

Variational Derivation of the Steady State

RYOICHI KIKUCHI

Hughes Research Laboratories, Malibu, California

(Received November 3, 1960; revised manuscript received July 27, 1961)

For a system undergoing Markoffian irreversible change of state, a concept of *persistence* (shown to arise naturally from a physical microscopic description of the system) is introduced; it is defined as the probability that a certain state persists during a short time interval without drifting away from the state. This concept is used in deriving a theorem which says that "the steady state of a Markoffian irreversible process corresponds to the maximum of the persistence." This theorem holds even outside the range of validity of Prigogine's principle of minimum entropy production, namely, even when the steady state is not close to the equilibrium state or even when the process is nonlinear. The proposed function is derived from the path probability for irreversible statistical dynamics previously reported by the author. The theorem is first applied to a system of two level atoms interacting simultaneously with a heat bath and a radiation field. Next, Kohler's variational derivation of the Boltzmann transport equation is shown to follow from the theorem.

I. INTRODUCTION

THE question, "which state variable has its extremum when the stationary state is reached?," first recorded by Ehrenfest,¹ is partly answered by Prigogine's principle² of minimum entropy production ($\min dS/dt$). This is not a complete answer, however, because the validity of this principle is restricted by the following conditions³⁻⁹:

- (1) Phenomenological laws are linear,
- (2) Onsager's reciprocal relations¹⁰ hold, and
- (3) Phenomenological coefficients are constant.

A large amount of effort was spent by many workers in trying to find a better variational principle which is free from these restrictions, as are evidenced by the large number of publications on this problem.¹¹⁻²²

¹ P. and T. Ehrenfest, *Encyk. Math. Wiss.* Vol. 4, Heft 6, p. 82 (1911). See also S. R. de Groot, *Thermodynamics of Irreversible Processes* (Interscience Publishers, Inc., New York, 1952), Chap. X.

² I. Prigogine, *Etude Thermodynamique des Phénomènes Irréversibles* (Editions Desoer, Liège, 1947).

³ K. G. Denbigh, *Trans. Faraday Soc.* **48**, 389 (1952).

⁴ A. E. Nielsen, *Acad. roy. Belg., Bull. classe Sci.* **40**, 539 (1954).

⁵ M. J. Klein and P. H. E. Meijer, *Phys. Rev.* **96**, 250 (1954).

⁶ I. Prigogine, *Thermodynamics of Irreversible Processes* (C. C. Thomas, Springfield, Illinois, 1955).

⁷ M. J. Klein, *Transport Processes in Statistical Mechanics*, edited by I. Prigogine (Interscience Publishers, Inc., New York, 1958), p. 311.

⁸ M. J. Klein, *Termodinamica dei Processi Irreversibili* (Zanichelli, Bologna, 1960), p. 198.

⁹ H. B. Callen, *Phys. Rev.* **105**, 360 (1957); *Transport Processes in Statistical Mechanics*, edited by I. Prigogine (Interscience Publishers, Inc., New York, 1958), p. 327.

¹⁰ L. Onsager, *Phys. Rev.* **37**, 405 (1931); **38**, 2265 (1931).

¹¹ R. Haase, *Z. Naturforsch.* **6a**, 522 (1951); *Ergeb. der exakt. Naturw.* **26**, 56 (1954).

¹² P. Glansdorff, *Physica* **19**, 737 and 1029 (1953); *Acad. roy. Belg. Bull. classe Sci.* **39**, 971 (1953).

¹³ I. Prigogine, *Acad. roy. Belg. Bull. classe Sci.* **40**, 471 (1954).

¹⁴ I. Prigogine and H. C. Mel, *Acad. roy. Belg. Bull. classe Sci.* **40**, 588 (1954).

¹⁵ H. C. Mel, *Acad. roy. Belg. Bull. classe Sci.* **40**, 834 (1954).

¹⁶ P. Glansdorff and I. Prigogine, *Physica* **20**, 773 (1954).

¹⁷ I. Prigogine and R. Balescu, *Acad. roy. Belg. Bull. classe Sci.* **41**, 917 (1955); **42**, 256 (1956).

¹⁸ H. Wergeland, *Transport Processes in Irreversible Statistical Mechanics*, edited by I. Prigogine (Interscience Publishers, Inc., New York, 1958), p. 324.

The present paper answers this question by applying the path probability method of statistical dynamics for Markoffian systems proposed by the author previously.²³ In Sec. II the concept of persistence is introduced and a theorem is proposed that the persistence is the variational function sought after. This theorem is free from any one of the three restrictions listed above. In Sec. III the theorem is applied to a simple problem of two-level atoms analyzed by Klein.^{7,8} In Sec. IV, the case of electrons in a crystal under an external electric field is discussed to show that Kohler's variational derivation^{20,24} of the Boltzmann transport equation results from this theorem.

II. PERSISTENCY

In order to prepare for the derivation of a function which takes a maximum value in the steady state, the path probability method²³ developed previously will be summarized briefly. Consider an assembly composed of L systems, L being an arbitrarily large number. States of a system are labeled by $i=1, 2, \dots$. A state of the assembly at time t is specified by a set of probability parameters $\{p_i(t)\}$ ($i=1, 2, \dots$) such that $Lp_i(t)$ is the number of systems (in the assembly) which are in the state i . The set of numbers $\{p_i(t)\}$ may be regarded as a vector and will be denoted by $\mathbf{p}(t)$, interchangeably.

The process in which the state of the assembly changes during a short time interval τ from t to $t+\tau$ is described by a set of parameters $\{P_{ijr}(t; t+\tau)\}$; $LP_{ijr}(t; t+\tau)$ is the number of such systems (in the

¹⁹ P. H. E. Meijer and J. I. Bowen, *Physica* **26**, 478 (1960); J. I. Bowen and P. H. E. Meijer, *ibid.* **26**, 485 (1960).

²⁰ M. Kohler, *Z. Physik* **124**, 772 (1948).

²¹ J. M. Ziman, *Can. J. Phys.* **34**, 1256 (1956).

²² S. Ono, *Sci. Papers Coll. Gen. Educ., Univ. Tokyo* **5**, 87 (1955). See also S. Ono, *Advances in Chemical Physics*, edited by I. Prigogine (Interscience Publishers, Inc., New York, 1961) Vol. **3**, p. 267.

²³ R. Kikuchi, *Ann. of Phys.* **10**, 127 (1960). This paper will be referred to as ICP. Applications of the method will be found in the following publications: R. Kikuchi, *Ann. Phys.* **11**, 306, 328 (1960); J. Phys. Chem. Solids **20**, 17, 35 (1961).

²⁴ A. H. Wilson, *Theory of Metals* (Cambridge University Press, New York, 1956), Chap. X.

assembly) that were in the state i at t and are in the state j at $t+\tau$, the transition having occurred via the r th mechanism.²⁵ The definition of p_i and P_{ijr} leads to the compatibility relations:

$$\begin{aligned} p_i(t) &= \sum_j \sum_r P_{ijr}(t; t+\tau), \\ p_j(t+\tau) &= \sum_i \sum_r P_{ijr}(t; t+\tau). \end{aligned} \quad (2.1)$$

The probability $p_i(t)$ will be called a *state* parameter and $P_{ijr}(t; t+\tau)$ a *path* parameter, since the former specifies a state of the assembly and the latter a path for the change of a state to another.

The path probability method for Markoffian systems²³ provides the recipe of writing the conditional probability²⁶ $\mathcal{G}\{P_{ijr}(t; t+\tau)\}$ that any one of the paths specified by the path parameters $\{P_{ijr}(t; t+\tau)\}$ is realized when the initial state $\{p_i(t)\}$ is known. Namely, as will be seen in the subsequent sections and also in the Appendixes, $\mathcal{G}\{P_{ijr}(t; t+\tau)\}$ is written as a product of the following two factors: the number of different possible paths consistent with the set of path parameters $\{P_{ijr}(t; t+\tau)\}$, given the initial state $\{p_i(t)\}$; and the *a priori* probability for a certain path whose path parameters are $\{P_{ijr}(t; t+\tau)\}$. The general concept behind \mathcal{G} is similar to equilibrium statistical mechanics; when the partition function is written as $\sum_{\epsilon} G(\epsilon) \times \exp(-\epsilon/kT)$, a term in the sum is proportional to the probability that any one of the states of energy ϵ is realized and is composed of two factors, $G(\epsilon)$, which is the number of different states consistent with the energy, and $\exp(-\epsilon/kT)$, which can be regarded as the *a priori* probability (unnormalized) for a certain state of energy ϵ . It is also instructive to observe that the path probability method is an extension of the cluster-variation treatment²⁷ of one-dimensional order-disorder problem; when the spatial axis in the latter is regarded as the time axis, the construction of the free energy in the equilibrium theory is directly translated into the construction of the conditional path probability $\mathcal{G}\{P_{ijr}(t; t+\tau)\}$. This point is further discussed in Sec. V.

Starting from the path probability thus constructed, which we will write as $\mathcal{G}^{(I)}\{P_{ijr}(t; t+\tau)\}$ in order to distinguish it from $\mathcal{G}^{(II)}$ and $\mathcal{G}^{(III)}$ to be introduced below, we ask the most probable path for the short time interval τ when the initial state $\{p_i^{(1)}(t)\}$ and the final state $\{p_j^{(2)}(t+\tau)\}$ are specified.²⁸ This most probable path is obtained by maximizing $\mathcal{G}^{(I)}$, and this maximum value of $\mathcal{G}^{(I)}\{P_{ijr}(t; t+\tau)\}$ corresponding to

²⁵ There can be more than one mechanism of changing from the state i to j , for instance by exchanging energies with either one of many heat reservoirs in contact with the system. See Sec. III.

²⁶ In defining \mathcal{G} in ICP, the initial state $\{p_i(t)\}$ was not specified. The definition adopted here is more convenient for the present work.

²⁷ R. Kikuchi, Phys. Rev. **81**, 988 (1951).

²⁸ The subscript i on p indicates the i th "component" of \mathbf{p} , or the probability of finding a system in the i th state, whereas the superscript (1) labels the whole vector, or the whole set of numbers.

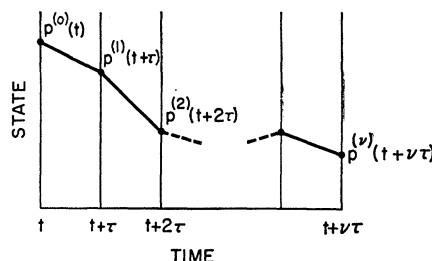


FIG. 1. A sequence of states of the assembly (schematic).

the most probable path will be written as $\mathcal{G}^{(II)}[\mathbf{p}^{(1)}(t); \mathbf{p}^{(2)}(t+\tau)]$, or $\mathcal{G}^{(II)}[\mathbf{p}^{(1)}; \mathbf{p}^{(2)}]$ for short; here the vectors $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(2)}$ stand for the sets $\{p_i^{(1)}\}$ and $\{p_j^{(2)}\}$, respectively, and t and $t+\tau$ in the arguments of $\mathcal{G}^{(II)}$ will not be written explicitly when they are understood. Hence $\mathcal{G}^{(II)}[\mathbf{p}^{(1)}(t); \mathbf{p}^{(2)}(t+\tau)]$ is the probability of going from $\mathbf{p}^{(1)}(t)$ to $\mathbf{p}^{(2)}(t+\tau)$ in a short time interval τ along the most probable path.

Consider a certain sequence of states (of the assembly), $\mathbf{p}^{(0)}(t), \mathbf{p}^{(1)}(t+\tau), \mathbf{p}^{(2)}(t+2\tau), \dots$, for a finite time interval as illustrated in Fig. 1. The probability that the assembly follows this sequence, or the path, is written as

$$\mathcal{G}[\mathbf{p}^{(0)}(t); \mathbf{p}^{(1)}(t+\tau); \dots] = \prod_{\nu} \mathcal{G}^{(II)}[\mathbf{p}^{(\nu)}; \mathbf{p}^{(\nu+1)}]. \quad (2.2)$$

If the assembly approaches a unique steady state described by $\mathbf{p}^{(s)}$ asymptotically as the natural development (see Fig. 2), we can approximate for a sufficiently long time interval $N\tau$,

$$\mathcal{G}[\text{Natural path for the time interval } N\tau] \cong \{\mathcal{G}^{(III)}[\mathbf{p}^{(s)}; \mathbf{p}^{(s)}]\}^N, \quad (2.3)$$

because the initial transient part of the path has little effect. Here $\mathcal{G}^{(III)}$ is a special case of $\mathcal{G}^{(II)}$ when the two arguments of the latter are identical.

The natural development of the physical system corresponds to the largest probability; in other words, the natural path \mathcal{G} in (2.3) is greater than \mathcal{G} for any path other than the natural one, provided that the natural path leads to a unique steady state.²⁹ As a

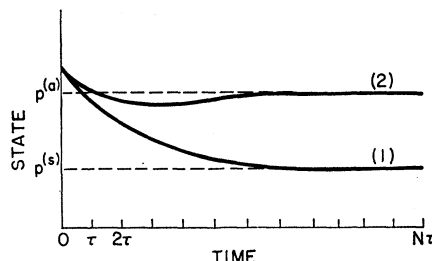


FIG. 2. A schematic diagram of the natural path (1) whose steady state is $\mathbf{p}^{(s)}$, and an artificial path (2) which approaches a state $\mathbf{p}^{(a)}$ asymptotically.

²⁹ It is possible that the natural path leads to an oscillation of \mathbf{p} and does not approach a unique state $\mathbf{p}^{(s)}$. However, this kind of oscillating case is excluded from the present consideration. Examples of the oscillating or the cyclic processes are discussed in reference 17.

special choice, when the natural path is compared with an artificial path which approaches a certain state $\mathbf{p}^{(a)}$ asymptotically (Fig. 2), the maximum property of the natural path stated above has the consequence:

$$\{\mathcal{G}^{(\text{III})}[\mathbf{p}^{(a)}; \mathbf{p}^{(a)}]\}^N \leq \{\mathcal{G}^{(\text{III})}[\mathbf{p}^{(s)}; \mathbf{p}^{(s)}]\}^N. \quad (2.4)$$

Since $\mathcal{G}^{(\text{III})}[\mathbf{p}; \mathbf{p}]$ is the variational function playing the central role in answering the question discussed in Sec. I, it seems meaningful to give it a name. This quantity is the conditional probability of going from the state (of the assembly) at t specified by the vector \mathbf{p} (or by a set of state parameters $\{p_i\}$) to the same state \mathbf{p} at time $t+\tau$; in other words, it is the probability that the system stays in a particular state \mathbf{p} during τ without drifting away from it.³⁰ Therefore it is probably appropriate to call $\mathcal{G}^{(\text{III})}$ the *persistence*. Then the inequality (2.4) is the proof of the following theorem:

"If the assembly possesses a unique steady state, it is the state for which the *persistence* is a maximum."

This is an answer to Ehrenfest's question.^{1,31}

III. THE TWO-LEVEL ATOM PROBLEM

As an example let us discuss the problem which was treated by Klein^{7,8} and was analyzed in Sec. III of ICP.²³ The assembly consists of L atoms, each of which has two energy levels 0 and ϵ . It is in contact with a heat bath of temperature T and also with a radiation field with a monochromatic radiation of frequency ϵ/h . An atom can make a transition between its two states either by exchanging energy with the heat bath or with the radiation field.

The same notation as in Sec. III of ICP will be used. The normalized probability for finding an atom in the lower and upper state at time t will be denoted by $p_l(t)$ and $p_u(t)$, respectively. For the change of state, the path parameters are introduced. In the notation of the previous section, they are P_{ll} , P_{lu} , P_{ul} , P_{uu} , P_{ulr} , and P_{uur} . Here, the subscripts r and b indicate that energy is exchanged with the radiation field or with the heat bath, respectively. These path parameters were written in ICP as P_i ($i=1, 2, \dots, 6$) in this order. The compatibility relations are

$$p_l(t) = P_1 + P_2 + P_3, \quad (3.1)$$

$$p_u(t) = P_4 + P_5 + P_6,$$

$$p_u(t+\tau) - p_u(t) - P_2 - P_3 + P_4 + P_5 = 0, \quad (3.2)$$

where P_i is used for $P_i(t; t+\tau)$.

As was shown in ICP, and as will be discussed further in the Appendixes, the conditional path probability is written as

³⁰ Although a finite interval τ is used as a unit of time, the discreteness along the time axis does not do any harm, since τ can be taken arbitrarily small.

³¹ It may be noted here that persistence is a quantity derived from microscopic statistical dynamics and not a quantity definable within macroscopic thermodynamics of irreversible processes.

$$\begin{aligned} & \mathcal{G}^{(\text{I})}\{P_i(t; t+\tau)\} \\ &= [Lp_l(t)]! [Lp_u(t)]! \prod_{i=1}^6 [LP_i(t; t+\tau)]!^{-1} \\ & \times (\tau\theta_b)^{L(P_3+P_4)} (\tau\theta_r)^{L(P_2+P_5)} (1-\tau\theta_1)^{LP_1} (1-\tau\theta_6)^{LP_6} \\ & \times \exp[L\epsilon\beta(P_4-P_3)/2]. \quad (3.3) \end{aligned}$$

Here the combinatorial factors represent the number of different paths specified by $\{P_i\}$ given the initial state, and the remainder may be considered as the *a priori* probability that one of the paths occurs. Further, $\theta_b(\theta_r)$ is the quantum-mechanical transition probability between the two levels by exchanging energy with the heat bath (with the radiation field), and $\theta_1(\theta_6)$ is the probability that an atom in the lower state (the upper state) makes a transition in unit time by either of the transition mechanisms. For the path from time t to $t+\tau$, θ_1 and θ_6 do not depend on $\{p_i(t+\tau)\}$,³² and are determined in terms of known quantities (at time t) from the relations

$$\begin{aligned} \tau\theta_1 &= [P_2^{(n)}(t; t+\tau) + P_3^{(n)}(t; t+\tau)]/p_l(t), \\ \tau\theta_6 &= [P_4^{(n)}(t; t+\tau) + P_5^{(n)}(t; t+\tau)]/p_u(t), \end{aligned} \quad (3.4)$$

where $P_i^{(n)}(t; t+\tau)$ is the path probability for the natural path for the short time interval τ ; the natural path is the path the assembly takes as the natural development and is derived mathematically by maximizing the path probability, keeping the initial state (but not the final state) fixed, as will be shown.

In order to apply the theorem proved in the previous section, $\mathcal{G}^{(\text{I})}\{P_i(t; t+\tau)\}$ in (3.3) is first maximized with respect to P_i ($i=2, 3, 4$, and 5) under the subsidiary condition (3.2), P_1 and P_6 being regarded as the dependent variables through (3.1). It may be noted that $p_l(t)$ and $p_u(t)$ are fixed and θ_b , θ_r , θ_1 , and θ_6 are constant. Multiplying (3.2) by the Lagrange multiplier λ , adding the product to (3.3), and taking the variation of the sum with respect to the four independent variables, we obtain

$$\begin{aligned} P_2 &= \tau\theta_r e^{-\lambda} P_1 / (1 - \tau\theta_1), \\ P_3 &= \tau\theta_b e^{-\epsilon\beta/2} e^{-\lambda} P_1 / (1 - \tau\theta_1), \\ P_4 &= \tau\theta_b e^{\epsilon\beta/2} e^{\lambda} P_6 / (1 - \tau\theta_6), \\ P_5 &= \tau\theta_r e^{\lambda} P_6 / (1 - \tau\theta_6). \end{aligned} \quad (3.5)$$

For the natural path, $\mathcal{G}^{(\text{I})}$ of (3.3) is maximized without the subsidiary condition (3.2), or the $P_i^{(n)}$'s for the natural path are derived from (3.5) by putting $\lambda=0$. Neglecting higher order terms in τ , (3.4) thus leads to the expressions for θ_1 and θ_6 as³³

$$\begin{aligned} \theta_1 &= \theta_r + \theta_b e^{-\epsilon\beta/2}, \\ \theta_6 &= \theta_r + \theta_b e^{\epsilon\beta/2}. \end{aligned} \quad (3.6)$$

³² They may depend on $\{p_i(t)\}$, as is the case for the Fermi statistics. See the next section.

³³ Both θ_1 and θ_6 were incorrectly written as $\theta_r + \theta_b$ in (3.5) of ICP. This error did not affect the results of ICP.

Here and in the previous equations $\beta = 1/kT$.

Equation (3.5) gives the most probable path corresponding to the maximum of $\mathcal{G}^{(I)}$ for the given initial state at t and the final state at $t+\tau$. This maximum $\mathcal{G}^{(I)}$, which we call $\mathcal{G}^{(II)}$, can then be written from (3.3) and (3.5) as

$$L^{-1} \ln \mathcal{G}^{(II)}[p_u(t); p_u(t+\tau)] = \tau \theta_1 p_l(t) (e^{-\lambda} - 1) + \tau \theta_6 p_u(t) (e^{\lambda} - 1) + \lambda [p_u(t+\tau) - p_u(t)] + O(\tau^2), \quad (3.7a)$$

where the Lagrange multiplier λ is to be determined from (3.2) by use of (3.4) and (3.5) as

$$p_u(t+\tau) - p_u(t) = \tau \theta_1 p_l(t) e^{-\lambda} - \tau \theta_6 p_u(t) e^{\lambda} + O(\tau^2). \quad (3.7b)$$

Before proceeding to writing down the persistency from (3.7), it seems worthwhile to examine the meaning of these equations. The fact that $\lambda=0$ for the natural path of the system suggests that λ is a parameter which characterizes fluctuation of the system away from the natural path. When we disregard the case of large fluctuations which are always of minor importance, we may assume λ is small and may neglect higher order terms of λ . Thus, Eqs. (3.7a) and (3.7b) are expanded as

$$(\tau L)^{-1} \ln \mathcal{G}^{(II)}[p_u(t); p_u(t+\tau)] = \lambda [\dot{p}_u(t) - \theta_1 p_l(t) + \theta_6 p_u(t)] + \lambda^2 [\theta_1 p_l(t) + \theta_6 p_u(t)]/2 + O(\tau, \lambda^3), \quad (3.8a)$$

$$\dot{p}_u(t) = \theta_1 p_l(t) - \theta_6 p_u(t) - \lambda [\theta_1 p_l(t) + \theta_6 p_u(t)] + O(\tau, \lambda^2). \quad (3.8b)$$

When λ is eliminated from these two, we obtain

$$(\tau L)^{-1} \ln \mathcal{G}^{(II)}[p_u(t); p_u(t+\tau)] = -[\dot{p}_u(t) - \theta_1 p_l(t) + \theta_6 p_u(t)]^2/2[\theta_1 p_l(t) + \theta_6 p_u(t)] + O(\tau, \lambda^3). \quad (3.9)$$

It should be noticed that (3.9) is a generalization of the expression Onsager and Machlup worked out³⁴ for the case of small fluctuations from an equilibrium state. Namely, when we assume, for instance, that the system is not interacting with the radiation field and that $p_u(t)$ is close to the equilibrium value, (3.9) becomes identical with the expression which Onsager and Machlup derived. It is further seen from the meaning of (3.9) that the value of \dot{p}_u for the natural path is derived by maximizing (3.9) with respect to $\dot{p}_u(t)$ keeping $p_u(t)$ fixed. This procedure is a generalization of what Onsager called the "principle of least dissipation of energy."^{10,34}

Returning to the quantity of main concern in this section, the persistency $\mathcal{G}^{(III)}$, we see that it is a special case of (3.7) for which $p_u(t)$ and $p_u(t+\tau)$ are equal. When λ is eliminated, the persistency is then written as

$$(\tau L)^{-1} \ln \mathcal{G}^{(III)}[p_u; p_u] = -[(\theta_1 p_l)^{\frac{1}{2}} - (\theta_6 p_u)^{\frac{1}{2}}]^2 + O(\tau). \quad (3.10)$$

Or, if we disregard $O(\lambda^3)$, the persistency may be derived from (3.9) by putting $\dot{p}_u=0$ as

$$(\tau L)^{-1} \ln \mathcal{G}^{(III)}[p_u; p_u] = -(\theta_1 p_l - \theta_6 p_u)^2/2(\theta_1 p_l + \theta_6 p_u) + O(\tau, \lambda^3). \quad (3.11)$$

Here $O(\lambda^3)$ means quantities of order $\lambda^3 = [(\theta_1 p_l - \theta_6 p_u) \div (\theta_1 p_l + \theta_6 p_u)]^3$, which is derived from (3.8b) by putting $\dot{p}_u=0$.

The theorem of the previous section says that the expression (3.11) takes its maximum value with respect to $p_u (=1-p_l)$ in the steady state. From the quadratic expression we see that the maximum is achieved when

$$\theta_1 p_l = \theta_6 p_u. \quad (3.12)$$

In order to compare this relation with the result for the steady state obtained from the kinetic equation approach, we derive the kinetic equation for the natural path either by putting $\lambda=0$ in (3.7b) or by maximizing (3.9) with respect to \dot{p}_u , keeping p_u fixed as

$$\dot{p}_u^{(n)} = \theta_1 p_l^{(n)} - \theta_6 p_u^{(n)}. \quad (3.13)$$

Since the steady state is characterized by $\dot{p}_u^{(n)}=0$, we see that the steady state derived from the kinetic equation (3.13) is identical with (3.12) obtained by the maximum persistency procedure. This agreement demonstrates the correctness of the "maximum persistency" theorem proved in the previous section. It must be emphasized that (3.12) holds even when the temperature of the heat bath is low and the $\min(dS/dt)$ principle fails.^{7,8}

It is instructive to compare the result with the entropy production. It is made of two parts: The entropy production in the system and that in the heat bath. The former is derived from the definition of entropy in the equilibrium theory, and the latter is by definition the energy flowing into the heat bath divided by its temperature. The entropy production in the universe is the sum of these two parts and is given in Eq. (6) of Klein's paper,⁷ which is written in the present notation as

$$(kL)^{-1} dS/dt = (\theta_6 p_u - \theta_1 p_l) \ln(p_u/p_l) + \epsilon \beta \theta_b (p_u e^{\epsilon \beta/2} - p_l e^{-\epsilon \beta/2}). \quad (3.14)$$

Since dS/dt is defined for the natural development of the universe, (3.13) has been used in deriving (3.14).

When the temperature of the heat bath is high, (3.14) is expanded as

$$(kT)^{-1} dS/dt = [p_u - p_u^{(s)} + O(\beta^3)]^2 (\theta_1 + \theta_6)^3 / \theta_1 \theta_6 + (\epsilon \beta)^2 \theta_r \theta_b / 2(\theta_r + \theta_b) + O(\beta^3), \quad (3.15)$$

where $p_u^{(s)}$ is the value of p_u at the steady state derived from (3.12), i.e., $p_u^{(s)} = \theta_1 / (\theta_1 + \theta_6)$. The minimum of (3.15) with respect to p_u gives the right result $p_u = p_u^{(s)}$ only when β is small. In order to compare (3.15) with (3.11), we may write the latter as

$$(\tau L)^{-1} \ln \mathcal{G}^{(III)}[p_u; p_u] = -(p_u - p_u^{(s)})^2 \times (\theta_1 + \theta_6)^3 / 4\theta_1 \theta_6 + \dots \quad (3.16)$$

³⁴ L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953).

Equations (3.15) and (3.16) look close but differ by a factor of four and also in the higher order terms of β . The relation between the two will be discussed in a subsequent paper.³⁵

IV. BOLTZMANN TRANSPORT EQUATION

In this section we discuss a system of conduction electrons in a crystal which is in an external electric field (external magnetic field and temperature gradient are not considered here). The state of an electron specified by its wave vector \mathbf{k} changes by way of two mechanisms; the external electric field changes \mathbf{k} (quasi-) continuously, while the collisions with phonons change \mathbf{k} discontinuously. Since the electron spin is assumed not to change during the whole process, it is not indicated explicitly. The notations of Wilson's book²⁴ will be used whenever possible.

We consider a system consisting of N electrons and at the same time an assembly consisting of L systems. A state of the assembly is specified by the distribution function $f(\mathbf{k}, t)$ which is the probability for finding a system in which the state \mathbf{k} is occupied by an electron at time t . The probability for finding a system in which the state \mathbf{k} is empty is denoted by $\bar{f}(\mathbf{k}, t)$: or

$$\bar{f}(\mathbf{k}, t) = 1 - f(\mathbf{k}, t). \quad (4.1)$$

For the change of state, the path distribution function $P(\mathbf{k}, t; \mathbf{k}', t + \tau)$ is defined as the probability for finding such a system in which an electron existed in the state \mathbf{k} at t and has moved to \mathbf{k}' at $t + \tau$, where τ is a small time interval. In the terminology of Sec. II, f and P are the state parameter and the path parameter, respectively. We may omit t and $t + \tau$ in f and P when no confusion is expected.

Due to the external electric field \mathcal{E} , the electron of the state \mathbf{k} changes into $\mathbf{k} + \tau\boldsymbol{\kappa}$ during τ . This drift may be regarded as a mapping of the whole \mathbf{k} space. The discontinuous changes of \mathbf{k} due to collisions with phonons are superposed on this (quasi-) continuous mapping. The drift $\boldsymbol{\kappa}$ is the same for all \mathbf{k} and

$$\boldsymbol{\kappa} = -e\mathcal{E}/\hbar, \quad (4.2)$$

where $-e$ is the charge of an electron.

Taking into consideration the $\boldsymbol{\kappa}$ drift, we may write the compatibility relations between f and P as

$$\begin{aligned} f(\mathbf{k}, t) &= \sum_{\mathbf{k}'} P(\mathbf{k}, t; \mathbf{k}' + \tau\boldsymbol{\kappa}, t + \tau), \\ f(\mathbf{k}, t) &= \bar{P}(\mathbf{k}, t; \mathbf{k} + \tau\boldsymbol{\kappa}, t + \tau) \\ &\quad + \sum_{\mathbf{k}'} P(\mathbf{k}', t; \mathbf{k} + \tau\boldsymbol{\kappa}, t + \tau). \end{aligned} \quad (4.3)$$

Here $\bar{P}(\mathbf{k}, t; \mathbf{k} + \tau\boldsymbol{\kappa}, t + \tau)$ is the probability for finding a system in which the state \mathbf{k} at t and its drifted state $\mathbf{k} + \tau\boldsymbol{\kappa}$ at $t + \tau$ are both empty. The prime on \sum'

indicates the condition $\mathbf{k}' \neq \mathbf{k}$. In writing Eq. (4.3) the second and higher order scattering processes are neglected; in other words, when an electron is scattered out from a state \mathbf{k} , no other electron will be simultaneously scattered into the drifted position of the same state $\mathbf{k} + \tau\boldsymbol{\kappa}$.

The conditional path probability $\mathcal{G}^{(1)}\{P(\mathbf{k}, t; \mathbf{k}', t + \tau)\}$ is made of three factors. The first factor \mathcal{G}_1 is the number of different paths specified by the path parameters $\{P(\mathbf{k}, t; \mathbf{k}', t + \tau)\}$ given the initial state $\{f(\mathbf{k}, t)\}$ and is constructed as follows. We assume that τ is small enough so that in one system of N electrons, at most one collision occurs during τ . Then the L systems in the assembly are classified by the pattern of a collision occurred in the system; $LP(\mathbf{k}; \mathbf{k}' + \tau\boldsymbol{\kappa})$ is the number of systems in which the collision $\mathbf{k} \rightarrow \mathbf{k}' + \tau\boldsymbol{\kappa}$ occurred, and $L(1 - \sum_{\mathbf{k}} \sum_{\mathbf{k}'} P(\mathbf{k}; \mathbf{k}' + \tau\boldsymbol{\kappa}))$ is the number of systems in which no collision occurred during τ . The number of ways of sorting the systems based on the collision patterns is then

$$\begin{aligned} g^{(1)} &= L! [L(1 - \sum_{\mathbf{k}} \sum_{\mathbf{k}'} P(\mathbf{k}; \mathbf{k}' + \tau\boldsymbol{\kappa}))]^{-1} \\ &\quad \times \prod_{\mathbf{k}} \prod_{\mathbf{k}'} [LP(\mathbf{k}; \mathbf{k}' + \tau\boldsymbol{\kappa})]^{-1}. \end{aligned} \quad (4.4a)$$

For each sorting of this kind, we can distribute holes and electrons on each \mathbf{k} state among the systems whose \mathbf{k} state did not participate in the collisions; this number of ways of distribution is

$$\begin{aligned} g_{\mathbf{k}}^{(2)} &= \{L[1 - \sum_{\mathbf{k}'} (P(\mathbf{k}; \mathbf{k}' + \tau\boldsymbol{\kappa}) + P(\mathbf{k}'; \mathbf{k} + \tau\boldsymbol{\kappa}))]\}^{-1} \\ &\quad \times [LP(\mathbf{k}; \mathbf{k} + \tau\boldsymbol{\kappa})]^{-1} [L\bar{P}(\mathbf{k}; \mathbf{k} + \tau\boldsymbol{\kappa})]^{-1}. \end{aligned} \quad (4.4b)$$

The first factor \mathcal{G}_1 of the path probability is then obtained as the product of $g^{(1)}$ and all the $g_{\mathbf{k}}^{(2)}$'s for different \mathbf{k} 's divided by the number of ways of realizing the initial configurations of the assembly specified by $\{f(\mathbf{k}, t)\}$. The latter quantity is

$$g_i = L! \prod_{\mathbf{k}} \{[Lf(\mathbf{k}, t)]^{-1} [L\bar{f}(\mathbf{k}, t)]^{-1}\}. \quad (4.4c)$$

Thus \mathcal{G}_1 is, in the logarithmic form,³⁶

$$\begin{aligned} L^{-1} \ln \mathcal{G}_1\{P\} &= \sum_{\mathbf{k}} [\mathcal{L}f(\mathbf{k}, t) + \mathcal{L}\bar{f}(\mathbf{k}, t) + \mathcal{L}(1 - \sum_{\mathbf{k}'} [P(\mathbf{k}; \mathbf{k}' + \tau\boldsymbol{\kappa}) \\ &\quad + P(\mathbf{k}'; \mathbf{k} + \tau\boldsymbol{\kappa})]) - \sum_{\mathbf{k}} \sum_{\mathbf{k}'} \mathcal{L}P(\mathbf{k}; \mathbf{k}' + \tau\boldsymbol{\kappa}) \\ &\quad - \mathcal{L}(1 - \sum_{\mathbf{k}} \sum_{\mathbf{k}'} P(\mathbf{k}; \mathbf{k}' + \tau\boldsymbol{\kappa})) - \sum_{\mathbf{k}} [\mathcal{L}P(\mathbf{k}; \mathbf{k} + \tau\boldsymbol{\kappa}) \\ &\quad + \mathcal{L}\bar{P}(\mathbf{k}; \mathbf{k} + \tau\boldsymbol{\kappa})], \end{aligned} \quad (4.5)$$

where, in order to avoid repetition of long formulas, an

³⁵ R. Kikuchi and P. Gottlieb, following paper [Phys. Rev. 124, 1691 (1961)].

³⁶ The derivation of (4.5) in this section looks different from the corresponding method in Appendixes II and III of ICP, where an arbitrary ordering of all the \mathbf{k} states was introduced in counting \mathcal{G}_1 . The two methods, however, give the same results when τ is small.

operator \mathfrak{L} was introduced as

$$\mathfrak{L}x = x \ln x. \quad (4.6)$$

The expression $L^{-1}k \ln g_1(t; t+\tau)$, proportional to (4.5), was called the path entropy of a system for the time interval τ in ICP, in analogy with the entropy of the equilibrium theory, which is k times the logarithm of the number of different configurations having the same energy.

The second factor in $\mathcal{G}^{(1)}$ comes from the *a priori* probability of collisions, which is

$$L^{-1} \ln \mathcal{G}_2\{P\} = \sum_{\mathbf{k}} P(\mathbf{k}; \mathbf{k} + \tau\kappa) \ln(1 - \tau W(\mathbf{k})) + \sum_{\mathbf{k}} \sum' P(\mathbf{k}; \mathbf{k}' + \tau\kappa) \ln(\tau W(\mathbf{k}, \mathbf{k}')). \quad (4.7)$$

Here $W(\mathbf{k}, \mathbf{k}')$ for $\mathbf{k} \neq \mathbf{k}'$ is the quantum-mechanical transition probability per unit time from \mathbf{k} to \mathbf{k}' due to a collision, and $W(\mathbf{k})$ is the total probability that an electron leaves the state \mathbf{k} in unit time due to any collision, and is to be determined analogous to the method described in the previous section from the relation

$$\tau W(\mathbf{k}) = \sum' P^{(n)}(\mathbf{k}; \mathbf{k}' + \tau\kappa) / f(\mathbf{k}, t), \quad (4.8)$$

where $P^{(n)}(\mathbf{k}, \mathbf{k}' + \tau\kappa)$ is the path probability for the natural path.

The third factor in $\mathcal{G}^{(1)}$ comes from the path probability in the heat bath and was proved in ICP to be the energy flow into the heat bath during τ divided by $2kT$, so that writing the energy of the \mathbf{k} state as $E_{\mathbf{k}}$,

$$L^{-1} \ln \mathcal{G}_3\{P\} = \sum_{\mathbf{k}} \sum' P(\mathbf{k}; \mathbf{k}' + \tau\kappa) (E_{\mathbf{k}} - E_{\mathbf{k}'}) \beta / 2. \quad (4.9)$$

The conditional path probability $\mathcal{G}^{(1)}$ is obtained as a product of three \mathcal{G}_i 's, or $L^{-1} \ln \mathcal{G}^{(1)}(t; t+\tau)$ is written as the sum of (4.5), (4.7), and (4.9). After this is obtained, the procedure of applying the theorem of Sec. II is the same as demonstrated in the previous section. First $\mathcal{G}^{(1)}\{P(\mathbf{k}; \mathbf{k}' + \tau\kappa)\}$ is maximized with respect to $\{P(\mathbf{k}; \mathbf{k}' + \tau\kappa)\}$, keeping $\{f(\mathbf{k}, t)\}$ and $\{f(\mathbf{k}, t+\tau)\}$ fixed, or under the condition

$$f(\mathbf{k} + \tau\kappa, t+\tau) - f(\mathbf{k}, t) - \sum' [P(\mathbf{k}'; \mathbf{k} + \tau\kappa) - P(\mathbf{k}; \mathbf{k}' + \tau\kappa)] = 0. \quad (4.10)$$

Using the Lagrange multiplier $\lambda(\mathbf{k})$ for (4.10), we derive the value of P for the maximum of $\mathcal{G}^{(1)}\{P(\mathbf{k}; \mathbf{k}' + \tau\kappa)\}$ as

$$P(\mathbf{k}; \mathbf{k}' + \tau\kappa) = \tau U(\mathbf{k}, \mathbf{k}') \times \exp[\lambda(\mathbf{k}) - \lambda(\mathbf{k}')] + O(\tau^2), \quad (4.11)$$

where we used the abbreviation

$$U(\mathbf{k}, \mathbf{k}') = W(\mathbf{k}, \mathbf{k}') f(\mathbf{k}, t) \bar{f}(\mathbf{k}', t) \times \exp[\beta(E_{\mathbf{k}} - E_{\mathbf{k}'}) / 2]. \quad (4.12)$$

Since the natural path $P^{(n)}(\mathbf{k}; \mathbf{k}' + \tau\kappa)$ is a special case of (4.11) with $\lambda(\mathbf{k}) = 0$, the relation (4.8) leads to the form of $W(\mathbf{k})$ as

$$W(\mathbf{k}) = \sum_{\mathbf{k}'} W(\mathbf{k}, \mathbf{k}') \bar{f}(\mathbf{k}', t) \exp[\beta(E_{\mathbf{k}} - E_{\mathbf{k}'}) / 2]. \quad (4.13)$$

It may be noticed that this $W(\mathbf{k})$ depends on t through $\bar{f}(\mathbf{k}', t)$.³² When (4.11) and (4.13) are used in $\mathcal{G}^{(1)}\{P(\mathbf{k}; \mathbf{k}' + \tau\kappa)\}$, the conditional path probability of going from $\{f(\mathbf{k}, t)\}$ to $\{f(\mathbf{k}, t+\tau)\}$ is written as

$$L^{-1} \ln \mathcal{G}^{(1)}[\{f(\mathbf{k}, t)\}; \{f(\mathbf{k}, t+\tau)\}] = \tau \sum_{\mathbf{k}} \sum' U(\mathbf{k}, \mathbf{k}') [\exp(\lambda(\mathbf{k}) - \lambda(\mathbf{k}')) - 1] + \sum_{\mathbf{k}} \lambda(\mathbf{k}) [f(\mathbf{k} + \tau\kappa, t+\tau) - f(\mathbf{k}, t)] + O(\tau^2), \quad (4.14)$$

where $\lambda(\mathbf{k})$ is to be derived from (4.10) as

$$f(\mathbf{k} + \tau\kappa, t+\tau) - f(\mathbf{k}, t) = \tau \sum_{\mathbf{k}'} [U(\mathbf{k}', \mathbf{k}) \exp(\lambda(\mathbf{k}') - \lambda(\mathbf{k})) - U(\mathbf{k}, \mathbf{k}') \exp(\lambda(\mathbf{k}) - \lambda(\mathbf{k}'))]. \quad (4.15)$$

It is seen that (4.14) is analogous to (3.7).

The persistency is a special case of (4.14) when

$$f(\mathbf{k}, t) = f(\mathbf{k}, t+\tau). \quad (4.16)$$

In this case we may drop t and $t+\tau$ and write $\mathcal{G}^{(1)}$ in (4.14) as $\mathcal{G}^{(11)}[\{f(\mathbf{k})\}; \{f'(\mathbf{k})\}]$, whose behavior will be examined near the steady state. Without loss of accuracy we may regard $\lambda(\mathbf{k})$ as small, since it vanishes in the steady state, so that we expand the exponentials in (4.14) and (4.15). From (4.15),

$$f(\mathbf{k} + \tau\kappa) - f(\mathbf{k}) = \tau \sum_{\mathbf{k}'} \{U(\mathbf{k}', \mathbf{k}) [1 + \lambda(\mathbf{k}') - \lambda(\mathbf{k})] - U(\mathbf{k}, \mathbf{k}') [1 + \lambda(\mathbf{k}) - \lambda(\mathbf{k}')] \}, \quad (4.17)$$

and using this in (4.14) and retaining terms up to the second order, we obtain

$$(\tau L)^{-1} \ln \mathcal{G}^{(11)}[\{f(\mathbf{k})\}; \{f(\mathbf{k})\}] = -\frac{1}{4} \sum_{\mathbf{k}} \sum' [U(\mathbf{k}, \mathbf{k}') + U(\mathbf{k}', \mathbf{k})] \times [\lambda(\mathbf{k}) - \lambda(\mathbf{k}')]^2 + \dots \quad (4.18)$$

At this point it is clear that the "maximum persistency" theorem holds because the steady state value of $\lambda(\mathbf{k})$, which is zero, does make (4.18) a maximum. However, this interpretation of the theorem in Sec. II does not give us any interesting information. In order to derive Kohler's form of the variation function, we eliminate $\lambda(\mathbf{k})$ from (4.18) and (4.17). This is done as follows.

We treat only the case of small external field as is usually done.²⁴ Then κ of (4.2) is small and we define $\Phi(\mathbf{k})$ as

$$f(\mathbf{k}) = f_0(\mathbf{k}) + \beta \Phi(\mathbf{k}) f_0(\mathbf{k}) \bar{f}_0(\mathbf{k}), \quad (4.19)$$

where $f_0(\mathbf{k})$ is the equilibrium Fermi distribution which depends on \mathbf{k} through the energy $E_{\mathbf{k}}$, and $\Phi(\mathbf{k})$ is a small quantity proportional to $|E|$. Neglecting $O(E^2)$, we can expand (4.12) as

$$U(\mathbf{k}, \mathbf{k}') = V(\mathbf{k}, \mathbf{k}') \{1 + \beta[\Phi(\mathbf{k})\bar{f}_0(\mathbf{k}) - \Phi(\mathbf{k}')f_0(\mathbf{k}')]\}, \quad (4.20)$$

where

$$V(\mathbf{k}, \mathbf{k}') = V(\mathbf{k}', \mathbf{k}) = W(\mathbf{k}, \mathbf{k}') f_0(\mathbf{k}) \bar{f}_0(\mathbf{k}') \times \exp[\beta(E_{\mathbf{k}} - E_{\mathbf{k}'})/2]. \quad (4.21)$$

Use of (4.19) and (4.20) in (4.17) leads to

$$\kappa \cdot \text{grad}_{\mathbf{k}} f_0(\mathbf{k}) = \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') [\beta \Phi(\mathbf{k}') + 2\lambda(\mathbf{k}') - \beta \Phi(\mathbf{k}) - 2\lambda(\mathbf{k})]. \quad (4.22)$$

If we write the value of $\Phi(\mathbf{k})$ in the steady state as $\Phi_s(\mathbf{k})$, it satisfies the Boltzmann equation:

$$\kappa \cdot \text{grad}_{\mathbf{k}} f_0(\mathbf{k}) = \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') [\beta \Phi_s(\mathbf{k}') - \beta \Phi_s(\mathbf{k})], \quad (4.23)$$

since $\lambda(\mathbf{k})$ vanishes in the steady state. Subtract (4.23) from (4.22) to obtain

$$0 = \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') \{ \beta [\Phi(\mathbf{k}') - \Phi_s(\mathbf{k}')] + 2\lambda(\mathbf{k}') - \beta [\Phi(\mathbf{k}) - \Phi_s(\mathbf{k})] - 2\lambda(\mathbf{k}) \}. \quad (4.24)$$

This shows that $\beta[\Phi(\mathbf{k}) - \Phi_s(\mathbf{k})] + 2\lambda(\mathbf{k})$ is the value of $\Phi_s(\mathbf{k})$ when there is no field and $\kappa=0$. Therefore we conclude

$$2\lambda(\mathbf{k}) = \beta[\Phi_s(\mathbf{k}) - \Phi(\mathbf{k})], \quad (4.25)$$

because $\Phi_s(\mathbf{k})$ vanishes when $\kappa=0$.

When we substitute (4.25) in (4.18) and neglect higher order terms, we can write

$$\begin{aligned} (\tau L)^{-1} \ln \mathcal{G}^{(\text{III})}[\{f(\mathbf{k})\}; \{f(\mathbf{k})\}] \\ = -(\beta^2/8) \sum_{\mathbf{k}} \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') \{ [\Phi_s(\mathbf{k}) - \Phi_s(\mathbf{k}')]^2 \\ - 2[\Phi_s(\mathbf{k}) - \Phi_s(\mathbf{k}')][\Phi(\mathbf{k}) - \Phi(\mathbf{k}')] \\ + [\Phi(\mathbf{k}) - \Phi(\mathbf{k}')]^2 \}. \end{aligned} \quad (4.26)$$

At this point it may be noticed that because of the symmetry of $V(\mathbf{k}, \mathbf{k}')$ indicated in (4.21), when (4.23) is multiplied with $\Phi(\mathbf{k})$ and summed over \mathbf{k} it leads to

$$\begin{aligned} 2 \sum_{\mathbf{k}} \Phi(\mathbf{k}) \kappa \cdot \text{grad}_{\mathbf{k}} f_0(\mathbf{k}) = \beta \sum_{\mathbf{k}} \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') \\ \times [\Phi_s(\mathbf{k}') - \Phi_s(\mathbf{k})][\Phi(\mathbf{k}) - \Phi(\mathbf{k}')]. \end{aligned} \quad (4.27)$$

When (4.27) is used for the middle term of (4.26), we

arrive at the final form of the variation function:

$$\begin{aligned} (8/\tau L) \ln \mathcal{G}^{(\text{III})}[\{f(\mathbf{k})\}; \{f(\mathbf{k})\}] \\ + \beta^2 \sum_{\mathbf{k}} \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') [\Phi_s(\mathbf{k}) - \Phi_s(\mathbf{k}')]^2 \\ = -4\beta \sum_{\mathbf{k}} \Phi(\mathbf{k}) \kappa \cdot \text{grad}_{\mathbf{k}} f_0(\mathbf{k}) - \beta^2 \sum_{\mathbf{k}} \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') \\ \times [\Phi(\mathbf{k}) - \Phi(\mathbf{k}')]^2. \end{aligned} \quad (4.28)$$

On the left-hand side, $\Phi_s(\mathbf{k})$ is a still unknown function but is a constant in the same sense that $p_u^{(s)}$ is a constant in (3.13). Therefore we can apply the theorem of Sec. II and conclude that the right-hand side of this expression is the variation function we have been after which takes a maximum value in the steady state. It is easy to verify that the Boltzmann equation (4.23) is derived by varying (4.28) with respect to $\Phi(\mathbf{k})$.

Equation (4.28) is equivalent to the variation principle stated in Wilson's book²⁴; namely, when $\Phi(\mathbf{k})$ satisfies

$$\begin{aligned} -2 \sum_{\mathbf{k}} \Phi(\mathbf{k}) \kappa \cdot \text{grad}_{\mathbf{k}} f_0(\mathbf{k}) \\ = \beta \sum_{\mathbf{k}} \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') [\Phi(\mathbf{k}) - \Phi(\mathbf{k}')]^2, \end{aligned} \quad (4.29)$$

the right-hand side of (4.28) is modified so that the expression to be maximized is

$$\beta^2 \sum_{\mathbf{k}} \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') [\Phi(\mathbf{k}) - \Phi(\mathbf{k}')]^2. \quad (4.30)$$

This is the variation function used by Kohler,²⁰ who derived it by reasoning backward from the Boltzmann equation purely on mathematical basis.

The physical meaning of the variation function is now clear. As is seen on the left-hand side of (4.28), except for an additional constant it is the persistency $\mathcal{G}^{(\text{III})}[\{f(\mathbf{k})\}; \{f(\mathbf{k})\}]$, which was defined in the theorem of Sec. II and was proved (starting from first principles) to take a maximum value in the steady state. Thus Kohler's variation procedure is now put on the sound physical basis.

It is instructive to compare (4.28) with the entropy production. The latter quantity is made up of two parts, the one in the system, dS_s/dt , and the other in the heat bath, dS_b/dt . The first part is calculated using the entropy expression known in the equilibrium statistical mechanics³⁷ as

$$\frac{dS_s}{dt} = -k \sum_{\mathbf{k}} \frac{d}{dt} [f(\mathbf{k}, t) \ln f(\mathbf{k}, t) + \bar{f}(\mathbf{k}, t) \ln \bar{f}(\mathbf{k}, t)]. \quad (4.31)$$

This is to be supplemented by the expression of $df(\mathbf{k}, t)/dt$, which is the sum of two terms due to the drift and the collisions, respectively, and can be

³⁷ See for example D. ter Haar, *Elements of Statistical Mechanics* (Rinehart Publishing Company, New York, 1954), p. 76.

written as

$$df(\mathbf{k}, t)/dt = -\kappa \cdot \text{grad}_{\mathbf{k}} f_0(\mathbf{k}) + \beta \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') [\Phi(\mathbf{k}', t) - \Phi(\mathbf{k}, t)]. \quad (4.32)$$

The entropy production in the heat bath, dS_b/dt , is the energy flowing into the heat bath, i.e., the system of phonons in the present case, divided by the temperature of the bath. Since the rate of this energy flow is expressed by the second term of (4.32), we obtain

$$dS_b/dt = -T^{-1} \sum_{\mathbf{k}} E_{\mathbf{k}} \beta \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') \times [\Phi(\mathbf{k}', t) - \Phi(\mathbf{k}, t)]. \quad (4.33)$$

Combining the last three equations and after mathematical transformations, we obtain for the rate of entropy production of the universe

$$d(S_s + S_b)/dt = (\beta^2/2k) \sum_{\mathbf{k}} \sum_{\mathbf{k}'} V(\mathbf{k}, \mathbf{k}') \times [\Phi(\mathbf{k}, t) - \Phi(\mathbf{k}', t)]^2. \quad (4.34)$$

This is the entropy production whose minimum with respect to $\Phi(\mathbf{k})$ would correspond to the Boltzmann equation if the "min(dS/dt)" principle should hold. Comparison of the right-hand side of (4.28) with (4.34) indicates that the latter is a part of the former (with the opposite sign) and thus the minimum of (4.34) does not give the steady state. This shows the failure of the min(dS/dt) principle for the present case.³⁸

V. DISCUSSION

It is instructive to notice the similarity between the variational theorem proposed in this paper and the principle of minimum free energy in the equilibrium statistical mechanics for one-dimensional cooperative phenomena. It will be shown in this discussion that if we regard time t as the coordinate along the one-dimensional axis and the time interval τ as the lattice constant, the two variational procedures become identical.

When the cluster-variation treatment²⁷ is used for the one-dimensional cooperative phenomena of nearest-neighbor interaction, the free energy of the system is first written in terms of two kinds of probability parameters, $\{p_i(R)\}$ and $\{P_{ij}(R, R+a)\}$; $p_i(R)$ is the probability for finding the state i at the lattice point R , and $P_{ij}(R; R+a)$ for the states i and j at R and $R+a$, a being the lattice constant. $P_{ij}(R; R+a)$ and $p_i(R)$ satisfy the compatibility relations similar to (2.1). In writing the free energy $F\{P_{ij}\}$ for a system of N lattice points, we may introduce the concept of an

assembly composed of L systems, and write the equations:

$$\exp(-L\beta F\{P_{ij}\}) = (\mathcal{G}_1 \mathcal{G}_2)^N, \quad (5.1)$$

where

$$\mathcal{G}_1 = (Lp_i)! \prod_{i,j} (LP_{ij})!^{-1}, \quad (5.2)$$

and

$$\mathcal{G}_2 = \exp(-L\beta \sum_{i,j} \epsilon_{ij} P_{ij}). \quad (5.3)$$

Here ϵ_{ij} is the energy for the $i-j$ bond. For (5.1) it was assumed that all the lattice points are equivalent and

$$p_i(R) = p_i(R+a). \quad (5.4)$$

\mathcal{G}_1 is the number of different configurations having the same set of parameters $\{P_{ij}\}$, and \mathcal{G}_2 can be regarded as the *a priori* probability that a change of state described by $\{P_{ij}\}$ appears. Equation (5.1) corresponds to $\mathcal{G}^{(1)}$ of this paper with a restriction of stationarity, e.g., (4.16). The *a priori* transition probabilities, θ in Sec. III and $W(\mathbf{k}, \mathbf{k}')$ in Sec. IV, correspond to $\exp(-\beta \epsilon_{ij})$ of (5.3), except for a constant factor.

In equilibrium statistical mechanics, the equilibrium state is derived by minimizing the free energy $F\{P_{ij}\}$ or by maximizing the whole quantity in (5.1) with respect to $\{P_{ij}\}$ under the "stationarity" restriction (5.4). This procedure exactly corresponds to the variational theorem of maximum persistency.

The key to the success of the present treatment lies in the introduction of the *path parameters* (denoted by $\{P_{ijr}(t; t+\tau)\}$ in Sec. II). It should be noticed that all the previous attempts (listed in Sec. I) which aimed at the same goal as the present work started with the *state parameters* (written as $\{p_i(t)\}$ in Sec. II). The inadequacy of using only $\{p_i(t)\}$ but not $\{P_{ijr}(t; t+\tau)\}$ can be well understood when we remember, as was pointed out above, that these two kinds of probability variables correspond to the probability for the point $\{p_i(R)\}$ and that for the pair $\{P_{ij}(R, R+a)\}$, respectively, in the one-dimensional cooperative processes of equilibrium statistical mechanics. It is known in the theory of order-disorder processes, that as far as only the point parameters $\{p_i(R)\}$ are used, the theory cannot go beyond the Bragg-Williams approximation; whereas when the pair probabilities $\{P_{ij}(R, R+a)\}$ are introduced, the one-dimensional cooperative phenomena of nearest-neighbor interaction can be solved *rigorously*, as was shown in Sec. IIA of reference 27. Since the Markov process in the time sequence problem corresponds to the nearest-neighbor interaction in the spatial problem, it is understandable that the Markov process can be solved rigorously when the path parameters $\{P_{ijr}(t; t+\tau)\}$ are used.

As the last comment, it should be emphasized here that the path probability method for Markoffian processes²³ used in this paper starts from a microscopic description and first principles and is constructed on

³⁸ Equations (4.33) and (4.34) are essentially the basis of Ziman's interpretation (reference 21) of Kohler's variational procedure. Ziman's proposal, however, does not seem to provide any more meaning than an interpretation of individual terms in the expression of entropy production.

the basis as sound as that for the equilibrium statistical mechanics.³⁹

APPENDIX I. DERIVATION OF $\mathcal{G}^{(1)}$ (MATHEMATICAL PART)

The path probability $\mathcal{G}^{(1)}$ written in terms of the path parameters is the starting expression on which all the subsequent derivations in the present work are based. Therefore, although the expression was proved in ICP, it seems worthwhile to supplement the explanation in more detail how $\mathcal{G}^{(1)}$ is constructed. As an example, (3.3) will be discussed here.

This is a mathematical expression for the following problem of probability theory. Three different kinds of independent events happen with the *a priori* probabilities $1 - \tau(\theta_r + \theta_b e^{-\epsilon\beta/2})$, $\tau\theta_r$, and $\tau\theta_b e^{-\epsilon\beta/2}$. The probability that out of LP_1 events, these three kinds happen LP_1 times, LP_2 times, and LP_3 times, respectively, [see the compatibility relations in (3.1)] is given by the multinomial expression as

$$\mathcal{P}_1 = [1 - \tau(\theta_r + \theta_b e^{-\epsilon\beta/2})]^{LP_1} (\tau\theta_r)^{LP_2} (\tau\theta_b e^{-\epsilon\beta/2})^{LP_3} \times (LP_1)! [(LP_1)! (LP_2)! (LP_3)!]^{-1}. \quad (\text{AI.1})$$

Similarly, when there are another three kinds of independent events which happen with the *a priori* probabilities $\tau\theta_b e^{\epsilon\beta/2}$, $\tau\theta_r$, and $1 - \tau(\theta_r + \theta_b e^{\epsilon\beta/2})$, the probability that out of LP_u events they happen LP_4 times, LP_5 times, and LP_6 times, respectively, is given as

$$\mathcal{P}_2 = (\tau\theta_b e^{\epsilon\beta/2})^{LP_4} (\tau\theta_r)^{LP_5} [1 - \tau(\theta_r + \theta_b e^{\epsilon\beta/2})]^{LP_6} \times (LP_u)! [(LP_4)! (LP_5)! (LP_6)!]^{-1}. \quad (\text{AI.2})$$

The product of these two probability expressions \mathcal{P}_1 and \mathcal{P}_2 is (3.3). The correspondence between this mathematical problem and the problem for atoms in Sec. III is obvious and does not need further elaboration.

The physics part of the derivation of (3.3) lies in how to identify the *a priori* probabilities $\tau\theta_r$, $\tau\theta_b e^{-\epsilon\beta/2}$, etc. θ_r and θ_b are the quantum-mechanical transition probabilities of the respective processes as were defined in Sec. III. For the interaction with the heat bath, however, θ_b itself is not sufficient; when the energy is to be supplied from the bath (the process for P_3), θ_b is to be decreased by the factor $e^{-\epsilon\beta/2}$; and when the energy is given to the bath (the process for P_4), θ_b is to be increased by the factor $e^{\epsilon\beta/2}$. These two Boltzmann-like factors are understandable because the ratio of populations in the upper and the lower levels in equilibrium is $e^{-\epsilon\beta} = e^{-\epsilon\beta/2} / e^{\epsilon\beta/2}$. The proof of these factors was given in Sec. IV and Appendix I of ICP in

³⁹ Another advantage of using the path parameters and the path probability is that problems for which the kinetic equations are difficult to write down based on intuitive physical arguments can be treated by deriving approximate kinetic equations (for the natural path) starting from the path probability. All the problems of irreversible cooperative phenomena discussed in the papers listed in reference 23 are of this nature.

detail. An alternative, and probably more straightforward, derivation is shown in the following Appendix.

APPENDIX II. DERIVATION OF $\mathcal{G}^{(1)}$ (THE BOLTZMANN-LIKE FACTOR)⁴⁰

A heat bath is characterized as being very large compared with the system of interest, so that the former is practically in thermal equilibrium even when energy exchange occurs between the bath and the system. We may use as a heat bath any large system in thermal equilibrium. To be specific, suppose that the heat bath is made of M particles, each of which has energy states labeled with $i=1, 2, \dots$. Let $p_i(t)$ be the state probability, $P_{ij}(t; t+\tau)$ be the path probability, and Θ_{ij} ($i \neq j$) be the quantum mechanical transition probability between the i th and j th states. Then the path probability \mathcal{G}_b for the heat bath is derived using the multinomial expression described in the previous Appendix as (in the logarithmic form),

$$\ln \mathcal{G}_b = M \left[\sum_i \sum_{j \neq i} P_{ij} \ln(\tau \Theta_{ij}) + \sum_i P_{ii} \ln(1 - \tau \Theta_{ii}) + \sum_i p_i(t) \ln p_i(t) - \sum_i \sum_j P_{ij} \ln P_{ij} \right]. \quad (\text{AII.1})$$

Here $1 - \tau \Theta_{ii}$ is the probability that the i th state does not change during τ . The logarithm of the path probability for the system of Sec. III is written from (AI.1) and (AI.2) as

$$\ln \mathcal{G}_s = L \left[(P_2 + P_5) \ln(\tau\theta_r) + (P_3 + P_4) \ln(\tau\theta_b) + P_1 \ln(1 - \tau\theta_1) + P_6 \ln(1 - \tau\theta_6) + p_l \ln p_l + p_u \ln p_u - \sum_i P_i \ln P_i \right]. \quad (\text{AII.2})$$

In this expression, the effect of the heat bath is not included yet. The path probability for the whole universe is the product $\mathcal{G}_b \mathcal{G}_s$.

The interaction between the system and the heat bath is such that the energy flow from the system into the heat bath is equal to the energy increase in the heat bath. Therefore the path parameters $\{P_{ij}\}$ for the heat bath and $\{P_i\}$ for the system are not independent of each other, but are related by

$$\sum_i \sum_j M(E_j - E_i)(P_{ji} - P_{ij}) + L\epsilon(P_4 - P_3) = 0. \quad (\text{AII.3})$$

Here E_i is the energy of the i th level of a particle in the heat bath. In deriving the natural path in the heat bath we multiply (AII.3) with a Lagrange multi-

⁴⁰ In Sec. IV and Appendix I of ICP, the path entropy of the heat bath, written as $\mathcal{S}^{(b)}$, $\mathcal{S}_s^{(b)}$, \mathcal{S}_e , and \mathcal{S} interchangeably, should be added with the kinetic terms; in other words, they should be replaced by (k times the logarithm of) the path probability. Thus (AI.1) of ICP should read

$$\mathcal{S}(t; t+\tau) = kN \left[\sum_r \sum_s' P_{rs} \ln(\tau\theta_{rs}) + \sum_r P_{rr} \ln(1 - \tau \sum_s \theta_{rs}) - \sum_r \sum_s P_{rs} \ln P_{rs} \right],$$

where (2.5) and (2.6) of ICP should be used of P_{rs} . This modification does not lead to any change in the conclusions; (4.3) of ICP holds as before, but is not supported by a stronger condition than (4.2), namely by

$$\mathcal{S}_e = \mathcal{S}_s + O(\tau^2).$$

plier Λ , add the product to $\ln \mathcal{G}_b + \ln \mathcal{G}_s$ and then maximize the sum with respect to P_{ij} . Thus we obtain

$$P_{ij} = \tau \Theta_{ij} \dot{p}_i(t) e^{\Lambda(E_i - E_j)} + O(\tau^2). \quad (\text{AII.4})$$

The value of Λ is determined as follows. Using (AII.4), the change of $\dot{p}_i(t)$ in the heat bath is derived as

$$\begin{aligned} \dot{p}_i(t) &= \sum_j (P_{ji} - P_{ij}) / \tau \\ &= \sum_j [\dot{p}_j e^{\Lambda(E_j - E_i)} - \dot{p}_i e^{\Lambda(E_i - E_j)}] \Theta_{ij}. \end{aligned} \quad (\text{AII.5})$$

In the derivation, the quantum-mechanical reciprocity, $\Theta_{ij} = \Theta_{ji}$, was used. Since the definition of the heat bath requires that $\dot{p}_i(t)$ is arbitrarily small even when the system is interacting with it, (AII.5) vanishes. In the sum of (AII.5), Θ_{ij} is the kinetic quantity related to the transition among levels, whereas the equilibrium statistical mechanics tells us that the equilibrium

values of \dot{p}_i 's are independent of Θ_{ij} . Therefore the coefficients of Θ_{ij} 's vanish individually:

$$\dot{p}_j e^{\Lambda(E_j - E_i)} - \dot{p}_i e^{\Lambda(E_i - E_j)} = 0. \quad (\text{AII.6})$$

On the other hand, we know for the equilibrium distribution,

$$\dot{p}_j / \dot{p}_i = e^{-\beta(E_j - E_i)}, \quad (\text{AII.7})$$

where $\beta = 1/kT$. From (AII.6) and (AII.7) we can identify

$$\Lambda = \beta/2. \quad (\text{AII.8})$$

Thus, as far as the variations with respect to the quantities of the system, P_i for $i=2, 3, 4$ and 5 , are concerned, we may add only $L\epsilon\beta(P_4 - P_3)/2$ to (AII.2) and need not worry about the detail of the heat bath. This is the origin of the Boltzmann-like factors in (AI.1) and (AI.2).

Path Integral in Irreversible Statistical Dynamics

RYOICHI KIKUCHI AND PETER GOTTLIEB
Hughes Research Laboratories, Malibu, California

(Received December 23, 1960; revised manuscript received July 27, 1961)

Extending the path probability method for irreversible statistical dynamics proposed previously, the two-gate probability distribution for a stochastic Markoffian system is written in the form of a path integral. The most probable path is derived as a solution of the variational Euler-Lagrange equation. Introducing the concept of the anticausal path, a method to calculate the one-gate probability distribution function (for fluctuations away from the nonequilibrium steady state) is explained. The distribution function derived recently by Mathews, Shapiro, and Falkoff for a system of many levels is shown to follow. The concept of the pseudo-entropy \tilde{S} is introduced. \tilde{S} always increases in time as the system approaches its steady state. It is shown that the rate $d\tilde{S}/dt$, rather than dS/dt of Prigogine, is minimum in the steady state, although the latter can serve as an approximation when the steady state is near equilibrium.

I. INTRODUCTION

IN a previous publication by one of the authors¹ a theorem was proved which shows that a function (of state probability parameters² $\{p_i\}$) called the persistency, $\mathcal{G}^{(\text{II})}[\{p_i\}, \{p_i\}]$, takes its maximum value in the nonequilibrium steady state of a Markoffian system. This theorem has a wider range of validity than Prigogine's principle of minimum entropy production³ $[\min(dS/dt)]$, and the variation function $\mathcal{G}^{(\text{III})}$ is derived from the path probability of irreversible statistical dynamics proposed previously.⁴ In the present paper, the path probability is exploited further, and the relation between the proposed theorem and some other related principles will be discussed.

¹ R. Kikuchi, preceding paper [Phys. Rev. **124**, 1682 (1961)]. This paper will be called VS hereafter.

² The curly brackets indicate a set of numbers. $\{p_i\}$ may be written in the vector form as \mathbf{p} , p_i being regarded as the i th component of this vector. See Sec. II of VS.

³ I. Prigogine, *Etude Thermodynamique des Phénomènes Irréversibles* (Maison Desoer, Liège, 1947).

⁴ R. Kikuchi, Ann. Phys. **10**, 127 (1960).

When the fluctuation from equilibrium is small, the two-gate probability distribution has been written in the form of the path integral.⁵⁻⁷ Section II derives the path integral representation of the two-gate probability for the case when the irreversible steady state may not be close to the equilibrium state. In Sec. III the concept of the anticausal path is used to calculate the probability for a specific fluctuation from the steady-state distribution. The relation between the maximum persistency theorem of VS and Prigogine's principle of $\min(dS/dt)$ will be discussed in later sections with the help of the concept of pseudo-entropy.

II. PATH INTEGRAL REPRESENTATION

In deriving the maximum persistency theorem in the previous paper, first the path probability $\mathcal{G}^{(\text{I})}$ is written in terms of the path parameters¹ $\{P_{ij}(t; t+\tau)\}$. Then $\mathcal{G}^{(\text{I})}$ is maximized keeping the end states $\{p_i^{(1)}(t)\}$ and

⁵ L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953).

⁶ N. Hashitsume, Progr. Theoret. Phys. (Kyoto) **15**, 369 (1956).

⁷ D. Falkoff, Ann. Phys. **4**, 325 (1958).