The chemical Langevin equation

Daniel T. Gillespie

Citation: The Journal of Chemical Physics **113**, 297 (2000); doi: 10.1063/1.481811
View online: https://doi.org/10.1063/1.481811
View Table of Contents: http://aip.scitation.org/toc/jcp/113/1
Published by the American Institute of Physics

The chemical Langevin equation

Daniel T. Gillespie[a)]
Research Department, Code 4T4100D, Naval Air Warfare Center, China Lake, California 93555

(Received 6 December 1999; accepted 5 April 2000)

The stochastic dynamical behavior of a well-stirred mixture of N molecular species that chemically interact through M reaction channels is accurately described by the chemical master equation. It is shown here that, whenever two explicit dynamical conditions are satisfied, the microphysical premise from which the chemical master equation is derived leads directly to an approximate time-evolution equation of the Langevin type. This chemical Langevin equation is the same as one studied earlier by Kurtz, in contradistinction to some other earlier proposed forms that assume a deterministic macroscopic evolution law. The novel aspect of the present analysis is that it shows that the accuracy of the equation depends on the satisfaction of certain specific conditions that can change from moment to moment, rather than on a static system size parameter. The derivation affords a new perspective on the origin and magnitude of noise in a chemically reacting system. It also clarifies the connection between the stochastically correct chemical master equation, and the deterministic but often satisfactory reaction rate equation. [S0021-9606(00)50925-8]

I. INTRODUCTION

We consider here a well-stirred mixture of $N \geqslant 1$ molecular species $\{S_1,\ldots,S_N\}$ that chemically interact, inside some fixed volume Ω and at constant temperature, through $M \geqslant 1$ reaction channels $\{R_1,\ldots,R_M\}$. We specify the dynamical state of this system by $\mathbf{X}(t) \equiv (X_1(t),\ldots,X_N(t))$, where

$$X_i(t) \equiv \text{the number of } S_i \text{ molecules in the system at time } t \quad (i=1,\ldots,N). \tag{1}$$

Our goal is to describe the evolution of $\mathbf{X}(t)$ from some given initial state $\mathbf{X}(t_0)=\mathbf{x}_0$. (All boldface vectors in this paper are species indexed, with N components.)

The molecular populations $X_i(t)$ will actually be *random* variables, because we choose not to track the positions and velocities of all the molecules in the system. Indeed, we intend to rely heavily on the occurrence of many nonreactive molecular collisions to ''stir'' the system between successive reactive collisions. Under these conditions, it can be shown that there will exist for each reaction channel R_j a well-defined function a_j, henceforth referred to as the *propensity function* for R_j, which is such that

$$a_j(\mathbf{x})\,dt \equiv \text{the probability, given } \mathbf{X}(t)=\mathbf{x}, \text{ that one } R_j \text{ reaction will occur somewhere inside } \Omega \text{ in the next infinitesimal time interval } [t,t+dt) \quad (j=1,\ldots,M). \tag{2}$$

The propensity function a_j and the *state-change vector* $\boldsymbol{\nu}_j$, whose ith component is defined by

$$\nu_{ji} \equiv \text{the change in the number of } S_i \text{ molecules produced by one } R_j \text{ reaction} \quad (j=1,\ldots,M;\, i=1,\ldots,N), \tag{3}$$

together completely specify the reaction channel R_j.

Equation (2) provides the logical basis for ''stochastic chemical kinetics,'' as exemplified for instance by the chemical master equation[1] and the stochastic simulation algorithm.[2] It will be shown in this paper that Eq. (2) also leads directly to an approximating chemical Langevin equation, and hence also an approximating chemical Fokker–Planck equation.

Since all of our arguments will be based on premise (2), we shall begin in Sec. II by briefly reviewing the rationale for regarding that premise to be firmly grounded in kinetic theory. In Sec. III we shall review, for later reference, how the chemical master equation and several other important results follow from premise (2), either rigorously or heuristically. In Sec. IV we develop our main result: We show that, provided two dynamical conditions are satisfied, premise (2) implies an approximate Langevin equation for $\mathbf{X}(t)$, and hence also an approximate Fokker–Planck equation. We highlight some salient implications of this result in Sec. V. We conclude in Sec. VI by discussing the connection between our analysis and earlier works on this subject, noting in particular how our analysis seems to mollify some earlier objections to the use of Langevin and Fokker–Planck equations in chemical kinetics.

II. THE PROPENSITY FUNCTION

Early discussions of premise (2) regarded it as a kind of *ad hoc* stochastization of deterministic chemical kinetics,[3] but it was later demonstrated that Eq. (2) actually has a solid microphysical basis.[4] Its derivation, although not proceeding directly from the Liouville equation, uses arguments whose

a)Electronic mail: GillespieDT@navair.navy.mil

rigor is quite on par with the *Stosszahlansatz*[5] that underlies the Boltzmann transport equation. Recent numerical studies comparing the predictions of the chemical master equation with those of molecular dynamics simulations[6] have only confirmed the validity of the theoretical arguments underlying Eq. (2).

The derivation of Eq. (2) for a *bimolecular* gas-phase reaction is superficially similar to the elementary kinetic theory derivation of the mean bimolecular collision frequency; however, the former has a higher degree of rigor because it deals with probabilities instead of average numbers, and so can make use of the well-established laws of probability theory. The function a_j in Eq. (2) is found to have the mathematical form[2,4,7]

$$a_j(\mathbf{x}) = c_j h_j(\mathbf{x}). \tag{4}$$

Here, c_j is the *specific probability rate constant* for channel R_j, defined so that $c_j\,dt$ gives the probability that a randomly chosen pair of R_j reactant molecules will react accordingly in the next infinitesimal time interval dt. This probability $c_j\,dt$ is in turn equal to the probability that a randomly chosen combination of R_j reactant molecules will *collide* in the next dt, multiplied by the probability that a *colliding* pair of reactant molecules will actually *react* according to R_j. The first (collision) probability is found to be directly proportional to the average relative speed and collision cross section of a reactant pair, and inversely proportional to the total system volume Ω. The second (collision-conditioned reaction) probability is typically calculated as the probability that the collision energy will exceed a certain threshold value; it turns out to be a factor with the familiar Arrhenius exponential form. The final expression for the bimolecular c_j turns out to be the same as that for the conventional bimolecular ''reaction rate constant'' k_j, except divided by Ω, and also multiplied by 2 if the reactant species happen to be identical.

The function $h_j(\mathbf{x})$ in Eq. (4) is defined to be the number of distinct combinations of R_j reactant molecules available in the state $\mathbf{x}$. It can be easily deduced by simply inspecting the left hand side of reaction R_j. So, for example, if R_1 were the reaction $X_1 + X_2 \rightarrow 2X_1$ we would have $a_1(\mathbf{x}) = c_1 x_1 x_2$, and if R_2 were the inverse of that reaction we would have $a_2(\mathbf{x}) = c_2 x_1(x_1-1)/2$. [Note that, for these two hypothetical reactions, we would have from Eq. (3) that $\boldsymbol{\nu}_1 = (+1, -1, 0, \ldots, 0)$ and $\boldsymbol{\nu}_2 = -\boldsymbol{\nu}_1$.]

If R_j is a *monomolecular* reaction, Eq. (4) still applies, but c_j will then be equal to some quantum mechanically determined ''decay probability'' per unit time; c_j will then be independent of Ω, and in fact equal to the conventional monomolecular reaction rate constant k_j. The function $h_j(\mathbf{x})$ will be simply the population of the (single) reactant species. Thus, if R_1 were the reaction $S_2 \rightarrow S_3$ we would have $a_1(\mathbf{x}) = c_1 x_2$.

Genuinely *trimolecular* reactions do not physically occur in dilute fluids with any appreciable frequency. *Apparently* trimolecular reactions in a fluid are usually the combined result of two bimolecular reactions and one monomolecular reaction, and involve an additional short-lived species. For such an ''effectively trimolecular'' reaction, it can be shown[7] that Eq. (4) still applies, but with an *approximate* c_j that is proportional to Ω^{-2}.

In virtually all time-evolution equations, c_j and h_j occur together in the product from (4). That is why it is convenient to designate that product by a special symbol, a_j, and to give it a special name, the propensity function (it evidently measures the propensity for reaction R_j to occur in the next moment).

III. SOME NOTABLE CONSEQUENCES OF EQ. (2)

The evolution law (2) implies that the state vector $\mathbf{X}(t)$ is a jump-type Markov process on the non-negative N-dimensional integer lattice. The traditional way of analyzing such a process is to focus on its singly conditioned probability function,

$$P(\mathbf{x},t|\mathbf{x}_0,t_0) \equiv \text{Prob}\{\mathbf{X}(t)=\mathbf{x}, \text{ given that } \mathbf{X}(t_0)=\mathbf{x}_0\}. \tag{5}$$

(Whenever possible, we use an upper case letter to denote a random variable, and the corresponding lower case letter to denote a possible value of that random variable.)

To derive a time evolution equation for this function, we take a time increment dt that is so small that the probability for two or more reactions to occur in dt is negligible compared to the probability for only one reaction. We then use Eq. (2) and the addition and multiplication laws of probability theory to write the probability of the system being in state $\mathbf{x}$ at time $t+dt$ as the sum of the probabilities of all the mutually exclusive ways in which that can happen via either zero or one reaction in $[t,t+dt)$:

$$P(\mathbf{x},t+dt|\mathbf{x}_0,t_0) = P(\mathbf{x},t|\mathbf{x}_0,t_0)\times\left[1-\sum_{j=1}^{M} a_j(\mathbf{x})dt\right] + \sum_{j=1}^{M}\left[P(\mathbf{x}-\boldsymbol{\nu}_j,t|\mathbf{x}_0,t_0)a_j(\mathbf{x}-\boldsymbol{\nu}_j)dt\right].$$

A few simple algebraic rearrangements and a passage to the limit $dt \rightarrow 0$ then yields the *chemical master equation*:[1,4]

$$\frac{\partial}{\partial t}P(\mathbf{x},t|\mathbf{x}_0,t_0) = \sum_{j=1}^{M}\left[a_j(\mathbf{x}-\boldsymbol{\nu}_j)P(\mathbf{x}-\boldsymbol{\nu}_j,t|\mathbf{x}_0,t_0) - a_j(\mathbf{x})P(\mathbf{x},t|\mathbf{x}_0,t_0)\right]. \tag{6}$$

Equation (6) is an exact consequence of Eq. (2). If it can be solved for P, then we can in principle find out everything there is to know about the process $\mathbf{X}(t)$. But in practice, exact solutions of Eq. (6) can rarely be obtained.

Another exact consequence of Eq. (2) is the existence and form of the *next-reaction density function* $p(\tau,j|\mathbf{x},t)$,[2] which is defined by

$$p(\tau,j|\mathbf{x},t)\,d\tau \equiv \text{probability that, given } \mathbf{X}(t)=\mathbf{x}, \text{ the next reaction in } \Omega \text{ will occur in the infinitesimal time interval } [t+\tau,t+\tau+d\tau), \text{ and will be an } R_j \text{ reaction.} \tag{7}$$

Since $\Sigma_j a_j(\mathbf{x})dt$ is the probability that *some* reaction will occur in the next dt, then an elementary probability argu-

ment shows that $\exp(\Sigma_j a_j(\mathbf{x})\tau)$ is the probability that a time τ will elapse without any reaction occurring. This probability multiplied by the probability in Eq. (2) then gives the probability in Eq. (7); thus, the function p is given by

$$p(\tau,j|\mathbf{x},t)=a_j(\mathbf{x})\exp\left(\sum_{k=1}^{M}a_k(\mathbf{x})\tau\right)$$
$$(0\leqslant\tau<\infty;j=1,...,M). \tag{8}$$

This formula provides the basis for the *stochastic simulation algorithm*,[2] in which one uses rigorous Monte Carlo techniques to generate random pairs (τ,j) according to the joint density function (8); those random pairs in turn allow one to construct ''unbiased realizations'' of the process $\mathbf{X}(t)$. Such realizations are fully consistent with the chemical master equation (6), since Eqs. (6) and (8) are both exact consequences of premise (2). But note that the stochastic simulation algorithm is *not* a procedure for numerically solving the chemical master equation; that is usually a vastly more challenging task.

If Eqs. (6) and (8) are exact consequences of the dynamical premise (2), a ''semiexact'' consequence is the chemical Kramers–Moyal equation. Its somewhat shady lineage arises from the fact that it requires us to regard the components of $\mathbf{X}(t)$ as real numbers, instead of the integers that Nature stipulates. This is not an entirely innocuous presumption, since the propensity functions usually do not make sense when their arguments are nonintegers, especially such between 0 and 1. Intuitively, we expect that this real number assumption should be warranted if all the components of $\mathbf{X}(t)$ are very large compared to 1. But that just shows that we are moving here into the realm of a ''large number'' approximation. A second assumption we have to make for this calculation is that the function $f_j(\mathbf{x})\equiv a_j(\mathbf{x})P(\mathbf{x},t|\mathbf{x}_0,t_0)$ must be analytic (infinitely differentiable) in the real variable $\mathbf{x}$. Given these two assumptions, we can use Taylor's theorem to write

$$f_j(\mathbf{x}-\boldsymbol{\nu}_j)=f_j(\mathbf{x})+\sum_{n=1}^{\infty}\sum_{\substack{m_1,...,m_N=0\\[m_1+\cdots+m_N=n]}}^{n}\frac{1}{m_1!\cdots m_N!}$$
$$\times(-\nu_{j1})^{m_1}\cdots(-\nu_{jN})^{m_N}\frac{\partial^n f_j(\mathbf{x})}{\partial x_1^{m_1}\cdots\partial x_N^{m_N}}.$$

Substituting this into the master equation (6), reordering the summations, and rearranging a few factors, we immediately obtain the *chemical Kramers–Moyal equation*,

$$\frac{\partial}{\partial t}P(\mathbf{x},t|\mathbf{x}_0,t_0)$$
$$=\sum_{n=1}^{\infty}(-1)^n\sum_{\substack{m_1,...,m_N=0\\[m_1+\cdots+m_N=n]}}^{n}\frac{1}{m_1!\cdots m_N!}\frac{\partial^n}{\partial x_1^{m_1}\cdots\partial x_N^{m_N}}$$
$$\times\left\{\left[\sum_{j=1}^{M}(\nu_{j1}^{m_1}\cdots\nu_{jN}^{m_N})a_j(\mathbf{x})\right]P(\mathbf{x},t|\mathbf{x}_0,t_0)\right\}. \tag{9}$$

We shall return to this ''semirigorous'' consequence of premise (2) in Sec. VI.

If we multiply the chemical master equation (6) through by x_i, sum over all $\mathbf{x}$, and then re-index the first summation on the right-hand side, keeping in mind that the propensity functions will ensure that no component of the process can ever become negative, we can deduce the following formula for the time derivative of the *mean* of $X_i(t)$:

$$\frac{d\langle X_i(t)\rangle}{dt}=\sum_{j=1}^{M}\nu_{ji}\langle a_j(\mathbf{X}(t))\rangle\quad(i=1,...,N). \tag{10}$$

This *exact* consequence of Eq. (2) is sometimes taken as heuristic justification for assuming that, whenever fluctuations are not important, the species populations evolve *deterministically* according to the set of ordinary differential equations

$$\frac{dX_i(t)}{dt}=\sum_{j=1}^{M}\nu_{ji}a_j(\mathbf{X}(t))\quad(i=1,...,N), \tag{11}$$

where the components of $\mathbf{X}(t)$ are now regarded as *sure*, *real* variables. Equation (11) is essentially the macroscopic *reaction rate equation* of conventional chemical kinetics. But the reasoning we have used here to infer it obviously is not an honest derivation. The legitimacy of Eq. (11) would appear to require, at the very least, that the populations of all species be very large compared to 1.

The reaction rate equation (11) is more commonly written in terms of the species concentrations,

$$Z_i(t)\equiv X_i(t)/\Omega\quad(i=1,...,N). \tag{12}$$

If we examine the Ω dependence of Eq. (4) for monomolecular, bimolecular, and trimolecular reactions (see the discussion in Sec. II), we find that in each case we have, at least *approximately* for sufficiently large molecular population numbers,

$$a_j(\mathbf{x})=\Omega\tilde{a}_j(\mathbf{z}). \tag{13}$$

Here, $\tilde{a}_j$ is functionally identical to a_j except that the reaction probability rate constant c_j has been replaced by the reaction rate constant k_j, and any combinatorial factors arising from identical species are omitted. When Eq. (13) is substituted into Eq. (11), we obtain the more familiar ''concentration'' form of the reaction rate equation:

$$\frac{dZ_i(t)}{dt}=\sum_{j=1}^{M}\nu_{ji}\tilde{a}_j(\mathbf{Z}(t))\quad(i=1,...,N). \tag{14}$$

From a theoretical standpoint though, the difference between the two versions (14) and (11) of the deterministic reaction rate equation is minor. Neither form, at this stage anyway, can claim to be a rigorous consequence of Eq. (2).

Finally, we must take note of the intriguing effect that the variable change (12) and the scaling relation (13) has on the chemical Kramers–Moyal equation (9). Since the singly conditioned density function P of the (assumed real) random variables $\mathbf{X}(t)$ will be related to that of the random variables $\mathbf{Z}(t)$ in Eq. (12) by $\tilde{P}(\mathbf{z})=\Omega^N P(\mathbf{x})$, we multiply both sides of Eq. (9) by Ω^N to convert the P's to $\tilde{P}$'s. Then we complete the $x_i\rightarrow z_i\equiv x_i/\Omega$ transformation on the right-hand side of Eq. (9) by invoking Eq. (13) and taking note of the fact that the m_k's sum to n. The result is[8–10]

$$\frac{\partial}{\partial t}\tilde{P}(\mathbf{z},t|\mathbf{z}_0,t_0)$$

$$=\sum_{n=1}^{\infty}(-1)^n\left(\frac{1}{\Omega}\right)^{n-1}$$

$$\times\sum_{\substack{m_1,\ldots,m_N=0\\[m_1+\cdots+m_N=n]}}^{n}\frac{1}{m_1!\cdots m_N!}\frac{\partial^n}{\partial z_1^{m_1}\cdots\partial z_N^{m_N}}$$

$$\times\left\{\left[\sum_{j=1}^{M}(\nu_{j1}^{m_1}\cdots\nu_{jN}^{m_N})\tilde{a}_j(\mathbf{z})\right]\tilde{P}(\mathbf{z},t|\mathbf{z}_0,t_0)\right\}. \tag{15}$$

What is intriguing about this ''concentration'' form of the Kramers–Moyal expansion is this: If we go to the *thermodynamic limit*, in which the number of molecules in the system and the system volume Ω both go to ∞ in such a way that the species concentrations remain constant (a feat that theoreticians perform much more casually than either experimentalists or numerical simulators), the factors of Ω^{-1} tend to kill off the higher terms in Eq. (15). If Ω can be taken so large that only the $n=1$ term need be retained, then the resulting equation for $\tilde{P}$ would imply that the process $\mathbf{Z}(t)$ evolves according to the *deterministic* evolution equation (14).[9,10] A slightly less drastic truncation at the $n=2$ term would yield a Fokker–Planck equation, implying that $\mathbf{Z}(t)$ is approximately a continuous (or diffusion) Markov process; that approximation might be computationally convenient, since the Fokker–Planck equations that govern *continuous* Markov processes are generally more tractable[11] than the master equations that govern *jump* Markov processes. We shall return to these points later in Sec. VI.

We have enumerated here several known consequences of the fundamental premise (2), some of which are rigorous and exact, and some of which are not. We turn now to develop a new consequence of Eq. (2), one whose validity will be seen to depend entirely on the degree to which two specific dynamical conditions are satisfied.

IV. AN APPROXIMATING CHEMICAL LANGEVIN EQUATION

Suppose the system's state $\mathbf{X}(t)$ at the current time t is known to be $\mathbf{x}_t$. Let $K_j(\mathbf{x}_t,\tau)$, for any $\tau>0$, be the number of R_j reactions that occur in the subsequent time interval $[t,t+\tau]$. Since each of those reactions will increase the S_i population by ν_{ji}, the number of S_i molecules in the system at time $t+\tau$ will be

$$X_i(t+\tau)=x_{ti}+\sum_{j=1}^{M}K_j(\mathbf{x}_t,\tau)\nu_{ji}\quad(i=1,\ldots,N). \tag{16}$$

$K_j(\mathbf{x}_t,\tau)$ is of course a *random variable*. To compute it for arbitrary $\tau>0$ would be quite as difficult as solving the master equation (6). But we can obtain an excellent *approximation* to $K_j(\mathbf{x}_t,\tau)$ rather easily if we impose the following conditions.

Condition (i): Require τ to be *small* enough that the change in the state during $[t,t+\tau]$ will be so slight that none of the propensity functions changes its value ''appreciably;'' i.e., the propensity functions then satisfy

$$a_j(\mathbf{X}(t'))\cong a_j(\mathbf{x}_t),\quad\forall t'\in[t,t+\tau],\quad\forall j\in[1,M]. \tag{17}$$

Of course, we could easily satisfy Eq. (17) if we simply took τ to be so small that *zero* reactions would be likely to occur in $[t,t+\tau]$. But usually we need not be that extreme. Since the propensity functions generally depend on nonnegative integer powers of the molecular species populations, and the latter practically never change by more than two molecules in any one reaction event, then we need only require that the likely number of reactions occurring in $[t,t+\tau]$ be much smaller than the population of the least populous reactant species. So condition (i) can *always* be satisfied if all the reactant molecule populations are sufficiently large compared to 1.

Since the reactions that occur in the time interval $[t,t+\tau]$ do not appreciably change the values of any of the propensity functions, then all reaction events occurring in the time interval $[t,t+\tau]$ will be essentially *independent* of each other. So $K_j(\mathbf{x}_t,\tau)$ will simply be the number of times reaction channel R_j would occur in a duration τ if its propensity function remained constant at the value $a_j(\mathbf{x}_t)$. In light of our fundamental premise (2), this means, as is shown in Appendix A, that each $K_j(\mathbf{x}_t,\tau)$ will be a *statistically independent Poisson* random variable, $\mathcal{P}_j(a_j(\mathbf{x}_t),\tau)$. So the net effect of condition (i) is to allow Eq. (16) to be approximated by

$$X_i(t+\tau)=x_{ti}+\sum_{j=1}^{M}\nu_{ji}\mathcal{P}_j(a_j(\mathbf{x}_t),\tau)\quad(i=1,\ldots,N). \tag{18}$$

Next we impose a second condition on τ.

Condition (ii): Require τ to be *large* enough that the expected number of occurrences of each reaction channel R_j in $[t,t+\tau]$ be much larger than 1, i.e., (see Appendix A)

$$\langle\mathcal{P}_j(a_j(\mathbf{x}_t)),\tau)\rangle=a_j(\mathbf{x}_t)\tau\gg1,\quad\forall j\in[1,M]. \tag{19}$$

This condition obviously runs counter to condition (i), and it may very well happen that both conditions cannot be satisfied simultaneously; in that case, the scheme being devised here will fail. But there will be many practical circumstances in which conditions (i) and (ii) can be simultaneously satisfied. As with condition (i), satisfaction of condition (ii) is greatly facilitated by having large molecular population numbers: Since $a_j(\mathbf{x}_t)$, through the combinatorial function $h_j(\mathbf{x}_t)$, will typically be proportional to one or more components of $\mathbf{x}_t$, sufficiently large molecular populations can make it is possible for the inequalities (19) to hold even when τ is quite small.

As explained in Appendix A, condition (ii), or more precisely the inequality (19), allows us to *approximate* each

Poisson random variable $\mathcal{P}_j(a_j(\mathbf{x}_t),\tau)$ by a *normal* random variable with the same mean and variance. That brings Eq. (18) into the form

$$X_i(t+\tau)=x_{ti}+\sum_{j=1}^{M}\nu_{ji}\mathcal{N}_j(a_j(\mathbf{x}_t)\tau,a_j(\mathbf{x}_t)\tau)\quad(i=1,...,N),\tag{20}$$

where $\mathcal{N}(m,\sigma^2)$ denotes the normal random variable with mean m and variance σ^2. Notice that in replacing the *integer Poisson* random variables in Eq. (18) by the *real normal* random variables in Eq. (20), we in effect convert the molecular populations X_i from discretely changing integer variables to continuously changing real variables. Notice also that the M normals $\mathcal{N}_j$ in Eq. (20) will be *statistically independent*, this in consequence of the statistical independence of the M Poisson's $\mathcal{P}_j$ in Eq. (18).

The linear combination theorem for normal random variables,

$$\mathcal{N}(m,\sigma^2)=m+\sigma\mathcal{N}(0,1),$$

can now be invoked to bring Eq. (20) into the form

$$X_i(t+\tau)=x_{ti}+\sum_{j=1}^{M}\nu_{ji}a_j(\mathbf{x}_t)\tau+\sum_{j=1}^{M}\nu_{ji}[a_j(\mathbf{x}_t)\tau]^{1/2}\mathcal{N}_j(0,1)\quad(i=1,...,N),\tag{21}$$

where again the M normals $\mathcal{N}_j(0,1)$ are all statistically independent. We conclude by making some purely notational changes in this result: First, let us regard any time interval τ that satisfies *both* conditions (i) and (ii) as a *macroscopic infinitesimal*, and denote it simply by dt. Second, let us write the ''unit normal'' random variable $\mathcal{N}_j(0,1)$ as $\mathcal{N}_j(t)$, with the understanding that $N_j(t)$ and $N_{j'}(t')$ will be statistically independent if either $j\neq j'$ or $t\neq t'$. And finally, let us recall that $\mathbf{x}_t$ stands for $\mathbf{X}(t)$. Equation (21) becomes

$$X_i(t+dt)=X_i(t)+\sum_{j=1}^{M}\nu_{ji}a_j(\mathbf{X}(t))dt+\sum_{j=1}^{M}\nu_{ji}a_j^{1/2}(\mathbf{X}(t))N_j(t)(dt)^{1/2}\quad(i=1,...,N).\tag{22}$$

As explained in Appendix B, Eq. (22) has the canonical form of a ''standard-form Langevin equation'' for a multivariate continuous Markov process. Equation (22) thus implies the equivalent ''white-noise form'' Langevin equation

$$\frac{dX_i(t)}{dt}=\sum_{j=1}^{M}\nu_{ji}a_j(\mathbf{X}(t))+\sum_{j=1}^{M}\nu_{ji}a_j^{1/2}(\mathbf{X}(t))\Gamma_j(t)\quad(i=1,...,N),\tag{23}$$

where the $\Gamma_j(t)$ are temporally uncorrelated, statistically independent Gaussian white noises. Equation (22) also implies (see Appendix B) that the singly conditioned density function for $\mathbf{X}(t)$ obeys the (forward) Fokker–Planck equation

$$\frac{\partial}{\partial t}P(\mathbf{x},t|\mathbf{x}_0,t_0)=-\sum_{i=1}^{N}\frac{\partial}{\partial x_i}\left[\left(\sum_{j=1}^{M}\nu_{ji}a_j(\mathbf{x})\right)P(\mathbf{x},t|\mathbf{x}_0,t_0\right]+\frac{1}{2}\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2}\left[\left(\sum_{j=1}^{M}\nu_{ji}^2a_j(\mathbf{x})\right)P(\mathbf{x},t|\mathbf{x}_0,t_0)\right]+\sum_{\substack{i,i'=1\\ i<i'}}^{N}\frac{\partial^2}{\partial x_i\partial x_{i'}}\left[\left(\sum_{j=1}^{M}\nu_{ji}\nu_{ji'}a_j(\mathbf{x})\right)P(\mathbf{x},t|\mathbf{x}_0,t_0)\right].\tag{24}$$

V. SOME SALIENT IMPLICATIONS

Equation (22) was derived from Eq. (2) subject only to conditions (i) and (ii). Condition (i) allowed us to approximate the generally inscrutable random variables $K_j(\mathbf{x}_t,\tau)$ in Eq. (16) by the much more benign Poisson random variables $\mathcal{P}_j(a_j(\mathbf{x}_t),\tau)$, and condition (ii) allowed us to approximate those in turn by the normal random variables $\mathcal{N}_j(a_j(\mathbf{x}_t)\tau,a_j(\mathbf{x}_t)\tau)$. The immediate result of those two approximations was shown to be Eq. (22). Since it has the canonical form of the standard-form Langevin equation, it implies that the *jump* Markov process $\mathbf{X}(t)$ defined by the master equation (6) has been *approximated* as a *continuous* Markov process. The white-noise form Langevin equation (23) and the Fokker–Planck equation (24) then follow as straightforward consequences of Eq. (22) via continuous Markov process theory.[12]

The requirements that are imposed by conditions (i) and (ii) can be summarized as follows: The system must be such that there exists a *domain of macroscopically infinitesimal time intervals*, which is defined so that during any time interval dt in that domain, no propensity function will suffer a noticeable change in its value, yet every reaction channel can be expected to fire many more times than once. The validity of Eqs. (22)–(24) is solely contingent on the existence of such a domain, and of course on our willingness to confine our ''infinitesimal time increments'' to that domain. Large molecular populations for the reactant species will normally be conducive to the existence of a macroscopically infinitesimal time scale, but it is the existence of that time scale, and not the size of the system per se, that ultimately validates the continuous approximation underlying Eqs. (22)–(24).

The imposition of a lower limit on the infinitesimal dt, which is the effect of condition (ii), is not at all unprecedented in physics. It has obvious close ties to the ''coarse-grained time'' arguments used by Einstein[13] in his pioneering analysis of Brownian motion, although a careful examination of those arguments will reveal that they are actually much more heuristic than those employed here. In electromagnetic theory, electrical current is defined as the ratio of dq, the charge passing in time dt, to dt, in the limit that $dt\to 0$. But it is tacitly understood that the limit $dt\to 0$ is

''macroscopic'' in the sense that dt is not allowed to become so small that shot noise effects due to the quantized nature of electrical charge become important. In practice this constraint usually poses no problem: We usually *can* find a domain of macroscopically infinitesimal time intervals relative to which the definition of electrical current as the ''macroscopic derivative'' dq/dt turns out to be quite serviceable. A similar lower limit on the infinitesimal of length dl would be required to define the ''mass per unit length'' dm/dl of a metal wire, this in order to accommodate the atomistic structure of the metal.

When conditions (i) and (ii) are satisfied, the result (22) shows that each reaction channel R_j contributes to the dt-increment $X_i(t+dt)-X_i(t)$ in the population of species S_i a *deterministic* component $\nu_{ji}a_j(\mathbf{X}(t))dt$ and a zero-mean *randomly fluctuating* component $\nu_{ji}a_j^{1/2}(\mathbf{X}(t))N_j(t)(dt)^{1/2}$. To compare the expected sizes of these two components, we must first recall from continuous Markov process theory[12] that, over a *succession* of small time steps dt, the cumulative contribution of the factor $N_j(t)(dt)^{1/2}$ in the random component will be comparable to that of the factor dt in the deterministic component; in effect, the imbalance $(dt)^{1/2}\gg dt$ gets compensated over a succession of steps by the frequent changes in sign of the sample values of $N_j(t)=\mathcal{N}(0,1)$. It follows that the *ratio* of {the size of the *random* component of R_j's contribution to the dt increments} to {the size of the *deterministic* component of R_j's contribution to the dt increments} is simply $a_j^{1/2}(\mathbf{X}(t))/a_j(\mathbf{X}(t))$, or $a_j^{-1/2}(\mathbf{X}(t))$. Recalling from Sec. II that $a_j(\mathbf{X}(t))$ is, owing to the combinatorial function h_j in Eq. (4), directly proportional to products of the molecular populations of the R_j reactant species, we may conclude that *the relative fluctuations in the time-evolving species populations scale as the inverse square root of the reactant populations.* It is satisfying to see how this well-known rule-of-thumb in chemical kinetics emerges as a direct consequence of premise (2). In fact, a careful examination of our derivation in Sec. IV [see Eq. (18)] reveals that this quantitative relation between deterministic drift and random noise in a chemically reacting system follows simply from the fact that the standard deviation of a Poisson random variable is always equal to the square root of its mean.

An extrapolation of the foregoing line of reasoning to *infinitely* large reactant molecular populations shows that, in the thermodynamic limit, the contributions to the changes in the species populations from the random terms in Eqs. (22) and (23) will usually become *vanishingly small* compared to the contributions from the deterministic terms. In that case, Eq. (23) reduces straightaway to the conventional reaction rate equation (11). So we now have a direct logical linking of that cornerstone formula of conventional chemical kinetics to the fundamental stochastic premise (2). But it must be kept in mind that both the reaction rate equation (11) and the rule of thumb for fluctuation scaling follow *only* if conditions (i) and (ii) are satisfied, i.e., only if the system admits a domain of macroscopically infinitesimal time intervals.

VI. CONNECTIONS WITH EARLIER WORKS

A careful inspection of our chemical Fokker–Planck equation (24) reveals that it is precisely the equation that would be obtained by simply truncating the chemical Kramers–Moyal equation (9) at $n=2$. The temptation to make that truncation has long been recognized, but the legitimacy of doing so has been seriously doubted. Has our present derivation of Eqs. (22)–(24) allayed those doubts?

Gardiner has observed that ''a confused history [has arisen] out of repeated attempts to find a limiting form of the master equation in which a Fokker–Planck equation arises.''[14] Although it has been apparent since the 1905 paper of Einstein[13] that a perfunctory second-order truncation of the chemical Kramers–Moyal equation (9) will result in a Fokker–Planck equation, we really are not entitled to do that without reasoned justification. Arguing that there is no such justification, van Kampen in 1961 offered instead his ''system-size expansion'':[15,10,9] Starting with the chemical Kramers–Moyal equation (9), van Kampen makes the variable change (12) from $\mathbf{X}(t)$ to $\mathbf{Z}(t)=\mathbf{X}(t)/\Omega$, invokes the scaling relation (13), and then in effect assumes that the solution $\mathbf{Z}(t)$ to that equation will differ from the deterministic process $\mathbf{Z}^*(t)$ defined by the reaction rate equation (14) by the amount $\Omega^{-1/2}\mathbf{Y}(t)$. van Kampen then proceeds to deduce, in the limit of large Ω, a well-defined Fokker-Planck equation for the process $\mathbf{Y}(t)$, an equation in which the drift and diffusion functions both depend strongly on the deterministic solution $\mathbf{Z}^*(t)$.

In 1973, Kubo, Matsuo, and Kitahara[8] pointed out that the scaling relation (13) by itself transforms the chemical Kramers–Moyal equation (9) into the form (15), which in turn suggests that a straightforward second-order truncation might actually be reasonable in the limit of large Ω. But later in their paper, Kubo, Matsuo, and Kitahara[8] back away from that position, and instead voice agreement with van Kampen that some sort of explicit time dependence [like that provided by $\mathbf{Z}^*(t)$ in van Kampen's analysis] seems to be needed in the drift and diffusion functions, in order to ''reproduce a correct equilibrium solution.''[16]

From a present-day perspective, this dismissal of a second-order truncation of Eq. (15) for large Ω is difficult to understand. It appears to be motivated by the view that a Fokker–Planck equation ''is generally meaningless except under the simple situation where [the drift function] is linear in x and [the diffusion function] is constant.''[16] But in fact, many Fokker–Planck equations with nonlinear drift and diffusion functions are now known that give perfectly legitimate results.[17] Furthermore, the following result for a single-species chemical system has been demonstrated by the present writer:[18] If the population of the time-varying species can change by no more than one molecule in any one reaction event, then an order-consistent discretization of the second-order truncated Kramers–Moyal equation (9)—which is a finite-difference equation that could legitimately be used to numerically solve that partial differential equation—is precisely the corresponding chemical master equation. This result does require that the population of the time-varying species be large compared to 1; however, there is *no* requirement that the propensity functions of the various reaction channels be constant or linear in the species population.

Another indication that a second-order truncation of the

chemical Kramers–Moyal equation might actually be reasonable, at least for finite times and in the thermodynamic limit, can be found in the highly mathematical works of Kurtz.[19] As explicated by Gardiner,[20] Kurtz analyzes the difference between the process $\mathbf{X}(t)$ defined by the chemical master equation (6), and the process $\mathbf{X}'(t)$ defined by the Langevin equation corresponding to a second-order truncation of the Kramers–Moyal equation (9), i.e., our Eq. (22) [or (23)]. He proves that the difference between those two processes becomes proportional to $\log\Omega$ in the limit $\Omega\to\infty$. It follows that the difference between the corresponding concentration processes $\mathbf{Z}(t)$ and $\mathbf{Z}'(t)$, obtained by dividing by Ω, asymptotically approaches zero with increasing Ω like $(\log\Omega)/\Omega$, again for finite times.

One is tempted to conclude from all this that the early summary rejection of the second-order truncated Kramers–Moyal equation may have been too hasty.

A quite different approach to the chemical Langevin equation, often attributed to Grossmann,[21] seems to mediate between van Kampen's system size expansion and the second-order Kramers–Moyal truncation. Using arguments based on the fluctuation-dissipation theorem of linear irreversible thermodynamics, Grossmann[21] infers the structure of the noise terms that should be added to the deterministic reaction rate equation (14) to convert it into a stochastic Langevin equation. His result appears to be equivalent to a second-order Kramers–Moyal truncation, but with the important difference that the propensity functions in the second-order derivative terms now have argument $\mathbf{Z}^*(t)$ instead of $\mathbf{z}$, where $\mathbf{Z}^*(t)$ is again the solution of the deterministic reaction rate equation (14). This is the chemical Langevin equation that was used by Baras, Malek Mansour, and Pearson[6] in their numerical study of the differences between it, the chemical master equation, and molecular dynamics simulation. For a model chemical system with multiple stable states, they found good agreement between the chemical master equation and molecular dynamics, but significant differences between those and their Langevin equation. Baras, Malek Mansour, and Pearson observed that this failure of their Langevin equation is not really so surprising, since the deterministic function $\mathbf{Z}^*(t)$ that guides their noise terms is not globally accurate in the presence of multiple stable states.[22,23]

This last point is important, and deserves elaboration. In the presence of multiple stable states, the solution to the deterministic reaction rate equation will simply follow the gradient to the first-encountered stable state, and there come to rest. But the process described by the chemical master equation will be forever on the move, occasionally making spontaneous transitions between the stable states, and thus exploring a large region of the state space. It follows that any particular solution of the deterministic reaction rate equation cannot provide an accurate picture of the long-time behavior of a system with multiple stable states. We should therefore expect that any chemical Langevin or Fokker–Planck equation that makes significant use of the deterministic solution—and this would include not only the Grossmann equation but also the van Kampen equation—would probably not give a very accurate description of such a system. The aforementioned results of Baras, Malek Mansour, and Pearson[6] seem to confirm this suspicion. But note that the Langevin equation considered by Kurtz,[19] namely our Langevin equation (22), does not use the solution of the deterministic reaction rate equation, and therefore should not be expected to have this problem.

An unsettling footnote to all this is a comment by Gardiner to the effect that the second-order truncated Kramers–Moyal equation and van Kampen's system size expansion will usually give the same results ''to lowest order in $\Omega^{-1/2}\ldots$ and each will only be valid to this order.''[24]

To understand what our present analysis contributes to this ''confused'' history, we begin by making three observations. First, all the cited earlier works have focused strongly on what happens in the thermodynamic limit, often devoting considerable attention to the order in which the two limits $\Omega\to\infty$ and $t\to\infty$ ought to be taken. Second, the underlying motivation for that focus is essentially to prove that what emerges in the thermodynamic limit is the deterministic reaction rate equation; indeed, that result is virtually a foregone conclusion in the van Kampen and Grossman approaches, since they essentially assume it at the outset. But that of course poses a problem for systems with multiple steady states, since for those systems the determininistic reaction rate equation does *not* provide a good description of the system's long time behavior in the thermodynamic limit. And finally, all the earlier cited works have tacitly assumed that the *accuracy* of any Langevin or Fokker–Planck approximation will be determined by a single, static, system size parameter, typically Ω.

The focus of our efforts here is rather different. Our goal is to describe the evolution of a finite system over a finite time. An example would be the genetic-enzymatic reactions that take place inside a living cell,[25] where the molecular populations of some key reactants can at times be less than 100. The population levels in such systems are not available to be ''taken to the thermodynamic limit;'' they are whatever they are. For such systems as these, the stochasticity implicit in Eq. (2) can play an important role, and a careful stochastic analysis is essential. But passing to thermodynamic and infinite-time limits are academic exercises with little practical import.

For such systems, we have shown in this paper that circumstances may arise in which the exact chemical master equation (6) can be satisfactorily *approximated* by the Langevin equation (22) [or (23)] and hence also the Fokker–Planck equation (24). This possibility may in retrospect not be surprising in view of the ''large Ω'' implications of the Kramers–Moyal expansion (15),[8] the convergence results of Kurtz,[19] and the well-known fact that the deterministic reaction rate equation quite often *is* accurate. However, our present analysis has demonstrated, apparently for the first time, that it is not sufficient to predicate a Langevin approximation on the value of a single parameter like Ω. Instead, it is necessary to verify that the system evolution possesses a *macroscopically infinitesimal time scale*, or more specifically, that from moment to moment there exists a time interval during which none of the system's propensity functions will suffer a noticeable change of value, yet every reaction

channel will be expected to fire many more times than once. It is true that satisfying these conditions will nearly always require ''large'' molecular populations; however, the practical question of ''how large'' can be answered only by appealing directly to those dynamic conditions. And we must be prepared to find that, for *some* systems, there will be *some* time periods during which a macroscopically infinitesimal time scale will *not* exist, i.e., the shortest time interval in which all reaction channels can be expected to fire many more times than once is nevertheless so long that some propensity function will suffer a ''noninfinitesimal'' change in value during that interval. Whenever that happens, we must simply acknowledge that a Langevin or Fokker–Planck approximation is not feasible, and return to the exact chemical master equation or its companion stochastic simulation algorithm.

Finally, there are two current criticisms of Langevin and Fokker–Planck equations in a chemical kinetics context that need to be addressed directly.

First, it was pointed out long ago by van Kampen that noise in a *jump* Markov process is ''internal,'' in that ''it is inherent in the very mechanism by which the state of the system evolves and cannot be divorced from its equations of motion.''[26] By contrast, the noise in a *continuous* Markov process usually appears to be ''external,'' since it manifests itself in the canonical Langevin equation (B1) as a seemingly independent enhancement to the deterministic drift. These observations were originally delivered as a warning to some early investigators who were proposing to accommodate the stochasticity in chemical systems by simply tacking on some generic noise to the deterministic reaction rate equation. In that context, van Kampen's observation is quite valuable. But over the years, this observation seems to have evolved into a general criticism of *any* Langevin or Fokker–Planck equation in a chemical kinetics context. So it is important to underscore the fact that the chemical Langevin and Fokker–Planck equations (22)–(24) derived in this paper are *not* subject to that criticism: Our derivation in Sec. IV of the Langevin equation (22) demonstrates quite clearly that the drift and diffusion terms in that equation arise from a *common source*, namely the Poisson random variables in Eq. (18). And it is easy to see from the very forms of Eqs. (22)–(24) that the noise component in the process defined by those equations ''cannot be divorced'' from the deterministic component. Nor should this be surprising, since Eqs. (22)–(24) have been derived from the same fundamental premise (2) that gave rise to the chemical master equation (6).

We conclude by addressing some claims made in the recent work of Baras, Malek Mansour, and Pearson.[6] As mentioned earlier, that work examines numerically the time evolutions predicted for a multistable-state chemical system by a molecular dynamics method, the chemical master equation, and a chemical Langevin equation. Baras, Malek Mansour, and Pearson find excellent agreement between the predictions of molecular dynamics and the chemical master equation; however, they claim that their results ''demonstrate the failure of the Langevin approach.''[6] But in fact, their results demonstrate the failure of only the *Grossmann*-type chemical Langevin equation, whose noise terms utilize the solution of the deterministic reaction rate equation.[27] Baras, Malek Mansour, and Pearson never numerically examine the *Kurtz*-type chemical Langevin equation that is derived in this paper. Their restricted focus in this regard is evident from their statement that ''a Langevin formalism is always characterized by a macroscopic law of evolution to which a noise term is added [whose amplitude] is directly related to the macroscopic law through a fluctuation-dissipation theorem... .''[28] But a quite different perspective is provided by our present analysis: Our Langevin formalism is characterized by drift and noise terms that *both* follow directly from the fundamental premise (2) whenever the system possesses a macroscopically infinitesimal time domain. We require neither a ''macroscopic law'' nor a ''fluctuation-dissipation theorem,'' but only the well-known mathematical facts (see Appendix A) that a Poisson random variable has a variance that is equal to its mean, and approaches normality whenever that mean is large compared to 1.

A repetition of the study of Baras and co-workers[6] using the Langevin equation (22) instead of the Grossmann version should show the chemical Langevin equation in a fairer light, provided one additional precaution is taken. In their investigation, Baras and co-workers assume that by taking the total number of molecules (which is conserved in their system) to be 2000, they will satisfy all the conditions required for the Langevin approximation to apply. In light of our work here, that does not seem to be a safe assumption: Since their system has three time-varying species, then with a total molecular population of 2000 it seems quite possible that the population of at least one of those species might at times become so small that a macroscopically infinitesimal time domain will not exist for the system. In such a case, a Langevin approximation will not be justified. Of course, it may not be easy to continually monitor the system to ensure that conditions (i) and (ii) of Sec. IV are satisfied. But that will not be the first time that Nature has proved to be unaccommodating to our purposes.

The present work grew out of an ongoing effort to develop approximate methods of speeding up the stochastic simulation algorithm. Results of that broader effort will be reported on in the near future.

ACKNOWLEDGMENT

This work was supported in part by the Office of Naval Research, Grant No. N0001499WX20544.

APPENDIX A: THE POISSON RANDOM VARIABLE

The Poisson random variable $\mathcal{P}(a,t)$ is understood here to be the number of ''events'' that occur in a time interval t given that the probability for an event to occur in any infinitesimal time interval dt is $a\,dt$. Letting $Q(n;a,t)$ denote the probability that $\mathcal{P}(a,t)$ has the (integer) value n, it is easy to show that $Q(0;a,t)=e^{-at}$. And by the laws of probability, we have for any $n\geqslant 1$,

$$Q(n;a,t)=\int_{t'=0}^{t} Q(n-1;a,t')\times a\,dt'\times Q(0;a,t-t').$$

Using this recursion relation and the $n=0$ formula, one can easily establish by induction the well-known general formula

$$Q(n;a,t)=\frac{e^{-at}(at)^n}{n!} \quad (n=0,1,2,...). \tag{A1}$$

It can be shown from this formula that the mean and variance of $\mathcal{P}(a,t)$ are

$$\langle \mathcal{P}(a,t)\rangle = \mathrm{var}\{\mathcal{P}(a,t)\}=at. \tag{A2}$$

By invoking Stirling's factorial approximation along with the small-ϵ approximation for $\ln(1+\epsilon)$, one can prove that

$$\frac{e^{-at}(at)^n}{n!}\approx(2\pi at)^{-1/2}\exp\left(-\frac{(n-at)^2}{2at}\right) \quad \text{if } at\gg 1.$$

This implies that, when $at\gg 1$, the Poisson random variable $\mathcal{P}(a,t)$ can be approximated by the *normal* random variable with the same mean and variance:

$$\mathcal{P}(a,t)\approx\mathcal{N}(at,at) \quad \text{if } at\gg 1. \tag{A3}$$

APPENDIX B: THE MULTIVARIATE LANGEVIN AND FOKKER–PLANCK EQUATIONS

The general multivariate Langevin and Fokker-Planck equations have been derived in different ways by different writers.[9–12] We adopt here the approach and notation developed in Ref. 12. There it is shown that an N-component process $\mathbf{X}(t)$ which evolves in a *continuous, memoryless* manner—i.e., a continuous Markov process $\mathbf{X}(t)$—must for reasons of self-consistency obey an ''infinitesimal updating'' formula of the form

$$X_i(t+dt) = X_i(t)+A_i(\mathbf{X}(t),t)dt+\sum_{j=1}^{M} b_{ij}(\mathbf{X}(t),t)N_j(t)(dt)^{1/2} \quad (i=1,...,N). \tag{B1}$$

Here the N functions A_i and the NM functions b_{ij} are smooth but otherwise arbitrary. And $N_j(t)$ is a zero-mean, unit-variance normal random variable $\mathcal{N}(0,1)$, with $N_j(t)$ and $N_{j'}(t')$ statistically independent if either $j\neq j'$ or $t\neq t'$; thus, $\langle N_j(t)N_{j'}(t')\rangle$ equals 1 if *both* $j=j'$ and $t=t'$, and equals 0 otherwise.

Equation (B1) is called the *standard-form Langevin equation* for the continuous Markov process $\mathbf{X}(t)$. In the derivation of Eq. (B1) given in Ref. 12, the number M of statistically independent unit normals is assumed to be equal to the number of process components N. Although M must be no less than N to accommodate the most general such process, nothing in the derivation precludes a continuous Markov process $\mathbf{X}(t)$ being defined with M either greater or less than N. A comparison of Eq. (B1) with Eq. (22) shows the latter to be a standard-form Langevin equation with

$$A_i(\mathbf{x},t)=\sum_{j=1}^{M}\nu_{ji}a_j(\mathbf{x}), \quad b_{ij}(\mathbf{x},t)=\nu_{ji}a_j^{1/2}(\mathbf{x}). \tag{B2}$$

A derivative-type limit of the standard-form Langevin equation (B1) will formally bring that equation into the *white-noise form*,

$$\frac{dX_i(t)}{dt}=A_i(\mathbf{X}(t),t)+\sum_{j}^{M} b_{ij}(\mathbf{X}(t),t)\Gamma_j(t). \tag{B3}$$

Here, the $\Gamma_j(t)$ are temporally uncorrelated, statistically independent *Gaussian white noises*. They are formally defined by

$$\Gamma_j(t)\equiv\lim_{dt\to 0}\mathcal{N}(0,1/dt). \tag{B4}$$

This definition, together with the properties of temporal and statistical independence, can be shown to imply

$$\langle\Gamma_j(t)\Gamma_{j'}(t')\rangle=\delta(j,j')\delta(t-t'), \tag{B5}$$

where the first delta function is Kronecker's and the second is Dirac's. When Eq. (B2) is substituted into Eq. (B3), the result is Eq. (23).

Strictly speaking, genuinely stochastic continuous Markov processes, while indeed continuous, are *not* differentiable; therefore, all white-noise form Langevin equations are in the nature of heuristic stand-ins for their corresponding standard forms. This caveat applies especially to our chemical problem here, where the limit that transforms Eq. (22) into Eq. (23) must be tempered by the fact that dt in this case is, owing to our condition (ii), a *macroscopic* infinitesimal.

Finally, it is proven in Ref. 12 that the standard-form Langevin equation (B1) together with the Chapman–Kolmogorov equation[7,9–11] imply, without any extra assumptions, that the singly conditioned probability density function of $\mathbf{X}(t)$ obeys the partial differential equation

$$\frac{\partial}{\partial t}P(\mathbf{x},t|\mathbf{x}_0,t_0)=-\sum_{i=1}^{N}\frac{\partial}{\partial x_i}[A_i(\mathbf{x},t)P(\mathbf{x},t|\mathbf{x}_0,t_0)] +\frac{1}{2}\sum_{i=1}^{N}\frac{\partial^2}{\partial x_i^2}[D_i(\mathbf{x},t)P(\mathbf{x},t|\mathbf{x}_0,t_0)] +\sum_{\substack{i,i'=1\\ i<i'}}^{N}\frac{\partial^2}{\partial x_i\partial x_{i'}}[C_{ii'}(\mathbf{x},t)P(\mathbf{x},t|\mathbf{x}_0,t_0)], \tag{B6}$$

with

$$D_i(\mathbf{x},t)\equiv\sum_j b_{ij}^2(\mathbf{x},t), \quad C_{ii'}(\mathbf{x},t)\equiv\sum_j b_{ij}(\mathbf{x},t)b_{i'j}(\mathbf{x},t). \tag{B7}$$

Equation (B6) is called the *forward Fokker-Planck equation* for the continuous Markov process $\mathbf{X}(t)$. If the formulas (B2) are substituted into Eqs. (B6) and (B7), the result is Eq. (24).

[1] D. A. McQuarrie, J. Appl. Probab. **4**, 413 (1967).
[2] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[3] I. Oppenheim, K. E. Shuler, and G. H. Weiss, J. Chem. Phys. **50**, 460 (1969).
[4] D. T. Gillespie, Physica A **188**, 404 (1992).

[5] J. Kestin and J. R. Dorfman, *A Course in Statistical Thermodynamics* (Academic, New York, 1971).
[6] F. Baras, M. Malek Mansour, and J. E. Pearson, J. Chem. Phys. **105**, 8257 (1996).
[7] D. T. Gillespie, *Markov Processes: An Introduction for Physical Scientists* (Academic, San Diego, 1992).
[8] R. Kubo, K. Matsuo, and K. Kitahara, J. Stat. Phys. **9**, 51 (1973).
[9] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer, Berlin, 1985).
[10] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).
[11] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, Berlin, 1989).
[12] D. T. Gillespie, Am. J. Phys. **64**, 1246 (1996).
[13] A. Einstein, Ann. Phys. (Leipzig) **19**, 549 (1905); see also Ref. 9, pp. 3–6.
[14] In Ref. 9, p. 246.
[15] N. G. van Kampen, Can. J. Phys. **39**, 551 (1961).
[16] In Ref. 8, p. 68.
[17] In Ref. 7, a continuous Markov process with a linear drift function and a *quadratic* diffusion function is discussed on pp. 163–169, and *two* continuous Markov processes with *cubic* drift functions, one having a constant diffusion function and the other a *quadratic* diffusion function, are discussed on pp. 169–173. All of these processes are shown by both theoretical analysis and extensive numerical simulation to have well defined and well understood equilibrium behaviors.
[18] D. T. Gillespie, J. Chem. Phys. **72**, 5363 (1980); see also Ref. 7, Sec. 6.2.A.
[19] T. G. Kurtz, Math. Progr. Stud. **5**, 67 (1976); Stoch. Proc. Appl. **6**, 223 (1978).
[20] In Ref. 9, pp. 254–255.
[21] S. Grossmann, J. Chem. Phys. **65**, 2007 (1976).
[22] G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems* (Wiley–Interscience, New York, 1977),
[23] M. Malek Mansour, C. Van Den Broeck, G. Nicolis, and J. W. Turner, Ann. Phys. (N.Y.) **131**, 283 (1981).
[24] In Ref. 9, p. 252.
[25] H. McAdams and A. Arkin, Proc. Natl. Acad. Sci. USA **94**, 814 (1997).
[26] See the 1981 edition of Ref. 10, p. 247.
[27] See Eqs. (A2) of Ref. 6.
[28] See the ''Concluding Remarks'' of Ref. 6.