PHYSICAL REVIEW VOLUME 139, NUMBER 4A 16 AUGUST 1965

Statistical Mechanics of Irreversible Processes and the Principle of Minimum Entropy Production*

J. D. CURRIN
Department of Physics, University of Colorado, Boulder, Colorado
(Received 15 March 1965)

The transport properties of a system of N weakly interacting subsystems are derived using as the basic statistical assumption a principle of minimum entropy production. This replaces the assumption of "local equilibrium" where each subsystem is assumed to be described by a canonical distribution at the initial time.

1. INTRODUCTION

THE purpose of the statistical mechanics is to provide a description of a dynamical system in terms of a less-than-maximal set of observations. Consider a dynamical system consisting of an assembly of N subsystems whose Hamiltonians are $\mathcal{H}_1\cdots\mathcal{H}_N$ and which are weakly coupled by means of the interactions $\lambda\mathcal{H}_{ij}$ $i<j=1\cdots N$.[1] Let $\mathfrak{F}_i^{(\alpha)},\alpha=1\cdots k$, be a set of extensive commuting operators of the ith subsystem. In the weak coupling limit one obtains a useful statistical-mechanical description of the system in terms of the quantities

$$\mathfrak{F}^{(\alpha)}=\sum_i \mathfrak{F}_i^{(\alpha)} \tag{1.1}$$

by maximizing an entropy expression subject only to the constraint that the ensemble averages $\langle\mathfrak{F}^{(\alpha)}\rangle$ take prescribed values. One is thus led to the familiar canonical density

$$\rho=\exp-\{\sum_\alpha \mu^{(\alpha)}\mathfrak{F}^{(\alpha)}\}/\mathrm{Tr}\exp-\{\sum_\alpha \mu^{(\alpha)}\mathfrak{F}^{(\alpha)}\} \tag{1.2}$$

which describes the system at equilibrium. It is well known that the description in terms of the quantities (1.1) cannot treat transport processes within the system. For this, one needs a more detailed description.

This may be accomplished by augmenting the description of the system with the quantities $\mathfrak{F}_i^{(\alpha)}$. Maximization of the entropy expression under the constraint that the $\langle\mathfrak{F}_i^{(\alpha)}\rangle$ as well as $\langle\mathfrak{F}^{(\alpha)}\rangle$ take prescribed values then leads to a "local equilibrium ensemble" density

$$\rho=\exp-\{\sum_\alpha \mu^{(\alpha)}\mathfrak{F}^{(\alpha)}+\sum_\alpha\sum_i \mu_i^{(\alpha)}\mathfrak{F}_i^{(\alpha)}-\Theta\} \tag{1.3}$$

which in conjunction with the Liouville equation of the system

$$d\rho(t)/dt=-i[\mathcal{H},\rho(t)] \tag{1.4}$$

provides a means of treating transport processes within the system. This is the theory of irreversible processes which has been developed by Kubo,[2] Mori,[3] and others.

It is possible to give an alternative procedure to that just described. Rather than prescribing the values of $\langle\mathfrak{F}_i^{(\alpha)}\rangle$ and $\langle\mathfrak{F}^{(\alpha)}\rangle$ one instead chooses $\langle d\mathfrak{F}_i^{(\alpha)}/dt\rangle$ and $\langle\mathfrak{F}^{(\alpha)}\rangle$, where

$$d\mathfrak{F}_i^{(\alpha)}/dt=i[\mathcal{H},\mathfrak{F}_i^{(\alpha)}]. \tag{1.5}$$

It is this procedure which will be developed here. In doing this we will need to use the master equation of the system, which we discuss in the next section.

2. THE MASTER EQUATION

In the representation in which the set of operators $\mathfrak{F}_i^{(\alpha)}$ are diagonal we may label a typical state of the system as

$$|f_1,f_2,\cdots,f_N\rangle\equiv|f\rangle$$

with

$$\langle f|\mathfrak{F}_i^{(\alpha)}|f'\rangle=\langle f_i|\mathfrak{F}_i^{(\alpha)}|f_i\rangle\delta_{f,f'}, \tag{2.1}$$

where f_i is a set of quantum numbers for the ith subsystem.

At time t the average value of $\mathfrak{F}_i^{(\alpha)}$ is given by the ensemble average,

$$\begin{aligned}\langle\mathfrak{F}_i^{(\alpha)}\rangle_t&=\mathrm{Tr}\rho(t)\mathfrak{F}_i^{(\alpha)}\\&=\sum_f\langle f|\rho(t)|f\rangle\langle f|\mathfrak{F}_i^{(\alpha)}|f\rangle\\&=\sum_f\mathrm{Tr}\{\rho(t)|f\rangle\langle f|\}\,\mathrm{Tr}\{\mathfrak{F}_i^{(\alpha)}|f\rangle\langle f|\}\\&\equiv\sum_f\mathrm{Tr}\{\mathfrak{F}_i^{(\alpha)}|f\rangle\langle f|\mathrm{Tr}\{\rho(t)|f\rangle\langle f|\}\}\\&\equiv\mathrm{Tr}\sigma(t)\mathfrak{F}_i^{(\alpha)},\end{aligned} \tag{2.2}$$

where

$$\sigma(t)=\sum_f|f\rangle\langle f|\mathrm{Tr}\rho(t)|f\rangle\langle f|\equiv P\{\rho(t)\}.$$

Here P is a projection on $\rho(t)$ obeying

$$P\{P\{\rho(t)\}\}=P\{\rho(t)\}. \tag{2.3}$$

We will assume that at $t=0$

$$\rho(0)=\rho_1\rho_2\cdots\rho_N \tag{2.4}$$

and

$$\rho_i=\sum_{f_i}|f_i\rangle\langle f_i|\langle f_i|\rho_i|f_i\rangle \tag{2.5}$$

---

* This work supported by a grant from the U. S. Air Force Office of Scientific Research.

[1] We assume that $\mathcal{H}_{ij}$ has no diagonal matrix elements.

[2] R. Kubo, *Lectures in Theoretical Physics* (Interscience Publishers, Inc., New York, 1959), Vol. I, p. 120.

[3] H. Mori, I. Oppenheim, and J. Ross, *Studies in Statistical Mechanics* (North-Holland Publishing Company, Amsterdam, 1962), Vol. 1, p. 271 ff.

which leads to

$$\sigma(0)=\rho(0). \tag{2.6}$$

Making use of the Liouville equation

$$d\rho(t)/dt=-i[\mathcal{H},\rho(t)], \tag{2.7}$$

one may derive a dynamical equation for $\sigma(t)$. This is[4]

$$d\sigma(t)/dt=-\int_0^t d\tau P\{[\mathcal{H},G(t-\tau)\{[\mathcal{H},\sigma(\tau)]\}]\}, \tag{2.8}$$

where the operator valued functional $G(t)$ is given by the series expansion

$$G(t)\{A\}=A-i(1-P)\{[\mathcal{H},A]\}t+(i^2/2!)(1-P)\{[\mathcal{H},(1-P)\{[\mathcal{H},A]\}]\}+\cdots. \tag{2.9}$$

For the average value of $\mathfrak{F}_i^{(\alpha)}$ we obtain

$$\langle d\mathfrak{F}_i^{(\alpha)}/dt\rangle=\int_0^t d\tau\ \mathrm{Tr}[\mathcal{H},\mathfrak{F}_i^{(\alpha)}]G(\tau)\{[\mathcal{H},\sigma(t-\tau)]\}. \tag{2.10}$$

We define the coarse-grained time derivative

$$\langle d\mathfrak{F}_i^{(\alpha)}/dt\rangle=\frac{1}{T}\int_0^T\langle d\mathfrak{F}/dt\rangle_\tau d\tau \tag{2.11}$$

and write

$$\begin{aligned}\langle d\mathfrak{F}_i^{(\alpha)}/dt\rangle&=T^{-1}\int_0^T dt\int_0^t d\tau\ \mathrm{Tr}[\mathcal{H},\mathfrak{F}_i^{(\alpha)}]G(\tau)\{[\mathcal{H},\sigma(t-\tau)]\}\\&=\int_0^T d\tau\ \mathrm{Tr}[\mathcal{H},\mathfrak{F}_i^{(\alpha)}]G(\tau)\{[\mathcal{H},T^{-1}\int_0^{T-\tau}dt\sigma(t)]\}\\&=\int_0^T d\tau\ \mathrm{Tr}[\mathcal{H},\mathfrak{F}_i^{(\alpha)}]G(\tau)\{[\mathcal{H},\bar{\sigma}(T-\tau)]\}(1-\tau/T),\end{aligned} \tag{2.12}$$

where the coarse-grained density $\bar{\sigma}$ is given by

$$\bar{\sigma}(t)=(1/t)\int_0^t\sigma(\tau)d\tau.$$

At this point we wish to consider what we shall call a "stationary" process, namely, the case where $\bar{\sigma}(t)$ approaches a limiting value as t becomes large. For a closed system such as we are describing here this is possible only in the limit that the interactions between subsystems become vanishingly small. We may then allow T in Eq. (2.12) to become large without exceeding the macroscopic relaxation times of the system. To be specific we write the Hamiltonian as

$$\mathcal{H}=\mathcal{H}_0+\lambda\mathcal{H}_I,$$

where $\mathcal{H}_I$ is the interaction between subsystems, and λ^2 determines the macroscopic relaxation time of the system. We then may take the limit

$$T\to\infty,\quad \lambda\to 0,\quad \lambda^2T=\text{const.}$$

Equation (2.12) then becomes, to the lowest order in λ,

$$\langle d\mathfrak{F}_i^{(\alpha)}/dt\rangle=\int_0^\infty d\tau\ \mathrm{Tr}[\lambda\mathcal{H}_I,\mathfrak{F}_i^{(\alpha)}]\times e^{-i\mathcal{H}_0\tau}[\lambda\mathcal{H}_I,\bar{\sigma}(T\to\infty)]e^{i\mathcal{H}_0\tau}. \tag{2.13}$$

In writing Eq. (2.13) we have also made the usual assumption that the integrand vanishes sufficiently rapidly for τ large. We may write $\bar{\sigma}$ in the exponential form

$$\bar{\sigma}(T\to\infty)=e^{-\mathcal{K}}, \tag{2.14}$$

where $\mathcal{K}$ is a positive Hermitian operator. Then, using the identity

$$[A,e^B]=\int_0^1 dz e^B e^{-zB}[A,B]e^{zB} \tag{2.15}$$

Eq. (2.13) becomes

$$\langle d\mathfrak{F}_i^{(\alpha)}/dt\rangle=\int_0^\infty d\tau\int_0^1 dz\ \mathrm{Tr}e^{-\mathcal{K}}e^{z\mathcal{K}}[\lambda\mathcal{H}_I,\mathcal{K}]\times e^{-z\mathcal{K}}e^{i\mathcal{H}_0\tau}[\lambda\mathcal{H}_I,\mathfrak{F}_i^{(\alpha)}]e^{-i\mathcal{H}_0\tau}. \tag{2.16}$$

It is readily shown that the integrand of Eq. (2.16) is an even function of the time. This enables us to extend the region of integration from $-\infty$ to ∞.

The *fluxes* are derived from the quantities $\mathfrak{F}_i^{(\alpha)}$ by means of the relations

$$d\mathfrak{F}_i^{(\alpha)}/dt=\sum_j\lambda[\mathcal{H}_{ij},\mathfrak{F}_i^{(\alpha)}]\equiv\sum_j\mathcal{J}_{ij}^{(\alpha)}. \tag{2.17}$$

This leads to the expression

$$\langle\mathcal{J}_{ij}^{(\alpha)}\rangle=\tfrac{1}{2}\int_{-\infty}^\infty d\tau\int_0^1 dz\ \mathrm{Tr}e^{-\mathcal{K}}e^{z\mathcal{K}}(d\mathcal{K}/dt)e^{-z\mathcal{K}}\mathcal{J}_{ij}^{(\alpha)}(\tau) \tag{2.18}$$

with

$$\langle\mathcal{J}_{ij}^{(\alpha)}\rangle=-\langle\mathcal{J}_{ji}^{(\alpha)}\rangle. \tag{2.19}$$

The property (2.19) is restricted to operators $\mathfrak{F}_i^{(\alpha)}$ which satisfy

$$\langle f|(\mathfrak{F}_i^{(\alpha)}+\mathfrak{F}_j^{(\alpha)})|f\rangle=\langle f'|(\mathfrak{F}_i^{(\alpha)}+\mathfrak{F}_j^{(\alpha)}|f'\rangle \tag{2.20}$$

for states f, f' of the same total energy. We assume this to be true.

[4] R. Zwanzig, *Lectures in Theoretical Physics* (Interscience Publishers, Inc., New York, 1961), Vol. III, p. 106.

3. DETERMINATION OF $\mathcal{K}$

It can be shown that (2.16) vanishes if $\mathcal{K}$ is a function of "constants of the motion," $\mathfrak{F}_i^{(\alpha)}$,

$$\mathfrak{F}^{(\alpha)}=\sum_i \mathfrak{F}_i^{(\alpha)}. \tag{3.1}$$

We shall assume that $\mathcal{K}$ differs but little from the canonical distribution $\mathcal{K}_0$

$$\mathcal{K}_0=\sum_\alpha \mu^{(\alpha)}\mathfrak{F}^{(\alpha)}, \tag{3.2}$$

and thus write $\mathcal{K}$ as

$$\mathcal{K}=\mathcal{K}_0+\mathfrak{R}, \tag{3.2$'$}$$

where $\mathfrak{R}$ expresses a small deviation from the canonical distribution and is responsible for the transport properties of the system. Writing Eq. (2.16) to lowest order in $\mathfrak{R}$, we obtain

$$\langle d\mathfrak{F}_i^{(\alpha)}/dt\rangle=\tfrac{1}{2}\int_{-\infty}^{\infty} d\tau(d\mathfrak{R}/dt,\mathfrak{F}_i^{(\alpha)}(\tau)), \tag{3.3}$$

where

$$(A,B(\tau))\equiv\int_0^1 dz\, \mathrm{Tr}e^{-\mathcal{K}_0}e^{z\mathcal{K}_0}A\,e^{-z\mathcal{K}_0}B(\tau)$$
$$=(B,A(-\tau))$$
$$dA/dt=i[\lambda\mathcal{K}_I,A]$$
$$A(\tau)=e^{i\mathcal{K}_0\tau}A\,e^{-i\mathcal{K}_0\tau}.$$

We have not, as yet, indicated how the operator $\mathfrak{R}$ is to be chosen. We first make the assumption, alluded to in Sec. 1, that the values of the quantities $\langle d\mathfrak{F}_i^{(\alpha)}/dt\rangle$ are prescribed in advance. With these constraints we then require that the quantity

$$\langle d\mathfrak{R}/dt\rangle=\tfrac{1}{2}\int_{-\infty}^{\infty} d\tau(d\mathfrak{R}/dt,(d\mathfrak{R}/dt)(\tau)) \tag{3.4}$$

should be an extremum. Introducing the Lagrange multipliers $\mu_i^{(\alpha)}$, we must then demand that

$$\int_{-\infty}^{\infty} d\tau\{(\delta(d\mathfrak{R}/dt),(d\mathfrak{R}/dt)(\tau))+((d\mathfrak{R}/dt),\delta(d\mathfrak{R}/dt)(\tau))$$
$$+2\sum_i\sum_\alpha(\delta(d\mathfrak{R}/dt),\mu_i^{(\alpha)}\mathfrak{F}_i^{(\alpha)}(\tau))\} \tag{3.5}$$

should vanish. Using the symmetry property

$$((d\mathfrak{R}/dt),\delta(d\mathfrak{R}/dt)(\tau))$$
$$=(\delta(d\mathfrak{R}/dt),(d\mathfrak{R}/dt)(-\tau)), \tag{3.6}$$

we have

$$\int_{-\infty}^{\infty} d\tau\{(\delta(d\mathfrak{R}/dt),(d\mathfrak{R}/dt)(\tau))$$
$$+(\delta(d\mathfrak{R}/dt),\sum_i\sum_\alpha \mu_i^{(\alpha)}\mathfrak{F}_i^{(\alpha)}(\tau))\}=0 \tag{3.7}$$

which leads to

$$d\mathfrak{R}/dt=-\sum_\alpha\sum_i \mu_i^{(\alpha)}d\mathfrak{F}_i^{(\alpha)}/dt. \tag{3.8}$$

Equation (2.18) then gives for the fluxes,

$$\langle \mathcal{J}_{ij}^{(\alpha)}\rangle=\tfrac{1}{2}\int_{-\infty}^{\infty} d\tau\sum_\beta(\mathcal{J}_{ij}^{(\beta)},\mathcal{J}_{ij}^{(\alpha)}(\tau))(\mu_i^{(\beta)}-\mu_j^{(\beta)}). \tag{3.9}$$

These are the linear relations connecting the fluxes $\langle\mathcal{J}_{ij}^{(\alpha)}\rangle$ to the *affinities* $(\mu_j^{(\beta)}-\mu_j^{(\beta)})$ through the *kinetic coefficients*

$$L^{(\alpha\beta)}\equiv\tfrac{1}{2}\int_{-\infty}^{\infty} d\tau(\mathcal{J}_{ij}^{(\beta)},\mathcal{J}_{ij}^{(\alpha)}(\tau))$$
$$=L^{(\beta\alpha)}. \tag{3.10}$$

It is still necessary to specify the values of the "constants of motion" for the total system. For this we have at our disposal the operator $\mathcal{K}_0(\mathfrak{F}^{(\alpha)})$, which through the relations

$$\langle\mathfrak{F}^{(\alpha)}\rangle=\lim_{T\to\infty}T^{-1}\int_0^T \mathrm{Tr}\sigma(\tau)\mathfrak{F}^{(\alpha)}d\tau$$
$$=\mathrm{Tr}\{\bar{\sigma}(T\to\infty)\mathfrak{F}^{(\alpha)}\}$$
$$=\mathrm{Tr}e^{-(\mathcal{K}_0+\mathfrak{R})}\mathfrak{F}^{(\alpha)} \tag{3.11}$$

gives the values of the $\langle\mathfrak{F}^{(\alpha)}\rangle$. It is natural to require that if $\mathfrak{R}$ vanishes the values (3.11) should reduce to those given by the canonical density (1.2). We thus have

$$\mathcal{K}_0=\sum_\alpha \mu^{(\alpha)}\mathfrak{F}^{(\alpha)}-\Theta, \tag{3.12}$$

where Θ is a constant.

4. CONCLUSION

The condition (3.5) which determines $d\mathfrak{R}/dt$ is the essential statistical assumption of the theory. It will be recognized as a principle of minimum entropy production (MEP) and plays a role similar, in the present theory, to that of the principle of maximum entropy in the statistical mechanics of equilibrium states. In earlier treatments this principle is a derived result, the essential statistical assumption being that the process proceeds from an initial state of "local equilibrium." We have instead taken MEP as our fundamental statistical assumption and have thus avoided specifying the distribution of states at an initial time. Thus the affinities have been introduced in a natural way as Lagrange multipliers for the fluxes in much the same way as intensive variables appear as a result of constraints placed on the extensive variables in equilibrium statistical mechanics. This is appropriate in as much as the fluxes are fundamental quantities in irreversible phenomena.