THE

PHYSICAL REVIEW

A journal of experimental and theoretical physics established by E. L. Nichols in 1893

SECOND SERIES, VOL. 117, NO. 5 MARCH 1, 1960

Theory of Irreversibility*

NOBUHIKO SAITÔ†

Department of Chemistry, University of Oregon, Eugene, Oregon

(Received September 4, 1959; revised manuscript received January 4, 1960)

We consider an isolated system which is exposed to external perturbations such as electric fields. The current as a linear response to a single pulse of an electric field, namely the aftereffect function, is a sum of terms periodic in time in a finite system, and does not vanish after a sufficiently long time. The transfer function is defined by the Laplace transform of the aftereffect function with respect to time and thus has poles on the imaginary axis. By applying the idea of image charges in two-dimensional electrostatic problems to a certain property of the transfer function, it is shown that, in the limit of a large system, the poles on the imaginary axis are replaced by poles in the left-half plane and consequently the aftereffect function can really vanish. This fact yields the various aspects of irreversibility. It is also shown that the complex conductivity is expressed in terms of the residues of the poles of the transfer function distributed continuously on the imaginary axis, and, in particular, that the static conductivity is proportional to the residue of the pole at the origin which never exists in a finite system but appears in an infinite system. General reciprocal relations are given for both irreversible and reversible thermodynamics, and their connection to Onsager's and Maxwell's relations are discussed. An expression for the entropy production is given and a new interpretation of the H-function is proposed.

I. INTRODUCTION

We know that all molecular or atomic processes are described through reversible mechanics, in principle through quantum mechanics, but under a certain approximation through classical mechanics. However, we also know that there are many irreversible processes and almost all of the phenomena in our world are more or less irreversible. Irreversibility is rather a universal phenomenon in the macroscopic world. Thus, since the days of Maxwell and Boltzmann, statistical mechanical theories of irreversible processes have been developed in various ways by many investigators, having in mind applications to dilute or dense gases, liquids, crystal lattices, electrons in metals, *et cetera.*

These theories required, to some extent, special assumptions in their development; for example, molecular chaos in the kinetic theory of gases, a time-smoothing or coarse-graining procedure in the theories of Kirkwood[1] and others, and the random phase assumption in a quantum statistical treatment. The cause of irreversibility was attributed to these assumptions, but the relation of these assumptions to reversible mechanics is not clear. Green[2] determined the coefficients of the Fokker-Planck equation describing time dependent phenomenon, assuming the Markoffian character of the time evolution of the system. Thus irreversibility was introduced from the outset. Yamamoto[3] gave Green's assumption a statistical mechanical basis by introducing the assumption of local ergodicity. Recently Kubo[4] developed a general theory of irreversible processes for a system under external fields. This theory and its extension[5] to a system under thermal disturbances are very important in the sense that they give rigorous expressions for various kinetic coefficients. But again Kubo gave only a suggestive discussion on the vanishing of the aftereffect function after a sufficiently long time.[6] Lax[7] adopted the same formulation taking into account the effect of the surrounding heat reservoir in a rather conventional way so as to make the aftereffect function vanish. But irreversibility is expected in an isolated system and the *ad hoc* introduction

* This work was supported in part by a research grant from the National Science Foundation.

† On leave from Department of Applied Physics, Waseda University, Tokyo, Japan.

[1] J. G. Kirkwood, J. Chem. Phys. 14, 180 (1946).

[2] M. S. Green, J. Chem. Phys. 20, 1281 (1952).

[3] T. Yamamoto, Progr. Theoret. Phys. (Kyoto) 10, 11 (1953).

[4] R. Kubo, J. Phys. Soc. Japan 12, 570 (1957).

[5] R. Kubo, M. Yokota, and S. Nakajima, J. Phys. Soc. Japan 12, 1203 (1957); S. Nakajima, Progr. Theoret. Phys. (Kyoto) 20, 948 (1958).

[6] See also, U. Fano, Phys. Rev. 96, 869 (1954).

[7] M. Lax, Phys. Rev. 109, 1921 (1958).

of a decay term does not give any solution to the problem of irreversibility.

In 1955, Van Hove[8] made an important contribution to the theory of irreversibility. He introduced a perturbation of a special character which is expected to appear in the case of gases, crystals, *et cetera*, and was able to derive from the Schrödinger equation a master equation which is regarded as a fundamental equation leading to irreversibility. This theory is, as far as the author knows, a unique existing theory which has solved the problem of irreversibility. However, this theory is mathematically too complicated to make intuitively clear the essential reason for irreversibility.

In Secs. II and III we show how irreversibility appears from reversible quantum mechanics in the limit of a large system under an external field such as an electric field without introducing any assumptions. We will follow Kubo's theory and limit ourselves to linear processes. In order to see that the aftereffect functions of certain mechanical quantities vanish for a large system we notice the following property in electrostatics: It is generally possible to find image charges equivalent to a continuous charge distribution on a plane surface of a conductor, as far as the electric field outside the conductor is concerned. A two-dimensional analogy of this property is applied to a certain property of the transfer function which is defined by the Laplace transform of the aftereffect function. It is possible to find the image charges in the limit of an infinitely large system contrary to the case of finite system. The appearance of image charges just corresponds to the appearance of irreversibility. In this way irreversibility is understood very clearly and rigorously starting from a purely mechanical point of view.

In Sec. IV, the problems of electric conduction and polarizability are treated as illustrative examples of irreversibility, and in Sec. V, general reciprocal relations are discussed. In Sec. VI, the entropy production is calculated, and a new definition of entropy and a new interpretation of the H function are proposed. In Sec. VII, the connection between the vanishing of the aftereffect function and irreversibility is discussed.

II. TRANSFER FUNCTION

The present theory can be developed either classically or quantum mechanically.[9] But for the sake of generality, we will treat the problem by quantum mechanics. The Hamiltonian $\mathcal{H}$ of the system under consideration is assumed to be composed of two parts, $\mathcal{H}=\mathcal{H}_0-AF(t)$, the unperturbed Hamiltonian $\mathcal{H}_0$ and the perturbation $-AF(t)$. The scalar function $F(t)$ is in general an arbitrary function of time representing an external field. For example, for charged particles under an electric field in the x direction, $F(t)$ is the electric field, and A is given by

$$A=\sum_i e_i x_i, \tag{2.1}$$

where e_i is the charge of the ith particle and x_i is its x coordinate. In the following, we will use the terminology employed in electrical conduction, whenever possible, for the sake of concreteness. For example, current is used for $\langle\dot{A}\rangle$, an average of the time derivative of A.

The system is assumed to be in a thermal equilibrium at time $t=0$ described by the canonical density matrix $\rho_0=\exp\beta(\Omega-\mathcal{H}_0)$, where $\beta=1/kT$ and $e^{-\beta\Omega}=\mathrm{Tr}\times\exp(-\beta\mathcal{H}_0)$. Then we let the system be isolated and apply an electric field $F(t)$. Hence, after $t=0$, the density matrix of the system varies with time according to

$$i\hbar\partial\rho/\partial t=[\mathcal{H},\rho],\quad \mathcal{H}=\mathcal{H}_0-AF(t). \tag{2.2}$$

In general ρ can be expanded in terms of the perturbation $F(t)$ as follows

$$\rho=\rho_0+\Delta_1\rho+\Delta_2\rho+\cdots, \tag{2.3}$$

$$\Delta_1\rho=-\frac{1}{i\hbar}\int_0^t \exp\left(-\frac{i(t-t')\mathcal{H}_0}{\hbar}\right)[A,\rho_0]\times\exp\left(\frac{i(t-t')\mathcal{H}_0}{\hbar}\right)F(t')dt', \tag{2.4}$$

$$\Delta_2\rho=-\frac{1}{i\hbar}\int_0^t \exp\left(-\frac{i(t-t')\mathcal{H}_0}{\hbar}\right)[A,\Delta_1\rho]\times\exp\left(\frac{i(t-t')\mathcal{H}_0}{\hbar}\right)F(t')dt', \tag{2.5}$$

as was shown by Kubo.[4] The average of an arbitrary dynamical quantity B is given by

$$\begin{aligned}\langle B(t)\rangle&=\mathrm{Tr}\rho B\\&=B_0+\langle\Delta_1 B\rangle+\cdots,\end{aligned} \tag{2.6}$$

where

$$\begin{aligned}B_0&=\mathrm{Tr}\rho_0 B,\\ \langle\Delta_1 B\rangle&=\mathrm{Tr}\Delta_1\rho B\\&=-\frac{1}{i\hbar}\mathrm{Tr}\int_0^t[A,\rho_0]B(t-t')F(t')dt',\end{aligned} \tag{2.7}$$

$$B(t)=\exp(it\mathcal{H}_0/\hbar)B\exp(-it\mathcal{H}_0/\hbar). \tag{2.8}$$

This was also shown by Kubo. Now we assume that the system is composed of a finite number of particles in a finite volume, and has discrete eigenvalues ϵ_n ($n=1$, $2,\cdots$) of $\mathcal{H}_0$. In such a representation, $\langle\Delta_1 B(t)\rangle$ is

[8] L. Van Hove, Physica **21**, 517 (1955); **23**, 441 (1957).

[9] In order to develop the theory in classical mechanics, we have only to adopt a system of eigenfunctions of the self-adjoint operator L introduced by Kirkwood[1] and defined by

$$iLg=(\mathcal{H}_0,g),$$

for an arbitrary function g, where the bracket is the Poisson bracket and $\mathcal{H}_0$ is the unperturbed Hamiltonian.

written as follows

$$\langle \Delta_1 B(t) \rangle = \int_0^t \phi(t-t')F(t')dt', \tag{2.9}$$

$$\phi(t) = -(1/i\hbar) \sum_{n,m} [A,\rho_0]_{nm} B_{mn} e^{-it(\epsilon_n - \epsilon_m)/\hbar}. \tag{2.10}$$

Equation (2.9) shows that $\phi(t)$ is a response of B at time t to the pulse field applied at $t=0$. However, as $\phi(t)$ is a sum of terms periodic in time, we cannot expect that $\phi(t) \to 0$ as $t \to \infty$. For example, the current does not disappear after the electric field is turned off. In a finite system, irreversibility in a strict sense never exists.

Define the transfer function $\Phi(z)$, which is familiar in electrical engineering, by the Laplace transform of the aftereffect function

$$\Phi(z) = \int_0^\infty \phi(t) e^{-zt} dt \tag{2.11}$$

$$= \sum_{n,m} K_{mn}/(z + i\omega_{mn}), \tag{2.12}$$

where

$$K_{mn} = -(1/i\hbar)[A,\rho_0]_{nm} B_{mn}, \tag{2.13}$$

$$\omega_{mn} = (\epsilon_n - \epsilon_m)/\hbar. \tag{2.14}$$

The integral of Eq. (2.11) has a meaning only for $\Re z > 0$ but we can define $\Phi(z)$ over all the z plane by analytic continuation using Eq. (2.12). The function $\Phi(z)$ thus defined has poles at $-i\omega_{mn}$ on the imaginary axis. As long as $\Phi(z)$ has poles on the imaginary axis, the aftereffect function never vanishes. To achieve the vanishing of the aftereffect function, all the poles of $\Phi(z)$ on the imaginary axis should disappear and, at the same time, new poles should appear on the left half of the z plane. This is possible in the limit of a large system as we shall show in the following.

First of all,we notice the following property of $\Phi(z)$:

$$\Phi(z) = \sum \frac{K_{mn}}{z + i\omega_{mn}} = \left(\frac{\partial}{\partial \alpha} - i \frac{\partial}{\partial \beta} \right) \sum_{m,n} K_{mn} \ln r_{mn}, \tag{2.15}$$

where $z = \alpha + i\beta$, and r_{mn} is the distance between z and a pole at $-i\omega_{mn}$ in the z plane. (See Fig. 1.) This form of $\Phi(z)$, Eq. (2.15), suggests an analogy, in which $\sum -K_{mn} \ln r_{mn}$ and $\Phi(z)$ are interpreted, respectively, as the two dimensional "electrostatic potential"[10] and the "electric field" at the point z in a plane where "charges" K_{mn} lie at $-i\omega_{mn}$.[11]

[10] In the following electrostatic terms with "..." are used for the electrostatic analogs described in the text, to avoid the confusion with terms pertinent to the problems of electrical conduction or polarization.

[11] Strictly speaking, the amount of "charge" at $-i\omega_{mn}$ is $-(1/2)K_{mn}$, but we disregard this unimportant factor $-(1/2)$ when we refer to the electrostatic analog. Furthermore, the "charge" is not necessarily real.

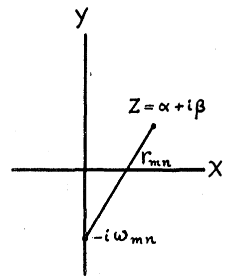

FIG. 1. Explanation of symbols.

The inverse transformation of Eq. (2.11) is given by

$$\phi(t) = \frac{1}{2\pi i} \int_c \Phi(z) e^{zt} dz, \tag{2.16}$$

where the path of the integration is along a line from $-i\infty$ to $+i\infty$ at the right-hand side of the imaginary axis. Thus, the value of $\Phi(z)$ used for the calculation of (2.16) is that for $\Re z > 0$.

In the limit of an infinitely large system,[12] the eigenvalues ϵ_n are presumably continuous, and the poles of $\Phi(z)$ are distributed continuously on the imaginary axis. In the electrostatic analogy, these continuously distributed poles are regarded as "charges" continuously distributed on the imaginary axis. We denote this continuous distribution of "charges" on the imaginary axis by $f(p)$, where p is the distance from the origin. In this electrostatic analogy, we assume that we can find a discrete or continuous distribution $F(-x, y)$ of "charges" in the left half of the z plane which is equivalent to this $f(p)$, in so far as the "electrostatic potential" in the right half of the z plane is concerned. More precisely, we replace the left half of the z plane by a "conductor" and assume that we can put an appropriate discrete or continuous "charge" distribution $-F(x,y)$ on the right-hand side of the "conductor" so as to make the distribution of induced "charges" on the surface of the "conductor" the same as the given distribution $f(p)$. The image "charge" distribution of $-F(x,y)$ about the imaginary axis is just $F(-x, y)$. Thus we have

$$\sum_{m,n} K_{mn} \ln r_{mn} \to \int_{-\infty}^{+\infty} f(p) \ln r_p dp$$

$$= \int\int F(-x, y) \ln r_{-x,y} dx dy, \tag{2.17}$$

$$r_p^2 = \alpha^2 + (\beta - p)^2,$$

$$r_{-x,y}^2 = (\alpha + x)^2 + (\beta - y)^2,$$

[12] By an infinitely large system, we mean that the volume of the system is infinitely large. However, in a system of very many interacting particles ($\sim 10^{20}$) in a finite volume, the eigenvalues are distributed practically continuously. We can include this case also in the discussion of infinitely large systems. This point of view is connected with the time-smoothing or coarse-graining procedure (see Sec. VII).

or

$$\Phi(z)=\sum\frac{K_{mn}}{z+i\omega_{mn}}\rightarrow\int_{-\infty}^{+\infty}\frac{f(p)}{z-ip}dp \tag{2.18}$$

$$=\int_0^\infty dx\int_{-\infty}^\infty dy\, F(-x,y)/[z-(-x+iy)]. \tag{2.19}$$

Substituting $\Phi(z)$ from Eq. (2.19) into Eq. (2.16), we have

$$\phi(t)=\int_0^\infty dx\int_{-\infty}^\infty dy\, F(-x,y)e^{-(x-iy)t}. \tag{2.20}$$

If $F(-x,y)$ does not allow any "charge" on the imaginary axis at $x=0$, we get $\phi(t)\rightarrow 0$ for $t\rightarrow\infty$. This is just what we want to show as a manifestation of irreversibility. Mathematically, $\Phi(z)$ of Eq. (2.18), which is obtained as a limit of $K_{mn}/(z+i\omega_{mn})$ has a branch cut on the imaginary axis. The function defined by Eq. (2.19) is the analytic continuation of the function (2.18), originally defined in the right half of the z plane, into the left half of the z plane on a Riemannian surface contiguous to the right half of the z plane across this branch cut. If the function $\Phi(z)$ thus defined has a pole (i.e., image "charge") on the (negative) real axis, we have a Debye-type dispersion or absorption, and if it has a pole apart from the real axis, we have a Lorentz-type resonance absorption. In general, we have a superposition of many Debye and Lorentz-type dispersions.[13]

III. PROPERTIES OF IMAGE "CHARGE"

From the definition Eq. (2.13), we have $K_{mn}=K_{nm}{}^*$ and $\omega_{mn}=-\omega_{nm}$. Therefore, we have $f(p)=f(-p)^*$ in the limit of the continuous distribution. Therefore, we have

$$F(-x,y)=F(-x,-y)^*. \tag{3.1}$$

This property is consistent with the physical requirement that the aftereffect function $\phi(t)$ is a real function, as is easily shown in Eq. (2.20).

In Eqs. (2.18) and (2.19), we put $z=\alpha+i\omega$, and let $\alpha\rightarrow+0$. Then, since

$$\lim_{\alpha\rightarrow+0}\frac{1}{\alpha+i(\omega-p)}=\pi\delta(\omega-p)-i\mathrm{P}\frac{1}{\omega-p}, \tag{3.2}$$

where P denotes the principal value of Cauchy, we have

$$\lim_{\alpha\rightarrow+0}\Phi(\alpha+i\omega)=\pi f(\omega)-i\mathrm{P}\int_{-\infty}^\infty\frac{f(p)}{\omega-p}dp \tag{3.3}$$

$$=\int_0^\infty dx\int_{-\infty}^{+\infty}dy\frac{F(-x,y)}{x+i(\omega-y)}. \tag{3.4}$$

This function will be called the generalized admittance

If $f(\omega)$ and $F(-x,y)$ are real, we have

$$\pi f(\omega)=\int_0^\infty dx\int_{-\infty}^{+\infty}dy\frac{x}{x^2+(\omega-y)^2}\times F(-x,y), \tag{3.5}$$

$$\mathrm{P}\int_{-\infty}^\infty\frac{f(p)}{\omega-p}dp=\int_0^\infty dx\int_{-\infty}^\infty dy\frac{\omega-y}{x^2+(\omega-y)^2}F(-x,y). \tag{3.6}$$

These relations (3.5) and (3.6) hold in general even if $f(\omega)$ and $F(-x,y)$ are not real, as is seen on differentiating Eq. (2.17) with regard to α and β, respectively. Integrating $f(\omega)$ in Eq. (3.5) with respect to ω, we have

$$\int_{-\infty}^{+\infty}f(\omega)d\omega=\int_0^\infty dx\int_{-\infty}^\infty dy\, F(-x,y). \tag{3.7}$$

Equation (3.5) is the electrostatic expression for the induced "charge" on the surface of the "conductor" due to a "charge" distribution $-F(x,y)$, and Eq. (3.7) shows that the total amount of the induced "charge" is numerically equal to the total amount of "charge" distributed outside the "conductor." That is, Eqs. (3.5) and (3.6) are, respectively, interpreted as the x and y components of the "electric field" on the imaginary axis due to "charges" on this axis or image "charges" in the left half of the z plane. As we will see later, if we take $\dot{A}$ for B, $F(-x,y)$ is real and if we take A for B, the corresponding $F(-x,y)$ is pure imaginary. In the former case, $\Phi(i\omega)$ corresponds to the complex conductivity, and in the latter case, to the complex polarizability. In either case, Eq. (3.5) stands for the dissipative part of the generalized admittance (real part of the complex conductivity or imaginary part of the complex polarizability), and Eq. (3.6) corresponds to its storage part [imaginary part of the complex conductivity or real part of the complex polarizability, see Eqs. (4.14) and (4.25)]. This fact, together with the electrostatic interpretation given above, is very important to the understanding of the nature of irreversibility as seen in Sec. IV.

Equation (3.5) shows that $f(p)$ is uniquely determined, if $F(-x,y)$ is given, but the inverse is not true, as is easily seen in many examples in electrostatics. Unfortunately, it is still unknown how to obtain an $F(-x,y)$ from a given $f(p)$. Furthermore, it is also unknown under what conditions an $F(-x,y)$ exists,

[13] See N. Saitô, *Introduction to Polymer Physics* (Syokabo, Tokyo, 1958), p. 219 (in Japanese).

In the case where a function $\phi(t)$ has a continuous spectrum, such as expressed by

$$\phi(t)=\int_0^\infty A(\omega)e^{i\omega t}d\omega,$$

we can see in many examples of Fourier transforms that $\phi(t)$ vanishes as $t\rightarrow\infty$. G. O. Hultgren of the California Institute of Technology showed that the electric field correlation in a hohlraum has the form of the above equation with

$$A(\omega)=\omega^3/(e^{\hbar\omega/kT}-1),$$

and $\phi(t)\rightarrow 0$ as $t\rightarrow\infty$, but this decay is of an overdamped type (private communication).

for a given $f(p)$, such that it satisfies Eq. (3.5) and vanishes in the neighborhood of the imaginary axis. However, as we will see later in Eqs. (4.14) and (4.25), we do not need to know $F(-x, y)$ to obtain the generalized admittance, such as the electric conductivity or polarizability, but instead we have only to know the existence of an $F(-x, y)$ with the above-mentioned properties. Furthermore, Eqs. (3.3) and (3.4) suggest that the existence of a $\lim_{\alpha\to+0}\Phi(\alpha+i\omega)$, such as the one given by Eq. (3.3), seems to be equivalent to the existence of an $F(-x, y)$ of the required character, because, in this case, we can make an analytic continuation of the $\Phi(z)$ defined in the right half of the z plane across the cut on the imaginary axis and the singularities of $\Phi(z)$ thus defined in the left half of the z plane give the distribution $F(-x, y)$. Thus we have the following statement. If a singularity is isolated, the corresponding "image charge" is unique. However, if a singularity is a branch point, a cut starting from this point yields the continuous "image charge" distribution. But the cut is, to some extent, arbitrarily deformable and thus the corresponding "image charge" distribution is not unique except at the branch point.

IV. ELECTRIC CONDUCTIVITY AND POLARIZABILITY

The foregoing discussion can be applied to any dynamical quantity B. In this section, we will study the average behavior of $I=\dot{A}$ and A, which correspond to a current and a polarization in the electric case, respectively. We will write J_{mn} instead of K_{mn} defined by (2.13) in the case of the current:

$$J_{mn}=(i/\hbar)[A,\rho_0]_{nm}I_{mn}, \tag{4.1}$$

where I is determined by Heisenberg's equation of motion, and its mn element is given by

$$I_{mn}=(\dot{A})_{mn}=(1/i\hbar)[A,\mathcal{H}_0]_{mn}=(1/i\hbar)A_{mn}(\epsilon_n-\epsilon_m).$$

Hence, we have

$$J_{mn}=\hbar^{-2}|A_{nm}|^2(\rho_{0m}-\rho_{0n})(\epsilon_n-\epsilon_m)=J_{nm}, \tag{4.2}$$

and $J_{mn}=J_{nm}$ is non-negative because

$$(\epsilon_n-\epsilon_m)(\rho_{0m}-\rho_{0n})=(\epsilon_n-\epsilon_m)(e^{-\beta\epsilon_m}-e^{-\beta\epsilon_n})e^{\beta\Omega}\geq 0. \tag{4.3}$$

The transfer function for I is given by

$$\Phi_I(z)=\sum J_{mn}/(z+i\omega_{mn}). \tag{4.4}$$

Now, on letting the system become infinitely large, we must keep the electron density constant, and we define[14] as a limit

$$\bar{\Phi}_I(z)=\lim_{\substack{V\to\infty,\, N\to\infty\\ N/V=\text{const}}}\frac{1}{V}\sum_{m,n}J_{mn}/(z+i\omega_{mn}). \tag{4.5}$$

[14] In the following a bar over a symbol is used to denote quantities referring to a unit volume in an infinitely large system such as defined by Eq. (4.5).

If the system has N electrons in a volume V whose linear dimension is L, J_{mn} given by Eq. (4.2) is of the order of $(NL)^2$, while the separation of two successive energy levels $\epsilon_{n+1}-\epsilon_n$ is of the order of $1/NL^2$. Hence, in the limit of an infinitely large system, the integral form of (4.5), which may be written as[15]

$$\bar{\Phi}_I(z)=\int_{-\infty}^{\infty}[\bar{f}_I(p)/(z-ip)]dp, \tag{4.6}$$

where $\bar{f}_I(p)$ is of the order of $(1/V)[(NL)^2/NL^2]=N/V$, and has a definite limit.

In Eq. (4.6) $\bar{f}_I(p)$ is non-negative and

$$\bar{f}_I(p)=\bar{f}_I(-p),\quad \bar{f}_I(p)\geq 0, \tag{4.7}$$

because of Eqs. (4.2) and (4.3). Thus the image "charge" distribution $\bar{F}_I(-x, y)$ has the property

$$\bar{F}_I(-x, y)=\bar{F}_I(-x, -y), \tag{4.8}$$

and from the general property described in Eq. (3.1), $\bar{F}_I(-x, y)$ is found to be real. In this way, Eq. (4.6) can be written for $\Re z>0$

$$\bar{\Phi}_I(z)=\int_0^{\infty}dx\int_{-\infty}^{\infty}dy\frac{\bar{F}_I(-x, y)}{z+x-iy}. \tag{4.9}$$

Thus we have

$$\begin{aligned}\langle\Delta\bar{I}(t)\rangle&=\int_0^t\bar{\phi}_I(t-t')F(t')dt',\\ \bar{\phi}_I(t)&=\frac{1}{2\pi i}\int\bar{\Phi}_I(z)e^{zt}dz\\ &=\int_{-\infty}^{\infty}\bar{f}_I(p)e^{ipt}dp\\ &=\int_0^{\infty}dx\int_{-\infty}^{\infty}dy\,\bar{F}_I(-x, y)e^{-(x-iy)t}.\end{aligned} \tag{4.10}$$

For a periodic field $F(t)=F_0e^{i\omega t}$, we have for $t\to\infty$,

$$\langle\Delta\bar{I}(t)\rangle=\left(\int_0^{\infty}dx\int_{-\infty}^{\infty}dy\frac{\bar{F}_I(-x, y)}{i\omega+x-iy}\right)F_0e^{i\omega t}. \tag{4.11}$$

This yields immediately for the complex conductivity

$$\sigma(i\omega)=\int_0^{\infty}dx\int_{-\infty}^{\infty}dy\frac{\bar{F}_I(-x, y)}{i\omega+x-iy}. \tag{4.12}$$

[15] In the case when the distribution of poles still remains discrete in the limit of an infinitely large system keeping the density constant, Eq. (4.5) has a limiting form

$$\bar{\Phi}_I(z)=\Sigma_l\,\bar{f}_I(p_l)/(z-ip_l), \tag{4.6'}$$

instead of Eq. (4.6).

Comparing (4.12) with (4.9) we have

$$\sigma(i\omega)=\lim_{\alpha\to+0}\bar{\Phi}_I(\alpha+i\omega)=\lim_{\alpha\to+0}\int_0^\infty\bar{\phi}_I(t)e^{-(\alpha+i\omega)t}dt, \tag{4.13}$$

because the value of $\bar{\Phi}_I(z)$ on the imaginary axis is to be taken as an analytic continuation from the right side of the z plane. The introduction of the convergence factor $e^{-\alpha t}$ in Eq. (4.13) is very natural in contrast to the usual introduction from physical considerations (see reference 4).

Equations (4.12) or (4.13) can be rewritten, according to Eq. (3.3), as

$$\sigma(i\omega)=\pi f_I(\omega)-i\mathrm{P}\int_{-\infty}^{\infty}[f_I(p)/(\omega-p)]dp. \tag{4.14}$$

Therefore, $\Re\sigma(i\omega)$ is positive in general as shown in Eq. (4.7), and as it should be. Especially the static conductivity $\sigma(0)$ is given by the "charge" at the origin $\bar{f}_I(0)$ [16]

$$\sigma(0)=\pi\bar{f}_I(0). \tag{4.15}$$

On the other hand, in a finite system $J_{mn}=0$ when $\omega_{mn}=0$, namely, when $\epsilon_n=\epsilon_m$. This means that in a finite system there is no "charge" at the origin. In an infinite system, the "charge" distribution becomes continuous and a positive "charge" can appear as a limit. This positive "charge" gives the static conductivity.[17]

Now we take A instead of I as a dynamical quantity and investigate its response to an electric field. The transfer function for $\langle\Delta A(t)\rangle$ is given by

$$\begin{aligned}\Phi_A(z)&=\sum D_{mn}/(z+i\omega_{mn}),\\ D_{mn}&=(i/\hbar)[A,\rho_0]_{nm}A_{mn}\\ &=(i/\hbar)|A_{nm}|^2(\rho_{0m}-\rho_{0n})\\ &=-D_{nm}=i\hbar J_{mn}/(\epsilon_n-\epsilon_m),\end{aligned} \tag{4.17}$$

or

$$\Phi_A(z)=\sum_{m,n}\frac{i(J_{mn}/\omega_{mn})}{z+i\omega_{mn}}. \tag{4.18}$$

Therefore, in an infinite system, the "charge" on the imaginary axis $\bar{f}_A(p)$ is related to $\bar{f}_I(p)$ by

$$ip\bar{f}_A(p)=\bar{f}_I(p). \tag{4.19}$$

Thus, if our system has a finite conductivity, $[\bar{f}_I(0)=\text{finite}]$, then $\bar{f}_A(0)$ is infinite. This means that the "charge" at the origin is infinite and cannot be replaced by finite image "charges" in the left half of the z plane. This singular part of the "charge" distribution around the origin is expressed by $-i\bar{f}_I(0)/p$ and its contribution to the aftereffect function $\bar{\phi}_A(t)$ is given by

$$\int_{-\infty}^{\infty}dp\frac{1}{2\pi i}\int\frac{-if_I(0)/p}{z-ip}e^{zt}dz=-i\int_{-\infty}^{\infty}f_I(0)\frac{e^{ipt}}{p}dp=\pi f_I(0),\quad t>0. \tag{4.20}$$

The contribution to $\bar{\phi}_A(t)$ from the remaining "charge" distribution on the imaginary axis may completely vanish after a long time. Thus we have for $t\to\infty$

$$\bar{\phi}_A(t)\to\pi\bar{f}_I(0)=\sigma(0). \tag{4.21}$$

This gives an average displacement of electrons in a conductor due to the pulse of an electric field, and is just the integral of $\bar{\phi}_I(t)$ with respect to t from 0 to ∞,

$$\int_0^\infty\bar{\phi}_I(t)dt=\int_0^\infty\left[\int_0^\infty dx\int_{-\infty}^{\infty}dy\,\bar{F}_I(-x,y)e^{-(x-iy)t}\right]dt=\int_0^\infty dx\int_{-\infty}^{\infty}dy\,\bar{F}_I(-x,y)/(x-iy)=\sigma(0), \tag{4.22}$$

where use is made of Eq. (4.10).

Again we return to Eqs. (4.17) and (4.19). We see that D_{mn} is pure imaginary. Thus, the "charge" distribution in an infinitely large system $\bar{f}_A(p)$ can be written equal to $i\bar{g}_A(p)$, where $\bar{g}_A(p)$ is real and

$$\bar{g}_A(p)=-\bar{g}_A(-p). \tag{4.23}$$

If this "charge" distribution is a continuous function of p, it can be replaced by an image "charge" distribution $\bar{F}_A(-x,y)$ (in case of an insulator). $\bar{F}_A(-x,y)$ is also pure imaginary and is put equal to $i\bar{G}_A(-x,y)$. Thus,

$$\bar{\Phi}_A(z)=\int_{-\infty}^{\infty}[i\bar{g}_A(p)/(z-ip)]dp=\int_0^\infty dx\int_{-\infty}^{\infty}dy\frac{i\bar{G}_A(-x,y)}{z+x-iy}. \tag{4.24}$$

In the same way as in Eq. (4.13) the complex polarizability is given by, referring to Eq. (3.3),

$$p(i\omega)=\lim_{\alpha\to+0}\bar{\Phi}(\alpha+i\omega)=\mathrm{P}\int_{-\infty}^{\infty}\frac{\bar{g}_A(p)}{\omega-p}dp+i\pi\bar{g}_A(\omega)=\int_0^\infty dx\int_{-\infty}^{\infty}dy\frac{\omega-y+ix}{x^2+(\omega-y)^2}\bar{G}_A(-x,y). \tag{4.25}$$

---

[16] A similar expression for the static conductivity was given by H. Nakano, Progr. Theoret. Phys. (Kyoto) **15**, 77 (1956); **17**, 145 (1957),

$$\sigma(0)=(\pi\hbar/kT)\sum_{\substack{m,n\\(\epsilon_n=\epsilon_m)}}\rho_{0n}|I_{mn}|^2. \tag{4.16}$$

This is just equal to $\pi\sum_{mn}J_{mn}\delta(\omega_{mn})$.

[17] In this connection, the present theory is seen to be quite similar to the Yang-Lee's theory of condensation [C. N. Yang and T. D. Lee, Phys. Rev. **87**, 404, 410 (1952); see also, T. L. Hill, *Statistical Mechanics* (McGraw-Hill Book Company, New York, 1956), p. 174]. An application of the idea of image charges to the condensation problem will be discussed in N. Saitô, J. Chem. Phys. (submitted).

In this case, we have $\bar{g}_A(0)=0$, because of Eq. (4.23). Hence, this system has no static conductivity, but it may have a dynamic conductivity in general.

Again from Eq. (4.19) we can express $\bar{\Phi}_I(z)$ as

$$\bar{\Phi}_I(z)=\int_{-\infty}^{\infty}\frac{f_I(p)}{z-ip}dp=\int_{-\infty}^{\infty}\frac{ipf_A(p)}{z-ip}dp$$
$$=z\int_{-\infty}^{\infty}\frac{f_A(p)}{z-ip}dp=z\bar{\Phi}_A(z), \tag{4.26}$$

or in particular we have

$$\sigma(i\omega)=i\omega p(i\omega). \tag{4.27}$$

This is a well-known relation.

Whether the system under consideration is a conductor or not depends on the distribution of "charges" on the imaginary axis. This distribution itself comes from the structure of the Hamiltonian $\mathcal{H}_0$, but we do not enter into details of this problem.

If the system does not have any irreversibility, the "charge" distribution $\bar{f}_A(p)$ on the imaginary axis for A is not continuous, but remains discrete in the limit of an infinitely large system. Thus, the transfer function $\bar{\Phi}_A(z)$ has a form given by Eq. (4.6′) and the polarizability is given by

$$p(i\omega)=\lim_{\alpha\to+0}\bar{\Phi}_A(\alpha+i\omega)$$
$$=\mathrm{P}\sum_l\frac{\bar{g}_A(p_l)}{\omega-p_l}+i\pi\sum_l\bar{g}_A(p_l)\delta(\omega-p_l). \tag{4.28}$$

An electrostatic interpretation of this result is given by Eqs. (3.5) and (3.6). The first term of Eq. (4.28) is the y component of the "electric field" at the point $i\omega$ on the imaginary axis. Even if the distribution of "charge" is discrete, this component "field" does not vanish, while the x component of the "field" on the imaginary axis, which gives the dissipative part of the polarizability, does.

Finally, one point should be noted in connection with the previous discussion. If we have several polarizabilities (we use the terminology used in electrical problems, but they may be elastic moduli in mechanical problems), and if there is a relationship between them which is known to be valid for a reversible case, this relationship still holds for an irreversible case, provided that the polarizabilities are introduced in a complex form. This is because the validity of the relationship in a reversible case originates in a relationship between D_{mn}'s for different polarizabilities, and the relationship in an irreversible case is obtained as its limit. For example, in the theory of elasticity, we have a relationship among modulus M entering in the expression for the velocity of the longitudinal wave, the bulk modulus κ and the rigidity G

$$M=\kappa+(4/3)G.$$

The same type of equations hold for visco-elastic materials if we use complex moduli $M(i\omega)$, $\kappa(i\omega)$, and $G(i\omega)$,

$$M(i\omega)=\kappa(i\omega)+(4/3)G(i\omega).$$

Of course, this relationship is also proved by phenomenological considerations, as shown by Marvin et al.[18]

V. RECIPROCAL RELATIONS

In a case where the Hamiltonian has several terms representing external forces,

$$\mathcal{H}=\mathcal{H}_0-\textstyle\sum_l A_lF_l(t), \tag{5.1}$$

the treatment is a simple extension of that of the previous sections. Thus, the density matrix is expanded as

$$\rho=\rho_0+\Delta_1\rho+\Delta_2\rho+\cdots, \tag{5.2}$$

where

$$\Delta_1\rho=\textstyle\sum_l\Delta_1\rho_l(t), \tag{5.3}$$

and

$$\Delta_1\rho_l=\frac{i}{\hbar}\int_0^t\exp\left(\frac{-i(t-t')\mathcal{H}_0}{\hbar}\right)[A_l,\rho_0]\times\exp\left(\frac{i(t-t')\mathcal{H}_0}{\hbar}\right)F_l(t')dt'. \tag{5.4}$$

The average deviation $\langle\Delta A_h\rangle$ and $\langle\Delta I_h\rangle$ from the equilibrium $\langle A_h\rangle$ and $\langle I_h\rangle=\langle\dot{A}_h\rangle$ are, respectively,[19]

$$\langle\Delta_1A_h\rangle=\textstyle\sum_l\langle\Delta_1A_{l\to h}\rangle, \tag{5.5}$$

$$\langle\Delta_1A_{l\to h}\rangle=\mathrm{Tr}\Delta_1\rho_lA_h=\sum_{n,m}\int_0^tD_{nm}{}^{(l\to h)}e^{-i(t-t')(\epsilon_n-\epsilon_m)/\hbar}F_l(t')dt', \tag{5.6}$$

where

$$D_{nm}{}^{(l\to h)}=(i/\hbar)[A_l,\rho_0]_{nm}(A_h)_{mn}=(i/\hbar)(A_l)_{nm}(A_h)_{mn}(\rho_{0m}-\rho_{0n})=-D_{mn}{}^{(h\to l)}, \tag{5.7}$$

and

$$(\Delta_1I_h)=\textstyle\sum_l\langle\Delta_1I_{l\to h}\rangle, \tag{5.8}$$

$$\langle\Delta_1I_{l\to h}\rangle=\mathrm{Tr}\Delta_1\rho_lI_h=\sum_{n,m}\int_0^tJ_{nm}{}^{(l\to h)}e^{-i(t-t')(\epsilon_n-\epsilon_m)/\hbar}F_l(t')dt', \tag{5.9}$$

where

$$J_{nm}{}^{(l\to h)}=\hbar^{-2}(A_l)_{nm}(A_h)_{mn}(\epsilon_n-\epsilon_m)(\rho_{0m}-\rho_{0n})=J_{mn}{}^{(h\to l)}=J_{nm}{}^{(h\to l)*}. \tag{5.10}$$

[18] R. S. Marvin, R. Aldrich, and H. S. Sack, J. Appl. Phys. 25, 1213 (1954).

[19] $l\to h$ means that F_l is an excitation and $\Delta_1\rho_h$, A_h, or I_h is its response.

The reciprocity expressed by Eq. (5.7) or (5.10) leads to various reciprocal relations. If a dynamical quantity, say A_l, is transformed to $\tilde{A}_l$ under time and magnetic field (H) reversal, $(A_l)_{mn}$ is transformed into $(\tilde{A}_l)_{nm}$, and, by virtue of the reversibility of mechanics in a finite system (microscopic reversibility), we have from Eq. (5.10)

$$\begin{aligned} J_{mn}^{(l\to h)}(A_l,A_h,H) &= J_{nm}^{(l\to h)}(\tilde{A}_l,\tilde{A}_h,-H) \\ &= J_{mn}^{(h\to l)}(\tilde{A}_l,\tilde{A}_h,-H). \end{aligned} \tag{5.11}$$

Further, if A_l is either an "α variable" or a "β variable" in the terminology of Casimir[20] we can write

$$A_l=\epsilon_l\tilde{A}_l, \tag{5.12}$$

where ϵ_l is 1 for an α variable and -1 for a β variable. Thus, we have

$$J_{mn}^{(l\to h)}(A_l,A_h,H)=\epsilon_l\epsilon_h J_{mn}^{(h\to l)}(A_l,A_h,-H). \tag{5.13}$$

In the limit of an infinitely large system, we have

$$\bar{f}_I^{(l\to h)}(p,H)=\epsilon_l\epsilon_h\bar{f}_I^{(h\to l)}(p,-H). \tag{5.14}$$

In the same way we have

$$\bar{f}_I^{(l\to h)}(p,H)=\epsilon_l\epsilon_h\bar{f}_I^{(h\to l)}(p,-H). \tag{5.15}$$

These two relations are not independent, since we have

$$J_{mn}^{(l\to h)}=i\omega_{mn}D_{mn}^{(l\to h)}, \tag{5.16}$$

or

$$\bar{f}_I^{(l\to h)}(p,H)=ip\bar{f}_A^{(l\to h)}(p,H), \tag{5.17}$$

which is a generalization of Eq. (4.19). Neither $\bar{f}_I^{(l\to h)}$ nor $\bar{f}_A^{(l\to h)}$ is necessarily real or pure imaginary. Their symmetric and antisymmetric parts defined by, for example,

$$\begin{aligned} \mathrm{Sym}\bar{f}_I^{(l\to h)}(\omega) &= \tfrac{1}{2}[\bar{f}_I^{(l\to h)}(\omega)+\bar{f}_I^{(h\to l)}(\omega)], \\ \mathrm{Ant}\bar{f}_I^{(l\to h)}(\omega) &= \tfrac{1}{2}[\bar{f}_I^{(l\to h)}(\omega)-\bar{f}_I^{(h\to l)}(\omega)] \end{aligned} \tag{5.18}$$

are related by

$$\mathrm{Sym}\bar{f}_I^{(l\to h)}(\omega)=i\omega\,\mathrm{Sym}\bar{f}_A^{(l\to h)}(\omega), \tag{5.19}$$

$$\mathrm{Ant}\bar{f}_I^{(l\to h)}(\omega)=i\omega\,\mathrm{Ant}\bar{f}_A^{(l\to h)}(\omega), \tag{5.20}$$

and it follows from Eq. (5.10) that $\mathrm{Sym}\bar{f}_I^{(l\to h)}(\omega)$ and $\mathrm{Ant}\bar{f}_A^{(l\to h)}(\omega)$ are real and even functions of ω, while $\mathrm{Ant}\bar{f}_I^{(l\to h)}(\omega)$ and $\mathrm{Sym}\bar{f}_A^{(l\to h)}(\omega)$ are pure imaginary and odd functions of ω.

The generalized admittances corresponding to $\bar{f}_I^{(l\to h)}$ and $\bar{f}_A^{(l\to h)}$, which will be denoted, respectively, by $\sigma^{(l\to h)}(i\omega,H)$ and $p^{(l\to h)}(i\omega,H)$, have the forms of Eqs. (4.14) and (4.25), respectively, and have from Eqs. (5.14) and (5.15) the following properties:

$$\begin{aligned} \sigma^{(l\to h)}(i\omega,H) &= \epsilon_l\epsilon_h\sigma^{(h\to l)}(i\omega,-H), \\ p^{(l\to h)}(i\omega,H) &= \epsilon_l\epsilon_h p^{(h\to l)}(i\omega,-H). \end{aligned} \tag{5.21}$$

These are Onsager's reciprocal relations.[21] They are not independent since

$$\sigma^{(l\to h)}(i\omega,H)=i\omega p^{(l\to h)}(i\omega,H). \tag{5.22}$$

This relation is proved by making use of Eq. (5.17), and

$$\int_{-\infty}^{\infty} f_A^{(l\to h)}(p)dp=0. \tag{5.23}$$

Equation (5.23) is obtained by observing from Eq. (5.7) that[22]

$$\begin{aligned} \sum_{n,m} D_{mn}^{(l\to h)} &= (i/\hbar)\,\mathrm{Tr}[A_l,\rho_0]A_h \\ &= (i/\hbar)\,\mathrm{Tr}\rho_0[A_h,A_l]=0, \end{aligned} \tag{5.24}$$

because, in general, A_h and A_l commute.

As we have already mentioned, if the system is not subject to any irreversible processes, the poles on the imaginary axis of the relevant transfer function are discrete. Nevertheless, the polarizabilities can be defined and are real quantities. Among them relations hold, such as Eq. (5.21), which are similar to Maxwell's relation in thermodynamics, except for reversal of the magnetic field.[23] Further, in the usual demonstration of the Onsager reciprocity, one makes use of microscopic reversibility and the regression of fluctuations which is assumed to be described by a phenomenological relation in irreversible processes. However, the connection between these two controversial properties is not clear. In the above proof, microscopic reversibility is introduced in the mechanical stage into a finite system to obtain reciprocal relations [Eqs. (5.14) and (5.15)]. These reciprocal relations lead to the Onsager relation when irreversibility appears by making the system infinitely large.

Finally, it is noted that as mentioned above, $\sigma^{(l\to h)}$ is given by

$$\sigma^{(l\to h)}(i\omega)=\pi f_I^{(l\to h)}(\omega)-i\mathrm{P}\int_{-\infty}^{\infty}\frac{\bar{f}_I^{(l\to h)}(p)}{\omega-p}dp. \tag{5.26}$$

Thus, the real part of $\mathrm{Sym}\sigma^{(l\to h)}(i\omega)$ is equal to $\pi\,\mathrm{Sym}\bar{f}_I^{(l\to h)}(\omega)$ and is responsible for the irreversible production of entropy, as will be shown in Sec. VI. This

[20] H. B. G. Casimir, Revs. Modern Phys. **17**, 343 (1945).

[21] L. Onsager, Phys. Rev. **37**, 405; **38**, 2265 (1931).

[22] See reference 24 for the transformation from the second expression to the third.

[23] Maxwell's relation in equilibrium thermodynamics is proved to hold under mechanical perturbations in the following way,

$$\begin{aligned} p^{(l\to h)}(H) &= \frac{\partial}{\partial F_l}\mathrm{Tr}\left[\rho(\mathcal{H})\frac{\partial\mathcal{H}(H)}{\partial F_h}\right] \\ &= \mathrm{Tr}\rho(\mathcal{H})\frac{\partial\mathcal{H}(H)}{\partial F_l\partial F_h}+\mathrm{Tr}\left[\frac{\partial\rho}{\partial\mathcal{H}}\frac{\partial\mathcal{H}}{\partial F_l}\frac{\partial\mathcal{H}}{\partial F_h}\right] \\ &= p^{(h\to l)}(H). \end{aligned} \tag{5.25}$$

Equation (5.25) holds only for static case ($\omega=0$) with no irreversible elements, and is also derived directly from Eq. (5.7). It is to be noted that by using microscopic reversibility in addition to Eq. (5.25) we get $p^{(l\to h)}(H)=\epsilon_l\epsilon_h p^{(h\to l)}(-H)$, a special case of (5.21).

is also consistent with the electrostatic interpretation given in Sec. IV that irreversibility is always connected with the x component of the "electric field." Furthermore, we can show by making use of Eq. (5.10) that

$$\begin{aligned}\Re \textstyle\sum_l \sum_h \sigma^{(l\to h)}(\omega) X_l X_h &= \Re \textstyle\sum_l \sum_h \mathrm{Sym}\sigma^{(l\to h)}(\omega) X_l X_h \\ &= \pi \textstyle\sum_l \sum_h \mathrm{Sym}\bar{f}_I^{(l\to h)}(\omega) X_l X_h \\ &= \pi \textstyle\sum_l \sum_h \bar{f}_I^{(l\to h)}(\omega) X_l X_h \geq 0 \end{aligned} \qquad (5.27)$$

is always positive, where the X_l's are arbitrary real quantities. Equation (5.27) will be used in Sec. VI to prove the positive definiteness of the entropy production.

VI. ENTROPY AND H FUNCTION

For a while, we assume that the Hamiltonian has a single perturbation term representing an external force, as in Sec. IV. The current is given by Eq. (4.10), and we can write the rate of the irreversible production of entropy per unit volume $(dS/dt)_{\mathrm{irr}}$ as

$$(dS/dt)_{\mathrm{irr}} = \langle \Delta \bar{I}(t) \rangle F(t)/T, \qquad (6.1)$$

according to the thermodynamics of irreversible processes, T being the absolute temperature. For a periodic field $F(t) = F_0 e^{i\omega t}$, both factors of Eq. (6.1) must be taken as their real parts. Namely, the average rate of the entropy production over one period after a sufficiently long time, is given by

$$\left(\frac{dS}{dt}\right)_{\mathrm{irr}} = \frac{1}{T}\frac{1}{2}\Re\langle \Delta I(t)\rangle^* F(t) = \frac{1}{2T}\Re\sigma(i\omega)F_0^2, \qquad (6.2)$$

where $\langle \Delta \bar{I}(t)\rangle$ is given by Eq. (4.11). Thus, the rate is always positive.

However, Eq. (6.1) tells us nothing about the relation between the H function and the entropy. The H function $\bar{\mathrm{H}}$ is defined by

$$\bar{\mathrm{H}} = \lim_{\substack{V\to\infty,\, N\to\infty \\ N/V = \mathrm{const}}} \frac{1}{V} \mathrm{Tr}\rho \ln\rho. \qquad (6.3)$$

This function is, as is well known, invariant under the canonical transformation, or under the motion of the system having a Hamiltonian $\mathcal{H}$. However, it would be instructive to see how $\bar{\mathrm{H}}$ remains invariant in the course of time.

$\bar{\mathrm{H}}$ is expanded according to the order of perturbation, as follows

$$\bar{\mathrm{H}} = \lim(1/V)[\mathrm{Tr}\rho_0 \ln\rho_0 + \Delta_1\mathrm{H} + \Delta_2\mathrm{H} + \cdots]. \qquad (6.4)$$

On the other hand, $\rho(t)$ has already been given by Eqs. (2.3)–(2.5). Therefore, we can write

$$\ln\rho = X_0 + \Delta_1 X + \Delta_2 X + \cdots, \quad X_0 = \ln\rho_0, \qquad (6.5)$$

Here $\Delta_1 X$, $\Delta_2 X$, $\cdots$ cannot be expressed explicitly in terms of $\Delta_1\rho$, $\Delta_2\rho$, $\cdots$, but they are related to each other by

$$\Delta_1\rho = \int_0^1 d\lambda\, e^{(1-\lambda)X_0} \Delta_1 X e^{\lambda X_0}, \qquad (6.6)$$

$$\Delta_2\rho = \int_0^1 d\lambda\, e^{(1-\lambda)X_0} \Delta_2 X e^{\lambda X_0} + \int_0^1 d\lambda\, e^{(1-\lambda)X_0}\Delta_1 X e^{\lambda X_0} \int^\lambda d\mu\, e^{-\mu X_0} \Delta_1 X e^{\mu X_0}. \qquad (6.7)$$

Hence, we have

$$\Delta_1\mathrm{H} = \mathrm{Tr}\Delta_1\rho \ln\rho_0 + \mathrm{Tr}\rho_0\Delta_1 X, \qquad (6.8)$$

$$\Delta_2\mathrm{H} = \mathrm{Tr}\Delta_2\rho \ln\rho_0 + \mathrm{Tr}\rho_0\Delta_2 X + \mathrm{Tr}\Delta_1\rho\Delta_1 X. \qquad (6.9)$$

From Eqs. (2.4) and (2.5) we have[24]

$$\mathrm{Tr}\Delta_1\rho = 0, \quad \mathrm{Tr}\Delta_2\rho = 0, \qquad (6.10)$$

$$\mathrm{Tr}\Delta_1\rho \ln\rho_0 = 0. \qquad (6.11)$$

Equations (6.10) and (6.11) yield, by virtue of Eqs. (6.6) and (6.7),

$$\mathrm{Tr}\rho_0\Delta_1 X = 0, \qquad (6.12)$$

$$\mathrm{Tr}\rho_0\Delta_2 X = -\int_0^1 d\lambda\,(1-\lambda)\,\mathrm{Tr}\Delta_1 X e^{\lambda X_0}\Delta_1 X e^{(1-\lambda)X_0}. \qquad (6.13)$$

Consequently, we have

$$\Delta_1\mathrm{H} = 0. \qquad (6.14)$$

In order to see the behavior of $\Delta_2\mathrm{H}$ we first investigate $\mathrm{Tr}\Delta_2\rho \ln\rho_0$. It follows from Eq. (2.5) that

$$\begin{aligned}\mathrm{Tr}\Delta_2\rho \ln\rho_0 &= -\frac{1}{i\hbar}\int_0^t \mathrm{Tr}\exp\left(-\frac{i(t-t')\mathcal{H}_0}{\hbar}\right) \\ &\quad\times[A,\Delta_1\rho]\exp[i(t-t')\mathcal{H}_0/\hbar]\ln\rho_0 F(t')dt' \\ &= -\frac{\beta}{i\hbar}\int_0^t F(t')dt' \sum_{n,m}(A_{mn}\Delta_1\rho_{nm} \\ &\quad - \Delta_1\rho_{mn}A_{nm})(\Omega-\epsilon_m). \end{aligned}\qquad (6.15)$$

[24] Equations (5.24) and (6.10)∼(6.12) are easily proved by making use of the relation $\mathrm{Tr}AB = \mathrm{Tr}BA$. However, this relation does not necessarily hold in general, as we can see in a simple example given by taking Tr of the identity $pq - qp = \hbar/i$. In order for this relation to hold, the infinite series obtained by taking Tr must be absolutely convergent, and the order of terms is changeable. Speaking in terms of classical mechanics, the integral of the Poisson bracket corresponding to the commutator $AB - BA$ over the whole phase space must vanish, or, in other words, the partial integration vanishes at the boundaries of the phase space.

$\Delta_1\rho$ has already been given by Eq. (2.4):

$$\Delta_1\rho_{nm}(t')=\frac{i}{\hbar}\int_0^{t'} e^{-i(t'-t'')(\epsilon_n-\epsilon_m)/\hbar} \times A_{nm}(\rho_{0m}-\rho_{0n})F(t'')dt''. \quad (6.16)$$

This is substituted into Eq. (6.15) to yield

$$\begin{aligned}
&\mathrm{Tr}\Delta_2\rho\ \ln\rho_0\\
&=-\frac{\beta}{\hbar^2}\int_0^t F(t')dt'\int_0^{t'} F(t'')dt''\\
&\quad\times\sum_{m,n}(e^{-i(t'-t'')(\epsilon_n-\epsilon_m)/\hbar}+e^{i(t'-t'')(\epsilon_n-\epsilon_m)/\hbar})\\
&\quad\times(\rho_{0m}-\rho_{0n})A_{mn}A_{nm}(\Omega-\epsilon_m)\\
&=-\frac{\beta}{2\hbar^2}\int_0^t F(t')dt\int_0^{t'} F(t'')dt''\\
&\quad\times\sum_{m,n}(e^{-i(t'-t'')(\epsilon_n-\epsilon_m)/\hbar}+e^{i(t'-t'')(\epsilon_n-\epsilon_m)/\hbar})\\
&\quad\times A_{mn}A_{nm}(\rho_{0m}-\rho_{0n})(\epsilon_n-\epsilon_m)\\
&=-(\beta/2)\sum_{m,n}\left|\int_0^t e^{-it'(\epsilon_n-\epsilon_m)/\hbar}F(t')dt'\right|^2 J_{mn}\leq 0. \quad (6.17)
\end{aligned}$$

Here, the third expression is obtained by adding to the second expression the same one with m and n exchanged and divided by two. The fourth expression is obtained by noting that the integrand is an even function of $t'-t''$, and the upper limit of t'' can be extended to t provided that the total integral is divided by two, and, further, by making use of the definition of J_{mn} given by Eq. (4.2). It is obvious without detailed calculation, though it can be shown from Eqs. (6.6) and (6.13), that the last two terms of Δ_2H in Eq. (6.9) just cancel its first term given by Eq. (6.17).

Equation (6.17) is differentiated with respect to t, with the result

$$\frac{d}{dt}\mathrm{Tr}\Delta_2\rho\ \ln\rho_0 =-\beta\sum_{m,n}J_{mn}\left[\int_0^t e^{-i(t-t')(\epsilon_n-\epsilon_m)/\hbar}F(t')dt'\right]F(t). \quad (6.18)$$

On adding Eq. (6.18) and Eq. (6.11), the following relation holds at least up to the second order,

$$-k\lim\frac{1}{V}\mathrm{Tr}\frac{d\rho}{dt}\ln\rho_0=\frac{1}{T}\langle\Delta\bar{I}(t)\rangle F(t). \quad (6.19)$$

For this derivation, use has been made of Eq. (4.10). This is just the rate of the increase of entropy given by Eq. (6.1). This relation between $-k\,\mathrm{Tr}\rho\ln\rho_0$ and entropy has already been noticed by Mori.[25]

[25] H. Mori, J. Phys. Soc. Japan **11**, 1029 (1956).

Usually, $-k\,\mathrm{Tr}\rho\ln\rho$ is interpreted as entropy. However, it is constant in time, as already mentioned, unlike the increase of the entropy in a closed system. Thus, we are tempted to adopt the following interpretation:

$$-\frac{d}{dt}k\,\mathrm{Tr}\rho\ \ln\rho=\frac{1}{T}\frac{d'Q}{dt}, \quad (6.20)$$

where $d'Q/dt$ means the rate of flow of heat from a heat reservoir to the system, instead of the usual interpretation

$$-\frac{d}{dt}k\,\mathrm{Tr}\rho\ \ln\rho=\frac{dS}{dt}. \quad (6.21)$$

In a quasi-static process, we have $d'Q=TdS$, and (6.20) is in accord with Eq. (6.21). In an isolated system, such as the one we have treated in the present article, $d'Q$ equals zero, and Eq. (6.20) is in accord with the requirement of mechanics. Furthermore, according to thermodynamics, we can write

$$(1/T)d'Q/dt=dS/dt-(dS/dt)_{\mathrm{irr}}. \quad (6.22)$$

In effect, we have shown that $-(d/dt)k\,\mathrm{Tr}\rho\ln\rho$ is divided into two terms and one of them is just equal to $dS/dt=(dS/dt)_{\mathrm{irr}}$.

$$-\lim\frac{k}{V}\mathrm{Tr}\frac{d\rho}{dt}\ln\rho_0=\frac{dS}{dt}. \quad (6.23)$$

It is a conjecture that Eq. (6.20) will hold in case of a system in contact with a heat reservoir.

It is easy to extend the above calculations to a case where the Hamiltonian has several external forces as is discussed in Sec. V. It will be sufficient here to write only the main results. As a generalization of Eqs. (6.18) and (6.19), we have

$$\begin{aligned}
&-\lim_{\substack{V\to\infty,\,N\to\infty\\ N/V=\mathrm{const}}}\frac{k}{V}\mathrm{Tr}\frac{d\rho}{dt}\ln\rho_0\\
&=\frac{1}{T}\lim\frac{k}{V}\sum_l\left[\sum_h\sum_{m,n}J_{mn}^{(h\to l)}\times\int_0^t\exp\left(-i\frac{(t-t')(\epsilon_n-\epsilon_m)}{\hbar}\right)F_h(t')dt'\right]F_l(t)\\
&=(1/T)\sum_l\langle\Delta\bar{I}_l(t)\rangle F_l(t), \quad (6.24)
\end{aligned}$$

where use is made of Eqs. (5.8) and (5.9) to get the third expression, which is the thermodynamical expression for the increase of the entropy in a closed system. In periodic excitations given by $F_l=F_l^{(0)}e^{i\omega_l t}$, we have after a sufficiently long time

$$\langle\Delta\bar{I}_l(t)\rangle=\sum_h\sigma_I^{(h\to l)}(i\omega_h)F_h^{(0)}e^{i\omega_h t}. \quad (6.25)$$

If $\omega_l \neq \omega_h$ the two excitations do not interfere with each other, and the entropy production is a sum of those of single excitations. In a case where two excitations, say l and h, have the same frequency $\omega=\omega_l=\omega_h$, then the contribution of these two excitations to the entropy is given, just as in Eq. (6.2), by

$$\begin{aligned}(dS/dt)_{l,h} &= (dS/dt)_{l,h,\mathrm{irr}} \\ &= (1/2T)\Re[\sigma^{(l\to l)}(\omega)(F_l^{(0)})^2+\sigma^{(h\to h)}(\omega)(F_h^{(0)})^2 \\ &\quad +2\,\mathrm{Sym}\sigma^{(l\to h)}(\omega)F_l^{(0)}F_h^{(0)}], \end{aligned} \tag{6.26}$$

which is always positive. If all the excitations have the same frequency ω, we have

$$\frac{dS}{dt}=\left(\frac{dS}{dt}\right)_{\mathrm{irr}}=\frac{1}{2T}\Re\sum_{l,h}\mathrm{Sym}\sigma^{(l\to h)}F_l^{(0)}F_h^{(0)}, \tag{6.27}$$

which is always positive [see Eq. (5.27)].

Finally, if we adopt the definition of entropy according to Eq. (6.23), i.e.,

$$S=-\lim_{\substack{V\to\infty,\,N\to\infty\\ N/V=\mathrm{const}}}\frac{k}{V}\,\mathrm{Tr}\rho\ln\rho_0, \tag{6.28}$$

its first order variation is zero, as is shown in Eq. (6.11). There is no entropy production in the first order. This means that the relation between the first order variations of thermodynamical quantities in quasi-static processes still holds for linear irreversible processes. In the thermodynamics of irreversible processes, one assumption is that one can use the Gibbs relation which describes most generally the relation between the first order variations of thermodynamical quantities. A special case of this assumption is thus proved.

VII. FURTHER DISCUSSION ON IRREVERSIBILITY

In the above sections, we have shown that the vanishing of an aftereffect function, which is expected in the limit of an infinitely large system, is responsible for various aspects of irreversibility, such as the positive definiteness of the real part of the conductivity, entropy production or the reciprocal relations. However, the relation between the vanishing of the aftereffect function and the reversibility of mechanics itself should be further elucidated. The former property is, in part, closely connected with the Poincaré cycle which becomes infinitely long when the system is made very large. However, this is not sufficient to get irreversibility. The free motion of a particle in an infinite space never comes back to its original starting point, but it obeys reversible mechanics.

If the direction of time is reversed, as well as the magnetic field and Coriolis force, if any, the motion of the system will just reversibly trace back the trajectory of the previous motion in a finite system. In a linear process, this is described by Eq. (2.9) for $t<0$. $\langle \Delta B_1(t)\rangle$ in this case stands for the deviation from the initial value of B at the moment of the reversal of time. $\phi(t)$ given by Eq. (2.10) can be used for $t<0$. The transfer function must be defined for $z<0$. Thus, in the limit of an infinitely large system, the image "charge" equivalent to the continuously distributed "charges" on the imaginary axis must be made on the right half of the z plane, contrary to the case discussed in Sec. II. In other words, even when we make the system return back to the past by time and magnetic field reversal, we again have $\phi(t)\to 0$ for $t\to -\infty$. Thus the system will never return to the initial state. This may be the proper meaning of irreversibility. *Umkehreinwand* and *Wiederkehreinwand* raised by Loschmidt and by Zermelo,[26] respectively, are thus completely eliminated. In an actual case of a finite system with many interacting particles, the distribution of poles on the imaginary axis is still discrete, but is almost continuous. We cannot disclose its discrete nature in macroscopic measurements which necessarily contain errors or include time and/or space averages. This effectively continuous distribution of poles for a finite system leads to irreversibility. These measurement processes are supposed to be provided in theroretical considerations by time-smoothing or coarse-graining procedures.

ACKNOWLEDGMENTS

The author wishes to express his deep appreciation to Professor T. Yamamoto and Professor T. L. Hill for their illuminating discussions. Thanks are also due to Professor M. Namiki of Waseda University for his valuable communications during the research, and Professor M. Takeo of Portland State University and Professor T. L. Hill for reading the manuscript. He is also grateful for the hospitality of the Department of Chemistry, University of Oregon.

[26] See, for example, S. Chandrasekhar, Revs. Modern Phys. **15**, 1 (1943); A. Münster, *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1959), Vol. III, Part 2, p. 216.