicity relation of the form $n_t = 2\,sn_B + n_0 E^{\varrho}$, where s is the probability of baryon excitation, n_B the average number of mesons emitted in their decay, and $n_0 E^{\varrho}$ represents the creation of mesons in the pionization process.

The available data are neither numerous nor accurate enough to determine the form for the multiplicity law uniquely. A lead is provided by the approximate invariance of the nucleon energy spectrum with atmospheric depth, which suggests that the average energy loss of nucleons in collisions with nitrogen is energy independent. In this simplest case, the multiplicity relation must be of the form

$$n_t = 2\,sn_B + n_0\,E^{1/2}. \tag{A.4}$$

The curve in Fig. 9 represents eq. (A.4). A fit with experiment requires the following values for the parameters:

$$2\,sn_B = 4.75 \pm 0.25$$
$$n_0 = 1/4$$

i.e.

$$n_t = 4.75 + \frac{E^{1/2}}{4} \quad (E \text{ is in GeV}).$$

This relation between energy and number of created particles is consistent with existing measurements. (Although it cannot be deduced from the data in a unique manner, the data also do not warrant as yet the assumption of a more complicated relationship).

The multiplicity relation for high energy pions is still less certain than that for nucleons.

In the text (eq. IV.1 a), the relation

$$\nu_t = sn_B + \nu_0\,E_{\mathrm{GeV}}^{1/2}, \tag{IV.1 a}$$

which corresponds to a high degree of inelasticity, has been used as yielding sufficient multiplication of pions in the lower atmosphere to reproduce the measured pion/proton ratio near sea level[14]. Agreement can be obtained for $\nu_0 = 0.7$.

Using again $sn_B \approx 2.4$, the formula gives a reasonable multiplicity of created particles for $\pi - p$ collisions from 8 to 800 GeV, the energy region

in which measurements are available. Apart from the term due to baryon isobars, the two multiplicity laws $n_t(E)$ for nucleons (A.4) and $\nu_t(E)$ for pions (IV.1) are then related by

$$n_t(E) = \nu_t\left(E\frac{m_\pi}{M_p}\right),$$

i. e. the pionization in nucleon-nucleon collisions behaves as if it were due to completely inelastic collisions between two target masses, not very different from those of free pions; in that case,

$$\frac{n_0}{\nu_0} = \sqrt{\frac{m_\pi}{M_p}} = 0.39, \tag{A.5}$$

and the fractional energy going to pionization in nucleon-nucleon collisions is

$$1 - \eta' = \frac{m_\pi}{M_p} \sim 15\ \%,$$

values which are rather similar to those arrived at in Section IV.

If such collisions lead to boson quanta of rest mass $M_F \approx 2\,M_p$, as suggested by HASEGAWA[22], the creation of particles by pionization in nucleon-nucleon collisions will become important only if $m_\pi \gamma_c \approx M_p$ or $E \geqslant 100$ GeV; this may be connected with the smallness of the isotropic low energy pion component in the C-system of nucleon-nucleon collisions at accelerator energies[3].

A.3 *Relative Contributions to the Secondary Cosmic Ray Flux of Mesons from Isobar Decay and Mesons from Pionization.*

The two different processes by which particles are created in this model have now been specified sufficiently well so that their relative contribution to pion production in the atmosphere can be evaluated reliably and in a straightforward manner.

If the nucleon energy spectrum is of the form $dE/E^{\gamma+1}$, the spectrum of pions produced in the pionization process will be given by

$$\left.\begin{aligned} g_F(E)\,dE &= n_0\left(\frac{\varepsilon_F^2}{2\,M_p}\right)^{\gamma-\varrho} \frac{(1+\beta_F)^{2(\gamma+\frac{1}{2}-\varrho)}\,(1-\beta_F)^{2(\gamma+\frac{1}{2}-\varrho)}}{2\,\beta_F\,(\gamma+\frac{1}{2}-\varrho)} \frac{dE}{(E^2)^{\gamma+\frac{1}{2}-\varrho}} \\ &\approx \frac{n_0\,dE}{(\gamma+\frac{1}{2}-\varrho)\,(E^2)^{\gamma+\frac{1}{2}-\varrho}} \left(\frac{2\,\varepsilon_F^2}{M_p}\right)^{\gamma-\varrho}. \end{aligned}\right\} \tag{A.6}$$

ε_F is the energy of the mesons in the rest system of the fireball and β_F is their velocity.

This formula applies to isotropic emission. In the case of extreme anisotropy, where half the particles are emitted at 0° and half at 180°, g_F must be multiplied by $(\gamma + \frac{1}{2} - \varrho) \simeq 1.67$. An anisotropic emission of particles from the fireball does not, therefore, invalidate the conclusions derived in this section.

The spectrum of pions produced in isobar decay is given by

$$\left.\begin{aligned} g_B(E)\,dE &= sn_B \left(\frac{\eta'\,\varepsilon_B}{M_B}\right)^{\gamma} \frac{(1+\beta_B)^{\gamma+1} - (1-\beta_B)^{\gamma+1}}{2\,\beta_B\,(\gamma+1)} \frac{dE}{E^{\gamma+1}} \\ &\approx \frac{sn_B\,dE}{(\gamma+1)\,E^{\gamma+1}} \left(\frac{2\,\eta'\,\varepsilon_B}{M_B}\right)^{\gamma}, \end{aligned}\right\} \quad (A.7)$$

where ε_B is the energy of decay pions in the rest system of the baryon isobar of mass M_B and β_B is their velocity.

For the lowest lying isobar, (the $T = 3/2$, $J = 3/2$ state),
$\varepsilon_B \approx 270$ MeV and it is probably higher for heavier isobars.
$\varepsilon_F \approx 470$ MeV, as shown in experiments [5] discussed earlier.
η' may be taken to be approximately constant and equal to 0.8 (see Section IV).

Assuming an exponent $\gamma = 1.67$ for the primary energy spectrum, the relative contribution of the two processes, $\frac{g_B}{g_F}$, as a function of pion energy has been plotted in Fig. 1 for various assumed values of the product sn_B (sn_B is the average number of pions per collision from the deexcitation of the forward isobar). The constants n_0 and ϱ which characterize the size of the fireball have been chosen such that the total multiplicity of created particles n_t agrees with the experimental data for targets of low atomic number and for incident nucleon energies of 30 GeV ($n_t = 6$) and of 2700 GeV ($n_t \approx 18$). The appropriate multiplicity relation is shown on each of the curves in Fig. 1.

From these curves one sees that isobar pions dominate in the atmosphere at all energies if one accepts the value $sn_B \approx 2$; they dominate above 10 GeV for values of sn_B as small as 0.5. The dominance of the isobar decay process as a contributor to the flux of secondary cosmic ray particles increases rather rapidly with energy; of course this holds not only for mesons but also for their decay products, i.e. for muons, γ-rays, neutrinos etc. *Thus,*

unless baryon excitation is much less frequent in high energy collisions than it is at laboratory energies, the secondary cosmic radiation consists essentially of the decay products of baryon isobars and their progeny.

Appendix B

Formulae for the General Case of an Arbitrary Set of Isobaric States and for Arbitrary Decay Schemes

The simple model on which this paper is based permits a fairly rigorous calculation of nucleon and meson intensities and spectra at all points in the atmosphere.

We derive here formulae for the nucleon flux, the pion production spectrum, and the charge composition of nucleons and of pions, for the general case of an arbitrary set of baryon isobars, each with its own excitation probability and with arbitrary decay chains. Although the formulae involve summations over running indices, they can be evaluated easily for specific cases.

B.1 *The Nucleon Flux.*

Let the isobar of type r be produced with an interaction length λ_r and let it carry a fraction η_r' of the energy of the incident nucleon. The physical isobaric states are numbered 0,1,2,3 . . . r . . . in ascending order of mass, so that 0 denotes the nucleon ground state. Direct transition between any two of these states is assumed to lead to emission of a single boson which we shall take to be a pion, but which may equally well be a boson isobar, which subsequently disintegrates into pions.

A collision of type i is defined by specifying the isobar which is produced as well as the specific chain, σ, by which it decays into a nucleon and a number of pions. Taking account of the Poisson fluctuations in the number of collisions of each type, a simple extension of the argument given in chapter II (eq. II.4) shows that the attenuation length of nucleons in the atmosphere is given by the relation

$$\frac{1}{\Lambda} = \frac{1}{\lambda}(1 - \langle \eta^{\gamma} \rangle), \tag{B.1}$$

where

$$\langle \eta^{\gamma} \rangle = \sum_{i=1}^{\infty} \frac{\lambda}{\lambda_i} \eta_i^{\gamma}, \tag{B.2}$$

$$\frac{1}{\lambda} = \sum_{i=0}^{\infty} \frac{1}{\lambda_i}, \tag{B.3}$$

and η_i is the fraction of the incident energy retained by the nucleon in the i'th type of collision.

The problem is to express $\langle \eta^\gamma \rangle$ in terms of the mass and decay modes of the isobars. The averaging process may be broken up into three independent parts.

α)

For a particular isobar r and for decay to the ground state via a definite set of intermediate states $\sigma = (\sigma_1, \sigma_2, \sigma_3, \ldots)$ one can average over the angular distributions of decay and obtains

$$\langle \eta^\gamma \rangle_{r,\sigma} = \eta_r'^{\gamma} A_{r,\sigma_1} A_{\sigma_1,\sigma_2} A_{\sigma_2,\sigma_3} \cdots A_{\sigma_n,0} \tag{B.4}$$

$$A_{l,m} = \mathfrak{E}_{l,m} \frac{(1+\beta'_{l,m})^{\gamma+1} - (1-\beta'_{l,m})^{\gamma+1}}{2\,\beta'_{l,m}\,(\gamma+1)} \tag{B.5}$$

and $\mathfrak{E}_{l,m}$ is the fractional energy in the rest system of isobar "l" which is carried away by isobar "m", and $\beta'_{l,m}$ is its velocity.*

β)

The average over different decay modes involving different intermediate states σ is obtained by summing all possible expressions of the type

$$(bA)_{r,\sigma_1} (bA)_{\sigma_1,\sigma_2} (bA)_{\sigma_2,\sigma_3} \cdots (bA)_{\sigma_n,0} = Y_{r,\sigma}, \tag{B.6}$$

where $b_{l,m}$ is the branching ratio for transition from the state "l" to the state "m". The various terms in $Y_{r,\sigma}$ are subject to the restriction $\sigma_j < \sigma_{j-1}$.

The result of summing over all possible decay chains σ is designated by

$$Y_r = \sum_{\sigma} Y_{r,\sigma}. \tag{B.7}$$

One obtains

$$\eta_r^\gamma = \eta_r'^{\gamma}\, Y_r. \tag{B.8}$$

* The expression for $A_{l,m}$ is given here for isotropic decay; it is easily calculated for specified non-isotropic emission of decay products.

γ)

As the final step one must average over contributions from different isobars which have to be weighted according to excitation probability, e.g. with a weight factor which is inversely proportional to the mean free path for exciting the particular baryon state

$$\langle \eta^{\gamma} \rangle = \sum_{r=0}^{\infty} \frac{\lambda}{\lambda_r} \eta_r^{\gamma} = \sum_{r=0}^{\infty} \frac{\lambda}{\lambda_r} \eta_r'^{\gamma} Y_r. \tag{B.9}$$

B.2 *The Nucleon Charge Ratio.*

In Section III it is shown that the charge composition of nucleons at a depth x can be written as

$$\left(\frac{p-n}{p+n}\right)_x = \delta_x = \delta_0 \, e^{-(x/\lambda)\langle 2\eta^{\gamma} w \rangle}. \tag{III.2}$$

The average $\langle \eta^{\gamma} w \rangle$ may now be written as

$$\langle \eta^{\gamma} w \rangle = \sum_{r=0}^{\infty} \frac{\lambda}{\lambda_r} \eta_r'^{\gamma} \sum_{\sigma} w_{r,\sigma} Y_{r,\sigma}, \tag{B.10}$$

where $w_{r,\sigma}$ is the probability of producing a neutron from the decay of an isobar r, if the isobar is produced by an incident proton and decays through a particular decay chain σ. An explicit expression for $w_{r,\sigma}$ in terms of branching ratios is given later (in eq. B.19).

The proton and neutron spectra in the atmosphere are given by eq. (III.5).

B.3 *The Production Spectrum of Pions due to Isobar Decay.*

If ε is the fractional energy in the isobar rest system, carried away by a pion, and $n^{\pm}$ is the number of positive or negative pions emitted in the decay of an isobar produced in a collision of a proton with an air nucleus, then the production spectrum of charge pions at a depth x is given by

$$P_{\pi^{\pm}}(x, E)\, dx = \frac{dx}{\lambda} \frac{N(0, E)}{2} e^{-x/\Lambda} \left[\langle (n^{+} + n^{-})(\eta' \varepsilon)^{\gamma} \rangle \pm \delta x \langle (n^{+} - n^{-})(\eta' \varepsilon)^{\gamma} \rangle \right]. \tag{B.11}$$

The problem is to find an explicit expression for $\langle n^{\pm} (\eta' \varepsilon)^{\gamma} \rangle$ in terms of isobar masses and their excitation probabilities and decay branching ratios.

Starting with any isobar r, a pion may be obtained from any transition

$l \to m$, $(m < r)$. The state l itself, however, is in general the result of a series of decay combinations of states intermediate between r and l. Thus, in order to evaluate $\langle n^{\pm}(\eta'\varepsilon)^{\gamma}\rangle$, one must first sum over all possible decay schemes which can lead from the isobar r to the isobar l.

α)

$$\left.\begin{aligned}[n^{\pm}(\eta'\varepsilon)^{\gamma}]_{r,l,m} = \eta_r'^{\gamma}\,\alpha_{l,m}\,b_{l,m}\sum_{\sigma} V^{(\pm)\,l,m}_{r,\sigma}\,(bA)_{r,\sigma_1} \\ \times (bA)_{\sigma_1,\sigma_2}(bA)_{\sigma_2,\sigma_3}\ldots(bA)_{\sigma_n,l}\end{aligned}\right\} \quad \text{(B.12)}$$

$\alpha_{l,m}$ is defined in analogy to $A_{l,m}$, i. e.

$$\alpha_{l,m} = \varepsilon_{l,m}^{\gamma}\,\frac{(1+\beta_{l,m})^{\gamma+1} - (1-\beta_{l,m})^{\gamma+1}}{2\,\beta_{l,m}\,(\gamma+1)}, \quad \text{(B.13)}$$

where $\varepsilon_{l,m}$ is the fractional energy in the rest system of isobar l, carried away by a pion in the transition to the baryon state m, and $\beta_{l,m}$ is its velocity. The b's are the various branching ratios; $V^{(\pm)\,l,m}_{r,\sigma}$ is a function of the isospins of the initial and intermediate isobaric states and has the following significance: If a proton is incident on a target consisting of an equal number of protons and neutrons and emerges in the excited baryon state r, and if then this state r decays by a particular chain σ to the intermediate state l, then $V^{(\pm)\,l,m}_{r,\sigma}$ expresses the probability that the pion from the transition $l \to m$ is positively or negatively charged.

So far we have averaged over the angles of decay and combinations of intermediate states, which can lead from isobar r to isobar l.

β) Next, one must sum over all possible states l and m.

$$\langle n^{\pm}(\eta'\varepsilon)^{\gamma}\rangle_r = \sum_{l=1}^{r}\sum_{m=0}^{l-1}\langle n^{\pm}(\eta'\varepsilon)^{\gamma}\rangle_{r,l,m}. \quad \text{(B.14)}$$

γ) Lastly, one averages over contributions from different isobars, r, weighted according to the relative excitation probabilities λ/λ_r and obtains

$$\langle n^{\pm}(\eta'\varepsilon)^{\gamma}\rangle = \sum_{r=0}^{\infty}\frac{\lambda}{\lambda_r}\langle n^{\pm}(\eta'\varepsilon)^{\gamma}\rangle_r$$

or explicitly

$$\left.\begin{aligned}\langle n^{\pm}(\eta'\varepsilon)^{\gamma}\rangle = \sum_{r=0}^{\infty}\frac{\lambda}{\lambda_r}\,\eta_r'^{\gamma}\sum_{l=1}^{r}\sum_{m=0}^{l=1}\alpha_{l,m}\,b_{l,m} \\ \sum_{\sigma} V^{(\pm)\,l,m}_{r,\sigma}\,(bA)_{r,\sigma_1}(bA)_{\sigma_1\sigma_2}\ldots(bA)_{\sigma_n,l}.\end{aligned}\right\} \quad \text{(B.15)}$$

We define

$$\langle \delta_\pi \rangle = \frac{\langle n^+ (\eta' \varepsilon)^\gamma \rangle - \langle n^- (\eta' \varepsilon)^\gamma \rangle}{\langle n^+ (\eta' \varepsilon)^\gamma \rangle + \langle n^- (\eta' \varepsilon)^\gamma \rangle} \tag{B.16}$$

and

$$\langle B \rangle = \langle n^+ (\eta' \varepsilon)^\gamma \rangle + \langle n^- (\eta' \varepsilon)^\gamma \rangle . \tag{B.17}$$

(As shown in Section II, all secondary components of the cosmic radiation in the atmosphere are proportional to the parameter $\langle B \rangle$).

The charge composition of pions produced by the nucleon component at atmospheric depth x can be written as

$$\frac{P_{\pi^+} - P_{\pi^-}}{P_{\pi^+} + P_{\pi^-}} = \langle \delta_\pi \rangle \delta_x , \tag{III.5}$$

where δ_x gives the nucleon charge composition at that level (eq. III.2).

The total number of pions of either charge in the complete deexcitation of isobar r (produced in proton collisions) is given by

$$n_r^\pm = \sum_\sigma n_{r,\sigma}^\pm = \sum_\sigma \sum_{l=1}^{r} \sum_{m=0}^{l-1} b_{l,m} V^{(\pm) l,m}_{r,\sigma} b_{r,\sigma_1} b_{\sigma_1,\sigma_2} \ldots b_{\sigma_n,1} . \tag{B.18}$$

For a neutron collision with a charge symmetric target the isospin functions transform as

$$V^{(+)} \to V^{(-)},$$
$$V^{(-)} \to V^{(+)}.$$

The quantity $w_{r,\sigma}$ appearing in eq. (B.10), i. e. the charge exchange probability for a nucleon excited to the state r and decaying via a particular mode σ, can now be expressed by

$$w_{r,\sigma} = q_r + \sum_{l=1}^{r} \sum_{m=0}^{l-1} b_{l,m} [V^{(+) l,m}_{r,\sigma} - V^{(-) l,m}_{r,\sigma}] b_{r,\sigma_1} b_{\sigma_1,\sigma_2} \ldots b_{\sigma_n,1} , \tag{B.19}$$

where q_r is the average charge difference between the incident proton and the resulting isobar r.

The expression for $\langle B \rangle$ (eqs. B.15, 17) can be greatly simplified in the special case that the isobar decays by cascade in small steps of comparable size such that $2 m_\pi < (M_l - M_m = \Delta) \ll M_B$.

Then

$$\mathfrak{E}_{l,m} = \frac{M_l^2 + M_m^2 - m_\pi^2}{2 M_l^2} \approx 1 \quad \text{and} \quad A_{l,m} \approx 1\,, \tag{B.20}$$

while

$$\left.\begin{aligned} \varepsilon_{l,m} &= \frac{M_l^2 - M_m^2 + m_\pi^2}{2 M_l^2} \approx \frac{\Delta}{M_l}, \quad \text{and, therefore,} \\ \alpha_{l,m} &= \frac{\left(\frac{2\,\Delta}{M_l}\right)^\gamma}{\gamma + 1}\,. \end{aligned}\right\} \tag{B.21}$$

Equations (B.15, 17) reduce then to

$$\langle B \rangle = \sum_{r=0}^{\infty} \frac{\lambda}{\lambda_r} \frac{[2\,(\eta' \varepsilon_r)]^\gamma}{(\gamma + 1)} (n_r^+ + n_r^-)\,, \tag{B.22}$$

where ε_r is the fractional energy in the isobar rest system carried away by a pion.

Appendix C

Kaon Production from Non-Strange Isobars and the Resulting Muon Flux

The kaon flux due to non-strange isobars can be calculated in a manner similar to the pion flux. (As explained in the text, the most prominent decay mode will be $N^* \to K + Y$ and therefore the flux of anti-kaons produced in this manner is expected to be small and will be neglected).

Since kaons do not multiply in subsequent interactions, it is only necessary to set $q_+ + q_- = 0$ in eq. (II.11) and replace the values for mass, lifetime, and interaction mean free path of pions by those of charged (i. e. positive) kaons. Apart from the factor denoting the branching ratio for this decay mode, one obtains the flux of charged kaons

$$F_k = \frac{S_0 \langle B_k \rangle x}{E^{\gamma+1}} \frac{x}{\lambda} e^{-\frac{x}{\lambda_k}} \sum_{i=0}^{\infty} \frac{\left(1 - \frac{\lambda_k}{\Lambda}\right)^i}{i!\,(1 + i + u_k)}, \tag{C.1}$$

$$\text{where } u_k = \frac{h_0 m_{k^+}}{c \tau_{k^+} E} = \frac{e_{k^+}}{E} = \frac{945}{E_{\text{GeV}}}\,; \tag{C.2}$$

$\langle B_k \rangle$ is defined in analogy with $\langle B \rangle$ (eqs. B. 15, 17)

$$\langle B_k \rangle = \frac{1}{2} s\,(\eta' \varepsilon_k)^{\gamma} \frac{(1+\beta_k)^{\gamma+1} - (1-\beta_k)^{\gamma+1}}{2\,\beta_k\,(\gamma+1)}, \tag{C.3}$$

where $\varepsilon_k = \dfrac{M_B{}^2 - M_Y{}^2 + m_k{}^2}{2\,M_B{}^2}$ is the fractional energy in the rest frame of the isobar, carried away by the kaon, and β_k is its velocity.

For $\lambda_k \approx \Lambda$

$$F_k = \frac{S_0 \langle B_k \rangle}{E^{\gamma}\,(E + \mathrm{e}_{k+})} \frac{x}{\lambda} e^{-\frac{x}{\Lambda}}. \tag{C.3 a}$$

In analogy with the procedure in Section II.4 one obtains the production spectrum of muons

$$P_{\mu_k}(x, E) = 0.69 \frac{\mathrm{e}_{k+}}{E} \frac{F_k[x, (r_k E)]}{x}, \tag{C.4}$$

where 0.69 represents the fraction of muons which arise directly from kaon decay without an intermediate pion, i. e. mostly $k_{\mu 2}$ decays.

$$r_k = \frac{2}{1 + \left(\dfrac{m_\mu{}^2}{m_k}\right)} \left[\frac{2\,\beta\,(\sigma+1)}{(1+\beta)^{\sigma+1} - (1-\beta)^{\sigma+1}}\right]^{\frac{1}{\sigma}} \approx (\sigma+1)^{\frac{1}{\sigma}} = 1.7$$

(the "best fitting" exponent σ is defined by

$$\left.\begin{aligned} -(\sigma+1) &= \frac{\partial \log F_k}{\partial \log E} = \gamma + 1 + \frac{\mathrm{e}_{k+}}{E + \mathrm{e}_{k+}} \\ &\approx 3.2 \quad (\text{for } E \approx 500\ \mathrm{GeV}). \end{aligned}\right\} \tag{C.5}$$

In anaolgy with (II.17) the resulting muon flux is

$$F_{\mu_k}(x, E) = 0.69 \frac{S_0 \langle B_k \rangle}{\lambda} \frac{\mathrm{e}_{k+}}{r_k{}^{\gamma+1}} \left(\frac{\lambda_k E}{x}\right)^{v} \frac{\Gamma(v+1)}{E'{}^{\gamma+2+v} \left(1 + \dfrac{\mathrm{e}_{k+}}{r_k E'}\right)}, \tag{C.6}$$

where $E' = \mathrm{E} + b(x - \bar{x})$ as defined in (II.17).

Appendix D

Corrections at the Low Energy End of the Nucleon Spectrum

In its simplest form the model treated here assumes that the fractional energy loss of nucleons (which is about 45 $^0/_0$) does not depend on the collision energy; therefore the primary power law is preserved throughout the atmosphere. This assumption can be valid only as long as the energy lies sufficiently high above the threshold for producing the most prominent isobaric states (i. e. $\gtrsim 10$ GeV). Below that energy, meson production drops and collisions tend to become more elastic. In this region the fractional loss of total energy becomes small, but the fractional loss of kinetic energy is known to remain of the order of 50 $^0/_0$. Thus, if the spectrum of secondary nucleons is written in the form

$$N\,dE \sim \frac{dE}{(E-M)^{\gamma+1}} \quad \text{rather than} \quad \frac{dE}{E^{\gamma+1}},$$

it will represent the nucleon flux down to lower values of the energy without appreciably changing the results obtained in the region of higher energies (eq. II.2). For this reason, the uncorrected nucleon spectrum is represented in Fig. 2a as a power law in kinetic, rather than total energy.

Before extrapolating the proton spectrum (eq. III.3) into the low energy region, there are two other corrections to be made:

a) Energy loss by ionization and
b) The production of recoil nucleons.

a) The loss of energy due to ionization.

The probability $p(y, x)$, that the average incident primary cosmic ray nucleon (charge composition δ_0) is a proton at depth y and also at the point of observation x, is given by

$$p(y, x) = \frac{1}{4}\left\{1 + \delta_0\, e^{-\frac{2xw}{\lambda}} + \delta_0\, e^{-\frac{2yw}{\lambda}} + e^{-\frac{2(x-y)w}{\lambda}}\right\}, \tag{D.1}$$

where W is the charge exchange probability for nucleon collisions in air. Assuming a constant rate of energy loss b GeV/(g/cm^2) for the fraction of the path in which the nucleon is charged, one obtains the mean energy loss ΔE for a proton arriving at x:

$$\left.\begin{aligned} \Delta E &= \frac{2\,b}{1+\delta_0\, e^{-\frac{2\,xw}{\lambda}}} \int_0^x dy\, p\,(x,\,y) \\ &= \frac{b}{2}\left\{ x + \frac{1+\delta_0}{2}\,\frac{\lambda}{w}\,\frac{1-e^{-\frac{2\,xw}{\lambda}}}{1+\delta_0\, e^{-\frac{2\,xw}{\lambda}}} \right\} \\ \Delta E &\approx \frac{b}{1}\left[x + \frac{1+\delta_0}{2}\,\frac{\lambda}{w} \right] \quad \text{for} \quad 2\,x \gg \frac{\lambda}{w}\,. \end{aligned}\right\} \qquad \text{(D.2)}$$

Using $b = 2$ MeV/(g/cm^2) and the constants given in the text ($\lambda = 75$ g/cm^2, $\delta_0 = 0.74$, $w = 0.30$), one finds that the average proton has lost

1.25 GeV when reaching sea level and
0.95 GeV when reaching mountain altitude (700 g/cm^2).

The corrected flux of protons in the atmosphere is therefore given by

$$N_p\,(x,\,E) = \frac{S_0\, e^{-x/\Lambda}}{2\,(E - M + \Delta E)^{\gamma+1}} \left[1 + \delta_0\, e^{-\frac{2\,xw}{\lambda}\left(1-\frac{\lambda}{\Lambda}\right)} \right]. \qquad \text{(D.3)}$$

This relation is plotted in Fig. 2.

b) Contributions from recoil nucleons.

In order to compare the flux of low energy protons near sea level with the calculated flux, one must add to the flux of extraterrestrial protons given by eq. (D.3) a contribution from recoils, i. e. nucleons of terrestrial origin which originally formed part of air molecules. These recoils receive kinetic energies upto ~ 3 GeV when the collision energy is low ($\lesssim 10$ GeV); in more energetic collisions, the energy which a recoiling baryon receives approaches a small constant value, and the resulting recoil nucleon has a maximum energy

$$E_{\max} = \left[\frac{M_B}{2\,\eta'\,M_P}\,(1 + P_\perp^2) + \frac{\eta'\,M_P}{2\,M_B} \right] \left[\frac{M_B^2 + M_P^2 - n_B^2\, m_\pi^2}{2\,M_B} \right] (1 + \beta_B\,\beta_P). \qquad \text{(D.4)}$$

Here, $P_\perp$ is the transverse momentum taken up by the baryon in the excitation process, and $1 - \eta'$ is the fractional collision energy used up in the creation of particles via the pionization process. The first bracket represents the energy in rest

mass units of the baryon in the L-system, and the second bracket represents the energy of the nucleon in the baryon rest system, on the assumption that it decays into n_B pions of equal energy. β_B and β_P are the corresponding velocities of the baryon and the nucleon.

Setting $P_\perp \approx 500$ MeV/c and using the values derived in the text ($\eta' = 0.75$, $M_B = 2200$ MeV and $n_B = 3.5$), one finds a recoil kinetic energy

$$T_{\max} \approx 3.0 \text{ GeV};$$

if no excitation takes place ($M_B = M_P$, $n = 0$, $\beta_P = 0$)

$$T_{\max} \approx 250 \text{ MeV}.$$

Thus it is to be expected that, when reaching sea level, recoil nucleons will make a contribution to the proton flux mainly in the non-relativistic region. Fig. 2 shows that, in the energy region 0–500 MeV, there is in fact an excess of observed protons over and above the flux of extra-terrestrial protons calculated according to eq. (D.3). The excess is of the correct order of magnitude to be attributed to particles of terrestrial origin.

References

(1a) W. J. Fickinger, E. Pickup, D. K. Robinson, E. O. Salant, Phys. Rev. **125**, 2082 (1962).

(1b) E. L. Hart, R. I. Louttit, D. Luers, T. W. Morris, W. J. Willis, S. S. Yamamoto, Phys. Rev. **126**, 747 (1962).

(2a) R. M. Sternheimer and S. J. Lindenbaum, Phys. Rev. **105**, 1874 (1957); **123**, 333, (1961).

(2b) M. Olsson and G. B. Yodh, Phys. Rev. Letters **10**, 353 (1963).

(3) G. Damgaard and K. Hansen, Reported at the Sienna Conference on Elementary Particles (1963).

(4a) B. Peters, Nuovo Cimento **23**, 88 (1962).

(4b) B. Peters, Proc. International Conference on High Energy Physics at CERN, 623 (1962).

(5a) V. V. Guseva, N. A. Dobrotin, N. G. Zelevinskaya, K. A. Kotelnikov, A. M. Lebedev, S. A. Slavatinsky, Journal Phys. Soc. Japan, **17**, Suppl. A-**3**, 375 (1962).

(5b) M. Schein, D. M. Haskin, E. Lohrmann, M. W. Teucher, Proc. Int. Conf. On Cosmic Radiation, I, 6, Moscow (1960).

(5c) P. L. Jain, Nuovo Cim. **22**, 1104 (1961) and Phys. Rev. **125**, 679 (1962).

(5d) L. F. Hansen and W. B. Fretter, Phys. Rev. **118**, 812 (1960).

(5e) L. Montanet, J. A. Newth, G. Petrucci, R. A. Salmeron, A. Zichichi, Nuovo Cimento **17**, 166 (1960).

(5f) S. Lal, Yash Pal, R. Raghavan, Nuclear Physics **31**, 415 (1962).

(6) C. J. Waddington, Progress in Nuclear Physics **8**, 3 (1960).

(7) D. Lal, Yash Pal, B. Peters, Proc. Indian Academy of Sciences **38**, 398 (1953).

(8) B. Peters, Nuovo Cim. **22**, 800 (1961).

(9) A. Ashmore, G. Cocconi, A. N. Diddens, A. M. Wetherell, Phys. Rev. Letters **5**, 576 (1960).

(10a) K. Sitte, Handbuch der Physik XLVI/I, p. 174, Springer Verlag (1961).

(10b) G. Bozoki, E. Fenyes, L. Janossy, Nuclear Physics **33**, 326 (1962).

(11) G. Von Dardel, D. H. Frisch, R. Mermod, R. H. Milburn, P. A. Piroué, M. Vivargent, G. Weber, K. Winter, Phys. Rev. Letters **5**, 333 (1960).

(12) G. Brooke, P. J. Hatman, P. J. Taylor, A. W. Wolfendale, J. Phys. Soc. Japan, Suppl. A **3**, 311 (1962). Also G. Brooke, Priv. comm.

(13) F. B. McDonald and W. R. Webber, Phys. Rev. **115**, 194 (1959).

(14) S. Lal, R. Raghavan, T. N. Rangaswamy, B. V. Sreekantan A. Subramanian. Private Communication (1963). These results are obtained by using an air Cerenkov counter in combination with a cloud chamber and a total absorption spectrometer.

(15a) M. Gardiner, D. G. Jones, F. E. Taylor, A. W. Wolfendale, Proc. Phys. Soc. **80**, 697 (1962).

(15b) P. J. Hayman and A. W. Wolfendale, Proc. Phys. Soc. **80**, 710 (1962).
(16) O. Minakawa, Y. Nishimura, M. Tsuzuki, H. Yamanouchi, H. Aizu, H. Hasegawa, Y. Ishii, S. Tokunaga, Y. Fujimota, S. Hasegawa, J. Nishimura, K. Niu, K. Niashikwa, K. Imaeda, M. Kazuno, Nuovo Cim. Suppl. Vol. II, No. 1. 125 (1959). See also ref. 5d.
(17a) P. J. Hayman and A. W. Wolfendale, Nature **195**, 166 (1962).
(17b) M. J. Campbell, H. S. Murdoch, K. W. Ogilvie, H. D. Rathgeber, Nuovo Cimento **28**, 885 (1963).
(17c) Y. Kamiya, H. Uneo, S. Sagisaka, Y. Sekido, Private Communication (1963).
(18a) G. B. Owen and J. G. Wilson, Proc. Phys. Soc., A **64**, 417 (1951).
(18b) I. Filosofo, E. Pohl, J. Pohl-Ruling, Nuovo Cimento **12**, 809 (1954).
(18c) R. B. Brode and M. J. Weber, Phys. Rev. **99**, 610 (1955).
(18d) J. Pine, R. J. Davisson, K. Greisen, Nuovo Cimento **14**, 1181 (1959).
(18e) J. E. R. Holmes, B. G. Owen A. L. Rodgers, Proc. Phys. Soc. **78**, 805 (1961).
(19) F. Ashton and A. W. Wolfendale, Proc. Phys. Soc. **81**, 522 (1963).
(20) G. Clark, I. Escobar, Murakami, K. Suga, Proc. Fifth Interamerican Seminar on Cosmic Rays, Vol. II, XLIII I, La Paz (1962).
(21) P. Cziok, T. Coghen, J. Gierula, R. Holynski, A. Jurak, M. Miesowicz, T. Saniewska and J. Pernegr, Nuovo Cimento **10**, 741 (1958).
(22) S. Hasegawa, INSJ 48, University of Tokyo (1962). See also the report of Hasegawa's work, Proc. International Conference on High Energy Physics, 644 (1963).
(23) E. Lohrmann, M. W. Teucher, M. Schein, Phys. Rev. **122**, 672 (1961).
(24) F. Abraham, J. Kidd, M. Koshiba, R. Levi Setti, C. H. Tsao, W. Wolter, C. L. Deney, R. L. Fricken, R. W. Huggett (Preprint) (1963).
(25) D. H. Perkins, Proc. Int. Conf. on Theoretical Aspects of Very High Energy Phenomena, p. 99, CERN June (1961).
(26a) K. Greisen and W. D. Walker, Phys. Rev. **90**, 915 (1953).
(26b) M. Cervasi, G. Fidecaro, L. Mezzetti, Nuovo Cimento **1**, 300 (1955).
(26c R. H. Brown, U. Camerini, P. H. Fowler, H. Heitler, D. T. King, C. F. Powell, Phil. Mag. **40**, 862 (1949).
(26d) H. L. Kasnitz and K. Sitte, Phys. Rev. **94**, 977 (1954).
(26e) L. Farrow, Phys. Rev. **107**, 1687 (1957).
(26f) J. G. Duthie, C. M. Fisher, P. H. Fowler, A. Kaddoura, D. H. Perkins, K. Pinkau and W. Wolter: Phil. Mag. **6**, 126 (1961). See also ref. 5f.
(27) P. K. Malhotra, Private communication (1963).
(28) G. Cocconi, Phys. Rev. **111**, 1699 (1958).

Færdig fra trykkeriet den 10. juni 1964.