Advance in Chemical Physics, Volume XXXIV
Edited by I. Prigogine, Stuart A. Rice

THE EXPANSION OF THE MASTER EQUATION

N. G. VAN KAMPEN

Institute of Theoretical Physics of the University, Utrecht, Netherlands

CONTENTS

I. INTRODUCTION

Macroscopic physics deals with macroscopic quantities q_ν, such as the positions and velocities of bodies, electrical charges and currents, amounts or concentrations of chemical compounds, temperatures at various points in a material, or local density and velocity of a fluid. They obey macroscopic laws

$$\dot{q}_\nu = F_\nu(q_1, q_2, \ldots) \tag{1.1}$$

which on a macroscopic level can be derived from general principles such as conservation laws, together with some specific assumptions of phenomenological nature, for example, those of Fourier and Fick. On the one hand this macroscopic picture is incomplete, because it has to introduce phenomenological coefficients, as many as there are phenomena, but this is of no concern to us here. On the other hand the macroscopic laws (1.1) are merely an approximation, valid when so many particles are involved that fluctuations are negligible. The present work is concerned with improving on this macroscopic approximation by taking into account the fluctuations. Of course the quantities that characterize the discreteness are essential: Boltzmann's constant, the elementary charge, and the masses of individual particles.

In order to study the corrections to (1.1) caused by the discrete nature of matter one must view the macroscopic phenomena as the outcome of the collective behavior of many particles. This does not merely require an investigation of the phenomenological assumptions mentioned above, but a reappraisal of the very definitions of the q_ν is needed.

In principle all information is contained in the microscopic equations of motion of all particles, but it hardly needs saying that an exact solution of these equations is beyond human means, excepting a small number of simple models.[1] Even the macroscopic laws (1.1) can only be derived from them with the help of simplifications and assumptions, which are no more reliable than the phenomenological assumptions used in the purely macroscopic approach. It is therefore sensible to embark upon a less ambitious program and to develop a theory which goes beyond the macroscopic description in that it includes fluctuations, but short-cuts the connection with the microscopic equations by an appeal to some suitably chosen semiphenomenological assumptions. This is the customary approach in noise theory; we propose to call it the *mesoscopic* level of description.

One popular mesoscopic approach consists in adding to (1.1) a fluctuating term

$$\dot{q}_\nu = F_\nu(q_1, q_2, \ldots) + l_\nu(t) \tag{1.2}$$

and making suitable assumptions concerning the statistical properties of the random functions $l_\nu(t)$ (see Section XI). It should be clear that this device changes the nature of the q_ν; they are now also stochastic quantities. The macroscopic values that enter into (1.1) are identified with the averages of the q_ν in (1.2). This approach was first used by Langevin in his treatment of the Brownian movement, and his success has led many authors to apply the same device to other systems.[2] However, we shall show in Section XI that in many cases it leads to wrong results.

A second approach starts out by introducing the probability distribution $P(q_1, q_2, \ldots; t)$ defined as follows: $P(q_1, q_2, \ldots; t)\, dq_1\, dq_2 \ldots =$ the joint

probability that at time t the first quantity has a value between q_1 and $q_1 + dq_1$, and the second one between q_2 and $q_2 + dq_2$, etc. Note that the q's have changed their nature again and are merely coordinates in some q-space on which the probability density P is defined. One then assumes that P obeys the Fokker-Planck equation

$$\frac{\partial P}{\partial t} = -\sum_{\nu} \frac{\partial}{\partial q_\nu} F_\nu(q)P + \sum_{\nu\mu} \frac{\partial^2}{\partial q_\nu \partial q_\mu} D_{\nu\mu} P \tag{1.3}$$

The F_ν are the same as in (1.1), and the new coefficients $D_{\nu\mu}$ are found from the fluctuation-dissipation theorem. Although this looks quite different from the Langevin approach, it is actually equivalent to it, and therefore subject to the same criticism (Section XXII).

A third mesoscopic approach is the basis of this article. It also starts out from the probability density P, but merely assumes that it obeys an equation of type

$$\dot{P} = \mathbf{W}P \tag{1.4}$$

where $\mathbf{W}$ is a linear operator acting on the q-dependence. Let $W(q|q')$ be the integral kernel of $\mathbf{W}$; then the requirement that the total probability must remain equal to unity tells

$$\int W(q|q')\, dq = 0$$

(Each q stands for the whole set of q_ν, and dq is a volume element in q-space.) Hence one may write (1.4) in the physically more transparent form

$$\dot{P}(q, t) = \int \{W(q|q')P(q', t) - W(q'|q)P(q, t)\}\, dq' \tag{1.5}$$

The kernel $W(q|q')$ for $q \neq q'$ represents the transition probability per unit time from q' to q and must be nonnegative. The second term represents the decrease of $P(q, t)$ due to transitions to other values q'. Equation (1.4) or (1.5) is called the "master equation".*

The assumption (1.4) implies that the stochastic process described by q is a Markov process. This is a strong assumption, which in most applications is only approximately true and with the conditions that a suitable coarse-grained time scale is used, and that the correct set of variables $q = \{q_1, q_2, \ldots\}$ is chosen. On the other hand, it is weaker than the assumptions needed in the two previously mentioned mesoscopic approaches. Moreover it is easier to assess on physical grounds. The transition probabilities $W(q|q')$ usually have a direct physical interpretation in terms of the microscopic quantities

* Throughout this article the term is used in its original sense:[3] an equation of the type (1.4) for a probability distribution.

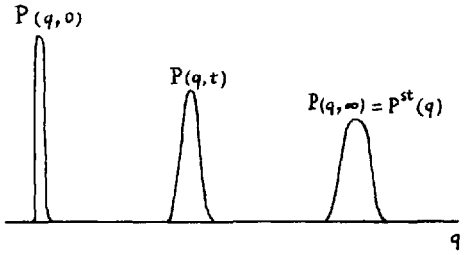

Fig. 1. The evolution of the probability density towards equilibrium.

describing the system, for instance, collision cross-sections or quantum mechanical matrix elements. We shall meet cases where the master equation (1.5) obviously holds, but neither the Langevin approach, nor the Fokker-Planck equation (1.3) leads to correct results.

The master equation purports to describe the entire behavior of the q and hence also the macroscopic equation (1.1) should follow from it. How is it possible for an equation governing the probability density in q-space to give rise to a deterministic set of equations for the q_v? The idea is that $P(q, t)$ is a sharp peak located at a rather well-defined point in q-space (Fig. 1). If the width of the peak may be neglected it is possible to consider its position in q-space as the macroscopic value of q_v. While P varies in time according to (1.5) the peak moves through q-space according to (1.1). Note that there is no contradiction between the fact that (1.5) is linear whereas (1.1) may well be nonlinear. The situation is analogous to the way in which a linear Schroedinger equation gives rise to a nonlinear classical equation of motion in the approximation in which the particle is heavy enough to neglect the spreading of the wave function.

The mathematical scheme describing this state of affairs was developed some time ago.[4–6] The present review is confined to the special but frequently occurring class of master equations in which the variable q takes only integral values. We therefore write n rather than q and the master equation is

$$\dot{P}(n, t) = \sum_{n'} \{W(n|n')P(n', t) - W(n'|n)P(n, t) \tag{1.6}$$

It may happen that n runs from $-\infty$ to $+\infty$, or from 0 to ∞, or only takes values in some finite range. The transition probabilities $W(n|n')$ need only be defined for $n \neq n'$ and are nonnegative. They are properties of the system and, of course, independent of the $P(n, t)$, which describe the special mesoscopic state considered.[7] It is possible to include the case that W depends on time (nonautonomous systems), but we shall not do so. Equation (1.6) may also be written by means of a matrix $\mathbf{W}$

$$\dot{P}(n, t) = \sum_{n'} \mathbf{W}_{nn'} P(n', t)$$

The master equation is "solved" if one can find the $P(n, t)$ that obey (1.6) and take arbitrarily prescribed initials values at $t = 0$. Obviously it suffices to consider the initial condition

$$P(n, 0) = \delta_{n,m} \tag{1.7}$$

for each m. The corresponding solution is

$$P(n, t \mid m, 0) = (e^{\mathbf{W}t})_{nm} \tag{1.8}$$

In order to evaluate this formal expression one has to diagonalize $\mathbf{W}$, but only in rare cases can that be done exactly. Hence it is necessary to have a systematic approximation scheme in the form of a power series expansion in some physical parameter. It appeared that the appropriate quantity is $\Omega^{-1/2}$, where Ω is a measure for the size of the system or the total number of particles involved. This scheme is demonstrated on a simple example in Section III, formulated in general in Sections IV and V, and subsequently applied to various problems.

It will appear that most of the problems treated in the literature can be readily handled with the Ω-expansion method. Many of the existing controversies and paradoxes[5] are caused by unsystematic approximations, in which terms are neglected according to the taste of the author. In addition it will be shown that the popular Langevin approach may lead to wrong results even in simple cases (Section XI), and the limitations of the Fokker–Planck equation are discussed in Section XXII. On the other hand, it must be stressed that the expansion is essentially based on the smallness of fluctuations and has only limited validity in unstable situations (Section XVIII) or phase transitions (Section XX).

II. PRELIMINARIES

The jump moments or derivate moments[8] are defined by

$$a_p(n) = \sum_{n'} (n' - n)^p W(n' \mid n) \qquad (p = 1, 2, \ldots) \tag{2.1}$$

Multiply (1.6) with n and sum

$$\begin{aligned} \frac{d}{dt}\langle n \rangle &= \sum_{nn'} \{nW(n \mid n')P(n') - nW(n' \mid n)P(n)\} \\ &= \sum_{nn'} (n' - n)W(n' \mid n)P(n) \\ &= \langle a_1(n) \rangle \end{aligned} \tag{2.2}$$

If $a_1(n)$ is a linear function this is identical with

$$\frac{d}{dt}\langle n\rangle = a_1(\langle n\rangle) \tag{2.3}$$

which permits us to determine $\langle n\rangle$ as a function of t. If, however, $a_1(n)$ is not linear, (2.3) is at best an approximation, which amounts to neglecting all fluctuations. We shall see that (2.3) is, indeed, the zeroth approximation in the Ω-expansion scheme, and is therefore to be identified with the macroscopic equation (1.1).* The exact identity (2.2) is not a closed equation for $\langle n\rangle$ but involves higher moments of n as well. To improve on the approximation (2.3) we expand in (2.2) the function $a_1(n)$ in $n - \langle n\rangle$ and break off after the second derivative:

$$\frac{d}{dt}\langle n\rangle = a_1(\langle n\rangle) + \tfrac{1}{2}a_1''(\langle n\rangle)\sigma_n^2 \tag{2.4}$$

where $\sigma_n^2 = \langle(n - \langle n\rangle)^2\rangle$. As this equation involves $\langle n^2\rangle$ we also multiply (1.6) with n^2 and sum

$$\begin{aligned}\frac{d}{dt}\langle n^2\rangle &= \sum_{nn'}(n'^2 - n^2)W(n'|n)P(n)\\ &= \langle a_2(n)\rangle + 2\langle n a_1(n)\rangle\end{aligned} \tag{2.5}$$

Combination with (2.2) yields the exact identity

$$\frac{d}{dt}\sigma_n^2 = \langle a_2(n)\rangle + 2\{\langle n a_1(n)\rangle - \langle n\rangle\langle a_1(n)\rangle\} \tag{2.6}$$

Making somewhat loose approximations similar to (2.3) we write for this

$$\frac{d}{dt}\sigma_n^2 = a_2(\langle n\rangle) + 2a_1'(\langle n\rangle)\sigma_n^2 \tag{2.7}$$

In Section V it will be shown that the pair of equations (2.4) and (2.7) together actually constitute a consistent approximation.[9,10]

The upshot is that in order to improve on (2.3) two coupled equations [(2.4) and (2.7)] are needed (unless $a_1(n)$ happens to be linear). That means that it is no longer possible to determine $\langle n\rangle$ from its initial value; one also needs to know the initial value of σ_n^2. All this is subject to the condition that σ_n^2 remains finite (of order n), otherwise there is no justification for omitting higher moments. This condition amounts to $a_1'(\langle n\rangle) < 0$, that is, the system must be stable (compare Section V).

* This statement requires a minor modification, see Section IV.

A special but important class of discrete Markov processes are the one-step or birth-and-death processes. They are defined by $W(n|n') = 0$ unless $n = n' \pm 1$, that is,

$$W(n|n') = r(n')\delta_{n,n'-1} + g(n')\delta_{n,n'+1}$$

$r(n)$ and $g(n)$ may be any two nonnegative functions, usually analytic; their names stem from recombination and generation of charge carriers in semiconductors.[11]

The master equation of a one-step process has the following form:

$$\dot{P}(n,t) = r(n+1)P(n+1,t) + g(n-1)P(n-1,t) - \{r(n) + g(n)\}P(n,t) \tag{2.8}$$

It is convenient to define the difference operator $\mathbf{E}$ by[12]

$$\mathbf{E}f(n) = f(n+1), \qquad \mathbf{E}^{-1}f(n) = f(n-1) \tag{2.9}$$

With its aid the master equation (2.8) may be written

$$\dot{P} = (\mathbf{E} - 1)r(n)P + (\mathbf{E}^{-1} - 1)g(n)P \tag{2.10}$$

The jump moments are

$$a_p(n) = (-1)^p r(n) + g(n) \tag{2.11}$$

The macroscopic rate equation (2.3) takes the form

$$\frac{d}{dt}\langle n\rangle = -r(\langle n\rangle) + g(\langle n\rangle) \tag{2.12}$$

and the coupled equations [(2.4) and (2.7)] are

$$\frac{d}{dt}\langle n\rangle = g(\langle n\rangle) - r(\langle n\rangle) + \tfrac{1}{2}\{g''(\langle n\rangle) - r''(\langle n\rangle)\}\sigma_n^2$$

$$\frac{d}{dt}\sigma_n^2 = g(\langle n\rangle) + r(\langle n\rangle) + 2\{g'(\langle n\rangle) - r'(\langle n\rangle)\}\sigma_n^2$$

Some general properties of one-step processes are listed in Section VI.

III. FIRST EXAMPLE: SPREADING OF AN EPIDEMIC

As a first example for demonstrating the expansion we choose a simple nonlinear one-step process, which describes the spreading of an epidemic in a population of Ω individuals.[6,13] If n is the number of infected individuals, the probability per unit time for a new infection to occur is proportional to n, and to the number $\Omega - n$ of uninfected. Thus $g(n) = \beta n(\Omega - n)$ with constant

β. Furthermore we take $r(n) = 0$, that is, no cure is possible. Hence

$$W(n|n') = \beta\delta_{n,n'+1}n'(\Omega - n') \tag{3.1}$$

The master equation is

$$\begin{aligned}\dot{P}(n, t) &= \beta(n - 1)(\Omega - n + 1)P(n + 1, t) - \beta n(\Omega - n)P(n, t)\\ &= \beta(\mathbf{E}^{-1} - 1)n(\Omega - n)P\end{aligned} \tag{3.2}$$

The more general problem with arbitrary $g(n)$ and $r(n) = 0$ has been treated by Weiss as a model for superradiance.[14] Actually such problems can be solved without approximations in a more or less closed form, but the result is too involved to be of much use, unless $g(n)$ is sufficiently simple.

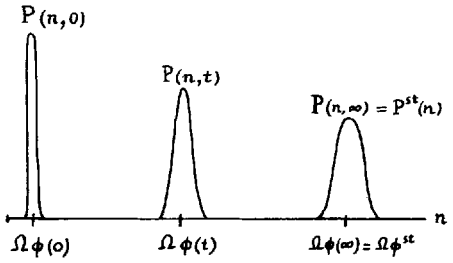

Fig. 2. The definition of the macroscopic part of a variable n.

One expects n to consist of a macroscopic part $\Omega\varphi(t)$ plus fluctuations of order $\Omega^{1/2}$. That is, $P(n, t)$ will be a sharp peak located roughly at $\Omega\varphi(t)$ with a width of order $\Omega^{1/2}$ (see Fig. 2). Hence we set

$$n = \Omega\varphi(t) + \Omega^{1/2}x \tag{3.3}$$

where x is the new variable and $\varphi(t)$ will be chosen presently. We shall call $\Omega\varphi(t)$ the "macroscopic part" and $\Omega^{1/2}x$ the "fluctuating part" of n, and refer to the new variables as the "Ω language." Accordingly the probability distribution of n now becomes a probability distribution Π of x,

$$\begin{aligned}P(n, t)\Delta n &= \Pi(x, t)\Delta x\\ \Pi(x, t) &= \Omega^{1/2}P(\Omega\varphi(t) + \Omega^{1/2}x, t)\end{aligned} \tag{3.4}$$

The following transformation formulas apply

$$\frac{\partial\Pi}{\partial x} = \Omega^{1/2}\frac{\partial P}{\partial n}, \qquad \frac{\partial\Pi}{\partial t} = \Omega^{1/2}\left\{\Omega\frac{d\varphi}{dt}\frac{\partial P}{\partial n} + \frac{\partial P}{\partial t}\right\}$$

Hence*

$$\Omega^{1/2}\frac{\partial P}{\partial t} = \frac{\partial \Pi}{\partial t} - \Omega^{1/2}\frac{d\varphi}{dt}\frac{\partial \Pi}{\partial x} \tag{3.5}$$

Moreover one has

$$\mathbf{E} = 1 + \Omega^{-1/2}\frac{\partial}{\partial x} + \frac{1}{2}\Omega^{-1}\frac{\partial^2}{\partial x^2} + \cdots \tag{3.6}$$

Substitute the new variables in the master equation (3.2) and cancel an overall factor $\Omega^{-1/2}$,

$$\frac{\partial \Pi}{\partial t} - \Omega^{1/2}\frac{d\varphi}{dt}\frac{\partial \Pi}{\partial x} = \beta\Omega^2\left\{-\Omega^{-1/2}\frac{\partial}{\partial x} + \frac{1}{2}\Omega^{-1}\frac{\partial^2}{\partial x^2}\right\} \times (\varphi + \Omega^{-1/2}x)(1 - \varphi - \Omega^{-1/2}x)\Pi \tag{3.7}$$

We absorb one factor Ω into the time variable (and for convenience also the β) by setting

$$\beta\Omega t = \tau \tag{3.8}$$

Then the largest terms are

$$-\Omega^{1/2}\frac{d\varphi}{d\tau}\frac{\partial \Pi}{\partial x} = -\varphi(1 - \varphi)\frac{\partial \Pi}{\partial x}$$

They can be made to cancel by subjecting φ to the equation

$$\frac{d\varphi}{d\tau} = \varphi(1 - \varphi) \tag{3.9}$$

This equation determines how the macroscopic part of n varies with time. Translating back to the original variables it takes the form

$$\frac{dn}{dt} = \beta n(\Omega - n) \tag{3.10}$$

which appears to be identical with the macroscopic rate equation (2.12).

The terms of order Ω^0 in (3.7) yield an equation for Π,

$$\frac{\partial \Pi}{d\tau} = -(1 - 2\varphi)\frac{\partial}{\partial x}x\Pi + \frac{1}{2}\varphi(1 - \varphi)\frac{\partial^2 \Pi}{\partial x^2} \tag{3.11}$$

This is a Fokker–Planck equation whose coefficients involve φ and therefore depend on time. Observe, however, that the coefficient of the first term is

* It is possible to arrive at (3.5) without the intervention of the dubious symbol $\partial P/\partial n$. Let t in (3.4) vary by δt and simultaneously x by $-\Omega^{1/2}\dot{\varphi}(t)\delta t$; this leads immediately to (3.5).

linear in x, and that the second term does not depend on x; we shall indicate these features of a Fokker–Planck equation by calling it *linear*. (Of course, all Fokker–Planck equations are linear in the unknown function—in this case Π.) Equation (3.11) governs the fluctuations in n of order $\Omega^{1/2}$ about the macroscopic part.

The strategy for solving the master equation (3.2) with initial condition (1.7) now emerges. First solve (3.9) with initial value $\varphi(0) = m/\Omega$. Then solve (3.11) with initial $\Pi(x, 0) = \delta(x)$. Then

$$P(n, t\,|\,m, 0) = \Omega^{-1/2}\Pi\left(\frac{n - \Omega\varphi(\tau)}{\Omega^{1/2}}, \tau\right)$$

In this solution terms of relative order $\Omega^{-1/2}$ have been neglected.

IV. THE GENERAL EXPANSION METHOD

The basic idea is that there is a parameter Ω measuring the size of the system, such that for large Ω the fluctuations are relatively small. It is then possible to expand in descending powers of Ω, as will be outlined in five steps.*

First step: specifying the dependence of the transition probabilities on Ω. It is assumed that the way in which $W(n\,|\,n')$ depends on Ω has the following form:

$$W(n\,|\,n') = f(\Omega)\left[\Phi_0\left(\frac{n'}{\Omega}; n - n'\right) + \Omega^{-1}\Phi_1\left(\frac{n'}{\Omega}; n - n'\right) + \cdots\right] \tag{4.1}$$

Each function Φ_j has a Taylor expansion with respect to its first argument, but is of course a discrete function of its second argument, which is the jump size. The factor $f(\Omega)$, usually some power of Ω, is innocuous because it can be absorbed in the time variable. The jump moments (2.1) are transformed accordingly,

$$a_p(n) = f(\Omega)\alpha_p\left(\frac{n}{\Omega}\right) \tag{4.2}$$

In the following we suppose for simplicity that $\Phi_1, \Phi_2, \ldots$ vanish. They are not hard to include when they occur, as in Section IX, but cumbersome in the general treatment. When they do not vanish it is not strictly true that (2.3) is identical with the macroscopic law, inasmuch as a_1 involves higher orders in $1/\Omega$, which do not belong to a macroscopic description. The macroscopic law is determined by the first jump moment of Φ_0 alone,[16] but in the next approximation (2.4) both Φ_0 and Φ_1 have to be used for a_1.

* Previously we have used the Kramers–Moyal expansion as a convenient intermediate step,[4,5] but we shall avoid it here, since its role has been misconstrued.[15]

Substituting (4.1) in (1.6) and changing the summation variable from n' to $\nu = n - n'$ one obtains for the master equation

$$\frac{\partial P(n,t)}{\partial t} = f(\Omega) \sum_{\nu} \left\{ \Phi_0\left(\frac{n-\nu}{\Omega}; \nu\right) P(n-\nu, t) - \Phi_0\left(\frac{n}{\Omega}; -\nu\right) P(n,t) \right\} \tag{4.3}$$

Second step: postulating the way in which P depends on Ω. One expects $P(n, t)$ to be a sharp peak located at some point $\Omega\varphi(t)$ with a width of order $\Omega^{1/2}$. Hence one transforms the variable n to a new variable x as in (3.3). This transforms $P(n, t)$ into $\Pi(x, t)$ according to (3.4). Substitute this in the master equation:

$$\frac{\partial \Pi}{\partial t} - \Omega^{1/2} \frac{d\varphi}{dt} \frac{\partial \Pi}{\partial x} = f(\Omega) \Bigg[\sum_{\nu} \Phi_0(\varphi(t) + \Omega^{-1/2}x - \Omega^{-1/2}\nu; \nu) \times \Pi(x - \Omega^{-1/2}\nu, t) - \sum_{\nu} \Phi_0(\varphi(t) + \Omega^{-1/2}x; -\nu)\Pi(x,t) \Bigg] \tag{4.4}$$

The factor [ ] vanishes to lowest order in $\Omega^{-1/2}$. To obtain the next order it is convenient to write it in the form

$$\left[\quad \right] = \sum_{\nu} \left\{ -\Omega^{-1/2}\nu \frac{\partial}{\partial x} + \frac{1}{2}\Omega^{-1}\nu^2 \frac{\partial^2}{\partial x^2} - \cdots \right\} \Phi_0(\varphi + \Omega^{-1/2}x; \nu)\Pi(x,t) \tag{4.5}$$

Third step: extracting the largest terms to obtain the macroscopic equation. The lowest order in [ ] is $\Omega^{-1/2}$; it can be combined with the term of order $\Omega^{1/2}$ on the left if we define a scaled time τ by*

$$f(\Omega)t = \Omega\tau \tag{4.6}$$

Then the largest terms are of order $\Omega^{1/2}$ on both sides,

$$-\Omega^{1/2} \frac{d\varphi}{d\tau} \frac{\partial \Pi}{\partial x} = \Omega\left(-\Omega^{-1/2} \frac{\partial \Pi}{\partial x} \right) \sum_{\nu} \nu \Phi_0(\varphi; \nu)$$

Since both terms involve Π only through the factor $\partial\Pi/\partial x$, it is possible to satisfy this equation by choosing for φ a solution of

$$\frac{d\varphi}{d\tau} = \sum_{\nu} \nu \Phi_0(\varphi, \nu) = \alpha_1(\varphi) \tag{4.7}$$

* In many of our examples we shall find $f(\Omega) = \Omega$ so that $t = \tau$.

This is the equation for the macroscopic part of n, that is, the macroscopic rate equation.

Fourth step: the next order determines the fluctuations. The terms of order Ω^0 in (4.4) are

$$\frac{\partial \Pi}{\partial \tau} = -\left\{\sum_\nu \nu \Phi_0'(\varphi; \nu)\right\} \frac{\partial}{\partial x} x\Pi + \frac{1}{2}\left\{\sum_\nu \nu^2 \Phi_0(\varphi; \nu)\right\} \frac{\partial^2 \Pi}{\partial x^2}$$
$$= -\alpha_1'(\varphi) \frac{\partial}{\partial x} x\Pi + \frac{1}{2}\alpha_2(\varphi) \frac{\partial^2 \Pi}{\partial x^2} \tag{4.8}$$

The prime indicates differentiation with respect to φ. This is again a linear Fokker–Planck equation with time-dependent coefficients, which governs the fluctuations in n of order $\Omega^{1/2}$ about the macroscopic part $\Omega\varphi$.

Final step: collecting the results in order to solve (1.6) with initial condition (1.7). First solve (4.7) with initial condition

$$\Omega\varphi(0) = m \tag{4.9}$$

and call the solution $\varphi(\tau | m/\Omega)$. Next solve (4.8) with initial condition

$$\Pi(x, 0) = \delta(x) \tag{4.10}$$

and call the solution $\Pi(x, \tau | 0, 0)$. Then

$$P(n, t | m, 0) = \Omega^{-1/2} \Pi\left(\frac{n - \Omega\varphi(\tau | m/\Omega)}{\Omega^{1/2}}, \tau \middle| 0, 0\right) \tag{4.11}$$

Note that one has to the same order

$$P(n, t | m, 0) = \Omega^{-1/2} \Pi\left(\frac{n - \Omega\varphi(\tau | m/\Omega - c\Omega^{1/2})}{\Omega^{1/2}}, \tau \middle| c, 0\right) \tag{4.12}$$

where c is an arbitrary number of order 1.

This program can be carried out by a number of integrations (see Appendix). It is simpler, however, and in many cases sufficient to determine only $\langle n \rangle$ and $\sigma_n^2 = \langle n^2 \rangle - \langle n \rangle^2$ as functions of t. The relevant formulas are derived in the next section.

Higher orders can be added and have the effect of modifying the equation for $\Pi(x, t)$ (see Section VIII). However, we shall be mainly concerned with the approximation to order Ω^0 as given here. This will be called the *linear noise approximation* since to this order the fluctuations are governed by the linear Fokker–Planck equation (4.8). It is the approximation on which the familiar theory of noise in electrical networks[11] is based.

V. THE EQUATIONS FOR THE MOMENTS

Without actually solving (4.8) one may deduce directly from it (by multiplying with x and integrating)

$$\frac{d}{d\tau}\langle x\rangle = \alpha_1'(\varphi)\langle x\rangle \tag{5.1}$$

Observe that this is identical with the "variational equation" belonging to (4.7), that is, the equation for the difference between two neighboring solutions of (4.7). The fact that this must be so can be gleaned from (4.12); a slight variation (of order $\Omega^{-1/2}$) in the initial value of φ can be compensated by the initial value of $\langle x\rangle$. An important consequence is the following: Since the variational equation of (4.7) determines the stability of the macroscopic solution $\varphi(t)$, it follows that the macroscopic stability also determines whether or not the average $\langle x\rangle$ of the fluctuations grows with time.

One also deduces directly from (4.8)

$$\frac{d}{d\tau}\langle x^2\rangle = 2\alpha_1'(\varphi)\langle x^2\rangle + \alpha_2(\varphi) \tag{5.2}$$

In both (5.1) and (5.2) terms of order $\Omega^{-1/2}$ have been neglected. With the choice of (4.9) or (4.10) for the initial values one has at $t = 0$

$$\langle x\rangle_0 = 0, \qquad \langle x^2\rangle_0 = 0 \tag{5.3}$$

Hence $\langle x\rangle$ remains zero at all $t > 0$, so that

$$\langle n\rangle_t = \Omega\varphi(t) + \mathcal{O}(1) \tag{5.4}$$

To the present order, therefore, the macroscopic part of n is also its average.

Furthermore, if the macroscopic solution $\varphi(\tau)$ is stable, and therefore also (5.1), it follows from (5.2) that $\langle x^2\rangle$ remains finite as well. Consequently x remains of order unity at all times, which constitutes the *a posteriori* justification of the Ansatz (3.3). Note that the stability is crucial for our approximation scheme: if $\langle x^2\rangle$ grows exponentially in time, the separation of powers of Ω becomes invalid after a time of order long Ω. In Section XVIII we shall meet an example where $\langle x^2\rangle$ grows linearly with time.

It is possible to improve the equation for $\langle n\rangle$ by one order without going beyond the linear noise approximation. To this end we rewrite the exact equation (2.2) in the Ω language in order to display the powers of Ω,

$$\begin{aligned}\frac{d}{d\tau}(\varphi + \Omega^{-1/2}\langle x\rangle) &= \langle \alpha_1(\varphi + \Omega^{-1/2}x)\rangle \\ &= \alpha_1(\varphi) + \Omega^{-1/2}\alpha_1'(\varphi)\langle x\rangle + \tfrac{1}{2}\Omega^{-1}\alpha_1''(\varphi)\langle x^2\rangle + \mathcal{O}(\Omega^{-3/2})\end{aligned}$$

Since φ obeys (4.7) by definition,

$$\frac{d}{d\tau}\langle x\rangle = \alpha_1'(\varphi)\langle x\rangle + \tfrac{1}{2}\Omega^{-1/2}\alpha_1''(\varphi)\langle x^2\rangle + \mathcal{O}(\Omega^{-1}) \tag{5.5}$$

Although this equation for $\langle x\rangle$ involves $\langle x^2\rangle$ there is a factor $\Omega^{-1/2}$; hence the approximation (5.2) for $\langle x^2\rangle$ suffices. Rewriting the result in the original variables we see that the two coupled equations (2.4) and (2.7) determine $\langle n\rangle$ to order Ω^0 and σ_n^2 to order Ω^1.

The conclusion reached here in a slightly devious manner can also be obtained by simply adding the next order correction to (4.8) and then computing the first and second moments to the desired order (compare Section VIII).

VI. ONE-STEP PROCESSES

One-step processes have been defined by processes that obey the master equation (2.8) or (2.10). However, it is necessary to specify in addition the range of n. There are three possibilities: (*a*) all integers, $-\infty < n < \infty$; (*b*) half-infinite range, $n = 0, 1, 2, \ldots$; (*c*) finite range, $n = 0, 1, 2, \ldots, N$. If the range consists of several intervals with gaps between them, a one-step process cannot have transitions between them, so that the process decomposes into several independent processes.

If $r(n)$ and $g(n)$ are constants and n ranges from $-\infty$ to $+\infty$, the one-step process is identical with the (unsymmetric) random walk. The master equation can then easily be solved explicitly and no Ω expansion is needed. If r and g are constant and n has a limited range, for example, $n = 0, 1, 2, \ldots, \infty$, then (2.8) cannot be valid for all n. It can at best hold for $n = 1, 2, \ldots$, whereas for $n = 0$ it must have a slightly different form. We shall then call the boundary at $n = 0$ *artificial*. The random walk with one or two artificial boundaries can still be solved explicitly and will therefore not be considered.

If $r(n)$ and $g(n)$ are linear functions of n, there must be at least one boundary to prevent them from becoming negative. Again this makes a modification of (2.8) necessary. The following particular case is of special interest and will be called a *natural boundary*. Suppose again $n = 0, 1, 2, \ldots$. Then $n = 0$ is a natural boundary if (Fig. 3)

(*a*) $$r(0) = 0 \tag{6.1a}$$

(*b*) the modified equation at $n = 0$ is

$$\dot{P}(0, t) = r(1)P(1, t) - g(0)P(0, t) \tag{6.1b}$$

Note that this is identical with (2.8) for $n = 0$ if one knows that $P(-1, t) = 0$.

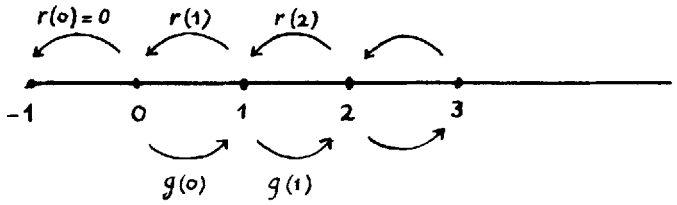

Fig. 3. The one-step process with an absorbing boundary. The boundary is "natural" when $r(0) = 0$.

Although artificial boundaries often create considerable difficulties for solving the master equation,[17] natural boundaries do not—for the following reason. Consider the master equation (2.8) with natural boundary (6.1b). Now solve (2.8) for $-\infty < n < \infty$, paying no attention to the boundary. Then take an initial state in which $P(n, 0) = 0$ for $n < 0$, for instance (1.7) with some nonnegative m. It will now be true that $P(n, t)$ for $n < 0$ remains zero at all $t > 0$; owing to (6.1a) no probability spills over to negative n. Hence (6.1b) is automatically satisfied, since it is implied in (2.8) if one has $P(-1, t) = 0$.

An upper boundary $n = N$ is called natural if $g(N) = 0$ and the modified equation for $P(N, t)$ is obtained from (2.8) by setting $P(N + 1, t) = 0$. (In a way infinity might also be considered as a natural boundary.) The one-step process with linear or constant $r(n)$ and $g(n)$ and no other than natural boundaries can be solved explicitly, for example with the aid of generating functions.

If $r(n)$ or $g(n)$ or both are nonlinear functions, for instance polynomials, the definition (6.1) of a natural boundary remains valid. Explicit solutions of such master equations are rare, but it is always possible to find the stationary, that is, time-independent solution. For this purpose write (2.10) in the form

$$0 = (\mathbf{E} - 1)\{r(n)P^{st}(n) - \mathbf{E}^{-1}g(n)P^{st}(n)\} \tag{6.2}$$

It follows that $\{\ \}$ must be constant

$$r(n)P^{st}(n) - g(n-1)P^{st}(n-1) = J \tag{6.3}$$

J is the net probability flow from n to $n - 1$. Using (6.3) one can construct the successive $P^{st}(n)$, starting from a single one, for instance $P^{st}(0)$, which then serves as a normalizing factor.

If there is a natural boundary, for instance at $n = 0$, one finds on substituting $n = 0$ in (6.3) that J must vanish:

$$r(n)P^{st}(n) = g(n-1)P^{st}(n-1) \tag{6.4}$$

It then follows directly that

$$P^{st}(n) = \frac{g(n-1)g(n-2)\cdots g(0)}{r(n)r(n-1)\cdots r(1)} P^{st}(0) \tag{6.5}$$

The normalizing factor $P^{st}(0)$ is subsequently found from

$$[P^{st}(0)]^{-1} = \sum_{n=0}^{N} \frac{g(n-1)g(n-2)\cdots g(0)}{r(n)r(n-1)\cdots r(1)} \tag{6.6}$$

When the upper bound N is infinite it may happen that the sum does not converge. In that case every solution $P(n, t)$ continues to spread out indefinitely, in the same way as in the familiar random walk.

It should be emphasized that (6.4) is simply a mathematical identity for one-step processes. It has to be distinguished from detailed balance, which for one-step processes reads

$$r(n)P^{eq}(n) = g(n-1)P^{eq}(n-1) \tag{6.7}$$

Here P^{eq} is the thermal equilibrium distribution and is known *a priori* from the familiar phase space argument of equilibrium statistical mechanics. On the one hand, detailed balance is not restricted to one-step processes; on the other hand it only applies to closed physical systems, without magnetic field or overall rotation.[18] The identity (6.4) also holds for open systems, for example, the photoconductor mentioned in the next section, and for population problems.

VII. SEMICONDUCTOR

As a second example of a nonlinear one-step process consider the following model of an intrinsic semiconductor. A crystal has a nearly empty conduction band and a nearly full valence band. Let n denote the number of electrons that by thermal fluctuations have been excited into the conduction band. The probability per unit time for an excitation to occur is $g(n) = \beta\Omega$, where Ω is the volume of the crystal and β a constant (see Fig. 4). The probability for a recombination is proportional to the number of excited electrons and to the density n/Ω of the available holes: $r(n) = \gamma n^2/\Omega$. Thus the macroscopic rate equation is

$$\frac{dn}{dt} = \beta\Omega - \frac{\gamma}{\Omega} n^2 \tag{7.1}$$

On the mesoscopic level the process is specified by the transition probabilities:

$$W(n|n') = \beta\Omega\delta_{n,n'+1} + \frac{\gamma}{\Omega} n'^2 \delta_{n,n'-1} \tag{7.2}$$