REVIEWS OF MODERN PHYSICS VOLUME 27, NUMBER 3 JULY, 1955

Foundations of Statistical Mechanics*

D. TER HAAR

Department of Natural Philosophy, St. Salvator's College, St. Andrews, Scotland

TABLE OF CONTENTS

INTRODUCTION

In recent years there have been many attempts (44K, 46B, 46G, 46K, 47B1, 47B2, 48B1, 48B2, 49B1, 49B2, 49B3, 49K1, 49K2, 52K, 52R, 52S2)[1] to justify the use of statistical mechanics in describing physical systems. As far as classical statistical mechanics, that is, statistical mechanics before the advent of quantum mechanics, is concerned, the Ehrenfests (11E2) in their well-known survey article in the *Enzyklopaedie der mathematischen Wissenschaften* have discussed the foundations of statistical mechanics and, as we shall see, their conclusions are still valid in more or less the same form as stated by them.[2] However, no such account of the state of the foundations of statistical mechanics including quantum mechanics is available, and it is the purpose of the present paper to attempt to give such an account. The present author has discussed in a recently published monograph (54H1)[3] some of the problems to be covered here, but, on the one hand, since that monograph was written a number of new developments have arisen, and, on the other hand, the emphasis in the present paper will be more on the present status of the foundations of statistical mechanics that on the historical development of these foundations as was the case in the earlier discussion. The present discussion is thus to a large extent complementary to the discussion in ESM. This does, of course, not mean that now we are to neglect completely the historical aspects of the development of statistical mechanics.

One might argue that the proof of the pudding is in the eating, and that the fact that statistical mechanics has been able to predict accurately and successfully the behavior of physical systems under equilibrium conditions—and even under certain circumstances in nonequilibrium conditions[4]—should be a sufficient justification for the methods used. This is especially the case nowadays since the introduction of quantum mechanics has cleared up such seemingly unsurmountable difficulties as the paradox of the specific heats or Gibbs' paradox. However, it is only logical that many physicists have not felt satisfied by this argument and have therefore attempted to show that the statistical formalism follows in a straightforward manner from classical or quantum mechanics. In the present survey we shall discuss in how far these authors have succeeded. Even if success has been limited, these attempts have been extremely important and very helpful in showing both the possibilities and the limitations of the statistical approach.

---

* Part of this paper was written while the author was on leave of absence at Purdue University and he would like to express his thanks to Dr. Lark-Horovitz for his hospitality. He would also like to express his thanks to Dr. N. G. van Kampen for some helpful criticism.

[1] References are placed at the end of the paper in chronological order. Each reference number consists of the year of publication and the first letter of the first author's surname; e.g., M. J. Klein, Phys. Rev. **87**, 111 (1952) is referred to as 52K.

[2] R. Kurth (55K) has also considered the foundations of classical thermostatistics from a more axiomatic point of view. I should like to express to him my thanks for showing me his paper before publication.

[3] We shall use, as far as feasible, the notation of this publication, and we refer to it as ESM.

[4] For a discussion of the statistical mechanics of nonequilibrium processes we refer to a recent review article by Montroll and Green (55M) and its bibliography.

Before we can start the discussion of the various papers on the fundamental ideas of statistical mechanics, we must remind ourselves that we shall mainly be concerned with the following two questions.

(A) Why is it possible to describe the behavior of almost all physical systems by considering only equilibrium situations?

(B) Why is it possible to describe the behavior of an actual physical system by considering a large number of identical systems (in ensemble theory) and identifying the average behavior of this group of systems with the behavior of the physical system in which we are interested?

In the following discussion it will become clear that these two questions are intimately related. The first one arises not only in statistical mechanics, but also in thermodynamics and kinetic theory, and a related problem is the question,

(A1) How can one define an equilibrium situation?

The second question is of a truly statistical nature, and is related to the question,

(B1) How can we construct an ensemble such that it will "represent" an actual, given physical system?

This question of *representative ensembles* which will play an important role in our discussion was especially stressed by Tolman (38T, 40T).

At this point we must introduce a definition of statistical mechanics, and we shall do this by using the definition given by Kramers in his presidential address to the International Conference on Statistical Mechanics at Florence in 1949 (49K3): "For a physicist nowadays Statistical Mechanics is just that branch of physics which deals with the atomistic interpretation of the thermal properties of matter and radiation; for this reason one might also call it 'thermostatistics'."

In our discussion we shall not be concerned explicitly with radiation, but restrict ourselves to systems containing particles. However, in discussing transitions from one state to another, the existence of a radiation field, which in many cases will be the agent enabling the transitions to take place, will tacitly be assumed.

Statistical mechanics has its origin in the kinetic theory of gases which was developed in the nineteenth century. It is probably justifiable to call Krönig the father of the kinetic theory, even though more than a century earlier Bernoulli (38B) had related the properties of a gas to the properties of individual particles. Krönig (56K) draws attention to the fact that although the orbit of any atom in a gas will be so irregular that it is impossible to follow the atom, one can use the theory of probability to reduce this completely chaotic behavior to the ordered behavior of a gas. The kinetic theory of gases was further developed by Clausius (57C1, 57C2, 58C, 62C, 70C1, 70C2), Maxwell (60M, 67M, 68M1, 68M2), and Boltzmann (68B, 72B, 96B, 98B).[5] In his *Gastheorie* Boltzmann was especially concerned with question (A) and he had introduced in 1872 (72B) his famous H-theorem in order to show that any nonequilibrium situation would develop in such a way that it would approach an equilibrium situation. The answer to question (A1) is in this case that the equilibrium situation is the most probable situation, that is, the most probable situation compatible with a few restricting conditions. In its original unrestricted form the H-theorem, if correct, proved that any system will tend toward equilibrium, if equilibrium does not initially exist.[6] This would mean that provided one waited a sufficiently long time one would find the system in an equilibrium situation and, moreover, this equilibrium situation would persist for ever after. From this it would follow that equilibrium would be the rule and nonequilibrium the exception so that question (A) is answered.

However, it was soon realized that the H-theorem in its original unrestricted form was not an absolute proof, but was based on certain assumptions about the number of collisions which a given particle will undergo during a specified time interval. Moreover, Loschmidt (76L, 77L) and Zermelo (96Z) showed conclusively that the H-theorem in its original form could not be true. In his later papers Boltzmann was therefore careful to stress the statistical aspects of the H-theorem. This means that the H-theorem is a statement about the most probable behavior of a system. Fluctuations around the equilibrium situation are, however, no longer prohibited. The statistical aspects of the H-theorem were especially stressed by the Ehrenfests (07E, 11E2) both in general and by considering simple models.[7]

In the first part of the present article we shall discuss this stage of the discussion of the foundations of statistical mechanics. In Sec. A(1) we shall discuss the unrestricted H-theorem. The arguments of Loschmidt and Zermelo against the H-theorem in its original form will be discussed in Sec. A(2), while we discuss in Sec. A(3) the statistical aspects of the H-theorem.

Apart from introducing the H-theorem to prove (A) Boltzmann also tried to prove that the average behavior of a system is the same as its equilibrium behavior. The contents of this statement are that the time average taken over an infinite period of any phase function, that is, a function depending on the values of the coordinates and velocities which completely determine the physical situation, or phase, of the system under consideration, should be equal to the value of this phase function at equilibrium. It can easily be seen that this

---

[5] We may refer here to the Bibliographical notes at the end of chapters I and II and at the end of Appendix I of ESM for more historical details.

[6] We exclude clearly hypothetical systems such as completely perfect gases within idealized walls. In such a system there would be no mechanism to alter the distribution function, and a nonequilibrium situation would persist.

[7] The importance of the statistical nature of the H-theorem is not properly taken into account by Sartre (54S) in a recent note.

second way of attacking (A) is equivalent to the first one. Firstly, if the H-theorem is correct, it follows that any nonequilibrium situation will develop into an equilibrium situation which will then persist. In taking a time average of a phase function over an infinite period we thus obtain the value of this phase function at equilibrium. Secondly, if the time average is equal to the value at equilibrium, it means that the system must be in an equilibrium state during most of the time and hence return to equilibrium from any nonequilibrium state.[8] In order to prove that average behavior and behavior at equilibrium are the same, it is necessary to calculate time averages and one then meets with serious difficulties. Boltzmann tried to get over these difficulties by assuming that most physical systems are *ergodic*.[9] An ergodic system is one which has a representative point in Γ space[10] which passes through every point of the energy surface corresponding to the energy of the system. The evaluation of the time average is in this case the same as taking the average over the energy surface, that is, taking the average over a collection of systems which possess all the same energy. This particular collection of systems, or *ensemble*, to use the term coined by Gibbs, is a so-called microcanonical ensemble. Since according to our assumptions the representative point of any system in the ensemble will pass through every point on the energy surface, the orbits described by the representative points of the systems in the ensemble will all be identical, the only difference between the systems being the exact moment at which a point of the orbit is passed.[11] It follows then that the time average which is the average taken over the one and only orbit on the energy surface where different points are passed at different times will give the same result as the average taken over the microcanonical ensemble which is the average taken over the same orbit, but now considering all points at the same time.

In considering averages over the energy surface, or over a microcanonical ensemble, Boltzmann departed from kinetic theory and really and truly entered the realm of statistical mechanics, if we use Gibbs' definition of statistical mechanics in the preface to his famous monograph (02G).[12] It must, however, be noticed that ensemble theory enters, so to speak, by the back door, as it is only introduced as a mathematical trick to calculate the behavior of one isolated system.

Around the turn of the century with the development of measure theory it was realized that ergodic systems would never occur, but it was hoped that most physical systems would be quasi ergodic. A quasi-ergodic system is one which has an orbit in Γ space which covers the energy surface everywhere densely, even though not actually passing through every point of it. It was also hoped that quasi ergodicity would be sufficient to ensure the equality of time averages and averages taken over a suitably chosen ensemble. This was, for instance, the point of view taken by the Ehrenfests (11E2). In 1913 Rosenthal (13R) and Plancherel (13P) independently proved the impossibility of ergodic systems, while in 1923 Fermi (23F1) showed that a certain class of systems was quasi ergodic—without, however, proving the equality of time averages and ensemble averages.[13] This equality had been proved by Rosenthal (14R2), but his proof was not completely rigorous.[14] More recently this problem of the equivalence of time and ensemble averages has been studied mainly by mathematicians. It is usually called the *ergodic theorem*. In 1931 and 1932 Birkhoff (31B1, 31B2, 32B) and von Neumann (32N1, 32N2) have shown that, provided certain plausible mathematical conditions were fulfilled, the two kinds of averages would give the same result, while Oxtoby and Ulam (41O) recently showed that a very wide class of systems would satisfy these mathematical conditions. The ergodic theorem will be discussed in Sec. B. In Sec. B(1) we shall discuss the developments up to about 1930, while modern developments are discussed in Sec. B(2).

The approach via the ergodic theorem is still con-

---

[8] This equivalence remains if one substitutes the H-theorem in its statistical form for its unrestricted form as can easily be seen.

[9] Ergodic from the Greek *ergon* (=work, used here in the sense of energy) and *hodos* (=path): the representative point in Γ space passes through all points of the energy surface. The term ergodic was first introduced by Boltzmann in 1887 (87B). The assumption of ergodicity is called by Maxwell the assumption of the *continuity of path*.

[10] A system with S degrees of freedom can be described by S (generalized) coordinates and S (generalized) momenta. The values of these $2S$ quantities at a given time will define a point, the so-called representative point, in a $2S$-dimensional space, called Γ space (Γ for gas), or phase space. A $2S-1$-dimensional hypersurface in Γ space, defined by the equation $\epsilon=$constant, where ϵ is the energy of the system, is called an energy surface.

[11] We use here the fact that the direction of the orbit at any point of Γ space is uniquely defined from the Hamiltonian equations of motion. From this it follows too that an orbit can never return to a point in Γ space through which it has passed—unless we are dealing with a strictly periodic orbit.

[12] We may at this point perhaps advise every student of statistical mechanics to read and reread carefully this preface, as it expresses more clearly than anywhere else the basic ideas of rational thermodynamics, as Gibbs calls it. It is in this preface also that the term statistical mechanics is coined. In view of the difficulties encountered by kinetic theory at the time that Gibbs' monograph appeared (paradox of specific heats, the understanding of Planck's radiation law, $\cdots$) Gibbs tried to build up a rational system based on a few axioms and not necessarily connected with natural phenomena. As Gibbs put it, "It is well known that while theory would assign to the gas (of diatomic molecules) six degrees of freedom per molecule, in our experiments on specific heat we cannot account for more than five. Certainly, one is building on an insecure foundation, who rests his work on hypotheses concerning the constitution of matter. Difficulties of this kind have deterred the author from attempting to explain the mysteries of nature, and have forced him to be contented with the more modest aim of deducing some of the more obvious propositions relating to the statistical branch of mechanics. Here, there can be no mistake in regard to the agreement of the hypotheses with the facts of nature, for nothing is assumed in that respect. The only error into which one can fall, is the want of agreement between the premises and the conclusions, and this, with care, one may hope, in the main, to avoid."

[13] We shall see later on, however, that this equality holds indeed for all quasi-ergodic systems.

[14] In his proof a double transition to the limit is taken, as was pointed out by Rosenthal himself to Epstein (see 36E, p. 478).

sidered by many physicists to be the most satisfactory, if not the only one, and the one providing the best justification for the use of statistical methods. It is felt by these people that mechanics can only deal with isolated systems and that the statistical approach is a necessary mathematical device without any further physical interest. It seems to the present author that this point of view fails to take into account the physical reason for a statistical approach.[15] The reason is that the enormously large number of particles in most physical systems will make it impossible for us to determine the values of all the integrals of motion. In general, we shall only determine the energy, the total linear momentum, and perhaps one or two more. We then must use this very inadequate knowledge about the system to predict its future behavior. Since our knowledge is clearly insufficient to predict the future of the system with complete certainty, we must have recourse to statistical methods and this is the point where *representative ensembles* enter.[16] Instead of considering one system, we consider a large collection of systems all possessing the same values of those quantities of which we know the value, but otherwise differing widely. In order to construct such a representative ensemble satisfactorily, we must know what weight to give to the various systems in the ensemble, or, in other words, we must make assumptions about *a priori* probabilities.

Once the need for representative ensembles is recognized and accepted, question (B) is solved in ensemble theory by showing that the great majority of the systems in the ensembles considered behave in practically the same way and, moreover, give just those values for phase functions which one should expect from a system at equilibrium.[17] However, question (B1) remains to be answered. One always finds that a canonical ensemble is the representative ensemble of a system in temperature equilibrium. In order to prove this one introduces a generalized H-theorem and shows that, if the representative ensemble at one moment is not a canonical ensemble, the situation at a later moment will be such that we are forced to use a representative ensemble which resembles much more closely a canonical ensemble. This question of the approach to equilibrium described by ensembles and the question of the representative ensembles will be discussed in part C, as far as classical ensemble theory is concerned.

The transition from classical to quantum statistics does not introduce any fundamental changes. As a matter of fact, one could develop the two cases, that is, the classical and the quantum-mechanical case, at the same time to emphasize the close similarity. There are, however, certain differences and for that reason we have preferred to discuss the two cases separately. In the quantum-mechanical case one can again approach the problem either via ensemble theory or via an ergodic theorem. We shall discuss these two alternative methods of approach in parts D and E, respectively.

In concluding this Introduction, I should like to make a few remarks concerning some general aspects of the second law of thermodynamics and its place in statistical mechanics.[18] The fact that in all practical applications of thermodynamics the second law holds, that is, the entropy increases, is due to the combination of two effects. First of all, if the entropy at a certain moment is smaller than its equilibrium value, the probability that it will increase is overwhelmingly larger than that it will decrease. Secondly, our observations are always made in such a way that starting from a given situation we watch the future development, but it is impossible for us to start from a given situation and watch the preceding time interval. The fact that we can do the one, but not the other is related to the fact that we have a memory of the past and we can thus possess knowledge of what happened at an earlier time, but not of what will happen at a later moment. To this extent the irreversibility of the second law of thermodynamics is physiological and basic.

Another point is that, as far as one can judge at this moment, all observations regarding phenomena in the universe are reconcilable with the idea that the universe as a whole is developing from some reasonable, though thermodynamically or statistically unlikely, state in the distant past. This may be thought to be connected with cosmogonical ideas involving a more or less singular beginning at, say, three to five billion years ago. On the other hand, it can also be argued, as was, for instance, done by Boltzmann (95B), that the fact that our world seems to be developing from a less probable to a more probable state is not necessarily in contradiction to the idea that the universe as a whole is in thermodynamic equilibrium. To quote Boltzmann: "We assume that the whole universe is, and rests for ever, in thermal equilibrium. The probability that one (only one) part of the universe is in a certain state is the smaller the further this state is from thermal equilibrium but this probability is greater, the greater is the universe itself. If we assume the universe great enough, we can make the probability of one relatively small part being in any given state (however far from the state of equilibrium) as great as we please. We can

---

[15] We might also add that completely isolated systems are of no interest to a physicist, since it is impossible to perform experiments with or on them. Of course, as long as we are dealing with classical systems, we can always consider idealized experiments and thus still talk about isolated systems. Moreover, Pauli (49P) has pointed out that also in the case of quantum-mechanical systems the influence of the act of observation on the system can be neglected, as long as we are dealing with statistical mechanics, although it is of paramount importance in the case of quantum mechanics.

[16] See Secs. 24, 86, 112, 117, and 120 of Tolman's monograph (38T) and part B of ESM for a discussion of representative ensembles; compare also the discussion in Secs. C2 and D3 of the present paper.

[17] The proof of this statement can be found in any textbook on statistical mechanics dealing with ensemble theory (02G, 38T, 54H1).

[18] I should like to express my thanks to Professor R. E. Peierls for some clarifying and critical remarks on these points.

also make the probability great that, though the whole universe is in thermal equilibrium, our world is in its present state. It may be said that the world is so far from thermal equilibrium that we cannot imagine the improbability of such a state. But can we imagine on the other side how small a part of the whole universe this world is? Assuming the universe great enough, the probability that such a small part of it as our world should be in its present state is no longer small.

If this assumption were correct, one would return more and more to thermal equilibrium; but because the whole universe is so great, it might be probable that at some future time some other world might deviate as far from equilibrium as our world does at present. Then the aforementioned H-curve[19] would form a representation of what takes place in the universe. The summits of the curve would represent the worlds where visible motion and life exists."

In giving this quotation we are not subscribing to Boltzmann's point of view, only mentioning it. In judging this point of view on its merits one must investigate the difficult problem in how far it is possible to talk about thermal equilibrium in an open system such as our universe.

A. THE H-THEOREM IN CLASSICAL STATISTICAL MECHANICS

(1) The Kinetic Aspect of the H-Theorem; Stosszahlansatz

In 1872 Boltzmann (72B) introduced his famous H-theorem in order to prove that any nonequilibrium distribution will tend to an equilibrium distribution. Let us briefly summarize his reasoning. He was concerned with the case of an isolated system, and in particular with the case of a monatomic gas and no external forces, although he later extended his considerations to the case of polyatomic gases under the influence of external forces (75B).

We are interested to compute the rate of change due to collisions of the distribution function f, which in the present simple case is a function of the three (Cartesian) components u, v, and w of the velocity $\mathbf{c}$ of an atom, and of the time t only. We must therefore consider, on the one hand, the rate at which atoms with velocities between $\mathbf{c}$ and $\mathbf{c}+\mathbf{dc}$ change their velocities, and, on the other hand, consider the rate at which atoms with velocities within the selected range are produced through collisions. In formula

$$(\partial f/\partial t)dudvdwdt=-A+B. \qquad \text{(A1.01)}$$

In Eq. (A1.01) $fdudvdw$ is the number of atoms per unit volume which have velocities with components in the intervals $(u, u+du)$, $(v, v+dv)$, and $(w, w+dw)$, A is the number of atoms per unit volume with velocities inside the selected range which during a time interval dt change their velocities, and B is the number of atoms per unit volume which during a time interval dt change their velocities in such a way that after collisions their velocities lie inside the prescribed range. It can be shown (see, e.g., ESM Sec. I.4) that A is given by the equation

$$A=f(u,v,w)dudvdwdt \times\int f(u_1,v_1,w_1)du_1dv_1dw_1\int ad\omega. \qquad \text{(A1.02)}$$

In Eq. (A1.02) $d\omega$ is an element of solid angle around the so-called line of centers ω which is the vector from the one colliding atom to the other at the moment of impact, a is a quantity which depends only on the absolute magnitude of the relative velocity c_{rel} and on the angle θ between the relative velocity vector and ω, and the integration is over all values of u_1, v_1, and w_1, and over all possible directions of the line of centers. Equation (A1.02) is obtained by multiplying the number of atoms with velocities within the specified range ($f(u,v,w)dudvdw$) with the number of atoms which collide with one atom in a time interval dt. Consider atoms having a velocity between $\mathbf{c}_1$, with components u_1, v_1, and w_1, and $\mathbf{c}_1+\mathbf{dc}_1$. *Provided we may assume that there is no correlation between velocities and positions of different atoms*, the number of collisions between these atoms and one particular atom in the case where the line of centers lies within a solid angle $d\omega$ will be equal to it: $af(u_1,v_1,w_1)du_1dv_1dw_1dtd\omega$, where the quantity a is a normalizing constant depending on c_{rel} and θ. Equation (A1.02) now follows by integration over all possible values of u_1, v_1, and w_1, and over all possible directions of ω.

Similarly, we have for B the equation

$$B=dt\int' du'dv'dw'f(u',v',w') \times du_1'dv_1'dw_1'f(u_1',v_1',w_1')\int' a'd\omega'. \qquad \text{(A1.03)}$$

The integral signs are primed to indicate that we now may integrate only over those values of $\mathbf{c}'$ and $\mathbf{c}_1'$ and those directions of ω' which are such that the velocities after the collision, which are completely determined by $\mathbf{c}'$, $\mathbf{c}_1'$, and ω', satisfy the condition that one of them lies between $\mathbf{c}$ and $\mathbf{c}+\mathbf{dc}$. Expressing the primed quantities in terms of the unprimed ones we can write instead of Eq. (A1.03) (see, e.g., ESM Sec. I.4)

$$B=dudvdwdt\int du_1dv_1dw_1f(u',v',w') \times f(u_1',v_1',w_1')\int ad\omega, \qquad \text{(A1.04)}$$

where u', v', w', u_1', v_1', and w_1' are functions of u, v, w,

---

[19] This is essentially a curve giving the entropy as a function of time; compare Sec. A(3).

u_1, v_1, and w_1 through the equations expressing the conservation of linear momentum and energy.

Combining Eqs. (A1.01), (A1.02), and (A1.04) we get for the rate of change of the distribution function due to collisions,

$$\frac{\partial f}{\partial t}=-\int (ff_1-f'f_1')a\,du_1dv_1dw_1d\omega, \quad \text{(A1.05)}$$

where we have used the short-hand notation

$$f=f(u,v,w),\quad f'=f(u',v',w'),\quad f_1=f(u_1,v_1,w_1),\quad f_1'=f(u_1',v_1',w_1'). \quad \text{(A1.06)}$$

Introducing a quantity H by the equation[20]

$$H=\int f\ln f\,dudvdw, \quad \text{(A1.07)}$$

we get for the rate of change of H from Eq. (A1.05)

$$\frac{dH}{dt}=\int (f'f_1'-ff_1)(\ln f+1)\times a\,dudvdwdu_1dv_1dw_1d\omega. \quad \text{(A1.08)}$$

By observing the equivalence of $\mathbf{c}$, $\mathbf{c}_1$, $\mathbf{c}'$, $\mathbf{c}_1'$ in Eq. (A1.08), one can write this equation in the form (see, e.g., ESM Sec. I.5)

$$\frac{dH}{dt}=\frac{1}{4}\int (f'f_1'-ff_1)\ln(ff_1/f'f_1')\times a\,dudvdwdu_1dv_1dw_1d\omega. \quad \text{(A1.09)}$$

Since

$$(p-q)\ln(p/q)>0,\quad \text{if}\quad p\neq q,\quad \text{and}\quad =0,\quad \text{if}\quad p=q, \quad \text{(A1.10)}$$

and since a is a positive quantity, we have from Eq. (A1.09)

$$dH/dt\leqslant 0, \quad \text{(A1.11)}$$

where the equality in Eq. (A1.11) only holds, if for *any* two velocities $\mathbf{c}$ and $\mathbf{c}_1$ we have $ff_1=f'f_1'$.

It follows thus from Eq. (A1.11) that, provided our basic assumptions are correct, H will steadily decrease until it reaches a minimum value which is reached as soon as the distribution function satisfies the relation

$$ff_1=f'f_1' \quad \text{(A1.12)}$$

for any pair of velocities $\mathbf{c}$ and $\mathbf{c}_1$. The minimum or equilibrium value of H is in most cases reached rapidly, if H differs appreciably from this value. The relaxation time in the case of a gas at room temperature and one atmosphere of pressure is, e.g., of the order of 10^{-7} sec (ESM p. 390).

One can derive a formula for the equilibrium distribution function either by using Eq. (A1.12) or by taking as the equilibrium condition that for equilibrium H is minimum, in both cases under the restricting conditions of given total number of particles, given total energy, and given total linear momentum. The result is the so-called generalized Maxwell distribution.

Unfortunately the case of a real gas is not easy to treat in detail and, although it is possible to derive the H-theorem, it is not easy to discuss critically the various assumptions made in arriving at Eq. (A1.11) and the consequences of the removal of these restricting conditions. However, it is possible to introduce simplified models (11E2, 53H, 54G, 55G, 55H2) and we shall especially consider a model which is essentially the same as the model used by Lorentz (09L) in his discussion of the electron theory of metals. In this model we have two kinds of particles. The first kind (the lattice points, or metallic ions) are fixed in space, and we shall assume them to be randomly distributed in space with a density of n per unit volume. The second kind (the electrons) are traveling through the lattice and their density is N per unit volume. We neglect electron-electron encounters and, furthermore, assume that the electron-lattice collisions are elastic and isotropic. This means (a) that on such a collision the electron will not change the absolute magnitude of its velocity, and (b) that if $p(\theta,\varphi;\theta',\varphi')$ is the probability that an electron traveling in the direction determined by the polar angles θ and φ changes to a direction characterized by θ' and φ', $p(\theta,\varphi;\theta',\varphi')$ will be independent of θ' and φ'. Since the absolute magnitude of the velocity does not change during the collision, we can simplify our model by assuming that all electrons are moving with the same speed c. Finally, we shall denote by σ the cross section for an electron-lattice collision.

In order to consider the approach to equilibrium, we introduce a distribution function. Let $f(\theta,\varphi)d\omega/4\pi$ ($d\omega=\sin\theta d\theta d\varphi=$ element of solid angle) be the number of electrons per unit volume with velocities in directions within the solid angle $d\omega$. We have then the normalization condition

$$\int f(\theta,\varphi)d\omega=4\pi N, \quad \text{(A1.13)}$$

where the integration extends over the unit sphere.

According to our assumption about the model we have for the number of electrons per unit volume, $N_{\omega,\omega'}dtd\omega d\omega'$, which change their directions from within an element of solid angle $d\omega$ to within a solid angle $d\omega'$ during a time interval dt:

$$N_{\omega,\omega'}dtd\omega d\omega'=f(\theta,\varphi)nc\sigma(d\omega/4\pi)(d\omega'/4\pi)dt, \quad \text{(A1.14)}$$

where we have used the assumption of isotropic scattering.

Using Eqs. (A1.13) and (A1.14), we can easily compute the rate of change of $f(\theta,\varphi)$, and we get

$$df(\theta,\varphi)/dt=\alpha\left[\int f(\theta',\varphi')d\omega'/4\pi-f(\theta,\varphi)\right]=\alpha[N-f], \quad \text{(A1.15)}$$

---

[20] In his first paper on the H-theorem Boltzmann used E (entropy) for H.

with

$$\alpha = nc\sigma. \tag{A1.16}$$

From Eq. (A1.15), we get

$$f(\theta,\varphi) = N - Ae^{-\alpha t}, \quad A = A(\theta,\varphi) = N - f_{t=0}(\theta,\varphi), \tag{A1.17}$$

that is, an exponential approach to the equilibrium value, if at $t=0$ a nonequilibrium situation were present.

Once again, as in Eq. (A1.11) we see a uniform approach to the equilibrium situation. However, once again we have introduced a basic assumption in writing down Eq. (A1.14), and again this assumption is one regarding the number of collisions, or a so-called *Stosszahlansatz*. As collisions play the dominant role in our discussion, we call this method of treating the problem the *kinetical* treatment. We shall show, however, in the next section that the unrestricted *Stosszahlansatz* leads to contradictions and that we therefore rather should base our discussion on the so-called *statistical* treatment, which pays due regard to possible fluctuations.

(2) Reversibility and Recurrence Paradox

Let us restate the H-theorem in a slightly different way. We are considering systems made up of a large number of identical particles and we have discussed the distribution function, and its rate of change, corresponding to such a system. However, a different way of looking at the behavior of the system is by depicting its situation, or phase, at each moment by a representative point in Γ space (see ESM p. 100 for a definition of Γ space). Each situation which, on the one hand, corresponds to a given distribution will, on the other hand, correspond to one point in Γ space. We have thus a one-to-one correspondence between, on the one hand, points in Γ space, and, on the other hand, values of the distribution function.[21] Consider now a series of instants $\cdots, t_1, t_2, \cdots, t_n, \cdots$. The phase of our system at t_n will correspond to a point P_n in Γ space and the orbit described by the representative point of the system will pass through the sequence of points

$$\cdots, P_1, P_2, \cdots, P_{n-1}, P_n, P_{n+1}, \cdots. \tag{A2.01}$$

As each point corresponds to a distribution function, we can evaluate the value of H corresponding to each point. Let H_n be the value of H in the situation corresponding to P_n. If we accept the H-theorem in its unrestricted form, it then follows that the following inequalities should hold

$$\cdots \geqslant H_1 \geqslant H_2 \geqslant \cdots \geqslant H_{n-1} \geqslant H_n \geqslant \cdots, \tag{A2.02}$$

where the equal signs only hold, if equilibrium has been reached.

In the general case H is given instead of by Eq. (A1.07) by the equation

$$H = \int f \ln f d\omega, \tag{A2.03}$$

where the integration extends over the whole of μ space (ESM p. 30), where f is now a function of the s generalized coordinates q_k and the s generalized momenta p_k (s= number of degrees of freedom of one particle), and where $d\omega$ is a volume element in μ space,

$$d\omega = \prod_{k=1}^{s} dp_k dq_k. \tag{A2.04}$$

Consider now two situations corresponding to the representative points P_i and P_i' which differ only in that all the p_k have the same absolute values, but with opposite signs, while the q_k in the two situations are the same. Since the Hamiltonian $\mathcal{H}(p_k,q_k)$ of the system is a homogeneous quadratic polynomial in the p_k, it is invariant against the transformation $q_k \rightarrow q_k'$, $p_k \rightarrow -p_k'$, which is the transformation from P_i to P_i'. We see then from the canonical equations of motion,

$$\dot{q}_k = \partial\mathcal{H}/\partial p_k, \quad \dot{p}_k = -\partial\mathcal{H}/\partial q_k, \tag{A2.05}$$

that the transformation from P_i to P_i' corresponds to a reversal of the time axis. If we now consider a system going through the sequence (A2.01), then, on transformation, we obtain a system which goes through the sequence

$$\cdots, P'_{n+1}, P_n', P_{n-1}', \cdots, P_2', P_1', \cdots. \tag{A2.06}$$

From the definition of H, Eq. (A2.03), it follows easily that

$$H_i = H_i'. \tag{A2.07}$$

This can be seen by performing the transformation under the integral sign and then changing to new variables,

$$\begin{aligned} H' &= \int f'(p',q') \ln f'(p',q') d\omega' \\ &= \int f(-p,q) \ln f(-p,q) d\omega \\ &= \int f(p'',q'') \ln f(p'',q'') d\omega'' = H, \end{aligned} \tag{A2.08}$$

where $p_k'' = -p_k$, $q_k'' = q_k$.

From Eqs. (A2.07) and (A2.02), it follows that the sequence (A2.06) corresponds to a sequence of H-values which satisfy the inequalities

$$\cdots \leqslant H'_{n+1} \leqslant H_n' \leqslant H'_{n-1} \leqslant \cdots \leqslant H_2 \leqslant H_1 \leqslant \cdots. \tag{A2.09}$$

We see thus that for each system which shows a steady decrease of H we can construct a system for which H

---

[21] This is, of course, only true as long as we restrict ourselves to the discussion of the relatively unimportant specific phases. If we use generic phases to each distribution corresponds a set of points in Γ [compare Sec. A (3)].

is steadily increasing. This fact constitutes the so-called *Umkehreinwand* or *reversibility paradox* of Loschmidt's (76L, 77L).

We can easily see the difficulty by considering the simple model introduced at the end of Sec. A(1). By reversing the direction of time we get a steady departure from the equilibrium distribution as can be seen from Eq. (A1.17).

Another difficulty was pointed out by Zermelo (96Z) who used a theorem of Poincaré's (90P; see also 97B). Poincaré had shown that if a system enclosed in a finite volume passes through the sequence (A2.01) from t_1 to t_{n+1}, say, this sequence will be repeated as accurately as we wish it to be repeated after a finite time interval.[22,23] More specifically, for any finite length element Δs in Γ space we can find a time interval T such that the sequence at t_1+T, t_2+T, $\cdots$, t_n+T, $t_{n+1}+T$ will be given by the points

$$\cdots, P_1'', P_2'', \cdots, P_n'', P''_{n+1}, \cdots, \qquad \text{(A2.10)}$$

where

$$|P_i - P_i''| < \Delta s. \qquad \text{(A2.11)}$$

The time interval T can be extremely long. For the case of 10^{18} atoms within 10^{-6} m^3 moving with an average velocity of 5.10^6 m sec^{-1}, Boltzmann (96B, see also 43C) estimated that it would take more than $10^{10^{19}}$ years to reproduce the positions within 10A and the velocities within 10^4 m sec^{-1}.

If Δs is chosen sufficiently small, we have for the H-values corresponding to the sequence (A2.10)

$$H_i'' \approx H_i, \qquad \text{(A2.12)}$$

and hence we have the following sequence at t_1, t_2, $\cdots$, t_n, $\cdots$, t_1+T, $\cdots$, t_n+T:

$$\cdots, H_1, H_2, \cdots, H_n, \cdots, H_1'', H_2'', \cdots, H_n'', \cdots. \qquad \text{(A2.13)}$$

Using Eqs. (A2.02) and (A2.12) we see that in going from t_n to t_1+T we have an increase in H. This paradox is the so-called *Wiederkehreinwand* or *recurrence paradox*.

In the next section we shall show how statistical considerations can clarify the situation. However, we wish to emphasize at this moment that in so far as the reversibility and recurrence paradoxes invalidate the H-theorem in its unrestricted form, that is, the statement that H can *never* decrease, they show that the *Stosszahlansatz* cannot be true under all circumstances. We shall illustrate this by considering two simple cases.

The first example is drawn from the consideration of the simple model of the previous section. Consider an arbitrary time interval Δt during the evolution of the system and denote by $f(\theta,\varphi)$ the distribution function at the beginning of this time interval. Denote further once again by $N_{\omega,\omega'}\Delta t d\omega d\omega'$ the number of collisions leading during this time interval to the change of a velocity within the solid angle $d\omega$ to a velocity within the solid angle $d\omega'$. Consider now the corresponding evolution with the time direction reversed and consider the corresponding time interval. The distribution function at the beginning of the time interval in this second case may be denoted by $f'(\theta,\varphi)$. If $N'_{\omega',\omega}\Delta t d\omega d\omega'$ is the number of collisions leading from ω' to ω within Δt of the reversed evolution, we have

$$N'_{\omega',\omega} = N_{\omega,\omega'} \qquad \text{(A2.14)}$$

for any pair of directions ω and ω'. From Eq. (A1.14) it then follows that, *if the Stosszahlansatz were valid for both evolutions*, we would have

$$f(\theta,\varphi) = f'(\theta',\varphi'), \qquad \text{(A2.15)}$$

independent of θ, φ, θ', and φ', or

$$f(\theta,\varphi) = \text{constant}. \qquad \text{(A2.16)}$$

We see thus that *the Stosszahlansatz cannot be valid for at least one of the two evolutions*, unless we are dealing with the equilibrium situation.

The second example is the following one.[24] Consider a stream of point particles moving all in the same direction, say, the $+x$-direction and being scattered by a system of hard spheres which are randomly distributed over a plane perpendicular to the x-axis. The result of the scattering will be that the point particles after their encounter with the plane of spheres will form a system of particles the velocities of which are isotropically distributed. Consider now the evolution of the system obtained from the previous system by reversing the direction of the velocities of all the particles. Clearly after a while all the particles will be moving in the $-x$-direction, that is, the system will proceed from a disordered condition to a highly ordered pattern. The reason for this can easily be seen when we consider the collisions of the particles of the second system with the spheres. Although the system seemed to be completely random, all collisions with the spheres occur on that half of the spheres which is turned to the $-x$-axis and clearly the collisions are not taking place in a random manner. For this second system we see thus that the *Stosszahlansatz* would certainly not be satisfied.

(3) Statistical Aspects of the *H*-Theorem

We mentioned before that the nature of the situations studied by statistical mechanics is such that we are led to statistical considerations. The first man to use a truly statistical approach was Boltzmann (77B) and at that point kinetic theory changed into statistical

---

[22] We refer to the literature (96Z, 43C, ESM p. 341, which latter proof is not very rigorous) for a proof of Poincaré's theorem. The proof follows essentially from the fact that a finite region in Γ space will sweep through the available part of phase space, which is finite in the case of a system enclosed inside a volume, within a finite period.

[23] For the relationship between Poincaré's theorem and Birkhoff's ergodic theorem we refer to a monograph by Wintner (41W, pp. 90–91).

[24] I am greatly indebted to Dr. H. M. James for providing me with this example.

mechanics even though it was another twenty odd years before Gibbs coined the expression.

In his statistical considerations Boltzmann was concerned first of all with the answer to question (A1), especially for the case of an isolated system, and he suggested two possible definitions of an equilibrium situation.

(a) The equilibrium situation is that one which is the *most probable* one for a given energy.

(b) The equilibrium situation is the *average* situation in which the system will find itself during an infinite time interval.

Let us first in this section consider the approach via the most probable situation. In order to prove that the most probable situation corresponds to the equilibrium situation characterized in the case of a system of independent particles, such as Boltzmann was considering, by the Maxwell-Boltzmann distribution we proceed as follows.

Consider an isolated system of N independent particles. Each particle will have a representative point $Q^{(k)}$ in μ space, where k numbers the particles in the system ($k=1, \cdots, N$). The phase of the whole system, and thus its representative point in Γ space, is then determined by the N points $Q^{(1)}, \cdots, Q^{(N)}$. We now divide μ space into very small *but finite* cells of equal volume ω, which we number consecutively, $\omega_1, \omega_2, \cdots, \omega_i, \cdots$. *This division is one of the fundamental steps in the argument.* The size of ω is such that, on the one hand, the dimensions of a cell are small compared to the smallest macroscopically measurable dimensions, but, on the other hand, the number of representative points $Q^{(k)}$ contained in each of them is large.[25]

Let now N_i be the number of representative points in the ith cell. The situation $\mathbf{Z}$[26] is then completely described by giving the N_i. The description of the situation by the N_i instead of by the actual position of the $Q^{(k)}$ corresponds to our experimental limitations, since the size of the cells has been chosen such that we cannot distinguish by experimental means between different points within one cell.

The relation between the representative point of the system in Γ space and the situation $\mathbf{Z}$ described by the N_i is as follows.

(a) To each point in Γ space corresponds one situation $\mathbf{Z}$.

(b) For each $\mathbf{Z}$ there exists a volume in Γ space such that each point of this volume corresponds to the same $\mathbf{Z}$. Such a region we shall call a $\mathbf{Z}$-star, and its volume is given by the equation

$$W(\mathbf{Z})=[N!/\Pi_i N_i!]\omega^N. \tag{A3.01}$$

Equation (A3.01) can be derived if we remember that the collective of the numbers N_i remains unchanged, and hence $\mathbf{Z}$ remains the same, under the following two operations. (a) Each of the N points $Q^{(k)}$ sweeps through its own cell, resulting in the filling out of a volume $\Omega=\omega^N$ in Γ space. (b) Any permutation of the N points such that only points in different cells are permuted.

We now introduce a quantity $H(\mathbf{Z})$ by the equation

$$H(\mathbf{Z})=-\ln W(\mathbf{Z}). \tag{A3.02}$$

Substituting expression (A3.01) for $W(\mathbf{Z})$ into (A3.02), using Stirling's formula for the factorial in the form

$$\ln x!=x\ln x-x \tag{A3.03}$$

and neglecting additive constants, we get

$$H(\mathbf{Z})=\Sigma_i N_i \ln N_i. \tag{A3.04}$$

Comparing Eqs. (A2.03) and (A3.04) we see that $H(\mathbf{Z})$ is the same as Boltzmann's H with the proviso that instead of integrating we take a sum over the cells which we introduced in μ space.

It can easily be shown (e.g., ESM Sec. I.7) that, if we define the equilibrium distribution N_i^e as that distribution for which $W(\mathbf{Z})$ is maximum, it follows that in the case where the total energy E and the total number of particles N is fixed, corresponding to the restricting conditions

$$N=\Sigma N_i, \tag{A3.05}$$

$$E=\Sigma N_i E_i, \tag{A3.06}$$

where E_i is a representative energy for the ith cell, the N_i^e obey the equation

$$N_i^e=\exp(\mu-\beta E_i), \tag{A3.07}$$

which is the so-called Maxwell-Boltzmann distribution. In identifying the equilibrium distribution with the distribution for which $W(\mathbf{Z})$ is maximum, we have defined the probability for a situation as the corresponding volume in Γ space.

In order to prove that the equilibrium distribution is the one for which $W(\mathbf{Z})$ is maximum, or $H(\mathbf{Z})$ minimum, we should like to introduce an H-theorem, that is, show that dH/dt is always negative, unless $H=H_{\min}$, or something of this kind. However, as soon as we start discussing the behavior of $H(\mathbf{Z})$ as a function of time, we realize that any function which depends on the coordinates and momenta of the particles in the system through the N_i will be a discontinuous function, since whenever one of the representative points in μ space leaves one cell and enters another, the two N_i concerned alter by unity. For the time dependence of $H(\mathbf{Z})$, we

---

[25] This division seems on first sight artificial in classical statistics, while it occurs naturally in quantum statistics. Uhlenbeck (27U) discussing this step remarks: "It is as if Boltzmann had a premonition of the occurrence of discrete quantum states in μ space." There is, however, more to it than this and to a large extent one may say that even in classical statistics the division is natural as it corresponds to the limitations of our macroscopic measurements. Compare in this connection the discussion of the coarse grained ensemble density in Sec. C(1) and the discussion at the end of the present section.

[26] We use here the Ehrenfests' notation (11E2): $\mathbf{Z}$ for *Zustandsverteilung*.

get thus a step function[27] and its time derivative will only have three possible values, $-\infty$, 0, and $+\infty$.[27a]

From the step function we now obtain the so-called *H-curve* by selecting a discrete set of points which are separated by constant time intervals. The length of the time interval τ, say, is chosen in such a way that it is small compared to experimental time intervals, but sufficiently large so that during τ a large number of collisions occur.[28]

We can now use the properties of the H-curve to illustrate the following points.[29] We shall use the simple model introduced in Sec. A(1) for this illustration. It can be shown (54G, 55G) that this model is only one of a larger class of models all showing the same behavior.[30]

I. It is possible to find a function Δ which, on the one hand, measures the departure from equilibrium and, on the other hand, is directly related to the entropy of the system, or H.[31]

II. If the history of a system is followed in time, the equilibrium distribution, for which Δ is equal to its minimum value, will be realized much more often than any other distribution. Furthermore, the chance of observing another distribution is so small that the equilibrium distribution is the same as the average distribution.

III. If the system is in a situation corresponding to $\Delta > \Delta_{\min}$, the development of the system will be such that Δ will most probably decrease.

IV. The previous statement will be true whether we read the Δ curve—which is obtained from the step function $\Delta(t)$ in the same way as the H-curve is obtained from $H(t)$—from $t=-\infty$ to $t=+\infty$, or from $t=+\infty$ to $t=-\infty$.

V. The Δ curve will practically always be in the neighborhood of $\Delta_{\min}$.

VI. The time between the reoccurrence of a Δ value different from $\Delta_{\min}$ increases steeply with increasing Δ.

VII. If $\mathfrak{D}_n$ is the average of all Δ values at $t_0+n\tau$, starting from Δ_0 at t_0,[32] the $\mathfrak{D}$-curve, that is, the sequence of values $\mathfrak{D}_1, \mathfrak{D}_2, \cdots, \mathfrak{D}_n, \cdots$ will decrease monotonically from Δ_0 approaching $\Delta_{\min}$ asymptotically.

VIII. By far the most of the Δ curves will follow the $\mathfrak{D}$-curve for an appreciable period, but practically none of them will follow the $\mathfrak{D}$-curve at all times.

We shall call the sequence of values $\Delta_1', \Delta_2', \cdots, \Delta_n', \cdots$, which would follow from Δ_0 by an application of the unrestricted H-theorem based on the *Stosszahlansatz*, the *Stosszahl* curve.[33] We then have the following point.

IX. The *Stosszahl* curve and the $\mathfrak{D}$-curve are identical.

Let us now consider the Lorentz model. The function Δ mentioned in I will be defined by the equation

$$\Delta = \sum_{v=-m}^{+m} (f_v - f_v{}^e)^2, \tag{A3.08}$$

where the $f_v{}^e$ are defined by Eq. (A3.11). In Eq. (A3.08) we have introduced the division of μ space—which in this case reduces to the unity sphere—into finite cells. Each cell is an element of solid angle of extension $\delta\omega$, where

$$\delta\omega = 4\pi/(2m+1). \tag{A3.09}$$

The $2m+1$ expressions f_v which are functions of time and which together determine the state of the system are defined in such a way that f_v is the number of electrons moving in the direction specified by the vth element of solid angle. Only $2m$ of the f_v are independent as they satisfy the relation

$$\sum_{v=-m}^{+m} f_v(t) = N. \tag{A3.10}$$

The equilibrium values $f_v{}^e$ of the f_v are all equal and given by the equation

$$f_v{}^e = N/(2m+1). \tag{A3.11}$$

As one can show that the f_v are practically always in the neighborhood of $f_v{}^e$ [compare Eq. (A3.16)] so that we can assume $f_v - f_v{}^e \ll f_v{}^e$ we have up to second-order terms in $(f_v - f_v{}^e)/f_v{}^e$

$$H = \Sigma_v f_v \ln f_v = \Sigma_v f_v{}^e \ln f_v{}^e + \Delta/2N = H_{\text{eq}} + \Delta/2N. \tag{A3.12}$$

From Eqs. (A3.08) and (A3.12) we see first of all that it makes no difference whether we use H or Δ to describe the departure from equilibrium, and secondly that Δ is a non-negative function of the f_v which is only zero when all the f_v are equal to their equilibrium values.

As far as point II is concerned, we can first of all calculate the relative occurrence of a state of the system corresponding to a nonequilibrium value of Δ. We find in this way for the normalized probability $w(\Delta)d\Delta$ that Δ has a value in the interval $(\Delta, \Delta+d\Delta)$ the expression[34]

$$w(\Delta)d\Delta = [(2m+1)/2N]^m [\Delta^{m-1}/(m-1)!] \times \exp\{-(2m+1)\Delta/2N\} d\Delta. \tag{A3.13}$$

---

[27] It must be noted that this fact does not depend on the character of the intermolecular forces but is solely a consequence of the introduction of finite cells ω.

[27a] More mathematically expressed: the time derivative is everywhere zero except where H is discontinuous when it is not defined.

[28] Compare the choice of the size of the cells in μ space.

[29] The numbering of the points, and the contents of some of them, is different from that of the Ehrenfests (11E2) or that in ESM.

[30] See also 51B2 and 52H.

[31] We have not proved that apart from additive and multiplying constants, H given by Eq. (A2.03) is the entropy of the system. For such a proof we refer to the literature (see, e.g. ESM pp. 22 and 43).

[32] In considering points VII and VIII (and also to some extent point III), it must be borne in mind that H and thus Δ does not determine the situation uniquely. For a given value of Δ there are thus several possible situations $\mathbf{Z}$.

[33] In the Ehrenfests' discussion where H was considered instead of Δ, this curve is called the H-theorem curve (see also 54G, 55G).

[34] For a detailed discussion and derivation of the formulas in the remainder of this section, we refer to Appendix I.

The function $w(\Delta)$ takes the place of the $W(\mathbf{Z})$ of the beginning of the section and we see that, indeed, $w(\Delta)$ reaches its maximum value for $\Delta=0$, that is, for the equilibrium situation.

We now make the assumption that the time $t(\mathbf{Z})$ spent in a situation described by $\mathbf{Z}$ will be proportional to $W(\mathbf{Z})$.[35] In that case the time average $\bar{G}$ of a phase function G will be given by the equation

$$\bar{G}=[\Sigma_{\mathbf{Z}}G(\mathbf{Z})W(\mathbf{Z})]/[\Sigma_{\mathbf{Z}}W(\mathbf{Z})]. \qquad \text{(A3.14)}$$

The assumption which we have just made about the proportionality of $t(\mathbf{Z})$ and $W(\mathbf{Z})$ is necessary in order to continue the argument. It is also an assumption which is closely related to assumptions made in such theories as those which discuss the outcome of lotteries. In the latter case since each draw is supposedly independent of all the preceding ones, the proportionality of the probability for a certain result to the number of times such a result occurs in an infinite series of draws is fundamental and, indeed, used to define the probability. However, when one is dealing with processes which are in fact determined by the preceding history, the proportionality of time spent in a situation and the probability for this situation must be assumed. Moreover, it cannot be used to define the probability since the processes considered are not even Markovian. The difficulty lies in the fact that $t(\mathbf{Z})$ is concerned with the long-time behavior of a system, but as the orbit is completely determined once one point of it is known in Γ space, the longer we follow the orbit the more data we obtain from which to determine completely the orbit and hence we are no longer dealing with purely statistical situations. We may refer in this connection to the discussion given by Tolman (p. 148 of 38T; see also 50B). Our assumption is thus closely connected with the Markovization of the processes giving the development of a system (compare 49S, 54G, 55G, 55H2).

We must also draw attention to the fact that the proportionality of $t(\mathbf{Z})$ and $W(\mathbf{Z})$ is *not* proved by the ergodic or quasi-ergodic theorem—as one might have hoped—since this theorem only proves the proportionality of $t(\mathbf{Z})$ and $W_E(\mathbf{Z})$, but not the proportionality of $W_E(\mathbf{Z})$ and $W(\mathbf{Z})$. It remains thus one of the basic assumptions in the approach to the H-theorem in the form in which we discuss it at the moment and, moreover, one which seems to be no nearer to a proof notwithstanding the vast amount of work done either on simplified models or on the ergodic theorem.

To continue our discussion of points II to IX, we shall confine ourselves mainly to those phase functions which depend on the p's and q's only through Δ. For such functions Eq. (A3.14) can be written in the form

$$\bar{G}=\int w(\Delta)G(\Delta)d\Delta, \qquad \text{(A3.15)}$$

since $w(\Delta)$ is normalized. We can, for instance, calculate the average value of Δ and the dispersion around this average value, and we find

$$\Delta_{\text{Av}}=N, \quad \langle(\Delta-\Delta_{\text{Av}})^2\rangle_{\text{Av}}=N^2/m. \qquad \text{(A3.16)}$$

In order to discuss the variation of Δ with time, we introduce the time interval τ and calculate the probability $w(\Delta,\Delta')$ that Δ changes its value from Δ to Δ' during this time interval τ. Once these transition probabilities are obtained, we find the average value of Δ' after τ when at the beginning of the interval the value was Δ from the equation

$$\Delta'_{\text{Av}}=\int w(\Delta,\Delta')\Delta' d\Delta', \qquad \text{(A3.17)}$$

provided $w(\Delta,\Delta')$ is normalized with respect to Δ', and for the average rate of change of Δ [and thus of H; compare Eq. (A3.12)] we have

$$(d\Delta/dt)_{\text{Av}}=(\Delta_{\text{Av}}'-\Delta)/\tau. \qquad \text{(A3.18)}$$

In the special case of the Lorentz model, we find for $w(\Delta,\Delta')$ the expression (see Appendix I)

$$w(\Delta,\Delta')=(16\pi NA\Delta)^{-\frac{1}{2}} \times\exp\{-[\Delta'-\Delta+A\Delta(2m+1)]^2/16NA\Delta\}, \qquad \text{(A3.19)}$$

where A is given by the equation

$$A=2\alpha\tau, \qquad \text{(A3.20)}$$

with α given by Eq. (A1.16). From Eqs. (A3.17) to (A3.19) we now get

$$\Delta_{\text{Av}}'=(1-A)\Delta, \qquad \text{(A3.21)}$$

$$(d\Delta/dt)_{\text{Av}}=-A\Delta/\tau. \qquad \text{(A3.22)}$$

Point III follows immediately from Eq. (A3.22). We see here the same exponential approach to equilibrium as was found in Sec. A(1) (Eq. (A1.17)). We shall see presently the close connection which exists between Eqs. (A1.15) and (A3.22).

In order to prove point IV, we must calculate the average value Δ''_{start} of Δ which after a time interval τ leads to a value Δ. This quantity is given by the equation

$$\Delta''_{\text{start}}=\left[\int \Delta'' w(\Delta'')w(\Delta'',\Delta)d\Delta''\right]\Big/\left[\int w(\Delta'')w(\Delta'',\Delta)d\Delta''\right], \qquad \text{(A3.23)}$$

where the denominator arises from the fact that it is not equal to unity [$w(\Delta'')w(\Delta'',\Delta)$ is only normalized

---

[35] We do not wish to stress here the point that this assumption really involves two steps (see ESM p. 350). The first step is the assumption that the area $W_E(\mathbf{Z})$, cut out of the energy surface on which the representative point of our system is situated, will be proportional to $W(\mathbf{Z})$. The second step consists in assuming $t(\mathbf{Z})$ to be proportional to $W_E(\mathbf{Z})$. It may be mentioned here that Einstein (03E, 10E) made the same (combined) assumption as we are making here.

if integrated over both Δ and Δ'']. From Eqs. (A3.23), (A3.13), and (A3.19), we find

$$\Delta''_{\text{start}}=(1-A)\Delta, \tag{A3.24}$$

and point IV has been proven.[36]

Point V follows from the expression we found for $w(\Delta)$ and the assumption that the chance of finding a value of during a sequence is proportional to $w(\Delta)$. In order to see point VI, we use a formula of Chandrasekhar's (43C) for the recurrence time $\Theta(\Delta)$ of a state characterized by Δ,[37]

$$\Theta(\Delta)=\tau[1-w(\Delta)]/[w(\Delta)-w(\Delta)w(\Delta,\Delta)], \tag{A3.25}$$

or

$$\Theta(\Delta)\cong\tau/w(\Delta), \tag{A3.26}$$

from which point VI follows [see Eq. (A3.13)].

In order to calculate $\mathfrak{D}_n$, that is, the average of all Δ values at $t_0+n\tau$, starting from Δ_0 at t_0, we must compute first of all the normalized probability $w(\Delta_n;\Delta_0,\Delta_1,\Delta_2,\cdots,\Delta_{n-1})$ for a sequence of values Δ_0 at t_0, Δ_1 at $t_0+\tau$, Δ_2 at $t_0+2\tau$, $\cdots$, Δ_{n-1} at $t_0+(n-1)\tau$, and Δ_n at $t_0+n\tau$. This quantity is given by the equation (compare 55G)[38]

$$w(\Delta_n;\Delta_0,\Delta_1,\Delta_2,\cdots,\Delta_{n-1}) =w(\Delta_0,\Delta_1)w(\Delta_1,\Delta_2)\cdots w(\Delta_{n-1},\Delta_n). \tag{A3.27}$$

For $\mathfrak{D}_n$ we now get

$$\mathfrak{D}_n=\int \Delta_n w(\Delta_n;\Delta_0,\Delta_1,\Delta_2,\cdots,\Delta_{n-1}) \times d\Delta_1 d\Delta_2\cdots d\Delta_n, \tag{A3.28}$$

from which follows that

$$\mathfrak{D}_n=(1-A)^n\Delta_0, \tag{A3.29}$$

and putting $t=n\tau$ we get

$$\mathfrak{D}_n\cong\Delta_0 e^{-At/\tau}=\Delta_0 e^{-2\alpha t}. \tag{A3.30}$$

Point VII follows immediately from Eqs. (A3.29) or (A3.30).

We can also calculate the dispersion σ_n of Δ_n for which we have the equation

$$\sigma_n=\int (\mathfrak{D}_n-\Delta_n)^2 w(\Delta_n;\Delta_0,\Delta_1,\cdots,\Delta_{n-1}) \times d\Delta_1\cdots d\Delta_n, \tag{A3.31}$$

or

$$\sigma_n=0. \tag{A3.32}$$

From Eq. (A3.32) and the fact that nonequilibrium values of Δ recur with a frequency given by $\Theta(\Delta)^{-1}$, point VIII follows.

Let us now consider finally the *Stosszahl* curve. We note that from Eq. (A3.08) it follows that

$$d\Delta/dt=\Sigma_v 2(f_v-f_v{}^e)(df_v/dt). \tag{A3.33}$$

Provided the *Stosszahlansatz* is valid, the rate of change of f_v with time is given by the equation [compare Eq. (A1.15); the proof of Eq. (A3.34) is given in Appendix I]

$$df_v/dt=[A/\tau][f_v{}^e-f_v]. \tag{A3.34}$$

Combining Eqs. (A3.33) and (A3.34), we get for the *Stosszahl* curve the differential equation

$$d\Delta/dt=-A\Delta/\tau, \tag{A3.35}$$

and we see from comparing Eqs. (A3.30) and (A3.35) that point IX is also proved for the model considered.

Let us now summarize the situation as regards the present position of the H-theorem in classical statistics. We notice, first of all that, although the situation has improved by the study of simplified models since the Ehrenfests (11E2) considered this problem in great detail, much remains to be done. As far as the approach to equilibrium and the fluctuations around the equilibrium situation are concerned, the study of simplified models has shown that the situation is much as one would have expected (points II to VI)—provided we may treat the process as a true Markov process. We have discussed the behavior of the system mainly from the point of view of the fluctuations, but we could have done the same for the return to equilibrium. Siegert (49S) has, for instance, shown that starting from any Δ at $t=t_0$, the probability for finding a value of Δ at $t=\infty$ will always be given by $w(\Delta)$. We have tacitly assumed that we are always dealing with a situation which has essentially forgotten its initial state. This may be done provided the relaxation time of the system is sufficiently short. The relaxation time t_{rel} is essentially given by τ/A which, if we assume $n=10^{30}$ m^{-3}, $c=10^4$ m sec^{-1}, $\sigma=10^{-20}$ m^2, $m=10^6$, leads to $t_{\text{rel}}\sim10^{-8}$ sec (compare ESM p. 390). This fast relaxation accompanied by long recurrence times for appreciable deviations from equilibrium[39] are the reasons why there are very few exceptions to the second law of thermodynamics, or, to put it slightly differently, why fluctuations very seldom play an important role in the behavior of macroscopic physical systems. As Smoluchowski (12S) puts it, a system will appear to behave irreversibly if its initial state is characterized by an average time of recurrence which is long compared to the period available to the experimenter. In this connection, we also see confirmed Tolman's suggestion (38T, p. 179) that the relaxation time would be very short compared to the period of the Poincaré cycle (Sec. A(2)), provided the size of the elementary cells

---

[36] It must be remarked here that in an earlier paper (53H) we used for the proof of the equivalence of the sequence with $t=-\infty\rightarrow t=+\infty$ and the sequence with $t=+\infty\rightarrow t=-\infty$ the relation $w(\Delta)w(\Delta,\Delta')=w(\Delta')w(\Delta',\Delta)$ which holds also in the present model.

[37] The objections raised by Bartlett (50B, 53B) against Chandrasekhar's formula are no longer valid, if we consider the sequences which are continually Markovized (see 50B).

[38] It may be noted that in writing down Eq. (A3.27) use is made of the assumption of continued Markovization.

[39] For an average deviation of 10^{-8} $f_v{}^e$ from $f_v{}^e$, we get $\Theta(\Delta)$ of the order of 10^{10^9} years!

were not too large. Since A is inversely proportional to $2m+1$, the number of the cells, we see indeed that t_{rel} increases with m, and we saw a moment ago that t_{rel} took on a reasonable value for a reasonable value of m.

The main problem still remaining is the justification of the Markovization (point VII of the Ehrenfests (11E2), or point IX of ESM). This seems an extremely difficult problem to tackle, but it might be possible, certainly with the powerful calculating machines available these days, to calculate the H-curve of a given system for a variety of initial conditions using the exact equations of motion.

In conclusion we wish to point out that even though to some extent the ergodic, or better, the quasi-ergodic theorem has solved the question of the equivalence of time averages and ensemble averages, because of the incompleteness and inexactness of available experimental data we can only define a situation up to a limited accuracy. Even if we are dealing with an isolated system so that its representative point will stay on an energy surface, we do not know on which energy surface, and the only truly representative ensemble will be one which consists of all the systems whose representative points at t_0 fill up a **Z**-star. Hence, the discussion of the H-theorem still remains germane to the problem we are discussing.

B. THE CLASSICAL ERGODIC THEOREM

(1) The Ergodic and Quasi-Ergodic Theorem

We mentioned in the Introduction that many authors prefer the approach via the ergodic theorem to the one using the H-theorem. At the end of Sec. A(3) we mentioned our reasons against this point of view and we shall return to this discussion in Part C. However, the ergodic theorem has played such an important role in the development of the subject of statistical mechanics that the omission of a discussion of this problem in an article such as the present one would be unforgivable.

We mentioned at the beginning of Sec. A(3) that Boltzmann considered two different definitions of an equilibrium situation, namely, as the most probable situation or as the average situation. We considered the first possibility and we shall now consider the second one. Boltzmann (71B1, 71B2; see also 68B) and Maxwell (79M) thought that one could prove in a more or less straightforward manner the theorem: The average behavior of a system of independent particles will correspond to the Maxwell-Boltzmann distribution. The average is taken in the sense of a time average over an infinite period.

The reason for believing in this theorem was that Boltzmann expected the existence of so-called *ergodic* systems. An ergodic system is one such that the orbit of its representative point in Γ space will go through every point of its energy surface. Once the existence of ergodic system is accepted one can easily see the advantage of considering time averages. First of all it has to be borne in mind that one always measures time averages in any physical experiment. Moreover, as relaxation times are usually small as compared to the period over which an experimentally determined quantity is measured, it does not matter whether we take the averages over a finite interval or over an infinite period. Secondly, if one could prove that actual physical systems were ergodic the following equalities would hold for any phase function $\varphi(p,q)$:

$$\langle\varphi\rangle_{\mathrm{Av}}=\overline{\langle\varphi\rangle_{\mathrm{Av}}}=\langle\bar{\varphi}\rangle_{\mathrm{Av}}=\bar{\varphi}, \tag{B1.01}$$

where the bar (—) indicates a time average and $\langle\,\rangle_{\mathrm{Av}}$ an average taken over a microcononical ensemble. The second and third terms in Eq. (B1.01) indicate the time average of the ensemble average and the ensemble average of the time average, respectively. We shall presently indicate a proof of Eq. (B1.01), but at the moment only remark that from these equations it follows that one can calculate instead of time averages ensemble averages and, since those can be calculated much more easily, we have a clear gain.

The time average of $\bar{\varphi}$ of a phase function is defined by the equation

$$\bar{\varphi}=\lim_{T\to\infty}\frac{1}{2T}\int_{t-T}^{t+T}\varphi(p,q)dt, \tag{B1.02}$$

where p's and q's are the values at the time t of the coordinates of the representation point of the system is Γ space. In order to see whether or not the average behavior of one system corresponds to a Maxwell-Boltzmann distribution, one selects a set of phase functions which together determine the distribution and compares their time averages with the values to be expected for a Maxwell-Boltzmann distribution.

The ensemble average $\langle\varphi\rangle_{\mathrm{Av}}$ is taken over a microcanonical ensemble. This average is given by the equation

$$\langle\varphi\rangle_{\mathrm{Av}}=\left[\int\varphi(p,q)\sigma(p,q)d\mathcal{S}\right]\bigg/\left[\int\sigma(p,q)d\mathcal{S}\right], \tag{B1.03}$$

where the two integrations extend over the whole of the energy surface $\mathcal{S}$, and where $\sigma(p,q)$ is the surface density on $\mathcal{S}$ corresponding to a microcanonical ensemble. This surface density is given by the equation[40]

$$\sigma(p,q)=\left\{\sum_{k=1}^{S}\left[\left(\frac{\partial\mathcal{H}}{\partial p_k}\right)^2+\left(\frac{\partial\mathcal{H}}{\partial q_k}\right)^2\right]\right\}^{-\frac{1}{2}}, \tag{B1.04}$$

where S is the number of degrees of freedom of the system, and thus the number of dimensions of Γ space, and $\mathcal{H}$ is the Hamiltonian of the system.

From Liouville's theorem (ESM p. 102) it follows

---

[40] The expression inside the braces in Eq. (B1.04) is nothing but the absolute value of the "velocity" of the representative points in Γ space.

(see, e.g., ESM p. 355) that the only stationary ensembles on an energy surface are those for which the surface density along an orbit satisfies Eq. (B1.04).

We can now proceed to the discussion of Eqs. (B1.01). The first equality follows immediately from the fact that a microcanonical ensemble is a stationary ensemble. The second equality is due to the interchangeability of averages. There remains the last equality which holds for ergodic system, since $\bar{\varphi}$ is the same for all systems in the ensemble. This can be seen as follows. Since for an ergodic system the orbit of the representative point passes through every point of $\mathcal{S}$, and since it follows from the canonical equations of motion (A2.05) that in each point the direction of the orbit is uniquely determined, therefore there is only one orbit on $\mathcal{S}$. The only difference for different systems in the ensemble lies in the exact moment at which a point of the orbit is passed.[41] As all systems follow the same orbit, it follows from Eq. (B1.02) that $\bar{\varphi}$ is the same for all systems and the last equality of Eq. (B1.01) follows.

For an ergodic system one can even go one step further and prove that the following equation holds (71B):

$$\lim_{T\to\infty}\frac{dt}{T}=\frac{\sigma d\mathcal{S}}{\int \sigma d\mathcal{S}}, \qquad \text{(B1.05)}$$

where dt is the time spent by the representative point within a surface element $d\mathcal{S}$ of $\mathcal{S}$ during the period T. From Eq. (B1.05) we can, in principle, calculate the frequency of the occurrence of nonequilibrium situations.

The problem is now reduced to a proof of the ergodicity of mechanical systems. However, one can easily see that it is unlikely that any system is ergodic (38T, p. 67). We remind ourselves that the system of Eqs. (A2.05) possesses $2S$ integrals of motion. One of these is the energy and one fixes the time along the orbit. By choosing a point P on the energy surface, we fix the values of the $2S-2$ other constants of motion. Choosing different values for these constants, but the same value for the energy, we find another point Q on the same energy surface, but there is no orbit connecting P with Q. In order to complete the proof one should, however, show that the $2S-2$ constants of motion are not constant almost everywhere on the energy surface—which seems unlikely.

We enter the realm of measure theory[42] with the expression "almost everywhere," and it is by using this theory that one can prove the nonexistence of ergodic systems (13R, 13P; see also 52R). The crux of this proof is that the points of an orbit form a set of measure zero on the energy surface, the measure of which is not zero.[43] We can see this as follows (52R). Consider a point P of the orbit of the representative point and consider a small region $\mathcal{A}$ of finite measure around P. As the orbit can clearly not stay indefinitely in $\mathcal{A}$, if it is to cover $\mathcal{S}$ completely, the only part of the orbit in $\mathcal{A}$ will be a collection of separate segments,[44] corresponding to a succession of finite intervals of time. As the set of such time intervals is enumerable, the segments of the orbit inside $\mathcal{A}$ will also be enumerable and thus have measure zero.

Although the impossibility of ergodic systems was only proved by Rosenthal and Plancherel in 1913, the Ehrenfests (11E2, especially footnote 89a; see also 91K, p. 484) were apparently well aware of the improbability of ergodic systems and of the fact that the points of an orbit and all the points of an energy surface have different measures. In order to avoid this difficulty it was suggested that mechanical systems might be *quasi ergodic*. The representative point of a quasi-ergodic system[45] will come arbitrarily near every point of the energy surface without, however, passing through each point.

The introduction of the possibility of quasi-ergodic systems did not, however, solve the problem how to prove Eq. (B1.01), although Rosenthal (14R2; see, however, reference 14) gave a not completely rigorous proof of Eq. (B1.05) based on the quasi ergodicity of mechanical systems and also proved the equivalence of time and ensemble averages. Fermi (23F1) proved that a certain class of systems, the so called *Kanonische Normalsysteme*,[46] were quasi ergodic. A sketch of his proof is given in Appendix II.

In order to advance it was necessary to find an independent proof of the equivalence of time and ensemble averages and that is the purpose of more modern developments of the ergodic theorem which are discussed in the next section. The term *ergodic theorem* is nowadays used to indicate a proof of the equivalence of time and ensemble averages, and the term quasi-ergodic theorem would probably have been more relevant from a historical point of view.

(2) Recent Developments[47]

Modern ergodic theory started with the paper by Birkhoff (31B2) who considered the following time

---

[41] We have excluded exceptional orbits which are strictly periodic.

[42] It may be noted here that many mathematical statisticians (see, e.g., 53D) consider statistics purely as a branch of measure theory. For a discussion of measure theory we refer to a recent text book by Halmos (50H; see also 53D).

[43] The measure of a point set $\mathcal{M}$ on $\mathcal{S}$ is defined as follows: Let $f(P)$ be a function which is 1, if P belongs to $\mathcal{M}$ and to 0 otherwise. The Lebesgue integral $\int fd\mathcal{S}$ extending over the whole of $\mathcal{S}$ is then called the Lebesgue measure of $\mathcal{M}$ on $\mathcal{S}$ and denoted by $\mathfrak{M}\,\mathcal{M}$.

[44] They have to be separate, as for a nonperiodic orbit no point of $\mathcal{S}$ is ever passed twice.

[45] It may be noted here that in view of the impossibility of ergodic systems in the original sense of the word, modern authors often refer to quasi-ergodic systems as ergodic systems.

[46] It is interesting to note that Einstein (02E, 03E) when he, independently of Gibbs, introduced ensemble theory, restricted himself to these systems (see also 11K2).

[47] See also 20M1, 20M2, 27H2.

averages:

$$\bar{f}(P;t_0,T)=\frac{1}{T}\int_{t_0}^{t_0+T} f(P_t)dt, \qquad \text{(B2.01)}$$

where P_t is the point passed at t of an orbit which passed through P at t_0. He showed, first of all, that for practically all points P of the energy surface $\mathcal{S}$ the limit of $\bar{f}(P;t_0,T)$ exists for $T\rightarrow\infty$. Moreover, Birkhoff showed that $\bar{f}(P;t_0,T)$ in the limit of $T\rightarrow\infty$ is independent of t_0 for practically all points P. The proof of this property of time averages on an energy surface is given in Appendix III. "Practically all" is understood here, as surmized by Birkhoff in 1922 (22B; see also 26S2), in the sense that the exceptional orbits would form a set of measure zero on $\mathcal{S}$.

Secondly, Birkhoff showed that $\bar{f}(P;\infty)$, i.e., the time average taken over an infinite period, is constant almost everywhere on $\mathcal{S}$, provided the group of transformations $P\rightarrow P_t$ is metrically transitive, or in other words, provided $\mathcal{S}$ is metrically indecomposable.[47a] The constancy of $\bar{f}(P;\infty)$ follows from the fact that, if it were not fulfilled, one could find a value F of $\bar{f}(P;\infty)$ such that the conditions $\bar{f}(P;\infty)<F$ and $\bar{f}(P;\infty)\geqslant F$ would define two sets of positive measure on $\mathcal{S}$ which would both be invariant against the transformations $P\rightarrow P_t$.[48]

Birkhoff went even further and proved the equivalent of Eq. (B1.05), namely, that for practically all points P on $\mathcal{S}$

$$\lim_{T\rightarrow\infty}\frac{dt(P;\mathcal{M})}{T}=\frac{\mathfrak{M}\mathcal{M}}{\mathfrak{M}\mathcal{S}}, \qquad \text{(B2.02)}$$

where $dt(P;\mathcal{M})$ is the fraction of time a representative point of the orbit going through P spends in the region $\mathcal{M}$ during a period T. Once again Eq. (B2.02) is only true provided the group of transformations $P\rightarrow P_t$ is metrically transitive.[49]

Birkhoff's proof is slightly stronger than the one given by von Neumann (32N1, 32N2) who showed that the convergence of $\bar{f}(P;t_0,T)$ was true in the mean, while Birkhoff showed that it was true for practically all P.

The problem of the ergodic theorem is now reduced to the problem of proving that in general energy surfaces are metrically indecomposable. In 1941 Oxtoby and Ulam (41O) proved this for a quite general class of surfaces, which were polyhedra of dimension three or more. They thus showed, in their own words, "that the ergodic hypothesis in its modern form of metrical transitivity is at least free from any objection on topological grounds." According to Gamow (49K1, p. 54; the italics are Gamow's) this implied "that in a certain sense *almost every* continuous transformation is metrically transitive." Gamow unfortunately does not give a justification for his statement. However, the energy surfaces considered by Oxtoby and Ulam are not quite physical systems, and it has often been felt therefore that, although Birkhoff's ergodic theorem had brought the problem a great deal nearer solution, one was still faced with a hypothesis. As Birkhoff and Koopman (32B) put it: "The quasi-ergodic hypothesis has been replaced by its modern version: the hypothesis of metrical transitivity." However, it has apparently been overlooked that the metrical indecomposability of the energy surface had been proven for a certain class of physical systems of which some are known to exist. In order to see this, we first observe that, for any system which is quasi ergodic, the energy surface is metrically indecomposable; thus justifying the conjecture of the Ehrenfests that, for quasi-ergodic systems, Eq. (B1.05) would hold. Indeed, quasi ergodicity implies that every orbit on the energy surface will pass through any region $\mathcal{A}$ of positive measure. Since an orbit is produced by the collective of the transformations $P\rightarrow P_t$ with t ranging from $-\infty$ to $+\infty$, it follows that the condition of metrical indecomposability is satisfied for the energy surface of any quasi-ergodic system. Secondly we remember that Fermi (23F1) proved the quasi ergodicity of *Kanonische Normalsysteme* and that Poincaré (92P) has shown that in the reduced three-body problem one is dealing with such a system.[50] Of course, in this case the number of degrees of freedom is very small and it would be of interest to show that other physical systems also belong to the same class. *Kanonische Normalsysteme* are characterized by the following properties. After eliminating the uniform integrals of motion such as the components of the total momentum, it is possible to introduce sets of canonically conjugate variables x_i and y_i, such that (i) the energy E is independent of the time; (ii) there exists in the system a parameter α, such that the Hamiltonian $\mathcal{H}$ can be expressed as a power series in α,

$$\mathcal{H}=\mathcal{H}_0+\alpha\mathcal{H}_1+\alpha^2\mathcal{H}_2+\cdots; \qquad \text{(B2.03)}$$

(iii) the first term $\mathcal{H}_0$ in Eq. (B2.03) does not depend on the x_i, but the other terms may depend on both the

---

[47a] The group of transformations $P\rightarrow P_t$ is called metrically transitive and $\mathcal{S}$ is called metrically indecomposable, if $\mathcal{S}$ cannot be decomposed into two parts $\mathcal{S}_1$ and $\mathcal{S}_2$ both of positive measure and both invariant under every transformation of the group.

[48] We do not wish to discuss here the complications due to the existence of other integrals of motion such as linear momentum, angular momentum, and so on. These introduce what is called by Rosenfeld an inessential decomposition of $\mathcal{S}$, while we are only concerned with the essential indecomposability of $\mathcal{S}$. For our discussion we shall assume that there are no uniform integrals of motion apart from the energy and, moreover, we shall only consider metrical indecomposability in the physical sense. For a discussion of complications which we avoid this way we refer to the literature (see, e.g., 52R). We must also refer to the work of Grad (52G1, 52G2) who has considered in detail statistical mechanics of systems with integrals other than energy.

[49] From Eq. (B2.02) we see the relation between Birkhoff's ergodic theorem and Poincaré's recurrence theorem (compare 41W pp. 90–91). This equation expresses the fact that the representative point will spend a finite period in each region of positive measure.

[50] The parameter α occurring in Eq. (B2.03) is then the ratio m_2/m_1, where the three masses, m_1, m_2, and m_3, satisfy the inequalities $m_3\ll m_2<m_1$.

x_i and the y_i; (iv) all the $\mathcal{H}_j$ are periodic with a common period in the x_i.

Before we discuss briefly Hopf's ergodic theorem, we want to draw attention to two points. The first one is that Einstein and other writers (02E, 03E, 06P, 11K2) explicitly state that they believe all physical systems to belong to the class of *Kanonische Normalsysteme*. Secondly, we want to point out that, in the case of a system satisfying the condition of metrical indecomposability of its energy surface, the various integrals of motion are nearly[51] constant over the whole of the energy surface and that thus Boltzmann's and Maxwell's assumption of ergodicity, which involved as we saw exactly constant integrals of motion almost everywhere on the energy surface, was not as far removed from actual fact as one might think.

The ergodic theorem of Birkhoff's is restricted to the discussion of orbits on one energy surface. However, as we have emphasized before, it is more in accordance with the physical reality to consider a set of energy surfaces corresponding to energies within a finite interval, say,

$$E_0-\delta<E<E_0+\delta, \tag{B2.04}$$

(compare ESM p. 99) where $\delta\ll E_0$ in order that we can still speak of a system with a (more or less) well-defined energy. For such an *energy shell* Hopf (37H;[52] see also 30H2, 32H1, 32H2, 32H3, 32H4, 34H, 52R) has given a slightly modified ergodic theorem which again proves the equality of time and ensemble averages. The necessary condition for the validity of Hopf's ergodic theorem is that not only each energy surface in the shell is metrically indecomposable, but that also almost every energy surface of each product space, which is obtained by introducing a phase space the coordinates of which consist of a pair of sets of coordinates of Γ space,[53] is metrically indecomposable. If this condition is satisfied, any distribution in the energy shell will ultimately become a more or less uniform one (compare the discussion in Sec. C). Hopf's ergodic theorem does not introduce any new essential feature. The main importance lies probably—apart from its mathematical interest—in the fact that the quantum-mechanical counterpart of the ergodic theorem is more closely related to Hopf's than to Birkhoff's theorem, as we shall see in Part E.

C. THE *H*-THEOREM IN CLASSICAL ENSEMBLE THEORY

(1) Fine-Grained and Coarse-Grained Densities

Up to now we have been considering only the behavior of isolated systems. As we have already mentioned a few times, this approach does not do justice, in our opinion, to truly statistical considerations. It is rather the approach appropriate to the kinetic theory of gases, and it is, therefore, not surprising that both the original form of the *H*-theorem—which we discussed in Part A—and the classical ergodic theorem (Part B) found their origin in Boltzmann's work which found its culmination in his *Gastheorie* (96B, 98B). When Gibbs developed his statistical mechanics and introduced the ensemble theory, a new element entered the discussion. Gibbs tried to show (02G, especially Chapter XII; see also 07L) that an ensemble of systems would evolve in such a way that it would approach to a micro- or a macrocanonical ensemble.[54] In order to do this one must introduce a slightly different *H*-theorem and this will be discussed in the present section. We shall base our discussion on grand ensembles, partly because one can easily adopt the discussion to the case of petit ensembles, and partly because it is, in our opinion, the logical consequence of statistical considerations to consider these ensembles, first introduced by Gibbs in the last chapter of his monograph (02G, Chapter XV). Once the importance of the idea of representative ensembles is accepted, one is led automatically to grand ensembles as we shall see in the next section (see also ESM p. 235 and 55H1).

To simplify our discussions we shall consider only systems containing one kind of particles. We shall denote by ν the number of particles in a system. A system containing ν particles each with s degrees of freedom can be described by νs generalized coordinates (q) and νs generalized momenta (p). Its phase space, or Γ space, has thus $2s\nu$ dimensions. If we consider a grand ensemble, we are dealing with systems of a varying number of degrees of freedom. Let $D(\nu;p,q)d\Omega_\nu$ denote the number of systems in the ensemble with ν particles and with a representative point (the point in its phase space with coordinates equal to the values of the generalized coordinates and momenta is called the representative point) within the volume $d\Omega_\nu$ of its phase space. Here we have

$$d\Omega_\nu=\prod_{k=1}^{s\nu} dp_k dq_k. \tag{C1.01}$$

We have indicated the dependence of D on the q_k and p_k by p, q. Let N_{ens} be the number of systems in the ensemble, so that we have

$$N_{\text{ens}}=\sum_{\nu=1}^{\infty}\int D(\nu;p,q)d\Omega_\nu, \tag{C1.02}$$

where the integration extends each time over the whole of phase space. We now introduce the density of the

---

[51] We do not wish to make our statement more rigorous as this would involve a lengthy discussion.

[52] This reference gives an extensive bibliography on the mathematical aspects of the quasi-ergodic theorem.

[53] The phase space is so to speak duplicated. For a more extensive discussion we refer to Hopf's papers (see also 52R).

[54] We often use the term *macrocanonical* for ensembles called by Gibbs *canonical*. The reasons for this deviation from orthodoxy have been given elsewhere (54H2).

ensemble,[55] $\rho(\nu;\, p,q)$ by the equation

$$\rho = D/N_{\text{ens}}, \tag{C1.03}$$

and from Eqs. (C1.02) and (C1.03) it follows that ρ satisfies the normalization condition

$$\sum\nolimits_\nu \int \rho(\nu;\, p,q)\, d\Omega_\nu = 1. \tag{C1.04}$$

The index of probability η is given by the equation

$$\eta = \ln\rho. \tag{C1.05}$$

The average value $\langle G\rangle$ of any phase function $G(\nu;\, p,q)$ is given by the equation

$$\langle G\rangle = \sum\nolimits_\nu \int \rho G\, d\Omega_\nu. \tag{C1.06}$$

Up to this point we have not yet taken into account the implications of the fact that in a system containing only one kind of particles we are dealing with ν identical entities, and that we must make up our minds whether or not to consider situations which differ only in the permutation of some of the particles as being different. If we consider them to be different, we are dealing with *specific phases*, but if we consider them to be the same, we are dealing with *generic phases*. As is easily seen each generic phase contains $\nu!$ specific phases. Although, even before the introduction of quantum mechanics, there were strong reasons to prefer the generic phases to the specific phases (see, e.g., 02G, or ESM p. 142), the decisive argument in favor of generic phases is that only they lead to formulas which are the limiting cases of the quantum-mechanical ones. In the following we shall only deal with generic phases and, moreover, assume that integrations such as occur in Eqs. (C1.02), (C1.04), or (C1.06) are extended only over all different generic phases; that is, of all $\nu!$ different specific phases, only one is taken into account.

We now consider the properties of a quantity σ defined by the equation

$$\sigma = \sum\nolimits_\nu \int \rho \ln\rho\, d\Omega_\nu. \tag{C1.07}$$

We see, first of all, that σ is the average value of the index of probability, η. Secondly we shall show that σ has the following properties.

(a) If our ensemble is such that all systems have the same number of particles N and that the energy ϵ of all systems lies in the interval $E \leqslant \epsilon \leqslant E+\delta E$, σ will be minimum if ρ is given by the equation

$$\left.\begin{aligned} &\rho = \text{constant}\ \delta(\nu-N)\ \text{for}\ E\leqslant\epsilon\leqslant E+\delta E;\\ &\rho = 0 \qquad\qquad\qquad\quad \text{for}\ E\geqslant\epsilon\ \text{or}\ E+\delta E<\epsilon,\end{aligned}\right\} \tag{C1.08}$$

where $\delta(x)$ is the Dirac delta function (35D, Secs. 20, 21).

(b) If our ensemble is such that all systems have the same number of particles, but only the average energy is given, that is, ρ must satisfy the condition[56]

$$\sum\nolimits_\nu \int \epsilon\rho\, d\Omega_\nu = E, \tag{C1.09}$$

then σ will be minimum, if ρ satisfies the equation

$$\rho = \delta(\nu-N)\, e^{\beta(\psi-\epsilon)}, \tag{C1.10}$$

where β and ψ are constants which can be determined from Eqs. (C1.04) and (C1.09).

(c) If only the average number of particles and the average energy of the systems in the ensemble are given so that ρ must satisfy Eq. (C1.09) and the condition

$$\sum\nolimits_\nu \int \rho\, d\Omega_\nu = N, \tag{C1.11}$$

σ will be minimum, if ρ satisfies the equation

$$\rho = \exp[-q+\nu\mu-\beta\epsilon]. \tag{C1.12}$$

Once again β, μ, and q are constants, that is, they do not depend on the p_k, q_k, or ν, and they can be determined from Eqs. (C1.04), (C1.09), and (C1.11).

Before we prove these properties of σ, we may just remind ourselves that the densities given in the three cases correspond, respectively, to an energy-shell ensemble,[57] a macrocanonical ensemble, and a canonical grand ensemble. It follows from the usual considerations (e.g., ESM Secs. 5.3 and 6.1) that β, which is equal to the inverse of the modulus of the ensemble, is equal to $1/\mathbf{k}T$ ($\mathbf{k}$: Boltzmann's constant; T: absolute temperature), that ψ is the free energy of the system represented by the macrocanonical ensemble, that μ is equal to β times the partial free energy or the partial thermal potential, that q is Kramers' q-potential—which for a homogeneous system is equal to βpV (p: pressure V: volume)—and that in both case (a) and case (c) $\langle\eta\rangle$ is equal to $-S/k$ (S: entropy).[58]

The proof of the minimum properties of σ rests mainly on the fact that the function y, given by the equation

$$y = xe^x - e^x + 1, \tag{C1.13}$$

is positive for $x\neq 0$ and zero for $x=0$. This property of y follows most easily if one sees (i) that $y(0)=0$, and (ii) that $dy/dx(=xe^x)$ has the same sign as x.

Compare now in the three cases (a), (b), and (c) two densities ρ_1 and ρ_2 where ρ_1 in each case is the one which should lead to the minimum value of σ, that is, ρ_1 is given, respectively, by Eqs. (C1.08), (C1.10), and

---

[55] Gibbs calls D the density (in phase) and ρ the coefficient of probability.

[56] The summation over ν is here trivial as only one term of the sum contributes.

[57] This ensemble is sometimes called a microcanonical ensemble, although the microcanonical ensemble is really only the limiting case of an energy-shell ensemble.

[58] This is another example of the advantage of the generic densities over the specific densities, as this relation between S and ρ does not hold for the specific density.

(C1.12) while ρ_2 is given by the equation

$$\rho_2=\rho_1 e^{\Delta\eta}, \tag{C1.14}$$

where $\Delta\eta$ may be any function of the p_k and q_k and in case (c) of ν. Both ρ_1 and ρ_2 satisfy Eq. (C1.04),

$$\sum\int\rho_1 d\Omega=\sum\int\rho_2 d\Omega=1, \tag{C1.15}$$

while in cases (b) and (c) both satisfy Eq. (C1.09),

$$\sum\int\epsilon\rho_1 d\Omega=\sum\int\epsilon\rho_2 d\Omega=E. \tag{C1.16}$$

In case (c) both satisfy Eq. (C1.11),

$$\sum\nolimits_\nu\int\rho_1 d\Omega=\sum\nolimits_\nu\int\rho_2 d\Omega=N. \tag{C1.17}$$

Finally we have

$$\sigma_2-\sigma_1=\sum\int(\rho_2\ln\rho_2-\rho_1\ln\rho_1)d\Omega. \tag{C1.18}$$

From Eq. (C1.18), one can derive for all three cases the equation[59]

$$\sigma_2-\sigma_1=\sum\int\rho_1[\Delta\eta e^{\Delta\eta}-e^{\Delta\eta}+1]d\Omega\geqslant 0, \tag{C1.19}$$

the last inequality being a consequence of the properties of the function given by Eq. (C1.13).

We have now proved that σ is minimum for an energy shell, macrocanonical, or canonical grand ensemble under certain circumstances (we shall investigate in the next section the physical significance of these circumstances). Comparing Eq. (C1.07) for σ with Eq. (A2.03) for H, we might think for a moment that we might be able to prove that $d\sigma/dt$ will always be negative and thus be able to show a tendency for the establishment of canonical ensembles. However, it is easy to show that $d\sigma/dt=0$. This was pointed out soon after the publication of Gibb's monography by Burbury (03B). In order to see this we write

$$\sigma(t'')=\sum\int\rho''\ln\rho'' d\Omega''=\sum\int\rho'\ln\rho' J d\Omega'$$

$$=\sum\int\rho'\ln\rho' d\Omega'=\sigma(t'), \tag{C1.20}$$

where $\rho''[\rho']$ is a short-hand notation for $\rho(p'',q'';t'')$ $[\rho(p',q';t')]$, where the relation between p'', q'' and p', q' is such that a representative point p', q' at time t' will have moved to p'', q'' by t'',[60] where J is the Jacobian of the transformation from p'', q'' to p', q' which by virtue of Liouville's theorem (e.g., ESM p. 102) is equal to 1.

Although we see that σ is constant, there is still in a certain sense an approach to a stationary state. We shall first illustrate this by an example due to Gibbs (02G; see also the discussion in 03B, 04B1, 04B2, 06E2, 11E2). Consider a container with a liquid, say water, in which is put some coloring material and let us assume that this coloring material is nondiffusible and consists of colloidal particles. It is a well-known empirical fact that, if we start from a state where the coloring material is unevenly distributed, practically any kind of stirring will produce a situation where the color distribution is, as far as our eye can see, uniform. That means that stirring will produce an "equilibrium" state. However, if we look at the system very closely, we will still find that, in microscopic volumes, part of the space is occupied by the water and part occupied by the colloidal particles. Although the coarse distribution is uniform the finer distribution is still uneven.

From this example it follows that it might be advantageous to introduce apart from the *fine-grained* density ρ a *coarse-grained* density P defined as follows. Divide for each ν the corresponding Γ space into finite, but small, cells[61] $\Omega_i^{(\nu)}$ of volume $W(\Omega_i^{(\nu)})$, and let $\mathrm{P}_i^{(\nu)}$ be the average of ρ over $\Omega_i^{(\nu)}$,

$$\mathrm{P}_i^{(\nu)}=\left[\int\rho d\Omega_\nu\right]\Big/W(\Omega_i^{(\nu)}), \tag{C1.21}$$

where the integration extends over the cell $\Omega_i^{(\nu)}$.[62]

We now introduce the coarse-grained density $\mathrm{P}(\nu;p,q)$ by putting it constant in each cell $\Omega_i^{(\nu)}$ and equal to $\mathrm{P}_i^{(\nu)}$. From Eq. (C1.04) we now get, as can easily be verified,

$$\sum\nolimits_\nu\sum\nolimits_i \mathrm{P}_i^{(\nu)}W(\Omega_i^{(\nu)})=1, \tag{C1.22}$$

or

$$\sum\int\mathrm{P}d\Omega=1, \tag{C1.23}$$

and we see that P is normalized.

Instead of σ we now introduce a function Σ by the equation

$$\Sigma=\sum\nolimits_\nu\sum\nolimits_i W(\Omega_i^{(\nu)})\mathrm{P}_i^{(\nu)}\ln\mathrm{P}_i^{(\nu)}, \tag{C1.24}$$

---

[59] The derivation of Eq. (C1.19) follows in case (a) by adding to the right-hand side of Eq. (C1.18) the expression $\sum\int(1-\ln\rho_1)\times(\rho_1-\rho_2)d\Omega$, which is zero by virtue of Eq. (C1.15) and the fact that $\rho_1=$constant. In case (b) one adds to the right-hand side of Eq. (C1.18) the expression $\sum\int[\beta(\psi-\epsilon)+1](\rho_1-\rho_2)d\Omega$ which is zero by virtue of Eqs. (C1.15) and (C1.16). Finally in case (c) one adds the expression $\sum\int\{q-\nu\mu+\beta\epsilon-1\}(\rho_2-\rho_1)d\Omega$ which is zero by virtue of Eqs. (C1.15) to (C1.17).

[60] As we are dealing with systems with only one constituent the representative point of a system will stay in the same Γ space and there will not be a change in ν. The situation is, however, much more complicated if we consider systems in which chemical reactions can take place. These might be tackled using the second quantization methods developed by Schönberg (52S1, 53S1, 53S2).

[61] Compare introduction of cells in Sec. A3.

[62] We could have slightly generalized the definition of the coarse-grained density by also dividing the possible ν-values into intervals containing a few integers each. However, in order not to complicate our formulas we have not done this, as it is not necessary for our arguments in the case of systems containing only one constituent.

which expressed in the coarse-grained density P gives

$$\Sigma=\sum\int \mathrm{P}\ \ln\mathrm{P}d\Omega, \qquad (\mathrm{C}1.25)$$

or, if we remember that lnP is constant over each cell $\Omega_i^{(\nu)}$ and that the integration of P over $\Omega_i^{(\nu)}$ is the same as that of ρ over $\Omega_i^{(\nu)}$,

$$\Sigma=\langle\ln\mathrm{P}\rangle. \qquad (\mathrm{C}1.26)$$

The properties of σ of cases (a), (b), and (c) now hold for Σ, if we everywhere change ρ to P. However, now Σ is no longer constant and we can, indeed, show that Σ will decrease; thus showing a tendency toward the establishment of canonical ensembles—at any rate in so far as the coarse-grained density is concerned. We may remark at this point that although the distinction between fine-grained and coarse-grained densities is assumed in Gibbs' monograph (02G), the two are not clearly distinguished in the discussion and the same is to a large extent true also for the discussion of Burbury (03B, 04B2) and Bumstead (04B1); we may further refer to discussions by the Ehrenfests (06E2, 11E2), Poincaré (06P), Lorentz (07L), and Kroò (11K2).

Let us now consider the change of Σ with time. In order to do this we have to anticipate some of the arguments of the next section where representative ensembles are discussed. Let us assume that we have made some observations about a physical system at t'. Since these observations will never give us the maximum possible information, we can construct an ensemble, the average properties of which at t' correspond to the observed properties of the system under observation at t'. Because of experimental limitations, we shall at most be able to give $\rho(\nu;p,q)$ changing from one cell to another, if we have chosen the size of the cells in accordance with the experimental limitations, as we shall assume to have done. We choose thus the fine-grained density constant in each cell and have then at t'

$$\mathrm{P}'=\rho', \qquad (\mathrm{C}1.27)$$

and for Σ,

$$\Sigma'=\sum\int \mathrm{P}'\ \ln\mathrm{P}'d\Omega=\sum\int \rho'\ \ln\rho' d\Omega. \qquad (\mathrm{C}1.28)$$

If the situation at t' already corresponded to an equilibrium situation, Σ would have had its minimum value and no change could be expected, since the three ensembles corresponding to minimum Σ in cases (a), (b), and (c) are all three stationary ensembles (see, e.g., ESM pp. 105 and 137). Let us therefore assume that ρ' does not correspond to a stationary ensemble. At a later[63] time t'' Eq. (C1.27) will no longer hold and we have

$$\rho''\neq\mathrm{P}'', \qquad (\mathrm{C}1.29)$$

because of the fact that, although ρ will stay constant in extensions in phase of unchanging volume $W(\Omega_i^{(\nu)})$, the shape of these extensions will change and at a later moment each of the cells will be covered by points which at t' belonged to many different cells.

We write Eq. (C1.29) in the form

$$\rho''=\mathrm{P}''e^{\Delta}, \qquad (\mathrm{C}1.30)$$

where Δ is a function of ν and the p's and q's. For Σ we have now

$$\Sigma''=\sum\int \mathrm{P}''\ \ln\mathrm{P}''d\Omega. \qquad (\mathrm{C}1.31)$$

and for the change of Σ we have[64]

$$\Sigma'-\Sigma''=\sum\int \mathrm{P}''[\Delta e^{\Delta}-e^{\Delta}+1]d\Omega>0. \qquad (\mathrm{C}1.32)$$

We see now that Σ'' will be less than Σ' because of the fact that ρ'' and P'' are no longer everywhere equal. Comparing the present situation with the case of the coloring material in a liquid, we may expect that, as time marches on, ρ and P will differ more and more and that Σ will continue to decrease until equilibrium distributions of P have been reached, that is,

$$\text{case (a)}\quad \left.\begin{array}{ll}\mathrm{P}=\delta(\nu-N)\cdot\text{constant}, & E\leqslant\epsilon\leqslant E+\delta E;\\ \mathrm{P}=0, & \epsilon<E \text{ or } E+\delta E<\epsilon;\end{array}\right\} \qquad (\mathrm{C}1.33)$$

$$\text{case (b)}\quad \mathrm{P}=\delta(\nu-N)e^{\beta(\psi-\epsilon)}; \qquad (\mathrm{C}1.34)$$

$$\text{case (c)}\quad \mathrm{P}=\exp[-q+\nu\mu-\beta\epsilon]. \qquad (\mathrm{C}1.35)$$

As soon as the distribution (C1.33) to (C1.35) are reached, Σ has reached its minimum value and will therefore no longer decrease. In that case we would thus, on observation, reach the conclusion that the state of the system under consideration would best be represented by a microcanonical or a canonical grand ensemble—which are stationary ensembles. This presents us with another justification for using these ensembles to describe systems in thermodynamical equilibrium.[65]

The time needed to reach the equilibrium distributions will, in general, be of the order of the relaxation time, and not—as suggested by the Ehrenfests (11E2)—of the order of the recurrence times of the Poincaré cycle (Sec. A2). It seems to us that they forget to take into account the indistinguishability of the constituent

---

[63] Strictly speaking we should say "at *another* time," leaving open the question whether t'' is later or earlier than t'. However, since we have set up our ensemble to represent a system observed at t', we are only interested in predictions which might be verified by experiment, that is, in *later* times t''. In this connection it is interesting to read Burbury's and the Ehrenfests' remarks on this point (03B, 04B2, 06E2; see also 02G, 04B1).

[64] Equation (C1.32) is obtained by subtracting Eq. (C1.31) from Eq. (C1.28) changing in the integrand of (C1.28) ρ' to ρ'' which can be done since $d\rho/dt=0$ [compare Eq. (C1.20)], and adding the expression $\sum\int(\mathrm{P}''-\rho'')d\Omega$, which is zero by virtue of Eqs. (C1.04) and (C1.23).

[65] For the usual justification see, e.g., Gibbs' monograph (02G, especially Chapters IV, X, and XV) or ESM, Secs. 5.3, 5.7, and 6.1 (compare also ESM Secs. 5.6 and 6.3).

particles in the system. This means that the Poincaré period must be divided by $N!$ and time intervals of the order of the relaxation time ensue.

To conclude this section there are two points which we should like to mention. First of all we have proved that, if we have set up a representation ensemble at t', the quantity Σ will decrease from its value at t' to a smaller value at t'', but we have *not* proved, but only made plausible by appealing to the case of coloring material in water that Σ will continue to decrease. The situation is very similar to the one encountered in Sec. A(3), where we discussed the statistical aspects of the H-theorem. It would be worth while to investigate the behavior of Σ as a function of time in more detail along lines analogous to the discussion of Sec. A(3). It is interesting to note that Gibbs (02G) himself did not state that Σ' would monotonically decrease, but only that

$$\lim_{t\to\infty} \Sigma(t) \leqslant \Sigma(t'), \tag{C1.36}$$

a property which he did not prove, but which for a special case was proven by Poincaré (06P) and Kroò (11K2). We may also refer to the discussion of Tolman's (38T, especially Secs. 49 and 51) and the discussion in the next section.

The second point to which we wish to draw attention is that the decrease of Σ resulted from the fact that we came from a state where ρ and P were equal to a state where ρ was no longer everywhere equal to P. As P is obtained by experimental observation we may say that at t' we "know" ρ. However, at t'' we no longer possess the same amount of specific information. Since Σ can be considered to be a "coarse-grained" σ and since σ is related to the entropy by the relation

$$\sigma = \langle \eta \rangle = -S/k, \tag{C1.37}$$

we see once more how *increase of lack of knowledge* corresponds to an increase in entropy.[66]

(2) Representative Ensembles

In the preceding section we have considered the behavior of ensembles and we have thus started the discussion of question (B) of the Introduction, that is, the question: Why is it possible to describe the behavior of a system by considering the average behavior of a system in an ensemble. This is perhaps the most important question in statistical mechanics, but for some queer reasons it has not been considered as carefully as it deserved. This may be partly due to the fact that Gibbs himself introduced ensembles to use them rather for statistical considerations than to illustrate the behavior of physical systems—even though he did not adhere strictly to his original intentions as set out in the introduction to his monograph. As a matter of fact, there are some indications that Gibbs had in mind the possibility that the average behavior of a system in a micro- or macrocanonical ensemble might represent the actual behavior of a physical system in equilibrium [see, e.g., the remarks made by Bumstead (04B1)]. In the same year as Gibbs and independently, Einstein (02E, 03E) introduced the ensemble theory and he definitely had in mind this possibility of the equivalence of the average behavior in an ensemble and the actual behavior of a physical system. However, the Ehrenfests (11E2), Ornstein (08O), and Uhlenbeck (27U) considered the ensemble theory mainly as a mathematical trick to lighten the calculation of average values of phase functions. Uhlenbeck, for instance, definitely puts the ensemble theory on a par with the Darwin-Fowler method (22D1, 22D2, 22D3, 23F2, 23F3, 23D1, 23D2, 25F1, 26F2, 26F3; for a brief account see ESM, Appendix IV or 48S, Chapter VI). It was not until Tolman (38T, 40T) introduced the concept of *representative ensembles* that ensemble theory got a thorough physical foundation, but his point of view has not always been properly taken into account. We shall discuss his approach in some detail in the present section as far as classical statistical mechanics is concerned, reserving for Sec. D(3) the corresponding discussion of the quantum-mechanical case.

We have mentioned several times before in this paper that owing to severe limitations to our possibilities, we can only obtain very inadequate knowledge about the initial condition, say, of the physical systems which we happen to study. If we remind ourselves that one mole of gas contains 6.10^{23} atoms, we see immediately that instead of obtaining values for the 36.10^{23} coordinates and momenta which we should have to know in order to characterize the situation completely (assuming each atom to be a point particle with only three degrees of freedom) we can only obtain a few relations between them, for instance, by measuring the pressure, the linear momentum, and the angular momentum of the system. Moreover, we certainly will not know the total number of particles in the system; we have only to remind ourselves that Avogadro's number is only known to a few parts in ten thousand (51B1, 51D). Having obtained this very scant information we now want to make predictions about the future behavior of the system from this information. Clearly the only possible way to do so is by using probability theory and statistical methods. If we could have obtained the values of all the coordinates and momenta at one moment—and if we could have solved the equations of motion—the behavior of the system would be completely determined, as we would then be dealing with a problem of classical mechanics with a sufficient number of boundary conditions to determine the orbit of the representative point of the system in phase space. However, as we cannot obtain these values we are reduced to making statements about the most probable behavior of the system. In order to determine this most probable behavior we compare various systems which all possess

---

[66] We do not wish to pursue this correspondence between entropy and lack of detailed information or extent of ignorance, but refer to the literature; see, e.g., ESM p. 160 and papers by Szilard (29S) and Brillouin (51B4; see also 48W, 49S1, 54M1, 54G1, 55G1).

the same value of those quantities which have been measured, but which otherwise may differ widely. This means, that we construct an ensemble to *represent* our systems. There remains the question how we can construct such a representative ensemble, and this introduces the problem of *a priori probabilities*. In constructing a classical ensemble we use the assumption of equal *a priori* probabilities for equal volumes in Γ space.[67] As far as petit ensembles are concerned this leads to a straightforward prescription, but in the case of the construction of a representative grand ensemble we must remember that we are dealing with Γ spaces of differing dimensions and that those with larger numbers of dimensions have thus larger *a priori* weights. We shall return presently to discuss the plausibility or justification of our choice for the *a priori* probability.

In order to see what kind of ensemble will best represent a system under consideration, we must first of all bear in mind that in practically all circumstances we only measure the properties of a small subsystem. In measuring the pressure of a gas, only the part of the gas near the wall takes part in exerting the force on the wall which is registered by our instruments. In determining the density of a system by optical method only part of the system is traversed by the light beam (compare, e.g., the case of critical opalescence). In measuring the electrical conductivity of a semiconductor, a probe is put on the piece of material and, indeed, is often used to find out differences between situations at one place or another, and so on. This entails that we are really interested only in predicting the future behavior of a subsystem. For the sake of simplicity we shall assume that the complete system—which may be either the whole of the physical system in our apparatus or even this system plus part of the surroundings—is isolated so that it cannot exchange either energy or particles with its surroundings. Let ρ be the density of the ensemble describing one subsystem and ρ' the density of the ensemble describing the "rest system," that is, the complete system minus the subsystem. Let furthermore ϵ and ν denote the energy and number of particles in the subsystem and ϵ' and ν' those in the rest system. The densities ρ and ρ' will then satisfy the relations

$$\sum\int\rho d\Omega=1, \tag{C2.01}$$

$$\sum\int\rho' d\Omega'=1, \tag{C2.02}$$

$$\sum\int(\epsilon+\epsilon')\rho\rho' d\Omega d\Omega'=\text{constant}, \tag{C2.03}$$

$$\sum(\nu+\nu')\int\rho\rho' d\Omega d\Omega'=\text{constant}, \tag{C2.04}$$

where Eqs. (C2.03) and (C2.04) express the fact that in an isolated system the total energy and the total number of particles is constant. Let us, furthermore, assume that an equilibrium situation has been reached, which means that Σ is minimum for the complete system, or,

$$\sum\int(\rho\rho')\ln(\rho\rho')d\Omega d\Omega'=\text{minimum}. \tag{C2.05}$$

In writing down Eqs. (C2.03) to (C2.05) we have used the fact that the ensemble of the complete system is the product of the ensembles of the subsystem and the rest system. Combining Eqs. (C2.01) to (C2.04) we can write instead of Eqs. (C2.03) and (C2.04),

$$\sum\int\epsilon\rho d\Omega+\sum\int\epsilon'\rho' d\Omega'=\text{constant}, \tag{C2.06}$$

$$\sum\nu\int\rho d\Omega+\sum\nu'\int\rho' d\Omega'=\text{constant}. \tag{C2.07}$$

The equilibrium density now follows from the minimum condition (C2.05) which must be satisfied under the conditions (C2.01), (C2.02), (C2.06), and (C2.07). This leads in the well-known way (method of Lagrangian multipliers, see, e.g., ESM p. 20 or 444) to the following equilibrium expression for ρ;

$$\rho=\exp[-q+\nu\mu-\beta\epsilon]. \tag{C2.08}$$

We may remark here that, as Eq. (C2.05) really only holds for the coarse-grained density, we really only obtain expression (C2.08) for P.

Secondly, we may draw attention to the fact that our assumption about the *a priori* probabilities is used in writing down Eqs. (C2.01) to (C2.04) *without* introducing apart from ρ a density function which would indicate the *a priori* weights of the various volumes in phase space.

We have proved here in a manner slightly different from the one used in the previous section that the canonical grand ensemble will represent any subsystem in equilibrium. We must, however, pay some attention to the problem of nonequilibrium systems and we will follow Tolman's discussion (38T, especially Secs. 79b, 51c, and 112) to a large extent. After that we conclude this section by discussing the choice of equal *a priori* probabilities for equal volumes in phase space.

Let us assume that our system is in a nonequilibrium situation and that we have ascertained this by making suitable observations on the system at t'. In order to predict its future behavior we construct an ensemble to represent the system, using, on the one hand, our observational information and, on the other hand, the assumption of equal *a priori* probabilities (or equal weights) for equal volumes in phase space. As we discussed in the previous section we can only determine at most the coarse-grained density P′ and, using the

---

[67] This assumption was tacitly introduced when we wrote down expression (A3.01) for the probability of a situation Z.

assumption of equal weights for equal volumes, we are thus led to choose our fine-grained density ρ' according to Eq. (C1.27). At a later time t'' the fine-grained density ρ'' which follows from ρ' will no longer be constant in each of the cells into which we have divided phase space. As we saw earlier this results from the fact that P'' and ρ'' are no longer equal. From Eq. (C1.32) we see that $\Sigma'-\Sigma''$ is the larger, the more P'' and ρ'' differ, and we can thus *expect* a continued decrease of Σ until the distributions of Eqs. (C1.33) to (C1.35) are reached.

In order to see this approach to equilibrium we might make suitable observations on the system. From the forgoing discussion we would expect that these observations would lead to coarse-grained densities such that Σ is a steadily decreasing function of time. However, in making that assumption we, first of all, have tacitly assumed that the mixing of the various cells in phase space is a thorough one corresponding to the mixing of coloring material in water, and, secondly, we have forgotten to take into account the fact that each time we perform an observation this will add to the amount of available information and hence our choice of a representative ensemble at t''', say, will no longer be as simple as at t', since the information obtained at t' and t'' must be taken into account. As far as the mixing is concerned, we see that we are back to the question treated in Hopf's ergodic theorem and we see that, indeed, the approach via the ergodic theorem or via the ensemble theory, once we get down to the basic essentials, do not differ as much as one would expect beforehand.

Let us now consider the problem of *a priori* probabilities. Let us restrict ourselves for the sake of simplicity to petit ensembles. Consider two regions $\mathfrak{A}_0$ and $\mathfrak{B}_0$ of Γ space. Let those two regions at t_0 be filled with representative points of an ensemble and let $p(\mathfrak{A}_0)$ and $p(\mathfrak{B}_0)$ be the *a priori* probabilities for the two regions. At $t_1, t_2, \cdots$ the representative points which at t_0 filled up $\mathfrak{A}_0$ and $\mathfrak{B}_0$ will fill up regions $\mathfrak{A}_1$ and $\mathfrak{B}_1$, $\mathfrak{A}_2$ and $\mathfrak{B}_2 \cdots$ with *a priori* probabilities $p(\mathfrak{A}_1)$ and $p(\mathfrak{B}_1)$, $p(\mathfrak{A}_2)$, and $p(\mathfrak{B}_2)\cdots$. One can clearly only tolerate such a choice of *a priori* probabilities that

$$p(\mathfrak{A}_0)/p(\mathfrak{B}_0)=p(\mathfrak{A}_1)/p(\mathfrak{B}_1)=p(\mathfrak{A}_2)/p(\mathfrak{B}_2)=\cdots, \qquad \text{(C2.09)}$$

and these relations are satisfied, if we take

$$p(\mathfrak{A})=W(\mathfrak{A}), \quad p(\mathfrak{B})=W(\mathfrak{B}), \qquad \text{(C2.10)}$$

where $W(\mathfrak{A})$ or $W(\mathfrak{B})$ is the volume of the region $\mathfrak{A}$ or $\mathfrak{B}$ in Γ space. We remark here that Eq. (C2.10) will only satisfy Eq. (C2.09) if we are dealing with a phase space made up of canonically conjugated p's and q's—one of the reasons for choosing the canonical coordinates to describe the phase of a system.

Instead of Eq. (C2.10) one could choose the more general form

$$p(\mathfrak{A})=\int F(\epsilon,\varphi_i)d\Omega, \qquad \text{(C2.11)}$$

where the integration extends over the region $\mathfrak{A}$ and where $F(\epsilon,\varphi_i)$ is a once and for all chosen function of the energy ϵ and the other constants of motion φ_i.

The reason for putting $F(\epsilon,\varphi_i)$ equal to 1, which means choosing Eq. (C2.10) instead of the more general form (C2.11) was originally historical and was based on the assumption that ergodic systems existed. Since we are nowadays convinced that mechanical systems are quasi ergodic, which means essentially that the various possible values of φ_i all occur with practically equal frequency in each part of phase space, we can immediately restrict ourselves to what the Ehrenfests call the ergodic choice,

$$p(\mathfrak{A})=\int F(\epsilon)d\Omega. \qquad \text{(C2.12)}$$

However, since the density ρ of the representative ensembles of equilibrium systems are functions of the energy only, the choice (C2.12) instead of (C2.10) would correspond to a density

$$\rho'(\epsilon)=\rho(\epsilon)/F(\epsilon), \qquad \text{(C2.13)}$$

instead of the $\rho(\epsilon)$ given by Eqs. (C1.08) or (C1.10). The $F(\epsilon)$ would occur in all integrals, but the final results would be the same.

We may mention here that the Ehrenfests' remark that the ergodic choice is often inexpedient is no longer relevant as it was based on difficulties encountered by thermostatistics before the introduction of quantum mechanics.[68]

SUMMARY OF THE SITUATION IN CLASSICAL THERMOSTATISTICS

Before we start the discussion of the status of H-theorem and ergodic theorem in quantum statistics, we wish to summarize briefly the contents of Parts A, B, and C. There are essentially three different methods of approaching the problem of how to justify the method of statistical mechanics: (1) the utilitarian approach; (2) the formalistic approach; and (3) the physical approach,[69] although the last two can overlap to some extent.

The utilitarian approach justifies the statistical method by its results and feels no need for a further justification. In this approach calculations made by any method, be it the Darwin-Fowler method, the use of microcanonical ensembles, the use of macrocanonical ensembles, or the use of grand ensembles, are welcome as long as they lead to useful results. The fact that for physical systems with many degrees of freedom all these calculations lead to the same results makes it a

---

[68] We may also refer to a paper by Ehrenfest (14E) discussing the possible choices of $F(\epsilon,\varphi_i)$.

[69] The reader will forgive me, I hope, for this classification which shows clearly my own prejudices and preference.

matter of expediency to choose the method ultimately used.[70]

The formalistic approach usually argues that a mechanical system should be an isolated one and thus possess a definite value of its energy as well as of the number of particles in it, since these are the only systems with which classical mechanics is concerned. The only stationary ensemble to be considered is thus the microcanonical petit ensemble although one may use the macrocanonical or grand ensembles as a mathematical trick to ease the calculations. Moreover, the microcanonical ensemble is only introduced once it is realized that probably the time average of any phase function, if calculated for one system, is the same as the average taken over a microcanonical ensemble. In order to prove this one must introduce the ergodic or better the quasi-ergodic theorem, and adherents to the formalistic school of thought place the ergodic theorem in the center of their considerations.

The physical approach emphasizes the fact that we can only obtain a limited amount of information about the systems in which we are actually interested and that we thus must work with representative ensembles. In this case, one must prove that the development of a nonequilibrium situation will be such that the ensemble representing the system under observation will resemble a canonical ensemble more and more closely.

The considerations trying to justify the assumptions of the last two methods of approach fall into three groups. One can either try to use the straightforward H-theorem, but as we saw in Part A, one runs into trouble. The straightforward H-theorem would provide an alternative to the ergodic theorem, but the H-theorem in its statistical form does not do this, unless one is willing to introduce the various assumptions about transition probabilities without further proofs. However, the statistical form of the H-theorem does serve a useful purpose in that it shows the relative frequency with which fluctuations will occur and the rate at which nonequilibrium situations may be expected to return to equilibrium.

One is then led to the ergodic theorem and as we saw in Part B the problem is at the present time reduced to a proof of the metrical indecomposability of the relevant phase spaces. If one could prove that physical systems are in general quasi ergodic, the ergodic theorem would have been proved but this has not been done. It is interesting to note that both Einstein (02E, 03E) and Poincaré (06P) explicitly state that they expect that the energy is the only time-independent, uniquely valued, analytical integral of a physical system. If one introduces a slight amount of reality in the considerations by allowing the energy to vary over a small, but finite range, one has recourse to Hopf's ergodic theorem where the mixing of various energy surfaces takes place. Once again the metrical indecomposability of the phase spaces involved has not yet been proved.

If one prefers the approach via the representative ensembles, one uses the H-theorem in the form suitable for ensemble theory, that is, where H is identified with the average value of the natural logarithm of the coarse-grained density. The approach to equilibrium follows easily, but the considerations are based on a special choice for *a priori* probabilities, and as we saw at the end of Sec. C(2) this choice can only be justified, if the systems under consideration are quasi ergodic.

We see thus that all methods have in common the lack of proof of the quasi ergodicity of physical systems. Once this proof is given, one can either use the formalistic approach since the (quasi) ergodic theorem is then proved, or the physical approach since the choice of *a priori* probabilities is justified from first principles. Concluding we can thus say that although there seems to be a large difference between the formalistic and the physical approach, this difference is in fact not so enormous, and both break down on the same point.

D. THE *H*-THEOREM IN QUANTUM STATISTICS

(1) The *H*-Theorem in the Elementary Method

It is well known that statistical mechanics was relatively little affected by the change over from classical to quantum mechanics. The same methods which were used to good effect in classical thermostatistics could be taken over in quantum thermostatistics. If anything, statistical methods were more suited to the new situation. Since quantum mechanics started from statistical considerations, namely, Planck's discussion of black-body radiation (00P1, 00P2; see also 43P for an account by Planck himself of the reasons for introducing the quantum of action), it is not surprising that the questions raised by the introduction of quantum theory into statistical discussions have been the subject of papers both before the advent of Bohr's old quantum theory (13B1, 13B2) and before and after the development of modern quantum theory, the so-called wave or matrix mechanics (25H, 25B, 26S1).[71] In the beginning the discussion was mainly on the subject of *a priori* weights (see Sec. D3), but after the start of wave mechanics, many authors (e.g., 27N1, 27N2, 29D, 32N3, 35D, 36D, 40H) have shown how Gibbs' ensembles can be introduced to advantage in quantum mechanics. In many ways thermostatistics fits more easily into quantum mechanics than into classical mechanics, since quantum mechanics involves statistical predictions. For instance, even if we have the maximum amount of information about the physical system under consideration, that is, when we are dealing with that von Neumann calls a *pure case* (*Reiner Fall*), we can only verify the predic-

---

[70] It is interesting to note that even Tolman uses this utilitarian point of view when he justifies the assumption of equal *a priori* probabilities for equal volumes in Γ space.

[71] For an account of the development of both the old and the modern quantum theory we must refer to the literature (e.g., 19S, 33H, 38K).

tions which can be made from the appropriate wave function by using an ensemble.[72]

In quantum statistics as in classical statistics there are different ways of tackling the problem with which we are concerned in the present paper. Once again we are confronted with questions (A) and (B) of the Introduction and once again we can try to consider first of all isolated systems, either by using the quantum-mechanical analogy of the H-theorem or by considering a quantum-mechanical ergodic theorem. Alternatively, we can consider ensembles and the question of representative ensembles. In this part we shall begin by considering [Sec. D(1)] the question of isolated systems by using the so-called *elementary method* (see ESM Ch. IV) and, firstly, show how one derives the equilibrium distribution in this case and, secondly, discuss the H-theorem. After that [Sec. D(2)] we shall discuss the H-theorem in ensemble theory, while we conclude this chapter with a discussion of representative ensembles and the choice of *a priori* weights [Sec. D(3)]. In the next part (E) we shall discuss the quantum-mechanical ergodic theorem.

Before we discuss the elementary method, we wish to consider briefly Ornstein and Kramer's consideration (27O) of the equilibrium of a system of Fermi-Dirac particles (compare also papers by Nordheim (28N) and Halpern and Doermann (39H)). In their paper they show how one can derive the Fermi distribution from purely (or better nearly purely) kinetic arguments. Their argument runs as follows. Consider a system of independent particles and let each particle be capable of occupying energy levels ϵ_k which, for the sake of simplicity, we assume to be nondegenerate. Consider now a "collision" between two particles such that the energies before the collision were ϵ_i and ϵ_j and after the collision $\epsilon_{i'}$ and $\epsilon_{j'}$. Conservation of energy gives us the relation

$$\epsilon_i+\epsilon_j=\epsilon_{i'}+\epsilon_{j'}. \tag{D1.01}$$

Let the transition probability for this transition be denoted by a, the transition probability for the inverse transition ($i'j'\rightarrow ij$) by a', the number of transitions per unit time in the one direction by A, and the number of transitions in the opposite direction by A'. The quantities A and A' satisfy the equations

$$A=an_in_j(1-n_{i'})(1-n_{j'}), \tag{D1.02}$$

$$A'=a'n_{i'}n_{j'}(1-n_i)(1-n_j), \tag{D1.03}$$

where $n_i(n_j,n_{i'},n_{j'})$ is the average number of particles in the energy state $\epsilon_i(\epsilon_j,\epsilon_{i'},\epsilon_{j'})$, and where Eqs. (D1.02) and (D1.03) follow from the fact that the number of transitions is not only proportional to the average number of particles in the states i and j, but also to the chance of finding states i' and j' unoccupied, since otherwise the transition cannot take place because of the Pauli principle (25P, 27P).

If the system is in equilibrium, we have

$$A=A'. \tag{D1.04}$$

We shall now assume that

$$a=a', \tag{D1.05}$$

and from Eqs. (D1.02) to (D1.05) we then get

$$m_im_j=m_{i'}m_{j'}, \tag{D1.06}$$

where

$$m_i=n_i/(1-n_i). \tag{D1.07}$$

Equation (D1.06) must be satisfied for any pair of states ϵ_i, ϵ_j in equilibrium. Just as the Maxwell distribution followed from Eq. (A1.12) and the condition of conservation of energy, so we get from Eqs. (D1.06) which have to be satisfied under the condition (D1.01) that in equilibrium

$$m_i=\exp(\mu-\beta\epsilon_i), \tag{D1.08}$$

or, using Eq. (D1.07),

$$n_i=1/[\exp(-\mu+\beta\epsilon_i)+1], \tag{D1.09}$$

the Fermi distribution (26F1).

It must be noted here (i) that Eq. (D1.02) is purely based on a kinetic argument and not on any quantum-mechanical formalism, and (ii) that Eq. (D1.05) is introduced as an assumption—an assumption which takes the place of the *Stosszahlansatz* in Sec. A1. We may mention here that Ornstein and Kramers actually discussed whether Eq. (D1.05) could be derived from first principles and mentioned that it followed from considerations by Heisenberg (26H) and Jordan (26J) that only the much weaker relation

$$\textstyle\sum_{ij}(a_{ij;i'j'}-a_{i'j';ij})=0 \tag{D1.10}$$

instead of relation (D1.05) could be derived from quantum mechanics. In Eq. (D1.10) $a_{ij;i'j'}(a_{i'j';ij})$ is the transition probability up to now denoted by $a(a')$, and the summation is over all pairs ϵ_i, ϵ_j for which $\epsilon_i+\epsilon_j$ is the same. The question in how far Eq. (D1.05) can be derived from Eq. (D1.10) is connected with questions discussed in Sec. D(3) and in Appendix V, to which we refer.[73]

We have not yet proved that a distribution different from the one given by Eq. (D1.09) will tend toward this distribution. This follows easily, if we use the expression for the entropy of a system of independent Fermi-Dirac particles[74] [see, e.g., ESM p. 406, Eq. (A3.06) and com-

---

[72] This is, for instance, discussed in detail by Kemble (37K; especially Sec. 14b); see also Sec. D(3).

[73] It is interesting to note that Eq. (D1.10) is sufficient to derive from Eq. (D1.04) and condition (D1.01) the Maxwell-Boltzmann distribution when, of course, we do not have the factors $(1-n_{j'})$, etc. in Eqs. (D1.02) and (D1.03); compare in this connection also 54K Sec. 14, 53T and 54L.

[74] Such particles are sometimes called *fermions* and the Bose-Einstein particles, *bosons*.

pare also Eq. (D1.28)] and write

$$H=-S/k=\sum_i[(1-n_i)\ln(1-n_i)+n_i\ln n_i]. \quad \text{(D1.11)}$$

The rate of change of H is given by the equation

$$\begin{aligned} dH/dt&=\sum_i \ln[n_i/(1-n_i)]dn_i/dt\\ &=\sum_{i,j,i'j'}\ln[n_i/(1-n_i)]\\ &\quad\times\{a_{i'j';ij}n_{i'}n_{j'}(1-n_i)(1-n_j)\\ &\qquad -a_{ij;i'j'}n_in_j(1-n_{i'})(1-n_{j'})\}. \end{aligned} \quad \text{(D1.12)}$$

In Eq. (D1.12) we sum firstly over all energy levels ϵ_i, secondly over all energy levels ϵ_j with which collisions can take place, and thirdly over all pairs of energy levels $\epsilon_{i'}$ and $\epsilon_{j'}$ which can be obtained from a collision between particles in states ϵ_i and ϵ_j. In order to get a symmetrical expression, where the summation is both over all pairs ϵ_i and ϵ_j and over all pairs $\epsilon_{i'}$ and $\epsilon_{j'}$ we write

$$\begin{aligned} dH/dt=\sum_{ij,i'j'} a[&n_{i'}n_{j'}(1-n_i)(1-n_j)\\ &-n_in_j(1-n_{i'})(1-n_{j'})]\\ &\times\{\ln[n_in_j/(1-n_i)(1-n_j)]\}, \end{aligned} \quad \text{(D1.13)}$$

where we have used Eq. (D1.05). If we now take the mean of the expression on the right-hand side of Eq. (D1.13) and the expression obtained from this one by exchanging ij and $i'j'$, we get [compare the discussion of Eq. (A1.09)]

$$\begin{aligned} dH/dt=\tfrac{1}{2}\sum_{ij,i'j'} &a\{\ln[n_in_j(1-n_{i'})\\ &\times(1-n_{j'})/(1-n_i)(1-n_j)n_{i'}n_{j'}]\}\\ &\times[n_{i'}n_{j'}(1-n_i)(1-n_j)-n_in_j(1-n_{i'})\\ &\qquad\times(1-n_{j'})]\leqslant 0, \end{aligned} \quad \text{(D1.14)}$$

where the equal sign only holds, if the n_i satisfy Eq. (D1.09). We see here, thus, once again the approach to an equilibrium distribution proved by using an assumption about the number of transitions. We leave a discussion of the derivation of Eq. (D1.14) until we have considered the approach via the elementary method.

In the elementary method one takes the energy levels together in groups with all approximately the same energy. Let the ith group contain Z_i levels, let there be N_i particles in the system with energies within the group, and let E_i be the (approximate) value of the energy of the levels in the group. We first of all want to find the equilibrium distribution of the N_i and once again we shall, as in Sec. A(3), define this as the most probable distribution, and we must thus find the *a priori* probability $W(N_i)$ for a given distribution, corresponding to the $W(\mathbf{Z})$ of Eq. (A3.01). At this point we must differentiate between Boltzmann (Bo), Fermi-Dirac (F-D) and Bose-Einstein (B-E) statistics. We must first of all remind ourselves that quantum mechanics introduces two aspects different from those of classical mechanics, namely, diffraction effects and symmetry effects.[75] The first aspect entered already in the old quantum mechanics and is responsible for the existence of the quantization of energy, angular momentum, and so on. The second aspect enters when one takes into account that apparently nature only permits wave functions of a system of identical particles which are either completely symmetric in all particles (B-E) or completely antisymmetric (F-D).[76] At first sight one might think therefore that only B-E or F-D statistics would occur. However there are cases such as, for instance, a system of particles arranged on a crystal lattice, where identical particles can be distinguished, e.g., in the case of a crystal, by the lattice site they occupy (compare 49R, Chap. III, Sec. 1). In such a case we are dealing with Boltzmann statistics—with one slight difference which we shall discuss presently. B-E statistics were introduced by Bose (24B)[77] for the case of light quanta who used them to derive Planck's radiation law. Einstein (24E2, 25E1, 25E2) applied B-E statistics to a perfect gas, and showed how they led to the so-called Einstein condensation—a phenomenon which may have a bearing on the λ transition of liquid helium (compare the discussion in Ch. IX of ESM). Fermi (26F1) introduced Pauli's exclusion principle (25P, 27P) into the statistical discussion, while Dirac (26D) discussed the connection between statistics and wave mechanics. Ehrenfest and Uhlenbeck (27E, see also 27U) showed how also the Boltzmann statistics can be incorporated in quantum statistics.

We can now proceed to evaluate $W(N_i)$. Let n_j be the number of particles in the energy level ϵ_j and let $W(n_j)$ be the probability for a distribution of the n_j. We then have

$$W(N_i)=\sum W(n_j), \quad \text{(D1.15)}$$

where the summation extends over all n_j-distributions such that for each group

$$\sum n_j=N_i, \quad \text{(D1.16)}$$

where the ϵ_j over which is summed belong to the group Z_i. Our problem is now reduced to finding $W(n_j)$, that is, the probability that N particles are distributed over the various energy levels which are all assumed to be nondegenerate in such a way that these are n_j particles in level ϵ_j.

In the case of B-E statistics for any given distribution there exists just one wave function and hence

$$W_{\text{B-E}}(n_j)=1. \quad \text{(D1.17)}$$

In the case of F-D statistics only completely antisymmetric wave functions are possible—which means that each level can be either occupied by one particle

---

[75] For a discussion of the importance of this distinction see 54H3.

[76] For a discussion of the reasons why some particles obey B-E and others F-D statistics we refer to the literature (e.g., 39B2, 40P1, 40P2, 49W2).

[77] This paper was translated by Einstein who added the following remark: "Boses Ableitung der Planckschen Formel bedeutet nach meiner Meinung einen wichtigen Fortschritt. Die hier benutzte Methode liefert auch die Quantentheorie des idealen Gases, wie ich an anderer Stelle ausführen will."

or unoccupied, but cannot be occupied by more than one particle—and we have thus either one possible wave function, or none,

$$\left.\begin{aligned} W_{\text{F-D}}(n_j)&=1, \quad \text{if all } n_j \text{ are 0 or 1;} \\ W_{\text{F-D}}(n_j)&=0, \quad \text{if at least one } n_j>1. \end{aligned}\right\} \quad \text{(D1.18)}$$

Finally in the case of Bo statistics any of the $N!$ permutations of the arguments of the N particles in the wave function will lead to another permitted wave function. As a permutation of the n_j-particles in the same level would not change the wave function we might expect

$$W'_{\text{Bo}}=N!\prod_j (1/n_j!). \quad \text{(D1.19)}$$

This does not take into account that, although we can sometimes label our particles, we cannot keep track of each separate atom and are thus overestimating the weight of each n_j-distribution by a factor $N!$ which is the number of different microsituations corresponding to the same macrosituation.[78] We must thus use in the case of Bo statistics instead of Eq. (D1.19), the expression

$$W_{\text{Bo}}(n_j)=\prod_j (1/n_j!). \quad \text{(D1.20)}$$

We can now use Eqs. (D1.15) to (D1.18) and (D1.20) to calculate $W(N_i)$. In the case of Bo statistics one finds easily (compare ESM, p. 74) that

$$W_{\text{Bo}}(N_i)=\prod_i \frac{1}{N_i!}\sum\frac{N_i!}{\prod n_j!}=\prod_i (Z_i^{N_i}/N_i!). \quad \text{(D1.21)}$$

In the case of B-E statistics we must find the number of different ways in which N_i particles can be distributed over Z_i levels and we get

$$W_{\text{B-E}}(N_i)=\prod_i [(N_i+Z_i-1)!/N_i!(Z_i-1)!]. \quad \text{(D1.22)}$$

Finally in the case of F-D statistics we must find the number of different ways in which N_i particles can be distributed over Z_i levels without having more than one particle in any level, and we get

$$W_{\text{F-D}}(N_i)=\prod_i [Z_i!/N_i!(Z_i-N_i)!]. \quad \text{(D1.23)}$$

We now assume that all Z_i, N_i, and Z_i-N_i (in the F-D case) are large numbers so that we can use Eq. (A3.03) for the factorials. We then get

$$\begin{aligned} \ln W(N_i)=\sum_i \{ (N_i+\alpha Z_i) \ln[(Z_i/N_i)+\alpha] \\ -\alpha Z_i \ln(Z_i/N_i)\}, \end{aligned} \quad \text{(D1.24)}$$

where

$$\alpha_{\text{Bo}}=0, \quad \alpha_{\text{B-E}}=1, \quad \alpha_{\text{F-D}}=-1, \quad \text{(D1.25)}$$

and where we have omitted a term $\Sigma N_i(=N)$ in the Bo case, which is unimportant for our discussion. From Eq. (D1.24) it can be seen that in the limit where $N_i/Z_i\rightarrow 0$, that is, for highly diluted systems, or systems at very high temperatures, the three expressions become the same.

We find the equilibrium distribution by maximizing $\ln W(N_i)$ under the conditions of given total energy and given total number of particles,

$$\sum N_i=N, \quad \sum N_i E_i=E, \quad \text{(D1.26)}$$

and the result is [compare the discussion in Sec. A(3)],

$$\ln[(Z_i+\alpha N_i)/N_i]=-\mu+\beta E_i, \quad \text{(D1.27)}$$

leading to the well-known expressions for the Boltzmann, Fermi-Dirac, and Bose-Einstein distributions.

We now must prove the approach to equilibrium.[79] Using the relation $H=-S/k$ and the Boltzmann-Planck relation $S=k\ln W$, we have for our H-function

$$\begin{aligned} H=\sum_i\{\alpha Z_i \ln(Z_i/N_i) \\ -(N_i+\alpha Z_i)\ln[(Z_i+\alpha N_i)/N_i]\}. \end{aligned} \quad \text{(D1.28)}$$

Let $N_{ij;i'j'}$ be the number of transitions per unit time where the groups Z_i and Z_j each lose one particle and the groups $Z_{i'}$ and $Z_{j'}$ each gain one particle. Such transitions will be possible only if energy is conserved or if

$$E_i+E_j=E_{i'}+E_{j'}. \quad \text{(D1.29)}$$

It will be shown in Appendix V that the $N_{ij;i'j'}$ satisfy the equations

$$N_{ij;i'j'}=A_{ij;i'j'}N_iN_j(Z_{i'}+\alpha N_{i'})(Z_{j'}+\alpha N_{j'}), \quad \text{(D1.30)}$$

while we shall assume [compare Eq. (D1.05)] that

$$A_{ij;i'j'}=A_{i'j';ij}. \quad \text{(D1.31)}$$

We get now, in a manner completely analogous to the one leading to Eq. (D1.14),[80]

$$\begin{aligned} dH/dt&=\sum_i (dN_i/dt)\ln[N_i/(Z_i+\alpha N_i)] \\ &=\tfrac{1}{2}\sum_{ij;i'j'} A[N_{i'}N_{j'}(Z_i+\alpha N_i)(Z_j+\alpha N_j) \\ &\quad -N_iN_j(Z_{i'}+\alpha N_{i'})(Z_{j'}+\alpha N_{j'})] \\ &\quad \times\ln[N_iN_j(Z_{i'}+\alpha N_{i'})(Z_{j'}+\alpha N_{j'})/ \\ &\qquad N_{i'}N_{j'}(Z_i+\alpha N_i)(Z_j+\alpha N_j)]\leqslant 0. \end{aligned} \quad \text{(D1.32)}$$

The equal sign only holds when Eq. (D1.27) is satisfied.

Once again we have arrived at an equation showing the steady decrease of H until equilibrium has been reached. Both Eq. (D1.32) and Eq. (D1.14) are derived by using a *Stosszahlansatz* and by considering transition probabilities. The discussion is thus on the same level as that of Sec. A(3) and we could also, in the present, case consider the frequency with which we can expect fluctuations to occur. Just as the use of Eq. (A3.01) for the probability of the occurrence of a situation was based on the choice of equal *a priori* probabilities for

---

[78] For the importance of the omission of the factor $N!$ see also Gibbs' monograph (02G, Chap. XV) and a paper by Nordheim (24N).

[79] We follow here the considerations of Pauli (28P); Compare also the considerations by Nordheim (28N). Nordheim is especially interested in the distribution of electrons in metals (compare also 29P).

[80] It must be noted that $Z_i+\alpha N_i$ is never negative, since in the F-D case $N_i\leqslant Z_i$.

equal volumes in phase space, so the use of Eqs. (D1.21) to (D1.23) is based on the assumption of equal *a priori* probabilities for all nondegenerate quantum states as can be seen from the derivation of these equations. This assumption will be discussed in Sec. D(3) when we also shall discuss how far Eq. (D1.31) can be justified from first principles.[81]

(2) *H*-Theorem in Quantum Ensemble Theory[82]

The considerations of the preceding section corresponded to the considerations of Sec. A(3) in the classical case. We shall now be concerned with the quantum-mechanical analogy to Part C. First of all, in the present section we shall consider the fine-grained density. We shall then introduce a coarse-grained density and consider again a quantity Σ which is related to the Σ of Sec. C(1), and we draw attention to a few points where slight differences from the classical case occur. After that we shall briefly consider the influence of measurements.[83] The question of representative ensembles and of *a priori* probabilities is deferred until Sec. D(3).

It is well known that the so-called density matrix $\boldsymbol{\rho}$[84] introduced by von Neumann (27N1; see also 27N2, 29D, 30D1, 30D2, 31D, 32N3, 33P, 35D, 36D, 37K, 40H, 54T, ESM Chapter VII) replaces in quantum statistics the role of the ensemble density ρ in classical statistics. This density matrix also plays an important part in many quantum-mechanical discussions, but this is not the place to discuss the use of the density matrix for anything but statistical purposes.

The (fine-grained) density matrix $\boldsymbol{\rho}$ is introduced as follows. Let ψ^k be the normalized wave function describing the kth system of the ensemble and let φ_n be a complete set of orthonormal functions in the Hilbert space of the wave functions.[85] We can then expand the ψ^k as follows:

$$\psi^k=\sum_n a_n{}^k\varphi_n, \tag{D2.01}$$

where

$$a_n{}^k=\int\varphi_n{}^*\psi^k d\tau, \tag{D2.02}$$

where a star indicates the conjugate complex and where $d\tau$ is a volume element of the coordinate space.[86]

We can use the $a_n{}^k$ instead of the ψ^k to describe the kth system. The two representations are equivalent and, while the ψ^k satisfy the Schrödinger equation

$$\mathbf{H}\psi^k=i\hbar\dot{\psi}^k, \tag{D2.03}$$

the $a_n{}^k$ satisfy the transformed Schrödinger equation

$$i\hbar\dot{a}_n{}^k=\sum_m H_{nm}a_m{}^k, \tag{D2.04}$$

where the H_{mn} are the matrix elements of the Hamiltonian operator $\mathbf{H}$,

$$H_{nm}=\int\varphi_n{}^*\mathbf{H}\varphi_m d\tau. \tag{D2.05}$$

The physical significance of the $a_n{}^k$ is that they are probability amplitudes, $|a_n{}^k|^2$ being the probability that the kth system is characterized by the function φ_n. From the normalization of the ψ^k and the fact that the φ_n form a complete orthonormal set, it follows that

$$\sum_n |a_n{}^k|^2=1. \tag{D2.06}$$

We can now introduce the density operator $\boldsymbol{\rho}$ or *density matrix* by defining it through its elements,

$$\rho_{mn}=\frac{1}{N}\sum_{k=1}^{N}a_m{}^k a_n{}^{k*}, \tag{D2.07}$$

where N is the number of systems in the ensemble.

The average value $\langle G\rangle$ of a physical quantity G, which now corresponds to an operator $\mathbf{G}$, is given by the equation

$$\langle G\rangle=N^{-1}\sum_k\int\psi^{k*}\mathbf{G}\psi^k d\tau, \tag{D2.08}$$

which is twice an average, firstly an average over the wave function, and secondly an average over the ensemble.

Introducing the a_n, we get[87]

$$\begin{aligned}\langle G\rangle&=N^{-1}\sum_k\sum_{m,n}a_m{}^k a_n{}^{k*}G_{nm}\\&=\sum_{m,n}\rho_{mn}G_{nm}=\mathrm{Tr}(\boldsymbol{\rho}\mathbf{G}).\end{aligned} \tag{D2.09}$$

An example of Eq. (D2.09) is the normalization of $\boldsymbol{\rho}$. Putting $\mathbf{G}=1$, we have

$$1=\mathrm{Tr}\boldsymbol{\rho}, \tag{D2.10}$$

an equation which also follows directly from Eq. (D2.06).

The index of probability $\mathbf{n}$ is again related to $\boldsymbol{\rho}$ by the equation

$$\boldsymbol{\rho}=\exp\mathbf{n}, \tag{D2.11}$$

where the right-hand side is an abbreviated notation for the infinite exponential power series:

$$\exp\mathbf{n}=\sum(n!)^{-1}\mathbf{n}^n. \tag{D2.12}$$

At this point we may draw attention to the fact that the introduction of grand ensembles in quantum statistics does not involve any extra complications.

---

[81] Recently Stueckelberg and co-workers (52S2, 54I) have based the decrease of H on the unitarity of the so-called $\mathbf{S}$-matrix.

[82] Compare 38T, Chap. XII, part A, which we follow to a large extent in our discussion.

[83] For a thorough discussion of this subject we refer to von Neumann's book (32N3) and a recent paper by Groenewold (46G2).

[84] We denote all matrices by bold face characters.

[85] For the sake of simplicity we shall assume that the φ_n form a discrete set so that in Eq. (D2.01) we are dealing with a sum. It is not difficult to extend the formulas to the case where the sum has to be replaced by an integral (a Stieltjes integral, if necessary) but we leave this to the reader.

[86] We restrict ourselves in our formulas to the case where the wave functions are scalars. It is not difficult to extend our formulas to the case where they are spinors or undors (39B1).

[87] TrA is a shorthand notation for the trace, or spur, of an operator, that is, the sum of its diagonal elements: $\mathrm{Tr}\mathbf{A}=\sum_n A_{nn}$.

We can assume that second quantization has taken place[88] and that thus all wave functions are expressed as products of Jordan-Klein and Jordan-Wigner matrices. The formalism remains unchanged.

We now introduce a quantity σ by the equation

$$\sigma = \langle \eta \rangle, \tag{D2.13}$$

or, using Eqs. (D2.11) and (D2.09)

$$\sigma = \mathrm{Tr}(\boldsymbol{\rho} \ln \boldsymbol{\rho}). \tag{D2.14}$$

As in Sec. C(1) we can show that σ possesses a number of minimum properties (compare 36D). In order to show this, we first choose as our complete orthonormal set φ_n a complete set of characteristic functions of the Hamiltonian operator and of the number operator.[89] We shall indicate the state corresponding to ν particles while the energy of the system is $E_k(\nu)$[90] by a double index k, ν and matrix elements between two states, say k, ν_1 and l, ν_2 by $(k,\nu_1; l,\nu_2)$ for instance $H(k,\nu_1; l,\nu_2)$.

We have now the following properties of σ.

(a) If our ensemble is such that all systems have the same number of particles N and that the energy of all systems lies within a given interval E, $E+\delta E$, σ will be minimum if $\boldsymbol{\rho}$ is given by the equation

$$\left.\begin{aligned} &\rho(k,\nu_1; l,\nu_2) = c\delta_{kl}\delta_{\nu_1 N}\delta_{\nu_2 N} \\ &\qquad\qquad \text{for} \quad E \leqslant E_k(N) \leqslant E+\delta E; \\ &\rho(k,\nu_1; l,\nu_2) = 0 \quad \text{otherwise,} \end{aligned}\right\} \tag{D2.15}$$

where δ_{kl} is the Kronecker symbol and where c is a constant.

(b) If our ensemble is such that all systems have the same number of particles N but that only the average energy is given, that is, that $\boldsymbol{\rho}$ must satisfy the condition

$$\mathrm{Tr}(\boldsymbol{\rho}\mathbf{H}) = E, \tag{D2.16}$$

σ will be minimum, if $\boldsymbol{\rho}$ satisfies the equation

$$\rho(k,\nu_1; l,\nu_2) = \delta_{kl}\delta_{\nu_1 N}\delta_{\nu_2 N} \exp\{\beta[\psi - E_k(N)]\}. \tag{D2.17}$$

(c) If only the average number of particles and the average energy of the systems in the ensemble are given so that $\boldsymbol{\rho}$ must satisfy Eq. (D2.16) and the condition

$$\mathrm{Tr}(\boldsymbol{\rho}\boldsymbol{\nu}) = N, \tag{D2.18}$$

where $\boldsymbol{\nu}$ is the number operator, σ will be minimum, if $\boldsymbol{\rho}$ satisfies the equation[91]

$$\rho(k,\nu_1; l,\nu_2) = \delta_{kl}\delta_{\nu_1\nu_2} \exp[-q + \nu_1\mu - \beta E_k(\nu_1)]. \tag{D2.19}$$

We shall prove only the last statement since the proofs of the two other statements are completely analogous. As in Sec. C(1), the densities given by Eqs. (D2.15), (D2.17), and (D2.19) correspond to our energy-shell ensemble, a macrocanonical ensemble, and a canonical grand ensemble and β, ψ, and μ have here the same meaning as there.

The proof of (c) is as follows (compare 36D Sec. 8): Compare two density matrices $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$, where $\boldsymbol{\rho}_1$ is given by Eq. (D2.19) and $\boldsymbol{\rho}_2$ by the equation

$$\boldsymbol{\rho}_2 = \boldsymbol{\rho}_1 \exp(\Delta\boldsymbol{\eta}). \tag{D2.20}$$

The quantity $\sigma_2 - \sigma_1$ is given by the equation

$$\sigma_2 - \sigma_1 = \mathrm{Tr}(\boldsymbol{\eta}_2 \exp\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1 \exp\boldsymbol{\eta}_1). \tag{D2.21}$$

The right-hand side of Eq. (D2.21) can be written as

$$\begin{aligned} &\mathrm{Tr}[(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1) \exp\boldsymbol{\eta}_2] + \mathrm{Tr}[\boldsymbol{\eta}_1(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1)] \\ &\quad = \mathrm{Tr}[(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1) \exp\boldsymbol{\eta}_2] + \mathrm{Tr}[(-q + \boldsymbol{\nu}\mu - \beta\mathbf{H})(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1)] \\ &\qquad = \mathrm{Tr}[(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1) \exp\boldsymbol{\eta}_2], \end{aligned}$$

where we have used the fact that both $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ satisfy Eqs. (D2.16), (D2.18), and the normalization condition (D2.10), and where we have used Eqs. (D2.19) and (D2.11). We have thus

$$\sigma_2 - \sigma_1 = \mathrm{Tr}[(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1) \exp\boldsymbol{\eta}_2]. \tag{D2.22}$$

The right-hand side of Eq. (D2.22) is positive, except when $\boldsymbol{\eta}_2 = \boldsymbol{\eta}_1$. This can be seen as follows. We first of all note that for any Hermitean operator $\mathbf{A}$ the following inequality holds (compare 38P):

$$[\exp\mathbf{A}]_{kk} \geqslant \exp A_{kk}, \tag{D2.23}$$

the equal sign only holding when $\mathbf{A}$ is on diagonal form.

One proves Eq. (D2.23) by considering a unitary transformation $\mathbf{U}$ which brings $\mathbf{A}$ onto diagonal form with diagonal elements (characteristic values) A_k. We then have

$$A_{kk} = \Sigma_k |U_{kl}|^2 A_l, \quad \Sigma_k |U_{kl}|^2 = 1, \tag{D2.24}$$

and

$$\begin{aligned} [f(\mathbf{A})]_{kk} &= \Sigma_k |U_{kl}|^2 f(A_l) \\ &= \Sigma |U_{kl}|^2 \{f(A_{kk}) + (A_l - A_{kk}) f'(A_{kk}) \\ &\quad + \tfrac{1}{2}(A_l - A_{kk})^2 f''(A_{kk} + \theta_{kl}[A_l - A_{kk}])\} \; (0 \leqslant \theta_{kl} \leqslant 1) \\ &= f(A_{kk}) + \tfrac{1}{2}\Sigma |U_{kl}|^2 (A_l - A_{kk})^2 f'', \end{aligned} \tag{D2.25}$$

where we have used Eqs. (D2.24). If $f(\mathbf{A})$ is the exponential function, Eq. (D2.23) follows.

We now choose a representation in which $\boldsymbol{\eta}_2$ is on diagonal form. Since from Eq. (D2.10) it follows that

$$\mathrm{Tr}(\exp\boldsymbol{\eta}_1) = \mathrm{Tr}(\exp\boldsymbol{\eta}_2)(=1), \tag{D2.26}$$

we get

$$\begin{aligned} &\mathrm{Tr}[(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1) \exp\boldsymbol{\eta}_2] \\ &\quad = \mathrm{Tr}[\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1) \exp\boldsymbol{\eta}_2 - \exp\boldsymbol{\eta}_2 + \exp\boldsymbol{\eta}_1] \\ &\quad = \sum[(\eta_{2i} - \eta_{1ii}) \exp\eta_{2i} - \exp\eta_{2i} + (\exp\boldsymbol{\eta}_1)_{ii}] \\ &\qquad \geqslant \sum[(\eta_{2i} - \eta_{1ii}) \exp\eta_{2i} - \exp\eta_{2i} + \exp\eta_{1ii}] \\ &\quad = \sum(\exp\eta_{1ii})[(\eta_{2i} - \eta_{1ii} - 1) \exp(\eta_{2i} - \eta_{1ii}) + 1] \geqslant 0, \end{aligned}$$

where we have first used Eq. (D2.23) and then used the properties of the function y given by Eq. (C1.13).

---

[88] For an account of the method of second quantization (28J; see also 27J2, 34F) we refer to Kramers' monograph (38K Sec. 72).

[89] Once again we shall restrict ourselves to the case where only one kind of particle is present. In the present discussion it would, however, be simple to extend the discussion to the case of a number of different kinds of particles.

[90] We shall assume for the sake of simplicity that all energy levels are nondegenerate.

[91] In this case it is really unnecessary to make a particular choice for the φ_n. In general Eq. (D2.19) would read $\boldsymbol{\rho} = \exp[-q + \boldsymbol{\nu}\mu - \beta\mathbf{H}]$.

Once again, however, σ will not change with time. This follows from the fact that we can transform from one instant (t') to another (t'') by a unitary transformation. This means that, if $\boldsymbol{\varrho}'$ and $\boldsymbol{\varrho}''$ denote the density matrices at t' and t'', they are related by the equation

$$\boldsymbol{\varrho}''=\mathbf{U}^\dagger\boldsymbol{\varrho}'\mathbf{U}, \qquad \text{(D2.27)}$$

where $\mathbf{U}^\dagger$ is the Hermitean conjugate of $\mathbf{U}$ and where

$$\mathbf{U}^\dagger\mathbf{U}=\mathbf{U}\mathbf{U}^\dagger=\mathbf{1}. \qquad \text{(D2.28)}$$

Any average value $\langle G\rangle$ is invariant against a unitary transformation. This follows since $\mathrm{Tr}(\mathbf{AB})=\mathrm{Tr}(\mathbf{BA})$, and thus

$$\begin{aligned}\mathrm{Tr}(\boldsymbol{\varrho}''\mathbf{G}'')&=\mathrm{Tr}(\mathbf{U}^\dagger\boldsymbol{\varrho}'\mathbf{U}\mathbf{U}^\dagger\mathbf{G}'\mathbf{U})=\mathrm{Tr}(\mathbf{U}^\dagger\boldsymbol{\varrho}'\mathbf{G}'\mathbf{U})\\ &=\mathrm{Tr}(\mathbf{G}'\mathbf{U}\mathbf{U}^\dagger\boldsymbol{\varrho}')=\mathrm{Tr}(\mathbf{G}'\boldsymbol{\varrho}')=\mathrm{Tr}(\boldsymbol{\varrho}'\mathbf{G}'). \qquad \text{(D2.29)}\end{aligned}$$

Since σ is the average value of $\mathbf{n}$ we see that $\sigma'=\sigma''$ as in the classical case.

We shall now introduce as in Sec. C(1) a coarse-grained density, or in this case a coarse-grained density matrix $\mathbf{P}$. We choose now as our complete orthonormal set of functions the characteristic functions of the Hamiltonian $\mathbf{H}$ of the system. Furthermore, we shall divide the stationary states into groups, as before, but now we shall assume the grouping to be in such a way as to correspond to the inaccuracy of our observation, this is, we shall assume that by means of our available observational methods we can discriminate between the different groups, but not within them.[92] Let S_i be the number of levels in the ith group. We now define the coarse-grained density matrix $\mathbf{P}$ by its matrix elements in the particular representation which we are using at the moment, as follows:

$$\mathrm{P}_{kl}=\delta_{kl}\sum\nolimits_j \rho_{jj}/S_i, \qquad \text{(D2.30)}$$

where the energy level E_k belongs to the ith group and where the summation is taken over all the states of the ith group. Since ρ_{jj} is the probability of finding a system of the ensemble in the state characterized by φ_j, we see that $S_i\mathrm{P}_{kk}$ is the probability of finding a system of the ensemble in a state belonging to the ith group. Equation (D2.30) defines the coarse-grained density in a particular representation; its matrix elements in any other representation follow by the usual transformation rules.[93]

From Eqs. (D2.30) and (D2.10) it follows that $\mathbf{P}$ is also normalized,

$$\mathrm{Tr}\,\mathbf{P}=1. \qquad \text{(D2.31)}$$

We now define a quantity Σ by the equation

$$\Sigma=\mathrm{Tr}(\mathbf{P}\ln\mathbf{P})=\sum\nolimits_k \mathrm{P}_{kk}\ln\mathrm{P}_{kk}, \qquad \text{(D2.32)}$$

as long as we stay in the representation which we chose originally. Using Eq. (D2.30) we can also write

$$\Sigma=\sum \rho_{kk}\ln\mathrm{P}_{kk}=\mathrm{Tr}(\rho\ln\mathbf{P})=\langle\ln\mathbf{P}\rangle. \qquad \text{(D2.33)}$$

There are several ways of studying the behavior of Σ in time. Possibly the crudest way is the one used by Born and Green (48B1, 48B2, 49B1, 49B2, 49B3) who, in our opinion, do not properly distinguish between fine-grained and coarse-grained densities and, moreover, do not take into account the fact that coarse-grained densities are introduced as a necessary step to keep our formalism in accordance with our experimental possibilities. In their work they define as the entropy a quantity which is not proportional to either $-\sigma$ or $-\Sigma$, but to a hybrid of the two, which we shall call $-\sigma\Sigma$. The quantity $\sigma\Sigma$ can be obtained from Σ by putting all S_i equal to one, or from σ by assuming $\boldsymbol{\varrho}$ to be always diagonalized,

$$\sigma\Sigma=\sum \rho_{kk}\ln\rho_{kk}. \qquad \text{(D2.34)}$$

We shall see later on that $\sigma\Sigma$ and Σ are sometimes the same, namely, for an ensemble representing a system on which a measurement has just been made.

The second step in their discussion consists in observing that one is usually interested in systems which are in contact with their surroundings so that the Hamiltonian of the system consists of two parts,

$$\mathbf{H}=\mathbf{H}_0+\mathbf{V}, \qquad \text{(D2.35)}$$

where $\mathbf{H}_0$ is the Hamiltonian of the system if interaction with the outside world is neglected, and $\mathbf{V}$ the interaction energy operator. It is a well-known result from quantum-mechanical perturbation theory that transitions between characteristic states of $\mathbf{H}_0$ can occur under the influence of $\mathbf{V}$. Choose now for the φ_n the characteristic functions of $\mathbf{H}_0$ and of any other operators which commute with $\mathbf{H}_0$. For reasons which will become clear presently [compare Eqs. (D2.37) and (D2.41)], we are especially interested in the case where the energy characteristic values are degenerate. To indicate this degeneracy we shall give two indices to the φ_n: $\varphi_{k,l}$, where the first index indicates the energy value E_k, and the second index distinguishes between the various states belonging to the same E_k.

Since we are only dealing with the diagonal elements of $\boldsymbol{\varrho}$ which are, as we saw a moment ago, the probabilities of finding a system of the ensemble in a particular state, we can use the quantum-mechanical formulas for the time dependence of these probabilities, and we find[94]

$$\begin{aligned}\rho(kl,kl;t)=\rho(kl,kl;t_0)&+\sum\nolimits_{k'l'} J(kl,k'l')\\ &\times\{\rho(k'l',k'l';t_0)-\rho(kl,kl;t_0)\}, \qquad \text{(D2.36)}\end{aligned}$$

where

$$J(kl,k'l')=\delta_{kk'}(4\pi^2/h)\,|V(kl,k'l')|^2(t-t_0). \qquad \text{(D2.37)}$$

In Eq. (D2.37) the $\delta_{kk'}$ shows that J is zero unless energy is conserved, that is, transitions only take place

---

[92] There is a difference between the grouping here and in Sec. D(1). Previously we grouped the energy levels of the constituent particles but now we group the energy levels of the whole system.

[93] For a discussion of the choice of the coarse-grained density matrix see also a recent paper by van Kampen (54K).

[94] Compare the discussion in Appendix V.

between states of the same energy, h is Planck's constant, and $V(kl,k'l')$ is the matrix element of $\mathbf{V}$ which satisfies the equation

$$V(kl,k'l')=V^*(k'l',kl), \tag{D2.38}$$

because $\mathbf{V}$ is a Hermitean matrix. By virtue of Eq. (D2.38) we have the relation

$$J(kl,k'l')=J(k'l',kl). \tag{D2.39}$$

Substituting from Eq. (D2.36) into Eq. (D2.34), we get up to the first order in J[95]

$$\sigma\Sigma(t)=\sigma\Sigma(t_0)-\tfrac{1}{2}\sum_{kl,\,k'l'}J(kl,k'l')Q(kl,k'l') \tag{D2.40}$$

with

$$\begin{aligned}Q(kl,k'l')=&[\rho(kl,kl;t_0)-\rho(k'l',k'l';t_0)]\\ &\times\ln[\rho(kl,kl;t_0)/\rho(k'l',k'l';t_0)].\end{aligned} \tag{D2.41}$$

As Q is never negative and J always positive and, moreover, a monotonically increasing function of t, we have in $\sigma\Sigma$ a monotonically decreasing function.

One can object to this argument on the grounds that Eq. (D2.37) is only valid for small values of $t-t_0$, but Green has shown that the argument can be changed so as to be valid for all values of $t-t_0$. The most important objection is, however, against the use of the function $-\sigma\Sigma$ as a measure for the entropy. As was pointed out by Pauli (49P) as soon as this choice is made everything else follows as can be seen from the work of Klein (31K) which we discuss in Appendix VI (see also later in this section).

A second method of discussing the behavior of Σ the one due to Pauli (28P). Let P_i be the probability of finding a system of the ensemble in the ith group. We then have the following relation[96]:

$$P_i=S_iP_{kk}, \tag{D2.42}$$

where P_{kk} is one of the S_i equal diagonal elements belonging to the ith group. From Eq. (D2.32) we then have in terms of the P_i

$$\Sigma=\sum_i\sum(P_i/S_i)\ln(P_i/S_i)=\sum_i P_i\ln(P_i/S_i), \tag{D2.43}$$

where in the second member the first summation is over all groups and the second summation over the S_i members of the group.

The time derivation of Σ is now given by the equation

$$d\Sigma/dt=\sum_i(\ln P_i-\ln S_i)(dP_i/dt), \tag{D2.44}$$

where we used the fact that $\sum P_i=1$, or $\sum dP_i/dt=0$ as follows from Eqs. (D2.31) and (D2.42).

Let N_{ij} be the average transition probability from group S_i to group S_j. In Appendix V we shall prove that N_{ij} is given by the equation

$$N_{ij}=A_{ij}S_jP_i, \tag{D2.45}$$

and we shall assume

$$A_{ij}=A_{ji}. \tag{D2.46}$$

We then get

$$dP_i/dt=\sum_j(N_{ji}-N_{ij})=\sum_j A_{ij}[S_iP_j-S_jP_i], \tag{D2.47}$$

and in much the same manner as we derived Eq. (D2.40) we get now

$$d\Sigma/dt=\tfrac{1}{2}\sum_{i,\,j}A_{ij}[P_jS-P_iS_j]\ln(P_iS_j/P_jS_i). \tag{D2.48}$$

Once again it follows from the form of the right-hand side of equation (D2.48) that $d\Sigma/dt$ will never be positive.

We may at this point draw attention to the fact that $d\Sigma/dt$ will only be zero if

$$P_jS_i=P_iS_j,\quad\text{or,}\quad P_j/S_j=P_i/S_i, \tag{D2.49}$$

or, when use is made of Eqs. (D2.45) and (D2.46),

$$N_{ij}=N_{ji}, \tag{D2.50}$$

and we see that at equilibrium there are as many transitions from group S_i to group S_j as the other way round. This is an example of the so-called *principle of detailed balancing* which we shall discuss in more detail in Appendix VII.

We shall return presently to the discussion of this second method of considering the variation with time of Σ, but we want first to discuss a third method which is completely analogous to the consideration of the variation with time of Σ in Sec. C(1).[97]

Let us assume that we have made observations on our system at t' which will give us the coarse-grained density $\mathbf{P}$. The fine-grained density $\boldsymbol{\varrho}$ at t' will then be given by the equation [compare the discussion in Sec. D(3)].

$$\boldsymbol{\varrho}'=\mathbf{P}', \tag{D2.51}$$

while we have for Σ'

$$\Sigma'=\mathrm{Tr}(\mathbf{P}'\ln\mathbf{P}')=\mathrm{Tr}(\boldsymbol{\varrho}'\ln\boldsymbol{\varrho}'). \tag{D2.52}$$

If the situation at t' did not correspond to an equilibrium situation so that $\boldsymbol{\varrho}'$ (or $\mathbf{P}'$) did not satisfy one of the Eqs. (D2.15), (D2.17), or (D2.19), we cannot expect Eq. (D2.51) to hold at a later time t'', but we would have

$$\boldsymbol{\varrho}''\neq\mathbf{P}'', \tag{D2.53}$$

and as a result

$$\Sigma''<\Sigma'. \tag{D2.54}$$

Let us consider in slightly more detail how Eq. (D2.54) follows from Eq. (D2.53).[98] We are concerned with the expression

$$\begin{aligned}\Sigma'-\Sigma''&=\mathrm{Tr}(\mathbf{P}'\ln\mathbf{P}')-\mathrm{Tr}(\mathbf{P}''\ln\mathbf{P}'')\\ &=\mathrm{Tr}(\boldsymbol{\varrho}'\ln\boldsymbol{\varrho}')-\mathrm{Tr}(\boldsymbol{\varrho}''\ln\mathbf{P}'')\\ &=\sum_k\rho_{kk}'\ln\rho_{kk}'-\sum_n\rho_{nn}''\ln\mathrm{P}_{nn}'',\end{aligned} \tag{D2.55}$$

---

[95] We use here Eq. (D2.39) (i) to show that the sum

$$\sum J(kl,k'l')[\rho(k'l',k'l';t)-\rho(kl,kl;t)]$$

is zero, and (ii) to obtain the sum on the right-hand side of Eq. (D2.40) [compare the derivation of Eq. (D1.14); hence the factor $\tfrac{1}{2}$].

[96] Compare the discussion of Eq. (D2.30).

[97] I should like to express my thanks to Professor O. Klein for a discussion of some points connected with this method of approach.

[98] The discussion on p. 373 of ESM is too short and rather misleading.

where we have used the fact that the coarse-grained density is always a diagonal matrix [see Eq. (D2.30)]. We add now to the right-hand side of equation (D2.55) first of all the expression

$$\textstyle\sum_n \rho_{nn}'' \ln\rho_{nn}'' - \sum_k \rho_{kk}' \ln\rho_{kk}', \qquad \text{(D2.56)}$$

which by Klein's lemma (31K) is never positive and only equal to zero, if $\boldsymbol{\varrho}''$ is a diagonal matrix as is proved in Appendix VI. We also add to the right-hand side of Eq. (D2.55) the expression $\mathrm{Tr}\,\mathbf{P}'' - \mathrm{Tr}\,\boldsymbol{\varrho}''$ which is zero since both $\boldsymbol{\varrho}''$ and $\mathbf{P}''$ are normalized. The final result is

$$\begin{aligned}\Sigma' - \Sigma'' = \textstyle\sum_n (\rho_{nn}'' \ln\rho_{nn}'' - \rho_{nn}'' \ln \mathrm{P}_{nn}'' \\ - \rho_{nn}'' + \mathrm{P}_{nn}'') \geqslant 0, \qquad \text{(D2.57)}\end{aligned}$$

the last inequality following from the properties of the function y given by Eq. (C1.13), since substituting $x = \ln\alpha - \ln\beta$ and multiplying by β gives for $\beta > 0$

$$\left.\begin{aligned}\alpha \ln\alpha - \alpha \ln\beta - \alpha + \beta > 0, \quad \alpha \neq \beta; \\ = 0, \quad \alpha = \beta.\end{aligned}\right\} \qquad \text{(D2.58)}$$

The fact that we expect Σ to continue to decrease is based on considerations similar to those discussed in Sec. C(1). This decrease will continue until a stationary state is reached where $\mathbf{P}$ is given by one of the three Eqs. (D2.15), (D2.17), or (D2.19), and where $\boldsymbol{\varrho} = \mathbf{P}$.

We must now discuss the different ways of considering the time dependence of Σ. As we mentioned before, we are not convinced of the validity of Born and Green's method of approach and we shall not discuss it any further at this point. The second method of approach is the one used by Pauli. This method is actually closely connected with the simple approach of Sec. D(1). This we can see as follows. We write Σ of Eq. (D2.43) in the form

$$\Sigma = \textstyle\sum_i P_i \ln P_i - \sum_i P_i \ln S_i. \qquad \text{(D2.59)}$$

Now, P_i is the probability of finding a system in one of the states corresponding to the ith group, and S_i is the number of states in this group. The number of states in a group is, however, just equal to the function $W(N_i)$ discussed in Sec. D(1) as will be seen after a moment's thought, and using the relation between H and $\ln W$ which led to Eq. (D1.28) we can write Eq. (D2.59) in the form

$$\Sigma = \textstyle\sum_i P_i H_i + \sum_i P_i \ln P_i = \langle H \rangle + \Sigma_i P_i \ln P_i, \qquad \text{(D2.60)}$$

where H_i is the value of H if the system is in a state belonging to the ith group,[99] and $\langle H \rangle$ the average of H over the ensemble.

Using Eq. (D2.60) we see first of all that if we are dealing with an ensemble representing a system which by measurement has just been found to be in one particular state so that one of the P_i is equal to one and the others equal to zero, this equation would reduce to the form

$$\Sigma = H. \qquad \text{(D2.61)}$$

Secondly, taking the time derivation of Eq. (D2.60), we get

$$d\Sigma/dt = \langle dH/dt \rangle + \textstyle\sum_i (dP_i/dt) \ln P_i, \qquad \text{(D2.62)}$$

and we see that the decrease of Σ is due to two facts (i) the general decrease in H for each system of the ensemble, and (ii) the tendency to decrease $\Sigma_i P_i \ln P_i$ by a more uniform distribution of the systems in the ensemble over the different groups of states.

Let us now consider Eq. (D2.57). We saw that Σ decreased firstly because of the fact that $\boldsymbol{\varrho}$ and $\mathbf{P}$ were no longer the same, a cause for the decrease of Σ which also operated in the classical case, and secondly because the expression $\Sigma \rho_{kk} \ln\rho_{kk}$ decreased (Klein's lemma). For the second cause there does not exist a classical counterpart and we must discuss it briefly. It is called by Tolman (38T) the *quantum-mechanical spreading of fine-grained probability*. There is no classical counterpart since the quantity $\Sigma \rho_{kk} \ln\rho_{kk}$ is not a trace of a quantum-mechanical operator. The more states are occupied the lower the value of this sum becomes. This decrease is associated with the irreversible disturbance due to observations on the system.[100] The effect will thus be most pronounced in the case where observations are such as to approach the limits set by the Heisenberg relation, but will be unimportant in those cases where the disturbance of the system by the observation may be neglected. In the latter case the growing difference between $\boldsymbol{\varrho}$ and $\mathbf{P}$ will be responsible for the decrease of Σ.

It is of interest to note that some authors (36D, 37S) only consider this influence of measurements on the system without properly paying attention to the role of fine-grained and coarse-grained densities and to the fact that the approach to equilibrium should be independent of whether or not the system is actually observed.[101]

In the next section we shall discuss how one can set up representative ensembles in quantum statistics, but some aspects of the consequences of observations on a physical system we wish to discuss briefly now, especially the importance of the pure case.

In classical statistics we introduce ensembles since our knowledge about the physical systems under consideration is practically always far removed from the maximum attainable knowledge. In quantum statistics the situation is, however, complicated due to the statistical aspects inherent in quantum mechanics. In the ideal (and always idealized) case we may, in classical statistics, know the values of all constants of motion so that the problem of the behavior of the system reduces to a problem of classical mechanics. In the

---

[99] We may remark here that we could have derived an equation similar to Eq. (D2.60) in the classical case. The discussion would have been analogous to the present discussion.

[100] Compare, for instance, the discussion by Bohr (33B1, 33B2).

[101] Compare the criticism raised by Pauli and Fierz (37P). We also may refer to a paper by Davydov (47D).

ideal (and always idealized) case in quantum statistics, we may know that one system is in a characteristic state of a quantum-mechanical operator for which for the sake of simplicity we shall choose the energy operator[102] so that we are dealing with the system in a stationary state. We are dealing with a *pure case* (*Reiner Fall*) and all the systems in the representative ensemble will possess the same wave function, namely, the one corresponding to the characteristics state in question. Only one of the two averaging processes described by Eq. (D2.08) remains, namely, the quantum-mechanical average. We still must use statistical considerations, although our knowledge is maximal.

In general, however, the situation will not be so favorable and the construction of the representative ensemble is much more complicated. Both averages in Eq. (D2.08) must be taken into account. We are dealing with a *mixed case* (*Gemisch*).

In the pure case, if the state of the system is characterized by one of the functions forming the complete orthonormal set φ_n, say φ_l, the ψ^k are all equal to φ_l, and we have

$$a_n{}^k=\delta_{nl}, \tag{D2.63}$$

and hence

$$\rho_{mn}=\delta_{mn}\delta_{ml}. \tag{D2.64}$$

In this case we have for σ

$$\sigma=\mathrm{Tr}(\boldsymbol{\varrho}\ \ln\boldsymbol{\varrho})=0. \tag{D2.65}$$

In a mixed case, there are more than one ρ_{mn} which are different from zero and σ is smaller than zero.[103] This is probably the reason by Elsasser (37E) calls σ the *mixture index*.[104] The decrease of σ when we go from a pure case to a mixed case once again shows the relation between entropy—which is proportional to $-\sigma$—and lack of information as the decrease of detailed information involves the decrease of σ (compare also the discussion in 39K1, 39K2).

In the next section we shall see that the representative ensemble of a subsystem which is part of a large isolated system will be a canonical grand ensemble. This means that while the larger system may be represented by a pure case, the subsystems will certainly correspond to a mixture. This fact has been the source of a long discussion started off by a paper by Einstein, Podolski, and Rosen (35E) in which the paradox which bears their names was first stated. We have not got the space to discuss this paradox and refer to the literature (35B, 35S1, 35S2, 35S3, 35S4, 36M, 36F, 36S, 51B3; compare also 27N1).

(3) Representative Ensembles

Once again the situation is very similar to the situation in classical statistics. We have only incomplete data from which to predict the future behavior of our system and we must have recourse to representative ensembles. Once again the question of *a priori* probabilities crops up, and as we really need the $a_n{}^k$'s, also the question of the *a priori* phases, as we shall see presently. We use the assumption of equal *a priori* probabilities for all nondegenerate states[105] and the assumption of random *a priori* phases, where we shall see presently what we mean by the term random in this connection. Our discussion will now proceed analogous to the discussion in Sec. C(2). We assume again that our system is part of a larger, isolated system and instead of Eqs. (C2.01) to (C2.04) we have now

$$\mathrm{Tr}\ \boldsymbol{\varrho}=1, \tag{D3.01}$$

$$\mathrm{Tr}\ \boldsymbol{\varrho}'=1, \tag{D3.02}$$

$$\mathrm{Tr}(\mathbf{H}\boldsymbol{\varrho})+\mathrm{Tr}(\mathbf{H}'\boldsymbol{\varrho}')=\text{constant}, \tag{D3.03}$$

$$\mathrm{Tr}(\boldsymbol{\nu}\boldsymbol{\varrho})+\mathrm{Tr}(\boldsymbol{\nu}'\boldsymbol{\varrho}')=\text{constant}, \tag{D3.04}$$

where the unprimed quantities refer to the subsystem and the primed quantities to the rest system. If equilibrium has been attained Σ should be a minimum, or,

$$\mathrm{Tr}[(\boldsymbol{\varrho}\boldsymbol{\varrho}')\ \ln(\boldsymbol{\varrho}\boldsymbol{\varrho}')]=\text{minimum}. \tag{D3.05}$$

Using the method of undetermined multipliers, one then finds (see, e.g., 40H for a discussion[106]) for the equilibrium expression for

$$\boldsymbol{\varrho}=\exp[-q+\boldsymbol{\nu}\mu-\beta\mathbf{H}], \tag{D3.06}$$

in accordance with what we discussed in the previous section. As in Sec. C(2), Eq. (D3.06) really only holds for the coarse-grained density $\mathbf{P}$ as Eq. (D3.05) only holds for Σ and not for σ.

The situation is also completely analogous to the classical case as far as nonequilibrium situations are concerned and we refer to the discussion of Sec. C(2) which can be immediately applied to the present case.

We wish to make one more remark about the setting up of a representative ensemble after an observation has been made on a system. We can only distinguish between states belonging to different groups as we discussed before and we can thus only determine the coarse-grained density, or, the P_{kk}. In accordance with the assumption of equal *a priori* probabilities for all nondegenerate states we should then put all diagonal elements of $\boldsymbol{\varrho}$ corresponding to one group equal to each other and equal to the corresponding P_{kk}. As far as the off-diagonal elements of $\boldsymbol{\varrho}$ are concerned, since these are

---

[102] It is easy to extend the discussion to include the case where the system is in a characteristic state of two or more commuting operators.

[103] Equation (D2.65) is a necessary and sufficient condition for a pure case (33P) and is equivalent to the usual condition $\boldsymbol{\varrho}^2=\boldsymbol{\varrho}$.

[104] In his paper Elsasser discusses how one can use measurements to set up representative ensembles. Instead of making assumptions about *a priori* probabilities Elsasser chooses his ensembles in such a way that they make σ minimum. In this way he, of course, arrives at the same results as we do, in the more conventional way, since the normal method leads to minimum values of σ.

[105] If a state is g-fold degenerate, it is counted as g nondegenerate states.

[106] One varies both the characteristic values and the characteristic functions of $\boldsymbol{\varrho}$ in this case.

the average values of the product of two a_n^k we have

$$\rho_{mn}=\frac{1}{N}\sum a_m^k a_n^{k*} \quad m\neq n$$
$$=\frac{1}{N}\sum r_m^k r_n^k \exp[i(\phi_m^k-\phi_n^k)], \qquad \text{(D3.07)}$$

where r_m^k is the modulus of a_n^k and ϕ_m^k its phase.

Using the assumption of random *a priori* phases, we may expect the average of $\exp[i(\phi_m^k-\phi_n^k)]$ to be zero, and we have thus for the fine-grained density of our ensemble

$$\rho_{kl}=P_{kk}\delta_{kl}, \qquad \text{(D3.08)}$$

and we see that, indeed, as we remarked when discussing Eqs. (D2.34), σ, Σ, and $\sigma\Sigma$ are all three equal.

We come now to a discussion of our choice of *a priori* probabilities and *a priori* phases. When we discussed the analogous situation in classical statistics we showed first of all that the *a priori* weights should be such as to be invariant against the movement of the representative point along its orbit. We thus showed that in the case of quasi-ergodic systems, the *a priori* weights would be functions of the energy only and could thus be taken to be equal for equal volumes in phase space. The situation in quantum statistics is both simplified and complicated as compared with the situation in classical statistics. We shall first show that an ensemble constructed without any conditions according to the prescription of equal *a priori* probabilities and random *a priori* phases for different nondegenerate quantum states will be a stationary ensemble and will, moreover, correspond to a similar ensemble if we change our representation. Secondly we shall briefly discuss in this connection the importance of the so-called adiabatic invariance of the weights of the different quantum states. We shall then briefly discuss the relation between our choice of *a priori* probabilities and the condition of nondegeneracy of energy states which is the ergodic condition (see Part E). Finally we shall show how in the classical limit, the condition of equal *a priori* probabilities and random *a priori* phases from different nondegenerate states leads to the condition of equal *a priori* probabilities for equal volumes in phase space.[107]

Consider the so-called *uniform ensemble* whose density matrix is given by the equation

$$\rho_{kl}=\rho_0\delta_{kl}, \quad \text{or,} \quad \boldsymbol{\varrho}=\rho_0\mathbf{1}, \qquad \text{(D3.09)}$$

where $\mathbf{1}$ is the unit matrix. Equation (D3.09) will be satisfied by $\boldsymbol{\varrho}$ independent of the choice of the complete orthonormal system, as we can easily show by performing a unitary transformation $\mathbf{U}$,

$$\boldsymbol{\varrho}'=\mathbf{U}^\dagger\boldsymbol{\varrho}\mathbf{U}=\mathbf{U}^\dagger\rho_0\mathbf{1}\mathbf{U}=\rho_0\mathbf{U}^\dagger\mathbf{1}\mathbf{U}=\rho_0\mathbf{U}^\dagger\mathbf{U}=\rho_0\mathbf{1}. \qquad \text{(D3.10)}$$

The uniform ensemble can be considered to have been obtained by applying the prescription of equal *a priori* probabilities and random *a priori* phases for different states, since we have from Eq. (D2.07)

$$\rho_{mn}=(a_m a_n^*)_{\text{Av}}, \qquad \text{(D3.11)}$$

where the average is taken over all the systems in the ensemble. Putting

$$a_m=r_m\exp(i\alpha_m), \qquad \text{(D3.12)}$$

where r_m is the modulus and α_m the phase of a_m, we have from Eq. (D3.11)

$$\rho_{mn}=\{r_m r_n \exp[i(\alpha_m-\alpha_n)]\}_{\text{Av}}. \qquad \text{(D3.13)}$$

Equal *a priori* probabilities means

$$r_m=\text{independent of } m=\sqrt{\rho_0}, \qquad \text{(D3.14)}$$

so that Eq. (D3.13) reduces to

$$\rho_{mn}=\rho_0\{\exp[i(\alpha_m-\alpha_n)]\}_{\text{Av}}, \qquad \text{(D3.15)}$$

which because of the random distribution of the phases reduces to Eq. (D3.09).

The uniform ensemble is also a stationary ensemble. This follows immediately from Eq. (D2.27) [compare Eq. (D3.10)].

We see here that the condition of equal *a priori* probabilities and random *a priori* phases for different states is invariant both against the passage of time and against transformations from one representation to another. The assumption of random *a priori* phases is necessary in order to assure the establishment of these invariances, as the density matrix given by the equation

$$\rho_{kl}=\rho_0\delta_{kl}+A_{kl}(1-\delta_{kl}) \qquad \text{(D3.16)}$$

would satisfy the requirement of equal *a priori* probability but would not be a stationary density or a density invariant against a unitary transformation.

We may remark here that Dirac (29D; see also 27N1) singles out the uniform ensemble since it remains invariant against arbitrary slight perturbations of the Hamiltonian and thus gives the *a priori* distribution of quantum statistics. He also draws attention to the fact that it leads—as we just saw—to equal *a priori* probabilities for all discrete states.

The equality of *a priori* weights for nondegenerate states follows also from the so-called *adiabatic principle* to which Ehrenfest's name is usually attached. We cannot enter into a detailed discussion of this very important subject and must refer to the literature[108] (06E1, 11E1, 14E, 16E, 17B1, 17B2, 17B3, 17B4, 17E1, 18B, 23B2, 27U, and especially 23E1, where further references can be found). This principle states that if the external parameters are varied adiabatically, the relative weights of the different states cannot change. From this it follows directly that in the case of multiply

---

[107] We refer to a paper by Uhlenbeck (35U) for a discussion of some other aspects of the transition from quantum to classical statistics.

[108] Some of Ehrenfest's papers also discuss the importance of this principle for the choice of the *a priori* weights in classical statistics.

periodic systems each nondegenerate quantum state has the same *a priori* weight.

We come now to the important question whether, as in the classical case, the choice of *a priori* probabilities is determined completely, if we are dealing with ergodic systems. We saw in Sec. C(3) that this was the case in classical statistics due to the fact that the other constants of motion take up their possible values with practically equal frequency in different parts of phase space, or, as it is sometimes put, are practically constant over an energy surface. This latter property is also true for an ergodic system in quantum mechanics as we can see easily. If all energy states are nondegenerate, the energy characteristic functions will form a complete orthonormal set. Since a constant of motion corresponds to an operator which commutes with the energy operator, it follows from quantum theory (e.g., 38K Sec. 38) that the energy characteristic functions are also characteristic functions of all the operators corresponding to the integrals of motion. It then follows that in each energy state—which, because of the nondegeneracy corresponds to one and only one characteristic function—the integrals of motion have well-defined constant values.

Unfortunately, however, this does not ensure that our choice of *a priori* probabilities and of *a priori* phases follows. As far as the *a priori* probabilities are concerned, we hardly need the ergodicity of the systems under consideration, since as we saw a moment ago, the adiabatic principle allows only those *a priori* weights which we have used throughout. The choice of *a priori* phases is, however, more difficult to justify from first principles—and that is what we are trying to do at the moment.

Let us consider a moment the plausibility of the existence of ergodic systems. We know from quantum theory that degenerate levels are the rule rather than the exception when we are dealing with simple systems such as atoms and molecules, and it seems unlikely that the situation would be different when we come to complicated systems such as are considered in statistical mechanics. In this connection it is of interest to remind ourselves of the fact that it has not even been possible to prove that the ground state of any system is nondegenerate—a property which is important for a discussion of the third law of thermodynamics, the so-called Nernst heat theorem.[109] We must therefore also face the possibility that physical systems are not ergodic. Let us briefly consider in how far the ergodic theorem of Part E, that is, the equality of time and ensemble averages can still be saved. This equality is based on Eq. (E.26) which is proved in Appendix IV. From the discussion in that appendix we see that Eq. (E.26) can also be proved, if we may average over the phases on the right-hand side of Eq. (IV.07).[110] I would like to suggest that this averaging is in accordance with the basic ideas of quantum mechanics. It is well known that in no physical quantities the (initial) phases of the wave functions play a role, and it is therefore tempting to suggest that each wave function be considered to be an average (or better a mixture) of a large number of wave functions, only differing in their phases, but otherwise being the same. This would not alter any results obtained from the normal wave mechanics and would ensure the ergodic theorem as well as justifying the manner in which we set up our representative ensembles. We transfer in this way the assumption of random *a priori* phases from statistical mechanics to quantum mechanics itself.

We conclude this section with a discussion of how our classical *a priori* probabilities are the limiting case of the assumption regarding *a priori* weights and *a priori* phases made in quantum statistics. The transition could easily be justified in the old (model) quantum mechanics, at any rate, as long as one was dealing with multiply periodic systems (compare 18B), since in the old quantum theory the energy levels were just obtained by dividing phase space into cells of volume h^s (h: Planck's constant; s: number of degrees of freedom) and assigning one level to each cell.

In modern quantum mechanics, the situation is more complicated. One often finds the cryptic statement that the assignment of a volume h^s in phase space for each state is connected with the Heisenberg relations (27H1) which are then written in the form

$$\Delta p \cdot \Delta q \geqslant h, \tag{D3.17}$$

where p and q are a canonically conjugate momentum-coordinate pair, and where Δp and Δq are the ranges of accuracy with which they can be measured simultaneously. It is not explained in these cases, why the volume would not be $(h/4\pi)^s$, as the right-hand side of inequality (D3.17) in the most rigorous form should be $h/4\pi$ and not h (30H1).

However, one can probably justify the choice of a volume h^s in phase space for each state as follows (ESM p. 60; compare also 35D Sec. 37, 49M). In classical mechanics the phase of the particle is given by the representative point in phase space. This point can be considered to be the combination of a point in coordinate—or q-space—and a point in momentum—or p-space. In quantum mechanics we must, however, consider probability densities. The two probability densities in p- and in q-space are not independent. The probability density in q-space is determined by the wave function ψ and is $|\psi|^2$, which is normalized,

$$\int |\psi|^2 dq = 1, \tag{D3.18}$$

where dq stands for a volume element in q-space. The probability density in p-space is obtained by first transforming the wave function $\psi(q)$ in coordinate representation to the probability amplitude $A(p)$ in

---

[109] For a discussion see 30S, 51S, 16S, and ESM Appendix III.

[110] Compare also Eq. (E.08).

momentum space. The probability density in p-space is then $|A|^2$. It follows from transformation theory that A and ψ are each other's Fourier transforms, if we consider q and p/h as the variables,[111] and that thus

$$\int |A|^2 dp = h^s. \tag{D3.19}$$

We can now say that the total volume of phase space corresponding to one state of the system is obtained by multiplying the probability densities in p- and q-space and integrating over the whole of phase space, and we get from Eqs. (D3.18) and (D3.19) as the final result a volume h^s.

From the foregoing discussion we can conclude that in the limit of $h \to 0$, indeed, equal volumes of phase space—which contain equal numbers of cells of volume h^s—correspond to equal *a priori* probabilities, once we have accepted the equal *a priori* probabilities for all nondegenerate states. The necessity of random *a priori* phases in this connection is not so apparent. It follows, however, from the considerations of the uniform ensemble. We should like to have equal weights for nondegenerate states independent of the representation we are using, and this can only be ensured if we use the additional assumption of random *a priori* phases as we saw earlier in the present section.

E. THE ERGODIC THEOREM IN QUANTUM STATISTICS

We saw earlier that apart from the approach via the H-theorem one could justify the use of statistical methods by using the ergodic (or quasi-ergodic) theorem and proving the equivalence of equality of time and ensemble averages. The same choice of method of approach exists in quantum statistics, and we do now wish to consider the quantum-statistical ergodic theorem. In analogy with the old ergodic theorem [Sec. B(1)] one can assume the existence of ergodic systems. Such systems are defined as such systems where all states E_k within a given energy range

$$E \leqslant E_k \leqslant E + \delta E \tag{E.01}$$

can be reached from all other states within the range—even if only via other intermediate states.[112]

Let P_k denote once again the probability of finding the system in the kth state, and let w_{kl} be the transition probability for a transition from the kth to the lth state. We then have the following equation governing the temporal behavior of the P_k:

$$dP_k/dt = -\sum_l P_k w_{kl} + \sum_l P_l w_{lk} = \sum_l w_{kl}(P_l - P_k), \tag{E.02}$$

where we have used the fact that $w_{kl} = w_{lk}$.[113] The equilibrium solution of Eq. (E.02) which is attained exponentially is

$$P_k = \text{constant} = 1/n, \tag{E.03}$$

if n is the number of states in the range (E.01). From this it follows that the average time spent in any of the n states is the same and to calculate the time average of any quantity we can take the mean over the states of the range (E.01), or over a representative microcanonical ensemble [compare the density for such an ensemble given by Eq. (D2.15)]. We have thus proved that for ergodic systems the two averages are the same. The problem remains whether or not ergodic systems exist. In the present case it seems very likely that this is the case. In a closed system radiation must have been included and this should make transitions from one state to another possible.[114]

There is, however, also in quantum mechanics a different method of approach which is analogous to Birkhoff's or Hopf's ergodic theorem. This method consists of proving directly the equivalence of time and ensemble averages. The first proof was given by von Neumann (29N) and it was later simplified by Pauli and Fierz (37P) and by Fierz (private communication).[115]

As in the classical case we can distinguish two different ergodic theorems according to whether or not we allow a "coarsening" of the energy of the system. In the quantum-statistical case the second ergodic theorem is the more important one, but we shall first discuss briefly the first ergodic theorem and we shall follow largely the exposition given by Rosenfeld (52R).

In the first case we are dealing with a pure case. Let $\psi(t)$ be the wave function describing the system under consideration. We choose now for our complete orthonormal set φ_n the characteristic functions of the energy operator $\mathbf{H}$ and denote by E_n the corresponding characteristic values. We now write

$$\psi(0) = \sum_n a_n \varphi_n, \tag{E.04}$$

with

$$a_n = r_n \exp(i\alpha_n), \tag{E.05}$$

where r_n is the modulus and α_n the phase of a_n. For $\psi(t)$ we then have in the usual way

$$\psi(t) = \sum r_n \exp(i\alpha_n) \exp(-iE_n t/\hbar) \varphi_n, \tag{E.06}$$

and the density matrix as a function of time is given by the equation

$$\rho_{mn}(t) = r_m r_n \exp[i(\alpha_m - \alpha_n)] \times \exp[-i(E_m - E_n)t/\hbar]. \tag{E.07}$$

---

[111] It is at this point that Planck's constant enters—one might say through the de Broglie relations, or through the commutation relations which also lead to the Heisenberg relation.

[112] See 33J, Sec. 2.2; 40M, p. 56; 46J, p. 152.

[113] Van Kampen (54K) has shown that Eq. (E.02) can also be derived under conditions where $w_{kl} \neq w_{lk}$.

[114] We may mention in this connection the recent paper by M. J. Klein (52K) who makes some explicit assumptions about the interactions allowing the transitions to take place.

[115] I am greatly indebted to Professor M. Fierz for putting at my disposal his latest considerations which were discussed at the Les Houches Summer School of 1953.

The average value of any physical quantity A is the quantum-mechanical average given by the equation

$$\langle A\rangle=\sum_{mn} r_m r_n \exp[i(\alpha_m-\alpha_n) -i(E_m-E_n)t/\hbar]A_{mn}, \qquad \text{(E.08)}$$

where A_{mn} is given by the equation

$$A_{mn}=\int \varphi_m{}^*\mathbf{A}\varphi_n d\tau. \qquad \text{(E.09)}$$

We can now take the time average of $\langle A\rangle$ and we see that, if there is no degeneracy, that is, if no two characteristic energy values are the same, we have

$$\overline{\langle \mathrm{A}\rangle}=\sum_n \rho_{nn}A_{nn}, \qquad \text{(E.10)}$$

which is independent of the initial phases (the α_n).

We see that $\overline{\langle \mathrm{A}\rangle}$ will only be independent of the initial conditions, in this case, the initial phases, if there is no degeneracy. The requirement of nondegeneracy thus takes the place of the requirement of metrical transitivity in the classical case.

That Eq. (E.10) gives $\overline{\langle \mathrm{A}\rangle}$ as a statistical average can be seen when we introduce the so-called projection operators $\mathbf{R}(\varphi)$. This operator is defined by the equation

$$\mathbf{R}(\varphi)\chi=\varphi\int \varphi^*\chi d\tau, \qquad \text{(E.11)}$$

which should be valid for any pair of functions φ and χ. The geometric meaning of the projection operator is simple, if we may use for a moment the language of the Hilbert space. The integral $\int \varphi^*\chi d\tau$ then corresponds to the scalar product of the two vectors χ and φ in Hilbert space and the right-hand side of Eq. (E.11) gives thus the "projection" of χ on φ if φ is normalized, and thus corresponds to a unit vector in Hilbert space.[116]

One can easily see that, in the case where the system is characterized by the wave function ψ, the operator $\mathbf{R}(\psi)$ is diagonal and its diagonal elements are equal to the diagonal elements of the density matrix so that we can write Eq. (E.10) in the form

$$\overline{\langle \mathrm{A}\rangle}=\mathrm{Tr}(\mathbf{R}(\psi)\mathbf{A}). \qquad \text{(E.12)}$$

From Eq. (E.07) it follows also that $\mathbf{R}(\psi)$ can be obtained by averaging the density matrix over the different values of the initial phases, or,

$$\mathbf{R}(\psi)=\varrho_{\mathrm{Av}}, \qquad \text{(E.13)}$$

the subscript "Av" indicating here the averaging over the values of the α_n, and Eq. (E.12) or (E.10) can also be written in the form

$$\overline{\langle \mathrm{A}\rangle}=\mathrm{Tr}(\varrho_{\mathrm{Av}}\mathbf{A}). \qquad \text{(E.14)}$$

Loosely speaking we may interpret this result as meaning that the temporal change of the wave function corresponds to a uniform distributing of the α_n over all possible values, corresponding to the classical picture of the trajectory completely covering the energy surface.

If we are dealing with an isolated system of constant energy the $\psi(0)$ is one of the φ_n and we are dealing with a pure case with a microcanonical density matrix. In the general case discussed at this moment time averaging has the effect of replacing the pure case by a mixture obtained by averaging over the initial phases.

The discussion of the first ergodic theorem is mainly of importance for the discussion of the influence of observations on the state of the system and the discussion of the change from a pure to a mixed case by observation. We have discussed some aspects of this at the end of Sec. D(2) and are not going to discuss it any further at this point.

The second ergodic theorem deals as does Hopf's ergodic theorem with an energy shell. At the same time it meets the difficulty that we often use to describe a system a number of physical quantities which do not correspond to commuting operators so that it is not possible to have a wave function which is a characteristic function of all these operators. It is then necessary to introduce instead of these microscopic operators so-called macroscopic operators as was first done by von Neumann.[117] Roughly speaking a macroscopic operator can be obtained from its microscopic counterpart as follows. Let A be a continuous variable which can take on any value between $-\infty$ and $+\infty$ and let macroscopic measurements be such that only the intervals

$$k\leqslant A\leqslant k+1, \quad k=0, \pm 1, \pm 2, \cdots \qquad \text{(E.15)}$$

can be distinguished. Let further $f(x)$ be defined by the equation

$$f(x)=k, \quad k\leqslant x<k+1, \quad k=0, \pm 1, \pm 2, \cdots. \qquad \text{(E.16)}$$

The macroscopic variable corresponding to A will then be $f(A)$. As is emphasized by von Neumann,[118] macroscopic variables can always be measured simultaneously. It is instructive to quote his actual remarks: "During a macroscopic simultaneous measurement of coordinate and momentum (or of another pair of quantities which cannot be measured simultaneously due to the Heisenberg relations) we measure indeed two physical quantities simultaneously and accurately, but these two quantities are not exactly coordinate and momentum. We measure, for instance, the positions of two needles or the positions of two images on a photographic plate and nothing can prevent us from measuring these simultaneously and as accurately as we wish, but the coupling between them and the quantities in which we are

---

[116] We use the notation **R** instead of **P** to emphasize the relation with the density matrix.

[117] We must refer here to a paper by van Kampen (54K) which was published after this was written and which discusses in great detail and rigor the introduction of these macroscopic operators.

[118] Compare also the discussion given by von Weizsäcker (49W1).

interested (q_k and p_k) is a loose one and the looseness of the coupling is determined by the Heisenberg relations."

We first of all introduce energy shells in such a way that we group the energy levels together as before and assign to each group of S_i levels one representative energy value E_i. Inside each shell we introduce (phase) cells [compare the discussion in Sec. A(3)] which each contain s_ν states, say. In each cell the macroscopic variables A, B, $\cdots$ in which we are interested will have a constant value. This means that we have introduced macroscopic operators for the energy, A, B, and so on, in such a way that they all commute and that their characteristic values are s_ν-fold degenerate. If ω_r are the characteristic functions of the macroenergy, and the macro-operators corresponding to A, B, $\cdots$, they correspond inside a cell to the same characteristic values. We further have

$$\sum\nolimits_\nu s_\nu = S_i, \tag{E.17}$$

where ν runs from 1 to N_i, if N_i is the number of cells in the shell.

Let now P_ν be the probability of finding the system in the νth cell and let p_ν be the probability of finding the system in one of the states belonging to the νth cell.[119] We have then [compare Eq. (D.242)]

$$P = s_\nu p_\nu. \tag{E.18}$$

We now define a quantity $\Sigma(t)$ by the equation [compare Eq. (D2.43)]

$$\Sigma = \sum\nolimits_\nu P_\nu \ln(P_\nu / s_\nu). \tag{E.19}$$

We must bear in mind that the number of cells in an energy shell, that is, the number of states which can be distinguished by an observer will be of the order of magnitude of (at most) 10^{20} say. On the other hand, $-\Sigma$ which is a measure of the entropy of the system in units $\mathbf{k}$ ($=$Boltzmann's constant) will also be of the order of magnitude 10^{20}, that is, if the order of Avogadro's number, which means that the s_ν will be enormously big,[120]

$$s_\nu \sim \exp(10^{20}) \text{ and also } S_i \sim \exp(10^{20}). \tag{E.20}$$

As a result P_ν will be of the order of magnitude 10^{-20} and $\ln s_\nu$ will be extremely large compared to $-\ln P_\nu$ so that we can write instead of Eq. (E.19)

$$\Sigma(t) = -\sum\nolimits_\nu P_\nu \ln s_\nu. \tag{E.21}$$

On the other hand consider the quantity

$$\langle\Sigma\rangle_{\mathrm{Av}} = -\sum\nolimits_\nu (s_\nu / S_i) \ln s_\nu. \tag{E.22}$$

For $\langle\Sigma\rangle_{\mathrm{Av}}$ we can easily prove the following inequalities:

$$\ln S_i \geqslant \sum\nolimits_\nu (s_\nu/S_i) \ln(s_\nu/S_i) + \sum\nolimits_\nu (s_\nu/S_i) \ln S_i = -\langle\Sigma\rangle_{\mathrm{Av}}, \tag{E.23}$$

$$-\langle\Sigma\rangle_{\mathrm{Av}} \geqslant \sum\nolimits_\nu (1/N_i) \ln(S_i/N_i) = \ln S_i - \ln N_i. \tag{E.24}$$

The first inequality follows since $0 < s_\nu/S_i < 1$, and the second inequality since a sum of the kind $\Sigma a_\nu \ln a_\nu$ under the condition $\Sigma a_\nu = 1$ is minimum for all a_ν being equal. By Eq. (E.20), $\ln S_i$ is of the order 10^{20} while $\ln N_i$ is only of the order 50, so that we can neglect $\ln N_i$ with respect to $\ln S_i$ and write

$$\langle\Sigma\rangle_{\mathrm{Av}} = -\ln S_i. \tag{E.25}$$

Since S_i is the number of states in the energy shell, $-\langle\Sigma\rangle_{\mathrm{Av}}$ can be interpreted (apart from a factor $\mathbf{k}$) as the entropy of the microcanonical ensemble [compare the discussion just before Eq. (D1.28)].

We can prove the ergodic theorem, if we can prove that $\langle\Sigma\rangle_{\mathrm{Av}}$ is the time average of $\Sigma(t)$ given by Eq. (E.21) or, that

$$\bar{P}_\nu = s_\nu / S_i, \tag{E.26}$$

a relation which is strongly reminiscent of Eq. (E.03).

It is possible, that Eq. (E.26) holds on the average. The average is here taken over all possible divisions of an energy shell into cells. The way in which the various divisions are weighted is the one introduced by Weyl (25W). In Appendix IV we shall discuss the proof of Eq. (E.26). A necessary condition for the proof is that there are no degenerate energy levels.[121]

We have now proved the equivalence of time and ensemble averages in the case of the function Σ. However, as soon as Eq. (E.26) holds, one can also prove the equivalence of time and ensemble averages for other variables. This follows from the considerations of von Neumann,[122] but can also be seen by our argument similar to the one discussed in the Introduction and which runs as follows. From the fact that $\bar{\Sigma}$ and $\langle\Sigma\rangle_{\mathrm{Av}}$ are the same, it follows that the system spends most of the time in the equilibrium state and that thus the time average of any variable will correspond to the equilibrium value. We must also mention that in proving Eq. (E.26) we have really proved more than just the equivalence of $\bar{\Sigma}$ and $\langle\Sigma\rangle_{\mathrm{Av}}$, but have made a statement about the relative periods spent in the different cells. Equation (E.26) corresponds to Eq. (B1.05) of the classical case.

We may, in conclusion, say a few words about a recent paper by Klein (52K). His point of view is similar to the one of Born and Green in that only systems in a well-defined state are considered. In that case it is necessary to introduce an interaction between the system and the outside world. This interaction, provided it satisfies certain mixing conditions, will ensure the equivalence of time and ensemble averages. However, this method of approach does not, in our opinion, do justice to the fact that an observer is dealing

---

[119] The p_ν would be the coarse-grained density of an ensemble where all systems corresponded to the same wave function, namely, the wave function of the system which we are considering.

[120] I am indebted to Professor M. Fierz for these considerations.

[121] It must be noted here that if we use, with Fierz, Eq. (E.21) for $\Sigma(t)$, we only need the condition of nondegeneracy, but not the extra restriction of no resonances. This is very satisfactory, since, while the energy differences have a physical meaning (they enter into expressions for relaxation times, for instance), the second differences do not have such a physical meaning. I am indebted to Professor Fierz for calling my attention to this point.

[122] See Appendix IV.

with macroscopic variables. We refer to his paper for a detailed discussion of his very interesting analysis.

SUMMARY OF THE SITUATION IN QUANTUM THERMOSTATISTICS

Once again we can briefly summarize the results of the two preceding parts (D and E). There is no need to discuss again the utilitarian approach as there is no difference between the classical and the quantum-mechanical case.

In the case of the formalistic approach, once again one investigates the necessary conditions for the equality of time and ensemble averages. In the quantum-mechanical case the condition is the one of nondegeneracy of all energy characteristic states.

In the case of the physical approach, one works again with representative ensembles and shows that the development of a nonequilibrium situation will be such that the ensemble representing the system under consideration will approach a canonical ensemble when time goes on. The problem here reduces to finding the rules according to which the representative ensembles can be set up. These rules are the assumption of equal *a priori* probabilities for all nondegenerate states and of random *a priori* phases for the probability amplitudes. Once these assumptions are made, one can either use the quantum-mechanical version of the statistical H-theorem, or one can use the H-theorem as applied to ensembles. The discussion is very similar to the one applicable in the classical case.

In the classical case we saw that in the last instance the basic ideas of both the physical and the formalistic approach were the same. Once one has proved that physical systems are quasi ergodic, the ergodic theorem follows, but one can also justify from first principles the choice of *a priori* probabilities. In the quantum-mechanical case, the situation is not so simple. One can justify from first principles the assumption of equal *a priori* weights, but one is left with the assumption of random *a priori* phases. On the other hand, it seems unlikely that physical systems are really ergodic in the quantum-mechanical sense, and in that case the quantum-mechanical ergodic theorem must also use an assumption about the distribution of the phases. Since in quantum-mechanics the (initial) phases do not play a physical role, one might with some justification say that one should really always average over these phases. In that case the quantum-mechanical ergodic theorem is always true, and the assumption of random *a priori* phases is also justified. As in the classical case we see that the two different methods of approach are not differing by as much as one would expect at first sight.

APPENDIX I. THE LORENTZ MODEL[123]

In Sec. A(1) we introduced a model of a physical system which possessed the following properties. One kind of particle is fixed in space, randomly distributed with a density of n per unit volume. A second kind of particle is traveling through the lattice colliding elastically and isotropically with the first kind, but not interacting with its own kind. Its density in space is N per unit volume and its speed is c. The cross section for a collision is denoted by σ.

For our discussion we assume phase space, which in the present case is the unit sphere, to be divided into $2m+1$ finite cells, each of extension $\delta\omega$,

$$\delta\omega = 4\pi/(2m+1), \tag{I.01}$$

and numbered by an index v which runs through the sequence $-m, -m+1, \cdots, 0, \cdots, m-1, m$. We denote by f_v the number per unit volume of particles of the second kind—which in order to keep the analogy with Lorentz's model of a metal we shall call electrons—moving in the direction specified by the vth element of solid angle. The f_v satisfy the condition

$$\textstyle\sum_v f_v = N, \tag{I.02}$$

and their equilibrium values $f_v{}^e$ are given by the equation

$$f_v{}^e = N/(2m+1). \tag{I.03}$$

To characterize the state of the system, we introduce Δ by the equation

$$\Delta = \textstyle\sum_v (f_v - f_v{}^e)^2. \tag{I.04}$$

We can now ask for the probability $w(\Delta)d\Delta$ of finding a value of Δ in the interval $\Delta, \Delta+d\Delta$. We use the method introduced by Markoff (12M; see also 43C, 46C) to find a probability distribution for a quantity which is a sum of a number of similar terms of functions, the probability distribution of which is given. We mention the result (see 43C) but give no proof. If

$$\Phi = \textstyle\sum_k \phi_k(q_i), \tag{I.05}$$

where the q_i are stochastic variables such that

$$p(q_i)\Pi dq_i \tag{I.06}$$

is the probability of finding the q_i within ranges q_i, q_i+dq_i, the probability $w(\Phi)d\Phi$ to find Φ between Φ and $\Phi+d\Phi$ is given by the equation

$$w(\Phi)d\Phi = (d\Phi/2\pi)\int_{-\infty}^{+\infty} \exp(-i\rho\Phi)A(\rho)d\rho, \tag{I.07}$$

where $A(\rho)$ is determined from

$$A(\rho) = \int\cdots\int \exp(i\rho\textstyle\sum\phi_k)p(q_i)\prod dq_i. \tag{I.08}$$

In our case Eq. (I.04) takes the place of Eq. (I.05), and the probability of finding a certain set of values for the f_v is a compound probability. Using Eq. (I.02), we find

$$p(f_v) = [N!/\Pi f_v!](2m+1)^{-N}, \tag{I.09}$$

---

[123] See 54G; compare also 55G, 55H2.

which for large values of the f_v which are in the neighborhood of f_v^e takes the form[124]

$$p(f_v)=[(2m+1)/2\pi N]^m(2m+1)^{\frac{1}{2}} \times\exp[-(2m+1)\sum(f_v-f_v^e)^2/2N], \qquad (\text{I}.10)$$

which is normalized.

For the function $A(\rho)$ we get

$$A(\rho)=\int\cdots\int p(f_v)\exp(-i\rho\Delta)df_{-m}\cdots df_m, \qquad (\text{I}.11)$$

where the integral is $2m$-fold since by virtue of Eq. (I.02) only $2m$ of the f_v are independent. Since one of the f_v must be eliminated, for which f_0 is chosen to preserve the symmetry, the integration is rather tedious, but luckily straightforward and one find

$$A(\rho)=[1-2iN\rho/(2m+1)]^{-m}. \qquad (\text{I}.12)$$

Substituting from Eq. (I.12) into Eq. (I.07) one finds by contour integration

$$w(\Delta)d\Delta=[(2m+1)/2N]^m[\Delta^{m-1}/(m-1)!] \times\exp\{-(2m+1)\Delta/2N\}d\Delta, \qquad (\text{I}.13)$$

which is Eq. (A3.13).

From Eq. (I.13) one gets

$$\Delta_{\text{Av}}=\int\Delta w(\Delta)d\Delta=2mN/(2m+1), \qquad (\text{I}.14)$$

and

$$(\Delta^2)_{\text{Av}}=\int\Delta^2 w(\Delta)d\Delta=2m(2m+2)N^2/(2m+1)^2, \qquad (\text{I}.15)$$

from which Eqs. (A3.16) follow if we neglect in the final result 1 against $2m$, since $m\gg1$.

We now must evaluate the probability $w(\Delta,\Delta')$ that Δ changes its value from Δ to Δ' during a time interval τ. The change in Δ is related to the changes in the values of the f_v by

$$\Delta'-\Delta=2\sum_v(f_v-f_v^e)(f_v'-f_v) =2\sum_v f_v(f_v'-f_v), \qquad (\text{I}.16)$$

since from Eq. (I.02) it follows that $\sum f_v'=\sum f_v$.

Let $x_{vv'}$ be the actual number of electrons passing from cell v to cell v' during a time interval τ. If we could use a *Stosszahlansatz*, we would have

$$x_{vv'}=af_v, \qquad (\text{I}.17)$$

where

$$a=n\sigma\tau c/(2m+1), \qquad (\text{I}.18)$$

where we have used the assumption of isotropic scattering [compare Eq. (A1.15)].

We are, however, interested in fluctuations and instead of Eq. (I.17) we must consider a distribution of the $x_{vv'}$ around an average value which will be given by Eq. (I.17). If the distribution of the fixed particles in space is random we may assume this distribution to be a Bernoullian one.

The time interval τ will be chosen in such a way that the probability of any one electron undergoing more than one collision during τ is negligibly small, or

$$\tau\ll1/n\sigma c. \qquad (\text{I}.19)$$

If we chose N so large that aN is still large compared to one so that we have the inequalities

$$N\gg aN\gg1, \qquad (\text{I}.20)$$

we can use the Gaussian form for the $x_{vv'}$-distribution,

$$p(x_{vv'})=(2\pi af_v)^{-\frac{1}{2}}\exp[-(x_{vv'}-af_v)^2/2af_v]. \qquad (\text{I}.21)$$

The change of the f_v during a time interval τ is governed by the $x_{vv'}$ and we have

$$f_v'-f_v=\sum_{v'}(x_{v'v}-x_{vv'}). \qquad (\text{I}.22)$$

Combining Eqs. (I.16) and (I.22) we see that we have once again a problem of the kind which led to $w(\Delta)$. The expression $w(\Delta)w(\Delta,\Delta')$ now takes the place Φ of Eq. (I.05) and the ϕ_k are expressions in the f_v and the $x_{vv'}$. From Eq. (I.07) we now have

$$w(\Delta)w(\Delta,\Delta')=(1/4\pi^2) \times\int\int\exp[-i\sigma\Delta-i\rho(\Delta'-\Delta)]A(\rho,\sigma)d\rho d\sigma, \qquad (\text{I}.23)$$

where

$$A(\rho,\sigma)=\int\cdots\int df_{-m}\cdots df_m dx_{-m,-m}\cdots dx_{mm} \times p(f_v)p(x_{-m-m})\cdots p(x_{mm}) \times\exp[i\sigma\sum(f_v-f_v^e)^2+2i\rho\sum f_v(f_v'-f_v)]. \qquad (\text{I}.24)$$

The integration over the $(2m+1)^2$ variables $x_{vv'}$ is straightforward after we have substituted from Eqs. (I.21) and (I.22). The integration over the $2m$ variables f_v $(v\neq0)$ is more complicated. We introduce new variables α_v through

$$\alpha_v=(f_v-f_v^e)/f_v^e, \qquad (\text{I}.25)$$

and neglect cubic terms in the exponent with respect to quadratic terms. We are then left with a quadratic expression in the exponent and can integrate. The final result leads to

$$A(\rho,\sigma)=[(2m+1)/2]^m[4Na\rho(2N\rho+2m+1) -2Ni\sigma+2m+1]^{-m}, \qquad (\text{I}.26)$$

and from Eqs. (I.23) and (I.13) we get the normalized expression

$$w(\Delta,\Delta')=(32\pi Na\Delta)^{-\frac{1}{2}} \times\exp[-(\Delta'-\Delta+2(2m+1)a\Delta)^2/32Na\Delta], \qquad (\text{I}.27)$$

which is Eq. (A3.19).

---

[124] To derive Eq. (I.10), one uses the Stirling formula in the form $\ln x!=x\ln x-x+\frac{1}{2}\ln x+\frac{1}{2}\ln 2\pi$ and expands in terms of $(f_v-f_v^e)/f_v^e$.

Equation (A3.21) follows from the equation

$$\Delta_{\mathrm{Av}}'=\int \Delta' w(\Delta,\Delta')d\Delta', \tag{I.28}$$

using Eq. (I.27).

In footnote 36 we mentioned that the following relation held:

$$w(\Delta)w(\Delta,\Delta')=w(\Delta')w(\Delta',\Delta). \tag{I.29}$$

We must draw attention to the fact that in proving Eq. (I.29) one neglects higher-order terms in a which is permitted because of inequality (I.19).

Using Eqs. (A3.27), (A3.28), (A3.31), and (I.27), Eqs. (A3.19) and (A3.32) follow, again up to first-order terms in a. Similarly, Eq. (A3.30) follows from Eq. (A3.29) by neglecting terms in a^2 and higher-order terms.

If the *Stosszahlansatz* would be valid, one can use for the rate of change of the f_v Eqs. (I.22) and (I.17) which gives us

$$\begin{aligned} df_v/dt=(f_v'-f_v)/\tau&=\textstyle\sum_{v'}(x_{v'v}-x_{vv'}) \\ &=(a/\tau)[\textstyle\sum_{v'}f_{v'}-(2m+1)f_v], \end{aligned} \tag{I.30}$$

whence Eq. (A3.34) follows by using the fact that $\Sigma f_{v'}=\Sigma f_{v'}{}^e=(2m+1)f_v{}^e$.

APPENDIX II. PROOF OF THE EXISTENCE OF QUASI-ERGODIC SYSTEMS[125]

Fermi's proof of the quasi ergodicity of *kanonische Normalsysteme* with more than two degrees of freedom consists of two parts. The first part is a proof that the energy surfaces $\mathcal{S}$ are the only family of surfaces in phase space which have the property that any orbit starting from a point of one of the surfaces of the family will always stay on the same surface. The second part then consists of showing that, if there are two arbitrary regions σ and σ^* on $\mathcal{S}$, there will be orbits passing through both σ and σ^*.

The proof of the first part is slightly complicated and runs as follows. A *kanonisches Normalsystem* is a system where one can introduce canonically conjugate coordinates x_i and y_i such that:

(a) the energy is time independent;

(b) the Hamiltonian $\mathcal{H}$ of the system can be developed in a power series of a parameter α,

$$\mathcal{H}=\mathcal{H}_0+\alpha\mathcal{H}_1+\alpha^2\mathcal{H}_2+\cdots; \tag{II.01}$$

(c) the Hamiltonian and thus all $\mathcal{H}_i$ are periodic in all the x_i. We can then choose our x_i such that the period is the same for all x_i, say, 2π;

(d) the first term in the series expansion in Eq. (II.01), $\mathcal{H}_0$, does not depend on the x_i.

Let us assume now that the equations

$$\Phi(x,y;\alpha)=0 \tag{II.02}$$

describe a family of surfaces $\mathcal{R}_\alpha$ and that the functions Φ are uniquely valued, analytical and periodic in the x_i. We must show that each of the $\mathcal{R}_\alpha$ coincides with an energy surface $\mathcal{S}$.

We expand Φ in powers of α and have

$$\Phi=\Phi_0+\alpha\Phi_1+\alpha^2\Phi_2+\cdots. \tag{II.03}$$

If the $\mathcal{R}_\alpha$ are given we can still to a large extent choose our Φ_i arbitrarily and especially it is possible to prove[126] that one can for given $\mathcal{R}_\alpha$ choose Φ_0 not to depend on the x_i. Moreover, one can choose the Φ_i arbitrarily outside $\mathcal{R}_0$.

As the $\mathcal{R}_\alpha$ are surfaces such that an orbit starting from one of its points will stay on the same surface, the equation

$$\{\mathcal{H},\Phi\}=0 \tag{II.04}$$

must be a consequence of Eq. (II.02), where $\{A,B\}$ is the Poisson bracket defined by

$$\begin{aligned} \{A,B\}=\textstyle\sum_i[(\partial A/\partial y_i)(\partial B/\partial x_i) \\ -(\partial A/\partial x_i)(\partial B/\partial y_i)]. \end{aligned} \tag{II.05}$$

From Eq. (II.04) and the series expansions (II.01) and (II.03) it follows that the following equations must hold:

$$\{\mathcal{H}_0,\Phi_0\}=0, \tag{II.06}$$

$$\{\mathcal{H}_0,\Phi_1\}+\{\mathcal{H}_1,\Phi_0\}=0,\ \cdots. \tag{II.07}$$

Equation (II.06) is trivial as neither $\mathcal{H}_0$ nor Φ_0 contain the x_i.

As both Φ_1 and $\mathcal{H}_1$ are periodic in all the x_i, we can write

$$\Phi_1=\textstyle\sum A_m(y_i)\exp\theta_m, \tag{II.08}$$

$$\mathcal{H}_1=\textstyle\sum B_m(y_i)\exp\theta_m, \tag{II.09}$$

with

$$\theta_m=i(m_1x_1+m_2x_2+\cdots+m_nx_n), \tag{II.10}$$

where m is a short-hand notation for all the n m_i and where the summations in Eqs. (II.08) and (II.09) are over all n m_i.

From Eqs. (II.07) to (II.09) we get

$$\textstyle\sum\exp\theta_m[A_m\sum_i m_i\omega_i-B_m\sum_j m_j\chi_j]=0, \tag{II.11}$$

or

$$A_m\textstyle\sum_i m_i\omega_i=B_m\sum_j m_j\chi_j, \tag{II.12}$$

where

$$\omega_i=\partial\mathcal{H}_0/\partial y_i,\quad \chi_j=\partial\Phi_0/\partial y_j. \tag{II.13}$$

If we exclude the possibility that the B_m can be zero,[127] it follows from Eq. (II.12) that in any point of $\mathcal{R}_0$ where $\Sigma m_i\omega_i$ is zero, $\Sigma m_j\chi_j$ must also be zero.

Assume now $n=3$. The discussion is equally valid for $n>3$, but not for $n=2$ (compare 23F1). In general at most one of the three ratios ω_2/ω_1, ω_3/ω_2 will be constant in $\mathcal{R}_0$. We may thus assume the first two ratios to be varying on $\mathcal{R}_0$. In that case $\mathcal{R}_0$ is covered densely

---

[125] See 23F1.

[126] See Sec. 3 of 23F1; one uses the fact that the $\mathcal{R}_\alpha$ are surfaces on which an orbit stays.

[127] For a discussion of this case see 92P.

with points where ω_2/ω_1 takes on rational values so that one can find m_1 and m_2 such that

$$m_1\omega_1+m_2\omega_2=0. \tag{II.14}$$

It then follows that in these points

$$m_1\chi_1+m_2\chi_2=0, \tag{II.15}$$

so that

$$\omega_2/\omega_1=\chi_2/\chi_1. \tag{II.16}$$

As Eq. (II.16) holds in a point set which is everywhere dense on $\mathcal{R}_0$, it holds identically on $\mathcal{R}_0$. The same is true for equation

$$\omega_3/\omega_1=\chi_3/\chi_1. \tag{II.17}$$

From Eqs. (II.13), (II.16), (II.17) and the fact that $\mathcal{H}_0$ and Φ_0 do not depend on the x_i, it follows that on $\mathcal{R}_0$

$$\Phi_0=\mathcal{H}_0+c_0, \tag{II.18}$$

where c_0 is a constant. As Φ_0 can be chosen arbitrarily outside $\mathcal{R}_0$, we can choose it in such a way that Eq (II.18) holds everywhere in phase space.

If our system is nondegenerate, the equation

$$\sum m_i\omega_i=0 \tag{II.19}$$

cannot hold identically on $\mathcal{R}_0$ and we can divide Eq. (II.12) by $\Sigma m_i\omega_i$ which by virtue of Eq. (II.18) is equal to $\Sigma m_j\chi_j$. The result is

$$A_m=B_m \text{ except when all } m_i=0. \tag{II.20}$$

It follows that on $\mathcal{R}_0$

$$\Phi_1(x_i,y_i)=\mathcal{H}_1(x_i,y_i)+f_1(y_i), \tag{II.21}$$

and once again we can choose ϕ_1 outside $\mathcal{R}_0$ such that Eq. (II.21) holds everywhere in phase space.

One can now easily prove by complete induction from Eqs. (II.18) and (II.21) that

$$\Phi_i=\mathcal{H}_i+c_i, \tag{II.22}$$

where the c_i are constants. The proof is along the same lines as the proof of Eqs. (II.18) and (II.21).

Equations (II.22) lead to the result that Eq. (II.02) is equivalent to the equation

$$\mathcal{H}(x_i,y_i;\alpha)=\sum c_n\alpha^n=c, \tag{II.23}$$

and the $\mathcal{R}_\alpha$ thus coincide with energy surfaces.

The second part of Fermi's proof is much less complicated. Let σ and σ^* be two arbitrary regions on $\mathcal{S}$. Let σ' be that part of $\mathcal{S}$ which is covered by the orbits which start somewhere in $\mathcal{S}$. If σ' would cover all of $\mathcal{S}$ our point would have been proved. If σ' does not cover all of $\mathcal{S}$, let σ'' be that part of $\mathcal{S}$ which is not covered, and let $\mathcal{B}$ be the boundary surface between σ' and σ''. There can be no orbit which contains points both in σ' and σ'', as can be seen as follows. Let P' in σ' be passed at t' by an orbit which at t'' passes through P'' in σ''. Since the solutions of the equations of motion are analytical functions, we can find small regions η' in σ' around P' and η'' in σ'' around P'' such that the orbit passing through an arbitrary point Q' of η' at t' will pass through η'' at t''. Since, however, η' lies in σ' there will be points Q' lying on an orbit starting from σ and we have arrived at a contradiction.

Let now P be a point of $\mathcal{B}$ and once again P' and P'' be points of σ' and σ'', respectively. If we consider orbits passing at t through P, P', P'', these orbits will pass at t_1 through points P_1, P_1', P_1'', at t_2 through P_2, P_2', P_2'', and so on, where P_1', P_2', $\cdots$ are all lying in σ' and P_1'', P_2'', $\cdots$ are all lying in σ''. We can choose P' and P'' so close to P that P_1' and P_1'' are close to P_1, P_2', and P_2'' close to P_2, and so on. It then follows that P, P_1, P_2, $\cdots$ must all lie on $\mathcal{B}$ and $\mathcal{B}$ thus contains an orbit. We saw, however, that the only surface on which an orbit could stay was an energy surface. Hence, $\mathcal{B}$ cannot exist, or, σ' must cover $\mathcal{S}$ completely. That means that σ^* will be part of σ' and there will be orbits starting from σ which pass through σ^*. As both σ and σ^* can be taken arbitrarily small we have proved the quasi ergodicity of the systems considered.

APPENDIX III. PROOF OF THE CLASSICAL ERGODIC THEOREM[128]

Birkhoff's ergodic theorem proves the equivalence of the average taken over an energy surface and a time average taken for practically all orbits on the energy surface, provided the energy surface is metrically indecomposable. Let $f(P)$ be a well-behaved phase function,[129] that is, a function of the phase represented by a point P on the energy surface $\mathcal{S}$.

Consider the function $\bar{f}(P;t_0,T)$ defined by the equation

$$\bar{f}(P;t_0,T)=\frac{1}{T}\int_{t_0}^{t_0+T} f(P_t)dt, \tag{III.01}$$

where P_t is the point passed at t of an orbit which passed through P at t_0. We shall first of all prove that for almost all orbits the following limit exists:

$$\bar{f}(P;t_0,\infty)=\lim_{T\to\infty}\bar{f}(P;t_0,T). \tag{III.02}$$

Secondly we shall show that this limit is independent of t_0. Finally we shall show that this limit is constant almost everywhere on $\mathcal{S}$.

To prove the first part we divide the time scale in finite segments and write

$$T=n\tau, \tag{III.03}$$

and

$$\bar{f}_n(P;t_0)=\bar{f}_0(P;t_0,n\tau). \tag{III.04}$$

If P were a phase point for which the limit of Eq. (III.02) would not exist, the lower bound $L(P)$ of $\bar{f}_n(P;t_0)$ and the upper bound $U(P)$ of $\bar{f}_n(P;t_0)$ would be different and one could find two quantities α and β

---

[128] The proof we give here is due to Kolomogoroff (see 52R, 49K1).

[129] Well-behaved is understood here in the sense of summable over the energy surface which is assumed to be of finite volume.

such that

$$L(P)<\alpha<\beta<U(P). \tag{III.05}$$

Moreover, if the set $\mathfrak{D}$ of phase points for which the limit of Eq. (III.02) did not exist were of positive measure, one could find a subset $\mathfrak{D}'$ of $\mathfrak{D}$ also of positive measure and a set of values α and β such that the inequalities (III.05) would be satisfied for all points P of $\mathfrak{D}'$. This, however, leads to a contradiction, as follows.

Let P_k be the phase point at $t_0+k\tau$ and let $x_k(P)$ be defined by the equation

$$x_k(P)=\frac{1}{\tau}\int_{t_k}^{t_{k+1}} f(P_t)dt. \tag{III.06}$$

By shifting the origin one sees easily that

$$x_k(P)=x_0(P_k). \tag{III.07}$$

The time average $\bar{f}_n(P;t_0)$ can be expressed in the x_k as follows:

$$\bar{f}_n(P;t_0)=\frac{1}{n}\sum_{k=0}^{n-1} x_k(P). \tag{III.08}$$

We now integrate $\bar{f}_n(P;t_0)$ over a set of points $\mathfrak{D}_0^{(n)}$ which is a subset of $\mathfrak{D}'$ such that for any point of $\mathfrak{D}_0^{(n)}$ we have

$$\bar{f}_n(P;t_0)>\beta. \tag{III.09}$$

The result of the integration is

$$\begin{aligned} n\beta\mathfrak{M}(\mathfrak{D}_0^{(n)})&<n\int \bar{f}_n(P;t_0)d\omega_0 \\ &=\sum_{k=0}^{n-1}\int x_k(P)d\omega_0=\sum_{k=0}^{n-1}\int x_0(P_k)d\omega_k, \end{aligned} \tag{III.10}$$

where $\mathfrak{M}(\mathfrak{D}_0^{(n)})$ is the measure of the set $\mathfrak{D}_0^{(n)}$, where $d\omega_0^{(n)}$ indicates integration on the set $\mathfrak{D}_0^{(n)}$ and $d\omega_k^{(n)}$ integration over the set $\mathfrak{D}_k^{(n)}$ obtained from $\mathfrak{D}_0^{(n)}$ by the transformation $P\rightarrow P_k$.

Suppose that the sets $\mathfrak{D}_k^{(n)}$ are nonoverlapping and let their sum be $\mathfrak{D}^{(n)}$,

$$\mathfrak{D}^{(n)}=\sum \mathfrak{D}_k^{(n)}. \tag{III.11}$$

From Eq. (III.10) we now have

$$\int x_0(P)d\omega^{(n)}>\beta\mathfrak{M}(\mathfrak{D}^{(n)}), \tag{III.12}$$

where $\int d\omega^{(n)}$ indicates integration over $\mathfrak{D}^{(n)}$ and where we have used the fact that $\mathfrak{M}(\mathfrak{D}_k^{(n)})=\mathfrak{M}(\mathfrak{D}_0^{(n)})$ for all k, as a consequence of Liouville's theorem.

It can be shown that one can find such sums $\mathfrak{D}^{(n)}$ of nonoverlapping sets for each value of n in such a way that they exhaust $\mathfrak{D}'$ and from inequality (III.12) it thus follows that

$$\int x_0(P)d\omega'>\beta\mathfrak{M}(\mathfrak{D}'), \tag{III.13}$$

where $\int d\omega'$ indicates integration over the whole of $\mathfrak{D}'$.

One can similarly prove the inequality

$$\int x_0(P)d\omega'<\alpha\mathfrak{M}(\mathfrak{D}'), \tag{III.14}$$

and combining inequalities (III.13) and (III.14) with the assumption $\mathfrak{M}(\mathfrak{D}')>0$ we arrive at a contradiction with our choice of $\alpha<\beta$. It thus follows that $\mathfrak{D}$ is of measure zero and that the limit of Eq. (III.02) exists for practically all orbits.

The first part of the proof is now concluded by considering

$$A=\left|\frac{1}{T}\int_{t_0}^{t_0+T} f(P_t)dt-\frac{1}{n\tau}\int_{t_0}^{t_0+n\tau} f(P_t)dt\right|, \tag{III.15}$$

where now Eq. (III.03) does not necessarily hold, but n is the largest integer contained in T/τ. We can split A into two parts,

$$A=A_1+A_2, \tag{III.16}$$

with

$$A_1=\left|\left(\frac{1}{T}-\frac{1}{n\tau}\right)\int_{t_0}^{t_0+n\tau} f(P_t)dt\right|, \tag{III.17}$$

$$A_2=\left|\frac{1}{T}\int_{t_0}^{t_0+T} f(P_t)dt-\frac{1}{T}\int_{t_0}^{t_0+n\tau} f(P_t)dt\right|. \tag{III.18}$$

Since $\bar{f}_n(P;t_0)$ had a limiting value for $n\rightarrow\infty$, say, F, we have for A_1 the limit

$$A_1\rightarrow\frac{n\tau-T}{T}F, \tag{III.19}$$

and we see that $A_1\rightarrow 0$ for $T\rightarrow\infty$.

For A_2 we have the inequality

$$A_2=\frac{1}{T}\left|\int_{t_0+n\tau}^{t_0+T} f(P_t)dt\right|\leqslant\frac{\tau}{T}|x_n(P)|, \tag{III.20}$$

and for any well-behaved function $f(P)$ $x_n(P)$ will be bounded and thus $A_2\rightarrow 0$ for $T\rightarrow\infty$. This concludes the first part of the argument.

The fact that $\bar{f}(P;t_0,\infty)$ is independent of t_0 follows easily from the equations

$$\begin{aligned} \lim\frac{1}{T}\int_{t_0}^{t_0+T} f(P_t)dt&=\lim\frac{1}{T}\int_{t_1}^{t_1+T} f(P_t)dt \\ &+\lim\frac{1}{T}\int_{t_0}^{t_1} f(P_t)dt-\lim\frac{1}{T}\int_{t_0+T}^{t_1+T} f(P_t)dt, \end{aligned} \tag{III.21}$$

and for any well-behaved function $f(P)$ the last two limits will be equal to zero.

The final part of the proof of Birkhoff's ergodic theorem follows from the fact that, if $\bar{f}(P;t_0,\infty)$ were not constant practically everywhere on $\mathcal{S}$ it would be possible to find a value F of $\bar{f}(P;t_0,\infty)$ such that the conditions $\bar{f}(P;t_0,\infty)<F$ and $\bar{f}(P;t_0,\infty)\geqslant F$ would define two sets of positive measure on $\mathcal{S}$ which would both be invariant against the transformations $P\rightarrow P_t$, since $\bar{f}(P;t_0,\infty)$ is invariant against such transformations. This would, however, be in contradiction to the assumed metrical indecomposability of $\mathcal{S}$.

APPENDIX IV. PROOF OF THE QUANTUM-MECHANICAL ERGODIC THEOREM[130]

In Sec. E we found that the quantum mechanical ergodic theorem was proved once we could show that the time average $\bar{P}_\nu$ of the probability P_ν of finding the system in the νth cell is practically always equal to the ratio of the number of states s_ν in the cell to the number of states S_i of the corresponding energy shell. In our discussion we had used the fact that both s_ν and S_i are of the order of magnitude of $\exp(10^{20})$ and we shall use this fact also in the following. The expression "practically always" is meant in the sense that Eq. (E.26) holds for practically all subdivisions of phase space into phase cells, where the weight of different subdivisions will be defined presently.

Let ψ be the wave function of our system, let φ_σ be a complete orthonormal system corresponding to the Hamiltonian operator $\mathbf{H}$, and let ω_τ be a complete orthonormal system corresponding to the macroscopic operators discussed in part E. The set of the ω_τ will depend on the manner in which we have chosen our phase cells. We can express ψ in terms of the φ_σ or in terms of the ω_τ,

$$\psi=\textstyle\sum_\sigma r_\sigma\varphi_\sigma, \tag{IV.01}$$

$$\psi=\textstyle\sum_\tau t_\tau\omega_\tau. \tag{IV.02}$$

The φ_σ and ω_τ are related to each other through a unitary transformation,

$$\omega_\tau=\textstyle\sum_\sigma U_{\tau\sigma}\varphi_\sigma, \quad \varphi_\sigma=\sum_\tau U_{\sigma\tau}{}^*\omega_\tau, \tag{IV.03}$$

and the different ways in which the phase cells can be chosen is reflected in the different transformation matrices $\mathbf{U}$.

From Eqs. (IV.01) to (IV.03) it follows that

$$t_\tau=\textstyle\sum_\sigma r_\sigma U_{\tau\sigma}{}^*, \tag{IV.04}$$

and since the φ_σ are the characteristic functions of the energy operator we have

$$r_\sigma=|r_\sigma|\,\exp(iE_\sigma t/\hbar). \tag{IV.05}$$

The quantity P_ν is given by the equation

$$P_\nu=\textstyle\sum|t_\tau|^2, \tag{IV.06}$$

where the summation extends over the s_ν levels of the νth cell. From Eqs. (IV.04) to (IV.06) we get

$$P_\nu=\textstyle\sum_{\tau\sigma\rho}|r_\sigma|\,|r_\rho|\,U_{\tau\sigma}{}^*U_{\tau\rho}\exp[i(E_\sigma-E_\rho)t/\hbar], \tag{IV.07}$$

and for the time average

$$\bar{P}_\nu=\textstyle\sum_{\tau\sigma}|r_\sigma|^2|U_{\tau\sigma}|^2, \tag{IV.08}$$

where we used the fact that no two energy levels are the same.[131]

Consider now

$$|\bar{P}_\nu-s_\nu/S_i|=|\textstyle\sum_\sigma|r_\sigma|^2(C_\sigma-s_\nu/S_i)|, \tag{IV.09}$$

where

$$C_\sigma=\textstyle\sum_\tau|U_{\tau\sigma}|^2, \tag{IV.10}$$

and where we have used the fact that

$$\textstyle\sum_\sigma|r_\sigma|^2=1. \tag{IV.11}$$

From Eqs. (IV.09) and (IV.11) it follows that

$$|\bar{P}_\nu-s_\nu/S_i|<\max|C_\nu-s_\nu/S_i|. \tag{IV.12}$$

If we can prove that, for practically all transformations matrices $\mathbf{U}$, C_σ is practically equal to s_ν/S_i our theorem is proved. We must thus find the probability distribution of the quantity C_σ given by Eq. (IV.10). The $U_{\tau\sigma}$ are the components of unitary unit vectors in Hilbert space and the C_σ are the square of the length of the projection of these vectors on the subspace corresponding to the s_ν states of the phase cell under consideration. If we now assume that the probability of finding divisions corresponding to the unitary unit vectors $\mathbf{U}$ lying within a given solid angle in Hilbert space is proportional to the measure of this solid angle—or to the surface area cut out of the unit sphere in Hilbert space—the determination of the probability $W(C)dC$ that C_σ lies between C and $C+dC$ is reduced to a problem in multidimensional geometry. Similar problems are discussed by von Neumann (29N) and Pauli and Fierz (37P), and we shall only give the final result for $W(C)$ which is

$$W(C)=k\cdot C^{s_\nu}(1-C)^{S_i-s_\nu}, \tag{IV.13}$$

where k is a normalizing constant and where we have made use of Eqs. (E.20).

From Eq. (IV.13) we see, first of all, that the most probable value for C_σ is s_ν/S_i, and secondly, that the maximum is extremely sharp. In fact, the function $W(C)$ is reduced to half its value at a distance $S^{-\frac{3}{2}}$ from its maximum. From this it follows that for "practically all" divisions C_σ will have a value "practically" equal to s_ν/S_i and that thus Eq. (E.26) is "practically always" satisfied.

APPENDIX V. TIME-PROPORTIONAL TRANSITIONS[132]

In this appendix we wish to sketch the derivation of Eqs. (D2.45) and (D1.30) which played such an important role in our discussion of the quantum-mechanical H-theorem.

We, first of all, remember that in ordinary quantum-mechanical perturbation theory it follows from the

---

[130] See 29N, 37P, 52R; I am also indebted to Professor M. Fierz for correspondence on this topic.

[131] It must be noticed that the additional condition of no degeneracy of resonances is not needed.

[132] Compare 38T, Secs. 99, 100.

method of the variation of the constants (26D, 27D, 28P, 38K Sec. 53) that if $a_k(0)$ are the amplitudes at $t=0$, the amplitudes at a later time t will be given by the formula

$$a_n(t)=\sum_k V_{kn}\{\exp[i(E_n-E_k)t/\hbar]-1\} \times a_k(0)/(E_k-E_n), \qquad (\mathrm{V.01})$$

where the summation is over all these states which were represented at $t=0$, where the state n was not represented at $t=0$, where V_{kn} is the matrix element of the perturbation operator $\mathbf{V}$ through which the transitions take place [compare the discussion of the Born-Green considerations in Sec. D(2)], and where E_k is the energy value of the state k.

Consider now the situation where we want to know the number of transitions from one group S_i of energy levels to another group S_j of energy levels. Let us first of all assume that at $t=0$ we know from observation that the system is in one of the S_i states of the ith group. In accordance with our basic assumption of equal *a priori* probabilities and random *a priori* phases we have at $t=0$ for the density matrix representing our system

$$\left.\begin{aligned}\rho_{kl}=\langle a_k a_l^*\rangle&=\delta_{kl}/S_i,\\ &\text{if state } k \text{ belongs to the } i\text{th group,}\\ &=0, \quad \text{otherwise.}\end{aligned}\right\}(\mathrm{V.02})$$

These equations follow since if both k and l belong to the ith group and are unequal $\langle a_k a_l^*\rangle$ is zero when we average over the phases; the remainder of the equations follows easily.[133]

The probability P_j' of finding the system at t in a state of the jth group follows from Eq. (V.01) and we get

$$\begin{aligned}P_j'&=\sum_n a_n^*(t)a_n(t)\\ &=\sum_{k,l,n} V_{kn}^*V_{ln}\{\exp[-i(E_k-E_n)t/\hbar]-1\}\\ &\quad\times\{\exp[i(E_l-E_n)t/\hbar]-1\}\\ &\quad\times a_k^*(0)a_l(0)/(E_n-E_k)(E_n-E_l). \qquad (\mathrm{V.03})\end{aligned}$$

Averaging over the representative ensemble and using Eqs. (V.02), we get for the P_j the expression

$$P_j(t)=(4/S_i)\sum_{n,k}|V_{kn}|^2 \times[\sin^2\tfrac{1}{2}(E_k-E_n)t/\hbar]/(E_k-E_n)^2. \qquad (\mathrm{V.04})$$

In Eqs. (V.03) and (V.04) the states k and l belong to the ith group and the states n to the jth group. To evaluate the sum in Eq. (V.04) we replace it by a double integration and get (compare 38T Sec. 99)

$$P_j(t)=(2\pi/\hbar S_i)\sum_{kn}|V_{kn}|^2\sigma_n(E)\sigma_k(E)\Delta E\, t, \qquad (\mathrm{V.05})$$

where $\sigma_n(\sigma_k)$ are the density of the energy levels in the jth (ith) group and where ΔE is the range of energy corresponding to the jth group.

Equation (V.05) can be written in the form

$$P_j=T_{ij}t/S_i, \qquad (\mathrm{V.06})$$

and, if we had been interested in a situation where originally observation had shown the system to be in the jth group and we would like to know the probability of finding the systems in a state of the ith group, we would have got the result

$$P_i=T_{ji}t/S_j, \qquad (\mathrm{V.07})$$

and from Eq. (V.05) and the Hermitean character of $\mathbf{V}$ it follows that

$$T_{ij}=T_{ji}. \qquad (\mathrm{V.08})$$

If observation had only shown that at $t=0$ the different groups were occupied with probabilities P_i, P_j, $\cdots$, we should instead of Eq. (V.02) for $\boldsymbol{\rho}$ have used the equation

$$\rho_{kl}=\delta_{kl}P_i/S_i \text{ (state } k \text{ belonging to the } i\text{th group),} \qquad (\mathrm{V.09})$$

and by an analysis similar to the one just given we would have found that the rate of transition N_{ij} from the ith group to the jth group would be given by the equation

$$N_{ij}=T_{ij}P_i/S_i, \qquad (\mathrm{V.10})$$

which reduces to Eq. (D2.45) by the substitution

$$A_{ij}=T_{ij}/S_iS_j, \qquad (\mathrm{V.11})$$

while from Eqs. (V.08) and (V.11) Eq. (D2.46) follows.

In order to derive Eq. (D1.30) we remind ourselves that we are dealing with a system containing N practically independent particles so that the Hamiltonian of the system can be written in the form

$$\mathbf{H}=\mathbf{H}_0+\mathbf{V}, \qquad (\mathrm{V.12})$$

where

$$\mathbf{H}_0=\sum_i\mathbf{H}_i, \qquad (\mathrm{V.13})$$

$\mathbf{H}_i$ being the Hamiltonian of the ith particle, and where $\mathbf{V}$ is the interaction operator which is neglected to a first approximation, but which is necessary in order that transitions can take place between different states of the system.[134] The characteristic functions of $\mathbf{H}_0$ are of the form

$$\Phi_k=\prod_i \varphi_i, \quad E_k=\sum_i \epsilon_i, \qquad (\mathrm{V.14})$$

where the φ_i are the characteristic functions of the $\mathbf{H}_i$, E_k the characteristic value of $\mathbf{H}_0$ corresponding to Φ_k, and the ϵ_i the characteristic values of the $\mathbf{H}_i$. A wave function $P\Phi_k$ obtained from Φ_k by a permutation of the arguments of the N particles is another characteristic function of $\mathbf{H}_0$ belonging to E_k.

If we are dealing with a system of Fermi-Dirac particles we can only allow those wave functions which are antisymmetrical in the arguments of all N particles and we have thus only one acceptable combination of the $P\Phi_k$ in this case, namely,

$$\Phi_{\text{F-D}}=C\sum(-)^P P\Phi_k, \qquad (\mathrm{V.15})$$

---

[133] One takes into account the fact that $\boldsymbol{\rho}$ is normalized.

[134] Compare the analogous position in the case of Boltzmann's H-theorem, where the Maxwell-Boltzmann distribution which really presupposes independent particles is established through collisions, that is, by means of their interactions.

when $(-)^P$ is $+1$ or -1 according as the permutation is even or odd.

In the case of a system of Bose-Einstein particles the only acceptable wave function is completely symmetrical, or,

$$\Phi_{\text{B-E}} = C\sum P\Phi_k, \tag{V.16}$$

where in Eqs. (V.15) and (V.16) the summations extend over all $N!$ permutations, and where in both cases the factor C occurs as a normalization constant.

In the following we shall restrict ourselves to a system of Bose-Einstein particles. The case of Fermi-Dirac particles is simpler, but otherwise analogous. We first of all assume that $\mathbf{V}$ is of the form

$$\mathbf{V} = \sum \mathbf{V}_{\alpha\beta}, \tag{V.17}$$

where α and β number the N particles, and where the summation extends over all pairs of the system $(\alpha<\beta)$. This means that we assume that only two-body collisions are of importance and that three-body or higher-order collisions can be neglected. As we are already assuming that the system is so dilute that to a first approximation $\mathbf{V}$ can be neglected this is a reasonable assumption.[135]

The wave function Φ_k may contain n_1 factors φ_1, n_2 factors $\varphi_2, \cdots$, n_i factors $\varphi_i, \cdots$, where the φ_i are a set of ortho-normalized characteristic functions. In order to evaluate the factor C in Eq. (V.16), one must combine only those permutations which do not lead to products of orthogonal φ_i with the same argument[136] so that we finally get for C the equation

$$|C|^2 = [N!n_1!n_2!n_3!\cdots n_i!\cdots]^{-1}. \tag{V.18}$$

In a similar manner we can evaluate the matrix element V_{kl} corresponding to a transition from the original state to a state where instead of n_i factors φ_i, n_j factors φ_j, $n_{i'}$ factors $\varphi_{i'}$, and $n_{j'}$ factors $\varphi_{j'}$ we have n_i+1 factors φ_i, n_j+1 factors φ_j, $n_{i'}-1$ factors $\varphi_{i'}$, and $n_{j'}-1$ factors $\varphi_{j'}$. Introducing the two integrals

$$I_1 = \int \varphi_i^*(\alpha)\varphi_j^*(\beta)\mathbf{V}_{\alpha\beta}\varphi_{i'}(\alpha)\varphi_{j'}(\beta)d\tau_\alpha d\tau_\beta, \tag{V.19}$$

$$I_2 = \int \varphi_i^*(\beta)\varphi_i^*(\alpha)\mathbf{V}_{\alpha\beta}\varphi_{i'}(\alpha)\varphi_{j'}(\beta)d\tau_\alpha d\tau_\beta, \tag{V.20}$$

one finds

$$V_{kl} = |I_1+I_2|^2(n_i+1)(n_j+1)n_{i'}n_{j'}. \tag{V.21}$$

In the Fermi-Dirac case we would have found instead of Eq. (V.21) the expression[137]

$$V_{kl} = |I_1-I_2|^2 n_{i'}n_{j'}(n_i-1)(n_j-1). \tag{V.22}$$

Up to now we have not introduced the groups of energy levels which we considered in Sec. D(1). We do this now and then ask for the number of transitions which take place per unit time from Z_i and Z_j to $Z_{i'}$ and $Z_{j'}$, where therefore N_i, N_j, $N_{i'}$, $N_{j'}$, change to N_i-1, N_j-1, $N_{i'}+1$, $N_{j'}+1$. Once again we use the method of the variation of constants and the analysis very similar to the one leading to Eq. (V.06) so that we shall only sketch it.

The probability of $P_{\mathrm{f}}(t)$[138] of finding the system at t in one of the states corresponding to N_i-1, N_j-1, $N_{i'}+1$, $N_{j'}+1$, if the state was originally in one of the states corresponding to N_i, N_j, $N_{i'}$, $N_{j'}$, is given by the expression

$$\begin{aligned} P_{\mathrm{f}} = \sum_{\mathrm{f}}|a_{\mathrm{f}}(t)|^2 = \sum_{\mathrm{foo'}} V_{\mathrm{fo'}}{}^*V_{\mathrm{fo}}\rho_{\mathrm{oo'}}(0) \\ \times\{\exp[-i(E_{\mathrm{f}}-E_{\mathrm{o}})t/\hbar]-1\} \\ \times\{\exp[-i(E_{\mathrm{f}}-E_{\mathrm{o'}})t/\hbar]-1\}/ \\ (E_{\mathrm{o'}}-E_{\mathrm{f}})(E_{\mathrm{o}}-E_{\mathrm{f}}), \end{aligned} \tag{V.23}$$

where $\rho_{\mathrm{oo'}}(0)$ is the density matrix at $t=0$,

$$\rho_{\mathrm{oo'}}(0) = a_{\mathrm{o'}}{}^*(0)a_{\mathrm{o}}(0), \tag{V.24}$$

and where the a_{f}'s and a_{o}'s are again the probability amplitudes.

Again, we assume that we have made a measurement at $t=0$ giving us the information that at $t=0$ one of the initial states was realized so that we can write for the density matrix

$$\left.\begin{aligned} \rho_{\mathrm{oo'}}(0) &= (1/g_{\mathrm{o}})\delta_{\mathrm{o'o}}, \\ &\quad \text{if state o satisfies our requirements,} \\ &= 0, \quad \text{otherwise,} \end{aligned}\right\} \tag{V.25}$$

where g_{o} is the number of states satisfying our requirements.

In Eq. (V.23) we can now introduce expressions (V.21) and (V.22) for V_{kl} and write for $E_{\mathrm{f}}-E_{\mathrm{o}}$,

$$E_{\mathrm{f}}-E_{\mathrm{o}} = \epsilon_i+\epsilon_j-\epsilon_{i'}-\epsilon_{j'}. \tag{V.26}$$

Writing further

$$n_i = N_i/Z_i, \quad n_j = N_j/Z_j, \quad n_{i'} = N_{i'}/Z_{i'}, \quad n_{j'} = N_{j'}/Z_{j'}, \tag{V.27}$$

that is, replacing the $n_i, \cdots$ by their mean values in the corresponding group, and integrating we finally get for $P_{\mathrm{f}}(t)$ the expression

$$\begin{aligned} P_{\mathrm{f}}(t) = (2\pi/\hbar)[|I_1\pm I_2|^2/\Delta E] \\ \times N_iN_j(Z_{i'}\pm N_{i'})(Z_{j'}\pm N_{j'})t, \end{aligned} \tag{V.28}$$

---

[135] There are very few cases where three-body collisions—to restrict ourselves to those—have been considered. They enter into the calculation of the third virial coefficient (39B3, 40B, 41M) but up to now only the classical case has been studied and the difficulties to be expected in a quantum-mechanical consideration are formidable.

[136] We only sketch the argument here and refer to Tolman (38T Sec. 100) for a detailed derivation.

[137] This formula can be found in papers by Jordan (25J, 27J1), Ornstein and Kramers (27O), and Bothe (28B); the further development is due to Tolman.

[138] The subscript f indicates the final state, while the subscripts o and o′ indicate the original states.

where ΔE is the width of the energy groups. From Eq. (V.28) Eq. (D1.30) follows easily, and Eq. (D1.31) is now a consequence of the Hermitean nature of the operator $\mathbf{V}$.

We have now derived Eqs. (D1.30) and (D2.45) which played such an important role in the discussion of the quantum-mechanical H-theorem. We have used (i) the quantum-mechanical picture of time dependent transitions, and (ii) the statistical assumption of equal *a priori* probabilities and random *a priori* phases. The second point has been discussed in Sec. D(3), but with regard to the first point we should like to conclude this appendix with a quotation from Pauli's classical paper (28P): "The statistical laws for the frequency of transitions between stationary states, which are of the same character as the laws for radioactive decay, guarantee in themselves such a sufficient degree of randomness as is necessary for the statistical interpretation of the second law of thermodynamics."[139]

APPENDIX VI. KLEIN'S LEMMA[140]

In Sec. D(2) we used the fact that expression (D2.56) would never be positive. This property will now be proved. We are dealing with two density matrices $\boldsymbol{\varrho}'$ and $\boldsymbol{\varrho}''$ representing a system at t' and t'' with matrix elements ρ_{kl}' and ρ_{kl}'', where

$$\rho_{kl}'=\rho_{kk}'\delta_{kl}, \tag{VI.01}$$

but $\boldsymbol{\varrho}''$ is not necessarily a diagonal matrix. The diagonal elements ρ_{kk}' (ρ_{kk}'') are, as we saw in Sec. D(2), the probabilities of finding the system at t' (t'') in the kth state. They are thus never negative and the expression Q_{kn} given by the equation

$$Q_{kn}=\rho_{kk}'[\ln\rho_{kk}'-\ln\rho_{nn}''-1]+\rho_{nn}'' \tag{VI.02}$$

is never negative as a consequence of Eqs. (D2.58).

The ρ_{kl}'' are determined through Eq. (D2.07) by the probability amplitudes at t'' which in turn follow from the probability amplitudes at t' by integration of the Schrödinger equation. It is well known that one can obtain the probability amplitudes at t'', $a_n^k(t'')$, from those at t', $a_n^k(t')$, by a unitary transformation $\mathbf{U}$,

$$a_n^k(t'')=\sum_l a_l^k(t')U_{nl}. \tag{VI.03}$$

Using Eq. (VI.03), we get from Eq. (D2.07)

$$\rho_{nn}''=\sum_k|U_{nk}|^2\rho_{kk}'+\sum_{l\neq k}U_{nl}^*U_{nk}\rho_{kl}', \tag{VI.04}$$

or using Eq. (VI.01),

$$\rho_{nn}''=\sum_k|U_{nk}|^2\rho_{kk}'. \tag{VI.05}$$

Multiplying Eq. (VI.02) by the positive quantity $|U_{nk}|^2$ and summing over all values of n and k, we get

$$\sum_{k,n}|U_{nk}|^2\rho_{kk}'\ln\rho_{kk}'-\sum_{k,n}|U_{nk}|^2\rho_{kk}'\ln\rho_{nn}'' \\ +-\sum_{k,n}|U_{nk}|^2\rho_{kk}'+\sum_{k,n}|U_{nk}|^2\rho_{nn}''\geqslant 0. \tag{VI.06}$$

Using Eq. (VI.05) and the property of unitary matrices that

$$\sum_n U_{nk}^*U_{nl}=\delta_{kl}, \tag{VI.07}$$

Eq. (VI.06) reduces to

$$\sum_k\rho_{kk}'\ln\rho_{kk}'-\sum_n\rho_{nn}''\ln\rho_{nn}'' \\ -\sum_k\rho_{kk}'+\sum_n\sigma_{nn}''\geqslant 0, \tag{VI.08}$$

or

$$\sum_k\rho_{kk}'\ln\rho_{kk}'-\sum_n\rho_{nn}''\ln\rho_{nn}''\geqslant 0, \tag{VI.09}$$

since as a result of the normalization of $\boldsymbol{\varrho}'$ and $\boldsymbol{\varrho}''$ [Eq. (D2.10)] we have

$$\sum_k\rho_{kk}'=\sum_n\rho_{nn}''. \tag{VI.10}$$

We have now proved Klein's lemma. It was stressed by Pauli (49P) that the essential content of this lemma is that $\mathrm{Tr}(\boldsymbol{\varrho}\ln\boldsymbol{\varrho})$ increases on the substitution of $\boldsymbol{\varrho}$ by a diagonalized matrix in such a way that one just puts the off-diagonal elements equal to zero. As soon as this is realized, Born and Green's proof of the decrease of $\sigma\Sigma$ is really trivial and could have been replaced by a reference to Klein's lemma.

APPENDIX VII. THE PRINCIPLE OF DETAILED BALANCING[141]

In Sec. D(2) we mentioned that Eq. (D2.50) expressed the fact that at equilibrium as many transitions from the ith groups of energy levels to the jth group as the other way round. This is an example of a principle which is valid in a great many cases and which states that at equilibrium the number of processes destroying situation A and creating situation B will be equal to the number of processes producing A and destroying B. Other instances of situations where the principle of detailed balancing applies were met in Sec. A(1) [Eq. (A1.12)] in the case of a gas of spherical molecules, Sec. A(3) (equation in footnote 36) in the case of some general models of a gas, and Sec. D(1) [Eq. (D1.04) and Eq. (D1.31) combined with Eqs. (D1.30) and (D1.27)].

From these examples one might get the impression that detailed balancing would always hold. In the case of classical statistics, Lorentz (87L) showed that it could not hold in the case of polyatomic molecules which could not be considered to be spheres (see also 98B). He showed how one could still prove an H-theorem where now cycles of transitions had to be considered instead of collisions and their inverse collisions.[142]

---

[139] After this was written a paper by van Kampen (54K) was published in which the symmetry of the transition matrix is discussed. Van Kampen draws special attention to the fact that, if magnetic fields or Coriolis forces are present, this symmetry must be replaced by a slightly weaker condition which does not follow from the Hermiticity of the Hamiltonian. Van Kampen also discusses in his paper the relation between the Onsager relations (31O1, 31O2, 45C) and the symmetry of the transition matrix.

[140] See 31K; 38T, Secs. 101 and 106.

[141] See 25F2, 25T, 38T Sec. 50, ESM p. 381.

[142] For a discussion of this more general H-theorem we refer the reader to Tolman's monograph (38T). On p. 119 of this monograph one finds an example of a collision which does not possess an inverse collision.

The same difficulty arises in quantum statistics.[143] Hamilton and Peng (44H1; see also 44H2, 52S2) have shown that in the system composed of particles with spin and electromagnetic radiation detailed balancing does not apply.[144]

The principle is of great importance for its applications, for instance, in the consideration of rate processes (11K1, 15M, 23C2). If, for instance, we wish to calculate the number of collisions in a gas which lead to excited states of a molecule—which may be a chemically active state—we can calculate the number of collisions leading to deactivation and use detailed balancing to give us the first number. Since the second number is more easily computed than the first, we have a clear gain.

In 1925 an article by Lewis (25L1) evoked a discussion (25F2, 25L2, 25T) of the principle of detailed balancing, called by Lewis the principle of entire equilibrium or the law of reversibility to the last detail.[145] Fowler and Milne especially give an impressive list of applications. They point out that the principle was to a large extent inspired by Einstein's classical paper (17E2) on transition probabilities. They also point out that the principle is really an extension of the ideas of Kirchhoff (60K) while it was probably first formulated by Richardson (14R1, 24R).

In 1916 Langmuir (16L) applied the principle to the problem of evaporation and condensation, stating: "Since evaporation and condensation are in general thermodynamically reversible phenomena, *the mechanism of evaporation must be the exact reverse of that of condensation*,[146] even down to the smallest detail."

In 1921 Klein and Rosseland (21K) used the principle of detailed balancing in their discussion of inelastic collisions between atoms and electrons. Franck and Cario (22F, 22C1, 22C2, 23C1, 23F4, 24F4) applied it also to collision processes. Fowler (24F1, 24F2, 24F3) applied it to the phenomena of capture and loss of electrons by high velocity α particles. Becker (23B1), Kramers (23K), and Milne (24M) applied it to the photoelectric process of ionization and capture. Eddington (22E, 24E1) used it in his investigations of the stellar absorption coefficient. Pauli (23P) and Einstein and Ehrenfest (23E2) discussed with its aid the scattering of radiation by electrons. Dirac (24D) used it to discuss multiple collision processes. Lewis (25L2) used it to discuss Planck's radiation law. From this list it can be seen that the principle is certainly one of the more important ones and one which can be applied to a large number of different problems.

BIBLIOGRAPHY

38B D. Bernoulli, *Hydrodynamica* (Dulsecker, Argentorati, 1738).

56K A. Krönig, Ann. Physik **99**, 315 (1856).

57C1 R. Clausius, Ann. Physik **100**, 353 (1857).
57C2 R. Clausius, Phil. Mag. **14**, 108 (1857).

58C R. Clausius, Ann. Physik **105**, 239 (1858).

60K G. Kirchhoff, Ann. Physik **109**, 148 (1860).
60M J. C. Maxwell, Phil. Mag. **19**, 19 (1860).

62C R. Clausius, Ann. Physik **115**, 2 (1862).

67M J. C. Maxwell, Trans. Roy. Soc. (London) **157**, 49 (1867).

68B L. Boltzmann, Wien Ber. **58**, 517 (1868).
68M1 J. C. Maxwell, Phil. Mag. **35**, 129 (1868).
68M2 J. C. Maxwell, Phil. Mag. **35**, 185 (1868).

70C1 R. Clausius, Ann. Physik **141**, 124 (1870).
70C2 R. Clausius, Phil. Mag. **40**, 122 (1870).

71B1 L. Boltzmann, Wien. Ber. **63**, 397 (1871).
71B2 L. Boltzmann, Wien. Ber. **63**, 679 (1871).

72B L. Boltzmann, Wien. Ber. **66**, 275 (1872).

75B L. Boltzmann, Wien. Ber. **72**, 427 (1875).

76L J. Loschmidt, Wien. Ber. **73**, 139 (1876).

77B L. Boltzmann, Wien. Ber. **76**, 373 (1877).
77L J. Loschmidt, Wien. Ber. **75**, 67 (1877).

79M J. C. Maxwell, Trans. Cambridge Phil. Soc. **12**, 547 (1879).

87B L. Boltzmann, J. Math. **100**, 201 (1877).
87L H. A. Lorentz, Wien. Ber. **95**, 115 (1877).

90P H. Poincaré, Acta Math. **13**, 67 (1890).

91K Lord Kelvin, *Collected Works* (Cambridge University Press, Cambridge, 1891), Vol. IV.

92P H. Poincaré, *Méthodes Nouvelles de la Mécanique Céleste* (Gauthiers-Villars, Paris, 1892), Vol. I.

95B L. Boltzmann, Nature **51**, 413 (1895).

96B L. Boltzmann, *Vorlesungen über Gastheorie* (Barth, Leipzig, 1896), Vol. I.

96Z E. Zermelo, Ann. Physik **57**, 485 (1896).

97B L. Boltzmann, Wien. Ber. **106**, 12 (1897).

98B L. Boltzmann, *Vorlesungen über Gastheorie* (Barth, Leipzig, 1898), Vol. II.

00P1 M. Planck, Verhandl. deut. physik. Ges. **2**, 202 (1900).
00P2 M. Planck, Verhandl. deut. physik. Ges. **2**, 237 (1900).

02E A. Einstein, Ann. Physik **9**, 417 (1902).
02G J. W. Gibbs, *Elementary Principles in Statistical Mechanics* (Yale University Press, New Haven, 1902).

03B S. H. Burbury, Phil. Mag. **6**, 251 (1903).
03E A. Einstein, Ann. Physik **11**, 170 (1903).

04B1 H. A. Bumstead, Phil. Mag. **7**, 8 (1904).
04B2 S. H. Burbury, Phil. Mag. **8**, 43 (1904).

06E1 P. Ehrenfest, Physik. Z. **7**, 528 (1906).
06E2 T. and P. Ehrenfest, Wien. Ber. **115**, 89 (1906).
06P H. Poincaré, J. phys. **5**, 369 (1906).

07E P. and T. Ehrenfest, Physik. Z. **8**, 311 (1907).
07L H. A. Lorentz, *Abhandlungen über theoretische Physik* (Teubner, Leipzig, 1907), p. 202.

08O L. S. Ornstein, thesis, Leiden (1908).

---

[143] The statement in ESM that in quantum mechanics the principle of detailed balancing always holds is incorrect.

[144] See also the recent paper by van Kampen (54K) which shows that the same applies to situations where magnetic fields are present. See also a paper by Klein (55K1) showing that detailed balancing does not hold generally for nonequilibrium steady states (compare also 54L2).

[145] Tolman (24T, 25T) sometimes uses the expression "principle of microscopic reversibility."

[146] Langmuir's italics. We may remark here that Langmuir's statement only holds at equilibrium (compare also footnote 144).

09L H. A. Lorentz, *The Theory of Electrons* (Teubner, Leipzig, 1909), Note 29.

10E A. Einstein, Ann. Physik **33**, 1275 (1910).

11E1 P. Ehrenfest, Ann. Physik **36**, 91 (1911).
11E2 P. and T. Ehrenfest, Encykl. math. Wiss. **4**, No. 32 (1911).
11K1 P. Kohnstamm and F. E. C. Scheffer, Proc. Amsterdam Acad. Sci. **13**, 789 (1911).
11K2 J. Kroò, Ann. Physik **34**, 907 (1911).

12M A. A. Markoff, *Wahrscheinlichkeitsrechnung* (Teubner, Leipzig, 1912).
12S M. von Smoluchowski, Physik. Z. **13**, 1069 (1912).

13B1 N. Bohr, Phil. Mag. **26**, 1 (1913).
13B2 N. Bohr, Nature **92**, 231 (1913).
13P M. Plancherel, Ann. Physik **42**, 1061 (1913).
13R A. Rosenthal, Ann. Physik **42**, 796 (1913).

14E P. Ehrenfest, Physik. Z. **15**, 657 (1914).
14R1 O. W. Richardson, Phil. Mag. **27**, 476 (1914).
14R2 A. Rosenthal, Ann. Physik **43**, 894 (1914).

15M R. Marcelin, Ann. Physik **3**, 120 (1915).

16E P. Ehrenfest, Ann. Physik **51**, 327 (1916).
16L I. Langmuir, J. Am. Chem. Soc. **38**, 2221 (1916).
16S O. Stern, Ann. Physik **49**, 823 (1916).

17B1 J. M. Burgers, Ann. Physik **52**, 195 (1917).
17B2 J. M. Burgers, Proc. Amsterdam Acad. Sci. **20**, 149 (1917).
17B3 J. M. Burgers, Proc. Amsterdam Acad. Sci. **20**, 158 (1917).
17B4 J. M. Burgers, Proc. Amsterdam Acad. Sci. **20**, 163 (1917).
17E1 P. Ehrenfest, Proc. Amsterdam Acad. Sci. **19**, 576 (1917).
17E2 A. Einstein, Physik. Z. **18**, 121 (1917).

18B J. M. Burgers, thesis, Leiden (1918).

19S A. Sommerfeld, *Atombau und Spektrallinien* (Vieweg, Brunswick, 1919).

20M1 R. von Mises, Physik. Z. **21**, 225 (1920).
20M2 R. von Mises, Physik. Z. **21**, 256 (1920).

21K O. Klein and S. Rosseland, Z. Physik **4**, 46 (1921).

22B G. D. Birkhoff, Acta Math. **43**, 113 (1922).
22C1 G. Cario, Z. Physik **10**, 185 (1922).
22C2 G. Cario and J. Franck, Z. Physik **11**, 161 (1922).
22D1 C. G. Darwin and R. H. Fowler, Phil. Mag. **44**, 450 (1922).
22D2 C. G. Darwin and R. H. Fowler, Phil. Mag. **44**, 823 (1922).
22D3 C. G. Darwin and R. H. Fowler, Proc. Cambridge Phil. Soc. **21**, 262 (1922).
22E A. S. Eddington, Monthly Notices Roy. Astron. Soc. **83**, 32 (1922).
22F J. Franck, Z. Physik **9**, 259 (1922).

23B1 R. Becker, Z. Physik **18**, 325 (1923).
23B2 N. Bohr, Z. Physik **13**, 117 (1923).
23C1 G. Cario and J. Franck, Z. Physik **17**, 202 (1923).
23C2 J. A. Christiansen and H. A. Kramers, Z. physik Chem. **104**, 451 (1923).
23D1 C. G. Darwin and R. H. Fowler, Proc. Cambridge Phil. Soc. **21**, 391 (1923).
23D2 C. G. Darwin and R. H. Fowler, Proc. Cambridge Phil. Soc. **21**, 730 (1923).
23E1 P. Ehrenfest, Naturwiss. **11**, 543 (1923).
23E2 A. Einstein and P. Ehrenfest, Z. Physik **19**, 301 (1923).
23F1 E. Fermi, Physik. Z. **24**, 261 (1923).
23F2 R. H. Fowler, Phil. Mag. **45**, 1 (1923).
23F3 R. H. Fowler, Phil. Mag. **45**, 497 (1923).
23F4 J. Franck, Ergeb. exakt. Naturw. **12**, 112 (1923).
23K H. A. Kramers, Phil. Mag. **46**, 836 (1923).
23P W. Pauli, Z. Physik **18**, 272 (1923).

24B S. N. Bose, Z. Physik **26**, 178 (1924).
24D P. A. M. Dirac, Proc. Roy. Soc. (London) **A106**, 581 (1924).
24E1 A. S. Eddington, Monthly Notices Roy. Astron. Soc. **84**, 104 (1924).
24E2 A. Einstein, Sitzber. preuss. Akad. Wiss., Physik.-math. Kl., p. 261 (1924).
24F1 R. H. Fowler, Phil. Mag. **47**, 257 (1924).
24F2 R. H. Fowler, Phil. Mag. **47**, 416 (1924).
24F3 R. H. Fowler, Proc. Cambridge Phil. Soc. **22**, 253 (1924).
24F4 J. Franck, Naturwissenschaften **12**, 1066 (1924).
24M E. A. Milne, Phil. Mag. **47**, 209 (1924).
24N L. Nordheim, Z. Physik **27**, 65 (1924).
24R O. W. Richardson, Proc. Phys. Soc. (London) **36**, 383 (1924).
24T R. C. Tolman, Phys. Rev. **23**, 693 (1924).

25B M. Born and P. Jordan, Z. Physik **34**, 858 (1925).
25E1 A. Einstein, Sitzber. preuss. Akad. Wiss., Physik.-math. Kl., p. 3 (1925).
25E2 A. Einstein, Sitzber. preuss. Akad. Wiss., Physik.-math. Kl., p. 18 (1925).
25F1 R. H. Fowler, Proc. Cambridge Phil. Soc. **22**, 861 (1925).
25F2 R. H. Fowler and E. A. Milne, Proc. Natl. Acad. Sci. **11**, 400 (1925).
25H W. Heisenberg, Z. Physik **33**, 879 (1925).
25J P. Jordan, Z. Physik **33**, 649 (1925).
25L1 G. N. Lewis, Proc. Natl. Acad. Sci. **11**, 179 (1925).
25L2 G. N. Lewis, Proc. Natl. Acad. Sci. **11**, 423 (1925).
25P W. Pauli, Z. Physik **31**, 776 (1925).
25T R. C. Tolman, Proc. Natl. Acad. Sci. **11**, 436 (1925).
25W H. Weyl, Math. Z. **23**, 271 (1925).

26D P. A. M. Dirac, Proc. Roy. Soc. (London) **A112**, 661 (1926).
26F1 E. Fermi, Z. Physik **36**, 902 (1926).
26F2 R. H. Fowler, Phil. Mag. **1**, 845 (1926).
26F3 R. H. Fowler, Proc. Roy. Soc. (London) **A113**, 432 (1926).
26H W. Heisenberg, Z. Physik **40**, 501 (1926).
26J P. Jordan, Z. Physik **40**, 661 (1926).
26S1 E. Schrödinger, Ann. Physik **79**, 361 (1926).
26S2 A. Smekal, Encykl. math. Wiss. **5**, No. 28 (1926).

27D P. A. M. Dirac, Proc. Roy. Soc. (London) **A114**, 243 (1927).
27E P. Ehrenfest and G. E. Uhlenbeck, Z. Physik **41**, 24 (1927).
27H1 W. Heisenberg, Z. Physik **43**, 172 (1927).
27H2 P. Höflich, Z. Physik **41**, 636 (1927).
27J1 P. Jordan, Z. Physik **41**, 711 (1927).
27J2 P. Jordan and O. Klein, Z. Physik **45**, 751 (1927).
27N1 J. von Neumann, Nachr. Akad. Wiss. Göttingen, Math.-physik. Kl., p. 245 (1927).
27N2 J. von Neumann, Nachr. Akad. Wiss. Göttingen, Math.-physik. Kl., p. 271 (1927).
27O L. S. Ornstein and H. A. Kramers, Z. Physik **42**, 481 (1927).
27P W. Pauli, Z. Physik **41**, 91 (1927).
27U G. E. Uhlenbeck, thesis, Leiden (1927).

28B W. Bothe, Z. Physik **46**, 327 (1928).
28J P. Jordan and E. Wigner, Z. Physik **47**, 361 (1928).
28N L. Nordheim, Proc. Roy. Soc. (London) **A119**, 689 (1928).
28P W. Pauli in *Probleme der Modernen Physik*, Sommerfeld Festschrift; P. Debye, ed. (Hirzel, Leipzig, 1928) p. 30.

29D P. A. M. Dirac, Proc. Cambridge Phil. Soc. **25**, 62 (1929).
29N J. von Neumann, Z. Physik **57**, 30 (1929).
29P R. Peierls, Ann. Physik **3**, 1055 (1929).
29S L. Szilard, Z. Physik **53**, 840 (1929).

30D1 P. A. M. Dirac, Proc. Cambridge Phil. Soc. **26**, 361 (1930).
30D2 P. A. M. Dirac, Proc. Cambridge Phil. Soc. **26**, 376 (1930).
30H1 W. Heisenberg, *Die Physikalischen Prinzipien der Quantentheorie* (Hirzel, Leipzig, 1930).
30H2 E. Hopf, Math. Ann. **103**, 710 (1930).
30S F. Simon, Ergeb. exakt. Naturw. **9**, 222 (1930).

31B1 G. D. Birkhoff, Proc. Natl. Acad. Sci. **17**, 650 (1931).
31B2 G. D. Birkhoff, Proc. Natl. Acad. Sci. **17**, 656 (1931).
31D P. A. M. Dirac, Proc. Cambridge Phil. Soc. **27**, 240 (1931).
31K O. Klein, Z. Physik **72**, 767 (1931).
31O1 L. Onsager, Phys. Rev. **37**, 405 (1931).
31O2 L. Onsager, Phys. Rev. **38**, 2265 (1931).

32B G. D. Birkhoff and B. O. Koopman, Proc. Natl. Acad. Sci. **18**, 279 (1932).
32H1 E. Hopf, Proc. Natl. Acad. Sci. **18**, 93 (1932).
32H2 E. Hopf, Proc. Natl. Acad. Sci. **18**, 204 (1932).
32H3 E. Hopf, Proc. Natl. Acad. Sci. **18**, 333 (1932).
32H4 E. Hopf, Sitzber. preuss. Akad. Wiss., Physik.-math. Kl., p. 182 (1932).
32N1 J. von Neumann, Proc. Natl. Acad. Sci. **18**, 70 (1932).
32N2 J. von Neumann, Proc. Natl. Acad. Sci. **18**, 263 (1932).
32N3 J. von Neumann, *Mathematische Grundlagen der Quanten mechanik* (Verlag Julius Springer, Berlin, 1932).

33B1 N. Bohr, Nature **131**, 421 (1933).
33B2 N. Bohr, Nature **131**, 457 (1933).
33H *Handbuch der Physik* **24**, Part I (Verlag Julius Springer, Berlin, 1933).
33J P. Jordan, *Statistische Mechanik auf Quanten theoretischer Grundlage* (Vieweg, Brunswick, 1933).
33P W. Pauli, *Handbuch der Physik* **24**, Part I, 149 (Verlag Julius Springer, Berlin, 1933).

34F V. Fock, Z. Physik **75**, 622 (1934).
34H E. Hopf, J. Math. Phys. **13**, 51 (1934).

35B N. Bohr, Phys. Rev. **48**, 696 (1935).
35D P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, New York, 1935).
35E Einstein, Podolsky, and Rosen, Phys. Rev. **47**, 777 (1935).
35S1 E. Schrödinger, Naturwissenschaften **23**, 807 (1935).
35S2 E. Schrödinger, Naturwissenschaften **23**, 823 (1935).
35S3 E. Schrödinger, Naturwissenschaften **23**, 844 (1935).
35S4 E. Schrödinger, Proc. Cambridge Phil. Soc. **31**, 555 (1935).
35U G. E. Uhlenbeck, J. Math. Phys. **14**, 10 (1935).

36D M. Delbrück and G. Molière, Abhandl. preuss. Akad. Wiss., Physik.-math. Kl. No. 1 (1936).
36E P. S. Epstein in *Commentary on the Scientific Writings of J. Willard Gibbs* (Yale University Press, New Haven, 1936).
36F W. H. Furry, Phys. Rev. **49**, 393 (1936).
36M H. Margenau, Phys. Rev. **49**, 240 (1936).
36S E. Schrödinger, Proc. Cambridge Phil. Soc. **32**, 446 (1936).

37E W. M. Elsasser, Phys. Rev. **52**, 987 (1937).
37H E. Hopf, *Ergodentheorie* (Verlag Julius Springer, Berlin, 1937).
37K E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1937).
37P W. Pauli and M. Fierz, Z. Physik **106**, 572 (1937).
37S T. Sakai, Proc. Phys.-Math. Soc. Japan **19**, 172 (1937).

38K H. A. Kramers, *Grundlagen der Quantentheorie* (Akademische Verlag, Leipzig, 1938).
38P R. Peierls, Phys. Rev. **54**, 918 (1938).
38T R. C. Tolman, *The Principles of Statistical Mechanics* (Oxford University Press, New York, 1938).

39B1 F. J. Belinfante, Physica **6**, 849 (1939).
39B2 F. J. Belinfante, Physica **6**, 870 (1939).
39B3 J. de Boer and A. Michels, Physica **6**, 97 (1939).
39H O. Halpern and F. W. Doermann, Phys. Rev. **55**, 1077 (1939).
39K1 E. C. Kemble, Phys. Rev. **56**, 1013 (1939).
39K2 E. C. Kemble, Phys. Rev. **56**, 1146 (1939).

40B J. de Boer, thesis, Amsterdam (1940).
40H K. Husimi, Proc. Phys.-Math. Soc. Japan **22**, 264 (1940).
40M J. E. and M. G. Mayer, *Statistical Mechanics* (John Wiley and Sons, Inc., New York, 1940).
40P1 W. Pauli, Phys. Rev. **58**, 716 (1940).
40P2 W. Pauli and F. J. Belinfante, Physica **7**, 177 (1940).
40T R. C. Tolman, Phys. Rev. **57**, 1160 (1940).

41M E. W. Montroll and J. E. Mayer, J. Chem. Phys. **9**, 626 (1941).
41O J. C. Oxtoby and S. M. Ulam, Ann. Math. **42**, 874 (1941).
41W A. Wintner, *The Analytical Foundations of Celestial Mechanics* (Princeton University Press, Princeton, 1941).

43C S. Chandrasekhar, Revs. Modern Phys. **15**, 1 (1943).
43P M. Planck, Naturwiss. **31**, 153 (1943).

44H1 J. Hamilton and H. W. Peng, Proc. Roy. Irish Acad. **A49**, 197 (1944).
44H2 W. Heitler, *Quantum Theory of Radiation* (Oxford University Press, New York, 1944).
44K N. Krylov, Nature **153**, 709 (1944).

45C H. B. G. Casimir, Revs. Modern Phys. **17**, 343 (1944).

46B M. Born and H. S. Green, Proc. Roy. Soc. (London) **A188**, 10 (1946).
46C H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1946).
46G1 H. S. Green, Proc. Roy. Soc. (London) **A189**, 103 (1946).
46G2 H. J. Groenewold, Physica **12**, 405 (1946).
46J H. and B. S. Jeffreys, *Method of Mathematical Physics* (Cambridge University Press, Cambridge, 1946).
46K J. G. Kirkwood, J. Chem. Phys. **14**, 180 (1946).

47B1 M. Born and H. S. Green, Proc. Roy. Soc. (London) **A190**, 455 (1947).
47B2 M. Born and H. S. Green, Proc. Roy. Soc. (London) **A191**, 168 (1947).
47D B. Davydov, J. Phys. U.S.S.R. **11**, 33 (1947).

48B1 M. Born, Ann. Physik **3**, 107 (1948).
48B2 M. Born and H. S. Green, Proc. Roy. Soc. (London) **A192**, 166 (1948).
48S E. Schrödinger, *Statistical Thermodynamics* (Cambridge University Press, Cambridge, 1948).
48W N. Wiener, *Cybernetics* (John Wiley and Sons, Inc., New York, 1948).

49B1 M. Born, *Natural Philosophy of Cause and Chance* (Oxford University Press, New York, 1949).
49B2 M. Born, Nuovo cimento **6**, Suppl., 161 (1949).
49B3 M. Born and H. S. Green, *A General Kinetic Theory of Liquids* (Cambridge University Press, Cambridge, 1949).
49K1 A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover Publications, New York, 1949).
49K2 J. G. Kirkwood, Nuovo cimento **6**, Suppl., 233 (1949).
49K3 H. A. Kramers, Nuovo cimento **6**, Suppl., 158 (1949).
49M J. E. Moyal, Proc. Cambridge Phil. Soc. **45**, 99 (1949).
49P W. Pauli, Nuovo cimento **6**, Suppl., 166 (1949).
49R G. S. Rushbrooke, *Introduction to Statistical Mechanics* (Oxford University Press, New York, 1949).
49S A. J. F. Siegert, Phys. Rev. **76**, 1708 (1949).
49S1 C. E. Shannon and W. Weaver, *Mathematical Theory of Communication* (University of Illinois Press, Urbana, Illinois, 1949).
49W1 C. F. von Weizsäcker, *Zum Weltbild der Physik* (Hirzel, Zürich, 1949), p. 80.
49W2 G. Wentzel, *Quantum Theory of Fields* (Interscience Publishers, Inc., New York, 1949).

50B M. S. Bartlett, Nature **165**, 727 (1950).
50H P. R. Halmos, *Measure Theory* (D. Van Nostrand Company, Inc., New York, 1950).

51B1 J. A. Bearden and H. M. Watts, Phys. Rev. **81**, 73 (1951).
51B2 R. Bellmann and T. Harris, Pacific J. Math. **1**, 179 (1951).

51B3 D. Bohm, *Quantum Theory* (Prentice-Hall, Inc., New York, 1951).
51B4 L. Brillouin, J. Appl. Phys. **22**, 334 (1951).
51D J. W. M. DuMond and E. R. Cohen, Phys. Rev. **82**, 555 (1951).
51S F. E. Simon, Z. Naturforsch. **6a**, 397 (1951).

52G1 H. Grad, J. Phys. Chem. **56**, 1039 (1952).
52G2 H. Grad, Comm. Pure Appl. Math. **5**, 455 (1952).
52H T. E. Harris, Trans. Am. Math. Soc. **73**, 471 (1952).
52K M. J. Klein, Phys. Rev. **87**, 111 (1952).
52R L. Rosenfeld, Mimeographed Notes of Lectures given at the Summer School for Theoretical Physics at Les Houches, 1952.
52S1 M. Schönberg, Nuovo cimento **9**, 1139 (1952).
52S2 E. C. G. Stueckelberg, Helv. Phys. Acta **25**, 577 (1952).

53B M. S. Bartlett, Proc. Cambridge Phil. Soc. **49**, 263 (1953).
53D J. L. Doob, *Stochastic Processing* (John Wiley and Sons, Inc., New York, 1953).
53H D. ter Haar and C. D. Green, Proc. Phys. Soc. (London) **A66**, 153 (1953).
53S1 M. Schönberg, Nuovo cimento **10**, 419 (1953).
53S2 M. Schönberg, Nuovo cimento **10**, 697 (1953).
53T J. S. Thomson, Phys. Rev. **91**, 1263 (1953).

54G C. D. Green, thesis, St. Andrews University, Scotland (1954).
54G2 A. Gamba, J. Appl. Phys. **25**, 1549 (1954).
54H1 D. ter Haar, *Elements of Statistical Mechanics* (Rinehart & Company, Inc., New York, 1954).
54H2 D. ter Haar, Am. J. Phys. **22**, 638 (1954).
54H3 Hirschfelder, Curtiss, and Bird, *Molecular Theory of Gases and Liquids* (John Wiley and Sons, Inc., New York, 1954).
54I Inagaki, Wanders, and Piron, Helv. Phys. Acta **27**, 71 (1954).
54K N. G. van Kampen, Physica **20**, 603 (1954).
54L P. T. Landsberg, Phys. Rev. **96**, 1420 (1954).
54L1 J. P. Lloyd and G. E. Pake, Phys. Rev. **94**, 579 (1954).
54M1 D. K. C. MacDonald, J. Appl. Phys. **25**, 619 (1954).
54S L. Sartre, J. phys. radium **15**, p. 3S (April, 1954).
54T T. Takabayasi, Progr. Theoret. Phys. **11**, 341 (1954).

55G C. D. Green and D. ter Haar, Physica **21**, 63 (1955).
55G1 A. Gamba, Nuovo cimento **1**, 358 (1955).
55H1 D. ter Haar, Am. J. Phys. (to be published).
55H2 D. ter Haar and C. D. Green, Proc. Cambridge Phil. Soc. **51**, 141 (1955).
55K R. Kurth (submitted for publication in Revs. Modern Phys.).
55K1 M. J. Klein, Phys. Rev. **97**, 1446 (1955).
55M E. W. Montroll and M. S. Green, Ann. Rev. Phys. Chem. **6** (to be published) (1955).