Les Houches 2 - 1952

QUANTUM MECHANICS

Notes on lectures by

J. M. Luttinger
Department of Physics
University of Wisconsin

Delivered at the Ecole d'Eté de
Physique Théoriques, Les Houches, France

July 15 - August 20, 1952

Table of Contents

III) Special Systems

IV) Approximation Methods

V) Spin and Identical Particles

VI) Scattering Problems

VII) Relativistic Theory of the Electron

1. Mathematical Preliminaries

We shall begin our discussion by studying vectors in an n-dimensional space, later generalizing our results for the case of n being infinitely large. By a vector we shall mean an ordered array of n complex numbers. That is, a vector ψ is given by giving the n numbers $x_1, x_2, x_3, \ldots\ldots x_n$, which we shall call its components. We will often write

$$\psi = (x_1, x_2, \cdots, x_n)$$

We define several operations:

1. Multiplication by a constant λ:

$$\lambda\psi \equiv (\lambda x_1, \lambda x_2, \cdots, \lambda x_n)$$

2. Addition of two vectors: If $\psi = (x_1, \cdots, x_n)$ and $\phi = (y_1, \cdots, y_n)$

$$\psi + \phi \equiv (x_1 + y_1, \cdots, x_n + y_n)$$

3. Scalar product of two vectors: The scalar product of two vectors ψ and ϕ is denoted by (ϕ, ψ) and defined by

$$(\phi, \psi) \equiv \sum_{i=1}^{n} y_i^* x_i$$

The star indicates the complex conjugate. The scalar product has the following properties, at once evident from the definition:

$$(\phi, \psi_1 + \psi_2) = (\phi, \psi_1) + (\phi, \psi_2)$$

$$(\phi, \psi) = (\psi, \phi)^*$$

$$(\phi, \lambda\psi) = \lambda(\phi, \psi)$$

$$(\lambda\phi, \psi) = \lambda^*(\phi, \psi)$$

$$(\psi, \psi) = \sum_i x_i^* x_i = \sum_i |x_i|^2 \geq 0$$

Therefore (ψ,ψ) is real and ≥ 0. We shall call $\sqrt{(\psi,\psi)}$ the <u>norm</u> of the vector ψ. Only if the vector is identically zero can the norm be zero.

4. <u>Schwartz Inequality</u>: Consider the vector

$$\chi = \phi + \lambda\psi$$

$$(\chi,\chi) = (\phi+\lambda\psi, \phi+\lambda\psi) = (\phi,\phi) + \lambda(\phi,\psi) + \lambda^*(\psi,\phi) + \lambda^*\lambda(\psi,\psi)$$

$$= (\psi,\psi)\left(\lambda + \frac{(\psi,\phi)}{(\psi,\psi)}\right)\left(\lambda^* + \frac{(\psi,\phi)^*}{(\psi,\psi)}\right) + (\phi,\phi) - \frac{(\psi,\phi)(\phi,\psi)}{(\psi,\psi)} .$$

Choose $\lambda = -(\psi,\phi)/(\psi,\psi)$. Then we obtain

$$(\chi,\chi) = (\phi,\phi) - \frac{(\psi,\phi)(\phi,\psi)}{(\psi,\psi)} \geq 0$$

since the norm of any vector is greater than or equal to zero. The equality can only hold if $\chi = 0$. Thus we have $(\psi,\psi)(\phi,\phi) \geq |(\psi,\phi)|^2$, equality holding if $\chi = 0$ i.e. if $\phi = -\lambda\psi$ This is known as Schwartz's inequality. When two vectors are proportional to each other we say they are <u>parallel</u>, and thus the equality sign only holds when ϕ is parallel to ψ.

By analogy with ordinary vector analysis where the cosine of the angle between the vectors $\vec{A}$ and $\vec{B}$ is given by

$$\cos\theta_{A,B} = \frac{\vec{A}\cdot\vec{B}}{\sqrt{\vec{A}\cdot\vec{A}}\,\sqrt{\vec{B}\cdot\vec{B}}}$$

we can define

$$\cos^2\theta_{\psi,\phi} = \frac{|(\psi,\phi)|^2}{\left(\sqrt{(\psi,\psi)(\phi,\phi)}\right)^2}$$

as the angle between the vectors ψ and ϕ. Schwartz's inequality then requires that

$$\cos^2\theta_{\psi,\phi} \leq 1$$

so that the angle $\theta_{\psi,\phi}$ so defined is real. When the scalar product of two vectors is zero, i.e., when $(\psi,\phi)=0$ then $\theta_{\psi,\phi}=90°$ and we say that the vectors are perpendicular or orthogonal.

Unit or Basic Vectors

Consider the set of n vectors

$$e_1 = (1, 0, 0, \cdots, 0)$$
$$e_2 = (0, 1, 0, \cdots, 0)$$
$$\vdots$$
$$e_n = (0, 0, \cdots, 1).$$

Then any vector may clearly be written as

$$\psi = \sum_i x_i e_i$$

For these vectors we clearly have $(e_j, e_i) = \delta_{ij}$ where δ_{ij} is the Kronecker symbol defined so that

$$\delta_{ij} = \begin{matrix} 1 & i=j \\ 0 & i \neq j \end{matrix}$$

We shall call any set of vectors e_i which satisfy the requirements; (i) that any vector is a linear combination of them, and (ii) that their scalar products $(e_j, e_i) = \delta_{ij}$ a set of basic vectors or unit vectors. Requirement (i) we shall call the completeness requirement and requirement (ii) the orthogonality requirement.

Now suppose we have two bases e_i and e_i'. Since e_i' is complete we can always write

$$e_j = \sum_i e_i' S_{ij}$$

The S_{ij} are given by

$$S_{ij} = (e_i', e_j) = \frac{(e_i', e_j)}{\sqrt{(e_i', e_i')(e_j, e_j)}}$$

so that the S_{ij} are the generalization of the notion of direction cosines in ordinary vector analyses.

If we write

$$e'_j = \sum_i e_i U_{ij} \qquad \text{we have}$$

$$U_{ij} = (e_i, e'_j) = (e'_j, e_i)^* = (S_{ji})^* = S^*_{ji}$$

Thus

$$e'_j = \sum_i e_i S^*_{ji}$$

The S_{ij} must satisfy the following condition:

Since

$$(e'_i, e'_j) = \delta_{ij} = \Big(\sum_k S^*_{ik} e_k , \sum_l S^*_{jl} e_l\Big)$$

$$= \sum_{k,l} S_{ik} S^*_{jl} (e_k, e_l)$$

$$= \sum_{k,l} S_{ik} S^*_{jl} \delta_{kl}$$

$$\delta_{ij} = \sum_k S_{ik} S^*_{jk}$$

It is very convenient to introduce the notation of matrices at this point. We shall denote the square array of elements S_{ij} by the letter S, i.e.

$$S = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & & & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{pmatrix}$$

We define also the <u>transpose of a matrix</u> as the matrix with rows and columns interchanged, and denote it by $\tilde{S}$. That is

$$(\tilde{S})_{ij} = S_{ji}$$

By the Adjoint of a matrix we shall mean the matrix obtained by taking the transpose of the original matrix and then the complex conjugate of every element. If we denote the adjoint by S^+ we have

$$(S^+)_{ij} = S^*_{ji}$$

Finally we define the product of two matrices S and U as a matrix T whose elements are defined by

$$(SU)_{ij} = T_{ij} = \sum_k S_{ik} U_{kj}$$

In terms of this notation we can write our relationship between the elements of S

$$\sum_k S_{ik} S^*_{jk} = \sum S_{ik} (S^+)_{kj} = (SS^+)_{ij} = \delta_{ij}$$

Or, introducing the unit matrix I definied as the matrix with elements $I_{ij} = \delta_{ij}$ we get $SS^+ = I$

The matrix I plays the same role in matrix multiplication as the number 1 in ordinary multiplication. That is, for any matrix S we have

$$IS = SI = S$$

This is very easily proved for the elements

$$(IS)_{ij} = \sum_k I_{ik} S_{kj} = \sum \delta_{ik} S_{kj} = S_{ij}$$

$$(SI)_{ij} = \sum_k S_{ik} I_{kj} = \sum S_{ik} \delta_{kj} = S_{ij}$$

We can use the matrix I to define the reciprocal of a matrix U, which we shall denote by U^{-1}. U^{-1} is defined by $U^{-1}U = UU^{-1} = I$, if we can find a matrix which satisfies this requirement. It is not necessary for any arbitrary matrix U to have a reciprocal, but it is easily shown that if the determinant of the matrix U $\neq 0$ i.e., if

$$\det(U) = \begin{vmatrix} U_{11} & \cdots & U_{1n} \\ \vdots & & \\ U_{n1} & \cdots & U_{nn} \end{vmatrix} \neq 0$$

then U has a reciprocal. We can restate our condition on S then by simply saying that $S^{-1} = S^{+}$ A matrix which satisfies this condition is said to be a unitary matrix.

Linear Transformations

Let us now consider an operation in the vector space which transforms every vector into another vector. In particular let $\psi \to \psi'$ We may write this symbolically as

$$\psi' = A\psi$$

indicating the result of this operation--which we call A--applied to the vector ψ gives the vector ψ'. We shall call A a linear operator if it satisfies the requirements:

$$A(a\psi) = a\,A\psi$$

$$A(\phi + \psi) = A\phi + A\psi$$

for any two vectors ϕ and ψ, and any complex number a. Any linear operator is completely characterized as soon as we know its effect on the basis vectors, since

$$A\psi = A\sum x_i e_i = \sum x_i \, Ae_i$$ by the linearity.

From this we see at once that a linear operator is known completely as soon as a certain set of numbers a_{ij} are known. By completeness we must have

$$Ae_i = \sum e_j a_{ji}$$ where the a_{ji} are given as soon as we know the operator A. They are clearly given by

$$a_{ji} = (e_j, Ae_i)$$

and are called the matrix elements of the operator A.

<u>Effect of a change of basis</u> on the matrix representing A.

If we use instead of the basis e_i the basis e_i' the new matrix elements will be a_{ij}', and will be given by

$$a_{ij}' = (e_i', Ae_j') = \sum_{k,l} (S_{ik}^{+} e_k, A S_{jl}^{+} e_l)$$

$$= \sum_{k,l} S_{ik} S_{jl}^{+} (e_k, Ae_l)$$

$$= \sum_{k,l} S_{ik} a_{kl} (S^{+})_{lj} = \sum_{k,l} S_{ik} a_{kl} (S^{-1})_{lj}$$

on using the unitary character of S. Thus $a_{ij}' = (SaS^{-1})_{ij}$ by the rules of matrix multiplication. In terms of the matrices we have

$$a' = SaS^{-1}$$

Such a trnasformation of a matrix with a unitary matrix is called a <u>unitary transformation</u> or <u>canonical transformation</u>.

Returning to Aψ we have.

$$A\psi = \sum_{i,j} x_j a_{ij} e_i$$

or $$(A\psi)_i = \sum_j a_{ij} x_j = \psi_i' = x_i'$$

It is often very convenient to interpret this as a matrix equation. For this purpose we imagine ψ to be a matrix with one column and n rows

$$\psi = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

and A = (a_{ij}) a matrix with n columns and n rows. Then the above equation is just the equation $\psi' = A\psi$ using the ordinary rules of matrix multiplication.

<u>Sum</u> of two linear operators A and B: If A sends $\psi \to \psi'$ B sends $\psi \to \psi''$ then we define A + B as that operator which sends $\psi \to \psi' + \psi''$. Or $(A+B)\psi \equiv A\psi + B\psi$. If we call

A + B = C , then the matrix elements of C are

$$C_{ij} = (e_i, C e_j) = (e_i, (A+B) e_j) = (e_i, A e_j + B e_j)$$

$$= (e_i, A e_j) + (e_i, B e_j)$$

$$= a_{ij} + b_{ij}$$

Product of two linear operators A and B: If A sends $\psi \to \psi'$ and B sends $\psi' \to \psi''$, then the operator C = B A is the operator which sends $\psi \to \psi''$. The matrix elements of C are given as follows

$$C_{ij} = (e_i, C e_j) = (e_i, B(A e_j)) = \sum_k a_{kj} (e_i, B e_k)$$

$$= \sum_k b_{ik} a_{kj}$$

which is the ordinary rule for matrix multiplication of the matrix B and A. So the matrix representing the product of two linear operators is just the product of their matrices.

Special Operators.

We define the adjoint A^+ of an operator A as the operator whose matrix elements are represented by the adjoint of the matrix representing A. That is, if $(A)_{ij} = a_{ij}$ then $(A^+)_{ij} = a^*_{ji}$.

we have at once that

$$(\phi, A\psi) = \sum y_j^* a_{ji} x_i = \sum (a^*_{ji} y_j)^* x_i = (A^+\phi, \psi),$$

which is a fundamental property of the adjoint operator. By means of this we can at once find $C^+ = (AB)^+$

$$(\phi, C\psi) = (\phi, AB\psi) = (A^+\phi, B\psi) = (B^+A^+\phi, \psi)$$

$$= (C^+\phi, \psi)$$

so that

$$(AB)^+ = B^+A^+$$

Similarly

$$(\lambda A)^+ = \lambda^* A^+$$

$$(A+B)^+ = A^+ + B^+$$

$$(A^+)^+ = A$$

An operator which is its own adjoint is said to be self-adjoint or hermitian. An hermitian operator satisfies:

$$A^+ = A$$

If we have any operator A, then the combinations $A + A^+$ and $i(A - A^+)$ are clearly always hermitian. Hermitian operators also have the property that for any vector ψ the scalar product $(\psi, A\psi)$ is real

$$(\psi, A\psi) = (A\psi, \psi)^* = (\psi, A^+\psi)^+ = (\psi, A\psi)^*$$

This fundamental relationship will, as we shall see, give hermitian operators an all important role in quantum theory. An operator U is said to be unitary if the matrix representing it is a unitary matrix, i.e., if it satisfies $UU^+ = U^+U = 1$. The most fundamental property of unitary operators is that they have scalar products invariant:

$$(U\phi, U\psi) = (\phi, U^+U\psi) = (\phi, \psi).$$

They are thus the generalization of rotations in ordinary vector analysis.

Characteristic values and vectors.

If we have an operator A we can ask if there are any vectors which are left parallel to themselves as a result of the transformation with A. That is, are there any vectors for which

$$A\psi = \lambda\psi$$

where λ is a complex number? If such a vector exists it is called a characteristic vector of A, and the corresponding λ is called a characteristic value of A. (We shall also use the words eigen-vector and eigen-value, as both are current in the literature). Written out in components this equation takes the form

$$\sum_j a_{ij} x_j = \lambda x_i$$

or

$$\sum_j (a_{ij} - \lambda \delta_{ij}) x_i = 0$$

This is a set of n homogeneous linear equations for the n unknowns x_i, and a solution not identically zero can only exist if the determinant of the coefficients vanishes, that is if

$$\det(a_{ij} - \delta_{ij}\lambda) = 0$$

or

$$\det(A - \lambda I) = 0$$

This is known as the secular equation, and is an equation of the n^{th} order for λ, and therefore we obtain n λ's as its roots. For each root λ_i we obtain a corresponding ψ_i which clearly can only be determined to within a multiplicative constant. The set of all possible λ_i will be called the <u>spectrum</u> of the operator A. If A is hermitian, then all the λ_i are real. This follows at once by considering

$$(\psi_i, A\psi_i) = (\psi_i, \lambda_i \psi_i) = \lambda_i (\psi_i, \psi_i)$$

Now for an hermitian operator ($\psi_i, A\psi_i$) is real, while for any vector (ψ_i, ψ_i) is real, and therefore λ_i is real. The characteristic vectors of an hermitian operator A have a remarkable property: Consider two characteristic vectors ψ_i and ψ_j having characteristic values λ_i, λ_j :

$$(\psi_i, A\psi_j) = (\psi_i, \lambda_j \psi_j) = \lambda_j (\psi_i, \psi_j)$$

$$= (A\psi_i, \psi_j) = (\lambda_i \psi_i, \psi_j) = \lambda_i (\psi_i, \psi_j)$$

Therefore $\lambda_j (\psi_i, \psi_j) = \lambda_i (\psi_i, \psi_j)$ or

$$(\lambda_j - \lambda_i)(\psi_i, \psi_j) = 0$$

Now if $\lambda_i \neq \lambda_j$ we get immediately that (ψ_i, ψ_j) = 0. Thus the characteristic vectors of distinct characteristic values are orthogonal. If all the λ_i are distinct then all the ψ_i are orthogonal. In this case we speak of the operator A as non-degenerate. If on the other hand some of the roots of the secular equation are not distinct - several of the λ_i equal - we say the operator A is <u>degenerate</u>. We say that a particular λ is d-fold degenerate if λ is a d-fold root of the secular equation. 1-fold degeneracy is the same as non-degeneracy. We can also express d-fold degeneracy by saying that and different characteristic vectors belong to the same characteristic value λ. Any linear combination of them will also be a characteristic vector. Let

$$\phi = \sum_{i=1}^{d} \mu_i \psi_i$$

$$A\phi = A\sum \mu_i \psi_i = \sum \mu_i \lambda \psi_i = \lambda \sum \mu_i \psi_i$$

$$= \lambda \phi$$

By choosing this linear combination correctly, we can always arrange that the characteristic vectors are all orthogonal.

As an example consider the case of triple degeneracy with the vectors ψ_1, ψ_2, ψ_3 belonging to the same λ. Let

$$\phi_1 = \psi_1$$

$$\phi_2 = \psi_2 - \phi_1 (\phi_1, \psi_2)/(\phi_1, \phi_1)$$

$$\phi_3 = \psi_3 - \frac{(\phi_1, \psi_3)\phi_1}{(\phi_1, \phi_1)} - \frac{(\phi_2, \psi_3)\phi_2}{(\phi_2, \phi_2)}$$

$$(\phi_1, \phi_2) = (\psi_1, \psi_2) - (\psi_1, \psi_2) = 0$$

$$(\phi_1, \phi_3) = (\psi_1, \psi_3) - (\psi_1, \psi_3) = 0$$

$(\phi_2, \phi_3) = 0$ by direct calculation.

This is not the only way of producing orthogonal characteristic vectors, but it is one technique which always works. It is known as the Gram-Schmidt orthogonalization process. We shall usually choose the arbitrary constant which multiplies a characteristic vector in such a way that its norm is unity, i.e. such that $(\psi_i, \psi_i) = 1$. Then we shall speak of the vectors as being normalized. We have for a normalized set of orthagonal characteristic vectors the relationships:

$$(\psi_i, \psi_j) = \delta_{ij}$$

Then

$$(UAU^{-1})_{ij} = \sum_{k,l} U_{ik}\, a_{kl}\, (U^{-1})_{lj} = \sum U_{ik}\, a_{kl}\, U^*_{jl}$$

$$= \sum x_k^*(i)\, a_{kl}\, x_l(j) = \sum x_k^*(i)\, \lambda_j\, x_k(j)$$

$$= \delta_{ij}\, \lambda_j$$

Just as we saw that the characteristic values of an hermitian operator are real, we can obtain a property of the characteristic values of unitary operators. Let $U\psi_i = \lambda_i \psi_i$

We have

$$(U\psi_i, U\psi_i) = (\psi_i, \psi_i)$$

$$= (\lambda_i \psi_i, \lambda_i \psi_i) = \lambda_i \lambda_i^* (\psi_i, \psi_i),$$

or

$$\lambda_i^* \lambda_i = 1$$

All the λ_i lie on the unit circle

$$\lambda_i = e^{i\phi_i}$$

where ϕ_i is real.

Functions of an operator

We shall define a function $f(A)$ of the operator A as an operator whose matrix has the form

$$f(A) = \begin{pmatrix} f(\lambda_1) & & & 0 \\ & f(\lambda_2) & & \\ & & \ddots & \\ 0 & & & f(\lambda_n) \end{pmatrix}$$

When A has the form

$$A = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Clearly from this we see that $(f(A))^+ = f^*(A^+)$

Using this, we can express a unitary operator U in terms of an hermitian operator H,

$$U = e^{iH}$$

Since

$$U^{-1} = e^{-iH} \quad \text{and}$$

$$U^+ = e^{-iH^+} = e^{-iH} = U^{-1}$$

This representation of a unitary operator will be very common when we come to quantum mechanics.

Simultaneous Diagonalization of Hermitian Operators

We now ask the question, under what circumstances can we find a basis ψ_i such that two hermitian operators A and B are both in diagonal form? We shall prove that a necessary and sufficient condition for this to be possible is that AB = BA.

Necessity:

$$A\psi_i = a_i\psi_i$$

$$B\psi_i = b_i\psi_i$$

where a_i, b_i are the characteristic values of A and B. Multiply the first equation by B, the second by A and subtract giving

$$(BA - AB)\psi_i = (a_i B\psi_i - b_i A\psi_i) = 0 .$$

But since the ψ_i form a basis, we have that (BA − AB) ψ = 0 for arbitrary ψ and therefore AB = BA.

Sufficiency: If AB = BA. Let us first choose ψ_i such that B is diagonal. Thus B ψ_i = $b_i\psi_i$, and AB ψ_i = ~~AB~~ $A b_i\psi_i = b_i(A\psi_i) = B(A\psi_i)$

Two cases arise:

(a) B is non-degenerate, i.e. all b_i are distinct. If this is so then to each b_i there is only one ψ_i. However we have shown that if ψ_i is a characteristic vector belonging to b_i, so is $A\psi_i$. Thus $A\psi_i$ must be proportional to ψ_i, or $A\psi_i = a_i \psi_i$, and A is automatically diagonal.

(b) B is degenerate, say there are d_i characteristic vectors belonging to b_i. Call these vectors $\psi_{i,\alpha}\ (\alpha = 1, \cdots, d_i)$ Then since $A\psi_{i,\alpha}$ also belongs to b_i we must have

$$A\psi_{i,\alpha} = \sum_{\beta=1}^{d_i} \psi_{i,\beta} A^{(i)}_{\beta\alpha}$$

where the $A^{(i)}_{\beta\alpha}$ are just numbers. Now all we have to do is choose instead of the $\psi_{i,\alpha}$ those combinations $\phi_{i,\gamma} = \sum_\alpha C^{(i)}_{\gamma\alpha}\psi_{i,\alpha}$ which make A diagonal, i.e.

$$A\phi_{i,\gamma} = a_{i,\gamma}\phi_{i,\gamma}$$

This change of basis leaves B diagonal - since we have just taken linear combinations of characteristic vectors of B belonging to the same b_i - and reduces A to diagonal form.

In general if we have a sequence of operators A,B,C, etc. which commute they can be made simultaneously diagonal. If the set A,B,C, etc. has the property that the basis necessary to diagonalize them simultaneously is non-degenerate, then we say that it is a <u>complete set of commuting variables</u>.

<u>Invariance of characteristic values under a Canonical transformation</u>:

If A has the characteristic values a_i, then so does $A^1 = SAS^{-1}$. We have

$A\psi_i = a_i\psi_i$, but $A = S^{-1}A^1S$, and therefore

$(S^{-1}A'S)\psi_i = a_i\psi_i$, [redacted] $A'(S\psi_i) = a_i(S\psi_i)$ or

$A'\phi_i = a_i\phi_i$ where $\phi_i = S\psi_i$, Since ψ_i is a basis, and S is unitary, ϕ_i is also a basis

$$((\phi_i, \phi_j) = (S\psi_i, S\psi_j) = (\psi_i, \psi_j) = \delta_{ij}),$$

and therefore in this basis A^1 is diagonal and has exactly the same characteristic values as A. There is one combination of the matrix elements of an operator which is invariant under a canonical transformation, and which is of particular importance in quantum mechanics. We define the <u>trace</u> of the matrix of A as the sum of the diagonal elements of A.

$$tr(A) = \sum_{i=1}^{n} A_{ii}$$

From this definition we see at once that

$$tr(AB) = \sum_i (AB)_{ii} = \sum_{i,k} A_{ik} B_{ki} = \sum B_{ki} A_{ik} = \sum_k (BA)_{kk}$$
$$= tr(BA).$$

Thus $tr(A^1) = tr(SAS^{-1}) = tr(S^{-1}SA) = tr(A)$.

The trace of a matrix is invariant under a canonical transformation. If in particular we choose the representation in which A is diagonal, then we see that $tr(A) = \sum_i a_i$, which, since the characteristic values are invariant, exhibits the invariance of the trace explicitely.

Dirac Notation

It will sometimes be very convenient to introduce a notation due to Dirac instead of that which we have used up to the present. If a set of basis vectors are characteristic vectors of the operator A, then $\Psi(a')$ shall denote the vector corresponding to the characteristic value a' of A. If A is degenerate then we shall use as our basis the characteristic vectors of a complete set of commuting operators A_1, A_2, A_3 . . ., and write for the basis vectors $\Psi(a_1', a_2', a_3', \dots)$. Let us group all these indices together with a label α', and write simply $\Psi(\alpha')$. Let the component of an arbitrary vector Ψ be denoted by $(\alpha|)$, that is

$$\Psi = \sum_{\alpha'} \Psi(\alpha')(\alpha'|)$$

$$(\alpha'|) = (\Psi(\alpha'), \Psi)$$

We shall use the empty space in $(\alpha'|)$ to place any label which may be placed on the vector Ψ. For example, consider Ψ to be the characteristic vector of some other complete set of commuting operators, the characteristic values of which we label with the numbers β', *i.e.*

$$\Psi = \Psi(\beta')$$

Then we write

$$\Psi(\beta') = \sum_{\alpha'} \Psi(\alpha')(\alpha'|\beta')$$

$$(\alpha'|\beta') = (\Psi(\alpha'), \Psi(\beta')).$$

We can just as well expand the $\Psi(\alpha')$ in the $\Psi(\beta')$

$$\Psi(\alpha') = \sum_{\beta'} \Psi(\beta')(\beta'|\alpha')$$

$$(\beta'|\alpha') = (\Psi(\beta'), \Psi(\alpha')) = (\Psi(\alpha'), \Psi(\beta'))^*$$

$$= (\alpha'|\beta')^*.$$

$$\Psi(\beta') = \sum_{\alpha'} \Psi(\beta'')(\beta''|\alpha')(\alpha'|\beta'), \text{ and}$$

therefore $\sum_{\alpha'} (\beta''|\alpha')(\alpha'|\beta') = \delta_{\beta'\beta''}$. Similarly

$$\sum_{\beta'} (\alpha''|\beta')(\beta'|\alpha') = \delta_{\alpha'\alpha''}.$$

These relationships just express the unitary character of the transformation of bases. For the matrix elements of an operator we shall use the notation

$$(\beta'|A|\beta'') \equiv (\Psi(\beta'), A\,\Psi(\beta'')),$$

which is equivalent to

$$A.\Psi(\beta') = \sum_{\beta''} \Psi(\beta'')(\beta''|A|\beta').$$

For the product of two operators A, C we have

$$\beta'|AC|\beta'') = \sum_{\beta'''} (\beta'|A|\beta'')(\beta''|C|\beta''') .$$

The transformation formula for matrix elements on changing the basis from $\Psi(\beta')$ to $\Psi(\alpha')$ is easily found

$$(\alpha'|C|\alpha'') = (\Psi(\alpha'), C\,\Psi(\alpha''))$$

$$= \sum_{\beta',\beta''} \big((\beta'|\alpha')\Psi(\beta'), (\beta''|\alpha'')C\,\Psi(\beta'')\big)$$

$$= \sum_{\beta',\beta''} (\alpha'|\beta')(\Psi(\beta'), C\,\Psi(\beta''))(\beta''|\alpha'')$$

$$= \sum_{\beta',\beta''} (\alpha'|\beta')(\beta'|C|\beta'')(\beta''|\alpha'')$$

which is a very symmetric formula.

Generalization to vector space with infinitely many dimensions (Hilbert Space).

Two types of generalization are possible. We may assume simply that $n \rightarrow \infty$ in all the work we have done up tell now. Then, if the various sums involved converge, we have a theory exactly equivalent to our old one. This is the theory of a vector space with denumerably many dimensions. On the other hand, it is possible to imagine a space in which we have a non-denumerable infinity of independent vectors or, said another way, a space where we have to label the basic vectors with a continuous index instead of a descrete one. We shall see, that, with suitable definitions, all the machinery of our finite vector space can be carried over to this case. Clearly when we have a continuous index the natural thing to do is to replace sums over the index by integrals. Call the continuous variable x. Therefore any vector Ψ we shall write

$$\Psi = \int e(x)\,dx\,(x|)$$

where e(x) is the basic vector belonging to the value of the variable x, and the range of integration goes over the range of ~~continuous~~ variation of x. Let us consider now how we must define the unit operator I for continuous vector spaces. I is defined by

$$\Psi = I\,\Psi$$

for any vector Ψ. Writing out the components we have

$$(x'|) = \int (x'|I|x)\,dx\,(x|)$$

Now for an arbitrary vector $(x|)$ is an arbitrary <u>function</u>

of x, say $f(x)$. We must have

$$f(x') = \int (x'|I|x)\, f(x)\, dx$$

for any f(x). This is only possible if the quantity $(x'|I|x) = \delta(x-x')$, where $\delta(x-x')$ is the so-called Dirac delta function. It has the property that $\delta(x-x') = 0$ if $x \neq x'$

$$\int \delta(x-x')\,dx' = 1$$

for any interval of x' which includes the point x. This is clearly the generalization to continuous variables of Kronecher δ symbol. We also see at once from

$$(x|) = (e(x), \Psi) = \int (e(x), e(x'))\,dx'(x'|)$$

that $(e(x), e(x')) = \delta(x-x')$

The characteristic value problem takes the following form:

$$A\Psi = \lambda\Psi$$

$$\int e(x'')\,dx'(x''|A|x)\,dx\,(x|) = \lambda \int e(x'')\,dx''(x''|)$$

Taking the scalar product with e(x') we get

$$\int (x'|A|x)\,dx\,(x|) = \lambda(x'|)$$

The quantity $(x|)$ is now an unknown function of x, to be determined by the solution of this equation. One very often writes $(x|) = \psi(x)$ and calls ψ the wave function, for reasons that will become clear when we discuss the application of these ideas to quantum mechanics. Thus we have

$$\int (x'|A|x)\,\psi(x)\,dx = \lambda\,\psi(x').$$

As before we can take our basis vectors to be the eigen-vectors of a complete set of commuting variables. Now, however we have the possibility of the characteristic values being continuously distributed instead of discretely as in our old work. Consider two sets of complete, commuting operators $A_1 \; . \; . \; . \; . \; A_n$ and $B_1 \; . \; . \; . \; . \; B_m$, both of which have continuous spectra, labeled by the variables α' and β' respectively. Then we must have

$$\Psi(\beta') = \int \Psi(\alpha')\,d\alpha'\,(\alpha'|\beta')$$

$$\Psi(\alpha') = \int \Psi(\beta')\,d\beta'\,(\beta'|\alpha')$$

$$\Psi(\beta') = \int \Psi(\beta'')\,d\beta''\,(\beta''|\alpha')\,d\alpha'\,(\alpha'|\beta')$$

or

$$\int (\beta''|\alpha')\,d\alpha'\,(\alpha'|\beta') = \delta(\beta''-\beta').$$

Similarly

$$\int (\alpha''|\beta')\,d\beta'\,(\beta'|\alpha') = \delta(\alpha''-\alpha')$$

, which express the unitary character of the transformation of bases. It is straightforward to generalize these relationships for the transformation from a continuous to a descrete spectrum, or vis versa. The transformation rule for matrix elements is easily shown to be

$$(\alpha'|C|\alpha'') = \int (\alpha'|\beta')\,d\beta'\,(\beta'|C|\beta'')\,d\beta''\,(\beta''|\alpha'')$$

, while for

the matrix element of the product of two operators

$$(\alpha'|BC|\alpha'') = \int (\alpha'|B|\alpha''')d\alpha'''(\alpha'''|C|\alpha'').$$

The rule for the transformation of components of a vector is also simple to obtain:

$$\Psi = \int \Psi(\alpha')d\alpha'(\alpha'|) = \int \Psi(\beta')d\beta'(\beta'|)$$

Taking the scalar product of this equation with $\Psi(\beta')$ we get

$$(\beta'|) = \int (\beta'|\alpha')d\alpha'(\alpha'|)$$

If one writes

$$(\alpha'|) = \psi(\alpha')$$
$$(\beta'|) = \phi(\beta')$$

we obtain

$$\phi(\beta') = \int (\beta'|\alpha')d\alpha' \, \psi(\alpha'),$$

Example: Fourier transform as an example of a change of basis in vector space. Fourier's theorem tells us that if we have any function $\psi(x)$, then the quantity

$$\phi(k) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx}\psi(x)\,dx \quad \text{satisfies}$$

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikx}\phi(k)\,dk.$$

Let us define $(k|x) = \frac{1}{\sqrt{2\pi}} e^{ikx}$ and $(x|k) = (k|x)^*$

$$(x|k) = \frac{1}{\sqrt{2\pi}} e^{-ikx}$$

Then we may write these equations as

$$\phi(k) = \int (k/x)\,\psi(x)\,dx$$

$$\psi(x) = \int (x/k)\,\phi(k)\,dk$$

which are just of the form of a transformation of basis in vector space. We still must prove that

$$\int (k/x)\,dx\,(x/k') = \delta(k-k')$$ and

$$\int (x/k)\,dk\,(k/x') = \delta(x-x')$$ in order to

establish the unitary nature of the transformation. Consider the former:

$$\int (k/x)\,dx\,(x/k') = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i(k-k')x}\,dx$$

Strictly speaking this integral is not convergent, but if we put in any "convergence factor" we obtain the same result.

For example, consider the function ($\varkappa \equiv k - k'$)

$$\delta_\alpha(\varkappa) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\alpha|x|}\,e^{i\varkappa x}\,dx \qquad \alpha > 0,$$

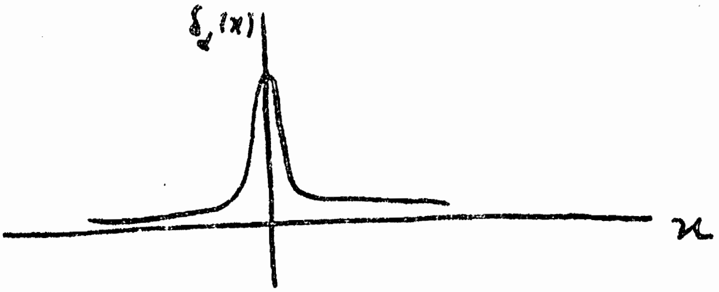
$$= \frac{1}{2\pi}(2)\int_0^{\infty} e^{-\alpha x}\cos\varkappa x\,dx$$

$$= \frac{1}{\pi}\,\frac{\alpha}{\alpha^2+\varkappa^2}.$$

Clearly

$$\int_{-\infty}^{\infty}\delta_\alpha(\varkappa)\,d\varkappa = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\alpha\,d\varkappa}{\alpha^2+\varkappa^2} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{dt}{1+t^2} = 1$$

which is one requirement for the δ-function. If we plot $\delta_\alpha(x)$ against x we get:



As $\alpha \rightarrow 0$, $\delta_\alpha \rightarrow 0$ except at $x = 0$, where it approaches infinity as $1/\pi\alpha$. Thus as $\alpha \rightarrow 0$ $\delta_\alpha(x) \rightarrow \delta(x)$ which is the result we want.

As an example of the transformation of an operator, consider the simple operator $\mathbf{X}$ which is diagonal with characteristic values x in the original representation. That is,

$$(x'|\mathbf{X}|x'') = x'\,\delta(x'-x'')$$

What are the matrix elements of this operator in the k basis?

$$(k'|\mathbf{X}|k'') = \int (k'|x')dx'(x'|\mathbf{X}|x'')dx''(x''|k'')$$

$$= \frac{1}{2\pi}\iint_{-\infty}^{+\infty} e^{i(k'x'-k''x'')}\,x'\,\delta(x'-x'')\,dx'dx''$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} x' e^{i(k'-k'')x'}\,dx'$$

$$= \frac{1}{i}\frac{\partial}{\partial k'}\,\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i(k'-k'')x'}\,dx'$$

$$= \frac{1}{i}\frac{\partial}{\partial k'}\,\delta(k'-k'') = -\frac{1}{i}\frac{\partial}{\partial k''}\,\delta(k'-k'').$$

In terms of the effect on the coordinates of a general vector these relationships become a little clearer.

Thus the operation X on simply changes the coordinate from $\psi(x)$ to $x\,\psi(x)$, it is equivalent to simple multiplication with X in the basis where $\mathbf{X}$ is diagonal. In the other basis on the other hand

$$\begin{aligned}
X\,\Psi &= \int \Psi(k')\,dk'\,(k'|X|k'')\,dk''\,\phi(k'') \\
&= \int \Psi(k')\,dk'\left(\frac{-1}{i}\right)\frac{\partial}{\partial k''}\delta(k'-k'')\,dk''\,\phi(k'') \\
&= \int \Psi(k')\,dk'\,\delta(k'-k'')\,dk''\,\frac{1}{i}\,\frac{\partial\phi(k'')}{\partial k''} \\
&= \int \Psi(k')\,dk'\left(\frac{1}{i}\,\frac{\partial\phi(k')}{\partial k'}\right), \quad \text{on}
\end{aligned}$$

integrating by parts. Thus the operator $\mathbf{X}$ on Ψ sends $\phi(k) \rightarrow \frac{1}{i}\frac{\partial\phi(k)}{\partial k}$. Multiplication by the variable in one bases is the same as performing the operation $\frac{1}{i}\frac{\partial}{\partial k}$ in the other basis.

II. Fundamental Assumptions of Quantum Mechanics

The assumptions of quantum mechanics fall into three classes, which we may call geometrical, kinimatical and dynamical. The first consist of assumptions about how one describes a "state" of a system in quantum mechanics, the second gives us relationships between quantities which represent a position and those which represent motion, and the last tells us how one calculates for a definite system its behavior in time.

A. Geometrical Assumptions

(1) Each possible physical state of a system is represented – in some way – by a vector Ψ in a Hilbert space.

(2) Each observable physical quantity (position, momentum, angular momentum, energy, etc.) is represented by an hermitian operatoryoperating in this space.

(3) the only numerical values which an observable can attain are the characteristic values of the operator associated with that observable. Because an observable is represented by an hermitian operator, these numbers are of course real.

(4) If a system is in a state $\Psi = \sum_a \Psi(a)(a|)$ where $\Psi(a)$ are the normalized characteristic vectors of an observable A, then the quantity $|(a|)|^2$ gives the probability of finding the value a for the quantity A. (Of course, if a is degenerate we must sum over all the $|(a|)|^2$ belonging to the same a).

Assumption (4) has an immediate corollary: Consider any observable A and state Ψ, then the scalar product

$$(\Psi, A\Psi) = \sum_a a|(a|)|^2$$

$$= \sum_a a P_a$$

tate which specifies definite values of A_1, A_2 . . . . A_n. The measurement of A_1 and A_2 is compatible (that is, a knowledge of A_1 does not destroy a knowledge of A_2) if the corresponding operators commute. Conversely, since two operators cannot be simultaneously diagonalized if they do not commute, we cannot assign definite values simultaneously to two observables whose corresponding operators do not commute. In other words, information about the state of a system with respect to one observable may destroy information with respect to another. Later on we shall investigate the extent of this interference of two measurements.

So far we have spoken only of discrete characteristic values. If the characteristic values are continuous, all we have said remains unchanged, except that now one must interpret $|(a|)|^2$ as a "probability density," that is, the probability of A being between a and a + da is $|(a|)|^2 da$. With this interpretation we clearly retain

$$\bar{A} = (\Psi, A\Psi) \qquad \text{since}$$

$$(\Psi, A\Psi) = \int (\Psi(a')\,da' (a'|),\; a\,\Psi(a)(a|))$$

$$= \int \delta(a-a')(a'|)^*(a|)\,a\,da'\,da = \int a|(a|)|^2 da$$

$$= \bar{A}$$

B. Kinematical Assumptions.

One can say that the essence of the new assumption in quantum mechanics is that one cannot simultaneously measure the position and the momentum of a particle. In terms of our above assumptions this means that the operators representing the co-ordinate and the momentum cannot commute. Denote these operators by q and p, respectively: in general we shall use the same symbol for the operator as we do for the observable itself. This means that $pq \neq qp$ or

$$qp - pq = iC$$

where $C \neq 0$. C is clearly hermetian since

$$-iC^{+} = (qp)^{+} - (pq)^{+} = p^{+}q^{+} - q^{+}p^{+} = pq - qp = -iC$$

$$C^{+} = C$$

The quantity qp - pq is called the <u>commutator</u> of q and p, and is often written (q, p). Exactly how to choose the operator C is something that can only be given by experiment--different choices of C would give us different possible quantum mechanics. <u>The choice of C which gives us quantum mechanics as we know it is that C is simply a constant multiple of the unit operator.</u> The value of the constant depends on experiment and on the set of units used; we denote it by $\hbar$. Thus $C = \hbar I$ where $\hbar$ = 1.05435 x 10^{-27} erg-sec in the c.g.s. system of units. The quantity $2\pi\hbar$ is known as Planck's constant. Thus the basic kinematical assumption of quantum mechanics is

$$qp - pq = i\hbar I$$

We shall very often write this simply as $qp - pq = i\hbar$ dropping the explicit writing of the unit operator. Consider the case where the co-ordinate is say a cartesian co-ordinate of a particle. In this case, to the best of our knowledge the co-ordinate can have any value from - ∞ to + ∞. Thus in the representation with q diagonal we must have for the matrix elements of q

$$(q'|q|q'') = q'\delta(q'-q'')$$

where q', q" are any numbers from + ∞ to - ∞. From this we can see at once that p also has any value from - ∞ to + ∞. As a preliminary, let us prove the lemma

$$f(q)p - pf(q) = i\hbar\frac{\partial f}{\partial q}$$

where f (q) is any function of q. Consider first f = q^2. Then

$$(q^2p - pq^2) = q(qp - pq) + (qp - pq)q = q(i\hbar) + i\hbar(q) = 2i\hbar q$$

$$(q^3p - pq^3) = q(q^2p - pq^2) + (qp - pq)q^2 = 3i\hbar q^2$$

$$(q^np - pq^n) = n i\hbar q^{n-1} = i\hbar \frac{\partial q^n}{\partial q} .$$

Thus for any function that can be represented as a power series, this relationship is true for every term and therefore

$$f(q)\,p - p\,f(q) = i\hbar \frac{\partial f(q)}{\partial q}$$

Similarly

$$q\,g(p) - g(p)\,q = i\hbar \frac{\partial g(p)}{\partial p} .$$

Now let $\Psi(p')$ be a characteristic function of p belonging to the characteristic value p', i.e.

$$p\,\Psi(p') = p'\,\Psi(p') .$$

Consider now the vector $\Phi = e^{ikq}\Psi(p')$, where k is any real number. Then

$$p\,\Phi = p\,e^{ikq}\Psi(p'),$$

but

$$p\,e^{ikq} - e^{ikq}p = -i\hbar \frac{\partial(e^{ikq})}{\partial q} = \hbar k\, e^{ikq}$$

Thus

$$p\,\Phi = (e^{ikq}p + \hbar k\, e^{ikq})\Psi(p')$$

$$= (p' + \hbar k)\, e^{ikq}\,\Psi(p')$$

$$= (p' + \hbar k)\,\Phi .$$

Therefore if p' is a possible value of p, so is p' + k, where k is any real number, so that the possible values of p are all the real numbers from - ∞ to + ∞.

In the representation with q diagonal we can easily calculate the matrix elements of p. Consider

$$(q'|qp-pq|q'') = i\hbar(q'|I|q'') = i\hbar\,\delta(q'-q'')$$

$$= \int\{(q'|q|q''')dq'''(q'''|p|q'')$$

$$- (q'|p|q''')dq'''(q'''|q|q'')\}$$

$$= \int\{q'\delta(q'-q''')(q'''|p|q'')$$

$$- q''(q'|p|q''')\delta(q'''-q'')\}dq'''$$

$$= (q'-q'')(q'|p|q'').$$

Or

$$(q'|p|q'') = i\hbar\,\frac{\delta(q'-q'')}{(q'-q'')}$$

The function on the right hand side is actually the negative of the derivative of the delta function, i.e.

$$\frac{\delta(q'-q'')}{(q'-q'')} = -\frac{\partial}{\partial q'}\,\delta(q'-q'') = -\delta'(q'-q'').$$

To see this, let us consider $x\,\delta(x)$ and its derivative. Since $\delta(x) \neq 0$ only for $x = 0$, the quantity $x\,\delta(x)$ gives zero when placed in any integral with any regular function. Therefore $x\,\delta(x) = 0$. Differentiating we obtain

$$x\,\delta'(x) + \delta(x) = 0, \quad \text{or}$$

$$\delta'(x) = -\frac{\delta(x)}{x}.$$

Finally

$$(q'|p|q'') = \frac{\hbar}{i}\,\delta'(q'-q'')$$

From these matrix elements we can calculate at once the effect of p on any state

$$\Psi = \int \Psi(q')\,dq'\,(q'|)$$

$$p\Psi = \int \Psi(q')\,dq'\,(q'|p|q'')\,dq''\,(q''|)$$

$$= \int \Psi(q')\,dq'\,\frac{\hbar}{i}\,\delta'(q'-q'')\,dq''\,(q''|)$$

$$= -\int \Psi(q')\,dq'\,\frac{\hbar}{i}\,\frac{\partial}{\partial q''}\,\delta(q'-q'')\,dq''\,(q''|)$$

$$= \int \Psi(q')\,dq'\,\delta(q'-q'')\,\frac{\hbar}{i}\,\frac{\partial(q''|)}{\partial q''}\,dq''$$

$$= \int \Psi(q')\,dq'\,\frac{\hbar}{i}\,\frac{\partial(q'|)}{\partial q'}$$

We can formulate this result very simply by saying the effect of p on the representative $(q'|)$ of a state is found by simply carrying out the differentiation operation $\frac{\hbar}{i}\frac{\partial}{\partial q'}$ on it. For brevity one usually says that p is <u>equivalent</u> to the differential operator $\frac{\hbar}{i}\frac{\partial}{\partial q}$.

In terms of these results it becomes very straightforward to find the transformation from the basis with q diagonal to that with p diagonal.

$$\Psi(p') = \int \Psi(q')\,dq'\,(q'|p')$$

$$p\,\Psi(p') = p'\,\Psi(p') = \int \Psi(q')\,dq'\,\frac{\hbar}{i}\,\frac{\partial(q'|p')}{\partial q'}$$

$$= \int \Psi(q')\,dq'\,p'\,(q'|p').$$

Or,

$$\frac{\hbar}{i}\frac{\partial(q'|p')}{\partial q'} = p'(q'|p')$$

$$(q'|p') = c\, e^{iq'p'/\hbar}, \quad c \text{ a constant.}$$

In order to find the constant, we must make use of the normalization

$$\int (q'|p')dp'(p'|q') = \delta(q'-q'')$$

$$c^2 \int e^{i(q'-q'')p'/\hbar} dp' = c^2\hbar \int e^{i(q-q')k} dk, \quad k = \frac{p'}{\hbar}.$$

But from our old work on δ-functions we know that

$$\int e^{i(q'-q'')k}\, dk = 2\pi\, \delta(q'-q''),$$

and therefore we have

$$c^2(2\pi\hbar) = 1.$$

$$c = \frac{1}{\sqrt{2\pi\hbar}}.$$

Therefore

$$(q'|p') = \frac{1}{\sqrt{2\pi\hbar}}\, e^{iq'p'/\hbar}$$

gives the final formula for the transformation coefficients from the "q" representation to the "p" representation. The former is usually called the <u>configuration space</u> representation and the latter the <u>momentum space</u> representation. Since for any state we have from our general formulae

$$(p') = \int (p'|q')dq'(q'|)$$

$$= \frac{1}{\sqrt{2\pi\hbar}} \int e^{-ip'q'/\hbar}(q'|)\,dq'$$

we see that--apart from the quantity $\hbar$ ~~to~~ which is necessary to make the exponent dimensionless--the transformation from co-ordinate to momentum space is simply given by taking the fourier transform.

In the case of several degrees of freedom the <u>commutation rules</u> (as the rules for calculating the commutator of any two quantities are called) are easy to find. Say we have n <u>independent</u> co-ordinates $q_1, q_2, \ldots q_n$, with corresponding momenta $p_1, p_2, \ldots p_n$. Now independence just means that we can measure one without affecting the other in any way. Thus we take

$$(q_i, q_j) = 0$$

$$(q_i, p_j) = 0 \qquad i \neq j$$

$$(p_i, p_j) = 0$$

As before, we have of course $(q_i, p_i) = i\hbar$, or we may write in general

$$(q_i, p_j) = i\hbar\,\delta_{ij}\,.$$

Heisenberg's Uncertainty Relationships.

Consider our expression for $(q'|p')$. We know that $|(q'|p')|^2 dp'$ gives the probability of p having a value between p' and p' + dp' if the particle is definitely at the point q'. However

$$|(q'|p')|^2 = \left|\frac{1}{\sqrt{2\pi\hbar}}\,e^{ip'q'/\hbar}\right|^2 = \frac{1}{2\pi\hbar}\,, \qquad \text{a constant.}$$

Thus if we know q exactly, all values of p are equally probable. In other words an exact knowledge of q means no information whatever on p. This is an extreme form of Heisenberg's uncertainty principle. Suppose now we have a state in which q is not specified, but in which we have some knowledge of the position of the particle. How much can we know about its momentum? Call the state in question Ψ. Let

$$\bar{q} = (\Psi, q\Psi)$$

$$\bar{p} = (\Psi, p\Psi)$$

e the expectation values of q and p respectively, for this state.

efine $A = q - \bar{q}$, $B = p - \bar{p}$. Clearly the quantity $\overline{A^2}$ ives a measure of the deviation of q from its expectation value, and $\overline{B^2}$ ives a measure of the deviation of p from its expectation value.

f $\overline{A^2} = 0$ then Ψ is a characteristic vector of q belonging o the value $\bar{q}$. That is

$$\Psi, A^2\Psi) = (A\Psi, A\Psi) \geq$$

he equality only is possible if $A\Psi \equiv 0$ or $q\Psi = \bar{q}\Psi$

In general

$$\overline{A^2} = (\Psi, A^2\Psi) = (\Psi, (q^2 - 2q\bar{q} + \bar{q}^2)\Psi)$$

$$= \overline{q^2} - \bar{q}^2 \equiv (\Delta q)^2.$$

imilarly

$$\overline{B^2} = \overline{p^2} - \bar{p}^2 \equiv (\Delta p)^2.$$

hen we have

$$(\Delta q)^2(\Delta p)^2 = (\Psi, A^2\Psi)(\Psi, B^2\Psi)$$

$$= (A\Psi, A\Psi)(B\Psi, B\Psi)$$

y the Schwartz inequality though, for any two vectors $A\Psi$ and $B\Psi$ e have

$$(A\Psi, A\Psi)(B\Psi, B\Psi) \geq |(A\Psi, B\Psi)|^2,$$

he equality only holding when $A\Psi$ is parallel to $B\Psi$. Thus

$$(\Delta q)^2(\Delta p)^2 \geq |(A\Psi, B\Psi)|^2 = |(\Psi, AB\Psi)|^2$$

$$= \left|\left(\Psi, \left\{\frac{AB+BA}{2} + \frac{i(AB-BA)}{2}\right\}\Psi\right)\right|^2.$$

Now
$$\frac{AB-BA}{i} = \frac{qp-pq}{i} = C$$

($C = \hbar I$, of course, but temporarily we will keep it more general in order to obtain an uncertainty relationship which is true for any two quantities whose commutator is given by $i\,C$).

$$(\Delta q)^2(\Delta p)^2 \geq \left| \left(\Psi, \frac{AB+BA}{2}\Psi\right) + \frac{i}{2}(\Psi, C\Psi)\right|^2$$

Since $AB + BA$ and C are both hermitian the scalar products are both real, and we have

$$(\Delta q)^2(\Delta p)^2 \geq \frac{1}{4}\left\{(\Psi,(AB+BA)\Psi) + \overline{C}^2\right\}$$

$$\geq \frac{1}{4}\overline{C}^2$$

the first term being clearly positive.

Thus
$$|\Delta q|\cdot|\Delta p| \geq \frac{1}{2}\overline{C}$$

is the most general form of the uncertainty principle. The condition for equality is that

$$(\Psi,(AB+BA)\Psi) = 0$$ and

$$A\Psi = \lambda B\Psi .$$

From the first of these we have

$$(A\Psi, B\Psi) + (B\Psi, A\Psi) = 0$$

$$\lambda^*(B\Psi, B\Psi) + \lambda(B\Psi, B\Psi) = 0 \quad , \text{ or}$$

$$\lambda^* = -\lambda \qquad \lambda = \gamma/i$$

where γ is real. The condition for equality is thus

$$iA\Psi = \gamma B\Psi \quad , \gamma \text{ real.}$$

$$F\Psi = F'\Psi$$

In the coordinate representation this gives simply

$$F\Psi = F\int \Psi(q)\,dq\,(q|)$$

$$= \int \Psi(q)\,dq\,F\left(q, \frac{\hbar}{i}\frac{\partial}{\partial q}\right)(q|)$$

since we showed that the effect of p on a state function is the same as $\frac{\hbar}{i}\frac{\partial}{\partial q}$ on a representative. Thus

$$F\left(q, \frac{\hbar}{i}\frac{\partial}{\partial q}\right)(q|) = F'(q|)$$

is a differential equation which, when solved with the condition that $(q|)$ be well behaved (for example, $(q|)$ must be such that Ψ is normalizable) gives the characteristic values F'. Writing, as is usual,

$(q|) = \psi(q)$ we obtain

$$F\left(q, \frac{\hbar}{i}\frac{\partial}{\partial q}\right)\psi(q) = F'\psi(q)$$

In particular, if F is say the Hamiltonian of the particle in a potential V (q) we have

$$F = H = \frac{p^2}{2m} + V(q)$$

The possible values of the hamiltonian are just the possible energies of the system, say E. Our equation becomes

$$\left[\frac{1}{2m}\left(\frac{\hbar}{i}\frac{\partial}{\partial q}\right)^2 + V(q)\right]\psi(q) = E\,\psi(q)$$

or

$$\frac{\partial^2\psi}{\partial q^2} + \frac{2m}{\hbar^2}(E-V)\,\psi(q) = 0,$$

which is the well known Schroedinger equation for calculating the energy levels of a system. The normalization requirement on Ψ reduces, incidentally, to

$$\delta_{\varepsilon',\varepsilon''} = \int \psi^*_{\varepsilon'}(q)\,\psi_{\varepsilon''}(q)\,dq$$

$$\delta(\varepsilon'-\varepsilon'') = \int \psi^*_{\varepsilon'}(q)\,\psi_{\varepsilon''}(q)\,dq$$

or the discrete and continuous spectrum respectively.

C. Dynamical Assumptions

So far we have only considered a system at a given instant of ime. The question which dynamics answers is how this system changes ith time. There are two points of view possible:

(1) We can assume that a system is described by a fixed tate vector, but that as time goes on the observables are represented y operators which vary with time, the time variation being given by the lassical equations of motion of the observables. That is, since the bserved values of any physical quantity will vary with time, we just hange the operator so that in the state of the system it has this new alue. This point of view is known as the Heisenberg representation.

(2) We can imagine that the relationship between observables nd operators is fixed, but as time goes on the state vector changes, hus predicting different probabilities for the values of the observable. his latter point of view is known as the Schrödinger representation. e shall see immediately that they are related to each other very imply by means of a unitary transformation.

Let us first take the point fo view (1) and from it investigate he point of view (2). For simplicity we will write everything for the ase of one degree of freedom, though everything goes through for the eneral case. For a system with the hamiltonian $H(q,p)$ the quations of motion for q and p are Hamilton's:

$$\dot q = +\frac{\partial H}{\partial p}$$

$$\dot p = -\frac{\partial H}{\partial q}$$

which are now taken to be the equations of motion of the operators as well. Since by our old work we had

$$(q,H) = (qH - Hq) = i\hbar \frac{\partial H}{\partial p}$$

$$(H,p) = (Hp - pH) = i\hbar \frac{\partial H}{\partial q},$$

we may rewrite these equations as:

$$i\hbar \frac{dq}{dt} = (q,H)$$

$$i\hbar \frac{dp}{dt} = (p,H)$$

These equations must be integrated bearing in mind the relationship $qp - pq = i\hbar I$. When this is done we have a complete solution of the dynamics problem in the Heisenberg representation.

Now a formal solution of these equations can be given in general. Consider

$$q = e^{iHt/\hbar} q_0 e^{-iHt/\hbar}$$

$$p = e^{iHt/\hbar} p_0 e^{-iHt/\hbar}$$

where q_0, p_0 are independent of time.

$$qp - pq = e^{iHt/\hbar} q_0 e^{-iHt/\hbar} e^{iHt/\hbar} p_0 e^{-iHt/\hbar}$$

$$- e^{iHt/\hbar} p_0 e^{-iHt/\hbar} e^{iHt/\hbar} q_0 e^{-iHt/\hbar}$$

$$= e^{i/\hbar\, Ht} (q_0 p_0 - p_0 q_0) e^{-i/\hbar\, Ht} = e^{i/\hbar\, Ht} \hbar i I e^{-\frac{i}{\hbar} Ht}$$

$$= i\hbar$$

so that if q_0, p_0 satisfy the commutation rules, so will q, p. All that is left to do is to show that q and p are solutions of the equations of motion. Clearly

$$\frac{dq}{dt} = \frac{d(e^{iHt/\hbar})}{dt} q_0 e^{-iHt/\hbar} + e^{iHt/\hbar} q_0 \frac{d(e^{-iHt/\hbar})}{dt}$$

$$= \frac{i}{\hbar} H e^{iHt/\hbar} q_0 e^{-iHt/\hbar} - e^{iHt/\hbar} q_0 e^{-iHt/\hbar} \left(-\frac{i}{\hbar} H\right)$$

$$= \frac{i}{\hbar} (Hq - qH) = \frac{1}{i\hbar} (qH - Hq).$$

This is just the equation of motion for q. Similarly for p. Since $q(t=0) = q_0$, we have satisfied the differential equations, the commutation relations and the initial conditions, and have therefore obtained a complete solution of the problem in the Heisenberg representation. The time dependence of any operator will then clearly be given by

$$F(q,p) = e^{iHt/\hbar} F(q_0, p_0) e^{-\frac{iHt}{\hbar}}, \text{ or}$$

$$i\hbar \frac{dF}{dt} = (F, H).$$

From this we can at once obtain the Schrödinger representation. Consider the expectation value of any observable F--if we know how to calculate this for an arbitrary F we know everything about the system. In general we will have, in the Heisenberg representation

$$\bar{F} = (\Psi_0, F(q,p) \Psi_0)$$

where Ψ_0 is the constant state vector which represents the state of the system at time $t = 0$, and therefore for all time in the Heisenberg representation. Using our expression for F, we get

$$\bar{F} = (\Psi_0,\, e^{iHt/\hbar} F(q_0, p_0)\, e^{-iHt/\hbar}\, \Psi_0)$$

$$= (\Psi,\, F(q_0, p_0)\, \Psi),$$

where

$$\Psi \equiv e^{-iHt/\hbar}\, \Psi_0$$

This is, however, just what we would obtain if we treated the operators as constant but let the state vectors depend on time according to

$$\Psi(t) = e^{-iHt/\hbar}\, \Psi_0 \; .$$

Differentiating this with respect to time we obtain

$$\frac{\partial \Psi}{\partial t} = -\frac{iH}{\hbar}\, e^{-iHt/\hbar}\, \Psi_0 = \frac{1}{i\hbar} H \Psi$$

or,

$$i\hbar \frac{\partial \Psi}{\partial t} = H \Psi$$

This is the famous Schrödinger equation for the variation of a state in time, in the Schrödinger representation.

Since H is hermitian the quantity $S = e^{-\frac{iHt}{\hbar}}$ is unitary, and therefore the transformation from the Heisenberg to the Schrödinger representation is in fact simply a change of basis.

We have assumed throughout that H was not an explicit function of the time. If it is, i.e. if H = $H(q, p, t)$ then the entire argument goes through if we replace S by the operator U which is a solution of

$$i\hbar \frac{\partial U(t)}{\partial t} = H(t)\, U$$

subject to the initial condition

$$U(t=0) = 1 \; .$$

Because of the role the hamiltonian plays in determining the amics of the system, characteristic vectors of H (that is, states with efinite energy) play a very special role in quantum mechanics. If e state has a definite energy at time t = o, say E, then at time t

$$\Psi = e^{-\frac{iHt}{\hbar}}\Psi_0 = e^{-\frac{iEt}{\hbar}}\Psi_0$$

that a state of definite energy only has its phase affected by the ssage of time, and therefore represents exactly the same probability stribution for any observable as it initially had. Thus states of efinite energy are, in quantum mechanics, stationary states of the ystem. In particular, if a system has energy E at t = o, it will have nergy E at all times.

III. Special Systems

As an example of the application of the ideas of the previous two sections, let us consider several special systems which can be treated exactly and are of considerable physical importance.

(1) The harmonic oscillator

Consider a one dimensional harmonic oscillator with mass m and spring constant k. The kinetic energy $T = p^2/2m$, the potential energy $V = kq^2/2$ so that the hamiltonian is

$$H = \frac{p^2}{2m} + \frac{kq^2}{2}$$

What energy levels can this system have?

It is convenient to introduce instead of the operators p and q a new operator A along with its adjoint A^+, the real and imaginary parts of which are connected with p and q respectively. That is

$$p = \left(\frac{mk\hbar^2}{4}\right)^{1/4}(A+A^+)$$

$$q = i\left(\frac{\hbar^2}{4mk}\right)^{1/4}(A-A^+)$$

$$H = \frac{1}{2m}\left(\frac{mk\hbar^2}{4}\right)^{1/2}(A+A^+)^2 + \frac{k}{2}\left(\frac{\hbar^2}{4mk}\right)^{1/2} i^2(A-A^+)^2$$

$$= \frac{1}{2}\hbar\omega\,(AA^+ + A^+A)$$

where $\omega \equiv \sqrt{k/m}$, is the classical angular frequency of the oscillator. We can very easily obtain the commutation rules for A and A^+ :

$$(q,p) = \sqrt{\frac{\hbar^2}{4}}\; i\,(A-A^+, A+A^+)$$

$$= \frac{\hbar i}{2}\, 2(A,A^+) = i\hbar(A,A^+) = i\hbar, \text{ or}$$

$$(A,A^+) = 1$$

Thus H = $\frac{1}{2}\hbar\omega(2N+1)$, where N is the operator defined by $N = A^+A$.

We need only find the characteristic values of N. Now suppose we have a characteristic vector of N, say Ψ, belonging to the characteristic value ν.

$$N\Psi = \nu\Psi$$

Consider

$$(A, N) = (A, A^+)A = A, \text{ or}$$

$$AN = (N+1)A.$$

Similarly,

$$(A^+, N) = A^+(A^+, A) = -A^+, \text{ or}$$

$$A^+N = (N-1)A^+$$

Multiplying by A we get

$$AN\Psi = (N+1)A\Psi = \nu A\Psi$$

or

$$N(A\Psi) = (\nu-1)(A\Psi)$$

Similarly multiplying by A^+ gives

$$N(A^+\Psi) = (\nu+1)(A^+\Psi)$$

This if Ψ has the characteristic value ν

$A\Psi$ has the value $\nu-1$
$A^2\Psi$ " " " $\nu-2$, etc.
$A^+\Psi$ " " " $\nu+1$
$(A^+)^2\Psi$ " " " $\nu+2$, etc. .

On the other hand, it is clear that the possible characteristic values of N must be greater than or equal to zero since any characteristic value ν is given by

$$\nu = (\Psi_\nu, A^+A\Psi_\nu) = (A\Psi_\nu, A\Psi_\nu) \geq 0.$$

The series of values $\nu, \nu-1, \nu-2, \ldots$ must therefore terminate. The only way in which it can terminate is if when we apply the operator A to the last of corresponding characteristic vectors, we get nothing. That is, the characteristic vector corresponding to the lowest value a (say Ψ_0) must satisfy

$$A \Psi_0 = 0$$

But if this is so then $A^+ A \Psi_0 = 0$ or $N \Psi_0 = 0$ and therefore

$$N \Psi_0 = 0 \Psi_0$$

The lowest characteristic value of N is thus zero. Further, all the positive integers are characteristic values, since the vector $(A^+)^n \Psi_0$ gives

$$N\left((A^+)^n \Psi_0\right) = n\left((A^+)^n \Psi_0\right).$$

Lastly, these are the only possible characteristic values, since a non-integral value, say $\lambda = n + \alpha$, $0 < \alpha < 1$ would lead (via the vector $A^{n+1} \Psi_\lambda$) to the characteristic value $\alpha - 1 < 0$, which is impossible. The possible values of H (say E_n) are therefore

$$E_n = \frac{\hbar \omega}{2}(2n+1) \qquad n = 0, 1, 2, \ldots$$

It is very easy to work out the matrix elements of various operators involved, in this representation. We know that

$$A \Psi_n = C_n \Psi_{n-1},$$

where the factor C_n has been put in since we are assuming that the Ψ_n are normalized. However,

$$n = (\Psi_n, A^+ A \Psi_n) = (A \Psi_n, A \Psi_n) = c_n^2 (\Psi_{n-1}, \Psi_{n-1})$$

$$= C_n^2$$

$$C_n = \sqrt{n}.$$

Th~~is~~us $A\Psi_n = \sqrt{n}\,\Psi_{n-1}$. The matrix elements of A are therefore

$$(n'|A|n'') = (\Psi_{n'}, A\Psi_{n''}) = \sqrt{n''}\,(\Psi_{n'}, \Psi_{n''-1}) = \sqrt{n''}\,\delta_{n',n''-1}$$

The matrix A has the form

$$A = \begin{array}{c|cccccc} n' \backslash n'' & 0 & 1 & 2 & 3 & \cdots & \\ \hline 0 & 0 & \sqrt{1} & 0 & 0 & \cdots & \\ 1 & 0 & 0 & \sqrt{2} & 0 & \cdots & \\ 2 & 0 & 0 & 0 & \sqrt{3} & \cdots & \\ 3 & 0 & 0 & 0 & 0 & \sqrt{4} & \cdot \\ \vdots & \cdot & \cdot & \cdot & \cdot & \cdot & \sqrt{5}\ \ddots \end{array}$$

Since A$^+$ is the adjoint of A it has the form

$$A^+ = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \\ \sqrt{1} & 0 & 0 & 0 & 0 & \\ 0 & \sqrt{2} & 0 & 0 & 0 & \cdots \\ 0 & 0 & \sqrt{3} & 0 & 0 & \\ 0 & 0 & 0 & \sqrt{4} & 0 & \\ & & \vdots & & & \end{pmatrix}$$

which is clearly equivalent to

$$(n'|A^+|n'') = \sqrt{n''-1}\;\delta_{n',n''+1}$$

which is in turn the same as

$$A^+\Psi_n = \sqrt{n+1}\;\Psi_{n+1}\,.$$

From these matrix elements we can at once find the matrix elements for our original p and q.

$$p\,\Psi_n = \left(\frac{mk\hbar^2}{4}\right)^{1/4}\left[\sqrt{n}\,\Psi_{n-1} + \sqrt{n+1}\,\Psi_{n+1}\right]$$

$$q\,\Psi_n = i\left(\frac{\hbar^2}{4mk}\right)^{1/4}\left[\sqrt{n}\,\Psi_{n-1} - \sqrt{n+1}\,\Psi_{n+1}\right]$$

$$p = \left(\frac{mk\hbar^2}{4}\right)^{1/4}\begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0 & \\ \sqrt{1} & 0 & \sqrt{2} & 0 & 0 & \\ 0 & \sqrt{2} & 0 & \sqrt{3} & 0 & \cdots \\ 0 & 0 & \sqrt{3} & 0 & \sqrt{4} & \\ & & \cdots & & & \end{pmatrix}$$

$$q = i\left(\frac{\hbar^2}{4mk}\right)^{1/4}\begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0 & \\ -\sqrt{1} & 0 & \sqrt{2} & 0 & 0 & \cdots \\ 0 & -\sqrt{2} & 0 & \sqrt{3} & 0 & \\ 0 & 0 & -\sqrt{3} & 0 & \sqrt{4} & \\ & & \cdots & & & \end{pmatrix}$$

which are clearly both hermitian as they must be.

(2) Properties of Angular Momentum

Let us now study the properties of angular momentum in quantum mechanics. Consider first one particle. Classically the angular momentum vector $\vec{L}$ is given by

$$\vec{L} = \vec{r} \times \vec{p}$$

where $\vec{r}$ is the position vector of the particle, $\vec{p}$ is its momentum vector. In components

$$L_x = y p_z - z p_y$$
$$L_y = z p_x - x p_z$$
$$L_z = x p_y - y p_x$$

Since y, p_z etc. commute, we obtain well defined operators for L_x, L_y, L_z by simply replacing the classical quantities by their corresponding operators in quantum mechanics. We can at once obtain commutation rules for the L's from those of the co-ordinates and momenta:

$$\begin{aligned}(L_x, L_y) &= (y p_z - z p_y, z p_x - x p_z) \\ &= y p_x (p_z, z) + x p (z, p_z) = i\hbar (x p_y - y p_x) \\ &= i\hbar L_z\end{aligned}$$

Similarly $$(L_y, L_z) = i\hbar L_x$$
$$(L_z, L_x) = i\hbar L_y$$

One often writes these relationships in the compact form

$$\vec{L} \times \vec{L} = i\hbar \vec{L}$$

Now in quantum mechanics we have not only the possibility of angular momentum due to orbital motion, but it is possible for a particle to possess an <u>intrinsic</u> angular momentum (as, for example, the electron's spin). To take into account this possibility we must broaden the classical definition of angular momentum in such a way that it no longer refers to the specific formulae expressing $\vec{L}$ as a function of $\vec{r}$ and $\vec{p}$. For this purpose <u>it is natural to define in quantum</u>

mechanics any vector J which satisfies the commutation rules $\vec{J} \times \vec{J} = i\hbar \vec{J}$ as an angular momentum vector. This is possible because in quantum mechanics as soon as we know the commutation rules we can determine the properties of the operator. We shall later see that the vector $\vec{L}$ is only one specific case of an operator satisfying these commutation rules.

Possible values of an angular momentum

Since J_x, J_y, J_z do not commute it is only possible to measure one of them at a time; in other words we can only choose a representation in which one of them is diagonal, let us say J_z. We can consider along with the components the magnitude of the angular momentum

$$J^2 \equiv J_x^2 + J_y^2 + J_z^2$$

Clearly

$$\begin{aligned}(J_z, J^2) &= (J_z, J_x^2) + (J_z, J_y^2)\\ &= (J_z, J_x)J_x + J_x(J_z, J_x) + (J_z, J_y)J_y + J_y(J_z, J_y)\\ &= i\hbar\{J_y J_x + J_x J_y - J_x J_y - J_y J_x\}\\ &= 0,\end{aligned}$$

and similarly $(J_x, J^2) = (J_y, J^2) = 0$, so that we well as one component of $\vec{J}$ we may also simultaneously give the magnitude of $\vec{J}$. We shall now find the simultaneous characteristic values of J_z and J^2. Let us write for the characteristic values J_z' and $J^{2\prime}$ the expressions

$$J_z' = m\hbar$$

$$J^{2\prime} = j(j+1)\hbar^2,$$

where we have as yet made no assertions whatever about m or j. Write the characteristic vectors as $\Psi(j,m)$, that is

$$J_z \Psi(j,m) = \hbar m \, \Psi(j,m)$$

$$J^2 \Psi(j,m) = \hbar^2 j(j+1) \Psi(j,m) \, .$$

It is, just as in the case of the harmonic oscaillator, convenient to introduce complex operators. Put

$$J_+ \equiv J_x + i J_y$$

$$J_- \equiv J_x - i J_y = J_+{}^{+}$$

$$(J_+, J_z) = (J_x, J_z) + i(J_y, J_z) = i\hbar\{-J_y + iJ_x\}$$

$$(J_+, J_z) = -\hbar J_+ \quad , \text{ and similarly}$$

$$(J_-, J_z) = \hbar J_-$$

Therefore $\quad J_z J_+ - J_+ J_z = \hbar J_z \quad$ or

$$J_z J_+ = J_+ (J_z + \hbar)$$

Now consider the vector $J_+ \Psi(j,m)$. Since J_x and J_y commute with J^2 so does J_+. Therefore

$$J_+ J^2 \Psi(j,m) = J_+ j(j+1) \hbar^2 \Psi(j,m)$$

$$J^2 (J_+ \Psi(j,m)) = j(j+1)\hbar^2 (J_+ \Psi(j,m))$$

Further

$$\begin{aligned} J_z (J_+ \Psi(j,m)) &= J_+ (J_z + \hbar) \Psi(j,m) \\ &= J_+ (m+1)\hbar \, \Psi(j,m) \\ &= (m+1)\hbar \, (J_+ \Psi(j,m)) \, . \end{aligned}$$

Thus $J_+ \Psi(j,m)$ is again a characteristic vector belonging to the

$$J_+ \Psi(j,m) = C\,\Psi(j,m+1),$$

where C is a normalization constant to insure that if the function $\Psi(j,m)$ is normalized so is $\Psi(j,m+1)$. We can easily calculate C :

$$C^2\,(\Psi(j,m+1),\Psi(j,m+1)) = (J_+\Psi(j,m), J_+\Psi(j,m)),\ \text{or}$$

$$C^2 = (\Psi(j,m), J_-J_+\Psi(j,m)).$$

On the other hand

$$J_-J_+ = (J_x - iJ_y)(J_x + iJ_y) = J_x^2 + J_y^2 + i(J_xJ_y - J_yJ_x)$$

$$= J^2 - J_z^2 - \hbar J_z$$

$$J_-J_+\Psi(j,m) = (j(j+1) - m^2 - m)\hbar^2\Psi(j,m)$$

$$= \hbar^2(j-m)(j+m+1)\Psi(j,m)$$

Therefore

$$C^2 = \hbar^2(j-m)(j+m+1)$$

$$C = \hbar\sqrt{(j-m)(j+m+1)}$$

$$J_+\Psi(j,m) = \hbar\sqrt{(j-m)(j+m+1)}\;\Psi(j,m+1).$$

The matrix elements of J_+ are

$$(j'm'|J_+|j''m'') = \hbar\sqrt{(j''-m'')(j''+m''+1)}\;\delta_{j'j''}\,\delta_{m',m''+1}$$

Since $J_- = (J_+)^+$ we have for the matrix elements of J_-

$$(j'm'|J_-|j''m'') = (j''m''|J_+|j'm')^*$$

$$= \hbar\sqrt{(j''+m'')(j''-m''+1)}\;\delta_{j'j''}\,\delta_{m'+1,m''}$$

This is of course equivalent to

$$J_- \Psi(j,m) = \hbar\sqrt{(j+m)(j-m+1)}\;\Psi(j,m-1)$$

So far we have only investigated the effects of the operators, now we consider the possible values of j, m themselves. Since we have that

$$(\Psi(j,m), J_-J_+\Psi(j,m)) = (J_+\Psi(j,m), J_+\Psi(j,m)) \geq 0$$

or

$(j-m)(j+m+1) \geq 0$. Therefore $m \leq j$, [redacted] since we can clearly always take j positive. [J^2 has only positive characteristic values]. Similarly from

$$(\Psi(j,m), J_+J_-\Psi(j,m)) = (J_-\Psi(j,m), J_-\Psi(j,m)) \geq 0,$$

we get $(j+m)(j-m+1) \geq 0$ or $m \geq -j$

That means that for a fixed j m is bounded from above and below. Call the highest possible value of m = $\bar{m}$, the lowest $\underline{m}$. Then since $J_+\Psi(j,\bar{m})$ would have m = $\bar{m}$ + 1 we only avoid a contradiction if this state vanishes, i.e., if

$$J_+\Psi(j,\bar{m}) = \hbar\sqrt{(j-\bar{m})(j+\bar{m}+1)}\;\Psi(j,\bar{m}+1) \equiv 0$$

$$(j-\bar{m})(j+\bar{m}+1) = 0 \quad , \quad \text{or} \quad \bar{m} = j$$

Similarly

$$J_-\Psi(j,\underline{m}) = \hbar\sqrt{(j+\underline{m})(j-\underline{m}+1)}\;\Psi(j,\underline{m}-1)$$

must vanish to avoid a contradiction, that is

$$(j+\underline{m})(j-\underline{m}+1) = 0 \qquad \underline{m} = -j .$$

On the other hand, if -j is an allowed value of m, so is $-j+1$, $-j+2$, etc. (obtained by successive application of J_+ to $\Psi(j,m)$). This series must terminate with + j since we know that +j is the highest possible value of m. Thus we must have (where n is some non-negative integer) :

$$-j + n = j \quad \text{or} \quad j = \frac{n}{2} .$$

Therefore the possible values of j and m are completely known:

j has any half integral value $j = 0, \frac{1}{2}, 1, 3/2$, etc

m has, for fixed j, the possible values $-j, -j+1, \cdots, +j$.

Clearly there are 2j+1 possible m values for a given j, so that a state with a definite total angular momentum squared $j(j+1)\hbar^2$ is 2 j+1 fold degenerate. This is often called **spacial degeneracy**.

Some special cases are of considerable interest.

(a) $\underline{j = 0}$. Then clearly $J'^2 = 0$, $m = 0$; the only possible state is $\Psi(0,0)$. $J_+ \Psi(0,0) = 0$ and therefore $J_x \Psi(0,0) = 0$ and $J_y \Psi(0,0) = 0$. Thus we can simultaneously diagonalize J_x, J_y, J_z in this case: they can all have only the value zero.

(b) $\underline{j = \frac{1}{2}}$. Then $J'^2 = \hbar^2(\frac{1}{2})(\frac{1}{2}+1) = \frac{3}{4}\hbar^2$, and $m = \pm 1/2$.

The matrix elements of J_+ and J_- are easily seen to be

$$(\tfrac{1}{2}, \tfrac{1}{2} | J_+ | \tfrac{1}{2}, -\tfrac{1}{2}) = \hbar \sqrt{(\tfrac{1}{2} + \tfrac{1}{2})(\tfrac{1}{2} - \tfrac{1}{2} + 1)} = \hbar ,$$

all others vanishing. Similarly

$$(\tfrac{1}{2}, -\tfrac{1}{2} | J_- | \tfrac{1}{2}, \tfrac{1}{2}) = \hbar$$

all others vanishing.

$$J_+ = \hbar \begin{array}{cc} \begin{array}{cc} 1/2 & -1/2 \end{array} & \\ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} & \begin{array}{c} 1/2 \\ -1/2 \end{array} \end{array}$$

$$J_- = \hbar \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} ,$$

the labeling telling the m values of the matrix element .

$$J_x = \frac{1}{2}(J_+ + J_-) = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$J_y = \frac{1}{2i}(J_+ - J_-) = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

$$J_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

Squaring we easily see

$$J_x^2 = \frac{\hbar^2}{4} * \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \frac{\hbar^2}{4}$$

$$J_y^2 = \frac{\hbar^2}{4}$$

$$J_z^2 = \frac{\hbar^2}{4}$$

so that $J_x^2 + J_y^2 + J_z^2 = \frac{3}{4}\hbar^2$, as it must.

One often introduces the matrices

$$\sigma_x \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

which are known as the <u>Pauli spin matrices</u>. In terms of them

$$J_x = \frac{\hbar}{2}\sigma_x, \quad J_y = \frac{\hbar}{2}\sigma_y, \quad J_z = \frac{\hbar}{2}\sigma_z, \quad \vec{J} = \frac{\hbar}{2}\vec{\sigma}.$$

The Pauli matrices have the properties

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$$

$$\sigma_x\sigma_y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = i\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = i\sigma_z$$

$$\sigma_y\sigma_z = i\sigma_x$$

$$\sigma_z\sigma_x = i\sigma_y$$

$$\sigma_y \sigma_x = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = -i\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = -i\sigma_z = -\sigma_x \sigma_y$$

$$\sigma_z \sigma_y = -\sigma_y \sigma_z$$

$\sigma_x \sigma_z = -\sigma_z \sigma_x$, and of course

$$\vec{\sigma} \times \vec{\sigma} = 2i\vec{\sigma}$$

ich follow at once from the general commutation rules for angular menta.

This case is realized in practice. It is a well known fact that in der to explain the properties of the electron one must associate with it t only an orbital angular momentum but also an intrinsic angular mentum $j = 1/2$, which is present always. This is known as the ectron spin.

So far we have dealt with the most general case of the angular mentum of a particle, whether it be due to orbital motion, intrinsic spin, a combination of both. Let us now discuss the case of orbital angular mentum alone. We shall see that in this case the situation simplifies mewhat and we get only integral values of the total angular momentum.

We have $\vec{L} = \vec{r} \times \vec{p}$. We shall diagonalize L_z, L^2 as efore. Put $L_z' = m\hbar$, $L^2 = l(l+1)\hbar^2$. It is convenient to use a presentation in which the coordinates (x, y, z) are diagonal:

$$\Psi(l,m) = \int \Psi(xyz)\,dx\,dy\,dz\,(xyz|l,m)$$

om our previous work we know that

$$L_z \Psi(l,m) = m\hbar\, \Psi(l,m)$$

$$L^2 \Psi(l,m) = l(l+1)\hbar\, \Psi(l,m) \quad \text{are}$$

uivalent to

$$L_z(xyz|l,m) = m\hbar(xyz|l,m)$$

$$L^2(xyz|l,m) = l(l+1)\hbar(xyz|l,m),$$

where we replace x, y, z by simple multiplication with x, y, z and p_x, p_y, p_z with $\frac{\hbar}{i}\frac{\partial}{\partial x}, \frac{\hbar}{i}\frac{\partial}{\partial y}, \frac{\hbar}{i}\frac{\partial}{\partial z}$ respectively.

Thus

$$L_z = \frac{\hbar}{i}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right), \text{ etc.}$$

It is much more convenient to work with spherical coordinates instead of rectangular coordinates. Put

$$x = r\sin\theta\cos\varphi$$

$$y = r\sin\theta\sin\varphi$$

$$z = r\cos\theta$$

Clearly the operator

$$\frac{\partial}{\partial\varphi} = \frac{\partial x}{\partial\varphi}\frac{\partial}{\partial x} + \frac{\partial y}{\partial\varphi}\frac{\partial}{\partial y} + \frac{\partial z}{\partial\varphi}\frac{\partial}{\partial z}$$

$$= -r\sin\theta\sin\varphi\frac{\partial}{\partial x} + r\sin\theta\cos\varphi\frac{\partial}{\partial y}$$

$$= \left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) = \frac{i}{\hbar}L_z$$

$$L_z = \frac{\hbar}{i}\frac{\partial}{\partial\varphi}$$

In an exactly similar manner we may verify that

$$L_x + iL_y = L_+ = \hbar e^{i\varphi}\left(\frac{\partial}{\partial\theta} + i\cot\theta\frac{\partial}{\partial\varphi}\right)$$

$$L_x - iL_y = L_- = -\hbar e^{-i\varphi}\left(\frac{\partial}{\partial\theta} - i\cot\theta\frac{\partial}{\partial\varphi}\right).$$

(The best way to establish this is by directly checking the commutation rules for angular momenta.) Let us write $(xyz|l,m) \equiv \psi_{l,m}(r,\theta,\varphi)$.

We must have

$$\frac{\hbar}{i}\frac{\partial}{\partial\varphi}\psi_{l,m}(r\theta\varphi) = m\hbar\,\psi_{l,m}(r\theta\varphi) \quad \text{or}$$

$$\frac{\partial\psi_{lm}}{\partial\varphi} = im\,\psi_{l,m}$$

Therefore the φ dependence of $\psi_{l,m}$ is given by

$$\psi_{l,m} \sim e^{im\varphi}$$

Put

$$\psi_{l,m} = C_{lm}\,R_{lm}(r)\,\Theta_l^m(\theta)\,e^{im\varphi},$$

where Θ_l^m is still to be determined by the requirement that $\Psi_{l,m}$ is also a state with definite L^2. From our previous work with an arbitrary angular momentum we know that

$$(L_x + iL_y)\,\Psi_{l,m} = \hbar\sqrt{(l-m)(l+m+1)}\;\Psi_{l,m+1}$$

$$(L_x - iL_y)\,\Psi_{l,m} = \hbar\sqrt{(l+m)(l-m+1)}\;\Psi_{l,m-1}\,, \text{ or}$$

$$e^{i\varphi}\left(\frac{\partial}{\partial\theta} + i\cot\theta\,\frac{\partial}{\partial\varphi}\right)\psi_{l,m} = \sqrt{(l-m)(l+m+1)}\;\psi_{l,m+1}$$

$$-e^{-i\varphi}\left(\frac{\partial}{\partial\theta} - i\cot\theta\,\frac{\partial}{\partial\varphi}\right)\psi_{l,m} = \sqrt{(l+m)(l-m+1)}\;\psi_{l,m-1}$$

This relationship holds for normalized $\Psi_{l,m}$. The normalization condition in terms of $\psi_{l,m}$ is clearly

$$(\Psi_{l,m}, \Psi_{l,m}) = \int |(xyz|l,m)|^2\,dx\,dy\,dz$$

$$= \int |\psi_{lm}|^2\,r^2 dr\,\sin\theta\,d\theta\,d\varphi$$

$$= C^2\int_0^\infty |R_{lm}|^2 r^2 dr \int_0^\pi |\Theta_l^m|^2 \sin\theta\,d\theta \int_0^{2\pi} |e^{im\varphi}|^2 d\varphi$$

$$= 1$$

we take C = $\frac{1}{\sqrt{2\pi}}$

$$\int_0^\infty |R_{lm}|^2 r^2\, dr = 1$$

$$\int_0^\pi |\Theta_l^m|^2 \sin\theta\, d\theta = 1$$

e have clearly normalized $\Psi_{l,m}$. Substituting now in the above elations we get

$$\left(\frac{d}{d\theta} - m\cot\theta\right)\Theta_l^m = \sqrt{(l-m)(l+m+1)}\;\Theta_l^{m+1}$$

$$\left(\frac{d}{d\theta} + m\cot\theta\right)\Theta_l^m = -\sqrt{(l+m)(l-m+1)}\;\Theta_l^{m-1}$$

or the normalized Θ_l^m.

hese functions are easy to construct. Consider the first of these quations. We may rewrite it as

$$(\sin\theta)^m \frac{d}{d\theta}\left\{(\sin\theta)^{-m}\,\Theta_l^m\right\} = \sqrt{(l-m)(l+m+1)}\;\Theta_l^{m+1}$$

utting $\Theta_l^m = (\sin\theta)^m P_l^m$ we obtain

$$\frac{1}{\sin\theta}\frac{d}{d\theta}P_l^m = \sqrt{(l-m)(l+m+1)}\;P_l^{m+1}\quad , \text{ or}$$

$$\frac{d P_l^m}{d(\cos\theta)} = -\sqrt{(l-m)(l+m+1)}\;P_l^{m+1}.$$

ake $m = -l$. Then

$$P_l^{-l+1} = -\frac{1}{\sqrt{(2l)(1)}}\frac{d}{d(\cos\theta)}P_l^{-l}$$

$$P_l^{-l+2} = -\frac{1}{\sqrt{(2l-1)(2)}}\frac{d P_l^{-l+1}}{d(\cos\theta)} = \frac{1}{\sqrt{(2l)(2l-1)(1)(2)}}\left(\frac{d}{d(\cos\theta)}\right)^2 P_l^{-l},$$

tc.. Finally we may write

$$P_\ell^{-\ell+k} = (-)^k \frac{1}{\sqrt{(2\ell)(2\ell-1)\cdots(2\ell-k+1)(1)(2)\cdots(k)}} \left(\frac{d}{d(\cos\theta)}\right)^k P_\ell^{-\ell}$$

ere k is any positive integer $\leq 2\ell$.

can easily calculate $P_\ell^{-\ell}$ and P_ℓ^{ℓ}. Since

$$\left(\frac{d}{d\theta} - \ell\cot\theta\right)\Theta_\ell^{\ell} = 0 \quad \text{and}$$

$$\left(\frac{d}{d\theta} - \ell\cot\theta\right)\Theta_\ell^{-\ell} = 0$$

obtain

$$\Theta_\ell^{\ell} = C_1(\sin\theta)^\ell \quad , \quad P_\ell^{\ell} = C_1$$

$$\Theta_\ell^{-\ell} = C_2(\sin\theta)^\ell \quad , \quad P_\ell^{-\ell} = C_1\sin^{2\ell}\theta.$$

can easily be determined from the normalization. Before we do this, wever, we may use the relationships obtained here to show that ℓ must be integer. If we put $k = 2\ell$ in our previous equation, we obtain

$$P_\ell^{\ell} = (\text{Constant})\left(\frac{d}{d(\cos\theta)}\right)^{2\ell}\sin^{2\ell}\theta$$

$$\text{Constant} = \left(\frac{d}{d(\cos\theta)}\right)^{2\ell}(1-\cos^2\theta)^\ell$$

w if ℓ is an integer this equation is clearly satisfied, since then may expand by the binomial theorem; the highest term we obtain is s$^{2\ell}\theta$, and on 2ℓ differentiations with respect to cos θ we tain a constant. On the other hand, if is not integral the series es not terminate and we get a contradiction. For example, for $\ell = 1/2$ would have

$$\text{Const} = \frac{d}{d(\cos\theta)}\sqrt{1-\cos^2\theta} = \frac{-\cos\theta}{\sqrt{1-\cos^2\theta}} = -\cot\theta$$

ch is impossible. Thus ℓ may only have integral values for orbital ular momentum.

To obtain final formulae we now need the normalization:

$$= \int_0^\pi |\Theta_\ell^{-\ell}|^2 \sin\theta \, d\theta = C_2^2 \int_0^\pi \sin^{2\ell}\theta \sin\theta \, d\theta = C_2^2 \int_0^\pi \sin^{2\ell+1}\theta \, d\theta$$

$$= C_2^2 \, 2^{2\ell+1} \frac{(\ell!)^2}{(2\ell+1)!} ,$$

$${}_2 = \sqrt{\frac{(2\ell+1)!}{2^{2\ell+1}(\ell!)^2}}$$

$$P_\ell^{-\ell+k} = (-)^k \sqrt{\frac{(2\ell-k)!}{(2\ell)!\,k!}} \left(\frac{d}{d(\cos\theta)}\right)^k \sqrt{\frac{(2\ell+1)!}{2^{2\ell+1}(\ell!)^2}} \sin^{2\ell}\theta$$

$$= (-)^k \sqrt{\frac{(2\ell-k)!\,(2\ell+1)!}{(2\ell)!\,(\ell!)^2\,2^{2\ell+1}\,k!}} \left(\frac{d}{d(\cos\theta)}\right)^k (1-\cos^2\theta)^\ell .$$

put $-\ell + k = m$, $k = \ell + m$. Then we have finally

$$P_\ell^m = (-)^m \sqrt{\frac{2\ell+1}{2}\,\frac{(\ell-m)!}{(\ell+m)!}} \, \frac{1}{2^\ell \ell!} \left(\frac{d}{d(\cos\theta)}\right)^{\ell+m} (\cos^2\theta - 1)^\ell$$

$$\Theta_\ell^m = \sin^m\theta \, P_\ell^m .$$

can show very easily, by just expanding $(\cos^2\theta - 1)^\ell$ — that

$$\Theta_\ell^{-m} = (-1)^m \, \Theta_\ell^m .$$

Combination of Angular Momentum

Suppose we have two systems, one with angular momentum vector $\vec{J}_1$ and the other with $\vec{J}_2$. We take the systems independent, so that all components of $\vec{J}_1$ commute with all components of $\vec{J}_2$. Then we can simultaneously diagonalize J_1^2, J_2^2, J_{1z}, J_{2z}, and a state is specified by

$$\Psi(j_1, j_2; m_1, m_2).$$

The total number of states with a given j_1 and j_2 is obviously $(2j_1+1)(2j_2+1)$. This is, however, not the only representation of interest. If we define the total angular momentum $\vec{J}$ by

$$\vec{J} = \vec{J}_1 + \vec{J}_2 .$$

Then J^2 commutes with J_1^2, J_2^2, and so does J_z:

$$(J^2, J_1^2) = (J_1^2 + J_2^2 + \vec{J}_1 \cdot \vec{J}_2 + \vec{J}_2 \cdot \vec{J}_1,\ J_1^2) = 0$$

since $(\vec{J}_1, J_1^2) = 0$

$$(J_z, J_1^2) = (J_{1z}, J_1^2) + (J_{2z}, J_1^2) = 0 .$$

Further

$$\vec{J} \times \vec{J} = (\vec{J}_1 + \vec{J}_2) \times (\vec{J}_1 + \vec{J}_2) = \vec{J}_1 \times \vec{J}_1 + \vec{J}_2 \times \vec{J}_2 + \vec{J}_1 \times \vec{J}_2 + \vec{J}_2 \times \vec{J}_1$$

$$= \vec{J}_1 \times \vec{J}_1 + \vec{J}_2 \times \vec{J}_2 = i\hbar(\vec{J}_1 + \vec{J}_2) = i\hbar \vec{J},$$

so that as could be expected $\vec{J}$ also has the commutation rules of an angular momentum. Therefore J^2 and J_z commute, and we can take the set J_1^2, J_2^2, J^2, J_z as the quantities to be simultaneously diagonalized. Since $\vec{J}$ has the commutation rules of an angular momentum the characteristic values of J^2 will be of the form $j(j+1)\hbar^2$ and those of J_z of the form $m\hbar$, $-j \le m \le j$. We shall investigate the transformation from one representation to the other now in some detail. Call a state with fixed j_1, j_2, j, m $\Phi(j_1, j_2; j, m)$. We may certainly write

$$\Phi(j_1 j_2; j m) = \sum_{m_1, m_2} \Psi(j_1 j_2; m_1 m_2)(j_1 j_2; m_1 m_2 | j_1 j_2; j m)$$

$$J_z \Phi(j_1 j_2; j m) = m \hbar \Phi(j_1 j_2; j, m)$$

$$= \sum_{m_1, m_2} (J_{1z} + J_{2z}) \Psi(j_1 j_2; m_1, m_2)(j_1 j_2 m_1 m_2 | j_1 j_2 j m)$$

$$= \sum_{m_1, m_2} \hbar (m_1 + m_2) \Psi(j_1 j_2; m_1 m_2)(j_1 j_2 m_1 m_2 | j_1 j_2 j m).$$

Therefore only those terms in the sum contribute if

$m_1 + m_2 = m$, i.e. $(j_1 j_2 m_1 m_2 | j_1 j_2 j m) = 0$ unless $m_1 + m_2 = m$

We may then write

$$\Phi(j_1 j_2 j m) = \sum_{\substack{m_1, m_2 \\ m_1 + m_2 = m}} \Psi(j_1 j_2; m_1 m_2)(j_1 j_2 m_1 m_2 | j_1 j_2 j m).$$

From this we have that the largest possible value of m is

$$\max.(m) = \max(m_1) + \max.(m_2)$$

$$= j_1 + j_2.$$

However, each j will always mean a state with $m = j$ is possible, so that we obtain at once the result that

$$j_{max} = j_1 + j_2 .$$

For this state we have

$$\Phi(j_1 j_2; j_{max}, j_{max}) = \Psi(j_1 j_2; j_1, j_2),$$

since there is only one state with $m_1 = j_1$, $m_2 = j_2$. Now consider those states with $m = j_{max} - 1$. There are two ways to obtain this, i.e.

(1) $m_1 = j_1$ $m_2 = j_2 - 1$

(2) $m_1 = j_1 - 1$ $m_2 = j_2$

One of these states must certainly belong to $j = j_{max}$, since it has all m's from $-j_{max}$ to $+j_{max}$. Therefore the other must belong to $j = j_{max}-1$ since we need this high an angular momentum to obtain $m = j_{max}-1$. Thus we have established that a state with $j = j_1+j_2-1$ exists for the composite system. We can continue the process in exactly the same way: Consider those states with $m = j_{max}-2 = j_1+j_2-2$. There are three ways of obtaining this

(1) $m_1 = j_1 \qquad m_2 = j_2-2$

(2) $m_1 = j_1-1 \qquad m_2 = j_2-1$

(3) $m_1 = j_1-2 \qquad m_2 = j_2$.

But one of these must belong to $j = j_1+j_2$, another to $j = j_1+j_2-1$ and therefore we have that a state with $j = j_1+j_2-2$ also exists. We can continue in this fashion till we get a state with

$$j = j_1+j_2-n \quad \text{and} \quad j_1-n = -j_1$$

(where we have assumed $j_1 \le j_2$), since for n greater than this we would need an $m_1 < -j_1$. Thus $n = 2j_1$, $j_{min} = |j_2-j_1|$. The possible values of j are therefore

$$j = j_1+j_2,\ j_1+j_2-1,\ \dots,\ |j_2-j_1|,$$

where we have dropped the restriction that $j_1 \le j_2$. We can very easily see that the total number of states is just correct

Number of states $= \sum_{j=j_{min}}^{j_{max}} (2j+1)$, the factor $2j+1$ being just the number of states with a given j and different m. This is an arithmetical series and is easily summed:

$$\text{Number of States} = 2\left(\frac{j_{max}+j_{min}}{2}\right)(j_{max}-j_{min}+1)+(j_{max}-j_{min}+1)$$

$$= (j_{max}+j_{min}+1)(j_{max}-j_{min}+1)$$

$$= (2j_1+1)(2j_2+1),$$

which is just the original value for the number of states.

The explicit calculation of the transformation coefficients is in general very tedious. To illustrate the general technique, however, we give it for a very simple case. Consider the case of $j_1 = j_2 = \frac{1}{2}$. Then $j = 1, 0$. Denote the states $\Phi(\frac{1}{2}, \frac{1}{2}; j, m)$ by $\Phi(j, m)$ the states $\Psi(\frac{1}{2}, \frac{1}{2}; m_1, m_2)$ by $\Psi(m_1, m_2)$, since we shall hold j_1, j_2 fixed throughout the calculation. There are four possible states:

m_1	m_2		j	m
$\frac{1}{2}$	$\frac{1}{2}$		1	1
$\frac{1}{2}$	$-\frac{1}{2}$		1	0
$-\frac{1}{2}$	$\frac{1}{2}$	or	1	-1
$-\frac{1}{2}$	$-\frac{1}{2}$	.	0	0

By our previous work we know

$$\Phi(1,1) = \Psi(\tfrac{1}{2}, \tfrac{1}{2}),$$

since this is the only state with m = 1.

But

$$J_- \Phi(j,m) = \sqrt{(j+m)(j-m+1)}\ \Phi(j, m-1)$$

$$J_- \Phi(1,1) = \sqrt{2}\ \Phi(1,0)$$

$$\Phi(1,0) = \frac{1}{\sqrt{2}} J_- \Phi(1,1) = \frac{1}{\sqrt{2}} (J_{1-} + J_{2-}) \Phi(1,1)$$

$$= \frac{1}{\sqrt{2}} (J_{1-} + J_{2-})\ \Psi(\tfrac{1}{2}, \tfrac{1}{2}) = \frac{1}{\sqrt{2}} (\Psi(-\tfrac{1}{2}, \tfrac{1}{2}) + \Psi(\tfrac{1}{2}, -\tfrac{1}{2}$$

$$J_- \Phi(1,0) = \sqrt{2}\ \Phi(1,-1)$$

$$\Phi(1,-1) = \frac{1}{\sqrt{2}} (J_{1-} + J_{2-}) \frac{1}{\sqrt{2}} (\Psi(-\tfrac{1}{2}, \tfrac{1}{2}) + \Psi(\tfrac{1}{2}, -\tfrac{1}{2}))$$

$$= \frac{1}{2} (\Psi(-\tfrac{1}{2}, -\tfrac{1}{2}) + \Psi(-\tfrac{1}{2}, -\tfrac{1}{2})) = \Psi(-\tfrac{1}{2}, -\tfrac{1}{2}),$$

as of course it must.

There is only one state left, $j=0$. Since for this state $m=0$; it must be a linear combination of $\Psi(-\frac{1}{2},\frac{1}{2})$ and $\Psi(\frac{1}{2},-\frac{1}{2})$, and (since it has a different j) it must be orthogonal to $\Phi(1,0)$. Clearly the only possibility is

$$\Phi(0,0) = \frac{1}{\sqrt{2}}\left(\Psi(-\tfrac{1}{2},\tfrac{1}{2}) - \Psi(\tfrac{1}{2},-\tfrac{1}{2})\right),$$

which satisfies all requirements and is normalized. Exactly the technique used here will work in all cases, but it gets very cumbersome when j_1 and j_2 become at all large. General formulae have been given by Wigner.

(3) The Hydrogen Atom

As a final example, let us consider the spectrum of a hydrogen like atom. The model we take is that of a charge $-e$ moving in a potential Ze/r, where r is the distance from the (fixed) nucleus to the electron. The hamiltonian is clearly

$$H = \frac{p_x^2+p_y^2+p_z^2}{2m} - \frac{Ze^2}{r}, \quad r = +\sqrt{x^2+y^2+z^2},$$

and we shall ask after this characteristic values of this H. It is convenient to rewrite the expression for $p^2 = p_x^2+p_y^2+p_z^2$. To do this let us consider first the expression for the total angular momentum:

$$L^2 = L_x^2+L_y^2+L_z^2 = (yp_z - zp_x)^2 + (zp_x - xp_z)^2 + (xp_y - yp_x)^2$$

$$= r^2p^2 + \vec{r}\cdot\vec{p}\,(\vec{r}\cdot\vec{p} - i\hbar).$$

Define $$p_r \equiv \frac{1}{r}(\vec{r}\cdot\vec{p} - i\hbar).$$

$$p_r^2 = \frac{1}{r}(\vec{r}\cdot\vec{p} - i\hbar)\frac{1}{r}(\vec{r}\cdot\vec{p} - i\hbar).$$

However, $(\vec{r}\cdot\vec{p} - i\hbar)\frac{1}{r} = \frac{1}{r}(\vec{r}\cdot\vec{p})$, and therefore

$$p_r^2 = \frac{1}{r^2}(\vec{r}\cdot\vec{p})(\vec{r}\cdot\vec{p} - i\hbar),$$ so that

$$L^2 = r^2 p^2 - r^2 p_r^2 .$$

The quantity p_r is the momentum conjugate to r, since

$$(r, p_r) = \left(r, \frac{1}{r}(\vec{r}\cdot\vec{p})\right) = \frac{1}{r}(r, \vec{r}\cdot\vec{p}) = \frac{1}{r} i\hbar \left(x\frac{\partial r}{\partial x} + y\frac{\partial r}{\partial y} + z\frac{\partial r}{\partial z}\right) = i\hbar$$

as it must for conjugate momentum and coordinate. Further we would expect

$$p_r = m\dot{r} = m\frac{1}{i\hbar}[r, H] = -\frac{1}{2i\hbar}[p^2, r]$$

$$= -\frac{1}{2i\hbar}\left\{\vec{p}\cdot(\vec{p}, r) + (\vec{p}, r)\cdot\vec{p}\right\}$$

$$(\vec{p}, r) = \frac{\hbar}{i}\frac{\vec{r}}{r}$$

$$p_r = \frac{1}{2}\left(\vec{p}\cdot\frac{\vec{r}}{r} + \frac{\vec{r}}{r}\cdot\vec{p}\right) = \frac{1}{2}\left(2\frac{\vec{r}}{r}\cdot\vec{p} - \frac{2i\hbar}{r}\right)$$

$$= \frac{\vec{r}}{r}\cdot p - \frac{i\hbar}{r},$$

just as it must.

Classically we would have $p_r = \frac{\vec{r}}{r}\cdot\vec{p}$, but this is not hermitian. If we symmetrize (add the adjoint and divide by two) we again come back to exactly this same definition. We may therefore write

$$r^2 p^2 = L^2 + r^2 p_r^2$$

$$p^2 = p_r^2 + \frac{1}{r^2}L^2,$$ or finally

$$H = \frac{1}{2m}p_r^2 + \frac{1}{2mr^2}L^2 - \frac{Ze^2}{r}.$$

Now direct computation shows that $\vec{L}$ commutes with r and p_r (in fact it commutes with any observable which is rotationally invariant).

$$(L_x, r) = (y p_z - z p_y, r) = \frac{\hbar}{i}\left(y\frac{\partial r}{\partial z} - z\frac{\partial r}{\partial y}\right) = \frac{\hbar}{i}\left(\frac{yz}{r} - \frac{yz}{r}\right) = 0,$$

Similarly for L_y, L_z and for the commutators with p. Thus we get that $\vec{L}$ and therefore L^2 commutes with H and can therefore be simultaneously diagonalized. Taking L^2 diagonal then gives as possible values for L^2, $\hbar^2 l(l+1)$, $l = 0, 1, 2, \ldots$. Let a state with definite l

be Ψ_l. Then the characteristic value problem

$$H \Psi_l = E_l \Psi_l$$

reduces to

$$H_l \Psi_l = E_l \Psi_l ,$$

where

$$H_l \equiv \frac{1}{2m} p_r^2 + \frac{\hbar^2}{2mr^2} l(l+1) - \frac{Ze^2}{r},$$

and $(r, p_r) = i\hbar$, which is a problem with one degree of freedom, and must in principle be solved for each $l = 0, 1, 2, \cdots$. If we solve this problem we will get a set of states each belonging to a certain energy and l value. Let us label these states with an index n, which can be either discrete or continuous. That is

$$H_l \Psi_{nl} = E_{n,l} \Psi_{nl} .$$

One way to solve this problem would be to go into the basis with r diagonal. Then we put $p_r = \frac{\hbar}{i}\frac{\partial}{\partial r}$, and try to solve the resulting ordinary differential equation. This is the standard method used in wave mechanics, and is found in all the text books. We shall, however, give an alternate treatment which is more closely related to the methods used for the harmonic oscillator and the angular momentum. Define an operator B_l by the equation

$$B_l = \frac{1}{\sqrt{2m}} p_r - i\left(\frac{(l+1)\hbar}{\sqrt{2m}\, r} - \frac{Ze^2\sqrt{2m}}{2\hbar(l+1)}\right)$$

$$B_l^+ = \frac{1}{\sqrt{2m}} p_r + i\left(\frac{(l+1)\hbar}{\sqrt{2m}\, r} - \frac{Ze^2\sqrt{2m}}{2\hbar(l+1)}\right).$$

Then

$$B_l B_l^+ = H_l + \eta_l , \quad \eta_l = \frac{(Ze^2)^2 m}{2\hbar^2(l+1)^2},$$

as may be seen by direct computation. The B_l, B_l^+ have a simple commutation rule:

$$(B_l, B_l^+) = -\frac{\hbar^2}{2m} \frac{2(l+1)}{r^2} .$$

By means of this we see at once a very important result

$$B_l^+ B_l = B_l B_l^+ + \frac{\hbar^2}{2m} \frac{2(l+1)}{r^2} = H_l + \eta_l + \frac{\hbar^2}{2m} \frac{2(l+1)}{r^2} = H_{l+1} + \eta_l .$$

Further we have

$$B_l^+ H_l = H_l B_l^+ + (B_l^+, H_l)$$

$$= H_l B_l^+ + (B_l^+, B_l B_l^+) = [H_l + (B_l^+, B_l)] B_l^+$$

$$= \left[H_l + \frac{\hbar^2}{2m} \frac{2(l+1)}{r^2}\right] B_l^+ = H_{l+1} B_l^+ , \text{ or}$$

$$B_l^+ H_l = H_{l+1} B_l^+ .$$

By taking the adjoint we get $H_l B_l = B_l H_{l+1}$ or $B_l H_{l+1} = H_l B_l$.

Now consider the characteristic value problem:

$$H_l \Psi_{nl} = E_{nl} \Psi_{nl}$$

Multiply by B_l^+

$$B_l^+ H_l \Psi_{nl} = E_{nl} B_l^+ \Psi_{nl}$$

$$H_{l+1} (B_l^+ \Psi_{nl}) = E_{nl} (B_l^+ \Psi_{nl}).$$

But

$$H_{l+1} \Psi_{n,l+1} = E_{n,l+1} \Psi_{n,l+1}$$

Therefore we must take

$$E_{n,l+1} = E_{n,l} = E_n$$ (the energy is independent of

and

$$B_l^+ \Psi_{n,l} = C \Psi_{n,l+1}$$

Put $l_0 + 1 = n$, then $n = 1, 2, \cdots$

$$E_n = \frac{-m(ze^2)^2}{2\hbar^2 n^2},$$

which is the famous Bohr formula for the hydrogen levels. Each of these levels is degenerate in l, i.e., there are states with $l = 0, 1, \cdots, n-1$ having the same energy. Since L_z can also be simultaneously diagonalized, there are to each of these l states $2l+1$ states with $m = -l, -l+1, \cdots, +l$ respectively.

We can easily, if we wish, obtain expressions for the radial functions by first solving the equation

$$B_{l_0}^{+} \Psi_{n l_0} = 0,$$

which is first order and very simple, and then obtaining the others by lowering with the B_l. It is also very easy to obtain the matrix elements of p_r and $\frac{1}{r}$ since we have those for B_l and B_l^{+}. However we shall not go into these questions here.

For (2) $E_n > 0$ and therefore $C^2 = E_n + \eta_l$ is always greater than zero. This means that no matter what l we take there is a solution, and therefore we have no restriction onthe value of the energy. That is, all positive values of the energy are allowed.

IV Approximation Methods

There are only very few problems in quantum mechanics for which exact solutions can be given, and one is generally driven to some form of approximation. We shall consider three types of approximation, useful in different connections:

(A) Perturbation method

(B) Variational method

(C) W.K.B. method

(A) Perturbation Theory

In perturbation theory we envisage a situation like this: The hamiltonian H (though it could be any other operator) consists of two parts

$$H = H_0 + H_1$$

H$_0$ represents an operator for which we can solve the characteristic value problem, H$_1$ represents an operator which is in some sense small compared to H$_0$, that is, we imagine the states of the system do not iffer too much from those of H$_0$. An example would be the anharmonic oscillator; that is, if we try to correct the motion of an oscillator for small nonquadratic terms, say λq^3 terms, in the potential energy.

We shall call the term H_1 "the perturbation," and shall iscuss three cases:

(1) All the levels of H_0 are nondegenerate.

(2) Degeneracy is present in the levels of H$_0$

(3) Transitions caused by a perturbation.

(1) Nondegenerate perturbation theory

The characteristic value problem has the form

$$(H_0 + H_1)\,\Psi = E\,\Psi$$

Now let us choose a basis with H$_0$ diagonal, and let Ψ^0 be some state of H$_0$, i.e.

$$H_0\,\Psi^0 = E^0\,\Psi^0$$

hen the true state Ψ and the state Ψ^0 will in general be connected by a unitary transformation U

$$\Psi = U\,\Psi^0,$$

ince all that is involved is a change of basis. Then

$$(H_0 + H_1)\,U\,\Psi^0 = E\,U\,\Psi^0 \qquad \text{or}$$

$$U^{-1}(H_0 + H_1)\,U\,\Psi^0 = E\,\Psi^0.$$

That is, we must find U such that $U^{-1}(H_0+H_1)U$ is diagonal in the H_0 representation. The diagonal elements will then be the allowed energies. Call this diagonal matrix W.

$$U^{-1}(H_0+H_1)U = W$$

or

$$(H_0+H_1)U = UW.$$

The unitary condition on U is of course

$$U^+U = 1.$$

We now try to solve these equations by an appxoximate method. If H_1 were zero clearly we would take $U=1$. Therefore it is natural to write for

$$U = 1 + U_1 + U_2 + \cdots + U_n + \cdots ,$$

where U_1 involves H_1 linearly

U_2 " H_1 quadratically, etc.

Similarly put

$$W = W_0 + W_1 + W_2 + \cdots + W_n + \cdots ,$$

W_0 being the diagonal matrix of the unperturbed levels, W_1 involving H_1 linearly, etc. Substituting we get

$$(H_0+H_1)(1+U_1+U_2+\cdots) = (1+U_1+U_2+\cdots)(W_0+W_1+W_2+\cdots)$$

$$(1+U_1^+ + U_2^+ + \cdots)(1+U_1+U_2+\cdots) = 1$$

Equating all the terms of the same order in H_1 we get

$$\left.\begin{aligned} H_0 &= W_0 \\ 1 &= 1 \end{aligned}\right\} \quad \text{zeroth order}$$

$$\left.\begin{aligned} H_1 + H_0U_1 &= W_1 + U_1W_0 \\ U_1^+ + U_1 &= 0 \end{aligned}\right\} \quad \text{first order}$$

$$\left.\begin{aligned} H_1U_1 + H_0U_2 &= W_2 + U_1W_1 + U_2W_0 \\ U_1^+U_1 + U_2^+ + U_2 &= 0 \end{aligned}\right\} \quad \text{second order}$$

$$\left.\begin{aligned} H_1 U_{n-1} + H_0 U_n &= \sum_{k=0}^{n} W_{n-k} U_k \\ \sum_{k=0}^{n} U_k^+ U_{n-k} &= 0 \end{aligned}\right\}$$

n^{th} order, with $U_0 \equiv 0$.

These equations can be solved successively. The zeroth order set is trivial, since H_0 is diagonal by choice of basis, and W_0 is just the diagonal matrix of the unperturbed levels. The first order set give

$$H_1 + W_0 U_1 = W_1 + U_1 W_0$$

If we take the diagonal matrix elements we get

$$(m|H_1|m) = (m|W_1|m) + (m|U_1 W_0 - W_0 U_1|m)$$

But $(m|U_1 W_0 - W_0 U_1|m) = (E_{0m} - E_{0m})(m|U_1|m) = 0$, where E_{0m} is the characteristic value of the m^{th} unperturbed level. Therefore $(m|W_1|m) = (m|H_1|m)$, which is the first order energy correction.

The nondiagonal elements give:

$$(m|H_1|n) = (m|U_1 W_0 - W_0 U_1|n) \qquad n \neq m$$

$$= (E_{0n} - E_{0m})(m|U_1|n) \qquad \text{or}$$

$$|U_1|n) = -\frac{(m|H_1|n)}{(E_{0m} - E_{0n})} \qquad n \neq m.$$

ary condition gives

$$|U_1^+|n) + (m|U_1|n) = 0$$

nondiagonal elements this is identically satisfied by the above.

diagonal elements

$$|U_1|m)^* + (m|U_1|m) = 0.$$

herefore we may take for the numbers $(m|U_1|m)$ <u>any</u> purely imaginary et of numbers (different choices correspond to different <u>phases</u> for the tate vectors). For simplicity we choose them zero, i.e. $(m|U_1|m) = 0$. hus we have a complete correction both to the energy and the state vector hich is accurate to the first order. For the second order we get, taking iagonal elements,

$$(m|H_1U_1|m) = (m|W_2|m) + (m|U_1W_1|m)$$

$$\sum_l (m|H_1|l)(l|U_1|m) = (m|W_2|m) + \underbrace{(m|U_1|m)}_{=0}(m|W_1|m), \quad \text{or}$$

$$(m|W_2|m) = -\sum_{l\neq m} \frac{(m|H_1|l)(l|H_1|m)}{E_{0l}-E_{0m}},$$

hich is the second order energy correction. Taking nondiagonal elements

$$(m|H_1U_1 - U_1W_1|n) = (m|U_2W_0 - W_0U_2|n) \qquad m \neq n$$

$$\sum_l (m|H_1|l)(l|U_1|n) - (m|U_1|n)(n|W_1|n) = (E_{0n}-E_{0m})(m|U_2|n)$$

$$(m|U_2|n) = \frac{1}{(E_{0n}-E_{0m})}\left\{-\sum_{l\neq n}\frac{(m|H_1|l)(l|H_1|n)}{E_{0l}-E_{0n}} + \frac{(m|H_1|n)(n|H_1|n)}{E_{0m}-E_{0n}}\right\}$$

$$= \frac{1}{E_{0m}-E_{0n}}\left\{\sum_{l\neq m,n}\frac{(m|H_1|l)(l|H_1|n)}{E_{0l}-E_{0n}} + \frac{(m|H_1|n)\left((m|H_1|m)-(n|H_1|n)\right)}{E_{0m}-E_{0n}}\right\}$$

$$m \neq n.$$

he unitary condition gives:

$$(m | U_2^{+} | n) + (m | U_2 | n) + \sum_{l} (m | U_1^{+} | l)(l | U_1 | n) = 0$$

r

$$n | U_2 | m)^{*} + (m | U_2 | n) + \sum_{l \neq m,n} \frac{(l | H_1 | m)^{*}(l | H_1 | n)}{(E_{0l} - E_{0m})(E_{0l} - E_{0n})} = 0.$$

he nondiagonal elements automatically satisfy this, as they must since hey are determined independently of it. For the diagonal elements we ave

$$m | U_2 | m)^{*} + (m | U_2 | m) = - \sum_{l \neq m} \frac{|(m | H_1 | l)|^{2}}{(E_{0l} - E_{0m})^{2}} \quad ,$$

hus only the real part of $(m | U_2 | m)$ is determined. We can hoose the imaginary part zero if we wish and then we have

$$(m | U_2 | m) = - \frac{1}{2} \sum_{l \neq m} \frac{|(m | H_1 | l)|^{2}}{(E_{0l} - E_{0m})^{2}} \quad ,$$

hich completes the calculation to the second order. Clearly we can terate the method and obtain--though the formulae get very complex--he correction to any order.

(2) Degenerate Perturbation Theory

Clearly the method we have used in the previous section breaks down if to two different states m and n have identical unperturbed energies, i.e., if $E_{0m} = E_{0n}$ since then some of the energy denominators vanish and the corrections come out infinite. In this case we must proceed differently. Since we can have different states with the same energy we need to label states with two indices, say, n, α. Different n mean different energy, different α mean different states with same energy. That is, the states are denoted by $\Psi^0_{n\alpha}$, with the corresponding energy being E_{0n}. We wish to solve

$$(H_0 + H_1)\Psi = E\Psi$$

As before this is equivalent to finding a unitary matrix U in the H_0 representation such that

$$(H_0 + H_1)U = UW,$$

where W is the diagonal matrix of the energy values of H. The unitary condition is of course $U^+ U = 1$ as before. Now we put

$$U = U_0 + U_1 + U_2 + \cdots$$

$$W = W_0 + W_1 + \cdots ,$$

and obtain

$$H_0 U_0 = U_0 W_0$$

$$H_1 U_0 + H_0 U_1 = U_0 W_1 + U_1 W_0 , \text{ etc} .$$

H_0 is just W_0, so our first equation is

$$W_0 U_0 = U_0 W_0 .$$

If W_0 were non-degenerate, the only solution of this equation would be $U_0 = 1$, which is what we used previously. However, if W_0 is degenerate this is not at all the case. Since

$$\sum_{l,\gamma} (m\alpha|W_0|l\gamma)(l\gamma|U_0|n\beta) = \sum_{l,\gamma} (m\alpha|U_0|l\gamma)(l\gamma|W_0|n\beta),$$

and $(m\alpha|W_0|l\gamma) = E_{0m}\,\delta_{ml}\,\delta_{\alpha\gamma}$, this reduces to

$$(E_{0m} - E_{0n})(m\alpha|U_0|n\beta) = 0 .$$

Since $E_{0m} \neq E_{0n}$, $(m\alpha|U_0|n\beta)=0$ for $m \neq n$, but the quantities $(m\alpha|U_0|m\beta)$ are completely (apart from the unitary condition) undetermined. We shall determine them in such a way as to eliminate the difficulties we had with vanishing energy denominators. Consider now the second equation,

$$H_1 U_0 + H_0 U_1 = U_0 W_1 + U_1 W_0 \quad \sim$$

$$H_1 U_0 - U_0 W_1 = U_1 W_0 - W_0 U_1 .$$

Taking matrix elements we get:

$$\sum_{\gamma} \{(m\alpha|H_1|n\gamma)(n\gamma|U_0|n\beta) - (m\alpha|U_0|m\gamma)(m\beta|W_1|m\beta)\,\delta_{\gamma\beta}\,\delta_{mn}\}$$

$$= (E_{0n} - E_{0m})(m\alpha|U_1|n\beta) .$$

The diagonal components of this (with respect to m, n) enable us to determine the first order energy corrections:

$$\sum_{\gamma} \{ (m\alpha|H_1|m\gamma) - \delta_{\alpha\gamma}(m\beta|W_1|m\beta)\} (m\gamma|U_0|m\beta) = 0 .$$

This gives us for each m a characteristic value problem with the submatrix $(m\alpha|H_1|m\gamma)$ of H_1 to be diagonalized. Solution gives us the matrix $(m\gamma|U_0|m\beta)$, and the d_m numbers (where d_m is the degeneracy of the mth level) $(m\beta|W_1|m\beta)$, which are the first order corrections to the d_m levels E_{0m}.

By taking the nondiagonal components we get

$$\sum_{\gamma} (m\alpha|H_1|n\gamma)(n\gamma|U_0|n\beta) = (E_{0n} - E_{0m})(m\alpha|U_1|n\beta) \qquad m \neq n$$

$$(m\alpha|U_1|n\beta) = -\sum_{\gamma} \frac{(m\alpha|H_1|n\gamma)(n\gamma|U_0|n\beta)}{E_{0m} - E_{0n}} \qquad m \neq n .$$

The unitary condition gives

$$U^{+}U = U_0^{+}U_0 + U_1^{+}U_0 + U_0^{+}U_1 + \cdots = 1$$

$$U_0^{+}U_0 = 1$$

$$U_1^{+}U_0 + U_0^{+}U_1 = 0$$

etc..

That is, the matrix U_0 must be chosen unitary, which is always possible.

The second equation gives

$$\sum_{l,\gamma} \{ (m\alpha|U_1^{+}|l\gamma)(l\gamma|U_0|n\beta) + (m\alpha|U_0^{+}|l\gamma)(l\gamma|U_1|n\beta) \} = 0$$

or,

$$\sum_{\gamma} \{ (m\alpha|U_1^{+}|n\gamma)(n\gamma|U_0|n\beta) + (m\alpha|U_0^{+}|m\gamma)(m\gamma|U_1|n\beta) \} = 0$$

Finally,

$$\sum \{ (m\gamma|U_0|n\beta)(n\gamma|U_1|m\alpha)^{*} + (m\gamma|U_0|m\alpha)^{*}(m\gamma|U_1|n\beta) \} = 0.$$

If we take $n \neq m$ we see that the above definition of U_1 satisfies this. For $m = n$, on the other hand, we get

$$\sum_{\gamma} \{ (m\gamma|U_0|m\beta)(m\gamma|U_1|m\alpha)^* + (m\gamma|U_0|m\alpha)^*(m\gamma|U_1|m\beta) \} = 0$$

All we need do is take $(m\gamma|U_1|m\alpha) \equiv 0$ to satisfy this.

It is a straightforward matter to push this approximation to the second order, but we shall not go into it here.

(3) Time dependent perturbation theory

Suppose we have a system with hamiltonian $H = H_0 + H_1$, At time t = 0 (say) the system is in some definite state Ψ_i of H_0, with energy E_i. That is, $H_0 \Psi_i = E_i \Psi_i$, where, for the purposes of this section, we shall drop the subscript o on the unperturbed energy and state. We can ask the question, what is the state at some later time t ? We shall work in the Schroedinger representation. Then, at any instant of time,

$$i\hbar \frac{\partial \Psi}{\partial t} = (H_0 + H_1)\Psi$$

It is convenient to make a transformation of basis. Put

$$\Phi = e^{iH_0t/\hbar}\,\Psi \quad \text{or}$$

$$\Psi = e^{-iH_0t/\hbar}\,\Phi .$$

Then

$$i\hbar \frac{\partial \Psi}{\partial t} = H_0 + i\hbar\, e^{-iH_0t/\hbar} \frac{\partial \Phi}{\partial t},$$

and therefore

$$i\hbar \frac{\partial \Phi}{\partial t} = \left(e^{\frac{iH_0t}{\hbar}} H_1\, e^{-\frac{iH_0t}{\hbar}} \right) \Phi .$$

This is often called the "interaction" representation, since the entire time dependence of depends on the size of the "interaction" H_1. If we assume that H_1 is in some sense small, we can hope to solve this equation by a perturbation method. Since $\Psi = \Psi_i$ at t = 0 we also have $\Phi = \Psi_i$ at t = 0. Put

$$\Phi = \sum_f (f|U|i)\, \Psi_f \quad ,$$

where the sum extends over all the states Ψ_f of the unperturbed system. $H_0 \Psi_f = E_f \Psi_f$.

The equation for U then becomes:

$$i\hbar \frac{d}{dt}(f|U|i) = \left(\Psi_f,\ e^{\frac{iH_0 t}{\hbar}} H_1 e^{-\frac{iH_0 t}{\hbar}} \sum_g (g|U|i)\, \Psi_g \right)$$

$$= \sum_g (f|H_1|g)(g|U|i)\, e^{i(E_f - E_g)t/\hbar}$$

Now put

$$(g|U|i) = \delta_{gi} + (g|U_1|i) + (g|U_2|i) + \cdots \quad ,$$

where U_1 is of the first order, U_2 of the second order, etc. in H_1. Equating orders on the right and left hand side will give a series of simple equations for determining U_1, U_2, etc.

$$i\hbar \frac{d}{dt}(f|U_1|i) = \sum_g (f|H_1|g)\, \delta_{gi}\, e^{i\frac{(E_f - E_g)t}{\hbar}}$$

$$i\hbar \frac{d}{dt}(f|U_2|i) = \sum_g (f|H_1|g)(g|U_1|i)\, e^{\frac{i(E_f - E_g)t}{\hbar}} \quad , \text{etc..}$$

Let us consider the first equation:

$$i\hbar \frac{d}{dt}(f|U_1|i) = (f|H_1|i)\, e^{i(E_f - E_i)t/\hbar}$$

$$(f|U_1|i) = \frac{1}{i\hbar} \int_0^t (f|H_1|i)\, e^{\frac{i(E_f - E_i)t}{\hbar}}\, dt \, ,$$

nce $U_1 = 0$ at $t = 0$. The requirement that Φ be normalized .e., that U be a unitary matrix) only requires U_1 to be anti-hermitian

$f|U_1|i) = -(i|U_1|f)^*$, which it clearly is. We can sily substitute this in the second equation and obtain results accurate the second order, but we shall not go into this here. If we are terested in the probability P_f of finding the system in the state f ter a time t we clearly need

$$P_f = |(f|U_1|i)|^2$$

consider two cases:

) H_1 independent of time.

) H_1 dependent on time.

(a) In this case we have

$$(f|U_1|i) = \frac{1}{i\hbar}(f|H_1|i)\frac{\hbar\left[e^{i(E_f-E_i)\frac{t}{\hbar}}-1\right]}{i(E_f-E_i)}$$

or

$$P_f = 2|(f|H_1|i)|^2\frac{1-\cos(\frac{E_f-E_i}{\hbar}t)}{(E_f-E_i)^2}.$$

w consider $w_{f\leftarrow i} \equiv \frac{\partial P_f}{\partial t}$. This is the probability per unit time a transition from i to f.

$$w_{f\leftarrow i} = \frac{2}{\hbar}|(f|H_1|i)|^2\frac{\sin\frac{E_f-E_i}{\hbar}t}{E_f-E_i}$$

v consider the situation after a very long time t. The time factor will general oscillate very much, sometimes being zero, but never having re than a limited value, unless $E_f \cong E_i$. Unless this is the se there will clearly be no permanent transitions to the state f. lytically we can formulate the situation as follows: We assert that

$$\lim_{t\to\infty} \frac{\sin \frac{(E_f - E_i)t}{\hbar}}{E_f - E_i} = \pi\,\delta(E_f - E_i)\,,$$ where E_f is in the continuum. (For E_f not in the continuum we have a problem in degenerate perturbation theory, which must be handled in a different fashion.) To see this, consider

$$\lim_{t\to\infty} \int F(E_f)\,\frac{\sin\left(\frac{E_f - E_i}{\hbar}\right)t}{E_f - E_i}\,dE_f = \lim_{t\to\infty} \int \frac{\sin x}{x}\,F\left(E_i + \frac{\hbar}{t}x\right)dx$$

$$= \int_{-\infty}^{\infty} \frac{\sin x}{x}\,F(E_i)\,dx = \pi\,F(E_i)\,, \text{ where } x = \frac{(E_f - E_i)t}{\hbar}\,.$$

This is, however, exactly equivalent to the above assertion. Thus if we wait long enough we will only find transitions to those states which conserve energy.

$$w_{f\leftarrow i} = \frac{2\pi}{\hbar}\,\left|(f|H_1|i)\right|^2\,\delta(E_f - E_i)\,.$$

It is usual to write the formula in a different way, since we are usually interested in the number of transitions <u>into</u> any of a whole class of final states. This would be, for example, if we were interested in all transitions which emit a particle into a certain solid angle with an energy between E_f and $E_f + d\,E_f$. Say the number of states having the desired property and with energy between E_f and $E_f + d\,E_f$ is $\rho_f(E_f)\,dE_f$. Then the total probability per unit time of achieving one of these states is

$$w_{f\leftarrow i} = \int dE_f\, \rho_f(E_f)\, \frac{2\pi}{\hbar}\, |(f|H_1|i)|^2\, \delta(E_f - E_i)$$

$$= \frac{2\pi}{\hbar}\, |(f|H_1|i)|^2\, \rho_f(E_i)\,.$$

A similar formula is easily obtained when the initial states of interest form a continuum, but we shall not go into this here.

We shall make use of this formula when we come to discuss the theory of scattering.

(b) If H_1 is a monatonic slowly varying function of time (slowly varying with respect to the time $\hbar/E_f - E_i$, then clearly the situation is unchanged from the time independent case, and one only gets transitions to states of the same energy. This is often called the adiabatic theorem. On the other hand if H_1 varies rapidly, the above results are seriously altered. One cannot discuss this problem in complete generality, but can only discuss special cases. One special case of considerable interest is that which obtains when H_1 depends periodically on the time, say with frequency ω. This would occur, for example, if an atom were placed in an oscillating electric field. Let us write

$$H_1 = V e^{-i\omega t} + V^{+} e^{+i\omega t},$$

as is necessary if H_1 is to be hermitian. Then

$$(f|V_1|i) = \frac{1}{i\hbar}\int_0^t \left\{ (f|V|i)\, e^{-i\omega t} + (f|V^{+}|i)\, e^{+i\omega t} \right\} e^{\frac{i(E_f - E_i)t}{\hbar}}\, dt$$

$$= \frac{\hbar}{(i\hbar)\, i} \left\{ \frac{(f|V|i)\left(e^{i\frac{E_f - E_i - \hbar\omega}{\hbar}t} - 1\right)}{E_f - E_i - \hbar\omega} + \frac{(f|V^{+}|i)\left(e^{i\frac{E_f - E_i + \hbar\omega}{\hbar}t} - 1\right)}{E_f - E_i + \hbar\omega} \right\}.$$

The same analyses we used in the time independent case can be applied to each of these terms separately, which means that we get two possible results. Either the first term is important $E_f = E_i + \hbar\omega$ or the second $E_f \hat{=} E_i - \hbar\omega$. The first case corresponds to absorption of energy by the system, the latter corresponds to induced emission.

The corresponding formulae are

$$w_{f\leftarrow i} = \frac{2\pi}{\hbar} |(f|V|i)|^2 \rho_f(\epsilon_i + \hbar\omega) \qquad \text{(Absorption)}$$

and

$$w_{f\leftarrow i} = \frac{2\pi}{\hbar} |(f|V^{+}|i)|^2 \rho_f(E_i - \hbar\omega) \qquad \text{(Induced Emmission)}.$$

B. The Variational Method.

Suppose we consider the expectation value of the energy of a system in any state Ψ which is not necessarily a stationary state. Allowing for the possibility of Ψ not being normalized, we have

$$\bar{H} = \frac{(\Psi, H\Psi)}{(\Psi, \Psi)}$$

Now consider the expansion of Ψ in states of H. Let $H\Psi_i = E_i\Psi_i$, and

$$\Psi = \sum_i \Psi_i (i|)$$

$$H\Psi = \sum \Psi_i E_i (i|)$$

$$(\Psi, H\Psi) = \sum_i E_i |(i|)|^2$$

$$(\Psi, \Psi) = \sum |(i|)|^2 .$$

Thus

$$\bar{H} = \frac{\sum_i E_i |(i|)|^2}{\sum |(i|)|^2} .$$

Now assume that the energy values E_i have been listed in order of their increase, i.e. $E_1 \leq E_2 \leq E_3 \cdots$, Then

$$\bar{H} = \frac{E_1 \sum_i |(i|)|^2 + (E_2 - E_1)|(2|)|^2 + (E_3 - E_1)|(3|)|^2 + \cdots}{\sum_i |(i|)|^2}$$

$$= E_1 + (E_2 - E_1)\frac{|(2|)|^2}{\sum |(i|)|^2} + \cdots ,$$

However, $E_2 - E_1$, $E_3 - E_1$, etc ≥ 0, and therefore

$$\bar{H} \geq E_1 .$$

Thus if there is a lowest energy state for a system, the expectation value of the energy for any state is always greater than or equal to it. Clearly if $\Psi = \Psi_1$, $\bar{H} = E_1$, so that the equality can be attained. Further, it is only attained if $\Psi = \Psi_1$, since from the above formula equality is present only if $(2|) = (3|) = \cdots = 0$ (no degeneracy).

Not only does $\bar{H}$ have this property for the lowest characteristic value, but it is actually stationary for the other characteristic values. Consider

$$\Psi = \Psi_i + \Phi$$

where $\bar{\Phi}$ is a vector having no component in the i direction. That is $(\Phi, \Psi_i) = 0$.

$$\bar{H} = \frac{(\Psi_i + \Phi,\ H(\Psi_i + \Phi))}{(\Psi_i + \Phi,\ \Psi_i + \Phi)}$$

$$= \frac{E_i(\Psi_i, \Psi_i) + (\Phi, H\Psi_i) + (\Psi_i, H\Phi) + (\Phi, H\Phi)}{(\Psi_i, \Psi_i) + (\Psi_i, \Phi) + (\Phi, \Psi_i) + (\Phi, \Phi)}$$

However using $(\Psi_i, \Phi) = (\Phi, \Psi_i) = 0$ and

$$(\Phi, H\Psi_i) = E_i(\Phi, \Psi_i) = 0$$

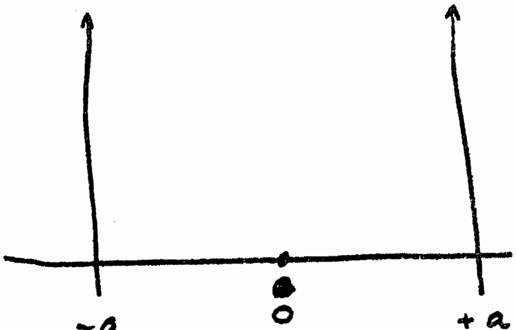
$$(\Psi_i, H\Phi) = E_i(\Psi_i, \Phi) = 0$$

we get

$$\bar{H} = \frac{E_i(\Psi_i, \Psi_i) + (\Phi, H\Phi)}{(\Psi_i, \Psi_i) + (\Phi, \Phi)}$$

This means that $\bar{H}$ differs from E_1 by something quadratic if Ψ differs from Ψ_i by something linear. This means in turn that a relatively poor guess for the state vector will give a much better answer for the energy. This is the practical use of the variational method. One guesses a Ψ which seems physically reasonable, and computes $\bar{H}$ with it. As a very simple example of the technique, consider the following problem:

A particle of mass m is places in a one dimensional box of length 2a, with infinitely high potential walls. What is the lowest possible energy value? Clearly the probability of finding the



particle out of the box is zero, and for the ground state it should be a maximum at the center. Assume the simplest curve which will fit the boundary conditions, a parabola. That is, put

$$(41) = \psi(x) = (x-a)(x+a) = x^2 - a^2 \qquad |x| \le a$$

$$= 0 \qquad |x| \ge a.$$

For $|x| \le a$, H (operating on the representative) is

$$\frac{p^2}{2m} \rightarrow \frac{-\hbar^2}{2m}\frac{d^2}{dx^2},$$ and therefore

$$\bar{H} = \frac{(\Psi, H\Psi)}{(\Psi,\Psi)} = \frac{\int_{-a}^{a} \psi \left(-\frac{\hbar^2}{2m}\right) \psi'' dx}{\int_{-a}^{a} \psi^2 dx}$$

$$= -\frac{\hbar^2}{2m}(2)\frac{\int_{-a}^{a}(x^2-a^2)dx}{\int_{-a}^{a}(x^2-a^2)^2 dx} = \frac{5}{2}\frac{\hbar^2}{2ma^2}$$

This problem can easily be solved exactly, and if we do we find for the lowest state

$$E = \left(\frac{\pi}{2}\right)^2 \frac{\hbar^2}{2ma^2}$$

$$\left(\frac{\pi}{2}\right)^2 = 2.4674$$

$$\frac{5}{2} = 2.500,$$

so that we see even with a very crude approximation we get an almost 1% accuracy.

It is sometimes possible to compute $\frac{1}{H}$, and when this is so the variational method can become a more systematic variational-iteration method. To see how this works, consider a state vector Ψ which gives us approximately the lowest energy value. Let us imagine Ψ expanded in stationary states of H:

$$\Psi = \sum_i \Psi_i (i|) .$$

Then

$$\frac{1}{H}\Psi = \sum_i \Psi_i \frac{1}{E_i}(i|) = \frac{1}{E_1}\left[\Psi_1 (1|) + \frac{E_1}{E_2}\Psi_2 (2|) + \cdots\right]$$

However $\frac{E_1}{E_2}, \frac{E_1}{E_3}, \ldots < 1$, so that the state $\frac{1}{H}\Psi$ contains less of states with higher energy than Ψ does, and therefore $\overline{H}$ for this state will be closer to E_1 than $\overline{H}$ for Ψ. By applying H^{-1} often enough, we can in principle come arbitrarily close to E_1. In practice it becomes very difficult to iterate more than once or twice.

A lower bound

Suppose λ is any real number. Consider the expression

$$\overline{(H-\lambda)^2} = \frac{(\Psi,(H-\lambda)^2\Psi)}{(\Psi,\Psi)} = \frac{\sum (E_i-\lambda)^2|(i|)|^2}{\sum |(i|)|^2} \geq (E_n-\lambda)^2 ,$$

where E_n is the characteristic value of H closest to λ. Now there are two cases to consider:

(1) $\lambda < E_n$, Then we obtain

$$\lambda < E_n \leq \lambda + \underset{+}{\sqrt{\overline{(H-\lambda)^2}}}$$

(2) $\lambda > E_n$. Then we obtain

$$\lambda > E_n \geq \lambda - \underset{+}{\sqrt{\overline{(H-\lambda)^2}}} .$$

For the ground state problem (1) is never interesting since $\lambda + \underset{+}{\sqrt{\overline{(H-\lambda)^2}}} \geq \overline{H}$, so that the simple variational principle $E_1 \leq \overline{H}$ is always sharper. On the other hand, if we know a $\lambda > E_1$, but closer to E_1 than to E_2, we may use (2) to obtain a lower bound for E_1. This will be the case if $E_1 < \lambda \leq \frac{E_1+E_2}{2}$. Choose

$$\lambda = \frac{E_1+E_2}{2}$$

Then

$$E_1 \geq \frac{E_1+E_2}{2} - \underset{+}{\sqrt{\overline{\left(H - \frac{E_1+E_2}{2}\right)^2}}}$$

$$\underset{+}{\sqrt{}} \geq \frac{E_2-E_1}{2}$$

$$\overline{\left(H - \frac{E_1+E_2}{2}\right)^2} \geq \left(\frac{E_2-E_1}{2}\right)^2 \quad \text{or,}$$

$$E_1 E_2 \geq \bar{H}(E_1+E_2) - \overline{H^2}$$

$$E_1 \geq \bar{H} - \frac{\overline{H^2} - \bar{H}^2}{E_2 - \bar{H}}$$

Thus if we know something about the position of the second level we can give a lower bound for E_1 by the variational method.

A simpler formula, though not as general a one, may be obtained as follows: <u>If</u> the trial vector Ψ builds a reasonably good approximation to Ψ_1, and we know (either experimentally or by other work) that $\bar{H}$ is closer to E_1 than to E_2, we may put $\lambda = \bar{H}$. Then we obtain

$$\bar{H} \geq E_1 \gtrsim \bar{H} - \sqrt{\overline{(H-\bar{H})^2}}$$

Write $\overline{(H-\bar{H})^2} = \overline{H^2} - \bar{H}^2 = (\Delta H)^2$, a measure of the "goodness" of Ψ as a characteristic vector of H. Then

$$\bar{H} \geq E_1 \gtrsim \bar{H} - \Delta H ,$$ for ΔH not too large.

C. The W.K.B. Method

This method is applicable to those problems which are either inherently one dimensional in nature or which by suitable treatment (such as the hydrogen atom problem) may be reduced to one dimensional problems. Let us imagine then that we have a hamiltonian

$$H = \frac{p_x^2}{2m} + V(x)$$

$V(x)$ is the effective protential of the problem, in the hydrogen case it would be

$$V(x) = -\frac{Ze^2}{x} + \frac{l(l+1)\hbar^2}{2m x^2} \quad , \qquad \text{for example}.$$

It is convenient to work in the coordinate representation. Put $(x|) = \psi(x)$, then the Schroedinger equation reads

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V(x)\psi = E\psi \qquad \text{or}, \qquad \text{or}$$

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E-V)\psi = 0.$$

Let us write $f(x) = \frac{2m}{\hbar^2}(E - V(x))$, so that the differential equation reads

$$\psi'' + f(x)\psi = 0.$$

Now if $V(x)$ is a "slowly" varying function--a notion which will be made precise as we go on--then so is $f(x)$. It is to this situation that the W.K.B. method applies. We may write

$$\psi = A e^{i\phi}$$

where A and ϕ are real. Then

$$\psi' = (A' + i\phi' A) e^{i\phi}$$

$$\psi'' = (A'' + 2i\phi' A' + A(i\phi'' - \phi'^2)) e^{i\phi}$$

The differential equation becomes:

$$A'' - A\phi'^2 + i(2\phi'A' + A\phi'') + fA = 0.$$

Equating real and imaginary parts we get

$$\left.\begin{aligned} A'' - A\phi'^2 + fA &= 0 \\ 2\phi'A' + A\phi'' &= 0 \end{aligned}\right\} \text{Exact}$$

Now if f were constant, A would be constant, so that a slowly varying f must give rise to a slowly varying A. Therefore we would expect that $|A''| << |fA|$. Assuming this, the first equation becomes

$$\phi'^2 = f$$

$$\phi' = \pm\sqrt{f}$$

$$\phi = \pm \int_{x_0}^{x} \sqrt{f}\, dx ,$$

where x_0 is some fixed value of x.

From the second equation we may at once calculate A.

$$\frac{2A'}{A} + \frac{\phi''}{\phi'} = 0$$

$$2 \log A + \log \phi' = \text{Constant}$$

$$A^2\phi' = C$$

$$A = \frac{c}{\sqrt{\phi'}} = \frac{c}{f^{1/4}} , \quad \psi = \frac{c}{f^{1/4}} e^{\pm i \int_{x_0}^{x} \sqrt{f}\, dx} .$$

In order for our approximations to be valid, we must have $|A''| << |Af|$. To see what this means, it is best to introduce the "de Broglie" wavelength λ of the particle

$$\lambda \equiv \frac{2\pi\hbar}{p_{\text{classical}}} = \frac{2\pi\hbar}{\sqrt{2m(E-V)}}.$$

Put $\bar{\lambda} = \frac{\lambda}{2\pi} = \frac{\hbar}{\sqrt{2m(E-V)}}$. We shall loosely refer to $\bar{\lambda}$ as the de Broglie wavelength also. We say that the condition for validity of the W.K.B. method is that the change in the amplitude A of ψ in one wavelength $\bar{\lambda}$ is small compared to A.

$$\frac{(\Delta A)_{\bar{\lambda}}}{A} \cong \frac{A'\bar{\lambda}}{A} << 1 , \qquad \frac{(\Delta A')_{\bar{\lambda}}}{A'} \cong \frac{A''\bar{\lambda}}{A'} << 1$$

For, if this is so, then

$$\frac{A''}{A} = \frac{A''}{A'}\,\frac{A'}{A} << \frac{1}{\bar{\lambda}^2} = f$$

Therefore the condition of validity of the W.K.B. approximation is that

$$\frac{A'\bar{\lambda}}{A} << 1 \qquad \sim \qquad \left| \cdot \frac{f'}{f^{3/2}} \right| << 1$$

or finally

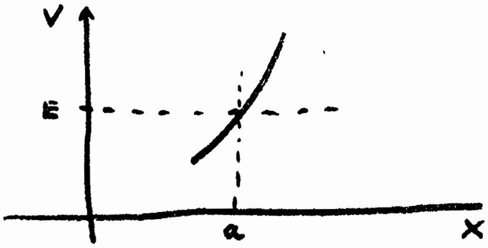
$$\left| \frac{\hbar \frac{\partial V}{\partial x}}{\sqrt{2m}\,(E - V(x))^{3/2}} \right| << 1$$

This is certainly true when $V(x)$ varies slowly enough and $E - V(x) \neq 0$. However, when $E - V = 0$ the method always breaks down. The W.K.B. method may never be used in the neighborhood of a classical turning point. We have assumed throughout the discussion so far that f was > 0. If $f < 0$ it is clear that we will have solutions of the form

$$\psi = \frac{C'}{(-f)^{1/4}}\, e^{\pm \int_{x_0}^{x} \sqrt{-f}\, dx}$$

Now in wave mechanics we need one solution which joins continuously to itself throughout the range of x, and therefore if the potential is such that there are regions of f greater than zero and less than zero we must find a way of joining our two types of solutions continuously to each other. It is clear that the W.K.B. method itself cannot be used, since it breaks down at a turning point, $f = 0$, where f changes sign. The treatment is straightforward but rather involved, and we omit it here. What is necessary is to find an approximate solution in the neighborhood of $f = 0$, and join it smoothly to the W.K.B. solution in the $f > 0$ region and $f < 0$ region, which gives a connection between their coefficients. We may, however, state the results very simply:

(1) Barrier to the right.



Let a be turning point, $f(a) = 0$,

such that $f(x) > 0 \quad x < a$

$f(x) < 0 \quad x > a$

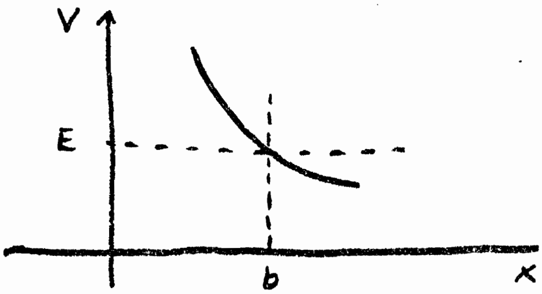
Then

$$\frac{2}{(f)^{1/4}}\cos\left[\int_x^a \sqrt{f}\,dx - \frac{\pi}{4}\right] \longleftrightarrow \frac{1}{(-f)^{1/4}}\, e^{-\int_a^x \sqrt{-f}\,dx}$$

$$\frac{1}{f^{1/4}}\sin\left[\int_x^a \sqrt{f}\,dx - \frac{\pi}{4}\right] \longleftrightarrow -\frac{1}{(-f)^{1/4}}\, e^{+\int_a^x \sqrt{-f}\,dx}$$

$x < a \qquad\qquad x > a$

(2) Barrier to the left.



Let b be a turning point, $f(b) = 0$,

such that $f(x) < 0 \quad x < b$

$f(x) > 0 \quad x > b$

Then

$$\frac{1}{(-f)^{1/4}}\, e^{-\int_x^b \sqrt{-f}\,dx} \longleftrightarrow \frac{2}{(f)^{1/4}}\cos\left[\int_b^x \sqrt{f}\,dx - \frac{\pi}{4}\right]$$

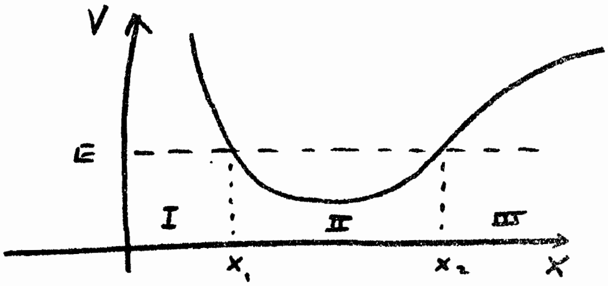
$$\frac{1}{(-f)^{1/4}}\, e^{+\int_x^b \sqrt{-f}\,dx} \longleftrightarrow -\frac{1}{\sqrt[4]{f}}\sin\left[\int_b^x \sqrt{f}\,dx - \frac{\pi}{4}\right]$$

$x > b$

These formulae may be used for many purposes, we consider two of the simplest.

(1) Approximate energy states

Suppose we have a situation in which the potential energy has a minimum, and ask what the allowed energy states are for a particle "trapped" in this well. We must clearly have



$$\psi_I = \frac{A}{(-f)^{1/4}} e^{-\int_x^{x_1}\sqrt{-f}\,dx},$$

the other solution increasing exponentially.

Using the connection formula we have (x_1 is a type b point)

$$\psi_{II} \to \frac{2A}{(f)^{1/4}} \cos\left[\int_{x_1}^{x}\sqrt{f}\,dx - \frac{\pi}{4}\right]$$

$$= \frac{2A}{(f)^{1/4}} \cos\left[-\int_{x}^{x_2}\sqrt{f}\,dx - \int_{x_2}^{x_1} - \frac{\pi}{4}\right], \text{ or}$$

$$= \frac{2A}{(f)^{1/4}} \cos\left[\int_{x}^{x_2}\sqrt{f}\,dx - \int_{x_1}^{x_2}\sqrt{f}\,dx + \frac{\pi}{4}\right].$$

Now this must go over into a decaying exponential in III, which we see from the connection formulae is only possible if

$$\psi_{II} \sim \frac{1}{f^{1/4}} \cos\left[\int_{x}^{x_2}\sqrt{f}\,dx - \frac{\pi}{4}\right],$$

since x_2 is a type a turning point. Therefore we must have

$$-\frac{\pi}{4} - \left(-\int_{x_1}^{x_2}\sqrt{f}\,dx + \frac{\pi}{4}\right) = n\pi \qquad n = 0, 1, \ldots$$

r $$\int_{x_1}^{x_2}\sqrt{f}\,dx = (n+\tfrac{1}{2})\pi\,,$$

This is a condition for obtaining the possible energy values. We may put it in a more familiar form, using the expression for the classical momentum p_c

$$\frac{p_c^2}{2m}+V=E\quad,\quad p_c=\sqrt{2m(E-V)}=\sqrt{f}\,\hbar$$

Thus

$$\int_{x_1}^{x_2} p_c\,dx = (n-\tfrac{1}{2})\hbar\pi = (n+\tfrac{1}{2})\frac{h}{2}$$

or

$2\int_{x_1}^{x_2} p_c\,dx =$ [redacted] $= (n+\tfrac{1}{2})h$

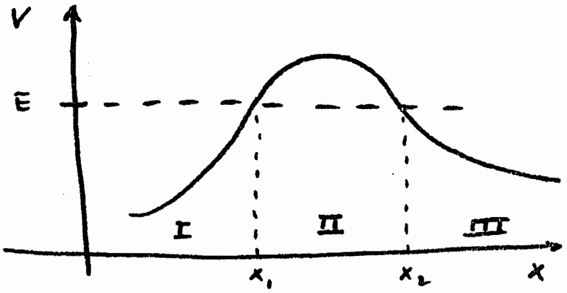
However, $2\int_{x_1}^{x_2} p_c\,dx = \oint p_c\,dx\,,$ the integral over one complete cycle of motion, so that

$$\oint p_c\,dx = (n+\tfrac{1}{2})h\,,$$

which are the famous Bohr-Sommerfeld quantum rules.

(2) Penetration through a barrier.

Suppose we send a stream of particles against a potential barrier. Classically all the particles would be reflected at x_1 It is well known however that this is not the case in quantum mechanics, and we shall now investigate this quantitatively by means of the W.K.B. approximation.



In general to the left of x_1 there will be transmitted and reflected waves. However for $x > x_2$ we get only transmitted waves, that is, only waves moving to the right. This means that for $x > x_2$ the solution must have the form

$$\psi_{III} = \frac{A}{(f)^{1/4}}\, e^{i\left(\int_{x_2}^{x}\sqrt{f}\,dx - \frac{\pi}{4}\right)}$$

the phase factor $-\pi/4$ being inserted to make application of the connection formulae simpler. In order to apply these formulae write

$$\psi_{III} = \frac{A}{(f)^{1/4}}\left\{\cos\left(\int_{x_2}^{x}\sqrt{f}\,dx - \frac{\pi}{4}\right) + i\sin\left(\int_{x_2}^{x}\sqrt{f}\,dx - \frac{\pi}{4}\right)\right\},$$

so that (since x_2 is a turning point of type b)

$$\psi_{II} = \frac{A}{(-f)^{1/4}}\left\{\frac{1}{2}e^{-\int_{x}^{x_2}\sqrt{-f}\,dx} - i\,e^{\int_{x}^{x_2}\sqrt{-f}\,dx}\right\}.$$

Let us write

$$\int_{x}^{x_2} = -\int_{x_1}^{x} - \int_{x_2}^{x_1} = -\int_{x_1}^{x} + G$$

$$G \equiv \int_{x_1}^{x_2}\sqrt{-f}\,dx$$

$$\psi_{II} = \frac{A}{(-f)^{1/4}}\left\{\frac{1}{2}e^{+\int_{x_1}^{x}\sqrt{-f}\,dx}\,e^{-G} - i\,e^{-\int_{x_1}^{x}\sqrt{-f}\,dx}\,e^{+G}\right\}.$$

To obtain the wave function in region I we must use the connection formulae for a type a turning point.

$$\psi_{I} = \frac{A}{(f)^{1/4}}\left\{-\frac{1}{2}e^{-G}\sin\left[\int_{x}^{x_1}\sqrt{f}\,dx - \frac{\pi}{4}\right] - 2i\,e^{+G}\cos\left[\int_{x}^{x_1}\sqrt{f}\,dx - \frac{\pi}{4}\right]\right\}$$

$$= \frac{-iA}{(f)^{1/4}}\left\{\left(e^{G} - \frac{1}{4}e^{-G}\right)e^{i\left(\int_{x}^{x_1}\sqrt{f}\,dx - \frac{\pi}{4}\right)} + \left(e^{G} + \frac{1}{4}e^{-G}\right)e^{-\left(\int_{x}^{x_1}\sqrt{f}\,dx - \frac{\pi}{4}\right)}\right\}.$$

The first term represents the reflected wave, the second the incident wave. By the transmission coefficient T we usually mean the ratio of transmitted particles / unit time to incident particles / unit time.

$$T = \frac{v_t \, |\psi_t|^2}{v_i \, |\psi_i|^2} = \frac{1}{(e^{G} + \frac{1}{4} e^{-G})^2} .$$

For a high, wide barrier $G >> 1$, and T $\approx e^{-2G}$, a widely used formula.

V. Spin and Identical Particles

Many experimental facts have led to the result that in order to describe the behavior of an electron one must introduce not only its space coordinates but also must assign to it an intrinsic angular momentum or spin. The most prominent of these effects are the gyromagnetic effect, the anomalous Zeeman effect, the existence of doublets in the alkali metal spectra and finally the Stern-Gerlach experiment. We shall only discuss the last one. Consider an atom, say hydrogen, which has for its ground state an S state, that is to say, a state with angular momentum $l = 0$. Then we know that since the degeneracy of a state is $2l+1$ that this state should be nondegenerate. However, if one passes a beam of such atoms through a homogeneous magnetic field followed by an inhomogeneous one, ~~then~~ one finds that the beam splits into two beams, indicating that the ground state is actually doubly degenerate. The interpretation of this, borne out by all the other experiments mentioned, is the following: The electron has an intrinsic angular momentum $j = 1/2$ associated with it. Since the degeneracy is $2j+1$ we get a degeneracy of $2(\frac{1}{2})+1 = 2$. which is the observed. Further, we must associate with the angular momentum a magnetic moment of magnitude $\frac{e\hbar}{2mc}$ (which is known as the Bohr magneton). Then a homogeneous magnetic field splits the two degenerate levels, adding to one $\frac{e\hbar}{2mc} H$ and subtracting from the other the same amount; this corresponding to spin being oriented parallel and antiparallel to the field respectively. The inhomogeneous field then will exercise a force on the magnetic moment, the sign of the force depending on how it is oriented. Thus we get the observed splitting of the beam. The ratio of magnetic moment in Bohr magnetons to the angular momentum of a system is called its gyromagnetic ratio. For electrons we see that for the spin the gyromagnetic ratio is 2.

It is an easy matter to show that for magnetic moments arising from orbital effects the ratio is one.

Since we then need this extra coordinate, the state of a system will have to have some spin variable as well as others to describe it. For the sake of definiteness, consider the coordinate representation, where x, y, z are diagonal. To specify the spin state of the system we need to give an additional quantum number which specifies the orientation of the spin. Since J_x, J_y, J_z do not commute, we can only specify one of these, and conventionally one takes J_z. J_z takes on the values $\pm \frac{1}{2} \hbar$. We shall write $J_z = S_z \hbar$, $S_z = \pm 1/2$. Then a wave function is a function of <u>four</u> variables

$$\psi(x, y, z, S_z)$$

where $S_z = \pm 1/2$. The interpretation of this wave function is the following

$$|\psi(x, y, z, \tfrac{1}{2})|^2 dx\, dy\, dz$$

is the probability of the electron being at the point x, y, z within the volume element $dx\, dy\, dz$, and with its angular momentum parallel to the direction z. $\psi(x, y, z, -1/2)$ is the same with the spin direction reversed. One very often combines these two functions into a column matrix ψ and writes

$$\psi = \begin{pmatrix} \psi(xyz\, 1/2) \\ \psi(xyz\, -1/2) \end{pmatrix}$$

If we do this then we may think of all operators in coordinate space as now represented by 2×2 matrices:

$$p_x^2 = \begin{pmatrix} \frac{\hbar^2}{i^2} \frac{\partial^2}{\partial x^2} & 0 \\ 0 & \frac{\hbar^2}{i^2} \frac{\partial^2}{\partial x^2} \end{pmatrix}, \text{ etc., for example.}$$

Operators involving the spin angular momentum on the other hand will not necessarily be multiples of the unit matrix [redacted] $(\vec{J}_{Spin} = \hbar \vec{S})$

$$S_x = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad S_y = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad S_z = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

as we saw in the section on angular momentum. An electron placed in a magnetic field (say in the z- direction) would then have a term $2\mu_B H S_z$ added to its hamiltonian. This type of formalism was first introduced by Pauli and Darwin. We often write this in a slightly different way. Define

$$\chi(\tfrac{1}{2}) = \begin{pmatrix}1\\0\end{pmatrix} \quad , \quad \chi(-\tfrac{1}{2}) = \begin{pmatrix}0\\1\end{pmatrix}.$$

Then

$$\psi = \psi(xyz,\tfrac{1}{2})\chi(\tfrac{1}{2}) + \psi(xyz,-\tfrac{1}{2})\chi(-\tfrac{1}{2}).$$

The quantities $\chi(\pm 1/2)$ are often called the spin functions of the electron, and we say that the most general wave function of a system is a linear combination--the coefficients depending on x, y, z alone--of the spin functions.

Pauli wave equation for an electron with spin

Suppose we consider for a moment an electron in an external electromagnetic field, and ignore spin. Then the hamiltonian, as we know from classical mechanics, is

$$\mathcal{H} = \frac{(\vec{p} + \frac{e}{c}\vec{A})^2}{2m} - e\phi$$

where $\vec{A}$ and ϕ are the scalar and vector potentials respectively. Now suppose we include spin. Spin gives rise to extra terms in the hamiltonian due to the intrinsic magnetic moment it gives the electron. Call this magnetic moment $-\vec{\mu}$, the minus since the electronic charge is negative. Then classically (neglecting terms in v/c) the extra energy is

$$\mathcal{H}' = \vec{\mu} \cdot \vec{H}$$

where $\vec{H}$ is the magnetic field. Now $\vec{\mu}$ has in general the magnitude μ_B and the direction of the spin, so we may write $\vec{\mu} = \mu_B \vec{\sigma}$ where $\vec{\sigma}$ is the matrix vector previously introduced:

$$\sigma_x = \begin{pmatrix}0 & 1\\1 & 0\end{pmatrix} \qquad \sigma_y = \begin{pmatrix}0 & -i\\i & 0\end{pmatrix} \qquad \sigma_z = \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}$$

$$\mathcal{H}' = \mu_B \vec{\sigma} \cdot \vec{H}.$$

The total hamiltonian becomes

$$\mathcal{H} = \frac{(\vec{p} + \frac{e}{c}\vec{A})^2}{2m} - e\phi + \mu_B \vec{\sigma} \cdot \vec{H} .$$

If we want an equation correct to terms in $\frac{v}{c}$ Thomas has shown that we must add the term

$$+ \frac{1}{2mc} \mu_B \left[\vec{\sigma} \cdot (\vec{p} + \frac{e}{c}\vec{A}) \times \vec{E} \right] .$$

This term consists of two contributions, one arising because of the fact that a particle moving in an electric field sees a magnetic field which is $O(\frac{v}{c})$ as large, and the second because another relativistic effect concerning a moving spin and known as the Thomas precession. We cannot go into the details of this here.

In the special case where there is only a central field present $\vec{A} = 0$, $\phi = \phi(r)$, the Thomas term becomes

$$-\frac{\mu_B}{2mc} \vec{\sigma} \cdot (\vec{p} \times \nabla\phi) = \frac{\mu_B}{2mc} \vec{\sigma} \cdot (\vec{p} \times \frac{\vec{r}\phi'}{r})$$

$\phi' = \frac{d\phi}{dr}$. Putting $\vec{S} = \frac{1}{2}\vec{\sigma}$, $\vec{L} = \vec{r} \times \vec{p}$

we obtain

$$- \frac{e\hbar}{2m^2c^2} \frac{1}{r} \frac{d\phi}{dr} \vec{L} \cdot \vec{S}$$

which is just [redacted] of the form of a spin-orbit coupling.

The operators in this hamiltonian are of course of the form $p_x \to \frac{\hbar}{i}\frac{\partial}{\partial x}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $A_1 = A_x(xyzt)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, etc. acting on the two component wave function.

The same formalism given here can be used to treat protons and neutrons, which also have an intrinsic spin of 1/2, but one must change the magnetic moment term which is not simply a magneton for these particles.

Identical Particles

By identical particles we mean particles which cannot be distinguished by any inherent property. That is, they have the same mass, charge, spin, etc. The quantum theory of systems of identical particles differs very widely from the classical theory for the following reasons. In classical mechanics each particle has a definite trajectory, and therefore one can in principle disginguish between particles which are identical in everything but path by simply following them along ~~these~~ their paths, being careful never to lose trace. In quantum mechanics, on the other hand, it is impossible exactly to localize particles, and thus when two identical particles are present it is impossible to say which of them is the one that a measurement has been made on. Thus identity or indistinguishability of particles is a much more significant thing in quantum mechanics.

Let us consider the Schrödinger equation for a system of n identical particles

$$i\hbar \frac{\partial \psi(1,2,\cdots,n)}{\partial t} = \mathcal{H}(1,2,\cdots,n)\,\psi(1,2,\cdots,n).$$

where each of the numbers represents all of the coordinates (positional and spin) of one of the particles. The hamiltonian H is symmetrical in its arguments, since the identity of the particles just means that they can be substituted for each other without changing H. Now consider an operator P which produces some permutation of the numbers 1,. . . . ,n. If we apply this permutation to the Schrödinger equation we obtain:

$$i\hbar \frac{\partial (P\psi(1,2,\cdots))}{\partial t} = P\left[H(1,2,3,\cdots)\,\psi(1,2,3,\cdots)\right]$$

$$= \mathcal{H}(1,2,\cdots)\,(P\psi(1,2,\cdots))$$

since H is symmetric. Thus if we have any solution of the Schrödinger equation $\psi(1,\cdots,n)$, we can obtain another by simply permuting the indices $(1\cdots n)$. In particular, if we are interested in the stationary states, then the Schrödinger equation reads

$$\mathcal{H}\psi = E\psi$$

and of course if ψ is a solution so is $P\psi$. Since for n particles there are in general $n!$ permutations, we see that for n identical particles there would appear to be an $n!$-fold degeneracy. Now the interesting, and at first sight rather surprising, result is that actually only one of these solutions--which one depending on the nature of the particles--occurs in nature.

Since any linear combination of solutions is a solution, the most general solution is of the form:

$$\psi = \sum_P a_P \, P\psi(1,2,\cdots),$$

where the sum is over all $n!$ permutations. Two of these are of particular interest.

(1) Symmetrical solutions. These are the solutions which are invariant under any permutation of the indices. Clearly they are obtained if we put $a_P = 1$

$$\psi_s = \sum_P P\psi(1,2,\cdots).$$

If we apply a permutation to this we get the same terms in different order.

(2) Anti-symmetric solutions. These are the solutions which change sign on having a pair of particles interchanged. If we write $a_P = \varepsilon_P$ where $\varepsilon_P = +1$ for an even permutation (a permutation which can be made up of an even number of interchanges of pairs) and $\varepsilon_P = -1$ for an odd permutation (a permutation made up of an odd number of pair interchanges) then

$$\psi_A = \sum_P \varepsilon_P \, P\psi(1,2,\cdots).$$

If we apply any interchange of a pair to this then $P_{pair}(P_{odd}) = P_{even}$

$$P_{pair}(P_{even}) = P_{odd}$$

so that we get the same thing with reversed sign.

In the case of two particles we would have, for example

$$\psi_S = \psi(1,2) + \psi(2,1)$$
$$\psi_A = \psi(1,2) - \psi(2,1),$$

while for three particles

$$\psi_S = \psi(1,2,3) + \psi(231) + \psi(312) + \psi(213) + \psi(132) + \psi(321)$$
$$\psi_A = \psi(123) + \psi(231) + \psi(312) - \psi(213) - \psi(132) - \psi(321),$$

as may be verified by direct calculation.

It is an experimental fact that only symmetric or antisymmetric wave functions occur in nature. Electrons, protons, neutrons are all found to have antisymmetric wave functions, systems of alpha particles, i.e. O^{16} nuclei, etc. to have symmetric ones. In the former case we often say that the particles are Fermi-Dirac, in the latter we call them Einstein-Bose. One can easily understand that the alpha particle (say) is an Einstein-Bose particle, since it is made up of two protons and two neutrons. Thus interchange of two α's is the interchange of two pairs of protons and two pairs of neutrons, and the sign of the wave function must be unchanged.

If the particles in question can be treated as each having its own hamiltonian and not interacting with each other, then the requirement of antisymmetry takes on a very simple form.

Write

$$\mathcal{H}(1,2,\cdots,n) = \mathcal{H}(1) + \mathcal{H}(2) + \cdots + \mathcal{H}(n),$$

and let the stationary states of $H(1)$ be

$$\mathcal{H}(1)\, v_\alpha(1) = E_\alpha\, v_\alpha(1).$$

Then the state $\psi(1,2,\cdots n) = v_\alpha(1)\, v_\beta(2) \cdots v_\nu(n)$ will be a stationary state with energy $E = E_\alpha + E_\beta + \cdots + E_\nu$. The corresponding antisymmetric function is

$$\psi_A = \sum_P \varepsilon_P P\, v_\alpha(1)\, v_\beta(2) \cdots v_\nu(n)$$

$$= \begin{vmatrix} v_\alpha(1) & v_\alpha(2) & \cdots & v_\alpha(n) \\ v_\beta(1) & v_\beta(2) & \cdots & v_\beta(n) \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ v_\nu(1) & v_\nu(2) & & v_\nu(n) \end{vmatrix} ,$$

by the definition of a determinant. We see at once that if any two states (say α and β) making up this state are the same, then two rows are equal. Two equal rows though mean that a determinant vanishes, and therefore it is impossible to have a state in which two particles are in the same one particle states. This is the original form of the exclusion principle of Pauli, which provided the key to the understanding of atomic structure.

VI. Scattering Problems

The problems which we consider in this section are of the following type. A beam of incident particles (electrons, neutrons, γ -rays, etc.) are incident on an ~~atoms~~ atom or nucleus or any other scattering center. We ask: what is the number of particles scattered out of the beam in a given direction, per unit time? Clearly this number will be proportional to the incident flux J (the number which arrive per unit area, per unit time) and to the size of the solid angle $\Delta\Omega$ about the direction θ, φ into which we consider the scattering as going. The proportionality factor is called the differential scattering cross section $\sigma(\theta, \varphi)$. In other words, the differential scattering cross section is the number of particles scattered in a given direction per unit solid angle, per unit incident flux. We also define a <u>total scattering cross section σ</u> as the total number of particles scattered per unit time (regardless of angle) per unit incident flux. Clearly

$$\sigma = \int \sigma(\theta,\varphi)\,d\Omega = \int_0^{2\pi} d\varphi \int_0^{\pi} d\theta\, \sigma(\theta,\varphi) \sin\theta .$$

We shall first consider the simplest situation, scattering by a central force which is fixed. We will characterize this interaction by a potential energy function $V(r)$. For electrons scattered on protons it would be

$$V(r) = -\frac{e^2}{r},$$

or for electrons on an atom it would in general be a screened coulomb potential

$$V(r) = \frac{-Ze^2}{r} e^{-qr}$$

In dealing with this problem it is usually most convenient to work in the representation with the coordinate diagonal, and in polar coordinates. Write $(xyz|) = \psi(r,\theta,\varphi)$. We may always write our solution as a sum of states of different angular momentum

$$\psi(r\theta\varphi) = \sum_{m,l} A_l \psi_l(r)\, \textcircled{H}_l^m(\theta,\varphi) = \sum_{m,l} A_l \frac{u_l}{kr} \textcircled{H}_l^m(\theta,\varphi).$$

Then from our previous work with the hydrogen atom we know that must satisfy (since they give the probability of being between r and $r+dr$

$$\left[-\frac{\hbar^2}{2M}\frac{d^2}{dr^2} + \frac{\hbar^2 l(l+1)}{2Mr^2} - V(r)\right] u_l = E u_l ,$$

where M is the mass of the particle. Introduce the notation

$$U(r) = \frac{2M}{\hbar^1} V(r) \qquad k^2 = \frac{2ME}{\hbar^2} .$$

Then this equation takes the form

$$\frac{d^2 u_l}{dr^2} + \left(k^2 - U(r) - \frac{l(l+1)}{r^2}\right) u_l = 0 .$$

Clearly k is real, since if a particle can come from ∞ its energy must be > 0 ($V \to 0$ as $r \to \infty$). In general, the function Θ_ℓ^m has the form

$$\Theta_\ell^m = P_\ell^m(\theta) \frac{e^{im\varphi}}{\sqrt{2\pi}}$$

From this we see at once that if we identify the polar axis (z-axis) with the direction of propagation of the incident beam, we can restrict ourselves to the m = o values alone. This is because for a symmetrical potential and propagation along the z axis no angle φ is distinguished, and therefore the wave function must be independent of φ. We need only take

$$\Theta_\ell^0 = \frac{1}{\sqrt{2\pi}} P_\ell^0 = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{2\ell+1}{2}} \frac{1}{2^\ell \ell!} \left(\frac{d}{d(\cos\theta)}\right)^\ell (\cos^2\theta - 1)$$

$$= C_\ell P_\ell(\cos\theta), \text{ where}$$

$$P_\ell(\cos\theta) = \frac{1}{2^\ell \ell!} \left(\frac{d}{d(\cos\theta)}\right)^\ell (\cos^2\theta - 1)^\ell$$

These polynomials are known as the Legendre Polynomials.

From the ortho-normalization of the Θ_ℓ^0 we obtain at once that

$$\int_0^\pi P_\ell(\cos\theta) P_{\ell'}(\cos\theta) \sin\theta \, d\theta = \frac{2}{2\ell+1} \delta_{\ell\ell'} .$$

The most general solution of our problem will then clearly be

$$\psi(r\theta\varphi) = \sum_{\ell=0}^{\infty} A_\ell \frac{u_\ell}{kr} P_\ell(\cos\theta)$$

where the A_l are still to be determined. Now if the incident wave is a plane wave moving in the $+Z$ direction we have $\psi_i = e^{ikz}$. The scattered wave we denote by ψ_s. We must have

$$\psi = \psi_i + \psi_s .$$

However, we know that the scattered wave must represent a radially outgoing wave in any direction θ. Therefore as $r \to \infty$

$$\psi_s \longrightarrow \frac{e^{ikr}}{r} f(\theta) .$$

In order to investigate the asymptotic behavior of the waves in question, let us look at differential equations for u_l. If $V \to 0$ more rapidly than $1/r$ (the coulomb field requires special investigation) then for very large r this equation becomes approximately

$$\frac{d^2 u_l}{dr^2} + k^2 u_l = 0 ,$$ so that

$$u_l \sim \sin(kr + \eta_l) \quad \text{as} \quad r \to \infty ,$$ where η_l is some constant determined by the condition that $u_l = 0$ at $r = 0$.

On the other hand, the plane incident wave may also be written

$$e^{ikz} = e^{ikr\cos\theta} = \sum C_l f_l(kr) P_l(\cos\theta) ,$$

the f_l corresponding to the u_l/kr for $V = 0$, the quantities C_l determined by the condition that these elementary solutions add up to a plane wave. Now it is actually very easy to obtain the f_l by solving

$$\frac{d^2(rf_l)}{dr^2} + \left(k^2 - \frac{l(l+1)}{r^2}\right)(rf_l) = 0 ,$$ and taking the solution which is ~~zero~~ finite at $r = 0$.

These are $f_l = \left(\frac{\pi}{2kr}\right)^{1/2} J_{l+1/2}(kr)$, $C_l = (2l+1)i^l$ where J_ν is the Bessel function of order ν. However, for our purposes these explicit solutions are unnecessary. We can conceive of the expression

$$e^{ikr\cos\theta} = \sum C_l f_l P_l(\cos\theta)$$

as an expansion of $e^{ikr\cos\theta}$ in Legendre polynomials. To find the coefficients we simply multiply by $\sin\theta\, P_{\ell'}(\cos\theta)$ and integrate:

$$c_\ell f_\ell \frac{2}{2\ell+1} = \int_0^\pi e^{ikr\cos\theta} P_\ell(\cos\theta)\sin\theta\, d\theta$$

$$= \int_{-1}^{1} e^{ikr\mu} P_\ell(\mu)$$

$$c_\ell f_\ell = \frac{2\ell+1}{2}\left\{ \frac{e^{ikr\mu}}{ikr} P_\ell(\mu)\Big]_{-1}^{1} - \frac{e^{ikr\mu}}{(ikr)^2} P_\ell'(\mu)\Big]_{-1}^{1} + \cdots \right\},$$

on repeated integration by parts. Clearly the first term is $O(\frac{1}{r})$, the second $O(\frac{1}{r^2})$, etc. Therefore asymptotically

$$c_\ell f_\ell \longrightarrow \frac{2\ell+1}{2} \frac{e^{ikr} - e^{-ikr}(-1)^\ell}{ikr},$$

since $P_\ell(1) = 1$ $P_\ell(-1) = (-1)^\ell$. We may write this

$$c_\ell f_\ell \longrightarrow (2\ell+1)\, i^\ell \frac{\sin(kr - \frac{1}{2}\ell\pi)}{kr}.$$

For purposes of comparison it is therefore convenient to write $\eta_\ell = -\frac{1}{2}\ell\pi + \delta_\ell$. The quantities $\delta_\ell = 0$ if there is no scattering potential present. Now the condition that we must have, is that

$$\psi_s = \psi - \psi_i \longrightarrow \frac{e^{ikr}}{r} f(\theta) \quad \text{as} \quad r \to \infty.$$

$$\psi - \psi_i \xrightarrow[r\to\infty]{} \sum_{\ell=0}^{\infty} \left\{ A_\ell \frac{\sin(kr - \frac{\ell\pi}{2} + \delta_\ell)}{kr} - (2\ell+1)\, i^\ell \frac{\sin(kr - \frac{\ell\pi}{2})}{kr} \right\} P_\ell(\cos\theta)$$

$$= \sum \frac{P_\ell(\cos\theta)}{2ikr} \Big\{ \left(A_\ell e^{i(\delta_\ell - \frac{\ell\pi}{2})} - (2\ell+1)\right) e^{ikr} - \left(A_\ell e^{-i(\delta_\ell - \frac{\ell\pi}{2})} - (2\ell+1) e^{i\pi\ell}\right) e^{-ikr} \Big\}.$$

The coefficient of e^{-ikr} must vanish, giving

$A_\ell = (2\ell+1) i^\ell e^{i\delta_\ell}$, which leaves for the scattered wave

$$\psi_s \to \frac{e^{ikr}}{r} \frac{1}{2ik} \sum_\ell (2\ell+1)(e^{2i\delta_\ell}-1) P_\ell(\cos\theta) .$$

This means that

$$f(\theta) = \frac{1}{2ik} \sum (2\ell+1)(e^{2i\delta_\ell}-1) P_\ell(\cos\theta).$$

As soon as we know $f(\theta)$ however, we know the differential scattering cross section. The incident wave e^{ikz} represents (since $|\psi_i|^2 = 1$) a beam with one particle/unit volume in it. The velocity of this beam is given by $\frac{1}{2}Mv^2 = E = \frac{\hbar^2}{2M}k^2$, so that $v = \frac{\hbar k}{M}$.

Thus the incident flux $J = (1)\frac{\hbar k}{M}$. On the other hand $|\psi_s|^2 = \frac{|f(\theta)|^2}{r^2}$ (as $r \to \infty$) is equal to the number of scattered particles per unit volume at a distance r from the origin. These particles are also moving with a velocity v, so that the number of them incident on an area A in a unit time is

$$\frac{v\,|f(\theta)|^2 A}{r^2}$$

$A = r^2 d\Omega$ however, for the number incident on an area which the solid angle $d\Omega$ intercepts, so that the number scattered into $d\Omega$ is simply

$$v\,|f(\theta)|^2\, d\Omega$$

and the incident flux is v, so that

$$\sigma(\theta,\varphi) = |f(\theta)|^2$$

The entire problem of calculating the scattering due to a center of force is just reduced to the problem of calculating the phase shifts δ_ℓ. In neral this can only be done numerically, but we shall come to speak of approximation method valid at high energies later.

Using the expression just obtained we may obtain a very simple xpression for the total cross section

$$\sigma = \int d\Omega\, |f(\theta)|^2 = \frac{1}{4k^2} \sum_{l,l'} (2l+1)(2l'+1)(e^{2i\delta_l} - 1)(e^{-2i\delta_{l'}} - 1) \int d\Omega\, P_l P_{l'} .$$

But $\int d\Omega\, P_l P_{l'} = \frac{4\pi}{2l+1} \delta_{ll'}$, and therefore

$$\sigma = \frac{4\pi}{k^2} \sum_{l=0}^{\infty} (2l+1) \sin^2 \delta_l .$$

Some remarks on the phase shifts δ_l .

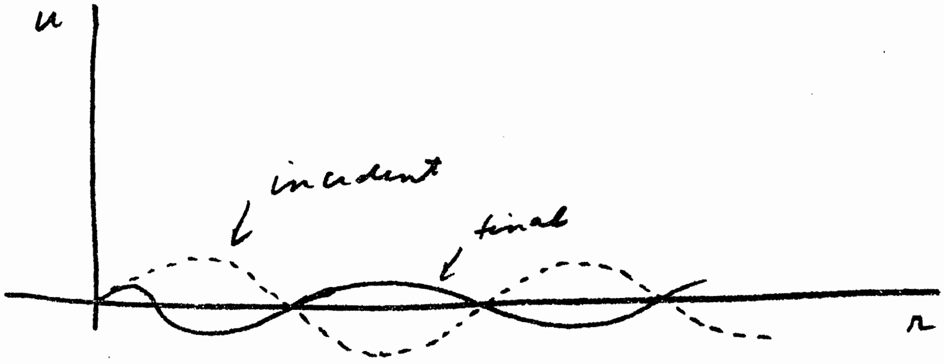
(1) The lower the energy, the less the highest l necessary o produce a good approximation is. This statement is true only if the otential is significantly different from zero in a region of radius R round the origin. Classically the argument goes as follows: Only hose particles which approach within a distance R of the nucleus will et scattered. That means that the highest angular momentum of a particle ith velocity v which gets scattered is

$$l_{max} = MvR = \hbar kR \quad , \text{ or } \quad l_{max} \sim kR .$$

hus for very low energy only the $l = 0$ scattering will be important, nd since $P_0(\cos\theta) = 1$, the scattering at low energies will be pherically symmetric, the criterion being $k \ll \frac{1}{R}$, or the avelength $\lambda \gg R$.

(2) An attractive potential $V(r)$ tends to bring the wave unctions in closer to the origin, and therefore δ_l will be positive (same value of $\sin(kr - \frac{l\pi}{2} + \delta_l)$ will occur for smaller r) while repulsive potential will push the wave function out and give a negative δ_l.

This provides a very simple qualitative description of the Ramsauer-Townsend effect. If the energy is low enough we need consider only $l = 0$. on the other hand, if the attraction is strong enough we can



pull the wave function so far in that $\delta_0 = 180°$ (see figure). Then $\sin^2\delta_0 = 0$ and $\sigma \sim 0,$ perfect transmission. This is actually found for electrons on rare gas atoms, and is known as the Ramsauer-Townsend effect. For a repulsive potential it is easy to see that the wave function can never be pushed out so far that $\delta_0 = -180°$.

The Born Approximation

We shall now consider a method of approximation which is valid when the incident particle has very high energy. Suppose the hamiltonian of the incident particle is $\mathcal{H}_0 = \frac{p^2}{2M}$. Let the energy of interaction with the scatterer be H_1 (for a center of force this will be $V(r)$). Now the incident wave is a particle with momentum $\vec{p}_i$ and energy $E = p_i^2/2M$, while the final particle after the scattering has momentum $\vec{p}_f$ and of course the same energy. We can ask, what is the probability per unit time of a particle appearing in a state $\vec{p}_f$ for a given incident flux? Now if the incident particles are very energetic, they will in general not be deviated very much, and therefore we may treat H_1 as a perturbation. This is known as the Born approximation. If we do that, however, we can make use of our time dependent perturbation theory to give us the answer at once.

$$w_{f\leftarrow i} = \frac{2\pi}{\hbar} \left|(f|H_1|i)\right|^2 \rho_f(E_i)$$

The calculation of the matrix element of H_1 is straightforward, but the density of states requires some care. To see this most clearly, let us take instead of an infinite space, a very large cubic volume of side L $(V = L^3)$. We can take as boundary conditions on $(xyz|) = \psi$ that it be periodic of period L (in x, y or z), it being clear that since $V \to \infty$ just what boundary conditions we use are unimportant. If we do this, then for the initial wave function we take

$$\psi_i = \frac{1}{\sqrt{V}} e^{i(\vec{p}_i\cdot\vec{r})/\hbar},$$

a state with definite momentum $\vec{p}_i$, and <u>one</u> particle present, since $\int dx\,dy\,dz\,|\psi_i|^2 = 1$. The momenta $\vec{p}_i$ are now restricted by the boundary condition

$$\psi_i(x+L, y, z) = \psi_i(x, y, z), \text{ etc.},$$ etc. which give

$$e^{ip_{ix}L/\hbar} = 1 \qquad p_{ix} = \frac{2\pi\hbar}{L} n_1,$$ where

$$n_1 = 0, \pm 1, \pm 2, \text{ etc.}$$

Similarly

$$p_{iy} = \frac{2\pi\hbar}{L} n_2$$

$$p_{iz} = \frac{2\pi\hbar}{L} n_3 .$$

Clearly as $L \to \infty$ these approach the continuum of possible values, as they must. The final state must also be a state of this type

$$\psi_f = \frac{1}{\sqrt{V}} e^{i\vec{p}_f\cdot\vec{r}/\hbar} \qquad p_{fx} = \frac{2\pi\hbar}{L} m_1, \text{ etc.}$$

Now we have discrete states, and we can ask how many are there with an energy between E and $E + \Delta E$, and with $\vec{p}$ in a given solid ngle $\Delta\Omega$. Certainly the number with p_x between p_x and $p_x + \Delta p_x$, p_y between p_y and $p_y + \Delta p_y$, p_z between p_z and $p_z + \Delta p_z$ s given by

$$\left(\frac{L}{2\pi\hbar}\Delta p_x\right)\left(\frac{L}{2\pi\hbar}\Delta p_y\right)\left(\frac{L}{2\pi\hbar}\Delta p_z\right),$$

ince $\frac{L}{2\pi\hbar}\Delta p_x$ tells us by how much we must increase n_1 to get change Δp_x. This expression becomes

$$\frac{L^3}{(2\pi\hbar)^3}\Delta p_x\,\Delta p_y\,\Delta p_z \longrightarrow \frac{V}{(2\pi\hbar)^3}dp_x\,dp_y\,dp_z$$

s we make the volume larger and Δp infinitesimal and equal dp. ntroducing polar coordinates we have $dp_x\,dp_y\,dp_z = p^2\,dp\,d\Omega$, so that the density of tates becomes:

$$\rho_f(E) = \frac{V}{(2\pi\hbar)^3}p^2\,dp\,d\Omega\Big/dE$$

$$= \frac{VMp\,d\Omega}{(2\pi\hbar)^3}$$

$p = \sqrt{2ME}$

$p^2dp = Mp\,dE$

$p = p_i \sim p_f$.

On the other hand, the matrix element is

$$(f|H_1|i) = \int \frac{e^{-i\vec{p}_f\cdot\vec{r}/\hbar}}{\sqrt{V}}H_1\frac{e^{i\vec{p}_i\cdot\vec{r}/\hbar}}{\sqrt{V}}\,dx\,dy\,dz$$

$$= \frac{1}{V}\int e^{-i\vec{p}_f\cdot\vec{r}/\hbar}H_1\,e^{+i\vec{p}_i\cdot\vec{r}}\,dx\,dy\,dz$$

$$\equiv \frac{1}{V}H_1(f\leftarrow i)$$

$$w_{f\leftarrow i} = \frac{2\pi}{\hbar}\,\frac{1}{V^2}\,|H_1(f\leftarrow i)|^2\,\frac{VMp}{(2\pi\hbar)^3}\,d\Omega$$

$$= \frac{1}{(2\pi)^2\hbar^4}\,\frac{1}{V}\,Mp\,|H_1(f\leftarrow i)|^2\,d\Omega.$$

is the number of transitions going into $d\Omega$ per unit time, so that the number per unit solid angle $W_{f\leftarrow i}$ is

$$W_{f\leftarrow i} = \frac{1}{(2\pi)^4\hbar^4}\,\frac{1}{V}\,Mp\,|H_1(f\leftarrow i)|^2.$$

Now the differential scattering cross section is given by

$$\sigma(\theta,\varphi) = \frac{W_{f\leftarrow i}}{J}$$

where J is the incident flux. As before

$$J = \text{(incident particle density)(incident particle velocity)}$$

$$= \frac{1}{V}\sqrt{\frac{2E}{M}} = \frac{1}{V}\,\frac{p}{M}$$

Thus

$$\sigma(\theta,\varphi) = \frac{1}{(2\pi)^2\hbar^4}\,M^2\,|H_1(f\leftarrow i)|^2$$

In this final formula the volume V dropped out, as it of course must. This is the Born approximation expression for the scattering cross section. If H_1 is a simple potential $V(xyz)$ we have

$$H_1(f\leftarrow i) = \int e^{i(\vec{p}_i-\vec{p}_f)\cdot\vec{r}/\hbar}\,V(xyz)\,dx\,dy\,dz.$$

This is just the $\vec{p}_i-\vec{p}_f/\hbar$ fourier component of $V(xyz)$ so that in the Born approximation the scattering cross section gives us direct information on the fourier transform of the potential. If the potential is a function of r alone $V(xyz) = V(r)$, we can simplify this further

$$H_1(f \leftarrow i) = \int e^{iKr\cos\vartheta} V(r)\, r^2 dr \sin\vartheta \, d\vartheta \, d\phi$$

where $\vec{K}$ is the vector $\vec{p}_i - \vec{p}_f / \hbar$ and ϑ is the angle between $\vec{K}$ and $\vec{r}$. Integration over ϑ and ϕ give

$$H_1(f \leftarrow i) = 4\pi \int_0^\infty V(r) \frac{\sin Kr}{Kr} r^2 dr,$$

so that the expression for $\sigma(\theta, \varphi)$ becomes

$$\sigma(\theta, \varphi) = \left(\frac{2M}{\hbar^2}\right)^2 \left[\int_0^\infty V(r) \frac{\sin Kr}{Kr} r^2 dr\right]^2 .$$

The quantity K is related to the scattering angle and incident momentum as follows:

$$K^2 = \frac{(\vec{p}_i - \vec{p}_f)^2}{\hbar^2} = \frac{p_i^2 + p_f^2 - 2 p_i p_f \cos\theta}{\hbar^2},$$

where Θ is the scattering angle, that is, the angle between $\vec{p}_i$ and $\vec{p}_f$. However $p_i = p_f = p$, so that

$$K^2 = \frac{2p^2(1-\cos\theta)}{\hbar^2} = \frac{4p^2 \sin^2\frac{\theta}{2}}{\hbar^2}$$

$$K = \frac{2p \sin\frac{\theta}{2}}{\hbar} .$$

s a very simple example of the method, let us consider the case of a hielded coulomb potential

$$V(r) = \frac{Ze^2}{r} e^{-qr}.$$

Then $$\int_0^\infty V(r) \frac{\sin Kr}{Kr} r^2 \, dr = \frac{Ze^2}{K} \int_0^\infty e^{-qr} \sin Kr \, dr$$

$$= \frac{Ze^2}{K^2+q^2}, \text{ so that}$$

$$\sigma(\theta, \varphi) = \left(\frac{2M}{\hbar^2}\right)^2 \frac{Z^2 e^4}{\left[\frac{4p^2 \sin^2\frac{\theta}{2}}{\hbar^2} + q^2\right]^2}$$

f we let $q \rightarrow 0$ (no shielding), we get the scattering in a oulombian field:

$$\sigma(\theta, \varphi) = \left(\frac{MZe^2}{2p^2}\right)^2 \frac{1}{\sin^4\frac{\theta}{2}},$$

which happens to gree with the exact Rutherford scattering formula.

If the potential is not so strongly repulsive that there would e classically a reflection of the particle, one can give a simple and ough criterion for the validity of the Born approximation. The Born pproximation should be valid if the phases δ_l are not too great. ow we know that if the potential is not too rapidly varying then we can ssociate with a particle the wavelength $\bar{\lambda} = \hbar/p_{class}$ or

$$\frac{1}{\bar{\lambda}} = \sqrt{\frac{2M}{\hbar^2}}\sqrt{E-V},$$

if V is the effective potential f the particle. This means that the phase change in going from ∞ to he origin is $\delta' = \int_0^\infty \sqrt{\frac{2M}{\hbar^2}}\sqrt{E-V}\, dr$, which for a free particle it would be $\delta'' = \int_0^\infty \sqrt{\frac{2M}{\hbar^2}}\sqrt{E}\,dr$. The phase shift as a result of the potential is therefore

$$\delta = \delta' - \delta'' = \int_0^\infty \sqrt{\frac{2M}{\hbar^2}}\left\{\sqrt{E-V} - \sqrt{E}\right\}dr .$$

The condition for validity of the Born approximation is therefore, very roughly,

$$\int_0^\infty \sqrt{\frac{2M}{\hbar^2}}\left\{\sqrt{E-V} - \sqrt{E}\right\}dr \ll 1$$

As an example of the use of this, consider the potential $V(r) = -\frac{\alpha}{r^n}$. Then

$$\delta = \sqrt{\frac{2M}{\hbar^2}}\int_0^\infty \left\{\sqrt{E+\frac{\alpha}{r^n}} - \sqrt{E}\right\}dr .$$

$$= \sqrt{\frac{2M}{\hbar^2}}\sqrt{E}\left(\frac{\alpha}{E}\right)^{\frac{1}{n}}\int_0^\infty dx\left\{\sqrt{1+\frac{1}{x^n}} - 1\right\} \qquad r^n = \alpha/E\, x^n$$

The integral is a dimensionless pure number and generally of the order of unity. Therefore $\delta \approx \sqrt{\frac{2M}{\hbar^2}}\sqrt{E}\left(\frac{\alpha}{E}\right)^{1/2} \ll 1$ as the criterion for the Born Approximation. Strictly speaking this integral diverges for $2 \leqslant n \leq 1$, (The latter because plane waves are not the valid solutions at infinity.) However, we can obtain an estimate simply by putting $n = 1$ and $\alpha = Ze^2$ for how good the Born approximation is in the coulomb case:

$$\delta \approx \sqrt{\frac{2M}{\hbar^2}}\;\frac{Ze^2}{\sqrt{E}} \ll 1$$

or

$$\frac{2Ze^2}{\hbar v} \ll 1,$$

Another criterion which is often used is the following:

$$\left(\frac{M}{\hbar p}\right)^2 \left| \int_0^\infty V(r)\,(e^{2ikr}-1)\,dr \right|^2 \ll 1,$$

p is the momentum of the particle, $k = p/\hbar$. This is usually obtained by studying the change in the wave function near the origin, and requiring that it be small. As an example of its use consider again the coulomb potential, then

$$\left(\frac{M}{\hbar p}\right)^2 (Ze^2)^2 \left| \int_0^\infty \frac{1}{r}(e^{2ikr}-1)\,dr \right|^2 \ll 1 .$$ Put $2kr = x$

or

$$\left(\frac{M}{\hbar p}\, Ze^2\right)^2 \left| \int_0^\infty \frac{1}{x}(e^{ix}-1)\,dx \right|^2 \ll 1$$

The integral is once more a pure number, and roughly unity, so that we have

$$\left(\frac{M}{\hbar p}\, Ze^2\right)^2 \ll 1$$

or $\frac{Ze^2 M}{\hbar p} \ll 1$ or $\frac{Ze^2}{\hbar v} \ll 1$, which is the same as our previous result.

Coulomb Scattering

We now consider the exact treatment of scattering in a coulomb field. If we consider the initial beam as moving along the z axis, then the problem has cylindrical symmetry, and the wave function cannot depend on φ. Since r, θ may be expressed in terms of r and z we may write $\psi = \psi(r, z)$. It turns out to be convenient to use so-called parabolic coordinates ξ and η: $\xi \equiv r - z$, $\eta \equiv r + z$. The Schrödinger equation has the form

$$-\frac{\hbar^2}{2m}\nabla^2\psi + \frac{\gamma}{r}\psi = E\psi$$

$\gamma = Z_1 Z_2 e^2$, Z_1, Z_2 being the charge (in units of e) on the scattered particle and scatterer respectively. Using

$$\nabla^2\psi = \frac{4}{\xi+\eta}\left[\frac{\partial}{\partial\xi}\left(\xi\frac{\partial\psi}{\partial\xi}\right) + \frac{\partial}{\partial\eta}\left(\eta\frac{\partial\psi}{\partial\eta}\right) + \frac{1}{\xi\eta}\frac{\partial^2\psi}{\partial\varphi^2}\right],$$

$$r = \frac{\xi+\eta}{2}$$

we obtain

$$\frac{\partial}{\partial\xi}\left(\xi\frac{\partial\psi}{\partial\xi}\right) + \frac{\partial}{\partial\eta}\left(\eta\frac{\partial\psi}{\partial\eta}\right) + \left(\frac{\xi+\eta}{4}\frac{2m}{\hbar^2}E - \frac{m\gamma}{\hbar^2}\right)\psi = 0 .$$

We now assert that the solution of the scattering problem may be written in the form

$$\psi = e^{ikz}f(\xi) = e^{ik\frac{(\eta-\xi)}{2}}f(\xi) , \quad E = \frac{\hbar^2 k^2}{2m} ,$$

since we shall show that with solution we can satisfy the differential equation and the boundary conditions at the origin and at infinity. Inserting this solution we obtain

$$\xi f'' + (1 - ik\xi)f' - nkf = 0 ,$$

where

$$n = \frac{\gamma m}{\hbar^2 k} = \frac{\gamma}{\hbar v} ,$$

v being the velocity of the incoming particle. Now this equation is exactly of the form of the confluent hypergeometric equation

$$zF'' + (b - z)F' - aF = 0$$

the only solution of which that is finite at the origin being

$$F = F(a,b,z) = \sum_{s=0}^{\infty}\frac{\Gamma(a+s)\,\Gamma(b)}{\Gamma(b+s)\,\Gamma(a)}\frac{z^s}{s!}$$

Γ being the well known gamma function. Thus

$$f(\xi) = C\, F(-in, 1, ik\xi)$$

We must show that this has the proper asymptotic behavior for large r that is, for large ξ (except right on the z axis).

$$F(-in,1,ik\xi) \xrightarrow[\xi\to\infty]{} \frac{e^{n\pi/2}}{\Gamma(1+in)}\left\{ e^{-in\log k\xi}\left[1-\frac{n^2}{2k\xi}\right] + \frac{f_c(\theta)}{r}\, e^{i(kr - n\log(2kr))}\right\},$$

where

$$f_c(\theta) = n\,\frac{\Gamma(1+in)}{i\,\Gamma(1-in)}\;\frac{e^{-in\log\sin\frac{\theta}{2}}}{2k\sin^2\frac{\theta}{2}} .$$

In terms of ψ this gives

$$\psi \xrightarrow[r\to\infty]{} \frac{C\,e^{\frac{n\pi}{2}}}{\Gamma(1+in)}\left\{ e^{i(kz - n\log k(r-z))}\left[1-\frac{n^2}{2k(r-z)}\right] + \frac{f_c(\theta)}{r}\, e^{i(kr - n\log(2kr))}\right\}.$$

Thus we see that the coulomb wave function does not approach asymptotically the form

$$\psi \xrightarrow[r\to\infty]{} e^{ikz} + \frac{f(\theta)}{r}\, e^{ikr},$$

but there is a modification of each term by a factor. This simply means that the coulomb force is so long ranged that no matter how far out you go the wave is still distorted. It is easily shown that no physically observable effects, such as the current, are effected by the presence of these extra factors, and so we can ignore them in calculating cross sections. The incident current then has the magnitude

$$j_i = \frac{\hbar k}{m}\,\frac{|C|^2}{|\Gamma(1+in)|^2}$$

and the outgoing current per unit solid angle is

$$j_0 = \frac{\hbar k}{m} \frac{|C|^2 |f_c(\theta)|^2}{|\Gamma(1+in)|^2}$$

Thus the differential scattering cross section is

$$\sigma(\theta) = |f_c(\theta)|^2 = \frac{n^2}{\left(2k \sin^2\frac{\theta}{2}\right)^2} = \left(\frac{Z_1 Z_2 e^2}{2mv^2}\right)^2 \frac{1}{\sin^4\frac{\theta}{2}},$$

which is just the Rutherford formula.

Exchange effects in scattering

If we scatter two identical particles against each other then there will arise effects--due to the symmetry character of the wave function--which have no analog in classical physics. Let us consider, for example, the scattering of alpha particles by alpha particles. Here the wave function must be symmetric. Let $\chi(\vec{r}_1, \vec{r}_2)$ be any solution to the problem. Then the symmetrized solution will be

$$\psi = \chi(\vec{r}_1, \vec{r}_2) + \chi(\vec{r}_2, \vec{r}_1).$$

If we introduce relative coordinates and the coordinate of the center of mass:

$$\vec{r} = \vec{r}_1 - \vec{r}_2$$
$$R = \frac{\vec{r}_1 + \vec{r}_2}{2},$$

this function takes the form

$$\psi = e^{i\vec{P}\cdot\vec{R}/\hbar}\left[f(\vec{r}) + f(-\vec{r})\right]$$

where $\vec{P}$ is the momentum of the center of gravity and $f(\vec{r})$ is the solution of the Schrödinger equation corresponding to the "reduced mass" $\mu\left(=\frac{m}{2}\right)$ and the energy left over when the center of mass energy is subtracted off. For the center of mass at rest we get

$$\psi = f(\vec{r}) + f(-\vec{r}).$$

Now asymptotically

$$f(\vec{r}) \xrightarrow[r\to\infty]{} e^{ikz} + \frac{f(\theta)}{r} e^{ikr}$$ for the

scattering problem. Thus

$$\psi \xrightarrow[r\to\infty]{} e^{ikz} + e^{-ikz} + \frac{f(\theta)+f(\pi-\theta)}{r} e^{ikr} .$$

We do not divide by $\sqrt{2}$ when constructing the symmetric function, since we want two particles to be present in the original state. Thus the differential scattering cross section for symmetric states is

$$\sigma_s(\theta) = |f(\theta)+f(\pi-\theta)|^2$$

The same analysis for states which are spacially antisymmetric gives

$$\sigma_A(\theta) = |f(\theta)-f(\pi-\theta)|^2$$

For electrons which are unpolarized being scattered by electrons which are unpolarized we have 3 states with parallel spin (spin symmetric, therefore spacially antisymmetric) to each state with antiparallel spin (spin antisymmetric, therefore space symmetric). The scattering cross is thus

$$\sigma(\theta) = \frac{3}{4}|f(\theta)-f(\pi-\theta)|^2 + \frac{1}{4}|f(\theta)+f(\pi-\theta)|^2$$

$$= |f(\theta)|^2 + |f(\pi-\theta)|^2 - \frac{1}{2}\left(f(\theta)f^*(\pi-\theta) + f^*(\theta)f(\pi-\theta)\right) .$$

The first two terms are just the classical result, the last is the effect of exchange. If we apply this formula to coulomb scattering of two electrons we obtain

$$\sigma(\theta) = \left(\frac{Z_1 Z_2 e^2}{2\mu v^2}\right)^2 \left\{ \frac{1}{\sin^4\frac{\theta}{4}} + \frac{1}{\cos^4\frac{\theta}{2}} - \frac{\cos\left(n \log \tan^2\frac{\theta}{2}\right)}{\sin^2\frac{\theta}{2}\cos^2\frac{\theta}{2}} \right\}$$

$$E = \frac{\hbar^2 k^2}{2\mu}, \quad n = \frac{Z_1 Z_2 c^2 \mu}{\hbar^2 k}, \quad v = \frac{\hbar k}{\mu} .$$

which is the famous Mott formula.

VII Relativistic theory of the electron

A free particle in the special theory of relativity has an energy which is given by

$$E = c\sqrt{m^2c^2+p^2}$$

where p is its momentum. The natural thing to do to obtain a relativistic quantum mechanics would be to use for p the operator which represents it, and write

$$\left(E - c\sqrt{m^2c^2+p^2}\right)\psi = 0 , \qquad \vec{p} = \frac{\hbar}{i}\nabla$$

as the wave equation of free particle states. This equation is however completely unsymmetrical in p and E, whereas we know that these together actually build a four vector in relativity. This leads, among other things, to the impossibility of generalizing it relativistically when an external electromagnetic field is present. We can improve this situation if we multiply by $E + c\sqrt{m^2c^2+p^2}$. Then we obtain

$$\left[E^2 - c^2(m^2c^2+p^2)\right]\psi = 0,$$

which is a relativistically invariant form. This is known as the Klein-Gordon equation, and with the proper interpretation can be used as the basis for a quantum theory of particles without spin. We shall not go into this here, because this equation does not describe electrons. We can at the same time find a relativistic equation and obtain one more closely related to ordinary quantum mechanics if we proceed in the following fashion. We made the above equation relativistically invariant by arranging things so that the squares of E and p come in symmetrically. It was Dirac's brilliant idea to arrange things so that E and p occur linearly. Let us write

$$\sqrt{m^2c^2+p^2} = \left(\alpha_x p_x + \alpha_y p_y + \alpha_z p_z + mc\beta\right),$$

where the α's and β are to be determined in such a way that the squares of both sides are equal. That is, we are finding an operator extraction of the square root. Squaring we get

$$\begin{aligned} m^2c^2 + p^2 = \; & \alpha_x^2 p_x^2 + \alpha_y^2 p_y^2 + \alpha_z^2 p_z^2 + p_x p_y (\alpha_x \alpha_y + \alpha_y \alpha_x) \\ & + p_y p_z (\alpha_y \alpha_z + \alpha_z \alpha_y) \\ & + p_z p_x (\alpha_z \alpha_x + \alpha_x \alpha_z) \\ & + \left(p_x(\alpha_x \beta + \beta \alpha_x) + p_y(\alpha_y \beta + \beta \alpha_y) + p_z(\alpha_z \beta + \beta \alpha_z)\right) mc \\ & + m^2 c^2 \beta^2 , \end{aligned}$$

From this it follows at once that

$$\begin{aligned} & \alpha_x^2 = \alpha_y^2 = \alpha_z^2 = \beta^2 = 1 \\ & \alpha_x \alpha_y + \alpha_y \alpha_x = 0 \\ & \alpha_x \beta + \beta \alpha_x = 0 , \\ & \text{etc.} \end{aligned}$$

That is, the square of each operator is unity, but they all anti-commute with each other.

We shall take

$$(E - c(\alpha_x p_x + \alpha_y p_y + \alpha_z p_z + mc\beta))\psi = 0$$

as our relativistic wave equation. Before we go on to prove the invariance let us give a representation of the α's and β. Others are possible, but this one is in common use. It is straightforward to show that unless we take α, β to be at least 4 x 4 matrices we cannot satisfy the above relationships, and further that any matrices larger than 4 x 4 which satisfy them are equivalent to direct products of four by four matrices. One possibility is

$$\alpha_x = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \qquad \alpha_y = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}$$

$$\alpha_z = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \qquad \beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

as may be verified by direct calculation. One may write these in another way:

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \alpha_x = \begin{pmatrix} 0 & \sigma_x \\ \sigma_x & 0 \end{pmatrix}, \ etc.,$$

where all the elements are two by two matrices. Of course if the α 's and β are 4 x 4 matrices, the wave function must be represented by a column vector of four rows:

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}$$

We notice from this representation that it is possible to choose the α 's and β to be hermitian.

Relativistic Invariance

To put space and time on the same footing let us consider not just the stationary solutions corresponding to a given energy, but those which vary with time. We know that for a stationary state with energy E we have

$$i\hbar \frac{\partial \psi}{\partial t} = E\psi ,$$

so that our euqation is equivalent to:

$$\left[i\hbar \frac{\partial}{\partial t} - c\frac{\hbar}{i}\left(\alpha_x \frac{\partial}{\partial x} + \alpha_y \frac{\partial}{\partial y} + \alpha_z \frac{\partial}{\partial z}\right) - mc^2\beta \right]\psi = 0$$

Put $\alpha_0 = -1$, $p_0 = \frac{\hbar}{ci}\frac{\partial}{\partial t}$, $x_0 = ct$, $p_0 = \frac{\hbar}{i}\frac{\partial}{\partial x_0}$.

Then this equation takes the form

$$(\alpha_0 p_0 + \alpha_x p_x + \alpha_y p_y + \alpha_z p_z + mc\beta)\psi = 0 .$$

Further, let us write $\alpha_0 p_0 + \alpha_x p_x + \alpha_y p_y + \alpha_z p_z = \alpha_k p_k$ implying the summation convention for repeated indices, $k = 0, x, y, z$. We have then

$$(\alpha_k p_k + mc\beta)\psi = 0 .$$

This is known as the Dirac Equation.

Now what do we mean by Lorentz invariance? We mean that if we write down the wave equation in a different Lorentz frame (i.e., transform the p_k), then the solutions of the new equation can be put in one to one correspondence with the solutions of the original equation in such a way that corresponding solution may be assumed to represent the same state. Now the most general Lorentz transformation consists of two parts, a translation with velocity v, and a rotation of coordinates. We shall show that the above is true for both. Consider first a rotation of coordinates, say about the x-axis through an angle θ.

Call the new coordinates x_k' the old x_k .

$$t = t' \quad \text{or} \quad x_0 = x_0'$$

$$x = x' \qquad y = y'\cos\theta + z'\sin\theta \,, \qquad z = -y'\sin\theta + z'\cos\theta \,.$$

Then, since $p_k = \frac{\hbar}{i}\frac{\partial}{\partial x_k}$, we have

$$p_0 = p_0'$$

$$p_x = p_x'$$

$$p_y = p_y'\cos\theta + p_z'\sin\theta$$

$$p_z = -p_y'\sin\theta + p_z'\cos\theta$$

The new wave equation is

$$(\alpha_0 p_0' + \alpha_x p_x' + (\alpha_y\cos\theta - \alpha_z\sin\theta)p_y' + (\alpha_y\sin\theta + \alpha_z\cos\theta)p_z' + mc\beta)\psi = 0 \,.$$

Now can we find a new wave function ψ' which satisfies the same equation for the new system as ψ for the old? That is, can we find a matrix γ such that

$$\psi' = \gamma\psi \quad \text{and}$$

$$(\alpha_k p_k' + mc\beta)\psi' = 0$$

where γ depends only on the Lorentz transformation and not on the state?

We assert that $\gamma = \exp(-\frac{1}{2}\theta\alpha_y\alpha_z)$ does this. For if we put this in we have (after multiplying with γ)

$$e^{-\frac{1}{2}\theta\alpha_y\alpha_z}(\alpha_0 p_0' + \alpha_x p_x' + (\alpha_y\cos\theta - \alpha_z\sin\theta)p_y' + (\alpha_y\sin\theta + \alpha_z\cos\theta)p_z' + mc\beta)e^{+\frac{1}{2}\theta\alpha_y\alpha_z}\psi' = 0$$

But $\gamma\alpha_0\gamma^{-1} = \alpha_0$, $\alpha_0 = -1$

$$\left.\begin{aligned} \gamma\alpha_x\gamma^{-1} &= \alpha_x \\ \gamma\beta\gamma^{-1} &= \beta \end{aligned}\right\}$$ since α_x, β anticommute with α_y and α_z and therefore commute with $\alpha_y\alpha_z$.

Finally since $(\alpha_y\alpha_z)^2 = (\alpha_y\alpha_z)(\alpha_y\alpha_z) = -\alpha_y^2\alpha_z^2 = -1$, we get

$$\gamma = e^{-\frac{1}{2}\theta\alpha_y\alpha_z} = \cos\frac{\theta}{2} - \alpha_y\alpha_z\sin\frac{\theta}{2}$$

$$\gamma^{-1} = e^{+\frac{1}{2}\theta\alpha_y\alpha_z} = \cos\frac{\theta}{2} + \alpha_y\alpha_z\sin\frac{\theta}{2}.$$

$$\begin{aligned}
\gamma(\alpha_y\cos\theta - \alpha_z\sin\theta)\gamma^{-1} &= \left(\cos\frac{\theta}{2} - \alpha_y\alpha_z\sin\frac{\theta}{2}\right)(\alpha_y\cos\theta - \alpha_z\sin\theta)\left(\cos\frac{\theta}{2} + \alpha_y\alpha_z\sin\frac{\theta}{2}\right) \\
&= \cos^2\frac{\theta}{2}(\alpha_y\cos\theta - \alpha_z\sin\theta) \\
&\quad + \cos\frac{\theta}{2}\sin\frac{\theta}{2}\Big[-\alpha_y\alpha_z(\alpha_y\cos\theta - \alpha_z\sin\theta) \\
&\qquad + (\alpha_y\cos\theta - \alpha_z\sin\theta)\alpha_y\alpha_z\Big] \\
&\quad - \sin^2\frac{\theta}{2}\left(\alpha_y\alpha_z(\alpha_y\cos\theta - \alpha_z\sin\theta)\alpha_y\alpha_z\right) \\
&= \cos\theta(\alpha_y\cos\theta - \alpha_z\sin\theta) + \sin\theta(\alpha_y\sin\theta + \alpha_z\cos\theta) \\
&= \alpha_y
\end{aligned}$$

Similarly $\gamma(\alpha_y\sin\theta + \alpha_z\cos\theta)\gamma^{-1} = \alpha_z$, so that we have

$(\alpha_k p_k' + mc\beta)\psi' = 0$, which was to be proven.

Now consider a translation:

$$x = \frac{x' - v t'}{\sqrt{1 - v^2/c^2}} = \frac{x' - \frac{v}{c} x_0'}{\sqrt{}}$$

$$x_0 = c t = \frac{-\frac{v}{c} x' + x_0'}{\sqrt{}}$$

$$y' = y$$

$$z' = z ,$$

These are the well known equations for a Lorentz transformation due to a translation along the X axis with velocity -v. If we define

$$\cosh \vartheta = \frac{1}{\sqrt{1 - v^2/c^2}} , \quad \sinh \vartheta = \frac{v/c}{\sqrt{1 - v^2/c^2}} ,$$

then $\cosh^2 \vartheta - \sinh^2 \vartheta = 1$, as they should, and our relationships have the form

$$x = \cosh \vartheta \cdot x' - \sinh \vartheta \, x_0'$$

$$x_0 = -\sinh \vartheta \, x' + \cosh \vartheta \, x_0'$$

$$y = y'$$

$$z = z'$$

Making this Lorentz transformation we obtain for the p_k :

$$p_0 = p_0' \cosh \vartheta + p_x' \sinh \vartheta$$

$$p_x = p_0' \sinh \vartheta + p_x' \cosh \vartheta$$

$$p_y = p_y'$$

$$p_z = p_z'$$

Consequently we have

$$(\alpha_0 \cosh\vartheta + \alpha_x \sinh\vartheta)p_0' + (\alpha_0 \sinh\vartheta + \alpha_x \cosh\vartheta)p_x'$$

$$+ \alpha_y p_y' + \alpha_z p_z' + mc\beta)\psi = 0$$

Now put $\psi' = \bar{\gamma}\,\psi$, where

$$\bar{\gamma} = \exp(-\tfrac{1}{2}\vartheta\alpha_x) = \cosh\frac{\vartheta}{2} - \alpha_x \sinh\frac{\vartheta}{2}$$

Substituting for ψ and multiplying the resultant equation by $\bar{\gamma}^{-1}$ we obtain

$$\bar{\gamma}^{-1}\left[(\alpha_0 \cosh\vartheta + \alpha_x \sinh\vartheta)p_0' + (\alpha_0 \sinh\vartheta + \alpha_x \cosh\vartheta)p_x' + \alpha_y p_y' + \alpha_z p_z' + mc\beta\right]\bar{\gamma}^{-1}\psi' = 0.$$

However,

$$e^{+\frac{1}{2}\vartheta\alpha_x}\,\alpha_y\, e^{+\frac{1}{2}\vartheta\alpha_x} = e^{+\frac{1}{2}\vartheta\alpha_x}\, e^{-\frac{1}{2}\vartheta\alpha_x}\,\alpha_y = \alpha_y$$

ince α_y anticommutes with α_x . Similarly

$$\bar{\gamma}^{-1}\alpha_z\bar{\gamma}^{-1} = \alpha_z$$

$$\bar{\gamma}^{-1}\beta\,\bar{\gamma}^{-1} = \beta$$

Finally $e^{\frac{1}{2}\vartheta\alpha_x}(\alpha_0 \cosh\vartheta + \alpha_x \sinh\vartheta)e^{\frac{1}{2}\vartheta\alpha_x}$

$$= e^{\vartheta\alpha_x}(\alpha_0 \cosh\vartheta + \alpha_x \sinh\vartheta)$$

$$= -e^{+\vartheta\alpha_x}e^{-\vartheta\alpha_x} \quad , \text{ since } \alpha_0 = -1$$

$$= -1 = \alpha_0$$

$$e^{\frac{1}{2}\vartheta\alpha_x}(\alpha_0\sinh\vartheta+\alpha_x\cosh\vartheta)e^{\frac{1}{2}\vartheta\alpha_x}$$

$$=e^{\vartheta\alpha_x}(-\sinh\vartheta+\alpha_x\cosh\vartheta)$$

$$=(\cosh\vartheta+\alpha_x\sinh\vartheta)(-\sinh\vartheta+\alpha_x\cosh\vartheta)$$

$$=\alpha_x(\cosh^2\vartheta-\sinh^2\vartheta)=\alpha_x,$$

so that once again we have

$$(\alpha_k p_k'+mc\beta)\psi'=0$$

Thus since the most general Lorentz transformation can be built up out of translations and rotations, we have proved Lorentz invariance in general. It is sometimes rather convenient to have another notation for the Dirac equation. Let us put

$$\gamma_1=-i\beta\alpha_x$$

$$\gamma_2=-i\beta\alpha_y$$

$$\gamma_3=-i\beta\alpha_z$$

$$\gamma_4=\beta$$

These matrices are hermitian if the α and β are. From the commutation rules of the latter we obtain at once the very symmetric ones for the γ

$$\gamma_\mu\gamma_\nu+\gamma_\nu\gamma_\mu=2\delta_{\mu\nu}\qquad(\mu,\nu=1,2,3,4)$$

Now suppose we return to the Dirac equation and multiply it by $-i\beta$. Then we have

$$-i\beta\left(\frac{\hbar}{ic}\frac{\partial}{\partial t}+\alpha_x\frac{\hbar}{i}\frac{\partial}{\partial x}+\alpha_y\frac{\hbar}{i}\frac{\partial}{\partial y}+\alpha_z\frac{\hbar}{i}\frac{\partial}{\partial z}+mc\beta\right)\psi=0, \text{ or}$$

$$\left\{\frac{\hbar}{i}\left(\gamma_1\frac{\partial}{\partial x}+\gamma_2\frac{\partial}{\partial y}+\gamma_3\frac{\partial}{\partial z}+\gamma_4\frac{\partial}{\partial(ict)}\right)-mci\right\}\psi=0.$$

Put

$$x = x_1$$

$$y = x_2$$

$$z = x_3$$

$$ict = x_4$$

Then we have

$$\left\{ \frac{\hbar}{i} \left(\gamma_\mu \frac{\partial}{\partial x_\mu} \right) - mci \right\} \psi = 0$$

or

$$\left(\gamma_\mu \frac{\partial}{\partial x_\mu} + \varkappa \right) \psi = 0 \qquad \varkappa = \frac{mc}{\hbar},$$

the reciprocal compton wavelength of the particle. It proves useful to introduce an auxillary wave function $\bar{\psi}$ which is defined by

$$\bar{\psi} = \psi^* \beta, \qquad \psi^* = (\psi_1^*, \psi_2^*, \psi_3^*, \psi_4^*)$$

ψ^* being the transpose complex conjugate of ψ. By taking the transpose complex conjugate of the Dirac equation and multiplying on the right by β we obtain

$$\frac{\partial \bar{\psi}}{\partial x_\mu} \gamma_\mu - \varkappa \bar{\psi} = 0.$$

Physical Interpretation of the Dirac Equation

As is well known, in quantum mechanics we need not only a Schrödinger equation, but also a definition of the probability density, i.e., the probability that a particle will be in a certain volume per unit volume. In ordinary nonrelativistic quantum mechanics this was just $\psi^*\psi$. The question is, what do we take in Dirac theory? We shall show that the following definition

$$P = \sum_{i=1}^{4} \psi_i^*(x)\, \psi_i(x) \equiv \text{[redacted]}\ (\psi^*\psi)$$

(the scalar product notation will only be used in this sense from now on) is satisfactory. For a probability density to be satisfactory, we must have that

$$\int_{\text{all space}} P\,dx\,dy\,dz = 1\,, \text{ or}$$

$$\frac{\partial}{\partial t}\int P\,dx\,dy\,dz = 0 = \int \frac{\partial P}{\partial t}\,dx\,dy\,dz.$$

Further it must be positive. The latter is from the definition obvious; we now prove the former. Taking the Dirac equation in the form

$$\left[\frac{\partial}{\partial t} + c\left(\alpha_x \frac{\partial}{\partial x} + \alpha_y \frac{\partial}{\partial y} + \alpha_z \frac{\partial}{\partial z}\right) + \frac{mc^2 i}{\hbar}\beta\right]\psi = 0,$$

we get

$$\left[\frac{\partial \psi^*}{\partial t} + c\left(\frac{\partial \psi^*}{\partial x}\alpha_x + \frac{\partial \psi^*}{\partial y}\alpha_y + \frac{\partial \psi^*}{\partial z}\alpha_z\right) - \frac{mc^2 i}{\hbar}\psi^*\beta\right] = 0.$$

Take the scalar product of the first with ψ^* of the second with ψ, and add. We get

$$\frac{\partial}{\partial t}(\psi^*\psi) + c\left[\left(\psi^*\alpha_x \frac{\partial \psi}{\partial x}\right) + \left(\frac{\partial \psi^*}{\partial x}\alpha_x \psi\right) + \text{etc.}\right] = 0$$

$$\frac{\partial}{\partial t}(\psi^*\psi) + c\left[\frac{\partial}{\partial x}(\psi^*\alpha_x\psi) + \dots\right] = 0$$

Write

$$\vec{S} = +c\,(\psi^*\vec{\alpha}\,\psi)$$

$$P = (\psi^*\psi),$$

then

$$\frac{\partial P}{\partial t} + \nabla\cdot\vec{S} = 0$$

This is the differential form of the law of conservation of probability. $\vec{S}$ is the probability current density.

$$\int \frac{\partial P}{\partial t}\, dx\, dy\, dz = \int \nabla\cdot\vec{S}\, dx\, dy\, dz = \int_{\infty} \vec{S}\cdot d\vec{A} \longrightarrow 0$$

if the particle is confined to a finite region of space (i.e. if it does not have infinite probability of being found at infinity). Thus P is a satisfactory probability density.

These relationships may be written in a form which is a little more clearly invariant.

$$P = (\psi^*\psi) = (\bar{\psi}\beta\psi) = (\bar{\psi}\gamma_4\psi)$$

$$\vec{S} = +c(\psi^*\vec{\alpha}\psi) = +c(\bar{\psi}\beta\vec{\alpha}\psi) = +ic(\bar{\psi}\vec{\gamma}\psi).$$

Write $S_\mu = \left(\frac{\vec{S}}{c}, iP\right)$. Then

$$S_\mu = i(\bar{\psi}\gamma_\mu\psi).$$

If we multiply this by the electric charge $-e$ we get the current four vector

$$J_\mu = -ie(\bar{\psi}\gamma_\mu\psi),$$

which satisfies the equation of continuity

$$\frac{\partial J_\mu}{\partial x_\mu} = 0$$

To show that J_μ (or S_μ) really transform as four vectors under Lorentz transformations, we can proceed as follows:

Write the Dirac equation in the form

$$\gamma_\mu \frac{\partial\psi}{\partial x_\mu} + \kappa\psi = 0$$

Then Lorentz invariance means that when we make a transformation

$$x_\mu' = a_{\mu\nu} x_\nu \quad \text{with}$$

$$x_\mu x_\mu = x_\mu' x_\mu' \quad \text{so that}$$

$$a_{\mu\rho} a_{\mu\sigma} = \delta_{\rho\sigma}$$ (which is the most general Lorentz transformation), we get

$$\gamma_\mu \frac{\partial \psi}{\partial x_\mu} + \varkappa \psi = 0 \quad \text{is equivalent to}$$

$$\gamma_\mu \frac{\partial \psi'}{\partial x_\mu'} + \varkappa \psi' = 0 ,$$ where $\psi' = S\psi$ and S depends only on the coefficients $a_{\mu\nu}$. Since

$$\frac{\partial}{\partial x_\mu} = \frac{\partial x_\nu'}{\partial x_\mu} \frac{\partial}{\partial x_\nu'} = a_{\nu\mu} \frac{\partial}{\partial x_\nu'} ,$$ this implies that

$$\gamma_\mu a_{\nu\mu} \frac{\partial \psi}{\partial x_\nu'} + \varkappa \psi = 0 \quad \text{or}$$

$$\gamma_\mu a_{\nu\mu} S^{-1} \frac{\partial \psi'}{\partial x_\nu'} + S^{-1} \varkappa \psi' = 0 .$$

Multiplying this by S we obtain

$$(S \gamma_\mu S^{-1}) a_{\nu\mu} \frac{\partial \psi'}{\partial x_\nu'} + \varkappa \psi' = 0$$

This is only possible if

$$S \gamma_\mu S^{-1} a_{\nu\mu} = \gamma_\nu \quad \text{or}$$

$$S^{-1} \gamma_\nu S = a_{\nu\mu} \gamma_\mu .$$

That means that under the transformation S the γ's transform like four vectors.

Consider the equation

$$\frac{\partial \bar{\psi}}{\partial x_\mu}\gamma_\mu - \varkappa\bar{\psi} = 0,$$

which must take the form

$$\frac{\partial \bar{\psi}'}{\partial x'_\mu}\gamma_\mu - \varkappa\bar{\psi}' = 0$$

after a Lorentz transformation.

Then $\bar{\psi}' = \bar{\psi}\bar{S}$. What is $\bar{S}$?

$$a_{\nu\mu}\frac{\partial \bar{\psi}}{\partial x'_\nu}\gamma_\mu - \varkappa\bar{\psi} = 0$$

$$\frac{\partial \bar{\psi}'}{\partial x_\nu'}(\bar{S}^{-1}\gamma_\mu a_{\nu\mu}) - \varkappa\bar{\psi}'\bar{S}^{-1} = 0$$

$$\frac{\partial \bar{\psi}'}{\partial x_\nu'}(\bar{S}^{-1}\gamma_\mu \bar{S})a_{\nu\mu} - \varkappa\bar{\psi}' = 0$$

Therefore we must have

$$(\bar{S}^{-1}\gamma_\mu\bar{S})\,a_{\nu\mu} = \gamma_\nu$$

$$\bar{S}\gamma_\nu\bar{S}^{-1} = a_{\nu\mu}\gamma_\mu = S^{-1}\gamma_\nu S$$

By comparison* we see that $\bar{S} = S^{-1}$ so that

$$\bar{\psi}' = \bar{\psi}S^{-1}$$

Now consider the quantities

$$J = (\bar{\psi}\psi)$$

$$J_\mu = (\bar{\psi}\gamma_\mu\psi)$$

$$\left.\begin{array}{ll} J_{\mu\nu} = (\bar{\psi}\gamma_\mu\gamma_\nu\psi) & \mu\neq\nu \\ J_{\mu\nu\lambda} = (\bar{\psi}\gamma_\mu\gamma_\nu\gamma_\lambda\psi) & \mu\neq\nu\neq\lambda \\ J_{\mu\nu\lambda\rho} = (\bar{\psi}\gamma_\mu\gamma_\nu\gamma_\lambda\gamma_\rho\psi) & \mu\neq\nu\neq\lambda\neq\rho \end{array}\right\} \text{Zero otherwise.}$$

---

* From above we have at once that $S\bar{S}\gamma_\nu = \gamma_\nu S\bar{S}$. Thus $S\bar{S}$ commutes with all the γ_ν, and it is a trivial matter to show that it is therefore a multiple of the unit matrix.

and how they change under a Lorentz transformation. The first is just a scalar since

$$J' = (\bar{\psi}'\psi') = (\bar{\psi} S^{-1} S \psi) = (\bar{\psi}\psi) = J .$$

The second is a four vector, since

$$J_\mu' = (\bar{\psi}'\gamma_\mu \psi') = (\bar{\psi} S^{-1}\gamma_\mu S \psi) =$$
$$= a_{\mu\nu}(\bar{\psi}\gamma_\nu\psi) = a_{\mu\nu} J_\nu ,$$

which is just the transformation rule for a four vector. Thus we have proved that the current really does build a four vector.

Similarly $J_{\mu\nu}$ is an antisymmetric tensor of the second rank, $J_{\mu\nu\lambda}$ is an antisymmetric tensor of the third rank (it therefore has 4 independent components, and is equivalent to a pseudo-vector), and finally $J_{\mu\nu\lambda\rho}$ is an antisymmetric tensor of the fourth rank and has only one independent component. This latter is actually a pseudoscalar. Consider this component, call it N ;

$$N = (\bar{\psi}\gamma_1\gamma_2\gamma_3\gamma_4\psi)$$

or $N = (\bar{\psi}\gamma_5\psi)$, where $\gamma_5 \equiv \gamma_1\gamma_2\gamma_3\gamma_4$.

This follows at once since

$$N' = (\bar{\psi}'\gamma_5\psi') = (\bar{\psi} S^{-1}\gamma_5 S \psi)$$
$$= (\bar{\psi}\, S^{-1}\gamma_1 S\; S^{-1}\gamma_2 S\; S^{-1}\gamma_3 S\; S^{-1}\gamma_4 S\,\psi)$$
$$= a_{1\nu}\, a_{2\mu}\, a_{3\lambda}\, a_{4\rho}\,(\bar{\psi}\gamma_\nu\gamma_\mu\gamma_\lambda\gamma_\rho\psi)$$

The terms which do not satisfy $\nu \neq \mu \neq \lambda \neq \rho$ are identically zero. If $\lambda = \rho$ say, then $\gamma_\lambda \gamma_\rho = \gamma_\lambda^2 = 1$ and $a_{3\lambda} a_{4\lambda} = \delta_{3,4} = 0.$ Now the terms satisfying this condition are all equal, apart from a sign. This is +1 if $\nu \mu \nu \rho$ is an even permutation of $1\,2\,3\,4$, -1 if an odd permutation. Thus

$$N' = \mathrm{Det}\,(a_{\mu\nu})(\bar{\psi}\,\gamma_1\gamma_2\gamma_3\gamma_4\,\psi) = \mathrm{Det}(a_{\mu\nu})\,N$$

$\mathrm{Det}\,(a_{\mu\nu}) = +1$ for a proper Lorentz transformation

$\mathrm{Det}\,(a_{\mu\nu}) = -1$ " an improper " " ,

so that we have established the pseudoscalarity of N. Incidentally γ_5 also has the property

$$\gamma_5^2 = \gamma_1\gamma_2\gamma_3\,\gamma_4 . \gamma_1\gamma_2\gamma_3\gamma_4 = 1$$

$$\gamma_5\gamma_\mu + \gamma_\mu\gamma_5 = 0$$

so that there are actually five matrices γ_μ, γ_5 which satisfy

$$(\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu) = 2\delta_{\mu\nu}.$$

Free Particle Solutions

Let us find the solutions corresponding to the free particle Dirac equation. If we write the equation in the form

$$(E - c\,\vec{\alpha}\cdot\vec{p} - mc^2\beta)\psi = 0,$$

we can immediately separate out the space and time dependence by writing

$$\psi = u\, e^{i\vec{p}'\cdot\vec{r}/\hbar}\, e^{-\frac{iEt}{\hbar}},$$

where u is a column matrix independent of x, y, z or t, and $\vec{p}'$ is the vector giving the momentum of the free particle. This reduces our equation to

$$(E - c\,\vec{\alpha}\cdot\vec{p}' - mc^2\beta)\,u = 0.$$

Since this is a 4 x 4 matrix equation we should be able to find four solutions. Call these u^1, u^2, u^3, u^4. Since each has 4 components, we may build a matrix

$$U = \begin{pmatrix} u_1^1 & u_1^2 & u_1^3 & u_1^4 \\ u_2^1 & u_2^2 & u_2^3 & u_2^4 \\ u_3^1 & u_3^2 & u_3^3 & u_3^4 \\ u_4^1 & u_4^2 & u_4^3 & u_4^4 \end{pmatrix}.$$

Since each column satisfies the above equation, so will U, i.e.,

$$(E - c\vec{\alpha}\cdot\vec{p}' - mc^2\beta)U = 0.$$

U however may easily be given. Consider

$$U = (E + c\vec{\alpha}\cdot p' + mc^2\beta)N$$

where N is a constant ~~matrix~~ to provide normalization.

$$(E - c\vec{\alpha}\cdot\vec{p}' - mc^2\beta)U = (E^2 - c^2(\vec{\alpha}\cdot\vec{p}' + mc\beta)^2)N$$

$$= (E^2 - c^2(p'^2 + m^2c^2))N = 0, \text{ if}$$

$$E = \pm c\sqrt{p'^2 + m^2c^2}.$$

If we take the upper sign we get the usual relativistic energy momentum relationship, if we take the lower we get negative energies for free particles. To obtain all four solutions, however, one must take both signs. This difficulty is not so prominent here, but when we come to discuss external fields it becomes so, since we cannot simply exclude the negative energy states as the external fields cause transitions between them and states of positive energy. We will return to this question later. Let us call $+c\sqrt{p'^2 + c^2m^2} = c\Omega(p')$ Then we may clearly write

$$u_\mu^1 = (U_{\mu 1}), \text{ etc}$$

[redacted] $$u_\mu^1 = (E\,\delta_{\mu 1} + c(\vec{\alpha}\cdot\vec{p}' + mc\beta)_{\mu 1})N_1$$

Taking the representation we had previously we see*

$$u^1 = N \begin{pmatrix} mc+\Omega \\ 0 \\ p_z' \\ p_x'+ip_y' \end{pmatrix} \qquad u^2 = N \begin{pmatrix} 0 \\ mc+\Omega \\ p_x'-ip_y' \\ -p_z' \end{pmatrix}$$

$$u^3 = N \begin{pmatrix} p_z' \\ p_x'+ip_y' \\ -\Omega-mc \\ 0 \end{pmatrix} \qquad u^4 = N \begin{pmatrix} p_x'-ip_y' \\ -p_z' \\ 0 \\ -\Omega-mc \end{pmatrix}$$

The first two solutions corresponding to a positive energy, $c\Omega$ the second two to negative energy $-c\Omega$. If we choose

$$N = \frac{\pm 1}{{}_{+}\sqrt{(mc+\Omega)^2+p'^2}} \qquad (\pm \text{ depending on the sign of the energy}),$$

then the solutions are normalized in the sense that $(u^*u)=1$. This may all be verified by direct calculation. If we go to the limit of vanishing velocity we have

$$u^1 \rightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad u^2 \rightarrow \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$u^3 \rightarrow \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \qquad u^4 \rightarrow \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

---

* The independence of our four solutions can readily be checked by [redacted] noting that $(u^{i*}u^j)=0$ for $i \neq j$.

so that we might imagine that at low velocities Dirac theory goes over to Pauli spin theory, the first two solutions corresponding (since their energy is positive) to the spin "up" and spin "down" cases. We shall verify this in detail later.

External Forces

We shall now consider how the Dirac equation is modified by the pressure of an external electromagnetic field. Let this field be given by its four potential A_μ [$A_1 = A_x$, $A_2 = A_y$, $A_3 = A_z$, $A_4 = i\phi$, $\vec{A}$ and ϕ being the ordinary vector and scalar potentials of the electromagnetic field. It is well known from Maxwell theory that this builds a four vector.] Now in classical mechanics to obtain the motion in an electromagnetic field we just replace p_μ by (for an electron with charge $-e$) $p_\mu + \frac{e}{c} A_\mu$ This makes sense, since they are both four vectors. We shall do the same for the Dirac case. The field free equation is

$$\frac{i}{\hbar} \gamma_\mu p_\mu \psi + \kappa \psi = 0 \quad .$$

Therefore when a field is present we take as our equation:

$$\frac{i}{\hbar} \gamma_\mu (p_\mu + \frac{e}{c} A_\mu) \psi + \kappa \psi = 0$$

or

$$\left[\gamma_\mu \left(\frac{\partial}{\partial x_\mu} + \frac{ie}{\hbar c} A_\mu \right) + \kappa \right] \psi = 0 \, .$$

This we shall take as the basic equation for a particle of charge -e in a given electromagnetic field A_μ . One important property which this equation must possess, in addition to that of Lorentz invariance,* is that of gauge invariance. Since the electromagnetic fields are given by the curl of the 4-vector potential

---

*The Lorentz invariance is of course trivial, since and transform identically.

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x_\mu} - \frac{\partial A_\nu}{\partial x_\mu}$$

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & H_z & -H_y & -iE_x \\ -H_z & 0 & H_x & -iE_y \\ H_y & -H_x & 0 & -iE_z \\ iE_x & iE_y & iE_z & 0 \end{pmatrix},$$

the 4 potential is only determined to within a 4 gradient. That is, if we add to A_μ the quantity $\frac{\partial f}{\partial x_\mu}$, where f is any scalar, we get the same fields, since

$$\frac{\partial}{\partial x_\mu}\left(\frac{\partial f}{\partial x_\nu}\right) - \frac{\partial}{\partial x_\nu}\left(\frac{\partial f}{\partial x_\mu}\right) = \frac{\partial^2 f}{\partial x_\mu \partial x_\nu} - \frac{\partial^2 f}{\partial x_\nu \partial x_\mu} \equiv 0 .$$

Therefore if we replace A_μ by $A_\mu + \frac{\partial f}{\partial x_\mu} = A_\mu'$, the solutions of our equation (ψ') should be in one to one correspondence with those of our original equation. The transformed equation is:

$$\left[\gamma_\mu\left(\frac{\partial}{\partial x_\mu} + \frac{ie}{\hbar c}A_\mu'\right) + \kappa\right]\psi' = \left[\gamma_\mu\left(\frac{\partial}{\partial x_\mu} + \frac{ie}{\hbar c}\left(A_\mu + \frac{\partial f}{\partial x_\mu}\right)\right) + \kappa\right]\psi' = 0$$

Now put $\psi' = e^{-i\frac{e}{\hbar c}f}\psi$, then we see at once that

$$\left[\gamma_\mu\left(\frac{\partial}{\partial x_\mu} + \frac{ie}{\hbar c}A_\mu\right) + \kappa\right]\psi = 0$$

so that the required one to one correspondence exists, and we have obtained our original equation. All physical properties of the system must be the same if we use either ψ or ψ'. As an example

$$j_\mu' = -ie(\bar{\psi}'\gamma_\mu\psi') = -ie\, e^{\frac{ie}{\hbar c}f} e^{-\frac{ie}{\hbar c}f}(\bar{\psi}\gamma_\mu\psi) = j_\mu .$$

Taking the complex conjugate as before, and multiplying with β we can easily prove that satisfies

$$\frac{\partial\bar{\psi}}{\partial x_\mu}\gamma_\mu - \frac{ie}{\hbar c}\bar{\psi}\gamma_\mu A_\mu - \kappa\bar{\psi} = 0.$$

If we wish to write our equation in terms of a hamiltonian we have to multiply by β and find H such that $i\hbar\frac{\partial\psi}{\partial t} = H\psi$.

This is easily seen to be

$$\mathcal{H} = c\left[\vec{\alpha}\cdot\left(\vec{p}+\frac{e}{c}\vec{A}\right)+mc\beta\right]-e\phi .$$

Incidentally, the equation we have given is not the only one which satisfies the conditions of Lorentz and gauge invariance. We could add a term

$$\lambda\, F_{\mu\nu}\,\gamma_\mu\gamma_\nu\psi .$$

The resulting equation describes a particle which (we shall not prove it here) has an intrinsic magnetic moment depending on λ. It is possible that the proton or neutron satisfy such an equation, but for the electron (if we neglect very small terms due to interaction with the radiation field) we need not consider terms of this type.

Non-Relativistic Spin Theory as First Approximation

In performing this reduction it is convenient to use the representation which has β (or γ_4) diagonal. In this representation we shall find that the first two components of ψ (in the non-relativistic limit) are large, while the other two are small. Let us write again

$$\psi = \begin{pmatrix}\psi_1\\ \psi_2\\ \psi_3\\ \psi_4\end{pmatrix}.$$

In the non-relativistic case the energy differs only slightly from mc^2. Let us put

$$\psi_1 = \varphi_1 e^{-\frac{i}{\hbar}mc^2t} \qquad \psi_3 = \chi_1 e^{-i/\hbar\, mc^2t}$$

$$\psi_2 = \varphi_2 e^{-\frac{i}{\hbar}mc^2t} \qquad \psi_4 = \chi_2 e^{-i/\hbar\, mc^2t}$$

We may write the Dirac equation as:

$$i\hbar\frac{\partial\psi}{\partial t} = \left\{ c\left[\vec{\alpha}\cdot\left(\vec{p}+\frac{e}{c}\vec{A}\right)+mc\beta\right] - e\phi\right\}\psi$$

$$mc^2\begin{pmatrix}\varphi\\ \chi\end{pmatrix} + i\hbar\begin{pmatrix}\partial\varphi/\partial t\\ \partial\chi/\partial t\end{pmatrix}$$

$$= \left\{ c\left[\begin{pmatrix}0 & \vec{\sigma}\\ \vec{\sigma} & 0\end{pmatrix}\cdot\left(\vec{p}+\frac{e}{c}\vec{A}\right)+mc\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\right] - e\phi\right\}\begin{pmatrix}\varphi\\ \chi\end{pmatrix},$$

$$\varphi = \begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix}, \quad \chi = \begin{pmatrix}\chi_1\\ \chi_2\end{pmatrix}.$$

Or

$$i\hbar\frac{\partial\varphi}{\partial t} = \left(c\left(\vec{p}+\frac{e}{c}\vec{A}\right)\cdot\vec{\sigma}\,\chi - e\phi\,\varphi\right)$$

$$2mc^2\chi + i\hbar\frac{\partial\chi}{\partial t} = \left(c\left(\vec{p}+\frac{e}{c}\vec{A}\right)\cdot\vec{\sigma}\,\varphi - e\phi\chi\right).$$

Now since χ comes in the second equation multiplied by $2mc^2$ while φ comes multiplied at most by terms $\sim cp$ this equation is only possible if the components of χ are much less than those of φ. The Zeroth approximation would be to put $\chi = 0$, except in the term which is multiplied by $2mc^2$. That is

$$2mc^2\chi \cong c\left(\vec{p}+\frac{e}{c}\vec{A}\right)\cdot\vec{\sigma}\,\varphi$$

$$\chi = \frac{1}{2mc}\left(\vec{p}+\frac{e}{c}\vec{A}\right)\cdot\vec{\sigma}\,\varphi.$$

Substituting this in the first equation we obtain

$$i\hbar\frac{\partial\varphi}{\partial t} = \left\{\frac{(\vec{\sigma}\cdot\vec{\pi})(\vec{\sigma}\cdot\vec{\pi})}{2m} - e\phi\right\}\varphi,$$

where $\vec{\Pi} \equiv \vec{p} + \frac{e}{c}\vec{A}$.

To evaluate this we make use of the following identity:

$$(\vec{\sigma}\cdot\vec{B})(\vec{\sigma}\cdot\vec{C}) = \sigma_x^2 B_x C_x + \cdots + \text{[illegible]} + \sigma_x\sigma_y B_x C_y + \sigma_y\sigma_x B_y C_x + \cdots$$

$$= (B_x C_x + B_y C_y + B_z C_z) + i\sigma_z (B_x C_y - C_x B_y) + \cdots$$

$$= \vec{B}\cdot\vec{C} + i\,\vec{\sigma}\cdot(\vec{B}\times\vec{C})$$

Now put $\vec{B} = \vec{C} = \vec{\pi}$

$$(\vec{\sigma}\cdot\vec{\pi})(\vec{\sigma}\cdot\vec{\pi}) = \pi^2 + i\,\vec{\sigma}\cdot(\vec{\pi}\times\vec{\pi})$$

$$\vec{\Pi}\times\vec{\Pi} = (\vec{p} + \tfrac{e}{c}\vec{A})\times(\vec{p} + \tfrac{e}{c}\vec{A}) = \frac{e}{c}(\vec{p}\times\vec{A} + \vec{A}\times\vec{p})$$

$$(\vec{p}\times\vec{A} + \vec{A}\times\vec{p})f = \frac{\hbar}{i}\left(\nabla\times(\vec{A}f) + \vec{A}\times\nabla f\right)$$

$$= \frac{\hbar}{i}\left((\nabla\times\vec{A})f + (\nabla f)\times\vec{A} + \vec{A}\times(\nabla f)\right)$$

$$= \frac{\hbar}{i}(\nabla\times\vec{A})f = \frac{\hbar}{i}\vec{H}f,$$

$$\vec{\sigma}\cdot(\vec{\Pi}\times\vec{\Pi}) = -\frac{i\hbar e}{c}\,\vec{\sigma}\cdot\vec{H}$$

where $\vec{H}$ is the magnetic field.

Thus

$$i\hbar\frac{\partial\varphi}{\partial t} \cong \left\{\frac{(\vec{p} + \frac{e}{c}\vec{A})^2}{2m} + \frac{e\hbar}{2mc}\vec{\sigma}\cdot\vec{H} - e\phi\right\}\varphi$$

This is exactly the Pauli equation for a particle of spin 1/2, magnetic moment $\mu_B = \frac{e\hbar}{2mc}$ in a given electromagnetic field. Thus we see that the Dirac equation gives us the spin and magnetic moment of the electron without any assumptions at all. It is very easy to push this approximation further, and obtain relativistic corrections. To push it one step further we have

$$\chi = \frac{1}{2mc^2}\left[c(\vec{\sigma}\cdot\vec{\pi})\varphi - (i\hbar\frac{\partial}{\partial t} + e\phi)\chi\right]$$

$$\cong \frac{\vec{\sigma}\cdot\vec{\pi}}{2mc}\varphi - (i\hbar\frac{\partial}{\partial t} + e\phi)\frac{1}{2mc^2}\cdot\frac{1}{2mc}\vec{\sigma}\cdot\vec{\pi}\varphi$$

$$= \frac{\vec{\sigma}\cdot\vec{\pi}}{2mc}\varphi - \frac{1}{4m^2c^3}(i\hbar\frac{\partial}{\partial t} + e\phi)(\vec{\sigma}\cdot\vec{\pi})\varphi .$$

Substituting in the first equation yields:

$$i\hbar\frac{\partial\varphi}{\partial t} \cong c(\vec{\pi}\cdot\vec{\sigma})\left\{\frac{\vec{\pi}\cdot\vec{\sigma}}{2mc}\varphi - \frac{1}{4m^2c^3}(i\hbar\frac{\partial}{\partial t} + e\phi)(\vec{\sigma}\cdot\vec{\pi})\varphi\right\}$$
$$- e\phi\varphi$$

$$= \frac{1}{2m}(\vec{\sigma}\cdot\vec{\pi})(\vec{\sigma}\cdot\vec{\pi})\varphi - e\phi\varphi - \frac{1}{4m^2c^2}\vec{\sigma}\cdot\vec{\pi}(i\hbar\frac{\partial}{\partial t} + e\phi)\vec{\sigma}\cdot\vec{\pi}\varphi .$$

The last term yields the first relativistic correction. It is a straightforward matter to multiply it out. If we divide the terms as we did with $(\vec{\sigma}\cdot\vec{B})(\vec{\sigma}\cdot\vec{C})$, we find on a little reduction

$$(1 + \frac{1}{4m^2c^2}\pi^2)(i\hbar\frac{\partial\varphi}{\partial t} + e\phi\varphi)$$

$$= \left\{\frac{1}{2m}\pi^2 + \mu_B\vec{\sigma}\cdot\vec{H} + \mu_B\frac{1}{2mc}(\vec{E}\times\vec{\pi})\cdot\vec{\sigma}\right.$$
$$\left. - i\mu_B\frac{1}{2mc}\vec{E}\cdot\vec{\pi}\right\}\varphi$$

The factor multiplying $i\hbar \frac{\partial \varphi}{\partial t} + e\phi\varphi$ in brackets is just the result of the mass change of the particle with velocity. The term $\mu_B \frac{1}{2mc} (\vec{E} \times \vec{\pi}) \cdot \vec{\sigma}$ is the celebrated Thomas correction, which we have already discussed. The final term was first discovered by Darwin and had no place in the original Pauli spin theory. It is purely imaginary, and therefore means that the hamiltonian which we use is not hermitian. This is, however, not a contradiction since this is only for a two component function φ, while all averages must actually be ocmputed with the correct 4 rowed ψ. When the latter is done, all averages do turn out real as they must, and the Darwin term makes a significant contribution. It actually is possible to write down an equivalent hamiltonian which requires only 2 x 2 matrices for all averages. This has been done by R. Becker (Göttingen Nachtrichten, 1944) and by Foldy and Wouthuysen (Phys.Rev. 78, 29 (1950)) but we shall not go into it here.

The Relativistic Theory of the Hydrogen Atom

We shall now consider the predictions which the Dirac theory makes with respect to the energy levels of an electron placed in ~~a~~ the ~~central~~ field of a nucleus. That is, we shall discuss the Dirac equation for the case of $\vec{A} = 0$, $\phi = Ze/r = \phi(r)$. We will look for the stationary states, that is, we put

$$i\hbar \frac{\partial \psi}{\partial t} = E\psi$$

The Dirac equation becomes

$$E\psi = (c(\vec{\alpha} \cdot \vec{p} + mc\beta) - e\phi(r))\psi .$$

Now just as in the nonrelativistic case it is convenient to go over to spherical coordinates, since we have a central field. This may be done in the following way. Define a quantity α_r by

$$\alpha_r = \frac{\vec{\alpha}\cdot\vec{r}}{r} .$$

Then, clearly, since $\frac{\vec{r}}{r}$ is a unit vector $\alpha_r^2 = 1$.

Thus we may write

$$(\vec{\alpha}\cdot\vec{p}) = \alpha_r^2(\vec{\alpha}\cdot\vec{p}) = \alpha_r(\alpha_r(\vec{\alpha}\cdot\vec{p}))$$

$$= \alpha_r \frac{1}{r}(\vec{\alpha}\cdot\vec{r})(\vec{\alpha}\cdot\vec{p})$$

However, for any two vectors $\vec{B}$ and $\vec{C}$ (which do not necessarily commute with each other, but which do commute with $\vec{\alpha}$) we have

$$(\vec{\alpha}\cdot\vec{B})(\vec{\alpha}\cdot\vec{C}) = \alpha_x^2 B_x C_x + \cdots$$

$$+ \alpha_x\alpha_y B_x C_y + \alpha_y\alpha_x B_y C_x + \cdots$$

$$= \vec{B}\cdot\vec{C} + \alpha_x\alpha_y (B_x C_y - B_y C_x) + \cdots$$

Define a matrix vector $\vec{\Sigma}$ given by

$$\Sigma_x = \frac{1}{i}\alpha_y\alpha_z = \left(\begin{array}{c|c} \sigma_x & 0 \\ \hline 0 & \sigma_x \end{array}\right),$$ etc. Then

$$(\vec{\alpha}\cdot\vec{B})(\vec{\alpha}\cdot\vec{C}) = \vec{B}\cdot\vec{C} + i\vec{\Sigma}\cdot(\vec{B}\times\vec{C}) .$$

$\vec{\Sigma}$ is clearly just the 4x4 generalization of the 2x2 spin matrix vector $\vec{\sigma}$. Applying this to our case we obtain

$$(\vec{\alpha}\cdot\vec{r})(\vec{\alpha}\cdot\vec{p}) = \vec{r}\cdot\vec{p} + i\vec{\Sigma}\cdot(\vec{r}\times\vec{p}).$$

However, we also have that the orbital angular momentum $\vec{L}$ is given by $\vec{r}\times\vec{p}$, so that finally

$$(\vec{\alpha}\cdot\vec{r})(\vec{\alpha}\cdot\vec{p}) = \vec{r}\cdot\vec{p} + i\vec{\Sigma}\cdot\vec{L}$$

Using this we have

$$\vec{\alpha}\cdot\vec{p} = \alpha_r\frac{1}{r}(\vec{r}\cdot\vec{p}) + i\alpha_r\frac{1}{r}(\vec{\Sigma}\cdot\vec{L})$$

$$= \alpha_r\frac{1}{r}(\vec{r}\cdot\vec{p} - i\hbar) + i\alpha_r\frac{1}{r}(\vec{\Sigma}\cdot\vec{L} + \hbar)$$

$$\vec{\alpha}\cdot\vec{p} = \alpha_r p_r + i\hbar \frac{1}{r} \alpha_r \beta k$$

where $p_r = \frac{1}{r}(\vec{r}\cdot\vec{p} - i\hbar)$, which is our old definition of the radial momentum, and

$$\hbar k \equiv \beta(\vec{\Sigma}\cdot\vec{L} + \hbar).$$

The reason for choosing the factor β in the definition of k is that the resulting quantity is a constant of the motion. To see this we must show that $(\mathcal{H}, k) = 0$

$$\mathcal{H} = c(\vec{\alpha}\cdot\vec{p} + mc\beta) - e\phi(r)$$

Since $\vec{L}$, $\vec{\Sigma}$ and β commute with $\phi(r)$, we have

$$(\mathcal{H}, k) = c(\vec{\alpha}\cdot\vec{p}, \beta(\vec{\Sigma}\cdot\vec{L} + \hbar))$$

$$= c\{(\vec{\alpha}\cdot\vec{p}, \beta(\vec{\Sigma}\cdot\vec{L})) + \hbar(\vec{\alpha}\cdot\vec{p}, \beta)\}$$

$$= c\{[(\vec{\alpha}\cdot\vec{p})(\vec{\Sigma}\cdot\vec{L}) + (\vec{\Sigma}\cdot\vec{L})(\vec{\alpha}\cdot\vec{p})]\beta + 2\hbar(\vec{\alpha}\cdot\vec{p})\beta\},$$

since β anticommutes with $\vec{\alpha}$ and of course commutes with $\vec{\Sigma}$.

But by direct calculation (since $\vec{p}\cdot\vec{L} \equiv 0$)

$$(\vec{\alpha}\cdot\vec{p})(\vec{\Sigma}\cdot\vec{L}) = i\,\vec{\alpha}\cdot(\vec{p}\times\vec{L})$$

$$(\vec{\Sigma}\cdot\vec{L})(\vec{\alpha}\cdot\vec{p}) = i\,\vec{\alpha}\cdot(\vec{L}\times\vec{p}),$$

and therefore

$$(\vec{\alpha}\cdot\vec{p})(\vec{\Sigma}\cdot\vec{L}) + (\vec{\Sigma}\cdot\vec{L})(\vec{\alpha}\cdot\vec{p}) = i\,\vec{\alpha}\cdot(\vec{p}\times\vec{L} + \vec{L}\times\vec{p})$$

But $\vec{p}\times\vec{L} + \vec{L}\times\vec{p} = \frac{-2\hbar}{i}\vec{p}$, as is easily verified by writing out the components. Therefore

$$(\mathcal{H}, k) = c[-2\hbar(\vec{\alpha}\cdot\vec{p})\beta + 2\hbar(\vec{\alpha}\cdot\vec{p})\beta] \equiv 0.$$

In order to see the physical significance and the characteristic values of this constant, it is necessary to consider the total angular momentum of our system. Since the particle has spin 1/2 we would expect that not the orbital angular momentum $\vec{L}$ but the orbital plus spin angular momentum will be conserved. Define

$$\vec{J} = \vec{L} + \frac{1}{2}\hbar\vec{\Sigma} ,$$

We shall now show that this is a constant of the motion.

$$[\mathcal{H}, J_x] = [\, c(\vec{\alpha}\cdot\vec{p}) + mc^2\rho - e\phi ,\; L_x + \tfrac{1}{2}\hbar\Sigma_x]$$

$$= c\,[(\vec{\alpha}\cdot\vec{p}),\; L_x + \tfrac{1}{2}\hbar\Sigma_x]$$

$$= c\,[(\vec{\alpha}\cdot\vec{p}), L_x] + \frac{\hbar c}{2}[(\vec{\alpha}\cdot\vec{p}), \Sigma_x]$$

$$= \frac{\hbar c}{i}[\alpha_y p_z - \alpha_z p_y] + \frac{\hbar c}{i}[p_y\alpha_z - \alpha_y p_z] \equiv 0 ,$$

and similarly for the other two components. Thus $\vec{J}$ is a constant, and of course so is J^2.

$$J^2 = (\vec{L} + \tfrac{1}{2}\hbar\vec{\Sigma})^2 = L^2 + \hbar\,\vec{L}\cdot\vec{\Sigma} + \tfrac{3}{4}\hbar^2$$

But $(\vec{\Sigma}\cdot\vec{L})^2 = L^2 + i\,\vec{\Sigma}\cdot(\vec{L}\times\vec{L}) = L^2 - \hbar\,\vec{\Sigma}\cdot\vec{L}$, so that

$$J^2 = (\vec{L}\cdot\vec{\Sigma})^2 + 2\hbar(\vec{L}\cdot\vec{\Sigma}) + \hbar^2 - \tfrac{1}{4}\hbar^2$$

$$= (\vec{\Sigma}\cdot\vec{L} + \hbar)^2 - \tfrac{1}{4}\hbar^2 = [\rho(\vec{\Sigma}\cdot\vec{L} + \hbar)]^2 - \tfrac{1}{4}\hbar^2$$

$$= \hbar^2 k^2 - \tfrac{1}{4}\hbar^2$$

Thus k is, apart from constants, essentially the magnitude of the total angular momentum. Since $j(j+1)\hbar^2$ ~~has~~ are the characteristic values of J^2 (where $j = 1/2, 3/2, 5/2, \cdots$) , we obtain for k the values

$$\hbar^2 k^2 = \hbar^2 (j)(j+1) + \frac{\hbar^2}{4} = \hbar^2\left(j^2+j+\tfrac{1}{4}\right) = \hbar^2\left(j+\tfrac{1}{2}\right)^2$$

$$k = \pm\left(j+\tfrac{1}{2}\right) = \pm 1, \pm 2, \pm 3, \ldots$$

In terms of k the hamiltonian now takes the form

$$\mathcal{H} = c\left(\alpha_r p_r + i\hbar \frac{1}{r}\alpha_r \beta k + mc\beta\right) - e\phi(r).$$

Since we have $\alpha_r^2 = 1 \qquad \beta^2 = 1$

and $\alpha_r\beta + \beta\alpha_r = 0$

$(p_r, \alpha_r) = 0 = (k, \alpha_r)$ (direct calculation),

we may choose α_r and β to be any two numerical matrices whose squares are one and which anticommute. Let us take

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\alpha_r = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

$$\alpha_r\beta = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

$$\mathcal{H} = \begin{pmatrix} mc^2 - e\phi & c\left(-ip_r - \frac{\hbar}{r}k\right) \\ c\left(ip_r - \frac{\hbar}{r}k\right) & -mc^2 - e\phi \end{pmatrix}.$$

Put $\psi = \begin{pmatrix} f(r) \\ g(r) \end{pmatrix} \qquad p_r = \frac{\hbar}{i}\frac{\partial}{\partial r},$

then the Schrödinger equation $\mathcal{H}\psi = E\psi$ becomes

$$c\left(\hbar\frac{\partial}{\partial r} + \frac{\hbar}{r}k\right)g + \left(E - mc^2 + e\phi\right)f = 0$$

$$c\left(\hbar\frac{\partial}{\partial r} - \frac{\hbar}{r}k\right)f - \left(E + mc^2 + e\phi\right)g = 0$$

Put $\gamma_1 = \dfrac{mc^2 - E}{\hbar c}$, $\gamma_2 = \dfrac{mc^2 + E}{\hbar c}$

$$\alpha = \frac{e^2}{\hbar c} \cong \frac{1}{137}, \quad \phi = \frac{Ze}{r}.$$

($\gamma_1 > 0$ for bound states, since $E < mc^2$ for them).

$$\left(\frac{\partial}{\partial r} + \frac{k}{r}\right) g - \left(\gamma_1 - \frac{Z\alpha}{r}\right) f = 0$$

$$\left(\frac{\partial}{\partial r} - \frac{k}{r}\right) f - \left(\gamma_2 + \frac{Z\alpha}{r}\right) g = 0.$$

To solve these euqations, first consider ~~x~~ r very large. Then approximately

$$g' - \gamma_1 f = 0$$

$$f' - \gamma_2 g = 0$$

$$g'' - \gamma_1 f' = 0, \quad g'' - \gamma_1 \gamma_2 g = 0, \quad \text{or} \quad g'' - \gamma^2 g = 0,$$

where $\gamma \equiv \underset{+}{\sqrt{\gamma_1 \gamma_2}}$ is real. Similarly $f'' - \gamma^2 f = 0$.

These equations are easily solved, giving

$$g = \mu e^{-\gamma r} + \nu e^{\gamma r}$$

$$f = \bar{\mu} e^{-\gamma r} + \bar{\nu} e^{\gamma r}$$

The solutions $e^{\gamma r}$ are uninteresting, since they are non-normalizable, so that approximately for very large r

$$g \sim e^{-\gamma r}$$

$$f \sim e^{-\gamma r}$$

Let us therefore put $\gamma r = \rho$ and

$$g = e^{-\rho} G(\rho)$$

$$f = e^{-\rho} F(\rho)$$

This gives us

$$\frac{dG}{d\rho} - G + \frac{k}{\rho} G - \left(\frac{\gamma_1}{\gamma} - \frac{Z\alpha}{\rho} \right) F = 0$$

$$\frac{dF}{d\rho} - F - \frac{k}{\rho} F - \left(\frac{\gamma_2}{\gamma} + \frac{Z\alpha}{\rho} \right) G = 0 \ .$$

Now let us look for power series solutions in the form

$$F = \rho^s (a_0 + a_1 \rho + a_2 \rho^2 + \dots) \qquad a_0 \neq 0$$

$$G = \rho^s (b_0 + b_1 \rho + b_2 \rho^2 + \dots) \qquad b_0 \neq 0$$

Substituting, and putting the coefficient of ρ^{s+n-1} equal to zero we obtain

$$\left. \begin{aligned} & b_n(n+s+k) - b_{n-1} + Z\alpha\, a_n - \frac{\gamma_1}{\gamma} a_{n-1} = 0 \\ & a_n(n+s-k) - a_{n-1} - Z\alpha\, b_n - \frac{\gamma_2}{\gamma} b_{n-1} = 0 \end{aligned} \right\} n \neq 0$$

$$\left. \begin{aligned} & b_0(s+k) + Z\alpha\, a_0 = 0 \\ & a_0(s-k) - Z\alpha\, b_0 = 0 \end{aligned} \right\} n = 0 \ .$$

Therefore we must have

$$\begin{vmatrix} s+k & Z\alpha \\ -Z\alpha & s-k \end{vmatrix} = 0$$

$$s^2 - k^2 + Z^2\alpha^2 = 0$$

$$s = \pm \sqrt{k^2 - Z^2\alpha^2} = + \sqrt{k^2 - Z^2\alpha^2} \ .$$

We take the upper sign since for the lower sign the wave function would not be normalizable--it would diverge at the origin.

If we multiply our first equation by $\gamma \left(= \sqrt{\gamma_1 \gamma_2}\right)$, and the second by γ_1, and subtract we obtain

$$b_n\left(\gamma(n+s+k)+Z\alpha\gamma_1\right) = a_n\left(\gamma_1(n+s-k)-Z\alpha\gamma\right)$$

If the series do not terminate, then their behavior at infinity is determined by the coefficients with n very large. For n very large we have from the above

$$b_n n\gamma \cong a_n\gamma_1 n \quad \text{or} \quad b_n\gamma \cong a_n\gamma_1 .$$

From the original expressions though we have

$$\left.\begin{array}{l} b_n n - b_{n-1} - \frac{\gamma_1}{\gamma} a_{n-1} \cong 0 \\ a_n n - a_{n-1} - \frac{\gamma_2}{\gamma} b_{n-1} \cong 0 \end{array}\right\} n \text{ very large} ,$$

or

$$\left.\begin{array}{l} b_n \cong \frac{2}{n} b_{n-1} \\ a_n \cong \frac{2}{n} a_{n-1} \end{array}\right\} n \text{ very large} .$$

Therefore these series converge as

$$\sum \frac{1}{n!}(2\rho)^n \longrightarrow e^{2\rho}$$

Thus f or g go as $e^{-\rho}$(F or G), we will always get functions which are non-normalizable at infinity unless they terminate. Let us say they terminate for $n = N$. Then $a_{N+1} = 0$ $b_{N+1} = 0$. The condition for this to happen is that

$$-b_N - \frac{\gamma_1}{\gamma} a_N = 0$$

and

$$-a_N - \frac{\gamma_2}{\gamma} b_N = 0 ,$$

which is equivalent.

$b_N = -\frac{\gamma_1}{\gamma} a_N$. Substitution in the relationship between a_n and b_n gives

$$-\frac{\gamma_1}{\gamma}\left(\gamma(N+s+k) + Z\alpha\gamma_1 \right) = \gamma_1(N+s-k) - Z\alpha\gamma$$

$$2\gamma(N+s) = Z\alpha(\gamma_2-\gamma_1) = Z\alpha\frac{2E}{\hbar c} = \frac{2Z\alpha}{\hbar c}E \quad (\text{thus } E>0).$$

Squaring we obtain (since $\gamma = \frac{1}{\hbar c}\left((mc^2)^2 - E^2 \right)^{1/2}$):

$$\left((mc^2)^2 - E^2 \right)(N+s)^2 = (Z\alpha)^2 E^2$$

$$E^2 = \frac{(mc^2)^2 (N+s)^2}{(N+s)^2 + (Z\alpha)^2}$$

$$E = \frac{mc^2}{\underset{+}{\sqrt{1+\left(\frac{Z\alpha}{N+s}\right)^2}}}$$

$$E_{Nk} = \frac{mc^2}{\underset{+}{\sqrt{1+\left(\frac{Z\alpha}{N+\underset{+}{\sqrt{k^2-Z^2\alpha^2}}}\right)^2}}}$$

Here $N = 0, 1, 2, \ldots$

$k = \pm 1, \pm 2, \ldots$

For all $N \neq 0$, takes on $\pm$ values, for $N = 0$, however, takes on only positive values. To see why this is so we notice that there are two sets of relationships connecting a_0 and b_0 when $N = 0$. The ordinary ones for determining s:

$$b_0(s+k) + Z\alpha a_0 = 0$$

$$a_0(s-k) - Z\alpha b_0 = 0$$

and that coming from termination

$$b_0 = -\frac{\gamma_1}{\gamma} a_0$$

From the first we have

$$\frac{b_0}{a_0} = -\frac{Z\alpha}{s+k},$$

from the second

$$\frac{b_0}{a_0} = -\frac{\gamma_1}{\gamma}$$

Since $s < |k|$, the first ratio is positive for negative k, negative for positive k, while the second ratio is always negative. Therefore we must (for $N = 0$) choose $k > 0$. We can see the meaning of this and the connection with the nonrelativistic theory if we expand our energy ($(Z\alpha)^2$ is small $\sim (\frac{1}{137})^2$ for hydrogen):

$$E_{Nk} \cong mc^2 \left(1 - \frac{1}{2} \frac{Z\alpha^2}{(N+|k|)^2}\right)$$

$$= mc^2 - \frac{mZ^2e^4}{2\hbar^2(N+|k|)^2}$$

The nonrelativistic formula is

$$E_n = -\frac{mZ^2e^4}{2\hbar^2 n^2} \qquad n = 1, 2, \ldots$$

These certainly agree (apart from the rest energy mc^2) if we take

$$n = N + |k| \qquad N = 0, 1, \ldots$$
$$k = \pm 1, \pm 2, \ldots$$

Now we can see what happens to the degenerate levels with a given n, when relativistic effects are taken into account

$\underline{n = 1}$ $\quad N + |k| = 1$, but $|k| \geq 1$.

Therefore $\quad N = 0 \quad k = +1 \quad j = |k| - 1/2 = 1/2$.

That is, the state $n = 1$ which corresponds to $l = 0$ is originally nondegenerate and now acquires the spin degeneracy. Since there is only one way to add spin 1/2 to $\ell = 0$ = to give j = 1/2, we see why for $N = 0$ we only get one of the two possible k values. This is a $^2S_{\frac{1}{2}}$ state in the nonrelativistic limit.

$\underline{n = 2}$ $\quad 2 = N + |k|$. Either $N = 0,\ k = 2 \quad j = 3/2$

or $\quad N = 1,\ k = \pm 1 \quad j = 1/2$

The degeneracy for j = 1/2 comes because there are two ways of making j = 1/2: either we can add 1/2 to $l = 0$ or subtract 1/2 from $\ell = 1$. On the other hand, there is only one way of making j = 3/2, by adding 1/2 to l = 1. These states are respectively

N	k	j	ℓ	
0	2	$3/2$	1	$^2P_{3/2}$
1	-1	$1/2$	1	$^2P_{1/2}$
1	$+1$	$1/2$	0	$^2S_{1/2}$

The identification of l for the last two states proceeds as follows: If ℓ is approximately a good quantum number these (nonrelativistically $\beta \cong 1,\ \vec{\Sigma} \to \vec{\sigma}$),

$$\hbar k = \vec{L}\cdot\vec{\sigma} + \hbar = \frac{1}{\hbar}\left[(\vec{L} + \tfrac{1}{2}\hbar\vec{\sigma})^2 - L^2 - (\tfrac{1}{2}\hbar\vec{\sigma})^2 + \hbar^2\right]$$

$$k = j(j+1) - \ell(\ell+1) + \frac{1}{4}$$

$$\left.\begin{aligned} &= l+1 \quad \text{if} \quad j = l + 1/2 \\ &= -l \quad \text{if} \quad j = l - 1/2 \end{aligned}\right\} \text{ the only two possibilities.}$$

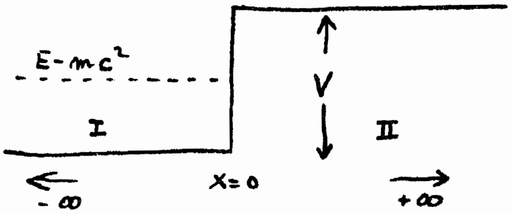
Thus when $j = 1/2 \quad k = -1 \quad l = 1$

$j = 1/2 \quad k = +1 \quad l = 0$.

Since for the last two states ${}^2P_{1/2}$, ${}^2S_{1/2}$, $n=2$ we have the same N and $|k|$, the relativistic energy will be the same, and therefore the Dirac theory predicts that these levels will be exactly degenerate. Very recently (1947) Lamb and Retherford have found that these levels are actually split by about 1000 megacycles. The explanation of this phenomenon lies outside the scope of these lectures, but we simply mention that its origin lies in the interaction of the electron with the fluctuating electromagnetic field which is always present in the vacuum.

The Klein Paradox, Negative Energy States

If we consider a Dirac electron which is reflected from a potential discontinuity, the presence of negative energy states produces some very startling results. Imagine the following situation: An electron is incident on a potential energy jump of V. The electron has kinetic energy $E-mc^2 < V$. Then in classical mechanics and in non-relativistic quantum mechanics we would expect the electron to be totally reflected.



This is so also for the Dirac electron, as long as is not too great. For $V > E + mc^2$ transmission begins to take place. We shall not treat the problem in full detail, but indicate how the solution goes. The solution in region I must be of the form

$$\psi_I = \Big(\underset{\substack{\uparrow \\ \text{incident wave}}}{A u_I e^{ipx/\hbar}} + \underset{\substack{\uparrow \\ \text{reflected wave}}}{B v_I e^{-ipx/\hbar}} \Big) e^{-iEt/\hbar}$$

In II it must have the form

$$\psi_{II} = C u_{II} e^{ip'x/\hbar} e^{-iEt/\hbar} .$$

The quantities u_I, v_I, u_{II} are normalized Dirac vectors which must be solutions of

$$(c\alpha_x p + mc^2\beta) u_I = E u_I$$

$$(-c\alpha_x p + mc^2\beta) v_I = E v_I$$

$$(c\alpha_x p' + mc^2\beta) u_{II} = (E-V) u_{II} .$$

Put $mc^2 = E_0$. Then $pc = \sqrt{E^2 - E_0^2}$

$$p'c = \sqrt{(E-V)^2 - E_0^2} .$$

Since only the matrices α_x, β appear we can satisfy the commutation rules with 2 x 2 matrices and we need only 2 component functions (the others would just give the same results again).

Take

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \alpha_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Then direct solution gives

$$u_I = N(p) \begin{pmatrix} 1 \\ \frac{cp}{E+E_0} \end{pmatrix} \qquad N(p) = \frac{1}{\sqrt{1 + \frac{c^2p^2}{(E+E_0)^2}}}$$

$$v_I = N(p) \begin{pmatrix} 1 \\ \frac{-cp}{E+E_0} \end{pmatrix}$$

$$u_{II} = N'(p') \begin{pmatrix} 1 \\ \frac{cp'}{E-V+E_0} \end{pmatrix} \qquad N'(p') = \frac{1}{\sqrt{1 + \frac{c^2p'^2}{(E-V+E_0)^2}}}$$

Matching boundary conditions we have

$$\psi_I = \psi_{II} \quad \text{at} \quad x = 0,$$

and therefore

$$(A+B)\,N(p) = C\,N'(p')$$

$$(A-B)\,\frac{c\,p\,N(p)}{E+E_0} = C\,\frac{c\,p'\,N'(p')}{E-V+E_0}$$

Dividing, we obtain

$$\frac{A+B}{A-B} = \frac{p}{p'}\,\frac{E-V+E_0}{E+E_0}$$ n

$$\frac{B}{A} = \frac{\frac{p}{p'}\,\frac{E-V+E_0}{E+E_0} - 1}{\frac{p}{p'}\,\frac{E-V+E_0}{E+E_0} + 1}.$$

The reflection coefficient may be defined as the ratio of reflected to incident intensity, that is

$$R \equiv \left(\frac{B}{A}\right)^2.$$

This expression is in general a very complicated function of p, but to show the essential property we want, let us consider the extreme limit of $V \to \infty$. Then $cp' \longrightarrow -V$ (Closer investigation* shows we must take the negative root), and

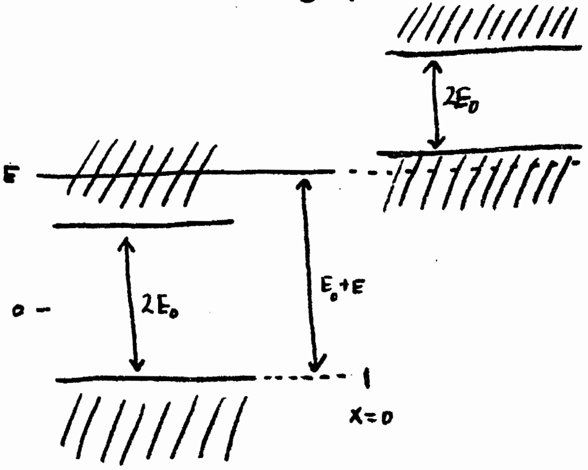
$$\frac{B}{A} \longrightarrow \frac{\frac{pc}{E+E_0} - 1}{\frac{pc}{E+E_0} + 1} = \frac{pc - E - E_0}{pc + E + E_0}$$

$$R = \left(\frac{B}{A}\right)^2 = \frac{E - pc}{E + pc} < 1.$$

---

* Going back to the equation for u_{II} in this limit.

Thus we see that if the potential barrier is infinite we still get transmission! This is the famous Klein paradox. A closer investigation shows that as soon as $V > E + E_0$ transmission takes place. This can easily be understood in terms of the negative energy states. Consider first the case of no barrier. Then electrons will be transmitted if $E > E_0$ or if $E < -E_0$. The shaded region gives possible states which allow transmission. Now suppose that $V = E_0 + E$. Then the lower band is raised so that $E > 0$ lies in it, and once again we get transmission, but <u>via an electron in a negative energy state</u>. Thus the negative energy states produce in fields which are strong enough, effects which are certainly in contradiction with experience.



Another effect, which has been calculated by Oppenheimer, is that a hydrogen atom would exist for only about 10^{-10} seconds, the decay taking place by the electron dropping into a negative energy state and emitting a light quantum. This is also in contradiction with experiment. The question arises, what can we do about these negative energy states, since clearly they are not real physically? Dirac has given an extremely beautiful interpretation, which we sketch very roughly here.

Suppose we consider the Dirac equation

$$i\hbar \frac{\partial\psi}{\partial t} = \left\{ c\left[\vec{\alpha}\cdot\left(\vec{p} + \frac{e}{c}\vec{A}\right) + mc\beta\right] - e\phi \right\}\psi = E\psi .$$

Let us take a representation in which $\vec{\alpha}$ is real β purely imaginary. This can clearly be done by just taking our previous representation and interchanging β and α_y. Now take the complex conjugate* of this equation

$$E\psi^* = \left\{ c\left[\vec{\alpha}\cdot\left(-\vec{p}+\frac{e}{c}\vec{A}\right) - mc\beta\right] - e\phi \right\}\psi^* \quad \text{or}$$

$$-E\psi^* = \left\{ c\left[\vec{\alpha}\cdot\left(\vec{p}-\frac{e}{c}\vec{A}\right) + mc\beta\right] + e\phi \right\}\psi^*$$

Now if we consider a solution ψ corresponding to a negative energy E we see that (in this representation) ψ^* corresponds to the same absolute value of the energy but with a plus sign, and for a particle with the same mass and opposite charge. Thus the states associated with negative energy really correspond to states of positive energy for particles with the opposite charge. Such particles are found in nature and are known as positrons. We are still not out of the difficulty, because we would still have to find a way for electrons to be prevented from going over to positron states. Dirac solved this very simply, by saying that normally all positron states are filled up and therefore by the exclusion principle we cannot make transitions to them. Positrons then appear when an electron has been knocked out of one of these states. We see that to obtain a consistent picture of the Dirac equation, we have been driven to a many body theory. This is characteristic of all relativistic quantum mechanics, and can really be well formulated only if we use the quantum theory of wave fields. If we do, it is possible to build up a theory of electrons and positrons which is very natural and in complete agreement with experiment.

---

* Not the complex conjugate transpose. Here ψ^* means $-i$ replaces i everywhere in ψ.

MECANIQUE STATISTIQUE

Notes on lectures by

Prof. L. ROSENFELD

Department of Physics
University of Manchester

Delivered at the Ecole d'Eté de
Physique Théorique, Les Houches, FRANCE
28 Juillet - 28 Août 1952

INTRODUCTION

1.- Aim and scope of statistical thermodynamics.

The aim of "statistical thermodynamics" is to give an atomistic interpretation of heat phenomena. In the first phase of its historical development, the atomic conception was restricted to the structure of ponderable matter and the atomistic treatment of thermal properties of matter accordingly became known as "statistical mechanics". Later on, however, the phenomenological concepts and laws of thermodynamics were extended to radiqtion phenomena, and the interpretation of these phenomena in terms of electromagnetic fields provided the basis for a statistical treatment of the thermodynamics of radiation on quite the same lines of that of matter. It seems therefore advisable to replace the traditional denomination of "statistical mechanics" by a name which indicates more correctly the wider scope of the theory.

In its broadest aspect, the atomistic view of matter and radiation consists in regarding these physical agencies as composed of elements - the atoms (or molecules) of matter and the monochromatic waves of electromagnetic radiation - obeying well-defined laws of motion or propagation. The physical quantities describing the properties of macroscopic observation are then interpreted as average values of corresponding atomic, or elementary, quantities, taken over large numbers of constituent elements. Macroscopic relations between physical quantities can then in principle be derived, by suitable averaging processes, from the elementary laws of motion or propagation. In itself, the taking of averages involves only a trivial use of statistical conceptions. But just in the analysis of heat phenomena such conceptions are found to play a quite essential part and their introduction into the atomistic picture thus appears as a deep-lying feature of the whole theory.

2.- Statistical aspect of the second law of thermodynamics.

The central problem is the interpretation of the second law of thermodynamics. This law can be conveniently analysed into two statements, relating to controlled (quasi-statisc) and spontaneous transformations respectively :

(a) there is a quantity called entropy which belongs to each state of thermal equilibrium state of the system in a controlled way is a well defined function of the thermal parameters characterizing the transformation ;

(b) if a thermally isolated system undergoes a spontaneous transformation from one state of equilibrium to another, its entropy increases, thus indicating the irreversible character of the transformation.

Now, the essentially statistical aspect of entropy and thermodynamical irreversibility is well illustrated by such processes as the mixing of two gases. For the increase of entropy in this case, when considered from the atomistic point of view, can only appear as a function of the numbers of atoms involved in the mixing processes : in fact, it is simply related to the greater probability of the disordered state represented by the mixture, as compared with that of the separated gases. But more generally, it is found that heat in its various manifestations must be regarded as the result of some disordered form of motion or propogation of the elementary constituents. Entropy is a statistical measure of this disorder, and its increase expresses a transition to a state of greater probability.

However, it is important to realize that entropy is a statistical concept of a very specific kind : it expresses a definite relation between statistical and thermal characteristics of the system. In other words, the statistical element is introduced into the definition of entropy by the fact that one considers only thermal transformations of the system performed in a quasi-static, i.e. macroscopically controlled, way. This physical aspect of the entropy concept is not always properly appreciated in the so-called theory of information. Entropy is indeed a measure of a certain type of "information" pertaining to the microscopic state of the system ; but it only refers to that information which can be obtained or lost by operations involving exchanges of heat between the system and other bodies (or, in particular, keeping the system thermally isolated). The statistical aspect of thermodynamics thus arises from the fact that the definition of the macroscopic states of the system entails a limitation of our control of the microscopic behaviour of its elementary constituents.

It is just this circumstance which leads to the irreversible thermal evolution of an isolated system when its atomistic elements are left to follow the course determined by their own laws. The derivation of this fundamental property is the oblect of the " ergodic theorems" which forms the backbone of statistical thermodynamics. Asymptotic irreversibility is a very general feature of statistical "chains" of events : it is exhibited by results of lotteries just as well as by assemblies of atoms. It is especially remarkable in the latter case, however, in view of the reversible character of the elementary laws of motion : the conciliation (in a statistical sense) of macroscopic irreversibility with microscopic reversibility is the essential achievement of the statistical analysis of the second law of thermodynamics by means of the ergodic theorems.

3. Statistical thermodynamics and quantum theory.

In the preceding general outline of statistical thermodynamics, the exact nature of the atomic systems and radiation fields considered has been left open. It is well-known that, when one attempted to describe them in terms of ordinary mechanics and electromagnetic theory, the resulting thermodynamics led to serious inconsistencies : this was in fact the origin of the discovery of the quantum of action and of the ensuing development of the quantum theory of both radiation fields and atomic systems.

It is clear, therefore, that in order to obtain a consistent formulation, we must take as basic assumption the quantal description of the elementary constituents. But it should be equally clear that this does not make the least difference for the fundamental argument of statistical theory, outlined above. Since the quantal laws, just as the classical ones, are essentially reversible with respect to time, the problem of deriving the macroscopic irreversibility by introducing a suitable statistical element into the theory remains unchanged and is again solved by ergodic theorems.

The present course will take as its leit-motiv the complete parallelism, in regard of the general structure of the theory, between classical and quantal statistics of atomistic systems. A special emphasis on this aspect of the subject is needed, because it has not always been clearly recongized. The issue has been obscured by the fact that quantum theory itself, in contrast to classical theory, introduces a statistical element at the miscroscopic level ; and it has sometimes been confusendly argued that it is the elementary quantal statistics which provides the basis of macroscopic irreversibility. In reality, we have here two completely distinct statistical features, which are not only logically independent of each other, but also without physical influence upon each other. The question whether the elementary law of change is deterministic (as in classical physics) or statistical (as in quantum theory) is entirely irrelevant for the validity of the ergodic theorems.

The thermodynamical irreversibility is thus related, not to any statistical feature of the elementary systems, but to another, quite independent, statistical feature, which specifies the point of view of the macroscopic pbserver. It is interesting to note that the logical relationship between the macroscopic and microscopic modes of description of the system is one of complementarity, in the sense that the two points of view are mutually exclusive : a complete knowledge of the microscopic state of the elementary constituents precludes the use of such macroscopic concepts as temperature and entropy ; conversely, the macroscopic description implies a statistical indeterminacy in the knowledge of the microscopic state. But this complementarity must be carefully distinguished from the similar relation between conjugate quantities pertaining to the elementary quantal systems.

While it is important to realize the essential identity of structure of classical and quantal thermodynamics, it is no less necessary to keep in mind the main points of difference between the two theories :

(a) as already stated, the concrete consequences of quantal thermodynamics, while containing those of classical theory as limiting cases, are free from the inconsistencies of the latter. The law of equipartition of energy, which is the source of these inconsistencies, is in fact reduced to a law of asymptotic validity. There are well-known examples of systems whose behaviour even under ordinary experimental conditions illustrates extreme cases of quantal statistics : the pure radiation field on the one hand, the metallic state of matter on the other.

(b) There is in quantum theory a type of irreversible process unknown to classical theory : the measuring process. The measurement of a certain quantity pertaining to a system may involve an irreversible change of state of the system, with a corresponding increase of entropy.

(c) From a more formal point of view, the development of statistical thermodynamics on the basis of the quantal laws is neater than on the assumptions of classical theory for the elementary systems. The proof of the ergodic theorems, in particular, is somewhat easier. But the main advantage of quantum theory is that it makes possible, in a much more compendious way than classical theory, a unified treatment of both matter and radiation (and in fact of any physical system whatsoever). This is, of course, a consequence of the elimination of the dualism between matter and force and its replacement by the conception of the complementary manifestations of particle and field for every constituent of the physical world.

:.- Outline of the fundamental argument.

In order to bring out the similarities and differences just enumerated between classical and quantal thermodynamics, it will be necessary to present the fundamental argument successively from each point of view. This we shall try to do with a minimum of repetition by first giving a full treatment of classical theory and then going more rapidly over the argument again on the basis of quantum theory, just stressing the points of agreement and difference.

The argument may be conveniently divided into three parts :

(1) : The first part, which is of a preparotory character, is concerned with isolated systems. The systems of actual interest are usually not isolated, but they can always be conceived as embedded in a larger system, which may be treated as isolated to any desired approximation (just by making it large enough). Once the statistical treatment of isolated systems is developped, that of actual physical systems in interaction with their surroundings can be derived from it by studying the behaviour of a small part of an isolated system.

The interactions of a physical system composed of a large number of elementary constituents with its surroundings are of two kinds ; there are dynamical interactions, involving exchange of energy, momentum and angular momentum; there is also a flow of the atomic constituents themselves into and out of the system, with an accompanying exchange of mass and perhaps electric charge or other substantial properties. Strictly speaking, the two types of exchange cannot be separated; in particular, according to quantum theory, no physical system in interastion with others can be regarded as <u>chosed</u>, i.e. as having a fixed number of elementary constituents. However, the concept of closed system is, of course, quite a useful approxilation in many cases, and it is convenient to study sush systems first.

(2) The second part of the theory will accordingly be devoted to <u>closed systems</u> in dynamical interaction with their surroundings. It will be sufficient to restrict the analysis to the exchange of energy, which is the only one of interest for the theory of heat, and which is typical of the kind of treatment to be applied to other dynamical processes when necessary. It is, in fact, by considering energy exchanges that the concept of temperature can be defined : if we want to give any system a definite temperature, we must bring it in thermalcontact with a "thermostat", i.e. allow it to exchange energy with a system of sufficiently large heat capacity to maintain its temperature constant to any desired approximation. We may then regard the total system formed by our closed system and the thermostat as an isolated system. From the statistics of the total system that of the closed system is then derived by eliminating the atomistic variables pertaining to the thermostat and reducing the representation of the latter to a single macroscopic parameter, which plays the part of the temperature.

(3) The method just oulined for taking account of the dynamical interactions of closed systems with the surroundings can immediatly be extended to the exchanges of elementary constituents between the system and its surroundings.. This is the third and last stage of the theory; it is concerned with <u>open systems</u>, i.e. systems with an indefinite number of elements. The procedure consists in treating an open system as a small part of a closed system, which forms so to speak a reservoir of elements of all kinds taking part in the exchange. The elimination of the variables pertaining to the reservoir yields a statistical distribution

function for open systems, containing macroscopic parameters related to the average numbers of elementary constituents present in the systems.

Let us know take up in succession the three stages of the theory and consider the argument a little more closely in each of them.

5. - Isolated systems : the ergodic theorems.

Consider the evolution of an isolated system in the course of time from a certain instant at which all connections with the surroundings have been severed. We know that after a period of adjustment, or "relaxation", it will tend to a stationary state in which it will persist indefinitely. From the atomistic point of view, the value of any macroscopic quantity pertaining to the system will be interpreted as the average of a definite microscopic qiantity over the interval of time necessary for the measurement. Such a time interval may be extremely short in the macroscopic scale and yet cover the passage of the system through a large number of successive microscopic configurations : it is then called "physically infinitesiaml". In the present case, the time average will be the same over any interval, except perhaps during the period of relaxation : the macroscopic value pertaining to the stationary state of the system may thus be represented by an average extending over an infinite time, an idealized concept which turns out to be more amenable to a general analysis. In fact, this average over infinite time embodies all the microscopic configurations actually taken by the system but is obviously independent of the order in which they have been successively taken : it is therefore equivalent to a definite <u>statistical</u> average, i.e. an average defined by a distribution law indicating the relative frequencies of occurence of the various configurations of the system in the course of its evolution. Such a statistical distribution law, fromwhich the time has disappeared completely, characterizes the stationnary state of the isolated system. The derivation of the distribution law for the isolated system is thus the fundamental problem of statistical thermodynamics.

The above argument, however, is incomplete in as much as it takes for granted the empirically observed irreversible tendency of the isolated system to a stationnary state. This behaviour, as already pointed out, is by no means obvious in view of the reversibility of the elementary law governing the change of configuration of the system with time. The first task is therefore to define the precise meaning of the macroscopic irreversibility, i.e. to find the conditions under which the time average of any quantity pertaining to the system exists independently of its initial state. This is the object of the <u>ergodic theorems</u>. It then turns out that the solution of the second problem, viz. the determination of the statistical average equivalent to the time average, is an immediate consequence of these theorems.

At this stage, it will be necessary for the first time to make a distinction between classical and quantal systems. We must first describe more accurately in each case what we have hitherto loosely called a "configuration" of the system.

A configuration or <u>phase</u> of a classical system at any time is given by a set of values (p_t, q_t) of the 2f canonical variables describing the systam. Such a set corresponds to a point in a Cartesian 2f-dimensional phase space.

The temporal evolution of the system from any given initial phase is represented by a well-defined trajectory in phase space, corresponding to a unique solution of the Hamiltonian equations of motion.

This trajectory remains on a surface of constant energy in phase space : this expresses the condition that the system be isolated. Morever we shall make the essential assumpion that all the phases of our

A configuration or <u>state</u> [+] of a quantal system at any time t is described by a wave-function $\Psi(q,t)$ depending on a set of variables q. This wave-function corresponds to a vector in a functionnal Hilbert space.

The temporal evolution of the system from any given initial state is represented by a well-defined precession of the state vector in Hilbert space, corresponding to a unique solution of the Schrodinger wave equation.

Owing to thenperturbation involved in the operation of isolating the system, the initial state -

---

[+] To avoid confusion with the macroscopic states of the system we shall when necessary speak of <u>atomistic states</u>.

system are confined to a finite region of phase space.

vector will in general not coincide with any of the eigen-vectors of the energy of the isolated system, but will be expressed as a superposition of these eigen-vectors with definiteamplitudes and phases.

Birkhopp's fundamental ergodic theorem states that the time average of any quantity exists for <u>almost all</u> trajectories on any energy surface. The condition for the time-average to be independent of the initial phase is expressed by a certain topological property of the trajectory, which can be very roughly described by saying that the trajectory is not confined to a part of the energy surface but in a certain sense covers the whole of its surface.

It is easy to see that the time-average of any quantity exists in any state of the system. The condition for the time-average to be independent of the initial phases is that there be no degeneracy of the energy of the system.

This condition mainly excludes the existence of other uniform integrals of the motion of the system than the energy

A system, for which the above condition is fulfilled for almost all trajectories will therefore exhibit the characteristic irreversible behaviour of macroscopic experience, provided that the determination of the initial state in which it is isolated from its surroundings, is sufficiently inaccurate to prevent the exceptional trajectories from "showing up". In other words, we are not

A system for which the above condition is fulfilled will therefore asymptotically exhibit the behaviour of a statistical mixture of its stationary states (i.e. eigenstates of its energy). The coefficients of the mixture are the squares of the amplitudes of the corresponding eigenvectors in the initial state : the statistical mixture is thus seen to arise from the act of isolating the system interested in following up the system along any particular trajectory, but as part of the definition of the macroscopic (initial) state of the system we introduce a statistical uncertainty in the initial phases. The macroscopic time-average is obtained as a result of this additional averaging over the initial phase, which has the effect of wiping out the possibility of finding the system just on an exceptional trajectory.

Since any phase can of course be regarded as a possible initial one, we must extend the statistical uncertainty of phase determination to the whole region of phase-space occupied by the trajectories. We thus see that the statistical average equivalent to the time-average must not be conceived as taken over the energy surface as a mathematically continuous multiplicity of phases but over a coarse subdivision of this surface into finite "cells", representing the physically infinitesiaml domains allowing for the uncertainty of phase determination. The ergodis law of distribution which characterizes the isolated system, and which is immediately derived from Birkhoff's theorem, actually refers to this coarse distribution on the energy surface.

from its surroundings, and thus to introduce the macroscopic element in the definition of the isolated system.

One may say that this corresponds to a certain "coarseness" in the determination of the energy of the isolated system. The ergodic law of distribution of an isolated quantal system, which is directly given by the above statistical mixture, actually refers to this coarse distribution over the energy of system.

Of course, the particular way in which the coarseness is introduced (the mode of subdivision of the energy surface into "cells") is largely arbitrary. For purposes of calculation, it will even be allowed to ignore it, i.e. to treat the cells as mathematically infinitesimal; but it is essential to remember that one is actually dealing with a coarse distribution in order to cope successfully with all paradoxes: raised by the concept of irreversibility.

Of course, the particular way in which the coarseness in introduced (the choice of the amplitudes of the energy eigenvectors in the expression for the initial state vector) is largely arbitrary. It will not enter into any actual calcualtion, but though purely formal its introduction is quite essential for the consistency of the argument.

A convenient way of expressing the coarseness of the energy determination is to divide the succession of eigenvalues of the energy into neighbouring groups, or "shells" all members of which are attributed the same weight (i.e? amplitude.)

The ergodic theorem just discussed is so to speak the core of statistical mechanics. As already stated, the distribution laws for closed systems (and open) are ultimately based upon it. But as it stands, it does not quite cover all cases of isolated systems which may be of interest and it needs an extension for this purpose both in the classical and in the quantal case. It will be noticed that although the argument in the two cases run quite parallel to each other, there is in the final step a certain discrepancy.

The classical treatment of the isolated system leads to a "coarse" but detailed distribution among the phases of a sharply defined energy surface.

in the quantal treatment of the isolated system, a statistical distribution over the energies is obtained, which gives no information whatever over any distribution over "phases" defined by the dynamical variables of the system.

It is possible, however, to restore the parallelism completely and to establish a <u>second ergodic theorem</u> which has now the same form in the two cases . To this end,

let us, in the classical case, extend the definition of a (macroscopically) isolated system by allowing a certain "coarseness" in the definition of its energy also. Instead of considering a "cell" of initial phases on an energy surface this amounts to considering such a cell in an energy "shell", i.e. in a domain of phase space comprised between two neighbouring energy surfaces. <u>The second ergodic theorem</u> due to Hopf, states the conditions for the existence of a time-average under these more general circumstances. For almost all pairs of energy surfaces within the shell we must have ergodicity not only on each of them, but also on the composite surface formed by them in the 4f-dimensional phase space obtained by taking together two identical sets of variables $(q^{(1)}, p^{(1)}; q^{(2)}, p^{(2)})$ Moreover, the existence of the asymptotic time average must now be understood in the sense of convergence in the mean : what we can prove is only that the time average of the fluctuations tends to zero as the time interval over which it is taken increases indefinitely.

we must try, in the quantal case, to define an analogue to the cell subdivision of phase space of the classical theory. The notion of phase space has of course no meaning in quantum theory, since the q's and p's are not commutable. But it is always possible to construct some set of quantities Q, P which do commute between themselves and also with the energy, and which can thus be used to define a subdivision of the energy shells already introduced into smaller"celles". We may then study the fluctuations in the course of time of the expectation values of the quantities Q,P (or any function of them). As shown by Von Neumann, an ergodic theorem entirely similar to Hopf's can be enunciated for these fluctuations : their time average over increasingly large intervals tends asymptotically to zero.

The value around which the fluctuations occur, and which thus represents the asymptotic time average of the quantity considered, appears in the form of a very simple statistical average : it corresponds to a uniform coarse distribution over the cells of an energy shell ; i.e. the relative statistical weights of these cells are given by

their volumes in phase space.	their degrees of degeneracy with respect to the energy.

We thus arrive at the following general picture of the statistical behaviour of isolated systems. The macroscopic concept of isolating the system from its surroundings is represented by the introduction of a "coarse" subdivision of "phase space" into cells. The precise mode of subdivision is irrelevant ; the only characteristic of the cells entering into the fianl result is their weight, defined either by their classical phase space volume or their quantal energy degeneracy. The second ergodic theorem then asserts the equivalence (in the sense just explained) of the asymptotic time average with a statistical average, corresponding to a uniform coarse distribution over the cells of an energy shell.

It is clear that the distribution law derived from the first ergodic theorem is a limiting case of the law just enunciated. Although the latter is both simpler and more general, the former nevertheless finds a more direct application in important cases, and it is essential to keep in mind the distinction, sometimes not clearly realised, between the first and second ergodic theorems.

6.- Closed systems : the canonical distribution.

A closed system, interacting with a "thermostat" is not limited to any energy shell ; its statistical distribution law will extend over the whole domain of its possible atomistic states, and will in particular assign definite statistical weights to the different energy shells. The mechanism by which the closed system isdistributed in energy is its exchange of energy with the thermostat, and the resulting distribution law will thus depend in some way on quantities describing the state of the thermostat. Our aim will now be to show that the thermostat is represented in the statistical law for the closed system by just one parameter, whose physical significance is directly related to the concept of temperature.

The argument, due to Boltzmann, is quite straightforward : the total system consisting of the closed system and the thermostat is regarded as isolated and its ergodic law of distribution, as given by the <u>first</u> ergodic theorem, is considered. Although it is quite essential for the whole argument that there should exist an interaction between the closed system and the thermostat, it is immaterial how large this interaction is; it is therefore permissible to assume it to be infinitesiaml in comparison with the energy content of both the closed system and the thermostat. The energy of the total system is thus the sum of the energies of the two component parts; the ergodic distribution of the total system accordingly takes the form of an integral over all divisions of the total energy into two parts, and to each such division corresponds a statistical distribution which is a product of two factors, pertaining to the closed system and the thermostat, respectively. The form of the"law of composition" of the distribution in energy of the two components of our system allows a direct application of the central limit theorem of the calculus of probabilities, which yields the asymptotuc form of the distribution of energy of any small part of the total system. This has the well-known expression called the Boltzmann or canonical distribution ; the thermostat is represented in it only by a parameter, called the modulus of the distribution; which is related to the energy of the total system and is interpreted as the absolute temperature (in dynamical measure). Moreover, the distribution depends on the macroscopic parameters defining the external conditions to which the closed system is subjected (e.g. the volume of the container if it is a gas).

The knowledge of the distribution law suffices, in principle to derive all the macroscopic properties of the closed system in terms of its temperature and external parameters. In fact, it is a well-known consequence of the two laws of thermodynamics that there exists a <u>characteristic function</u> of any set of macroscopic cariables, representing the various physical aspects of the system from which all the other quantities can be derived. If the set of variables consists of the temperature and external parameters of the extensive type (such as the volume), the characteristic function is the <u>free energy</u> of the system. Its expression in terms of the canonical distribution law is readily established; the procedure being to verify that the proposed statistical expression has the characteristic properties of the free energy, i.e. that its variation in a quasi static transformation of the sytem has the expected relation to the work done and heat exchanged in the transformation. In isothermal transformations, the variations of free energy gives directly the amount of work yielded by the system, a quantity which can be immediately expressed in terms of statistical averages. The case of a diabatic transformations is more delicate, since it involves the isolation of the system from its environment ; the effect of a change of the external parameters on the statistical distribution of such an isolated system requires careful investigation. It can be shown, however, that the distribution remains canonical, with an appropriate variation of its modulus. It then becomes apparent that the corresponding variation of the free energy with respect to the modulus is that given by Helmholtz'formula. This completes the identification of the statistical expression dor the free energy, i.e. the statistical foundation of thermodynamics, so far as reversible processes are concerned.

Owing to the choice of temperature as the fundamental thermal parameters, the entropy does not enter explicitly into the preceding considerations. It can be derived in the usual way from the free energy, and its statistical expression is found to be very simply related to the canonical distribution density : it is, apart from the sign, the canonical average of the logarithm of this density. If now the system undergoes an irreversible transformation, the initial distribution density will differ in an arbitrary way from the canonical form corresponding to the final values of the parameters, and the initial entropy, with the mimus sign, will appear as the average logarithm of this arbitrary distribution density. By taking account of the monotonic character of the logarithm function, it can be shown that such an average logarithm of a distribution density is always larger than that corresponding to the canonical distribution in the fianl state of the system. The law of increase of entropy is thus obtained as a consequence of the statistical expression for the entropy.

7.- Open systems : the chemical potentials.

It will be seen that the argument leading to the canonical distribution law for closed systems is essentially concerned with the statistical distribution, between the various parts of a large system, of a quantity, the energy, which satisfies a law of conservation and a law of additivity. These two properties are all that is needed for the application of the central limit theorem which directly gives the asymptotic distribution law. It is clear, therefore, that the same argument can be immediatly extended to open systems, since the quantities whose distributions have to be considered in this case ; the numbers of constituent elements of various kinds, have the two required properties. Moreover, just as in the case of energy exchanges for closed systems, the abstract argument is exactly adapted to a clear physical situation, viz. the exchange of elementary constituents between the system considered and its surroundings, which play the part of an arbitrarily large reservoir of such constituents.

The distribution law for the number of elements of any kind has thus a form very similar to the canonical distribution law for the energy ; in fact, it arises from a Poisson distribution around an average value of this number of elements. For each kind of element, a new macroscopic parameter thus appears, related to the average number of elements in much the same way as the modulus of the canonical distribution is related to the average energy; it is readily interpreted as the chemical potential introduced by Gibbs in the thermodynamics of open systems.

In dealing with systems of identical elements, one must observe that a microscopic state of the systems is physically determined by the set of values of the variables pertaining to all the constituent elements, irrespective of the individuality of these elements. Thus, two microscopic states differing only by some permutation of the groups of variables belonging to the various elements are physically indistinguishable. To express this fact, Gibbs introduced a distinction between specific and generic phases : a specific phase is defined by the assignment of the values of the variables for each individual element ; a generic phase is the set of specific phases corresponding to the same values of all the variables. In classical statistics, it is quite essent ial to take into account the distinction between specific and generic phases in all questions in which the numbers of elements of the systems occur explicitely, as in the thermodynamics of open systems or in the estimate of the time of relaxation neccessary for the establishment of statistical equilibriul. Many confuse and inconclusive discussions arose from the failure to pay sufficient regard to this point. In quantal statistics, on the other hand, all such difficulties vanish, since qunatum theory only deals with generic phases.

The quantization of the numbers of elemrnts gives rise to novel aspects of the statistics of open systems. It is well-known that this quantization obeys very different rules according as the elements in question are <u>fermions</u>, subject to the exclusion principle, or <u>bosons</u> satisfying a principle of symmetry. The general distribution law for open systems accordingly predicts very different behaviours for systems of fermions or bosons. The corresponding theories are usually called Fermi-Dirac and Bose-Einstein"statistics"; it should be clear that the statistical basis of these theories is the same, the difference arising from the specific properties of the elementary constituents in each case.

CLASSICAL STATISTICS

I.- General properties of Mechanical systems

1.- Notion as isomorphism in phase-space. We consider a mechanical system, the state of which is described by f generalised coordinates. q_1 and their conjugate momenta p_i. The set (q_i, p_i) $i = 1, \ldots, f$ defines a point P in a 2f-dimensional phase space, to which we assign for a reason soon to become apparent, a euclidean metric. The measure of an element of phase-space is accordingly defined as its euclidean volume $d\mu = dq_1 \ldots dq_f \, dp_1 \ldots dp_f$.

The succession of states occupied by the system in the course of time form a curve or trajectory in phase space, whose parametric equations can be written in Hamiltonian form

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} \qquad (1)$$

with the help of the Hamiltonian $H(q;p)$. The system is assumed to be conservative, i. e. $H(q;p) = E$ is a constant of the motion. In other words, the trajectory determined by the initial phase P_0 is contained in the energy surface $H(P) = E = H(P_0)$ passing through P_0. We assume that all the energy surfaces of our system are contained within a finite domain of phase space : physically, this simply means that no single coordinate or momentum ever becomes infinite. This assumption therefore, is a quite natural one to make : it is fundamental for the validity of the whole ergodic theory.

It is convenient to vizualize the motion, i.e. the passage from a phase P_0 to a phase P_t as a transformation of phase space into itself. The transformations P_0 P_t obviously form a continuous group, of parameter t, whose infinitesimal transformation is just given by the Hamiltonian equations (1). In other words, the motion is an automorphism of phase space. Integrals, like the Hamiltonian, are invariants of this group; the corresponding surfaces, like the energy surface, are invariant domains.

2.- Liouville's theorem For a statistical description of a stationary state of the system, we want a measure in phase-space which is invariant for the motion group. The measure $d\mu$ introduced above has this property : this is the statement of Liouville's theorem. The most striking way of proving this theorem makes use of an anlogy between the motion in phase-space regarded as a euclidean space and a hydro dynamical of flow, whose velocity v is directly given at any point by the Hamiltonian equations (1). The theorem follows from the observation that this flow satisfies the condition of incompressibility div $\vec{v} = 0$, i.e.

$$\sum_{i=1}^{f} \left(\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) = 0 .$$

This means in fact that any domain D_0 of phase space will be transformed by the motion into a domain D_t of perhaps quite different shape, but of the same volume.

From the invariant measure $d\mu$ in the 2f-dimensional phase space it is easy to derive an invariant measure on the (2f-1) dimensional energy surface. To this end, consider an energy-shell i.e. the space between two neighbouring energy surface E, E + dE. Let $d\Sigma$ be the Euclidean measure of an element of the surface H(P) = E, and dn an element of lenght along the direction normal to the surface element $d\Sigma$. The invariant measure $d\mu_E$ on the energy surface is then :

$$d\mu_E = \frac{d\mu}{dE} = d\Sigma \cdot \frac{dn}{dE}$$

i.e. the euclidean element $d\Sigma$ is weighted with the factor dn/dE. The latter is just $|\text{grad } H(P)|^{-1} = \left\{ \sum_i \left[\left(\frac{\partial H}{\partial q_i} \right)^2 + \left(\frac{\partial H}{\partial p_i} \right)^2 \right] \right\}^{-\frac{1}{2}}$ it is numerically equal to the inverse of the modulus of the velocity $\vec{v}$ at point P. In the following, we shall always use the invariant measure $d\mu_E$, and we shall drop the index E when no confusion is to be feared.

3. Physically equivalent phases and uniform phase functions. In the preceding sections we have singled out the energy integral as

defining in 2f-dimensional space a manifold of (2f-1) dimensions which completely contains a given trajectory. It is necessary to explain the physical reason for doing so and not pursuing the reduction of the number of dimensions of this manifold any further with the help of the other time-independent integrals. For this purpose, a closer consideration of the physical interpretation of the formalism is needed.

It generally occurs that the same physical state of the system is represented by more that one phase. An example is offered by angular variables : phases in which the values of some angular variables differ by 2 describe the same physical state. Another case, of less formal character and of fundamental importance in atomistic physics, is that of systems consisting of identical elements. A phase is then defined by the sets of canonical coordinates pertaining to all the elements, enumerated in a certain order; two phases differing only by this order of enumeration are indistinguishable from the physical point of view. To express this situation, one calls the phases as just defined <u>specific phases</u>, and one denotes by <u>generic phase</u> the set of all specific phases corresponding to the same values of all coordinates, taken in any order. A physical state of the system is thus described by a generic phase.

Phase functions representing physical quantities must have the same value for all phases corresponding to the same physical state of the system. Thus must be periodic with respect to any angular variable +, and if they refer to a system of identical elements they must be symmetrical with respect to permutations of the sets of variables pertaining to different elements. Phase functions possessing this property will be called <u>uniform</u>.

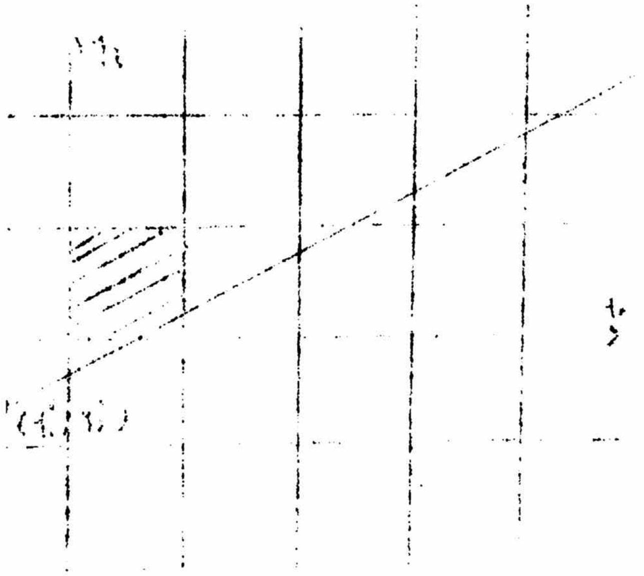
+ Strictly speaking, angular variables do not satisfy the finiteness dondition enunciated above (§1). But they nevertheless can be used just on account of the periodicity property of all phase functions of physical significance. See the example discussed in section 4 below.

Now, if we consider a set of (2f-1) independent integrals not containing the time, which fix the shape of a trajectory, they will not in general be all uniform. We can assign the value of any uniform integral which means that we can control it by physical means; but it has no physical meaning to assign a value to a non-uniform integral. The latter can therefore in any case not be used to reduce the manifold containing the trajectory. As regards the uniform integrals, the assignment of a definite value to any one of them is purely a matter of fixing the conditions under which we wish to consider the system under investigation.

Thus, we define an isolated system by fixing the value of the energy integral. We might wish to fix the value of the momentum and angular momentum integrals : this would mean that we consider the system as free to move in empty space. In this case, the number of dimensions of the surface on which the trajectory is contained would be reduced by 6 more units. It will be more in accordance with usual conditions, however, to leave these integrals indetermined by imagining the system in contact with an infintely heavy body, with which the elements of the system interact elastically. We shall therefore base the following argument on the consideration of the energy surface; nothing essential would be changed, of course, by the assumption of a "surface" of a somewhat smaller number of dimensions, corresponding to the assignment of fixed values to other uniform integrals as well.

4.- Example. The above considerations may be illustrated by the simple example of a system of two uncoupled rotations around fixed axes. Let us take as coordinates the asimuths q_1, q_2 and assume for simplicity the two moments of inertia equal to unity; the conjugate momenta p_1, p_2, which represent the angular momenta around the axes are uniform integrals, to which we assign the values ω_1 , ω_2 (numerically equal, in our case, to the angular velocities of rotation); this also fixes the energy $E = \frac{1}{2}(\omega_1^2 + \omega_2^2)$. The manifold containing the trajectory thus reduces to the plane q_1, q_2 and the trajectory is a straight line in this plane.

$$q_1 = \omega_1 t + q_1^0 \quad , \quad q_2 = \omega_2 t + q_2^0$$



The plane is divided by the lines $q_1 = m. 2\pi$, $q_2 = n. 2\pi$ (m, n integers) into a chequer of squares all containing phases physically equivalent to each other. For all computation of averages of physically significant phases-functions, we may accordingly restrict the manifold containing the trajectory to a single square, by transferring to this square all the segments of the trajectory contained in the other squares. The equivalent trajectory thus obtained consists of a set of parallel segments within the single square; this set is finite or infinite according as the quantity ω_1 / ω_2 is rational or irrational.

Now, we may choose as a third time-independent integral the function

$$q_2 p_1 - q_1 p_2 = M$$

For the initial phase one may write $M_0 = \omega_1 q_2^0 - \omega_2 q_1^0$, but this integral is not uniform. In fact, it takes a different value on each of the finite or infinite set of segments composing the trajectory : for on the segment originally in the square containing the phase $(q_1^0 + m. 2\pi , q_2^0 + n. 2\pi)$, the value of M differs from M_0 by $2\pi (n\omega_1 - m\omega_2)$.

5. Metrical indecomposability of energy surface. A question of primary importance for the establishment of a statistical distribution on the energy surface is to characterize from the metrical point of view the set of phases occupied by the system in the course of time, i. e. to know the measure of this set on the energy surface. Our simple example (§ 4) suggests that; apart from exceptionnal cases arising from some "degeneracy" (ω_1/ω_2 rational), any trajectory will in a certain sense "fill" the whole energy surface. We must try to give to this intuitive description of the physical situation a rigorous mathematical formulation.

Boltzmann and Maxwell assumed that the trajectory actually passes through every point of the energy surface : this is the famous ergodic hypothesis. It is easy, however, by the modern methods of the theory of sets, to disprove this hypothesis (Rosenthal and Plancherel). Consider a neighbourhood S of a phase occupied at some time by the system; it can be choosen sufficiently small to prevent the trajectory from remaining inside it at all times. Then it is clear that the part of the trajectory inside the neighbourhood will consist of a set of separate segments. Moreover, the time intervals during which the system is inside S will form a succession of separate finite segments of the time axis. According to a well-known property of sets of points, the set of such time intervals is enumerable. The segments of the trajectory inside S thus form an enumerable set, whose measure on the energy surface is zero.

A correct formulation of the "ergodic" situation is obtained from the consideration of the automorphism defined by the motion of the system. The set of phases through which a trajectory passes is clearly an invariant set with respect to this group, and we are concerned with its measure on the energy surface. Let us call the energy surface metrically indecomposable (and the group metrically transitive) if it cannot be expressed as the sum of two invariant sets both of positive measure. This means, then, that the set of phases forming a trajectory either is of measure zero (this is an exceptional case, such as that of a periodic motion; with a closed trajectory), or has the same measure as the whole energy surface.

At first sight, the condition of metrical indecomposability of the energy surface would seem impossible to fulfil. Let us in fact consider any time-independent integral $I(P)$, different from the Hamiltonian; the phase-function $I(P)$ therefore cannot have the same value over the whole energy surface. But then it is always possible to find a number I such that the invariant sets of phases for which $I(P) > I$ and $I(P) \leqslant I$, respectively, are both of positive measure.[+] At this point it is necessary to remember that there may be

---

[+] For the proof, see Khinchin, p. 30, footnote.

different phases corresponding to the same physical state of the system; we might then have an invariant subdivision of the energy surface into parts of positive measure, such that no two phases of any part are physically equivalent (e.g. the chequer of squares in our example, §4) such a subdivision, while formally violating the condition of metrical indecomposability as formulated above, would nevertheless represent, physically, an"ergodic"situation.

We must accordingly modify the definition of metrically indecomposable sets so as to cope with this case. We call an invariant subdivision of the energy surface into two parts of positive measure <u>essential</u> when all physically equivalent phases belong to the same part. The surface will then be <u>metrically indecomposable in the physical sense</u> when it does not allow of any essential subdivision. This modification will not affect the above argument in respect of uniform integrals : these will bring about an essential subdivision of the energy surface. But we cannot say anything about the effect of non-uniform integrals; in fact, in the example of § 4, it can be shown that the non-uniform integral M in the general case (ω_1/ω_2 irrational) does not disturb the metrical indecomposability of the (reduced) energy surface.

The next question would be, how from the structure of the Hamiltonian could one draw conclusions regarding the metrical indecomposability of the corresponding reduced manifold. This problem is not solved, however; in this sense, the assumption of the metrical indecomposability of the reduced manifold remains an hypothesis. The progress with respect to the original ergodic hypothesis lies in its precise mathematical formulation.

As we have just seen, we can only assume metrical indecomposability for the manifold reduced by taking account of <u>all</u> uniform integrals. In practice, however, the uniform integrals distinct from the energy will usually have the same value over a very large part of the energy surface, and it will be possible to neglect the domains of very small measure in which they differ from this dominant value. We may therefore still restrict ourselves to the consideration of the energy surface and speak, in this approximate sense, of its metrical indecomposability.

6.- Poincaré's theorem. From the discussion of the "shape" of the trajectories we now pass to the consideration of the way in which they are followed by the system in the course of time. This question is dominated by a theorem enunciated by Poincaré and proved rigorously much later by Caratheodory : for almost all trajectories, the system returns after a sufficiently long time arbitrarily near to its initial phase. An essential condition for the validity of Poincaré's theorem is the finiteness of the energy surface containing the trajectories.

The proof consists in showing that the set of exceptional trajectories, i.e. the set of those phases near which the system never returns, is of measure zero. We must first formulate in a precise way what we mean by the "return" of the system near some phase P_0. Let us choose some time interval τ and consider the sequence (P) of phases P_0, P_1, P_2, ... successively occupied by the system at times t_0, $t_1+\tau$ $t_0+2\tau$, ... We shall then say that the system returns near P_0 if every neighbourhood ς_{P_0} of P_0, however small, contains at least a point of the sequence (P). The exceptional phases will thus be those for which a neighbourhood can be found containing no point of the sequence (P). We must prove that the set of exceptional phases is of measure zero.

To this end, we cover the energy surface with a net, dividing it into intervals U_i, and so fine-meshed that for every point P and every neighbourhood ς_P of P, there is at least one interval U_k containing P and contained within ς_P : this means that we must actually have an infinitely fine mesh, and an enumerable infinity of intervals U_i . Let now D_i be the set of phases of U_i whose time-sequence (P) never returns to U_i or its boundary; the set D_i is therefore open and accordingly measurable. Moreover, the sum $D = \sum_i D_i$ is the set of exceptional phases : in fact, every point of D is clearly an exceptional phase, and conversely, it will readily be seen that any exceptional phase must belong to one of the sets D_i.

Consider now one of the D_i's and the sequence of sets $D_i^{(\ell)}$, $\ell =1,2,\ldots$ into which D_i is successively transformed at times $t_o+\tau$, $t_o+2\tau$,.... $t_o+\ell\tau$,.... No two sets of this sequence can have common points, otherwise D_i would have some phase in common with a $D_i^{(\ell)}$, i.e. there would be a point of D_i whose sequence (P) would return to D_i, in contradiction to the definition of this set. The sequence D_i, $D_i^{(\ell)}$ is thus an enumerable infinity of distinct sets, all of the same measure, (according to Liouville's theorem); the measure of their sum, which is the sum of their measures, must be finite, since it cannot exceed the total measure of the energy surface : therefore, the measure of each of them must be zero. From $\mu(D_i) = 0$ it follows that $\mu(D) = 0$, which we wanted to prove.

While a strictly periodic behaviour of a mechanical system would require very special conditions, Poincaré's theorem shows that in the general case the evolution of such a system still has a remarkable feature of near-periodicity. in that almost every state of the system will recur indefinitely at irregular intervals in the course of time, at any rate approximately. This approximation can be as close as one wishes, provided one waits long enough. Such a microscopic behaviour of a system, just as much as the reversibility of its motion in time, contrasts with macroscopic irreversibility. In the early period of elaboration of statistical mechanics, both features were emphasized by Boltzmann's opponents as throwing doubt on the consistency of the latter's statistical derivation of the second law of thermodynamics : the argument drawn from the microscopic reversibility was put forward by Loschmidt, while that based on Poincaré's theorem is known as Zermelo's objection. How such objections are overcome will appear in the course of the following argument.

Whether the recurrence phenomenon predicted by Poincaré's theorem can or cannot be observed depends entirely on the order of magnitude of the corresponding recurrence times. These times vary enormously according to the structure of the system and, of course, increase very rapidly with the number of degrees of freedom. We may get a rough idea of the situation by considering a simple example.

Consider a vessel containing 2N molecules of an ideal gas under normal conditions of temperature and pressure. The probability of finding an excess of molecules in one half of the vessel comprised between δ. N and ($\delta + d\delta$) N is, asymptotically,

$$\Pi(\delta)\,d\delta \simeq \frac{1}{\sqrt{\pi N}}\, e^{-\frac{(\delta N)^2}{N}}\, d\delta . N$$
$$\simeq \sqrt{\frac{N}{\pi}}\, e^{-\delta^2 N}\, d\delta \quad ,$$

if we assume that this inequality has arisen in the course of random fluctuations. Now, if it is the average time interval needed for the relative excess δ to change by $|d\delta|$, the probability of occurence of an excess in the interval (δ ,δ+ dδ) in the course of the temporal evolution of the system may be expressed as the fraction dt/T of the average time T between two successive occurrences of the excess δ . The ergodic theorem, as we shall see in the next chapter, allows us to equate the two probabilities just considered and in this way obtain an estimate of the recurrence time T for the excess δ :

$$T = \frac{1}{\Pi(\delta)} \left| \frac{dt}{d\delta} \right| ,$$

The average time variation of the excess δ is of the form

$$\delta \sim e^{-t/\tau} ,$$

where the relaxation time τ is determined by the mechanics of the irreversible diffusion process leading to the establishment of a uniform density. One has, therefore,

$$\frac{1}{\delta} \left| \frac{d\delta}{dt} \right| = \frac{1}{\tau} ,$$

and

$$T = \sqrt{\frac{\pi}{\delta^2 N}}\; e^{\delta^2 N} . \tau .$$

If D is the diffusion coefficient, the order of magnitude of τ is related to the linear dimensions L of the volume in which the process takes place by a formula of the type

$$\tau \approx \frac{L^2}{D}$$

If the volume is of macroscopic dimensions, $L \gtrsim$ 1cm, say, one has $\tau \gtrsim$ 1 sec and $N \approx 10^{19}$. Even an extremely small relative excess of density, $\delta \gtrsim 10^{-6}$, would not recur spontaneously before times of the fantastic order of magnitude $T \gtrsim e^{10^7}$ sec. But if the volume is of microscopic dimensions, e.g. $L \gtrsim 10^{-5}$ cm, an excess of density of 1% will have a recurrence time of the order of 10^{-9} sec.

II. The Ergodic Theorems.

1.- Existence of the time average. Let us consider the evolution of a system starting at time t_0 from a certain phase P_0. As explained in the introduction, the measurement of some macroscopic quantity corresponding to the microscopic phase function $f(P)$ yields in the first instance a time average

$$F\,(P_0;t_0\,,T) = \frac{1}{T}\int_{t_0}^{t_0+T} f\,(P_t)\,dt, \qquad (1)$$

taken along the trajectory defined by the initial phase P_0; in this formula, the symbol P_t denotes the set of canonical coordinates of a point of this trajectory, expressed as functions of the time and the initial conditions P_0, t_0. One is led to expect that this time of average has a limit for $T \to \infty$, which is independent of the initial conditions and represents the value of the macroscopic quantity for the system in its given stationary state. We must now discuss the mathematical justification of this surmise.

The first step is described by Birkhoff's ergodic theorem, which states that the limit

$$F\,(P_0) = \lim_{T\to\infty} F\,(P_0;\,t_0,\,T) \qquad (2)$$

exists for almost all trajectories, and is independent of the initial time t_0. We first prove the theorem for the case that the time interval T varies by finite increments of duration T, i.e. we take $T = n\,\tau$ and investigate the limit of the sequence

$$F_n\,(P_0;\,t_0) \equiv F\,(P_0;\,t_0,\,n\tau\,) \qquad (3)$$

as $n\to\infty$. Let P_0 be an exceptional phase, i.e. such that $F_n(P_0;t_0)$ has no limit; this means that the lower bound $\underline{F}\,(P_0)$ and the upper bound $\overline{F}(P_0)$ are different . We can then choose a pair of members α, β $(\alpha<\beta\,)$ between $\underline{F}(P_0)$ and $\overline{F}(P_0)$, i.e. such that

$$\underline{F}(P_0) < \alpha \qquad\qquad \overline{F}(P_0) > \beta \qquad (4)$$

Now, if the set D of exceptional phases were of positive measure, it is easily seen [*] that one could find a pair $(\alpha\,,\,\beta\,)$ for which the

---

[*] For details, see Khinchin, p. 19-27

conditions (4) are satisfied for all the phases P_o of a subset D^* of positive measure. By showing that this last property $\mu(D^*) > 0$ is contradictory to the inequality $\alpha < \beta$, we deduce that the set D of exceptional phases is of measure zero.

The contradiction is elicited as follows. Let us consider the sequence of times $t_k = t_o + k\tau$ and the corresponding phases $P_k = P_{t_k}$ (k being an integer of any sign), we define

$$f_k(P_o) = \frac{1}{\tau}\int_{t_k}^{t_{k+1}} f\,(P_t)\,dt \qquad (5)$$

the time average over the interval (t_k, t_{k+1}). By a change of origin of the times, we see that

$$f_k(P_o) = f_o\,(P_k)\,, \qquad (6)$$

Now, the time average $F_n(P_o;\,t_o)$ is expressed as

$$F_n(P_o;\,t_o) = \frac{1}{n}\sum_{k=0}^{n-1} f_k\,(P_o),$$

and if we integrate this over any set of phases $D_o^{(n)}$, we get

$$n\int_{D_o^{(n)}} F_n(P_o;\,t_o)\,d\mu = \sum_{k=0}^{n-1}\int_{D_o^{(n)}} f_k\,(P_o)\,d\mu = \sum_{k=0}^{n-1}\int_{D_k^{(n)}} f_o\,(P)\,d\mu,$$

where $D_k^{(n)}$ is the transform of the set $D_o^{(n)}$ when $P_o \to P_k$. If now $D_o^{(n)}$ is a subset of D^* such that for any P_o of $D_o^{(n)}$ one has $F_n\,(P_o;\,t_o) > \beta$, we obtain the inequality

$$\sum_{k=0}^{n-1}\int_{D_k^{(n)}} f_o\,(P)\,d\mu > n\beta\mu\,(D_o^{(n)})\,.$$

Suppose further that the sets $D_k^{(n)}$ are non-overlapping, and call their sum

$$D^{(n)} \equiv \sum_{k=0}^{n-1} D_k^{(n)}$$

Since, by Liousville's theorem, $\mu\left(D_k^{(n)}\right) = \mu\left(D_o^{(n)}\right)$, we then have

$$\int_{D^{(n)}} f_o\,(P)\,d\mu > \beta\mu\,(D^{(n)})$$

Now it can be shown that such sums of non-overlapping sets $D^{(n)}$ can actually be specified for each value of n in such a way that they together exhaust D^*. We can then sum the inequqlities of the type just derived for all n, and we get in the limit $n \to \infty$

$$\int_{D^*} f_o\,(P)\,d\mu > \beta\mu\,(D^*)\,.$$

A similar argument leads to $\int_{D^*} f_0(P)\, d\mu < \alpha \mu(D^*)$,

and these two ine ualities with $\mu(D^*) > 0$, do indeed contradict the inequality $\alpha < \beta$.

To complete the proof of Birkhoff's theorem, we have to co pare the average for an arbotrary time interval T with that for the interval of the sequence n τ nearest to T. We have

$$\left| \frac{1}{T} \int_{t_0}^{t_0+T} f(P_t)\, dt - \frac{1}{n\tau} \int_{t_0}^{t_0+T} f(P_t)\, dt \right| \to 0,$$

and

$$\left| \frac{1}{n\tau} \int_{t_0}^{t_0+T} f(P_t)\, dt - \frac{1}{n\tau} \int_{t_0}^{t_0+nT} f(P_t)\, dt \right| = \frac{1}{nT} \left| \int_{t_0+nT}^{t_0+T} f(P_t)\, dt \right| \leq \frac{1}{n} \left| f_n(P_0) \right|,$$

with the notation (5). It is readily verified * that

$$\lim_{n \to \infty} \frac{1}{n} \left| f_n(P_0) \right| = 0$$

almost everywhere. The proof is of the familiar type : one shows by making use of (6) and of Liouville's theorem, that the set of phases P_0 for which $|f_n(P_0)| > r\varepsilon$ is of measure zero. Hence, the time average $F(P_0; t_0, T)$ has a limit for $T \to \infty$ almost everywhere.

Finally, it must be shown that this limit is independent of the initial time t_0. We have

$$\lim \frac{1}{T} \int_{t_0}^{t_0+T} = \lim \frac{1}{T+t_1-t_0} \int_{t_0}^{t_1+T} = \lim \frac{1}{T} \int_{t_0}^{t_1+T}$$

since the difference of the last two expressions tends to zero like $(t_1-t_0) / T$. Further, the difference

$$\frac{1}{T} \int_{t_0}^{t_1+T} - \frac{1}{T} \int_{t_1}^{t_1+T} = \frac{1}{T} \int_{t_0}^{t_1}$$

also tends to zero, which complete the proof.

2;- Equivalence of time average and statistical average.

The time average whose existence is established by Birkhoff's theorem does not yet correspond, in general, to the physical notion of a macroscopic quantity attached to an isolated system, since it may still depend on the particular trajectory followed by the system and have different values for different trajectories even on the same energy surface. It is easy, however, to indicate a general condition sufficient to ensure the constancy of all time averages almost everywhere on the energy surface : it is the metrical indecomposability of this surface, at least in the physical sense. For if the time average F(P) is not almost everywhere constant, it is possible to find a value F of F(P) such that the conditions $F(P) > F$ and $F(P) \leq F$ define two invariant sets of positive measure, effecting an essential decomposition of the energy surface.

If the energy surface is metrically indecomposable, the constant value of the time average can be expressed as a statistical average over this surface. Roughly speaking, the trajectory along which the time average is taken "fills" the whole energy surface : the time average can thus be considered, if we disregard the temporal succession of the phases as an average over the energy surface, with a definite weighting of each surface element. This statistical weight turns out to have an extremely simple specification : if we use the <u>invariant measure</u> on the surface, the statistical distribution equivalent to the time average is <u>uniform</u>; in other words, the amount of time spent by the system in any region of the surface is proportional to the invariant measure of this region. This uniform statistical distribution is called <u>ergodic</u> and the fundamental corollary of Birkhoff's theorem which express this situation can be formulated as follows :

If the energy surface Ω is metrically indecomposable, the time average F of the phase function f(P) is given by the ergodic average

$$\bar{f}\big|_E \equiv \frac{1}{\mu(\Omega)} \int_\Omega f(P)\, d\mu_E \,. \qquad (7)$$

In the first place, it is clear that the ergodic average of f(P) is the same as that of the time average $F(P_0;\ t_0,\ T)$ over any finite time interval T; indeed

$$\frac{1}{\mu(\Omega)} \int_\Omega F(P_0; t_0, T)\, d\mu_E = \frac{1}{T} \int_{t_0}^{t_0+T} dt\, \frac{1}{\mu(\Omega)} \int_\Omega f(P_t)\, d\mu_E$$

$$= \frac{1}{T} \int_{t_0}^{t_0+T} dt\, \frac{1}{\mu(\Omega)} \int_\Omega f(P_0)\, d\mu_E$$

in virtue of the invariance of the energy surface Ω ; and this is

$$= \frac{1}{T} \int_{t_0}^{t_0+T} f\Big]_E \, dt = f\Big]_E$$

Since, owing to the constancy of F ,

$$F = \frac{1}{\mu(\Omega)} \int_\Omega F \, d\mu_E ,$$

it remains to prove that

$$\frac{1}{\mu(\Omega)} \int_\Omega \left[F - F(P_0; t_0, T) \right] d\mu_E = 0$$

This is readily [x] seen to be a consequence of the ergodic theorem $F(P_0;\ t_0,\ T) \rightarrow F$ for $T \rightarrow \infty$ for almost all P_0.

The last step in established the physical equivalence between time average and ergodic average is an obviuous extension of the preceding theorem ;

If the energy surface Ω is metrically indecomposable in <u>the physical sense</u>, the time average F of any <u>uniform</u> phase function f(P) is equal to its ergodic average :

$$F = f\Big]_E \tag{8}$$

The equation (8) is valid for <u>almost</u> all phases of the energy surface. For the physical interpretation, it is therefore necessary to introduce a further averaging in order to eliminate the possibility that the initial phase would happen to be an exceptional one. Instead of starting from a well-defined initial phase, we thus assume an initial distribution of "copies" of our system over a finite domain of the energy surface. The time average F must then be averaged again over this domain ; any exceptional phases contained in it will not give any contribution to the average, and the result will again be expressed by an equation perfectly similar to (8). The replacement of an initial phase by an initial "cell" of finite (and arbitrary) size has a great importance for the physical interpretation of the theory. Before we discuss it, however, we shall proceed to an extension of the ergodic theory which first displays its full scope

3.- <u>Hopf's ergodic theorem.</u> The restriction of the phases to an energy surface is a strong idealisation of an isolated system. It is more realistic to allow a certain margin to the definition of the energy, i.e. to regard as possible all trajectories contained within

---

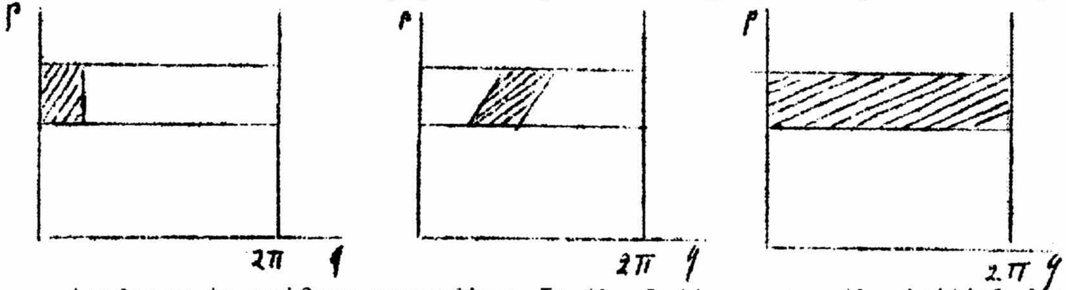
[x] See Khinchin, p. 31- 32.

an energy shell, consisting of all energy surfaces in the nergy interval (E, E + dE). The ergodic average over an energy shell is even somewhat simpler of expression than that on the energy surface, since it corresponds to a distribution of uniform density with respect to the simpler measure in phase-space given by Liouville's theorem : denoting the energy shell by (E) and the corresponding ergodic average by $\overline{f}\rceil_{(E)}$, we have

$$\overline{f}\rceil_{E} = \frac{1}{\mu[(E)]}\int_{(E)} f(P)\,d\mu, \qquad d\mu = dq_1 \cdots dq_f\, dp_1 \cdots dp_f\,; \quad (9)$$

here $\mu[(E)] = \int_{(E)} d\mu$ represents the measure of the energy shell.

The condition for the equivalence of time average and ergodic average over an energy sehll is not simply that all the energy surfaces within the shell be metrically indecomposable. An additional condition is needed, involving mutual relationships between these surfaces. The situation will be made clear by a simple example. Consider a motion of a single degree of freedom, defined by an angle variable $q = \omega t \pmod{2\pi}$ and the conjugate action variable p , which is a constant of the motion. We may assume that the value of p fixes the energy; in the "phase space" (p, q,) the "energy surface" is then the line p = const., or rather the segment $0 \leqslant Q \leqslant 2$ on this line. An energy shell will be a strip limited by two such lines. While each energy line is metrically indecomposable, the situation with respect to an energy strip will be radically different according as the integral p is or is not independent of ω . In the first case, any initial distribution will simply be displaced along the strip without any tendency to uniform spreading. In the latter case, the initial distribution, while remaining of the same total area, will spread out in the form of a more and more intricate subdivision of partial domains stretching over the whole energy strip : this illustrates the "mixing" mechanism by which the ergodic distribution is established asymptotically.



The general analysis of this "mixing" over an energy-shell has been carried out by Hopf. It introduces two new elements into the problem. In the first place, it is necessary to give a precise mathematical definition of the mixing process; morever, we must formulate the additional condition which guarantees this process to take place. The limiting process corresponding to the mixing over an energy shell is somewhat weaker than the convergence almost everywhere which obtains for the time average on an energy surface : it has the character of "convergence in the mean". To define this, let us start, at time $t_0 = 0$ from some arbitrary distribution characterized by a density $g(P)$; the statistical average of a phase-function $f(P)$ for this distribution is given by

$$(f, g) \equiv \int f(P) . g(P) \, d\mu ,$$

the integration being extended to the whole phase-space (the limitation to an energy shell is contained in the form of the density function $g(P)$. At time t, the distribution density has become $g(P_t)$, i.e. the transform of $g(P)$ by the automorphism $P \rightarrow P_t$; this is some function of P and t which we denote by $\mathcal{T}_t \, g(P)$. The average of $f(P)$ at time t is accordingly $(f, \mathcal{T}_t g)$. The limiting form $\overline{g}(P)$ of the distribution density is then defined, in the sense of convergence in the mean, by the condition

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \left| (f, \mathcal{T}_t g) - (f, \overline{g}) \right|^2 dt = 0 \qquad (10)$$

for any pair of functions $f(P)$, $g(P)$. This means, physically, that the time average of the statistical fluctuations of the phase average of $f(P)$ around its limiting value $(f, \overline{g})$ tends to zero as the time T increases indefinitely. Such a formulation is all that is needed for physical applications.

We must now state the condition for the existence of a limiting distribution $\overline{g}(P)$ satisfying (10); as our example shows, this condition must express a relation between different energy surfaces : in its simplest form it must refer to a pair of energy surfaces. It is convenient, for the consideration of such a pair, to "duplicate" in a certain sense the phase space, i.e. to introduce a phase space whose coordinates consist of a pair of sets of coordinates of the original phase space : this new phase space may be described as the "direct product" of the original one.into itself. A pair of energy surfaces of the original phase space thus appears as a single energy surface of the product space. We are now in a position to enunciate [x] Hopf's ergodic theorem : if every energy surface of the original phase space is metrically indecomposable, and almost every energy surface of the product space has the same property, every distribution tends to a limiting distribution of uniform density in the sense of equation (10) This theorem completes the foundation of the concept of ergodic distribution, on which the whole structure of statistical thermodynamics rests.

4.- The ergodic distribution. Let us now discuss from a more physical point of view the main features of the ergodic distribution and of the mixing process by which it comes about. In considering this process we immediately face the paradox of an essentially reversible mechanism leading to an irreversible situation; but we have also in hand the means of solving this paradox. If we analyse at any instant the distirbution arising from any initial one by letting the size of the elements of phase space decrease indefinitely, we shall find that a given infinitesiaml element is either occupied or empty, and the fraction of occupied elements, by Liouville's theorem, will always remain the same : from this point of view, there is thus no mixing at all, but a perfectly reversible evolution of the distribution. If however, we fix our attention upon an element of phase space of arbitrary but finite size, the density of distribution in this element will tend asymptotically to a constant value : as Ehrenfest expressed it, the irreversible behaviour pertains to a course distribution in phase. The degree of "coarseness", i.e. the size of the cells of phase space with respect to which the distribution is defined, is arbitrary; but it is essential that some finite subdivision into cells be assumed. It is this cell-structure which represents the lack of definition of the state of the system necessary for the application of a statistical mode of description of its behaviour. From the mathematical point of view, we need not specify the size of the cells and we may even ultimately regard them as infinitesimal for the purpose of practical calcu-

---

[x] We shall not give the proof of Hopf's theorem, since it does not involve any essentially new feature.

lation; but the limiting process of letting the cell-size tend to zero must then carried out after the process of letting the time of evolution of the system tend to infinity. If we should reverse the order of the two limiting processes, we should obtain the reversible behaviour of a "fine" distribution.

The next problem concerns the time of relaxation of the mixing process, i.e. the average time which must elapse before any given distribution has pratically reached its asymptotic uniformity. At first sight, it would seen that we are landing into another paradox; our example of section 3 suggests that we must expect the mixing to proceed during a time sufficient to allow the system to return several times near phases previously occupied, in accordance with Poincaré's theorem. But the occurence of such repetitions of initial configurations is just the kind of behaviour characteristic of the purely mechanical evolution of the system as opposed to its statistical, irreversible evolution. The fantastic order of magnitude of the times of recurrence accounts for the failure to observe such repetitions in usual cases; and if the mixing process should really have to involve them, the whole statistical interpretaion of irreversibility would collapse.

To clear up the situation, it must be remembered that the usual systems to which thermodynamics is applied are composed of a large number of identical elements : they are either bodies built up of atoms or radiation fields consisting of proper oscillations. The physical states of such systems are represented by generic phases; the latter do not correspond to single "cells" of phase space but to regions called "constellations" by Ehrenfest containing many cells and intrically imbricated through each other over each energy shell. The measure of a constellation, according to the ergodic theorems, gives the relative probability of occurence of the state represented by the corresponding generic phase in the course of time. States very near to equilibriul will thus occuoy a much larger part of the energy shell than states deviating appreciably from the equilibrium conditions. Thus, if a trajectory starts from such an "improbable" constellation, it will soon enter into more and more porbable ones and when it will have reached the equilibriul constellation, it will stay in it most of the time , although it will continually cross less probably ones, appearing as "fluctuations" from the equilibrium state. The return of a very improbable state, i.e. the return of the trajectory to a very improbable constellation, would require a time of a fantastic order of magnitude; but the time of relaxation is defined as that needed to reach the most probable constellation, starting from a less probable one; and this is extremely short as soon as the number of elements is reasonably large. The importance of the consideration of generic phases for the consistency of the statistical theory in all cases in which its application is significant was, of course, fully recognized both by Gibbs and by Boltzmann. But the mode of exposition adopted by Gibbs tended to obscure this issue. Gibbs was worried by the discrepancies between theory and experiment, especially with respect to the consequences of the equipartition of energy for the specific heats; these discrepancies seemed to be connected with the assignment of the number of degrees of freedom of the atomic systems; in the hope of throwing light on the origin of the difficulty, he accordingly set himself the task to derive as rigorously as possible all those statistical properties which apply to the most general mechanical systems, without any restriction as to the number of degrees of freedom. Hence the emphasis in his book upon the derivation of such results, while the discussion of systems of identical elements was confined to the last chapter. Boltzmann strongly objected to the tendency of Gibbs' book :"I can understand statistics applied to a gas, he said,[x] but I see no point in applyingit to a sewing machine". Gibbs' distrust of detailed atomic models was part of an attitude of mind shared by many physicists towards the end of the XIXth century : it was an essentially idealistic reaction against the mechanistics school which had flourished during the second half of the century, and of which Kelvin is the typical representative. Boltzmann remained to the last a staunch supporter of the mechanistic ideal; and in spite of the fact that this ideal has now been recongized to be too narrow, it did at the time give Boltzmann the right inspiration. His views on statistical mechanics were perfectly clear and precise, although he lacked the mathematics, we now have for their appropriate expression. Yet he had to face widespread scepticism and opposition and did not live to see the triumph of his ideas.

---

[x] This utterance was reported in conversation by Ehrenfest who had been a pupil of Boltzmann's.

In fact, the objections raised against Boltzmann's position are only interesting insofar as they show the kind of misunderstanding of the significance of the statistical approach one might fall into. Loschmidt observes that if at a certain time all velocities (and magnetic fields) of the system are inverted, the system will so to speak, retrace its steps : if, therefore, its evolution went from less to more probable configurations, it will now exhibit an opposite trend. Zermelo, availing himself of Poincaré's theorem, pointed out that any configuration, however improbable, must be expected to repeat itself with arbitrary accuracy in the course of the time. Both assertions are, of course, perfectly true : they can be vizualized by following the path of the system among the entanglement of the various constellations. But they are beside the point : it is the course distribution which has an irreversible evolution, and we have seen that such an irreversible trend is perfectly compatible with the reversibility of the mechanism by by the "mixing" of the course distribution is brought about.

5. - Systems with weak interactions. We may follow up the last argument more quantitavely. We introduce the phase space of a single element of the system, the "μ-space" in Ehrenfest's terminology. The total phase space of a system of N elements, or "Γ-space", is the direct product of N identical μ-spaces. Let us subdivide the μ-space into cells of arbitrarily small, but finite size ω_i ; this defines a subdivision of Γ-space into cells corresponding to the specific phases of the system. Thus, consider the specific phases for which there are N_i elements in the μ-cell ω_i : they occupy a cell in Γ-space of measure $\omega_1^{N_1}\omega_2^{N_2}\cdots$, or $\prod_i \omega_i^{N_i}$. To find the measure of the corresponding generic phase, we have simply to multiply this by the number of ways in which the N elements can be arranged in groups of N_1, N_2,......, i.e. $N! / N_1! N_2! \ldots$; the measure of the generic phase is thus

$$\Omega(N_1, N_2, \ldots) = N! \prod_i \frac{\omega_i^{N_i}}{N_i!}$$

To go on with explicit calculations, we must restrict ourselves to a particular type of system : we assume that the interactions between its elements can be neglected ; examples of such systems are the ideal gas and the pure radiation field. It is important to note that in order for the equilibrium distribution of the system to be ergodic, it is essential that the elements do interact ; otherwise, the numbers of elements in the various energy shells of μ-space would obviously remain inaltered in the course of time.

But the magnitude of the interaction is irrelevant and in suitable circumstances can be ragarded as negligible in comparison with the proper energies of the elements. In this case, we may assign a well-defined energy E_i to all elements in the μ-cell ω_i and the total energy of the system is an additive function of the E_i's :

$$E = \sum_i N_i E_i \qquad (12)$$

If the total energy E is given, this is a condition imposed on the distribution of the elements among the cells ω_i. Another condition is, of course

$$N = \sum_i N_i \qquad (13)$$

We may now readily evaluate the distribution N_i which, under the conditions (12) (13) gives the constellation of maximum measure (11). If the N_i's are large, the asymptotic expression for this distribution is

$$\overline{N_i}\big|_E \simeq \frac{N}{Z}\, \omega_i\, e^{-\beta E_i} \qquad (14)$$

where

$$Z \equiv \sum_i \omega_i\, e^{-\beta E_i} \qquad (15)$$

and the parameter β must be determined in terms of E from the condition (12). In evaluation, the "sum over states" Z it is permissible to treat the μ-cells as infinitesimal, and to write

$$Z = \int e^{-\beta H(p,q)}\, d\mu , \qquad (16)$$

where H(p,q) is the Hamiltonian of a single element and the integration is extended to the whole μ-space.

Let us now evaluate the measure $\Omega(\overline{N_1}|_E, \overline{N_2}|_E, \cdots)$ of the equilibrium constellation, and compare it with the measure of the energy shell in Γ space in which it is contained. We have

$$\log \Omega(\overline{N_1}|_E, \overline{N_2}|_E, \cdots) \simeq N \log Z + \beta E , \qquad (17)$$

so that the problem is reduced to the computation of log Z. To this end, we decompose the integration in μ space into an integration over all the values of the energy and another over each energy surface in turn, using, of course, the invariant measure in accordance with $d\mu = d\mu_E . dE$. For the invariant measure of the energy surface E we write

$$\mu(E) = e^{\sigma(E)} ;$$

thus

$$Z = \int_{-\infty}^{+\infty} e^{\sigma(E) - \beta E}\, dE$$

The exponent of the integrand has a maximum for an energy E_m given by

$$\left(\frac{d\sigma}{dE}\right)_m = \beta$$

Expanding it near the maximum, we have, to a sufficient approximation,

$$\sigma(E) - \beta E = \sigma(E_m) - \beta E_m + \frac{1}{2}\left(\frac{d^2\sigma}{dE^2}\right)_m (E - E_m)^2 ;$$

the coefficient $\frac{1}{2}(d^2\sigma / dE^2)_m$ is assumed to be negative. This defines a Gaussian distribution of the energy around the extremal values E_m. The mean square fluctuation $\overline{\Delta E^2}$ of the energy is $|(d^2\sigma / dE^2)_m|^{-1}$, which gives the physical meaning of the latter quantity. To this approximation, we get for log Z

$$\log Z \approx \log \mu(E_m) - \beta E_m + \log\sqrt{2\pi\,\overline{\Delta E^2}}$$
$$\approx \log \mu([E_m]) - \beta E_m ,$$

if we denote by $[E_m]$ an energy shell in μ-space of thickness $\sqrt{2\pi\,\overline{\Delta E^2}}$ around the energy surface E_m. Inserting the last formula into (17) and noting that $NE_m \approx E$, we see that $\Omega(\overline{N_1}]_E, \overline{N_2}]_E, \dots)$ takes [illegible] form $\mu([E_m])^N$ of the measure of the energy shell in Γ space consisting of the shells $[E_m]$ in all the μ-spaces. In this sense, we may say that, asymptotically, the equilibrium constellation is equivalent to the whole energy shell in Γ space. The distribution (1[illegible]), which completely describes the equilibrium constellation, may therefore be used to compute, to a sufficient approximation, the ergodic averages of all additive phase functions for the special type of system here considered and, in particular, to define the thermodynamical functions pertaining to it. We shall come back to this last point in a moment, after having discussed it for the most general isolated systems.

6.- Thermodynamics of isolated systems. Isolated systems do not lend themselves very well to a discussion of thermal quantities, for the definition of temperature requires, from the physical point of view, the possibility of an exchange of energy between the system and its surroundings. Nevertheless, one can speak of the entropy of an isolated system and may thus ask for the atomistic interpretation of this quantity. The temperature will thus not receive any direct physical interpretation, but will appear as a derived concept. We shall give later a more satisfactory treatment of this problem; but we may just as well take it up at this stage in order to present the general point of view from which it is considered.

Let us first recall the peculiar way in which macroscopic variables enter into the fundamental thermodynamics relations. Each non - thermal physical aspect is represented by a pair of variables (a, A) whose

mutual

relationship is characterized by saying that if the magnitude of a is controlled by some agency external to the system, the work done *on* the system to change a by da is Ada. In the atomistic treatment, we may take account of these macroscopic variables (a, A) by introducing into the Hamiltonian of our sysyem suitable terms of potential energy which will contain the parameters a. The associated variables A, called "forces" in an extended sense, will then be defined as the statistical averages of the phase functions $\overline{(\partial/\partial a)\, H(p,q;a)}$. For an isolated system, we must take the ergodic averages :

$$A = \overline{\frac{\partial H}{\partial a}}\Big|_E$$

The thermal transformations are also formally represented by a pair of associated variables, the entropy and the temperature, but these must, of course, be treated separately and in a quite different way.

For an isolated system, the *independent* variables (i.e. those which are controlled by external agencies) are the energy E and the parameters a. The entropy then plays the part of the *characteristic function*, from which all other quantities are obtained by differentiation and algebraic cpmbination. If θ denotes the absolute temperature, one has

$$dS = \frac{1}{\theta}\, dE - \frac{A}{\theta}\, da, \qquad (18)$$

so that

$$\frac{1}{\theta} = \left(\frac{\partial S}{\partial E}\right)_a, \quad -\frac{A}{\theta} = \left(\frac{\partial S}{\partial a}\right)_E . \qquad (19)$$

We shall always give the temperature, the dimension of an energy; the entropy will accordingly be dimensionless. Now, we can readily indicate an expression pertaining to the atomistic description of the system and which formally satisfies the relations (19) characteristic of the entropy. Let Ω represent the part of phase space "enclosed" by the energy surface E, i.e. containing all the energy surfaces corresponding to energy values $\leqq$ E. We may then take

$$S = \log \mu(\Omega) \qquad (20)$$

The verification of the suitability of this definition must be limited to the second relation (19), since the first one can only be regarded as a formal definition of the temperature. In order to calculate the derivative $\partial S/\partial a$ at constant E, let us introduce the characteristic set function $\mathcal{E}(\Omega)$ of the set Ω, i.e. a function equal to unity for all phases of Ω and zero for all other phases; in the present case, we may express $\xi(\Omega)$ very simply as the "step" function $\xi(x)$ of the argument $x = E - H(p,q,a)$:

$$\xi(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

The derivative of $\xi(x)$ is the distribution $\delta(x)$. This allows us to write

$$\mu(\Omega) = \int \xi\left[E - H(p,q;a)\right] d\mu$$

the integration extending to the whole phase space. We further get

$$\frac{\partial \mu(\Omega)}{\partial a} = -\int \frac{\partial H}{\partial a}\, \delta\left[E - H(p,q;a)\right] d\mu$$

$$= -\int dE'\, \delta(E-E') \int_{E'} \frac{\partial H}{\partial a}\, d\mu_{E'} = -\int \frac{\partial H}{\partial a}\, d\mu_E = -\overline{\frac{\partial H}{\partial a}}\bigg|_E \mu(E)$$

Therefore $$\frac{\partial \log \mu(\Omega)}{\partial a} = -A \cdot \frac{\mu(E)}{\mu(\Omega)}$$

But clearly $$\mu(E) = \frac{\partial \mu(\Omega)}{\partial E},$$

and therefore $$\frac{\mu(E)}{\mu(\Omega)} = \frac{\partial \log \mu(\Omega)}{\partial E} = \frac{1}{\theta}$$

whence finally $$\frac{\partial \underset{S}{\cancel{\log \mu(\Omega)}}}{\partial a} = -\frac{A}{\theta}$$

In the special case of systems with weak interactions, discussed in the preceding section, we may, of course, use the definition (20) of the entropy; but we may just as well take for this purpose a somewhat different definition, viz.

$$S = \log \Omega\left(\overline{N_1}\big|_E, \overline{N_2}\big|_E, \ldots\right), \tag{21}$$

whose explicit form is given by (17). The "force" A is now expressed as an average over the most probable distribution (14)

$$A = \frac{N}{Z} \int \frac{\partial H}{\partial a}\, e^{-\beta H}\, d\mu,$$

the Hamiltonian H(p,q;a) now referring to a single element, and the integration being extended to the corresponding μ-space. From (21) and (17) we derive

$$\left(\frac{\partial S}{\partial E}\right)_a = \beta$$

which gives the interpretation of the parameter β as the inverse of the absolute temperature, and

$$\left(\frac{\partial S}{\partial a}\right)_E = -\beta\frac{N}{Z}\int\frac{\partial H}{\partial a}\, e^{-\beta H} d\mu = -\beta A,$$

which shows that the definition (21) of S is as acceptable as (20). It is essential for this equivalence that the system considered has a large number of degrees of freedom; for the equation (17) is an asymptotic one. This ambiguity in the entropy definition for large systems was discussed in detail by Gibbs, and also from a more physical point of view by Lorentz, who called it the "insensibility" of the entropy definition. The reason for this latitude in the choice of the function S is clear. We have seen already that the measure of $\Omega(\overline{N_1}|_\varepsilon, \overline{N}|_\varepsilon)$ is pratically the same as that of an energy shell in Γ -space; and the latter is not significantly different from the measure of the whole domain Ω when the number N of elements is very large. To give a simple example, consider an ideal monoatomic gas, whose Hamiltonian consists of the kinetic energy of translation $\frac{1}{2m}\sum_{i=1}^{3N} p_i^2$ (m being the mass of an atom) and of a potential energy expressing that the system is confined to a volume V. In computing $\mu(\Omega)$ the integrations over the position and momentum coordinates can be performed separately, the former gives a factor V^N, the latter the volume of a 3N-dimensional sphere of radius $\sqrt{2m\,E}$, which is proportional to $E^{(3N/2)}$. The measure $\mu(E)\Delta E$ of an energy shells differs from $\mu(\Omega)$ only by the replacement of $E^{(3N/2)}$ by $(3N/2)\, E^{\frac{3(N-1)}{2}}\Delta E$, which for very large N is quite insensible.

Our two definitions of the entropy have a common feature : they both express the entropy as the logarithm of the measure of a domain of phase space, i.e. of the relative probability of occurence, in the course of time, of the phases contained in this domain. The logarithmic character of the link between entropy and probability is easy to understand : if we unite two independent isolated systems into a single system, their entropies should be added, the probabilities of their configurations multiplied. The definition (21), usually adopted for systems with weak interactions, corresponds to a more detailed discriminination of the configurations than the general definition (20).

The law of increase of entropy appears as a statistical law. If (by removing constraints) we allow the isolated system to pass from an initial state of equilibrium to the state of equilibrium corresponding to the new conditions , the values of the entropy in the final state will necessary be larger than the initial one ; but the behaviour of the statistical function log. $\Omega(N_1, N_2....)$ during the transition will, as already stated, exhibit continual fluctuations (even after the new state of equilibrium is reached). The concept of entropy, however, essentially refers to the state of equilibrium under neglect of such fluctuations. The statistical interpretation thus indicates the limit of validity of the entropy concept ; if the fluctuations form a prominent feature of the phenomenon, as in Brownian motion, the thermodynamical concepts become meaninglesss. The second law is not "violated", it ceases to be applicable. On the other hand, the statistical treatment is still perfectly adapted to this kind of phenomena; e.g. it can be used to evaluate the average amplitude of the observed fluctuations. In this sense, the scope of statistical mechanics is wider than that phenomenological thermodynamics.

III. Statistics of Closed Systems.

1.- System of given temperature. To ensure that a system S has a definite temperature, we must imagine it in "thermal contact" with a "thermostat", i.e. some other system T whose heat capacity is so large that the exchange of heat with the system S does not sensibly alter its temperature. The two systems S and T, freely exchanging energy with each other, have then a common temperature, which is entirely fixed by the thermostat. From the atomistic point of view, we have a total system ϱ consisting of the system S under investigation and the thermostat T, such that the number of degrees of freedom of S is much smaller than that of T. While the energy of both S and T continually varies, we may assume that the total energy of ϱ remains constant, i. e. that ϱ is isolated. The problem is thus to derive from the statistical distribution of the isolated system ϱ that of a small part S of ϱ.

For this purpose, we consider the phase-spaces of S and T, with measures $d\mu_S$, $d\mu_T$ and the product space of ϱ with measure $d\mu = d\mu_S d\mu_T$. Moreover, although the systems S and T are essentially interacting, we may assume that the magnitude of the interaction is negligible, so that the total energy E is composed additively of the energies E_S, E_T of the systems S and T. We again introduce in the S- and T- spaces and in the product spaces the domains Ω_S, Ω_T, Ω enclosed by the energy surfaces E_T, E_S, E. Now, we are interested in the averages of phase functions pertaining to the system S, i.e. depending only on the phase coordinates P_S. We may write down these averages immediatly as ergodic averages in the total ϱ-space; we shall then seek to transform these expressions so as to eliminate the irrelevant phase coordinates P_T the thermostat will be represented in the final result by only one parameter, which play the part of the temperature.

For the transformation we have in view, it will be convenient to write the ergodic average in the form

$$\overline{f}_E = \frac{1}{\mu(E)} \frac{d}{dE} \int_\Omega f(P)\, d\mu \qquad (1)$$

For a phase function $f(P_S)$ belonging to S we may carry out the integration with respect to the variables P_T; for each phase P_S, to which corresponds an energy $E_S = H_S(P_S)$, and accordingly an energy $E_T = E - H_S$, this gives a a factor $\mu_T[\Omega_T(E - H_S)]$. The integration over P_S extends over the domain enclosed by the surfaces corresponding to the largest value of E_S, viz. E; but we may actually extend it over the whole S - space since for larger values of E_S, the factor μ_T vanishes. Thus,

$$\int f(P_S)\, d\mu = \int f(P_S)\, \mu_T[\Omega_T(E - H_S)]\, d\mu_S .$$

The derivation with respect to E transforms $\mu_T(\Omega_T)$ into the invariant measure of the limiting energy surface $\mu_T(E - H_S)$ and we get from (1)

$$\overline{f(P_S)}\Big|_E = \int f(P_S)\, \frac{\mu_T(E - H_S)}{\mu(E)}\, d\mu_S \qquad (2)$$

We have now expressed the ergodic average of $f(P_S)$ as an average over the phase space of the system S alone, each element of this space being affected with the statistical weight $\mu_T(E - H_S)/\mu(E)$; since this weight factor only depends on the energy H_S, one may say that it affects the energy surfaces or energy shells in S - space. As a result of the contact with the thermostat, the energy E_S of the system S is not fixed, but has the equilibrium distribution given by $\mu_T(E - E_S)/\mu(E)$: we must therefore take all ergodic averages $\overline{f}\big|_{E_S}$ and add them up with the weights corresponding to the distribution of E_S.

2.- Canonical distribution. Our next task is to find a simpler expression for the energy distribution in S - space. This means essentially that we want to transform the factor $\mu_T(E_T)$, for the factor $\mu(E)$ is easily evaluated in terms of it; in fact, by an argument similar to the above, we find

$$\mu(E) = \int dE_S\, \mu_S(E_S)\, \mu_T(E - E_S] \qquad (3)$$

the integration extending over all values of the energy E_S

Now, we note that the structure of the thermostat is entirely arbitrary, apart from the requirement that it be a very large system. We avail ourselves of this latitude to obtain an asymptotic expression for $\mu_T(E_T)$: to this end, we assume the thermostat to consist of a large number of parts T_1, T_2,... whose interactions we may neglect. By making repeated use of a formula of the type (3), we get

$$\mu_T(E_T) = \int dE_{T_1} \int dE_{T_2} \cdots \int dE_{T_{n-1}}\, \mu_{T_1}(E_{T_1})\, \mu_{T_2}(E_{T_2}) \cdots \mu_{T_n}(E_T - E_{T_1} - E_{T_2} - \cdots - E_{T_{n-1}})$$

The formal analogy of this equation with the "law of composition" of probabilities suggested to Khinchin an elegant application to our problem of the "central limit theorem" of the theory of probability.

Consider n stochastic variables x_1, $x_2, \dots x_n$ with independent distributions $u_1(x_1)dx_1$, $u_2(x_2)dx_2, \dots$ The distribution of the sum $x = x_1 + x_2 \pm x_n$ obeys the law of composition.

$$u(x)\,dx = \int dx_1 \int dx_2 \cdots \int dx_{n-1}\, u_1(x_1)\, u_2(x_2) \cdots u_n(\ldots x - x_1 - x_2 - \cdots x_{n-1}) \tag{4}$$

Now, provided that the distribution laws $u_i(x_i)$ satisfy certain conditions the most important being that the mean fluctuations of the variable x_i are finite, the distribution law for x has the asymptotic form (for $n \to \infty$

$$u(x)\,dx \simeq \frac{1}{\sqrt{2\pi B}}\, e^{-\frac{(x-\bar{x})^2}{2B}} \tag{5}$$

where $\bar{x} = \sum_n \bar{x}_n$ is the sum of the average values $\bar{x}_i = \int x_i u_i(x_i)\, dx_i$, i.e. the average value of x, and $B = \sum_n b_n$ is the sum of the mean quadratic fluctuations

$$b_i = \int (x_i - a_i)^2 u_i(x_i)\, dx_i$$

We cannot, however, immediately identify $\mu_{T_i}(E_{T_i})\, dE_{T_i}$ with a distribution law $u_i(x_i)dx_i$ because the measures $\mu_{T_i}(E_{T_i})$, though finite, are not necessarily bounded : they generally increase as some power of the energy. This obstacle is, of course, easily overcome by putting

$$u_i(E_{T_i}) = \frac{1}{Z_i(\beta)}\, e^{-\beta E_{T_i}}\, \mu_{T_i}(E_{T_i}) \tag{6}$$

where

$$Z_i(\beta) = \int e^{-\beta E_{T_i}}\, \mu_{T_i}(E_{T_i})\, dE_{T_i} \tag{7}$$

secures the normalization, and is an arbitrary positive parameter. In fact, the u_i's so defined are bounded and satisfy the law of composition (4); it can be verified in detail+ that they satisfy all the conditions for the validity of the asymptotic formula (5).

The law of distribution $u(E_T)$ derived from the set (6) by the law of composition (4) has exactly the same form as each of the u_i's, the normalization factor $Z(\beta)$ being the product of the Z_i's. We readily get

$$\bar{E}_T = -\frac{d \log Z}{d\beta} \tag{8}$$

and

$$B_T = \frac{d^2 \log Z}{d\beta^2} \tag{9}$$

+ Voir Khinchin, p. 86

It is easily seen [+] that equation (8) umiquely determines β in terms of $\overline{E}_t$: this makes it possible to fix the parameter β by relating it to the given average energy of the thermostat. We shall henceforth assume that β has the value resulting from this equation (8) : it is this parameter which will "represent" the thermostat in the final result, and we shall see that it is simply related to the temperature.

Taking for $\mu (E_T)$ its asymptotic form (7) we get for $\mu_T(E_T)$ the asymptotic expression

$$\mu_T (E_T) \simeq \frac{Z_T(\beta)}{\sqrt{\quad}} e^{\beta E_T - \frac{(E_T - \overline{E}_T)^2}{2 B_T}} \qquad -(10)$$

In applying this formula to the argument $E_T = E - E_S$, we note that $(E_T - \overline{E}_T)^2 = (E_S - \overline{E}_S)^2$. Now, excepting the unliky cases when E_S happens to be widely different from its average $\overline{E}_S$, the quadratic fluctuations of the system S will be of a much smaller order of magnitude that those of the thermostat whose average is given by B_T : we shall accordingly neglect them altogether and write

$$\mu_T (E - E_S) \simeq \frac{Z_T(\beta)}{\sqrt{2\pi B_T}} e^{\beta (E - E_S)}$$

For the factor μ (E) we may write down a completely similar expression since according to (3) the systems S and T are linked together by the same law of composition as the various parts of T : we may therefore treat the total system S as we have just treated the thermostat. Thus,

$$\mu (E) \simeq \frac{Z(\beta)}{\sqrt{2\pi B}} e^{\beta E}$$

and

$$Z(\beta) = Z_T(\beta) \, Z_S(\beta)$$

while B differs from B_T only by the contribution from the small system S which can be neglected. For the energy distribution of the system S we therefore get the very simple expression

$$\frac{\mu_T (E - E_S)}{\mu (E)} \simeq \frac{1}{Z_S(\beta)} e^{-\beta E_S} \qquad (11)$$

in which the thermostat only appears through the parameter β.

In its most general aspect, the property expressed by

[+] Voir Khinchin, p. 77

formula (11) may be stated as follows : any small[+] part of a large system, whose interaction with the system is weak, is distributed in energy according to the law (11). In this form, the theorem is due to Boltzmann, who recognized its fundamental importance for the statistics of non-isolated systems. The law of distribution (11), under the name of <u>canonical distribution law</u>, was extensively studied by Gibbs; the insufficient emphasis put by Gibbs on Boltzmann's theorem is no doubt responsible for the wisdespread confusion about the physical significance of the canonical distribution. The above derivation, due to Khinchin, has the merit of elucidating the deep-lying relation of Boltzmann's theorem to the general principles of statistics.

[+] It need hardly be emphasized that the "small" part can be any macroscopic system; it is small only with respect to its surroundings, which constitute the "thermostat". On the other hand, the small part could also be just an atom of an ideal gas : Boltzmann's theorem then gives an independent derivation of the theory of systems of weakly coupled identical elements of Ch II, § 5.

3. - Energy distribution . The canonical distribution law has important consequences for the average value of the energy. In the first place, the study of this quantity will allow us to establish a very simple relation between the thermostat parameter β and the absolute temperature as usually difined. Further, we shall derive a general property known as "equipartition of the energy". Finally, it is also possible, at least for a certain class of systems, to derive the explicit law of distribution of the energy around the mean and, in particular, the mean quadratic fluctuation of the energy.

The canonical average of the energy of our system S is given by :

$$\overline{E}_s = - \frac{d \log Z_s}{d\beta}$$

If, in particular, the system S is an ideal monoatomic gas, we have (cf. Ch. 11, §6)

$$\mu(E_s) \sim E_s^{\frac{3N}{2}-1}$$

whence

$$Z_s(\beta) = \int e^{-\beta E_s} \mu(E_s)\, dE_s$$

$$\sim \int e^{-\beta x} x^{\frac{3N}{2}-1} dx \sim \beta^{-\frac{3N}{2}}$$

and

$$\overline{E}_s = \frac{3N}{2} \frac{1}{\beta} . \qquad (12)$$

The thermodynamical value of the energy of such a gas is $\frac{3}{2} R \mathcal{T}$ per mole, where R is the gas constant and $\mathcal{T}$ the absolute temperature in the Kelvin scale. This may be written $\frac{3}{2} N\theta$, if N is the number of atoms per mole and $\theta = k\mathcal{T}$, with $k = R/N$, is the measure of the absolute temperature in an energy scale. The conversion factor k is a universal constant to which the name of Boltzmann has been linked (although Boltzmann himself did not introduce it explicitely). The comparison with the statistical formula fixes the relation of the thermostat parameter β with the absolute temperature :

$$\beta = \frac{1}{\theta} \qquad (13)$$

The simple formula (12) is a special case of a more general property. If we can isolate in the Hamiltonian of the system any group of terms H' homogeneous and quadratic in some of the p's, i.e. such that :

$$H' = \frac{1}{2}\sum_{i=1}^{k} p_i \frac{\partial H'}{\partial p_i}$$

we may perform the calculation of the canonical average H' explicitly. In fact, putting H= H' + H'', we may write :

$$\overline{H'} = \frac{1}{Z}\int d\mu'' e^{-\beta H''}\int H' e^{-\beta H'} d\mu' ,$$

where $d\mu = d\mu' d\mu''$ with $d\mu' = dp_1 \ldots dp_k$. Now, for any of the relevant p_i's we have

$$\int p \frac{\partial H'}{\partial p} e^{-\beta H'} dp = -\frac{1}{\beta}\int p \frac{\partial}{\partial p}\left(e^{-\beta H'}\right) dp = +\frac{1}{\beta}\int e^{-\beta H'} dp$$

whence

$$\overline{H'} = \frac{k}{2}\frac{1}{\beta} = k.\frac{1}{2}\theta. \tag{14}$$

The same argument applies to the case that the quadratic homogeneous part of the Hamiltonian belongs to the potential energy, i.e. has this property with respect to some of the q's.

The general formula (14) expresses the law of equipartition of the energy : for a system of temperature θ , every degree of freedom for which the kinetic or potential part of Hamiltonian is homonogeneous and quadratic in the corresponding canonical variable contributes $\frac{1}{2}\theta$ to the average energy of the system. In particular, every degree of freedom of harmonic oscillation contributes θ to the average energy.

If the system S can itself be analysed into a large number of component parts with weak interactions, e.g. if it is an ideal gas, we can apply to the measure of its own energy surfaces $\mu_s(E_s)$ the asymptotic formula that we had derived for the thermostat. The law of distribution of the values E_s of the energy then takes the simple form :

$$w(E_s) = \mu_s(E_s)\frac{\mu_T(E-E_s)}{\mu(E)} = \frac{Z_s(\beta)}{\sqrt{2\pi B_s}} e^{\beta E_s - \frac{(E_s-\bar{E}_s)^2}{2B_s}} . Z_T(\beta) e^{\beta(E-E_s) - \frac{(E_s-\bar{E}_s)^2}{2B_T}} . \frac{\sqrt{2\pi B}}{Z(\beta)} e^{-\beta E}$$

$$= \frac{1}{\sqrt{2\pi B_s^*}} e^{-\frac{(E_s-\bar{E}_s)^2}{2B_s^*}} ,$$

where B_s is defined by

$$\frac{1}{B_s^*} = \frac{1}{B_s} + \frac{1}{B_t} = \frac{B}{B_s B_t}$$

We thus obtain a Gaussian distribution of the energy with a mean quadratic fluctuation B_S^* which is always smaller than B_S : the exchange of energy with the thermostat thus results in a negative correlation of the energy distribution between the two parts of the total system.

The value of the mean quadratic fluctuation B_S is easily computed. One has

$$B_S = \overline{(E_S - \overline{E}_S)^2} = \overline{E_S^2} - (\overline{E}_S)^2$$

$$\overline{E}_S = \frac{-d \log Z}{d\beta}, \qquad \overline{E_S^2} = + \frac{1}{Z} \frac{d^2 Z}{d\beta^2}, \quad \frac{d\overline{E}_S}{d\beta} = -\frac{1}{Z}\frac{d^2 Z}{d\beta^2} + \frac{1}{Z^2}\left(\frac{dZ}{d\beta}\right)^2$$

i. e.

$$B_S = -\frac{d\overline{E}_S}{d\beta} = \frac{d\overline{E}_S}{d\theta}\,\theta^2 \qquad (15)$$

This formula exhibits a fundamental feature common to all fluctuation phenomena : while the product $\theta \frac{d\overline{E}_S}{d\theta}$ is of macroscopic order of magnitude, there is an extra factor θ which contains the essentially atomistic constant k. In fact, fluctuation phenomena are outside the scope of macroscopic physics : they are essentially linked to the atomic theory of matter and radiation. Accordingly, the observation of their macroscopic effects, such as Brownian motion, light scattering, etc., affords a proof of the consistency of the atomic picture and a possibility of determination of atomic dimensions.

4.- Thermodynamics of closed systems. We shall now follow a line of argument parallel to that developped for isolated systems. We shall look for an atomistic interpretation of the characteristic function corresponding to the choice of the temperature and the extremal parameters as independent variables : this is the free energy of the system, defined as[+]:

$$F = \overline{E} - \theta S \qquad (16)$$

In fact, we find

$$dF = - S\, d\theta + A\, da, \qquad (17)$$

i.e.

$$S = -\left(\frac{\partial F}{\partial \theta}\right)_a, \quad A = + \left(\frac{\partial F}{\partial a}\right)_\theta \qquad (18)$$

From (16) and (17), we derive, by making use of the general transforma-

---

[+] We may now drop the index

tion formula : $\theta \frac{\partial f}{\partial \theta} = - \beta \frac{\partial f}{\partial \beta}$,

the <u>Helmholtz relation</u> between mean energy and free energy

$$\overline{E} = F - \theta \frac{\partial F}{\partial \theta} = \frac{\partial}{\partial \beta} (\beta F) \tag{19}$$

Now, a direct calculation gives for the canonical average of the force $(\partial H / \partial a)$

$$A = - \frac{1}{\beta} \frac{\partial \log Z}{\partial a} \quad ;$$

this formula, together with the previous one

$$\overline{E} = - \frac{\partial \log Z}{\partial \beta} \quad ,$$

shows by comparison with the second equation (18) and with (19), respectively, that we must take for F the expression

$$F = - \theta \log Z \; , \tag{20}$$

except for an arbitrary linear function of θ (without physical meaning) Formula (20) completes the atomistic interpretation of the thermal quantities pertaining to closed systems, viz. temperature and free energy. The entropy now appears as a derived quantity :

$$S = \log Z + \beta \overline{E} \tag{21}$$

It is not surprising that this expression for the entropy is formally similar to that obtained in Ch. II, § 6 for the systems of identical elements with weak interactions : for, as already noted (see footnote at the end of § 2), the theory of such systems can be regarded as a special application of that of closed systems.[+] However, the scope of formula (21) is much wider : it applies to any kind of system without restriction. If the system is sufficiently large to allow the use of the asymptotic expression for the measure of its energy surfaces, we may compare the value (21) of the entropy with the logarithm of the measure of the surface corresponding to the <u>mean</u> energy $\overline{E}$, viz.

$$\log \mu (\overline{E}) = \log \left\{ \frac{Z(\beta)}{\sqrt{2\pi B}} e^{\beta \overline{E}} \right\} = S - \log (\sqrt{2 \pi B})$$

The logarithm of the fluctuation term is of a negligible order of magnitude, and we thus get a new interpretation of the entropy in terms of the

---

[+] The system S then represents a single element, and the expression (21) accordingly gives the entropy <u>per element</u>.

measure of a set of phases : this set is not, however, the domain $\Omega(\overline{E})$ enclosed by the energy surface $\overline{E}$, but the surface itself. We have here another illustration of the "insensibility" of the entropy definition.

With the help of (21), we easily verify the well-known thermodynamic property, on which the phenomenological definition of the entropy rests, that the inverse temperature β is an integrating factor of infinitesimal quantity of heat supplied to the system in a quasi-static transformation. This quantity of heat is

$$Q = d\overline{E} - A da$$

whence

$$\begin{aligned} \beta\delta Q &= d\,(\beta\overline{E}) - \overline{E}\,d\beta - \beta A da \\ &= d\,(\beta\overline{E}) + \frac{\partial \log Z}{\partial\beta}\,d\beta + \frac{\partial \log Z}{\partial a}\,da \\ &= d\,(\beta\overline{E}) + \log Z\,) = d\,S \end{aligned}$$

Finally, let us introduce with Gibbs the concept of <u>probability exponent</u> which will give us not only a compact notation for canonical averages, but also exhibit a new aspect of our entropy definition. The canonical average of any phase function

$$\overline{f} = \frac{1}{Z}\int f\,(P)\;{}^{-\beta H}\;d\mu$$

may be written as

$$\overline{f} = \int f(P)\,e^{\eta(P)}\,d\mu \qquad (22)$$

with

$$\eta(P;\,\theta\,,a) = \frac{1}{\theta}\left[F\,(\theta\,,a) - H\,(P;a)\right] \qquad (23)$$

The function η is called by Gibbs the "probability exponent", since

$$n\,(P;\,\theta\,,\,a) = e^{\eta} \qquad (24)$$

represents the density of probability in phase space for the canonical distribution. Now, the entropy, according to (16), is just minus the canonical average of the probability exponent :

$$S = -\,\overline{\eta}\;;$$

this may be written, with the notation (24)

$$S = -\int n \log n\;\;d\mu \qquad (25)$$

Again, we find here a formal analogy with Boltzmann's H-function for systems of weakly coupled elements. The expression (25) may be applied to quite general systems : it does not then refer however, to any actual distribution of physical elements, but to a purely fictitious distribution of "copies" of the system under investigation.

The statistical foundation of thermodynamics just outlined is not yet logically complete. The basic differential equation (17) refers essentially to *quasi-static transformations*. We have still to investigate the behaviour of the entropy when the system is subjected to irreversible transformations. But before this, we have a more immediate task. In the preceding discussion, all macroscopic quantities have been as a matter of course represented by canonical averages : this procedure, however, implies the assumption that the quasi-static transformations do not disturb the canonical form of the equilibrium distribution. Quasi-static transformations are of two distinct types : isothermal or adiabatic. For isothermal transformations, the assumption just mentioned is obviously justified, since the system remains in contact with the same thermostat during the transformation. But the case of adiabatic transformations requires a closer investigation. In fact, in such a transformation, the contact with the thermostat is initially severed, and at the end the temperature of the system has varied, i.e. it is in equilibrium with another thermostat : it is is then far from obvious that the distribution, while both external parameters and temperature varied, nevertheless remained canonical. This is what we shall now proceed to show.

5.- *Permanence of canonical distribution in quasi-static adiabatic transformations*. An adiabatic transformation can be analysed as a succession of steps in which, the system being isolated, some external parameter (or set of parameters) a is varied suddenly by a very small amount da, after which the system is left to itself for a certain time. A second sudden variation of a is then applied and the process is repeated with suitably chosen intervals between the successive variations so as to approximate any continuous variation a(t) of the parameter a. In order for the transformation to be quasi-static, we must, after each small variation of a, allow sufficiant time to elapse for the system for the system to reach the state of equilibrium corresponding to the new value of the parameter : a quasi-static adiabatic transformation corresponds in this sense to an infinitely show variation of a.

For our present purpose, it is of course sufficient to consider a single step, leading from a_o to $a = a_o + da$, and to show that if the initial distribution was canonical, with a certain temperature (or, in Gibbs' terminology, "modulus") θ_o, the final distribution is again canonical, with a slightly different modulus. The variation of a has the effect of deforming the energy surfaces, so that a distribution which was originally ergodic ceases to have this property when referred to the new energy surfaces. Eventually, however, it will become ergodic on these new surfaces, and the new phase density will be given by the ergodic average of the old one, taken on the new energy surfaces. The initial phase density was given by $e^{\eta(\theta_0,a_0;pq)}$; the new one is accordingly $\overline{e^{\eta(\theta_0,a_0;pq)}}\Big|_{E(a)}$ We want to show that this quantity can again be written in the form of a canonical distribution of suitable modulus θ .

To this end, let us first expend the distribution density $e^{\eta(\theta_0,a_0;pq)}$ around the new value a of the parameter :

$$e^{\eta(\theta_0,a_0;pq)} = e^{\eta(\theta_0,a;pq)}\left[1-\frac{da}{\theta_0}\left(A-\frac{\partial H}{\partial a}\right)\right]$$

and accordingly, since the factor $e^{\eta(\theta_0,a;pq)}$ is constant on the new energy surface E(a)

$$\overline{e^{\eta(\theta_0,a_0;pq)}}\Big|_{E(a)} = e^{\eta(\theta_0,a;pq)}\left[1-\frac{da}{\theta_0}\left(A-\overline{A}\Big|_{E(a)}\right)\right]$$

Now, for systems with a large number of degrees of freedom, the canonical distribution has a sharp maximum for the average energy $\overline{E}$ and we may use the expansions

$$\overline{A}\Big|_E = \overline{A}\Big|_{\overline{E}} + \frac{\partial \overline{A}|_{\overline{E}}}{\partial \overline{E}}(E-\overline{E}) + \frac{1}{2}\frac{\partial^2 \overline{A}|_{\overline{E}}}{\partial \overline{E}^2}(E-\overline{E})^2$$

$$A = \overline{A}\Big|_{\overline{E}} \qquad + \frac{1}{2}\frac{\partial^2 \overline{A}|_{\overline{E}}}{\partial \overline{E}^2}\overline{(E-\overline{E})^2}$$

The difference $A - \overline{A}\Big|_E$ will thus involve, besides a term linear in $E - \overline{E}$ a term containing

$$(E-\overline{E})^2 - \overline{(E-\overline{E})^2} = (E-\overline{E})^2 - B = \frac{\partial B}{\partial \overline{E}}(E-\overline{E})$$

owing to the small factor $\partial B / \partial \overline{E}$, this term will be negligible in comparison with the direct one, and we may therefore write

$$\overline{e^{\eta(\theta_0,a_0;pq)}}\Big|_{E(a)} \simeq e^{(\theta_0,a;pq)}\left[1+\frac{\partial a}{\theta_0}\,\frac{\partial \overline{A}|_{\overline{E}}}{\partial \overline{E}}(E-\overline{E})\right]$$

It is now clear that the correction term in $E - \overline{E}$ can be compensated by a suitable change of modulus; for if we pass from θ_o to $\theta = \theta_0 + d\theta$, we have, using Helmholtz's formula,

$$e^{\eta(\theta_0, a; pq)} = e^{\eta(\theta, a; pq)} \left[1 - \frac{d\theta}{\theta_0^2} (E - \bar{E})\right]$$

We conclude that the final distribution is indeed the canonical one of modulus θ, provided that the variation d θ of modulus is related to the variation da by the equation

$$\frac{d\theta}{\theta_0} - \frac{\partial \overline{A}|_{\bar{E}}}{\partial \bar{E}} \, da = 0$$

Replacing $\overline{A}|_{\mathrm{E}}$ by A and θ_0 by θ, and noting, at constant a,

$$\frac{\partial A}{\partial \bar{E}} = \frac{\partial A}{\partial \theta} \Big/ \frac{\partial \bar{E}}{\partial \theta} ,$$

the last equation may be written

$$\frac{1}{\theta} \frac{\partial \bar{E}}{\partial \theta} \partial\theta - \frac{\partial A}{\partial \theta} \, da = 0$$

It is readily verified[*] that this equation just expresses the constancy of the entropy, dS = 0, during the transformation.

6.- Irreversible transformations. The last step in establishing the statistical foundation of thermodynamics consists in verifying that our interpretation of the entropy minus as the average probability exponent of the canonical distribution satisfies the part of the second law referring to irreversible transformations. We must show that for any transformation such that the surrendings of our system to their initial state after it is performed, the entropy of the final state of the system is larger (or at any rate not smaller) than that of the initial state; if we formulate the law in terms of average probability exponents, the direction of the inequality is reversed. It must be observed that a probability exponent which would correspond to an equilibrium state under certain conditions represents an arbitrary density distribution if these conditions are altered. We shall thus have to introduce, following Gibbs, arbitrary probability exponents $\varpi(P)$ and study the behaviour of their average values.

$$\overline{\varpi} = \int \varpi \, e^{\varpi} \, d\mu$$

under the various transformations satisfying the above-mentionned condition.

---

[*] In fact, from (18) it follows that

$$-\frac{\partial A}{\partial \theta} = \frac{\partial S}{\partial a}$$

and from (19) that

$$\frac{1}{\theta} \frac{\partial E}{\partial \theta} = -\frac{\partial^2 F}{\partial \theta^2} = \frac{\partial S}{\partial \theta}$$

All such transformations reduce to two types : (a) the adiabatic transformations, during which the system is thermally isolated, (b) the heat exchange, in which two initially separated systems are brought into thermal contact and separated again. We shall discuss these two types in succession, closely following Gibbs' elegant treatment of the problem.

(a) Adiabatic transformations. During an adiabatic transformation, the distribution changes independently in each energy shell : we shall thus compare distributions with the same total number of "copies" of the system in the same energy shell, and prove that the distributions with probability exponent ϖ_0 constant in each energy shell have a smaller average probability exponent than the others. We call this statement the first lemma of Gibbs. The assumption is

$$\int_{[E]} e^{\varpi} d\mu = \int_{[E]} e^{\varpi_0} d\mu$$

for each energy shell $[E]$. For the average of ϖ_0 over a shell we may thus write

$$\int_{[E]} \varpi_0 e^{\varpi_0} d\mu = \varpi_0 \int_{[E]} e^{\varpi_0} d\mu = \varpi_0 \int_{[E]} e^{\varpi} d\mu = \int_{[E]} \varpi_0 e^{\varpi} d\mu ,$$

and therefore $\bar{\varpi}_0 = \int \varpi_0 e^{\varpi} d\mu$

We must now prove that

$$\bar{\varpi} - \bar{\varpi}_0 = \int (\varpi - \varpi_0) e^{\varpi} d\mu \geqslant 0$$

Gibbs uses here a general procedure applicable to any inequality of the form $-\int x e^{\varpi} d\mu \geqslant 0$. He refers this inequality to a property of the exponential function :

$$e^{x} - 1 - x \geqslant 0 \text{ for any } x.$$

To this end, he uses the normalization conditions to show that the appropriate terms $e^{x} - 1$ associated to $-x$ have the average 0. In the present instance $-x = \varpi - \varpi_0$; from the normalization conditions

$$\int e^{\varpi} d\mu = \int e^{\varpi_0} d\mu = 1$$

it indeed follows that

$$\int e^{\varpi} \left[e^{\varpi_0 - \varpi} - 1 \right] d\mu = 0$$

Adding this to the expression $\int (\varpi - \varpi_0) e^{\varpi} d\mu$, we immediately conclude that this quantity is always $\geqslant 0$.

Gibbs's lemma can now be applied successively to each step da of the adiabatic variation of the parameter a from a_q to a_n (§5).

Calling $\varpi_1, \varpi_2, \ldots, \varpi_n$ the successive probability exponents, we have

$$\bar\varpi_1 \geqslant \bar\varpi_2 \geq \cdots\cdots \geqslant \bar\varpi_n$$

If the transformation is quasi-static, the differences $\bar\varpi_1 - \bar\varpi_2$ are of the second order in da, and since the number of steps is inversely proportional to da, the difference $\bar\varpi_1 - \bar\varpi_n$ after a finite variation of a is infinitesimal : the entropy in this case remains constant. If, on the other hand, the transformation is irreversible, the successive differences will in general be of order da, and for a finite variation of a there will be a finite increase of the entropy.

(b) Heat exchange. To discuss the case of heat exchange we have to consider a system composed of two distinct parts S_1, S_2; we accordingly distinguish the two phase spaces of these parts and the toatl phase space which is their direct product. Between the respective elements of measure we have the relation

$$d\mu_1 \cdot d\mu_2 = d\mu$$

The distribution e^{ϖ} defines the distributions in the S_1, and S_2 spaces as

$$e^{\varpi_1} = \int e^{\varpi} d\mu_2 \qquad e^{\varpi_2} = \int e^{\varpi} d\mu_1 \ ;$$

these partial distributions are, of course, normalized to unity :

$$\int e^{\varpi_1} d\mu_1 = \int e^{\varpi_2} d\mu_2 = \int e^{\varpi} d\mu = 1$$

Gibbs' second lemma states that

$$\bar\varpi \geq \bar\varpi_1 + \bar\varpi_2$$

the inequality holding only if the two distributions ϖ_1, ϖ_2 are entirely independent, so that $\varpi = \varpi_1(P_1) + \varpi_2(P_2)$

To prove the lemma, note that

$$\bar\varpi_1 = \int \varpi_1 e^{\varpi_1} d\mu_1 = \int \varpi_1 e^{\varpi} d\mu \ , \quad \bar\varpi_2 = \int \varpi_2 e^{\varpi} d\mu$$

and therefore

$$\bar\varpi - \bar\varpi_1 - \bar\varpi_2 = \int (\varpi - \varpi_1 - \varpi_2) e^{\varpi} d\mu$$

The normalization conditions give

$$\int e^{\varpi} \left[e^{\varpi_1 + \varpi_2 - \varpi} - 1 \right] d\mu = 0$$

whence the announced inequality follows. Moreover, the equality is seen to hold only if $\varpi - \varpi_1 - \varpi_2 = 0$, as stated.

With the help of this lemma, the discussion of the thermal contact of two systems is immediate. Before the contact, we have independent distributions with exponents ϖ_1, ϖ_2 and the exponent of the total distribution is $\varpi = \varpi_1 + \varpi_2$. As a result of the thermal contact (during which the total system is isolated) the exponent changes from ϖ to ϖ' and by the first lemma

$$\bar\varpi \geq \bar\varpi'$$

Now, we had on the one hand

$$\bar\varpi = \bar\varpi_1 + \bar\varpi_2,$$

and on the other, by the second lemma

$$\bar\varpi' \geq \bar\varpi_1' + \bar\varpi_2',$$

ϖ_1', ϖ_2' being the exponents of probability of the two systems in their final states. Therefore

$$\varpi_1 + \varpi_2 \geq \varpi_1' + \varpi_2' \qquad (26)$$

i.e. the heat exchange has caused the sum of the entropies of the two systems to increase.

It is interesting to analyse somewhat more closely the role of the temperature in the phenomena of heat exchange. The basis for such an analysis is supplied by a <u>third lemma of Gibbs</u> : If θ is a positive parameter, the quantity $\bar\varpi + (\bar E / \theta)$ is minimum for the canonical distribution of modulus θ. For the latter distribution, the quantity in question is just F / θ ; the lemma thus expresses a minimum property of the free energy of a system of given temperature, which is parallel to the maximum property of the entropy of an isolated system. We have to prove

$$\int \left(\varpi + \frac{E}{\theta}\right) e^{\varpi} d\mu - \frac{F}{\theta} \geq 0 , \quad \text{i.e.} \quad \int (\varpi - \eta)\, e^{\varpi} d\mu \geq 0 ,$$

η being the probability exponent of the canonical distribution of modulus θ. By the usual procedure, we derive from the normalization conditions

$$\int e^{\varpi} \left[e^{\eta - \varpi} - 1 \right] d\mu = 0$$

whence the preceding inequality follows.

Suppose one of the two systems just considered is a "thermostat" of temperature θ_2. Besides the inequality (26) we may now write, by the third lemma

$$\bar\varpi_2' + \frac{\bar E_2'}{\theta_2} \leq \bar\varpi_2 + \frac{\bar E_2}{\theta_2}$$

since the initial state was described by a canonical distribution of modulus θ_2. Therefore

$$\bar\varpi_1 - \bar\varpi_1' \geq \frac{\bar E_2' - \bar E_2}{\theta_2} , \qquad (27)$$

or in terms of thermal quantities

$$S_1' - S_1 \geq \frac{Q}{\theta_2} ,$$

i.e. the entropy increase of a system brought in contact with a thermostat of temperature θ_2 is at least Q/θ_2, where Q is the quantity of heat passing tfrom the thermostat to the system. Using

$$\bar{E}_2 - \bar{E}'_2 = \bar{E}'_1 - \bar{E}_1$$

we can also write the inequality (27) in the form

$$\bar{\varpi}_1 + \frac{\bar{E}_1}{\theta_2} \geqslant \bar{\varpi}'_1 + \frac{\bar{E}'_1}{\theta_2}$$

showing how the quantity $\bar{\varpi} + (\bar{E}/\theta_2)$ pertaining to the system S_1 decreases after thermal contact with the thermostat, until, by repeated or prolonged contact, the canonical distribution of modulus θ_2 is reached, and with it the minimum value of the quantity in question, which is then the free energy of the system.

We may finally consider a system S_0 undergoing a cycle of transformations in which it comes successively in contact with various thermostats of temperature θ_1, θ_2.... Assuming S_0 to be initially at temperature θ_0, we have

$$\bar{\varpi}_0 + \bar{\varpi}_1 + \bar{\varpi}_2 + \ldots \geqslant \bar{\varpi}'_0 + \bar{\varpi}'_1 + \bar{\varpi}'_2 + \ldots$$

$$\bar{\varpi}'_i + \frac{\bar{E}'_i}{\theta_i} \geq \bar{\varpi}_i + \frac{\bar{E}_i}{\theta_i} \qquad (i = 0, 1, 2, \ldots)$$

whence

$$\frac{\bar{E}_1 - \bar{E}'_1}{\theta_1} + \frac{\bar{E}'_2 - \bar{E}_2}{\theta_2} + \ldots \leq 0 \qquad (28)$$

the term pertaining to the system S_0 has disappeared since $\bar{E}'_0 = \bar{E}_0$ owing to the cyclic character of the transformation. The inequality (28) coincides with an important theorem proved by Clansius and used by him for establishing on a phenomenological basis the concept of entropy. If there are only two thermostats, we have the kind of idealized engines considered by Carnot. Calling Q_1, Q_2 the algebraic quantities of heat supplied by the thermostats during a cycle, inequality (28) takes the form

$$\frac{Q_1}{\theta_1} + \frac{Q_2}{\theta_2} \leq 0$$

If the heat transfers are irreversible and not accompanied by any mechanical work, e.g. if heat is conducted or radiated through the system S_0, we have $Q_1 = -Q_2$ and the inequality merely expresses the fact that heat passes irreversibly from the hot to the cold thermostat. If an amount of mechanical work $W = Q_1 + Q_2$ is supplied by the engine ($\theta_1 > \theta_2$; $Q_1 > 0$), we find for the efficiency the wellknown expression $\frac{W}{Q_1} \leq \frac{\theta_1 - \theta_2}{\theta_1}$, the equality corresponding to Carnot's reversible engine.

IV.- STATISTICS OF OPEN SYSTEMS.

1. Phase functions and averages for open systems. Our last object of investigation is the statistical description of systems consisting of a variable number of elements. Since the main application of this theory is the study of chemical reactions, we shall speak of systems of "molecules". We start from the μ -spaces of the single molecules, from which we construct by direct product the Γ-spaces corresponding to any given numbers of molecules of the various species occuring in the system considered. In the following, we shall only treat the case of a mixture of two distinct constituents, which is sufficiently typical. We denote by P_{N1N2} the set of coordinates of the Γ -space corresponding to N_1 molecules of the first species and N_2 molecules of the other.

The definition of phase functions requires some care. In general, a physical quantity pertaining to a system of N molecules of the same species may be expressed in terms of other quantities attached either to a single molecule or to a pair of molecules or, more generally, to a cluster of any number of molecules. It can thus be written in the form

$$f\,(P_N) = \sum_{i=1}^{N} f^{(1)}(P_i) + \sum_{i=1}^{N}\sum_{k=1}^{N} f^{(2)}(P_i, P_k) + \dots\dots \qquad (1)$$

where P_i denotes a phase in the μ -space of the i-th molecule. The extension of this definition to the case of a mixture of molecules is immediate ; it gives a well-defined meaning to the notation $f\,(P_{N1N2})$.

In order to compute the statistical average of such a phase function for an open system, we follow the same line of argument as for closed systems ; we consider our system in conjuction with the requisite number of very large reservoirs of molecules of the different species, so that the total system may be regarded as closed. For the closed system, which we assume to have a definite temperature, we may write down the canonical average of the phase function. there then remains the task of eliminating the variables pertaining to the reservoirs. The exchange of molecules between the system S containing a mixture of different species and any reservoir R containing one definite species must be imagined to take place through a suitable semi-permeable membrane : the physical existence of such membranes, however, is not relevant to the argument; it suffices that the concept of semi-permeable membrane be logically consistent with the basic assumptions of the atomic theory of matter.

The phase functions referred to the closed system $S + R_1 + R_2$ are themselves statistical averages over all passible distributions of the molecules between the system S and the appropriate reservoir. With respect to each species of molecules, the system S can be regarded as a small part of the corresponding large reservoir ; the law of distribution of molecules in S therefore takes the asymptotic form of a Poisson distribution.[*]

$$w(N) = \frac{\bar{N}^N}{N!} e^{-\bar{N}} \qquad (2)$$

$\bar{N}$ being the average number present in the system S. Thus, the phase function whose canonical average is needed is

$$\sum_{N_1,N_2} w(N_1)\, w(N_2)\, f(P_{N_1N_2}), \qquad (3)$$

with the definition (1) of $f(P_{N1N2})$. Here, $f(P_{N1N2})$ de pends only on the variables of the system S. We have also to consider the total Hamiltonian of the closed system $S + R_1 + R_2$, which we may take to be the sum of the Hamiltonians of the three parts : $H + H_{R1} + H_{R2}$. For any distribution of the molecules, characterized by the numbers N_1, N_2, we have, more explicitly,

$$H(P_{N1N2}) + H_{R1}(P_{N1}') + H_{R2}(P_{N2}') \qquad (4)$$

where P_{N1}' denotes the phase of the reservoir R_i containing N_i' molecules. If $\mathcal{N}_i$ represents the total number of molecules of species i, we have $N_i' = \mathcal{N}_i - N_i$, so that the phase function of the more general type (4) is also specified by the numbers N_1, N_2 of molecules in the system .

The statistical weight of the phase P_{N1N2} of the system S is therefore proportional to

$$d\mu_{N_1N_2}\, e^{-\beta H(P_{N_1N_2})} \int e^{-\beta[H_{R_1}(P_{N_1'}) + H_{R_2}(P_{N_2'})]}\, d\mu_{N_1'}\, d\mu_{N_2'},$$

i.e.

$$d\mu_{N_1N_2}\, e^{-\beta H(P_{N_1N_2})}\, Z_{R_1}(\mathcal{N}_1 - N_1)\, Z_{R_2}(\mathcal{N}_2 - N_2),$$

where $Z_{R_i}(N_i')$ represents the "sum over states" for the reservoir R_i containing N_i' molecules. Denoting the corresponding sum over states for the system by $Z(N_1, N_2)$, we finally get for the canonical average of any phase function pertaining to the system S the expression

---

[*] We might even take a Gaussian distribution, but the choice of the Poisson distribution is somewhat more general and fits in better with the thermodynamical formulae.

$$\bar{f} = \frac{\sum_{N_1,N_2} w(N_1)w(N_2) Z_{R_1}(\mathcal{N}_1-N_1) Z_{R_2}(\mathcal{N}_2-N_2) \int f(P_{N_1,N_2}) e^{-\beta H(P_{N_1,N_2})} d\mu_{N_1 N_2}}{\sum_{N_1,N_2} w(N_1)w(N_2) Z_{R_1}(\mathcal{N}_1-N_1) Z_{R_2}(\mathcal{N}_2-N_2) Z(N_1,N_2)}$$

In this formula, the parameters defining the external conditions under which the system is considered are, besides the temperature and the non-thermal parameters of type a, the average numbers $\bar{N}_1$, $\bar{N}_2$ of molecules in the system (which, however, are not necessarily all independent).

2.- Chemical potentials. The problem now confronting us is to find some simple asymptotic expression for the dependence of the function $Z_{R_i}(N_i')$, representing the reservoirs in formula (5), on the numbers N_i' of molecules contained in them. Since the total numbers $\mathcal{N}_i$ can be made arbitrarily large, we are actually interested only in small relative deviations $(N_i' - \bar{N}_i') / \bar{N}_i'$ of the numbers N_i' from their average values $\bar{N}_i' = \mathcal{N}_i - \bar{N}_i$. Even the absolute fluctuations $N_i' - \bar{N}'_i = -(N_i - \bar{N}_i)$ may be restricted in magnitude; for the contributions of terms corresponding to large fluctuations are in any case cut down by the distribution factors $w(N_i)$. The situation here is entirely similar to that we had when discussing the energy fluctuations in the interaction of a closed system with a thermostat. We have thus to compare each $Z_{R_i}(N_i')$ with its average value $Z_{R_i}(\bar{N}_i')$ for relatively small differences $N'_i - \bar{N}_i'$, and in order first to obtain the dependence of the latter function on the average number $\bar{N}_i'$ we shall make use of the fact that log. $Z_{R_i}(\bar{N}_i')$ is directly connected with the free energy of the system R_i. The following argument is essentially due to Tolman [*].

To present the matter as simply as possible, we shall first consider in general terms the question of how the macroscopic variables describing a system of identical elements depend on the number of these elements. From this point of view the variables fall into two classes. Some of them, called intensive variables, are independent of the number of elements they characterize properties which belong to any part of the system and are uniform throughout; such are, for instance, the pressure and the temperature. Other variables, the extensive ones, are proportional to the number of elements; they describe those properties of the homogeneous system which are due to additive contributions from the various par of it; examples of extensive variables are the volume, the mean energy, and the characteristic functions : entropy and free energy. Extensive variables give rise to

---

[*] Cf R. Tolman, Phys. Rev. 57, 1160, 1940

corresponding "densities", which are again intensive variables : the density is the partial derivative of the extensive variable with respect to the number of elements, provided that the other independent variables, kept constant in the derivation, are all intensive.

It is always possible to have all independent variables intensive, for each pair a, A consists of an extensive and an intensive variable, either of which can be chosen as representative of the corresponding physical phenomenon. Thus, we may take either the volume or the pressure as the mechanical variable of a body of isotropics structure; there are accordingly two different "free energies" $F(\theta, V)$ and $G(\theta, p)$ which are usually associated, in a rather arbitrary fashion, with the names of Helmholtz and Gibbs, respectively. In general, the passage from the characteristic function $F(\theta, a)$ to the function $G(\theta, A)$ is affected by a "Legendre transformation" :

$$G(\theta, A) = F\{\theta, a(A)\} - a(A) . A,$$

where $a(A)$ is the inversion of the function $A(a)$ computed from $F(\theta, a)$ Thus, $G(N; \theta, p) = F(N; \theta, V) + Vp$, and the corresponding density is

$$g(\theta, p) = \left(\frac{\partial G}{\partial N}\right)_{\theta, p} \qquad (6)$$

so that

$$G(N; \theta, p) = N g(\theta, p) \qquad (7)$$

From the connexion

$$\log Z(N; \theta, p) = - \beta G(N; \theta, p)$$

between the statistical function log Z and the characteristic function G we derive immediately

$$\log Z(N; \theta, p) = - N \beta g(\theta, p)$$

i.e. log Z is an extensive function. This formula, it is true, is not strictly applicable to the case of our reservoirs, because the variation of the number of molecules in such a reservoir does not occur in an externally controlled way, at constant pressure, but as the result of random fluctuations. The pressurein the reservoir is regulated by the mechanical equilibrium conditions across the semi-permeable membrane through which the exchange of molecules with the system S takes place; its value is thus a function of the average numbers $\overline{N}_1$, $\overline{N}_2$ of molecules in the system S, and every deviation from these averages brings about a perturbation of the equilibrium. For small deviations, however, we may write :

$$\log \frac{Z_{R_i}(N_i',\theta,p_i)}{Z_{R_i}(\bar{N}_i',\theta,p_i)} = -(N_i'-\bar{N}_i')\beta g_i(\theta,p_i)$$

$$= (\bar{N}_i - N_i)\beta g_i(\theta,p_i),$$

where the reservoir pressure p_i is a function of the pressure p of the system $\mathcal{S}$ and of the average numbers $\bar{N}_1$, $\bar{N}_2$. Apart from a factor independent of N_i, and which therefore cancels out in the average (5) of the phase function f, we have the required asymptotic expression

$$Z_{R_i}(\mathcal{N}_i - N_i) \sim e^{-\beta g_i N_i}$$

Combining this with the factor $\mu'(N_i)$, given by (2); we may say that in the average (5), each number N_i is weighted by a factor of the form

$$\frac{1}{N_i!} e^{\beta \zeta_i N_i}$$

where the function $\zeta_i(\theta, p; \bar{N}_1, \bar{N}_2) = -g_i + \theta \log \bar{N}_i$ represents the reservoir R_i in the same way as the parameter θ represents the thermostat. In view of its fundamental importance for the determination of chemical equilibria, it is usually called the chemical potential of the molecular species i in the system $\mathcal{S}$.

The average (5) of the phase function f now takes the form

$$\bar{f} = \frac{\sum_{N_1,N_2} \frac{1}{N_1!N_2!} e^{\beta(\zeta_1 N_1 + \zeta_2 N_2)} \int f(P_{N_1N_2}) e^{-\beta H(P_{N_1N_2})} d\mu_{N_1N_2}}{\sum_{N_1,N_2} \frac{1}{N_1!N_2!} e^{\beta(\zeta_1 N_1 + \zeta_2 N_2)} Z(N_1,N_2)} . \quad (8)$$

The phase integrations indicated by $d\mu_{N_1N_2}$ are over specified phases, and the factor $(N_1!\,N_2!)^{-1}$ has the effect of reducing them to integrations over generic phases, in which each physical configuration of the system is counted only once. The introduction of this reduction factor was presented by Gibbs as a matter of convention, justified by its success, and it has since given rise to much confuse discussion : the present derivation[**] shows how naturally it comes in when a physical point of view is adopted, in contrast to the more abstract attitude of Gibbs.

[**] On this point even Tolman's paper, which we followed in all other particulars, is not sufficiently explicit.

---

[*] It is easily found (see : Tolman's paper) that the effect of the perturbation of the equilibrium is of the order $(N_i' - \bar{N}') / \bar{N}_i'$

According to formula (8), the average f can be described as a canonical average over generic phases for each possible set of numbers of molecules, weighted by an exponential distribution factor exp $[\beta(\xi_1 N_1 + \xi_2 N_2)$ there is a complete analogy with the concept of canonical average itself which is an ergodic average over each energy surface, weighted by an exponential factor exp $[-\beta H]$, Hence the somewhat awkward nomenclature proposed by Gibbs : the ordinary canonical distribution of copies of a closed system forms a "petit ensemble"; the distribution just found for an open system is a "grand ensemble". It is often, more simply called a grand canonical distribution.

3.- Thermodynamics of open systems. The statistical interpretation of the thermodynamical functions for open systems is a straight forward extension of the theory of closed systems, but it adds an aspect of fundamental importance in chemistry, viz. the way in which the characteristic functions depend on the numbers of molecules of various species present in the system; We are here concerned, of course, with the equilibrium states of the system, and therefore with the average numbers of molecules. Let us assume, for definiteness, that the independent mechanical variable is the pressure; the characteristic function will thus be the Gibbs free energy.

Let us put

$$\mathcal{Z}(\bar{N}_1, \bar{N}_2) = \frac{1}{N_1!\,N_2!} Z(N_1, N_2) \qquad (9)$$

For a system with fixed numbers of molecules N_1, N_2, we may define the free energy by

$$G(N_1, N_2) = -\theta \log \mathcal{Z}(N_1, N_2): \qquad (10)$$

this differs from our previous definition only by a constant term without interest so long as the numbers of molecules do not vary. Let us now consider the denominator of our fundamental formula (8) :

$$e^{-\Omega} = \sum_{N_1, N_2} \mathcal{Z}(N_1, N_2)\, e^{\beta(\xi_1 N_1 + \xi_2 N_2)} \quad , \qquad (11)$$

and differentiate it N_1, N_2 with respect to all macroscopic variables: this differentiation includes the average numbers $\bar{N}_1$, $\bar{N}_2$, but leaves out the numbers N_1, N_2, over which a summation is performed. We get

$$-d\Omega = e^{\Omega}\sum_{N_1,N_2} \mathcal{Z}(N_1,N_2)\, e^{\beta(\xi_1 N_1+\xi_2 N_2)}\Big[d\log \mathcal{Z}(N_1,N_2) + N_1\, d(\beta\xi_1) + N_2\, d(\beta\xi_2)\Big]$$

$$= -d\,\overline{[\beta G(N_1,N_2)]} + \overline{N_1}\, d(\beta\xi_1) + \overline{N_2}\, d(\beta\xi_2),$$

where the averages refer to the distribution of the numbers N_1, N_2 given by the probability law

$$e^{\Omega}\, \mathcal{Z}(N_1, N_2)\, e^{\beta(\xi_1 N_1+\xi_2 N_2)}$$

The last equation may also be written

$$d\{\Omega + \beta\xi_1\bar{N}_1 + \beta\xi_2\bar{N}_2] = d\,\overline{[\beta G(N_1,N_2)]} + \beta\xi_1\, d\bar{N}_1 + \beta\xi_2\, d\bar{N}_2 \qquad (12)$$

which shows that the characteristic function in terms of the independent variables θ, p, $\bar{N}_1$, $\bar{N}_2$ is $\Omega + \beta(\xi_1\bar{N}_1 + \xi_2\bar{N}_2)$.

If, in the right hand side of equation (11) we replace the sum over the N_i's by the single term corresponding to the values $\bar{N}_1$, $\bar{N}_2$ of these numbers, we derive from it an approximate form of the characteristic function,

$$\Omega + \beta(\xi_1\bar{N}_1 + \xi_2\bar{N}_2) \approx \beta\, G(\bar{N}_1, \bar{N}_2) \qquad (1.)$$

according to (10). To this approximation, which is justified provided that the numbers $\bar{N}_1$, $\bar{N}_2$ are very large, we thus see that the characteristic function is still the freeeenergy [*], taken for the average numbers of molecules, even when these average numbers are varied independently. Now; however, we see that it is quite essential to introduce the factor $(N_1!\, N_2!)^{-1}$, whose role is to reduce the "sum over states" to the physically distinct states of the system, represented by generic phases. Indeed, it is only when the integration is reduced to generic phases that the free energy of a system of identical elements preserves its extensice character when a subdivision of the system into parts is taken into consideration. For if we make such a subdivision into two parts containing N_A and N_B elements, respectively, and consider the sums over state Z ($N_A + N_B$), Z (N_A), Z (N_B), integrated over all specific phases of the total system and of the two parts separately, we have

$$Z\,(N_A + N_B) = \frac{(N_A + N_B)!}{N_A!\, N_B!}\, Z\,(N_A)\, Z\,(N_B)$$

---

[*] The factor β can evidently be retained or omittes without changing the property of the function of being a "characteristic" one.

i.e. precisely

$$\mathcal{Z}(N_A + N_B) = \mathcal{Z}(N_A)\,\mathcal{Z}(N_B).$$

The argument is immediately extended to the more general case we are considering of homogeneous systems containing several species of molecules. We must only generalise the concept of extensive and intensive functions : these will denote homogeneous functions of the numbers of molecules, of degree 1 and 0, respectively.

Using the approximation (13) we get from (12) the fundamental relation

$$\zeta_i = \left(\frac{\partial G}{\partial \bar{N}_i}\right)_{\theta, p;\{\bar{N}_i\}} \tag{14}$$

the symbol $\{\bar{N}_i\}$ indicating that all $\bar{N}_k$'s except $\bar{N}_i$ are kept constant in the derivation. The extensive character of G allows us to write

$$G = \sum_i \bar{N}_i \frac{\partial G}{\partial \bar{N}_i} = \sum_i \bar{N}_i \zeta_i \tag{15}$$

Comparing (15) with (13), we see that, to the approximation considered, $\Omega \approx 0$. This means that the denominator of formula (8) pratically reduces to unity, which considerably simplifies this formula.

An interesting feature of the free energy for a mixture of molecules is that it does not reduce to the expression for a single species if the different kinds of molecules are identified. Take e.g. the simple case of a mixture of two species of molecules in the ideal gas state. We have

If we identify the molecules $Z_1, = Z_2 = Z$, the two first terms reduce to the expected form - $(N_1 + N_2) \log Z$, but the last one is different from log $(N_1 + N_2)$. This "Gibbs paradox" shows how essential it is for a consistent treatment of systems of various kinds of elements that theses elements be distinguished by discontinuous criteria. Of course, the selection of those marks which will be used to distinguish different species is a matter of convention, to be decided according to the circumstances of the concrete problem at hand. Thus, in ordinary chemical reactions, isotopes must be treated as identical elements, whereas in questions of isotope separation they will naturally be distinguished into different species by taking into consideration the mass differences which are neglected in principle for the definition of chemical species.

4.- Characteristic functions involving the temperature. In the preceding section, we have chosen as independent variables, besides the temperature, the pressure and the numbers of molecules : we have seen that the corresponding characteristic function is the free energy $G(\theta, p, \bar{N})$ in the sense of Gibbs. Always keeping the temperature as the independent thermal parameter, we have in principle 3 other possible combinations for the mechanical and chemical variables, according as we choose the volume instead of the pressure and the chemical potentials instead of the numbers of molecules. It is easy to construct the characteristic functions for all these cases, by applying the suitable Legendre transformations. To begin with, we can eliminate the choice of θ, p, ζ as independent variables, for the characteristic function would then become $G - \sum_i \bar{N}_i \zeta_i$, i.e. by (15), identically zero. There accordingly remain 3 possibilities, viz.[*]

$$
\begin{aligned}
G(\theta, p, \bar{N}) &= \sum_i \bar{N}_i \zeta_i \\
F(\theta, V, \bar{N}) &= \sum_i N_i \zeta_i - pV \qquad (16)\\
\Omega(\theta, V, \zeta) &= -pV
\end{aligned}
$$

The property of being a characteristic function is expressed in the three cases by similar equations :

$$
\begin{aligned}
dG &= -S\,d\theta + V\,dp + \sum_i \zeta_i\, d\bar{N}_i \\
dF &= -S\,d\theta - p\,dV + \sum_i \zeta_i\, d\bar{N}_i \qquad (17)\\
d\Omega &= -S\,d\theta - p\,dV - \sum_i \bar{N}_i\, d\zeta_i
\end{aligned}
$$

Thus, we see that the chemical potentials can be defined by

$$\zeta_i = \left(\frac{\partial F}{\partial \bar{N}_i}\right)_{\theta, V;\{\bar{N}_i\}} \qquad (18)$$

just as well as by (14). On the other hand, we have

$$\bar{N}_i = -\left(\frac{\partial \Omega}{\partial \zeta_i}\right)_{\theta, V;\{\zeta_i\}} \qquad (19)$$

The statistical definitions of the characteristic functions Ω, F, G, are likewise quite parallel : the present common feature of

---

[*] The function Ω in (16) has a different meaning from the function denoted by the same letter in section 3. Henceforth, Ω will always denote the function defined in (16).

being simply related to the normalisation factors of the respective statistical distributions. The definition of Ω is in fact given by

$$e^{-\beta\Omega} = \sum_{N_1 N_2 \cdots} \frac{e^{\beta \sum_i \zeta_i N_i}}{N_1!\, N_2! \cdots} \int e^{-\beta H(P_{N_1 N_2 \cdots};\, V)}\, d\mu_{N_1 N_2 \cdots} \tag{20}$$

where the Hamiltonian is expressed in terms of the volume as the external mechanical variable; the proof is immediately obtained* by computing the differential of Ω with respect to the independent variables θ, V and the ζ_i's. From (20) and (16) we derive for the function F the equation

$$e^{-\beta F} = \sum_{N_1 N_2 \cdots} \frac{e^{\beta \sum_i \zeta_i (N_i - \bar{N}_i)}}{N_1!\, N_2! \cdots} \int e^{-\beta H(P_{N_1 N_2 \cdots};\, V)}\, d\mu_{N_1 N_2 \cdots} \tag{21}$$

if we neglect the fluctuations of the numbers N_i around their averages $\bar{N}_i$ this expression reduces to

$$e^{-\beta F} \approx \frac{1}{\bar{N}_1!\, \bar{N}_2!} \int e^{-\beta H(P_{\bar{N}_1 \bar{N}_2 \cdots};\, V)}\, d\mu_{\bar{N}_1 \bar{N}_2 \cdots} \tag{22}$$

Now, the corresponding rigorous and approximate expressions for the function G are the same as those for F, except that the Hamiltonian must now be expressed in terms of p. This corresponds, for the macroscopic quantities, to the passage from the "energy" to the "enthalpy" of the system. From the atomistic point of view, we have

$$H(P;\, p) = H(P;\, V) + p\bar{V} \tag{23}$$

the physical meaning of this relation is that in this passage we change the definition of the mechanical system considered. In fact, if the volume is given, the Hamiltonian is simply the energy of the system of molecules enclosed in a fixed container. If the pressure is given, we must imagine that, for instance, one wall of the container is a movable piston, upon which the external pressure is exerted: this piston is now part of the system and contributes a term pV to the Hamiltonian (where V is now regarded as a function of p and the other independent macroscopic variables.)

---

* The situation is very similar to the above discussion of equation (11); but there, we had chosen the pressure as independent variable, which eventually led to the result that the other characteristic function Ω defined by (11) vanishes.

Seconde Partie

Chapter I.- QUANTAL DESCRIPTION OF PHYSICAL SYSTEMS.

1.- The states of a physical system. We shall here briefly recall the main features of the quantal mode of description of physical systems in the form best adapted to the application of statistical considerations; we shall especially have in view the definition of an invariant measure in the "space" in which the states of such systems are represented.

As is well-known, the most general way of characterizing the state of a system is by a complex vector in a Hilbert-space, i.e. a linear space in which the operation of scalar product of two vectors is defined. This operation plays a fundamental part in the physical interpretation : it makes possible the normalization of state vectors, by equating the scalar product of the vector by itself to unity; for two such normalized vectors, the square of the modulus of the scalar product gives the probability of ascertaining one of the states when the system is in the other.

Finally, one can specify the Hilbert space somewhat further by assuming that it contains complete orthogonal systems of state vectors; such systems form enumerable sequences of vectors. Every state vector can be expressed as a superposition of all the vectors of any complete orthogonal system. Let (f,g) denote the scalar product of any two vectors f, g; it is defined in such a way that $(g,f) = (f,g)^*$, the complex conjugate of (f,g). An orthogonal system φ_i of normalized state vectors is defined by the equations $(\varphi_i, \varphi_k) = \delta_{ik}$. If it is complete, we may write any state vector f in the form

$$f = \sum_i (f, \varphi_i)\, \varphi_i . \tag{1}$$

---

[*] For such a treatment, see von Neumann's treatise.

2.- Physical quantities and operators in Hilbert space.

Physical quantities are represented by linear operators in Hilbert space : the quantity A acts upon the state vector f to transform it into another state vector Af, whose physical interpretation is fixed by saying that the scalar product (Af,f) represents the expectation value of the quantity A in the state f. In order that this expectation value be real, the operator A must satisfy the condition (Af, f) = (f, Af) for any f; such operators are called Hermitian. Physical quantities always correspond to Hermitian operators.

Besides the expectation value of a physical quantity in a given state, one can define the distribution of the possible values of this quantity in the state in question. This is the problem of main interest to us, since the distribution will give us the specification of measure we need for the study of statistical averages of the most general kind. We shall therefore go into some detail about it, without, however, aiming at a complete treatment.[*] We start with a class of Hermitian operators such that it is possible to find states in which they have a definite value : the eigenstates and corresponding eigenvalues of the operators. An eigenstate φ_i of the operator A, with eigenvalue a_i, is defined by

$$A\varphi_i = a_i\varphi_i \qquad (2)$$

From equation (2) it follows that any two eigenstates corresponding to distinct eigenvalues are orthogonal. Further, if an eigenvalue is degenerate, i.e. corresponds to a finite set of eigenstates, such a set can always be "orthogonalized". The total set of eigenstates can therefore be regarded as a complete orthogonal system; this means that the eigenvalue like the eigenfunctions, form an enumerable sequence.

Operators with an enumerable set of "discrete" eigenvalues are not the most general type associated with physical quantities; but the further discussion of their properties will lead quite naturally to the required generalization. If we expand the state vector f according to (1) in terms of the eigenfunctions of the operator A, we get for the expectation value of A in state f

---

[*] For such a treatment, see Von Neumann's treatise.

$$(Af, f) = \sum_i a_i \left|(f, \varphi_i)\right|^2$$

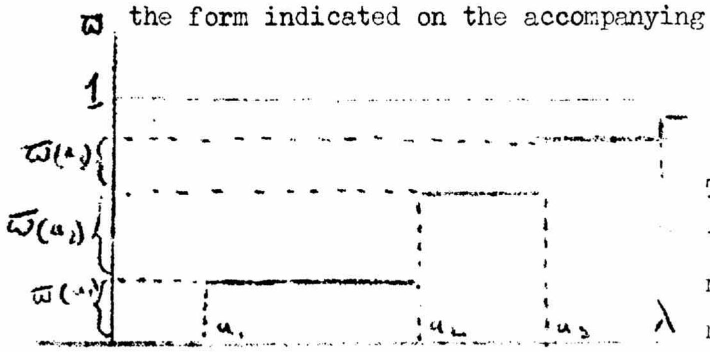
Assuming, for simplicity, all eigenvalues non-degenerate,[*] this relation expresses the fact that the quahtity

$$\varpi(a_i ; f) = \left|(f, \varphi_i)\right|^2$$

represents the probability of finding the eigenvalues a_i of A for a system in the state defined by f; one has, of course, the right normalization for probabilities $\sum_i \varpi(a_i ; f) = 1$. For our class of operators with discrete eigenvalues, the statistical weights $\varpi(a_i ; f)$ may be used to fix a measure in the sense of Stieltjes; i.e., we may write the expectation walue of A in the state f in the form of a Stieljes integral extended over the whole range of real values.

$$(Af, f) = \int_{-\infty}^{+\infty} \lambda \, d\varpi(\lambda ; f) \qquad (3)$$

where the set function $\varpi(\lambda ; f)$ is a discontinuous step function of the form indicated on the accompanying graph :



$$\varpi(\lambda, f) = \sum_{\text{all } a_i \leq \lambda} \varpi(a_i ; f) \qquad (4)$$

The interest of formula (3) is that it can be upheld with a determination of the measure $\varpi(\lambda ; f)$ more general than that give by (4)

In order to arrive at this extension, we shall write formula (3) in a form in which the arbitrary state vector f does not appear explicitely. For this purpose, we associate with any state vector φ the <u>projection operator</u> P_φ defined by

$$P_\varphi f = (f, \varphi)\, \varphi \qquad (5)$$

With this notation, we may write, for any operator A with discrete eigenvalues

$$A = a_i P_{\varphi_i} , \qquad (6)$$

or, symbolically,

$$A = \int_{-\infty} \lambda \, d\, U(\lambda) \qquad (7)$$

---

[*] The way to take account of any degeneracy is quite obvious. Our fundamental formulae (4) and (8) cover this caze as well.

with
$$U(\lambda) = \sum_{\text{all } a_i \leq \lambda} P_{\varphi_i} \qquad (8)$$

The meaning of the symbolical expressions (7) and (8) is given (3) and (4), we may say that for any state vector f, the measure $\varpi(\lambda; f)$ is the expectation value of the operator $U(\lambda)$ in this state. The normalization relation may be written

$$\int_{-\infty}^{+\infty} d\varpi(\lambda; f) = 1,$$

or, in operator form, $$\int_{-\infty}^{+\infty} dU(\lambda) = 1$$

this explains the name "decomposition of unity" given by Von Neumann to the operator function $U(\lambda)$.

Now, decomposition of unity need not be restricted to the form (8), corresponding to discontinuous step functions of the type (4). We may have more general set functions $\varpi(\lambda; f)$, e.g. continuous in some interval. In fact, the properties which entirely characterize a decomposition of unity completely are that the operator $U(\lambda)$ be Hermitian and idempotent (i.e. $U^2 = U$). According to (7), a given decomposition of unity completely defines an operator A; equation (7) expresses the "spectral decomposition" of A. The points of discontinuity of the decompostion of unity $U(\lambda)$ of A give discrete eigenvalues of A; intervals in which $U(\lambda)$ is continuous, but not constant, form the "continuous spectrum" of A. Not all Hermitian operators have a decomposition of unity; but it may be assumed that those operators which represent physical quantities belong to this class. We may then conclude that the decomposition of unity of the operator representing a physical quantity defines the probability distribution, for any state, of all the possible values of this quantity.

3.- Projection operators and traces. Projection operators may sometimes be used with advantage instead of the state vectors to which they belong. Thus, the expectation values of the quantity A in the state f may be expressed in terms of the projection operator P_f as

$$(Af, f) = \text{tr}\,(P_f\, A). \qquad (9)$$

The operation indicated by tr is taking the <u>trace</u> of the operator following it. If φ_i is any complete orthogonal system, the trace of the operator A is defined by :

$$\mathrm{tr}\, A = \quad (A\varphi_i, \varphi_i);$$

it is readily verified that this definition is independent of the choice of the sytem φ_i; The notation (9) is especially convenient for expressing the distribution density of the values of an operator ; the probability of finding a value between λ and $\lambda + d\lambda$ when the system is in state f is

$$d\varpi(\lambda ; f) = \mathrm{tr}\left[P_f' \, d\,U(\lambda)\right] \tag{10}$$

If the operator A has a degenerate eigenvalue a_i with a set of orthogonal eigenfunctions $\varphi_{i,m}$ the corresponding term in the spectral decomposition of A is $a_i U_i$ with

$$U_i = \sum_m P_{\varphi_{i,m}}$$

and the probability of finding this value in state f is accordingly $\mathrm{tr}\,(P_f U_i)$. Moreover, since $\mathrm{tr}\, P_f = 1$ for any normalised state vector $\mathrm{tr}\, U_1$ represents the degree of degeneracy of the eigenvalue a_i.

4.- Temporal evolution of a system. The way in which the state of a system varies in the course of time may be described by assigning the variation to the state vector. The fundamental equation of "motion" of a state vector ψ is

$$i\hbar \frac{\partial\psi}{\partial t} = H\psi, \tag{11}$$

the operator H denoting the Hamiltonian of the system. In the representation thus adopted, which is called the " Schrödinger" representation the time variation is entirely ascribed to the state vectors, while the physical quantities are fixed operators in Hilbert space. This situation is quite analogous to usual onein classical theory ; the time dependent state vectors correspond to the classical trajectories; while the fixed operators correspond to the fixed phase functions which represent classical quantities. The time dependence of the latter is indirectly given by the variation of their argument, just as the time dependence of the expectation values of operators results from the variation of the state vectors.

A further analogy with the classical case appears when the motion is regarded as a transformation of the state vector or the phase, respectively. The Schrödinger equation (11) defines an automorphism of Hilbert space just as the Hamiltonian equations define an automorphism of phase space. In Hilbert space, the canonical transformations are those which preserve the fundamental operation of scalar product of state vectors : they are the so-called unitary transformations. Now, equation (11) can be immediately integrated in operator form :

$$\psi(t) = e^{-\frac{i}{\hbar}Ht}\,\psi(0) \qquad (12)$$

and this form shows explicitely the unitary transformation by which $\psi(0)$ is changed into $\psi(t)$. The corresponding transformation of the projection operators is

$$P_{\psi(t)} = e^{-\frac{i}{\hbar}Ht}\,P_{\psi_0}\,e^{\frac{i}{\hbar}Ht},$$

and the analogue of Liouville's theorem here reduces to the trivial statement of the invariance of the trace tr $P_{\psi(t)}$.

Chapter II.- THE ERGODIC THEOREM.

1.- *The first ergodic theorem.* The discussion of the equivalence between time averages and statistical averages in quantum theory is formally parallel to the classical one. It is still convenient to distinguish between strictly "isolated" systems, whose state is described at any time by the vector ψ (t) arising from an initial one $\psi(0)$ by the undisturbed time evolution of the system, and systems whose initial state is more "coarsely" defined : there will correspond an ergodic theorem to each of these cases, like in classical theory.

From the point of view of the physical interpretation, however, the first isolated system of quantum theory is rather different from that of classical theory. While in the latter theory any initial phase fixes an energy surface in which the trajectory is entirely contained, an arbitrary chosen initial state vector $\psi(0)$ will in general define a statistical distribution of the system over all the possible values of the energy. The ergodic average of classical theory accordingly corresponds to detailed statistical distribution over all the phases of a definite energy surface; but there is no connexion between the distributions on different energy surfaces; the parallel ergodic theorem of quantum theory will give us a statistical distribution over all values of the energy, but without any reference to the distribution of other physical quantities. Such a difference clearly lies in the nature of the question. Although, as we shall see, it will be levelled out when the conditions of the second ergodic theorem are introduced, there nevertheless remains an important field of application for the first ergodic theorem ; we shall meet later with a specific example of such an application.

Let us consider the time evolution of an arbitrary state vector $\psi(0)$, as determined by the Schrodinger equation. We shall assume that the Hamiltonian H has only discrete eigenvalues E_i with eigenfunctions E_i. We have initially a statistical distribution of the energy values :

$$\psi(0) = \sum_i (\psi(0), \varphi_i)\varphi_i ,$$

the coefficients $(\psi(0), \varphi_i)$ are complex numbers with definite amplitudes and phases :

$$(\psi(0), \varphi_i) \equiv r_i e^{i\alpha_i}$$

At time t, the state vector has become

$$\psi(t) = \sum_i r_i e^{i\alpha_i} e^{-\frac{i}{\hbar}E_i t} \varphi_i ;$$

the probability density $\varpi(E_i;\psi) = r_i^2$ of the energy distribution is unchanged (this corresponds to the fixed energy surface of the classical case), but the phases vary with time.

The expectation value of the quantity A at time t may be written

$$(A\psi_t,\psi_t) = \sum_{i,k} r_i r_k e^{i(\alpha_i-\alpha_k)}(A\varphi_i,\varphi_k) e^{-\frac{i}{\hbar}(E_i-E_k)t}$$

It depends on the time only through the last exponential factor. Now, this factor has always an average value with respect to time : it is zero or one according as $E_i \neq E_k$ or $E_i = E_k$. Therefore, the time average $\overline{A}|_t$ of the expectation value $(A\psi_t, \psi_t)$ always exists; its character is essentially different, however, according as the eigenvalues of the energy are or are not degenerate. If there is no degeneracy, we have

$$\overline{A}|_t = \sum_i r_i^2 (A\varphi_i,\varphi_i) ; \tag{1}$$

if, on the other hand, to take a simple illustration, the two eigenvalues E_l, E_m are equal, their contribution to the average $\overline{A}|_t$, besides the diagonal terms of the general type (1), will include a cross-term

$$2 r_l r_m \cdot \mathrm{Re}\left[e^{i(\alpha_l-\alpha_m)}(A\varphi_l,\varphi_m)\right]$$

(Re = real part of) depending on the initial phases α_l, α_m.

We therefore see that in order ot obtain a time average of type (1), independent of the initial phases, we require an assumption about the eigenvalues of the energy, the absence of degeneracy : this assumption is the analogue of the classical hypothesis of the metrical indecomposability of the energy surface. Remembering the meaning of the coefficients r_i^2 and using the trace notation, we may write formula (1) in the form

$$\overline{A}|_t = \mathrm{tr}(U_\psi A), \tag{2}$$

with

$$U_\psi = \sum_i \varpi(E_i;\psi) P_{\varphi_i} \tag{3}$$

Formula (2) expresses the time average of A as a statistical average, which is the quantal analogue of the ergodic average defined by Birkhoff's therorem, The statistical operator U_ψ depends, as already explained, on ly on the probability density of the energy distribution in the initial state ψ, but not on the initial phases of the energy eigenfunctions in state ψ. As a matter of fact, the operator U_ψ simply results from averaging the projection operator P_ψ over these phases. Infact, from

$$P_\psi f = (f,\psi)\psi = \sum_{ik} r_i r_k e^{i(\alpha_i - \alpha_k)} (f,\varphi_k)\varphi_i$$

we get by averaging over the α's

$$\overline{P_\psi}_\alpha f = \sum_i r_i^2 P_{\varphi_i} f = U_\psi f$$

In other words, the time variation of the state vector has the effect in the long run, of uniformly distributing the initial phases over all possible values ; this somewhat loose statement is analogous to the classical picture of the trajectory "filling up" the energy surface. The averaging over the phases means that a "pure state", represented for the statistical purposes by a projection operator P_ψ is replaced by a statistical assembly, represented by the operator U_ψ.

2.- <u>Macroscopic quantities and coarse distribution</u>. As already pointed out, the statistical operator U_ψ only describes a distribution over the eigenvalues of the energy, without any further detail. If one tries in quantum theory to approximate the calssical description of the state of a system by a number of different physical quantities, one immediatly meets with the limitations in the assignment of definite values to such quantities which result from the non-commutability of the corresponding operators. A way out of this difficulty has been suggested by von Neumann[*]. The idea is to represent the macroscopic quantities by commutable operators constructed by an appropriate "smoothing out" procedure from those which give the idealized representation of the same quantities in a strict quantal description of the system.

Thus, the energy of the system is strictly represented by the Hamiltonian

$$H = \sum_i E_i P_{\varphi_i}$$

---

[*] J. Von Neumann, Z. Physik 57, 30, 1929.

we again assume the Hamiltonian to have only discrete eigenvalues. Let us now subdivide the energy-axis into a sequence intervals I_a, each containing a certain number $\mathcal{S}_a$ of eigenvalues E_i; let us ascribe to each interval I_a a single value $\mathcal{E}_a$, which will be some average of the E_i's contained in I_a, and a projection operator

$$\mathcal{U}_a \equiv U(I_a) = \sum_i^{(a)} P_{\varphi_i} \tag{4}$$

the summation $\sum^{(a)}$ extending to all eigenfunctions φ_i which pertain to the E_i's in I_a. We may then define a new operator $\mathcal{H}$, which we shall call the "macroscopic energy", by its spectral decomposition

$$\mathcal{H} = \sum_a \mathcal{E}_a \mathcal{U}_a \tag{5}$$

We may describe this procedure as a "coarse" subdivision of the Hilbert space into <u>energy shells</u>, represented by the projection operators $\mathcal{U}_a$. To each shell corresponds a degenerate value of the "macroscopic" energy, of degree of degeneracy $\mathcal{S}_a = \mathrm{tr}\,\mathcal{U}_a$.

Any other macroscopic quantity $\mathcal{A}$ must commute with the macroscopic energy $\mathcal{H}$: its eigenfuncrions ω_k are therefore linearly related to the φ_i's. In particular, one may re-write formula (4) in terms of the P_{ω_k}:

$$\mathcal{U}_a = \sum_k^{(a)} P_{\omega_k} \tag{6}$$

the summation extends over all the P_{ω_k} belonging to the internal I_a. Now, it may happen that the ω_k's belonging to I_a correspond to several distinct eigenvalues of $\mathcal{A}$: in this case, such eigenvalues will characterize a certain subdivision of the internal I_a into smaller intervals i_a, i.e. of the energy shell $\mathcal{U}_a$ into <u>cells</u>

$$u_a^{(\nu)} = \sum_k^{(a,\nu)} P_{\omega_k} \tag{7}$$

(the summation extends over all the P_{ω_k}'s belonging to the interval $i_a^{(\nu)}$ of degeneracy $s_a^{(\nu)} = \mathrm{tr}\,u_a^{(\nu)}$.

By using a sufficient number of suitable macroscopic quantities $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, ... one finally arrives at a subdivision of each energy shell $\mathcal{U}_a$ into N_a cells of type (7), in such a way that *

$$\mathcal{U}_a = \sum_{\nu=1}^{N_a} u_a^{(\nu)} \quad ; \quad \mathcal{S}_a = \sum_{\nu=1}^{N_a} s_a^{(\nu)} \tag{8}$$

---

[*] We take the ω_k's to be the set of eigenfunctions common to all the macroscopic variables considered.

We shall not here raise the question whether this cell subdivision can actually be brought down to limits sufficiently near to these imposed by the uncertainty relations; von Neumann has shown that it is indeed possible to construct a set of "macroscopic" coordinates and momenta $\mathcal{Q}_k$, P_k satisfactory in this respect. All macroscopic quantities of physical interest can then be expressed as functions $f(\mathcal{Q}_k, \mathcal{P}_k)$.

3.- Von Neumann's ergodic theorem. The coarse cell distribution in Hilbert space just defined forms the basis for the formulation of a second ergodic theorem, due to Von Neumann, and widely analogous to the classical ergodic theorem of Hopf. We envisage a macroscopic quantity

$$\mathcal{A} = \sum_a \sum_{\nu=1}^{N_a} a_a^{(\nu)} u_a^{(\nu)}$$

and its expectation value in any state $\psi(t)$ at time t :

$$\mathcal{A}(\psi_t) = tr(P_{\psi_t}\mathcal{A}) = \sum_a \sum_{\nu=1}^{N_a} a_a^{(\nu)} w_a^{(\nu)}(t) \qquad (9)$$

$$w_a^{\nu}(t) = tr(P_{\psi_t} u_a^{(\nu)})$$

The statement of Von Neumann's ergodic theorem is that there exists a statistical average $\overline{\mathcal{A}}|_\psi$ equivalent to the time average of $\mathcal{A}(\psi_t)$ in as much as the mean quadratic fluctuation

$$\frac{1}{T}\int_0^T \left|\mathcal{A}(\psi_t) - \overline{\mathcal{A}}|_\psi\right|^2 dt \qquad (10)$$

tends to zero in the sense of probability, provided that certain conditions (which we shall formulate in the course of the analysis) are fulfilled by the system. Convergence "in the sense of probability" means that the probability that the quantity in question does not tend to zero can be made as small as one likes.

The expression for the statistical average $\overline{\mathcal{A}}|_\psi$ is a direct extension to our coarse distribution of that given by the first ergodic theorem

$$\overline{\mathcal{A}}|_\psi = tr(U_\psi \mathcal{A})$$

$$U_\psi = \sum_a \frac{tr(P_{\psi_0} V_a)}{tr\, V_a} V_a \qquad (11)$$

In fact, by Chapter I, §3, the quantity $W_a = tr(P_{\psi_0} V_a)$

is the total probability for the initial state ψ_0 of finding the macroscopic energy in the interval I_a (i.e. with the value $\mathcal{E}_a$), and $\mathrm{tr}[U_a]^{-1} U_a = U_a/S_a$ is the mean projection operator in this interval. With the notation introduced so far, we can write (11) as

$$\overline{\mathcal{A}}\rceil_\psi = \sum_a \sum_{\nu=1}^{N_a} a_a^{(\nu)} s_a^{(\nu)} \frac{W_a}{S_a} \tag{12}$$

The form of the expression

$$\left|\mathcal{A}(\psi_t) - \overline{\mathcal{A}}\rceil_\psi\right|^2 = \left|\sum_a \sum_\nu a_a^{(\nu)} \left[w_a^{(\nu)} - \frac{s_a^{(\nu)} W_a}{S_a}\right]\right|^2$$

suggests a comparison with the average of the quantity $\mathcal{A}^2$

$$\overline{\mathcal{A}^2}\rceil_\psi = \sum_a \sum_\nu \left(a_a^{(\nu)}\right)^2 s_a^{(\nu)} \frac{W_a}{S_a}$$

By an application of Schwarz's inequality we indeed find

$$\left|\mathcal{A}(\psi_t) - \overline{\mathcal{A}}\rceil_\psi\right|^2 \leqq \overline{\mathcal{A}^2}\rceil_\psi \sum_a \sum_\nu \frac{S_a}{s_a^{(\nu)} W_a} \left(w_a^{(\nu)} - \frac{s_a^{(\nu)} W_a}{S_a}\right)^2$$

We shall have proved the theorem if we show that the factor of $\overline{\mathcal{A}^2}\rceil_\psi$ in the right hand side of this inequality, after time averaging, tends to zero in the sense of probability.

The only quantity depending on thetime is

$$w_a^{(\nu)}(t) = \sum_k^{(a,\nu)} \left|(\psi_t, \omega_k)\right|^2$$

For ψ_t we may write, using the same notation as in § 1,

$$\psi_t = \sum_i r_i e^{i\beta_i} \varphi_i, \quad \beta_i = \alpha_i - \frac{1}{\hbar} E_i t$$

We thus have

$$w_a^{(\nu)}(t) = \sum_k^{(a,\nu)} \left|\sum_i r_i e^{i\beta_i} (\varphi_i, \omega_k)\right|^2$$

$$= \sum_{ij}^{(a)} r_i r_j e^{i(\beta_i - \beta_j)} C_{ij}^{(a,\nu)}$$

with

$$C_{ij}^{(a,\nu)} = \sum_k^{(a,\nu)} (\varphi_i, \omega_k)(\omega_k, \varphi_j) \tag{13}$$

The $C_{ij}^{(a,\nu)}$ thus defined obviously vanish unless φ_i φ_j both belong to the interval I_a; hence the restriction of the summation over i, j, to this interval. Moreover, if φ_i, φ_j belong to I_a,

$$\sum_{\nu=1}^{N_a} C_{ij}^{(a\nu)} = \delta_{ij}$$

and therefore
$$W_a = \sum_{\nu=1}^{N_a} w_a^{(\nu)}(0) = \sum_i^{(a)} r_i^2 \tag{14}$$

We thus have to take the time average of quantities of the form
$$x_a^{(\nu)} \equiv \left(w_a^{(\nu)} - \frac{s_a^{(\nu)} W_a}{S_a}\right)^2 = \left\{\sum_{ij}^{(a)} r_i r_j e^{i(\beta_i-\beta_j)}\left[C_{ij}^{(a\nu)} - \frac{s_a^{(\nu)}}{S_a}\delta_{ij}\right]\right\}^2$$
We see that these time averages reduce to expressions independent of the phases α_i, α_j only if the two following conditions aresatisfied :

a) no eigenvalue E_i of the Hamiltonian H is degenerate,

b) there are no energy "resonances", i.e. no two energy differences $E_i - E_j$ are equal. Note here again the parallelism with the conditions of validity of Hopf's ergodic theorem : besides the metrical indecomposability of each energy surface (corresponding to condition a), we had to assume metrical indecomposability for almost energy surfaces in the product space, which is analogous to our present condition (b). Under conditions (a) and (b), we get for the time average of $x_a^{(\nu)}$ the expression
$$\overline{x_a^{(\nu)}} = \left\{\sum_i^{(a)} r_i^2\left[C_{ii}^{(a,\nu)} - \frac{s_a^{(\nu)}}{S_a}\right]\right\}^2 + \sum_{\substack{i,j\\ i\neq j}}^{(a)} r_i^2 r_j^2 \left|C_{ij}^{(a,\nu)}\right|^2$$
Applying Schwarz'z inequality to the first term and using (14) we have
$$\overline{x_a^{(\nu)}} \leq W_a \sum_i^{(a)} r_i^2\left[C_{ii} - \frac{s_a^{(\nu)}}{S_a}\right]^2 + \sum_{\substack{i,j\\ i\neq j}}^{(a)} r_i^2 r_j^2 \left|C_{ij}^{(a,\nu)}\right|^2$$
and accordingly
$$\sum_a\sum_{\nu=1}^{N_a}\frac{S_a}{s_a^{(\nu)} W_a}\overline{x_a^{(\nu)}} \leq \sum_a\left\{\sum_i^{(a)} r_i^2 M_{ii}^{(a)} + \sum_{\substack{i,j\\ i\neq j}}^{(a)}\frac{r_i^2 r_j^2}{W_a}M_{ij}^{(a)}\right\} \tag{15}$$
where
$$M_{ij}^{(a)} = \sum_{\nu=1}^{N_a}\frac{S_a}{s_a^{(\nu)}}\left|C_{ij}^{(a,\nu)} - \frac{s_a^{(\nu)}}{S_a}\delta_{ij}\right|^2 \tag{16}$$

In the expression on the right hand side of (15), the factors which depend on the state vector considered, viz. the r_i's and W_a, are separated from the coefficients $M_{ij}^{(a)}$, in which the influence of the coarse subdivision of the Hilbert space into cells is concentrated. It is clear that the right hand side of (15) will always be $\leq 2\xi\sum W_a$, i.e. $\leq 2\xi$ for any state vector, provided only that all $M_{ij}^{(a)}$ are themselves $\leq \xi$. The crucial step in the proof of the ergodic theorem is the study of the distribution of the values of the $M_{ij}^{(a)}$'s for all possible cell subdivisions. This will lead to an estimate of the probability that the parameter ξ be larger than any given value ξ_0 and will show under which condition this probability can be sufficiently reduced.

Von Neumann's original argument has been simplified and sharpened by Pauli and Fierz[*], to whose paper we refer the reader for all details. The starting point is the obserbation that the (φ_i, ω_k) which enter into the expression for the $C_{ij}^{(a\nu)}$ can be regarded, for any interval I_a, as defining a set of S_a orthonormal unit vectors in a com plexe space of S_a dimensions; each $C_{ij}^{(a\nu)}$ represents the scalar product of the projections of two such vectors on a of $\Delta_a^{(\nu)}$ dimensions. With the help of this geometrical model, the "moments" of the distribution of the M_{ij}, i.e. the average values of the various powers $\left[M_{ij}^{(a)}\right]^n$ taken over all possible orientations of the vector set just described, can be valuated. From the moments, an estimate of the probability distribution readily follows.

The probability that, for any interval I_a, the upper bound of the $M_{ij}^{(a)}$ be larger than ξ_0 is found to be of the form

$$\text{Const}\ e^{-b\sqrt{\frac{S_a}{N_a}}\xi_0 + 2\log S_a} \tag{17}$$

where b is a number of the order of unity; this estimate is valid for $\xi_0 > \frac{2N_a}{S_a - 2}$. We this see that for not too small values of ξ_0, the probability (17) becomes vanishingly small provided only that

$$\frac{S_a}{N_a} \gg (2\log S_a)^2$$

This is the essential condition to be fulfilled by our coarse cell subdivision of the Hilbert space in order to ensure the validity of the ergodic theorem. It means that the average number of eigenfunctions in any coarse cell must be large: a quite réasonable requirement.

It must be admitted that von Neumann's ergodic theorem looks rather forbiffing and ill-suited to practical applications. Fortunately, as we shall see, its use is not required for the solution of actual physical problems. It has essentially the character of an existence theorem, by means of which statistical distributions more nearly approximating situations of physical interest can be derived. Thus, we shall presently see, in the next chapter, how von Neumann's theorem supplies the basis for distribution, which describes the statistical behaviour of systems of definite temperature.

---

[*] W. Pauli and M. Fierz, Z. Physik 106, 572, 1937.

Chapter III.- STATISTICS OF CLOSED SYSTEMS.

1.- Statistics of composite systems. The course we shall follow from now on will be closely parallel to the development of the calssical theory : we shall study successively closed systems in thermal contact with their surroundings and open systems , and in each case the argument will be essentially the same as in classical theory : only the formal aspect will undergo the changes required by the quantal mode of description. Thus, for the discussion of a closed system in contact with a thermostat we first need a formal treatment of composite systems, whose various parts are assumed to be in weak interaction with each other.

Let us consider, e.g., two systems, each described in its own Hilbert space by the Hamiltonian $H^{(1)}$, $H^{(2)}$ and state vectors $\psi^{(1)}$, $\psi^{(2)}$. The composite system resulting from the juxtaposition of these two systems is described in a Hilbert space which is the direct product of the two individual spaces. Its Hamiltonian is

$$H = H^{(1)} \times 1^{(2)} + H^{(2)} \times 1^{(1)},$$

where $1^{(k)}$ denotes the unit operator in the Hilbert space of system (k) and the cross indicates the direct product of operators pertaining to different Hilbert spaces. The state vectors of the composite system are products $\psi^{(1)} \times \psi^{(2)}$ and the corresponding projection operators direct products $P\ (1) \times P\ (2)$. The operator which governs the statistical behaviours of the composite system is the decomposition of unity of the Hamiltonian : this is [illegible] we must try to determine.

Let $E_i^{(k)}$, $\varphi_i^{(k)}$ be the eigenvalues and eigenfunctions of the Hamiltonian $H^{(k)}$; and $P_{\varphi_i}(k)$ the projection operator of $\varphi_i^{(k)}$) in the Hilbert space of system (k). The energy eigenvalues of the total system will be of types $E_i^{(1)} + E_j^{(2)}$ with corresponding projection operators $P_{\varphi_i}(1) \times P_{\varphi_j}(2)$ (or sums of such operator products if there is degeneracy). In particular, the projection operator $U(I)$ belonging to an interval I of possible values of the total energy will be a sum

$$U(I) = \sum_{(I)} P_{\varphi_i}(1) \times P_{\varphi_j}(2)$$

extended over all pairs $\varphi_i^{(1)}$, $\varphi_j^{(2)}$ belonging to eigenvalues $E_i^{(1)} + E_j^{(2)}$ contained in I. Quite generally, therefore, the decomposition of unity of the total energy is

$$dU(\lambda) = \int_{-\infty}^{\infty} dU^{(1)}(\lambda') \cdot dU^{(2)}(\lambda-\lambda')$$

where $dU^{(k)}(\lambda)$ represents the decomposition of unity of the Hamiltonian $H^{(k)}$, and the integration is performed with respect to λ'.

The expression for $dU(\lambda)$ is immediately generalised to a system composed of N parts with additive Hamiltonian

$$H = \sum_k 1^{(1)} \times 1^{(2)} \times \dots \times H^{(k)} \times \dots \times 1^{(N)}.$$

The decomposition of unity of this operator is

$$dU(\lambda) = \underset{(\lambda_1, \lambda_2, \dots \lambda_N.}{\iint \dots \int} dU^{(1)}(\lambda_1) \times dU^{(2)}(\lambda_2) \times \dots \times dU^{(N-1)}(\lambda_{N-1}) \times dU^{(N)}\Big(\lambda - \sum_{k=1}^{N-1} \lambda_k\Big) \qquad (1)$$

From this formula we derive a "law of decomposition" for the traces[*]:

$$tr[dU(\lambda)] = \underset{(\lambda_1, \lambda_2, \dots \lambda_{N-1})}{\iint \dots \int} \prod_{k=1}^{N-1} tr[dU^{(k)}(\lambda_k)] \cdot tr\Big[dU^{(N)}\Big(\lambda - \sum_{k=1}^{N-1} \lambda_k\Big)\Big] \qquad (2)$$

which is entirely analogous to the law of composition of the invariant measures of energy surfaces in the classical case (Part I, Ch. III, §2). In fact, at this point, Khinchin's argument can be taken up again and repeated without essential modification.

For each component of the system, we associate with the "weight" $tr\,[dU^{(k)}(\lambda)]$ of the energy interval $(\lambda, \lambda + d\lambda)$ normalized probability distribution (in the sense of a Stieltjes measure)

$$d\varpi^{(k)}(\lambda) = \frac{e^{-\beta\lambda}\, tr[dU^{(k)}(\lambda)]}{Z^{(k)}(\beta)} \qquad (3)$$

which satisfies a law of composition of the same form as (2). The normalization factor is

$$Z^{(k)}(\beta) = \int_{-\infty}^{\infty} e^{-\beta\lambda}\, tr[dU^{(k)}(\lambda)] = tr[e^{-\beta H^{(k)}}], \qquad (4)$$

If the total system has a large number of components, we get for its probability distribution the asymptotic expression

$$d\varpi(\lambda) \simeq \frac{1}{\sqrt{2\pi B}}\, e^{-\frac{(\lambda-\bar{E})^2}{2B}}\, d\lambda, \qquad (5)$$

where

---

[*] We use the same symbol tr to indicate the trace taken in any one of our Hilbert spaces.

$$\bar{E} = \sum_k \bar{E}^{(k)}, \quad \bar{E}^{(k)} = \frac{1}{Z^{(k)}(\beta)} \int_{-\infty}^{\infty} \lambda e^{-\beta\lambda} \, tr[dU^{(k)}(\lambda)] = -\frac{d}{d\beta} \log Z^{(k)}(\beta) \quad (6)$$

represents the average value of the energy in this distribution, and

$$B = \sum_k \ell^{(k)}, \quad \ell^{(k)} = \frac{1}{Z^{(k)}(\beta)} \int_{-\infty}^{+\infty} (\lambda - \bar{E}^{(k)})^2 e^{-\beta\lambda} \, tr[dU^{(k)}(\lambda)] = \frac{d^2}{d\beta^2} \log Z^{(k)}(\beta) \quad (7)$$

is the sum of the mean square fluctuations of the energies of the components.

2.- The canonical distribution. We can now be very brief in establishing the canonical distribution law in its quantal form for a systen S in contact with a thermostat T, for the argument is little else than a repetition of the classical one. To the total system $\mathcal{S}$ we apply Von Neumann's ergodic operator $\mathcal{U}_\psi$. For the initial state ψ in which we consider the system $\mathcal{S}$, we may choose an eigenstate of its energy. This reduces essentially to the projection operator U_a belonging to the interval I_a in which the corresponding eigenvalue Λ is contained; this projection operator will be more conveniently denoted by $\Delta U(\Lambda)$, and

$$\mathcal{U}_\psi = \frac{\Delta U(\Lambda)}{tr\{\Delta U(\Lambda)\}} .$$

Now, assuming a weak interaction between system and thermostat, we may write

$$dU(\lambda) = \int dU_S(\lambda') \times dU_T(\Lambda - \lambda') \quad ;$$

and in particular,

$$\Delta U(\Lambda) = \int dU_S(\lambda') \times \Delta U_T(\Lambda - \lambda') \quad ;$$

i.e. the "coarseness" of the energy definition of the total system is entirely referred to the thermostat; it does not affect at all the energy λ' of the system S , for which we shall obtain a "fine" distribution.

In fact, any operator A_S pertaining to the system S gives rise to an operator in a special form in the Hilbert space of the total system, viz.

$$A = A_S \times 1_T,$$

whose average is given by

$$\bar{A}_S = tr(\mathcal{U}_\psi A) = \int tr[A_S dU_S(\lambda')] \frac{tr\, \Delta U_T(\Lambda - \lambda')}{tr\, \Delta U(\Lambda)} \; ..$$

For both the thermostat and the total system we may use the asymptotic form (5) of the distribution law, and we may treat the energy interval $\Delta\lambda$ defining the thickness of the energy shell as a "physically infinitesimal" quantity. We then get in excatly the same way as in the classical case

$$\frac{tr[\Delta U_T(\Lambda-\lambda')]}{tr(\Delta U(\Lambda))} \simeq \frac{1}{Z_s(\beta)}\, e^{-\beta\lambda'}$$

The statistical operator of the canonical distribution is thus

$$U_\beta = \frac{1}{Z_s(\beta)} \int e^{-\beta\lambda'}\, d\, U_s(\lambda')$$

i.e. (dropping the index S)

$$U_\beta = \frac{1}{Z(\beta)}\, e^{-\beta H}, \text{ with } Z(\beta) = tr\left[e^{-\beta H}\right]; \tag{8}$$

in this formula H is the Hamiltonian of the system and β the inverse of the absolute temperature ; it is uniquely related to the average energy of of the thermostat by the equation

$$-\frac{d}{d\beta} \log Z_T(\beta) = \bar{E}_T$$

The canonical average of any quantity is expressed by means of the operator U_β in the form

$$\bar{A} = tr\,(U_\beta A). \tag{9}$$

The statistical interpretation of thermodynamics can simply be taken over from the classical theory. The free energy is $F = -\theta \log Z(\beta)$ and the entropy can be put in the form

$$S = -tr\left[U_\beta \log U_\beta\right]. \tag{10}$$

One point, however, requires special consideration : it is the proof of the permanence of the canonical distribution during quasi-static adiabatic transformations.

4.- Quasi-static adiabatic transformations. Let us start from a system of given temperature, with the statistical operator

$$U_{\beta_0} = \frac{e^{-\beta_0 H(\alpha_0)}}{tr\left[e^{-\beta H(\alpha_0)}\right]}$$

We assume the Hamiltonian H(a) to have, for any value of the external parameter a, discrete and non-degenerate eigenvalues $E_i(a)$ with eigenfunctions $E_i(a)$. If after breaking the contact with the thermostat, we suddenly change the external parameter from a_0 to $a = a_0 + da$, each of the state vectors $\varphi_i(a_0)$ will evolve according to the Schrödinger equation pertaining to the new Hamiltonina $H(a)$. According to the <u>first</u> ergodic theorem[*], the statistical operator $P_{\varphi_i(a_0)}$ corresponding to this state vector will eventually take the equilibrium form

$$\sum_j tr\,(P_{\varphi_i(a_0)} P_{\varphi_j(a)})\, P_{\varphi_j(a)}$$

But the state vector $\varphi_i(a_0)$ was initially represented in the statistical operator U_{β_0} by its projection operator $P_{\varphi_i(a_0)}$ with the coefficient $\frac{1}{Z(\beta_0, a_0)} e^{-\beta_0 E_i(a_0)}$. The statistical operator corresponding to the new equilibrium is therefore

$$U(a) = \frac{1}{Z(\beta_0, a_0)} \sum_{i,j} e^{-\beta_0 E_i(a_0)}\, tr(P_{\varphi_i(a_0)} P_{\varphi_j(a)})\, P_{\varphi_j(a)}$$

$$= \frac{\sum_j tr\,[e^{-\beta_0 H(a_0)} P_{\varphi_j(a)}]\, P_{\varphi_j(a)}}{tr\,[e^{-\beta H(a_0)}]} \quad . \qquad (11)$$

We have now to evaluate the traces occuring in the expression (11) for $U(a)$ in terms of the new Hamiltonian; i.e. we must substitute for $H(a_0)$ its value in terms of $H(a)$:

$$H\,(a_0) = H\,(a\ \ ) - da \frac{\partial H}{\partial a} .$$

This calculation demands some care, because the two terms in the expression for $H(a_0)$ do not generally commute. Since we only require a result accurate to the first order in da, we may write

$$e^{-\beta_0 [H(a) - da \frac{\partial H}{\partial a}]} = e^{-\beta_0 H(a)} + \beta_0\, da \sum_{n=0}^{\infty} \frac{1}{n!} (-\beta_0)^{n-1} \sum_{m=0}^{n-1} H^m \frac{\partial H}{\partial a} H^{n-1-m} ;$$

but $\left(\sum_{m=0}^{n-1} H^m \frac{\partial H}{\partial a} H^{n-1-m} \varphi_j, \varphi_j \right) = n\, E_j^{n-1} \left(\frac{\partial H}{\partial a} \varphi_i, \varphi_j \right),$

---

[*] There is no question of introducing here any coarse energy shells, and applying Von Neumann's ergodic theorem, as L. Broer erroneously tried to do in his Amsterdam thesis (1945). As already emphasized, we are here dealing with a <u>fine</u> distribution of our system because we consider it to initially in contact with its surroundings.

and therefore

$$\left(e^{-\beta_0 H(\alpha_0)}\varphi_j(\alpha),\varphi_j(\alpha)\right)=e^{-\beta_0 E_j(\alpha)}\left[1+\beta_0 d\alpha\, A]_{E_j}\right]$$

with the notation

$$A]_{E_j}=\left(\frac{\partial H}{\partial \alpha}\varphi_j,\varphi_j\right)$$

for the expectation value of the force in the state $\varphi_j(a)$. For the denominator of (11) we thus have

$$\operatorname{tr}\left[e^{-\beta_0 H(\alpha_0)}\right]=\sum_j e^{-\beta_0 H(\alpha)}\left[1+\beta_0\, d\alpha\, A]_{E_j}\right]$$

$$=\operatorname{tr}\left[e^{-\beta_0 H(\alpha)}\right]\cdot\left[1+\beta_0\, d\alpha\cdot\bar{A}\right],$$

where $\bar{A}$ denotes the canonical average of the force for the values β_0, $\underline{a}$ of the thermodynamical variables. The operator (11) accordingly takes the form

$$U(\alpha)=\frac{1}{\operatorname{tr}\left[e^{-\beta_0 H(\alpha)}\right]}\,e^{-\beta_0 H(\alpha)}\left\{1-\beta_0\, d\alpha\left[\bar{A}-A]\right]\right.$$

We may write this more compactly by introducing the operator

$$A]=\sum_j A]_{E_j}\,P_{\varphi_j},$$

which commutes with the Hamiltonian H(a). Therefore,

$$U(\alpha)=\frac{1}{\operatorname{tr}\left[e^{-\beta_0 H(\alpha)}\right]}\,e^{-\beta_0 H(\alpha)}\left\{1-\beta_0\, d\alpha\left[\bar{A}-A]\right]\right\}$$

The operator $A]$, commuting with H, can be regarded as a function of H; let us assume that it can be expanded in the neighbourhood of the value A_E that it takes when H is replaced by its canonical average E :

$$A]=A]_{\bar{E}}+\left(\frac{\partial A]}{\partial H}\right)_{\bar{E}}(H-\bar{E})+\frac{1}{2}\left(\frac{\partial^2 A]}{\partial H^2}\right)_{\bar{E}}(H-\bar{E})^2+\dots\dots$$

Taking the canonical average, we get for $\bar{A}$ the similar expansion :

$$A=A]_{\bar{E}}+\qquad\qquad\qquad+\frac{1}{2}\left(\frac{\partial^2 A]}{\partial H^2}\right)_{\bar{E}}\overline{(H-\bar{E})^2}+\dots\dots$$

By the same argument as in the classical case, we can show that the difference $A]-\bar{A}$ is equal to $\left(\frac{\partial A]}{\partial H}\right)_{\bar{E}}(H-\bar{E})$ except for terms at most of the order of relative fluctuations which are negligible for macroscopic systems. Now, a change of modulus in the statistical operator

$$U_{\beta_0}(a) = \frac{e^{-\beta_0 H(a)}}{tr\,[e^{-\beta_0 H(a)}]}$$

yields

$$U_{\beta_0}(a) = U_\beta(a)\,[1 + d\beta\, H - \bar{E}]$$

We thus see that the new statistical operator $U(a)$ has approximately the canonical form $U_\beta(a)$ provided that the variation $d\beta$ of modulus is connected with the variation of external parameter da by the relation

$$d\beta + \beta_0\, da \left(\frac{\partial A]}{\partial H}\right)_{\bar{E}} = 0,$$

whose thermodynamical interpretation is the classical one.

5.- Irreversible transformations. Gibbs' classical treatment of irreversible transformations can be readily transposed to quantum theory[*] We consider arbitrary statistical operators U or rather, following Gibbs' example, the corresponding probability exponents η, defined by $U = e^{\eta}$ We have then to compare the average probability exponents $tr(\eta e^{\eta})$ under various conditions.

(a) Adiabatic transformations : As we have just seen, if we start from a statistical distribution of the form $e^{\eta} = \sum_k e^{\eta_k} P_{\varphi_k}$ and perform a small change of the external parameters; we arrive eventually at a distribution

$$e^{\eta^{(0)}} = \sum_i e^{\eta_i^{(0)}} P_{\varphi_i^{(0)}},$$

whose density fulfils the condition

$$e^{\eta_i^{(0)}} = tr\,(e^{\eta} P_{\varphi_i^{(0)}}) = \sum_k e^{\eta_k} w_{ki}\,, \qquad (12)$$

with

$$w_{ki} = |(\varphi_k, \varphi_i^{(0)})|^2\ ;$$

we have

$$\sum_i w_{ki} = \sum_k w_{ki} = 1\,. \qquad (13)$$

This is the analogue of the condition of Gibbs' first lemma. We accordingly prove that under condition (12)

$$tr\,(\eta^{(0)} e^{\eta^{(0)}}) \leq tr\,(\eta\, e^{\eta}) \qquad . \qquad (14)$$

In fact, using the relation (12)

---

[*] O. Klein, Z. Physik 72, 767, 1931.

$$\mathrm{tr}(\eta e^{\eta}) - \mathrm{tr}\,(\eta^{(0)} e^{\eta^{(0)}}) = \sum_k e^{\eta_k} \{ \eta_k - \sum_i \eta_i^{(0)} w_{ki} \},$$

or, by (13),

$$= \sum_{i,k} e^{\eta_k} w_{ki} (\eta_k - \eta_i^{(0)} \} . \qquad (15)$$

On the other hand, again using (13), the normalization conditions $\mathrm{tr}(e^{\eta}) = \mathrm{tr}\,(e^{\eta^{(0)}}) = 1$ may be written

$$\sum_{i,k} e^{\eta_k} w_{ki} = \sum e^{\eta_i^{(0)}} w_{ki} = 1 ,$$

whence

$$\sum_{i,k} e^{\eta_k} w_{ki} \left(e^{\eta_i^{(0)} - \eta_k} - 1 \right) = 0$$

Addingt this to the right hand side of (15) gives (14) in virtue of the identity $e^{x} - 1 - x \geqslant 0$. The discussion of adiabatic transformations on the basis of the inequality (14) is the same as in classical theory.

(b) Heat exchange. The theory of heat exchange is likewise contained in the analogue of Gibbs' second lemma : if a system is composed of two parts in weak interaction, any statistical distribution e^{η} pertaining to it has a spectral decomposition of the form

$$e^{\eta} = \sum_{i,k} e^{\eta_{ik}} P_{\varphi_i^{(1)}} \times P_{\varphi_k^{(2)}}$$

and defines the statistical operators $e^{\eta^{(1)}}$, $e^{\eta^{(2)}}$ for the component systems by

$$e^{\eta^{(1)}} = \sum e^{\eta_i^{(1)}} P_{\varphi_i^{(1)}} , \quad e^{\eta^{(2)}} = \sum_k e^{\eta_k^{(2)}} P_{\varphi_k^{(2)}}$$

with

$$e^{\eta_i^{(1)}} = \sum_k e^{\eta_{ik}} , \quad e^{\eta_k^{(2)}} = \sum_i e^{\eta_{ik}} \qquad (16)$$

We then have

$$\mathrm{tr}\,(\eta e^{\eta}) \geqslant \mathrm{tr}\,(\eta^{(1)} e^{\eta^{(1)}}) + \mathrm{tr}\,(\eta^{(2)} e^{\eta^{(2)}}) \qquad (17)$$

The proof is immediate : we have, with (16),

$$\mathrm{tr}(\eta e^{\eta}) - \mathrm{tr}\,(\eta^{(1)} e^{\eta^{(1)}}) - \mathrm{tr}(\eta^{(2)} e^{\eta^{(2)}}) = \sum_{i,k} e^{\eta_{ik}} (\eta_{ik} - \eta_i^{(1)} - \eta_k^{(2)})$$

whence (17) follows.

It is not necessary to pursue this kind of considerations any further : the parallelism with Gibbs' classical discussion is quite obvious. There is, however, in quantum theory, a kind of irreversible process which does not occur for classical systems : a measurement performed on a quantal system produces in general an irreversible change of the state of the system: it is interesting to investigate the accompanying increase of entropy.

6.- The process of measurement.[*] If we have a system in a "pure" state ψ , its statistical operator is represented by the projection operator P_ψ : strictly speaking, there is no thermodynamical analogy to this case, just as little as to the case of a classical mechanical system on a definite trajectory. The statistics of the pure state is a fundamental property of quantal systems, which has nothing to do with thermodynamics. As we have seen, the quantal model of a macroscopic situation, to which thermodynamics can be applied, involves an element of "coarseness", expressed by the change from the operator P_ψ to the statistical operator

$$\sum_{\alpha,\nu} \frac{tr(P_\psi U_\alpha^{(\nu)})}{tr\, U_\alpha^{(\nu)}} U_\alpha^{(\nu)}$$

This corresponds to passing from a single classical trajectory to a "tube of flow" in phase space, composing all trajectories defined by the phases of a finite cell.

However, the fact that a quantal system even in a well-defined state involves a statistical distribution has in one respect consequences of the same general kind as the effects considered in thermodynamics : a measurement performed on such a quantal system is a macroscopic process by which the "pure states" is transformed into a "mixture" of all the eigenstates of the quantity measured. The question therefore arises wether the entropy concept can be extended so as to apply to this specifically quantal type of irreversible transformation. The answer given by Von Neumann to this question is extremely simple : if a quantal system in is a state represented by the statistical operator U, its entropy may be defined as

$$S(U) = -\operatorname{tr}\left[U \log U\right] \tag{18}$$

This expression vanishes for any pure state, for which U is of the form P_ψ . In order to check the adequacy of the definition (18), we must therefore show, in the first place, that any two pure states of a quantal system have the same entropy in the sense of thermodynamics

---

[*] J. Von Neumann, Gottingen Nachrichten 1927, p. 273

i.e; that it is possible to pass from any pure state to any other by a quasi-static transformation which does not involve any heat exchange. This seems at first sight somewhat surprising, especially if the two states ψ, φ are orthogonal, since in that case no direct measurement performed on the system in state ψ will ever yield state φ. But as we shall see, the quasi-static transition from the one to the other becomes possible if we use an appropriate sequence of intermediate steps. It will suffice to discuss the case of two orthogoanl states, since any two others can be linked together by a third one, orthogonal to both of them.

Let us consider the sequence of (p + 1) states defined by

$$\psi_r = \cos\frac{\pi r}{2p} \cdot \psi + \sin\frac{\pi r}{2p}\,\varphi \,, \quad = 0, 1, 2, \ldots, p; \qquad (19)$$

the first state ψ_0 is our initial state ψ, the last one ψ_p our final state φ, orthogonal to ψ. We start from an assembly of copies of the system considered all in state ψ_0. If we perform on all these systems the measurement of some quantity A_1 which has ψ_1 as one of its eigenfunctions, we transform the original assembly into a mixture, a fraction $|(\psi_0, \psi_1)|^2 = \cos^2\frac{\pi}{2p}$ of which is in state ψ_1. Next we measure a quantity A_2, having ψ_2 as oneof its eigenfunctions; the new mixture resulting from this measurement contains a fonction ϖ_2 of systems in state ψ_2 which is at least the fraction of states ψ_2 contained in the state ψ_1; i.e.

$$\varpi_2 \geqslant |(\psi_0, \psi_1)|^2 \cdot |(\psi_1, \psi_2)|^2;$$

but from (19) we get quite generally

$$|(\psi_1, \psi_2)|^2 = |(\psi_r, \psi_{r+1})|^2 = \cos^2\frac{\pi}{2p} .$$

Therefore, $\varpi_2 \geqslant \cos^4\frac{\pi}{2p}$. Proceeding in the sale way, we finally arrive at a mixture containing a fraction ϖ_p of sysyems in state ψ_p for which we may write

$$\varpi_p \geqslant \cos^{2p}\frac{\pi}{2p}$$

Since $\lim\limits_{p\to\infty} \cos^{2p}\frac{\pi}{2p} = 1$, we see that by taking q sufficiently large number of intermediate steps, we can actually transform the assembly in the pure state ψ into an assembly in the pure state φ. Moreover, the operations invoved in the sequence of measurements do not give rise to any change of entropy : we have therefore proved that all pure states have the same entropy.

It is now easy to evaluate the increase of entropy associated with the transformation of a pure state ψ into the mixture resulting from the measurement of a quantity with eigenfunctions φ_i. The statistical operator of the mixture is

$$U = \sum_i \varpi_i P_{\varphi_i} \quad , \qquad \text{with } \varpi_i = |(\psi, \varphi_i)|^2$$

i,e. the mixture contains a fraction ϖ_i of states φ_i. In order to find the entropy of this mixture we must try to produce it in a quasi static way, starting from an assembly of systems all in state ψ . To this end, let us divide the N systems of the assembly into groups of ϖ_1 N, ϖ_2N,... systems. As we have just seen, we may without change of entropy transform each of these groups into the corresponding pure state fraction of the mixture U , i.e. the i-th group into ϖ_iN systems all in state We have then only to mix together in a quasi static way, all the separate groups thus obtained : this last operation, as is well-known, is accompanied by an increase of entropy,

$$-N \sum_i \varpi_i \log \varpi_i ,$$

i.e. if the entropy is referred to a single system, just the expression (18).

Chapter IV.- Statistics of open systems.

1.- Quantal systems of identical elements. Consider a system of N identical elements. Any physical quantity pertaining to this system is symmetrical with respect to the sets of coordinates representing the elements. This symmetry entails an essential degeneracy of the corresponding operator : from any eigenfunction a set of N! distinct eigenfunctions can be constructed by permuting the coordinates of the N elements in all possible ways; all theses eigenfunctions belong to the same eigenvalue of the operator. In particular, the Hamiltonian presents this degeneracy, which is the analogue in quantum theory of the set of specific phases constituting a generic phase.

The symetry degeneracy is removed, however, by an additional requirement imposed upon all state vectors of the system : these vectors must be either symmetrical or antisymmetrical in the coordinates of the identical elements. By means of any set of N! eigenfunctions of the kind just described one can construct only one symmetrical and one antisymmetrical combination : the one or the other must be chosen according to the nature of the elements. The requirement of antisymmetry is called the exclusion principle; since it implies that no two elements can have the same set of coordinates. The exclusion principle applies, in particular, to the fundamental constituents of matter, nucleons and electrons. It can be shown that elements composed of nucleons and electrons (such as nuclei or atoms or molecules) have symmetrical or antisymmetrical state vectors according as they contain an even or odd number of constituents.

In discussing quantal systems with a variable number of identical elements, a considerable formal simplification is achieved by treating the numbers of elements in this various possible state as quantal variables. Consider first a single element under the external conditions applying equally to all elements of the system : its behaviour is described by a Hamiltonian, which defines a complete set of non-degenerate stationary states of energies E_k of this element. We may now introduce the operator N_k characterizing the number of elements in state k : its eigenvalues are $N'_k = 0, 1, 2.....$ if the state vectors of the system are symmetrical, or $N'_k = 0, 1$ if the elements obey the exclusion principle. If the interactions between the elements are neglected, the Hamiltonina of the system will take the simple additive form

$$H = \sum_k E_k N_k \qquad (1)$$

valid for an unspecified total number of elements.

Any interaction of the elements between themselves or with other species of elements will effect transitions of the systems between states characterized by different numbers N'_k of elements in their various individual states k. The corresponding operators may be expressed in terms of elementary "annihilation" and " creation" operators a_k, a^+_k whose effect on an eigenvector $\psi(N'_1, N'_2, .., N'_k ....)$ of the variables N_1, N_2.... is defined by the relations

$$a_k \, \psi(\cdot\cdot N'_k \cdot\cdot) = \sqrt{N'_k} \; \psi(\cdot\cdot N'_k - 1 \cdot\cdot\cdot)$$
$$a^+_k \, \psi(\cdot\cdot N'_k \cdot\cdot) = \sqrt{N'_k + 1} \; \psi(\cdot\cdot\cdot, N'_k + 1, \cdot\cdot)$$

in the case of symmetrical state vectors, and

$$a_k \, \psi(\cdot\cdot N'_k \cdot\cdot) = N'_k \; \psi(\cdot\cdot\cdot 1 - N'_k \cdot\cdot)$$
$$a^+_k \, \psi(\cdot\cdot N'_k \cdot\cdot) = (1 - N'_k) \; \psi(\cdot\cdot\cdot 1 - N'_k \cdot\cdot)$$

in the case of antisymmetrical state vectors. In both cases, the operator N_k is given by

$$N_k = a^+_k \, a_k$$

The interaction Hamiltonian of the system may consist of various terms representing interactions between pairs, triples, etc... of elements. Thus, the interaction between pairs of elements has a Hamiltonian of the form

$$\sum_{k,\ell,k',\ell'} a^+_k a^+_\ell \left(k\ell \,\middle|\, V(P^{(1)}, P^{(2)}) \,\middle|\, k'\ell' \right) a_{k'} a_{\ell'}$$

where $V(P^{(1)}, P^{(2)})$ denotes the operator of potential energy between two elements of coordinates $P^{(1)}$, $P^{(2)}$; the matrice element of this operator must be taken between the two inital states k', ℓ' and the two final states k, ℓ . In the following we shall only discuss in detail systems with weak interactions, for which the simple formula (1) is a sufficient approximation to the Hamiltonian. But in general we may always regard the Hamiltonian operator as a function of the operator a_k, a^+_k, acting upon state vectors of the type $\psi(N'_1, N'_2, ... N'_k, ...)$

2.- Grand canonical distribution. The classical argument leading to the establishment of the grand canonical distribution for open systems can be extended to quantal systmes with no essential alteration. The system S under investigation will be described by an Hamiltonian $H(a_k^{(i)}, a_k^{(i)+}; a_{k'}^{(i)}, a_{k'}^{(i)+}; \dots ; p)$ where the operators $a_k^{(i)}, a_k^{(i)+}$ pertain to the i-th species of elements contained in the system. The total number of such elements is an operator

$$N^{(i)} = \sum_k N_k^{(i)}$$

We further introduce reservoirs $R^{(i)}$ of elements of each species; their respective Hamiltonians will be $H^{(i)}(b_l^{(i)}, b_l^{(i)+}, p^{(i)})$ where the annihilation and creation operators are denoted by $b_l^{(i)}, b_l^{(i)+}$ to indicate that they refer to elements in different external conditions from those of the system : the a's and b's commute. The numbers of elements in the states of the i-th reservoir will be denoted by $M_l^{(i)} = b_l^{(i)+} b_l^{(i)}$; their total number is accordingly $M^{(i)} = \sum_l M_l^{(i)}$. The total system consisting of system S and reservoirs may be regarded as a closed system, with fixed numbers

$$\mathcal{N}^{(i)} = N^{(i)} + M^{(i)} \qquad (2)$$

of elements of each species.

The statistical operator of the total system is that of a canonical distribution, $\sim \exp\left[-\beta (H + \sum_i H^{(i)})\right]$; the normalization is obtained by taking the trace of this operator with respect to the set of eigenvectors $\psi (N_k^{(i)'} M_l^{(i)'})$ of the numbers of elements in the various possible states, subject to the restrictions (2). The statistical operator pertaining to the system S is accordingly

$$\sim e^{-\beta H} \prod_i \mathrm{Tr} \left[e^{-\beta H^{(i)}} \right]$$

where the traces over the reservoir operators are again taken subject to the relations (2); this means that the resulting factors are operators depending on the $N^{(i)}$. At this stage, the argument of the classical treatment which determines the type of dependance of tr $\left[e^{-\beta H^{(i)}}\right]$ on $N^{(i)}$ can be taken over. In general, for any closed system of M elements, the extensive character of the free energy leads to a proportionality relation

$$\log \ \mathrm{tr} \left[e^{-\beta H} \right] \sim M$$

Hence, for a very large open system, the variation of log tr $\left[e^{-\beta H} \right]$ for a small variation of the number of elements from its average value will be approximately proportinnal to this variation. For the reservoirs we may therefore write, on account of (2)

$$\mathrm{tr}\left[e^{-\beta H^{(i)}}\right] = e^{\zeta^{(i)} N^{(i)}}$$

where $\zeta^{(i)}$ is a function of the macroscopic variables, including the average numbers $\overline{N}^{(i)}$ $(i = 1, 2, \dots)$. This gives the final form of the statistical operator of the open system S :

$$U_\beta = e^{\beta\left\{\Omega + \sum_i \zeta^{(i)} N^{(i)} - H\right\}} \tag{3}$$

in this formula, the normalization factor $e^{\beta\Omega}$ is defined by

$$e^{-\beta\Omega} = \mathrm{tr}\left\{ e^{\beta\left[\sum_i \zeta^{(i)} N^{(i)} - H\right]} \right\} ; \tag{4}$$

all quantal operators $N^{(i)}$, H are expressed in terms of the $N_k^{(i)}$, or more generally of the $a_k^{(i)}$, $a_k^{(i)+}$.

If the pressure p is chosen as the external mechanical parameter, we again find that Ω pratically reduces to zero. But it will be more convenient, in the following, to take the volume V as independent variable : then, just as in classical theory, Ω defined by (4) plays the part of a characteristic function $\Omega(\theta, V, \zeta^{(i)})$. It will be noticed that no mention has been made, in the preceding considerations, of the distinction, so important in classical theory, between specific and generic phases : this is simply because the quantal states as explained in § 1, always refer to generic phases; no factors $(N^{(i)}!)^{-1}$ appear explicitely because the reduction to generic phases is already included from the beginning in the specification of the symmetry or antisymmetry of the state vectors.

3.- Open systems with weak interaction. We shall now apply our general formulae (3) and (4) to open systems with weak interactions. For the moment, we consider systems with only one species of elements : there is then in (3) and (4) no summation over this index (i), which may be dropped altogether. The Hamiltonian is of the additive type (1). We first discuss the thermodynamics of such systems, which is entirely contained in the characteristic function (4). This function can now be put into a form involving only the states E_k of a single element : a transformation which is the analogue of the passage from Γ-space to μ - space in classical theory. In fact, we may write

$$e^{-\beta\Omega} = \sum_{N_1', N_2', \dots} e^{\beta \sum_k (\zeta - E_k) N_k'}$$

$$= \prod_k \sum_{N_k'} e^{\beta(\zeta - E_k) N_k'}$$

For systems with symmetrical state vectors, we must sum over all integral values of the N'_k; this gives

$$e^{-\beta\Omega} = \prod_k \frac{1}{1 - e^{\beta(\zeta - E_k)}} ;$$

for systems with antisymmetrical state vectors, we have only $N'_k = 0$ or 1, whence

$$e^{-\beta\Omega} = \prod_k \left[1 + e^{\beta(\zeta - E_k)}\right]$$

It will be convenient to condense the formulae corresponding to these two cases by introducing the symbol

$$\varepsilon = \begin{cases} +1 & \text{in the antisymmetrical case} \\ -1 & \text{in the symmetrical case} \end{cases} \qquad (5)$$

With this notation

$$-\beta\Omega = \sum_k \varepsilon \log\left[1 + \varepsilon e^{\beta(\zeta - E_k)}\right] \qquad (6)$$

The average number of elements is given by

$$\bar{N} = -\frac{\partial\Omega}{\partial\zeta}$$

i.e.

$$\bar{N} = \sum_k \frac{1}{e^{-\beta(\zeta - E_k)} + \varepsilon} \qquad (7)$$

If the average number $\bar{N}$ is given, equation (7) is an implicit definition of the chemical potential ζ in terms of the macroscopic variables. Since in any case $\bar{N}$ must not be negative, formulae (7) shows that the range of possible values of ζ is restricted in the symmetrical case to $\zeta \leq 0$, while there is no restriction to ζ in the antisymmetrical case. Moreover, one observes that if $\beta\zeta$ becomes negative and very large, the expression for $\bar{N}$ reduces to an asymptotic form common to both kinds of elements :

$$\bar{N} \simeq \sum_k e^{\beta(\zeta - E_k)} \qquad (8)$$

This formula has the same structures as the classical one,[*] the only diffe-

---

[*] From the classical formula for an open system with weak interaction

$$e^{-\beta\Omega} = \sum_N \frac{1}{N!} e^{\beta\zeta N} Z^N$$

where Z denotes the sum over the states in μ-space, one derives

$$\bar{N} = -\frac{\partial\Omega}{\partial\zeta} = e^{\beta\Omega} \sum_N \frac{1}{(N-1)!} e^{\beta\zeta N} Z^N$$

The substitution N→N + 1 in the sum over N shows that this sum is equal to

$$e^{-\beta\Omega} e^{\beta\zeta} Z$$

whence $\bar{N} = e^{\beta\zeta} Z$

rence being that the classical expression for the "sum over state" in μ -space is replaced by the corresponding quantal expression

$$Z = \sum_k e^{-\beta E_k} . \tag{9}$$

One usually expresses this situation, in a somewhat loose terminology, by saying that the limiting case of large nagative $\beta \zeta$ is that of "classical statistics". The general case is then described as that of "quantal degeneracy", and the formulae applying to systems with symmetrical or antisymmetrical state vectors are denoted as "Bose-Einstein or Fermi-Dirac statistics" respectively; the constituent elements of such systems are often called "bosons" or "fermions". The limit of "strong degeneracy", opposed to that of classical statistics, corresponds to $\beta \zeta \approx 0$ for systems of bosons, and $\beta \zeta \to \infty$ for systems of fermions.

It will be useful to examine somewhat more closely the formula (9) for the sum over states Z. It may be split into two factors, Z_t and Z_i reffering to the degreees of freedom of translation and internal motion, respectively. The factor Z_t can be evaluated in a quite general way. The momentum of an element is related to the wave number of the corresponding stationnary de Broglie wave by the relation $\vec{p} = h \vec{k}$, and the wave numbers are quantized according to the relations

$$k_x = n_x / 2L , \; \ldots ,$$

where n_x, n_y, n_z are positive integers, and L is the side of a large cube in which the system is assumed to be enclosed. This gives in the usual way the asymptotic expression for the element of measure in momentum space :

$$d\mu_p = \frac{1}{8} \, dn_x \, dn_y \, dn_z = L^3 \, dk_x \, dk_y \, dk_z = \frac{V}{h^3} \, dp_x \, dp_y \, dp_z$$

We get accordingly, if m denotes the mass of an element

$$Z \simeq \int d\mu_p \, e^{-\frac{\beta}{2m}(p_x^2 + p_y^2 + p_z^2)} = \frac{(2\pi m)^{3/2}}{h^3} V \theta^{3/2} \tag{10}$$

this differs from the classical value only by the constant factor h^{-3}. Such a factor has no influence, in the classical limit, on the thermodynamical significance of the sum over the states, we shall come back in the next section to its physical meaning in quantum theory.

The other factor Z_i of Z depends on the structure of the elements of the system and cannot be reduced to any universal form. Let ε_0 be the (negative) eigenvalue of the energy of the ground state of an element, and ε_1, ε_2, ... the energies of the successive excited states; the positive difference $\varepsilon_k - \varepsilon_0$ accordingly represent the energy quanta corresponding to the transitions from the excited states to the ground state. We may write

$$Z_i = e^{\beta|\varepsilon_0|}\sum_{k=0}^{\infty} e^{-\beta(\varepsilon_k-\varepsilon_0)} \qquad (11)$$

a convenient form to study the behaviour of Z_i with varying temperature. We see that at very low temperature, Z_i becomes exponentially infinite, like $e^{\beta|\varepsilon_0|}$; at high temperatures, Z_i also tends to infinity, generally like some power of the temperature. For instance, if the element considered performs harmonic oscillations of frequency ν about an equilibrium configuration of energy ε_0', we have, with $\varepsilon_0 = \varepsilon_0' - \frac{1}{2}h\nu$

$$Z_i = e^{\beta|\varepsilon_0|}\left[1-e^{-\beta h\nu}\right]^{-1},$$

and therefore

$$Z_i \simeq \theta/h\nu \quad \text{for} \quad \theta/h\nu \gg 1.$$

We are now in position to discuss under which physical conditions the limit of "classical statistics" is valid. According to formula (8), the formal condition for the validity of classical statistics, viz $-\beta\zeta \gg 1$, means

$$\log\frac{Z}{N} \gg 1 \quad , \qquad (12)$$

Let us first treat the cause of elements without internal structure, for which Z reduces to the translation part Z_t. Using formula (10), we may then write the condition (12) in the form

$$\log\left\{\frac{(2\pi m)^{3/2}}{h^3}\cdot\frac{\theta}{(\bar{N}/V)}\right\} \gg 1 \qquad (13)$$

which shows that "classical" statistics applies at sufficiently high temperatures or low densities. Conversely, we shall expect quantal degeneracy at low temperatures or high densities.

This conclusion is not modified if we include the internal structure of the elements. In fact, returning to formula (7), and putting

$E'_k = E_k - \varepsilon_0$, we derive from it the asymptotic form

$$\bar{N} \simeq e^{\beta(\zeta - \varepsilon_0)} \sum_k e^{-\beta E'_k}$$

identical with (8), on the condition that $-\beta(\zeta + |\varepsilon_0|) \gg 1$, which is slighty more stringent for the absolute value of ζ that our previous condition $-\beta\zeta \gg 1$. But then, according to (11) we get instead of (12) the condition

$$\log \frac{Z_c \cdot Z_i \, e^{-\beta|\varepsilon_0|}}{\bar{N}} \gg 1$$

in which the factor $e^{\beta|\varepsilon_0|}$ which renders Z_i singular at low temperatures neutralized. The only modification in the final formula (13) is therefore an alteration of the exponent of θ in the limiting case of large θ . The qualitative statement of the conditions for the validity of classical statistics remains the same.

4.- Chemical equilibria. As is well known, the two laws of thermodynamics do not suffice to fix completely the law of equilibrium of a system of several constituents in chemical reaction with each other. It is possible to derive by thermodynamical reasoning the general form of hte "law of mass action", but in the expression for this law a constant factor remains undetermined. This arises from the fact that the characteristic functions are only defined up to a linear function of the temperature (a constant for the entropy). In order to remedy thie defect, Nernst was led to supplement the classical scheme of thermodynamics by a further postulate which he called the third law. Nernst's postulate does not directly fix the value of the entropy constant; but it sets up a relation of universal character between the entropies of different systems, which suffices to remove any ambiguity from the law of mass action. Nernst assumes that when the temperature tends to the absolute zero, the entropies of all bodies tend to become equal.

The true significance of Nernst's postulate has only been revealed by quantum theory. In fact, ist validity is a direct consequence of the existence of stationary states, which we can always assume to form a discrete sequence. As the temperature tends to zero, the statistical distribution of any system will tend to be concentrated in its ground state, and according to our quantal definition the corresponding entropy will tend to zero, independently of the nature of the system.

This does not mean that we have *derived* Nernst's postulate from quantal statistics : for we might of course have added an arbitrary constant to our entropy definition. But we may say that Nernst's postulate fits in quite naturally in the frame of quantum theory : it has no longer the appearance of an *ad hoc* requirement, but it seen to be intimately connected the existence of stationary states expressed by Bohr's fundamental postulate of quantum theory.

From the preceding considerations it follows immediately that in order to get a complete determination of the law of mass action all we have to do is to stick to the definition we have adopted for the free energy, i.e. to consider this definition as "absolute", without any addition of a linear function of the temperature. Now, our definition consists in relating the free energy directly to the normalization factor of the statistical distribution of equilibrium. This means, precisely, that we take as the "unnormalized" statistical operator that factor which contains the atomistic operators (which we would call classically the "phase functions" viz. the Hamiltonian and the numbers of elements : we thus obtain a unique meaning for the normalization factor, and consequently for the characteristic function. In other words, all we have to do to secure the fulfilment of Nernst's postulate is to work consistently with actual probabilities correctly normalized to unity in the sense just explained.

Considering, in particular, the limiting case of "classical statistics" (which covers the most usual applications to chemical reactions), we have just to work with the "absolute" value of the sum over states, as given by the formula (10), to get the precise determination of the "chemical constants" entering into the law of chemical equilibrium This implies that the chemical constants will essentially involve the quantum of action, represented by the factor h^{-3} in (10) and similar factors arising from the sum over states of internal motion Z_i. We here meet with a striking illustration of the essential part played by the quantum of action in chemistry : we are, in fact, reminded that the stability of atoms and molecules can only be understood on the basis of quantum theory.

There are phenomena of an essentially different character from ordinary chemical reactions, in which elements of various kinds can be created and annihilated ; e.g., in interaction with a radiation field pairs of electrons of opposite signs can be produced or can annihilate into radiation. The statistics of open systems we have developped covers such phenomena as well. Let us consider, as a simple, the equilibrium between positons and negatons; the radiation field can be left out of consideration ; it can be regarded as the "thermostat" supplying all necessary energy. We have now a statistical operator of the general type (3), with two species of elements, which we will distinguish by the indices +, -. The average numbers $\bar{N}_+$, $\bar{N}_-$ are not fixed, but connected by the relation

$$\bar{N}_- - \bar{N}_+ = \text{const.}$$

expressing that the total charge (usually an excess of negatons) is conserved (i.e. that the positrons and negatons appear and vanish in pairs) We have simply to write down the equilibrium condition

$$dG = \frac{\partial G}{\partial \bar{N}_+} d\bar{N}_+ + \frac{\partial G}{\partial \bar{N}_-} d\bar{N}_- = \zeta_+ d\bar{N}_+ + \zeta_- d\bar{N}_- = 0$$

with the additional requirement

$$d\bar{N}_+ - d\bar{N}_- = 0$$

This gives immediately

$$\zeta_+ + \zeta_- = 0$$

the chemical potential of either the negatons or the positrons can be chosen arbitrarily; the other is then completely fixed by the above relation. It must be noted that the chemical potentials we have used are not the most appropriate to the present problem : in fact, they are referred to a Hamiltonian representing only the kinetic energy of the particles, whereas we are here concerned with an essentially relativistic effect, for which it is more convenient to use the ordinary relativistic definition of the energy which includes the rest mass. The chemical potential ζ^0 referred to this determination of the Hamiltonian is simply related to the potential ζ by

$$\zeta^0 = \zeta + mc^2$$

so that the equilibrium condition takes the form

$$\zeta^0_+ + \zeta^0_- = 2mc^2 \qquad (14)$$

The astrophysical implications of this relation have been discussed by

Chandrasekhar and Rosenfeld[2]. The most immediate inference is the following : we see from (14) that the positron and negatron systems are in a kind of reciprocal relationship with respect to quantal degeneracy. If (as is the case in white dwarfs) the neagaton gas is strongly degenerate, the positon gas is in a state corresponding to classical statistics : it cannot, therefore, essentially modify the stellar equilibrium, which is predominantly determined by the pressure of the degenerate negaton gas.

5. Statistiacl distribution of open systems with weak interactions.

Hitherto we have discussed the thermodynamical consequences of the general statistical operator (3) for open systems with weak interactions. Let us now turn to the more detailed consideration of the statistical distribution itself, or, what amounts to the same, the average value

$$\bar{A} = \mathrm{Tr}\left[A\,\rho_\beta\right] \tag{15}$$

of any operator A. If we express A in terms of the operator a_k, a_k^*, it is always possible in principle to eliminate from (15) all reference to the numbers of elements and to reduce $\bar{A}$ to an expression involving only the stationary states of a single element. Just as in the classical case, however a simple result is only obtained for additive quantities of the form

$$A = \sum_i \sum_k A_k^{(i)} N_k^{(i)}$$

To avoid trivial complications, let us take the case of a single species of elements. If A is of the type $A = \sum_k A_k N_k$, we get

$$\bar{A} = e^{\beta\Omega} \sum_k \left\{ A_k \,\mathrm{Tr}\left[N_k\, e^{\beta(\zeta - E_k) N_k}\right] . \prod_{l \neq k} \mathrm{Tr}\left[e^{\beta(\zeta - E_l) N_l}\right] \right.$$

$$= \sum_k A_k \frac{\mathrm{Tr}\left[N_k\, e^{\beta(\zeta - E_k) N_k}\right]}{\mathrm{Tr}\left[e^{\beta(\zeta - E_k) N_k}\right]}$$

This has indded the form of an average over the stationary states of a single element, with a distribution function $f(E_k)$:

$$\bar{A} = \sum_k A_k\, f(E_k) \tag{16}$$

The function $f(E_k)$ represents the average number $\overline{N_k}$ of elements in state k. An explicit expression for it is easily obtained. We start from

$$F(E_k) \equiv \mathrm{tr}\left[e^{\beta(\zeta - E_k)N_k}\right] = \left[1 + \varepsilon\, e^{\beta(\zeta - E_k)}\right]^{\varepsilon}, \qquad (17)$$

with the notation (5) for distinguishing bosons and fermions. The distribution function $f(E_k)$ is the logarithmic derivative of $F(E_k)$ with respect to $\beta(\zeta - E_k)$, i.e.

$$f(E_k) = \frac{1}{e^{-\beta(\zeta - E_k)} + \varepsilon} \qquad (18)$$

The limiting case of classical statistics requires no special comment; using (8) and (9) to eliminate ζ, we get

$$f(E_k) \simeq \frac{\bar{N}}{Z}\, e^{-\beta E_k}$$

On the other hand, the behaviour of the distribution for strong quantal degeneracy is quite different for bosons and fermions. In the case of bosons, we shall have an accumulation of elements in the ground state, leading to a peculiar "condensation" phenomenon; in the case of fermions the tendency will be for the elements to fill up all the lowest stationary states. A thorough discussion of these aspects, which have physical applications of fundamental importance, is outside the scope of this course.

6. Statistics of the radiation field. The only point which remains to be settled is the position of the radiation field with respect to the statistics of open systems. A radiation field within an enclosure with reflecting walls can be described as a system of independent proper oscillations, whose wave numbers $\vec{k}_i$ are determined by the boundary conditions; moreover, to each wave-number belong two independent modes of polarization. Each proper oscillation *, of frequency $\nu_i = c\,|\vec{k}_i|$, is quantized as a harmonic oscillator of that frequency, i.e. its energy has the eigenvalues $(N_i' + \frac{1}{2})\, E_i$ where $E_i = h\,\nu_i$ and N_i' is a non-negative integer. Although there is no interaction between different proper oscillations, the system is nevertheless ergodic, provided that we introduce into the enclosure a "grain of coal dust"

---

* The single index i represents the set of quantum numbers characterizing the proper oscillation, including its polarization.

("Kohlestäubchen) to secure by absorption and emission the necessary exchange of energy between the various oscillations. In other words, we treat the radiation field as a closed system of proper oscillations in contact with a thermostat represented by the grain of coal dust.

The total Hamiltonian is simply the sum of the Hamiltonians H_i of the proper oscillations. Leaving out the zero-point energy of the oscillators, we get for the free energy of the systems

$$e^{-\beta F} = \mathrm{tr}\left[e^{-\beta H}\right] = \prod_i \mathrm{tr}\left[e^{-\beta H_i}\right]$$

$$= \prod_i \sum_{N_i'} e^{-\beta E_i N_i'} = \prod_i \frac{1}{1-e^{-\beta E_i}},$$

i.e.

$$-\beta F = -\sum_i \log\left[1-e^{-\beta E_i}\right] \qquad (19)$$

This expression presents a formal analogy with that for the characteristic function Ω of a system of bosons, given by (6) : it would correspond to the value $\zeta = 0$ of the chemical potential.

The bosons in question are the quanta of oscillation of the field, usually called photons : they represent the "particle aspect" of the radiation field. In fact, to each wave number $\vec{k}_i$ and mode of polarization wa may associate photons of energy $E_i = h\nu_i$. Moreover, in virtue of the laws of the radiation field, the proper oscillation has a momentum $N_i' E_i/c$ in the direction $\vec{k}_i$: we may thus ascribe to each photon belonging to this oscillation a momentum of magnitude $p_i = E_i/c$ and direction $\vec{k}_i$, i.e. $\vec{p}_i = h\vec{k}_i$; The rest mass of the photon is accordingly zero. We may further introduce creation and annihilation operators for photons, and define the operator N_i for the number of photons in state i. The Hamiltonian of the field then takes the form $H = \sum_i E_i N_i$, corresponding to a system of particles in weak interaction. In this representation, the radiation field is conceived as an open system of bosons, the grain of coal dust playing the part of a reservoir of photons. Thus, the exchange of energy and the exchange of elements are here the same process, and this circumstance is expressed by the fact that the chemical potential vanishes.

The statistical operator of the radiation field

$$U_\beta = \frac{e^{-\beta H}}{\mathrm{Tr}\left[e^{-\beta H}\right]} \qquad \text{with } H = \sum_i E_i N_i$$

gives rise to a distribution function

$$f\,(E_k) = \frac{1}{e^{-\beta E_k} - 1}$$

for the computation of averages of operators additive in the photon numbers The asymptotic expression for the element of measure in the momentum space of the photons is twice the usual one, so as to account for the two independent modes of polarization. Thus from (19) we get for the free energy the asymptotic value

$$F = \theta \cdot \frac{8\pi V}{h^3} \int_0^\infty p^2 dp \log\left(1 - e^{-\beta c p}\right)$$

$$= \frac{8\pi V \theta^4}{(hc)^3} \int_0^\infty x^2 dx \log\left(1 - e^{-x}\right)$$

$$= -\frac{\sigma}{3} V \theta^4 \qquad \text{with } \sigma = \frac{8\pi}{(hc)^3} \int_0^\infty \frac{x^3 dx}{e^x - 1}$$

This formula contains the thermodynamics of the radiation field, viz. its entropy density

$$s = \frac{4}{3} \sigma \theta^3 ,$$

its energy density

$$\bar{\varepsilon} = \frac{\bar{E}}{V} = \sigma \theta^4 \qquad \text{(law of Stefan-Boltzmann)}$$

and the radiation pressure

$$p = 1/3 \; \bar{\varepsilon}$$

CONTENTS

INTRODUCTION

Page

CLASSICAL STATISTICS.

I.- General properties of mechanical systems.

II.- The Ergodic Theorems.

III.- Statistics of closed systems.

IV.- Statistics of open systems.

QUANTUM STATISTICS.

I.- Quantal description of Physical systems.

II.- The ergodic theorems.

III.- Statistics of closed systems.

IV.- Statistics for open systems.

CHAPITRES CHOISIS DE PHYSIQUE NUCLEAIRE

par

le Professeur Mario VERDE
Université de Turin

Université de Grenoble
Ecole d'Eté de Physique Théorique

Les Houches (Haute Savoie)

15 Aout - 10 Septembre 1952

ERRATA

- P. 1, 18º ligne avant la fin, lire : le spin , le moment ...

- P. 3, dernière ligne, lire "approximation" au lieu de "équation".

- P. 5 et suivantes, lire τ_ζ au lieu de τ_z

- P. 7, 12ème ligne avant la fin, lire "de la fonction d'onde spatiale".

- P. 8, 14 º ligne, lire $\vec{P}$ au lieu de p

- P. 9, 5º ligne avant la fin, lire "flux de particules" au lieu de "nombre de particules".

- P. 10 et suivantes : lire τ_ζ au lieu de τ_z

- P. 11, 6º ligne avant la fin, lire 16 cas au lieu de 18 cas

- P. 13, 8º ligne avant la fin, lire $U = \frac{2\mu V}{\hbar^2}$ au lieu de $(r) = \frac{2\mu V}{\hbar^2}$

  2º ligne avant la fin lire $L_z = 0$

- P. 14, à partire de la 6º ligne avant la fin , lire

$$u_l^r(0) \simeq \frac{2^l}{2l+1}\,\frac{l!}{(2l)!}\,(kr)^{l+1} = \frac{(kr)^{l+1}}{(2l+1)!!}$$

$$u_l^s(0) = \frac{1}{2^l}\,\frac{(2l)!}{l!}\left(\frac{1}{kr}\right)^l = (2l-1)!!\left(\frac{1}{kr}\right)^l$$

$$u_0^r(0) \simeq \frac{kr}{1}$$

$$u_0^s(0) \simeq 1$$

nous utiliserons aussi les deux fonctions

$$u_l^+ = u_l^s + i\,u_l^r \underset{r\to\infty}{\simeq} e^{i\left(kr - l\frac{\pi}{2}\right)}$$

$$u_l^- = u_l^s - i\,u_l^r \underset{r\to\infty}{\simeq} e^{-i\left(kr - \frac{l\pi}{2}\right)}$$

- P. 15, 2º ligne lire :

$$u_l^2 \frac{d u_l^1}{d(kr)} - u_l^1 \frac{d u_l^2}{d(kr)} = \text{cte} = 1$$

4º ligne, lire :

$$kr\,\psi^i = \sum_l 2\sqrt{(2l+1)\pi}\; i^l\, \frac{u_l^+ - u_l^-}{2i}\, Y_l^0(\theta,\varphi)$$

12º ligne, lire :

"et la section efficace différentielle est"

4º ligne avant la fin, lire : " $S_l = \frac{\lambda_l + i}{\lambda_l - i}$ "

- P. 16, 4º ligne avant la fin, lire : " On peut aussi déduire ces formules de l'équation .....

- P. 18, 1º ligne, lire $\bar{u}^2 u^{1\prime} - u^1 \bar{u}^{2\prime} = - \frac{\bar{k}^{l+1}}{k^l}$

- P. 22, 10º ligne, lire : "où ν est indépendant de λ"

- P. 23, 8º ligne, lire : $\delta J_S = \delta J_H - 2\int_0^\infty \cdots$

9º ligne, lire : " il résulte de $\delta J_S = 0$ que ...

11º ligne, lire $(J_S)_{\text{ext}} = (J_H)_{\text{ext}} = k\lambda_{\text{extrémal}}$

- p. 25, 12º ligne, lire : $S_{12} K(r) = \left\{ 3\, \frac{(\bar{\sigma}^{(1)} \cdot \bar{r})(\bar{\sigma}^{(2)} \cdot \bar{r})}{r^2} - \bar{\sigma}^{(1)} \cdot \bar{\sigma}^{(2)} \right\} K(r)$

9º ligne avant la fin, lire $j-1, j, j+1$ au lieu de $l-1, l, l+1$

- P. 27, 7º ligne, lire : "de connaitre le paramètre ε ".

INTRODUCTION

Afin de décrire théoriquement les phénomènes nucléaires sur la base d'un schéma Hamiltonien, il est nécessaire de connaitre l'énergie d'interaction entre deux nucléons. Le vecteur d'état ψ du système doit, en effet, être solution de l'équation du mouvement :

$$-\frac{\hbar}{i}\frac{\partial\psi}{\partial t} = (H_0 + H_{int})\psi$$

On recherche donc la forme de l'interaction en partant d'un certain nombre de données expérimentales concernant les configurations de deux ou plusieurs nucléons. Cette méthode employée auparavant dans le cas atomique s'est montrée, comme on le sait, trés fructueuse.

Nous nous occuperons ici des tentatives les plus récentes qui ont été faites dans le but de déterminer la forme du potentiel nucléaire. Nous n'avons pas la possibilité de traiter en détail les nombreux travaux sur ce sujet. Nous préférons nous étendre sur certains aspects du problème renvoyant à la littérature pour d'autres moins récents.

Les derniers renseignements relatifs à l'interaction nucléaire proviennent des données expérimentales concernant le système de deux nucléons seulement.

Le système constitué par deux nucléons n'a qu'un seul état lié : l'état fondamental du deutéron. Nous connaissons pour cet état fondamental l'énergie de liaison, le moment angulaire total, le spin, le moment dipolaire magnétique et le moment quadrupolaire électrique. Bien qu'elles soient insuffisantes pour établir une forme définitive pour le potentiel nucléaire, ces données nous fournissent un certain nombre d'informations sur son caractère.

On peut, par exemple affirmer que la force entre neutron-proton dépend de l'orientation relative des spins et de l'angle entre le spin total et la ligne qui joint les centres de gravité des deux particules, l'état fondamental du deutéron n'ayant pas la symétrie sphérique.

Les renseignements les plus importants concernant les forces entre deux nucléons proviennent des états du spectre continu. On déduit de l'expérience les sections efficaces pour les chocs n-p et p-p en fonction de l'énergie et de l'angle de déviation. Il faut citer aussi les données relatives à la désintégration du deutéron en ses deux constituants lorsqu'on le soumet à l'action d'un champ électromagnétique externe , ou le processus inverse de la capture d'un neutron par un proton avec émissions de rayons γ .

Les expériences de choc aux basses énergies ne permettent pas d'établir la dépendance du potentiel nucléaire de la distance relative r : on peut en effet avec différentes formes de potentiel expliquer la variation de la section efficace totale en fonction de l'énergie et de la section différentielle qui est isotrope dans le système de référence centré autour du centre de gravité (en négligeant l'effet coulombien dans le cas p-p).

La forme du potentiel de Yukawa :

$$V \sim \frac{e^{-\kappa r}}{\kappa r}$$

permet d'expliquer d'une manière satisfaisante l'égalité des forces nucléaires entre n-p et p-p se trouvant dans le même état quantique. Les 3 autres formes de potentiel : puit de potentiel, potentiel exponentiel, potentiel gaussien ne peuvent l'expliquer. Les expériences plus récentes aux hautes énergies, liées à l'emploi de grands accélérateurs, ont fourni des renseignements importants. Dans le choc n-p, la distribution angulaire prise dans le système de référence du centre de gravité est anisotrope, mais presque parfaitement symétrique par rapport à la direction perpendiculaire à la trajectoire incidente
Le fait que pour des angles supérieurs à 90º, la valeur de la section efficace est élevée confirme l'hypothèse, faite pour d'autres raisons, du caractère de forces d'échange des forces nucléaires. Les types de forces ordinaires fournissent toujours aux énergies assez élevées un maximum en avant (à ce propos, penser au scattering coulombien).

Dans le choc p-p-, on a une isotropie remarquable qui s'étend à peu près dans l'intervalle 15º - 165º si l'on néglige l'action du potentiel électrostatique qui ne se fait sentir que vers 0º et 180º.
Si les états du système n-p étaient identiques à ceux du système p-p on pourrait conclure du comportement différent de la distribution angulaire dans les deux cas que l'interaction n-p est nettement différente aux

énergies élevées de celle entre deux protons.

Mais à cause du principe de Pauli, la moitié seulement des états est occupable dans le second cas. Deux protons dans des états de moment orbital impair doivent avoir nécessairement leurs spins parallèles, et leurs spins antiparallèles dans des états pairs.

On n'a encore fait aucune analyse quantitative indiquant si aux énergies élevées, les forces n-p et p-p- sont les mêmes, sauf pour les effets électromagnétiques. Le problème de l'indépendance de la charge aux énergies élevées pour les forces nucléaires n'est pas encore résolu.

On n'a pas encore pu trouver jusqu'ici une forme phénoménologique de potentiel nucléaire expliquant les diverses configurations possibles aux hautes énergies. Ces difficultés ont conduit certains auteurs à tenter de nier le schéma Hamiltonien et de renoncer aux potentiels nucléaires, sans toutefois donner une autre conception plus efficace. Nous n'avons encore actuellement aucun principe sur lequel on puisse bâtir une nouvelle dynamique nucléaire. On peut obtenir des renseignements moins directs, mais utiles, sur la nature du potentiel nucléaire en étudiant les phénomènes nombreux qui intéressent plus de deux nucléons. En partant des configurations d'un système compléxe, on peut en effet remonter à la nature de la force d'intéraction élémentaire. Cette question sera présentée dans la seconde partie de ces notes.

Etant donné les difficultés que présentent les problèmes à plusieurs dégrés de liberté, on est obligé d'employer dans l'étude mathématique des noyaux complexes des méthodes d'approximation qui consistent à faire des modèles pour les configurations nucléaires. Il n'est pas aussi sipmle de se donner un modèle d'édifice nucléaire que de se donner un modèle d'édifice atomique. Dans les atomes le champ électrique central intense du noyau détermine essentiellement le caractère des configurations électroniques. On peut accepter, comme première approximation, que les électrons se déplacent indépendamment les uns des autres en obéissant au principe de Pauli. Ceci conduit, pour les électrons, au schéma d'orbites successives se disposant en couches, autour du noyau, dont celles qui sont complètes montrent une stabilité particulière. Cette première ~~équation~~ approximation dans l'étude des systèmes atomiques permet d'expliquer le système périodique des éléments. Une seconde approximation dans laquelle on tient compte des forces d'interaction entre les électrons conduit à une description trés soignée des spectres atomiques.

Dans l'étude des noyaux, de nombreuses données expérimentales sont aussi favorables à l'idée d'une espèce de système périodique. Les noyaux comprenant 8, 20, 50, 82, 126 protons ou neutrons occupent une position privilégiée dans la nature, comme les gaz rares dans la classification périodique des éléments. L'usage a consacré pour ces nombres le qualificatif de "magiques". La stabilité exceptionnelle des noyaux magiques est confirmée d'une manière indirecte par le fait que leur abondance relative sur la croute terrestre est plus grande que celles des noyaux dont les masses sont assez proches.

L'expérience montre que l'énergie de liaison d'un nucléon ajouté à une configuration magique est sensiblement inférieure à la valeur moyenne de 8 Mev.

Les spins et les moments électriques et magnétiques confirment la régularité indiquée pour les noyaux magiques. Il faut accepter comme une bonne hypothèse de travail l'existence d'un champ central dans les noyaux, bien qu'il ne soit pas facile de comprendre le mécanisme responsable de cette symétrie centrale. Une explication trés suggestive par sa simplicité de la structure en couches complètes pour les états nucléoniques des 5 nombres magiques cités plus haut est fondée sur l'hypothèse que les nucléons soumis à un champ central sont aussi sujets à l'action d'un couplage du type spin-orbite, proportionnel au moment angulaire total. Un nucléon en mouvement, ayant un certain moment orbital sera sollicité différemment selon que son spin propre s'ajoute ou se retranche au moment orbital. La présence d'une telle force de couplage est décelée par des effets de polarisation. En faisant passer un faisceau de protons ou neutrons non polarisés au voisinage d'un noyau, on peut observer aprés un choc que les particules orientent leurs spins dans une direction privilégiée. Des expériences ont mis en évidence la polarisation des neutrons et des protons, donnant ainsi une preuve directe de l'existence d'un couplage spin-orbite proportionnel au moment angulaire. Pour conclure ces considérations sur notre programme, nous voulons ajouter que nous nous occuperons aussi, assez briévemment, du modèle dit de Wigner-Hund qui a joué un rôle important dans l'explication de certaines propriétés des noyaux.

FORME DU POTENTIEL PHENOMELOGIQUE NUCLEAIRE

Afin de limiter le choix des divers types possibles pour un potentiel nucléaire, on est conduit à poser certaines hypothèses simplicatrices qui ne peuvent être justifiées qu'à postériori.

Nous admettrons que l'interaction nucléaire dépend uniquement de variables liées aux nucléons, c'est-à-dire des coordonnées d'espace, de spins, et de leur charge. Pour déduire la dépendance du potentiel de la charge, il est avantageux d'emplyer le formalisme du spin isotopique introduit pour la première fois par Heisenberg.

On va attribuer à chaque état ψ une variable de charge τ qui ne peut prendre que deux valeurs ;

la valeur + 1 correspond à un état p (proton)
la valeur - 1 correspond à un état n (neutron)

Dans une représentation de Schroedinger, la fonction d'état sera non seulement fonction de $\bar{r}$, mais aussi de σ et τ. On écrira :

$$\psi(\bar{r}, \sigma, \tau)$$

l'opératuer de charge est :

$$q = e\,\frac{1+\tau_3}{2}$$

Dans l'espace formé par $\psi(\bar{r}, \sigma, +1)$ et $\psi(\bar{r}, \sigma, -1)$ les opérateurs q et τ_ζ seront alors représentés par les matrices

$$q = e\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \tau_\zeta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Par analogie avec le formalisme emplyé dans l'étude du spin ordinaire, on introduit les deux opérateurs τ_ξ et τ_η qui seront représenté dans dans le même espace par les matrices

$$\tau_\xi = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \tau_\eta = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

On a :

$$\tau_\xi^2 = \tau_\eta^2 = \tau_\zeta^2 = 1$$

$$+\tau_\xi\tau_\eta = -\tau_\eta\tau_\xi = i\,\tau_\zeta$$

etc...

Les deux opérateurs

$$\frac{1}{2}\tau^+ = \frac{\tau_\xi + i\tau_\eta}{2} \quad , \quad \frac{1}{2}\tau^- = \frac{\tau_\xi - i\tau_\eta}{2}$$

réprésentés par les matrices

$$\frac{1}{2}\tau^+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad \frac{1}{2}\tau^- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

font passer d'un état de charge à l'autre.

On peut encore parler du carré du spin isotopique :

$$\tau^2 = \tau_\xi^2 + \tau_\eta^2 + \tau_\zeta^2 = \frac{1}{2}(\tau^+\tau^- + \tau^-\tau^+) + \tau_\zeta^2 = 3$$

et attribuer à chaque nucléon un spin isotopique $T = \frac{1}{2}$ dont la composante ζ dans l'espace de charge ne peut prendre que les valeurs $+\frac{1}{2}$ et $-\frac{1}{2}$
L'espace de charge est purement symbolique et n'a aucun rapport avec l'espace ordinaire. Considérons deux nucléons dont les états de charge soient décrits par $(a_1\ b_1)$ et $(a_2\ b_2)$. Les états de charge pour le système des deux nucléons sont

$$\zeta^t_{+1} = a_1 a_2$$

$$\zeta^t_0 = \frac{1}{\sqrt{2}}(a_1 b_2 + a_2 b_1) \qquad \zeta^s = \frac{1}{\sqrt{2}}(a_1 b_2 - a_2 b_1)$$

$$\zeta^t_{-1} = b_1 b_2$$

Les trois premiers états correspondent au spin isotopique total $T = 1$ le dernier au spin isotopique $T = 0$
La charge totale

$$e\left(\frac{1+\tau_\zeta^1}{2} + \frac{1+\tau_\zeta^2}{2}\right)$$

est diagonalisée dans la base formée par les ζ.

Les états p-p et n-n sont donnés d'une manière univoque par $a_1\ a_2$ et $b_1\ b_2$. Ils appartiennent donc au spin isotopique $T = 1$, tandis que l'état n-p peut être un mélange des états ζ^t_0 et ζ^s_0. Il n'est possible de distinguer entre ζ^t_0 et ζ^s_0 que si le spin isotopique total est un bon nombre quantique

Il faut remarquer que les fonctions propres ζ^t de l'opérateur de charge sont symétriques par rapport à l'échange des coordonnées de charge des deux nucléons; la fonction ζ^s au contraire est antisymétrique.

Appelant P^τ_{12} l'opérateur relatif à cet échange, on aura donc

$$P_{12}^{\tau}\,\xi^{r} = +\xi^{r} \qquad P_{12}^{\tau}\,\xi = -$$

Considérons maintenant le produit scalaire

$$\vec{\tau}^{(1)}\cdot\vec{\tau}^{(2)} = \tau_{\xi}^{1}\tau_{\xi}^{2} + \tau_{\eta}^{1}\tau_{\eta}^{2} + \tau_{\zeta}^{1}\tau_{\zeta}^{2}$$

il est aussi diagonal dans l'espace ξ^{r}, ξ^{a}

On peut facilement voir que

$$(\vec{\tau}^{(1)}\cdot\vec{\tau}^{(2)})\,\xi^{r} = +\xi^{r}$$
$$(\vec{\tau}^{(1)}\cdot\vec{\tau}^{(2)})\,\xi^{a} = -3\,\xi^{a}$$

On aura donc $(\vec{\tau}^{(1)}\cdot\vec{\tau}^{(2)}) = 2\,P_{12}^{\tau} - 1$

L'opérateur d'échange pour la charge s'exprime donc simplement dans le formalisme du spin isotopique par

$$P_{12}^{\tau} = \frac{1 + (\vec{\tau}^{(1)}\vec{\tau}^{(2)})}{2} \qquad (P_{12}^{\tau})^{2} = 1$$

Le principe de Pauli s'exprimera dans le formalisme du spin isotopique en disant que les états ψ doivent être antisymétrique par rapport à l'échange simultané des coordonnées d'espace, de spin, et de charge des deux nucléons

$$P_{12}\,P_{12}^{\sigma}\,P_{12}^{\tau}\,\psi = -\psi$$

Dans le cas n-n et p-p, cela implique une restriction pour la symétrie de la fonction d'onde spatiale. Dans le cas n-p cette restriction, au contraire, n'existe pas, comme il faut s'y attendre puisqu'il s'agit de deux particules discernables.

On peut observer que, comme nous avons pu le faire dans le cas du spin isotopique, nous pouvons écrire

$$P_{12}^{\sigma} = \frac{1 + \vec{\sigma}^{(1)}\cdot\vec{\sigma}^{(2)}}{2} \qquad (P_{12}^{\sigma})^{2} = 1$$

et l'opérateur échangeant les coordonnées d'espace prend la forme

$$P_{12} = -P_{12}^{\sigma}\,P_{12}^{\tau} = -\frac{1 + \vec{\sigma}^{(1)}\vec{\sigma}^{(2)}}{2}\cdot\frac{1 + \vec{\tau}^{(1)}\vec{\tau}^{(2)}}{2}$$

Cet opérateur commute avec le spin total et avec le spin isotopique.

Nous pouvons retourner maintenant au problème de la forme du potentiel phénomélogique nucléaire. Que peut-on dire de la dépendance du potentiel nucléaire des coordonnées spatiales $\vec{r}_{1}$, $\vec{r}_{2}$ des nucléons et de leurs impulsions $\bar{p}_1$ $\bar{p}_2$?

Le potentiel doit dépendre seulement de la coordonnée relative $\bar{r}_1 - \bar{r}_2 = \bar{r}$ afin d'être invariant. par rapport aux translations spatiales, et de la vitesse relative afin d'être invariante pour une transformation de Galilée (nous nous bornons au cas non relativiste).

Pour ce qui concerne ce dernier point, vu que l'opérateur correspondant à une transformation infinitésimale de Galilée dans une direction $\bar{s}$ peut s'écrire :

$$\frac{i}{\hbar}\left\{ M\bar{r}_g - t\bar{P} \right\}\bar{s}$$

(où r_G est le vecteur du centre de gravité, P l'impulsion totale du système) et que r_G ne commute pas avec P, l'interaction ne peut dépendre de r_G ou de $\bar{P}$. L'opérateur de la vitesse relative

$$\frac{\bar{p}_1}{m_1} - \frac{\bar{p}_2}{m_2}$$

commutant au contraire avec $r_G$et p peut donc être admis. Dans le cas nucléaire, m_1 est assez proche de m_2. Nous pouvons donc supposer que le potentiel V dépend de l'impulsion relative $\bar{p}_1 - \bar{p}_2 = \bar{p}$.

Avant tout nous désirons établir la forme de V dans le cas d'une force indépendante de la vitesse et nous nous bornerons au début à un système de deux particules égales p-p ou n-n pour lesquelles il n'estpas nécessaire de faire intervenir le spin isotopique.

Le potentiel V peut être soit :

a) indépendent de $\vec{\sigma}_1$ et $\vec{\sigma}_2$

b) linéaire en $\vec{\sigma}_1$ et $\vec{\sigma}_2$

c) bilinéaire en $\vec{\sigma}_1$ et $\vec{\sigma}_2$

Les diverses possibilités de dépendance des coordonnées du spin pour V sont exhaustives puisque $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 1$

On a donc, dans les 3 cas :

a) 1

b) $\bar{v}_1 = \bar{\sigma}_1 + \bar{\sigma}_2$ $\qquad \bar{v}_2 = \bar{\sigma}_1 - \bar{\sigma}_2$

c) $(\bar{\sigma}_1 \cdot \bar{\sigma}_2)$, $\bar{v}_3 = \bar{\sigma}_1 \times \bar{\sigma}_2$

et $t_{ik} = \frac{1}{2}\left(\sigma_{in}^{(1)}\sigma_{nk}^{(2)} + \sigma_{in}^{(2)}\sigma_{nk}^{(1)}\right) + \frac{2}{3}\delta_{ik}(\sigma^{(1)}\cdot\sigma^{(2)})$

t_{ik} est un tenseur à 5 dimensions dont la trace est nulle.

Dans l'espace produit $R_1 \times R_1$ de deux espaces vectoriels R_1 on a trois sous espaces invariants, un scalaire R_o, un vecteur R_1, un tenseur symétrique R_2, soit en tout 9 composantes. Nous écrirons

$$R_1 \times R_1 = R_o + R_1 + R_2$$

La dépendance du pspin peut être complètement caractérisée par l'espace somme :

$$R = 2R_o + 3\,R_1 + R_2 \quad .$$

Il est encore utile de suivre le comportement lié aux inversions d'espace et de temps. Nous l'indiquerons par des indices $\pm$ placés en haut des lettres, à droite pour les inversions d'espace et à gauche pour les inversions de temps. Dans l'espace ordinaire σ est un vecteur axial qui change de signe pour une inversion par rapport au temps comme un moment angulaire; on peut donc écrire :

$$R = 2\ {}^{+}R_o^{+} + 2\ {}^{-}R_1^{+} + {}^{+}R_1^{+} + {}^{+}R_2^{+}$$

Chacun de ces sous espaces . des variables de spin doit être multiplié par un sous espace correspondant des variables de position, afin d'obtenir une quantité scalaire -c'est à dire indépendante des rotations d'espace et des inversions d'espace et de temps. Avec les coordonnées d'espace il est possible de former

- un scalaire ${}^{+}S_o^{+}$ qui peut être une fonction arbitraire $F(r)$ dépendant de $|\vec{r}| = r$
- un vecteur ordinaire ${}^{+}S_1^{-}$ qui doit être du type $\vec{r}\, f(r)$
- un tenseur symétrique ${}^{+}S_2^{+}$ dent la trace est nulle qui est du type

$$\left(x_i\, x_k - \frac{1}{3}\, \delta_{ik}\, r^2 \right) g\,(r)$$

etc....

Les fonctions $F\,(r)$ $f(r)$ $g(r)$ doivent être réelles si l'on veut que l'interaction soit Hermitique, ce qui est nécessaire pour la conservation du nombre de particules.

On a donc trois possibilités.

$$2\ {}^{+}R_o^{+}\ .\ {}^{+}S_o^{+} \quad \text{et} \quad {}^{+}R_2^{+}\ .\ {}^{+}S_2^{+}$$

Les trois interactions étant

$$1,\ P_{12}^{\sigma},\ t_{ik}\left(x_i x_k - \frac{1}{3} r^2\, \delta_{ik}\right)$$

à . une fonction multiplicative près ne dépendant que de la valeur absolue de r.

L'interaction du type $t_{ik}\,(x_i x_k - \frac{1}{3}\, r^2\, \delta_{ik})$

ou aussi

$$S_{12} = \left[\frac{3\,(\bar{\sigma}^{(1)} \cdot \bar{r})(\bar{\sigma}^{(2)} \cdot \bar{r})}{r^2} - (\sigma^{(1)} \cdot \sigma^{(2)}) \right]$$

s'appelle force tensorielle. Lorsque l'on désire étendre ce type d'interaction au cas où le potentiel dépend de l'impulsion p, le nombre des diverses possiblités n'augmente pas beaucoup si l'on se restreint à une dépendance au maximum linéaire en p, qui est un vecteur d'un espace ${}^{-}T_1^{-}$ Il suffit en effet d'ajouter le seul espace produit

$$2\;{}^{-}R_1^{+} \times {}^{-}T_1^{-} = 2\,({}^{+}U_2^{-} + {}^{+}U_1^{-} + {}^{+}U_0^{-})$$

qui lorsqu'on le combine avec l'espace ordinaire ne donne que deux autres interactions possibles :

$$2\;{}^{+}U_1^{-} \times {}^{+}S_1^{-}$$

qui sont

$$\left[(\bar{\sigma}_1 + \bar{\sigma}_2) \times p\right] \bar{r}$$

$$\left[(\bar{\sigma}_1 - \bar{\sigma}_2) \times p\right] \bar{r}$$

La première de ces deux interactions correspond au couplage spin orbite L.S, la seconde doit être exclue dans le cas de particules indiscernables parce qu'elle n'est pas symétrique en 1 et 2. L'introduction du spin isotopique augmente le nombre d'interactions possibles. Comme dans le cas du spin ordinaire, il faut ici distinguer pour le potentiel V entre 3 cas :

a) V indépendant de τ_1 et τ_2

b) V linéaire en τ_1 et τ_2

C) V bilinéaire en τ_1 et τ_2.

Il faut en plus inposer la restriction pour V de conserver la charge totale , c'est-à-dire de commuter avec $\tau_3^{(1)} + \tau_3^{(2)}$ ce qui conduit aux possibilités suivantes :

a) 1

b) $\tau_3^{(1)} + \tau_3^{(2)}$ et $\tau_3^{(1)} - \tau_3^{(2)}$

c) $\tau_3^{(1)}\, \tau_3^{(2)}$; $\tau_+^{(1)}\, \tau_-^{(2)}$ $\tau_-^{(1)}\, \tau_+^{(2)}$

Comme l'on a

$$\tau_-^{(1)}\tau_+^{(2)} + \tau_+^{(1)}\tau_-^{(2)} = 2\left(\tau_\xi^{(1)}\tau_\xi^{(2)} + \tau_\eta^{(1)}\tau_\eta^{(2)}\right)$$

on peut aussi choisir dans le cas c)

c) $\tau_3^{(1)}\tau_3^{(2)}$, P_{12}^τ , $\tau_-^{(1)}\tau_+^{(2)} - \tau_+^{(1)}\tau_-^{(2)}$

Ces opérateurs sont invariants pour les rotations et inversions de l'espace ordinaire. En ce qui concerne l'inversion du temps τ_3 reste inchangé car il conserve la charge, mais τ_+ et τ_- doivent être remplacés par leurs complexes conjugués τ_- et τ_+. Tous les opérateurs que nous avons écrits ci-dessus sont invariants par rapport à l'inversion du temps, sauf

$\frac{1}{2}(\tau_-^{(1)}\tau_+^{(2)} - \tau_+^{(1)}\tau_-^{(2)}$) qui change de signe.

Les interactions possibles sont donc données pour les produits de

$1,\ P_{12}^\tau, \tau_3^{(1)}+\tau_3^{(2)}$ et $\tau_3^{(1)}\tau_3^{(2)}$

avec

$1,\ P_{12}^\sigma,\ (\bar\sigma_1+\bar\sigma_2)\ x\bar{p}\ .\bar{r}$ $\qquad S_{12}$

donnant en tout 16 cas possibles et les produits

$(\tau_3^{(1)} - \tau_3^{(2)}\ .\ (\bar\sigma_1 - \bar\sigma_2)\ x\bar{p}\ .\ \bar{r}$

(qui intervient seulement dans le cas n-p) et

$(\tau_-^{(1)}\tau_+^{(2)} - \tau_+^{(1)}\tau_-^{(2)})\ \ (\bar\sigma_1\ x\ \bar\sigma_2)\ x\bar{p}.\bar{r}.$

(qui est aussi invariant pour l'inversion du temps). Ces deux derniers types d'interactions ne conservent pas le spin, tandis que les autres 1, P_{12}^σ, L.S, S_{12} admettent S^2 (carré du spin total) comme un bon nombre quantique. 1, P_{12}^σ, L.S. commutent aussi avec L^2 (carré du moment angulaire); S_{12} ne commute pas avec L^2, commutant seulement avec le moment total $(L + S)^2$ et dans ce cas L^2 n'est pas un bon nombre quantique.

La forme la plus générale du potentiel nucléaire - en se restreignant au cas où l'on a au maximum une dépendance linéaire en p, et à l'approximation non relativiste, peut être une combinaison arbitraire des 18 cas que nous avons indiqué plus haut. Si l'on veut que le spin isotopique total soit un bon nombre quantique, il faut se restreindre à 8 cas seulement, car, en ce qui concerne la dépendance du spin isotopique total, on ne peut prendre que 1 et P_{12}^τ, donnant les possibilités

$$\begin{matrix} 1 \\ P_{12} \end{matrix}\ ,\ \left\{\begin{matrix} 1 \\ P_{12}^\sigma \\ L.S \\ S \end{matrix}\right.$$

Les forces les plus étudiées en théorie sont les forces statiques. Les forces de Wigner et Majorana $(W + m P_{12}) V(r)$ sont indépendantes du spin et du spin isotopique. Les forces du type P_{12}^{σ} s'appellent forces de Bartlett, et celles du type P_{12}^{τ}, forces de Heisenberg.

CH. II. CHOCS ENTRE 2 NUCLEONS.

Afin d'interpréter dans le cadre d'un schéma Hamiltonien les configurations possibles pour 2 nucléons se trouvant dans un état d'énergie $E > 0$, nous devons nous occuper de la théorie du choc de 2 nucléons avec des interactions telles qu'on les a décrites dans le chapitre précédent.

Nous nous occuperons d'abord de la théorie du choc dans le cas d'un potentiel $V(r)$ qui tend vers zéro plus vite que $\frac{1}{r}$ quand r tend vers l'infini, puis d'un potentiel correspondant à la force tensorielle $S_{12}\ g(r)$ où $g(r)$ tend vers 0 comme $\frac{1}{r^{1+\varepsilon}}$ avec $\varepsilon > 0$ et enfin du couplage L.S.

La dependance du spin ne donne lieu à aucune difficulté, puisque le spin est toujours un bon nombre quantique prenant les valeurs 1 et 0 (correspondant respectivement à un état triple et un état simple). La présence d'un potentiel coulombien tendant vers 0 comme $\frac{1}{r}$ demande une étude séparée. Nous ne nous en occuperons que brièvement car il s'agit simplement d'une légère modification du cas où il n'y a pas de potentiel de Coulomb.

1.- Champ central.

Dans la réprésentation de Schroedinger et dans le système de référence lié au centre de gravité, l'équation du mouvement s'écrit

$$(\Delta + k^2)\psi = U\psi$$

$$k^2 = \frac{2\mu E}{\hbar^2}, \qquad (r) = \frac{2\mu V}{\hbar^2}$$

μ est la masse réduite.

On a approximativement $\mu \simeq \frac{M}{2}$ où M est la masse d'un nucléon. Comme le moment angulaire L^2 et sa composante L_z sont de bons nombres quantiques, nous pouvons écrire la solution comme une superposition d'états correspondants aux différentes valeurs de $L^2 = \ell(\ell+1)$ et $L_z = 0$, car nous prenons l'axe z comme direction d'incidence et que nous admettons une symétrie initiale autour de cet axe.

Posons

$$kr\,\psi = \sum u_l(r)\; Y_l^0(\theta, \varphi)$$

$Y_l^m(\theta,\varphi)$ sont les fonctions normées sur la sphère unité,

$Y_l^0 \sqrt{\frac{2l+1}{4\pi}}\; P_l(\cos\theta)$ où P_l est le polynome de Legendre d'ordre l.

L'équation radiale pour $u_l(r)$ s'écrit :

$$\left(\frac{d^2}{dr^2} + k^2 - \frac{l(l+1)}{r^2}\right) u_l = U u_l \tag{1}$$

où $U(r)$ tend vers 0 plus vite que $\frac{1}{r}$ quand r tend vers l'infini.

$$u_l \underset{r\to\infty}{\longrightarrow} a_l^r u_l^r + b_l^s u_l^s$$

où u_l^r et u_l^s sont les solutions régulières et singulière à l'origin e de

$$\left(\frac{d^2}{dr^2} + k^2 - \frac{l(l+1)}{r^2}\right) u = 0$$

$$\left.\begin{aligned} u_l^r &= \sqrt{\frac{\pi kr}{2}}\; J_{l+\frac{1}{2}}(kr) = (kr)^{l+1}\left(-\frac{1}{kr}\,\frac{d}{d(kr)}\right)^l \frac{\sin kr}{kr} \\ u_l^s &= (-1)^l \sqrt{\frac{\pi kr}{2}}\; J_{-l-\frac{1}{2}}(kr) = (kr)^{l+1}\left(-\frac{1}{kr}\,\frac{d}{d(kr)}\right)^l \frac{\cos kr}{kr} \end{aligned}\right\} \tag{2}$$

Pour r $\longrightarrow \infty$

$$\left.\begin{aligned} u_l^r &\simeq \sin\left(kr - \frac{l\pi}{2}\right) \\ u_l^s &\simeq \cos\left(kr - \frac{l\pi}{2}\right) \end{aligned}\right\} (3) \qquad (3)$$

Pour r $\longrightarrow$ 0

$$\begin{aligned} u_l^r(0) &\simeq \frac{2^l}{2l+1}\,\frac{l!}{(2l)!}\,(kr)^{l+1} = \frac{(kr)^{l+1}}{(2l+1)!!} \\ u_l^s(0) &\simeq \frac{1}{2^l}\,\frac{(2l)!}{l!}\left(\frac{1}{kr}\right)^l = (2l-1)!!\left(\frac{1}{kr}\right)^l \end{aligned} \tag{4}$$

$$u_0^s(0) \simeq 1$$

Nous utiliserons aussi les 2 fonctions :

$$u_l^+ = u_l^s + i\,u_l^r \underset{r\to\infty}{\simeq} e^{i\left(kr - \frac{l\pi}{2}\right)}$$

$$u_l^- = u_l^s - i\,u_l^r \underset{r\to\infty}{\simeq} e^{-i\left(kr - \frac{l\pi}{2}\right)}$$

On peut remarquer que

$$u_\ell^r \frac{d u_\ell^s}{d(kr)} - u_\ell^s \frac{d u_\ell^r}{d(kr)} = -1$$

Avec une onde plane incidente

$$kr\, e^{ikz} = \sum_\ell 2\sqrt{(2\ell+1)\pi}\; i^\ell \frac{u_\ell^+ - u_\ell^-}{2i} Y_\ell^0$$

afin que pour chaque direction θ, φ la différence $\psi^f - \psi^i$ soit proportionnelle à u_ℓ^+ qui re-présente une onde sortante, on doit avoir comme comportement de la fonction à l'infini

$$kr(\psi^f - \psi^i) \cong \sum (2\ell+1)\, i^\ell \frac{1}{2i} (S_\ell - 1) u_\ell^+ P_\ell(\cos\theta)$$

$$kr\,\psi^f \simeq \sum_\ell 2\sqrt{(2\ell+1)\pi}\; i^\ell \frac{S_\ell u_\ell^+ - u_\ell^-}{2i} Y_\ell^0$$

L'amplitude de diffusion est donnée par

$$f(\theta) = \frac{1}{k} \sum (2\ell+1) \cancel{i^\ell} \frac{1}{2i} (S_\ell - 1) P_\ell(\cos\theta) \quad (6)$$

et la section efficace est

$$\sigma(\theta)\, d\Omega = \frac{1}{k^2} \left| \sum (2\ell+1)\, i^\ell \frac{1}{2i} (S_\ell - 1) P_\ell \right|^2 d\Omega$$

pour un choc élastique il faut poser $|S_\ell|^2 = 1$ pour qu'on ait conservation du flux. Nous écrirons $S_\ell = e^{2i\delta_\ell}$ où δ_ℓ est une phase réelle.

$$\sigma(\theta) = \frac{1}{k^2} \left| \sum (2\ell+1) \cancel{i}\, e^{i\delta_\ell} \sin\delta_\ell\, P_\ell \right|^2 \quad (7)$$

Il faut évaluer δ_ℓ ou mieux,

$$e^{i\delta_\ell} \sin\delta_\ell = \frac{1}{\lambda_\ell - i} \qquad \lambda_\ell = \frac{i(S_\ell + 1)}{(S_\ell - 1)}$$

$$\lambda_\ell = \cot g\, \delta_\ell$$

Cela équivaut à chercher une solution de (1) dont le comprtement à l'infini soit :

$$u_\ell^f \simeq A(\lambda_\ell u_\ell^r + u_\ell^s)$$

et

$$S_\ell = \frac{\lambda_\ell - i}{\lambda_\ell + i}$$

(S_ℓ est le cas le plus élémentaire de la matrice S de Heisenberg et se réduit à un nombre).

$$\sigma_{tot} = \int \sigma(\theta)\, d\Omega = \frac{4\pi}{k^2} \sum_0^\infty (2\ell+1) \frac{1}{1+\lambda_\ell^2}$$

2.- Formule intégrale pour λ_ℓ

On peut écrire une formule intégrale qui exprime λ_ℓ en fonction du comportement asymptotique de la solution u_ℓ :

$$u_\ell \simeq A\left(\lambda_\ell u_\ell^r + u_\ell^s\right)$$

De

$$\frac{d}{dr}\left(u^r u' - u u^{r\prime}\right) = U u^r u$$

(nous omettrons par simplicité l'indice ℓ), on obtient en intégrant :

$$-kA = \int_0^\infty u^r U u\, dr$$

et de

$$\frac{d}{dr}\left(u^s u' - u u^{s\prime}\right) = u^s U u$$

$$kA\lambda = k\left(\frac{u}{u^r}\right)_0 + \int_0^\infty u^s U u\, dr \quad : \quad \text{ahseur}$$

donc

$$\lambda = \frac{-k\left(\frac{u}{u^r}\right)_0 \mp \int_0^\infty u^s U u\, dr}{\int_0^\infty u^r U u\, dr} \qquad (8)$$

qui est indépendante de la normalisation choisie pour u. On obtient pour S l'expression

$$S = \frac{k\left(\frac{u}{u^r}\right)_0 + \int_0^\infty u^- U u\, dr}{k\left(\frac{u}{u^r}\right)_0 + \int_0^\infty u^+ U u\, dr}$$

On peut aussi déduire de ces formules de l'équation intégrale équivalente à (1)

$$u = a u^r + \int_0^\infty \mathcal{G}(r, r')\, U(r')\, u(r')\, dr'$$

où $\mathcal{G}$ (r, r') est la fonction de Green choisie de manière à satisfaire au comportement asymptotique de u .

$$\mathcal{G}(r,r') = \begin{cases} -\frac{1}{k}\, u^2(r)\, u^1(r') & r<r' \\ -\frac{1}{k}\, u^2(r')\, u^1(r) & r>r' \end{cases}$$

On a pour $r \to \infty$

$$u \simeq a\, u^2 - \frac{1}{k}\, u^1(r) \int_0^\infty u^2(r')\, U(r')\, u(r')\, dr'$$

pour $r \to 0$

$$u(0) = a\, u^2(0) - \frac{1}{k}\, u^2(0) \int_0^\infty u^1(r')\, U(r')\, u(r')\, dr'$$

Donc

$$a = \left(\frac{u}{u^2}\right)_0 + \frac{1}{k} \int_0^\infty u^1(r')\, U(r')\, u(r')\, dr'$$

$$\lambda = \frac{-k \left(\frac{u}{u^2}\right)_0 - \int_0^\infty u^1(r)\, U(r)\, u(r)\, dr}{\int_0^\infty u^2\, U\, u\, dr}$$

3.- <u>Développement de λ en série de puissance de l'énergie.</u>

Pour simplifier les formules, il est avantageux de normaliser u de manière que

$$\int_0^\infty u^2\, U u\, dr = -k$$

ce qui signifie que l'on prend

$$u \underset{r\to\infty}{\simeq} \lambda u^2 + u^1$$

Afin d'obtenir un développement en série de puissances paires de k pour λ il est commode au lieu de partir de (8) de suivre la méthode indiquée par Bethe (2) et (3). Nous prendrons le développement autour de la valeur $\bar{k}$ de k? Nous écrirons $\bar{u}$ pour $u\ (\bar{k}\ r)$ et $\bar{\lambda}$ pour $\lambda\ (\bar{k})$. On a :

$$\bar{u}\, u'' - u\, \bar{u}'' + (k^2 - \bar{k}^2)\, u\, \bar{u} = 0$$

et aussi

$$\bar{u}_\infty\, u''_\infty - u_\infty\, \bar{u}''_\infty + (k^2 - \bar{k}^2)\, u_\infty\, \bar{u}_\infty = 0$$

où

$$u_\infty = \lambda\, u^2 + u^1$$
$$\bar{u}_\infty = \bar{\lambda}\, \bar{u}^2 + \bar{u}^1$$

Donc :

$$(\bar{k}^2 - k^2) \int_r^\infty (u\bar{u} - u_\infty \bar{u}_\infty)\, dr = \Big\{(\bar{u} u' - u \bar{u}') - (\bar{u}_\infty u'_\infty - u_\infty \bar{u}'_\infty)\Big\}\Big|_r^\infty$$

En passant à la limite pour $r = 0$, on obtient

$$k^{2l+1}\lambda - \bar{k}^{2l+1}\bar{\lambda} = (k\bar{k})^l (\bar{k}^2 - k^2) \int_0^{\infty\,\prime} (u\bar{u} - u_\infty \bar{u}_\infty)\, dr$$

le signe $\int'$ indiquait que le terme singulier de $u_\infty \bar{u}_\infty$ pour $r = 0$ ne doit pas être pris en considération. La démonstration dérive du fait que pour $r \to 0$

$$\bar{u}^2 u^{2\prime} - u^2 \bar{u}^{2\prime} = 0$$

$$\bar{u}^{2}u^{1\prime} - u^{1}\bar{u}^{2\prime} = -\frac{\bar{k}^{\ell+1}}{k^{\ell}}$$

$$u^{2}\bar{u}^{1\prime} - u^{2\prime}u^{1} = -\frac{k^{\ell+1}}{\bar{k}^{\ell}} \quad ; \quad [\text{cf } (4)]$$

et que $$\frac{d}{dr}\left(\bar{u}^{1}u^{1\prime} - u^{1}\bar{u}^{1\prime}\right) = (\bar{k}^{2} - k^{2})\,u^{1}\bar{u}^{1}$$

$k^{\ell}u_{\ell}(k, r)$ est une fonction paire de k. Il suffit en effet de considérer

$$k^{\ell}u_{\ell}(k, r) = (\lambda u^{1} + u^{2})\,k^{\ell}$$
$$u^{2}(-k, r) = (-1)^{\ell+1}u^{2}(k, r)$$
$$u^{1}(-k, r) = (-1)^{\ell}u^{1}(k, r)$$
$$\lambda(-k) = -\lambda(k)$$

Cette dernière relation étant due au fait que (1) a une seule solution régulière pour $r = 0$, et que l'équation différentielle ne contient pas que kr. Comme la collision est invariante par rapport à une inversion du temps $$S(-k)\,S(k) = 1$$

dont on déduit $$\lambda(-k) = -\lambda(k)$$

Afin d'obtenir en partant de ([illegible]) le développement que l'on désire, il suffit d'écrire le développement de la solution $u(k, r)$ autour de k.

$$k^{\ell}u_{\ell}(k, r) = k^{\ell}\bar{u}_{\ell} + (k^{2} - \bar{k}^{2})\,\bar{u}_{\ell}^{(1)} + (k^{2} - \bar{k}^{2})^{2}\,\bar{u}_{\ell}^{(2)} + \cdots$$

On obtient

$$k^{2\ell+1}\lambda_{\ell} = \bar{k}^{2\ell+1}\bar{\lambda}_{\ell} + (k^{2} - \bar{k}^{2})\,\bar{k}^{2\ell}\int_{0}^{\infty}(v_{\ell\infty}^{2} - u_{\ell}^{2})\,dr + (k^{2} - \bar{k}^{2})^{2}\,\bar{k}^{\ell}\int_{0}^{\infty}(\bar{u}_{\ell}\,\bar{u}_{\ell}^{(1)} - \bar{u}_{\ell\infty}\,\bar{u}_{\ell\infty}^{(1)})\,dr + \cdots \quad (10)$$

4.- Développement au voisinage de l'énergie zéro.

Aux basses énergies, avec des potentiel à courts rayons d'action, on peut se borner au moment angulaire L = 0. Le développement (10) peut s'écrire (4) (5), (on néglige pour la facilité de placer en bas des lettres l'indice o)

$$k\lambda = -\frac{1}{a} + \frac{1}{2}k^{2}\rho - k^{4}\rho^{3}P + \cdots \quad (11)$$

$$-\frac{1}{a} = \lim_{\bar{k}=0}(\bar{k}\bar{\lambda})$$

a est la longueur de diffusion pour l'énergie zéro

$$\rho = \int_{0}^{\infty}(u_{\infty}^{2} - u^{2})\,dr$$

$$u_{\infty} = \lim_{k\to 0}(\lambda\sin kr + \cos kr) = 1 - \frac{r}{a}$$

la quantité ρ est le rayon d'action effectif du potentiel. Cette dénomination est justifiée parce que dans le cas particulier d'un potentiel en forme de puit dont la profondeur est choisie de telle manière que $\frac{1}{a} = 0$ (résonnance pour l'énergie 0) le rayon d'action effectif coincide avec le rayon du puit.

En général le rayon d'action effectif diffère du rayon entrant dans l'expression du potentiel d'une manière d'autant plus sensible que la diminution de V pour $r \to \infty$ est moins rapide. Lorsque l'énergie est assez petite pour que l'on puisse se limiter au premier terme du développement (11), on peut poser :

$$\lambda = -\frac{1}{k\,a} \tag{12}$$

On peut penser que cette relation est encore valable lorsque k est imaginaire en admettant que la fonction k λ soit analytique au voisinage de l'énergie zéro. S (k) = $\frac{\lambda + i}{\lambda - i}$ admet pour 0 : $\lambda = -$ i. $k \simeq -\frac{i}{a}$ donne une valeur approchée pour ce zéro. Ces valeurs de k et λ correspondent à une solution U qui se comporte à l'infini comme $e^{-\frac{r}{a}}$. En effet, le développement (11) a été déduit avec la condition que

$$u_\infty \simeq \lambda u^2 + u^1 = \frac{1}{r_0}\left\{(\lambda+i)\,u^+ - (\lambda-i)\,u^-\right\}$$

Donc

a $>$ 0 signifie que le système possède un état lié d'énergie : - . .

$$E = -\frac{\hbar^2}{2\mu}\,\frac{1}{a^2}$$

a $<$ 0 correspond à un état virtuel; l'existence d'un état réel n'est pas possible, car le comportement à l'infini de la fonction est dans ce cas du type $e^{+\frac{r}{|a|}}$. Ces considérations suffisent à illustrer la signification physique de $\frac{1}{a}$.

<u>Application</u> :

Dans un choc n-p, en supposant une interaction d'un type central, la section efficace peut s'écrire pour des énergies assez faibles dans le cas du spin S = 1

$$\sigma_t = \frac{4\pi}{k^2}\,\frac{1}{1+\lambda_t^2} \simeq \frac{4\pi}{(1/a_t)^2}$$

a_t, dans cette approximation, représente le rayon du deutéron dans l'état fondamental ^3S $a_t \simeq 4,3 \times 10^{-13}$ cm

Aux basses énergies ($k a_t \ll 1$) on peut donc dire que la section efficace totale est celle qui correspond à la diffusion par une sphère rigide ayant un rayon égal à celui du deutéron. Dans le cas du spin S = 0

$$\sigma_s = \frac{4\pi}{(1/a_s)^2}$$

l'expérience donne la valeur de l'amplitude de la diffusion cohérente dans le parahydrogène

$$f = \frac{3a_t + a_s}{2} = -4 . 10^{-13} \text{ cm}$$

qui a donc une valeur négative. a_s est inférieur à 0, donc l'état ^1S du système n - p est virtuel. La section efficace totale est en l'absence de polarisation

$$\sigma = 4\pi \left(\frac{3}{4} a_t^2 + \frac{1}{4} a_s^2 \right) \simeq 20 . 10^{-24} \text{ cm}^2$$

Aux énergies plus élevées (jusque 10 Mev) on obtient une excellente approximation en retenant les deux premiers termes de (11) ce qui donne les formules suivantes meilleures :

$$k \lambda_t = -\frac{1}{a_t} + \frac{1}{2} \rho_t k^2$$

$$\gamma = \frac{1}{a_t} + \frac{1}{2} \rho_t \gamma^2 \qquad (13)$$

$$k \lambda_s = -\frac{1}{a_s} + \frac{1}{2} \rho_s k^2$$

$$-\frac{\hbar^2}{2\mu} \gamma^2 = 2,2 \text{ Mev}$$

est l'énergie de liaison du deutéron

On peut déduire a_t et a_s des deux grandeurs mesurées

$$f = \frac{3a_t + a_s}{2}$$

$$\sigma_0 = 4\pi \left(\frac{3}{4} a_t^2 + \frac{1}{4} a_s^2 \right)$$

section efficace pour les neutrons épithermiques.

ρ_t se déduit de la première relation (13), ρ_s en comparant la forme théorique de la section efficace totale avec la valeur expérimentale (de 0 à 6 Mev) ainsi que des données concernant l'effet photoélectrique. Nous donnons un tableau -(6) des données les plus récentes pour les diverses quantités que nous avons considérées

$a_s = -23,68 \pm 0,08 \times 10^{-13}$ cm

$a_t = 5,38 \pm 0,05 \times 10^{-13}$ cm

$\rho_s = 2,5 \pm 0,5 \times 10^{-13}$ cm

$\rho_t = 1,71 \pm 0,1 \times 10^{-13}$ cm

$\frac{1}{\gamma} = 4,332 \pm 0,025 \times 10^{-13}$ cm

On voit donc que par ce que l'on vient de dire la diffusion n - p jusque 10 Mev peut être décrite au moyen de deux paramètres seulement. On ne peut donc établir la dépendance de r du potentiel statique car il est toujours possible avec un potentiel de forme quelconque de fixer les deux paramètres (profondeur et largeur) de façon que les sections efficaces soient expliquées).

5.- Méthodes d'approximation pour le calcul des déphasages.

Pour apprécier si une forme de potentiel donnée conduit à des résultats valables, il faut pouvoir intégrer l'équation (1). Sauf pour certaines formes particulières de potentiel pour lesquelles l'intégration exacte est possible, par exemple le puit du potentiel, le potentiel de Hulthen pour l'état $l = 0$ qui est du type $\frac{e^{-\chi r}}{1-e^{-\chi r}}$, . Dans la grande majorité des cas, il est indispensable d'employer des méthodes d'approximation. On peut les ranger en 3 catégories :

Méthode variationnelle
Méthode d'approximation de Born
Méthode W K B.

Nous nous intéresserons surtout à la première mentionnant celle de Born et ne nous occupant pas de la mèthode W K B.

Principe variationnel :

Nous commencerons par le principe indiqué par Hulthen. (1)

Considérons l'opérateur différentiel :

$$L = \frac{d^2}{dr^2} + k^2 - \frac{l(l+1)}{r^2} - U$$

l'intégrale

$$J_H = k\lambda - \int_0^\infty u\, L_l(u)\, dr \tag{14}$$

est stationnaire (variation première $\delta J = 0$) si u est solution de $L_l(u) = 0$ et satisfait aux conditions aux limites

$$\begin{cases} u(k,0) = 0 \\ u(k,\infty) = \lambda u^2 + u^1 \end{cases} \qquad (15)$$

On a en effet : $$\delta J_H = -2 \int_0^\infty \delta u \, L(u) \, dr$$

Réciproquement si l'on veut que δJ_H soit nul pour des valeurs arbitraires de δu, il faut que :

$$L(u) = 0$$

Hulthen choisit

$$u = \lambda v^2 + v$$

où v est indépendant de λ.

$$v(k,0) = 0$$
$$v(k,\infty) = u^1$$

On peut alors montrer facilement que

$$J_H = -F + 2(k-N)\lambda - k\eta_\delta \lambda^2 \qquad (16)$$

où

$$N = -\int_0^\infty u^2 U v \, dr$$

$$k\eta_\delta = -\int_0^\infty u^2 U u^2 \, dr$$

$$F = \int_0^\infty v \, L(v) \, dr$$

On déduit de $\dfrac{\partial J_H}{\partial \lambda} = 0$

$$k\eta_\delta \lambda = k - N$$

et en substituant dans (15)

$$J_H = k\lambda - (F + N\lambda)$$

$$\lambda_{stat} = \lambda - \frac{1}{k}(F + \lambda N)$$

Si v dépend linéairement d'un certain nombre de paramètres variationnels c et si $\dfrac{\partial N}{\partial c}$ est indépendant de c, $\dfrac{\partial F}{\partial c}$ linéaire en c, la condition de stationarité $\dfrac{\partial I}{\partial c} = 0$ donne :

$$\frac{\partial F}{\partial c} + 2\lambda \frac{\partial N}{\partial c} = 0$$

λ est fonction linéaire des paramètres c pour l'équation

$$k\eta_\delta \lambda = k - N$$

et à cause de cela les équations qui déterminent c

$$k\,\eta_0 \frac{\partial F}{\partial c} + 2(k-N)\frac{\partial N}{\partial c} = 0$$

sont linéaires pour les paramètres inconnus.

(8) (9)

Méthode de Schwinger .

L'intégrale

$$J_S = J_H - \int_0^\infty \frac{1}{U}\left[L(u)\right]^2 dr \qquad (14')$$

est stationnaire si u est solution de L (u) = 0 et satisfait aux mêmes conditions aux limites que J_H, défini en (14)

On a en effet :

$$\delta J_s \,.\, \delta J_H - 2\int_0^\infty \frac{1}{u} L_1(u)\, \delta\left[L_1(u)\right] dr \qquad (15')$$

Réciproquement, il résulte de $\delta J_s = 0$ que

L (u) = 0

$(J_s)_{ext.} = (J_H)_{ext.} = k\,\lambda_{\text{extrémale}}$

Schwinger choisit :

$$u = \lambda u^2 + \int_0^\infty g(rr')\, U(r')\, y(r')\, dr' \qquad (16')$$

$$g(rr') < \begin{matrix} -\frac{1}{k}\, u^2(r)\, u^1(r') & r < r' \\ -\frac{1}{k}\, u^1(r')\, u^1(r) & r > r' \end{matrix}$$

y est la fonction variationnellle, mais il faut pour satisfaire aux conditions à l'∞ que y soit normé de façon que :

$$-\frac{1}{k}\int_0^\infty u^2\, U\, y\, dr = 1$$

La relation (14') peut alors s'écrire

$$J_s = -\int_0^\infty U y^2 dr + \int_0^\infty dr\, U(r)\, y(r) \int_0^\infty dr'\, g(rr')\, U(r')\, y(r') dr'$$

On peut déduire de (14') que si $u(r) < 0$ La valeur de k λ de Schwinger est > à celle de Hulthen

si $u(r) > 0$; la valeur de k λ de Schwinger est < à celle de Hulthen.

Elimination des conditions de normalisation.

Afin d'éliminer toute restriction de normalisation pour la fonction variationnelle, c'est-à-dire si l'on veut choisir

$$u = \lambda u^2 + v$$

avec $v(0) = 0$

$v(\infty) = c\, u^1$

il suffit de poser dans le cas de Hulthen

$$J_H = \frac{k\lambda}{c} - \frac{\int_0 u\,\mathcal{L}(u)\,dr}{c^2}$$

c peut aussi s'écrire

$$c = -\frac{1}{k}\int_0 u^2\,L_\infty(u)\,dr = -\frac{1}{k}\int_0^\infty u^2\,U u\,dr$$

puisque
$$L_{1\infty}(u) = \left(\frac{d^2}{dr^2} + k^2 - \frac{\ell(\ell+1)}{r^2}\right)u$$

Dans le cas de Schinwger il faut écrire

$$J_S = J_H - \frac{1}{c^2}\int_0^\infty \frac{1}{u}\left[\mathcal{L}(u)\right]^2 dr$$

ou aussi

$$J_S = \frac{-\int_0^\infty v\,y^2\,dr + \int_0^\infty dr\,U y(r)\int_0^\infty dr'\,g(r\,r')\,U(r')\,y(r')}{\left(\frac{1}{k}\int_0^\infty u^2\,U\,y\,dr\right)^2} \qquad (18)$$

Observation :

Le principe variationnel (18) peut se déduire directement de l'équation intégrale pour la fonction u qui s'écrit :

$$u = u^2\left(-\frac{\lambda}{k}\int_0^\infty u^2\,U\,u\,dr\right) + \int_0^\infty g(r,r')\,U(r')\,u(r')\,dr' \qquad (19)$$

où g est la fonction de Green. Il suffit de multiplier par $U(u)$ et d'intégrer de 0 à l'infini pour obtenir l'expression de k λ.

$$k\lambda = \frac{-\int_0 U u^2\,dr + \int_0 g(r\,r')\,U(r')\,u(r')\,dr'\,U(r)\,u(r)\,dr}{\left(\frac{1}{k}\int_0 u^2\,U\,u\,dr\right)^2}$$

On peut encore montrer directement que cette expression est un principe variationnel.

Approximation de Born.

On prend comme approximation d'ordre zéro pour la fonction u, l'état initial u = u (r)

dans le cas $\lambda^{(0)} = \infty$ ou encore S = 1

En première approximation on pose

$$u = u^2 + \int_0^\infty g(r\,r')\,U(r')\,u^2(r')\,dr'$$

Donc (comparer avec (19))

$$\lambda^{(1)} = \frac{1}{-\frac{1}{k}\int_0^\infty u^2\,U\,u^2\,dr}$$

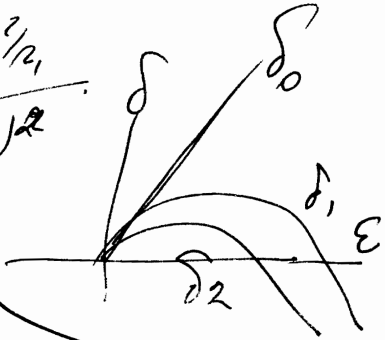
En seconde approximation

$\underline{Y^0_\ell X_m} \rightarrow \underline{Z^m_{\ell,\, J\,J\pm1}}$

$\underline{\ell-1}, \ell, \underline{\ell+1}. \; (Z^m_{\ell,J})$

$\underline{V_2 + V_4^{(a)}}$ $\underline{\mathcal{H}_0}, \underline{\mathcal{H}_1}.$ $S_{12}.$

$\dfrac{e^{-r/r_1}}{(r/r_1)^2}$



$+k^2)\Psi = U\Psi.$

$_{12}\, \underline{Z^m_{\ell J}} = 2\underline{Z^m_{\ell J}}$

$\underline{\text{variant par réflexion}}.$

rité d'être mine'. $Z^{m}_{\ell J}$; $Z^m_{\ell+1,J}, Z^m_{\ell-1,J}$

^m_J $\Psi_m = e^{ikz} X_m$

$Z^m_{J-1,J}$, $Z_{J+1,J}$. u_J w_J.

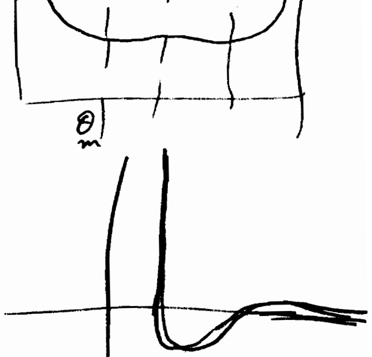
développe en J et m S_{12} $\sum_{J,m} \left(Z_{J,J} ; Z_{J-1,J} ; Z_{J+1,J} \right)$

$$\begin{pmatrix} L_{J-1} & 0 \\ & L_{J+1} \end{pmatrix} \begin{pmatrix} u_J \\ w_J \end{pmatrix} = \begin{pmatrix} f & g \\ g & h \end{pmatrix} \begin{pmatrix} u_J \\ w_J \end{pmatrix}.$$

$V\rho \begin{cases} 0 \end{cases}$ $r > 2{,}6 \cdot 10^{-13}\,\text{cm}$

$= -13{,}27\, \dfrac{1+P_x}{e}$ Mev $r <$

. P_x.



Christian-Noyes.

$\underline{n-p}$ jusqu'à 14 Mev. on a une sym. dans le c.g.

$\underline{E > 14 \text{ Mev. il y a une anisotropie}}$ e^{i}

$\dfrac{d\sigma}{d\Omega}$ sym / $\underline{90^\circ}$. nous avons

$\underline{p-p}$ $\underline{\text{Isotopie } 15^\circ\; 165^\circ}$ $\underline{\text{~~150~~} - 400 \text{ Mev}}.$

$\dfrac{1+P_\sigma P_x}{2}\,\dfrac{r_2}{2}\,\dfrac{}{3}\, e^{-r/r_1}$ $\dfrac{1+P_x}{2}$ Sin

$\dfrac{1-P_x}{2}\left(\dfrac{r_2}{r}\right)^2 e^{-r/r_2}\, S_{12}.$

$\underline{J, J_2, S^2}.$ $Z^m_{\ell J}$ J^2, J_2, S^2, L

$\underline{\text{Je développe } J \text{ donné}}$

$Z^m_{\ell-1,J}$, $Z^m_{\ell J}$, $Z^m_{\ell+1,J}$ $\underline{mqq}$.

$$S_{12,} = \begin{pmatrix} -2\,\frac{J-1}{2J+1} & \frac{6}{2J+1}\sqrt{J(J+1)} & 0 \\ \frac{6}{2J+1}\sqrt{J(J+1)} & -2\,\frac{J+2}{2J+1} & 0 \end{pmatrix}$$

$$SU = \begin{pmatrix}\cos\varepsilon & -\sin\varepsilon\\ \sin\varepsilon & \cos\varepsilon\end{pmatrix}\begin{pmatrix}e^{i\delta_\alpha} & 0\\ 0 & e^{i\delta_\beta}\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix}.$$

$$= \begin{pmatrix}\cos\varepsilon\, e^{i\delta_\alpha} & -\sin\varepsilon\, e^{i\delta_\beta}\\ \sin\varepsilon\, e^{i\delta_\alpha} & \cos\varepsilon\, e^{i\delta_\beta}\end{pmatrix}$$

$$S\begin{pmatrix}\cancel{\cos\varepsilon}\\ \cancel{\sin\varepsilon}\end{pmatrix}\begin{pmatrix}-\sin\varepsilon\\ \cos\varepsilon\end{pmatrix} = e^{i\delta_\beta}\begin{pmatrix}\cancel{\sin\varepsilon}\\ \cancel{\cos\varepsilon}\end{pmatrix}$$

~~$\cos\varepsilon, \sin\varepsilon)$~~ $e^{i\delta_\alpha}\begin{pmatrix}\cos\varepsilon\\ \sin\varepsilon\end{pmatrix}$, $e^{i\delta_\beta}\begin{pmatrix}-\sin\varepsilon\\ \cos\varepsilon\end{pmatrix}$ $\begin{pmatrix} \\ \end{pmatrix}$

$h = Sa.$

$h = e^{i\delta_\alpha}\begin{pmatrix}\cos\varepsilon\\ \sin\varepsilon\end{pmatrix}$

$\boxed{h = Sa}$

$\begin{pmatrix} \\ \end{pmatrix} - e^{i\delta}\begin{pmatrix}\cos\varepsilon\\ \sin\varepsilon\end{pmatrix}$ on peut choisir a comme $e^{+i\delta_\alpha}$.

$\begin{pmatrix}e^{i\delta_\alpha}\\ e^{i\delta_\alpha}\end{pmatrix}(S)$

$-e^{i\delta_\alpha}(1) = e^{+i\delta_\alpha}(\cos\varepsilon$

[illegible]

$$\lambda^{(2)} = \frac{1}{-\frac{1}{k}\int_0^\infty dr\, u^2\, U\left[u^2(r) + \int_0^\infty g(r,t)\, U(t)\, u^2(t)\, dt\right]}$$

Une étude soignée de la validité de l'équation de Born dans le cas nucléaire (10) permet de conclure que pour des énergies de [illegible]00 Mev le développement en série de Born ne peut être employé. Il fournit des valeurs essentiellement différentes de celles qu'on calcule avec plus d'exactitude par des intégrations numériques. Une solution de l'équation de (19) obtenue à partir de la méthode de Fredholm (quotient de deux séries de puissances du paramètres g fixant le potentiel) se montre peu utile pour les calculs numériques parce qu'elle converge trop lentement.

6.- Théorie de la collision dans le cas d'une force tensorielle.

En présence d'un potentiel du type tensoriel :

$$S_{12}\, K(r) = \left\{ 3\, \frac{(\sigma^{(1)} \cdot \vec{r})(\sigma^{(2)} \cdot \vec{r})}{r^2} - \vec{\sigma}^{(1)}\vec{\sigma}^{(2)} \right\} K(r)$$

la théorie du choc est légèrement plus compliquée car le moment orbital L^2 n'est plus un bon nombre quantique. Toutefois puisque le moment total $(L + S)^2$ celui de spin S^2 et $L_z + S_z$ sont de bons nombres quantiques, il est utile de se placer dans la base des fonctions propres $Z^m_{lJ}(\theta, \varphi, \sigma)$ qui correspondent à des valeurs déterminées de J^2, J_z, S^2, L^2. L'opérateur S_{12} sera diagonal par rapport à J^2, J_z S^2 mais non diagonal par rapport à L^2. La force tensorielle n'est efficace que pour les états triples car elle s'annule pour $S = 0$. Dans la suite, lorsque nous parlerons du spin, nous sous-entendrons toujours $S = 1$.

Pour un J donné, les valeurs possibles de ℓ sont $\ell - 1, \ell, \ell + 1$. Dans l'espace formé par $Z^m_{J-1,J}$, $Z^m_{J+1,J}$, $Z^m_{J,J}$ l'opérateur S_{12} est représenté par la matrice :

$$S_{12} = \begin{pmatrix} -2\,\frac{J-1}{2J+1} & \frac{6}{2J+1}\sqrt{J(J+1)} & 0 \\ \frac{6}{2J+1}\sqrt{J(J+1)} & -2\,\frac{(J+2)}{2J+1} & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Il est évident que l'état Z^m_{JJ} n'est pas mélangé avec les 2 autres $Z^m_{J-1,J}$ et $Z^m_{J+1,J}$ car ces deux derniers ont une parité opposée à celle de Z^m_{JJ} (S_{12} invariant pour les inversions d'espace).

Pour l'état $\overset{m}{Z}_{J,J}$ on peut répéter les considérations que nous avons faites auparavant pour un champ central et nous pourrons nous borner au sous espace de $\overset{m}{Z}_{J-1,J}$ et $\overset{m}{Z}_{J+1,J}$. Soient $u_J(r)$ et $w_J(r)$ les parties radiales correspondant aux deux états. On a le système d'équations différentielles

$$\begin{pmatrix} L_{J-1} & 0 \\ 0 & L_{J+1} \end{pmatrix} \begin{pmatrix} u_J \\ w_J \end{pmatrix} = \begin{pmatrix} f & g \\ g & h \end{pmatrix} \begin{pmatrix} u_J \\ w_J \end{pmatrix} \qquad (20)$$

L_J est l'opérateur habituel

$$L_J = \frac{d^2}{dr^2} + k^2 - \frac{\ell(\ell+1)}{r^2} \quad ? \quad (\ell = J)$$

g est proportionnelle à la partie radiale K de la force tensorielle f et g dépendent dans le cas plus général d'une superposition d'un potentiel central V à la force tensorielle de V et de K. En l'absence de force tensorielle (g =0) le système se réduit à 2 équations différentielles indépendantes entre elles. On a pour de grandes valeurs de r :

$$u_J \simeq a_1 e^{-i\left[kr - \frac{J-1}{2}\pi\right]} - b_1 e^{i\left[kr - \frac{J-1}{2}\pi\right]}$$

$$w_J \simeq a_2 e^{-i\left[kr - \frac{J+1}{2}\pi\right]} - b_2 e^{i\left[kr - \frac{J+1}{2}\pi\right]}$$

A l'onde incidente (a_1, a_2) correspond univoquement l'onde sortante (b_1, b_2) et la relation qui lie (b_1, b_2) à (a_1, a_2) est linéaire. Nous écrirons :

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

En général, S_{12} est différent de 0 , car aux paires (1, 0) peut correspondre la paire $(b_1\ b_2)$ avec b_2 différent de 0. On peut dire la même chose pour S_{21}. Nous écrirons pour abréger

$$b = S\,a$$

Au moyen d'une transofrmation orthogoanle

$$U = \begin{pmatrix} \cos\xi & -\sin\xi \\ \sin\xi & \cos\xi \end{pmatrix}$$

il est possible de combiner a et b

$$a' = U^{-1} a$$

$$b' = U^{-1} b.$$

b = S a devenant

$b' = (U^{-1} S U)\, a'$

U peut être choisi pour que $U^{-1} S U$ soit diagonal, La matrice S doit être unitaire et symétrique, L'unitarité est une conséquence de la conservation du flux de particules

$$|a_1|^2 + |a_2|^2 = |b_1|^2 + |b_2|^2$$

La symétrie est une conséquence de l'invariance des lois du choc par rapport à l'inversion du temps $^{(11)}$. La matrice U doit être orthogonale pour conserver l'unitarité et la symétrie de S. Nous appelons $e^{i\delta_\alpha}$ et $e^{i\delta_\beta}$ les deux valeurs propres de S (δ_α et δ_β sont réels) correspondant aux deux vecteurs propres ($\cos \varepsilon$, $\sin \varepsilon$) et ($-\sin \varepsilon$, $\cos \varepsilon$) Nous ppuvons alors parler de deux types de solutions.

Type α ayant le comportement asymptotique :-

$$u_j^\alpha \simeq \cos\varepsilon \left\{ e^{-i[kr - \frac{J-1}{2}\pi]} - e^{2i\delta_\alpha} e^{i[kr - \frac{J-1}{2}\pi]} \right\}$$

$$w_j^\alpha \simeq \sin\varepsilon \left\{ e^{-i[kr - \frac{J+1}{2}\pi]} - e^{2i\delta_\alpha} e^{i[kr - \frac{J+1}{2}\pi]} \right\}$$

Type β ayant le comportement asymptotique

$$u_j^\beta \simeq -\sin\varepsilon \left\{ e^{-i[kr - \frac{J+1}{2}\pi]} - e^{2i\delta_\beta} e^{i[kr - \frac{J-1}{2}\pi]} \right\}$$

$$w_j^\beta \simeq \cos\varepsilon \left\{ e^{-i[kr - \frac{J+1}{2}\pi]} - e^{2i\delta_\beta} e^{i[kr - \frac{J+1}{2}\pi]} \right\}$$

ou aussi

$$\left\{\begin{array}{l} u_j^\alpha \simeq 2i\cos\varepsilon\, e^{i\delta_\alpha} \sin\left[kr - \frac{J-1}{2}\pi + \delta_\alpha\right] \\ w_j^\alpha \simeq 2i\sin\varepsilon\, e^{i\delta_\alpha} \sin\left[kr - \frac{J+1}{2}\pi + \delta_\alpha\right] \end{array}\right. \qquad (21)$$

$$\left\{\begin{array}{l} u_j^\beta \simeq -2i\sin\varepsilon\, e^{i\delta_\beta} \sin\left[kr - \frac{J-1}{2}\pi + \delta_\beta\right] \\ w_j^\beta \simeq 2i\cos\varepsilon\, e^{i\delta_\beta} \sin\left[kr - \frac{J+1}{2}\pi + \delta_\beta\right] \end{array}\right. \qquad (21')$$

Les deux types de solutions α et β se caractérisent donc par le fait qu'elles ont le même déphasage (rendant la matrice S diagonale)

On a :

$$\left(\frac{u_j^\alpha}{w_j^\alpha}\right)_{r=\infty} = -\operatorname{cotg} \varepsilon_j \; ; \; \left(\frac{u_j^\beta}{w_j^\beta}\right)_{r=\infty} = \operatorname{tg} \varepsilon_j \qquad (22)$$

Le problème mathématique consiste à trouver deux solutions de (20) ayant le comportement asymptotique (21) et (21'). Les paramètres nécessaires pour la détermination de la section efficace sont donc : ε , δ_α et δ_β

Calcul de la section différentielle totale.

L'onde plane incidente

$$kr\Psi_{inc} = \sum_{0}^{\infty}{}_{\ell} \sqrt{4\pi(2\ell+1)} \; i^{\ell} u_{\ell}^{2} \; Y_{\ell}^{0}(\theta,\varphi) \chi_{m}$$

peut s'écrire dans le système des $Z_{\ell,j}^{m}$

$$kr\,\Psi_{in} = \sqrt{4\pi} \sum_{0}^{\infty}{}_{\ell} \sum_{\ell-1}^{\ell+1}{}_{j} (2\ell+1)^{\frac{1}{2}} i^{\ell} u_{\ell}^{2} (\ell 1 j m / \ell 1 0 m) Z_{\ell j}^{m}$$

$$= \sqrt{4\pi} \sum_{0}^{\infty}{}_{j} \sum_{j-1}^{j+1}{}_{\ell} (2\ell+1)^{1/2} i^{\ell} u_{\ell}^{2} (\ell 1 J m / \ell 1 0 m) Z_{\ell j}^{m} -$$

$$= \sqrt{4\pi} \sum_{0}^{\infty}{}_{j} \Big(c_{-}^{m} i^{j-1} u_{j-1}^{2} Z_{-}^{m} + c^{m} i^{j} \underset{\times u_{J}^{(2)}}{Z^{m}} + c_{+}^{m} i^{j+1} \underset{\times u_{J+1}^{(2)}}{Z_{+}^{m}} \Big)$$

En écrivant $Z_{\pm}^{m}$ pour $Z_{j\mp1,j}^{m}$

Z^{m} pour Z_{jj}^{m}

les coefficients C_{-}^{m}, C^{m}, C_{+}^{m} qui font passer de la base $Y_{\ell}^{0}\chi_{m}$ à la base Z sont donnés par

$c \backslash m$	1	0	-1
c_{-}^{m}	$\sqrt{\frac{J+1}{2}}$	$\sqrt{J}$	$\sqrt{\frac{J+1}{2}}$
c^{m}	$-\sqrt{\frac{2J+1}{2}}$	0	$\sqrt{\frac{2J+1}{2}}$
c_{+}^{m}	$\sqrt{\frac{J}{2}}$	$-\sqrt{J+1}$	$\sqrt{\frac{J}{2}}$

Nous pouvons remarquer que :

$$\sum_{m} |c_{-}^{m}|^{2} = \sum_{m} |c^{m}|^{2} = \sum |c_{+}^{m}|^{2}$$
$$= 2J+1$$

$$\sum_{m} c_{-}^{m} c_{+}^{m} = \sum_{m} c_{-}^{m} c^{m} = \sum_{m} c_{-}^{m} c_{+}^{m} = 0$$

De la condition

$$kr\,\Psi_{tot}^{in} = kr\,\Psi_{inc}^{in}$$

on déduit

$$kr\,\Psi_{scatt} = kr\left(\Psi_{tot}^{snt} - \Psi_{inc}^{snt}\right) \simeq$$

$$\frac{\sqrt{4\pi}}{2i} \left\{ (c_{-}^{m}, c_{+}^{m})(S-1)\begin{pmatrix} Z_{-}^{m} \\ Z_{+}^{m} \end{pmatrix} + c(s-1) Z \right\} e^{ikr}$$

En écrivant :

$$(\alpha_{-}^{m}, \alpha_{+}^{m}) = (c_{-}^{m}, c_{+}^{m})\, \mathcal{V}$$

$$\mathcal{V}^{-1} \begin{pmatrix} Z_{-}^{m} \\ Z_{+}^{m} \end{pmatrix} = \begin{pmatrix} Z_{-}' \\ Z_{+}' \end{pmatrix}$$

$$\delta_{\alpha} = \delta_{-} \qquad \delta_{\beta} = \delta_{+}$$

$$c = \alpha$$

Il en résulte pour l'amplitude de diffusion

$$f^{m}(\theta) = \frac{\sqrt{4\pi}}{k} \sum \left(\alpha_{-}^{m} e^{i\delta_{-}} \sin\delta_{-} Z_{-}' + \alpha_{+}^{m} e^{i\delta_{+}} \sin\delta_{+} Z_{+}' + \alpha^{m} e^{i\delta} \sin\delta\, Z \right)$$

d'où on deduit la section efficace différentielle (11). Pour la section efficace totale, nous remarquons que $\int |Z'^{m}_{\mp}|^2 d\omega = 1$

$$\int \left| f^{m}_{(\theta)} \right|^2 d\omega = \frac{4\pi}{k^2} \Sigma \left(|\alpha^{m}_{-}|^2 \sin\delta_- + |\alpha^{m}_{+}|^2 \sin\delta_+ + |\alpha^{m}|^2 \sin^2\delta \right)$$

et avec une onde initiale non polarisée

$$\sigma^{tr}_{tot} = \frac{1}{3} \sum_m \int \left| f^{m}_{(\theta)} \right|^2 d\omega = \frac{1}{3} \frac{4\pi}{k^2} \sum_j (2j+1) \left(\sin^2\delta_- + \sin^2\delta_+ + \sin^2\delta \right)$$

Nous remarquons que pour calculer la section efficace totale, il n'est pas nécessaire de connaitre le paramètre $\mathcal{E}$.

Méthode variationnelle dans le cas de la force tensorielle.

On peut généraliser facilement pour ce cas les principes variationnels que nous avons discutés dans le cas du champ central. Nous ne nous intéresserons brièvement qu'à ma méthode de Schwinger (12) appliquée pour les chocs n - p ou p - p. Le système (20) peut s'écrire sous la forme intégrale :

$$\begin{pmatrix} u \\ w \end{pmatrix} = \begin{pmatrix} a\, u^{r}_{-} \\ b\, u^{r}_{+} \end{pmatrix} + \int_0^\infty \begin{pmatrix} G_- & 0 \\ 0 & G_+ \end{pmatrix} \begin{pmatrix} f(r') & g(r') \\ g(r') & h(r') \end{pmatrix} \begin{pmatrix} u(r') \\ w(r') \end{pmatrix} dr'$$

$G(r,r')$ sont les fonctions de Green relatives aux 2 opérateurs

$$L_{j\mp1}(u) = 0 \qquad G_{\mp} = \begin{cases} -\frac{1}{k} u^{r}_{\mp}(r)\, u^{s}_{\mp}(r') & r < r' \\ -\frac{1}{k} u^{r}_{\mp}(r')\, u^{s}_{\mp}(r) & r > r' \end{cases}$$

Pour les solutions du type (21) ou (21') on a :

$$\lambda = \frac{a}{-\frac{1}{k}\int_0^\infty u^{r}_{-} (fu + gw)\, dr} = \frac{b}{-\frac{1}{k}\int_0^\infty u^{r}_{+} (gu + hw)\, dr}$$

$$\begin{pmatrix} u \\ w \end{pmatrix} = u^{r}_{\mp} \left[-\frac{\lambda}{k} \int_0^\infty u^{r}_{\mp} \begin{pmatrix} fu + gw \\ gu + hw \end{pmatrix} dr \right] + \int_0^\infty G_{\mp} \begin{pmatrix} f'u' + g'w' \\ g'u' + h'w' \end{pmatrix} dr' \qquad (24)$$

On a écrit f' pour f (r') et de même pour les autres grandeurs accentuées. En multipliant la première relation (24) par fu + gw et la seconde par gu + hw, on obtient en intégrant et sommant :

$$k\lambda = \frac{\int_0^\infty (u, w) \begin{pmatrix} f & g \\ g & h \end{pmatrix} \left[-\begin{pmatrix} u \\ w \end{pmatrix} + \int_0^\infty \begin{pmatrix} G_- & 0 \\ 0 & G_+ \end{pmatrix} \begin{pmatrix} f' & g' \\ g' & h' \end{pmatrix} \begin{pmatrix} u' \\ w' \end{pmatrix} dr' \right]}{\left(\frac{1}{k} \int_0^\infty u^{r}_{-} (fu + gw)\, dr \right)^2 + \left(\frac{1}{k} \int_0^\infty (gu + hw)\, u^{r}_{+}\, dr \right)^2}$$

Il est facile d'établir que $k\lambda$ est stationnaire pour des variations arbitraires de u et w qui satisfont aux conditions aux limites (21) et (21') et nulle pour r = 0.

La fonction d'essai la plus élémentaire consiste à prendre un seul paramètre variationnel, donné par le rapport de u et w pour $r = \infty$.

7.- Quelques données sur l'état lié du deutéron.

On peut aussi obtenir des renseignements pour la force n - p dans l'état triplet en partant de la configuration stable du deutéron. La présence d'un moment quadrupolaire électrique

$$a = +\,2{,}76 \quad .\ 10^{-27}\,\mathrm{cm}^2$$

démontre que l'état fondamental du deutéron ne peut être un état à symétrie sphérique. Avec un moment angulaire total J = 1 et un spin S = 1, l'état fondamental du deutéron doit être un mélange $^3S_1 + {}^3D_1$. Si χ_m sont les fonctions propres du spin (S = 1), $u(r)$, $w(r)$ les deux parties radiales, on peut écrire :

$$r\psi = \left[u + \frac{1}{\sqrt{8}}\, S_{12}\, w\right]\chi_m \tag{25}$$

où S_{12} est l'opérateur de la force tensorielle. On a :

$$\int r^2 |\psi|^2 \,dr\, d\omega = \int_0^\infty (u^2 + w^2)\,dr$$

et

$$a = \frac{1}{4}\,\frac{\langle \psi, (3z^2 - r^2)\psi\rangle_{m=1}}{\langle \psi, \psi\rangle_{m=1}} = \frac{1}{10}\,\frac{\int_0^\infty r^2 \left(uw - \frac{1}{2} w^2\right) dr}{\int_0^\infty (u^2 + w^2)\,dr}$$

Pour le moment dipolaire magnétique, on a, en négligeant les corrections relativistes :

$$\mu = \frac{\langle (\mu_p + \mu_n) S_z + \frac{1}{2} L_z \rangle_{m=1}}{\langle \psi, \psi\rangle_{m=1}} = (\mu_p + \mu_n) - \frac{3}{2}\left(\mu_p - \mu_n - \frac{1}{2}\right) P_d$$

mesuré en magnétons nucléaires

$$P_d = \frac{\int_0^\infty w^2\,dr}{\int_0^\infty (u^2 + w^2)\,dr}$$

De la valeur expérimentale $\mu = 0{,}857$, ($\mu_p = 2.793$ et $\mu_n = -1{,}923$), on déduit $P_D \cong 0.04$ un autre paramètre important pour la détermination du potentiel n - p dans l'état triplet est l'énergie de liaison

$$\varepsilon_D = 2{,}2 \text{ Mev}$$

Une étude systématique de potentiel phénoménologique a été faite récemment par Feshbach et Schwinger (14). Ces auteurs ont pris un potentiel central

$$V_c = V_0^c \frac{e^{-r/r_c}}{r/r_c}$$

et un potentiel tensoriel

$$V_t = \gamma V_0^c \frac{e^{-r/r_t}}{r/r_t} S_{12}$$

qui ont un comportement spatial du type Yukawa.

On a, en tout, quatre constantes V_0^c, γ, r_c, r_t. Deux de celles-ci sont fixées par les valeurs expérimentales de l'énergie de liaison et du moment quadrupolaire. Nous ne pouvons nous occuper ici des méthodes variationnelles qui permettent le calcul des parties radiales $u(r)$ et $w(r)$ de la fonction d'onde (25). Le calcul est beaucoup plus facile que dans le cas correspondant du spectre continu. On peut avoir une grande confiance en l'exactitude de la valeur propre correspondante (comparer avec le travail cité en (14)).

Il est possible d'exprimer V_0^c et γ en fonction de r_t et r_c Les valeurs de r_t et r_c peuvent être déterminées de manière à donner une valeur conforme à l'expérience pour le moment magnétique de deutéron (en fixant P_D), la diffusion pour l'état triplet n-p aux basses énergies (en fixant la portée effective ρ_t) et enfin l'effet photoélectrique : capture radiative par un neutron d'un proton. On trouve que P_D ne dépend pas du rapport r_c/r_t , mais que ρ_t limite fortement le choix de r_c et r_t .
Avec les valeurs :

$$\rho_t = 1{,}71 \,.\, 10^{-13}$$
$$P_D = 0.042$$

on a :

$$\frac{M r_c^2 V_c}{\hbar^2} = 1{,}12 \qquad \gamma = 1{,}79$$

r_c et r_t correspondent respectivement à des masses mésoniques de

$$\mu_c = 286 \qquad \mu_t = 280$$

Avec les valeurs

$$\rho_t = 1{,}\ 71$$
$$P_D = 0.038$$

$\frac{1}{\hbar^2} M r_c^2 V_c = 1{,}34$ $\gamma = 0{,}84$

$\mu_c = 326$ $\mu_t = 252$

La valeur de ρ_D ne peut être fixée exactement car des effets relativistes apportent des corrections au 3º chiffre décimal de ρ_D. La valeur de ρ_s n'est pas non plus connue avec une précision suffisante : $2.45 \neq 0{,}5$
On voit que la portée de la force tensorielle n'est pas égale à celle de la force centrale (comme on l'avait naturellement admis dans le travail de Rarita et Schwinger(15)), mais plus grande. Il n'est pas possible d'établir actuellement si cette conclusion dépend du choix fait pour la dépendance spatiale de V_t . Un autre comportement spatial pour V_t pourrait rétablir l'égalité des portées r_c et r_t☆.

La section efficace pour l'effet photoélectrique du deutéron, si on admet l'abseence de force dans l'état P dépend des paramètres qui fixent la force dans le cas du triplet. La valeur de ρ_t déduite de la section efficace expérimentale est en bon accord avec celle déduite de l'énergie de liaison et du scattering, mais la capture radiative est une preuve moins bonne de la validité de la théorie car la section efficace σ fait intervenir ρ_s dont on ne connait pas la valeur précise.

☆ Récemment Lévy a déduit en se basant sur la théorie mésonique symétrique pseudo-scalaire que l'interaction statique entre deux nucléons pouvait être décrite, pour des distances supérieures à une certaine valeur r_0 différente de zéro, par le potentiel

$$V_2 + V_4$$

$$V_2 = -\frac{3f^2}{4\pi}\left(\frac{\mu}{2M}\right)^2\left\{1+\left(1+\frac{3}{\mu r}+\frac{3}{(\mu r)^2}\right)S_{12}\right\}\frac{e^{-\mu r}}{r}$$

$$V_4 = -3\,\frac{f^2}{4\pi}\left(\frac{\mu}{2M}\right)^2\frac{1}{\mu r}\left\{\frac{2}{\pi}\,K_1(2\mu r)+\frac{\mu}{2M}\left(\frac{2}{\pi}\,K_0(\mu r)\right)^2\right\}$$

où μ est la masse du méson π

Ko et K_1 les fonctions de Hankel d'argument imaginaire d'ordre 0 et 1.

Les deux paramètres r_0 et f peuvent être fixés par l'énergie de liaison ε_0 du deutéron et la longueur de diffusion a_s

$\frac{f^2}{4\pi\hbar c} \sim 10$ $r_0 \sim 0{,}6 \cdot 10^{-13}$ cm

Pour des distances $r < r_0$, on évite de parler de potentiel; on introduit une condition aux limites pour la fonction d'onde à la distance $r = r_0$ correspondant à une barrière de potentiel infini. Avec le choix des paramètres r_0 et f indiqué plus haut, on peut déduire les quatre valeurs

$Q = 2{,}08 \ . \ 10^{-27}\ \mathrm{cm}^2$

$\rho_D = 0{,}05$

$\rho_t = 1{,}7$ en bon accord qualitatif

$\rho_s = 2{,}15$ avec les valeurs expérimentales

Remarquons que Q est sensiblement inférieur à la valeur expérimentale et que ρ_s et ρ_D ne sont pas encore bien connus.

Collision de deux nucléons en présence d'un potentiel coulombien.

En présence d'un potentiel coulombien

$$\frac{2\mu}{\hbar^2} V_{coul} = \frac{2\mu}{\hbar^2} \frac{e^2}{r} = \frac{2}{r} k \eta \qquad \eta = \frac{e^2}{\hbar v}$$

(cas du système p-p) la théorie du choc doit être modifiée. On ne peut traiter le potentiel coulombien comme le potentiel nucléaire qui tend vers 0 plus vite que $\frac{1}{r}$ lorsque $r \to \infty$ puisque le potentiel coulombien est du type $\frac{1}{r}$, il n'est plus possible d'écrire pour la solution de l'équation différentielle

$$\left(\frac{d^2}{dr^2} + k^2 - \frac{l(l+1)}{r^2} - \frac{2k\eta}{r} \right) u = V u$$

$$u = \lambda u^r + u^s$$

dans laquelle u^r et u^s sont les solutions régulière et singulière de

$$\left(\frac{d^2}{dr^2} + k^2 - \frac{l(l+1)}{r^2} \right) u = 0$$

La généralisation consiste à prendre les solutions régulières F et singulières G de l'équation

$$\left(\frac{d^2}{dr^2} + k^2 - \frac{l(l+1)}{r^2} - \frac{2k\eta}{r} \right) u = 0$$

F et G étant normalisés d'une manière analogue à u^s et u^r, c'est - à-dire

$$F \underset{r \to \infty}{\simeq} \sin\left(kr - \frac{\ell\pi}{2} - \eta \log 2kr + \sigma_\ell\right)$$

$$G \underset{r \to \infty}{\simeq} \cos\left(kr - \frac{\ell\pi}{2} - \eta \log 2kr + \sigma_\ell\right) \qquad (27)$$

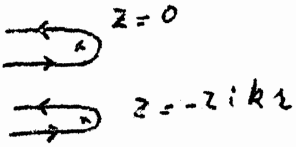
où
$$e^{-2i\sigma_\ell} = \frac{\Gamma(\ell+1-i\eta)}{\Gamma(\ell+1+i\eta)}$$

Les deux fonctions F et G n'ont pas d'expressions analytiques aussi simples que celles de u^2 et u^0. On peut trouver des tables trés soignées de leurs valeurs dans le travail cité en (17). Nous nous bornerons ici à dire qu'il est possible d'écrire les deux fonctions F et G sous une forme intégrale.

Ecrivons :
$$F^{(1)} = (2\ell+1)! \frac{1}{2\pi i} \oint_{c_1} \frac{e^{z}\, z^{i\eta-(\ell+1)}}{(z+2ikr)^{\ell+1+i\eta}}\, dz$$

$$F^{(2)} = (2\ell+1)! \frac{1}{2\pi i} \oint_{c_2} \frac{e^{z}\, z^{i\eta-(\ell+1)}}{(z+2ikr)^{\ell+1+i\eta}}\, dz$$

où c_1 et c_2 sont les chemins d'intégration de la figure



On a :
$$F = (kr)^{\ell+1} e^{ikr} C_\ell \left(F^{(1)} + F^{(2)}\right)$$

$$G = (kr)^{\ell+1} e^{ikr} C_\ell \left(F^{(1)} - F^{(2)}\right)$$

La constante C_ℓ est choisie pour satisfaire la normalisation donnée par (27)

$$C_\ell = e^{-\pi\eta/2}\, e^{-i\sigma_\ell}\, \frac{\Gamma(i\eta+\ell+1)}{(2\ell+1)!}\, 2^{\ell}$$

(facteur dé pénétration en présence d'une barrière coulombienne)

$F^{(1)} + F^{(2)} = {}_1F_1(\ell+1+i\eta, 2\ell+2, -2ikr)$ est la fonction hypergéométrique confluente.

Pour $r \to 0$, F s'annule comme $r^{\ell+1}$

G a une singularité logarithmique.

Bornons nous à écrire le développement pour le cas particulier $\ell = 0$

$$F(r) = c_0 k r (1 + k\eta r + \cdots)$$
$$G(r) = \frac{1}{c_0}\left(1 + 2k\eta r\left[\ln 2k\eta r + 2\gamma - 1 + h(\eta)\right] + \cdots\right)$$

avec $h(\eta) = -\ln\eta + \mathrm{Re}\,\frac{\Gamma'(-i\eta)}{\Gamma(-i\eta)}$.

γ = 0, 5772 est la constante d'Euler - ~~Machoroni~~ Macaroni

L'amplitude de diffusion due à un champ coulombien peut s'écrire

$$f_l(\theta) = \frac{1}{2ik}\sum_0^\infty (2l+1)\left(e^{2i\sigma_l} - 1\right) P_l(\cos\theta) \qquad (29)$$

$e^{2\pi i\sigma_l}$ ayant la signification habituelle $e^{2\pi i\sigma_l} = \frac{\Gamma(l+1+i\eta)}{\Gamma(l+1-i\eta)}$

En présence d'un champ nucléaire V, l'amplitude de diffusion est modifiée par la présence d'un facteur de phase nucléaire $e^{i\delta_l}$

$$f(\theta) = \frac{1}{2ik}\sum_0^\infty (2l+1)\left(e^{2i\sigma_l + 2i\delta_l} - 1\right) P_l(\cos\theta)$$

Il est commode d'écrire

$$f(\theta) = f_{coul}(\theta) + f_{nucl}(\theta)$$

où

$$f_{coul}(\theta) = \frac{1}{2ik}\sum (2l+1)\left(e^{2i\sigma_l} - 1\right) P_l(\cos\theta) = \frac{h}{k}\frac{1}{\sin^2\frac{\theta}{2}}\, e^{-i\eta\ln\sin^2\frac{\theta}{2} + i\sigma_0} \qquad (30)$$

et

$$f_{nucl}(\theta) = \frac{1}{2ik}\sum (2l+1)\, e^{2i\sigma_l}\left(e^{2i\delta_l} - 1\right) P_l(\cos\theta) \qquad (31)$$

Nous observons ici que le potentiel coulombien ayant une portée infinie, la série (29) converge trés lentement et il est indispensable d'arriver à connaitre la somme exacte donnée par (30). Pour le potentiel nucléaire, il suffit de retenir peut de termes de développement (31).

Nous ne voulons pas nous arrêter ici plus longtemps sur les sections efficaces différentielles et totales, mais nous nous limiterons à observer que dans le choc p-p de deux protons, particules identiques de spin $\frac{1}{2}$, en l'abscence de polarisation, il faut écrire pour la section efficace différentielle :

$$\frac{d\sigma}{d\Omega} = \frac{3}{4}\left|f(\theta) - f(\pi-\theta)\right|^2 + \frac{1}{4}\left|f(\theta) + f(\pi-\theta)\right|^2$$

Il faut en effet antisymétriser pour les états triples et symétriser pour l'état simple.

Les formules trouvées plus haut pour $\lambda = \cot g\,\delta$ peuvent être écrite dans le cas coulombien : il suffit de reprendre les mêmes considérations en posant à la place de u^{λ} et v^{λ} les fonctions F et G. En particulier, il est possible de donner un développement en série de puissances de l'énergie (3) et d'y inclure les effets de couplage spin-orbite.

Nous voulons écrire ici la forme (2) de k cotg δ_0 analogue à celle déjà trouvée pour le cas n - p. On a :

$$2\pi\eta \frac{k \cot g\, \delta}{e^{2\pi\eta}-1} + 2k\eta\, h(\eta) = -\frac{1}{a} + k^2 \int_0^\infty \left(\psi_0^{coul}\psi^{coul} - u_0 u\right) dr \quad (32)$$

ψ est la solution en l'abscence du champ nucléaire , l'indice 0 se rapporte à l'énergie $(k^2) = 0$. On obtient le développenent en série de k^2 en remplacant dans l'intégrale ψ et u par leur développement. On peut en particulier définir une portée effective

$$\rho_0 = 2\int_0^\infty (\psi_0^2 - u_0^2)\, dr$$

en complète analogie avec le cas (n-p) (jusque 405 Mev, seul le terme linéaire en l'énergie est important.)

Une discussion détailléed de la comparaison entre la formule théorique (32) et les valeurs expérimentales de σ peut être trouvée dans le travail de Blatt et Jackson(17). Ces auteurs concluent que dans l'état simplet l'interaction n - p et l'interaction p - p, ayant même portée intrinsèque et comportement spatial, sont légèrement différentes entre elles. Dans le cas d'un potentiel de Yukawa, par exemple, la force n-p est 1,6 % plus forte que la force p - p. Schwinger(19) a toutefois montré que l'indépendence de la charge d'interaction nucléaire peut être rétablie lorsque l'on tient compte des forces magnétiques. Entre deux dipoles magnétiques, dans l'état S, l'interaction est proportionnelle au produit $-(\mu_1\vec{\sigma}_1 \cdot \mu_2\vec{\sigma}_2)$. Dans un état singlet $\sigma_1 \cdot \sigma_2 = -3$. L'interaction $3\mu_1\mu_2$ change de signe lorsqu'on passe du système n - p au système p - p On a une légère attraction additive dans le premier cas. La valeur de ρ_1

$$\rho_1 = 2.76 \pm 0{,}07 \cdot 10^{-13}\,\text{cm}$$

obtenue pour un choc p - p avec un potentiel de Yukawa peut donner une valeur meilleure de ρ_1 pour le cas n - p.

9.- Tentatives théoriques d'interprétation des chocs entre deux nucléons aux hautes énergies.

Les données les plus frappantes fournies par les sections efficaces différentielles peuvent se résumer comme suit :(21)

1.) Choc n - p.

La distribution angulaire aprés le choc est isotrope dans le système du centre de gravité jusque des énérgies de 14 Mev. Pour des énergies plus grandes, l'anisotropie est de plus en plus marquée. Les sections efficaces différentielles ont une symétrie remarquable pour la direction à 90º par rapport au faisceaun incident.

2.) Choc p - p.

On a une isotropie remarquable dans l'intervalle angulaire trés étendu allant de 15º jusqu'à 165º. La valeur de la section efficace à 90º est de 2 à 4 fois plus grande que celle correspondant au cas n - p et est indépendante de l'énergie entre 150 Mev et 400 Mev. On n'a pas établi directement si la force nucléaire dans le cas n - p est égale à la force p -p dans le même état quantique.

Il y a eu deux tentatives théoriques pour essayer d'interpréter les données expérimentales en employant un potentiel **statique***. Christian et Noyes (23) obtiennent des résultats satisfaisants pour les sections efficaces en supposant :

pour un état simplet

$$V^s \begin{cases} = 0 & r > 2,6 \cdot 10^{-13}\,\mathrm{cm} \\ = -13,27\,\dfrac{1+P_x}{2}\ \mathrm{Mev} & r < 2,6 \cdot 10^{-13}\,\mathrm{cm} \end{cases}$$

pour un état triplet

$$V^T = -25,5\,\frac{1+P_x}{2}\,\frac{r_1}{r}\,e^{-r/r_1} + (-48,3)\,\frac{1+P_x}{2}\,\frac{r_1}{r}\,e^{-r/r_1}\,S_{12} + (-15,25)\,\frac{1-P_x}{2}\left(\frac{r_2}{r}\right)^2 e^{-r/r_2}\,S_{12}$$

$r_1 = 1.35 \ 10^{-13}$ cm

$r_2 = 1.6 \ 10^{-13}$ cm

Px est l'opératêur d'échange pour les coordonnées spatiales. La section efficace totale théorique est sensiblement plus grande que la section efficace expérimentale. Jastrow(24) prend le potentiel suivant :

$V = \infty$ pour $r < 0,6 \ 10^{-13}$ cm

$$V^s = -375 \frac{1+P_x}{2} e^{-\frac{r-r_0}{r_1}} \text{ Mev} \qquad r_1 = 0,40.10^{-13} \text{ cm}$$

$$V^t = -69\left[\frac{1+P_x}{2} + (0,3 + 0,7) P_x + 1,84 S_{12}\right] e^{-r/r_1} \text{ Mev}$$

On a un bon accord qualitatif avec les données expérimentales, mais on ne peut expliquer qu'à 345 Mev l'isotropie s'étende jusqu'à 15º. Le même potentiel permet d'interpréter le comportement de la section efficace différentielle dans le cas n - p. Une autre tentative due à Case et Pais (25) est basée sur l'hypothèse d'un potentiel linéaire dans la vitesse relative c'est - à - dire du type $(\vec{L}.\vec{S})$. Les calculs des sections efficaces ont été faites par ces auteurs en employant l'approximation de Born dans laquelle on ne peut avoir confiance(19). En tout cas, il est intéressant de suivre qualitativement les effets dus à la présence d'un tel potentiel.

Force $(\vec{L}.\vec{S})$

Pour un potentiel du type $(\vec{L}.\vec{S})\ V(r)$, la dépendance spatiale est encore arbitraire; en analogie avec le cas électromagnétique, on peut choisir pour $V(r)$ une dépendance du type :

$$V(r) = V_0 \frac{1}{x^2} \frac{d}{d(xr)} \left(\frac{e^{-xr}}{x^2}\right) \qquad (34)$$

La singularité à l'origine est utile, permettant d'obtenir à 90º une section efficace différentielle non négligeable, ce que donne un potentiel régulier. Un potentiel du type $(\vec{L}.\vec{S})$ n'agit pas dans les états S et nos considérations aux basses énergies sont donc toujours valables. Comme pour une force tensorielle, le potentiel $(\vec{L}.\vec{S})$ n'agit que dans les états triplets. Mais, à la différence de cette force, il conserve le moment orbital L^2. La théorie exacte du scattering donnant la section efficace différentielle en fonction du déphasage a été développée plus haut. Pour la conservation de L^2, les deux types de solutions ψ^{α} et ψ^{β} obtenues correspondent aux valeurs du paramètre $\varepsilon = 0, \frac{\pi}{2}$

Pour le calcul des déphasages, il est possible d'utiliser les méthodes variationnelles développées pour le champ central. Pour avoir

---

* Nous ne parlons pas de la tentative de Christian et Hart(22); qui ne peut expliquer le choc p - p qu'an introduisant une dépendance de la charge pour le potentiel

une idée qualitative de l'effet d'un couplage $(\vec{L}.\vec{s})$ sur la distribution angulaire en utilisant la première approximation de Born, il faut pouvoir calculer dans le cas n - p l'expression :

$$d\sigma_{LS}^{np} \cong \sum_{mm'} \left| (m| \; \vec{q}.\vec{\nabla}_{\vec{s}} \int V(r)\, e^{i\vec{s}\vec{r}}\, d\vec{r}\, |m') \right|^2$$

$$\vec{q} = [\vec{k} \times \vec{s}] \qquad \vec{r} = \vec{\imath}\, r \qquad x\vec{s} = \vec{k} - \vec{k}'$$

Le signe $\sim$ indique la proportionalité. Si :

$$q_s = \frac{\vec{q}.\vec{s}}{s} = \frac{\vec{s}\,(\vec{k}' \times \vec{k})}{s}$$

et

$$f\,(s) = \int V(r)\, e^{i\vec{s}\vec{r}}\, d\vec{r}$$

$$d\,\sigma_{LS}^{np} \cong (k \times k')^2 \left| \frac{1}{s} \frac{\partial f}{\partial s} \right|^2$$

c'est à - dire

$$d\,\sigma_{LS}^{np} = \left| \frac{k}{x} \right|^4 \sin^2\theta \left| \frac{1}{1 + 4\left(\frac{k}{x}\right)^2 \sin^2\frac{\theta}{2}} \right|^2$$

impliquant une contribution importante pour $\theta = \frac{\pi}{2}$

Il faut faire remarquer ici que dans l'approximation de Born il n'y a pas d'interférence pour la section différentielle entre la diffusio due au potentiel $(\vec{L}.\vec{S})\, V(r)$ et celles dues au champ central et à la force tensorielle. On peut donc ah ajouter les sections différentielles obtenues séparèment pour les trois champs. Case et Pais déclarent qu'il est possible d'ajouter la valeur de $\sigma\left(\frac{\pi}{2}\right)$ pour la diffusion p - p à 4 mb pour 350 Mev 2 mb pour 30 Mev, si l'on choisit dans 34 :

$$\frac{1}{x} = 1{,}1 \;.\; 10^{-13}\ \text{cm} \qquad V_0 = -\ 12\ \text{Mev.}$$

et une dépendance par rapport à l'échange des coordonnées spatiales du type pair. Il ne faut pas prendre ces valeurs à la lettre parce qu'elles se basent sur l'approximation de Born. Le signe - choisi pour V_0 correspond à une force répulsive dans l'état $3D_1$ du deutéron; selon les auteurs, cela ne donnerait lieu à aucune variation sensible de la valeur P_d = 0,04 indispensable pour obtenir des valeurs correctes pour le moment électromagnétiqud du deutéron. Un tel choix du signe semble toutefois donner lieu, dans la distribution angulaire pour les collisions n - p, à une légère asymétrie autour de 90º, dans le sens opposé à celle étudiée expérimentalement. Le choix du signe + pour Vo conduirait à avoir pour un ℓ déterminé, dés états moins liés pour

pour ceux ayant y une grande valeur de j; ceci est contraire - aux difficultés liées à la valeur de P_0 prés - à ce que nous suggérent les modèles centr traux pour les noyaux complexes.(Mayer).

Notre conclusion doit être qu'on ne peut établir avec sureté par les considérations que l'on vient de faire, si l'on peut être assuré de la présence d'un potentiel du type $(\vec{L}.\vec{S})$. Cette question pourra peut-être trouver une réponse dans l'étude expérimentale et théorique des effets de polarisation dus à la présence d'un couplage spin orbite dont nous allons nous occuper maintenant.

10.- Couplage spin orbite et polarisation des nucléons.

Si un faisceau de neutrons ou neutrons non polarisés tombe sur un noyau, et si l'interaction responsale du choc dépend de l'angle entre le moment du spin et le moment orbital, le spin des nucléons aprés le choc est en grande partie orienté dans une direction privilégiée : on a des effets de polarisation. Nous allons pour commencer traiter ici le cas le plus simple du choc d'un neutron contre un noyau de spin 0(26).

A chaque valeur ℓ du moment orbital correspondent deux valeurs $\ell + \frac{1}{2}$ et $\ell - \frac{1}{2}$ du moment total. L'amplitude de diffusion correspondant à l'onde plane incidente pourra être écrite :

$$f_m(\theta,\varphi) = \frac{1}{k}\sum_0^\infty e\,\overset{e^{ikz}\chi_m}{\sqrt{4\pi(2\ell+1)}}\left[e^{i\delta_\ell^+}\sin\delta_\ell^+\,Y^m_{\ell,\ell+\frac{1}{2}}(\theta,\varphi,\sigma) + e^{i\delta_\ell^-}\sin\delta_\ell^-\,Y^m_{\ell,\ell-\frac{1}{2}}(\theta,\varphi,\sigma)\right]$$

$Y^m_{\ell,\ell\pm\frac{1}{2}}$ sont les fonctions propres correspondant au moment total $\ell\pm\frac{1}{2}$ et à la composante m de $(L + S)_z$. Elles sont normées de telle manière que :

$$Y^m_{\ell,\ell+\frac{1}{2}} + Y^m_{\ell,\ell-\frac{1}{2}} = Y^0_\ell(\theta,\varphi)\,\chi_m$$

χ_m est la fonction propre du spin.

$Y^0_\ell(\theta,\varphi)$ est la fonction sphérique d'ordre ℓ (avec m = 0) normée sur la sphère unité.

Appelons $P_\ell^\pm$ les deux opérateurs de projection

$$P_\ell^+ = \frac{1}{2\ell+1}\left[(\ell+1) + (\vec{L}.\vec{\sigma})\right]$$

$$P_\ell^- = \frac{1}{2\ell+1}\left[\ell - (\vec{L}.\vec{\sigma})\right]$$

On peut facilement démontrer que

$$Y^m_{l,l\pm\frac{1}{2}}(\theta,\varphi,\sigma) = P_l^{\pm}\left(Y_l^0(\theta,\varphi)\,\chi_m\right)$$

Il est commode d'effectuer une rotation des axes de référence $x\,y\,z$ (où z est la direction d'incidence du neutron) d'un angle φ de telle manière que pour le nouveau système de référence ξ, η, ζ, [illegible] soit dans le plan de scattering. Dans ce système de référence $(\vec{L}.\vec{\sigma})$ s'écrit simplement

$$(\vec{L}.\,\vec{\sigma}) = i\,\sigma_\zeta \frac{\partial}{\partial\theta}$$

$$L_\xi = i\frac{\partial}{\partial\theta} \quad ; \quad L_\eta = i \cot g\,\theta \frac{\partial}{\partial\varphi}\,, \quad L_\zeta = -i\frac{\partial}{\partial\varphi}$$

L_η et L_ζ peuvent être posés égaux à zéro, puisque $Y_l^0(\theta,\varphi)$ est indépendant de φ. On peut donc écrire $f_m(\theta,\varphi)$ de la manière suivante

$$f_m(\theta,\varphi) = \frac{1}{k}\left(A + B\sigma_\zeta\right)\chi_m$$

avec

$$\left|\begin{array}{l} A = \sum_l \left[(l+1)\,e^{i\delta_l^+}\sin\delta_l^+ + l\,e^{i\delta_l^-}\sin\delta_l^-\right] P_l(\cos\theta) \\ B = i\sum_l \left[e^{i\delta_l^-}\sin\delta_l^- - e^{i\delta_l^+}\sin\delta_l^+\right] P_l^1(\cos\theta \\ P_l(\cos\theta) = \sqrt{\frac{4\pi}{l+1}}\;Y_l^0(\theta,\varphi) \qquad P_l^1(\cos\theta) = -\frac{\partial P_l^0}{\partial\theta} \end{array}\right. \qquad 36$$

Pour un état de spin initial $\quad \chi_{inc} = \sum a_m \chi_m$

$$(\chi_{inc}, \chi_{inc}) = 1$$

on a une amplitude de diffusion

$$f(\theta,\varphi) = \frac{1}{k}\left(A + B\sigma_\zeta\right)\chi_{inc}$$

La section efficace différentielle peut s'écrire :

$$\sigma(\theta,\varphi) = (f,f) = \frac{1}{k^2}\left[(A + B\sigma_\zeta)\chi_{inc}, (A + B\sigma_\zeta)\chi_{inc}\right]$$

dont on déduit

$$k^2\sigma(\theta,\varphi) = |A|^2 + |B|^2 + (\chi_{inc}, \sigma_\zeta\chi_{inc})(AB^* + A^*B)$$

Soient Θ et Φ les angles caractérisant la direction du vecteur polarisation incident $\overline{P} = (\chi_{inc}, \vec{\sigma}\chi_{inc})$

$(37)_\ell$

est la composante du vecteur polarisation perpendiculaire au plan du scattering. La section efficace peut s'écrire :

$$k^2\sigma(\theta,\varphi) = \left(|A|^2+|B|^2\right)\left(1+P_\perp\frac{AB^*+A^*B}{|A|^2+|B|^2}\sin\varphi\right) \qquad (38)$$

si φ est l'azimuth du plan du scattering par rapport au plan défini par $\vec{P}$ et $\vec{k}$ (ce qui revient à poser $\Phi = 0$ et si $P_\perp = P \sin\theta$ est la composante de P perpendiculaire à $\vec{k}$.) On déduit de (38) qu'il n'existe aucune dépendence pour σ de l'azimuth si $P_\perp = 0$, ou s'il n'existe aucun couplage spin orbite car dans ce cas :

$$\delta_\ell^+ = \delta_\ell^- \qquad \text{et } B \equiv 0$$

Il est important de calculer la valeur du vecteur polarisation aprés le choc défini de la manière suivante :

$$\vec{P}^{(s)} = \frac{\left[f(\theta)\,\vec{\sigma}\,f(\theta)\right]}{\left(f(\theta), f(\theta)\right]}$$

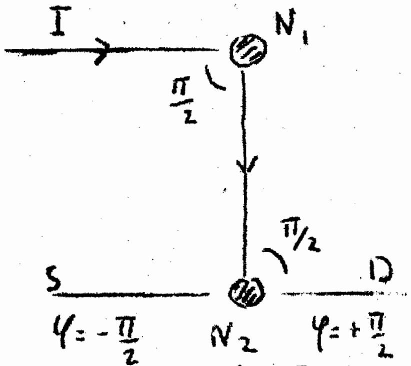
La composante ξ est :

$$P_\xi^{(s)} = \frac{\left(\chi_{inc}, (A^*+B^*\sigma_\xi)\,\sigma_\xi\,(A+B\sigma_\xi)\,\chi_{inc}\right)}{|A|^2+|B|^2+(AB^*+A^*B)\,P_\xi^{(i)}}$$

$P_\xi^{(i)}$ est la composante ξ du vecteur polarisation incident

$$P_\xi^{(s)} = \frac{(AB^*+A^*B)+(|A|^2+|B|^2)\,P_\xi^{(i)}}{|A|^2+|B|^2+(AB^*+A^*B)\,P_\xi^{(i)}}$$

On a donc, si on a un couplage spin orbite, un effet de polarisation même en présence d'un faisceau incident non polarisé $\vec{P}^{(i)} = 0$. On peut facilement démontrer que les composantes η et ξ du vecteur $\vec{P}^{(s)}$ sont proportionnelles à $P_\xi^{(i)}$ et $P_\eta^{(i)}$ et s'annulent donc si $\vec{P}^{(i)}$ est perpendiculaire au plan du scattering. En particulier, en l'abscence de polarisation initiale la polarisation aprés la collision est perpendiculaire au plan du scattering Pour mettre en évidence l'effet de polarisation dù à la présence d'un couplage spin-orbite lorsqu'on ne dispose que d'un faisceau non polarisé, il suffit d'avoir recours à une expérience de double collision.



On peut par exemple avoir une collision en N_1 suivie d'une collision en N_2 dans la direction perpendiculaire et mesurer la section efficace différentielle dans la direction N_2D ou de N_2S φ est l'angle azimuthal autour de la direction N_1N_2 ($\varphi = 0$ est donné par la direction perpendiculaire du plan I N_1 N_2). Les angles $\varphi = \pm\frac{\pi}{2}$ correspondent à N_2D et N_2S. Le rapport des deux sevtions efficaces est donné p par

$$R = \frac{1+Q^2}{1-Q^2}$$

où

$$Q = \left(\frac{AB^* + A^*B}{|A|^2 + |B|^2} \right)_{\theta = \frac{\pi}{2}}$$

L'expérience a été faite en envoyant des protons contre des noyaux d'Hélium (28); elle donne une valeur de $R \neq 0$, ce qui démontre la présence d'un couplage $(\vec{L}.\vec{S})$ Il faut évidemment modifier légèrement la théorie dans le cas des protons, pour tenir compte du champ coulombien : au lieu de considérer dans les formules les seuls déphasages nucléaires δ_ℓ il faut considérer la somme de δ_ℓ et de la phase coulombienne ζ_ℓ correspondante (Cf §8) On trouve expérimentalement que les niveaux $P_{1/2}$ et $P_{3/2}$ sont séparés dans le cas de He^5 par un intervalle d'énergie de l'ordre de 3 Mev ou le premier niveau $P_{3/2}$ (le plus bas)....

Une séparation aussi forte des niveaux $P_{3/2}$ et $P_{1/2}$ ne peut s'expliquer - au moins en se basant sur une évalaution qualitative - par une force tensorielle normale entre les nucléons ou un effet relativistic, du type Thomas et constitue une preuve en faveur du couplage $(\vec{L}.\vec{S})$ nucléaire. Une première évaluation a permis de conclure qu'un potentiel $(\vec{L}.\vec{S})$ $V(r)$ ayant une profondeur du même ordre de grandeur que celui des forces statiques peut donner lieu à des séparations de niveaux de l'ordre de quelques Mev comme on l'a observé expérimentalement. Des expériences de diffusion de neutrons contre He^4 ont confirmé le fait que les niveaux $P_{1/2}$ et $P_{3/2}$ se présentent dans He^5 à peu prés comme dans le cas du Li^5

Effet de polarisation avec des noyaux ayant un spin ≠ 0.

Nous voulons généraliser les considérations du § précédent au cas du choc de nucléons contre des noyaux de spin quelconque. Nous démontrerons aussi que les deux nucléons non polarisés peuvent donner lieu aprés collision à une polarisation toujours perpendiculaire au plan du scattering. La polarisation des nucléons incidents peut être mesurée en partant de la dépendance azimuthal de la section efficace différentielle. Nous devons pouvoir écrire les configurations correspondant à un état du spin S non polarisé ou du spin $\frac{1}{2}$ avec un certain dégré de polarisation. Dans le premier cas, no s observons que

$$\psi = \frac{1}{\sqrt{2S+1}} \sum_{-S}^{+S}{}_{\mu} \varepsilon_\mu \chi_\mu^S \tag{39}$$

où les ε_μ sont des opérateurs de projection; $\varepsilon_\mu^* \varepsilon_\nu = \delta_{\mu\nu}$. χ_μ sont les fonctions propres du spin ayant la composante μ le long de l'axe z. On voit aisément que la valeur moyenne du spin S est $(\psi, \vec{S}\psi) \equiv 0$ pour les ψ donnés par (39). Une configuration correspondant à un spin $\frac{1}{2}$ ayant la valeur $\overline{P}$ du vecteur polarisation peut s'écrire :

$$\varphi = a_- \, \varepsilon_- \, \overline{\chi}_- + a_+ \, \varepsilon_+ \, \overline{\chi}_+ \tag{40}$$

$\varepsilon_{\mp}$. sont les opérateurs de projection.

$\overline{\chi}_-$ et $\overline{\chi}_+$ les fonctions propres du spin le long de l'axe de polarisation.

$$\begin{pmatrix} \overline{\chi}_- \\ \overline{\chi}_+ \end{pmatrix} = \begin{pmatrix} \cos\frac{\theta}{2}\, e^{i\varphi/2} & -\sin\frac{\Theta}{2}\, e^{i\varphi/2} \\ \sin\frac{\Theta}{2}\, e^{i\varphi/2} & \cos\frac{\Theta}{2}\, e^{-i\varphi/2} \end{pmatrix} \begin{pmatrix} \chi_- \\ \chi_+ \end{pmatrix} \tag{41}$$

Θ, Φ sont les angles polaires de l'axe de polarisation. On a :

$$(\overline{\chi}_\pm \vec{S} \overline{\chi}_\pm) = \pm \frac{1}{2} \vec{n}$$

$$\vec{n} = (\sin\Theta\cos\varphi, \sin\Theta\sin\phi, \cos\Theta)$$

$$(\varphi, \vec{S}\varphi) = |a_-|^2 (\overline{\chi}_- \vec{S} \overline{\chi}_-) + |a_+|^2 (\overline{\chi}_+ \vec{S} \overline{\chi}_+)$$

$$= \frac{1}{2} \vec{n} \left(|a_+|^2 - |a_-|^2\right)$$

Pour la fonction (40)soit normée :

$$|a_+|^2 + |a_-|^2 = 1$$

on a :

$$\frac{(\varphi, \vec{S}\varphi)}{(\varphi, \varphi)} = \frac{1}{2} \vec{P} \qquad \vec{P} = \vec{n}\left(|a_+|^2 - |a_-|^2\right) \tag{42}$$

Nous allons considérer la collision faisant passer d'un état initial constitué par un noyau A de spin s_A et un nucléon P de spin $\frac{1}{2}$ à l'état final constitué d'une particule B de spin S_B et d'une particule Q de spin s_Q. L'état initial sera une superposition de configurations

$$Y_L^m(\theta, \varphi)\, \chi_a^{s_A}\, \chi_p^{s_p}$$

caractérisée par le moment orbital L (L + 1) et la projection Z du moment total.

$$M + a + p$$

$\chi_a^{S_A}, \chi_p^{S_P}$ sont les fonctions propres de spin A et P. Nous écrirons plus brièvement

$$Y_L^M(\omega_i)\,\chi_a\,\chi_p \tag{43}$$

L'amplitude de diffusion sera donnée par

$$\sum_{bq} \chi_b\,\chi_q\,(b\,q\,|F(\omega_F)|\,M\,a\,p) \tag{44}$$

χ_b et χ_q représentent les fonctions propres de spin $\chi_b^{S_B}$ et $\chi_q^{S_Q}$ des particules finales. Comme l'interaction du choc doit conserver le moment angulaire total, nous pouvons dire que ;

$$(b\,q\,|\,F(\omega_f)\,|\,M\,a\,p) \tag{45}$$

a une composante J_z du moment total égale à

$$J_z = M + a + p - b - q$$

Nous ne pouvons pas envisager qu'elle ait un moment orbital L (L + 1) car il n'est pas certain que L soit un bon nomre quantique. Nous allons étudier maintenant le comportement de F par rapport aux rotations spatiales.

Nous devons observer que (44) devra se transformer par rapport aux rotations de la même façon que (43). Soit R la rotation faisant passer de la direction $\overline{\omega_f}$ à la direction ω_f.

$$\omega_f = R\,\overline{\omega_f}.$$

On pourra écrire :

$$\begin{aligned}
Y_L^M(\omega_f) &= \sum_{M'} D^L_{MM'}(R)\, Y_L^{M'}(\overline{\omega_f}) \\
\chi_a &= \sum_{a'} D^{S_A}_{aa'}(R)\, \chi_{a'}. \\
\chi_p &= \sum_{p'} D^{S_P}_{pp'}(R)\, \chi_{p'} \\
\chi_b &= \sum_{b'} D^{S_B}_{bb'}(R)\, \chi_{b'} \\
\chi_q &= \sum_{q'} D^{S_A}_{qq'}(R)\, \chi_{q'}
\end{aligned} \tag{46}$$

On a :

$$D_{bb'} D_{qq'} X_{b'} X_{q'} (bq|F(\omega_f)|Map) = D^L_{MM'} D_{aa'} D_{qq'} X_b X_q (bq|F(\bar{\omega}_f)|M'a'p'$$

(sommation sur les indices répétés)

On a encore; étant donné l'unitarité des transformations D

$$(bq|F(\omega_f)|Map) = D^L_{MM'} D_{aa'} D_{pp'} D_{bb'} D_{qq'} \, f \, (b'q'|F(\bar{\omega}_f)|M'a'b') \quad (47)$$

Choisissant comme axe Z celui de la direction incidente, on a M = 0 et en prenant en particulier $\bar{\omega}_f \equiv 0$ la rotation R fera dans ce cas passer des angles polaires (0, 0) aux angles (θ, φ)

$$D^{1/2}_{pp'}(R) = \begin{pmatrix} \cos\frac{\theta}{2}\, e^{-i\frac{\varphi}{2}} & \sin\frac{\theta}{2}\, e^{-i\frac{\varphi}{2}} \\ -\sin\frac{\theta}{2}\, e^{-i\varphi/2} & \cos\frac{\theta}{2}\, e^{i\frac{\varphi}{2}} \end{pmatrix} \quad (48)$$

Les quantités

$$(b'q'|F(0)|M'a'p')$$

ne sont différentes de zéro que si

$$M' + a' + p' - b' - q' = 0$$

est développable en série de fonctions sphériques appartenant à la composante Z

$M' + a' + p' - b' - q' = 0$

On peut donc écrire (47) sous la forme :

$$(bq|F(\omega_f)|0, ap) =$$

$$D^L_{0,\, b'+q'-a'-p'} D_{aa'} D_{pp'} D_{bb'} D_{qq'} (b'q'|F(0)|b'+q'-a'-p', a'p')$$

que nous écrirons en abrégé

$$(bq|F(\omega)|ap) = D^L_{0,\, b'+q'-a'-p'} D_{aa'} D_{pp'} D_{bb'} D_{qq'} (b'q'|F(0)|a'p') \quad (49)$$

Cette relation nous donne le comportement de la partie angulaire de l'amplitude de diffusion par rapport aux rotations d'espace. Nous allons maintenant étudier le comportement de F pour une inversion de signe des nombres quantiques a p b q. Nous allons démontrer que :

$$(-b,-q\,|F(\omega)|\,-a-p) = \pm\, e^{i(\pi-2q)(a+b-p-q)}\,(bq|F(\omega)|ap) \qquad (50)$$

où le signe ± (indépendant de a, p, b, q) est fixé par la parité des configurations. On démontre ainsi en passant du schéma L S_A S_P de la base initiale à celui de J L S

$$Y_L^M(\omega_i)\,X_a X_p = C^{J,M+a+p}_{LM,S\,a+p}\; C^{S\,a+p}_{S_A a,S_P p}\; {}^{LS}Z_J^{M+a+p}$$

$C^{M+m}_{LM,Sm}$ sont les coefficients de Klebsch Gordon donnant la réduction de l'espace produit des deux moments angulaires L, S en les sous espaces irréductibles J. Si à la fonction d'onde ${}^{LS}Z_J^{M+a+p}$ correspond l'amplitude de diffusion

$$(LSJ/\ell sJ)\; {}^{\ell s}Z_J^{M+a+p}$$

où (L S J / ℓ s J) sont des coefficients indépendents des variables angulaires ou de spin, ayant posé

$${}^{\ell s}Z_J^{M+a+p} = C^{J,m+b+q}_{\ell m,J\,b+q}\; C^{J\,b+q}_{S_B b\,S_Q q}\; Y_\ell^m(\omega)\,X_b X_q$$

On peut écrire :

$$(bq|F(\omega)|Map) = (LSJ/\ell sJ)\, C^{J,M+a+p}_{LM,S\,a+p}\; C^{S\,a+p}_{S_A a,S_P p}\; C^{J\,m+b+q}_{\ell m,J\,b+q}\; C^{J,b+q}_{S_B b\,S_A q}\; Y_\ell^m(\omega)$$

On sait que

$$C^{J,M+m}_{L,M,S,m} = (-1)^{L+S-J}\, C^{J,-M-m}_{L,-M,S,-m}$$

Posant M = 0

m = a + p - b - q, on obtient la relation (50). Pour calculer la forme de la section efficace différentielle d'un faisceau polarisé de nucléons tombant sur un noyau de spin quelconque, non polarisé, il faut considérer l'amplitude de diffusion correspondant à l'état initial.

$$\frac{1}{\sqrt{2S_A+1}} \sum C_L\,\varepsilon_a\,\varepsilon_m\,a_m\,\overline{D}_{mp}(\Theta\varphi)\; Y_L^0(\theta,\varphi)\,X_a X_p$$

$\overline{D}_{mp}(\theta,\varphi)$ est la matrice définie en (41) : il faut tenir compte de (39) et (40). On peut écrire :

$$A_f = \frac{1}{\sqrt{2S_A+1}} \sum C_L\,\varepsilon_a\,\varepsilon_m\,a_m\,\overline{D}_{mp}\,X_b X_q\,(bq|F^L(\omega)|ap)$$

La section efficace est donnée par

$$(A_f,A_f) = \sum \frac{C_L^* C_{L'}}{2S_A+1}\,|a_m|^2\,\overline{D}^*_{mp}\,\overline{D}_{mp'}\,(bq|F^L(\omega)|ap)^*(bq|F^{L'}(\omega)|ap')$$

$$(2S_A+1)(Af,f) = \left|\sum c_L (lq|F^L(\omega)|ap)\right|^2 + \left\{\frac{1}{2} P \sin\Theta\, e^{i\varphi} \sum c_L^* c_{L'} (lq|F^L(\omega)|a,+\tfrac{1}{2})^* (l,q|F^{L'}(\omega)\,a,-\tfrac{1}{2}) + \text{complexe conjugué}\right)$$

La quantité

$$Q(\theta) = i\, e^{i\gamma} \sum c_L^* c_{L'} (lq|F^L|a\,\tfrac{1}{2})^* (lq|F^{L'}|a-\tfrac{1}{2})$$

est réelle, comme on peut le voir aisément grâce à (50) et indépendante de φ

Posons $\left|\sum c_L (lq|F^L|ap)\right|^2 = R(\theta)$

quantité réelle et indépendante de

On a :

$$(2S_A+1)(Af,Af) = R(\theta) + P \sin\theta \sin(\varphi-\gamma)\, Q(\theta) \qquad (51)$$

qui est une expression analogue à celle que nous avions trouvé précédemment indépendante du spin du noyau A initial et des particules finales.

Pour évaluer la polarisation dans le cas où la particule q est aussi un nucléon de spin ½ et où le faisceau incident n'est pas polarisé, il faut partir de l'amplitude de diffusion

$$A\,f = \frac{1}{\sqrt{2S_A+1}} \frac{1}{\sqrt{2}} \sum c_L \varepsilon_a \varepsilon_p \chi_b \chi_q (lq|F^L(\omega)|ap)$$

Appliquons l'opérateur spin de la particule q

$$S_z Af = \frac{1}{2\sqrt{2}} \frac{1}{\sqrt{2S_A+1}} \sum c_L \varepsilon_a \varepsilon_p \chi_b \left\{\chi_+ (l+\tfrac{1}{2}|F^L|ap) - \chi_-(l-\tfrac{1}{2}|F^L|ap)\right\}$$

$$(S_x+iS_y)Af = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2S_A+1}} \sum c_L \varepsilon_a \varepsilon_p \chi_b \chi_+ (l,-\tfrac{1}{2}|F^L|ap)$$

Donc $(Af,\ S_z\ Af) = 0$

$$(Af,(S_x+iS_y)Af) = \frac{1}{2}\frac{1}{2S_A+1} \sum c_L^* c_{L'} (l-\tfrac{1}{2}|F^L|ap)^* (l+\tfrac{1}{2}|F^{L'}|ap)$$

La quantité

$$\frac{1}{i} e^{i\gamma} \sum c_L^* c_{L'} (l,-\tfrac{1}{2}|F^L|ap)^* (l,+\tfrac{1}{2}|F^{L'}|ap) = \mathcal{Q}(\theta) \qquad (52)$$

est réelle et indépendante de γ, comme on peut le démontrer en partant de (12); donc

$$(Af,(S_x+iS_y)Af) = \frac{1}{2}\frac{i}{2S_A+1} e^{i\gamma} \mathcal{Q}(\theta) \qquad (53)$$

$$\langle S_x \rangle \simeq -\sin \varphi$$

$$\langle S_y \rangle \simeq \cos \varphi$$

$$\langle S_z \rangle \simeq 0$$

La polarisationn est donc un vecteur perpendiculaire à Z et au vecteur ($\sin \theta \cos \varphi$, $\sin \theta \sin \varphi$ $\cos \theta$) Elle est donc normale au plan du scattering. Nous observons que ces déductions sont indépendantes des spins des deux particules initiales S_A et S_P.

Polarisation de nucléons produits dans des réactions nucléaires.

Les nucléons produits dans les réactions nucléaires provenant du choc de deux noyaux peuvent présenter une polarisation normale (30) (31) au plan défini par la direction d'incidence et la direction d'émergence des nucléons. En particulier, il n'y a pas de polarisation dans la direction d'incidence.

On doit s'attendre, par exemple, à une polarisation pour les neutrons et protons engendrés par la réaction fameuse d + d. La valeur de l la polarisation dépend naturellement de l'énergie des particules incidentes et de l'angle entre les directions d'incidence et d'émergence (voir (30). et (31). Dans le cas de deux d non polarisés, il convient de prendre pour les configurations initiales une base liée au moment orbital L et au spin total S. Le principe de Pauli demande, puisqu'il s'agit de deux particules identiques de spin 1, que la parité de S soit opposée à celle de L. L'amplitude de diffassion peut s'écrire :

$$Af = \sum_{LS} c_L \frac{\varepsilon_m^S}{\sqrt{2S+1}} X_n X_q \left(l q \,|\, F^L(\omega) \,|\, m_1 S \right)$$

ε_m^S étant les opérateurs de projection dans l'espace des spins S. Il suffit de reprendre les développements que nous avons fait au paragraphe précédent pour obtenir la valeur de la polarisation. Il faut évaluer les coefficients :

$$\left(l q \,|\, F^L(0) \,|\, m_1 S \right)$$

qui sont liés auxphasesde la collision pour retourner à la valeur théorique de la polarisation. Cette évaluation est extrêmement difficile : il s'agit d'un problème de quatre corps qu'on ne peut résoudre que d'une manière approchée (32). La mesure de la polarisation des nucléons produit par la réaction d + d peut se déduire de l'étude des dépendances azimuthales de la section efficace différentielle pour la collision contre des noyaux de spin 0 par exemple, aux environs d'un niveau de résonnance (28) (33). On peut le déduire aussi d'effets électromagnétiques (34) (35); le mement symétrique du neutron peut en effet donner lieu dans le champ coulombien V d'un noyau à une interaction du type $(\vec{L}. \vec{\sigma})$.

$$H' = -\mu_n \frac{1}{2} \left(\frac{\hbar}{Mc}\right)^2 \frac{1}{r} \left(\frac{dV}{dr}\right)(\vec{L}. \vec{\sigma})$$

Pour mettre cet effet en évidence, il est utile de faire une mesure azimuthale de la section efficace des neutrons incidents dontre des noyaux lourds (Pb) afin d'avoir un couplage considérable et avoir de grands angles de scattering. Une expérience (35) utilisant les neutrons engendrés par la réaction d + d à permis d'établir ainsi l'existence d'une polarisation.

BIBLIOGRAPHIE :

(1) Eisenbud e Wigner.- Proc. Nat. Acad of America, 27, 281, 1941
(2) Bethe. P. R. 76, 38, 1949
(3) Kushnevink, e Preston, 64, 712, 1951
(4) Blatt e Jackson, P. R., 76, 18, 1949
(5) Salpeter, P. R. 82, 60, 1950
(6) Segré, Conference of Nuclear Physics of Chicago, p. 87, Sept. 1951.
(7) Hulthén, Ark. for Mat. och Phys.
(8) Schwinger Corso de Fisica Nucleare, Harward, 1949
(9) Kato P. R., 80, 475, 1950
(10) Jost e Pais, P.R. 82, 840, 1951
(11) P.R. 86, 399, 1952
(12) Röhrlich e Eisenstein, P.R. 75, 705, 1949
(13) Ramsey et al. P.R. 80, 483, 1950
(14) Feschbach e Schwinger, P.R. 84, 194, 1951
(15) Raritaet Schwinger, P.R. , 19
(16) Lévy (P.R. in press) Cf. Pais Report of Congress in Copenhagen, June 1952, p. 27
(17) Breit et al. Rev. Mod. Phys. 23, 147, 1951
(18) Blatt e Jackson, P.R. 76, 18, 1949 e Rev. Mod. Phys. 22, 77, 1950
(19) Schwinger, P.R. 78, 135, 1950
(20) Breit, Conference Nuc. Phys. at Chicago, p. 103, Sept. 1951
(21) Chamberlain et al. P. R., 83, 923, 1951
(22) Christian e Hart, P.R. 77, 441, 1950
(23) Christian e Noyes, P.R. 79, 85, 1950
(24) Jastrow, P.R. 79, 389, 1950
(25) Kase e Pais, P.R. 80, 203, 1950

(26) Lépore, P.R. 79, 137, 1950

(27) Adair, P.R. 86, 155, 1952

(28) Heunsinkveld e Freier, P.R. 85, 80, 1952

(29) Avery e Blanchard, P.R. 81, 35, 1951

(30) Wolfenstein, P.R. 75, 1664, 1949

(31) R.J. Blin-Stoyle, Proc. Phys. Soc. A 64, 700, 1951

(32) Beiduk et al. P.R. 77, 622, 1950

(33) Verde Nuovo Cimento IX, 376, 1952

(34) Schwinger, P.R. 73, 407, 1948

(35) Longley et al. , P.R. 86, 419, 1952.