

# Memory Effects in Irreversible Thermodynamics\*

ROBERT ZWANZIG

National Bureau of Standards, Washington, D. C.

(Received July 13, 1961)

A new generalization of Onsager's theory of irreversible processes is presented. The main purpose is to allow for memory effects or causal time behavior, so that the response to a thermodynamic force comes later than the application of the force. This is accomplished by a statistical mechanical derivation of an exact non-Markoffian kinetic equation for the probability distribution in the space of macroscopic state variables. The memory effect in the resulting transport equations is represented by a time convolution of the thermodynamic forces with memory functions. The latter are time-correlation functions in the rates of change of the phase functions corresponding to macroscopic quantities. The resulting transport equations are not restricted to small deviations from thermal equilibrium. Onsager's theory is shown to be the low-frequency limit of our causal theory.

## PROSPECTUS

THE subject of this article is a new generalization of Onsager's thermodynamic theory of irreversible processes.<sup>1</sup> The article has two purposes. One is to provide a theory that is applicable to situations where memory effects are important. The other is to establish criteria for the validity of Onsager's theory as a limiting case of a more general theory.

## SIMULTANEITY AND CAUSALITY IN IRREVERSIBLE THERMODYNAMICS

The familiar laws of irreversible thermodynamics have the characteristic form

$$d\alpha_j/dt = \sum_k L_{jk} F_k(\alpha_1, \alpha_2, \dots, \alpha_n), \quad (1)$$

where  $\alpha_j$  ( $j=1, 2, \dots, n$ ) is the deviation of the  $j$ th state variable from its value at thermal equilibrium,  $F_k$  is a thermodynamic force (defined as derivative of entropy with respect to  $\alpha_k$ ), and the  $L_{jk}$  are transport coefficients, satisfying the reciprocal relations  $L_{jk} = L_{kj}$ .<sup>2</sup> The forces are customarily assumed to be linear in the deviations.

We are concerned here with the following property of Onsager's theory: According to Eq. (1), the response of a system to an applied force is simultaneous with the application of the force.

As a general rule, such simultaneity in a macroscopic theory turns out to be an approximation to a causal behavior, where the response to a force comes *after* the application of the force.

In many familiar cases the approximation of simultaneity is as good as one needs to describe experimental observation. Under this category we mention the Navier-Stokes equation for fluid flow, Fick's law of diffusion, and Fourier's law of heat conduction. All of these are known to cover a wide range of experience. But there are occasions when a causal description is needed. Some examples are the decay of dielectric

polarization, the relaxation of stretched polymers, and ultrasonic absorption in molecular fluids.

Causal behavior is always associated with ignoring certain molecular variables. The time dependence of a completely specified state of a macroscopic system is governed by equations (Hamilton's, Schrödinger's, etc.) that show an instantaneous response.<sup>3</sup> A complete specification of the state of a macroscopic system requires knowledge of a very large number of molecular variables. In a macroscopic experiment one measures only a few of these, and one proceeds to deduce transport equations just as if the others did not exist. This is where causal behavior appears.

## ON THE VALIDITY OF ONSAGER'S THEORY

There is another reason, more methodological than physical, for seeking a causal generalization of Onsager's theory. In his derivation of Eq. (1), Onsager introduced certain hypotheses leading directly to an instantaneous response. Many subsequent rederivations (too many to cite individually) have been concerned mainly with justifying this simultaneity, by giving further support to Onsager's hypotheses, or by putting forth new and equivalent hypotheses.

Such attempts are of doubtful value. Our principal objection is that they do not show how it happens that certain substances, in certain experiments, behave causally instead of showing an instantaneous response.

A more satisfactory procedure is to derive a causal theory in the first place, without relying on unverified hypotheses, and then to investigate the validity of Onsager's theory as a limiting case. This is what we do here.

## SUMMARY OF RESULTS

Our results have the general structure of Eq. (2):

$$d\alpha_j(t)/dt = \sum_k \int_0^t ds K_{jk}(s) F_k(\alpha_1(t-s), \dots, \alpha_n(t-s)). \quad (2)$$

Some complicating features have been omitted in order

<sup>3</sup> We shall not be concerned with the extra complications of retarded interactions via transverse electromagnetic fields.

\* This work was supported in part by the Office of Scientific Research of the U. S. Air Force.

<sup>1</sup> L. Onsager, *Phys. Rev.* **37**, 405 (1931); **38**, 2265 (1931). See also S. R. deGroot, *Thermodynamics of Irreversible Processes* (North-Holland Publishing Company, Amsterdam, 1951).

<sup>2</sup> In the absence of magnetic fields.

to display the essential features of the time dependence in close analogy with Eq. (1). In the body of the article we give the complete and correct equations.<sup>4</sup>

Equation (2) was derived specifically to describe the free relaxation of a system that is initially ( $t=0$ ) in thermal equilibrium, subject to constraints fixing the numerical values  $\alpha_1(0)$ ,  $\alpha_2(0)$ ,  $\dots$ ,  $\alpha_n(0)$  of the  $n$  specified state variables.

The memory functions  $K_{jk}(s)$  are time-correlation functions of the kind that have appeared already in Onsager's work. We shall give explicit formulas for them later.<sup>4</sup> For the moment, we only observe that Onsager's transport coefficients  $L_{jk}$  are related to the memory functions in a simple way,

$$L_{jk} = \int_0^\tau ds K_{jk}(s). \quad (3)$$

The upper limit  $\tau$  is a time of macroscopic size, but not so large as the Poincaré recurrence time.

It is easy to see that qualitative conditions for the validity of Eq. (1) are as follows: (i) The thermodynamic forces should vary slowly enough that it is safe to replace  $\alpha(t-s)$  by  $\alpha(t)$  in the time integration. (ii) The memory functions should have a short enough relaxation time that the integrals from 0 to  $t$  can safely be replaced by integrals from 0 to some fixed macroscopic time  $\tau$ . More precise conditions can be obtained only by solving both Eq. (1) and Eq. (2), and by deciding when the two solutions do not differ appreciably.

The preceding qualitative requirements involve the structure of the thermodynamic forces and the time dependence of the memory functions. They thus involve specific properties of matter, rather than general statistical assumptions.

#### CONDITIONS AND ASSUMPTIONS USED IN THIS ARTICLE

We shall work entirely within the framework of classical statistical mechanics. Thus we shall not be able to say anything about the phenomena of superfluidity. Other than this, there seems to be no reason to expect anything strikingly different from a quantum mechanical treatment.

The derivation is based on several assumptions, listed in this paragraph. (i) A system known to be in thermal equilibrium subject to specified constraints can be described correctly by a phase space ensemble density that is microcanonical in the constraints. (ii) The molecular analogs of macroscopic state variables are phase functions depending on large numbers of molecules. (iii) These phase functions vary slowly in time, relative to the rates of individual molecular processes.

<sup>4</sup> See Eqs. (40), (39), (30), (31), and (28).

Some of the more formal parts of our results are independent of the second and third assumptions. The first is essential to our method.

The results of the derivation are limited to the following circumstances. (i) We consider only systems that are known to be initially in thermal equilibrium, subject to specified constraints. (ii) The system is not influenced by time-dependent external forces or boundary conditions.

This section ends with a few negative remarks. We shall not rely on assumptions about coarse graining, time smoothing, repeated phase randomization, or Markoffian behavior. Also, we shall not use diagrammatic techniques.

#### OUTLINE OF THE METHOD

The method followed here is a quasi-experimental one. We go through a sequence of steps, each of which has its counterpart in experimental activity.

The first step is to select a set of variables to describe the macroscopic state of the system. These variables are phase functions,  $A_1(x)$ ,  $A_2(x)$ ,  $\dots$ ,  $A_n(x)$ , where  $x$  denotes the position of the system in phase space, i.e., the set of all coordinates and momenta. We shall often abbreviate the notation to just  $A(x)$ .

The experimenter bases his choice of variables on past experience and physical intuition. We can hardly do otherwise here: No *a priori* choice is specifically indicated by theory. We shall require, however, that our phase functions have certain characteristic macroscopic properties. In particular, they should depend on the positions and velocities of very many molecules, and their time derivatives should be small, relative to the time derivatives of individual molecular quantities. We shall require also that the known integrals of motion (that is, total energy, linear and angular momentum,<sup>5</sup> and number of molecules) be included among the  $A(x)$ . A useful choice of phase functions is discussed by Green.<sup>6</sup>

The next step is to prepare a system initially ( $t=0$ ) having specified numerical values of the phase functions  $A(x)$ . We shall always require that the system be in thermal equilibrium, subject to whatever constraints are needed to maintain the specified values of  $A(x)$ . This has the effect of pinning down, implicitly, all properties of the system not included explicitly in  $A(x)$ . We know from experience that the future behavior of such initial states is experimentally reproducible.

Next we remove the constraints, and allow the system to relax freely. We do not impose time-dependent external forces or boundary conditions.

<sup>5</sup> The role of "intrinsic" angular momentum, i.e., the angular momentum of rotations of molecules about their centers, is expected to be negligible except in problems of orientational relaxation, for example in polar dielectrics. The angular momentum associated with translational molecular motions is taken into account by including hydrodynamic velocity fields among the  $A(x)$ .

<sup>6</sup> M. S. Green, J. Chem. Phys. 20, 1281 (1952); 22, 398 (1954).

Then we calculate the probability distribution of the  $A(x)$  as a function of time. This is accomplished by a formal solution of the molecular equations of motion. The theoretical distribution of the  $A(x)$  corresponds to the statistical distribution of observations on an experiment repeated very many times under macroscopically identical conditions.

Finally, we observe that the ensemble averages of the  $A(x)$  satisfy transport equations of the form of Eq. (2). We find that the ensemble averages correspond closely to experimental quantities, in that fluctuations from the ensemble average are macroscopically negligible. This corresponds to the observed reproducibility, on repetition, of actual experiments.

#### MOLECULAR PICTURE OF THE EXPERIMENTS TREATED HERE

Before embarking on a detailed mathematical theory, it will be useful to have in mind a picture of what happens in phase space during an experiment, and what corresponds in phase space to the repetition of an experiment.

The state of a system is specified in molecularly complete detail by the position  $x$  of the system point in phase space. In particular,  $x$  determines the numerical values  $a_j = A_j(x)$  ( $j = 1, 2, \dots, n$ ) of the chosen phase functions. We shall frequently denote the set  $a_1, a_2, \dots, a_n$  by just  $a$ .

The system point moves from its initial position  $x_0$  to the position  $x_t$  at time  $t$ , determined by the molecular equations of motion. The phase functions change from  $A(x_0)$  to  $A(x_t)$ . In the laboratory one measures only the numerical values  $A(x_t)$  as functions of time. One does not know the initial phase-space position  $x_0$ .

What corresponds to repeating an experiment? We note first that the equations  $A(x) = a$  define a surface in phase space; we call it  $S(a)$ . Any point  $x$  on the surface  $S(a)$  gives rise to the same set of numbers  $a$ . If we know only the initial values  $a_0 = A(x_0)$ , but not the exact position  $x_0$ , each repetition of the experiment consists in starting at a different point  $x$  on the same surface  $S(a_0)$ .

In this article we consider only experiments where the system is initially in thermal equilibrium, subject to constraints giving rise to specified values of the variables  $a$ . It is well known that the properties of such a system are correctly described by an ensemble of phase points uniformly (i.e., microcanonically) distributed on the surface  $S(a)$ .<sup>7</sup> The assumption of constrained thermal equilibrium in the initial state

<sup>7</sup> At this point our language should be more precise. Although we refer to the surface  $S(a)$ , we should really speak of a thin shell contained between the surfaces  $S(a)$  and  $S(a+da)$ , where  $da$  is very small. The microcanonical distribution is uniform within this shell. Since we shall use a delta function formula for the microcanonical distribution, we do not have to refer to the shell explicitly. Nevertheless, in order to avoid possible confusion, the shell should be kept in mind.

thus means that each repetition of an experiment consists in starting at a different point, *chosen at random*, on the same surface  $S(a)$ .

Each different initial point  $x_0$  gives rise to a different  $A(x_t)$ . That is, each repetition of an experiment, under identical macroscopic conditions, gives rise to a different outcome. A complete statistical description of the outcomes of repeated experiments is contained in the probability distribution  $g(a; t)$  in  $a$  space (the space of  $n$  dimensions, the points of which are labeled by  $a_1, a_2, \dots, a_n$ ). Specifically, the probability at time  $t$  of the event

$$a_j < A_j(x) < a_j + da_j, \quad (j = 1, 2, \dots, n), \quad (4)$$

is

$$g(a; t) da = g(a_1, a_2, \dots, a_n; t) da_1 da_2 \dots da_n. \quad (5)$$

Owing to our assumption of initial constrained thermal equilibrium, the initial value of the  $a$ -space distribution is

$$g(a; 0) = \delta(a - a_0) = \prod_{j=1}^n \delta(a_j - a_{j0}), \quad (6)$$

where  $a_0$  denotes the set of assigned initial values.

It is evident that  $g(a; t)$  represents the probability of transition from the initial surface  $S(a_0)$  to some arbitrary surface  $S(a)$  in the time  $t$ . (One should remember that "surface" means "thin shell.")

For a better picture of  $g(a; t)$ , some later results will be anticipated here. As time passes, the initial delta function will spread out in  $a$  space. The first moment of the distribution,

$$\alpha_j(t) = \int a_j g(a; t) da, \quad (j = 1, 2, \dots, n), \quad (7)$$

satisfies transport equations like Eq. (2); it corresponds to the statistical mean of observations on repeated experiments. The width of the distribution, specified for example by the second moments, remains of the order of magnitude of fluctuations from thermal equilibrium. As is well known, such fluctuations are macroscopically negligible when the  $A(x)$  are many-particle functions. Thus the distribution  $g(a; t)$  stays macroscopically sharp. This corresponds to the reproducibility of experiments.

This molecular picture of an experiment can be used also to elucidate the role of memory. Let us look at the surface  $S(\alpha(t))$  defined by the mean position of the distribution  $g(a; t)$ . If the phase-space distribution were approximately microcanonical on this surface, then we would be able to predict the future behavior, for  $t'$  later than  $t$ , with no more information than the values of  $\alpha(t)$ . We would not need to know the history of the distribution, and there would be no memory effect.

But there is no reason, in general, to expect the phase-space distribution to remain approximately microcanonical. Indeed, it may tend to concentrate on

some regions of the surface  $S(\alpha(t))$  and avoid other regions. In order to reproduce the experiment, starting at  $t$ , we have to take account of the possibility that the distribution is not microcanonical. To do this we must know how the distribution arose, or what its history is.

This explains why, in many instances, it is possible to eliminate long-time memory effects by expanding the set of phase functions  $A(x)$  to include more functions. By doing so, we have a subsurface  $S(\alpha'(t))$ . With a good choice of new phase functions, the phase space distribution on the new surface  $S(\alpha'(t))$  may be more nearly microcanonical. That is, the new  $S(\alpha'(t))$  may include only the region of the old  $S(\alpha(t))$  where phase points tended to concentrate.

#### DERIVATION OF A KINETIC EQUATION IN $a$ SPACE

As we have already observed, a complete statistical description of the outcomes of repeated experiments is contained in the time dependence of the probability distribution  $g(a; t)$  in  $a$  space. We shall derive a kinetic equation for  $g(a; t)$  in the next few sections.

The basic difficulty in finding the time dependence of  $g(a; t)$  is that there does not exist a one-to-one correspondence between motions in  $a$  space and in phase space. But the molecular equations of motion, which must form the starting point of our derivation, are given only in phase space.

The molecular equations of motion are frequently written in the form of Liouville's equation. As is universally known, the probability distribution in phase space,  $f(x; t)$ , satisfies the equation

$$i\partial f(x; t)/\partial t = Lf(x; t), \quad (8)$$

where  $L$  is the Liouville operator,<sup>8</sup> defined in terms of the Hamiltonian function  $H(x)$  and the Poisson bracket  $\{-, -\}_{\text{P.B.}}$ ,

$$Lf = i\{H, f\}_{\text{P.B.}} \quad (9)$$

The relation of the probability distribution in  $a$  space to that in phase space is

$$g(a; t) = \int dx \delta(A(x) - a) f(x; t). \quad (10)$$

[The reader should recall our conventions about the abbreviated notations  $A(x)$ ,  $\delta(A(x) - a)$ , and  $a$ .] The probability distribution in  $a$  space is clearly the total amount of the phase-space distribution contained on the surface  $S(a)$  specified by  $A(x) = a$ .

We shall also have use for the structure function  $W(a)$  of the surface  $S(a)$ ,

$$W(a) = \int dx \delta(A(x) - a), \quad (11)$$

<sup>8</sup> The  $i$  is customarily introduced so that  $L$  is Hermitian in the Hilbert space of phase functions.

which is the total volume of  $S(a)$ . The structure function provides a complete thermodynamic description of a system in thermal equilibrium subject to constraints giving rise to the specified values  $A(x) = a$ .<sup>9</sup>

The microcanonical distribution corresponding to our assumed initial state,  $g(a; 0) = \delta(a - a_0)$ , is

$$f(x; 0) = \delta(A(x) - a_0)/W(a_0). \quad (12)$$

As time passes,  $f(x; 0)$  will change to  $f(x; t)$ . We denote this symbolically by means of the Liouville time displacement operator  $\exp(-itL)$ ,

$$f(x; t) = \exp(-itL)f(x; 0). \quad (13)$$

The  $a$ -space distribution function at time  $t$  is therefore

$$g(a; t) = \int dx \delta(A(x) - a) e^{-itL} \delta(A(x) - a_0)/W(a_0). \quad (14)$$

#### THE DERIVATION CONTINUED: PROJECTION OPERATORS IN HILBERT SPACE

It is evident, from the preceding expression for  $g(a; t)$ , that we do not have to have full knowledge of the operator  $\exp(-itL)$ . All we really need is the matrix element of the operator between two microcanonical distributions. It is just this limitation that makes the rest of the derivation possible.

Let us consider the Hilbert space of phase-space distribution functions  $f(x; t)$ . This entire Hilbert space contains a smaller subspace, consisting of functions that depend on  $x$  only in the special way  $f = f(A(x))$ ; that is, they are functions of the specified phase functions  $A(x)$ . We observe that the microcanonical distribution belongs to this subspace, because of Eq. (12). In calculating  $g(a; t)$ , we are concerned only with matrix elements of  $\exp(-itL)$  taken between functions lying in this subspace.

The subspace of functions of  $A(x)$  can be described conveniently by means of a projection operator  $P$ , which selects from an arbitrary function  $G(x)$  that part,  $G_1(x) = PG(x)$ , which depends on  $x$  only through  $A(x)$ . The projection operator is given explicitly by

$$G_1(x) = PG(x) = \frac{\int dx' \delta(A(x') - A(x)) G(x')}{W(A(x))}. \quad (15)$$

The effect of  $P$  can be explained verbally as follows. For a given  $x$ , we determine the numerical values  $a = A(x)$ . We calculate the amount of  $G(x)$  contained on the corresponding surface  $S(a)$ ; this is the integral in the numerator. Then we spread this amount over the surface  $S(a)$ ; this is accomplished by the structure function  $W(A(x)) = W(a)$  in the denominator. In

<sup>9</sup> See, for example, A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover Publications, Inc., New York, 1949), particularly Chap. VII.

effect,  $P$  "averages" the function  $G(x)$  over the surface  $S(A(x))$ .

That  $P$  is in fact a projection operator can be seen easily by verifying that it satisfies the necessary and sufficient condition for a projection operator,  $P^2 = P$ .

Equation (10), expressing  $g(a; t)$  as an integral of  $f(x; t)$ , can be written

$$g(a; t) = W(a) [f_1(x; t)]_{A(x)=a}. \quad (16)$$

To find  $g(a; t)$ , we need the projection  $f_1(x; t)$  of  $f(x; t)$  onto the  $P$  subspace. The remaining part of  $f(x; t)$ , denoted by

$$f_2(x; t) = f(x; t) - f_1(x; t), \quad (17)$$

is irrelevant to the problem under discussion.

#### THE DERIVATION CONTINUED: A KINETIC EQUATION FOR $f_1$

In a previous article<sup>10</sup> we derived an equation for the time dependence of the projected part of a phase space distribution function. That discussion is immediately applicable here.

The method of derivation followed in reference 10 was to separate the distribution function into a "relevant" and an "irrelevant" part by means of a projection operator; to solve the equation for the irrelevant part, formally, in terms of the relevant part; and to substitute the solution back into the equation for the relevant part.

The resulting equation for  $f_1(x; t)$  [Eq. (11) of reference 10] is

$$i \frac{\partial f_1(t)}{\partial t} = PLf_1(t) - i \int_0^t ds PLe^{-is(1-P)L}(1-P)Lf_1(t-s) + PLe^{-it(1-P)L}f_2(0). \quad (18)$$

We have suppressed the  $x$  dependence for simplicity of notation.

The operator  $\exp[-it(1-P)L]$  is defined either by its series expansion or by its use in solving the equation

$$i \partial G(t) / \partial t = (1-P)LG(t) \quad (19)$$

for the initial condition  $G(0)$ . The solution is written

$$G(t) = e^{-it(1-P)L}G(0). \quad (20)$$

(We must be careful to note that the  $1-P$  in the exponent operates not only on the  $L$  immediately to its right, but on everything to its right in any expression where it occurs. This is more easily seen when one expands the exponential.)

The  $f_1$  equation contains a contribution from the initial value  $f_2(0)$ . This term vanishes in the present application, because our initial phase space distribution, being microcanonical, falls in the  $P$  subspace. Thus we have a closed kinetic equation for  $f_1$ .

Equation (18) shows a memory effect, or causal time behavior, which is described by the time convolution. This is a result of the conversion of the original Liouville equation for  $f(x; t)$  into an equation for the "relevant" part  $f_1(x; t)$ .

#### THE DERIVATION CONTINUED: AN EXACT KINETIC EQUATION FOR $g(a; t)$

We now transform the  $f_1$  equation into the corresponding equation for  $g(a; t)$ . The procedure is to evaluate  $f_1$  on a surface  $S(a)$ , using

$$[f_1(x; t)]_{S(a)} = g(a; t) / W(a), \quad (21)$$

which is another way of writing Eq. (16).

We shall want the value, on  $S(a)$ , of the projected part of an arbitrary function  $G$ . This is just the microcanonical average of  $G$  on the surface  $S(a)$ ,

$$[PG(x)]_{S(a)} = \frac{1}{W(a)} \int dx' \delta(A(x') - a) G(x') \quad (22)$$

because of the definition of  $P$ . The microcanonical average of  $G$  on  $S(a)$  will be denoted by  $\langle G; a \rangle$ .

The  $f_1$  equation, evaluated on  $S(a)$ , becomes

$$i \frac{1}{W(a)} \frac{\partial g(a; t)}{\partial t} = \langle Lf_1(t); a \rangle - i \int_0^t ds \langle Le^{-is(1-P)L}(1-P)Lf_1(t-s); a \rangle. \quad (23)$$

We still have to express  $f_1$  (appearing inside the averages) in terms of  $g(a; t)$ . This is accomplished by

$$f_1(x; t) = \int da' \delta(A(x) - a') g(a'; t) / W(a').$$

Now Eq. (23) is

$$\begin{aligned} \frac{1}{W(a)} \frac{\partial g(a; t)}{\partial t} = & -i \int da' \langle L\delta(A(x) - a'); a \rangle \frac{g(a'; t)}{W(a')} \\ & - \int_0^t ds \int da' \langle Le^{-is(1-P)L}(1-P)L\delta(A(x) - a'); a \rangle \\ & \times \frac{g(a'; t-s)}{W(a')}. \end{aligned} \quad (24)$$

Further simplifications are possible because of the Poisson bracket structure of  $L$ . The time rate of change of a phase function  $A(x)$  is

$$dA(x)/dt = \{A; H\}_{P.B.} = iLA. \quad (25)$$

Thus, when  $L$  operates on the delta function, for

<sup>10</sup> R. Zwanzig, J. Chem. Phys. 33, 1338 (1960). See also, R. Zwanzig, in *Lectures in Theoretical Physics*, edited by W. E. Brittin, B. W. Downs, and J. Downs (Interscience Publishers, Inc., New York, 1961), Vol. III, pp. 106-141.

example, it gives

$$\begin{aligned} L\delta(A(x)-a) &= -i \frac{dA}{dt} \delta'(A(x)-a) \\ &= i \sum_{j=1}^n \frac{dA_j}{dt} \frac{\partial}{\partial a_j} \delta(A(x)-a). \end{aligned} \quad (26)$$

(In the second line we have returned to the full subscript notation for the sake of clarity.)

Consequently, the equation for  $g(a; t)$  can be rewritten in an equivalent form which will prove to be more useful for approximations. The details of the transformation are given in an Appendix. The result of the transformation is

$$\begin{aligned} \frac{\partial g(a; t)}{\partial t} + \sum_{j=1}^n \frac{\partial}{\partial a_j} \{v_j(a)g(a; t)\} \\ = \int_0^t ds \int da' \sum_{j=1}^n \sum_{k=1}^n \frac{\partial}{\partial a_j} W(a) K_{jk}(a, a'; s) \\ \times \frac{\partial}{\partial a_k'} \frac{g(a'; t-s)}{W(a')}. \end{aligned} \quad (27)$$

In this equation several new symbols appear. These are

$$v_j(a) = \langle dA_j/dt; a \rangle, \quad (28)$$

and

$$\begin{aligned} K_{jk}(a, a'; s) \\ = \left\langle \frac{dA_j}{dt} e^{-is(1-P)L} (1-P) \frac{dA_k}{dt} \delta(A(x)-a'); a \right\rangle. \end{aligned} \quad (29)$$

The term containing  $v_j(a)$  comes from the first term on the right of Eq. (24), by applying Eq. (26) once. The memory term, containing  $K_{jk}(a, a'; s)$ , comes from the second term on the right of Eq. (24), by applying Eq. (26) twice.

The quantity  $v_j(a)$  is the average rate of change of  $A_j$  on the surface  $S(a)$ . The calculation of  $v_j(a)$  is a problem in statistical thermodynamics; no dynamical questions are involved. The memory function  $K_{jk}(a, a'; s)$  is a kind of correlation function in time and in  $a$  space. Its calculation involves the solution of a dynamical problem with the modified time displacement operator  $\exp[-is(1-P)L]$ .

#### THE DERIVATION CONCLUDED: AN APPROXIMATE KINETIC EQUATION IN $a$ SPACE

The preceding exact equation for  $g(a; t)$  is too complicated to be useful for any but the most formal applications. From it we may obtain, however, a useful approximate equation. The approximation is based on a characteristic property of macroscopic phase functions, namely, that they are varying only slowly in time.

The decisive role of this property has already been

noted by Green<sup>6</sup> and by Van Kampen.<sup>11</sup> We shall not repeat their arguments. It should be noted, however, that the rates of change of the phase functions considered by Green are controlled by a small parameter, the reciprocal of a wavelength characterizing the scale of spatial inhomogeneity in the system. By restricting the theory to a description of experiments involving only long-wavelength phenomena, the time derivatives of Green's phase functions can be made as small as desired.

The decision as to whether one can treat some specific time derivative as small must be based at present partly on intuitive grounds and partly on the *a posteriori* success of the resulting theory. We shall see later that there is in principle a direct theoretical check on this, but so far it is computationally impractical.

Regarding  $dA/dt$  as small, we shall keep only those terms in Eq. (27) that contain  $dA/dt$  to the first and second powers. Since the term containing  $v_j(a)$  is evidently of the first order in  $dA/dt$ , and gives no implicit higher order contributions, it will be kept intact. The term containing  $K_{jk}(a, a'; s)$  is already explicitly of second order in  $dA/dt$ , and gives higher order contributions implicitly because of the operator  $\exp[-is(1-P)L]$ . We want to dispose of these higher order contributions.

The method by which this is done is systematic but tedious. Details are given in an Appendix. The result is that to terms of the second order (but not to higher orders),  $K_{jk}(a, a'; s)$  is diagonal in  $a$  space,

$$K_{jk}(a, a'; s) = \delta(a-a') K_{jk}(a; s) + O(\dot{A}^3). \quad (30)$$

The coefficient  $K_{jk}(a; s)$  is a time-correlation function in the deviations of the time derivatives from equilibrium,

$$K_{jk}(a; s) = \langle [\dot{A}_j(s) - v_j(a)] [\dot{A}_k(0) - v_k(a)]; a \rangle, \quad (31)$$

where

$$\dot{A}_j(s) = e^{isL} \dot{A}_j(0) \quad (32)$$

is the value of  $dA/dt$  at time  $s$ , calculated with the ordinary time displacement operator. To this order, the modified operator  $\exp[-is(1-P)L]$  is not required.

The resulting approximate kinetic equation for  $g(a; t)$  is

$$\begin{aligned} \frac{\partial g(a; t)}{\partial t} + \sum_{j=1}^n \frac{\partial}{\partial a_j} \{v_j(a)g(a; t)\} \\ = \int_0^t ds \sum_{j=1}^n \sum_{k=1}^n \frac{\partial}{\partial a_j} \left\{ K_{jk}(a; s) W(a) \frac{\partial}{\partial a_k} \frac{g(a; t-s)}{W(a)} \right\} \\ + O(\dot{A}^3 g). \end{aligned} \quad (33)$$

The *only* approximation made in arriving at the preceding equation was the neglect of terms of the

<sup>11</sup> N. G. Van Kampen, *Physica* **20**, 603 (1954).

third order in  $dA/dt$ . Since we started with an exact equation for  $g(a; t)$ , and followed a systematic procedure for removing third and higher order terms, we could just as well keep the third order terms, and remove those of fourth and higher order. Thus it is possible, in principle, to see if third order terms have an appreciable effect. We have not done this for two reasons. First, the numerical investigation of the second order terms is so difficult, in any realistic case, that there is not much point in going further. Also, the second order theory leads to results, as far as is known, in good agreement with observation.

Equation (33) has very nearly the structure of a Fokker-Planck equation in  $a$  space. The only difference, though an essential one, is the presence of the time convolution. If we were to replace  $g(a; t-s)$  by  $g(a; t)$ , and to extend the time integration to infinity instead of stopping at  $t$ , we would arrive at the Fokker-Planck equation first derived by Green by means of a Markoffian assumption.<sup>6</sup>

#### DERIVATION OF TRANSPORT EQUATIONS FROM THE APPROXIMATE KINETIC EQUATION IN $a$ SPACE

So far we have obtained a statistical description of the outcomes of repeated experiments in the form of a kinetic equation for  $g(a; t)$ . In most situations such a complete statistical description is unnecessarily detailed. We know, for example, that many experiments can be repeated reproducibly, in the sense that the results are identical to within experimental error.

This is reflected in the customary use of transport equations having a precise, i.e., nonstatistical, meaning. Transport equations are descriptions of the average behavior observed in repeated experiments. Reproducibility on a macroscopic level means that fluctuations from average behavior are macroscopically unobserved. Our theoretical derivation of transport equations thus falls into two parts, a calculation of the average rates of change of the phase functions  $A(x)$ , and a discussion of fluctuations from the average.

The ensemble averages of the phase functions at time  $t$  are denoted by

$$\alpha_j(t) = \int da a_j g(a; t). \quad (34)$$

The time derivative of  $\alpha_j$  can be found from Eq. (33),

$$\begin{aligned} \frac{d\alpha_j(t)}{dt} = & - \int da a_j \sum_k \frac{\partial}{\partial a_k} \{v_k(a)g(a; t)\} \\ & + \int_0^t ds \int da a_j \sum_k \sum_l \frac{\partial}{\partial a_k} \\ & \times \left\{ K_{kl}(a; s) W(a) \frac{\partial}{\partial a_l} \frac{g(a; t-s)}{W(a)} \right\}. \end{aligned} \quad (35)$$

Integrations by parts lead us to

$$\begin{aligned} \frac{d\alpha_j(t)}{dt} = & \int da v_j(a)g(a; t) + \int_0^t ds \int da \\ & \times \sum_l \left\{ \frac{1}{W(a)} \frac{\partial}{\partial a_l} K_{jl}(a; s) W(a) \right\} g(a; t-s). \end{aligned} \quad (36)$$

Now we shall *assume* that the  $a$ -space distribution remains macroscopically sharp. This will be discussed further in a later paragraph. The assumption allows us to replace ensemble averages of functions by functions of ensemble averages, with an error that is macroscopically negligible, being of the order of thermal fluctuations. For example, we see that

$$\int da v_j(a)g(a; t) \cong v_j(\alpha(t)). \quad (37)$$

As a result of this assumption, the transport equation is

$$\begin{aligned} \frac{d\alpha_j(t)}{dt} = & v_j(\alpha(t)) + \int_0^t ds \sum_l \left\{ \frac{1}{W(\alpha(t-s))} \frac{\partial}{\partial \alpha_l} \right. \\ & \times K_{jl}(\alpha(t-s); s) W(\alpha(t-s)) \left. \right\}. \end{aligned} \quad (38)$$

The thermodynamic force is defined, in a conventional way, as

$$F_k(a) = \frac{\partial}{\partial a_k} \log W(a). \quad (39)$$

With this notation, the transport equation becomes

$$\begin{aligned} \frac{d\alpha_j(t)}{dt} = & v_j(\alpha(t)) + \int_0^t ds \sum_l \left\{ K_{jl}(\alpha(t-s); s) F_l(\alpha(t-s)) \right. \\ & \left. + \frac{\partial}{\partial \alpha_l} K_{jl}(\alpha(t-s); s) \right\}. \end{aligned} \quad (40)$$

At this point, the only assumptions made are (i) that contributions of order  $(dA/dt)^3$  may be neglected, and (ii) that the distribution in  $a$  space remains macroscopically sharp. We have already discussed the first of these; let us now turn to the second.

This question has been discussed, though in a more limited context, by Van Kampen.<sup>12</sup> We shall do no more here than refer to the essential points of his treatment.

Van Kampen considered only the conventional (Markoffian) Fokker-Planck equation in the linear approximation. (By this is meant that the forces are assumed to be linear in deviations from equilibrium,

<sup>12</sup> N. G. Van Kampen, *Physica* **23**, 707 (1957); **25**, 1294 (1959).

and the transport coefficients are assumed to be independent of the deviations.)

In the linear case, the solution of the Fokker-Planck equation is Gaussian in the quantities  $a - \alpha(t)$ . The width of  $g(a; t)$  is thus determined by the second moments. Van Kampen was able to solve for the time dependence of the second moments exactly. He found that the width of the Gaussian, initially equal to zero, grows until it reaches the width given by the size of spontaneous fluctuations from thermal equilibrium. Such fluctuations are known to be macroscopically negligible when the  $A(x)$  are sums, over all molecules in the system, of functions of one or two molecules.

We have not investigated what modifications have to be made in Van Kampen's argument in order to apply it to our more general kinetic equation. The calculation would be necessarily more complicated, both because of the memory and because one would like to account for nonlinear effects as well. Nevertheless, we regard Van Kampen's results as sufficiently conclusive; there is no apparent reason why the memory, or nonlinearities, should cause the width of the  $a$ -space distribution to be significantly larger.

In order to recover equations as simple as Onsager's, further assumptions are required. These are (i) the phase functions  $A(x)$  are chosen so that  $v_j(a)$  vanishes, and (ii) the memory functions  $K_{jk}(a; s)$  depend on only those  $a_j$  that are constants of the motion. Then we obtain

$$\frac{d\alpha_j(t)}{dt} = \int_0^t ds \sum_k K_{jk}(s) F_k(\alpha(t-s)), \quad (41)$$

where we have suppressed the dependence of  $K_{jk}$  on the constants of motion. If these requirements are not met, one must keep all of Eq. (40).

#### MEMORY EFFECTS AND FREQUENCY-DEPENDENT TRANSPORT COEFFICIENTS

There are two mathematically equivalent ways of looking at the effects of memory in irreversible thermodynamics. One is to work entirely with functions of time; the other is to work with functions of frequency, by means of Fourier transforms. In this section we shall illustrate the relation between the two approaches, and show that memory effects can be accounted for by introducing frequency-dependent transport coefficients.

We shall restrict our discussion, for the sake of simplicity, to the approximate Eq. (41). This equation will be compared with its Markoffian approximation,

$$\frac{d\alpha_j(t)}{dt} = \sum_k \int_0^\infty ds K_{jk}(s) F_k(\alpha(t)). \quad (42)$$

Let us take one-sided Fourier transforms of both

equations. We introduce the notations

$$\bar{\alpha}_j(\omega) = \int_0^\infty dt e^{i\omega t} \alpha_j(t), \quad (43)$$

$$\bar{K}_{jk}(\omega) = \int_0^\infty dt e^{i\omega t} K_{jk}(t), \quad (44)$$

$$\bar{F}_k(\omega) = \int_0^\infty dt e^{i\omega t} F_k(\alpha(t)). \quad (45)$$

It should be noted that

$$\int_0^\infty ds K_{jk}(s) = \bar{K}_{jk}(0). \quad (46)$$

The transform of Eq. (41) is

$$-i\omega \bar{\alpha}_j(\omega) - \alpha_j(0) = \sum_k \bar{K}_{jk}(\omega) \bar{F}_k(\omega), \quad (47)$$

and the transform of its Markoffian approximation is

$$-i\omega \bar{\alpha}_j(\omega) - \alpha_j(0) = \sum_k \bar{K}_{jk}(0) \bar{F}_k(\omega). \quad (48)$$

Evidently the entire effect of memory is contained in the frequency dependence of the transport coefficients  $\bar{K}_{jk}(\omega)$ . Onsager's theory, based on the Markoffian approximation, can be regarded as the low-frequency limit of our more general results.

It seems more natural, however, to work with functions of time, because the time dependence and not the frequency dependence is what one studies in an experiment of the sort we are considering.

The memory functions  $K_{jk}(a; s)$  obey reciprocal relations,

$$K_{jk}(a; s) = K_{kj}(a; s) \quad (49)$$

for fixed  $a$ , at each instant of time  $s$ . (We omit the extra discussion that must be made when magnetic fields are present.) Consequently, the Fourier transforms  $\bar{K}_{jk}(\omega)$  also obey reciprocal relations.

#### SOME NOTEWORTHY PROPERTIES OF THE TRANSPORT EQUATIONS DERIVED HERE

Equation (40) has several significant and potentially useful features, aside from its causal structure. These are as follows.

The thermodynamic forces are defined in terms of the structure function  $W(a)$ , by Eq. (39). They are not necessarily linear in the state variables  $\alpha$ . Furthermore, there is nothing in the present derivation that requires us to linearize them.

The memory functions  $K_{jk}(\alpha; s)$ , filling the role of transport coefficients, are functions of the state variables  $\alpha$ . In principle one can calculate, for example, the dependence of a viscosity coefficient on density and

velocity gradients. This goes beyond previous theories of transport coefficients.

The preceding nonlinearities, together with the causal behavior, ought to be enough to describe a remarkably diverse range of macroscopic phenomena. The author is not aware of any situations in rheology, for example, that evidently cannot be treated in this way. However, considerable further study of the properties of Eq. (40), and comparisons with experiment, are required before any quantitative claims should be made.

#### SURVEY OF THE VARIOUS EQUATIONS DERIVED HERE

In this article we have derived several equations having different ranges of applicability and validity. Here we shall survey these equations, and list, for the last time, the various conditions and assumptions that are involved.

The first condition concerned the *kind* of experiment to be discussed, namely, free relaxation of a system initially at thermal equilibrium subject to constraints. We did not include time-dependent external forces or boundary conditions. Along with this condition went the assumption that such an initial state can be described by a microcanonical ensemble.

At this point we derived a general kinetic equation for the probability distribution in  $a$  space, Eq. (27). This equation is exact.

Then we expanded the general kinetic equation in powers of the time derivatives  $dA/dt$ , and dropped terms of the third and higher orders, leading to Eqs. (33), (30), and (31). We assumed that  $dA/dt$  was sufficiently "small" to justify this. We observed that this "smallness" is a characteristic property of macroscopic variables.

Following this, we used the kinetic equation in  $a$  space to derive transport equations. First we obtained Eq. (36), which remains valid even when the  $a$ -space distribution is not sharp. [We could have found corresponding transport equations from the exact Eq. (27), but there seemed no reason to do so.] We assumed that the  $a$ -space distribution is sharp, so that we could replace averages of functions by functions of averages. This led directly to our principal results, Eqs. (40) and (39). (Again, we could have done this with the exact kinetic equation.)

This was followed by a discussion of the sharpness of  $g(a; t)$ . Here, an argument due to Van Kampen was used to justify (though not to prove) our assumption about the sharpness of  $g(a; t)$ . The argument was based on the many-body structure of macroscopic phase functions.

We displayed transport equations, Eq. (41), differing from Onsager's only in the presence of memory effects. In doing this we assumed that the average velocities  $v_j(\alpha)$  vanished, and that the memory functions  $K_{jk}(s)$

did not depend on the variables  $\alpha$ , except for those that are constants of motion.

Finally, we showed how Onsager's theory could be regarded as the low-frequency limit of our Eq. (41).

#### APPENDIX I

The transformation from Eq. (24) to Eq. (27) is described here. First we consider the term leading to  $v_j(a)$ .

By the definition of the ensemble average, we see that

$$\langle L\delta(A(x)-a'); a \rangle = \int dx \frac{\delta(A(x)-a)}{W(a)} L\delta(A(x)-a'). \quad (A1)$$

Because  $L$  is Hermitian, we may move  $L$  in front of the first delta function,

$$= \int dx [L^* \delta(A(x)-a)] \frac{1}{W(a)} \delta(A(x)-a'), \quad (A2)$$

and by Eq. (26), we obtain

$$= - \frac{i}{W(a)} \sum_{j=1}^n \int dx \left[ \dot{A}_j(x) \frac{\partial}{\partial a_j} \delta(A(x)-a) \right] \times \delta(A(x)-a'). \quad (A3)$$

On bringing the derivatives outside the integral sign, the preceding expression becomes

$$= - \frac{i}{W(a)} \sum_{j=1}^n \frac{\partial}{\partial a_j} \int dx \dot{A}_j(x) \delta(A(x)-a) \times \delta(A(x)-a'), \quad (A4)$$

and the delta functions can be rearranged to yield

$$= - \frac{i}{W(a)} \sum_{j=1}^n \frac{\partial}{\partial a_j} \int dx \dot{A}_j(x) \delta(A(x)-a) \delta(a-a'). \quad (A5)$$

According to our definition of  $v_j(a)$ , Eq. (28), this is

$$= - \frac{i}{W(a)} \sum_{j=1}^n \frac{\partial}{\partial a_j} [W(a) v_j(a) \delta(a-a')]. \quad (A6)$$

The final step is to put this into the integral over  $a'$  in Eq. (24), and to evaluate the integral,

$$-i \int da' \langle L\delta(A(x)-a'); a \rangle \frac{g(a'; t)}{W(a')} \\ = - \frac{1}{W(a)} \sum_{j=1}^n \frac{\partial}{\partial a_j} [v_j(a) g(a; t)]. \quad (A7)$$

Now we treat the memory term in exactly the same way. The first  $L$  is brought around, by its Hermitian property, to operate on  $\delta(A(x)-a)$ . This leads to one

$dA/dt$ ,

$$\begin{aligned} & \langle L e^{-is(1-P)L} (1-P) L \delta(A(x) - a'); a \rangle \\ &= - \frac{i}{W(a)} \sum_{j=1}^n \frac{\partial}{\partial a_j} W(a) \langle A_j e^{-is(1-P)L} \\ & \quad \times (1-P) L \delta(A(x) - a'); a \rangle. \end{aligned} \quad (\text{A8})$$

The second  $L$  gives

$$L \delta(A(x) - a') = i \sum_{k=1}^n \dot{A}_k \frac{\partial}{\partial a_k'} \delta(A(x) - a'). \quad (\text{A9})$$

When this is put into the average, we get

$$\begin{aligned} & \frac{1}{W(a)} \sum_i \sum_k \frac{\partial}{\partial a_j} \frac{\partial}{\partial a_k'} W(a) \\ & \quad \times \langle \dot{A}_j e^{-is(1-P)L} (1-P) \dot{A}_k \delta(A(x) - a'); a \rangle. \end{aligned} \quad (\text{A10})$$

In performing the integral over  $a'$ , we make use of

$$\begin{aligned} & \int da' \frac{\partial}{\partial a'} \delta(A - a') g(a') / W(a') \\ &= - \int da' \delta(A - a') \frac{\partial}{\partial a'} [g(a') / W(a')]. \end{aligned} \quad (\text{A11})$$

This leads directly to Eq. (27), with  $K_{jk}$  as defined in Eq. (29).

## APPENDIX II

The mathematical basis for identifying and eliminating higher order terms in the derivation of Eq. (30) is as follows.

We expand the exponential,

$$\begin{aligned} \exp[-is(1-P)L](1-P) &= (1-P) - is(1-P)L(1-P) \\ &+ \frac{1}{2}(is)^2(1-P)L(1-P)L(1-P) - \dots \end{aligned} \quad (\text{A12})$$

Every time a factor  $LP$  appears, we get a term of order  $dA/dt$ , because the projection of a function depends on position  $x$  only through  $A(x)$ , and because  $iLA = dA/dt$ . Consequently, the expansion reduces to

$$\begin{aligned} \exp[-is(1-P)L](1-P) &= (1-P) - is(1-P)L \\ &+ \frac{1}{2}(is)^2(1-P)L^2 + \dots + O(\dot{A}) \\ &= (1-P) \exp(-isL) + O(\dot{A}). \end{aligned} \quad (\text{A13})$$

This leads to

$$K_{jk}(a, a'; s) = \langle \dot{A}_j (1-P) e^{-isL} \dot{A}_k \delta(A - a'); a \rangle + O(\dot{A}^3). \quad (\text{A14})$$

We note further, for example by expansion, that

$$\begin{aligned} \exp(-isL) \dot{A}_k \delta(A - a') \\ = \delta(A - a') \exp(-isL) \dot{A}_k + O(\dot{A}^2). \end{aligned} \quad (\text{A15})$$

Thus the memory function is

$$\begin{aligned} K_{jk}(a, a'; s) &= \frac{1}{W(a)} \int dx \delta(A - a) \{ \delta(A - a') \\ & \quad \times \dot{A}_j (1-P) e^{-isL} \dot{A}_k \} + O(\dot{A}^3) \\ &= \delta(a - a') \langle \dot{A}_j (1-P) e^{-isL} \dot{A}_k; a \rangle + O(\dot{A}^3). \end{aligned} \quad (\text{A16})$$

To this order (but only to this order),  $K_{jk}(a, a'; s)$  has a diagonal singularity in  $a$  space.

As a further simplification, we observe that

$$\begin{aligned} \langle \dot{A}_j (1-P) e^{-isL} \dot{A}_k; a \rangle &= \langle \dot{A}_j e^{-isL} \dot{A}_k; a \rangle \\ &= \langle \dot{A}_j; a \rangle \langle e^{-isL} \dot{A}_k; a \rangle. \end{aligned} \quad (\text{A17})$$

The second term on the right is

$$\begin{aligned} & \int dx \frac{\delta(A - a)}{W(a)} e^{-isL} \dot{A}_k \\ &= \int dx \frac{1}{W(a)} [e^{isL} \delta(A - a)] \dot{A}_k. \end{aligned} \quad (\text{A18})$$

Again we use the expansion of the exponential to show that this becomes

$$\int dx \frac{1}{W(a)} \delta(A - a) \dot{A}_k + O(\dot{A}^2), \quad (\text{A19})$$

or

$$\langle e^{-isL} \dot{A}_k; a \rangle = \langle \dot{A}_k; a \rangle + O(\dot{A}^2). \quad (\text{A20})$$

But

$$\langle \dot{A}_k; a \rangle = v_k(a), \quad (\text{A21})$$

so that

$$\begin{aligned} \langle \dot{A}_j (1-P) e^{-isL} \dot{A}_k; a \rangle &= K_{jk}(a; s) \\ &= \langle [\dot{A}_j - v_j(a)] e^{-isL} [\dot{A}_k - v_k(a)]; a \rangle + O(\dot{A}^3), \end{aligned} \quad (\text{A22})$$

which is the result we have been striving for.