L. Rosenfeld

CLASSICAL STATISTICAL MECHANICS

1953

Centro Brasileiro de Pesquisas Físicas

CONTENTS

CHAPTER I

STATISTICAL THERMODYNAMICS

I. Survey of the main ideas in their historical development

CHAPTER II

CLASSICAL STATISTICS

I. General properties of mechanical systems.

I. <u>The ergodic theorems</u>

CHAPTER III

STATISTICS OF CLOSED SYSTEMS

CHAPTER IV

STATISTICS OF OPEN SYSTEMS

STATISTICAL THERMODYNAMICS

I. SURVEY OF THE MAIN IDEAS IN THEIR HISTORICAL DEVELOPMENT.

1. Introduction.

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 radiation 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 order to make a first acquaintance with the ideas and methods of statistical thermodynamics, it is instructive to retrace the history of their birth and growth; we shall accordingly proceed to survey the development of the theory, confining ourselves, however, to those points which afford the most significant illustration of the essential features. Thus we leave out all discussion of the origin of the concept of the atom in Greece, however interesting it would be to dwell upon the ancient atomists' profound speculations. Our starting point will be the revival of atomistic ideas in the time of Newton, within the framework of the new "experimental philosophy". In fact, the weakness of Greek science was its failure to understand the true meaning of the quantitative aspect of our account of physical phenomena. Modern science was born with the recognition of the paramount importance of this aspect; a recognition which could not be expected before the level of technology became such as to make possible accurate experimenting. We shall at first be concerned exclusively with the properties of material bodies: the study of radiation is a much later development, which we shall also discuss in due course.

2. Atomistic interpretations of Boyle's law.

One of the simplest phenomena to which atomic ideas could be applied was the so-called "spring" or elasticity of gases, newly discovered by Boyle. The problem was to account on atomistic lines for the quantitative relation of proportionality between pressure and density which could be deduced from Boyle's measurements. Newton's interpretation was based on a representation of the atoms as centers of force, a view inspired by his theory of gravitation and which remained preponderant until the beginning of the XIXth century. One imagined that atoms acted upon each other at any distance with forces varying with the distance according to some definite law. The atomic model of a gas was thus an essentially static one: the atoms were thought to be kept at definite positions as a result of their mutual interactions and their interaction with the wall of the container. The elasticity could be accounted for by postulating the existence of a central repulsion between the atoms. Newton* showed by a very simple argument that Boyle's law could be obtained by assuming that the repulsive force between any two atoms varies inversely to their distance. To see this, let us consider another sample of the same gas, with the same number of atoms and geometrically similar to the first, the scale of lengths being multiplied by some factor λ. The total forces exerted on two similar areas of the walls of the vessel will then be in the ratio λ^{-1}, and since the areas are in the ratio λ^{2}, the pressures will be in the ratio λ^{-3} Now, this is just the ratio of the inverse volumes or densities, which gives Boyle's law. This ingenious argument dispenses us from performing any detailed analysis of the mechanism of interaction

* Principia (1687)

of the gas atoms with the walls of the vessel. The model itself, however, involves an arbitrary element, viz. the entirely ad hoc assumption of the inverse distance repulsion between the atoms.

Some time later, Daniel Bernoulli* noticed that Boyle's law could be accounted for by a quite different atomic model, which was free from any arbitrary assumption. According to him, the pressure on the walls of the vessel is an essentially kinetic effect: the atoms of the gas are supposed to move in all directions: when they bounce against the walls of the vessel they transfer to them a certain amount of momentum. The total momentum thus received by the unit area of the wall during the unit of time is the pressure: although the process of momentum transfer is discontinuous, the collisions are imagined to take place in so quick succession that the pressure is experienced, at the macroscopic scale of observation, as a continuous force. Newton's similarity argument may be applied to this model as well. The geometrical factors λ^2, λ^3 of area and volume are the same, but now the dynamical factor λ^{-1} is replaced by a kinematical one: in the two samples of the gas, the velocities of the atoms are assumed to be the same, and the amount of momentum transferred to the wall in each collision is therefore the same, but the numbers of collisions per unit time are in the ratio λ^{-1}, since corresponding lengths of path of similar atoms reaching the wall during this time are in the ratio λ. Boyle's law is thus again established without any analysis of the mechanism of collision of the atoms against the walls: but moreover, this kinetic mechansim, as already noted, does not necessitate any specific assumption concerning the forces between the atoms.

3. Atomistic views on the nature of heat.

In spite of its greater simplicity, Bernoulli's kinetic gas model did not at first find favour, because Newton's static model fitted better into the general scheme of explanation of physical phenomena in terms of central forces, which was carried to its utmost consequences during the XVIIIth century. A striking illustration of the way in which "central force atomism" was developed is offered by the

---

* Hydrodynamica (1738)

interpretation of the heat phenomena, whose quantitative analysis is the work of XVIIIth century physicists, foremost among them Joseph Black. Heat was thought of as an imponderable fluid, the "caloric", whose atoms could be bound by specific central forces to those of ponderable matter. Names still in use, like "quantity of heat", "heat capacity" owe their origin to this <u>substantialistic</u> conception.

Yet here also there was a rival view of the nature of heat, based on a kinetic picture of matter. Heat could be conceived as the result, on our macroscopic scale of observation, of the individual motions of the atoms; if one takes the temperature as a measure of the average velocities of atomic motions, transfer of heat from a hot to a cold body would be a transfer of kinetic energy resulting from the interactions of the atoms of the two bodies. This theory of heat transfer is just as simple and accurate as the substantialistic one, according to which the transfer of heat is a flow of caloric from a higher to a lower "level" (indicated by the temperature). In fact, as was pointed out by Lavoisier and Laplace* the two theories, so far as heat transfer goes, are perfectly equivalent.

The superiority of the kinetic theory of heat became apparent, however, in another field, the exploration of which was initiated, during the first half of the XIXth century, by the introduction of steam as a source of power in industry. This focussed the attention on the transformations of heat into mechanical work and <u>vice versa</u>, and eventually led to the formulation of the <u>law of conservation of energy</u> as a universal principle, governing all transformations of physical agencies. Rumford was the first to point out how difficult it is for the caloric conception to account for the production of heat by friction, while the kinetic theory yields an obvious interpretation of this phenomenon. But it was above all Joule** who followed up the consequences of the kinetic theory of heat in this respect and established their truth by rigorous experimental tests: he showed how the law of conservation of mechanical energy, applied to all the

---

* <u>Mémoire sur la chaleur</u> (1780).

** Joule's great work was carried out in the period 1840-43.

motions of the system, including the intrinsic atomic motions manifested as heat, could be expressed as a law of equivalence between quantity of heat and mechanical work.

This momentous discovery was a striking demonstration of the heuristic power of the kinetic theory of atomic constitution, and it established the value of this theory as a principle of unification of our world picture: the laws governing the phenomena of heat were now "reduced" to those of mechanics. At this stage, a word of warning will not be superfluous: it would be wrong to conclude (as Joule did) that the experimental verification of the law of equivalence between heat and work proves the truth of the kinetic theory. As a matter of fact, the law of equivalence is entirely independent of any particular view on the nature of heat; this point was especially stressed by Robert Mayer*, who at the same time as Joule arrived at the same conclusion by an argument just as rigorous as Joule's and exclusively based on empirical relations formulated in terms of macroscopic quantities (such as pressure and specific heat).

Mayer always objected - quite rightly - to the phrase "heat is a mode of motion" in which one often sought to condense the kinetic conception. Indeed, heat is and remains a physical phenomenon qualitatively different from that of motion in bulk, and the establishment of a quantitative equivalence between the two phenomena in no way affects this qualitative difference. The law of equivalence supplies us with a quantitative estimate of transformations of quality in physical systems. Joule and Mayer never understood each other; each thought that the other was hopelessly wrong. In fact, they were both right: they emphasized different and equally important aspects of the equivalence between heat and work.

4. Use of statistical concepts in the kinetic theory of gases.

Let us now come back to the kinetic picture of an ideal gas as a system of atoms moving freely (except for occasional collisions), inside a vessel, against the walls of which they are reflected. We

---

* Mayer's first paper was published in 1842.

have seen that Boyle's law is an immediate property of this model, but we are now in a position to enquire about its further consequences, especially concerning the thermal properties of the gas. This investigation was, historically, initiated by Joule himself and pursued with greater mathematical skill by Clausius and Maxwell. The first property to be considered is, of course, that expressed by Charles' law, i.e. the proportionality of the (invariant) product of pressure and volume with the absolute temperature of the gas. A more detailed analysis of the atomic mechanism by which the pressure is produced will thus give us a precise kinetic interpretation of the concept of (absolute) temperature.

An atom of mass m, on being elastically reflected by the wall with a normal component of velocity v, yields to the wall a momentum $2mv$. Let N be the total number of atoms contained in the volume V and $f(v)dv$ the fraction of these atoms with a normal velocity component in the interval $(v, v+dv)$; the numberof collisions of the type considered on the unit of area of the wall per unit time will be $(N/V)v\,f(v)dv$. The pressure P will thus be obtained by integrating the expression $2(N/V)m\,v^2 f(v)dv$ over all values of v of one sign (indicating an initial motion <u>towards</u> the wall), or half the expression just written over all values of v. Therefore,

$$PV = Nm\left\langle v^2 \right\rangle ,$$

where the notation

$$\left\langle v^2 \right\rangle = \int f(v) v^2 dv$$

representing the <u>average value</u> of the function v^2 for the <u>distribution</u> of v given by $f(v)dv$. If K is the kinetic energy of an atom, we have obviously

$$\langle K \rangle = 3 \cdot \left\langle \frac{1}{2} m v^2 \right\rangle ,$$

and therefore

$$P\,V = \frac{2}{3} N \langle K \rangle .$$

Comparing with Charles' law, we see that the absolute temperature is a measure of the average kinetic energy of the gas. Besides yielding this kinetic interpretation of temperature, the above analysis illustrates the general fact that the definition of macroscopic quantities in terms of corresponding atomic quantities essentially involves a statistical average with respect to a distribution which characterizes the state of the atomic system. This statistical aspect of the atomic theory of matter was first clearly recognized by Maxwell, but it is inherent in the very conception of an interpretation of observable phenomena in terms of entities not accessible to direct observation. Obviously, such an interpretation would be meaningless if it required an accurate knowledge of the dynamical behaviour of the individual atoms.

5. Statistical interpretation of the first law of thermodynamics.

The example just treated is an extremely simple one, and the distribution function occurring in it is of a very special type. It shows us, however, how to proceed in the most general case. Let us consider any system of bodies in any physical state: from the atomic point of view, we have a dynamical system of an enormous number of degrees of freedom, whose state is described by canonical coordinates $q_1, q_2, \ldots, q_f;\ p_1, p_2, \ldots p_f$, and whose behaviour is determined by the Hamiltonian function of these coordinates. The Hamiltonian will further depend on certain macroscopic parameters $a_1, a_2, \ldots$ which define the external conditions to which the system is subjected; the values and variations of these parameters are under the observer's control, in contrast to those of the atomic coordinates q, p. We accordingly write the Hamiltonian as a function $E(q,p;a)$ and it will of course suffice to consider a single external parameter a (e.g. the volume V of the vessel in the case of a gas).

While the Hamiltonian determines the state, or phase, of the system on the atomic scale, its macroscopic state will be characterized by a distribution density $f(q,p;a,\theta)d\omega$, which gives the probability to find the system in the element $d\omega \equiv dq_1 \ldots dq_f\, dp_1 \ldots dp_f$ of phase space around the phase $(q_1 \ldots q_f;\ p_1 \ldots p_f)$. The distribution f depends also on the external parameters $a_1, a_2, \ldots$ and moreover on a parameter Θ (e.g. the temperature) which characterizes the state in question from the thermal point of view. The macroscopic quantity which is called in thermodynamics the <u>internal</u> <u>energy</u> U of the system is then expressed as

$$U = \langle E \rangle = \int E \cdot f \, d\omega \, . \tag{1}$$

According to the first law of thermodynamics, any variation dU of the internal energy can be referred either to external mechanical work $\delta\mathcal{T}$ performed <u>on</u> the system by varying the parameters <u>a</u>, or to a quantity of heat δQ supplied to the system by varying either <u>a</u> or Θ or both; i.e. we have

$$dU = \delta\mathcal{T} + \delta Q, \tag{2}$$

and $\delta\mathcal{T}$ is an expression of the form <u>A</u> da, <u>A</u> being some function of a, Θ. If, for example, <u>a</u> is the volume <u>V</u> of a gas, <u>A</u> represents the external pressure <u>P</u> which balances that of the gas; quite generally, <u>A</u> is called the (generalized) <u>force</u> associated with the parameter <u>a</u>. Now the Hamiltonian, as a function of <u>a</u>, is so defined that the work done <u>by</u> the system when <u>a</u> is varied is expressed by $-\frac{\partial E}{\partial a} da$; therefore, we have immediately

$$A \equiv \left\langle \frac{\partial E}{\partial a} \right\rangle , \qquad \delta\mathcal{T} = \left\langle \frac{\partial E}{\partial a} \right\rangle da \ ,$$

as the kinetic interpretation of the concepts of force and work. From this, the interpretation of the quantity of heat δQ follows readily. In fact, we get from (1)

$$dU = \int dE \cdot f d\omega + \int E \cdot df \cdot d\omega \, ;$$

the first term on the right is just $\delta\mathcal{T}$, therefore by (2) the second is δQ:

$$\delta Q = \int E \cdot df \cdot d\omega \, . \tag{3}$$

This formula is important in that it shows that the concept of heat is very directly connected with the change in the statistical distribution due to the change of external conditions.

6. <u>Statistical interpretations of the second law of thermodynamics and the problem of irreversibility</u>.

The second law of thermodynamics has two aspects. In the first place, it states the existence of a thermal quantity, the entropy, which (in contrast to the quantity of heat) is uniquely defined (except for an arbitrary additive constant) for each state of thermal equilibrium of a physical system. The entropy difference dS between two neighbouring states of equilibrium is defined as

$$dS = \frac{\delta Q}{T}, \tag{4}$$

δQ being the quantity of heat which must be supplied to the system in the initial state at temperature T, in order to bring it into the final state in a quasi-static way, i.e. by controlled variations of the macroscopic variables. The second law then makes the further statement that any spontaneous transformation of a thermally isolated system produces an increase of its entropy, indicating its irreversible character.

The comparison of formulae (4) and (3) shows that the kinetic interpretation of entropy will involve statistical conceptions in a much deeper sense than is implied by just taking statistical averages: a change of entropy is directly connected with a change of the statistical distribution itself. The fundamentally statistical meaning of the entropy concept is still more apparent in the case of irreversible processes. At first sight, we meet here with a paradoxical situation: how can an irreversible evolution be derived from a purel dynamical model, whose behaviour is essentially reversible in time? This is the central problem of statistical thermodynamics and its general solution is by no means easy. It is clear, however, that it can only be found in the properly statistical feature of the atomic picture; in fact, it will turn out that thermodynamic irreversibility is an example of a very general property of stochastic processes, viz. the tendency of such processes to "tend" (in a sense peculiar to statistical theory) towards the situation of greatest probability.

A simple example will illustrate what is meant by this. Consider the spontaneous mixing of two gases: the increase of entropy in this process, when considered from the atomistic point of view, can only appear as a function of the numbers of atoms involved; and it is in fact simply related to the greater probability of the disordered state represented by the mixture, as compared with that of the separated gases. More generally, heat in its various manifestations is the result of a <u>disordered</u> form of motion of the atoms (in contrast to the ordered bulk motions on the macroscopic scale); entropy is a statistical measure of this disorder, and its increase expresses a transition to a state of greater probability.

7. The Maxwell-Boltzmann equation:

The elucidation of the statistical aspect of the second law of thermodynamics was the joint work of Maxwell and Boltzmann; the contributions of both great physicists to this difficult investigation are equally essential, but Boltzmann was perhaps more actively engaged than Maxwell in developing the most general implications of the statistical point of view and in defending it against various objections. The first problem to be tackled was the simpler one of the ideal gas; afterwards, the case of general dynamical systems of atoms was treated and required for its solution the elaboration of new methods of approach. We shall now dwell at some length upon the discussion of the ideal gas according to the original method of Maxwell and Boltzmann; in spite of its limitations, this method is still of considerable interest and brings out the fundamental points very clearly.

Let us consider an ideal gas, i.e. an assembly of atoms assumed to be without interaction except during the short intervals during which they collide in the course of their random motions. The simplicity of this case lies in the circumstance that we can treat each atom separately and define a distribution function for its possible dynamical states independently of the other atoms. In other words, we may in this special case apply statistical considerations to the individual atoms and not only to the total system. In order to cover the case of irreversible transformations, we extend the definition of the distribution function to non-equilibrium states and we accordingly assume that the distribution density $f(\vec{x}, \vec{p}, t)d\omega$ depends on the time explicitly (and not only through the position vector $\vec{x}$ and momentum $\vec{p}$ of the atom); here $d\omega$ is the element of phase space of a single atom, i.e. the product of the volume element dv of ordinary space and the volume element dv_p of momentum space: $d\omega \equiv dv\, dv_p$. It is convenient to normalize the distribution density in such a way that $\int f d\omega = N$, the total number of atoms.

As already stated, we neglect the interactions between the atoms (for their mutual collisions a schematic treatment will prove sufficient); we may, however, without great complication take account of any external force (such as, e.g., gravitation) acting upon each atom separately and upon all of them alike. We shall assume that this force is defined by a potential $V(\vec{x})$ and we may thus write the Hamiltonian of an atom in the form

$$H = K + V, \quad K = \frac{1}{2m} p^2,$$

where, for simplicity, p^2 stands for $|\vec{p}|^2$. The equations of motion (except for the effect of collisions) are then

$$\frac{d\vec{p}}{dt} = - \operatorname{grad} V.$$

Our aim must now be to study the way in which the distribution function varies in time, as a result of the collisions, and to show that it does "tend" towards a definite form which is stationary, i.e. for which $\partial f/\partial t = 0$ for every phase. The definition of the entropy will suggest itself in the course of the argument. Let us consider the $f d\omega$ atoms originally occupying the element $d\omega$ of phase space; after a time interval dt, if no collision had ocurred, they would be found around the phase $\vec{x}' = \vec{x} + (\vec{p}/m)dt$, $\vec{p}' = \vec{p} - \operatorname{grad} V \cdot dt$ within a region whose volume $d\omega'$, to the first order in dt, is equal to $d\omega$ (as is easily seen by forming the Jacobian of the variable transformation). Thus the difference

$$df = f(\vec{x} + \frac{\vec{p}}{m}\, dt,\ \vec{p} - \operatorname{grad} V \cdot dt,\ t + dt) - f(\vec{x}, \vec{p}, t)$$

would vanish. However, as a result of the collisions, a certain number $a \cdot d\omega\, dt$ of atoms will enter into the element $d\omega$ and a number $b \cdot d\omega dt$ will be knocked out of it. Therefore,

$$\frac{df}{dt} = a - b. \qquad (5)$$

This is the fundamental equation of Maxwell and Boltzmann, which governs the variation of the distribution function with time. The left-hand side can be written more explicitly

$$\frac{df}{dt} \equiv \frac{\partial f}{\partial t} + \text{grad } f \cdot \frac{\vec{p}}{m} - \text{grad}_p\, f \cdot \text{grad } V, \qquad (6)$$

where $\text{grad}_p\, f$ denotes the vector with components $\partial f/\partial p_i$. The main task is to estimate the difference a-b on the right.

At this stage, statistical considerations must be introduced. We cannot hope to evaluate a-b exactly, since to do so we would require an exact solution of the equations of motion, <u>including</u> the effect of the collisions. But we are not actually interested in such an exact calculation; all we want to ascertain is the <u>average</u> trend of the distribution. For this purpose, it is sufficient to replace the difference a-b at time <u>t</u> by its average value. The latter is readily expressed in terms of the probability per unit time for the occurrence of a collision of a definite type: this probability is the only quantity representative of the collision processes which will enter into our argument; we shall not require any explicit expression for it, but only use a very general symmetry property which it must have.

In the first place, we must characterize the different types of elastic collisions. We need not make any definite assumption concerning the law of interaction of the atoms during a collision; we assign as space coordinates to the colliding atoms those of their center of mass, and any further description of the mode of collision is contained in the above-mentioned probability per unit time. The collision is then entirely characterized by the values of the momenta of the atoms before and after their interaction; these quantities $\vec{p}_1, \vec{p}_2; \vec{p}_1', \vec{p}_2'$ are only restricted by the conservation laws

$$\vec{p}_1 + \vec{p}_2 = \vec{p}'_1 + \vec{p}'_2\,, \qquad p_1^2 + p_2^2 = p_1'^2 + p_2'^2 \qquad (7)$$

In the barycentric system of reference, the momenta of the two atoms before and after the collision are equal and opposite, and their absolute value is not altered by the collision; the only effect of the latter is to alter the orientation of the common direction of the momentum vectors with respect to the plane defined by the initial momenta in the "rest" system (in which the gas is at rest). To determine the type of collision, we may accordingly choose as our 8 independent parameters the two angles which define this orientation and the two momenta in the initial state.

The probability per unit time for this type of collision will thus be of the form

$$w(\vec{p}_1\vec{p}_2 \rightarrow \vec{p}_1{}' \vec{p}_2{}')\, dv\, dv_{p_1} dv_{p_2}\, d\,\Phi'$$

where $d\,\Phi'$ is the element of solid angle for the direction of the final momenta in the barycentric system. We shall also have to consider the "inverse" collision, in which the initial and final states are interchanged; the corresponding probability per unit time is

$$w(\vec{p}_1{}' \vec{p}_2{}' \rightarrow \vec{p}_1 \vec{p}_2)\, dv\, dv_{p_1'}\, dv_{p_2'}\, d\Phi\ .$$

It is readily verified that, in virtue of the relations (7), the elements of variation of the parameters are the same: $dv_{p_1} dv_{p_2}\, d\Phi' = dv_{p_1'} dv_{p_2'}\, d\Phi'$. We assume, moreover, that the probability density <u>w</u> has also the same value for the inverse collisions; this assumption of <u>detailed balance</u> is only fulfilled for the simplest kinds of collision, but since we are here only concerned with the essential features of the theory, we may disregard the complications which arise when it has to be replaced by a less restrictive assumption.

We have now, for the average values of a and b,

$$a\,d\omega = \int w(\vec{p}_1{}'\vec{p}_2{}' \rightarrow \vec{p}_1\, \vec{p}_2)\, f(\vec{x}, \vec{p}_1{}', t)\, f(\vec{x}, \vec{p}_2{}' t)\, dv\, dv_{p_1'} dv_{p_2'} d\Phi\ ,$$

$$b\,d\omega = \int w(\vec{p}_1\vec{p}_2 \rightarrow \vec{p}_1{}'\, \vec{p}_2{}')\, f(\vec{x}, \vec{p}_1, t)\, f(\vec{x}, \vec{p}_2, t)\, dv\, dv_{p_1} dv_{p_2}\, d\Phi',$$

where we must take $\vec{p}_1 = \vec{p}$, and the integrations must be performed over the momentum parameters so as to leave $\vec{p}_1$ confined to the volume element dv_{p_1}. These formulae are often referred to by the name Boltzmann gave them: the "Stosszahlansatz", i.e. the assumption on the (average) numbers of collisions. For <u>b</u> we get immediately in somewhat simplified notation,

$$b = \int w(\vec{p}_1\vec{p}_2 \rightarrow \vec{p}_1{}' \vec{p}{}'_2)\, f(\vec{p}_1)\, f(\vec{p}_2)\, dv_{p_2} d\Phi', \quad [\vec{p}_1 \equiv \vec{p}] \tag{8}$$

where the integration extends over the whole range of momentum $\vec{p}_2$ and angular coordinates pertaining to the solid angle Φ'. In order to obtain the corresponding expression for a, we have only to make use of the detailed balance assumption, which gives

$$a = \int w(\vec{p}_1\vec{p}_2 \rightarrow \vec{p}_1' \, \vec{p}_2') \, f(\vec{p}_1') \, f(\vec{p}_2') \, dv_{p_2} d\Phi' . \quad [\vec{p}_1 \equiv \vec{p}\,] \tag{9}$$

8. The Boltzmann distribution.

If in the Maxwell-Boltzmann equation (5) a and b are understood to be the averages just calculated, this equation is no longer an exact dynamical relation, but can only claim to represent the average trend of the distribution function: this, however, is just what we are interested in. We shall use it first to derive an expression for a stationary distribution, and then show that no other distribution can be stationary; this second step will at the same time provide us with a complete kinetic interpretation of the second law.

Sufficient conditions of stationarity are obtained in equating to zero the two sides of equation (5) separately (and putting $\partial f/\partial t = 0$). Further, the condition a-b=o will be satisfied, according to (8) and (9), if we impose for any point $\vec{x}$ of space the more stringent condition

$$f(\vec{p}_1) \, f(\vec{p}_2) = f(\vec{p}_1') \, f(\vec{p}_2') \tag{10}$$

for any set of momenta satisfying the conservation equations (7). This will determine the way in which $f(\vec{x},\vec{p})$ depends on $\vec{p}$ for any value of $\vec{x}$. Then, the condition $df/dt = 0$ with $\partial f/\partial t = 0$, i.e. by (6)

$$\text{grad } f . \frac{\vec{p}}{m} - \text{grad}_p f \cdot \text{grad } V = 0 \tag{11}$$

will yield the dependence of f on $\vec{x}$.

Let us apply the condition (10) in the barycentric system, in which $\vec{p}_1 = -\vec{p}_2$, $\vec{p}_1' = -\vec{p}_2'$, the absolute values of $\vec{p}_1$ and $\vec{p}_1'$ or, if we like, the squares p_1^2, $p_1'^2$, being equal. This shows us that the product

$$f(\vec{p}\,) \, f(-\vec{p})$$

is independent of the direction of $\vec{p}$ and is a function of p^2 only. But in this system, we have obviously mirror symmetry in momentum space, i.e. $f(-\vec{p}) = f(\vec{p})$. We may therefore write $f(\vec{p}) = g(p^2)$. Now, let us write the condition (1) for a particular collision in which $p_1^2 = 0$, and therefore $p_2^2 = p_1'^2 + p_2'^2$. Putting for a moment $p_1'^2 \equiv \xi$, $p_2'^2 \equiv \eta$, this gives

$$f(\xi)\, g(\eta) = g(0)\, g(\xi+\eta).$$

This funcional equation is easily solved by taking the logarithmic derivative with respect to ξ:

$$\frac{g'(\xi)}{g(\xi)} = \frac{g'(\xi+\eta)}{g(\xi+\eta)},$$

which shows that this logarithmic derivative is independent of p^2 and reduces to a function of $\vec{x}$. Therefore, $f(\vec{p})$ is of the form

$$f(\vec{x},\vec{p}) = f_o(\vec{x}) \exp\left[- \beta(\vec{x})\, K \right],$$

where the variable p^2 has been replaced by the kinetic energy K.

We must now insert this expression for $f(\vec{x},\vec{p})$ into condition (11), which gives

$$\frac{\vec{p}}{m} \cdot \left[\frac{1}{f_o} \operatorname{grad} f_o + \beta \operatorname{grad} V - K \operatorname{grad} \beta \right] = 0$$

Since this equation must hold for all values of $\vec{p}$, we have separately

$$\operatorname{grad} \beta = 0 \qquad \text{and} \qquad \frac{1}{f_o} \operatorname{grad} f_o + \beta \operatorname{grad} V = 0.$$

Thus, β is a constant and $f_o(\vec{x}) =$ const. $\exp(-\beta V)$. Collecting all the results, we obtain a stationary distribution density of the form

$$f(\vec{x},\vec{p}) = \frac{N}{Z}\, e^{-\beta E(\vec{x},\vec{p};a)} \tag{12}$$

where E is the Hamiltonian of the atom, which also contains the external parameters <u>a</u>, and where we have put

$$Z \equiv \int e^{-\beta E}\, d\omega \tag{13}$$

in order to normalize the distribution. Formula (12) is called the <u>Boltzmann distribution</u>; it is a generalization of that first derived by Maxwell for the case of an ideal gas with no external forces.

Assuming provisionally that Boltzmann's formula represents the unique equilibrium distribution of the ideal gas, we must still ascertain the physical meaning of the parameter β. This will result from a simple relation between β and the average kinetic energy (which, as we know, is proportional to the absolute temperature. For any component p_x of the momentum $\vec{p}$, we have obviously

$$\left\langle \frac{p_x^2}{2m} \right\rangle = N \frac{\int_{\infty}^{\infty} \frac{p_x^2}{2m} \exp(-\beta \frac{p_x^2}{2m}) dp_x}{\int_{\infty}^{\infty} \exp(-\beta \frac{p_x^2}{2m}) dp_x} = -N \frac{\partial}{\partial \beta} \log \int_{\infty}^{\infty} e^{-\beta \frac{p_x^2}{2m}} dp_x =$$

$$= -N \frac{\partial}{\partial \beta} \log \left(\frac{2m}{\beta}\right)^{1/2} \int_{-\infty}^{\infty} e^{-x^2} dx = N \frac{\partial}{\partial \beta} \log \sqrt{\beta} ,$$

or finally

$$\frac{1}{N} \left\langle \frac{p_x^2}{2m} \right\rangle = \frac{1}{2} \beta^{-1} = \frac{1}{2} \theta, \tag{14}$$

if we denote by θ the inverse of β. Therefore,

$$\langle K \rangle = \frac{3}{2} N \theta . \tag{15}$$

Now, remembering the relation, derived above, $PV = \frac{2}{3} \langle K \rangle$, and combining it with Charles' law PV = RT (where R is the "gas constant"), we get

$$\langle K \rangle = \frac{3}{2} RT .$$

The comparison with (15) gives

$$\theta = kT, \tag{16}$$

with k ≡ R/N. We thus see that θ can be regarded as a dynamical measure of the absolute temperature.

However, for such an interpretation to be at all acceptable, it must be universal, i.e. independent of the nature of the gas (which is represented by the atomic mass m). Now, the universal character of the fundamental formula (14) may be verified by extending the preceding argument to a mixture of two ideal gases: on account of the additional condition of the form (10) for collisions between two molecules of different kinds, it will be found that the Boltzmann distributions for the two gases contain the same parameter β. This remarkable property is described as the equipartition of energy. We shall presently discuss its wider implications, but for the moment we are only concerned with its bearing on the ideal gas laws. In the first place, it immediately provides the kinetic interpretation of the universality of both the absolute temperature (embodied in the empirical fact that all gases have the same thermal dilation coefficient) and the gas constant. With regard to the latter, we conclude that it has the same value when referred to the same number of molecules of different gases. Hence, the coefficient <u>k</u> is (as it should be for the consistency of our interpretation of θ) a universal constant, called <u>Boltzmann's constant</u>. The property of R just formulated corresponds to the empirical law, discovered by Gay-Lussac, that R has the same value for all gases when referred to a mass of one mole: we infer from the comparison of the two statements that one mole of any gas contains a definite number <u>A</u> of molecules. This property had been postulated by Avogadro to account for the laws of chemical combination of gasses; the kinetic law of equipartition exhibits its close relationship to Gay-Lussac's law. The value of the universal constant A, called <u>Avogadro's number</u>, can be ascertained in an indirect way by a more detailed kinetic analysis of the non-equilibrium properties of gasses and the deviations from the ideal gaseous state*; we then get the numerical value of Boltzmann's <u>k</u> as R/A, where R now denotes the universal value of the gas constant referred to one mole. We quote the modern figures:

$$\begin{aligned} R &= 8.314 \text{ joules} \cdot \text{deg}^{-1} \cdot \text{mole}^{-1}, \\ A &= 6.023 \times 10^{23} \text{ mole}^{-1}, \\ k &= 1.38 \times 10^{-16} \text{ erg.deg}^{-1} . \end{aligned} \tag{17}$$

---

* The study of non-equilibrium states (eg.diffusion) yields the mean free path, or average length of path between collisions: this is inversely proportional to the product of the number N of molecules per unit volume and the square σ^2 of the "diameter" of a molecule. On the other hand, the total volume $\sim N\sigma^3$ of the molecules was first estimated by Loschmidt from the volume of the liquid into which a unit volume of the gas condenses; a better estimate can be derived from the volume correction of van der Waals' equation of state for non-ideal gases.

9. The H-theorem.

Boltzmann's proof of the uniqueness of the stationary distribution (12), i.e. essentially that the conditions (10) are necessary for the equilibrium, is one of the finest jewels of theoretical physics. The idea is the following: consider a functional $H[f\}$ of the distribution; its variation in time as a result of the collision will be some functional of both f and df/dt,

$$\frac{dH}{dt} = G\left[f, \frac{df}{dt}\right] ,$$

in which df/dt must be replaced by its expression a-b. If we use the average values (8) and (9) of a and b, we shall obtain the average trend of the functional H in the form

$$\frac{dH}{dt} = F\left\{ f(\vec{p}_1) \,;\, f(\vec{p}_1)\, f(\vec{p}_2) - f(\vec{p}_1{}')\, f(\vec{p}_2{}') \right. ,$$

where F again denotes the appropriate functional. If now we can construct H in such a way that F never vanishes unless the conditions (10) are fulfilled, we have obviously achieved the desired uniqueness proof. Such a functional H serves, so to speak, as an indicator of the deviations of the gas from equilibrium. On physical grounds, we would expect that once H would have reached its extreme value, it would remain constant, corresponding to the fact that when the gas has reached its stationary state, it remains in this state indefinitely. In other words, we expect the variation of H in time to be monotonic until a constant value corresponding to the equilibrium is attained.

Boltzmann has shown that the functional which represents essentially the average value of the logarithm of the distribution density exhibits the expected behaviour; this is the content of his famous H-theorem. In precise terms, let us take

$$H = \int f \log (f\omega)\, d\omega ; \qquad (18)$$

in this formula, ω represents a "unit cell" of phase space, of arbitrary volume. It has been introduced in order that the quantity whose logarithm is considered be dimensionless, and its inclusion means that H is only defined up to an arbitrary constant $N\omega$, which is of course of no influence on its time variation. We have then for the average time variation of H on account of the collisions

$$\frac{dH}{dt} \leqslant 0 \,, \tag{19}$$

the equality holding if and only if the stationarity conditions (10) are satisfied.

The proof makes an essential use of the assumption of detailed balance. Taking account of $\int (df/dt)\, d\omega = 0$ we may write

$$\frac{dH}{dt} = \int \frac{df}{dt} \log f \, d\omega = \int (a - b) \log f \, d\omega$$

$$= \int w(\vec{p}_1 \vec{p}_2 \rightarrow \vec{p}'_1 \vec{p}'_2) \left[f(\vec{p}_1{}') \, f(\vec{p}'_2) - f(\vec{p}_1) \, f(\vec{p}_2) \right] \log f(\vec{p}_1) \times$$

$$dv \, dv_{p_1} \, dv_{p_2} \, d\Phi' \,.$$

Since $\vec{p}_1$ and $\vec{p}_2$ play entirely similar parts,

$$\frac{dH}{dt} = \frac{1}{2} \int w \cdot (f_{p_1'} f_{p_2'} - f_{p_1} f_{p_2}) \log (f_{p_1} f_{p_2}) \, dv \, dv_{p_1} \, dv_{p_2} \, d\Phi' \,.$$

Interchanging the initial and final states, and taking account of the detailed balance assumption, we finally get

$$\frac{dH}{dt} = \frac{1}{4} \int w \cdot (f_{p_1'} f_{p_2'} - f_{p_1} f_{p_2}) \left[\log (f_{p_1} f_{p_2}) - \log (f_{p_1'} f_{p_2'}) \right] dv \, dv_{p_1} \, dv_{p_2} \, d\Phi'$$

The truth of the statement (19) is now apparent, owing to the monotonic character of the logarithmic function.

The H - theorem, however, has a much deeper significance than that of a mere formal instrument to establish the uniqueness of the equilibrium distribution. The property of the H functional which it expresses immediately invites comparison with the property of the entropy expressed by the second law of thermodynamics and suggests the existence of a close relationship between the H functional and the entropy. We shall first show that the quantity -H, taken for the stationary distribution, may be identified with the thermodynamical entropy of the system in the corresponding state of equilibrium. Strictly speaking, this stationary value $-H_{stat}$, which is a dimensionless quantity, should be multiplied by the conversion factor <u>k</u> to give the entropy in the usual thermodynamic units; but it is more convenient to use the dynamic measure θ of the temperature and accordingly treat the entropy as a dimensionless quantity.

To see this, let us write down the expression for $-H_{stat}$:

$$-H_{stat} = N \log (Z/N\omega) + \beta \langle E \rangle \tag{20}$$

and compare it with the thermodynamical formula

$$\theta S = U - F, \tag{21}$$

where $U \equiv \langle E \rangle$ and F denotes the <u>free energy.</u> We see that the identification of $-H_{stat}$ with S is equivalent with the identification

$$F = - N \theta \log (Z/N\omega) , \tag{22}$$

which, incidentally, gives a simple physical meaning to the normalization integral Z. Now, to justify equation (22), we have to show that the function F defined by it plays the part of a <u>characteristic function</u> for the independent variables θ, a, i.e. that for quasi-static transformations $dF = - S\, d\theta + A\, da$,
or
$$\frac{\partial F}{\partial \theta} = -S , \qquad \frac{\partial F}{\partial a} = \left\langle \frac{\partial E}{\partial a} \right\rangle .$$

The last formula is an immediate consequence of (22). As to the first, we may eliminate S from it by means of (21) and transform it into the Helmholtz formula

$$U = \frac{\partial}{\partial \beta} (\beta F) . \tag{23}$$

This last equation is readily verified to result from (22), which completes the proof of our kinetic interpretation of thermodynamical quantities.

Passing finally to spontaneous, irreversible processes, we see at once that the H-theorem implies the law of increase of the entropy of an isolated system. The fact that the H-theorem gives the <u>average</u> trend of the quantity H shows that the increase of the entropy is an essentially <u>statistical</u> property, and the irreversibility to which it corresponds has a statistical origin. The successive values actually taken by H in the course of time fluctuate around the averages following the trend of the H-theorem and, strictly speaking, this behaviour is reversible: even when the system has reached its stationary state, the fluctuations of H around its constant average value continue indefinitely, and we cannot exclude the spontaneous reappearance of any state, however far from the equilibrium. However, for systems of a large number of degrees of freedom, large fluctuations will be so seldom as to be practically impossible: for such systems, statistical predictions such as that expressed by the second law of thermodynamics become practical certainties. For smaller systems, on the other hand, as for instance emulsions of submicroscopic droplets, deviations from the thermodynamical behaviour do occur and the fluctuations eventually become so preponderant as to exclude entirely the use of thermodynamic concepts.

In this connexion, it is important to realize that the kinetic view-point is much wider than that of thermodynamics. The functional H is defined for every state of the gas, whether in equilibrium or not, whereas the entropy is only defined for equilibrium states. This naturally leads us to enquire whether the quantity H has any physical meaning also for non-equilibrium states. The answer given by Boltzmann to this question is extremely interesting; in particular, it affords a still more precise elucidation of the statistical character of entropy. Let us consider any distribution of the atoms from the point of view of its <u>probability of occurrence</u>. For the evaluation of this probability, we may assume that the <u>a priori</u> probability of any "cell" of finite volume of phase space is proportional to its volume; we have seen, in fact, that this volume remains unaltered when each of its points moves along the corresponding trajectory, even if the effect of the collisions is taken into account. Let us, therefore, divide the domain Ω of phase space, necessarily finite, within which any atom remains in the course of time, into arbitrary cells of finite volumes ω_i; a distribution is determined by the numbers N_i of atoms in these cells, and its probability is given by

$$P\ (N_i) = \frac{N!}{\prod_i N_i!} \cdot \prod_i \left(\frac{\omega_i}{\Omega}\right)^{N_i} , \qquad (24)$$

where $N = \sum_i N_i$ is the total number of atoms. If we assume that all the N_i's are sufficiently large integers (which obviously excludes only a few very unlikely distributions), we may write approximately (using Stirling's asymptotic formula)

$$\log P(N_i) \simeq N \log \frac{N\omega}{\Omega} - \sum_i N_i \log \frac{N_i\omega}{\omega_i} \quad .$$

In this formula, the sign $\simeq$ is used to denote the asymptotic character of the approximation; the arbitrary "unit cell" ω is again introduced for the sake of dimensional homogeneity. Now, if for the purpose of mathematical computation, we go over to infinitesimal cells $\omega_i \to d\omega$, the ratios N_i/ω_i will reduce to the corresponding values of the distribution density f and the term $\sum_i N_i \log (N_i\omega/\omega_i)$ in the last formula will just reduce to the functional H. We see, therefore, that apart from an arbitrary additive constant, the quantity $-H\{f]$ measures the logarithm of the probability for the occurrence of the distribution f in the course of time. If the distribution is a stationary one, the logarithm of its probability of occurrence represents the entropy of the corresponding equilibrium state.

From our present point of view, however, the equilibrium distribution, instead of being characterized by its stationarity, can be defined regardless of temporal succession in a purely statistical way, viz. as the distribution of maximum probability for the isolated system under consideration. To show this, we must find the extremes of the functional $H[f]$ for all distributions satisfying the conditions

$$\int f d\omega = N \qquad \text{and} \qquad \int E\, f\, d\omega = U \,, \tag{25}$$

which express that the gas is an isolated system. By the usual method of Lagrange multiplicators, we find

$$\int df \cdot \left[\log f + \beta\, E \log\alpha\right] d\omega = 0,$$

i.e.,

$$f = \alpha e^{-\beta E} \quad ,$$

where the parameters α, β are determined in terms of N and U by the accessory conditions (25). These give

$$\alpha = N/Z \quad \text{with } Z = \int e^{-\beta E} d\omega$$

and $\beta = (kT)^{-1}$ as before. It may be observed here that in all formulae involving the equilibrium distribution we have tacitly extended the integration over the momenta to the whole momentum space, although (as pointed out in the preceding argument) the phases actually occupied by an atom in the course of time are necessarily restricted to a finite domain Ω of phase space: owing to the exponential dependence of the distribution density on the kinetic energy, this procedure only involves a quite negligible error. Moreover, it is clear from either derivation that the exponential form of the distribution density is just an <u>asymptotic</u> one, valid only for systems of a large number of atoms.

We have thus established the asymptotic equivalence between the stationary distribution reached by the system in its temporal evolution and a purely statistical distribution over the available domain of phas space, from which every consideration of temporal succession has disappeared. In other words, instead of taking averages over the phases actually occupied in succession by the system, we take them over all the possible phases, each being weighted according to its probability of occurrence in the temporal evolution. The equivalence of the two points of view becomes of fundamental importance for the extension of statistical thermodynamics to more general systems of atoms, such as chrystals, in which the interactions between the atoms play an essential part. It is then no longer possible to study in detail, as we did for gases, the effect of the interactions on the statistical distribution of the dynamical states of the system, but it turns out that the equilibrium distribution can be found just by following up the idea of the equivalence between time averages and statistical averages. This is in fact the very foundation of statistical thermodynamics in its most general form, and we shall have to discuss it thoroughly later on. For the moment, we shall only outline the argument in the still imperfect, though substantially true, form in which it was originally developed by Boltzmann and Maxwell.

10. The ergodic theorem.

Since the Hamiltonian of the general system we are now considering is no longer separable into a sum of terms pertaining to the individual atoms, statistical considerations must not now be applied to the assembly of atoms, but to a (fictitious) assembly of "copies" of our total system, all placed in the same external circumstances. A phase or dynamical state of the system will be represented by a point in a phase space of 2f dimensions, if f is the number of degrees of freedom of the total system. A statistical distribution in this phase space will refer to our fictitious assembly of systems, i.e., it will indicate the probability of finding any one system in a given "cell" of phase space under the given external conditions. In the course of time, the representative point of a system describes a trajectory, whose differential equations are, in canonical form,

$$\dot{q}_i = \frac{\partial E}{\partial p_i} \quad , \quad \dot{p}_i = - \frac{\partial E}{\partial q_i} \qquad (i = 1,2,\ldots,f); \tag{26}$$

the canonical coordinates and momenta are denoted, as usual, by q_i, p_i and the Hamiltonian by $E(q,p)$ (this is now the Hamiltonian of the whole system). The energy is a constant of the motion, and its value E defines a "surface" in phase space

$$E(p,q) = E, \tag{27}$$

in which the trajectory is entirely contained. We assume that all energy surfaces are confined to a finite domain Ω of phase space.

The value of any macroscopic quantity pertaining to the system may be interpreted, from the atomistic point of view, as the average of a corresponding atomic quantity over the interval of time necessary for its measurement. If $A(q,p)$ is the atomic quantity in question, such an average would have the form

$$\frac{1}{T} \int_{t_o}^{t_o+T} A\left[q(t),p(t)\right] dt,$$

where the $q_i(t)$, $p_i(t)$ represent the solution of the equations of motion (26) for the conditions at the initial instant t_o. The time interval T of the measurement may be extremely short on a macroscopic scale and yet cover the passage of the system through a great many atomic configurations: it is then called "physically infinitesimal". Now, we know by experience how a system behaves from the instant at which all connexions with the surroundings have been severed: after a period of adjustment, or "relaxation", it soon reaches a stationary state in which it persists indefinitely. If we except the short period of relaxation, any time average will therefore have the same value over any interval: 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 more amenable to a general analysis. We denote this average over infinite time by

$$\langle A\rangle_t \equiv \lim_{T\to\infty}\frac{1}{T}\int_{t_o}^{t_o+T} A\left[q(t),p(t)\right]dt; \tag{28}$$

it is of course independent of the time t_o at which we suppose the system to have been isolated from its environment.

It is clear that the time average $\langle A\rangle_t$ is equivalent to a statistical average: for it embodies all the atomic configurations actually taken by the system and is of course independent of the order in which they have been successively taken. All we need to know is the relative frequencies of occurrence of the various configurations of the system in the course of its evolution: these will correspond to a definite statistical distribution on the energy surface, which will completely characterize the equilibrium state of the isolated system. Since this distribution is stationary, its density $\rho(q,p)$ per element of area $d\sigma$ will not contain the time explicitly and will thus satisfy the condition

$$\frac{d}{dt}\left\{\rho(q,p)\ d\sigma(q,p)\right\}_{\substack{q\,=\,q(t)\\ p\,=\,p(t)}} = 0, \tag{29}$$

where the q(t), p(t) again represent the trajectory.

In order to determine the equilibrium distribution, some knowledge is needed of the topological character of the trajectory. For the very large systems which interest us, one will expect that owing to the complexity of the interactions between the atoms practically every configuration compatible with the external conditions will occur sooner or later. The simplest assumption we can make to express this intuitive surmise is that the trajectory passes through every point of the energy surface: this assumption was adopted by Boltzmann as the basis of his argument and he called it the ergodic hypothesis. Strictly speaking, one should not introduce any hypothesis in addition to the equations of motion; but in Boltzmann's time the mathematical methods necessary for studying the topological properties of trajectories had not yet been developed and some guess had to be risked in order to make any progress at all. Boltzmann and Maxwell regarded the ergodic hypothesis as a natural generalization of the Stosszahlansatz of the kinetic theory of gases for describing the result of the "disorderly" motions of atomic systems.

If we accept the ergodic hypothesis, it is easy to see that the condition (29) uniquely determines the distribution density $\rho(q,p)$ on the energy surface. To this end, we start from the important property of dynamical systems expressed by Liouville's theorem, viz. that the volume of any domain D of phase space (defined in ordinary euclidean measure as $\int_D d\omega$, with $d\omega = dq_1 \ldots dq_f \, dp_1 \ldots dp_f$) is invariant with respect to the motion:

$$\frac{d}{dt} \int_D d\omega = 0$$

in virtue of the Hamiltonian equations (26). We shall leave the proof of this theorem for our later systematic treatment; let us now pursue the argument. Applying Liouville's theorem to an element of volume of the "shell" of phase space comprised between the energy surfaces E, E+dE, we obtain the invariant element $d\sigma \cdot dn$, where dn is the "thickness" of the shell counted along the normal to the element of area $d\sigma$. This thickness can be calculated in terms of the Hamiltonian from

$$\frac{dE}{dn} = \sqrt{\sum_{i} \left[\left(\frac{\partial E}{\partial q_i}\right)^2 + \left(\frac{\partial E}{\partial p_i}\right)^2 \right]} \equiv v(q,p);$$

the "velocity" $\sqrt{\sum_i (\dot{q}_i^2 + \dot{p}_i^2)}$ at the point (q,p) of the trajectory. Stationary distributions on the energy surface have therefore the general form $F(E, I_1, \ldots, I_{2f-2}) \times v(q,p)^{-1} d\sigma$, where F is a function of a set of (2f-1) time-independent integrals of the motion. Now, according to the ergodic hypothesis, the trajectory is completely characterized by the energy integral, and the function F accordingly reduces to a function of E only; this is just a normalization factor, equal to the inverse of the integral

$$\sum (E) = \int_E \frac{d\sigma}{v} .$$

We thus conclude that the equilibrium state of an isolated system is defined by the <u>ergodic distribution law</u>

$$\rho(q,p)\, d\sigma = \frac{1}{\sum(E)} \cdot \frac{d\sigma}{v(q,p)} ; \qquad (30)$$

the corresponding <u>ergodic average</u> of any quantity A(q,p) is

$$\langle A \rangle_E = \frac{1}{\sum(E)} \int_E A(q,p) \frac{d\sigma}{v(q,p)} . \qquad (31)$$

The equivalence between the time average $\langle A \rangle_t$ and the ergodic average $\langle A \rangle_E$ is now readily verified. We write

$$\langle A \rangle_t = \langle \langle A \rangle_t \rangle_E$$

and, taking account of (29), we interchange the order of the averaging processes on the right-hand side; this gives

$$\langle A \rangle_t = \langle \langle A \rangle_E \rangle_t = \langle A \rangle_E .$$

This equality $\langle A \rangle_t = \langle A \rangle_E$, known as the main <u>ergodic theorem</u> is indeed the true basis of statistical thermodynamics. But the ergodic <u>hypothesis</u>, by which Boltzmann tried to establish it, was later proved to be wrong: a correct formulation of the ergodic theorem and the conditions for its validity requires the ideas of the modern theory of sets and of integration due to Borel and Lebesgue. This is a quite recent development, which we shall discuss in due course. For the moment, there is only one important aspect of the ergodic theorem we want to stress, viz. its relation to the problem of irreversibility.

In the preceding argument, we considered a system whose initial state was described by a well-defined phase; it would have been more realistic to start from some distribution of initial phases in an energy shell, expressing the incompleteness of the determination of the dynamical state of/ the macroscopic system. The ergodic theorem states that any such arbitrary phase distribution will in the course of time spread evenly over the whole energy shell, so as to take the limiting form of the ergodic distribution; in other words, the motion of the phases in the shell can be compared with a mixing process. We here face in its most general form 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 distribution arising from any initial one by letting the size of the elements of phase space decrease indefinitely, we shall find that a given infinitesimal 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 <u>coarse</u> 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 calculation; but the limiting process of letting the cell-size tend to zero must then be 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.

11. The Canonical Distribution.

The next essential question that confronts us is the following. The systems we usually deal with are not isolated, but can freely exchange energy with their surroundings. In particular, in order to assign to any system S a definite temperature, we must keep it in thermal contact with a thermostat Ⓗ, i.e., some system with such a large heat capacity that the exchange of heat between it and the system S does not sensibly alter its own temperature; in other words, the thermostat acts as a practically infinite reservoir of heat at a definite temperature. There is thus a certain opposition, or complementarity, between an isolated system and a system of given temperature: while the former is characterized by a definite value of the energy, in the latter the energy is essentially undetermined; the concepts of (dynamical) energy and temperature are in this sense mutually exclusive. Nevertheless, the statistical law of equilibrium distribution of a system S of given temperature is, of course, implicitly contained in the ergodic theorem: for we may treat the total system S + Ⓗ formed by the given system and the thermostat as an isolated system.

There remains, however, the practical problem of reducing the ergodic distribution of the system S + Ⓗ to a more manageable form: for it involves an integration over the dynamical variables of the thermostat which is irrelevant to the physical situation; the only function of the thermostat is to fix the value of a single macroscopic parameter, viz. the temperature. Boltzmann showed how the elimination of the thermostat variables could be effected in a simple and general way, leaving an expression for the distribution density of the system S in terms of the variables of that system only. This density is now defined for each volume element $d\omega$ of the phase space of the system S, in accordance with the indetermination of the energy of the system. We quote it now without proof, reserving its derivation and closer discussion for later:

$$\begin{cases} \rho(p,q)\, d\omega = \dfrac{1}{Z}\, e^{-\beta E(q,p;a)}\, d\omega, \\ \text{with} \qquad Z = \displaystyle\int e^{-\beta E(q,p;a)}\, d\omega. \end{cases} \tag{32}$$

In this formula, the thermostat is represented by the parameter β, which is the inverse of the absolute temperature θ in dynamical units; the occurrence of other external parameters in the Hamiltonian has bee indicated explicitly. The formal resemblance with Boltzmann's earlier result for the distribution of the single molecules of a gas is not surprising: for the gas itself plays the role of thermostat with respect to each of its molecules.* The law (32), however, is of much wider validity: it applies to any dynamical system of a large number of degrees of freedom. We shall denote it by the name <u>canonical distribution</u>, which was given to it by Gibbs.

The establishment of the canonical distribution completes the statistical analysis of dynamical systems, at any rate insofar as these systems are <u>closed</u>, i.e. have a fixed number of constituents. The application of thermodynamics to chemical processes, which became of great importance with the growth of chemical industry in the last decades of the XIXth century, required the removal of this last restriction. The extension of both thermodynamical and statistical theory to <u>open</u> systems is the great work of Gibbs, who introduced the concept of <u>chemical potential</u> and carried Boltzmann's ideas to their last degree of generality. Yet the development of statistical thermodynamics in this first phase, however brilliant an achievement, was far from being a complete success: a dark cloud hung over it, portending as it were the radical transformation which atomic theory had still to undergo.

The root of the difficulties is the theorem of <u>equipartition of energy</u>, which is a direct consequence of the canonical distribution law, and holds therefore in a much wider sense than the special example we met above when discussing atomic collisions. In fact, the kinetic energy, being a quadratic and homogeneous function of the momentum coordinates, can be written in the form

$$K = \frac{1}{2} \sum_{i=1}^{f} p_i \frac{\partial K}{\partial p_i} \quad ,$$

---

*Hence also the possibility, at first sight paradoxical, of obtaining an equilibrium distribution for a gas at a definite temperature by treating it as an isolated system.

and for any one term of this sum we have

$$\int p_i \frac{\partial K}{\partial p_i} e^{-\beta K} dp_i = -\frac{1}{\beta} \int p_i \frac{\partial}{\partial p_i} (e^{-\beta K})\, dp_i$$

$$= + \frac{1}{\beta} \int e^{-\beta K} dp_i ;$$

the canonical average thus becomes

$$\left\langle \frac{1}{2}\, p_i \frac{\partial K}{\partial p_i} \right\rangle_\theta = \frac{1}{2}\,\theta :$$

this means that each degree of freedom of kinetic energy contributes $\frac{1}{2}\theta$ to the average energy of the system. By a similar calculation, the same conclusion can be extended to those degrees of freedom of potential energy with respect to which the potential function is quadratic and homogeneous: an important example is the case of harmonic oscillations around static equilibrium configurations, such as occur in crystals.

In terms of directly observable quantities, the law of equipartition yields definite predictions regarding <u>specific heats</u>. The translational motions of the molecules in gases, e.g., give rise to a contribution of $\frac{3}{2}$ R per mole to the specific heat of the gas; the oscillatory motions just mentioned in a crystal contribute 3R per mole to the specific heat. It should be stressed that in predictions of this kind atomic theory goes further than macroscopic thermodynamics: the latter establishes certain relations between physical quantities, but the values of some of them must be obtained directly from experience; atomic theory, on the other hand, supplies a dynamical interpretation for any physical quantity, from which the value of this quantity can be computed, at least in principle. The comparison of the above statements concerning specific heats with the empirical facts is therefore of great interest to assess the value of the statistical theory.

Now, such a comparison reveals a very complex state of affairs At first sight, the situation looks rather encouraging. Monoatomic gases do have a specific heat (at constant volume) of very nearly $\frac{3}{2}$ R per mole in a wide range of temperatures and pressures; as to the theoretical figure of 3R per mole for the vibrational specific heat of crystals, it corresponds quite nicely to the law of Dulong and Petit. However, the rapid decrease of the specific heat in both cases when the temperature becomes sufficiently low is in sharp contradiction with the independence of temperature indicated by the theory. Moreover the evidence from optical spectra shows that the atoms and molecules are complex structures with a considerable number of internal degrees of freedom: the corresponding contributions to the specific heat which the theory leads us to expect are strikingly absent up to the highest temperatures at which measurements have been made. We are thus faced, undoubtedly, with quite a serious failure of the equipartition theorem, but it is not a <u>complete</u> failure. It rather looks as if the various degrees of freedom of the atomic systems were gradually "frozen in" as the temperature decreases, so that the law of equipartition would only have asymptotic validity. The temperature at which the "freezing in" appears would seem to depend on some characteristic frequency of the mode of motion considered. The internal atomic motions of high frequency (corresponding to the frequencies of the accompanying light radiations) freeze in at much higher temperatures than the slower oscillatory motions of atoms in molecules and crystals.

CLASSICAL STATISTICS

I. General Properties of Mechanical Systems

1. Motion as isomorphism in phase-space. We consider a mechanical system, the state of which is described by f generalised coordinates q_i and their conjugate momenta p_i. The set (q_i, p_i) $i = 1, \ldots\ldots, f$ 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

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \quad , \quad \dot{p}_i = - \frac{\partial H}{\partial q_i} \qquad (1)$$

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_o is contained in the energy surface $H(P) = E = H(P_o)$ passing through P_o. 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_o to a phase P_t as a transformation of phase space into itself. The transformations $P_o \rightarrow P_t$ 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 analogy 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 $\operatorname{div} \vec{v} = 0$, 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 length 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

$$|\operatorname{grad} H(P)|^{-1} = \left\{ \sum_i \left[\left(\frac{\partial H}{\partial q_i}\right)^2 + \left(\frac{\partial H}{\partial p_i}\right)^2 \right] \right\}^{-1/2}$$

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 than one phase. An example is offered by angular variables: phases in which the values of some angular variables differ by 2π 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 symetrical with respect to permutations of the sets of variables pertaining to different elements. Phase functions possessing this property will be called <u>uniform</u>.

Now, if we consider a set of (2f-1) independent integrals not

---

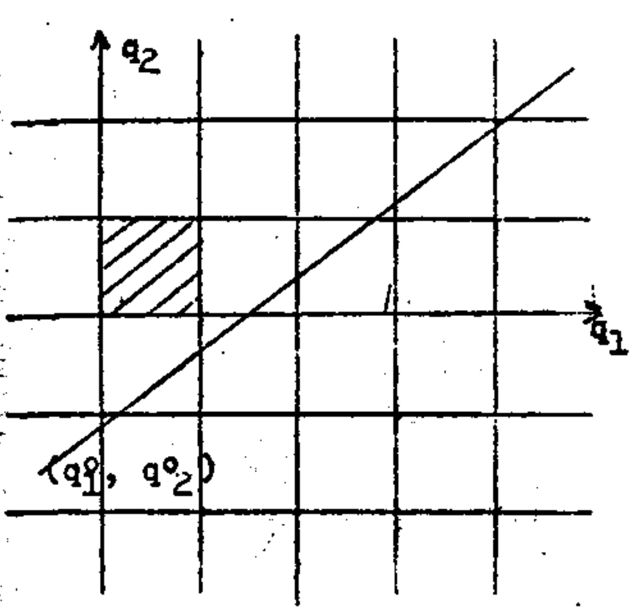
* Strictly speaking, angular variables do not satisfy the finiteness condition enunciated above (paragraph 1). But they nevertheless can be used just on account of the periodicity property of all phase functions of physical significance. See the example discussed in section 4 below.

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 infinitely 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 azimuths 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.

$$q_1 = \omega_1 t + q_1^{0} \quad , \; q_2 = \omega_2 t + q_2^{0}$$



The plane is divided by the lines $q_1 = m \cdot 2\pi$, $q_2 = n \cdot 2\pi$ (m,n integers) into a chequer of squares all containing phases physically equivalent to each other. For all computation of averages of physically significant phase-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

$$q_2\omega_1 - q_1\omega_2 = q_2p_1 - q_1p_2 = M$$

For the initial phase one may write $M_o = \omega_1 q_2^o - \omega_2 q_1^o$, 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 $(q_1^o + m \cdot 2\pi,\ q_2^o + n \cdot 2\pi)$, the value of M differs from M_o 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 exceptional 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 chosen 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 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 fulfill. 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 $I(P) \leq I$, respectively, are both of positive measure*. At

---

* For the proof, see Khinchin, p.30, footnote.

this point it is necessary to remember that there may be 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, paragraph 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 to uniform integrals: these will bring about an essential subdivision of the energy sufrace. But we cannot say anything about the effect of non-uniform integrals; in fact, in the example of paragraph 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 indecomposabilit 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_o. Let us choose some time interval τ and consider the sequence (P) of phases P_o, P_1, P_2 .... successively occupied by the system at times t_o-$t_o+\tau$ $t_o+2\tau$, ... We shall then say that the system returns near P_o if every neighbourhood δ_{P_o} of P_o, 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 δ_p of P, there is at least one interval U_k containing P and contained within δ_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.

The measure of D being the sum of the measures of the D_i's, the theorem will be established if we show that each exceptional set D_i is of measure zero. Let us consider the sequence of transforms $D_i^{(n)}$ of the set D_i at the times $t_0 + n\tau$: no two of these sets $D_i^{(n)}$ can overlap, since otherwise they would contain non-exceptional phases. The measure of the total set $\sum_n D_i^{(n)}$ is therefore the sum of the measures $\Omega(D_i^{(n)})$, which by Liouville's theorem are all equal to $\Omega(D_i)$. But owing to our assumption that the trajectory is contained within a finite domain of phase space, the measure $\Omega(\sum_n D_i^{(n)})$ must be bounded: this can only occur if $\Omega(D_i) = 0$.

7. The return time. A consequence of Poincaré's theorem is that almost any state of a dynamical system will recur to any given approximation in the course of time, and this recurrence will repeat itself an infinity of times. The evolution of the system will thus exhibit a "quasi-periodic" behaviour. The "quasi-period", or average time of recurrence of a given phase to a specified approximation will, however, not only depend on the demanded accuracy of reproduction of the original phase, but above all on the number of degrees of freedom of the system. The sensitiveness of the latter dependence accounts for the fact that the recurrence of a phase of a system of macroscopic dimensions is never observed, whereas the quasi-periodicity expressed by Poincaré's theorem is the rule for smaller systems. A simple example will illustrate this situation and give an idea of the orders of magnitude involved.

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 $\delta \cdot N$ and $(\delta + d\delta)\, N$ is, asymptotically

$$\pi(\delta)\, d\delta \simeq \frac{1}{\sqrt{\pi N}}\, e^{-\frac{(\delta N)^2}{N}}\, d\delta \cdot N \\ \simeq \sqrt{\frac{N}{\pi}}\; e^{-\delta^2 N}\, d\delta ,$$

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 occurrence of an excess in the interval $(\delta, \delta + d\delta)$ in the course of the 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, $L \approx 1$cm, say, one has $\tau \approx 1$ sec and $N \approx 10^{19}$. Even an extremely small relative excess of density, $\delta \approx 10^{-6}$, would not recur spontaneously before times of the fantastic order of magnitude $T \approx e^{10^7}$ sec. But if the volume is of microscopic dimensions, eg. $L \approx 10^{-5}$ 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_o from a certain phase P_o. 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

$$F(P_o;t_o,T) = \frac{1}{T}\int_{t_o}^{t_o+T} f\ (P_t)\ dt, \qquad (1)$$

taken along the trajectory defined by the initial phase P_o; 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_o, t_o. 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

$$F(P_o) = \lim_{T \to \infty} F(P_o;\ t_o,T) \qquad (2)$$

exists for almost all trajectories, and is independent of the initial time t_o. 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 limits of the sequence

$$F_n\ (P_o;\ t_o) \equiv F(P_o;\ t_o,\ n\tau) \qquad (3)$$

as $n \to \infty$. Let P_o be an exceptional phase, i.e. such that $F_n(P_o;\ t_o)$ has no limit; this means that the lower bound $\underline{F}\ (P_o)$ and the upper bound $\overline{F}(P_o)$ are different. We can then choose a pair of members α, β ($\alpha < \beta$) between $\underline{F}(P_o)$ and $\overline{F}(P_o)$, i.e. such that

$$\underline{F}(P_o) < \alpha \qquad \overline{F}(P_o) > \beta \qquad (4)$$

Now, if the set D of exceptional phases were of positive measure, it is easily seen* that one could find a pair (α, β) for which the 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

$$f_k(P_o) = \frac{1}{\tau}\int_{t_k}^{t_{k+1}} f(P_t)\, dt \tag{5}$$

the time average over the interval (t_k, t_{k+1}). By a change of origin of the times, we see that

$$f_k(P_o) = f_o(P_k) , \tag{6}$$

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 $P_o \rightarrow P_k$. 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 Liouville's theorem, $\mu(D_k^{(n)}) = \mu(D_o^{(n)})$, we then have

---

* For details, see Khinchin, p. 19-27

$$\int_{D^{(n)}} f_o\,(P)\ d\,\mu > \beta\,\mu\,(D^{(n)}),$$

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 inequalities 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 $\int_{D^*} f_o(P)\ d\mu < \alpha\,\mu\,(D^*)$,

and these two inequalities with $\mu\,(D^*) > 0$, do indeed contradict the inequality $\alpha < \beta$.

To complete the proof of Birkhooff's theorem, we have to compare the average for an arbitrary time interval T with that for the interval of the sequence n τ nearest to T. We have

$$\left|\ \frac{1}{T}\int_{t_o}^{t_o+T} f(P_t)\ dt - \frac{1}{n\tau}\int_{t_o}^{t_o+T} f(P_t)\ dt\ \right| \to 0,$$

and

$$\left|\ \frac{1}{n\tau}\int_{t_o}^{t_o+T} f(P_t)\ dt - \frac{1}{n\tau}\int_{t_o}^{t_o+n\tau} f(P_t)dt\right| = \frac{1}{n\tau}\left|\int_{t_o+n\tau}^{t_o+T} f(P_t)dt\right| \leqslant \frac{1}{n}\left|\ f_n(P_o)\right|$$

with the notation (5). It is readily verified that

$$\lim_{n\to\infty} \frac{1}{n}\ \left|f_n(P_o)\right| = 0$$

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_o for which $|f_n(P_o)| > n\,\varepsilon$ is of measure zero. Hence, the time average F $(P_o;\ t_o, T)$ has a limit for $T \to \infty$ almost everywhere.

Finally, it must be shwon that this limit is independent of the initial time t_o. We have

$$\lim \frac{1}{T}\int_{t_o}^{t_1+T} = \lim \frac{1}{T+t-t_o}\int_{t_o}^{t_1+T} = \lim\ \frac{1}{T}\int_{t_o}^{t_1+T}$$

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

$$\bar{f}\,|_E \equiv \frac{1}{\mu(\Omega)} \int_\Omega f(P)\, d\mu_E \, . \qquad (7)$$

In the first place, it is clear that the ergodic average of f(P) is the same as that of the/ time average $F(P_o;\, t_o,\, T)$ over any finite time interval T; indeed

$$\frac{1}{\mu(\Omega)}\int_\Omega F(P_o;\, t_o, T)\, d\mu_E = \frac{1}{T}\int_{t_o}^{t_o+T} dt\, \frac{1}{\mu(\Omega)}\int_\Omega f(P_t)\, d\mu_E$$

$$= \frac{1}{T}\int_{t_o}^{t_o+T} dt\, \frac{1}{\mu(\Omega)}\int_\Omega f(P_o)\, d\mu_E$$

in virtue of the invariance of the energy surface Ω ; and this is

$$\frac{1}{T}\int_{t_o}^{t_o+1} \bar{f}|_E\, dt = \bar{f}|_E$$

Since, owing to the constancy of F,

$$F = \frac{1}{\mu(\Omega)}\int_\Omega F\, d\mu_E,$$

it remains to prove that

$$\frac{1}{\mu(\Omega)}\int_\Omega \left[F - F(P_o;\, t_o,\, T)\right] d\mu_E = 0$$

This is readily* seen to be a consequence of the ergodic theorem $F(P_o;\, t_o, T) \to F$ for $T \to \infty$ for almost all P_o.

The last step established in the physical equivalence between time average and ergodic average is an obvious extension of the preceding theorem;

If the energy surface Ω is metrically indecomposable in *the physical sense*, the time average F of any *uniform* phase function f(P) is equal to its ergodic average;

$$F = \bar{f}|_E \qquad (8)$$

---

* See Kinchin, p. 31-32.

The equation (8) is valid for almost 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.

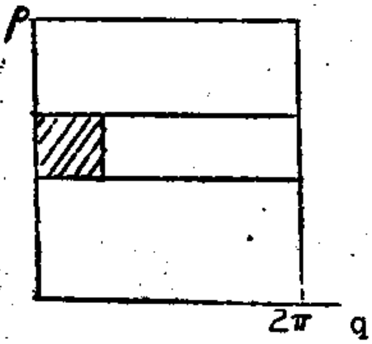
3. Hopf's ergodic theorem.

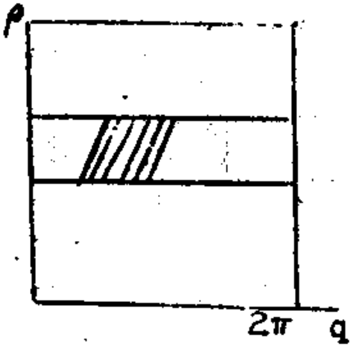
The restriction of the phases to an energy surface is a strong idealisation of an isolated system. It is more realistic to allow a certain margin to the definition of the energy, i.e. to regard as possible all trajectories contained within an energy shell, consisting of all energy surfaces in the energy 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 $\overline{f}|_{(E)}$, we have

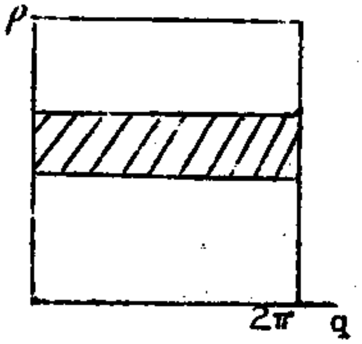
$$\overline{f}|_E = \frac{1}{\mu\,[(E)]} \int_{(E)} f(P)\, d\mu\,, \qquad d\mu = dq_1 \ldots dq_f\; dp_1 \ldots dp_f; \quad (9)$$

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 shell 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 ρ, which is a constant of the motion. We may assume that the value of ρ fixes the energy: in the "phase space" (p,q,) the "energy surface" is then the line p = const., or rather the segment $0 \leq Q \leq 2$ 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 districtuion 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 neccessary to give a precise mathematical definition of the mixing process; moreover, 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 $t_o = 0$ 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

$$(f,g) = \int f(P) \cdot g(P)\, d\mu,$$

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

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T |(f, \mathcal{T}_t g) - (f, \overline{g})|^2 \, dt = 0 \qquad (10)$$

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 existance 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 * Hopf's 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

---

* We shall not give the proof of Hopf's theorem, since it does not involve any essentially new feature.

theorem completes the foundation of the concept of ergodic distribution, on which the whole structure of statistical thermodynamics rests.

4. The ergodic districation.

Let us now discuss from a more physical point of view the main features of the ergodic districation 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 irreversibl situation; but we have also in hand the means of solving this paradox. If we analyse at any instant the distribution arising from any initial one by letting the size of the elements of phase space decrease indefinitely, we shall find that a given infinitesimal 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 coarse 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 calculation; but the limiting process of letting the cell-size tend to zero must then be 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 practically reached its asymptotic uniformity. At first sight, it would seem 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 occurrence 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 interpretation 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 <u>identical elements</u>: 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 intricately imbricated through each other over each energy shell. The measure of a constellation, according to the ergodic theorems, gives the relative probability of occurrence of the state represented by the corresponding generic phase in the course of time. States very near to equilibrium will thus occupy 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 probable ones and when it will have reached the equilibrium constellation, it will stay in it most of the time, although it will continually cross less probable 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*, "but I see no point in applying it 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 recognized 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.

In fact, the objections raised against Boltzmann's position

---

* This utterance was reported in conversation by Ehrenfest who had been a pupil of Boltzmann's.

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 visualized by following the path of the system among the entanglements of the various constellations. But they are beside the point: it is the coarse distribution which has an irreversible evolution, and we have seen that such an irreversible trend is perfectly compatible with the reversibility of the mechansim by the "mixing" of the coarse distribution is brought about.

5. Systems with weak interactions.

We may follow up the last argument more quantitatively. 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 $\omega_1^{N_1}\omega_2^{N_2}\ldots$, 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 $N_1, N_2, \ldots\ldots$, 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 that the equilibrium distribution of the system 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 unaltered in the course of time.

But the magnitude of the interaction is irrelevant and in suitable circumstances can be regarded 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:

$$E = \sum_i N_i E_i \tag{12}$$

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

$$N = \sum_i N_i \tag{13}$$

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

$$\bar{N_i}|_E \simeq \frac{N}{Z} \omega_i e^{-\beta E_i} \tag{14}$$

where

$$Z = \sum_i \omega_i e^{-\beta E_i} \tag{15}$$

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

$$Z = \int e^{-\beta H(p,q)} d\mu, \tag{16}$$

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 $\Omega(\bar{N}_1|_E, \bar{N}_2|_E, \dots)$ of the equilibrium constellation, and compare it with the measure of the energy shell in Γ space in which it is contained. We have

$$\log \Omega(\bar{N_1}|_E, \bar{N_2}|_E, \ldots) \simeq N \log Z + \beta E, \qquad (17)$$

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 $d\mu = d\mu_E \cdot dE$. For the invariant measure of the energy surface E we write

$$\mu(E) = e^{\sigma(E)} \quad ;$$

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,

$$\sigma(E) - \beta E = \sigma(E_m) - \beta E_m + \frac{1}{2} \left(\frac{d^2\sigma}{dE^2}\right)_m (E-E_m)^2 \quad ;$$

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 average value 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 $\Omega(\bar{N_1}|_E, \bar{N_2}|_E, \ldots)$ takes the 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 (14), 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 $A \cdot da$. In the atomistic treatment, we may take account of these macroscopic variables (a,A) by introducing into the Hamiltonian of our system 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

$$(\partial/\partial a)\, H\,(p,q;\, a)\ .$$

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 <u>independent</u> 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 <u>characteristic function</u>, from which all other quantities are obtained by differentiation and algebraic combination. If θ denotes the absolute temperature, one has

$$d\,S = \frac{1}{\theta}\,dE - \frac{A}{\theta}\,da, \tag{18}$$

so that

$$\frac{1}{\theta} = \left(\frac{\partial S}{\partial E}\right)_a , \quad - \frac{A}{\theta} = \left(\frac{\partial S}{\partial a}\right)_E . \tag{19}$$

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 smaller than E. We may then take

$$S = \log \mu(\Omega) \tag{20}$$

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 $\mathcal{E}(\Omega)$ very simply as the "step" function $\mathcal{E}(x)$ of the argument $x = E - H(p,q)$:

$$\mathcal{E}(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

The derivative of $\mathcal{E}(x)$ is the distribution $\delta(x)$. This allows us to write

$$\mu(\Omega) = \int \mathcal{E}\left[E - H(p,q;a) \right] d\mu$$

the integration extending to the whole phase space, We further get

$$\frac{\partial \mu(\Omega)}{\partial a} = - \int \frac{\partial H}{\partial a} \delta\left[E - H(p,q;a) \right] d\mu$$

$$= - \int dE' \, \delta(E-E') \int_{E'} \frac{\partial H}{\partial a} d\mu_{E'} = - \int \frac{\partial H}{\partial a} d\mu_{E} = - \frac{\partial \overline{H}}{\partial a}\Big|_E \mu(E)$$

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 $$\frac{\mu(E)}{\mu(\Omega)} = \frac{\partial}{\partial E}\left[\log \mu(\Omega) \right] = \frac{1}{\theta}$$

whence finally $$\frac{\partial}{\partial a}\left[\log \mu(\Omega) \right] = - \frac{A}{\theta}$$

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.

$$S = \log \Omega \left(\overline{N_1}\big|_E, \overline{N_2}\big|_E, \ldots \right) , \qquad (21)$$

whose explicit form is given by (17). The "force" A is now expressed as an average over the most probable distribution (14)

$$A = \frac{N}{Z} \int \frac{\partial H}{\partial a} d^{-\beta H} d\mu ,$$

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 $\Omega(\overline{N_1}|_E, \overline{N_2}|_E \ldots)$ is practically 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 $\sqrt{2mE}$, 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 a logarithm of the measure of a domain of phase space, i.e. of the relative probability of occurrence, 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 discrimination 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 necessarily be larger than the initial one; but the behaviour of the statistical function $\log \cdot \Omega(N_1, N_2 \ldots.)$ 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 meaningless. 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 of 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 $\mathcal{S}$ 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 $\mathcal{S}$ remains constant, i.e. that $\mathcal{S}$ is isolated. The problem is thus to derive from the statistical distribution of the isolated system $\mathcal{S}$ that of a small part S of $\mathcal{S}$.

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 $\mathcal{S}$ 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 $\Omega_S, \Omega_T, \Omega$ 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 immediately as ergodic averages in the total $\mathcal{S}$-space; we shall then seek to transform these expressions so as to eliminate the irrelevant phase coordinates P_T and the thermostat will be represented in the final result by only one parameter, which plays the part of the temperature.

For the transformation we have in view, it will be convenient to write the ergodic average in the form

$$\bar{f}|_E = \frac{1}{\mu(E)} \frac{d}{dE} \int_{\Omega} f(P)\, d\mu \qquad (1)$$

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 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)

$$f(\overline{P_S})|_E = \int f(P_S) \frac{\mu_T(E-H_S)}{\mu(E)}\, d\mu_S \qquad (2)$$

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 $\bar{f}|_{E_S}$ 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

$$\mu_T(E_T) = \int dE_{T_1} \int dE_{T_2} \cdots \int dE_{T_{m-1}} \mu_{T_1}(E_{T_1}) \mu_{T_2}(E_{T_2}) \cdot \mu_{T_n}(E_1 - E_{T_1} - E_{T_2} \cdots - E_{T_{n-1}})$$

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 probability.

Consider n stochastic variables $x_1, x_2, \ldots x_n$ with independent distributions $u_1(x_1)dx_1$, $u_2(x_2)dx_2, \ldots$ The distribution of the sum $x = x_1 + x_2 + x_n$ obeys the law of composition:

$$u(x)dx = \int dx_1 \int dx_2 \cdots \int dx_{n-1} \, u_1(x_1) \, u_2(x_2) \ldots u_n(x - x_1 - x_2 \ldots x_{n-1}) \qquad (4)$$

Now, provided that the distribution laws $u_i(x_i)$ satisfy certain cinditions 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 $n \to \infty$)

$$u(x)dx \simeq \frac{1}{\sqrt{2\pi B}} \, e^{-\frac{(x-\bar{x})^2}{2B}} \qquad (5)$$

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

$$b_i = \int (x_i - \alpha_i)^2 \, u_i(x_i) \, dx_i$$

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

$$u_i(E_{T_i}) = \frac{1}{Z_i(\beta)} \, e^{-\beta E_{T_i}} \mu_{T_i}(E_{T_i}) \qquad (6)$$

where

$$Z_i(\beta) = \int e^{-\beta E_{T_i}} \mu_{T_i}(E_{T_i}) \, dE_{T_i} \qquad (7)$$

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

$$\bar{E}_T = - \frac{d \log Z}{d\beta} \qquad (8)$$

and

$$B_T = \frac{d^2 \log Z}{d\beta^2} \qquad (9)$$

It is easily seen** that equation (8) uniquely determines β in terms of $\bar{E}_t$: 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 $u(E_T)$ its asymptotic form (6) we get for

---

* Voir Khinchin, p. 86

** Voir Khinchin, pg. 77

$\mu_T(E_T)$ the asymptotic expression

$$\mu_T(E_T) \simeq \frac{Z_T(\beta)}{\sqrt{2\pi B_T}} \; e^{\beta E_T - \frac{(E_T-\bar{E}_T)^2}{2B_T}} \tag{10}$$

In applying this formula to the argument $E_T = E - E_S$, we note that $(E_T - \bar{E}_T)^2 = (E_S - \bar{E}_S)^2$. Now, excepting the unlikely cases when E_S happens to be widely different from its average $\bar{E}_S$, the quadratic fluctuations of the system S will be of a much smaller order of magnitude than those of the thermostat whose average is given by B_T: we shall accordingly neglect them altogether and write

$$\mu_T(E - E_S) \quad \frac{Z_T(\beta)}{\sqrt{2\pi B_T}} \, e^{\beta(E-E_S)}$$

For the factor $\mu(E)$ 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

$$\frac{\mu_T(E-E_S)}{\mu(E)} \simeq \frac{1}{Z_S(\beta)} \; e^{-\beta E_S} \tag{11}$$

in which the thermostat only appears through the parameter β.

In its most general aspect, the property expressed by formula (11) may be stated as follows: any small* part of

---

* 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 which we have studied.

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 canonical distribution law, was extensively studied by Gibbs; the insufficient emphasis put by Gibbs on Boltzmann's theorem is no doubt responsible for the widespread 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.

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 defined. 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:

$$\bar{E}_S = \frac{d \log Z_S}{d\beta}$$

If, in particular, the system S is an ideal monoatomic gas, we have

$$\mu(E_S) \sim E_S^{\frac{3N}{2}-1}$$

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

$$\bar{E}_S = \frac{3N}{2}\,\frac{1}{\beta} \tag{12}$$

The thermodynamical value of the energy of such a gas is $\frac{3}{2}$ R T per mole, where R is the gas constant and T the absolute temperature in the Kelvin scale. This may be written $\frac{3}{2}$ N θ, if N is the number of atoms per mole and θ = kT, with k = R/N, is the measure of the absolute temperature in an energy scale. The conversion with the statistical formula fixes the relation of the thermostat parameter β with the absolute temperature:

$$\beta = \frac{1}{\theta} \tag{13}$$

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.

$$H' = \frac{1}{2} \sum_{i=1}^{k} p_i \frac{\partial H'}{\partial p_i}$$

we may perform the calculation of the canonical average H' explicitly. In fact, putting H = H', we may write:

$$\bar{H}' = \frac{1}{2} \int d\mu'' \, e^{-\beta H''} \int H' \, e^{-\beta H'} \, d\mu' \, ,$$

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

$$\int p \frac{\partial H'}{\partial p} e^{-\beta H'} dp = -\frac{1}{\beta} \int p \frac{\partial}{\partial p} (e^{-\beta H'}) \, dp = \frac{1}{\beta} \int e^{-\beta H'} \, dp$$

whence

$$\bar{H}' = \frac{k}{2} \frac{1}{\beta} = k \frac{1}{2} \theta \, . \qquad (14)$$

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 homogeneous 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:

$$W(E_s) = \mu_s(E_s) \frac{\mu_T(E-E_s)}{\mu(E)} = \frac{Z_S(\beta)}{\sqrt{2\pi B_s}} e^{\beta E_s - \frac{(E_s-\bar{E}_s)^2}{2B_s}} \cdot Z_T(\beta) \, e^{\beta(E-E_s)} - $$

$$\exp\left\{-\frac{(E_s \bar{E}_s)^2}{2B_T}\right\} \cdot \frac{\sqrt{2\pi B}}{Z(\beta)} e^{-\beta E} = \frac{1}{\sqrt{2\pi B_S^*}} e^{-\frac{(E_s-\bar{E}_s)^2}{2B_S^*}}$$

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

$$B_S = \overline{(E_S - \bar{E}_S)^2} = \overline{E_S^2} - (\bar{E}_S)^2$$

$$\bar{E}_S = \frac{-d \log Z}{d\beta}, \quad \overline{E_S^2} = + \frac{1}{Z}\frac{d^2 Z}{d\beta^2}, \quad \frac{d\bar{E}_S}{d\beta} = - \frac{1}{Z}\frac{d^2 Z}{d\beta^2} + \frac{1}{Z^2}\left(\frac{dZ}{d\beta}\right)^2$$

i.e.,

$$B_S = - \frac{d\bar{E}_S}{d\beta} = \frac{d\bar{E}_S}{d\theta}\,\theta^2 \tag{15}$$

This formula exhibits a fundamental feature common to all fluctuation phenomena: while the product $\theta \frac{d\bar{E}_S}{d\theta}$ 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 developed for isolated systems. We shall look for an atomistic interpretation of the characteristic function corresponding to the choice of the temperature and the external parameters as independent variables: this is the free energy of the system, defined as *

$$F = \bar{E} - \theta S \tag{16}$$

---

* We may now drop the index

In fact, we find

$$dF = -S\, d\,\theta \pm A\, da, \tag{17}$$

i.e.

$$S = -\left(\frac{\partial F}{\partial \theta}\right)_a, \quad A = \left(\frac{\partial F}{\partial a}\right)_\theta \tag{18}$$

From (16) and (17), we derive, by making use of the general transformation formula:

$$\theta \frac{\partial F}{\partial \theta} = -\beta \frac{\partial F}{\partial \beta},$$

the *Helmholtz relation* 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 $(\overline{\partial H/\partial a})$

$$A = -\frac{1}{\beta} \frac{\partial \log Z}{\partial a};$$

this formula, together with the previous one

$$\overline{E} = -\frac{\partial \log Z}{\partial \beta}$$

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 we have obtained for the systems of identical elements with weak interactions: for, as already noted (see footnote at the end of paragraph 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

[*] The system S then represents a single element, and the expression (21) accordingly gives the entropy *per element*.

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 $\bar{E}$, viz.

$$\log \mu(\bar{E}) = \log \left\{ \frac{Z(\beta)}{\sqrt{2\pi B}} e^{\beta \bar{E}} \right\} = S - \log (\sqrt{2\pi B})$$

S is log measure of energy shell as before.

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

$$\delta Q = d\bar{E} - A da$$

whence

$$\begin{aligned} \beta \delta Q &= d(\beta \bar{E}) - \bar{E}\, d\beta - \beta A da \\ &= d(\beta \bar{E}) + \frac{\partial \log Z}{\partial \beta}\, d\beta + \frac{\partial \log Z}{\partial a}\, da \\ &= d(\beta \bar{E}) + \log Z) = dS \end{aligned}$$

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

$$\bar{f} = \frac{1}{Z} \int f(P)\, e^{-\beta H}\, d\mu$$

may be written as

$$\bar{f} = f(P)\, e^{\eta(P)}\, d\mu \tag{22}$$

with

$$\eta(P;\theta, a) = \frac{1}{\theta} \left[F(\theta, a) - H(P;a) \right] \tag{23}$$

The function is called by Gibbs the "probability exponent", since

$$n(P;\theta, a) = e^{\eta} \tag{24}$$

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:

$$S = -\bar{\eta};$$

this may be written, with the notation (24)

$$S = -\int n \log n \, d\mu \tag{25}$$

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 still have 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 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 sufficient time to elapse 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 slow 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 $e^{\eta(\theta_o, a_o;\ p\ q)}$; the new one is accordingly

$$\overline{e^{\eta(\theta_o,\ a_o;\ p\ q}}\Big|_{E(a)}$$

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 expand the distribution density $e^{\eta(\theta_o,\ a_o;\ p\ q)}$ around the new value a of the parameter:

$$e^{\eta(\theta_o,\ a_o;\ p\ q)} = e^{\eta(\theta_o,\ a;\ pq)}\Big[1- \frac{da}{\theta_o}\ (A - \frac{\partial H}{\partial A}\ )$$

and accordingly, since the factor $e^{\eta(\theta_o,\ a;\ pq)}$ is constant on the new energy surface E(a)

$$e^{\overline{\eta(\theta_0,a_0;\ pq)}}\Big|_{E(a)} = e^{\eta(\theta_0,\ a;\ pq)}\left[1 - \frac{da}{\theta_0}(A - \overline{A}|_{E(a)})\right]$$

Now, for systems with a large number of degrees of freedom , the canonical distribution has a sharp maximum for the average $\overline{E}$ and we may use the expansions

$$\overline{A}|_E = \overline{A}|_{\overline{E}} + \frac{\partial \overline{A}|_{\overline{E}}}{\partial \overline{E}}(E-\overline{E}) + \frac{1}{2}\frac{\partial^2 \overline{A}|_{\overline{E}}}{\partial \overline{E}^2}(E-\overline{E})^2$$

$$A = \overline{A}|_{\overline{E}} \qquad\qquad + \frac{1}{2}\frac{\partial^2 \overline{A}|_{\overline{E}}}{\partial \overline{E}^2}\overline{(E-\overline{E})^2}$$

The difference $A - \overline{A}|_{\overline{E}}$ will thus involve, besides a term linear in $E - \overline{E}$ a term containing

$$(E - \overline{E})^2 - \overline{(E - \overline{E})^2} = (E - \overline{E})^2 - B$$

this term will be negligible in comparison with the first one, and we may therefore write

$$\overline{e^{\eta(\theta_0,\ a_0;\ pq)}}\Big|_{\overline{E}(a)} \simeq e^{(\theta_0,a;pq)}\left[1 + \frac{da}{\theta_0}\frac{\partial \overline{A}|_{\overline{E}}}{\partial E}(E-\overline{E})\right]$$

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_0 to $\theta = \theta_0 + d\theta$ we have, using Helmholtz's formula,

$$e^{\eta(\theta_0,a;\ pq)} = e^{\eta(\theta,a;pq)}\left[1 - \frac{d\theta}{\theta_0^{\,2}}(E-\overline{E})\right]$$

We conclude that the final distribution is indeed the canonical one of modulus θ, provided that the variation $d\theta$ of modulus is related to the variation da by the equation

$$\frac{d\theta}{\theta_0} - \frac{\partial \overline{A}|_{\overline{E}}}{\partial E}\ da = 0$$

Replacing $\overline{A}|_{\overline{E}}$ by A and θ_0 by θ, and noting, at constant a

$$\frac{\partial A}{\partial \overline{E}} = \frac{\partial A}{\partial \theta} \Big/ \frac{\partial \overline{E}}{\partial \theta}\ ,$$

the last equation may be written

$$\frac{1}{\theta}\frac{\partial \overline{E}}{\partial \theta}\partial\theta - \frac{\partial A}{\partial \theta}da = 0$$

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 in an irreversible passage of our system from one equilibrium state to another, 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 (P) and study the behaviour of their average values.

$$\overline{\overline{\omega}} = \int \overline{\omega} e^{\overline{\omega}} d\mu$$

under the various transformations satisfying the above-mentioned condition. 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

---

* In fact, from (18) it follows that

$$-\frac{\partial A}{\partial \theta} = \frac{\partial S}{\partial a}$$

and from (19) that

$$\frac{1}{\theta}\,\frac{\overline{E}}{\theta} = -\frac{\partial^2 F}{\partial \theta^2} = \frac{\partial S}{\partial \theta}$$

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 constant in each energy shell have a smaller average probability exponent than the others. We call this statement the <u>first lemma of Gibbs</u>. The assumption is

$$\int_{(E)} e^{\bar\omega}\, d\mu = \int_{(E)} e^{\bar\omega_0}\, d\mu$$

for each energy shell (E). For the average of $\bar\omega_0$ over a shell we may thus write

$$\int_{(E)} \bar\omega_0\, e^{\bar\omega_0}\, d\mu = \bar\omega_0 \int_{(E)} e^{\bar\omega_0}\, d\mu = \bar\omega_0 \int_{(E)} e^{\bar\omega}\, d\mu = \int_{(E)} \bar\omega_0\, e^{\bar\omega}\, d\mu$$

and therefore

$$\bar{\bar\omega}_0 = \int \bar\omega_0\, e^{\bar\omega}\, d\mu$$

We must now prove that

$$\bar{\bar\omega} - \bar{\bar\omega}_0 = \int (\bar\omega - \bar\omega_0)\, e^{\bar\omega}\, d\mu \geqslant 0$$

Gibbs uses here a general procedure applicable to any inequality of the form $-\int x e^{\bar\omega} d\mu \geqslant 0$. 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 $-x = \bar\omega - \bar\omega_0$; from the normalization conditions

$$\int e^{\bar\omega}\, d\mu = \int e^{\bar\omega_0}\, d\mu = 1$$

it indeed follows that

$$\int e^{\bar\omega} \left[e^{\bar\omega_0 - \bar\omega} - 1 \right] d\mu = 0$$

Adding this to the expression $\int (\bar\omega - \bar\omega_0)\, e^{\bar\omega}\, d\mu$, we immediately conclude that this quantity is always $\geqslant 0$.

Let us now consider the successive steps <u>da</u> of the adiabatic variation of the parameter <u>a</u> from a_0 to a_n. Let $\bar\omega_1, \bar\omega_2 \ldots, \omega_n$ be the successive probability exponents. If the transformation is quasi-static, the differences quasi-static, the differences $\bar{\bar{\omega}}_{i+1} - \bar{\bar{\omega}}_i$ are of the second order in d<u>a</u>, the difference $\bar{\bar{\omega}}_1 - \bar{\bar{\omega}}_n$ after a finite variation of <u>a</u> 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 d<u>a</u>, and for a finite variation of <u>a</u> there will be, by Gibbs' lemma, a finite increase of the entropy.

(b) <u>Heat exchange</u>. 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 total 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 $e^{\bar{\omega}}$ defines the distributions in the S_1, and S_2 spaces as

$$e^{\bar{\omega}_1} = \int e^{\bar{\omega}} \, d\mu_2 \qquad e^{\bar{\omega}_2} = \int e^{\bar{\omega}} \, d\mu_1 \;;$$

these partial distributions are, of course, normalized to unity;

$$\int e^{\bar{\omega}_1} d\mu_1 = \int e^{\bar{\omega}_2} d\mu_2 = \int e^{\bar{\omega}} d\mu = 1$$

<u>Gibbs' second lemma</u> states that

$$\bar{\bar{\omega}} \geq \bar{\bar{\omega}}_1 + \bar{\bar{\omega}}_2$$

the equality holding only if the two distributions $\bar{\omega}_1, \bar{\omega}_2$, are entirely independent, so that $\bar{\omega} = \bar{\omega}_1(P) + \bar{\omega}_2(P_2)$

To prove the lemma, note that

$$\bar{\bar{\omega}}_1 = \int \bar{\omega}_1 \, e^{\bar{\omega}_1} d\mu_1 = \int \bar{\omega}_1 \, e^{\bar{\omega}} d\mu, \quad \bar{\bar{\omega}}_2 = \int \bar{\omega}_2 \, e^{\bar{\omega}} d\mu$$

and therefore

$$\bar{\bar{\omega}} - \bar{\bar{\omega}}_1 - \bar{\bar{\omega}}_2 = \int (\bar{\omega} - \bar{\omega}_1 - \bar{\omega}_2) \, e^{\bar{\omega}} d\mu$$

The normalization conditions give

$$\int e^{\bar{\omega}} (e^{\bar{\omega}_1 + \bar{\omega}_2 - \bar{\omega}} - 1) \, d\mu = 0$$

whence the announced inequality follows. Moreover, the equality is seen to hold only if $\bar{\omega} - \bar{\omega}_1 - \bar{\omega}_2 = 0$, 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 $\bar\omega_1, \bar\omega_2$ and the exponent of the total distribution is $\bar\omega = \bar\omega_1 + \bar\omega_2$. As a result of the thermal contact (during which the total system is isolated) the exponent changes from $\bar\omega$ to $\bar\omega'$, and by the first lemma

$$\bar{\bar\omega} \geqslant \bar{\bar\omega}'$$

Now, we had on the one hand

$$\bar\omega = \bar\omega_1 + \bar\omega_2,$$

and on the other, by the second lemma

$$\bar{\bar\omega}' \geqslant \bar{\bar\omega}'_1 + \bar{\bar\omega}'_2,$$

$\bar\omega'_1, \bar\omega'_2$ being the exponents of probability of the two systems in their final states. Therefore

$$\bar{\bar\omega}_1 + \bar{\bar\omega}_2 \geqslant \bar{\bar\omega}'_1 + \bar{\bar\omega}'_2 \qquad (26)$$

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 $\bar{\bar\omega} + (\bar E / \theta)$ 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

$$\int (\bar\omega + E/\theta)\, e^{\bar\omega} \cdot d\mu - F/\theta \geqslant 0, \quad \text{i.e.} \int (\bar\omega - \eta)\, e^{\bar\omega}\, d\mu \geqslant 0,$$

having the probability exponent of the canonical distribution of modulus θ. By the usual procedure, we derive from the normalization conditions

$$\int e^{\bar\omega} \left[e^{\eta - \bar\omega} - 1 \right] d\mu = 0$$

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,

$$\bar{\bar{\omega}}_2 + \bar{E}_2 / \theta_2 \leq \bar{\bar{\omega}}'_2 + \bar{E}'_2 / \theta_2$$

since the initial state was described by a canonical distribution of modulus θ_2 . Therefore,

$$\bar{\bar{\omega}}_1 - \bar{\bar{\omega}}'_2 \geqslant \frac{\bar{E}_2-\bar{E}'_2}{\theta_2} \tag{27}$$

or in terms of thermal quantities

$$S'_1 - S_1 \geqslant \frac{Q}{\theta_2} \quad ,$$

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 from 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

$$\bar{\bar{\omega}}_1 + \frac{\bar{E}_1}{\theta_2} \geqslant \omega'_1 + \frac{\bar{E}'}{\theta_2}$$

showing how the quantity $\bar{\bar{\omega}} + (\bar{E} / \theta_2)$ 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_o undergoing a cycle of transformations in which it comes successively in contact with various thermostats of temperature θ_1, θ_2.... Assuming S_o to be initially at temperature θ_o, we have

$$\bar{\bar{\omega}}_o + \bar{\bar{\omega}}_1 + \bar{\bar{\omega}}_2 + \ldots\ldots \geqslant \bar{\bar{\omega}}'_o + \bar{\bar{\omega}}'_1 + \bar{\bar{\omega}}'_2 + \ldots\ldots$$

$$\bar{\bar{\omega}}'_i + \frac{\bar{E}'_i}{\theta_i} \geqslant \bar{\bar{\omega}}_i + \frac{\bar{E}_i}{\theta_i} \quad (i = 0, 1, 2, \text{-.})$$

whence

$$\frac{\bar{E}_2-\bar{E}'_1}{\theta_1} + \frac{\bar{E}_2-\bar{E}'_2}{\theta_2} + \ldots\ldots\ldots \leqslant 0 \tag{28}$$

the term pertaining to the system S_o has disappeared since $\bar{E}'_o=\bar{E}_o$

owing to the cyclic character of the transformation. The inequality (28) coincides with an important theorem proved by Clausius 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

$$\frac{Q_1}{\theta_1} + \frac{Q_2}{\theta_2} \leqq 0$$

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_o, 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 well known 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 $P_{N_1N_2}$ 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

$$f(P_N) = \sum_{i=1}^{N} f^{(1)}(P_i) + \sum_{i=1}^{N}\sum_{k=1}^{N} f^{(2)}P_i,P_k) +\dots \quad (1)$$

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 $f(P_{N_1N_2})$.

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 conjunction 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 possible 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 *

$$W(N) = \frac{\bar{N}^N}{N!} \; e^{-N} \qquad (2)$$

$\bar{N}$ 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 $f(PN_1N_2)$. Here, $f(PN_1N_2)$ depends 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

$$H(P_{N_1N_2}) + H_{R1}\,(P_{N_1'}) + H_{R2}(P_{N_2'}) \qquad (4)$$

where P_{N_i}' 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

---

* 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.

of molecules in the system S.

The statistical weight of the phase P_{N1N2} of the system S is therefore proportional to

$$d\mu_{N_1,N_2} e^{-\beta H(P_{N_1,N_2})} \int e^{-\beta\left[H_{R_1}(P_{N_1'}) + H_{R_2}(P_{N_2'})\right]} d\mu_{N_1'} d\mu_{N_2'},$$

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

$$\bar{F} = \frac{\sum\limits_{N_1,N_2} W(N_1)W(N_2)\, Z_{R_1}(\mathcal{N}_1 - N_1) Z_{R_2}(\mathcal{N}_2 - N_2) \int f(P_{N_1N_2}) e^{-\beta H(P_{N_1N_2})} d\mu_{N_1N_2}}{\sum\limits_{N_1,N_2} W(N_1)W(N_2) Z_{R_1}(\mathcal{N}_1 - N_1) Z_{R_2}(\mathcal{N}_2 - N_2)\, Z(N_1, N_2)}$$

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 <u>a</u>, the <u>average</u> numbers $\bar{N}_1$, $\bar{N}_2$ of molecules in the system (which, however, are not necessarily all independent).

2.- <u>Chemical potentials.</u>

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 <u>relative</u> deviations $-(N_i' - \bar{N}_i')/\bar{N}_i'$ of the numbers N_i' from their average values $N_i' = \mathcal{N}_i - \bar{N}_i$. Even the absolute fluctuation $N_i' - \bar{N}_i' = -(N_i - \bar{N}_i)$ may be restricted in magnitude; for the contributions of terms corresponding to large flucuations 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. $Z_{R_i}(N_i^!)$ 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 <u>intensive</u> 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 parts 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 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 isotropic structure; there are accordingly two different "free energies" F(θ, V) and G(θ, 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(θ, a) to the function G(θ, A) is affected by a "Legendre transformation":

---

* Cf R. Tolman, Phys. Rev. <u>57</u>, 1160, 1940

$$G(\theta, A) = F\left[\theta, a(A)\right] - a(A) \cdot A ,$$

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

$$g(\theta,\, p) = \left(\frac{\partial G}{\partial N}\right)_{\theta;\, p} \tag{6}$$

so that

$$G(N;\, \theta,\, p) = N\, g(\theta,p) \tag{7}$$

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 pressure in the reservoir is regulated by the mechanical equilibrium conditions across the semi-permiable 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 <u>small</u> deviations, however, we may write:

$$\log \frac{Z_{R_1}(N_1',\, \theta,\, p_1)}{Z_{R_1}(\overline{N}_1',\, \theta, p_1)} = -(N - \overline{N}_1')\, \beta\, g_1\,(\theta,p_1) =$$

$$= (\overline{N}_1 - N_1)\, \beta\, g_1(\theta,p_1) ,$$

where the reservoir pressure p_1 is a function of the pressure <u>p</u> of the system S and of the average numbers $\overline{N}_1$, $\overline{N}_2$. Apart from a factor independent of N_1, 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 $W(N_i)$, given by (2); we may say that in the average (5), each number N_i is weighted by a factor of the form

$$\frac{1}{/N_i!}\, e^{\beta \varphi_i N_i}$$

where the function $\varphi_i(\theta, p;\, \bar{N}_1, \bar{N}_2) = -g_i + \theta \log \bar{N}_i$ 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 <u>chemical potential</u> of the molecular species <u>i</u> in the system S.

The average (5) of the phase function f now takes the form

$$\bar{f} = \frac{\sum\limits_{N_1 N_2} \frac{1}{N_1!\,N_2!}\, e^{\beta(\varphi_1 N_1 + \varphi_2 N_2} \int f(P_{N_1 N_2})\, e^{-\beta H(P_{N_1 N_2})}\, d\mu_{N_1 N_2}}{\sum\limits_{N_1 N_2} \frac{1}{N_1!\,N_2!}\, e^{\beta(\varphi_1 N_1 + \varphi_2 N_2)}\; Z\,(N_1, N_2)} \tag{8}$$

The phase integrations indicated by μ_{N1N2} are over specific 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 given rise to much confused 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. According to formula (8) the average $\bar{f}$ can be described as a canonical average over generic phases for each possible set of numbers of molecules, weighted by an exponential distribution factor $e^{\beta(\varphi_1 N_1 + \varphi_2 N_2)}$ there is a complete analogy with the concept

---

[*] 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 $(N!_1 - \bar{N})/\bar{N}!_1$.

of canonical average itself which is an ergodic average over each energy surface, weighted by an exponential factor $\exp\left[-\beta H\right]$. 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):

$$e^{-\Omega} \equiv \sum_{N_1, N_2} \mathcal{Z}\,(N_1, N_2)\, e^{\beta(\zeta_1 N_1 + \zeta_2 N_2)} \qquad (11)$$

and differentiate it 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

$$-d\Omega = e^{\Omega}\sum_{N_1,N_2} \mathcal{Z}(N_1,N_2)e^{\beta(\zeta_1N_1+\zeta_2N_2)}\Big[d\log\mathcal{Z}(N_1N_2)+N_1d(\beta\zeta_1)+ + N_2\,d\,(\beta\zeta_2)\Big]$$

$$= d\,\overline{\left[\beta G\,(N_1,N_2)\right]} + \bar{N}_1\,d\,(\beta\zeta_1) + \bar{N}_2\,d(\beta\zeta_2),$$

where the averages refer to the distribution of the numbers N_1,N_2 given by the probability law

$$e^{\Omega}\,\mathcal{Z}\,(N_1N_2)\,e^{\beta(\zeta_1N_1+\zeta_2N_2)}$$

The last equation may also be written

$$d\left[\Omega+\beta\zeta_1\bar{N}_1+\beta\zeta_2\bar{N}_2\right] = d\,\overline{\left[\beta\;G(N_1N_2)\right]}+\beta\zeta_1 d\,\bar{N}_1+\beta\zeta_2 d\,\bar{N}_2$$

which shows that the characteristic function in terms of the independent variables θ, p_φ $\bar{N}_1$, $\bar{N}_2$ is $\Omega+\beta\,(\zeta_1\bar{N}_1+\zeta_2\bar{N}_2)$.

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,

$$\Omega+\beta\,(\zeta_1\bar{N}_1+\zeta_2\bar{N}_2)\approx\beta\,G(\bar{N}_1\;\bar{N}_2) \qquad (13)$$

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 free energy *, 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

---

* The factor β can evidently be retained or omitted without changing the property of the function of being a "characteristic" one.

of a system of identical elements preserves its extensive 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

$$Z\ (N_A + N_B) = \frac{(N_A+N_B)!}{N_A!\ N_B!}\ Z(N_A)\ Z(N_B)$$

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

$$\zeta_2 = \left[\frac{\partial G}{\partial \overline{N}_i}\right]_{\Theta,p\left\{\overline{N}_i\right\}} \tag{14}$$

the symbol $\left\{\overline{N}_i\right\}$ indicating that all $\overline{N}_k$'s <u>except</u> $\overline{N}_i$ are kept constant in the derivation. The extensive character of G allows us to write

$$G = \sum_i \overline{N}_i \frac{\partial G}{\partial \overline{N}_i} = \sum_i \overline{N}_i\, \zeta\ : \tag{15}$$

Comparing (15) with (13), we see that, to the approximation considered, $\Omega \approx 0$. This means that the denominator of formula (8) practically 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

$$\beta G = - N_1 \log Z_1 - N_2 \log Z_2 + \sum_i \log N_i!$$

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 før a consistent treatment of systems of various kinds of elements that these elements be distinguished by <u>discontinuous</u> 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 <u>in principle</u> 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 $G(\theta, p, \overline{N})$ 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 $G - \sum_i \overline{N}_i \zeta_i$, i.e. by (15), identically zero. There accordingly remain three possibilities, viz.*

---

* 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).

$$G(\theta, p, \bar{N}) = \sum_i \bar{N}_i \zeta_i$$
$$F(\theta, v, \bar{N}) = \sum_i N_i \zeta_i - p V \qquad (16)$$
$$\Omega(\theta, v\, \zeta) = - p V$$

The property of being a characteristic function is expressed in the three cases by similar equations:

$$dG = - Sd\,\theta + V\,dp + \sum_i \zeta_i\, d\,\bar{N}_i$$
$$dF = - Sd\,\theta - pdV = + \sum_i \zeta_i\, d\,\bar{N}_i \qquad (17)$$
$$d\Omega = - Sd\,\theta, - pdV - \sum_i \bar{N}_i\, d\zeta_i$$

Thus wee see that the chemical potentials can be defined by

$$\zeta_i = \left(\frac{\partial F}{\partial N_i}\right)_{\theta,\, V; \{\zeta_i\}} \qquad (18)$$

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 being simply related to the normalization factors of the respective statistical distributions. The definition of Ω is in fact given by

$$e^{-\beta\Omega} = {}_{N_1N_2\ldots} \frac{e^{\beta \sum_i \zeta_i N_i}}{N_1!N_2!\ldots} \int e^{-\beta H\,(P_{N_1N_2\ldots};V)}\, d\mu_{N_1 N_2\ldots} \qquad (20)$$

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

$$e^{-\beta F} = \sum_{N_1N_2\ldots} \frac{e^{\beta \sum_i \zeta_i (N_i \bar{N}_i)}}{N_1!\ N_2!\ldots} \int e^{-\beta\, H\,(P_{N_1N_2\ldots};V)}\, d\mu_{N_1N_2\ldots} \qquad (21)$$

---

* 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.

if we neglect the fluctuations of the numbers N_i around their averages $\bar{N}_i$ this expression reduces to

$$e^{-\beta F} \approx \frac{1}{N_1! N_2!} \int e^{-\beta H(P_{\bar{N}_1 \bar{N}_2 \cdots};\, V)} \, d\mu_{\bar{N}_1 \bar{N}_2 \cdots} \tag{22}$$

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

$$H\ (P;\ p) = H\ (P;\ V) + p\ V \tag{23}$$

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.)