

# The multivariate Langevin and Fokker–Planck equations

Daniel T. Gillespie

Citation: *Am. J. Phys.* **64**, 1246 (1996); doi: 10.1119/1.18387

View online: <http://dx.doi.org/10.1119/1.18387>

View Table of Contents: <http://ajp.aapt.org/resource/1/AJPIAS/v64/i10>

Published by the [American Association of Physics Teachers](#)

## Related Articles

Simulation of a Brownian particle in an optical trap

*Am. J. Phys.* **81**, 224 (2013)

Teaching the principles of statistical dynamics

*Am. J. Phys.* **74**, 123 (2006)

Game theory and physics

*Am. J. Phys.* **73**, 405 (2005)

Reappraising Einstein's 1909 application of fluctuation theory to Planckian radiation

*Am. J. Phys.* **72**, 1059 (2004)

Landau theory of domain walls for one-dimensional asymmetric potentials

*Am. J. Phys.* **71**, 1005 (2003)

## Additional information on *Am. J. Phys.*

Journal Homepage: <http://ajp.aapt.org/>

Journal Information: [http://ajp.aapt.org/about/about\\_the\\_journal](http://ajp.aapt.org/about/about_the_journal)

Top downloads: [http://ajp.aapt.org/most\\_downloaded](http://ajp.aapt.org/most_downloaded)

Information for Authors: <http://ajp.dickinson.edu/Contributors/contGenInfo.html>

## ADVERTISEMENT

# ***SHARPEN*** YOUR COMPUTATIONAL SKILLS.



**computing**  
in SCIENCE & ENGINEERING

Scientific Computing with GPUs



Subscribe for  
**\$49** | year

monotonically increasing function of its initial altitude. It is also found that if the initial altitude of the object is greater than a critical value, the object always strikes the ground with a speed that is higher than its terminal speed by a finite value, which increases asymptotically with the initial altitude. This is in contrast to the case of a falling object in a uniform atmosphere. The asymptotic ratio of the impact speed to the terminal speed increases with the terminal speed of the object.

Numerical solution of the equations of motion<sup>6</sup> for position and velocity as a function of time of an object falling through a nonuniform atmosphere is a trivial task and, therefore, we did not present them here.

- <sup>1</sup>J. W. Kittinger, "The long, lonely leap," *Nat. Geograph.* **118**, 854–873 (1960).
- <sup>2</sup>J. W. Kittinger, *The Long, Lonely Leap* (Dutton, New York, 1961).
- <sup>3</sup>C. Ryan, *The Pre-Astronauts, Manned Ballooning on the Threshold of Space* (Naval Institute, Annapolis, 1995).
- <sup>4</sup>G. R. Fowles and G. L. Cassiday, *Analytical Mechanics* (Saunders, New York, 1993), 5th ed., pp. 61–62.
- <sup>5</sup>J. B. Marion, *Classical Dynamics* (Academic, New York, 1970), 2nd ed., p. 53.
- <sup>6</sup>N. M. Shea, "Terminal speed and atmospheric density," *The Phys. Teacher* **31**, 176 (1993).
- <sup>7</sup>L. B. Loeb, *The Kinetic Theory of Gases* (Dover, New York, 1961), p. 394.
- <sup>8</sup>*U.S. Standard Atmosphere, 1962* (Government Printing Office, Washington, DC, December 1962).

## The multivariate Langevin and Fokker–Planck equations

Daniel T. Gillespie<sup>a)</sup>

Research and Technology Division, Naval Air Warfare Center, China Lake, California 93555

(Received 30 January 1996; accepted 4 June 1996)

A novel derivation of the Langevin equation that was recently presented in this journal for a univariate continuous Markov process is generalized here to the more widely applicable multivariate case. The companion multivariate forward and backward Fokker–Planck equations are also derived. The derivations require just a few modest assumptions, and are driven by a self-consistency condition and some established theorems of random variable theory and ordinary calculus. The constructive nature of the derivations shows *why* a multivariate continuous Markov process must evolve according to equations of the canonical Langevin and Fokker–Planck forms, and also sheds new light on some uniqueness issues. The need for self-consistency in the time-evolution equations of both Markovian and non-Markovian stochastic processes is emphasized, and it is pointed out that for a great many non-Markovian processes self-consistency can be ensured most easily through the multivariate Markov theory. © 1996 American Association of Physics Teachers.

### I. INTRODUCTION

A recent article in this journal<sup>1</sup> presented a derivation of the Langevin equation for a univariate (scalar) continuous Markov process. Here we generalize that derivation to the multivariate case in which the process has  $M \geq 1$  components, and we also derive the companion forward and backward Fokker–Planck equations. We shall presume here an acquaintance with certain parts of Ref. 1, specifically its Secs. II A–II C, so that we may avail ourselves of several important definitions and theorems introduced there; also, a familiarity with the comparatively simple derivation of the univariate Langevin equation given in the Appendix of Ref. 1 will afford a helpful perspective on our analysis here of the more complicated multivariate case.<sup>2</sup>

We begin with a quick review of the univariate results obtained in Ref. 1. If a function  $X$  of time  $t$  is continuous, memoryless, and stochastic—i.e., if  $X$  is a continuous Markov process—then its time evolution will be governed by an equation of the form

$$X(t+dt) = X(t) + A(X(t), t)dt + D^{1/2}(X(t), t)N(t)(dt)^{1/2}. \quad (1.1)$$

This is the (univariate) *standard form Langevin equation*. In it,  $dt$  is to be regarded as a real variable that is confined to the interval  $[0, \epsilon]$ , where  $\epsilon$  is an arbitrarily small positive number;  $A$  and  $D$  can be any two smooth functions, with  $D$  being non-negative; and  $N(t)$  is a normal random variable that has a mean 0 and variance 1, with  $N(t)$  and  $N(t')$  statistically independent if  $t \neq t'$ .

Equation (1.1) is essentially an "updating formula": Once the functions  $A$  and  $D$  have been specified, Eq. (1.1) tells us how to compute, from the value of the process at time  $t$ , its value at any infinitesimally later time  $t+dt$ . As was shown in Ref. 1, the functional form of this updating formula is a consequence of requiring  $X$  to be not only *continuous*, in the sense that  $X(t+dt) \rightarrow X(t)$  as  $dt \rightarrow 0$ , and *memoryless*, in the sense that the right side does not depend on the value of  $X$  at any time before  $t$ , but also *self-consistent*: It should make no difference (statistically and to first order in  $dt$ ) whether we compute the increment from  $t$  to  $t+dt$  by a *single* application of the updating formula, or by *successive* applications thereof to successive subintervals of  $[t, t+dt]$ . The Langevin equation can be used to derive time-evolution equations for the moments of  $X$ , and it can also be used to construct numerical simulations of  $X$ .

The appearance of the factor  $(dt)^{1/2}$  in the random term of

Eq. (1.1), which will seem quite bizarre if viewed from the restricted perspective of ordinary deterministic calculus,<sup>3</sup> reflects the fact that when  $D$  is not zero, the continuous function  $X$  is *not differentiable*. For heuristic reasons, though, it is often convenient to pretend that  $dX/dt$  exists. As was shown in Ref. 1, Eq. (1.1) can then formally be brought into the form

$$\frac{dX(t)}{dt} = A(X(t), t) + D^{1/2}(X(t), t)\Gamma(t). \quad (1.2)$$

Here, with  $N(m, \sigma^2)$  denoting the normal random variable with mean  $m$  and variance  $\sigma^2$ ,

$$\Gamma(t) \equiv \lim_{dt \rightarrow 0} N(0, 1/dt) \quad (1.3)$$

is called “Gaussian white noise.” Equation (1.2) is the (univariate) *white-noise form Langevin equation*. A careful examination of its derivation suggests, however, that it may be little more than a mnemonic for the standard form Langevin equation (1.1).

If  $P(x, t|x_0, t_0)dx$  gives the probability of finding  $X(t)$  between  $x$  and  $x+dx$ , given that  $X(t_0)=x_0$ , for any  $t_0 \leq t$ , then as was stated (but not proved) in Ref. 1, the function  $P$  satisfies the partial differential equation

$$\begin{aligned} \frac{\partial}{\partial t} P(x, t|x_0, t_0) = & -\frac{\partial}{\partial x} [A(x, t)P(x, t|x_0, t_0)] \\ & + \frac{1}{2} \frac{\partial^2}{\partial x^2} [D(x, t)P(x, t|x_0, t_0)]. \end{aligned} \quad (1.4)$$

This is called the (univariate) *forward Fokker–Planck equation*. It can be used to derive time-evolution equations for the moments of  $X$ , although the Langevin equation (1.1) is usually handier for that task. It can also be used to compute the stationary or “equilibrium” probability density function  $P(x, t \rightarrow \infty|x_0, t_0)$ , which for *some* functions  $A$  and  $D$  exists independently of  $t$ ,  $x_0$ , and  $t_0$ .

Not mentioned in Ref. 1 was the fact that the function  $P$  also satisfies the partial differential equation

$$\begin{aligned} -\frac{\partial}{\partial t_0} P(x, t|x_0, t_0) = & A(x_0, t_0) \frac{\partial}{\partial x_0} P(x, t|x_0, t_0) \\ & + \frac{1}{2} D(x_0, t_0) \frac{\partial^2}{\partial x_0^2} P(x, t|x_0, t_0). \end{aligned} \quad (1.5)$$

This equation, which is *not* a simple corollary of Eq. (1.4), is called the (univariate) *backward Fokker–Planck equation*. It turns out to be crucial for calculating the statistics of the time it takes the process to first exit a given interval  $[a, b]$  that contains the initial point  $x_0$ .

Our goal in this article will be to derive the forms of the Langevin and Fokker–Planck equations when the process  $X$  has any finite number  $M$  of components. Of course, those multivariate equations are “well known,”<sup>4,5</sup> having been deduced, in a slightly restricted form, at least as early as 1945 by Wang and Uhlenbeck.<sup>6</sup> But physics students often find the derivations of those equations to be intimidating and unconvincing, and there is even a tendency to suppose that those equations simply have to be “accepted” in the same way that we have to accept, say, the Schrödinger equation. We shall endeavor here to derive the multivariate Langevin and Fokker–Planck equations in a very careful way from just a

few modest assumptions, and thereby show *why* those equations have the forms that they do. In the course of our derivation, we shall illuminate the little-known fact that the correspondence between the Langevin and Fokker–Planck equations is not quite one-to-one.

Our work here will of course be quite mathematical, and a busy physicist may rightfully ask what is the payoff. It turns out that a *component* of a multivariate continuous Markov processes is often *not* by itself Markovian; hence, the multivariate generalization greatly enriches the theory by introducing processes that “have memory.” From a slightly different perspective, it is fair to say that an  $M$ -variate Langevin equation is the stochastic generalization of an  $M$ th order ordinary differential equation. Since most dynamical systems in physics find their deterministic descriptions in terms of ordinary differential equations of order greater than 1, a physicist simply cannot be content with the restriction to first order that is imposed by the univariate Langevin equation.

There is also an aesthetic payoff for going through all the mathematics: Once we have set down in clear terms what we should reasonably expect of a process that is “multivariate, continuous, and memoryless,” it is somehow very satisfying to watch how sheer mathematical necessity shapes the forms that the fundamental time-evolution equations must take.

The plan of this article is as follows. In Sec. II we define a multivariate continuous Markov process and derive the multivariate Langevin equation. That derivation is easily and quickly accomplished by invoking a certain theorem. The proof of that theorem, however, is not so easy and quick, and will occupy us in Sec. III. In Sec. IV we derive the companion forward and backward Fokker–Planck equations. In Sec. V we give examples of two simple bivariate continuous Markov processes. Finally, we conclude in Sec. VI by summarizing our results and pointing out some interesting implications for non-Markovian stochastic processes.

## II. THE MULTIVARIATE LANGEVIN EQUATION

For a *general*  $M$ -variate stochastic process

$$\mathbf{X}(t) \equiv [X_1(t), \dots, X_M(t)], \quad (2.1)$$

whether Markovian or not, each process component  $X_i$  is presumed to be a real random variable that depends on time  $t$ , a real sure variable. The fact that the independent variable  $t$  represents *time*, and not something else, is important. Time is something we perceive as being divided into a *past* and a *present*, which we can in principle know, and a *future*, which we can only speculate about. The overarching goal of stochastic process theory is to predict, as accurately as possible, future values of the process  $\mathbf{X}$  from a knowledge of its present and past values.

The defining feature of a *Markov* process is that any knowledge of past values of  $\mathbf{X}$ , beyond what is already implied by its present value, is of absolutely no help in predicting future values of  $\mathbf{X}$ . With that in mind, we now *define* an  *$M$ -variate continuous Markov process* to be any  $M$ -variate process  $\mathbf{X}$  that satisfies the following four conditions:

(i) The conditional increment in each component  $X_i$  from time  $t$  to any infinitesimally later time  $t+dt$ , namely, the random variable

$$\begin{aligned} \Xi_i(dt; \mathbf{x}, t) \equiv & X_i(t+dt) - X_i(t) \quad \text{given that } \mathbf{X}(t) = \mathbf{x} \\ & (i = 1, \dots, M), \end{aligned} \quad (2.2)$$

depends *only* on the values of  $dt$ ,  $\mathbf{x} \equiv (x_1, \dots, x_M)$  and  $t$ .<sup>7</sup>

(ii) The dependence of each random variable  $\Xi_i$  on its arguments  $dt$ ,  $\mathbf{x}$ , and  $t$  is “smooth.”

(iii) For all  $\mathbf{x}$  and  $t$ ,

$$\Xi_i(dt; \mathbf{x}, t) \rightarrow 0 \text{ as } dt \rightarrow 0. \quad (2.3)$$

(iv) Each random variable  $\Xi_i$  has a well-defined (finite) mean and variance.

Notice that condition (i) is what makes  $\mathbf{X}$  *memoryless* or *Markovian*, since it ensures that future values of  $\mathbf{X}$  depend on its past values *only* through its present value. And condition (iii) is what makes  $\mathbf{X}$  *continuous*, since it ensures, through definition (2.2), that  $X_i(t+dt) \rightarrow X_i(t)$  as  $dt \rightarrow 0$ . The other two conditions (ii) and (iv) are essentially “technical” requirements that force the random variables  $\Xi_i$  to be well behaved. The smoothness condition (ii) basically requires that all probability density functions for  $\Xi_i$ , in which  $dt$ ,  $\mathbf{x}$ , and  $t$  will naturally appear as independent parameters, shall be continuously differentiable with respect to those parameters. Condition (iv) is arguably unnecessary on the grounds that the infinitesimal nature of  $dt$  should ensure, through condition (iii), that sample values of  $\Xi_i(dt; \mathbf{x}, t)$  will be so close to 0 that the first and second moments of  $\Xi_i(dt; \mathbf{x}, t)$  will inevitably be finite; however, to guard against any surprises on that point, we call out condition (iv) explicitly.

With the foregoing definition of a multivariate continuous Markov process, we can derive the multivariate Langevin equation almost trivially by appealing to a certain theorem. But the proof of that theorem is not trivial, which is to say that this way of deriving the multivariate Langevin equation merely relegates all the detailed reasoning, all the hard work, to proving the theorem. We shall undertake that proof in Sec. III. For now, we simply want to state the theorem, clarify its assertions, and examine its implications. [The simpler univariate version of this theorem is given in Ref. 1 at Eq. (2.13).]

**Theorem:** The defining conditions (i)–(iv) of an  $M$ -variate continuous Markov process  $\mathbf{X}$  imply that the conditional increments  $\Xi_i$  defined in Eqs. (2.2) must have the analytical forms

$$\Xi_i(dt; \mathbf{x}, t) = A_i(\mathbf{x}, t)dt + \sum_{j=1}^M b_{ij}(\mathbf{x}, t)N_j(t)(dt)^{1/2} \quad (i=1, \dots, M). \quad (2.4)$$

Here, the  $M$  functions  $A_i$  and the  $M^2$  functions  $b_{ij}$  are all smooth in the variables  $\mathbf{x}$  and  $t$ , and  $N_1(t), \dots, N_M(t)$  are  $M$  statistically independent, temporally uncorrelated, unit normal random variables. Furthermore, any two different sets of the  $M^2$  functions  $b_{ij}$  that yield the *same*  $\frac{1}{2}M(M+1)$  functions

$$D_i(\mathbf{x}, t) \equiv \sum_{j=1}^M b_{ij}^2(\mathbf{x}, t) \quad (i=1, \dots, M), \quad (2.5a)$$

$$C_{ij}(\mathbf{x}, t) \equiv \sum_{k=1}^M b_{ik}(\mathbf{x}, t)b_{jk}(\mathbf{x}, t) \quad (i < j=1, \dots, M), \quad (2.5b)$$

will yield, through Eqs. (2.4), identical conditional increments  $\Xi_i(dt; \mathbf{x}, t)$ .

In the theorem, the assertion that  $N_j(t)$  is a “unit normal” random variable means that  $N_j(t)$  is normal with mean 0 and

variance 1:  $N_j(t) = N(0, 1)$ . The assertion that  $N_1(t), \dots, N_M(t)$  are “statistically independent” means that our ability to predict the value of any one of those random variables will not be altered by our learning the value of any of the others. And the assertion that  $N_j(t)$  is “temporally uncorrelated” means that  $N_j(t)$  and  $N_j(t')$  are likewise statistically independent whenever  $t \neq t'$ . These stipulated statistical independencies imply that the (ensemble) average  $\langle N_j(t)N_k(t') \rangle$  vanishes if *either*  $j \neq k$  or  $t \neq t'$ . Of course,  $\langle N_j(t)N_j(t) \rangle = \langle N_j^2(t) \rangle = 1$ , since  $N_j(t)$  has mean 0 and variance 1.

We shall comment on the assertions of the theorem regarding Eqs. (2.5) in a moment. But now let us take the final short step from the theorem to the multivariate Langevin equation: We substitute into the left side of Eq. (2.4) the definition (2.2) of  $\Xi_i$ , replace  $\mathbf{x}$  everywhere by  $\mathbf{X}(t)$  as required by that definition, and so obtain

$$X_i(t+dt) = X_i(t) + A_i(\mathbf{X}(t), t)dt + \sum_{j=1}^M b_{ij}(\mathbf{X}(t), t)N_j(t)(dt)^{1/2} \quad (i=1, \dots, M). \quad (2.6)$$

This is the *standard form multivariate Langevin* equation, the  $M$ -variate generalization of Eq. (1.1).

Equation (2.6) is evidently an “updating” formula for the process  $\mathbf{X}$ : Given specific forms for the  $M$  functions  $A_i$  and the  $M^2$  functions  $b_{ij}$ , Eq. (2.6) allows us to compute, from the values of all the components of the process at time  $t$ , the component values at any infinitesimally later time  $t+dt$ . It is obvious from the form of Eq. (2.6) that the prescribed updating procedure is *memoryless*, since no reference is made to values of the process at times earlier than  $t$ ; and also *continuous*, since  $X_i(t+dt)$  clearly approaches  $X_i(t)$  as  $dt \rightarrow 0$ ; and also *stochastic*, since we can never predict with certainty what the values of the  $M$  unit normal random variables  $N_j(t)$  will be. What is *not* obvious from Eq. (2.6) is *why*, as is clearly implied by the theorem, *every* continuous Markov process *must* evolve according to an updating formula with this specific structure—i.e., with  $dt$  entering in precisely the termwise manner indicated, and the  $N_j(t)$  all statistically independent, temporally uncorrelated unit normals. The answer to that question can only be inferred from the detailed proof of the theorem, which is given in Sec. III.

The  $M$ -variate Langevin equation (2.6) points up a subtle ambiguity in the univariate Langevin equation (1.1): Since the functions  $b_{ij}$  are not subject to any sign restrictions, we could have chosen the *negative* square root of the function  $D$  in Eq. (1.1), i.e., for  $M=1$ ,  $b_{11}=D^{1/2}$  can be negative. This minor ambiguity for  $M=1$  is really inconsequential; however, it is a hint of a much more substantial ambiguity that occurs for  $M>1$ , and which is directly addressed by the last part of the theorem: Changes in the  $M^2$   $b_{ij}$  functions that leave the  $\frac{1}{2}M(M+1)$  functions  $D_i$  and  $C_{ij}$  defined in Eqs. (2.5) unchanged have no effect on the conditional increments  $\Xi_i$ . That such changes in the  $b_{ij}$  functions are possible is due in part to the fact that there are  $\frac{1}{2}M(M-1)$  more  $b_{ij}$  functions than  $D_i$  and  $C_{ij}$  functions. Since such changes will not change the updating formula (2.6), then they will not change the continuous Markov process  $\mathbf{X}$  that is defined by that updating formula. Put another way, although an  $M$ -variate continuous Markov process is completely specified by the forms of the  $M$  functions  $A_i$  and the  $M^2$  functions  $b_{ij}$ , that specification will be *nonunique* in that all sets of  $b_{ij}$

functions that yield the same  $D_i$  and  $C_{ij}$  functions through Eqs. (2.5) will describe the same continuous Markov process  $\mathbf{X}$ .

On the other hand, whereas every set of smooth functions  $A_i$  and  $b_{ij}$  will yield a legitimate Markov process, not every set of smooth functions  $A_i$ ,  $D_i$ , and  $C_{ij}$  will be acceptable. It turns out that we can have for  $D_i$  and  $C_{ij}$  only those smooth functions that satisfy the  $\frac{1}{2}M(M+1)$  conditions

$$D_i(\mathbf{x}, t) \geq 0 \quad (i = 1, \dots, M), \quad (2.7a)$$

$$C_{ij}^2(\mathbf{x}, t) \leq D_i(\mathbf{x}, t)D_j(\mathbf{x}, t) \quad (i < j = 1, \dots, M). \quad (2.7b)$$

That the inequality (2.7a) must hold follows immediately from the definition (2.5a). That the inequality (2.7b) must hold follows the fact that

$$\begin{aligned} D_i(\mathbf{x}, t)D_j(\mathbf{x}, t) - C_{ij}^2(\mathbf{x}, t) \\ = \sum_{\substack{k,l=1 \\ (k < l)}}^M [b_{ik}(\mathbf{x}, t)b_{jl}(\mathbf{x}, t) - b_{il}(\mathbf{x}, t)b_{jk}(\mathbf{x}, t)]^2, \end{aligned} \quad (2.8)$$

a relation that can be deduced through straightforward algebra from the definitions (2.5).

As with the univariate Langevin equation, the reason why the  $M(dt)^{1/2}$ -terms in Eq. (2.6) do not render the usually much smaller  $dt$ -term there “negligible” is that each of the  $(dt)^{1/2}$ -terms is multiplied by a random variable  $N(0,1)$ , which, being about as often negative as positive, greatly diminishes the effect of those  $(dt)^{1/2}$ -terms over a succession of many  $dt$  increments. The cumulative, long-range contributions of the weak-but-steady  $dt$ -term and the strong-but-erratic  $(dt)^{1/2}$ -terms on the right side of Eq. (2.6) turn out to be comparable if the functions  $A_i$  and  $b_{ij}$  are of comparable magnitudes.

Another way of writing the Langevin equation (2.6), which is frequently encountered in the literature, can be obtained by appealing to the well-known<sup>1,8</sup> result in random variable theory that, for any two numbers  $\alpha$  and  $\beta$ ,

$$\alpha + \beta N(m, \sigma^2) = N(\alpha + \beta m, \beta^2 \sigma^2). \quad (2.9)$$

The implication of this result that  $(dt)^{1/2}N(0,1) = N(0,dt)$  allows us to write Eq. (2.6) as

$$\begin{aligned} X_i(t+dt) = X_i(t) + A_i(\mathbf{X}(t), t)dt + \sum_{j=1}^M b_{ij}(\mathbf{X}(t), t)dW_j(t) \\ (i = 1, \dots, M), \end{aligned} \quad (2.10)$$

where the  $M$  normal random variables

$$dW_j(t) \equiv N(0, dt) \quad (j = 1, \dots, M) \quad (2.11)$$

are stipulated to be statistically independent and temporally uncorrelated. These stipulations imply that the (ensemble) average  $\langle dW_j(t)dW_k(t') \rangle$  vanishes if either  $j \neq k$  or  $t \neq t'$ . Of course,  $\langle dW_j(t)dW_j(t) \rangle = \langle dW_j^2(t) \rangle = dt$ , since  $dW_j(t)$  has mean 0 and variance  $dt$ . Equations (2.10) and (2.6) are entirely equivalent to each other. Either can be used to derive time-evolution equations for the moments of the components of  $\mathbf{X}(t)$ , and either provides a basis for numerically simulating  $\mathbf{X}(t)$ . Equation (2.6) is arguably more convenient for those tasks, since it renders the  $dt$  dependence explicitly.

A third way of writing the multivariate Langevin equation is obtained by first algebraically rearranging Eq. (2.6) to read

$$\begin{aligned} \frac{X_i(t+dt) - X_i(t)}{dt} = A_i(\mathbf{X}(t), t) \\ + \sum_{j=1}^M b_{ij}(\mathbf{X}(t), t)N_j(t)(dt)^{-1/2} \\ (i = 1, \dots, M). \end{aligned} \quad (2.12)$$

Then, noting from the result (2.9) that  $(dt)^{-1/2}N(0,1) = N(0,1/dt)$ , we can formally take the limit  $dt \rightarrow 0$  of Eq. (2.12) to obtain

$$\begin{aligned} \frac{dX_i(t)}{dt} = A_i(\mathbf{X}(t), t) + \sum_{j=1}^M b_{ij}(\mathbf{X}(t), t)\Gamma_j(t) \\ (i = 1, \dots, M), \end{aligned} \quad (2.13)$$

where the  $M$  random variables

$$\Gamma_j(t) \equiv \lim_{dt \rightarrow 0} N(0, 1/dt) \quad (j = 1, \dots, M) \quad (2.14)$$

are stipulated to be statistically independent and temporally uncorrelated.  $\Gamma_j(t)$  is of course the Gaussian white noise process mentioned in Eq. (1.3). And Eq. (2.13) is the *white noise form multivariate Langevin equation*, the  $M$ -variate generalization of Eq. (1.2). The statistically independent, temporally uncorrelated nature of the  $\Gamma_j(t)$ 's implies that the (ensemble) average  $\langle \Gamma_j(t)\Gamma_k(t') \rangle$  vanishes if either  $j \neq k$  or  $t \neq t'$ . But note that  $\langle \Gamma_j(t)\Gamma_j(t) \rangle = \langle \Gamma_j^2(t) \rangle = \delta(0)$ , where  $\delta$  is the Dirac delta function; because,  $\Gamma_j(t)$  has a mean of 0 and variance of  $1/dt$  in the limit  $dt \rightarrow 0$ , and in that limit we can write from the definition of the Dirac delta function  $\delta(0)dt = 1$ , or  $1/dt = \delta(0)$ .

Some people with strict mathematical sensibilities will maintain that Eqs. (2.13) and (2.14) are nonsense, and that the proper inference to be drawn from Eq. (2.12) is that  $X_i(t)$  simply has no derivative unless  $b_{ij}$  vanishes for all  $j = 1$  to  $M$ . But it would be wrong to criticize continuous Markov process theory on the grounds that Eq. (2.13) is of dubious mathematical legitimacy, because the theory can be fully developed using either of the Langevin equations (2.6) or (2.10), without ever invoking the mathematically questionable notion of “white noise.” We shall adopt here a middle ground position, viewing the white noise Langevin equation as a sometimes convenient way of expressing the standard form Langevin equation.

If all the  $b_{ij}$  functions vanish identically, then it is clear from Eq. (2.13) that the  $M$ -variate continuous Markov process  $[X_1, \dots, X_M]$  would be the solution of a set of  $M$  coupled first-order ordinary differential equations. Since any  $M$ th-order ordinary differential equation can be written as a set of  $M$  coupled first-order ordinary differential equations, then it follows that multivariate continuous Markov process theory encompasses, as the special case  $b_{ij} \equiv 0$ , all of ordinary differential equation theory.

Finally, we note that a component  $X_i$  of the  $M$ -variate continuous Markov process  $\mathbf{X}$  will itself be Markovian only in the very special circumstance that none of the  $M+1$  functions  $A_i, b_{i1}, \dots, b_{iM}$  depends on any of the other components of  $\mathbf{X}$ . To prove this, we need not only the random variable result (2.9), but also another well-known<sup>1,8</sup> result: If two normal random variables  $N(m_1, \sigma_1^2)$  and  $N(m_2, \sigma_2^2)$  are statistically independent, then

$$N(m_1, \sigma_1^2) + N(m_2, \sigma_2^2) = N(m_1 + m_2, \sigma_1^2 + \sigma_2^2). \quad (2.15)$$

Using these two random variable results, we may write the summation in Eq. (2.6) as

$$\begin{aligned}\sum_j b_{ij} N_j &= \sum_j b_{ij} N(0,1) = \sum_j N(0, b_{ij}^2) = N(0, \sum_j b_{ij}^2) \\ &= (\sum_j b_{ij}^2)^{1/2} N(0,1) = D^{1/2} N,\end{aligned}$$

where  $D = \sum_j b_{ij}^2$  and  $N$  is a unit normal random variable. Therefore, if the  $M+1$  functions  $A_i, b_{i1}, \dots, b_{iM}$  all depend *only* on the component  $X_i$ , then Eq. (2.6) will take the form of the univariate Langevin equation (1.1), implying that  $X_i$  is itself Markovian.

### III. PROOF OF THE THEOREM

To prove the theorem at Eqs. (2.4) and (2.5), the theorem that made our derivation of the  $M$ -variate Langevin equation so easy, we begin exactly as in the univariate proof in Ref. 1: We divide the interval  $[t, t+dt]$  into  $n \geq 2$  subintervals of equal lengths  $dt/n$  by means of the points  $t_k = t + k(dt/n)$ , ( $k=0$  to  $n$ ). We then have, for each  $i$  from 1 to  $M$ ,

$$\begin{aligned}X_i(t+dt) - X_i(t) &= X_i(t_n) - X_i(t_0) \\ &= \sum_{k=1}^n [X_i(t_k) - X_i(t_{k-1})], \\ X_i(t+dt) - X_i(t) &= \sum_{k=1}^n [X_i(t_{k-1} + dt/n) - X_i(t_{k-1})].\end{aligned}\quad (3.1)$$

This relation shows that the conditional increment  $\Xi_i$  defined in Eq. (2.2) must satisfy

$$\begin{aligned}\Xi_i(dt; \mathbf{X}(t), t) &= \sum_{k=1}^n \Xi_i(dt/n; \mathbf{X}(t_{k-1}), t_{k-1}) \\ &\quad (i=1, \dots, M).\end{aligned}\quad (3.2)$$

Equation (3.2) expresses the *self-consistency condition*, that the conditional increment of the process over any infinitesimal time interval must be equal to the sum of the conditional increments over any set of successive subdivisions of that interval. The Markovian or “past-forgetting” nature of the process is embodied here in the fact that each increment depends only on the value of the process at the *beginning* of the associated interval, and not on any earlier process values.

Condition (ii) in our definition of a continuous Markov process tells us that  $\Xi_i$  is a smooth function of its arguments  $\mathbf{x}$  and  $t$ , and condition (iii) tells us that  $\mathbf{X}$  is continuous in  $t$ . So, since  $dt$  can be made so small that all the  $t_k$ 's are arbitrarily close to  $t$ , then we can *replace* in Eq. (3.2)

$$t_{k-1} \rightarrow t, \quad \mathbf{X}(t_{k-1}) \rightarrow \mathbf{X}(t) \equiv \mathbf{x}, \quad (3.3)$$

and thereby introduce only inconsequential errors of order  $>1$  in  $dt$ . Thus, at least to lowest order in  $dt$ , and for all  $n \geq 2$ ,

$$\Xi_i(dt; \mathbf{x}, t) = \sum_{k=1}^n \Xi_{ik}(dt/n; \mathbf{x}, t) \quad (i=1, \dots, M). \quad (3.4)$$

The terms  $\Xi_{i1}, \dots, \Xi_{in}$  in Eq. (3.4) are  $n$  copies of the random variable  $\Xi_i(dt/n; \mathbf{x}, t)$ . These copies are *statistically independent* of each other, since, as was mentioned earlier, the process increment over any interval depends only on the process value at the beginning of that interval. And condition (iv) assures us that these  $n$  statistically independent random

variables have well defined means and variances. Now, the central limit theorem tells us<sup>1,8</sup> that the sum of  $n$  statistically independent, identically distributed random variables with finite means and variances becomes *normal* in the limit  $n \rightarrow \infty$ . So, since  $n$  here can be taken arbitrarily large, then it must be the case that

$$\Xi_i(dt; \mathbf{x}, t) \text{ is normal } (i=1, \dots, M). \quad (3.5)$$

Next we turn to a consideration of the means, variances, and covariances of the  $M$  normal random variables  $\Xi_i$ . A well known result in random variable theory is that the mean and variance of the sum of any set of statistically independent random variables are equal to the sums of the respective means and variances. It thus follows from Eq. (3.4) that

$$\begin{aligned}\text{mean}\{\Xi_i(dt; \mathbf{x}, t)\} &= \sum_{k=1}^n \text{mean}\{\Xi_{ik}(dt/n; \mathbf{x}, t)\} \\ &= n \text{mean}\{\Xi_i(dt/n; \mathbf{x}, t)\},\end{aligned}\quad (3.6a)$$

$$\begin{aligned}\text{var}\{\Xi_i(dt; \mathbf{x}, t)\} &= \sum_{k=1}^n \text{var}\{\Xi_{ik}(dt/n; \mathbf{x}, t)\} \\ &= n \text{var}\{\Xi_i(dt/n; \mathbf{x}, t)\}.\end{aligned}\quad (3.6b)$$

We now assert that the covariances of  $\Xi_i$  and  $\Xi_j$  for  $i < j$  satisfy a like relation, namely,

$$\begin{aligned}\text{cov}\{\Xi_i(dt; \mathbf{x}, t), \Xi_j(dt; \mathbf{x}, t)\} &= \sum_{k=1}^n \text{cov}\{\Xi_{ik}(dt/n; \mathbf{x}, t), \Xi_{jk}(dt/n; \mathbf{x}, t)\} \\ &= n \text{cov}\{\Xi_i(dt/n; \mathbf{x}, t), \Xi_j(dt/n; \mathbf{x}, t)\}.\end{aligned}\quad (3.6c)$$

The justification for asserting the first line of Eq. (3.6c) is a result in probability theory that, if

$$X = \sum_{k=1}^n X_k, \quad Y = \sum_{l=1}^n Y_l, \quad (3.7a)$$

where  $X_k$  and  $Y_l$  are statistically independent when  $k \neq l$ , then

$$\text{cov}\{X, Y\} = \sum_{k=1}^n \text{cov}\{X_k, Y_k\}. \quad (3.7b)$$

This result is easily proved by first noting that

$$\begin{aligned}\text{cov}\{X, Y\} &\equiv \langle (\sum_k X_k)(\sum_l Y_l) \rangle - \langle \sum_k X_k \rangle \langle \sum_l Y_l \rangle \\ &= \sum_k \sum_l [\langle X_k Y_l \rangle - \langle X_k \rangle \langle Y_l \rangle] \\ &\equiv \sum_k \sum_l \text{cov}\{X_k, Y_l\},\end{aligned}$$

and then observing that  $\text{cov}\{X_k, Y_l\} = 0$  when  $k \neq l$  because of the hypothesized statistical independence of  $X_k$  and  $Y_l$  when  $k \neq l$ . In applying this result to deduce Eq. (3.6c), we have used the fact that  $\Xi_{ik}(dt/n; \mathbf{x}, t)$  and  $\Xi_{jl}(dt/n; \mathbf{x}, t)$  are indeed statistically independent of each other whenever  $k \neq l$ , since the two component increments then occur in *different*  $dt/n$  subintervals. But of course, statistical independence is *not* assumed when  $k = l$ , and the two component increments occur in the *same*  $dt/n$  subinterval.

To extract the important implications of Eqs. (3.6), we next invoke the following result from ordinary calculus:

**Lemma 1.** If  $h(z)$  is any smooth function of  $z$  that satisfies  $h(z) = nh(z/n)$  for all positive integers  $n$ , then it must be true that  $h(z) = Cz$ , where  $C$  is independent of  $z$ .

A proof of this lemma is given in the Appendix. Applying it to Eqs. (3.6), with  $dt$  playing the role of  $z$ , which is justified because condition (ii) requires  $\Xi_i(dt; \mathbf{x}, t)$  to be smooth in  $dt$ , we conclude that the means, variances, and covariances of the  $\Xi_i$ 's must all be *directly proportional* to  $dt$ :

$$\text{mean}\{\Xi_i(dt; \mathbf{x}, t)\} = A_i(\mathbf{x}, t)dt \quad (i=1, \dots, M), \quad (3.8a)$$

$$\text{var}\{\Xi_i(dt; \mathbf{x}, t)\} = D_i(\mathbf{x}, t)dt \quad (i=1, \dots, M), \quad (3.8b)$$

$$\text{cov}\{\Xi_i(dt; \mathbf{x}, t), \Xi_j(dt; \mathbf{x}, t)\} = C_{ij}(\mathbf{x}, t)dt \quad (i < j = 1, \dots, M). \quad (3.8c)$$

Here, the “coefficients”  $A_i$ ,  $D_i$ , and  $C_{ij}$  may depend (smoothly by hypothesis) on  $\mathbf{x}$  and  $t$ , but *not* on  $dt$ . The fact that the variance of any random variable is never negative, and the fact that the squared covariance of any two random variables is always bounded by the product of their variances, evidently dictate that the functions  $D_i$  and  $C_{ij}$  must satisfy the auxiliary relations

$$D_i(\mathbf{x}, t) \geq 0 \quad (i=1, \dots, M), \quad (3.9a)$$

$$C_{ij}^2(\mathbf{x}, t) \leq D_i(\mathbf{x}, t)D_j(\mathbf{x}, t) \quad (i < j = 1, \dots, M). \quad (3.9b)$$

We have thus far established that the defining conditions (i)–(iv) of an  $M$ -variate continuous Markov process  $\mathbf{X}$  imply that the conditioned increments  $\Xi_i(dt; \mathbf{x}, t)$  must be normal random variables whose means, variances, and covariances have the analytical forms indicated in Eqs. (3.8), wherein  $A_i$ ,  $D_i$ , and  $C_{ij}$  may be any smooth functions of  $x_1, \dots, x_M$ , and  $t$  that satisfy Eqs. (3.9). Now we shall show that formulas (2.4) endow the conditioned increments  $\Xi_i(dt; \mathbf{x}, t)$  with precisely these properties whenever the  $b_{ij}$  functions are chosen to satisfy Eqs. (2.5). For this, we appeal to the following result of random variable theory:

**Lemma 2.** If the  $M$  random variables  $Y_1, \dots, Y_M$  are defined by the relations

$$Y_i = \alpha_i + \sum_{j=1}^M \beta_{ij} N_j \quad (i=1, \dots, M), \quad (3.10)$$

where  $N_1, \dots, N_M$  are  $M$  statistically independent unit normals, then  $Y_1, \dots, Y_M$  are all *normal* random variables with

$$\text{mean}\{Y_i\} = \alpha_i \quad (i=1, \dots, M), \quad (3.11a)$$

$$\text{var}\{Y_i\} = \sum_{j=1}^M \beta_{ij}^2 \quad (i=1, \dots, M), \quad (3.11b)$$

$$\text{cov}\{Y_i, Y_j\} = \sum_{k=1}^M \beta_{ik}\beta_{jk} \quad (i < j = 1, \dots, M). \quad (3.11c)$$

The proof of this lemma is straightforward, and is given in the Appendix. Since formula (2.4) for  $\Xi_i(dt; \mathbf{x}, t)$  is of the form (3.10), with  $\alpha_i = A_i(\mathbf{x}, t)dt$  and  $\beta_{ij} = b_{ij}(\mathbf{x}, t)(dt)^{1/2}$ , then we conclude from the lemma that the random variables  $\Xi_i(dt; \mathbf{x}, t)$  given in Eqs. (2.4), with the functions  $D_i$  and  $C_{ij}$  as defined in Eqs. (2.5), are indeed *normal* random variables with

$$\text{mean}\{\Xi_i(dt; \mathbf{x}, t)\} = A_i(\mathbf{x}, t)dt,$$

$$\text{var}\{\Xi_i(dt; \mathbf{x}, t)\} = \sum_{j=1}^M b_{ij}^2(\mathbf{x}, t)dt \equiv D_i(\mathbf{x}, t)dt,$$

$$\text{cov}\{\Xi_i(dt; \mathbf{x}, t), \Xi_j(dt; \mathbf{x}, t)\} = \sum_{k=1}^M b_{ik}(\mathbf{x}, t)b_{jk}(\mathbf{x}, t)dt \equiv C_{ij}(\mathbf{x}, t)dt,$$

precisely as required by Eqs. (3.5) and (3.8).

The required smoothness of the functions  $D_i(\mathbf{x}, t)$  and  $C_{ij}(\mathbf{x}, t)$  is ensured through the definitions (2.5) by the hypothesized smoothness of the functions  $b_{ij}(\mathbf{x}, t)$ . Also, the earlier noted consequences (2.7) of definitions (2.5) are seen to be precisely the required auxiliary relations (3.9). Finally, we note that the hypothesis that the statistically independent unit normals  $N_j(t)$  be temporally uncorrelated guarantees that the conditioned increments in  $X_i$  over successive infinitesimal time intervals will be statistically independent, as is required by the past-forgetting condition (i). The theorem at Eqs. (2.4) and (2.5) is thus proved.

We can briefly summarize the thrust of the foregoing proof as follows: In order for the conditional increments  $\Xi_i(dt; \mathbf{x}, t)$  in Eq. (2.2) to have the *Markov* property (i) and the *continuity* property (iii), given the “reasonable behavior” stipulations (ii) and (iv), logical necessity requires—mainly by way of the self-consistency condition (3.2), the central limit theorem, and our lemma 1—that the  $M$  random variables  $\Xi_i(dt; \mathbf{x}, t)$  be *normal*, with means, variances, and covariances that are all *directly proportional* to  $dt$ . But it follows from our lemma 2 that any such set of  $M$  random variables can be written as a linear combination of  $M$  statistically independent unit normals  $N_j(t)$  in the manner of Eqs. (2.4).

#### IV. THE MULTIVARIATE FOKKER–PLANCK EQUATIONS

The  $M$ -variate generalizations of the univariate forward and backward Fokker–Planck equations (1.4) and (1.5) govern the behavior of the singly conditioned density function  $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$  of the random variable  $\mathbf{X}(t)$ . By definition,

$$P(\mathbf{x}, t | \mathbf{x}_0, t_0) d^M \mathbf{x} \equiv P(x_1, \dots, x_M, t | x_{01}, \dots, x_{0M}, t_0) dx_1, \dots, dx_M$$

gives the probability that  $X_i(t)$  will lie between  $x_i$  and  $x_i + dx_i$  simultaneously for all  $i=1$  to  $M$ , given that  $X_i(t_0) = x_{0i}$  for all  $i=1$  to  $M$ , and for any  $t_0 \leq t$ . But it is important to recognize that this function  $P$  is merely the first in an *infinite sequence* of conditioned joint probability density functions,  $P \equiv P^{(1)}, P^{(2)}, \dots$ , where  $P^{(n)}$  is defined by

$$P^{(n)}(\mathbf{x}, t | \mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1) d^M \mathbf{x} = \text{Prob}\{X_i(t) \in [x_i, x_i + dx_i] \text{ for } i=1 \text{ to } M, \text{ given that } \mathbf{X}(t_j) = \mathbf{x}_j \text{ for } j=1 \text{ to } n, \text{ with } t_1 \leq \dots \leq t_n \leq t\} \quad (n=1, 2, \dots). \quad (4.1)$$

Everything knowable about the process  $\mathbf{X}$  can in principle be calculated from this infinite set of conditioned density functions. For example, if  $\mathbf{X}(t_0) = \mathbf{x}_0$ , then the joint probability that  $\mathbf{X}(t_1)$  will be within  $d^M \mathbf{x}_1$  of  $\mathbf{x}_1$ , and then at a later time  $t_2$   $\mathbf{X}(t_2)$  will be within  $d^M \mathbf{x}_2$  of  $\mathbf{x}_2$ , can be calculated, ac-

cording to the multiplication law of probability theory, as the following product of the two functions  $P^{(1)}$  and  $P^{(2)}$ :

$$[P^{(1)}(\mathbf{x}_1, t_1 | \mathbf{x}_0, t_0) d^M \mathbf{x}_1] [P^{(2)}(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1; \mathbf{x}_0, t_0) d^M \mathbf{x}_2].$$

A stochastic process  $\mathbf{X}$ , whether it is Markovian or not, cannot be regarded as being known unless *all* of the functions  $P^{(n)}$  are specified. But we are *not* free to specify these  $P^{(n)}$  functions arbitrarily: The general result in probability theory that  $P_Z(z) = \int P_{Z|Y}(z|y) P_Y(y) dy$  requires that the set of functions  $\{P^{(n)}\}$  must satisfy

$$\begin{aligned} P^{(n)}(\mathbf{x}_{n+2}, t_{n+2} | \mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1) \\ = \int P^{(n+1)}(\mathbf{x}_{n+2}, t_{n+2} | \mathbf{x}_{n+1}, t_{n+1}; \mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1) \\ \times P^{(n)}(\mathbf{x}_{n+1}, t_{n+1} | \mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1) d^M \mathbf{x}_{n+1} \\ (\forall t_1 \leq t_2 \leq \dots \leq t_{n+2} \text{ and } \forall n = 1, 2, \dots), \end{aligned} \quad (4.2)$$

where it is understood that the integrations are carried out over all possible values of the  $M$  real variables  $\mathbf{x}_{n+1,1}, \dots, \mathbf{x}_{n+1,M}$ .

The daunting prospect of having to specify an infinite number of conditioned density functions that satisfy an infinite number of coupled integral equations simplifies enormously if  $\mathbf{X}$  is stipulated to be *Markovian*. The *memoryless* property of a Markov process implies that *only the most recent conditioning matters*; therefore, for a Markov process each function  $P^{(n)}$  will coincide with  $P^{(1)} \equiv P$  according to

$$P^{(n)}(\mathbf{x}, t | \mathbf{x}_n, t_n; \dots; \mathbf{x}_1, t_1) = P^{(1)}(\mathbf{x}, t | \mathbf{x}_n, t_n) \equiv P(\mathbf{x}, t | \mathbf{x}_n, t_n) \quad (t_n \leq t; n = 2, 3, \dots). \quad (4.3)$$

In that case, it is easy to show that *every one* of the required relations (4.2) reduces to the form

$$\begin{aligned} P(\mathbf{x}_3, t_3 | \mathbf{x}_1, t_1) = \int P(\mathbf{x}_3, t_3 | \mathbf{x}_2, t_2) P(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1) d^M \mathbf{x}_2 \\ (\forall t_1 \leq t_2 \leq t_3), \end{aligned} \quad (4.4)$$

a relation that is known as the ( $M$ -variate) *Chapman–Kolmogorov equation*. So, for a *Markov* process, we can specify *all* of the functions  $P^{(n)}$ , and be assured that they satisfy *all* of the conditions (4.2), simply by specifying a *single* function  $P$  that satisfies the *single* condition (4.4). Henceforth, we restrict our discussion to the Markovian case.

The Fokker–Planck equations are derived by making Taylor series expansions of carefully selected  $(\mathbf{x}_i, t_i)$ -versions of the Chapman–Kolmogorov equation (4.4). Specifically, the *forward* Fokker–Planck equation, which deals with the “ $t$ -part” of  $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$ , is derived by making a Taylor series expansion of Eq. (4.4) written in the form

$$\begin{aligned} P(\mathbf{x}, t + dt | \mathbf{x}_0, t_0) = \int P(\mathbf{x}, t + dt | \mathbf{x} - \boldsymbol{\xi}, t) \\ \times P(\mathbf{x} - \boldsymbol{\xi}, t | \mathbf{x}_0, t_0) d^M \boldsymbol{\xi} \quad (t > t_0). \end{aligned} \quad (4.5)$$

And the *backward* Fokker–Planck equation, which deals with the “ $t_0$ -part” of  $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$ , is derived by making a Taylor series expansion of Eq. (4.4) written in the form

$$\begin{aligned} P(\mathbf{x}, t | \mathbf{x}_0, t_0) = \int P(\mathbf{x}, t | \mathbf{x}_0 + \boldsymbol{\xi}, t_0 + dt_0) \\ \times P(\mathbf{x}_0 + \boldsymbol{\xi}, t_0 + dt_0 | \mathbf{x}_0, t_0) d^M \boldsymbol{\xi} \quad (t > t_0). \end{aligned} \quad (4.6)$$

The derivation of the multivariate *forward* Fokker–Planck equation begins by defining

$$f(\mathbf{x}) \equiv P(\mathbf{x} + \boldsymbol{\xi}, t + dt | \mathbf{x}, t) P(\mathbf{x}, t | \mathbf{x}_0, t_0), \quad (4.7)$$

so that the integrand in Eq. (4.5) is  $f(\mathbf{x} - \boldsymbol{\xi})$ . We expand  $f(\mathbf{x} - \boldsymbol{\xi})$  in a Taylor series as

$$\begin{aligned} f(\mathbf{x} - \boldsymbol{\xi}) = f(\mathbf{x}) + \sum_{n=1}^{\infty} \sum_{\substack{k_1, \dots, k_M=0 \\ [k_1 + \dots + k_M = n]}}^n \frac{1}{k_1! \dots k_M!} \\ \times (-\xi_1)^{k_1} \dots (-\xi_M)^{k_M} \frac{\partial^n f(\mathbf{x})}{\partial x_1^{k_1} \dots \partial x_M^{k_M}}. \end{aligned} \quad (4.8)$$

Substituting this expansion into Eq. (4.5), regrouping factors, and then inserting the explicit formula (4.7) for  $f(\mathbf{x})$ , we get

$$\begin{aligned} P(\mathbf{x}, t + dt | \mathbf{x}_0, t_0) = \int d^M \boldsymbol{\xi} P(\mathbf{x} + \boldsymbol{\xi}, t + dt | \mathbf{x}, t) P(\mathbf{x}, t | \mathbf{x}_0, t_0) \\ + \sum_{n=1}^{\infty} (-1)^n \sum_{\substack{k_1, \dots, k_M=0 \\ [k_1 + \dots + k_M = n]}}^n \frac{1}{k_1! \dots k_M!} \\ \times \int d^M \boldsymbol{\xi} \xi_1^{k_1} \dots \xi_M^{k_M} \frac{\partial^n}{\partial x_1^{k_1} \dots \partial x_M^{k_M}} \\ \times [P(\mathbf{x} + \boldsymbol{\xi}, t + dt | \mathbf{x}, t) P(\mathbf{x}, t | \mathbf{x}_0, t_0)]. \end{aligned}$$

Since  $\int d^M \mathbf{x}_2 P(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1) = 1$ , then the first term on the right integrates to give  $P(\mathbf{x}, t | \mathbf{x}_0, t_0)$ . Subtracting that term from both sides and then dividing through by  $dt$ , we get

$$\begin{aligned} \frac{P(\mathbf{x}, t + dt | \mathbf{x}_0, t_0) - P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{dt} \\ = \sum_{n=1}^{\infty} (-1)^n \sum_{\substack{k_1, \dots, k_M=0 \\ [k_1 + \dots + k_M = n]}}^n \frac{\partial^n}{\partial x_1^{k_1} \dots \partial x_M^{k_M}} \\ \times \left\{ \frac{1}{dt} \frac{1}{k_1! \dots k_M!} \left[ \int d^M \boldsymbol{\xi} \xi_1^{k_1} \dots \xi_M^{k_M} \right. \right. \\ \left. \left. \times P(\mathbf{x} + \boldsymbol{\xi}, t + dt | \mathbf{x}, t) \right] P(\mathbf{x}, t | \mathbf{x}_0, t_0) \right\}. \end{aligned} \quad (4.9)$$

Now we observe that the function  $P(\mathbf{x} + \boldsymbol{\xi}, t + dt | \mathbf{x}, t)$  inside the square brackets of Eq. (4.9), considered as a function of  $\boldsymbol{\xi} \equiv (\xi_1, \dots, \xi_M)$ , is just the joint density function of the  $M$  random variables  $\Xi_1(dt; \mathbf{x}, t), \dots, \Xi_M(dt; \mathbf{x}, t)$ ; because,  $P(\mathbf{x} + \boldsymbol{\xi}, t + dt | \mathbf{x}, t) d^M \boldsymbol{\xi}$ , being the probability that  $X_i(t + dt) \in [x_i + \xi_i, x_i + \xi_i + d\xi_i]$  for all  $i$  from 1 to  $M$  given that  $\mathbf{X}(t) = \mathbf{x}$ , is therefore, by virtue of the definition (2.2), the probability that  $\Xi_i(dt; \mathbf{x}, t) \in [\xi_i, \xi_i + d\xi_i]$  for all  $i$  from 1 to  $M$ . Accordingly, we have

$$\begin{aligned} \int d^M \boldsymbol{\xi} \xi_1^{k_1} \dots \xi_M^{k_M} P(\mathbf{x} + \boldsymbol{\xi}, t + dt | \mathbf{x}, t) \\ = \langle \Xi_1^{k_1}(dt; \mathbf{x}, t) \dots \Xi_M^{k_M}(dt; \mathbf{x}, t) \rangle. \end{aligned} \quad (4.10)$$

Substituting this into Eq. (4.9) and then passing to the limit  $dt \rightarrow 0$ , we obtain the so-called *forward Kramers–Moyal equation*,

$$\frac{\partial P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\partial t} = \sum_{n=1}^{\infty} (-1)^n \sum_{\substack{k_1, \dots, k_M=0 \\ [k_1 + \dots + k_M = n]}}^n \frac{\partial^n}{\partial x_1^{k_1} \dots \partial x_M^{k_M}} \times \{B_{k_1, \dots, k_M}^{(n)}(\mathbf{x}, t) P(\mathbf{x}, t | \mathbf{x}_0, t_0)\}, \quad (4.11)$$

wherein

$$B_{k_1, \dots, k_M}^{(n)}(\mathbf{x}, t) \equiv \lim_{dt \rightarrow 0} \frac{1}{dt} \frac{1}{k_1! \dots k_M!} \times \langle \Xi_1^{k_1}(dt; \mathbf{x}, t) \dots \Xi_M^{k_M}(dt; \mathbf{x}, t) \rangle \left( k_i \geq 0 \quad \forall i=1 \text{ to } M, \text{ and } \sum_{i=1}^M k_i = n \right). \quad (4.12)$$

The final step in our derivation of the forward multivariate Fokker–Planck equation is to evaluate the  $B$ -functions. This task is made relatively easy by our theorem in Sec. II. To compute the  $B^{(1)}$  functions, we simply take the average of Eq. (2.4); since  $\langle N_j(t) \rangle = 0$ , we get

$$\langle \Xi_i(dt; \mathbf{x}, t) \rangle = A_i(\mathbf{x}, t) dt \quad (i=1, \dots, M). \quad (4.13)$$

So the definition (4.12) gives

$$B_{1,0,0,\dots,0}^{(1)}(\mathbf{x}, t) = A_1(\mathbf{x}, t), \quad (4.14)$$

$$B_{0,1,0,\dots,0}^{(1)}(\mathbf{x}, t) = A_2(\mathbf{x}, t), \quad \text{etc.}$$

To compute the  $B^{(2)}$  functions, we use Eq. (2.4) to form the product  $\Xi_i(dt; \mathbf{x}, t) \Xi_j(dt; \mathbf{x}, t)$ , and then we take the average of that product; since  $\langle N_k(t) N_l(t) \rangle = \delta_{kl}$ , we get

$$\langle \Xi_i(dt; \mathbf{x}, t) \Xi_j(dt; \mathbf{x}, t) \rangle = \sum_{k=1}^M b_{ik}(\mathbf{x}, t) b_{jk}(\mathbf{x}, t) dt + o(dt) \quad (i, j=1, \dots, M),$$

where  $o(dt)$  denotes terms that go to zero with  $dt$  faster than  $dt$ . For the two cases  $i=j$  and  $i < j$ , this gives, by virtue of the definitions (2.5),

$$\langle \Xi_i^2(dt; \mathbf{x}, t) \rangle = D_i(\mathbf{x}, t) dt + o(dt) \quad (i=1, \dots, M), \quad (4.15a)$$

$$\langle \Xi_i(dt; \mathbf{x}, t) \Xi_j(dt; \mathbf{x}, t) \rangle = C_{ij}(\mathbf{x}, t) dt + o(dt) \quad (i < j=1, \dots, M). \quad (4.15b)$$

Consequently, Eqs. (4.12) give

$$\left. \begin{aligned} B_{2,0,0,\dots,0}^{(2)}(\mathbf{x}, t) &= \frac{1}{2} D_1(\mathbf{x}, t), \\ B_{0,2,0,\dots,0}^{(2)}(\mathbf{x}, t) &= \frac{1}{2} D_2(\mathbf{x}, t), \quad \text{etc.}, \\ B_{1,1,0,\dots,0}^{(2)}(\mathbf{x}, t) &= C_{12}(\mathbf{x}, t), \\ B_{1,0,1,\dots,0}^{(2)}(\mathbf{x}, t) &= C_{13}(\mathbf{x}, t), \quad \text{etc.} \end{aligned} \right\} \quad (4.16)$$

To compute the  $B^{(3)}$  functions, we start by using Eq. (2.4) to form the product  $\Xi_i(dt; \mathbf{x}, t) \Xi_j(dt; \mathbf{x}, t) \Xi_k(dt; \mathbf{x}, t)$ . But a glance at formula (2.4) reveals that every term in that product will be proportional to  $dt$  raised to at least the power 3/2. Clearly, all of those terms will give zero when substituted into the  $B$ -formula (4.12). Similar reasoning applied to the  $B^{(n)}$  functions for  $n > 3$  forces us to conclude that

$$B_{k_1, \dots, k_M}^{(n)}(\mathbf{x}, t) = 0 \quad \text{for all } n \geq 3. \quad (4.17)$$

Substituting the foregoing results for the  $B$ -functions into the forward Kramers–Moyal equation (4.11) and then regrouping terms, we obtain

$$\begin{aligned} \frac{\partial P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\partial t} &= - \sum_{i=1}^M \frac{\partial}{\partial x_i} [A_i(\mathbf{x}, t) P(\mathbf{x}, t | \mathbf{x}_0, t_0)] \\ &+ \frac{1}{2} \sum_{i=1}^M \frac{\partial^2}{\partial x_i^2} [D_i(\mathbf{x}, t) P(\mathbf{x}, t | \mathbf{x}_0, t_0)] \\ &+ \sum_{\substack{i,j=1 \\ [i < j]}}^M \frac{\partial^2}{\partial x_i \partial x_j} [C_{ij}(\mathbf{x}, t) P(\mathbf{x}, t | \mathbf{x}_0, t_0)]. \end{aligned} \quad (4.18)$$

This is the *multivariate forward Fokker–Planck equation*, the  $M$ -variate generalization of Eq. (1.4).

Our derivation of the multivariate *backward* Fokker–Planck equation proceeds from Eq. (4.6), and begins by defining

$$h(\mathbf{x}_0) \equiv P(\mathbf{x}, t | \mathbf{x}_0, t_0 + dt_0), \quad (4.19)$$

so that the *first factor* in the integrand of Eq. (4.6) can be written as  $h(\mathbf{x}_0 + \xi)$ . By expanding  $h(\mathbf{x}_0 + \xi)$  in a Taylor series about  $\mathbf{x}_0$  in the manner of Eq. (4.8), we bring Eq. (4.6) into the form

$$\begin{aligned} P(\mathbf{x}, t | \mathbf{x}_0, t_0) &= \int d^M \xi P(\mathbf{x}, t | \mathbf{x}_0, t_0 + dt_0) P(\mathbf{x}_0 + \xi, t_0 + dt_0 | \mathbf{x}_0, t_0) \\ &+ \sum_{n=1}^{\infty} \sum_{\substack{k_1, \dots, k_M=0 \\ [k_1 + \dots + k_M = n]}}^n \frac{1}{k_1! \dots k_M!} \int d^M \xi \xi_1^{k_1} \dots \xi_M^{k_M} \\ &\times \frac{\partial^n P(\mathbf{x}, t | \mathbf{x}_0, t_0 + dt_0)}{\partial x_{01}^{k_1} \dots \partial x_{0M}^{k_M}} P(\mathbf{x}_0 + \xi, t_0 + dt_0 | \mathbf{x}_0, t_0). \end{aligned}$$

Since  $\int d^M \mathbf{x}_2 P(\mathbf{x}_2, t_2 | \mathbf{x}_1, t_1) = 1$ , the first term on the right integrates to give  $P(\mathbf{x}, t | \mathbf{x}_0, t_0 + dt_0)$ . Subtracting that term from both sides and then dividing through by  $dt_0$ , we get

$$\begin{aligned} & - \frac{P(\mathbf{x}, t | \mathbf{x}_0, t_0 + dt_0) - P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{dt_0} \\ &= \sum_{n=1}^{\infty} \sum_{\substack{k_1, \dots, k_M=0 \\ [k_1 + \dots + k_M = n]}}^n \left\{ \frac{1}{dt_0} \frac{1}{k_1! \dots k_M!} \int d^M \xi \xi_1^{k_1} \dots \xi_M^{k_M} \right. \\ &\quad \times P(\mathbf{x}_0 + \xi, t_0 + dt_0 | \mathbf{x}_0, t_0) \left. \right\} \frac{\partial^n P(\mathbf{x}, t | \mathbf{x}_0, t_0 + dt_0)}{\partial x_{01}^{k_1} \dots \partial x_{0M}^{k_M}}. \end{aligned}$$

Finally, making use of Eqs. (4.10) and (4.12)–(4.17), we pass to the limit  $dt_0 \rightarrow 0$  and obtain

$$\begin{aligned}
& - \frac{\partial P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\partial t_0} \\
& = \sum_{i=1}^M A_i(\mathbf{x}_0, t_0) \frac{\partial}{\partial x_{0i}} [P(\mathbf{x}, t | \mathbf{x}_0, t_0)] \\
& + \frac{1}{2} \sum_{i=1}^M D_i(\mathbf{x}_0, t_0) \frac{\partial^2}{\partial x_{0i}^2} [P(\mathbf{x}, t | \mathbf{x}_0, t_0)] \\
& + \sum_{\substack{i,j=1 \\ [i < j]}}^M C_{ij}(\mathbf{x}_0, t_0) \frac{\partial^2}{\partial x_{0i} \partial x_{0j}} [P(\mathbf{x}, t | \mathbf{x}_0, t_0)]. \quad (4.21)
\end{aligned}$$

This is the *multivariate backward Fokker–Planck equation*, the  $M$ -variate generalization of Eq. (1.5).

## V. TWO SIMPLE EXAMPLES

We now consider briefly two simple examples of the *bivariate* Langevin and Fokker–Planck equations. The first is a mathematics application, and the second a physics application.

### A. A univariate continuous Markov process and its integral

A general univariate continuous Markov process  $X$  is governed by the Langevin equation (1.1). The time integral  $Y$  of that process  $X$  (or indeed of any process  $X$ ) can be defined by the updating formula

$$Y(t+dt) = Y(t) + X(t)dt. \quad (5.1)$$

By putting  $X = X_1$  and  $Y = X_2$  in Eqs. (1.1) and (5.1), we find that those two equations take the form of the multivariate Langevin equation (2.6) with  $M=2$  and

$$\left. \begin{aligned}
A_1(\mathbf{x}, t) &= A(x_1, t) = A(x, t), \\
A_2(\mathbf{x}, t) &= x_1 = x, \\
b_{11}(\mathbf{x}, t) &= D^{1/2}(x_1, t) = D^{1/2}(x, t), \\
b_{12}(\mathbf{x}, t) &= b_{21}(\mathbf{x}, t) = b_{22}(\mathbf{x}, t) = 0.
\end{aligned} \right\} \quad (5.2)$$

It therefore follows that  $X$  and  $Y$  together comprise a *bivariate* continuous Markov process.

The companion forward Fokker–Planck equation to the Langevin equations (1.1) and (5.1) is found by substituting Eqs. (5.2) into Eq. (4.18), utilizing the definitions (2.5). In that way we find that the joint density function  $P \equiv P(x, y, t | x_0, y_0, t_0)$  for  $X(t)$  and  $Y(t)$  is the solution of

$$\frac{\partial P}{\partial t} = - \frac{\partial}{\partial x} [A(x, t)P] - x \frac{\partial P}{\partial y} + \frac{1}{2} \frac{\partial^2}{\partial x^2} [D(x, t)P] \quad (5.3)$$

for the initial condition  $P(x, y, t_0 | x_0, y_0, t_0) = \delta(x - x_0) \times \delta(y - y_0)$ . The corresponding backward Fokker–Planck equation is similarly obtained from Eq. (4.21).

By integrating Eq. (5.3) over all  $y$  and using the fact that the normalization of  $P$  requires that  $P$  vanish at  $y = \pm\infty$ , we find that the marginal density function for  $X(t)$ ,

$$\begin{aligned}
\int_{-\infty}^{\infty} P(x, y, t | x_0, y_0, t_0) dy &\equiv P_X(x, t | x_0, y_0, t_0) \\
&= P_X(x, t | x_0, t_0), \quad (5.4a)
\end{aligned}$$

satisfies the univariate forward Fokker–Planck equation (1.4), just as we should expect. However, no analogous closed equation seems obtainable for the marginal density function for  $Y(t)$ ,

$$\int_{-\infty}^{\infty} P(x, y, t | x_0, y_0, t_0) dx \equiv P_Y(y, t | x_0, y_0, t_0), \quad (5.4b)$$

which depends on  $x_0$  as well as  $y_0$ . To find  $P_Y$ , we apparently must first solve Eq. (5.3) for  $P$  and then integrate out  $x$ . This points up the fact that  $Y$  by itself is *not* a Markov process: Since the  $Y$ -updating formula (5.1) involves a process other than  $Y$ , it is not of the canonical univariate Langevin form. The nonmemoryless nature of  $Y$  is reflected in the fact that the value  $X(t)$  on the right side of Eq. (5.1) tells us more about the past of  $Y$  than we can infer from  $Y$ 's current value  $Y(t)$  alone, e.g., if  $X(t) < 0$ , then we may infer that just before  $t$   $Y$  had a value greater than  $Y(t)$ . The non-Markovian nature of  $Y$  also manifests itself in the fact that  $Y$  has a “proper” time derivative, which is something that a genuinely stochastic Markov process does not have.

### B. The Kramers equation and the Smoluchowski equation

As was discussed in Ref. 1,<sup>9</sup> the velocity  $V$  of a Brownian particle of mass  $m$  immersed in a fluid at absolute temperature  $T$  and subject to no external forces obeys the equation

$$m \frac{dV(t)}{dt} = -\gamma V(t) + (2kT\gamma)^{1/2} \Gamma(t). \quad (5.5)$$

Here,  $\Gamma(t)$  is Gaussian white noise,  $k$  is Boltzmann's constant, and  $\gamma$  is the drag coefficient, which is related to the diffusion constant  $D$  through the Einstein relation,

$$\gamma = kT/D. \quad (5.6)$$

Equation (5.5) is of course Newton's second law for the particle under the *assumption* that the aggregate force of the fluid on the particle can be resolved into a linear drag component and a zero-mean randomly fluctuating component. The coefficient of the fluctuating force component, which arises from the requirement that  $V(t \rightarrow \infty)$  have an appropriate Maxwell–Boltzmann distribution, shows that the drag and fluctuating forces are intimately connected through their joint dependence on  $\gamma$ .

Suppose we impose on the Brownian particle a conservative external force field

$$F_{\text{ext}}(x) = -U'(x), \quad (5.7)$$

where the prime denotes differentiation with respect to the position variable  $x$ . Then it is reasonable to assume that we need merely to add this force to the right side of Eq. (5.5). Letting  $X(t)$  denote the position of the particle at time  $t$ , we therefore have

$$\frac{dV(t)}{dt} = -\frac{\gamma}{m} V(t) - \frac{1}{m} U'(X(t)) + \frac{(2kT\gamma)^{1/2}}{m} \Gamma(t), \quad (5.8a)$$

$$\frac{dX(t)}{dt} = V(t). \quad (5.8b)$$

These equations will be seen to be of the canonical Langevin form (2.13) with  $M=2$ ; hence,  $V$  and  $X$  together constitute a bivariate continuous Markov process. If we make the  $A_i$  and  $b_{ij}$  associations required to write Eqs. (2.13) as Eqs. (5.8)

and then compute the corresponding functions  $D_i$  and  $C_{ij}$  from Eqs. (2.5), we find that the canonical forward Fokker–Planck equation (4.18) gives the following time-evolution equation for the joint density function  $P \equiv P(v, x, t | v_0, x_0, t_0)$  of  $V(t)$  and  $X(t)$ :

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial v} \left[ \frac{\gamma v + U'(x)}{m} P \right] - v \frac{\partial P}{\partial x} + \frac{kT\gamma}{m^2} \frac{\partial^2 P}{\partial v^2}. \quad (5.9)$$

This equation (slightly rescaled) has been dubbed the *Kramers equation* by Gardiner.<sup>10</sup> The corresponding backward Fokker–Planck equation follows similarly from Eq. (4.21).

Since Eq. (5.8b) is not of the canonical univariate Langevin form, then  $X$  is not by itself a Markov process. And unless the external force  $-U'(x)$  is independent of  $x$ , as it would be though in the case of the uniform gravitational field  $U(x) = mgx$ , Eq. (5.8a) is also not of the canonical univariate Langevin form. So for a *general* external potential  $U(x)$ , *neither*  $V$  nor  $X$  is individually Markovian; yet the *pair*  $(V, X)$  is.

One can show by straightforward substitution and differentiation that a time-independent or *stationary* solution of the Kramers equation (5.9) is

$$P^*(v, x) \equiv K \exp \left( - \frac{\frac{1}{2}mv^2 + U(x)}{kT} \right), \quad (5.10)$$

where  $K$  is a normalizing constant. This is an especially satisfying result, being precisely what is required by equilibrium statistical mechanics. We may expect that the Kramers equation (5.9) will govern the precise way in which  $P$  approaches  $P^*$  as  $t \rightarrow \infty$ .

An interesting simplifying approximation can be made to the Kramers equation if the motion of the particle is viewed on a spatio-temporal scale in which  $V(t)$  *appears* to be constant. In such a case, Eq. (5.8a) can be approximated by setting  $dV/dt \approx 0$ . If we solve the thus approximated equation (5.8a) algebraically for  $V(t)$  and then substitute the result into Eq. (5.8b), we obtain, remembering Eq. (5.6),

$$\frac{dX(t)}{dt} \approx - \frac{1}{\gamma} U'(X(t)) + (2D)^{1/2} \Gamma(t). \quad (5.11)$$

This equation has (approximately) the canonical form (1.2) of a *univariate* Langevin equation for  $X$ , with  $A(x, t) = -U'(x)/\gamma$  and  $D(x, t) = 2D$ . The corresponding forward Fokker–Planck equation (1.4) for the density function  $P_X(x, t)$  of  $X(t)$  therefore reads

$$\frac{\partial P_X(x, t)}{\partial t} \approx \frac{1}{\gamma} \frac{\partial}{\partial x} [U'(x) P_X(x, t)] + D \frac{\partial^2 P_X(x, t)}{\partial x^2}. \quad (5.12)$$

This equation is called the *Smoluchowski equation*,<sup>11</sup> although its approximate nature is not always emphasized.<sup>12</sup> To find  $P_X(x, t)$  *exactly*, one could first solve the Kramers equation (5.9) for  $P$  and then integrate out  $v$ .

It is not hard to show that a stationary solution of the Smoluchowski equation (5.12) is

$$P_X^*(x) = K' \exp[-U(x)/kT], \quad (5.13)$$

which is the *exact* result predicted by the stationary solution (5.10) of the Kramers equation.

## VI. DISCUSSION

The key elements of our derivation of the  $M$ -variate Langevin equation in Sec. II were (a) the definition of a continuous Markov process in terms of the conditional increments  $\Xi_i(dt; \mathbf{x}, t)$  in Eq. (2.2), and (b) the theorem at Eqs. (2.4) and (2.5). Our proof of the theorem, which was given in Sec. III, showed that the mathematical form (2.4) of  $\Xi_i(dt; \mathbf{x}, t)$  is a consequence of the self-consistency relation (3.2) and several theorems of random variable theory and ordinary calculus (namely the central limit theorem, some elementary theorems on the additive properties of moments, and our lemmas 1 and 2). Our analysis thus showed that many seemingly arbitrary features of the Langevin equation (2.6), such as the special way in which the element of stochasticity enters through the last  $M$  terms, with each term being proportional to the *square root* of  $dt$  and a *statistically independent, temporally uncorrelated, zero-mean, normal* random variable, are really not arbitrary. All of those features emerge as inevitable consequences of the requirement that the process  $\mathbf{X}$  be continuous, memoryless, and self-consistent. The arguments in our derivation of the Langevin equation make room for, but are not logically contingent upon, the mathematically tenuous notion of white noise.

In Sec. IV we derived the associated  $M$ -variate forward and backward Fokker–Planck equations. Our derivation of the forward Fokker–Planck equation (4.18) proceeded from the Chapman–Kolmogorov equation (4.5) to the forward Kramers–Moyal equation (4.11) in the standard Taylor expansion way, but thereafter it was simplified by the fact that our theorem allows an easy calculation of the  $B$ -functions; e.g., our theorem makes it immediately clear why all derivative terms in the Kramers–Moyal equation of order  $\geq 3$  drop out for a continuous Markov process.<sup>13</sup>

Our analysis also revealed that two different Langevin equations can sometimes describe the *same* continuous Markov process, whereas two different Fokker–Planck equations always describe *different* continuous Markov processes. The underlying reason is that Eqs. (2.5) allow more than one set of  $b_{ij}$  functions to give the same  $D_i$  and  $C_{ij}$  functions. On the other hand, whereas *any* set of smooth  $b_{ij}$  functions will give rise to a legitimate Langevin equation, *only* those sets of smooth  $D_i$  and  $C_{ij}$  functions that satisfy conditions (2.7) will give rise to a legitimate Fokker–Planck equation.

Once the functions  $A_i$  and  $b_{ij}$  have been determined from the physics underlying the process  $\mathbf{X}$ , it may, of course, be very difficult to find even approximate analytical solutions to the resulting Langevin and Fokker–Planck equations. But the Langevin equation itself affords a generally straightforward way of making approximate *numerical simulations* of the process: If the values of the  $M$  components of  $\mathbf{X}$  are known at any time  $t$ , then the Langevin equation (2.6) implies that the component values at the slightly later time  $t + \Delta t$  can be estimated as

$$X_i(t + \Delta t) \doteq X_i(t) + A_i(\mathbf{X}(t), t) \Delta t + \sum_{j=1}^M b_{ij}(\mathbf{X}(t), t) (\Delta t)^{1/2} n_j \quad (i = 1, \dots, M). \quad (6.1)$$

Here,  $n_1, \dots, n_M$  are  $M$  sample values of the random variable  $N(0, 1)$ , which values can easily be generated on a digital computer.<sup>14</sup> The dot over the equal sign reminds us that this is usually a good approximation only if  $\Delta t$  is “suitably small.” In practice, the requirement is that the values of the

functions  $A_i$  and  $b_{ij}$  should not change “appreciably” over the interval  $t$  to  $t + \Delta t$ ; indeed, Eq. (6.1) will be *exact* if the functions  $A_i$  and  $b_{ij}$  are all constants. Equation (6.1) is essentially the stochastic generalization of the well-known Euler integration formula, the innovation evidently being the use of  $M$  unit normal random numbers at each time step.<sup>15</sup>

The view taken here of the Langevin equation as an *updating recipe* which is continuous, memoryless, and self-consistent has some important implications for those who would attempt a “non-Markovian generalization” of the Langevin equation. The fact that the updating recipe for a non-Markovian process is not memoryless does not absolve the recipe of the requirement that it be self-consistent; rather, it makes the self-consistency condition much more complicated: For a non-Markovian stochastic process, the updating recipe must be such that using it to compute the process increment from any given value  $\mathbf{x}$  at time  $t$  to its value at time  $t + dt$  always gives a statistically identical result, at least to lowest order in  $dt$ , as first using the recipe to compute the process value  $\mathbf{x}'$  at time  $t + \frac{1}{2}dt$  given the value  $\mathbf{x}$  at time  $t$ , and then using the recipe to compute the process value at time  $t + dt$  given the *two* values  $\mathbf{x}$  at  $t$  and  $\mathbf{x}'$  at  $t + \frac{1}{2}dt$ . Analogous consistency conditions arise with respect to subdivisions of the interval  $[t, t + dt]$  into three parts, four parts, etc. In the language of the density functions  $P^{(n)}$  in Eqs. (4.1), we are simply saying here that although the density function  $P^{(1)}$  for a non-Markovian stochastic process does not have to satisfy the Chapman–Kolmogorov equation (4.4), it *does* have to satisfy Eq. (4.2) for  $n = 1$ , and that, in turn, requires the specification of an *entire set* of functions  $\{P^{(n)}\}$  that satisfies Eqs. (4.2) for *all*  $n \geq 1$ .

So we see that the generalization of the Langevin equation to non-Markovian stochastic processes is not a task to be undertaken lightly; indeed, even the preliminary task of formally writing down all the consistency conditions that a non-Markovian update formula should satisfy is rather daunting. There is, however, one class of non-Markovian stochastic processes—and it is a very large and important class—for which self-consistency is apparently *not* a problem. This is the class of non-Markovian processes that are *components*, or *functions* of one or more components, of *multivariate Markov* processes.

Perhaps the simplest example of a *component* of a multivariate Markov process that is not itself Markovian is the integral  $Y$  of a univariate continuous Markov process  $X$ , which was discussed in Sec. V A. But since the updating formula (5.1) for that process  $Y$  is part of a bivariate Langevin equation, then there is no questioning the self-consistency of the resulting updating procedure. And while it might well be possible to construct a “generalized Langevin equation” for  $Y$ , i.e., a *single* updating formula that uses present and *past* values of  $Y$  but *no* values of  $X$ , it seems unlikely that such a formula would be more tractable or useful than the *pair* of updating formulas (1.1) and (5.1).

Also, *functions* of continuous Markov processes will sometimes be non-Markovian. For example, the *sine* of the *deterministic* continuous Markov process  $X(t) = t$  (Langevin equation  $dX/dt = 1$ ), although deterministic and continuous, is *not* Markovian. This is because the derivative of  $U(t) \equiv \sin X(t) = \sin t$  is  $\cos t = \pm[1 - \sin^2 t]^{1/2}$ , whence

$$\frac{dU(t)}{dt} = \pm[1 - U^2(t)]^{1/2},$$

and this cannot be regarded as a legitimate Langevin equation

since the sign of the right side changes with time in a way that cannot be determined from the value  $U(t)$  alone. But the procedure of updating  $U$  by first updating  $X$  through its Langevin equation and then taking  $U = \sin X$  should certainly result in a self-consistent process  $U$ ; indeed, that self-consistency can be seen quite explicitly in this case by noting that  $U$  satisfies an unambiguous *second-order* ordinary differential equation, namely,  $d^2U/dt^2 = -U$ , and therefore  $U(t) \equiv \sin t$  and its derivative  $U'(t) \equiv \cos t$  together form a *bivariate* continuous deterministic Markov process.

So, for a non-Markovian process that is either a component of a multivariate continuous Markov process or a function of one or more such components, self-consistency will be *guaranteed* by the self-consistency of the underlying continuous Markov process. But if one proposes to define a non-Markovian process by simply writing down a generalized Langevin equation, then it is incumbent upon one to demonstrate explicitly the self-consistency of that updating formula. The remarks above suggest that such a demonstration will not be easy.

## ACKNOWLEDGMENTS

The author thanks Don Lemons for a helpful observation. This work was sponsored by the Inhouse Laboratory Independent Research Program of the Office of Naval Research.

## APPENDIX: PROOFS OF THE TWO LEMMAS

Here, we prove the two lemmas used in Sec. III in our proof of the theorem at Eqs. (2.4) and (2.5). Lemma 1 allowed us to deduce the critical results (3.8) from Eqs. (3.6). The *converse* of Lemma 1 is trivial to prove, but the proof of the direct lemma is a little more involved.<sup>8</sup>

*Proof of Lemma 1.* Differentiate  $h(z) = nh(z/n)$  with respect to  $z$  to obtain  $h'(z) = nh'(z/n)(1/n) = h'(z/n)$ . Observe that the only way that this can be true for *all* positive integers  $n$ , given that  $h'$  is a continuous function, is to have  $h'(z) = h'(z/\infty) = h'(0) \equiv C$ , where  $C$  is independent of  $z$ . Also note that, since  $h(z/n) = h(z)/n$ , then by taking  $n$  arbitrarily large we may deduce that  $h(0) = 0$ . Finally recognize that the only smooth function  $h$  satisfying both  $h'(z) = C$  and  $h(0) = 0$  is the function  $h(z) = Cz$ . QED

Lemma 2, at Eqs. (3.10) and (3.11), showed that the asserted formula (2.4) for the conditional increment  $\Xi_i$  endows  $\Xi_i$  with precisely the properties demanded by Eqs. (3.5) and (3.8).

*Proof of Lemma 2.* We have

$$\begin{aligned} Y_i &= \alpha_i + \sum_{j=1}^M \beta_{ij} \mathbf{N}(0, 1) \quad [\text{by Eq. (3.10)}], \\ &= \alpha_i + \sum_{j=1}^M \mathbf{N}(0, \beta_{ij}^2) \quad [\text{by Eq. (2.9)}], \\ &= \alpha_i + \mathbf{N}\left(0, \sum_{j=1}^M \beta_{ij}^2\right) \quad [\text{by Eq. (2.15)}], \\ Y_i &= \mathbf{N}\left(\alpha_i, \sum_{j=1}^M \beta_{ij}^2\right) \quad [\text{by Eq. (2.9)}]. \end{aligned}$$

This establishes that the  $Y_i$ 's are indeed normal, and that they have the means and variances asserted by Eqs. (3.11a)

and (3.11b). To establish the asserted covariance formula (3.11c), we write

$$\begin{aligned}\text{cov}\{Y_i, Y_j\} &= \langle (Y_i - \langle Y_i \rangle)(Y_j - \langle Y_j \rangle) \rangle \quad [\text{by definition}], \\ &= \langle (Y_i - \alpha_i)(Y_j - \alpha_j) \rangle \quad [\text{by Eq.(3.11a)}], \\ &= \langle (\sum_k \beta_{ik} N_k)(\sum_l \beta_{jl} N_l) \rangle \quad [\text{by Eq.(3.10)}],\end{aligned}$$

$$\text{cov}\{Y_i, Y_j\} = \sum_k \sum_l \beta_{ik} \beta_{jl} \langle N_k N_l \rangle.$$

But since the  $N_i$ 's are statistically independent random variables with means 0 and variances 1, then  $\langle N_k N_l \rangle = \delta_{kl}$ , and the last equation reduces to Eq. (3.11c). QED

<sup>a</sup>e-mail: dtg@rattler.chinalake.navy.mil

<sup>1</sup>D. T. Gillespie, "The mathematics of Brownian motion and Johnson noise," *Am. J. Phys.* **64**, 225–240 (1996).

<sup>2</sup>There is an incorrect statement in Sec. II B of Ref. 1. It is *not* true, as is stated there just after Eq. (2.11), that of all random variables with finite means and variances *only* the normal preserves its class under statistically independent addition. Two other random variables that do likewise are the gamma and the chi-squared. While it is probably fair to say that "most" random variables do *not* have that property, there are apparently infinitely many that do: If one defines the random variable  $Y(n) \equiv X_1 + \dots + X_n$  where the  $X_i$ 's are statistically independent copies of some random variable  $X$  with a finite mean and variance, then clearly  $Y(n) + Y(m) = Y(n+m)$ , which shows that  $Y(n)$  preserves its class under statistically independent addition. But it does follow from the central limit theorem that, for *any* such  $X$ ,  $Y(n \rightarrow \infty)$  must be normal.

<sup>3</sup>Those who are bothered by the  $(dt)^{1/2}$  in Eq. (1.1) may need reminding that the square root operation is well defined for *all* real nonnegative numbers, no matter how "small" they may be; e.g., the square root of  $10^{-100}$  is  $10^{-50}$ , and the square root of 0 is 0. The reason why the  $A$ -term in Eq. (1.1) is not rendered negligible by the usually much larger  $D$ -term is discussed in Ref. 1, Sec. II C.

<sup>4</sup>N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992), 2nd ed.

<sup>5</sup>C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer-Verlag, New York, 1985).

<sup>6</sup>Ming Chen Wang and G. E. Uhlenbeck, "On the theory of Brownian motion II," *Rev. Mod. Phys.* **17**, 323–342 (1945).

<sup>7</sup>As in Ref. 1,  $dt$  is to be regarded as an ordinary real variable whose domain is restricted to some arbitrarily small interval  $[0, \epsilon]$ . Also, when possible we use upper and lower case letters to distinguish between a *random variable*  $X$  and its *possible values*  $x$ .

<sup>8</sup>D. T. Gillespie, *Markov Processes: An Introduction for Physical Scientists* (Academic, San Diego, 1992).

<sup>9</sup>See in Ref. 1 specifically Eqs. (3.3), (3.16), and (3.17).

<sup>10</sup>See Ref. 5, p. 156.

<sup>11</sup>S. Chandrasekhar, "Stochastic problems in physics and astronomy," *Rev. Mod. Phys.* **15**, 1–89 (1943); see especially pp. 41 and 57.

<sup>12</sup>See Ref. 5, pp. 196ff, for a detailed discussion of this point and literature references.

<sup>13</sup>Such a truncation of the infinite-term Kramers–Moyal equation does *not* occur for a *jump* Markov process; see, e.g., Ref. 8, Chap. 4.

<sup>14</sup>See, e.g., Ref. 1, Eqs. (2.48).

<sup>15</sup>Valid higher-order-in- $\Delta t$  improvements on formula (6.1) for nonconstant  $A_i$  and  $b_{ij}$  can be tricky to obtain; for an inconclusive discussion in the univariate case, see Ref. 8, Appendix E.

## Simple demonstrations for introducing spatial coherence

Lorenzo Basano and Carlo Pontiggia

*Università di Genova, Dipartimento di Fisica, Via Dodecaneso 33, 16146 Genova, Italy*

Emanuele Piano

*Università di Genova, Istituto di Fisica di Ingegneria, Piazzale Kennedy, Pad. D, 16129 Genova, Italy*

(Received 26 July 1995; accepted 4 February 1996)

When two quasiparallel identical beams, originating from the same laser, are superposed, the crossing volume becomes the seat of a system of interference fringes. The spatial coherence of the radiation filling this region may be evaluated by observing the interference pattern generated on a far screen. When the beams originate from two identical but *different* lasers, the fringe system in the crossing region will be washed out by mutual phase fluctuations: *the radiation field, however, retains its ability to produce an interference pattern* when it goes through a double slit. This remark may be the starting point for a simpler and intuitive approach to the van Cittert–Zernike theorem. Simple laboratory experiments are described which illustrate these results. © 1996 American Association of Physics Teachers.

### I. INTRODUCTION

This work describes an optical experiment devised to illustrate, in a simple way, the concepts of spatial coherence and field correlation for a quasimonochromatic radiation that was, therefore, not perfectly coherent in time; when required, the field correlation was measured through the phenomenon of interference from a double slit arrangement, whose purpose was to sample a pair of different portions of the radiation field we wanted to correlate.

We will consider two examples of light fields that can be simply described analytically and are also easy to implement experimentally, namely, the superposition of two plane waves of equal frequencies; the cases of coherent and incoherent superposition are considered separately.

For completeness, we recall that the expression for a polarized plane wave is

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t - \phi)}. \quad (1)$$