On the Theory of the Brownian Motion II

MING CHEN WANG AND G. E. UHLENBECK
H. M. Randall Laboratory of Physics, University of Michigan, Ann Arbor, Michigan

1. INTRODUCTION

IN 1930, Ornstein and one of us[1] tried to summarize and partially extend the existing theory of the Brownian motion for simple systems like the free particle and the harmonic oscillator. Since that time the theory has been developed and clarified to a considerable extent, so that it seems worth while again to try to summarize the theory. In this we will restrict ourselves to the case of the Brownian motion of a system of coupled harmonic oscillators, or in the electrical analogy to the theory of the thermal noise in a linear, passive network.[2] It is now clear that in this case we have to do with the theory of the so-called Gaussian random process, and we shall try, therefore, to present the theory of the Brownian motion against the background of the general theory of the random process.[3] This will also allow us to show the connection with some of the mathematical literature on this subject.

There are two approaches to the theory of the Gaussian random process. In the first the attention is focused on the actual random variation in time of the displacement, or voltage, or whichever variables of the system one is especially interested in. One usually[4] develops this variable in a Fourier series in time, of which the coefficients can vary in a random fashion. A fundamental notion is the notion of the spectrum of the random process, and the connection between the spectrum and the so-called correlation function is one of the basic theorems. For many purposes, and especially in the electrical case when the "noise" passes through *non-linear* circuit elements (like rectifiers for instance), this method is the most natural. Recently the method has been applied systematically to a whole series of problems by S. O. Rice,[5] and we shall call it, therefore, the *method of Rice* or the *Fourier series method*, and in the following we shall give only a short account of it.

The second method is the *method of Fokker-Planck* or the *diffusion equation method*. Macroscopically, for an ensemble of particles or systems, the variations which occur are like a diffusion process. The distribution function of the random variables of the system will, therefore, fulfill a partial differential equation of the diffusion type, and this is the basic equation of the method.[6] We shall discuss this method in more detail, mainly because, thanks to a recent article by Kramers,[7] one is now able to derive the distribution function for any of the random variables in the Brownian motion of a system of coupled oscillators. Thus it becomes completely clear that the two methods give identical results, and the relation between the two methods can then perhaps be better appreciated.

2. THE GENERAL RANDOM PROCESS

Roughly speaking, what we mean by a random process $y(t)$ is a process in which the variable y[8]

[1] G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930). We will refer to this paper as I.

[2] Only in the last section, where we shall mention some unsolved problems, we shall go beyond this restriction.

[3] Recently S. Chandrasekhar has also reviewed several aspects and applications of the general theory in Rev. Mod. Phys. **15**, 1 (1943). Although there will be some overlapping, we hope that our review will complement the exposition of Chandrasekhar.

[4] It sometimes is convenient to use a development in other sets of orthogonal functions.

[5] S. O. Rice, Bell Tel. J. **23**, 282 (1944); **25**, 46 (1945). We shall refer to these papers as Rice I and II. One finds here also references to previous applications of the method.

[6] For simple examples see I, and also R. Furth, Ann. d. Physik **53**, 177 (1917) and Riemann-Weber, 3rd ed. Vol. II, p. 177. In the mathematical literature the method has been analyzed by A. Kolmogoroff, Math. Ann. **104**, 415 (1931); **108**, 149 (1933); and by W. Feller, Math. Ann. **113**, 113 (1936); Trans. Am. Math. Soc. **48**, 488 (1940).

[7] H. A. Kramers, Physica **7**, 284 (1940).

[8] It may be that y is the displacement or velocity of a particle in Brownian motion or a fluctuating voltage or current when we have thermal noise. It may also denote a combination of two or more of such quantities, and we shall speak then of two-dimensional or more dimensional random processes. In the following, everything will be written as if y and t were continuous variables. This is *not* necessary; it may happen that either y or t or both y and t can assume only discrete values. We propose to let the words continuous and discrete refer only to the dependent variable y; while the words process and series refer to continuous and discrete t respectively. The well-

does not depend in a completely definite way on the independent variable t (=time), as in a *causal* process; instead one gets in different observations different functions $y(t)$, so that only certain probability distributions are directly observable. In fact the random process $y(t)$ is completely described (or defined) by the following *set of probability distributions:*

$W_1(yt)dy$ = probability of finding y in the range $(y, y+dy)$ at time t.

$W_2(y_1t_1; y_2t_2)dy_1dy_2$ = joint probability of finding y in the range (y_1, y_1+dy_1) at time t_1 *and* in the range (y_2, y_2+dy_2) at time t_2.

$W_3(y_1t_1;\ y_2t_2;\ y_3t_3)dy_1dy_2dy_3$ = joint probability of finding a triple of values of y in the ranges dy_1, dy_2, dy_3 at times t_1, t_2, t_3. (1)

And so on! The set of functions (1) must fulfill the following obvious conditions:

(*a*) $W_n \geqq 0$.

(*b*) $W_n(y_1t_1; y_2t_2\cdots y_nt_n)$ is a symmetric function in the set of variables $y_1t_1, y_2t_2\cdots y_nt_n$. This is clear since W_n is a *joint* probability.

(*c*) $$W_k(y_1t_1;\ \cdots y_kt_k) = \int\cdots\int dy_{k+1}\cdots dy_n W_n(y_1t_1\cdots y_nt_n)$$

since each function W_n must imply all the previous W_k with $k<n$. The set of functions (1) form, therefore, a kind of hierarchy; they describe successively the random process in more detail.[9]

To determine the functions W_n experimentally, it is clear that one needs a great number of records $y(t)$ obtained from a great number of experiments on "similarly prepared" systems (an ensemble of observations). To find then, for instance, $W_1(yt)$, one determines at a definite time t how often in the different experiments y occurs in a given interval $(y, y+\Delta y)$, etc. In most applications (and especially for the Brownian motion problems) we can make, however, a simplification because the processes are *stationary in time.* This means that the underlying "mechanism" which causes the fluctuations does not change in course of time. A shift of the t-axis will then not influence the functions W_n, and as a result the set (1) becomes:

$W_1(y)dy$ = probability of finding y between y and $y+dy$.

$W_2(y_1y_2t)dy_1dy_2$ = joint probability of finding a pair of values of y in the ranges dy_1 and dy_2, which are a time interval t apart from each other (t is therefore $=t_2-t_1$).

And so on again. These functions can now be experimentally determined from *one* record $y(t)$ taken over a sufficiently long time. One can then cut the record in pieces of length T (where T is long compared to all "periods" occurring in the process), and one may consider the different pieces as the different records of an ensemble of observations. In computing average values one has *in general* to distinguish between an ensemble average and a time average. However, for a stationary process these two ways of averaging will always give the same result, and one can, therefore, use either of them.

3. CLASSIFICATION OF RANDOM PROCESSES

The set of probability distributions (1) leads immediately to a method of classifying the random processes.

(*a*) We shall call a random process a *purely random process* when the successive values of y are not correlated at all. This means that:

$$W_2(y_1t_1; y_2t_2) = W_1(y_1t_1)\cdot W_1(y_2t_2)$$

---

known theory of Smoluchowski for the concentration fluctuations of a colloidal suspension is in our terminology an example of the analysis of a discrete random series. The general random walk problem is an example of the theory of continuous random series, etc. One finds a complete account of these examples in the article of Chandrasekhar.

[9] So far as we know, the first attempt to give a general theory of a random process is contained in two papers by L. S. Ornstein and H. C. Burger, Versl. Kon. Acad. Amst. **27**, 1146 (1919); **28**, 183 (1919). Here the set of distributions (1) and the property (*c*) are mentioned. Compare further A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Berlin, 1933), p. 27; H. Wold, "A study in the analysis of stationary time series, Diss. Uppsala (1938); B. Hostinky, Ann. Inst. H. Poincaré, **3**, 1 (1933); **7**, 69 (1937). The authors are aware of the fact that in the mathematical literature (especially in papers by N. Wiener, J. L. Doob, and others; cf. for instance Doob, Ann. Math. **43**, 351 (1942), also for further references) the notion of a random (or stochastic) process has been defined in a much more refined way. This allows for instance to determine in certain cases the probability that the random function $y(t)$ is of bounded variation, or continuous or differentiable, etc. However, it seems to us that these investigations have not helped in the solution of problems of direct physical interest, and we will, therefore, not try to give an account of them.

and analogously for the higher W_n. All the information about the process is then completely contained in the first distribution function W_1. When t is discrete, it is easy to give examples, but for continuous t, the purely random process can only be considered as a kind of limiting case; in any actual example, the y_1 and y_2 will surely be correlated when the time interval t_2-t_1 is small enough.

(*b*) In the next more complicated case, all the information about the process will be contained in W_2. Such processes are called *Markoff processes*. For the more precise definition it is useful first to introduce the notion of *conditional probabilities*. We will write $P_2(y_1|y_2, t)dy_2$ for the probability that given y_1 one finds y in the range (y_2, y_2+dy_2) a time t later. Of course, one finds P_2 from W_2 according to[10]

$$W_2(y_1y_2t)=W_1(y_1)P_2(y_1|y_2, t), \tag{2}$$

P_2 must further fulfill the obvious relations (which also follow from the properties (*a*), (*b*), (*c*), of Section 2):

$$P_2(y_1|y_2, t)\geqq 0, \tag{3a}$$

$$\int dy_2P_2(y_1|y_2, t)=1, \tag{3b}$$

$$W_1(y_2)=\int W_1(y_1)P_2(y_1|y_2, t)dy_1, \tag{3c}$$

while in the Brownian motion problems one also always has:[11]

$$\lim_{t\to\infty} P_2(y_1|y_2, t)=W_1(y_2). \tag{4}$$

Analogously one can introduce higher order conditional probabilities, and we will use an analogous notation; a bar will always separate the variables which are given from those for which the probability has to be found. All this holds, of course, still for any stationary random process. A Markoff process can now be defined more precisely by stating that for such a process the conditional probability that y lies in the interval (y_n, y_n+dy_n) at time t_n, given that y is equal to $y_1, y_2\cdots y_{n-1}$ at the times $t_1, t_2\cdots t_{n-1}$ (where $t_n>t_{n-1}\cdots>t_2>t_1$) depends besides on y_nt_n *only* on the value of y at the previous time t_{n-1}. Or in a formula, a Markoff process is defined by the equation:

$$P_n(y_1t_1, y_2t_2\cdots y_{n-1}t_{n-1}|y_nt_n) = P_2(y_{n-1}t_{n-1}|y_nt_n). \tag{5}$$

This makes it clear that all the W_n for $n>2$ can be found, when only W_2 is known. One derives for instance easily from (5) that:

$$W_3(y_1t_1, y_2t_2, y_3t_3)=\frac{W_2(y_1t_1, y_2t_2)W_2(y_2t_2, y_3t_3)}{W_1(y_2t_2)},$$

and so on. It is clear, therefore, that W_2 or P_2 completely describes the process. However, one cannot take P_2 as an arbitrary function of its variables. Besides the general relations (3) and (4), it must fulfill:

$$P_2(y_1|y_2, t)=\int dyP_2(y_1|y, t_0)P_2(y|y_2, t-t_0), \tag{6}$$

for all values of t_0 between zero and t. This follows immediately from the definition of a Markoff process and is called the *Smoluchowski equation*. It is the basic equation for the theory.

(*c*) In this way one can go on. The next class of processes will be completely described by giving W_3. However, in the physical applications there are very few examples studied of such higher order processes. Very often, when a process is *not* a Markoff process one can still consider it as a kind of "projection" of a more complicated Markoff process. Besides y, one then considers another dependent variable z (which may be, for instance, dy/dt or it may be a coordinate of another system), and it may be that for the two variables y, z combined, the process is then a Markoff process, so that:

$$P_2(y_1z_1|y_2z_2, t) = \int\int dydzP_2(y_1z_1|yz, t_0)P_2(yz|y_2z_2, t-t_0).$$

The $W_2(y_1y_2t)$ which one obtains by integrating $W_2(y_1z_1y_2z_2t)$ over z_1 and z_2 will then in general *not* be a Markoff process, and one can say that this is due to the fact that one has not given a complete enough description of the process.

[10] From now on we shall restrict ourselves to stationary processes.

[11] This property excludes, for instance, the existence of "hidden periodicities." In the theory of noise it excludes the presence of "signals."

Whether it is always possible to find the appropriate variables $z_1, z_2 \cdots$ (there may be more than one) so as to complete the given process to a Markoff process will, of course, depend on the physical "causes" of the fluctuation phenomena in question. As we shall see, for the theory of the Brownian motion, such a completion will *always* be possible, so that in some sense we will always have to do with Markoff processes.

4. THE RELATION BETWEEN THE SPECTRUM AND THE CORRELATION FUNCTION[12]

Suppose that one considers for a very long time T a stationary random process $y(t)$ whose average value is zero. Taking $y(t)=0$ outside the time interval T, one can develop the resulting function in a Fourier integral:

$$y(t)=\int_{-\infty}^{+\infty} dfA(f)e^{2\pi ift}, \tag{7}$$

where $A(f)=A^*(-f)$, since $y(t)$ is real. It is well known (Parzeval theorem) that:

$$\int_{-\infty}^{+\infty} y^2(t)dt=\int_{-\infty}^{+\infty} |A(f)|^2 df.$$

Using the fact that $|A(f)|^2$ is an even function of f and going to the limit $T\rightarrow\infty$, one can write this equation in the form:

$$\langle y^2\rangle_{\text{Av}}=\lim_{T\rightarrow\infty}\frac{1}{T}\int_{-\infty}^{+\infty} y^2(t)dt=\int_0^{\infty} dfG(f), \tag{8}$$

where

$$G(f)=\lim_{T\rightarrow\infty}\frac{2}{T}|A(f)|^2 \tag{9}$$

is the *spectral density.*

Consider next the average value:

$$\langle y(t)y(t+\tau)\rangle_{\text{Av}}=\lim_{T\rightarrow\infty}\frac{1}{T}\int_{-\infty}^{+\infty} y(t)y(t+\tau)dt. \tag{10}$$

By introducing the Fourier expansion (7) and using the Fourier integral theorem, one shows easily:

$$\langle y(t)y(t+\tau)\rangle_{\text{Av}}=\int_0^{\infty} dfG(f)\cos 2\pi f\tau, \tag{11a}$$

from which follows by inversion:

$$G(f)=4\int_0^{\infty} d\tau\langle y(t)y(t+\tau)\rangle_{\text{Av}}\cos 2\pi f\tau. \tag{11b}$$

This is the relation referred to in the title of this section. One can express it by saying that the *correlation function*

$$\rho(\tau)=\frac{\langle y(t)y(t+\tau)\rangle_{\text{Av}}}{\langle y^2\rangle_{\text{Av}}} \tag{12a}$$

and the *normalized spectrum:*

$$S(f)=\frac{G(f)}{\int_0^{\infty} dfG(f)} \tag{12b}$$

are each other's Fourier cosine transform, so that they are uniquely related to each other. For an almost pure random process, $\rho(\tau)$ is a function, which starting from unity drops very rapidly to zero, and as a result $S(f)=$constant except for very high frequencies. We call this a *white spectrum;* of course, the case that $S(f)=$constant for all f, which corresponds to a pure random process, is a limiting case, which will never occur in practice. When $S(f)$ has a sharp maximum around f_0, then $\rho(\tau)$ will look like a damped oscillation with roughly the frequency f_0.

5. SOME REMARKS ON THE THEORY OF DISCRETE RANDOM SERIES[13]

We will restrict ourselves to Markoff processes. The problem will then always be to determine $P(n|m, s\tau)$ when one knows $P(n|m, \tau)$. Here P is the analogue of the $P(y_1|y_2, t)$; y_1, y_2 can only have discrete values n, m and also the time t can only have discrete values $s\tau$ with $s=1, 2, 3, \cdots$. From now on we will drop the τ and write also $Q(n, m)$ for $P(n|m, \tau)$ in order to emphasize that

[12] This relation is contained in the paper of N. Wiener, Acta Math. **55**, 117 (1930) on generalized harmonic analysis. It was rediscovered by Khintchine, Math. Ann. **109**, 604 (1934). See also the dissertation of H. Wold for further references and for the formulation in the discrete case. In 1938, G. I. Taylor (Proc. Roy. Soc. **164**, 476 (1938)) gave a beautiful application of the theorem to the theory of turbulence. Cf. also Rice I, p. 310.

[13] The purpose of this section is only to present some of the ideas which are of importance for the understanding of the Fokker-Planck method. For a complete discussion compare Hostinky, see reference 9 and also M. Fréchet, *Traité du Calcul des Probabilités* (1938), Vol. I, Part II, Section 3. For the discussion of the important application to the theory of the concentration fluctuations, see Chandrasekhar, reference 3.

it is the basic probability which must be given from the "mechanism" or the "physical cause" of the random process. To find then $P(n|m, s)$ one can try to make successive use of the Smoluchowski equation:

$$P(n|m, s)=\sum_k P(n|k, s-1)Q(k, m). \quad (13)$$

However, for large values of s this is usually not practicable, and one has to look for other methods.

It is instructive to write (13) in a different way by remembering that:

$$\sum_m Q(k, m)=1,$$

or

$$Q(k, k)+\sum_m{}' Q(k, m)=1,$$

where the prime means that the value $m=k$ must be omitted. Using this and dropping in (13) the initial value n one can write Eq. (13) in the form:

$$\begin{aligned} P(m, s)-P(m, s-1) &= -P(m, s-1)\sum_k{}' Q(m, k) \\ &\quad +\sum_k{}' P(k, s-1)Q(k, m). \quad (13a) \end{aligned}$$

One can interpret this by saying that the rate of change of $P(m, s)$ with the time $(=s)$ is owing to the "gains" of P because of transitions from k to m minus the "losses" of P because of the transitions from m to all possible k. It is clear, therefore, that (13a) is completely analogous to the well known Boltzmann equation in the kinetic theory of gases.[14] One must solve such an equation for a given "initial" distribution; in our case this is the way the variable n comes in since:

$$P(m, 0)=\delta(n, m), \quad (14)$$

where $\delta(n, m)$ denotes the Kronecker symbol.

In many cases the process has the property that the dependent variable k can change in one step at most by ± 1. This means that $Q(k, m)=0$ except when $m=k$, $k\pm 1$, and Eq. (13) or (13a) becomes then a rather simple difference equation. To illustrate this we will consider two examples.

(*a*) Discrete random walk problem in one dimension. This is the simplest possible case; a point can move on a straight line in steps Δ; at each time moment s there is an equal chance that the point makes a step Δ to the right or to the left. If at $s=0$ the point is at the position $n\Delta$, what is the probability $P(n|m, s)$ that at time s the point is at the positions $m\Delta$. It is clear that the basic transition probability $Q(k, m)$ is given by:

$$Q(k, m)=\tfrac{1}{2}\delta(m, k-1)+\tfrac{1}{2}\delta(m, k+1).$$

Introducing this in (13) and dropping again the initial state n, one obtains the difference equation:

$$P(m, s)=\tfrac{1}{2}P(m+1, s-1)+\tfrac{1}{2}P(m-1, s-1), \quad (15)$$

which has to be solved with the initial condition (14). The solution is very easy to obtain; with $\nu=|n-m|$ one gets:

$$P(n|m, s)=\frac{s!}{\left(\frac{\nu+s}{2}\right)!\left(\frac{\nu-s}{2}\right)!}\left(\frac{1}{2}\right)^s. \quad (16)$$

The first probability distribution $W_1(n)$ should of course become independent of n, and will therefore be not strictly normalizable except when one limits the number of positions. This is also in accord with the general relation (3c). One sees easily that in this case $P(n|m, s)$, besides fulfilling the general Eq. (3b), fulfills in addition the special relation:

$$\sum_n P(n|m, s)=1,$$

and as a result the equation (analogous to (3c)):

$$W(m)=\sum_n W(n)P(n|m, s)$$

has the solution $W(n)=\text{constant}$.

(*b*) An example of Ehrenfest[15]—Suppose now that the fundamental transition probability $Q(k, m)$ has the form:

$$Q(k, m)=\frac{R+k}{2R}\delta(m, k-1)+\frac{R-k}{2R}\delta(m, k+1),$$

where R is a given integer. In the language of the random walk problem of the previous example, this means that there is an attractive center; the probabilities for making a step Δ to the right or left are not more equal but $\frac{1}{2}(1-k/R)$ and $\frac{1}{2}(1+k/R)$ so that the point will have the

[14] For the case that the molecules of the gas can only collide against fixed centers or against other molecules which have a *given* velocity distribution.

[15] For the description of the probability problem and for a more complete analysis see E. Schrödinger and F. Kohlrausch, Physik. Zeits. **27**, 306 (1926).

tendency to go to the position $k=0$. The difference equation now becomes:

$$P(m, s)=\frac{R+m+1}{2R}P(m+1, s-1)+\frac{R-m+1}{2R}P(m-1, s-1), \quad (17)$$

which has again to be solved with the initial condition (14). We did not succeed in finding the solution; it is possible, however, to calculate average values. For instance it is easy to show from (17):

$$\langle m(s)\rangle_{\text{Av}}=\sum_m mP(m, s)=\left(1-\frac{1}{R}\right)\langle m(s-1)\rangle_{\text{Av}}.$$

Since for $s=0$, $\langle m\rangle_{\text{Av}}=n$, one finds:

$$\langle m(s)\rangle_{\text{Av}}=n\left(1-\frac{1}{R}\right)^s, \quad (18a)$$

which shows how the average position of the point goes to zero. In the same way one gets:

$$\langle m^2(s)\rangle_{\text{Av}}=n^2\left(1-\frac{2}{R}\right)^s+\frac{1}{2}R\left[1-\left(1-\frac{2}{R}\right)^s\right]. \quad (18b)$$

$\langle m^2\rangle_{\text{Av}}$ will therefore go to $R/2$ for $s\rightarrow\infty$. This is in accord with the first probability distribution (which is also the stationary distribution, see Eq. (4)) for which one finds easily:

$$W_1(n)=\frac{(2R)!}{(R+n)!(R-n)!}\left(\frac{1}{2}\right)^{2R}.$$

One can verify that $W_1(n)$ fulfills the equation:

$$W_1(n)=\sum_k W_1(k)Q(k, n)$$

which is a special case of Eq. (3c).

6. THE GAUSSIAN RANDOM PROCESS; METHOD OF RICE

(a) Assumptions

The Gaussian random process is characterized by the fact that all the basic distribution functions (1) are Gaussian distributions, and one could take this fact as the defining property of the process. However, since as we shall see, the spectrum essentially determines everything, it is more natural to start (following Rice) with the Fourier development of the Gaussian random function $y(t)$.

Consider again the stationary random function $y(t)$ over a long time T, and suppose that $y(t)$ is repeated periodically with the period T. One can then develop $y(t)$ in a Fourier series:

$$y(t)=\sum_{k=1}^{\infty}(a_k\cos 2\pi f_k t+b_k\sin 2\pi f_k t), \quad (19)$$

where $f_k=k/T$. There is no constant term, since we will assume that the average value of y is zero. The coefficients a_k and b_l are random variables, and we will *assume*, that they are all independent of each other and Gaussianly distributed with average values zero, so that one has for the probability that the a_k and b_l are in certain ranges da_k, db_l the expression:

$$W(a_1a_2\cdots; b_1b_2\cdots)=\prod_k\frac{1}{\sigma_k\sqrt{2\pi}}\exp\left[-(a_k^2+b_k^2)/2\sigma_k^2\right], \quad (20)$$

where $\sigma_k^2=\langle a_k^2\rangle_{\text{Av}}=\langle b_k^2\rangle_{\text{Av}}=G(f_k)/T$. $G(f)$ is again the spectral density (cf. Eqs. (8) and (9)), since:

$$\langle y^2(t)\rangle_{\text{Av}}=\sum_k(\langle a_k^2\rangle_{\text{Av}}\cos^2 2\pi f_k t+\langle b_k^2\rangle_{\text{Av}}\sin^2 2\pi f_k t)=\frac{1}{T}\sum_k G(f_k)\cong\int_0^{\infty}dfG(f). \quad (21)$$

With these assumptions one is now able to derive all possible distribution functions for the Gaussian random function $y(t)$. As a preparation one needs:

(b) A Theorem about Gaussian Distributions

Suppose the variables $x_1x_2\cdots x_n$ are distributed according to:

$$W(x_1\cdots x_n)=\prod_{i=1}^{n}\frac{1}{\sigma_i(2\pi)^{\frac{1}{2}}}\exp\left\{-\frac{x_i^2}{2\sigma_i^2}\right\}.$$

Let $y_1y_2\cdots y_s$ $(s\leqslant n)$ be s linear combinations of the x_i:

$$y_k=\sum_{i=1}^{n}a_{ki}x_i, \quad k=1, 2\cdots s$$

where the a_{ki} are constants. One can prove easily[16] that the y_k will be distributed according

[16] See Note I of the appendix.

to the s-dimensional Gaussian distribution:

$$P(y_1\cdots y_s)=\frac{1}{(2\pi)^{s/2}B^{\frac{1}{2}}}\times\exp\left[-\frac{1}{2B}\sum_{k,l=1}^{s}B_{kl}y_ky_l\right]. \quad (22)$$

Here B_{kl} is the cofactor of the element b_{kl} in the matrix b_{kl}, where:

$$b_{kl}=\sum_{i=1}^{n}a_{ki}a_{li}\sigma_i^2=\langle y_ky_l\rangle_{\text{Av}}, \quad (23)$$

and B is the determinant of the matrix b_{kl}. As a special case take, for instance, $s=2$. One then gets the two-dimensional Gaussian distribution, which according to (22) can be written in the form:

$$P(y_1y_2)=\frac{1}{2\pi\sigma\tau(1-\rho^2)^{\frac{1}{2}}}\times\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{y_1^2}{\sigma^2}+\frac{y_2^2}{\tau^2}-\frac{2\rho}{\sigma\tau}y_1y_2\right\}\right], \quad (24)$$

where $\sigma^2=\langle y_1^2\rangle_{\text{Av}}$, $\tau^2=\langle y_2^2\rangle_{\text{Av}}$ and $\langle y_1y_2\rangle_{\text{Av}}=\sigma\tau\rho$; ρ is the correlation coefficient.

(c) Distribution Functions for $y(t)$

Using the general theorem mentioned above one can now derive from (19) any kind of distribution function referring to $y(t)$. The method is best explained by considering a few examples.

1. The distribution of y at fixed t. According to (19) y is for a given t a linear function of the basic variables a_k, b_l. We know, therefore, that the probability distribution for y will be Gaussian with a mean square value given by (21). The time t has disappeared, and this is as it should be since $W_1(y)$ must be independent of t, because the process is stationary. In the same way one can compute the distribution of the velocity $\dot{y}(t)$. This is also a linear function of the a_k, b_l, so the distribution will again be Gaussian with the mean square value:

$$\langle\dot{y}(t)^2\rangle_{\text{Av}}=4\pi^2\int_0^\infty f^2G(f)df. \quad (25)$$

2. The distribution of y and $\dot{y}$ at a fixed t. This will now be a two-dimensional Gaussian distribution, which is, however, especially simple since y and $\dot{y}$ at a given t are not correlated. One gets namely:

$$\langle y(t)\dot{y}(t)\rangle_{\text{Av}}=\sum_k 2\pi f_k\sin 2\pi f_kt\times\cos 2\pi f_kt(\langle -a_k^2\rangle_{\text{Av}}+\langle b_k^2\rangle_{\text{Av}})=0. \quad (26)$$

3. The joint distribution of $y(t_1)$ and $y(t_2)$. This will again be a two-dimensional Gaussian distribution. One gets:

$$\langle y(t_1)y(t_2)\rangle_{\text{Av}}=\frac{1}{T}\sum_k G(f_k)\cos 2\pi f_k(t_1-t_2)\cong\int_0^\infty G(f)\cos 2\pi f\tau df. \quad (27)$$

The correlation depends therefore only on $\tau=t_2-t_1$, as it should be since the process is stationary.

4. In this way one can go on. One can consider for instance the third distribution function $W_3(y_1t_1, y_2t_2, y_3t_3)$, which will be a three-dimensional Gaussian distribution depending only on t_2-t_1 and t_3-t_2. One can find the four-dimensional Gaussian distribution $W(y_1\dot{y}_1, y_2\dot{y}_2, \tau)$. One can bring in the acceleration $\ddot{y}(t)$ and its distribution functions, and so on.

7. FURTHER REMARKS ON THE METHOD OF RICE

1. We have seen that for a Gaussian random process all the distribution functions can be determined when one only knows the spectrum or the correlation function. In the actual problems of the Brownian motion this spectrum can be found from the so-called *Langevin equations* or in the electrical analogy from the *circuit equations with thermal noise sources*. For examples see Sections 9 and 10. It should be emphasized, however, that for many applications it is an advantage that one can leave open the question of the actual shape of the spectrum.

2. A disadvantage of not knowing the spectrum is that it does not allow a classification of the Gaussian processes, so that one does not know which distribution function describes the process completely. The different type of processes correspond to different type of spectra. For instance one can show[17] that a one-dimensional Gaussian process will be Markoffian only when the correlation function $\rho(t)=\exp(-\beta t)$ so

[17] This was first pointed out by J. L. Doob, Ann. Math. **43**, 351 (1942).

that according to (11) the spectrum must be $\sim 1/(\beta^2+\omega^2)$. To prove this one determines with the method of Rice the distribution functions $W_3(y_1y_2y_3)$ and $W_2(y_1y_2)$. One, therefore, knows also the conditional probability $P_3(y_1y_2|y_3)$. For a Markoff process this must be identical with $P_2(y_2|y_3)$ and one finds that this can only be the case when the correlation function $\rho(\tau)$ fulfills the functional equation:

$$\rho(t_3-t_1)=\rho(t_2-t_1)\rho(t_3-t_2). \tag{28}$$

The only non-singular solution of this equation is

$$\rho(\tau)=\exp(-\beta\tau).$$

3. This theorem can be generalized to n-dimensional Gaussian processes. The dependent variable y now denotes an n-dimensional vector with components $x_1, x_2, \cdots x_n$. Instead of a correlation *function* one gets a correlation *matrix:*

$$\mathbf{R}(\tau)=\begin{pmatrix}\langle x_1(t)x_1(t+\tau)\rangle_{\text{Av}} & \cdots & \langle x_1(t)x_n(t+\tau)\rangle_{\text{Av}}\\ \vdots & & \\ \langle x_n(t)x_1(t+\tau)\rangle_{\text{Av}} & \cdots & \langle x_n(t)x_n(t+\tau)\rangle_{\text{Av}}\end{pmatrix}. \tag{29}$$

From the stationarity of the process, follows:

$$\mathbf{R}(\tau)=\tilde{\mathbf{R}}(-\tau), \tag{30}$$

where $\tilde{\mathbf{R}}$ denotes the transposed matrix. Following the same reasoning as for the one-dimensional process one can show[18] that the n-dimensional Gaussian process is Markoffian only when $R(\tau)$ fulfills the matrix functional equation:

$$\mathbf{R}(t_3-t_1)=\mathbf{R}(t_2-t_1)\mathbf{R}(t_3-t_2), \tag{31}$$

where we have still assumed that $\mathbf{R}(0)=\mathbf{I}$, the unit matrix.[19] The only non-singular solution is

$$\mathbf{R}(\tau)=e^{\mathbf{Q}\tau}, \tag{32}$$

for $\tau>0$, $\mathbf{Q}$ is a constant matrix, which is in general not symmetric,[20] so that its eigenvalues may be complex. There is now of course a greater variety of possible spectra, corresponding to the different forms $\mathbf{Q}$ may have.

4. One should point out, that sometimes the distribution functions which one derives with Rice's method will have no meaning since some of the integrals over the spectrum are divergent. For instance when $G(f)\sim 1/(\alpha^2+f^2)$, the distribution functions in which the velocity $\dot{y}(t)$ appear have no meaning since (see Eq. (25)) $\langle\dot{y}^2\rangle_{\text{Av}}$ will not exist. In this case one may call the process *non-differentiable.* The degree of differentiability will be characteristic for the process and will depend again on the spectrum.

8. THE GAUSSIAN RANDOM PROCESS; METHOD OF FOKKER-PLANCK

(a) Basic Ideas

It is best to start with the discrete random series (cf. Section 5). Suppose that the basic transition probability $P(n|m,\tau)$ or $Q(n,m)$ has the property that in the time τ n can only change by zero or by ± 1. This was, for instance, the case in the examples (a) and (b) discussed in Section 5. Consider now for this case the limit in which n and the time $s\tau$ become continuous. The Smoluchowski equation will then become a partial differential equation of the first order in the time coordinate and of the second order in the space coordinate. For instance, in example (a), the Smoluchowski equation becomes Eq. (15) which may be written:

$$P(m,s)-P(m,s-1)=\tfrac{1}{2}[P(m+1,s-1)-2P(m,s-1)+P(m-1,s-1)].$$

In the limit that $s\tau=t$ and $m\Delta=x$ become continuous variables, this clearly goes over into:

$$(\partial P/\partial t)=D(\partial^2P/\partial x^2), \tag{33}$$

when $D=\text{Lim}\,\Delta^2/2\tau$. One gets, therefore, the well-known heat conduction or diffusion equation. In the same way one shows that in example (b) one gets from (17) in the limit the equation:

$$\frac{\partial P}{\partial t}=\beta\frac{\partial}{\partial x}(xP)+D\frac{\partial^2P}{\partial x^2}, \tag{34}$$

where $\beta=\text{Lim}\,\Delta/\tau R$.

In this limit the problem of finding the probability distribution $P(x_0|x,t)$ becomes then the problem of finding the *fundamental solution* of

[18] This result seems to be contained in a recent paper by J. L. Doob, Ann. Am. Stat. **15**, 229 (1944). See also Note II of the appendix, where we give some details of a more direct proof which we owe to Dr. M. Kac.

[19] This is no loss in generality, since it can always be achieved by using the proper linear combinations of the components of the n-dimensional vector y. In the physical language this means that we have used such coordinates that the energy is a sum of squares.

[20] For $\tau<0$ $\mathbf{R}(\tau)=\exp(-\tilde{\mathbf{Q}}\tau)$ in accordance with (30).

the partial differential equation of the diffusion (or parabolic) type into which the Smoluchowski equation has degenerated. We mean by this the solution which for $t=0$ becomes the Dirac singular function $\delta(x-x_0)$. This corresponds to the condition (14) in the discrete case and expresses again the fact that for $t=0$ one is certain that $x=x_0$. For (33) this solution is given by:

$$P(x_0|x,t)=\frac{1}{(4\pi Dt)^{\frac{1}{2}}}\exp\left[-(x-x_0)^2/4Dt\right]. \quad (35)$$

It is easy to show that this is the limit into which the solution (16) of the discrete case goes over. For Eq. (34) the fundamental solution is given by:[21]

$$P(x_0|x,t)=\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\exp\left[-(x-\bar{x})^2/2\sigma^2\right], \quad (36)$$

where $\langle x\rangle_{\text{Av}}=x_0\exp(-\beta t)$ and $\sigma^2=\langle(x-\bar{x})^2\rangle_{\text{Av}}=(D/\beta)[1-\exp(-2\beta t)]$. It is clear that these average values follow also in the limit from (18a) and (18b).

One should point out that one gets in the limit a diffusion equation *only* when $P(n|m,\tau)$ is such that in the time τ n can only change by zero or ± 1, or, less precisely, when in small times the space coordinate can only change with small amounts. *In the general case*, the Smoluchowski equation will become in the limit an integro-differential equation which is of the same type as the Boltzmann equation in the kinetic theory of gases.

(b) Assumptions

In the continuous case we will start from the Smoluchowski equation in the form:[22]

$$P(x|y,t+\Delta t)=\int dzP(x|z,t)P(z|y,\Delta t). \quad (37)$$

This *assumes*, therefore, that the process is a Markoff process. The moments of the change in the space coordinate in a small time Δt are given by:

$$a_n(z,\Delta t)=\int dy(y-z)^nP(z|y,\Delta t),$$

and we shall *assume* that for $\Delta t\to 0$, only the first and second moments become proportional to Δt so that the limits

$$\begin{aligned} A(z)&=\text{Lim}\,\frac{1}{\Delta t}a_1(z,\Delta t),\\ B(z)&=\text{Lim}\,\frac{1}{\Delta t}a_2(z,\Delta t), \end{aligned} \quad (38)$$

exist. This assumption expresses the fact that for these processes in small times the space coordinate can only change with small amounts. In the actual problems of the Brownian motion this assumption can be proved and the average values $A(z)$ and $B(z)$ can be calculated from the Langevin equations or in the electrical analogy from the circuit equations with thermal noise sources.[23] Just as in the method of Rice, these equations are, therefore, the real basis for the theory of the Brownian motion.

(c) Derivation of the Fokker-Planck Equation

Consider the integral

$$\int dyR(y)\frac{\partial P(x|y,t)}{\partial t},$$

where $R(y)$ is an arbitrary function, which goes to zero for $y\to\pm\infty$ sufficiently fast. Replacing the differential quotient by the limit of the difference quotient and using the Smoluchowski equation in the form (37) one can write:

$$\begin{aligned}\int dyR(y)\frac{\partial P}{\partial t}&=\text{Lim}\,\frac{1}{\Delta t}\int dyR(y)\\ &\times[P(x|y,t+\Delta t)-P(x|y,t)]\\ &=\text{Lim}\,\frac{1}{\Delta t}\Big[\int dyR(y)\int dzP(x|z,t)P(z|y,\Delta t)\\ &\qquad-\int dzR(z)P(x|z,t)\Big].\end{aligned}$$

In the double integral, interchange the order of integration and develop $R(y)$ in a Taylor series

[21] See, for instance, I, Section II. Equation (36) is also a special case of the solution derived in Note IV of the appendix.

[22] For simplicity we consider the process to be one dimensional, since the generalization to the n-dimensional case is obvious. We follow the notation and the exposition of Kolmogoroff, Math. Ann. **104**, 415 (1931).

[23] For examples see Sections 9 and 10.

in $(z-y)$. Because of (38) one can stop at the term with $(z-y)^2$ and one gets:

$$\int dyR(y)\frac{\partial P}{\partial t}=\int dzP(x\,|\,z,t)\times[R'(z)A(z)+\tfrac{1}{2}R''(z)B(z)].$$

Integrating partially and writing y for z one obtains:

$$\int dyR(y)\left[\frac{\partial P}{\partial t}+\frac{\partial}{\partial y}(AP)-\frac{1}{2}\frac{\partial^2}{\partial y^2}(BP)\right]=0.$$

Since this must hold for any function $R(y)$, the expression in the square brackets must be zero, which gives the general Fokker-Planck equation:

$$\frac{\partial P}{\partial t}=-\frac{\partial}{\partial y}[A(y)P]+\frac{1}{2}\frac{\partial^2}{\partial y^2}[B(y)P], \quad (39)$$

of which, of course, (33) and (34) are special cases. For an n-dimensional process one gets analogously:

$$\frac{\partial P}{\partial t}=-\sum_i\frac{\partial}{\partial y_i}[A_i(\mathbf{y})P]+\frac{1}{2}\sum_{k,l}\frac{\partial^2}{\partial y_k\partial y_l}[B_{kl}(\mathbf{y})P], \quad (39a)$$

where again the A_i and the B_{kl} are the first and second moments defined analogously as (38).

9. THE BROWNIAN MOTION OF A FREE PARTICLE[24]

(a) Assumptions; The Langevin Equation

For a free particle (mass m, velocity v) the equation of motion will be:

$$m(dv/dt)+fv=K(t), \quad (40a)$$

where f is the friction coefficient, and $K(t)$ is the fluctuating force, of which the average value is zero and which has a very sharp correlation function and therefore, a practically white spectrum. The spectral density of $K(t)$ is $4fkT$ where k is the Boltzmann constant and T the temperature of the surrounding medium. The analogous electrical problem is of course the (L, R) circuit, and the circuit equation is:

$$L(di/dt)+Ri=E(t), \quad (40b)$$

where $E(t)$ is a purely random fluctuating e.m.f. (the thermal noise source), which has a spectral density $4RkT$. We will combine these cases by writing (40a) and (40b) in the form:

$$(dy/dt)+\beta y=F(t) \quad (40c)$$

and by taking $4D$ as the spectral density of the purely random $F(t)$. This means that we *assume*:[25]

$$\langle F(t)\rangle_{\text{Av}}=0, \quad (41a)$$

$$\langle F(t_1)F(t_2)\rangle_{\text{Av}}=2D\delta(t_1-t_2). \quad (41b)$$

This is, however, not enough; besides being purely random, we must *assume* that $F(t)$ is *Gaussian.* This can be expressed in different ways. Either one can postulate the Gaussian distribution of the Fourier coefficients (see Eq. (20) where now $\sigma_k^2=\text{const.}=4D/T$) or one can assume the two properties:[26]

$$\langle F(t_1)F(t_2)\cdots F(t_{2n+1})\rangle_{\text{Av}}=0, \quad (42a)$$

$$\langle F(t_1)F(t_2)\cdots F(t_{2n})\rangle_{\text{Av}}=\sum_{\text{all pairs}}\langle F(t_i)F(t_j)\rangle_{\text{Av}}\cdot\langle F(t_k)F(t_l)\rangle_{\text{Av}}\cdots \quad (42b)$$

where the sum has to be taken over all the different ways in which one can divide the $2n$ time points $t_1\cdots t_{2n}$ into n pairs. It is easy to show the equivalence of these two definitions (see Note III of the appendix).

(b) The Spectrum of $y(t)$

Since $F(t)$ is Gaussian it is clear that $y(t)$ will also be a Gaussian random process with a spectrum:

$$G_y(f)=\frac{4D}{\beta^2+(2\pi f)^2}. \quad (43)$$

[24] Compare I, Sections II and III. The first complete derivation of the distribution functions obtained in Sections 9 and 10 was given by L. S. Ornstein and W. R. van Wyk, Physica 1, 235 (1934). The derivation from the Fokker-Planck or Kramers equation was found independently by Ming Chen Wang, Dissertation, Ann Arbor (1942) and by Chandrasekhar, reference 3.

[25] The second equation follows from (11b) since:

$$G_F(f)=4D=2\int_{-\infty}^{+\infty}d\tau\cos 2\pi f\tau\langle F(t)F(t+\tau)\rangle_{\text{Av}}.$$

[26] This is the starting point of some of the work of N. Wiener on the theory of the Brownian motion. The physical justification of the assumptions (41) and (42) comes from the Maxwell-Boltzmann distribution law, which in the theory of the Brownian motion is always *postulated* and not derived. Compare also I.

This corresponds to a correlation function $\rho(t) = \exp(-\beta t)$ and the second probability distribution is, therefore, the two-dimensional Gaussian distribution:

$$W_2(y_1y_2t) = \frac{\beta}{2\pi D(1-\rho^2)^{\frac{1}{2}}} \times \exp\left[-\frac{\beta}{2D(1-\rho^2)}\{y_1^2+y_2^2-2\rho y_1y_2\}\right], \quad (44)$$

since:

$$\langle y^2\rangle_{\text{Av}} = \int_0^\infty G_y(f)df = D/\beta. \quad (45)$$

According to the theorem of Doob $y(t)$ will be a Markoff process, so that $W_2(y_1y_2t)$ gives the complete description of the process.

(c) The Fokker-Planck Equation

The average values $A(y)$ and $B(y)$ can now be computed by means of (41) and one finds:

$$A(y) = -\beta y, \quad B(y) = 2D. \quad (46)$$

The proof is simple; integrating (40c) over a short time Δt one gets:

$$\Delta y = -\beta y\Delta t + \int_t^{t+\Delta t} d\xi F(\xi).$$

Therefore:

$$A(y) = \lim_{\Delta t\to 0}\frac{\langle\Delta y\rangle_{\text{Av}}}{\Delta t} = -\beta y,$$

since $\langle F\rangle_{\text{Av}} = 0$. Further:

$$\langle\Delta y^2\rangle_{\text{Av}} = \beta^2y^2\Delta t^2 + \int\int_t^{t+\Delta t} d\xi d\eta\langle F(\xi)F(\eta)\rangle_{\text{Av}},$$

and from (41b) one shows easily that the double integral is $2D\Delta t$, so that:

$$B(y) = \lim_{\Delta t\to 0}\frac{\langle\Delta y^2\rangle_{\text{Av}}}{\Delta t} = 2D.$$

In the same way it follows from (42) that all the higher moments of Δy go to zero in the limit $\Delta t\to 0$, so that all the assumptions of §8, *b* are fulfilled.[27] With the values given by (46) the Fokker-Planck equation becomes:

$$\frac{\partial P}{\partial t} = \beta\frac{\partial}{\partial y}(yP) + D\frac{\partial^2 P}{\partial y^2}. \quad (47)$$

This is identical with (34), so that the fundamental solution $P(y_0|y, t)$ is given by (36) (for the proof, cf. Note IV). For $t\to\infty$ one gets:

$$W_1(y) = \lim_{t\to\infty} P(y_0|y, t) = \left(\frac{\beta}{2\pi D}\right)^{\frac{1}{2}} \exp\left(-\frac{\beta y^2}{2D}\right),$$

in accordance with (45). For the second probability distribution:

$$W_2(y_1y_2t) = W_1(y_1)P(y_1|y_2, t)$$

one gets again Eq. (44). That $y(t)$ is a Markoff process has now, of course, been assumed from the beginning.[28]

10. THE BROWNIAN MOTION OF A SIMPLE HARMONIC OSCILLATOR

(a) The Langevin Equation

Suppose now that instead of (40c) we have the second-order differential equation:

$$\frac{d^2y}{dt^2} + \beta\frac{dy}{dt} + \omega_0^2 y = F(t). \quad (48)$$

This describes clearly the Brownian motion of a simple harmonic oscillator or the thermal noise in a (R, L, C) circuit. For the $F(t)$ we assume again the basic properties (41) and (42).

(b) The Spectrum and the Correlation Matrix

Since $F(t)$ is Gaussian, it is clear that $y(t)$ will also be a Gaussian random process with a spectrum:

$$G_y(f) = \frac{4D}{|-(2\pi f)^2 + 2\pi i\beta f + \omega_0^2|^2}, \quad (49)$$

[27] It should be emphasized perhaps again that from the physical point of view these assumptions (and, therefore, also (41) and (42)) are necessarily only approximations. The basic equation is always Boltzmann's integral equation. Only when in each collision the velocity of the particle can change very little, then the Boltzmann equation can be approximated by the diffusion Eq. (47). It is very instructive to compare the derivation above with the derivation of (47) in the well-known Rayleigh model (Scientific Papers Vol. 3, p. 473) for the Brownian motion of a heavy particle.

[28] We shall not discuss the distribution function for the displacement of the particle, since it follows from the velocity distribution (see, for instance, Doob, reference 17) and since it is also a special case of the distribution functions derived in Section 10.

from which follows according to (11a)

$$\langle y(t)y(t+\tau)\rangle_{\text{Av}} = \frac{2D}{\pi}\int_0^\infty \frac{\cos\omega\tau}{(\omega_0^2-\omega^2)^2+\beta^2\omega^2}d\omega = \frac{D}{\beta\omega_0^2}e^{-\beta\tau/2}\left(\cos\omega_1\tau+\frac{\beta}{2\omega_1}\sin\omega_1\tau\right), \quad (50a)$$

where $\omega_1^2=\omega_0^2-(\beta^2/4)$; the formula will always be written for the *underdamped* case; for the aperiodic case let $\omega_1\to 0$ and for the overdamped case put $\omega_1=i\omega'$.

Of course, $y(t)$ is *not* more a Markoff process. However, from the physical situation and also from the general theorem of Doob one must expect that $y(t)$ is the "projection" of the two-dimensional Gaussian Markoff process $[y(t), p(t)]$ where $p(t)=dy/dt$. The correlation function (50a) must be extended to the correlation matrix:

$$\begin{pmatrix} \langle y(t)y(t+\tau)\rangle_{\text{Av}} & \langle y(t)p(t+\tau)\rangle_{\text{Av}} \\ \langle p(t)y(t+\tau)\rangle_{\text{Av}} & \langle p(t)p(t+\tau)\rangle_{\text{Av}} \end{pmatrix},$$

and one finds easily that:

$$\langle p(t)y(t+\tau)\rangle_{\text{Av}} = 2\pi\int_0^\infty dfG_y(f)f\sin 2\pi f\tau, = \frac{2D}{\pi}\int_0^\infty \frac{\omega\sin\omega\tau}{(\omega_0^2-\omega^2)^2+\beta^2\omega^2}d\omega, = +\frac{D}{\beta\omega_1}e^{-\beta\tau/2}\sin\omega_1\tau, \quad (50b)$$

and:

$$\langle p(t)p(t+\tau)\rangle_{\text{Av}} = 4\pi^2\int_0^\infty dfG_y(f)f^2\cos 2\pi f\tau, = \frac{2D}{\pi}\int_0^\infty \frac{\omega^2\cos\omega\tau}{(\omega_0^2-\omega^2)^2+\beta^2\omega^2}d\omega, = \frac{D}{\beta}e^{-\beta\tau/2}\left(\cos\omega_1\tau-\frac{\beta}{2\omega_1}\sin\omega_1\tau\right). \quad (50c)$$

The complete description of the process will now be given by $W_2(y_1p_1, y_2p_2, t)$, which is a four-dimensional Gaussian distribution. We will not write it down since for the discussion it is easier to consider the conditional probability $P_2(y_1p_1|y_2p_2, t)$. Consider first, however,

(c) The Fokker-Planck Equation

Replacing the Langevin Eq. (48) by the simultaneous equations:

$$dy/dt=p \quad\text{and}\quad dp/dt+(\beta p+\omega_0^2y)=F(t), \quad (48a)$$

one finds easily for the average values occurring in the two-dimensional Fokker-Planck equation (cf. Eq. (39a)):

$$A_1=\text{Lim}\frac{\langle\Delta y\rangle_{\text{Av}}}{\Delta t}=p;$$

$$A_2=\text{Lim}\frac{\langle\Delta p\rangle_{\text{Av}}}{\Delta t}=-(\beta p+\omega_0^2y);$$

$$B_{11}=\text{Lim}\frac{\langle\Delta y^2\rangle_{\text{Av}}}{\Delta t}=0;$$

$$B_{12}=\text{Lim}\frac{\langle\Delta y\Delta p\rangle_{\text{Av}}}{\Delta t}=0;$$

$$B_{22}=\text{Lim}\frac{\langle\Delta p^2\rangle_{\text{Av}}}{\Delta t}=2D,$$

so that one gets:[29]

$$\frac{\partial P}{\partial t}=-p\frac{\partial P}{\partial y}+\frac{\partial}{\partial p}[(\beta p+\omega_0^2y)P]+D\frac{\partial^2 P}{\partial p^2}, \quad (51)$$

which has to be solved with the initial condition:

$$P(yp, 0)=\delta(y-y_0)\delta(p-p_0).$$

For the solution it is simpler to work with the independent variables:

$$z_1=p+ay; \quad z_2=p+by, \quad (52)$$

where:

$$a=\tfrac{1}{2}\beta+i\omega_1 \quad\text{and}\quad b=\tfrac{1}{2}\beta-i\omega_1.$$

Equation (51) is then transformed into the more symmetrical form:

$$\frac{\partial P}{\partial t}=b\frac{\partial}{\partial z_1}(z_1P)+a\frac{\partial}{\partial z_2}(z_2P)+D\left(\frac{\partial}{\partial z_1}+\frac{\partial}{\partial z_2}\right)^2P, \quad (51a)$$

and this is a special case of the equation solved in Note IV of the appendix. One finds that the

[29] This is a special case of the *equation of Kramers*, Physica **7**, 284 (1940). Kramers takes a general force $K(y)$ instead of the harmonic force $-\omega_0^2y$. His derivation is essentially the same as the one given above. One should emphasize perhaps, that with a general force $K(y)$ the process $[y(t), p(t)]$ is *still* Markoffian, but it is *not more* Gaussian, since the basic Langevin equation is then not *linear* anymore.

fundamental solution of (51a) is a two-dimensional Gaussian distribution in z_1 and z_2 with the average values:

$$\langle z_1\rangle_{\text{Av}}=z_{10}e^{-bt},\quad \langle z_2\rangle_{\text{Av}}=z_{20}e^{-at}, \tag{53}$$

and the variances:

$$\begin{aligned}\langle(z_1-\bar{z}_1)^2\rangle_{\text{Av}}&=\frac{D}{b}(1-e^{-2bt}),\\ \langle(z_2-\bar{z}_2)^2\rangle_{\text{Av}}&=\frac{D}{a}(1-e^{-2at}),\\ \langle(z_1-\bar{z}_1)(z_2-\bar{z}_2)\rangle_{\text{Av}}&=\frac{2D}{a+b}(1-e^{-(a+b)t}),\end{aligned} \tag{54}$$

where z_{10}, z_{20} are the initial values of z_1, z_2 corresponding to y_0 and p_0.

(d) Discussion

Since z_1 and z_2 are connected with p and y by the linear relations (52) it is clear that $P(p_0y_0|py,t)$ will also be a two-dimensional Gaussian distribution in p and y. One obtains from (53) and (54) for the average values and the variances the expressions:

$$\left.\begin{aligned}\langle p\rangle_{\text{Av}}&=\frac{p_0}{\omega_1}e^{-\frac{1}{2}\beta t}\left(\omega_1\cos\omega_1t-\frac{\beta}{2}\sin\omega_1t\right)\\&\qquad-\frac{\omega_0^2}{\omega_1}y_0e^{-\frac{1}{2}\beta t}\sin\omega_1t,\\ \langle y\rangle_{\text{Av}}&=\frac{p_0}{\omega_1}e^{-\frac{1}{2}\beta t}\sin\omega_1t\\&\qquad+\frac{y_0}{\omega_1}e^{-\frac{1}{2}\beta t}\left(\omega_1\cos\omega_1t+\frac{\beta}{2}\sin\omega_1t\right),\\ \langle(p-\bar{p})^2\rangle_{\text{Av}}&=\frac{D}{\beta}\Big[1-\frac{1}{\omega_1^2}e^{-\beta t}\\&\qquad\times(\omega_1^2+\tfrac{1}{2}\beta^2\sin^2\omega_1t\\&\qquad-\beta\omega_1\sin\omega_1t\cos\omega_1t)\Big],\\ \langle\omega_0^2(y-\bar{y})^2\rangle_{\text{Av}}&=\frac{D}{\beta}\Big[1-\frac{1}{\omega_1^2}e^{-\beta t}\\&\qquad\times(\omega_1^2+\tfrac{1}{2}\beta^2\sin^2\omega_1t\\&\qquad+\beta\omega_1\sin\omega_1t\cos\omega_1t)\Big],\\ \langle\omega_0(p-\bar{p})(y-\bar{y})\rangle_{\text{Av}}&=\frac{D\omega_0}{\omega_1^2}e^{-\beta t}\sin^2\omega_1t.\end{aligned}\right\} \tag{55}$$

One has, of course,

$$\bar{p}=\frac{d\bar{y}}{dt},$$

$$\frac{d^2\bar{y}}{dt^2}+\beta\frac{d\bar{y}}{dt}+\omega_0^2\bar{y}=0.$$

The center of the Gaussian distribution moves, therefore, like the harmonic oscillator starting from the initial values p_0, y_0. In the (p, y) plane one gets (in the periodic case) for the orbit the well known spirals. For small t:

$$\begin{aligned}\langle(p-\bar{p})^2\rangle_{\text{Av}}&\cong 2Dt,\\ \langle(y-\bar{y})^2\rangle_{\text{Av}}&\cong\tfrac{2}{3}Dt^3,\\ \langle(p-\bar{p})(y-\bar{y})\rangle_{\text{Av}}&\cong Dt^2.\end{aligned}$$

One sees, therefore, that the initial two-dimensional δ function $\delta(p-p_0)\delta(y-y_0)$ will become first a narrow ellipse elongated in the p direction. The distribution ellipse will then turn and broaden out till at the time $t=\pi/\omega_1$ it has become again a circle. This process will repeat on a larger and larger scale with the period π/ω_1 (see Fig. 1). The center of the distribution will come nearer and nearer to the origin and finally of course the $P(p_0y_0|py,t)$ will become the Maxwell-Boltzmann distribution.

11. THE BROWNIAN MOTION OF A SYSTEM OF COUPLED HARMONIC OSCILLATORS

The generalization to more complicated systems does not involve anything new, so that we will only give an outline of the main results. We will use for a change the electrical language and we will consider, therefore, an arbitrary linear network of n meshes. The circuit equations are then:[30]

$$\sum_{j=1}^{n}\left(L_{ij}\frac{d^2y_j}{dt^2}+R_{ij}\frac{dy_j}{dt}+G_{ij}Y_j\right)=\sum_{j=1}^{n}E_{ij};\qquad i=1,2\cdots n. \tag{56}$$

The E_{ii} is the fluctuating thermal e.m.f. in that part of the resistance of the ith mesh which is

[30] For the precise definition of the matrices L_{ij}, R_{ij}, G_{ij} see, for instance, E. A. Guillemin, *Communication Networks*, Vol. I, Chap. IV. All these matrices are *symmetrical*. Note, however, that R_{ij} $(i\neq j)$ does *not* need to be positive. It is negative, when in the resistance common to the ith and jth mesh the positive directions chosen for the currents are opposite to each other. The y_j are the mesh-charges.

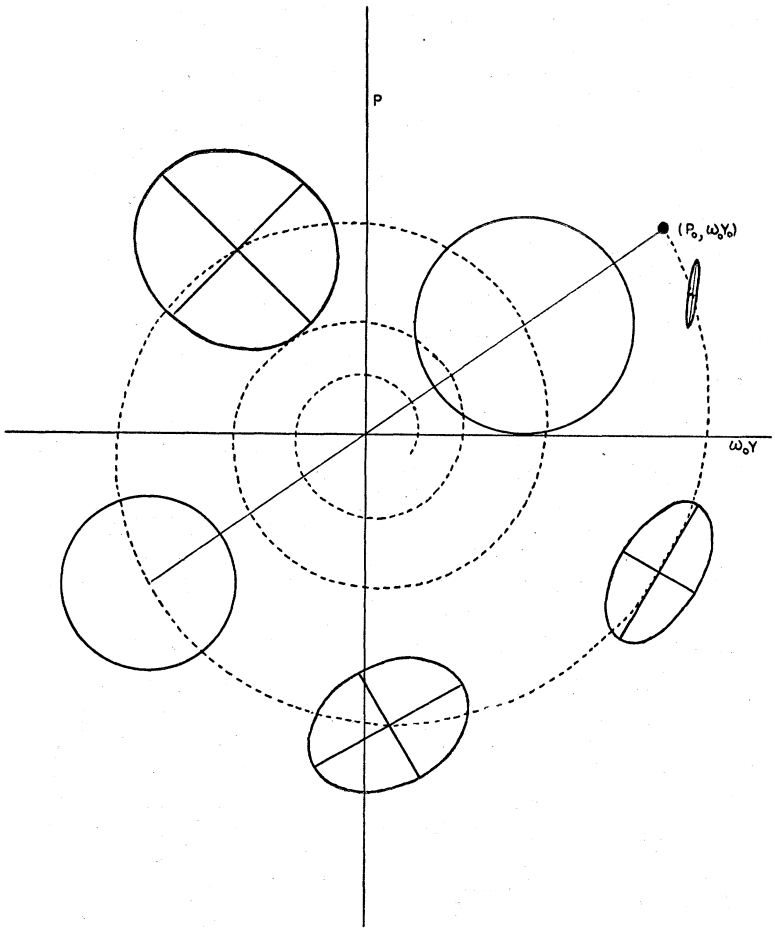

FIG. 1.

not in common with any other meshes. The E_{ij} $(i \neq j)$ is the fluctuating thermal e.m.f. in the resistance R_{ij}; $E_{ij}=E_{ji}$ if in R_{ij} the positive directions chosen for the currents are in the same direction; if they are opposite to each other then $E_{ij}=-E_{ji}$. The E_{ij} are again supposed to be Gaussian random processes with a constant spectrum. We assume especially:

$$\left.\begin{aligned} \langle E_{ij}\rangle_{\text{Av}}&=0\,; \\ \langle E_{ij}(t_1)E_{ji}(t_2)\rangle_{\text{Av}}&=2R_{ij}kT\delta(t_2-t_1)\,; \\ \langle E_{ij}(t_1)E_{ij}(t_2)\rangle_{\text{Av}}&=2\,|R_{ij}|\,kT\delta(t_2-t_1)\,; \\ \langle E_{ij}E_{kl}\rangle_{\text{Av}}&=0. \end{aligned}\right\} \tag{57}$$

In addition one needs, of course, assumptions analogous to (42).

Since the Eqs. (56) are *linear* it is clear that each of the $y_j(t)$ will be a Gaussian random process. From the physical situation one must expect in addition that the $2n$ variables $[y_1(t)\cdots y_n(t),\ dy_1/dt\cdots dy_n/dt]$ will form a $2n$-dimensional Markoff process, governed by the Fokker-Planck equation:

$$\frac{\partial P}{\partial t}=-\sum_{i=1}^{2n}\frac{\partial}{\partial x_i}(A_iP)+\frac{1}{2}\sum_{i,\,j=1}^{2n}\frac{\partial^2}{\partial x_i\partial x_j}(D_{ij}P), \tag{58}$$

where $x_1\cdots x_{2n}$ denote the variables $y_1\cdots y_n$, $dy_1/dt\cdots dy_n/dt$. From (56) and (57) one finds further that:

$$A_i=\textstyle\sum_k a_{ik}x_k, \tag{59}$$

where the $2n$ by $2n$ matrix $\mathbf{a}$ is of the form:

$$\mathbf{a}=\begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{L}^{-1}\mathbf{G} & -\mathbf{L}^{-1}\mathbf{R} \end{pmatrix}. \tag{60}$$

Finally one gets for the $2n$ by $2n$ matrix $\mathbf{D}$ the constant matrix:

$$\mathbf{D}=\begin{pmatrix}0 & 0\\ 0 & 2kT\mathbf{L}^{-1}\mathbf{R}\mathbf{L}^{-1}\end{pmatrix}. \tag{61}$$

To find the fundamental solution of (58), it is best to make first a linear transformation of the x_i (analogous to (52) in §10):

$$z_i=\sum_j c_{ij}x_j \tag{62}$$

where the matrix $\mathbf{c}$ is the matrix which diagonaluzes $\mathbf{a}$, so that:

$$\sum_j c_{ij}a_{jl}=\lambda_i c_{il}. \tag{63}$$

The eigenvalues λ_i are of course the $2n$ roots of the equation:

$$Det(a_{ij}-\lambda\delta_{ij})=0. \tag{64}$$

One can easily show that this equation is identical with

$$Det(L_{ij}\lambda^2+R_{ij}\lambda+G_{ij})=0, \tag{64a}$$

and it is well known that for a linear passive network the roots of this equation must have a negative real part, and the same must hold therefore for the λ_i. The Fokker-Planck equation now becomes:

$$\frac{\partial P}{\partial t}=-\sum_i \lambda_i\frac{\partial}{\partial z_i}(z_iP)+\frac{1}{2}\sum_{i,j}\sigma_{ij}\frac{\partial^2 P}{\partial z_i\partial z_j}, \tag{65}$$

where:

$$\boldsymbol{\sigma}=\mathbf{c}\mathbf{D}\tilde{\mathbf{c}}. \tag{66}$$

The fundamental solution of (65) is derived in Note IV of the appendix. One gets a $2n$-dimensional Gaussian distribution with the average values:

$$\langle z_i\rangle_{\text{Av}}=z_{i0}\exp(\lambda_i t) \tag{67}$$

and the variances:

$$\mu_{ij}=\langle(z_i-\bar{z}_i)(z_j-\bar{z}_j)\rangle_{\text{Av}}$$
$$=-\frac{\sigma_{ij}}{\lambda_i+\lambda_j}[1-\exp(\lambda_i+\lambda_j)t], \tag{68}$$

where z_{i0} are the initial values of the z_i. Transforming back to the original variables x_i one gets of course again a $2n$-dimensional Gaussian distribution. Combining the variables x_i in a column matrix $\mathbf{x}$, one can write for the average values the matrix equation:

$$\mathbf{x}=e^{\mathbf{a}t}\mathbf{x}_0. \tag{69}$$

This follows from (67) which can be written as

$$\bar{\mathbf{z}}=e^{\boldsymbol{\Lambda}t}\mathbf{z}_0,$$

where the diagonal matrix $\Lambda_{ij}=\lambda_i\delta_{ij}$. Now $\mathbf{z}=\mathbf{c}\mathbf{x}$ so that $\mathbf{x}=\mathbf{c}^{-1}\mathbf{z}$ and:

$$\bar{\mathbf{x}}=\mathbf{c}^{-1}\bar{\mathbf{z}}=\mathbf{c}^{-1}e^{\boldsymbol{\Lambda}t}\mathbf{z}_0=\sum_{n=0}^{\infty}\frac{t^n}{n!}\mathbf{c}^{-1}\boldsymbol{\Lambda}^n\mathbf{z}_0. \tag{70}$$

Equation (63) can be written in the matrix form:

$$\mathbf{c}\mathbf{a}\mathbf{c}^{-1}=\boldsymbol{\Lambda}, \tag{71}$$

from which follows that $\boldsymbol{\Lambda}^n=\mathbf{c}\mathbf{a}^n\mathbf{c}^{-1}$. Substituting this in (70) one obtains (69).

For the matrix of the variances:

$$b_{ij}=\langle(x_i-\bar{x}_i)(x_j-\bar{x}_j)\rangle_{\text{Av}},$$

one obtains from (68):

$$\mathbf{b}=\mathbf{c}^{-1}\boldsymbol{\mu}\tilde{\mathbf{c}}^{-1}. \tag{72}$$

Since the real parts of the λ_i are negative it is clear that for $t\to\infty$ all the average values $\bar{x}_i$ go to zero. The distribution function $P(x_{i0}|x_i,t)$ must become in the limit $t\to\infty$ the Maxwell-Boltzmann distribution law, which means that:

$$\lim_{t\to\infty}\mathbf{b}^{-1}=\begin{pmatrix}\frac{1}{kT}\mathbf{G} & 0\\ 0 & \frac{1}{kT}\mathbf{L}\end{pmatrix}. \tag{73}$$

To show this from (72), one starts from Eq. (68) which in the limit $t\to\infty$ can be written in the form:

$$\boldsymbol{\Lambda}\boldsymbol{\mu}+\boldsymbol{\mu}\boldsymbol{\Lambda}=-\boldsymbol{\sigma}.$$

Substituting $\boldsymbol{\mu}=\mathbf{c}\mathbf{b}\tilde{\mathbf{c}}$ and using (66) and (71) one finds that $\mathbf{b}$ has to fulfill (always in the limit $t\to\infty$) the equation:

$$\mathbf{a}\mathbf{b}+\mathbf{b}\tilde{\mathbf{a}}=-\mathbf{D}. \tag{74}$$

This determines, of course, the matrix $\mathbf{b}$ uniquely. From (72) follows also that $\mathbf{b}$ is symmetric so that we can put:

$$\mathbf{b}=\begin{pmatrix}\mathbf{X}_1 & \mathbf{X}_2\\ \mathbf{X}_2 & \mathbf{X}_3\end{pmatrix},$$

where $\mathbf{X}_1$ and $\mathbf{X}_3$ are symmetric n by n matrices.

From (74) follows then:

$$\mathbf{X}_2=0,\quad \mathbf{L}\mathbf{X}_3=\mathbf{G}\mathbf{X}_1,\quad \mathbf{R}\mathbf{X}_1\mathbf{G}+\mathbf{G}\mathbf{X}_1\mathbf{R}=2kT\mathbf{R}.$$

By inspection one sees that these equations are fulfilled by:

$$\mathbf{X}_1=kT\mathbf{G}^{-1};\quad \mathbf{X}_3=kT\mathbf{L}^{-1};\quad \mathbf{X}_2=0, \tag{75}$$

and since the solution of (74) is unique this must also be the only solution. Equation (75) is, of course, equivalent with (73).

The same distribution function follows, of course, from the method of Rice. One has to start then from the correlation matrix, for which one finds from the circuit Eqs. (56):

$$\langle y_r(t)y_s(t+\tau)\rangle_{\text{Av}}=\frac{kT}{\pi}\int_{-\infty}^{+\infty}d\omega e^{i\omega\tau}\times[\mathbf{Z}^{-1}(i\omega)\mathbf{R}\mathbf{Z}^{-1}(-i\omega)]_{rs}, \tag{76}$$

where $\mathbf{Z}(p)$ is the matrix:

$$\mathbf{Z}(p)=\mathbf{L}p^2+\mathbf{R}p+\mathbf{G}.$$

In order to see how in the method of Rice the Maxwell-Boltzmann distribution is reached, it is sufficient to prove the theorem of the equipartition of energy:

$$\left\langle y_r\frac{\partial V}{\partial y_r}\right\rangle_{\text{Av}}=kT, \tag{77}$$

where V is the potential energy

$$V=\tfrac{1}{2}\sum_{r,s}G_{rs}y_ry_s.$$

From (76) one obtains:

$$\left\langle y_r\frac{\partial V}{\partial y_r}\right\rangle_{\text{Av}}=\sum_s G_{rs}\langle y_ry_s\rangle_{\text{Av}}=\frac{kT}{\pi}\int_{-\infty}^{+\infty}d\omega[\mathbf{G}\mathbf{Z}^{-1}(i\omega)\mathbf{R}\mathbf{Z}^{-1}(-i\omega)]_{rr}. \tag{78}$$

To calculate the integral,[31] observe that:

$$\mathbf{Z}(i\omega)-\mathbf{Z}(-i\omega)=2i\omega\mathbf{R}.$$

Eliminating $\mathbf{R}$ in the integral (78) one finds:

$$\left\langle y_r\frac{\partial V}{\partial y_r}\right\rangle_{\text{Av}}=-\frac{kT}{\pi i}\int_{-\infty}^{+\infty}\frac{d\omega}{\omega}[\mathbf{G}\mathbf{Z}^{-1}(i\omega)]_{rr}, \tag{79}$$

where one has to take the principal value of the integral. Since the determinant of $\mathbf{Z}(i\omega)$ has no zeros in the lower half of the complex ω-plane, it can be easily seen that the integral in (79) is $-\pi i$ times the residue of the integrand at $\omega=0$. Since $\mathbf{Z}^{-1}(0)=\mathbf{G}^{-1}$ it is clear that the residue is unity, so that one obtains the equipartition theorem (77).

12. SOME UNSOLVED PROBLEMS

Since we may have created the false impression that with the derivation of the fundamental probability distribution all problems in the theory of the Brownian motion have been solved, it may be useful to list a number of unsolved or partially solved problems.

(a) The Approach to the Barometric Distribution

It should be emphasized that *only* for harmonic forces one gets the simple theory of the Gaussian random process. For a constant force the problem becomes already much more complicated. For instance for a gravitational field (directed towards the negative x axis) the Kramers equation becomes:

$$\frac{\partial P}{\partial t}=-p\frac{\partial P}{\partial x}+g\frac{\partial P}{\partial p}+\frac{\partial}{\partial p}\left[\beta pP+D\frac{\partial P}{\partial p}\right]. \tag{80}$$

The trouble is now, that one needs a *reflecting* boundary, say at $x=0$ in order to prevent the particles from disappearing towards $x=-\infty$. We feel sure that this means the condition:

$$P(0,p,t)=P(0,-p,t). \tag{81}$$

We have been unable to find the solution of (80) (for $x\geq 0$, $-\infty<p<+\infty$) which fulfills the condition (81) and which for $t=0$ becomes $\delta(x-x_0)\delta(p-p_0)$. One can show[32] that for the stationary case the Kramers equation and the boundary condition (81) determines *uniquely* the barometric distribution:

$$P(x,p,\infty)=C\exp\left(-\frac{\beta}{2D}p^2-\frac{g\beta}{D}x\right).$$

(b) First Passage Time Problems

One may ask for the probability that the random variable y starting from the value $y=y_0$

[31] For this proof we are indebted to Dr. J. Schwinger.

[32] A proof was communicated to us by Mr. M. Dresden.

reaches the value $y=y_1$ for the first time in a time between t and $t+dt$. In the usual theory of the Brownian motion (which is based on the ordinary diffusion Eq. (33))[33] such first passage time problems have been considered and solved by Smoluchowski and others.[34] The method of Smoluchowski can also be used for a one-dimensional Gaussian Markoff process $y(t)$. One can show, for instance, that the probability distribution $W(y_0, t)dt$ of the first passage times to reach $y=0$ starting from y_0 is given by:

$$W(y_0, t)dt = y_0\left(\frac{2\beta}{\pi D}\right)^{\frac{1}{2}} \exp\left(-\frac{\beta y_0^2}{2D}z^2\right)dz, \quad (82)$$

where

$$z = e^{-\beta t}(1-e^{-2\beta t})^{-\frac{1}{2}}.$$

However, the generalization to processes $y(t)$ which are "projections" of Markoff processes seems quite complicated to us. For instance, we have *not* succeeded in finding, even for a free particle, the distribution of the one-sided first passage times if the damping is small, so that one has to use the exact Kramers equation. To extend the method of Smoluchowski it is necessary to introduce the idea of an *absorbing* boundary say at $x=0$. We feel sure that this means the condition:

$$P(0, p, t) = 0 \text{ for all } p > 0 \quad (83)$$

when the particle has started from $x_0 > 0$. However, to find solutions of the Kramers equation with the boundary condition (83) seems even more difficult than to find solutions with the boundary condition (81).

(c) The Recurrence Time Problem

One may ask for the probability that the random variable y starting from $y=a$ returns to the value $y=a$ for the first time in a time between t and $t+dt$.[35] Or in other words, what is the distribution of the time intervals between successive a-values of the random function $y(t)$. A formal solution of this problem has been given by Rice (II, Section 3.4, Eq. (11), p. 64). However, even for the simple case of the harmonic oscillator the actual discussion of the solution has not been achieved.

(d) The Distribution of the Average Value

One may ask for the probability distributions of the random variable:

$$z(t) = \int K(s-t)y(s)ds, \quad (84)$$

where $K(x)$ is a given function. A special case of (84) is the *average* of the random process $y(t)$ over a time interval of length T. Of course, when $y(t)$ is a Gaussian random process then also $z(t)$ will be Gaussian, and the problem is trivial. But for other types of processes $y(t)$ the problem is quite difficult. Rice (II, Section 3.9) has discussed some of the average values of $z(t)$. Recently M. Kac and A. Siegert have succeeded in finding the complete solution for the case that $y(t)$ is the sum of the squares of two independent Gaussian processes, which have the same probability distributions.

(e) The Distribution of the Absolute Maximum of a Random Function $y(t)$ in a Given Time Interval T

For Markoff processes one can show that this problem is equivalent with the first passage time problem, so that it is of the same degree of difficulty.

APPENDIX

Note I. Proof of Eq. (22)

One uses the integral representation of the Dirac δ function:

$$\delta(x-x') = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dt \exp\left[it(x-x')\right],$$

[33] Kramers, reference 29, has shown that one obtains this equation (or the corresponding one if there is an outside force $K(y)$) from the Kramers Eq. (51) (with $-\omega_0^2 y$ replaced by $K(y)$) in the limit of strong damping.

[34] Cf. for instance, R. Furth, reference 6, and for the corresponding problems for random series M. Kac, Ann. Math. Stat. **16**, 62 (1945).

[35] For the definition of the mean recurrence time and the mean persistence time see the basic paper of Smoluchowski, Wien. Sitz. Ber. **124**, 339 (1915). Smoluchowski restricts himself mainly to discrete random series. Already in this case the question of the distribution of the recurrence times seems to be quite difficult. For the simple discrete random walk problem (example (*a*) of Section 5) the first passage time and the recurrence time problem can be solved exactly. But already for example (*b*) of Section 5 we have failed to find the solution.

which allows one to write:

$$P(y_1\cdots y_s)=\frac{1}{(2\pi)^{n/2+s}\sigma_1\sigma_2\cdots\sigma_n}$$
$$\times\int_{-\infty}^{+\infty}\cdots\int dx_1\cdots dx_n \exp\left[-\tfrac{1}{2}\sum_1^n (x_i^2/\sigma_i^2)\right]$$
$$\cdot\int_{-\infty}^{+\infty}\cdots\int dt_1\cdots dt_s \prod_{k=1}^{s}\exp\left[it_k(y_k-\sum_1^n a_{ki}x_i)\right].$$

Interchanging the integrations over the x_i with those over the t_k one can easily carry out the integrations over the x_i and one gets:

$$P(y_1\cdots y_s)=\frac{1}{(2\pi)^s}\int_{-\infty}^{+\infty}\cdots\int dt_1\cdots dt_s$$
$$\times\exp\left[i\sum_1^s y_k t_k-\tfrac{1}{2}\sum_{k,l=1}^{s} b_{kl}t_k t_l\right], \quad (85)$$

where the b_{kl} are given by (23). One sees, therefore, that $\exp(-\frac{1}{2}\sum b_{kl}t_k t_l)$ is the characteristic function of the probability distribution $P(y_1\cdots y_s)$. It is a standard result[36] that from (85) follows that P is an s-dimensional Gaussian distribution whose matrix is the inverse of the matrix b_{kl}, and this is just what is expressed by Eq. (22).

Note II. Proof of the General Theorem of Doob

Denote the $3n$ components of the vectors $\mathbf{y}(t_1)$, $\mathbf{y}(t_2)$, $\mathbf{y}(t_3)$ by $z_1z_2\cdots z_{3n}$. According to the general theorem (22) one then can write the third probability distribution in the form:

$$W_3(\mathbf{y}(t_1),\mathbf{y}(t_2),\mathbf{y}(t_3))=C_3\exp\left(-\tfrac{1}{2}\sum_{j,k=1}^{3n}\alpha_{jk}z_jz_k\right),$$

where the matrix α_{jk} is the inverse of the $3n$ by $3n$ matrix:

$$\mathbf{M}=\begin{pmatrix}\mathbf{R}(0) & \mathbf{R}(t_2-t_1) & \mathbf{R}(t_3-t_1)\\ \mathbf{R}(t_1-t_2) & \mathbf{R}(0) & \mathbf{R}(t_3-t_2)\\ \mathbf{R}(t_1-t_3) & \mathbf{R}(t_2-t_3) & \mathbf{R}(0)\end{pmatrix}. \quad (86)$$

[36] Comp. for instance H. Cramér, Random Variables and Probability Distributions (Cambridge Tracts No. 36, 1937), p. 110.

Note that $\mathbf{M}$ is symmetric because of (30). Analogously one has for the second probability distribution

$$W_2(\mathbf{y}(t_1),\mathbf{y}(t_2))=C_2\exp\left(-\tfrac{1}{2}\sum_{j,k=1}^{2n}\beta_{jk}z_jz_k\right),$$

where the $2n$ by $2n$ matrix β_{jk} is the inverse of

$$\begin{pmatrix}\mathbf{R}(0) & \mathbf{R}(t_2-t_1)\\ \mathbf{R}(t_1-t_2) & \mathbf{R}(0)\end{pmatrix}.$$

For a Markoff process $P(\mathbf{y}(t_1)\mathbf{y}(t_2)\,|\,\mathbf{y}(t_3))=W_3/W_2$ should be independent of $\mathbf{y}(t_1)$. This leads to the conditions:

$$\alpha_{jk}=0,\ \text{when } j=1,\ 2\cdots n \text{ and } k=2n+1,\ \cdots 3n, \quad (87)$$

$$\alpha_{jk}=\beta_{jk},\ \text{when } j=1,\ 2\cdots n \text{ and } k=1,\ 2\cdots 2n. \quad (88)$$

In order to calculate the inverse of $\mathbf{M}$ we resort to the following formal trick. We treat $\mathbf{M}$ not as a numerical matrix but as a matrix whose elements are matrices. The rules of multiplication are the same but in taking inverses one must be careful because of possible non-commutativity. We want then a 3 by 3 matrix $\mathbf{X}$, whose elements are n by n matrices such that:

$$\mathbf{M}\times\begin{pmatrix}\mathbf{X}_{11} & \mathbf{X}_{12} & \mathbf{X}_{13}\\ \mathbf{X}_{21} & \mathbf{X}_{22} & \mathbf{X}_{23}\\ \mathbf{X}_{31} & \mathbf{X}_{32} & \mathbf{X}_{33}\end{pmatrix}=\begin{pmatrix}\mathbf{I} & \mathbf{O} & \mathbf{O}\\ \mathbf{O} & \mathbf{I} & \mathbf{O}\\ \mathbf{O} & \mathbf{O} & \mathbf{I}\end{pmatrix}.$$

Using (86) this leads for instance to the three matrix equations:

$$\mathbf{R}(0)\mathbf{X}_{13}+\mathbf{R}(t_2-t_1)\mathbf{X}_{23}+\mathbf{R}(t_3-t_1)\mathbf{X}_{33}=\mathbf{O},$$
$$\mathbf{R}(t_1-t_2)\mathbf{X}_{13}+\mathbf{R}(0)\mathbf{X}_{23}+\mathbf{R}(t_3-t_2)\mathbf{X}_{33}=\mathbf{O},$$
$$\mathbf{R}(t_1-t_3)\mathbf{X}_{13}+\mathbf{R}(t_2-t_3)\mathbf{X}_{23}+\mathbf{R}(0)\mathbf{X}_{33}=\mathbf{I}.$$

Condition (87) says that $\mathbf{X}_{13}=0$; assuming further that $\mathbf{R}(0)=\mathbf{I}$, which as mentioned in footnote 19 is no loss in generality, these equations become:

$$\mathbf{R}(t_2-t_1)\mathbf{X}_{23}+\mathbf{R}(t_3-t_1)\mathbf{X}_{33}=\mathbf{0}, \quad (89a)$$

$$\mathbf{X}_{23}+\mathbf{R}(t_3-t_2)\mathbf{X}_{33}=\mathbf{0}, \quad (89b)$$

$$\mathbf{R}(t_2-t_3)\mathbf{X}_{23}+\mathbf{X}_{33}=\mathbf{I}. \quad (89c)$$

$\mathbf{R}$ is a non-singular matrix. One can eliminate $\mathbf{X}_{23}$ from the last two equations and one gets:

$$\{\mathbf{R}^{-1}(t_2-t_3)-\mathbf{R}(t_3-t_2)\}\mathbf{X}_{33}=\mathbf{R}^{-1}(t_2-t_3).$$

The matrix in the curly brackets cannot be singular since $DetR^{-1}(t_2-t_3)\neq 0$, and therefore:

$$\mathbf{X}_{33}=\{\mathbf{R}^{-1}(t_2-t_3)-\mathbf{R}(t_3-t_2)\}^{-1}\mathbf{R}^{-1}(t_2-t_3),$$

and:

$$\mathbf{X}_{23}=-\mathbf{R}(t_3-t_2)\{\mathbf{R}^{-1}(t_2-t_3)-\mathbf{R}(t_3-t_2)\}^{-1}\mathbf{R}^{-1}(t_2-t_3).$$

Substituting in (89a) one gets as the condition on $\mathbf{R}$:

$$\mathbf{R}(t_3-t_1)=\mathbf{R}(t_2-t_1)\mathbf{R}(t_3-t_2), \tag{90}$$

which is Eq. (31). One must calculate also the other elements of $\mathbf{X}$ and in the same way the matrix β_{kl} in order to show that the condition (88) is now automatically satisfied so that (90) is also the *only* condition to be imposed on the correlation matrix $\mathbf{R}$.

Note III. Proof that the Properties (42) Imply that $F(t)$ is a Gaussian Process

From (42a) follows immediately that the average values of all the *odd* powers of the Fourier coefficient:

$$a_k=\frac{2}{T}\int_0^T dt\cos 2\pi f_k t F(t)$$

are zero. One gets further:

$$\begin{aligned}\langle a_k^2\rangle_{\text{Av}}&=\frac{4}{T^2}\int\int_0^T dt_1dt_2\cos 2\pi f_k t_1 \times\cos 2\pi f_k t_2\langle F(t_1)F(t_2)\rangle_{\text{Av}}\\ &=\frac{8D}{T^2}\int\int_0^T dt_1dt_2\cos 2\pi f_k t_1 \times\cos 2\pi f_k t_2\delta(t_1-t_2)\\ &=\frac{8D}{T^2}\int_0^T dt_1\cos^2 2\pi f_k t_1=\frac{4D}{T}.\end{aligned} \tag{91}$$

From (42b) follows then that the average values of the *even* powers of a_k are given by:

$$\langle a_k^{2n}\rangle_{\text{Av}}=1\cdot 3\cdot 5\cdots(2n-1)\langle a_k^2\rangle_{\text{Av}}^n \tag{92}$$

since the number of ways in which we can divide the $2n$ time points t_1, $t_2\cdots t_{2n}$ into n pairs is $1\cdot 3\cdot 5\cdots(2n-1)$. Equation (92) is characteristic for the Gaussian distribution. One can show[37] that still more explicitly by calculating the characteristic function of the distribution function $W(a_k)$, which is given by:

$$\begin{aligned}\langle\exp(i\xi a_k)\rangle_{\text{Av}}&=\sum_{n=0}^{\infty}\frac{(i\xi)^m}{m!}\langle a_k^m\rangle_{\text{Av}}\\ &=\sum_{n=0}^{\infty}(-1)^n\frac{1}{n!}\left(\frac{2D}{T}\right)^n\xi^{2n}=\exp\left(-\frac{2D\xi^2}{T}\right)\end{aligned}$$

using (91) and (92). Therefore:

$$\begin{aligned}W(a_k)&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}d\xi\exp(-i\xi a_k)\\ &\cdot\exp\left(-\frac{2D\xi^2}{T}\right)=\left(\frac{T}{8\pi D}\right)^{\frac{1}{2}}\exp\left(-\frac{a_k^2T}{8D}\right).\end{aligned}$$

The same distribution one finds for:

$$b_k=\frac{2}{T}\int_0^T dt\sin 2\pi f_k t F(t),$$

and it is also easy to show that the different a_k and b_l are independent of each other, so that the complete distribution function $W(a_1a_2\cdots;b_1b_2\cdots)$ will be given by (20) with $\sigma_k^2=4D/T$.

Note IV. The Fundamental Solution of Equation (65)

The problem is to find the solution of:

$$\frac{\partial P}{\partial t}=-\sum_i\lambda_i\frac{\partial}{\partial y_i}(y_iP)+\tfrac{1}{2}\sum_{ij}\sigma_{ij}\frac{\partial^2P}{\partial y_i\partial y_j}, \tag{93}$$

which for $t=0$ becomes:

$$P=\delta(y_1-y_{10})\delta(y_2-y_{20})\cdots\delta(y_n-y_{n0}). \tag{94}$$

Introduce instead of P its Fourier transform:

$$\begin{aligned}f(\xi_1\cdots\xi_n,t)&=\int_{-\infty}^{+\infty}\cdots\int dy_1\cdots dy_nP(y_1\cdots y_n,t)\\ &\times\exp[-i\textstyle\sum_j\xi_jy_j].\end{aligned} \tag{95}$$

From (93) follows that f has to fulfill the linear, first-order partial differential equation:

$$\frac{\partial f}{\partial t}-\sum_i\lambda_i\xi_i\frac{\partial f}{\partial\xi_i}=-\tfrac{1}{2}f\sum_{ij}\sigma_{ij}\xi_i\xi_j. \tag{96}$$

[37] This was pointed out to us by Dr. A. Siegert.

The subsidiary equations are:

$$\frac{dt}{1}=-\frac{d\xi_1}{\lambda_1\xi_1}=-\frac{d\xi_2}{\lambda_2\xi_2}$$
$$=\cdots=-\frac{d\xi_n}{\lambda_n\xi_n}=-\frac{df}{\frac{1}{2}f\sum\sigma_{ij}\xi_i\xi_j}.$$

These can easily be integrated; one finds that the general solution of (96) is given by:

$$f(\xi_1\cdots\xi_n,t)$$
$$=\psi(\xi_1\exp(\lambda_1 t),\xi_2\exp(\lambda_2 t)\cdots\xi_n\exp(\lambda_n t))$$
$$\times\exp\left[+\frac{1}{2}\sum_{ij}\sigma_{ij}\frac{\xi_i\xi_j}{\lambda_i+\lambda_j}\right],\quad(97)$$

where ψ is an arbitrary function. Now for $t=0$ one sees from (94) and (95) that:

$$f(\xi_1\cdots\xi_n,0)=\exp\left[-i\sum_j\xi_j y_{j0}\right].$$

Therefore the arbitrary function ψ must be:

$$\psi(\xi_1\cdots\xi_n,0)=\exp\left[-\frac{1}{2}\sum_{ij}\sigma_{ij}\frac{\xi_i\xi_j}{\lambda_i+\lambda_j}-i\sum_j\xi_j y_{j0}\right],$$

and one obtains for f:

$$f=\exp\left[-i\sum_j\xi_j y_{j0}\exp(\lambda_j t)\right.$$
$$\left.+\frac{1}{2}\sum_{ij}\sigma_{ij}\frac{\xi_i\xi_j}{\lambda_i+\lambda_j}\{1-\exp[+(\lambda_i+\lambda_j)t]\}\right].\quad(98)$$

This is the Fourier transform of an n-dimensional Gaussian distribution with the average values

$$\langle y_i\rangle_{\text{Av}}=y_{i0}\exp(\lambda_i t),$$

and the variances:

$$\langle(y_i-\bar{y}_i)(y_j-\bar{y}_j)\rangle_{\text{Av}}=-\frac{\sigma_{ij}}{\lambda_i+\lambda_j}[1-\exp(\lambda_i+\lambda_j)t].$$