New Approach to Nonequilibrium Processes*

PETER G. BERGMANN AND JOEL L. LEBOWITZ
Department of Physics, Syracuse University, Syracuse, New York
(Received March 23, 1955)

In this paper we propose a new model for the description of irreversible processes, which permits the construction of a Gibbs-type ensemble and the employment of the general techniques of statistical mechanics. The internal dynamics of the system that is engaged in the process is assumed to be described fully by its Hamiltonian. Its interaction with the driving reservoirs is described in terms of impulsive interactions (collisions). The reservoirs themselves possess definite temperatures, are inexhaustible, and free of internal gradients. The ensemble obeys an integro-differential equation in Γ space, containing both the terms of the Liouville equation and a stochastic integral term that describes the collisions with the reservoirs. It is shown in this paper (1) that the ensemble will approach canonical distribution in the course of time in the presence of a single driving reservoir, (2) that it will approach a stationary nonequilibrium distribution in the presence of several reservoirs at different temperatures, and (3) that in the latter case, and for small temperature differences, Onsager's reciprocal relations are satisfied by the stationary distribution.

1. INTRODUCTION

THE purpose of this paper is to propose a statistical-mechanical model for nonequilibrium processes that leads naturally to the existence of a stationary state, approached by an ensemble in the course of time, which is independent of the initial ensemble distribution. In the stationary state, the ensemble is not in thermodynamic equilibrium. It remains in its state only because its surroundings maintain gradients inside the thermodynamic system; in the simplest case, these are temperature gradients.

In statistical mechanics, the concept of temperature is based on the canonical distribution with respect to the energy of the whole system. If the ensemble is not canonical, temperatures may still be definable on canonical or quasi-canonical distributions of partial energies; we speak of local temperature, kinetic temperature, temperature of the nuclear spins, etc. The first pioneering work in the theory of irreversible processes was based on the assumption that such partial temperatures could be defined significantly.[1,2] Other workers succeeded in formulating general rules by adopting modified versions of Boltzmann's assumptions: That is that certain fluctuating variables are always statistically uncorrelated to each other.[2] Later refinements of Boltzmann's "Stosszahlansatz," in which the single-particle occupation numbers are assumed to be uncorrelated, make similar assumptions about higher (n-particle) distribution functions.[3] None of these theories clarifies completely the nature of irreversibility of a process which in its internal dynamics obeys deterministic laws of motion. Some of the assumptions made, i.e., Boltzmann's original assumptions, are demonstrably inconsistent with the classical laws of motion. It appeared impossible, with a closed system, to overcome Zermelo's classical argument, either in classical or in quantum mechanics, according to which the solutions of Liouville's equation are either periodic or almost periodic and, hence, cannot exhibit true irreversibility.[4] The argument that Poincaré cycles have such enormous periods that in practice they cannot be observed is not wholly unsatisfactory. Nevertheless, the fact that a Poincaré cycle may be longer than the total period since the development of the present physical universe (10^9 to 10^{10} years) does not entitle us to the positive conclusion that the specific assumptions made by various workers[2,3] will lead to valid approximations for realistically observable periods of time.

The most satisfactory approach, so it appears to us, is to construct models that permit treatment in terms of Gibbs-type ensembles, in which the openness of the system, and with it the possibility of true irreversible behavior, is explicitly introduced. Klein and Prigogine have considered such a model,[5] which represents a perfect one-dimensional crystal. In this model the mean free path of phonons is infinite; hence there can be no internal gradient. The spectral distribution of phonons is everywhere the same; hence the Klein-Prigogine model does not lend itself to a study of internal gradients and internal dissipation. Our paper will present a new and somewhat more general attempt to construct a system that interacts with its surroundings, in such a fashion that Gibbsian ensembles can be constructed; their development in the course of time, i.e., their dynamics, becomes the subject of mathematical investigation.

For a truly stationary process to be possible, the reservoirs that maintain the gradients within the system must be inexhaustible. They must possess an infinite

[1] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).
[2] For other authors who have used similar approaches leading essentially to macroscopic (phenomenological) laws, see E. W. Montroll and M. S. Green, Ann. Rev. Phys. Chem. **5**, 449 (1954). The same review article also contains an extensive bibliography of work based on different approaches, *viz.*, kinetic theory, Markov chains, etc.
[3] M. Born and H. S. Green, *A General Kinetic Theory of Liquids* (Cambridge University Press, Cambridge, 1949); J. G. Kirkwood, J. Chem. Phys. **14**, 180 (1946); **15**, 72 (1947).
[4] P. and T. Ehrenfest, Encykl. d. Math. Wiss. 4, 2ii, September, 6 (1911).
[5] G. Klein and I. Prigogine, Physica **19**, 74, 89 (1953).

number of degrees of freedom, in such a manner that they present to the system to be considered always the same appearance. Hence their canonical coordinates cannot be fully included in the description of the process in phase space (or in Hilbert space, in a quantum description). What should be included is the (fluctuating) interaction between system and reservoirs. Again, if in this interaction we disregard the effects of fluctuations, i.e., if we simply add to the internal Hamiltonian of the system the interaction energy with the reservoirs, averaged at each point of system-phase space over the possible values of the reservoir coordinates, then the resulting Liouville equation in system-phase space will not differ from one with fixed external forces, and no disspation will result.

How, then, is the action of the reservoir(s) to be introduced? There appear to be two possibilities. They represent, respectively, continuous and impulsive interaction. In both it is assumed that any part of the surroundings that will interact with the system will be statistically independent of the state of the system at the beginning of such interaction. But once the interaction has begun, the statistical independence will be lost; in fact one can show that with a continuous interaction Hamiltonian the correlation function between system and reservoir(s) satisfies a number of differential equations that preclude continued statistical independence. At any instance the interaction term in the Hamiltonian will depend on the previous history. We have been unable as yet to discover a simple mathematical scheme that would represent such finite interaction.

The second possibility is to treat the interaction impulsively. If the reservoir interacts for brief spans of time only, but then strongly, it may be assumed that the net effect of such impulsive interaction will be to move the representative point in system-phase space a finite distance, that depends both on its original location and on the state of the reservoir just prior to interaction. If we further simplify the reservoir by assuming that it consists of a sensibly infinite number of disconnected and similar parts, that each such part interacts with the system but once, and that prior to interaction there is statistical independence between that reservoir component and the system, then we can average over the possible states of the reservoir (i.e., of that portion of the reservoir about to interact impulsively with the system). The result of this averaging operation will be the determination of a contingent probability $K(x',x)dx'dt$, which tells us the probability that the representative point of the system, known to be at the location x (in system-phase space) will be thrown into the volume element dx' (a finite distance from x) within the time interval dt. Once we know the function K, then the Liouville equation for the probability

density μ in system-phase space will take the form

$$\frac{\partial \mu(x,t)}{\partial t} + (\mu, H)_x$$

$$= \int_{x'} [K(x,x')\mu(x') - K(x',x)\mu(x)]dx'. \quad (1.1)$$

The remainder of this paper will be concerned with an investigation of the structure of the function K (Sec. 2), with the approach toward thermal equilibrium in the presence of a single reservoir at a definite temperature (Sec. 3), with a proof that in the presence of several reservoirs ensembles obeying Eq. (1.1) approach a stationary state in the course of time (Sec. 4), and with the Onsager reciprocal relations (Sec. 5).

Before we enter these somewhat more detailed discussions, it may be in order to justify further the form of Eq. (1.1). We are well aware of the fact that impulsive interaction represents an asymptotic limit, which is not always realistic. But it is possible to construct model interactions that approach this limit arbitrarily closely. By assuming this particular form of interaction, and by suppressing the influence that the several components of the reservoir exert on each other, we are able to introduce into the mathematical formulation the notion of the inexhaustible reservoir. We need not neglect recoil (i.e., the reaction of the system back on the reservoir); and we are able to maintain the reservoirs at definite temperatures (or other thermodynamic potentials) without allowing for thermal or other gradients within the reservoirs.

2. STOCHASTIC KERNEL

The expression $K(x',x)$ in Eq. (1.1) represents a transition probability density in the phase space of the system. It is a so-called contingent probability, dealing with the rate of transitions from the state x to the state x' (we shall always write the arguments of such transitions going from right to left) on the assumption that the system is known to be initially in the state x. Hence the value of K will be independent of the probability distribution of the ensemble of systems in the system phase space.

The likelihood of such a transition will depend on the distribution of the reservoir (or reservoir components) as well as on the intrinsic probability of a collision (something like the collision cross section) if both system and reservoir are known to be in suitable states for a collision of the type considered. We shall have to consider both these factors.

As for the distribution of the reservoir, we shall consider it as part of the basic concept of reservoir that its parts are in a canonical distribution, characterized by a definite temperature. Considering the components of such a reservoir in their own phase space (i.e., the phase space of any one of the infinitely many identical reservoir components, usually called the μ space of

such a composite), we may write this distribution in the standard form

$$\mu_r = \frac{1}{Z_r}e^{-E_r/kT} = \frac{1}{Z_r}e^{-\beta E_r}. \qquad (2.1)$$

The subscript r indicates that these quantities belong to a reservoir component (and not to our system).

Next, we come to the much more delicate problem of the intrinsic reaction constant (or collision cross section). All microscopic laws, both classical and quantum theoretical, are reversible with respect to the direction of time. That is to say, there exists a transformation involving (in a suitable coordinate system) a change in the signs of the time coordinate and of all the momentum coordinates with respect to which these laws are invariant. In classical mechanics, the Hamiltonian is an even function of the canonical momentum components; in quantum mechanics (in a wave mechanical representation), the same transformation is accomplished if we change the sign of the time coordinate and at the same time replace the wave function by its complex conjugate. Hence, under this transformation the Hamiltonian remains the same function of its arguments. Each point of phase space (or of Hilbert space) is mapped on another point. We shall denote this mapping or transformation by a bar: $x \rightarrow \bar{x}$. Given a point in classical phase space with the coordinates x (standing for both q_k and p_k), there will be a point $\bar{x}$. If a trajectory in phase space is given by the function $x(t)$, then $\bar{x}(-t)$ is a trajectory, also consistent with the equations of motion.

Microreversibility must also apply to the impulsive interactions between system and reservoir. If a certain initial state of the system x_s and initial state of the reservoir x_r go over into a final state x_s', x_r' through discontinuous impulsive interaction, then the initial state $\bar{x}_s'$, $\bar{x}_r'$ must go over into the final state $\bar{x}_s$, $\bar{x}_r$ by means of the exactly reversed impulsive interaction, and the cross sections for these two corresponding processes must be equal. We may assume that the energy of the reservoir component associated with the point x_r equals the energy associated with the point $\bar{x}_r$, and likewise the energies at the two points x_r' and $\bar{x}_r'$ equal each other. We may, therefore, conclude that the ratio between the transition rates from x to x' and from $\bar{x}'$ to $\bar{x}$, (contingent on the initial state of the system but not on that of the reservoir) equals $\exp[\beta(E_r'-E_r)]$, because of Eq. (2.1). But impulsive interactions conserve the energy; hence the difference in reservoir energies equals the negative difference in system energies. Accordingly, we shall find that the ratio of the contingent probabilities of transition of the system from x_s to x_s' and transition from $\bar{x}_s'$ to $\bar{x}_s$ equals $\exp[\beta(E_s-E_s')]$,

$$K(x',x)=e^{\beta(E_s-E_s')}K(\bar{x},\bar{x}'), \quad E_s=\bar{E}_s, \quad E_s'=\bar{E}_s'. \quad (2.2)$$

This result appears unexceptionable, but it is insufficient to achieve our purpose. We have, therefore, gone one step further and assumed that microreversibility (i.e., equal cross sections) holds not only for the transition from x to x' and the "properly reversed" transition $\bar{x}'$ to $\bar{x}$, but also for the "directly reversed" transition x' to x. This additional assumption is sufficient to assure the monotonic decrease of the Helmholtz potential of a system in thermal contact with a single temperature reservoir until the ensemble of systems becomes canonical. However, our assumption may not be necessary for this result to hold. In fact though one can conceive of many forms of interaction in which the strengthened form of symmetry between a transition and its "direct reverse" holds, because of some special *geometric* symmetry of the model, it appears likely that the assumption made in this paper will have to be generalized, i.e., weakened, so that the approach proposed here is applicable to a wider class of models and phenomena. This possibility is now being investigated.

In what follows we shall accept the strongest form of symmetry of interaction, that is to say, equality between the cross sections of any transition and its direct reverse. If there are several reservoirs present, each with its own peculiar coupling to the system, then the kernel K of Eq. (1.1) will be the sum of such individual kernels K_a,

$$K=\sum_a K_a, \qquad (2.3)$$

each representing the effects of one (the ath) reservoir. The kernel K_a will correspond to both the requirements of strong microreversibility (equality between the cross sections of direct reverses) and of canonical distribution of each reservoir. It will satisfy the equality:

$$K_a(x,x')=e^{\beta_a[H(x')-H(x)]}K_a(x',x),$$
$$K_a(x,x')=e^{\beta_a H(x')}L_a(x,x'), \quad L_a(x,x)=L_a(x'x)\geqslant 0. \qquad (2.4)$$

This condition on the stochastic kernel introduced into the Liouville equation (1.1) represents the point of departure of this paper. In the next section we shall show that with only one reservoir present the assumption (2.3) brings about the monotonic approach of the system governed by Eq. (1.1) toward the canonical distribution.

3. APPROACH TO THERMAL EQUILIBRIUM

We shall first consider an ensemble of systems each of which is in thermal contact with a reservoir at a definite temperature T. In that case the Liouville equation (1.1), with (2.4), assumes the form

$$\frac{\partial \mu}{\partial t}+(\mu,H)=\int_{x'}[e^{\beta H(x')}\mu(x')-e^{\beta H(x)}\mu(x)]L(x,x')dx'. \qquad (3.1)$$

We shall now examine the solutions of this integro-differential equation with arbitrary initial conditions.

Of course, we cannot require that the entropy of the system will always increase as a result of its interaction with the reservoir. After all, energy is being exchanged; depending on initial conditions, it is quite possible that the system on the average transfers energy to the reservoir. We shall show, instead, that as the result of the interaction the total entropy, i.e., the sum of the entropies of the system and the reservoir, will increase. True, we have no detailed description of the reservoir. But as our reservoir is inexhaustible and, at the same time, possesses a definite temperature, it is to be assumed that the change in its entropy equals the amount of energy transferred to it, divided through by its absolute temperature. In other words, the rate of change of total entropy equals

$$dS_{\text{total}}/dt = \sigma = \dot{S} - k\beta \dot{U}. \qquad (3.2)$$

Of course, the Hamiltonian of the system is assumed to be explicitly time-independent, i.e., no work is being performed on the system.

Our conjecture is, then, that the quantity $(S - k\beta U)$ will increase monotonically with time. If we multiply this expression by T and reverse the sign, we obtain the Helmholtz potential F,

$$F = U - ST = \int \mu(H + \beta^{-1} \ln\mu)dx, \qquad (3.3)$$

which is to decrease monotonically in the course of time. We shall now proceed to prove this assertion.

The form of Eq. (3.1) suggests that instead of the variable μ we introduce ν,

$$\nu = \mu e^{\beta H}. \qquad (3.4)$$

In terms of ν, Eqs. (3.1) and (3.3) read

$$\frac{\partial \nu}{\partial t} + (\nu, H) = e^{\beta H} \int_{x'} [\nu(x') - \nu(x)]L(x, x')dx',$$
$$\qquad (3.5)$$
$$F = \beta^{-1} \int_x e^{-\beta H} \nu \ln\nu \, dx,$$

respectively. The time derivative of the Helmholtz potential then becomes

$$\dot{F} = \beta^{-1} \int_x e^{-\beta H}(\ln\nu + 1)\dot{\nu} \, dx$$

$$= \beta^{-1} \int_x \int_{x'} [\ln\nu(x) + 1][\nu(x') - \nu(x)]L(x, x')dx'dx$$

$$+ \beta^{-1} \int_x e^{-\beta H}(\ln\nu + 1)(H, \nu)dx. \qquad (3.6)$$

It can be shown immediately that the second integral on the right-hand side vanishes. We make use of the fact that the integral over a Poisson bracket can be converted into a surface integral through integration by parts and hence vanishes when extended over the whole

phase space, provided only the variables under the integrand satisfy reasonable boundary conditions (the analogous statement in quantum mechanics is that the trace of a commutator vanishes). All we need to do then is to carry out the obvious transformation of the integrand

$$e^{-\beta H}(\ln\nu + 1)(H, \nu) \equiv ((\ln\nu + 1)H, e^{-\beta H}\nu)$$
$$= ((\ln\nu + 1)H, \mu), \qquad (3.7)$$

and the desired result follows as a matter of course. The rate of change in the Helmholtz potential is therefore determined wholly by the first integral on the right-hand side of Eq. (3.6). By interchanging the role of the two sets of variables of integration, x and x', and by adding the resulting expression to the original form of the integral, we obtain this expression for

$$\dot{F} = (2\beta)^{-1} \int_x \int_{x'} [\ln\nu(x) - \ln\nu(x')]$$
$$\times [\nu(x') - \nu(x)]L(x, x')dxdx' \qquad (3.8)$$
$$\leqslant 0.$$

The integrand on the right is manifestly negative-definite. Thus we have shown that the Helmholtz potential is bound to decrease with time. This decrease will be nonzero as long as there is any stochastic interchange between any two regions in phase space at which the ratio of the densities has not yet assumed the canonical value corresponding to the temperature $T = (\beta k)^{-1}$.

It remains to be shown that the Helmholtz potential assumes its absolute minimum for this canonical distribution. To examine the expression (3.3) for minima, we form its variation,

$$\delta F = \int \left[H + \frac{1}{\beta}(\ln\mu + 1) \right]\delta\mu \, dx, \qquad (3.9)$$

but subject to the restriction that the variations are not to violate the normalization requirement,

$$\int \delta u \, dx = 0. \qquad (3.10)$$

It follows that the only distributions stationary with respect to the variation (3.9), (3.10) are proportional, and therefore equal, to the canonical distribution. We can form the second variation as well, in order to determine whether the stationary solution represents a maximum, minimum, or saddle point. We have

$$\delta^2 F = \int [(H + \beta^{-1} \ln\mu)\delta^2\mu + \beta^{-1}(\delta\mu)^2]dx$$

$$= \beta^{-1} \int [(\delta\mu)^2 - \ln Z \delta^2\mu]dx \qquad (3.11)$$

$$= \beta^{-1} \int (\delta\mu)^2 dx > 0.$$

The absence of any additional stationary solutions guarantees that the canonical distribution is not only a relative but the absolute minimum. The value of this minimum is, of course $F = -\beta^{-1} \ln Z$.

Our conclusion is, then, that the Helmholtz potential decreases monotonically in the course of time and that its absolute minimum corresponds to the canonical distribution at the temperature of the (single) reservoir. This result is not quite equivalent to the assertion that the canonical distribution is reached asymptotically. It is possible, first of all, that certain regions of phase space do not interact with each other through the stochastic kernel. Such a contingency is equivalent to the existence of quantities that are absolute constants of the motion, in the sense that they change their value neither under the influence of the internal forces nor under the impact of collisions with the surroundings. Such absolute constants of the motion are known, e.g., the symmetry character of the wave function with respect to indistinguishable particles in quantum mechanics. The existence of such absolute constants of the motion will make it possible to subdivide the phase space into invariant subspaces, each of which is ergodic with respect to the motion of Eq. (3.1).

Another possibility is that the stochastic kernel L converges to zero near certain points or regions and that there is insufficient non-stochastic flux into them, (see the discussion at the end of Sec. 4). This possibility is well known in a special case, namely the interaction of a Knudsen gas with the walls. Molecules of very low velocity take a long time to reach the wall. If the number of collisions per second drops too rapidly with velocity, then the approach to canonical distribution will not be uniform or even certain. Apart from these two possibilities, our system will approach the canonical distribution.

4. APPROACH TO STATIONARY NONEQUILIBRIUM STATES

In this section we shall prove that for a wide class of kernels $K(x,x')$ the distribution $\mu(x,t)$ will approach asymptotically some stationary distribution $\mu_s(x)$, which is independent of the initial distribution $\mu(x,0)$,

$$\lim_{t \to \infty} \mu(x,t) = \mu_s(x). \qquad (4.1)$$

Our proof will be based on a general theorem by Doeblin,[6,7] which applies to all Markoffian processes.

If the initial distribution of the ensemble in phase space corresponds to a definite location, at the point x',

$$\mu(x,0) = \delta(x - x'), \qquad (4.2)$$

we shall denote the distribution at any subsequent time t by the symbol $P(x, x'/t)$. For an arbitrary initial

---

[6] W. Doeblin, Ann. Sci. Ecole Norm. Sup. (3) **57**, 61 (1940).
[7] J. L. Doob, *Stochastic Processes* (John Wiley and Sons, Inc., New York, 1953).

distribution $\mu(x,0)$, we then have

$$\mu(x,t) = \int_{\Gamma} P(x, x'/t)\mu(x',0)dx', \qquad (4.3)$$

$$P(x, x'/t) \geq 0, \quad \int_{\Gamma} P(x, x'/t)dx = 1, \quad t \geq 0. \qquad (4.4)$$

The probability that at the time t the system will be in some particular region in phase space A (i.e., A might be the region between two energy surfaces E_1 and E_2) will be denoted by $\mu(A,t)$; it is given by

$$\mu(A,t) = \int_A \mu(x,t)dx = \int_{\Gamma} P(A, x'/t)\mu(x',0)dx'. \qquad (4.5)$$

The integral

$$P(A, x'/t) = \int_A P(x, x'/t)dx \qquad (4.6)$$

represents the (contingent) probability that a system, initially at the point x', will be found in the domain A at the time t. It is usually called a transition probability. Obviously $P(\Gamma, x'/t) = 1$ for all t. It follows from the time independence of the Hamiltonian, and of the kernel $K(x,x')$ in Eq. (1.1) obeyed by $\mu(x,t)$, that $\mu(x, t+\tau)$ is determined by $\mu(x,\tau)$, again through Eq. (4.3); i.e.,

$$\mu(x, t+\tau) = \int_{\Gamma} P(x, x'/t)\mu(x',\tau)dx'$$

$$= \int \int P(x, x'/t)P(x', x''/\tau)\mu(x'',0)dx''dx'$$

$$= \int P(x, x''/t+\tau)\mu(x'',0)dx''. \qquad (4.7)$$

Because $\mu(x'',0)$ is arbitrary, it follows that

$$P(x, x''/t+\tau) = \int P(x, x'/t)P(x', x''/\tau)dx'. \qquad (4.8)$$

Equation (4.8) is a special case of the Chapman-Kolmogorov equation, which holds for all stationary Markov transition functions:

$$P(A, x''/t+\tau) = \int_{\Gamma} P(A, x'/t)P(dx', x''/\tau). \qquad (4.9)$$

In this equation, $P(dx', x''/\tau)$ is the probability that the system will pass from the initial point x'' to the "infinitesimal" set $(x', x'+dx')$ in time τ, and the integral is a Lebesgue integral. In the case where a transition density exists, $P(dx', x''/t) = P(x', x''/t)dx'$.

If the Hamiltonian of Eq. (1.1) were explicitly time-dependent, then we should have a nonstationary

Markov process, for which Eqs. (4.7), (4.8), and (4.9) do not hold. In that case, there could be no approach to stationary equilibrium in the strict sense, though it is not excluded that in the presence of a periodic Hamiltonian (e.g., spin resonance in the presence of dissipative processes) the solution of Eq. (1.1) may approach periodic behavior. In this paper we are not concerned with that possibility.

We shall now state Doeblin's theorem for the function $P(A, x'/t)$ that satisfies Eq. (4.9): If there exists an $\epsilon > 0$, an $s > 0$, and a set measure $\Psi(A)$ in Γ space, normalized so that $\Psi(\Gamma) = 1$, and if these quantities can be chosen so that

$$P(A, x'/s) < 1 - \epsilon, \qquad (4.10)$$

whenever $\Psi(A) < \epsilon$, for all x', then the limit

$$\lim_{t \to \infty} \{ P(A, x'/t) \}$$

exists. The convergence is uniformly exponentially fast.

When the above conditions for the existence of the limit are satisfied for some time s, then they will also be satisfied for all times $t > s$, because

$$P(A, x'/t) = \int_\Gamma P(A, x''/s) P(dx'', x'/t-s)$$
$$\leqslant \max\{ P(A, x''/s) \} < 1 - \epsilon. \qquad (4.11)$$

Intuitively, the conditions on $P(A, x/t)$ in Doeblin's theorem require that the stochastic process spread distributions that are concentrated initially in regions of measure zero into domains possessing a nonvanishing measure. This requirement excludes, for instance, the case of an isolated system, one that does not interact with a random reservoir. The ensemble distribution of such a system in its phase space obeys the unmodified Liouville equation, which maps distributions initially concentrated in a point on another point without spreading. Such a distribution, as is well known, will never become stationary if it is not stationary to begin with.

The requirement in Doeblin's theorem that $\psi(A)$ be normalized means that the measure of infinitely far regions in phase space is zero with respect to ψ. Hence with any transition function $P(A, x'/t)$ which satisfies the Doeblin conditions we must be able to find a finite region B in Γ space such that the probability of finding the system in B at any time $t > s$ is uniformly bounded away from zero (i.e., the bound is independent of the initial state of the system).

The assumptions made in Eq. (4.10) are sufficient but not necessary conditions for the asymptotic approach to a stationary distribution. As an illustration of this point we may use an example cited by Doob[7] for this purpose. Let

$$P(A, x'/t) = [2\pi(1-\rho^{2t})]^{-\frac{1}{2}} \int_A \exp\left[-\frac{(x-\rho^t x')^2}{2(1-\rho^{2t})} \right] dx,$$
$$0 < \rho < 1. \qquad (4.12)$$

Then

$$\lim_{t \to \infty} P(A, x'/t) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_A \exp(-\tfrac{1}{2}x^2) dx. \qquad (4.13)$$

This convergence is not uniform in x': if A is any finite set, $P(A, x'/t)$ can be made arbitrarily small for any t, by choosing x' sufficiently large. Hence the conditions under which Doeblin's theorem was proved are not satisfied in this case.

We shall now turn to the conditions to be satisfied by the transition rate $K(x,x')$, Eq. (1.1), in order that Doeblin's theorem be applicable to the distribution $P(A, x'/t)$ whose time dependence obeys the generalized Liouville equation (1.1). We shall expand $P(x, x'/t)$ into a series with respect to the number of collisions between the reservoir and the system during the time t.

The system can pass from x' to x in time t without any collisions only if $x = x_t'$, where x_t' is the position at time t of a system initially at x', according to its Hamiltonian equations of motion. The probability of zero collisions, $\eta(x',t)$, is

$$\eta(x',t) = \exp\left(-\int_0^t p(x_s') ds \right), \qquad (4.14)$$

where

$$p(x') = \int_\Gamma K(y,x') dy \qquad (4.15)$$

is the rate at which collisions with the reservoir occur at the point x'. Hence

$$P^0(x, x'/t) = \left[\exp\left(-\int_0^t p(x_s') dx \right) \right] \delta(x-x_t'). \qquad (4.16)$$

The probability that the system will experience exactly one collision in the interval t, given that at the beginning of the interval the system was at x' and at the end of the interval at x, will be a product of the following three probabilities: (1) there was no collision up to some time λ, $0 \leq \lambda \leq t$, which is given by $\eta(x',\lambda)$; (2) there was a collision during the time interval $(\lambda, \lambda + d\lambda)$, which caused the system to make a discontinuous jump from x_λ' to $x_{(\lambda - t)}$ [the probability for this event is $K(x_{(\lambda-t)}, x_\lambda') d\lambda$]; (3) there was no collision between $(t-\lambda)$ and t, $\eta(x, \lambda-t) = \eta(x_{(\lambda-t)}, t-\lambda)$. Since λ is arbitrary, we have to add the different possibilities. Hence

$$P^{(1)}(x, x'/t) = \int_0^t d\lambda \exp\left[-\int_{\lambda-t}^0 p(x_s) ds \right]$$
$$\times K(x_{\lambda-t}, x_\lambda') \exp\left[-\int_0^\lambda p(x_s') ds \right]. \qquad (4.17)$$

By analogous reasoning we find the probability for the system to have two collisions with the reservoir to

be

$$P^{(2)}(x, x'/t)$$

$$= \int_\Gamma dy \int_0^t d\beta \int_0^{t-\beta} d\lambda \left\{ \exp\left[-\int_{\lambda+\beta-t}^0 p(x_s)ds\right] \right.$$

$$\times K(x_{\lambda+\beta-t}, y_\beta) \exp\left[-\int_0^\beta p(y_s)ds\right]$$

$$\left. \times K(y,x_\lambda') \exp\left[-\int_0^\lambda p(x_s')ds\right] \right\}. \quad (4.18)$$

Following up this analysis, we can get all the terms $P^{(n)}(x, x'/t)$. The total transition density $P(x, x'/t)$ is the sum of all these different possibilities

$$P(x, x'/t) = \sum_{n=0}^\infty P^{(n)}(x, x'/t). \quad (4.19)$$

This expansion can also be obtained by means of an entirely different, nonprobabilistic, scheme. It is the result of a perturbation expansion of $\mu(x,t)$ in powers of $K(x,x')$. The same type of perturbation theory is commonly used in quantum mechanics for the wave function $\psi(q,t)$ which, like $\mu(x,t)$, obeys a first-order differential equation in time. Let us rewrite Eq. (1.1) in an operator form thus

$$\partial\mu/\partial t = (H,\mu) - p(x)\mu + \int K(x,x')\mu(x')dx'$$
$$= (\mathfrak{IC}+V)\mu, \quad (4.20)$$

where

$$\mathfrak{IC}\mu = (H,\mu) - p(x)\mu, \quad V\mu = \int K(x,x')\mu(x')dx'.$$

The operator $\mathfrak{IC}$ is singular, but V is regular, bounded, and positive definite, hence we shall expand $\mu(x,t)$ as a power series in V. For this purpose we shall first convert the integro-differential equation for $\mu(x,t)$ into an integral equation,

$$\mu(x,t) = \exp(t\mathfrak{IC})\mu(x,0) + \int_0^t \exp[(t-\lambda)\mathfrak{IC}]V\mu(x,\lambda)d\lambda.$$
$$(4.21)$$

Expanding now in V we get

$$\mu(x,t) = \exp(t\mathfrak{IC})\mu(x,0) + \int_0^t d\lambda\{\exp[(t-\lambda)\mathfrak{IC}]V\exp(\lambda\mathfrak{IC})\}\mu(x,0)$$

$$+ \int_0^t d\gamma\left\{\exp[(t-\gamma)\mathfrak{IC}]V\int_0^\gamma d\lambda[\exp((\gamma-\lambda)\mathfrak{IC})V\exp(\lambda\mathfrak{IC})]\right\}\mu(x,0) + \cdots. \quad (4.22)$$

Now $\exp(t\mathfrak{IC})f(x,0) = f(x,t)$ is the solution of the equation

$$\partial f(x,t)/\partial t = (H,f) - p(x)f. \quad (4.23)$$

More explicitly, this solution has the form

$$f(x,t) = \exp\left[-\int_0^t p(x_{s-t})ds\right]f(x_{-t},0). \quad (4.24)$$

Substituting this expression back into Eq. (4.21) for the case when $\mu(x,0) = \delta(x-x')$ and hence $\mu(x,t) = P(x, x'/t)$, we see that the expansions of Eqs. (4.22) and (4.19) are identical.

We can now prove that if $K(x,x')$ is bounded by an integrable function $v(x)$[8]

$$K(x,x') < Mv(x), \quad \int_\Gamma v(x)dx = 1, \quad (4.25)$$

then our $P(x, x'/t)$ satisfies the Doeblin conditions. There are no additional restrictions imposed on $K(x,x')$ if $v(x)$ is made a function of $H(x)$ alone, $v(x) = v(H(x))$. This last assumption implies that $v(x_t) = v(x)$ for all t.

It now follows from (4.17), (4.18) that

$$P^{(1)}(x, x'/t) < M\int_0^t v(x_{\lambda-t})d\lambda = Mtv(x),$$
$$(4.26)$$

$$P^{(2)}(x,x'/t) < \frac{(Mt)^2}{2!}v(x),$$

and in general

$$P^{(n)}(x, x'/t) < \frac{(Mt)^n}{n!}v(x), \quad n \geq 1. \quad (4.27)$$

Hence, from (4.19),

$$P(x, x'/t) < \exp\left[-\int_0^t p(x_s')ds\right]\delta(x-x_t')$$
$$+ (e^{Mt}-1)v(x). \quad (4.28)$$

In order that our system should approach a stationary state independent of its initial position, x', in phase space, we must, of course, assume that there are no trajectories in Γ space on which the system can move without ever getting into a region where it can collide with the reservoir. Otherwise if the system is initially in a state x' which lies on such a trajectory, it would

---

[8] This condition is somewhat more restrictive than absolutely necessary but it simplifies the proof considerably.

behave as if there were no reservoir, and the density $\mu(x,t)$ would always be $\delta(x-x_t')$ without ever becoming stationary. More precisely, we assume that there exists a time t_1, such that for any time $t>t_1$, the system will spend at least an interval τ, $\tau=\alpha(t-t_1)$, $\alpha>0$ in such a region in phase space where the probability per unit time, $p(x)$, of its undergoing a collision is greater than some positive number m. Hence

$$P(x, x'/t) < \exp[-m\alpha(t-t_1)]\delta(x-x_t')$$
$$+ (e^{Mt}-1)v(x), \quad t>t_1, \quad (4.29)$$

and

$$P(A, x'/t) < \exp[-m\alpha(t-t_1)]+(e^{Mt}-1)\Psi(A), \quad (4.30)$$

where

$$\Psi(A) = \int_A v(x)dx. \quad (4.31)$$

When $\Psi(A)<\epsilon$, then

$$P(A, x'/t) < \exp[-m\alpha(t-t_1)]+(e^{Mt}-1)\epsilon, \quad (4.32)$$

which is less than $(1-\epsilon)$ whenever

$$\epsilon < e^{-Mt}(1-\exp[-m\alpha(t-t_1)]), \quad t>t_1. \quad (4.33)$$

Thus whenever $K(x,x')$ satisfies (4.25), $P(A, x'/t)$ satisfies (4.10); Doeblin's theorem applies, and

$$\lim_{t\to\infty} P(A, x'/t) = P(A,x') \quad (4.34)$$

exists, The limiting distribution $P(A,x')$ will be independent of x' if there are no noninteracting regions in phase space. Otherwise, the stationary state will depend on the initial distribution between these invariant subspaces. This requirement of ergodicity is the same as in the case of a single reservoir discussed at the end of Sec. 3.

5. STATIONARY STATE WITH SEVERAL RESERVOIRS: ONSAGER RELATIONS

We shall now consider an ensemble of systems in contact with n reservoirs at temperatures $T_i=(k\beta_i)^{-1}$, $i=1, \cdots, n$. Equation (1.1) then assumes the form

$$\frac{\partial\mu}{\partial t}+(\mu,H)=\sum_{i=1}^{n}\int_{x'}[e^{\beta_iH(x')}\mu(x')-e^{\beta_iH(x)}\mu(x)]$$
$$\times L_i(x,x')dx'. \quad (5.1)$$

The rate of entropy production σ will be given, in analogy with Eq. (3.2), by

$$\sigma = \dot{S} - k\sum_{i=1}^{n}\beta_iJ_i, \quad (5.2)$$

where J_i is the average rate of energy flow from the ith reservoir to the system. The rate of change of the mean

energy U is equal to the sum of the flows

$$\dot{U} = \int_\Gamma H\frac{\partial\mu}{\partial t}dx = \sum_{i=1}^{n}J_i = \sum_{i=1}^{n}\int_x dx$$

$$\times \int_{x'} dx'[e^{\beta_iH(x')}\mu(x')-e^{\beta_iH(x)}\mu(x)]H(x)L_i(x,x')$$

$$= \sum_{i=1}^{n}\int_x dx\mu(x)\int_{x'} dx' \, e^{\beta_iH(x)}L_i(x,x')[H(x')-H(x)], \quad (5.3)$$

$$J_i = \int_x dx\mu(x)\int_{x'} dx' e^{\beta_iH(x)}L_i(x,x')[H(x')-H(x)]$$
$$= \int_x dx\mu(x)f_i(x). \quad (5.4)$$

It can be shown by methods analogous to those used in the derivation of Eq. (3.8) that $\sigma\geqslant 0$. The equality can hold only if all the β_i are equal to each other.

When the ensemble becomes stationary, $\partial\mu/\partial t=0$, the internal entropy remains constant, $\dot{S}=0$, and Eq. (5.4) reduces to

$$\sigma = -k\sum_{i=1}^{n}\beta_iJ_i\geqslant 0. \quad (5.5)$$

Since in the stationary state the mean energy U is constant, it follows that

$$\dot{U} = \sum_{i=1}^{n}J_i=0. \quad (5.6)$$

Here we may solve for some particular flow J_m in terms of the other flows

$$J_m = -\sum_{i\neq m}' J_i \quad (5.7)$$

in a stationary ensemble. Substituting this expression into Eq. (5.5), we get

$$\sigma = k\sum_{i\neq m}'(\beta_m-\beta_i)J_i=\sum_{i\neq m}' \gamma_iJ_i(\gamma_i), \quad (5.8)$$

where the J_i are now independent flows.

It follows from this expression for the total entropy production that in our model, the J_i and the γ_i, are the currents and the forces respectively, which are used in the Onsager theory.[1] Onsager's reciprocal relations state that in the linear approximation the dependence of J_l on γ_k is the same as that of J_k on γ_l. In other words, when the reservoir temperatures are close to each other (i.e., where the γ_i's are small), when we can set

$$J_l = \sum_{k\neq m}' A_{lk}\gamma_k, \quad (5.9)$$

then

$$A_{lk}=\left(\frac{\partial J_l}{\partial \gamma_k}\right)_{\gamma_i=0}=\left(\frac{\partial J_k}{\partial \gamma_l}\right)_{\gamma_i=0}=A_{kl}. \quad (5.10)$$

We shall now prove the validity of Eq. (5.10) for our model. Since the J_m chosen in Eq. (5.7) was arbitrary, the symmetry relations should be independent of that choice. They will be if and only if

$$J_{lk}=\left(\frac{\partial J_l(\beta_i)}{\partial \beta_k}\right)_{\beta_i=\beta}=\left(\frac{\partial J_k(\beta_i)}{\partial \beta_l}\right)_{\beta_i=\beta}=J_{kl}, \quad (5.11)$$

for $l, k=1, \cdots, m, \cdots n$. We shall, therefore, prove Eq. (5.11) rather than Eq. (5.10). Generally, $J_{lk}=-A_{lk}$, $l, k \neq m$.

From Eq. (5.3), we get

$$\frac{\partial J_l}{\partial \beta_k}=\int_x dx\left[\frac{\partial \mu(x)}{\partial \beta_k}f_l(x)\right], \quad l\neq k, \quad (5.12)$$

since $f_l(x)$ depends only on β_l. Solving for $\partial \mu/\partial \beta_k$, in the stationary state, from the time independent form of Eq. (5.1), we get

$$\left(\frac{\partial \mu}{\partial \beta_k}, H\right)=\sum_{i=1}^{n}\int_{x'}\left\{e^{\beta_i H(x')}\frac{\partial \mu(x')}{\partial \beta_k}\right.$$
$$-e^{\beta_i H(x)}\frac{\partial \mu(x)}{\partial \beta_k}\right\}L_i(x,x')dx'$$
$$+\int_{x'}\left\{\mu(x')e^{\beta_k H(x')}\left[H(x')L_k(x,x')+\frac{\partial L_k}{\partial \beta_k}\right]\right.$$
$$\left.-\mu(x)e^{\beta_k H(x)}\left[H(x)L_k+\frac{\partial L_k}{\partial \beta_k}\right]\right\}dx'. \quad (5.13)$$

We shall now evaluate this equation when all the β_i's equal to β. Under these conditions $\mu(x)=(1/Z)e^{-\beta H(x)}$, as derived in Sec. 3, and Eq. (5.13) reduces to

$$(\mu_k,H)=\int_x[\mathcal{L}(x,x')e^{\beta H(x')}\mu_k(x')]dx'-q(x)\mu_k(x)$$
$$+\frac{1}{Z}e^{-\beta H(x)}f_k{}^0(x), \quad (5.14)$$

where

$$\mu_k=\left(\frac{\partial \mu}{\partial \beta_k}\right)_{\beta_i=\beta}, \quad \mathcal{L}(x,x')=\sum_{i=1}^{n}[L_i(x,x')]_{\beta_j=\beta}=\mathcal{L}(x',x),$$

$$q(x)=\int_{x'}\mathcal{L}(x,x')e^{\beta H(x)}dx'=p(x)\,|\,\beta_i=\beta.$$

Multiplying Eq. (5.14) by $Ze^{\beta H(x)}\mu_l(x)$ and integrating

over Γ, we get

$$J_{kl}=\int_x f_k{}^0(x)\mu_l(x)dx$$
$$=Z\int_x\left\{e^{\beta H(x)}\mu_l(x)(\mu_k(x),H(x))\right.$$
$$-\int_{x'}\mathcal{L}(x,x')e^{\beta[H(x)+H(x')]}\mu_k(x')\mu_l(x)dx'$$
$$\left.+\mu_k(x)\mu_l(x)q(x)e^{\beta H(x)}\right\}dx. \quad (5.15)$$

Hence

$$J_{kl}-J_{lk}=Z\int_x e^{\beta H(x)}[\mu_l(x)(\mu_k(x),H(x))$$
$$-\mu_k(x)(\mu_l(x),H(x))]dx$$
$$=2Z\int_x e^{\beta H}\mu_l(\mu_k,H)dx$$
$$-\beta^{-1}Z\int_x (\mu_k\mu_l,e^{\beta H})dx. \quad (5.16)$$

The last term vanishes. In order to evaluate the remainder in (5.16), it is convenient to separate $\mu_i(x)$ into a symmetric and an antisymmetric part with respect to the momentum variables of the system:

$$\mu_i(x)=E_i(x)+O_i(x), \quad E_i(x)=E_i(\bar{x}), \quad O_i(x)=-O_i(\bar{x}). \quad (5.17)$$

A Poisson bracket between even functions is odd, and between an even and an odd function even, etc. Because H is an even function, its Poisson bracket with $E_i(x)$ is odd, and with $O_i(x)$ even. Since the integral of an odd function over the whole phase space vanishes, it follows that

$$J_{kl}-J_{lk}=2Z\int_\Gamma[E_l(O_k,H)+O_l(E_k,H)]e^{\beta H}dx$$
$$=2Z\int_\Gamma[O_l(e^{\beta H}E_k,H)-O_k(e^{\beta H}E_l,H)]dx. \quad (5.18)$$

We can now solve for $(e^{\beta H}E_k,H)$ from Eq. (5.14), remembering that it follows from the way we arrive at Eq. (2.3) that $L(x,x')=L(\bar{x},\bar{x}')$. Thus

$$(e^{\beta H}E_k,H)=\int_{x'}\mathcal{L}(x,x')e^{\beta[H(x)+H(x')]}O_k(x')dx'$$
$$-e^{\beta H(x)}q(x)O_k(x). \quad (5.19)$$

Substituting (5.19) into (5.18) we arrive at our desired result,

$$J_{kl}-J_{lk}=0. \quad (5.20)$$

This derivation of the Onsager relations requires no reference to fluctuation theory, nor does it involve an

assumption that the laws connecting the macroscopic variables are Markoffian.

6. CONCLUSION

In this paper we have developed in classical (non-quantum) mechanics a model for a stationary irreversible process that is capable of being treated with the methods of statistical mechanics due to Gibbs. Our model is based on the assumption that no essential features of the real process are lost if the interaction of the system with the driving reservoirs is pictured in terms of instantaneous impulsive interactions. The reservoirs themselves are described as infinitely large composites consisting of identical noninteracting components in canonical distribution. Thus each reservoir has a definite temperature, infinite heat capacity, and

vanishing internal heat conductivity. No special assumptions are made concerning the structure of the system; its internal dynamics are governed by some nonsingular Hamiltonian.

We have succeeded in showing that our model, with arbitrary initial ensemble distribution, will approach the canonical distribution if driven by a single reservoir, will approach a stationary (noncanonical) distribution if driven by several reservoirs at different temperatures, and in the stationary state will obey the Onsager relations if the driving temperature gradients are small.

Further work will be devoted to a more detailed investigation of the stochastic kernels that represent the action of the reservoirs, the introduction of more general thermodynamic forces than temperature gradients, and the transition to quantum mechanics.

Lagrangians Linear in the "Velocities"*

Ezra Newman† and Peter G. Bergmann
Department of Physics, Syracuse University, Syracuse, New York
(Received March 16, 1955)

Lagrangian linear in the first time derivatives are of sufficient importance in physics (particle fields, general relativity in the Palatini formulation, Einstein-Strauss type unified field theory, etc.) to warrant special consideration. Our treatment is patterned after Dirac's more general exploration of Lagrangians leading to algebraic relations between the canonical variables. In our case, the number of such constraints is at least as large as the number of configuration coordinates. The secondary constraints are free of canonical momentum densities. We have examined all the possibilities that may arise—incompatibility of the field equations, proper Cauchy-Kowalewski problems, and the appearance of arbitrary functions in the solutions. Appropriate quantization procedures for the compatible cases will be indicated.

1. INTRODUCTION

MANY physical theories are derivable from Lagrangians that are linear in the first time derivatives of the field variables. We have investigated in this paper the compatibility of the field equations of such a theory, methods of constructing a Hamiltonian formalism, and quantization procedures. The formalism developed is capable of handling such diverse theories as the Pauli-Fierz equations, gravitational theory in the Palatini form (i.e., considering the components of the affine connection as independent variables), the Einstein-Strauss unified field theory, and Maxwell theory with the vector potentials treated as variables independent of **E** and **H**.

Our treatment is based on Dirac's[1] study of theories for which the momenta canonically conjugate to the field variables are not all algebraically independent of the field variables and their spatial derivatives themselves. Dirac reduces all cases to Lagrangians that are

homogeneous of the first degree in the "velocities" (derivatives of the field variables with respect to the chosen time coordinate); any Lagrangian can be given this form by the device of parametrization. Lagrangians that are (inhomogeneously) linear in the velocities do not require this treatment. They can be discussed quite successfully without the introduction of a parameter. They are of sufficient importance in physics that they warrant a specialized treatment.

There are two general types of linear Lagrangians (in our sense). They all lead to differential equations that are free to accelerations and are linear in the velocities. The first type leads to equations in which the matrix of the coefficients of the velocities is nonsingular, the second to equations in which the same matrix is singular. The first case can be treated completely and in full generality. The second case has a number of subcases. Whenever the matrix of the coefficients is singular, then there exists a number of linear combinations of the (Lagrangian field) equations that are free of velocities. These combinations may be empty; if they are not, their time derivatives may be independent of the original field equations. In the latter case, new combinations

---

[1] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); **3**, 1 (1951).