“Über die mechanischen Grundlagen der Thermodynamik,” Ann. d. Phys. **33** (1910), 225-274.

On the mechanical foundations of thermodynamics

By Paul Hertz

Translated by D. H. Delphenich

Contents

Introduction

There can be no doubt that the kinetic theory of gases is incomplete to the extent that its methods admit no extension to the case of solid bodies. Its most important theorem – namely, that that all gases that exchange heat with the atmosphere through walls will possess the same number of molecules in a unit volume for the same pressure – can be proved only as long as the processes in the walls themselves are drawn into the sphere of consideration. However, the results of the theory of gases also make the success of such efforts quite probable from the outset. On the one hand, the theory of gases has found such a multitude of confirmations that it can probably be regarded as having been proved. On the other hand, laws flow from it – such as the law of entropy and the law of temperature equilibrium – that are conformed for the entire material world without exception, so that we gain from that knowledge a certain confidence that thermal processes can be mechanically-based for solid bodies, as well. Moreover, one will not appeal to this or that theory, since only the most general mechanical principles would be appropriate, corresponding to the great generality of the facts of experiment. One can then see no idle musings in the work of **Gibbs** (1), since one must confer an unconditional degree of truth to it.

It is hardly possible to find a fundamental formula for a basic idea in these questions that was not already obtained in the work of **Gibbs**. Thus, the current presentation will also not bring anything new with it, but only ease the transition to a theory that is known already (2).

In fact, it is by no means easy to arrive at a clear conception of that study. Above all, the concept of the total system leads to great difficulties in understanding. A given system possesses a well-defined state. However, we do not know it, but only how probable it is that the bodies before us are found in each possible state. These state probabilities shall now determine the entropy and temperature of the system. When regarded that way, the **Gibbs** definition will seem downright absurd. How can a quantity that pertains to a body depend, not upon the state that it has, but on the one that it might possibly have, and the probability that it might have that state? A small deviation might then lead back to a definition that is possible from the outset. An ensemble will be *fabricated* mathematically, and a function for its distribution that is defined up to an arbitrary constant will be chosen, but whose constant is chosen in such a way that the mean energy of the fabricated ensemble will be equal to the one that is given in reality. One will then arrive at definitions of the temperature and entropy of real bodies by means of the constants that determine that function (3). These functions will then depend upon the actual state of the body at their basis. One cannot object to anything in this process to begin with, since this way of defining the concept seems artificial, and one will understand it only after **Planck** mixes simplicity and relevance into it (4).

It is probably much simpler for one to consider only states that are actually chosen from systems. One will then arrive at the totality of all systems temporally, and from the phases that they go through, the *temporal ensemble*. It was the contribution of **Einstein** (5) to start with that physically-valuable concept, which he later abandoned, however (6). The concept of a temporal ensemble also appears in **Gibbs** (7), but somewhat in the background in comparison to the independent systems.

Here, we shall attempt to return the temporal ensemble to the foreground. However, the individual phases that belong to such an ensemble shall naturally possess the same energy. Thus, we shall consider only ensembles of phases with equal energy – the so-called *microcanonical ensembles* – while **Gibbs** preferred the canonical ones. *The goal of this paper is then to start with the temporal ensemble and microcanonical one, instead of the independent system ensemble and the canonical one.* Yet, later on, that will lead us

---

(1) **J. W. Gibbs**, *Elementary principles of statistical mechanics*. New York, 1902; German version by **Zermelo**: *Elementare Grundlagen der statistischen Mechanik.* Leipzig, 1905. In order to cite page numbers, the ones in the German edition will be given in parentheses here.

(2) The present paper can perhaps serve to provide an initial overview of statistical mechanics. Whoever is reading with that intent would do well to skip §§ 7 and 8.

(3) **M. Planck**, *Boltzmann-Festschrift,* pp. 117.

(4) *Loc. cit.*, pp. 121.

(5) **A. Einstein**, Ann. d. Phys. **9** (1902), 417; **11** (1903), pp. 170; **14** (1904), 359. In the sequel, these shall always be cited as I, II, III.

(6) When he considered canonical ensembles (cf., below), cf., II, pp. 181, “an assumption that has only a formal meaning.”

(7) **J. W. Gibbs**, pp. 169 (174); pp. 180 (185).

to glimpse a perfectly legitimate concept that relates to the independent system ensembles that can actually not be avoided. That is better left to its proper place in the discussion.

On the contrary, it seems difficult, if not impossible, to arrive at a physical meaning for the canonical ensemble. In its favor, it should be mentioned that it is easy to deal with than the microcanonical one, and since the former ultimately differs from the latter only slightly, the former can be replaced with the latter (1). It is also clear why the canonical ensemble seems simpler. That lies upon the same basis as the fact that a space element is a simpler invariant than a surface differential. However, it will now be shown that the laws of microcanonical ensembles can still be derived in the same way as those of canonical ones. The tool for that is multidimensional geometry, of which extensive use will be made.

Although the transition from the approximate to the rigorous treatment demands no advanced mathematical complications, on the other hand, it will indeed be required by the demands of the theory. A canonical ensemble can indeed yield a microcanonical one in the limit, and it can thus be replaced with one. One will then find the same distribution of phases over the energy surface in the microcanonical ensemble that prevailed in the canonical one. By contrast, the law of the energy gradient (which is then perpendicular to the energy surfaces) will drop out under the passage to the limit. However, that law determines the "modulus" of the canonical ensemble, which then seems to me to correspond to no physical meaning (2).

Now, these considerations also influence the standpoint that we assume in regard to the theory of independent ensembles. In connection with that, **Gibbs** found an explanation for the increase in entropy. Now, I, in fact, hold that the objections that have been raised against that to date (3) are unjustified; however, it seems conceivable to me that in order to verify that the **Gibbs** expression is actually entropy, the modulus must amount to temperature, although I can glimpse no physical concept in that. Therefore, I maintain that the theory of microcanonical temporal ensembles, which likewise goes back to **Gibbs**, has been completed (4).

In any event, it seems desirable to give a presentation that, first of all, begins with a presentation of the temporal ensemble and, above all, employs the microcanonical ensemble exclusively; that was the first viewpoint for the present paper. Secondly, the means shall be given here to single out the theories of **Boltzmann**, **Gibbs**, and **Einstein** collectively in the investigations that follow, and to show that they have basically the same content. That verification will be carried out as soon as one introduces some assumptions about the behavior of certain functions that are characteristic of the bodies, which are assumptions that do not refer to the its analytical structure, but only to the behavior of its maximum. It can be easily shown that these requirements will be fulfilled by ideal gases. However, on the basis of such assumptions, we will now infer a rigorous proof of the validity of the law of heat equilibrium, in addition, that, as the **Gibbs** book shows, cannot be carried out without such assumptions. Finally, some simpler

---

(1) **H. A. Lorentz**, *Gesammelte Abhandlungen*, 1, pp. 286, *et seq.*; Atti del 4 Congr. dei Mat. Roma 1, (1908), pp. 152.

(2) As long as one is, in fact, dealing with a complete system; cf., the **Einstein** papers.

(3) Cf., pp. 7, rem. 8.

(4) Still, it does not seem to me that the aforementioned gap will be filled in.

derivations of some laws that are known already will be given by appealing to multidimensional geometry.

Part I

Thermal processes

(The law of heat equilibrium)

§ 1. Mechanical foundations (1)

We assume that many bodies can be considered in regard to their thermal behavior, as well as purely mechanical systems, and restrict our examination to them. The kinetic theory of gases already shows that one and the same observable state can correspond to many molecular states that deviate by unobservable differences. We would like to assume that they will be traversed in sequence. Any observable state will then belong to a sequence of true states – viz., a state sequence. The observable state and the state sequence can be changed by external influences – e.g., a supply of heat. However, the mechanical laws that the individual states of the sequence follow will still be the same. In that case, we say that the mechanism remains the same. It can first be converted in the actual sense of mechanical operations – e.g., changing its volume, when one is dealing with a gas (2) – and we then say that the system now possesses another mechanism, or also, when the word "mechanism" applies to not only the properties of the system, but also to the system itself, another mechanism will now exist. There are then three kinds of changes to consider:

1. Changes in the true mechanical state, which are changes that collectively define a state sequence.
2. Changes of the state sequence.
3. Changes of the mechanism.

The true mechanical state of the systems will be determined by its generalized coordinates $q_1, q_2, \ldots, q_n$, and the generalized velocities $\dot{q}_1, \dot{q}_2, \ldots, \dot{q}_n$. The energy ε is a function of them. We refer to the differential quotients:

$$p_n = \frac{\partial \varepsilon}{\partial \dot{q}_n} \tag{1}$$

---

(1) **J. W. Gibbs**, § 1. – This paragraph brings nothing new with it, since it includes only a review of the longest-known laws. However, it might be of use as an exposition for what follows.

(2) However, as long as one includes the piston and its potential energy in the system, its displacement will not imply any change in the mechanism. In that case, no new coupling of the parts will come about as a result, but only a new configuration of the parts themselves. Here, as in all cases, the concept of equilibrium of the mechanism is a relative one (cf., § 11).

as the impulse, and from now on, we would like to determine the state of the system by the $2n$ state variables $q_1, q_2, \ldots, q_n, p_1, p_2, \ldots, p_n$. One refers to the totality of the quantities $q_1, \ldots.\ p_n$ as a *phase*. Energy is also given by the phase; i.e., ε is a function of the $q_1, \ldots.\ p_n$.

We can distinguish three kinds of changes in regard to the behavior of the energy ε:

1. q and p change, while ε remains constant.
2. ε changes, while ε remains the same function of q and p.
3. ε becomes another function of the q and p.

Now, these three cases correspond to the three that were just listed:

1. *Left to itself*, the system will continually assume other values q and p; i.e., it will run through other phases continually. Nevertheless, from the law of energy, the quantity ε will be conserved. *Each of these phases will be characterized by the quantities q and p.* The phases of equal energy that are traversed in that way shall be called a *phase manifold* or a *phase surface*.

2. One can devise a new phase manifold, another energy, and thus, a new observable state by means of *external* influences – namely, the supply of heat – without basing it upon the mechanical couplings ([1]). *Any phase manifold or any observable state is determined by* ε. ε thus remains constant under these processes, but $\varepsilon\,(q, p)$ is still the same function.

3. One can bring about a new mechanism by means of actual mechanical operations ([2]). Any mechanism is characterized by the function $\varepsilon\,(q, p)$ ([3]).

Since all changes can be reduced to ones of the first kind ([4]), we must first have laws for them. Now, it is known that one has ([5]):

$$\frac{\partial q_n}{\partial t} = \frac{\partial \varepsilon}{\partial p_n}, \tag{2}$$

---

[1] It is clear that different energies will also correspond to different observable states. Conversely, as long as the mechanism is the same, a change in the observable state will be possible only in conjunction with a change in energy. The fact that the phase surface is also given by the energy is the content of the hypothesis that was introduced in the second paragraph.

[2] At this point, one treats the second part. It must be once more emphasized that the differences that are being considered are only of the relative kind. As long as the externally-influenced systems are drawn into the sphere of consideration, there will be only changes of the first kind.

[3] The fact that the mechanical laws, and therefore, the mechanism, are known when one is given the function $\varepsilon\,(q, p)$ follows from equations (2) and (3).

[4] § 10. The changes of the second kind must generally be reduced to ones of the third kind to begin with. (§ 9).

[5] E. g., **J. W. Gibbs**, § 1.

(3) $$\frac{\partial p_n}{\partial t} = -\frac{\partial \varepsilon}{\partial q_n},$$

from which, it will follow that:

(4) $$\sum \frac{\partial}{\partial q_n}\frac{dq_n}{dt} + \frac{\partial}{\partial p_n}\frac{dp_n}{dt}.$$

These equations show that the changes that a system suffers are independent of the past history.

We will now address the first two kinds of changes, and accordingly consider a *fixed mechanism*. However, we will also deal with thermal changes in Part One, and thus changes of the second kind (1). It is therefore not preferable to restrict the consideration to the phases that possess one and the same energy, but rather one must draw one's attention to all possible phases of differing energies and their changes. These considerations define the content of *statistical mechanics*. However, in order to conveniently illustrate the laws of statistical mechanics, we appeal to a tool and find it in the *geometry of n-dimensional spaces.*

We imagine a $2n = m$-dimensional "phase space." We represent each state of the system by that point whose $m = 2n$ coordinates and impulses have the values $q_1, q_2, \ldots, q_n, p_1, p_2, \ldots, p_n$, or, as we would also like to write $x_1, x_2, \ldots, x_n$. A path curve then goes through each point of phase space. The path curve that goes through a point is then independent of the past history. That is, if changes of the first and second kind take the system back to the same phase then the sequence of following phases will also be the same.

We imagine an m-dimensional region g in phase space. We choose a time t and assign each point $x_1, x_2, \ldots, x_n$ of g with the point $x_1', x_2', \ldots, x_n'$, which represents a system at time t that possesses the phase $x_1, x_2, \ldots, x_n$ at time zero. Thus, the points x_1', $x_2', \ldots, x_n'$ that arise in that way will fill up an m-dimensional domain g'. We ask how the two domains relate to each other.

The answer is given by *Liouville's theorem*. From equation (4), which can also be written as:

(5) $$\sum_{\kappa=1}^{m} \frac{\partial \dot{x}_\kappa}{\partial x_\kappa} = 0,$$

it follows from a hydrodynamical argument that both domains are equal, so:

(6) $$\int_g dx_1\, dx_2 \ldots dx_m = \int_{g'} dx_1'\, dx_2' \ldots dx_m'.$$

---

(1) In order to bring this about, a change of the third kind must be performed, if only temporarily.

§ 2. The microcanonical temporal ensemble (1).

We restrict our consideration to systems whose coordinates cannot increase beyond all bounds. From a theorem that was proved by **Poincaré** (2) and **Zermelo** (3), after a finite time, such a system (most probably) must come very close to a previously-attained phase. It will then always assume the same phases, and indeed each of the ones that come under consideration, with a completely-defined frequency. Such an ensemble of phases that can run through a mechanism cyclically, each with a well-defined frequency, is called a *temporal ensemble* that belongs to a mechanism. It will be characterized by giving the frequency of its phases. Many temporal ensembles can belong to one mechanism.

One might wish to consider other ensembles. One then imagines a set of mutually-independent systems that are combined by some well-defined principle, all of which have the same mechanism, and follow its ensuing destiny. We can refer to such sets as *virtual ensembles*. Now, it is clear, with no further discussion, that any temporal ensemble can be assigned to a virtual one. We need only to imagine a set of systems, and as many of each phase as would correspond to the frequency of the temporal ensemble (4). These associated virtual ensembles possess a very remarkable property, as is immediately apparent. The number of systems that belong to a given phase does not change in time. If a virtual ensemble possesses that property then we would like to say that it is found in *statistical equilibrium*, and we can then say that: *The virtual ensemble that corresponds to a temporal ensemble is in statistical equilibrium.*

One would like to make the virtual ensemble the starting point of the kinetic considerations (5). Since most formulas refer to the case of statistical equilibrium, one requires a theorem whose effect is to say that every arbitrary (6), virtual ensemble will go to a statistical one. Attempts to prove that theorem (7) have led to many contradictions (8), which seem unjustified to me. By contrast, if one starts with the concept of a temporal ensemble then that theorem will be unnecessary. The provable theorem of the recurrence of phases and the determinacy of their frequencies accomplishes the same thing for the theory that is constructed from the concept of temporal ensemble as the theorem of statistical equilibrium should accomplish in the theory of virtual ensembles. That will imply a substantial simplification when one places the temporal ensemble at the center of attention, as in the **Einstein** process.

---

(1) **A. Einstein**, I, § 2; II, § 2. Cf., also **J. W. Gibbs**, chap. 10, pp. 169 (174), pp. 180 (185); **L. Boltzmann**, *Gastheorie*, **2**, pp. 98.

(2) **H. Poincaré**, Acta mathematica, **13** (1890), 1-270.

(3) **E. Zermelo**, Wied. Ann. **57** (1896).

(4) If an observable state belongs to one and only one temporal ensemble (cf., below) then one will obtain the virtual ensemble very simply in the following way: One takes a large number of bodies of equal observables states. From the laws of probability, each phase will then be attained as often as would correspond to its frequency in the temporal ensemble.

(5) **J. W. Gibbs**, § 1.

(6) A virtual ensemble that is not in statistical equilibrium does not correspond to any temporal ensemble.

(7) **J. W. Gibbs**, chap. 12; **H. A. Lorentz**, *Gesammelte Abhandlungen*, pp. 290.

(8) **E. Zermelo**, Jahresbericht der deutsch. Mathematiker-Vereinigung **15** (1906), 238; P. and T. Ehrenfest, Wiener Berichte **115** (1906).

That is still not the decisive factor. It is not obvious at first what sort of relationship a virtual ensemble can have to the individual bodies, but that is of no concern to us, so we assume that it is the image of its temporal ensemble. The fact that the observable state depends upon the sequence of phases that were traversed is understandable with no further explanation. However, the fact that it can be determined by the frequency of phases that are not in the system considered, but can be appended in some way, seems absurd at first. Thus, it would ease our understanding of things if we were to begin with the temporal ensemble. Later on, we will see that the virtual ensemble is surely quite indispensible, and that working with it possesses an actual physical sense.

That is again connected with the fact that all phases of our ensemble possess the same energy. If we denote the value of energy by ε^* then all points on the path will satisfy the equation:

$$\varepsilon(q_1, q_2, \dots, q_n, p_1, p_2, \dots, p_n) = \varepsilon^*, \tag{7}$$

i.e., it is found on a $2n$-1-dimensional surface in $2n$-dimensional space. That surface is briefly called the surface $\varepsilon = \varepsilon^*$, or even more briefly, the ε^**–surface*. From the great complexity of the system, it is, with no further discussion, plausible to assume that the curve almost completely describes that path in a multiple interlacing of the entire surface patch, and exhibits a behavior that is similar to that of the **Peano** curve (^{1}). A subset of the surface $\varepsilon = \varepsilon^*$ will then be filled by the temporal ensemble. Now, the surfaces of mathematical analysis are not very tractable as spaces, and the expression for the differential of volume is far simpler than that of the surface differential. Therefore, an ensemble that is distributed over a space will be much simpler than one that is distributed over a surface. **Gibbs** preferred such canonical spatial ensemble in his presentation. However, no physical sense can be ascribed to them. It has been said that they are only an analytical gimmick (^{2}). In fact, the canonical ensembles differ only slightly from the ones that are distributed over surfaces, and since they are easier to deal with, for the reasons that were given, one can appeal to them in order to derive these laws. We will see that we will not require such a method, and that we are capable of deriving laws that are true for surface ensembles from geometric considerations directly.

It would be convenient to introduce some simplified hypotheses. Our temporal ensemble (we must assume) fills up a subset of the ε^*-surface. With **Einstein** (^{3}), we now make the assumption that it covers the entire surface $\varepsilon = \varepsilon^*$. That is a hypothesis that we will actually first require in the second part of our investigations, but which we might introduce here already in order to work with specific constructions. The system will then run continually through all phases of the surface $\varepsilon = \varepsilon^*$. Therefore, the temporal ensemble will be determined completely by ε^*. Furthermore, ε^* is given naturally by the observable state, and conversely, when all external mechanical coordinates are fixed, the observable state will be determined by ε^*. The observable state, energy, and phase surface then correspond to each other.

(^{1}) The **Peano** curve goes through each point of the surface exactly; the one that is considered here comes arbitrarily close to each point.

(^{2}) **H. A. Lorentz**, *Ges. Abh.*, pp. 286 and 287, Atti del 4. Congr. dei Mat. Roma **1** (1908), pp. 152.

(^{3}) **A. Einstein**, II, § 1.

In order to have a complete knowledge of the ensemble, we must know the frequency that corresponds to each point of the surface $\varepsilon = \varepsilon^*$. We then next define the concept of *frequency* or *probability* precisely. Let *do* be an element of the surface $\varepsilon = \varepsilon^*$ that surrounds the point under scrutiny. We consider the system over a very large time interval T. All time intervals inside of T during which the system assumes a phase that is found in *do* might collectively yield the value τ. Therefore:

$$w \cdot do = \frac{\tau}{T} \tag{8}$$

should give the probability for a phase to exist in the surface element in question ([1]). Another definition that is consistent with that reads: Consider the system at $\mathfrak{N}$ different points in time, so it will be found in *do* at $\mathfrak{n}$ of them, and we will further have:

$$w \cdot do = \frac{\mathfrak{n}}{\mathfrak{N}}. \tag{8′}$$

Finally, one can consider Z systems that all have the same observable state ([2]). If one finds z of them in a phase that is inside of *do* then one will also have:

$$w \cdot do = \frac{z}{Z}. \tag{8″}$$

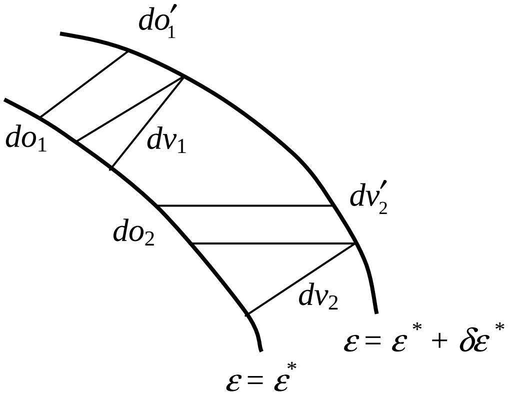

Figure 1.

It is easy to ascertain w. We consider systems in all possible phases and let do_1 and do_2 (Fig. 1) be two elements of the surface $\varepsilon = \varepsilon^*$ that go to each other – i.e., after a certain time, one might find those and only those systems in do_2 that were initially in do_1. Our system will run through do_2 as often as it does through do_1; i.e., one will have:

([1]) **A. Einstein**, II, pp. 172.

([2]) w is a function of q and p that does not depend upon only the mechanism, but also on the observable state of the surface $\varepsilon = \varepsilon^*$.

(9)
$$w_1 \cdot do_1 = w_2 \cdot do_2 .$$

We now imagine the surface $\varepsilon = \varepsilon^* + \delta\varepsilon^*$ that neighbors the ε^*-surface, where $\delta\varepsilon^*$ is a small quantity. Let do_1 be a surface element be a surface element of $\varepsilon = \varepsilon^* + \delta\varepsilon^*$ in the vicinity of do_1, and suppose that do_1' goes to do_2' during the time interval in which do_1 goes to do_2 . The skew cylinder do_1 do_2 then goes to $do_1' do_2'$, and therefore, if we understand dv_1 (dv_2, resp.) to mean the perpendicular distance between the two surfaces at the location of do_1 (do_2 , resp.) then, from **Liouville's** theorem (6):

$$do_1\, do_2 = dv_1\, dv_2,$$

or

(10)
$$do_1 \left(\frac{dv}{d\varepsilon}\right)_1 = do_2 \left(\frac{dv}{d\varepsilon}\right)_2 .$$

In combination with (6), that equation will yield the proportion:

$$w_1 : \left(\frac{dv}{d\varepsilon}\right)_1 = w_2 : \left(\frac{dv}{d\varepsilon}\right)_2 .$$

One then has, in general:

(11)
$$w = C\,\frac{dv}{d\varepsilon},$$

in which C is a constant that depends upon the mechanism and ε^*.

The probability for the phase of the system to belong to a region g is then:

(12)
$$W_g = C \cdot \int_g do \frac{dv}{d\varepsilon} .$$

If we apply that formula to the entire ε^*-surface then it will follow that (^{1}):

(13)
$$1 = C \cdot \int_{\varepsilon=\varepsilon^*} do \frac{dv}{d\varepsilon} ,$$

or, when written more briefly:

(13′)
$$1 = C \cdot \int_{\varepsilon^*} do \frac{dv}{d\varepsilon} .$$

If one, with **Einstein** (^{2}), finally sets:

(14)
$$\omega(\varepsilon^*) = \int_{\varepsilon=\varepsilon^*} do \frac{dv}{d\varepsilon}$$

then, from (12) and (13), one will have:

(^{1}) Integration over $2n$-1-dimensional manifolds shall be suggested by an integral sign, while integration over $2n$-dimensional ones shall be suggested by two integrals.

(^{2}) **A. Einstein**, III, pp. 365. Cf., also I, pp. 176; II, pp. 422.

$$(15) \qquad w = \frac{1}{\omega}\frac{dv}{d\varepsilon},$$

and the probability that a phase is found in g will amount to (¹):

$$(16) \qquad W_g = \frac{1}{\omega}\int_g \frac{dv}{d\varepsilon}\,do\,.$$

We will use yet another expression for ω. If one, with **Gibbs** (²), sets the total volume of all phases whose energy is less than ε^* equal to $V(\varepsilon^*)$, so:

$$(17) \qquad \int_{\varepsilon<\varepsilon^*}\int dx_1\cdots dx_m = V(\varepsilon^*),$$

for which, one can also write:

$$(17') \qquad \int^{\varepsilon^*}\int dx_1\cdots dx_m = V(\varepsilon^*),$$

then, from (14):

$$(18) \qquad \omega = \frac{dV}{d\varepsilon^*}\,.$$

An ensemble of states of equal energy whose distribution over all phases of that energy is given by the formulas (15) and (14) [(15), (17), and (18), resp.], is called a *microcanonical* ensemble. As we already mentioned, we will then concern ourselves with only the microcanonical ensembles (³).

If u is an arbitrary function of phase then a given system will take on other values of u during its continual travels about phase space. One can ask what the mean value $\bar{u}$ of u would be, which is, at the same time, a mean value in the microcanonical ensemble, and from (16), is given by (⁴):

$$(19) \qquad \bar{u} = \frac{1}{\omega}\int_{\varepsilon^*} u\cdot\frac{dv}{d\varepsilon}\,do\,.$$

A different form of the equation is often more convenient. One imagines the surface $\varepsilon = \varepsilon^* + \delta\varepsilon^*$, which is infinitely close to the ε^*-surface. Therefore:

$$do\,\frac{dv}{d\varepsilon}\,\delta\varepsilon = do\;\delta v$$

(¹) **A. Einstein**, *loc. cit.*

(²) **J. W. Gibbs**, formula 265.

(³) In the derivation of (15), we cannot restrict ourselves to the surface, but must appeal to space. To that extent, we are dealing with a spatial ensemble. However, once that derivation has been made, we will consistently remain on the surface. In that process, which corresponds to the method that was developed by **Gibbs** [pp. 116, (117)], no assumptions are made about the dependency of the system density upon the energy, and that dependency will not be regarded as characteristic of the ensemble. One can therefore not speak of a "canonical ensemble" in any proper sense of the phrase.

(⁴) **J. W. Gibbs**, formula 374.

is the small cylinder that lies between two surfaces, and whose basis is *do*. One then also has:

$$a = \frac{1}{\omega\cdot\delta\varepsilon^*}\int_{\varepsilon^*<\varepsilon<\varepsilon^*+\delta\varepsilon}\int u\,dx_1\cdots dx_m\,, \tag{20}$$

or more briefly:

$$a = \frac{1}{\omega\cdot\delta\varepsilon^*}\int_{\varepsilon^*}^{\varepsilon^*+\delta\varepsilon}\int u\cdot dx_1\cdots dx_m\,, \tag{20′}$$

in which the integration is taken between two surfaces (1).

§ 3. Mean value of a scalar product.

Kinetic energy is of especial importance for thermodynamics, and we, with **Gibbs**, will denote it by ε_p . We seek to determine its mean value. Since it is a homogeneous quadratic function of *p*, and since the potential energy is independent of *p*, one will have:

$$\varepsilon_p = \frac{1}{2}\sum_{\kappa=1}^{n} p_\kappa \frac{\partial\varepsilon}{\partial p_\kappa}\,. \tag{21}$$

We thus have to concern ourselves with the mean value of:

$$\sum_{\kappa=1}^{n} p_\kappa \frac{\partial\varepsilon}{\partial p_\kappa}\,.$$

We now generalize the statement of the problem to: *Divide the m state variables* $x_1, x_2, \ldots, x_m$ *into two groups:* $y_1, y_2, \ldots, y_s$ *and* $z_1, z_2, \ldots, z_\sigma$, *where* $s + \sigma = m$. *Find the mean value:*

$$\overline{\sum_{\kappa=1}^{s} y_\kappa \frac{\partial\varepsilon}{\partial y_\kappa}}\,.$$

We make some special preparations for that problem. First, we determine:

$$H = \overline{\sum_{\kappa=1}^{m} x_\kappa \frac{\partial\varepsilon}{\partial x_\kappa}}\,, \tag{22}$$

which is an expression that includes all of the state variables, and for which, from (19), the relations:

$$H = \frac{1}{\omega}h, \tag{23}$$

---

(1) **A. Einstein**, I, pp. 420; II, pp. 174. *u* is defined only on the surface, but it can be chosen between the surfaces arbitrarily as a continuous interpolation.

(24) $$h = \int \frac{do}{d\varepsilon / dv} \sum_{\kappa=1}^{m} x_\kappa \frac{\partial \varepsilon}{\partial x_\kappa}$$

will be true.

Here, we can apply the vector calculus only symbolically. We understand $\nabla\varphi$ to mean a vector in m-dimensional space with the components $\partial\varphi / \partial x_\kappa$, and understand $\mathfrak{r}$ to mean a vector with the components x_κ (which is then the ray that goes from the origin to the field point); we suggest the absolute value by vertical lines and scalar multiplication by parentheses. (24) then takes on the form:

(25) $$h = \int_{\varepsilon=\varepsilon^*} \frac{do}{|\nabla\varepsilon|} (\mathfrak{r}, \nabla\varepsilon).$$

However, $\nabla\varepsilon / |\nabla\varepsilon|$ is a vector of magnitude 1 that has a direction that is perpendicular to the surface element *do*. If one denotes it by $\mathfrak{e}_\nu$ then one will have:

$$h = \int_{\varepsilon^*} do\,(\mathfrak{r}, \mathfrak{e}_\nu),$$

or, more briefly:

(26) $$h = \int_{\varepsilon^*} do\,\mathfrak{r}_\nu .$$

This integral can now be converted into a "volume integral" that is extended over space that the ε^*–surface encloses using **Gauss's** theorem. If $d\tau$ denotes the "volume element" then it will follow from (26) that:

$$h = \int_{\varepsilon<\varepsilon^*} \int (\operatorname{div} \mathfrak{r})\, d\tau .$$

However, since:

$$\operatorname{div} \mathfrak{r} = m,$$

one will obtain, from (17):

(27) $$h = m \cdot V.$$

From (22), (23), and (20), one will then also have:

(28) $$\frac{1}{\delta\varepsilon^*} \int_{\varepsilon^*}^{\varepsilon^*+\delta\varepsilon^*} \int \left(\sum_{\kappa=1}^{m} x_\kappa \frac{\partial \varepsilon}{\partial x_\kappa} \right) dx_1 \ldots dx_m = m\,V$$

and

(29) $$H = \frac{h}{\omega} = \overline{\sum_{\kappa=1}^{m} x_\kappa \frac{\partial \varepsilon}{\partial x_\kappa}} = m\,\frac{V}{\omega}.$$

We now calculate:

(30) $$\hat{H} = \overline{\sum_{\kappa=1}^{s} y_\kappa \frac{\partial \varepsilon}{\partial y_\kappa}}.$$

If we again imagine the two neighboring surfaces $\varepsilon = \varepsilon^*$ and $\varepsilon = \varepsilon^* + \delta\varepsilon^*$ then, from (20), we will have:

$$(31) \qquad \begin{cases} \hat{H} = \dfrac{\hat{h}}{\omega}, \\ \hat{h} = \dfrac{1}{\delta\varepsilon^*} \displaystyle\int_{\varepsilon^* < \varepsilon < \varepsilon^* + \delta\varepsilon^*} \int \left(\sum_{\kappa=1}^{s} y_\kappa \frac{\partial \varepsilon}{\partial y_\kappa} \right) dy_1 \cdots dy_s \, dz_1 \cdots dz_\sigma . \end{cases}$$

We would now like to consider a fixed variable combination of the z that allows them to vary over a well-defined range $dz_1, dz_2, \ldots, dz_\sigma$. Let it be chosen in such a way that ε varies by a quantity that is small compared to $\delta\varepsilon^*$ under all possible fluctuations. We denote the magnitude that we obtain for the integral (31) with these restrictions by $d\hat{h}_z$; we then have:

$$(32) \qquad d\hat{h}_z = dz_1 \ldots dz_\sigma \, \frac{1}{\delta\varepsilon^*} \int_{\varepsilon^*}^{\varepsilon^* + \delta\varepsilon^*} \int \left(\sum_{\kappa=1}^{s} y_\kappa \frac{\partial \varepsilon}{\partial y_\kappa} \right) dz_1 \ldots dy_s .$$

Hence, all y are admissible that yield an ε between ε^* and $\varepsilon^* + \delta\varepsilon^*$ (always with our special choice of the z). However, from (28), one has:

$$\frac{1}{\delta\varepsilon^*} \int_{\varepsilon^*}^{\varepsilon^* + \delta\varepsilon^*} \int \left(\sum_{\kappa=1}^{s} y_\kappa \frac{\partial \varepsilon}{\partial y_\kappa} \right) dy_1 \ldots dy_s = s \int_{\varepsilon < \varepsilon^*} \int dy_1 \ldots dy_s ,$$

so

$$d\hat{h}_z = s \, dz_1 \ldots dz_\sigma \int_{\varepsilon < \varepsilon^*} \int dy_1 \ldots dy_s .$$

When we consider all possible combination of the z, we will get:

$$\hat{h} = s \int dz_1 \ldots dz_\sigma dz_1 \ldots dy_s .$$

so

$$(33) \qquad \hat{H} = \frac{sV}{\omega},$$

or, more thoroughly (1):

$$(34) \qquad \overline{\sum_{\kappa=1}^{s} y_\kappa \frac{\partial \varepsilon}{\partial y_\kappa}} = \frac{sV}{\omega} .$$

---

(1) Here, if we wished to connect with the **Einstein** method of proof then we would have to go over to space. One can avoid that, but one will then need theorems of the geometry of n-dimensional spaces that are immediately obvious, due to their analogy with ordinary geometry, but would still require a special proof.

§ 4. Kinetic energy and temperature.

The goal of our developments was to give an expression for the mean value of the kinetic energy. From (21), it is:

(35) $$\overline{\varepsilon_p} = \tfrac{1}{2}\overline{\sum_{\kappa=1}^{n} p_\kappa \frac{\partial \varepsilon}{\partial p_\kappa}};$$

hence, from (34) (1):

(36) $$\overline{\varepsilon_p} = \frac{n}{2}\frac{V}{\omega}$$

or

(37) $$\overline{\varepsilon_p} = \frac{n}{2}\mathfrak{t},$$

where one sets:

(38) $$\mathfrak{t} = \frac{V}{\omega}.$$

From (18), one can also write:

(39) $$\frac{1}{\mathfrak{t}} = \frac{d \ln V}{d\varepsilon^*}.$$

The component of the kinetic energy that is associated with the degree of freedom then amounts to:

$$\frac{\mathfrak{t}}{2} = \frac{1}{2}\frac{V}{\omega}.$$

If we further set:

(40) $$W = -\sum_{\kappa=1}^{n} q_\kappa \frac{\partial \varepsilon}{\partial q_\kappa}$$

then, from (34) and (35), one has:

(41) $$\overline{W} + 2\overline{\varepsilon_p} = 0.$$

The expression W can be referred to as the *virial*. (40) will then agree with the quantity in the theory of gases that is referred to by that word in the case for which all points of the system are uncoupled, and one appeals to **Cartesian** coordinates, and (41) will go to the equation that is known there (2). Likewise, when q_κ is an arbitrary coordinate, from (34) and (38), one will have:

(42) $$\overline{q_\kappa \frac{\partial \varepsilon}{\partial q_\kappa}} = \frac{1}{2}.$$

A mixture of two gases can be considered to be a single system. If q_h and q_k are the coordinates of a molecule in the first and second gases, respectively, then, from (42), one will have:

---

(1) **J. W. Gibbs**, formula (377).

(2) **L. Boltzmann**, *Gastheorie*, **2**, pp. 142.

(43) $$\overline{q_h \frac{\partial \varepsilon}{\partial q_h}} = \overline{q_k \frac{\partial \varepsilon}{\partial q_k}}.$$

Therefore, since one can choose the q to be **Cartesian** coordinates, the mean *vis vivas* of the molecules in each kind of gas will be equal ([1]).

We would next like to address the relationship:

(37) $$\overline{\varepsilon_p} = \frac{n}{2} t.$$

That equation says that the mean value of the kinetic energy is proportional to, firstly, the degrees of freedom, and secondly, a certain function t of the total energy. Therefore, the mean value $\overline{\varepsilon_p}$ can then refer to a temporal or a virtual ensemble; i.e., one can track one and the same system to very many time points (or over a long time) or observe an ensemble of systems of equally-observable states at a well-defined time point. Now, (37) corresponds completely to the equation between *vis viva* and temperature that is known from the kinetic theory of gases. We then expect to once more find the temperature that we know from experience in the function t, but we must first verify the validity of that suspicion.

In order to do that, it will be necessary to clarify the concept of thermal contact. From **Gibbs** ([2]) and **Einstein** ([3]), two systems Σ_1 and Σ_2 are in contact when they are combined into a system Σ of following mechanisms: The coordinates of Σ are the coordinates of both Σ_1 and Σ_2. Up to higher-order infinitesimals, the energy that belongs to a phase of Σ is equal to the sum of the energies ε_1 and ε_2 that would belong to the same coordinates in Σ_1 and Σ_2 if they were free (and one can also ascribe them to a bound state). One then has:

(44) $$\varepsilon = \varepsilon_1 + \varepsilon_2 .$$

In the uncoupled state, the first (second, resp.) system, with the impulses $p_1, p_2, \ldots, p_s$ ($p_{s+1}, p_{s+2}, \ldots, p_{s+\sigma}$, resp.), would possess the kinetic energy:

$$\tfrac{1}{2}\sum_{\kappa=1}^{s} \frac{\partial \varepsilon_1}{\partial p_\kappa} p_\kappa \qquad \left(\tfrac{1}{2}\sum_{\kappa=1}^{s} \frac{\partial \varepsilon_2}{\partial p_{\kappa+s}} p_{\kappa+s}, \text{resp.}\right).$$

We also ascribe this kinetic energy to a subsystem. One then has:

$$\varepsilon_{p1} = \tfrac{1}{2}\sum_{\kappa=1}^{s} p_\kappa \frac{\partial \varepsilon_1}{\partial p_\kappa},$$

but, from (44), one has:

---

[1] **L. Boltzmann**, *Gastheorie*, **1**, pp. 51; **2**, pp. 100, 124.

[2] **J. W. Gibbs**, pp. 121 (122).

[3] **A. Einstein**, I, pp. 420; II, pp. 174.

(45)
$$\begin{cases} \varepsilon_{p1} = \tfrac{1}{2}\sum\limits_{\kappa=1}^{s} p_\kappa \dfrac{\partial \varepsilon}{\partial p_\kappa}, \\ \varepsilon_{p2} = \tfrac{1}{2}\sum\limits_{\kappa=1}^{s} p_{\kappa+s} \dfrac{\partial \varepsilon}{\partial p_{\kappa+s}}. \end{cases}$$

It then follows, from (34) and (38):

(46)
$$\varepsilon_{p1} + \varepsilon_{p2} = \varepsilon_p ,$$

(47)
$$\begin{cases} \overline{\varepsilon_{p1}} = \dfrac{s}{2}\mathfrak{t}, \\ \overline{\varepsilon_{p2}} = \dfrac{\sigma}{2}\mathfrak{t}. \end{cases}$$

The kinetic energy per degree of freedom also amounts to $\mathfrak{t} / 2$ *for any subsystem.*

We can then prove that when two systems with the same $\mathfrak{t}$ are combined, a system with the same $\mathfrak{t}$ will result, so we can likewise show that both subsystems will possess the same mean *vis viva* after the combination that they did before (because the number of degrees of freedom remains the same), and it would become likely that the observable state would also not change by such a combination.

However, something that is very important about the observable state follows immediately from the theorems that were just conjectured. Namely, we can think of an ideal gas as being composed of bodies that have a very small size in comparison to the usual ones and are in a state of constant pressure. If it is in contact with an ordinary body then, due to the smallness of the gas, the combined system will include quantities that were previously present in the larger body. On the other hand, $\mathfrak{t}$ will make itself noticeable in a gas by way of the observable state, as will be shown later (1). The gas will then also serve to make the function $\mathfrak{t}$ that belongs to the larger body observable, and will thus be called a *thermometer* (2). Thus, if the theorem that was just assumed were true then it would give a measurable and immediately-observable property of two bodies that would not be changed when they are brought into contact if they are both the same in regard to it.

We then come to the derivation of the theorem (Combination Theorem):

If two bodies have the same $\mathfrak{t}$ *before contact then they will have the same* $\mathfrak{t}$ *after contact.*

We now put that next to the following, no-less-important, Separation Theorem:

---

(1) § 10.

(2) **A. Einstein**, I, pp. 425; II, pp. 176.

If a system that consists of bodies in contact is separated into its components then both of them will keep the same $\mathfrak{t}$ *that the combined system possessed previously.*

Now, if two bodies with equal $\mathfrak{t}$ are brought into contact temporarily and then separated again then they will also have the same $\mathfrak{t}$ after separation. However, since ε is supposed to be a single-valued function of t, the energy in both bodies will be the same as before, and likewise, from the assumptions of § 1, that of the observable state. Temporary contact between two bodies with the same t will have no effect upon the observable state as a result. From the proofs of the Combination and Separation Theorems, we can thus refer to $\mathfrak{t}$ as the *temperature.*

Einstein (1) sought to make both theorems (2) plausible. However, what he put forth cannot actually serve as a proof of plausibility, but implies a repudiation of the principles that he himself used as a basis. He sought to prove the separation theorem, although he regarded it as obvious that the gradual decomposition of a body would provoke only minor changes in the state of a body that was in contact with part of it. Moreover, any subdivision would imply the creation of a new *mechanism*, and it is not clear what sort of connection should exist between the states before and after the separation in a system that is coupled with part of the body that was cut away (3). Neither is it justified to regard the validity of the Combination theorem as a logical consequence of the Separation theorem just because the process that applies to both of them is invertible (4). The following consideration will show that the Separation Theorem and the Combination Theorem have nothing whatsoever to do with each other to begin with.

We first ask how the combination problem is really to be regarded then. We then start with the intuition that $\mathfrak{t}$ is a single-valued function $\mathfrak{t}(\varepsilon)$ of ε in any body, and

---

(1) **A. Einstein**, I, pp. 426; II, pp. 177.

(2) In other words, the theorems in his theory that correspond to the two above.

(3) For **Einstein**, a thermometer possesses a well-defined state probability that would be given, e.g., by the temporal ensemble of phases that is runs through. However, that state probability would not be given by the phase at an exact time point, but would depend upon the behavior of the body during a time interval. Therefore, even if the phase is not changed by the separation, and the later *phases* follow continuously from the previous ones, the *phase ensemble* can still become something different suddenly (cf., Fig. 2; one might imagine a metronome whose weight was suddenly displaced.)

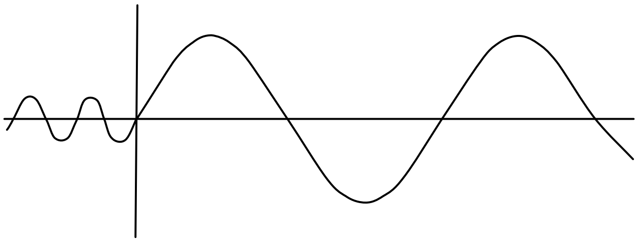

Figure 2.

Einstein imagined that the separation would be gradual. However, the two separate systems cannot amount to a unified mechanism, since they contradict a condition that **Einstein** himself expressed (II, § 1, final theorem). Thus, the mechanism will not change gradually, but two new mechanisms will appear in place of it at the moment of the complete separation.

(4) **A. Einstein**, I, pp. 426.

similarly, ε is a single-valued function $\varepsilon(t)$ of t. If the two systems Σ_1 and Σ_2 possess the same temperature t before the union then they will collectively have the energy $\varepsilon_1(t') + \varepsilon_2(t')$. Therefore, if $t_{12}(\varepsilon)$ denotes the temperature in the combined system Σ as a function of its energy then the temperature t″ that arises after the combination will be given by:

$$t'' = t_{12}\,\{\varepsilon_1(t') + \varepsilon_2(t')\}. \tag{48}$$

The solution of the problem of calculating t″ from t′ is rigorously *single-valued* and will lead to a proof of the Combination Theorem when the equation:

$$t'' = t_{12}\,\{\varepsilon_1(t') + \varepsilon_2(t')\} = t' \tag{49}$$

can be shown to be correct.

Things are completely different for the *Separation Theorem.* The energy of the combined system will oscillate continually between the two parts. The subsystems will keep the same energy after the separation that they had by chance at the moment of separation. Its temperature is then determined from that. The problem is then *not single-valued*, in full rigor. One cannot speak of *one* temperature for the separate bodies, but one can only ask what would be most probable one. It can arise from that fact that the processes of combination and separation are completely different from each other, and that deducing the one from the other would not be permissible.

§ 5. First attempt at a proof of the theorem of heat equilibrium

Since we regarded thermal contact in the previous paragraphs as the combination of two subsystems into a new one, we must next present some formulas that express V as a function $V_{12}(\varepsilon)$ of the energy in the combined system as long as the corresponding functions are known for the subsystems. We would like to choose the additive constant that generally appears in the energy function in such a way that the smallest energy that a mechanism can be assigned will possess the value zero, so, on the grounds of continuity, one will have:

$$V(0) = 0, \tag{50}$$

and the definition of V will directly imply the equations (1):

$$\left\{\begin{aligned} &a) && V_{12}(\varepsilon) = \int_0^{\varepsilon} \omega_1(\alpha) V_2(\varepsilon - \alpha)\,d\alpha, \\ &b) && V_{12}(\varepsilon) = \int_0^{\varepsilon} \omega_1(\varepsilon - \alpha) V_2(\alpha)\,d\alpha, \\ &c) && V_{12}(\varepsilon) = \int_0^{\varepsilon} V_1(\alpha)\,\omega_2(\varepsilon - \alpha)\,d\alpha, \\ &d) && V_{12}(\varepsilon) = \int_0^{\varepsilon} V_1(\varepsilon - \alpha)\,\omega_2(\alpha)\,d\alpha, \end{aligned}\right. \tag{51}$$

---

(1) **J. W. Gibbs**, *loc. cit.*, formulas (315) and (316).

which can go to each other by changing the notation and partial integration with the use of (18) and (51). By differentiating and applying just these equations, one will get:

(52)
$$\begin{cases} a) \quad \omega_{12} = \int_0^{\varepsilon} \omega_1(\varepsilon - \alpha)\omega_2(\alpha)\,d\alpha, \\ b) \quad \omega_{12} = \int_0^{\varepsilon} \omega_1(\varepsilon)\omega_2(\varepsilon - \alpha)\,d\alpha. \end{cases}$$

If t_1 (t_2, resp.) denotes the quantity t that is assigned to the first (second, resp.) mechanism as a function of ε_1 (ε_2, resp.) then one will have, from (38) and (51b):

$$V_{12} = \int_0^{\varepsilon} V_1(\varepsilon - \alpha)V_2(\alpha)\frac{1}{\mathrm{t}_1(\varepsilon - \alpha)}\,d\alpha,$$

and from (38) and (51d):

$$V_{12} = \int_0^{\varepsilon} V_1(\varepsilon - \alpha)V_2(\alpha)\frac{1}{\mathrm{t}_2(\alpha)}\,d\alpha\,.$$

If one takes the mean of the two integrals then it will follow that:

(53)
$$V_{12} = \int_0^{\varepsilon} V_1(\varepsilon - \alpha)V_2(\alpha)\frac{\mathrm{t}_1(\varepsilon - \alpha) + \mathrm{t}_2(\alpha)}{2\mathrm{t}_1(\varepsilon - \alpha)\mathrm{t}_2(\alpha)}\,d\alpha\,,$$

and from (38)and (55):

(54)
$$\omega_{12} = \int_0^{\varepsilon} V_1(\varepsilon - \alpha)V_2(\alpha)\frac{1}{\mathrm{t}_1(\varepsilon - \alpha)\mathrm{t}_2(\alpha)}\,d\alpha\,.$$

This integral cannot be evaluated, in general. However, we remark that, from (50), the function:

(55)
$$f(\alpha) = V_1\,(\varepsilon - \alpha)\,V_2\,(\alpha)$$

will possess a maximum. Call the place where that happens α_m .

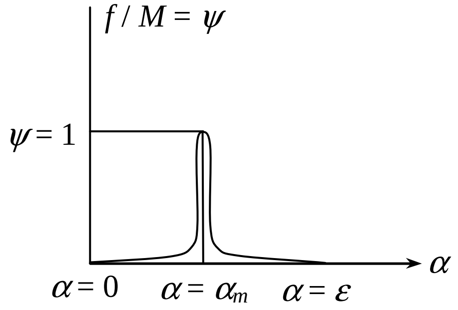

Figure 3.

We would now like to assume that the maximum is extremely steep; i.e., when f is divided by its maximum M, it shall drop to an exceptionally small value as soon as α also exceeds a small α_m, and therefore $\psi = f / M$ shall yield a curve of the kind that is represented in Fig. 3 geometrically. It will then be permissible to replace the t-dependent factor α with α_m in (53) and (54). Namely, if one understands ρ to mean the factor:

$$\rho = \frac{\mathrm{t}_1(\varepsilon-\alpha)+\mathrm{t}_2(\alpha)}{2\,\mathrm{t}_1(\varepsilon-\alpha)\,\mathrm{t}_2(\alpha)}$$

then one can set:

$$V_{12} = M \cdot \int_0^\varepsilon \frac{f}{M}\,\rho\,d\alpha\,.$$

However, since f / M differs from α_m essentially by zero, one can set:

$$V_{12} = M \cdot \int_0^\varepsilon \frac{f(\varepsilon)}{M}\,\rho\,(\alpha_m)\,d\alpha\,, \tag{56}$$

but that is:

$$V_{12} = \rho\,(\alpha_m)\int_0^\varepsilon f(\alpha)\,d\alpha\,, \tag{57}$$

or

$$V_{12} = \frac{\mathrm{t}_1(\varepsilon-\alpha_m)+\mathrm{t}_2(\alpha_m)}{2\,\mathrm{t}_1(\varepsilon-\alpha_m)\mathrm{t}_2(\alpha_m)}\int_0^\varepsilon V_1(\varepsilon-\alpha)V_2(\alpha)\,d\alpha\,. \tag{58}$$

One likewise has:

$$\omega_{12} = \frac{1}{\mathrm{t}_1(\varepsilon-\alpha_m)\mathrm{t}_2(\alpha_m)}\int_0^\varepsilon V_1(\varepsilon-\alpha)V_2(\alpha)\,d\alpha\,. \tag{59}$$

Now, it follows from (38) that:

$$\mathrm{t}_{12}\,(\varepsilon) = \tfrac{1}{2}[\mathrm{t}_{12}\,(\varepsilon-\alpha_m) + \mathrm{t}_{12}\,(\alpha_m)], \tag{60}$$

or when one sets:

$$\left\{\begin{array}{ll} a) & \varepsilon-\alpha_m = \varepsilon_1', \\ b) & \alpha_m = \varepsilon_2', \\ c) & \varepsilon_1'+\varepsilon_2' = \varepsilon, \end{array}\right. \tag{61}$$

that:

$$\mathrm{t}_{12}\,(\varepsilon) = \tfrac{1}{2}[\mathrm{t}_{12}\,(\varepsilon_1') + \mathrm{t}_{12}\,(\varepsilon_2')]. \tag{62}$$

However, α_m is defined by:

$$\left.\frac{df}{d\alpha}\right|_{\alpha=\alpha_m} = 0\,;$$

that is, from (55):

$$\frac{d}{d\alpha}\{V_1(\varepsilon-\alpha)V_2(\alpha)\}_{\alpha=\alpha_m} = 0 \tag{63}$$

or, from (18):

$$V_1\,(\varepsilon - \alpha_m)\;\omega_2\,(\alpha_m) = \omega_1\,(\varepsilon - \alpha_m)\;V_2\,(\alpha_m),$$

or, from (38):

$$\mathfrak{t}_1\,(\varepsilon - \alpha_m) = \mathfrak{t}_2\,(\alpha_m)\;. \tag{64}$$

From (61), we can write this as:

$$\begin{cases} \mathfrak{t}_1(\varepsilon_1') = \mathfrak{t}_2(\varepsilon_2'), \\ \varepsilon_1' + \varepsilon_2' = \varepsilon. \end{cases} \tag{65}$$

It then follows from (62) that:

$$\mathfrak{t}_{12}\,(\varepsilon) = \mathfrak{t}_1\,(\varepsilon_1') = \mathfrak{t}_2\,(\varepsilon_2')\;. \tag{66}$$

In order to then find the value of $\mathfrak{t}_{12}(\varepsilon)$ that corresponds to ε, one must solve the system of equations (65), and then obtain $\mathfrak{t}_{12}$ from (66); i.e., one must divide ε into two parts, such that the function $\mathfrak{t}_1$ for one of them is equal to the function $\mathfrak{t}_2$ for the other one. Physically speaking: One must distribute the energy over both systems in such a way that they both get the same value of $\mathfrak{t}$. That will then be the $\mathfrak{t}$ that is assigned to the total mechanism.

Now, if two mechanisms with the same $\mathfrak{t}'$ are given, and one sets:

$$\varepsilon_1(\mathfrak{t}') + \varepsilon_2(\mathfrak{t}') = \varepsilon$$

then $\varepsilon_1(\mathfrak{t}')$ and $\varepsilon_2(\mathfrak{t}')$ will satisfy the conditions (65) that relate to ε, since one has:

$$\mathfrak{t}_1\{\varepsilon_1(\mathfrak{t}')\} = \mathfrak{t}' = \mathfrak{t}_2\{\varepsilon_2(\mathfrak{t}')\}\;.$$

Thus, from (66), one will have:

$$\mathfrak{t}'' = \mathfrak{t}_{12}\{\varepsilon_1(\mathfrak{t}') + \varepsilon_2(\mathfrak{t}')\} = \mathfrak{t}_1\{\varepsilon_1(\mathfrak{t}')\} = \mathfrak{t}_2\{\varepsilon_2(\mathfrak{t}')\} = \mathfrak{t}';$$

i.e., (49) is proved. *If one combines two mechanisms with the same* $\mathfrak{t}'$ *into a new mechanism then it will likewise possess the same* $\mathfrak{t}'$.

One resolves the problem of decomposition in a different way. As long as the two bodies are coupled, the energy will continually oscillate from one to the other. Therefore, the probability that the two bodies possess an energy that is between α and $\alpha + d\alpha$ will amount to:

$$v(\alpha)\,d\alpha = \omega_1\,(\varepsilon - \alpha)\;\omega_2\,(\alpha)\,d\alpha,$$

so, from (38) and (55):

$$v(\alpha) = \frac{1}{\omega_{12}}\,\frac{1}{\mathfrak{t}_1(\varepsilon - \alpha)\,\mathfrak{t}_2(\alpha)}\;f\,(\alpha). \tag{67}$$

We assume that the second system possesses the same energy after the separation that it had immediately before the separation. No well-defined energy that the second body would have to possess can then be given with any certainty. By contrast, one can determine the most probable energy value that that the second system would assume.

That value will give the function v that is defined by (67) its maximum; however, when the maximum of the factor f that enters into (67) is exceptionally steep (cf., Fig. 3), the desired α will also give a maximum for f; i.e., it will coincide with the α_m that is determined by (61) and (62). If we assume that really the second body possesses the energy α_m after the separation, and the first one, the energy $\varepsilon - \alpha_m$ then, from (38), one will have $\mathfrak{t}_1 (\varepsilon - \alpha_m) = \mathfrak{t}_2 (\alpha_m)$; i.e., both bodies possess the same value of $\mathfrak{t}$, and from (64), that would be equal to the value of $\mathfrak{t}$ that prevailed before. Hence, if one divides a mechanism into two sub-mechanisms then each of them will take on a value of $\mathfrak{t}$ that existed before in the total mechanism.

At this point, we will first make use of the fact that the ensembles in question are temporal or virtual ensembles. Namely, if one defines V to be the space of all phases whose $\varepsilon < \varepsilon^*$ then, without assuming that all phases of equal ε will actually be reached eventually, and without making any sort of assumptions and conventions on the motion of the system, (51) must nevertheless be true, and when $\mathfrak{t}$ is defined by (38), instead of (37), so must (49). The combination theorem is therefore not at all a mechanical theorem; things are different for the separation theorem. For it, one must assume that all phases of the combined system will be reached eventually, and that will allow one to estimate the most probable phases in which the subdivision will take place.

Moreover, appealing to probability will lead to the concept of a virtual ensemble by itself. What does the probability that the energy α_m will come about in two bodies after separation mean then? It means nothing but the fact that we should repeat the experiment of separation very many times; in other words, that we should think in terms of an ensemble of independent system pairs and decompose each of them. According to our theorem, the second body will take on the energy α_m in the vast majority of pairs. We then have virtual ensembles. We now come to the distribution law in it. We have ascertained it here when we regarded it as the image of the temporal ensemble. Another theory of **Gibbs** proceeds in a different way. He presented general, statistical laws for system ensembles. At the moment of union, the distribution function can be computed from the distribution functions for the individual systems; it would not be at all necessary to examine how the distribution functions of the individual systems are arrived at. However, some theorems must now be introduced, from which, one can derive how the phase densities of the virtual systems change continuously at the moment of union. One sees that one is not dealing with contrived concepts, but with ones that are consistent with the nature of things. We shall go into that on another occasion. For now, the presentation that was given above is probably more convenient.

§ 6. Ideal gases.

In the previous paragraphs, we neglected something in connection with the assumptions about the character of the function $\psi = f / M$. There are three ways of doing that. Firstly, we must make the assumptions that we introduced before in an entirely indeterminate way more precise. That formulation came about in such a way that the propositions that were to be proved actually followed as consequences of mechanics and the assumed stipulations. The second problem then arose of deriving the laws of heat equilibrium mathematically on the basis of our special hypotheses. However, thirdly, one would have to verify that such a stipulation is actually realized in nature.

This last proof must obviously be linked with great difficulties, since the actual mechanism of the system is not known. We shall then restrict ourselves to carrying it out for ideal gases. Now, one must also prove the law of heat equilibrium directly for these bodies, and we thus seem to have achieved nothing by our considerations. However, that is not true. We have still shown by our investigation that the validity of those laws follows, not only from the special analytic character of the functions that appear in them, but also from certain ratios of magnitudes, and that it does not suffice to assume that the characteristic functions of the other bodies are equal to those of gases precisely, but only that the form of their peaks should correspond to them. We next provisionally show that the ideal gases satisfy our requirement in the indeterminate picture of the previous paragraphs.

If we choose the generalized coordinates q of the ideal gas to be its Cartesian coordinates, then understand p to mean the true mechanical quantities of motion, and call the mass of the molecule m then the energy ε, which consists of kinetic exclusively ([1]), will be given by the formula:

$$\varepsilon = \frac{1}{2m}\sum_{\kappa=1}^{n} p_\kappa^2 . \tag{68}$$

We now ask what volume the phases whose energy is less than ε^* assume. If v denotes the volume of the container then any variable q_κ will be likewise restricted to the volume v ([2]). If n denotes the number of degrees of freedom, so $n / 3$ denotes the number of molecules, then the positions of all molecules fill up a multidimensional space of size $v^{n/3}$. From (68), the p obey the condition:

$$\sum_{\kappa=1}^{n} p_\kappa^2 < \left(\sqrt{2m\varepsilon^*}\right)^2 ;$$

---

[1] That is true only approximately. In order to consider the influence of the walls, it will be necessary to introduce the potential energy between the wall and the molecule into the calculation. In the limit of rigidity, that can be regarded as a function that will become extremely large when the molecule comes close to the wall. Nevertheless, the potential energy can be neglected in the formula above. In any event, its magnitude will increase in proportion to the surface area of the container, and that of the kinetic energy will increase its volume, so the latter will predominate appreciably for a sufficient size of the container. We assume that this size of container has been attained. In many case, it can be preferable (§ 10) to choose some coordinate system besides a Cartesian one.

[2] As long as we prescribe no upper limit for the energy, any coordinate in infinite space is admissible, strictly speaking. Now, if ε^* is chosen to be the upper limit on the energy then the molecule can approach the wall only so closely that the potential energy:

$$< \varepsilon^* - \Sigma \frac{1}{2m} p_i^2$$

will be produced by it, in which the sum includes three terms; the wall can therefore be penetrated only very slightly. The q_κ of the volume that is available will then differ from v by only infinitely little.

i.e., they fill up the volume of an n-dimensional sphere of radius $\sqrt{2m\varepsilon^*}$. Such a sphere will possess the magnitude $\mathfrak{C}_n\left[\sqrt{2m\varepsilon^*}\right]^n$, in which $\mathfrak{C}_n$ means a constant that depends upon only n. Therefore, one will have:

$$V = c\, e^{n/2}\, v^{n/3}, \tag{69}$$

when one denotes a constant that depends upon only the type of gas and the quantum of gas by c and writes simply ε, instead of ε^*.

One then has (¹):

$$\ln V = \ln c + \frac{n}{2}\ln \varepsilon + \frac{n}{3}\ln v, \tag{70}$$

so, from (39) (²):

$$\varepsilon = \frac{n}{2}\,\mathfrak{t}. \tag{71}$$

Now, heat equilibrium already follows from this equation for two *equal* ideal gases. The expression $\frac{1}{2}\{\mathfrak{t}_1(\varepsilon - \alpha) + \mathfrak{t}_2(\alpha)\}$ that appears in (53) is equal to $\frac{1}{2}\mathfrak{t}_1(\varepsilon) = \frac{1}{2}\mathfrak{t}_2(\varepsilon)$ for all α, so it can be taken out of the integral. (58) and (59) then yield:

$$\mathfrak{t}_1 = \tfrac{1}{2}\mathfrak{t}_1(\varepsilon) = \tfrac{1}{2}\mathfrak{t}_2(\varepsilon),$$

so

$$\mathfrak{t}'' = \mathfrak{t}_{12}\{\varepsilon_1(\mathfrak{t}') + \varepsilon_2(\mathfrak{t}')\} = \mathfrak{t}_{12}\{2\varepsilon_1(\mathfrak{t}')\} = \mathfrak{t}_1\{\varepsilon_1(\mathfrak{t}')\} = \mathfrak{t}'.$$

However, we now come to the form of the function f that was introduced in the previous paragraphs, which shall be the exemplar for the character of f that we shall assume in the general case. If we then consider two different ideal gases with degrees of freedom n_1 and n_2, resp., constants [formula (69)] c_1 and c_2, resp., and volumes v_1 and v_2 then we will have:

$$f(\alpha) = C_{12}\,(\varepsilon - \alpha)^{\lambda_1}\alpha^{\lambda_2}, \tag{72}$$

in which we have set:

$$C_{12} = c_1 \cdot c_2\; v_1^{n_1/3} \cdot v_2^{n_2/3}, \tag{73}$$

$$\left\{\begin{aligned} \lambda_1 &= \frac{n_1}{2}, \\ \lambda_2 &= \frac{n_2}{2}. \end{aligned}\right. \tag{74}$$

However, α_m is given by:

(¹) **M. Planck**, *Boltzmann-Festschrift*, pp. 101.

(²) One will also get this from (37) when one considers the fact that ε_p is continually equal to ε.

(75)
$$\frac{\lambda_1}{\varepsilon-\alpha_m}=\frac{\lambda_2}{\alpha_m},$$

(76)
$$\left\{\begin{aligned}\alpha_m &= \frac{\lambda_2}{\lambda_1+\lambda_2}\varepsilon,\\ \varepsilon-\alpha_m &= \frac{\lambda_1}{\lambda_1+\lambda_2}\varepsilon.\end{aligned}\right.$$

One will further have:

(77)
$$\psi=\frac{f}{M}=\left(\frac{\varepsilon-\alpha}{\varepsilon-\alpha_m}\right)^{\lambda_1}\cdot\left(\frac{\alpha}{\alpha_m}\right)^{\lambda_2}.$$

If one now sets, to abbreviate:

(78)
$$\frac{\lambda_2}{\lambda_1}=\mu,$$

then it will follow that:

(79)
$$\psi=\left[\left(\frac{\varepsilon-\alpha}{\varepsilon-\alpha_m}\right)\left(\frac{\alpha}{\alpha_m}\right)^{\mu}\right]^{\lambda_1}.$$

If we assume that μ is a well-defined, finite number, while λ_1 increases beyond all limits, then ψ will be indeed be equal to unity for $\alpha=\alpha_m$, while the expression that is found in the square brackets in (79) will become a proper fraction when α also deviates only slightly from α_m, so ψ will be quite small. The function ψ will then have the form that is depicted in Fig. 3 for large λ_1 and finite μ. The maximum of the curve f is infinitely thin for an infinitely large number of degrees of freedom. We see that the requirement that the function f is subject to in the indeterminate picture of the previous paragraphs is fulfilled. A more precise formulation of that condition will be given in the later investigation.

In the meantime, we must still evaluate the integrals:

(80)
$$J=\int_0^{\varepsilon} f(\alpha)\,d\alpha$$

and

(51*c*)
$$V_{12}=\int_0^{\varepsilon} V_1(\alpha)\,\omega_2(\varepsilon-\alpha),$$

(51*b*)
$$\omega_{12}=\int_0^{\varepsilon} \omega_1(\alpha)\,\omega_2(\varepsilon-\alpha).$$

Partial integration then generally yields:

$$\int_0^{\varepsilon}(u-\xi)^a\xi^b d\xi=\frac{a}{b+1}\int_0^{\varepsilon}(u-\xi)^{a-1}\xi^{b+1}d\xi.$$

Therefore:

(81) $$\int_0^{\varepsilon} (u-\xi)^a \xi^b d\xi = \frac{a!b!}{(a+b+1)!} u^{a+b+1} .$$

Now, from (72), (80), and (81):

(82) $$J = C_{12} \cdot \frac{\lambda_1 ! \, \lambda_2 !}{(\lambda_1 + \lambda_2 + 1)!} \varepsilon^{\lambda_1 + \lambda_2 + 1} ,$$

and from (51c), (69), (18):

$$V_{12} = \lambda_2 \, c_1 \, c_2 \; v_1^{n_1/3} \, v_2^{n_2/3} \int_0^{\varepsilon} (\varepsilon - \alpha)^{\lambda_1} \alpha^{\lambda_2 - 1} d\alpha .$$

which is, from (73) and (81):

$$= \lambda_2 \cdot C_{12} \cdot \frac{\lambda_1 ! \, (\lambda_2 - 1)!}{(\lambda_1 + \lambda_2)!} \varepsilon^{\lambda_1 + \lambda_2} ,$$

so

(83) $$V_{12} = C_{12} \cdot \frac{\lambda_1 ! \, \lambda_2 !}{(\lambda_1 + \lambda_2)!} \varepsilon^{\lambda_1 + \lambda_2} .$$

Likewise, from (52b), (69), (18), (73), (81), one has:

$$\omega_{12} = C_{12} \cdot \lambda_1 \cdot \lambda_2 \frac{(\lambda_1 - 1)! \, (\lambda_2 - 1)!}{(\lambda_1 + \lambda_2 - 1)!} \varepsilon^{\lambda_1 + \lambda_2 - 1} ,$$

(84) $$\omega_{12} = C_{12} \cdot \frac{\lambda_1 ! \, \lambda_2 !}{(\lambda_1 + \lambda_2)!} \varepsilon^{\lambda_1 + \lambda_2} ,$$

such that, from (38), one will get:

$$\mathfrak{t}_{12} = \frac{\varepsilon}{\lambda_1 + \lambda_2} .$$

If *n* means the number of degrees of freedom in the combined system then, from (74):

$$\varepsilon = \frac{n}{2} \; \mathfrak{t}_{12} .$$

This equation, which in turn, can be derived from (37) immediately, can also serve as the proof of heat equilibrium. Namely, from (48), one has:

$$\mathfrak{t}'' = \mathfrak{t}_{12}\ \{\varepsilon_1(\mathfrak{t}')+\varepsilon_2(\mathfrak{t}')\} = \frac{n}{2}\left\{\frac{\mathfrak{t}'}{\frac{n_1}{2}}+\frac{\mathfrak{t}'}{\frac{n_2}{2}}\right\} = \mathfrak{t}'.$$

§ 7. Requirements on certain functions that are connected with the mechanism (1).

We would now like to present the conditions that f must satisfy in order for us to prove the law of heat equilibrium. Later, we shall show that the function f is, in fact, guaranteed to satisfy these conditions for ideal gases. It will then be shown that one does not necessarily need to assume a corresponding analytical behavior for the function f. It is sufficient to have a correspondence in regard to the proportions.

The ratio of two quantities shall be quite small, and indeed, that shall come about under the assumption that the number of degrees of freedom is very large. However, the concept of "quite small" should probably not be taken mathematically. We then propose that we are considering a large number of system pairs with an ever-increasing number of degrees of freedom, although the ratio of their degrees of freedom $n_2 / n_1 = \mu$ shall possess the same value through all of that. Each system pair $S_1^{(n_1)}$, $S_2^{(n_2)}$ is associated with functions $V_1^{(n_1)}, V_2^{(n_2)}$, resp., and a function $f^{(n)}$, where n means a quantity that will become infinite along with n_1, n_2 (e.g., n_1), and one can ask what the limiting value is that will be approached by a certain ratio that is defined by these functions when n increases. One must then still further assume that n has already attained the required magnitude for the systems and system pairs that nature commands to exist (2).

We first give the requirement for f in an imprecise fashion. If we set the factors that appear in (53) and (54) equal to:

(85) $$\rho(\alpha) = \frac{1}{2}\frac{\mathfrak{t}_1(\varepsilon-\alpha)+\mathfrak{t}_2(\alpha)}{\mathfrak{t}_1(\varepsilon-\alpha)\,\mathfrak{t}_2(\alpha)},$$

(86) $$\chi(\alpha) = \frac{1}{\mathfrak{t}_1(\varepsilon-\alpha)\,\mathfrak{t}_2(\alpha)}$$

then we will arrive at:

---

(1) § 7 and § 8 include mathematically-specialized explanations, and would be better skipped over by those who would seek a provisional orientation in statistical mechanics.

(2) One can, e.g., decompose a homogeneous body into very many small pieces, and then put them together them step-by-step. One then gets a sequence of bodies and a family of functions $V^{(\nu)}$ for an increasing number of components. The number of degrees of freedom would be a multiple of ν, and would become infinite with ν. Obviously, one has the recursion formula for the $V^{(\nu)}$:

$$V^{(\nu+1)} = \int_0^\varepsilon V^{(\nu)}(\varepsilon-\alpha)\frac{d}{d\alpha}[V^{(1)}(\alpha)]\ d\alpha.$$

It would be worthwhile to examine whether this chain of $V^{(\nu)}$ always satisfies the conditions that were posed in the text for an arbitrary starting function $V^{(1)}$.

1. The maximum of f will always become steeper with an increasing number of degrees of freedom.

2. The maximal values of ρ and χ are not so very different from the ones that they have for α_m .

3. The steepness of ρ and χ shall stay within measurable limits when f becomes ever steeper.

We now give these demands a mathematical form. We then set $n_1 = \lambda_2$; $n_2 = \lambda_2 = \lambda_1\,\mu$, and infer that:

1. For every pair of (still small) numbers h and b, there is a number Λ that has the following property: α_m can be surrounded by a region I of size b, while II is understood to be the remaining region, such that for all $\lambda_1 > \Lambda$ the inequality:

$$\int_{\Pi} f(\alpha)\,d(\alpha) < h \cdot \int_0^{\varepsilon} f(\alpha)\,d\alpha \tag{87}$$

will be true. That is, for a sufficiently large number of degrees of freedom, the integral J can be obtained, up to a sufficiently small remainder, by integration over an arbitrarily-small interval around the maximum.

2. There is a number Q that is independent of λ_1 such that for all λ_1, one has the inequalities:

$$\chi(\alpha) < Q \cdot \chi(\alpha_m), \tag{88}$$

$$\rho(\alpha) < Q \cdot \rho(\alpha_m). \tag{89}$$

3. There is a number S that is independent of λ_1 such that for all λ_1 and α one has the inequalities:

$$\chi'(\alpha) < S \cdot \chi(\alpha_m), \tag{90}$$

$$\rho'(\alpha) < S \cdot \rho(\alpha_m). \tag{91}$$

We would like to show that heat equilibrium will be fulfilled to any desired precision by a sequence of system pairs that satisfy these conditions. Let h and b be arbitrarily-chosen numbers, and let Λ be the quantity that corresponds to them from condition 1. We then divide the interval 0 to ε into the regions I and II that were mentioned in the first condition, and set, from (54), (55), and (86):

$$\omega_{12} = \int_0^{\varepsilon} f(\alpha)\,\chi(\alpha)\,d\alpha = K. \tag{92}$$

We then decompose:

$$K = K_1 + K_2 , \tag{93}$$

$$K_1 = \int_{\mathrm{I}} f(\alpha)\,\chi(\alpha)\,d\alpha , \tag{94}$$

$$K_{1\mathrm{I}} = \int_{\mathrm{II}} f(\alpha)\,\chi(\alpha)\,d\alpha\,. \tag{95}$$

All of the functions that appear here depend upon the index λ_1 . However, for $\lambda_1 > \Lambda$, $\lambda_2 > \Lambda\mu$, one will have:

$$K_{1\mathrm{I}} < Q\cdot\int_{\mathrm{II}} f(\alpha)\,\chi(\alpha_m)\,d\alpha\,,$$

and when we once more set:

$$\int_0^{\varepsilon} f(\alpha)\,\chi(\alpha)\,d\alpha = J, \tag{80}$$

from condition I (87):

$$K_{1\mathrm{I}} < \chi(\alpha_m)\cdot Q\,h\cdot J,$$

i.e.:

$$K_{1\mathrm{I}} = \eta_2\cdot\chi(\alpha_m)\cdot Q\cdot hJ, \tag{96}$$

$$\eta_2 < 1. \tag{97}$$

Moreover, from the mean value theorem of integral calculus, one will have:

$$K_1 = \chi(\alpha_m + \zeta)\int_{\mathrm{I}} f(\alpha)\,d\alpha\,, \tag{98}$$

where

$$|\,\zeta| < b. \tag{99}$$

Since:

$$\int_{\mathrm{I}} f(\alpha)\,d\alpha = (1 - h\,\eta_1)\,J, \tag{100}$$

where one should have:

$$\eta_1 < 1, \tag{101}$$

one will get the relation:

$$K_{\mathrm{I}} = \chi(\alpha_m + \zeta)\,(1 - h\,\eta_1)\,J. \tag{102}$$

Now, one has:

$$\chi(\alpha_m + \zeta) = \chi(\alpha_m) + \zeta\cdot\chi'(\alpha_m + \eta_1\,\zeta),$$

so from conditions 3 and 2, and from (99):

$$\chi(\alpha_m + \zeta) = \chi(\alpha_m)\,[1 + \eta_4\,b\cdot Q\,S],$$

where one should have:

$$\eta_3 < 1,$$

$$\eta_4 < 1.$$

Due to (102), one then has:

$$K_1 = (1 - h\,\eta_1)\,J\,(1 + \eta_4\,b\cdot Q\,S)\,\chi(\alpha_m)\,. \tag{103}$$

Therefore, from (93), (103), (96):

$$K = J\cdot\chi(\alpha_m)\left\{\begin{array}{c} 1 + h\cdot[Q\eta_2 - \eta_1], \\ +b[\eta_2\,Q\,S], \\ -b\,h[\eta_1\,\eta_4\,Q\,S] \end{array}\right\}. \tag{104}$$

Now, since, from condition 1, *h* and *b* can be chosen to be as small as one desires in order to insure a sufficient magnitude for Λ, one will have:

$$\lim_{\lambda_1=\infty} K = J \cdot \chi(\alpha_m) \,,$$

or, in more detail:

$$\omega_{12}^{(n)} = \chi^{(n_2)}(\alpha_m)\left[\int_0^{\varepsilon} V_1^{(n_1)}(\varepsilon-\alpha)\,V_2^{(n_2)}(\alpha)\,d\alpha\right][1+\beta^{(n)}], \tag{105}$$

$$\lim_{n=\infty} \beta^{(n)} = 0. \tag{106}$$

In exactly the same way, one can show that:

$$V_{12} = \rho^{(n_2)}(\alpha_m)\left[\int_0^{\varepsilon} V_{12}^{(n_1)}(\varepsilon-\alpha)\,V_2^{(n_2)}(\alpha)\,d\alpha\right][1+\gamma^{(n)}], \tag{107}$$

$$\lim_{n=\infty} \gamma^{(n)} = 0. \tag{108}$$

However, from (85), (86), and (64), it follows from this that:

$$\mathfrak{t}_{12}^{(n)} = \mathfrak{t}_{12}^{(n_2)}(\alpha_m)\,[1+\delta^{(n)}], \tag{109}$$

$$\lim_{n=\infty} \delta^{(n)} = 0, \tag{110}$$

or, more briefly:

$$\mathfrak{t}'' = \mathfrak{t}'.$$

Now, conditions 1 and 3 that were posed above are not both fulfilled in the case of ideal gases. Indeed, conditions 1 and 3 will apply in that case, but not the second one, since the functions χ and ρ become infinite at the points 0 and ε, and there will be no finite maxima for those functions. However, one can pose extended conditions that gases also satisfy and which essentially say that the conditions that were posed up to now are still true, except at critical points, but on the other hand, the influence of those points on the integration is not significant.

In fact, the presence of such critical points is not at all questionable physically. It would then have to almost completely exclude the possibility that the energy would be found completely in one or the other subsystem at some point in time. However, one will arrive at the aforementioned condition when one estimates the probability of such an event to be extremely slim.

In general, it is undeniable that we have gone beyond the scope of what we did up to now with that presentation. Up to now, we posed requirements on only the functions f, χ, and ρ, and their integrals, but not on the integral $\int f \cdot \chi \cdot d\alpha$, because that would be something that would indeed have to be developed, and any demand regarding it would seem to be a kind of *petitio principii*. However, one will admit that basically very little was assumed here besides that a complete shift of energy from one part into the other one would happen only quite rarely.

We alter the integrals V_{12} and ω_{12} when we set:

$$J(\varepsilon;\kappa) = \int_{\kappa}^{\varepsilon-\kappa} f(\alpha)\,d\alpha\,, \tag{111}$$

$$K(\varepsilon;\kappa) = \int_{\kappa}^{\varepsilon-\kappa} f(\alpha)\,\chi(\alpha)\,d\alpha\,, \tag{112}$$

$$L(\varepsilon;\kappa) = \int_{\kappa}^{\varepsilon-\kappa} f(\alpha)\,\rho(\alpha)\,d\alpha\,, \tag{113}$$

in which we understand κ to mean a positive number.

Moreover, we pose the extended system of requirements:

1. The condition 1 that was stated on pp. 29 shall be valid for the interval from 0 to ε.

2. Conditions 2 and 3 that are stated on pp. 29 shall be true for any region κ up to $\varepsilon-\kappa$; the (for every κ; there will be another Q and another S).

3. One has:

$$\lim_{\kappa=0}\frac{K^{(n)}(\varepsilon;\kappa)}{K^{(n)}(\varepsilon;0)} = 1,$$

$$\lim_{\kappa=0}\frac{L^{(n)}(\varepsilon;\kappa)}{L^{(n)}(\varepsilon;0)} = 1,$$

independently of the number of degrees of freedom, and thus, of n.

Loosely speaking, that means: By excluding the critical points, one will always obtain a region for which the previous conditions are valid. One can then come arbitrarily close in the integration of the total integral by reducing the excluded region.

It needs to be shown that heat equilibrium will also be true with these assumptions. Let δ be a small number. We impose two conditions upon κ in regard to its smallness. Firstly, let ([1]):

$$\int_{\kappa}^{\varepsilon-\kappa} f(\alpha)\,d\alpha = \left(1+\frac{\eta_1}{4}\delta\right)\int_{0}^{\varepsilon} f(\alpha)\,d\alpha\,. \tag{114}$$

$$|\,\eta_1\,| < 1.$$

---

[1] From condition 1, that will be possible.

Secondly, we choose κ to be small enough that for all n, one will have:

$$K(\varepsilon) = K(\varepsilon;\, 0) = K(\varepsilon;\, \kappa)\,[1 + \tfrac{1}{4}\,\eta_2\,\delta]. \tag{115}$$

It follows easily from the conditions I on pp. 29 and 32 that the quotient of the two integrals:

$$J(\varepsilon;\, \kappa) = \int_{\kappa}^{\varepsilon-\kappa} f(\alpha)\,d\alpha$$

and

$$J(\varepsilon;\, 0) = \int_{0}^{\varepsilon} f(\alpha)\,d\alpha$$

can be made as close to 1 as one desires for a sufficiently large choice of n ([1]). On the other hand, if b is an arbitrary quantity, and I is again understood to mean a region surrounding α_m that has the size 1 then since:

$$J_1 < J(\varepsilon;\, \kappa) < J,$$

and those integrals are positive, one will have:

$$\lim_{\kappa=\infty} \frac{J_1}{J(\varepsilon;\kappa)} = 1.$$

Therefore, the first condition of our previous system of demands is true for the interval κ to $\varepsilon - \kappa$; and from the second condition on pp. 32, the rest of the conditions, as well. Thus, all of the arguments that were presented before for the region 0 to ε can be repeated, and one choose n such that the relations:

$$K^{(n)}(\varepsilon;\, \kappa) = \left[\int_{\kappa}^{\varepsilon-\kappa} f(\alpha)\,d\alpha\right]\chi(\alpha_m)\left[1+\frac{\delta}{4}\eta_3\right]$$

$$\eta_3 < 1$$

exist. From (114) and (115), one then has:

$$K^{(n)}(\varepsilon) = [1 + \delta\,\eta_4]\left[\int_{0}^{\varepsilon} f(\alpha)\,d\alpha\right]\chi(\alpha_m) \tag{116}$$

$$\eta_3 < 1.$$

Since δ is subject to no restriction, that means:

---

[1] One needs for the quantity b that was spoken of in condition 1 on pp. 29 to take the value $\varepsilon - \kappa$ for a moment. The aforementioned interpretation will then follow easily.

$$\lim_{n=\infty} K^{(n)}(\varepsilon) = \chi(\alpha_m)\cdot\int_0^{\varepsilon} f(\alpha)\,d\alpha\,.$$

Likewise:

$$\lim_{n=\infty} L^{(n)}(\varepsilon) = \rho(\alpha_m)\cdot\int_0^{\varepsilon} f(\alpha)\,d\alpha\,.$$

It then finally follows from this that:

$$\mathfrak{t}'' = \mathfrak{t}'.$$

One then dispatches with the proof of the separation theorem very easily. There exists a probability of:

$$v\,(\alpha_1,\,\alpha_2) = \frac{1}{\omega_{12}}\int_{\alpha_1}^{\alpha_2} \omega_{12}(\varepsilon-\alpha)\,\omega_2(\alpha)\,d\alpha \tag{117}$$

that of two mutually-coupled subsystems, the second one will possess an energy between the limits α_1 and α_2 . The probability that the body will have a temperature between $\mathfrak{t}_1$ $(\varepsilon - \alpha_1)$ and $\mathfrak{t}_1$ $(\varepsilon - \alpha_2)$ by the separation will be just as large.

If we again choose two numbers *h* and *b* and the interval I that belongs to condition 1 (pp. 29) then the probability v_1 that α lies in I will be:

$$v_1 = \frac{K_1^{(n)}}{K^{(n)}}\,.$$

It now follows from (116) and (103) (1) that v_1 can be made arbitrarily close to unity by a suitable choice of *h* and *b* and the Λ that is associated with them. For sufficiently large *n*, the probability that $\mathfrak{t}_1$ and $\mathfrak{t}_2$ will equal $\mathfrak{t}'$ after the separation will come arbitrarily close to a certainty; as we might say briefly, the equation:

$$\mathfrak{t}' = \mathfrak{t}''$$

will also be true for the case of separation.

§ 8. Proof that the requirements that were posed above are fulfilled for ideal gases.

It would be desirable to show that the requirements that were posed in the previous paragraphs are fulfilled by the systems that exist in nature. However, since we do not know their true mechanisms, a solution to that problem must be ruled out. Moreover, one can think of carrying out the corresponding proof of a series of homogeneous bodies on the basis of the formulas that were stated in the previous paragraphs (pp. 28, rem.). That

---

(1) The derivation of equation (103) indeed employs only conditions 2 and 3 on pp. 29, but only for the region I. However, from the condition 2 on pp. 32, they must be true in the interior of I.

is also not possible. We would therefore like to set a more modest goal and seek to show that our requirements will be legitimate for ideal gases. Therefore, we are well aware that our considerations will achieve nothing new, since the case of ideal gases has been treated already in simpler ways (§ 6). However, a possible source of doubt will be omitted from the outset: The proof that we shall give is not sufficient to support our theory, but it is necessary for it to be valid.

From (71), (74), (85), (86), and (78), one has:

$$\rho(\alpha) = \left[\frac{\mu}{\alpha\cdot(\varepsilon-\alpha)}\right]\lambda_1^2, \tag{118}$$

$$\chi(\alpha) = \left[\frac{\alpha+\mu\cdot(\varepsilon-\alpha)}{2\cdot\alpha\cdot(\varepsilon-\alpha)}\right]\lambda_1. \tag{119}$$

Now, if *k* is an arbitrary finite number then the expressions in the square brackets in (118) and (119) will remain within a limit that is independent of λ_1, so condition 2 on pp. 32 is fulfilled.

We now address the first condition on pp. 29 (pp. 32, resp.). We pick a small number δ and consider the length of the interval in which $f > M\delta$ ($\psi > \delta$). In order to do that, we must solve the equation $\psi = \delta$. We now substitute:

$$\alpha - \alpha_m = \zeta \tag{120}$$

in (79) and get:

$$\psi = \left[\left(\frac{\varepsilon-\alpha_m-\zeta}{\varepsilon-\alpha_m}\right)\left(\frac{\zeta+\alpha_m}{\alpha_m}\right)^{\mu}\right]^{\lambda_1}.$$

One then has to solve the equation:

$$\frac{\varepsilon-\alpha_m-\zeta}{\varepsilon-\alpha_m}\left(1+\frac{\zeta}{\alpha_m}\right)^{\mu} = \delta^{1/\lambda_1}.$$

In the even that ζ is small in comparison to α_m (we will confirm that this is the case later on), we will get:

$$\left(\frac{\varepsilon-\alpha_m-\zeta}{\varepsilon-\alpha_m}\right)\left(1+\frac{\mu\zeta}{\alpha_m}\right)^{\mu} = \delta^{1/\lambda_1}$$

or

$$\zeta^2\frac{\mu}{\alpha_m(\varepsilon-\alpha_m)}+\zeta\left(\frac{1}{\varepsilon-\alpha_m}-\frac{\mu}{\alpha_m}\right) = 1 - \delta^{1/\lambda_1}.$$

However, from (75) and (78), one has the relation:

(121)
$$\frac{1}{\varepsilon - \alpha_m} - \frac{\mu}{\alpha_m} = 0,$$

so one will get:

$$\zeta^2 = \frac{\alpha_m(\varepsilon - \alpha_m)}{\mu}\left(1 - \delta^{1/\lambda_1}\right),$$

or with repeated use of (121):

$$\zeta^2 = \frac{\alpha_m^2}{\mu^2}\left(1 - \delta^{1/\lambda_1}\right).$$

If δ denotes the size of the region in which $f < M\,\delta$ then one will obtain:

(122)
$$b = 2 \cdot \frac{\alpha_m}{\mu}\sqrt{1 - \delta^{1/\lambda_1}},$$

or, from (76) and (78):

(123)
$$b = \frac{2\varepsilon}{1+\mu}\sqrt{1 - \delta^{1/\lambda_1}}.$$

However, for large λ_1, one will have:

$$\delta^{1/\lambda_1} = e^{(\ln b)/\lambda_1} = 1 + \frac{1}{\lambda_1}\ln\delta,$$

so:

(124)
$$b = \frac{2\varepsilon}{1+\mu}\sqrt{\frac{-\ln\delta}{\lambda_1}}$$

or

(125)
$$b = r\sqrt{\frac{-\ln\delta}{\lambda_1}},$$

in which we have set:

(126)
$$r = \frac{2\varepsilon}{1+\mu}.$$

If we understand I to mean the region in which $f > M\,\delta$, II, to mean the region in which $f < M\,\delta$, and understand J_{I} and J_{II} to mean the corresponding integrals $\int f(\alpha)\,d\alpha$, while J means the integral $\int f(\alpha)\,d\alpha$ when it is taken over the entire region 0 to ε, then we will have:

(127)
$$J_{\text{II}} < \delta \cdot M \cdot \varepsilon.$$

From (82), one has:

(82)
$$J = C_{12}\,\frac{\lambda_1!\,\lambda_2!}{(\lambda_1+\lambda_2+1)!}\,e^{\lambda_1+\lambda_2+1},$$

so, from (72) and (76):

$$M = C_{12} \cdot e^{\lambda_1+\lambda_2} \frac{\lambda_1^{\lambda_1}\lambda_2^{\lambda_2}}{(\lambda_1+\lambda_2)^{\lambda_1+\lambda_2}} .$$

Therefore, from (121)

$$\frac{J_{\mathrm{II}}}{J} < \delta(\lambda_1 + \lambda_2 + 1) \frac{\lambda_1^{\lambda_1}\lambda_2^{\lambda_2}}{\lambda_1!\lambda_2!} \frac{(\lambda_1+\lambda_2)!}{(\lambda_1+\lambda_2)^{\lambda_1+\lambda_2}} .$$

However, from **Stirling**'s formula, it will follow for large λ_1 and λ_2, if one neglects 1 in comparison to $\lambda_1 + \lambda_2$, that:

$$\frac{J_{\mathrm{II}}}{J} < \delta(\lambda_1 + \lambda_2) \sqrt{\frac{\lambda_1+\lambda_2}{2\pi \cdot \lambda_1\lambda_2}}$$

or

$$\frac{J_{\mathrm{II}}}{J} < \delta \sqrt{\frac{(\lambda_1+\lambda_2)^3}{2\pi\,\lambda_1\lambda_2}} ,$$

so, from (78):

(128)
$$\frac{J_{\mathrm{II}}}{J} < \delta \sqrt{\frac{(1+\mu)^3}{2\pi\,\mu}}\sqrt{\lambda_1} ,$$

or

(129)
$$\frac{J_{\mathrm{II}}}{J} < h,$$

(130)
$$h = \delta \cdot s \cdot \sqrt{\lambda_1} ,$$

(131)
$$s = \sqrt{\frac{(1+\mu)^3}{2\pi\,\mu}} .$$

It then follows from (125) and (130) that:

(132)
$$b = r \cdot \sqrt{\frac{1}{\lambda_1}\{\ln s - \ln h + \tfrac{1}{2}\ln \lambda_1\}} .$$

This equation represents a dependency between h and b. h is arbitrary in it. It will be associated with a number b that has the following property: If one understands I to mean an interval that surrounds α_m symmetrically, and II to mean the remaining region then:

(129)
$$J_{\mathrm{II}} < h \cdot J.$$

We once more recognize the relation (87) in the inequality (129). We further see that for a *particular* λ_1, a very large b will belong to a very small h ([1]). By contrast, if h is fixed

---

[1] Equation (132) gives only an upper limit for the length of the region I that enters into requirement 1.

(let it still be chosen to be small) then one can arrange that b is as small as one desires by increasing λ_1. Briefly, the larger that λ_1 is, the more that the integration involves only the immediate neighborhood of the maximum. Condition 1 [pp. 29 (pp. 32, resp.)] is then fulfilled.

We shall now address the proof that the third condition that was posed on pp. 32 is also fulfilled. For that, we consider the integrals that were defined by (112) and (113). In them, we will encounter the difficulty that the factors ρ and χ that appear in them possess no upper limit that is common to all κ. However, if we are dealing with an ideal gas then we will not necessarily have to deal with the individual factors in those integrals separately, but we can employ the expression for the undecomposed product that was given previously. Thus, from (92), (52b), (69), (73), (18), K will be the integral of the function:

$$f = C_{12}\,\lambda_2\,\lambda_2\,(\varepsilon-\alpha)^{\lambda_1-1}\alpha^{\lambda_2-1}$$

or

$$f = C_{12}\cdot\mu\,\lambda_1^2\,(\varepsilon-\alpha)^{\lambda_1-1}\alpha^{\lambda_2-1}. \tag{133}$$

It can now be shown that f also satisfies the conditions (1) (pp. 29 and pp. 32) in the interval 0 and ε. We only need to alter the proof that was given on p. 34 slightly. If M denotes the maximum of f, which is assumed to be at the location α_m, then we will get the formulas:

$$b = r\sqrt{\frac{-\ln\delta}{\lambda_1-1}}, \tag{134}$$

$$r = \frac{2\varepsilon}{1+\dfrac{\lambda_2-1}{\lambda_1-1}} \tag{135}$$

in place of (125) to (126). They associate each δ with a quantity b with the property that the inequality $f < \delta M$ is true outside of the symmetric region II around α_m with the size b. When one assumes that λ_1 is large, one can set:

$$b = \sqrt{\frac{-\ln\delta}{\lambda_1}}, \tag{136}$$

$$r = \frac{2\varepsilon}{1+\mu}. \tag{137}$$

One modifies the second part of the proof similarly. One finds:

$$\frac{K_{\text{II}}}{K} < h, \tag{138}$$

(139) $$h = \delta \cdot s \cdot \sqrt{\lambda_1 - 1}\,,$$

(140) $$s = \sqrt{\frac{\left|1 + \left(\dfrac{\lambda_2 - 1}{\lambda_1 - 1}\right)\right|^2}{2\pi \dfrac{\lambda_2 - 1}{\lambda_1 - 1}}}\,,$$

in place of (129) to (131). For large λ_1, that will yield the system:

(141) $$h = \delta s \sqrt{\lambda_1}\,,$$

(142) $$s = \sqrt{\frac{(1+\mu)^2}{2\pi\mu}}\,.$$

It will then follow once more from this that:

(143) $$b = r\sqrt{\frac{1}{\lambda_1}(\ln s - \ln h + \tfrac{1}{2}\ln \lambda_1)}\,,$$

which then shows that the integrand f of K also satisfies the first condition (pp. 32).

We now choose an arbitrarily small number h. From (138) and (143), a pair of numbers b and $\mathfrak{n}$ can be found such that for all $\lambda_1 > \mathfrak{n} / 2$ or $n > \mathfrak{n}$, one has:

$$\frac{K_{\text{II}}}{K} < h$$

or

$$1 - \frac{K_{\text{I}}}{K} < h.$$

Now, if κ_1 is a number such that I lies completely in the region between κ_1 and $\varepsilon - \kappa_1$ then when one applies the relation (112) one will have:

$$1 - \frac{K^{(n)}(\varepsilon;\kappa_1)}{K^{(n)}(\varepsilon;0)} < h$$

for all $n > \mathfrak{n}$.

One then has, *a fortiori*:

$$1 - \frac{K^{(n)}(\varepsilon;\kappa)}{K^{(n)}(\varepsilon;0)} < h$$

for all $n > \mathfrak{n}$ and all $\kappa > \kappa_1$. The number of functions $K^{(1)}$, $K^{(2)}$, … up to $K^{(n)}$ is, however, finite. Due to the continuity of $f^{(n)}$, one can also find a number κ_2 such that:

$$1 - \frac{K^{(n)}(\varepsilon;\kappa)}{K^{(n)}(\varepsilon;0)} < h$$

for all $\kappa < \kappa_1$ and for all $n < \mathfrak{n}$.

If κ_3 denotes a number that is smaller than κ_1 and κ_2 then one will have:

$$1 - \frac{K^{(n)}(\varepsilon;\kappa)}{K^{(n)}(\varepsilon;0)} < h$$

in which $\kappa < \kappa_2$, and n is arbitrary.

However, the quantity h can be chosen to be as small as one desires, so the K will satisfy the third condition that was posed on pp. 32. A corresponding statement will be true for the L. The ideal gas will then satisfy all of the requirements that were stated on pp. 32.

With that, we have proved that temperature equilibrium will be true for them. We have not needed that proof. However, it was the goal of our considerations to show that for solid bodies, one does not need to demand agreement with gases in regard to certain functions that are associated with them, but only that they behave similarly at their maxima.

§ 9. Isopycnic processes.

At the start of our analyses, we considered the motions that an autonomous system can perform. They are distinguished by the fact that they lead to only phases of equal energy. With the addition of external heat, however, a second kind of variation is possible: The system can be transferred to another phase manifold. As long as such changes are permissible, ε will cease to be constant, but it will still be the same function of q and p. That is, the observable states will change, but not the mechanism.

Strictly speaking, the second case is possible only by varying the mechanism. We must temporarily couple the given system with another one and then imagine that this coupling is resolved again (1). The mechanism will then be the same at the beginning and the end of the process in question, but not at the intermediate times. The investigations of the previous paragraph show that the can very probably change during such processes; in fact, that would always be true when a second system with a different temperature is used. Since we can now ignore the intermediate phases, we can regard the process as being one that assigns a set of energies to the system while keeping the mechanism constant. Processes for which energy is supplied to or withdrawn from one and the same unchanged mechanism are called *isopycnic* (2).

A great number of times, the energy of a mechanism can be changed in that way. If that happens often enough and very slightly each time then one can propose that the ε^*

---

(1) **A. Einstein**, II, pp. 179.

(2) **A. Einstein**, *loc. cit.*

that is added to the fixed mechanism will vary continuously. Naturally, V will also vary with ε^*. From (39), for two corresponding changes of V and ε^*, one will have:

$$d \ln V = \frac{d\varepsilon^*}{\mathfrak{t}},$$

or (1):

$$\ln V = \int \frac{d\varepsilon^*}{\mathfrak{t}} + \text{const.} \tag{144}$$

(To be continued)

(Received on 25 May 1910)

---

(1) **J. W. Gibbs**, formula (485).