A general method for the computation of probabilities in systems of first order chemical reactions

Xueying Zhang,[a] Katrien De Cock,[b] Mónica F. Bugallo,[c] and Petar M. Djurić[d]
COSINE Laboratory, Department of Electrical & Computer Engineering, Stony Brook University, Stony Brook, New York 11794-2350

We present a general method for the computation of molecular population distributions in a system of first-order chemical reactions. The method is based on the observation that the molecules in first-order reactions do not interact with each other. Therefore, the population distributions are a convolution of densities for one molecule. With this method one can study reactions involving any number of molecules. Such analysis is demonstrated on several examples, including an enzyme catalyst reaction and a first-order reaction chain. © *2005 American Institute of Physics*.
[DOI: 10.1063/1.1855311]

I. INTRODUCTION

In small biological systems, it is often more appropriate to model the chemical reactions in a stochastic way rather than with the traditional differential equations for evolution of concentrations. The reasons are (a) the number of molecules involved in a biological system can be very small;[1] (b) there are many situations in which molecules are not homogeneously distributed;[2,3] (c) the change of molecular population levels is a discrete integer amount; and (d) the population change is not a deterministic process.[4,5]

There are mainly two approaches for the stochastic study of the number of molecules in biochemical reactions: the first is based on the analysis of the master equation[6–8] and the second relies on Monte Carlo simulation methods.[9–11] Here, we adopt the first approach.

In 1967, McQuarrie summarized the stochastic study using the master equation approach.[7] For a reaction, it is assumed that in an infinitesimal time interval, the probability of having one reaction per unit reactant molecule combination is proportional to the length of the time interval. A probability difference function is first obtained based on the assumption, which leads to a differential-difference equation called the master equation of the reaction. The moment-generating function is then employed to transform the master equation into a partial differential equation. It can sometimes be solved to obtain the analytic solution. More often, the mean and variance of the number of molecules can only be obtained.

In 2000, Laurenzi introduced a different way of solving the master equation.[8] Instead of using the moment-generating function, he applied the Laplace transform. In this way, solving the partial differential equations is avoided, which is important for more complicated reactions. Instead, one needs to solve a set of linear equations. We will show how to obtain the solution of the master equation for any system of first-order reactions in yet another way.

In this paper we study systems of first-order reactions, i.e., combinations of (coupled) reactions of the type $S_i \rightarrow S_j$. Because in such systems molecules do not chemically interact, the analysis is much simpler than when second-order reactions are involved. Although this situation might occur only rarely in practice, our analysis is also useful for systems where the first-order reactions are dominant and for systems where second-order reactions can be approximated by first-order reactions (see Sec. IV for an example).

The independence of the molecules in a first-order reaction system is exploited to derive the population distributions of the molecules in the system. Instead of solving the master equation for the complete system, where the number of states grows with the number of molecules, we first solve the master equation for only one molecule and use its solution to find the population distributions for all species.

The structure of the paper is as follows. In Sec. II we solve the master equation for one molecule in a first-order reaction system. The resulting probability distribution functions are used in Sec. III to obtain the population distributions when more molecules are introduced. In Sec. IV we illustrate the general results of Sec. III with specific examples. Finally, in Sec. V we give some concluding remarks.

II. ONE MOLECULE

In this section we study the master equation for one molecule in a system of first-order reactions. The solution will then be used to solve more general cases in Sec. III.

Assume that there are M molecule species, $S_1, S_2, \ldots, S_M$ in the reaction system, and let the probability rate constant of the reaction $S_i \rightarrow S_j$ be denoted by c_{ij}. For a first-order reaction the specific probability rate constant is such that

$$c_{ij}\Delta t(i \neq j) \equiv \text{the probability for the first-order reaction}$$

[a] Electronic mail: sherry@ece.sunysb.edu
[b] Electronic mail: decock@ece.sunysb.edu
[c] Electronic mail: monica@ece.sunysb.edu
[d] Electronic mail: djuric@ece.sunysb.edu

$S_i \xrightarrow{c_{ij}} S_j$ to happen per molecule in an infinitesimally small time interval Δt.

The master equation describes the evolution of the molecules' population distribution in the chemical system as a function of time. When there is only one molecule in a first-order reaction system, the master equation gives the time evolution of the probability that this molecule has become a certain chemical species. Let $p^{(i)}(t)$ be the probability that the molecule is an S_i molecule at time t. For example, in the simple system of reactions

$$S_1 \underset{c_{21}}{\overset{c_{12}}{\rightleftarrows}} S_2 \xrightarrow{c_{23}} S_3, \tag{1}$$

the differential equation describing the time evolution of $p^{(2)}$ is the following first-order linear differential equation:

$$\frac{dp^{(2)}(t)}{dt} = c_{12} p^{(1)}(t) - (c_{21} + c_{23}) p^{(2)}(t). \tag{2}$$

We assume in the rest of this paper that the initial values for the probabilities are equal to zero: $p^{(i)}(0) = 0$, $i = 1, \ldots, M$. The injection of the molecule in the system is modeled by a source probability density function. For the given example, if the molecule begins as an S_2 molecule, the differential equation becomes

$$\frac{dp^{(2)}(t)}{dt} = c_{12} p^{(1)}(t) - (c_{21} + c_{23}) p^{(2)}(t) + f^{(2)}(t), \tag{3}$$

where $f^{(2)}(t)$ is the source's probability density, i.e., $f^{(2)}(t)\Delta t$ is the probability that a molecule from the source is injected in state S_2 in the interval $[t, t+\Delta t]$. If the injection time is known precisely, e.g., $t = \tau$, the source's probability density function is a delta function: $f^{(2)}(t) = \delta(t - \tau)$. Otherwise, the density function for the injection time of the molecule can have forms such as $ae^{-a(t-\tau)}u(t-\tau)$, $a\delta(t-\tau_1) + (1-a)\delta(t-\tau_2)$, and

$$\sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \delta(t - n\tau).$$

In general, the system can be represented by a Markov chain with a transition matrix

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1M} \\ c_{21} & c_{22} & \cdots & c_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ c_{M1} & c_{M2} & \cdots & c_{MM} \end{pmatrix}, \text{ where } c_{ii} = -\sum_{j=1, j\neq i}^{M} c_{ij}, \tag{4}$$

and the set of first-order linear differential equations has the following form:

$$\begin{pmatrix} \dfrac{dp^{(1)}(t)}{dt} \\ \dfrac{dp^{(2)}(t)}{dt} \\ \vdots \\ \dfrac{dp^{(M)}(t)}{dt} \end{pmatrix} = \mathbf{C}^{\mathsf{T}} \begin{pmatrix} p^{(1)}(t) \\ p^{(2)}(t) \\ \vdots \\ p^{(M)}(t) \end{pmatrix} + \begin{pmatrix} f^{(1)}(t) \\ f^{(2)}(t) \\ \vdots \\ f^{(M)}(t) \end{pmatrix}, \tag{5}$$

where

$$\sum_{i=1}^{M} p^{(i)}(t) = 1,$$

and

$$\begin{cases} \displaystyle\int_{t=0}^{\infty} f^{(i)}(t)dt = 1 & \text{if the molecule from source is injected} \\ & \text{into state } i, \\ f^{(i)}(t) = 0 & \text{otherwise}. \end{cases}$$

The equations can be solved by first applying the Laplace transform, followed by solving the resulting algebraic equation, and finally using the inverse Laplace transform. More specifically, after taking the Laplace transform on both sides of (5), we obtain

$$s\mathbf{p} = \mathbf{C}^{\mathsf{T}}\mathbf{p} + \mathbf{f}, \tag{6}$$

where $\mathbf{p}$ is an $M \times 1$ vector whose elements $p^{(i)}(s)$ are the Laplace transforms of $p^{(i)}(t)$, the sought-after probability density functions, and $\mathbf{f}$ is an $M \times 1$ vector with elements $f^{(i)}(s)$, which are the Laplace transforms of the source functions $f^{(i)}(t)$. Note that $\mathbf{f}$ has only one nonzero element.

After solving for $\mathbf{p}$ in (6), we get

$$\mathbf{p} = (s\mathbf{I} - \mathbf{C}^{\mathsf{T}})^{-1}\mathbf{f}. \tag{7}$$

Let

$$\mathbf{L} = (s\mathbf{I} - \mathbf{C}^{\mathsf{T}})^{-1} \tag{8}$$

and

$$\mathbf{G} = \mathcal{L}^{-1}(\mathbf{L}), \tag{9}$$

where $\mathcal{L}^{-1}$ is the inverse Laplace operator. If the elements of $\mathbf{G}$ are denoted by $g_{ij}(t)$, and $f^{(m)}(s)$ is the nonzero element of $\mathbf{f}$, the solutions $p^{(i)}(t)$ are given by

$$p^{(i)}(t) = g_{im}(t) \otimes f^{(m)}(t), \quad i = 1, 2, \ldots, M, \tag{10}$$

where the symbol $\otimes$ denotes convolution. Finding the inverse Laplace transforms of the elements of $\mathbf{L}$ may be tedious but it is straightforward.

This method is illustrated in details by example 1. As already mentioned, the technique of using the Laplace transform for solving the master equation was introduced by Laurenzi in Ref. 8. The difference between his and our approach is that Laurenzi was solving the many-molecule master equation, while here we only solve the one-molecule master equation.

In the deterministic framework, the concentrations of the species are found by solving the same differential equations

(5), without the stochastic sources, but taking into account the initial concentrations. Furthermore, the so-obtained results there yield only the means of the stochastic distributions.

III. MORE MOLECULES

In this section we treat the case when several molecules are injected into the system. First, in Sec. III A, we assume that all molecules are inserted by the same source, so that they all start as the same chemical species. In Sec. III B, the molecules can originate from different sources.

A. One source

Suppose that there are x molecules of species S_k injected into the system with the same probability density $f^{(k)}(t)$, and no molecules of other species. Denote the number of molecules at state S_i ($i=1,\dots,M$) at time t by $N^{(i)}(t)$, and let $K_j^{(i)}(t)=1$ if molecule j is at state S_i at time t and $K_j^{(i)}(t)=0$ if it is not. Thus, $N^{(i)}(t)=\Sigma_{j=1}^x K_j^{(i)}(t)$. Note that $K_1^{(i)}(t)$, $K_2^{(i)}(t),\dots,K_x^{(i)}(t)$ are independent Bernoulli random variables with the same probability $p^{(i)}(t)$, which can be computed by solving (5). Therefore, $N^{(i)}(t)$ has a binomial distribution. The probability that there are n molecules in state S_i at time t is

$$P_n^{(i)}(t) = \binom{x}{n}(p^{(i)}(t))^n(1-p^{(i)}(t))^{x-n}, \tag{11}$$

with mean $\mu_i(t)$ and variance $\sigma_i^2(t)$ equal to

$$\mu_i(t) = xp^{(i)}(t),$$

$$\sigma_i^2(t) = xp^{(i)}(t)(1-p^{(i)}(t)),$$

respectively. Moreover, the joint probability of finding n_i molecules in state S_i ($i=1,\dots,M$) at time t is equal to

$$P(n_1,\dots,n_M;t) = \frac{x!}{n_1!\cdots n_M!}(p^{(1)}(t))^{n_1}\cdots(p^{(M)}(t))^{n_M}, \tag{12}$$

where $\Sigma_{i=1}^M n_i=x$. Equation (11) is one of the marginals of (12).

When the number of molecules is large, i.e., when $xp^{(i)}(t)(1-p^{(i)}(t))\gg 1$, the binomial distribution (11) can be well approximated by a Gaussian distribution, or

$$P_n^{(i)}(t) \simeq \frac{1}{\sqrt{2\pi\sigma_i^2(t)}}e^{-[(n-\mu_i(t))^2/2\sigma_i^2(t)]}.$$

The purpose for employing the Gaussian approximation is to save computation time when the number of molecules is very large, as well as to make the analysis more tractable. Numerical examples of the comparison between analytic distributions and their Gaussian approximations can be found in Ref. 12.

The same purpose can be achieved if the binomial distribution (11) is approximated by a Poisson distribution. Conditions for the validity of such approximation are stricter. First, the number of molecules should be large. Second, the probability for a single molecule in state i should be small

[e.g., $p^{(i)}(t)<0.1$]. Finally, the product of x and $p^{(i)}(t)$ should be a sizable number. Under such conditions, the variance of the binomial distribution roughly equals the mean. Then (11) can be approximated by the Poisson distribution with parameter $\lambda_i=xp^{(i)}(t)$, or

$$P_n^{(i)}(t) \simeq \frac{\lambda_i^n}{n!}e^{-\lambda_i}.$$

B. More sources

If there are several sources which are independent from each other, the probability mass function for state S_i is the convolution of binomial distributions. Suppose there are J independent sources. Denote the number of molecules at state S_i ($i=1,\dots,M$) at time t by $N^{(i)}(t)$, and denote the number of molecules at state S_i that are injected from source j by $K_j^{(i)}(t)$ ($j=1,\dots,J$). Then, $N^{(i)}(t)=\Sigma_{j=1}^J K_j^{(i)}(t)$. Since $K_j^{(i)}(t)$ ($j=1,\dots,J$) are independent random variables, the probability mass function of the random variable $N^{(i)}(t)$ is the convolution of the probability mass functions of the random variables $K_j^{(i)}(t)$.[13]

For example, let x S_k molecules be injected by the source functions $f^{(k)}(t)$ and y S_l molecules by $f^{(l)}(t)$. Each of the x S_k molecules gives rise to the state probabilities $p_k^{(i)}(t)$ ($i=1,\dots,M$), whereas a molecule introduced by $f^{(l)}(t)$, to $p_l^{(i)}(t)$ ($i=1,\dots,M$). Then, the number of molecules at state S_i that are injected from source k is $K_k^{(i)}(t)$, which has a binomial distribution $\binom{x}{n}(p_k^{(i)}(t))^n(1-p_k^{(i)}(t))^{x-n}$. Similarly, the number of molecules at state S_i that are injected from source l is $K_l^{(i)}(t)$, which has a binomial distribution $\binom{y}{n}(p_l^{(i)}(t))^n(1-p_l^{(i)}(t))^{y-n}$. The probability mass function of $N^{(i)}(t)$, the total number of molecules at state S_i is equal to the convolution of two binomial distributions, i.e.,

$$P_n^{(i)}(t) = \binom{x}{n}(p_k^{(i)}(t))^n(1-p_k^{(i)}(t))^{x-n}$$

$$\otimes \binom{y}{n}(p_l^{(i)}(t))^n(1-p_l^{(i)}(t))^{y-n}, \tag{13}$$

where the convolution is along the n axis, that is

$$y_1(n) \otimes y_2(n) = \sum_{m=0}^{\infty} y_1(m)y_2(n-m).$$

The mean $\mu_i(t)$ and variance $\sigma_i^2(t)$ of $N^{(i)}(t)$ are

$$\mu_i(t) = xp_k^{(i)}(t) + yp_l^{(i)}(t), \tag{14}$$

$$\sigma_i^2(t) = xp_k^{(i)}(t)(1-p_k^{(i)}(t)) + yp_l^{(i)}(t)(1-p_l^{(i)}(t)), \tag{15}$$

respectively.

As we mentioned before, when the number of molecules is large, the binomial distribution can be well approximated by a Gaussian distribution with the same mean and variance. One property of Gaussian distributions is that the convolution of two Gaussian distributions yields a Gaussian distribution. Therefore, the convolution result in (13) can be approximated by a Gaussian distribution with mean (14) and variance (15). The Poisson distribution has the same prop-

erty: the convolution of Poisson distributions with parameters α and β, respectively, yields a Poisson distribution with parameter $\theta = \alpha + \beta$.

IV. EXAMPLES

In this section several examples of systems of first-order reactions are discussed. For some systems, an analytic solution for the population distribution is derived. This section also includes two numerical examples: the first is a one-directional reaction chain and the second an enzyme substrate reaction.

A. Example 1: The first-order irreversible reaction

Consider the first-order reaction $S_1 \xrightarrow{c} S_2$, where only molecules S_1 are injected into the system with probability density function $f^{(1)}(t) = \delta(t)$. The first step is to rewrite (5) with $M = 2$, $c_{12} = c$, $c_{21} = 0$, and $f^{(2)}(t) = 0$

$$\begin{pmatrix} \dfrac{dp^{(1)}(t)}{dt} \\ \dfrac{dp^{(2)}(t)}{dt} \end{pmatrix} = \begin{pmatrix} -c & c \\ 0 & 0 \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} p^{(1)}(t) \\ p^{(2)}(t) \end{pmatrix} + \begin{pmatrix} \delta(t) \\ 0 \end{pmatrix}. \tag{16}$$

Taking the Laplace transform leads to

$$s\begin{pmatrix} p^{(1)}(s) \\ p^{(2)}(s) \end{pmatrix} = \begin{pmatrix} -c & c \\ 0 & 0 \end{pmatrix}^{\mathsf{T}} \begin{pmatrix} p^{(1)}(s) \\ p^{(2)}(s) \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The matrix $\mathbf{L}$ from (8) is given by

$$\mathbf{L} = \begin{pmatrix} \dfrac{1}{s+c} & 0 \\ \dfrac{1}{s} - \dfrac{1}{s+c} & \dfrac{1}{s} \end{pmatrix}.$$

After taking the inverse Laplace transform of $\mathbf{L}$, we obtain

$$\mathbf{G} = \begin{pmatrix} e^{-ct} & 0 \\ 1 - e^{-ct} & u(t) \end{pmatrix},$$

where $u(t)$ is the unit step function. The final solution is

$$p^{(1)}(t) = e^{-ct} u(t), \tag{17}$$

$$p^{(2)}(t) = (1 - e^{-ct}) u(t). \tag{18}$$

In the rest of the paper we will drop the use of $u(t)$, and so we point out that the given solutions for $p(t)$ are valid only for $t \geq 0$. When $f^{(1)}(t)$ is some other density function

$$p^{(1)}(t) = \int_0^\infty f^{(1)}(\tau) e^{-c(t-\tau)} d\tau,$$

$$p^{(2)}(t) = \int_0^\infty f^{(1)}(\tau)(1 - e^{-c(t-\tau)}) d\tau.$$

Note that if an S_2 instead of an S_1 molecule was injected into the system with $f^{(2)}(t)$, $p^{(1)}(t) = 0$ and $p^{(2)}(t) = \int_0^t f^{(2)}(\tau) d\tau$.

By substituting the probabilities (17) and (18) in (11), we obtain the well-known result for $f_{(t)}^{(1)} = \delta(t)$

$$P_n^{(1)}(t) = \begin{pmatrix} x \\ n \end{pmatrix} e^{-nct}(1 - e^{-ct})^{x-n},$$

$$P_n^{(2)}(t) = \begin{pmatrix} x \\ n \end{pmatrix} (1 - e^{-ct})^n e^{-c(x-n)t}.$$

B. Example 2. The first-order reversible reaction

Consider the first-order reversible reaction $S_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_2$ which starts with x S_1 molecules and y S_2 molecules. Each of the x S_1 molecules gives rise to the following probabilities, which are obtained by solving (5) with $M = 2$, $f^{(1)}(t) = \delta(t)$, and $f^{(2)}(t) = 0$:

$$p_1^{(1)}(t) = \frac{1}{c_1 + c_2}(c_2 + c_1 e^{-(c_1+c_2)t}),$$

$$p_1^{(2)}(t) = \frac{c_1}{c_1 + c_2}(1 - e^{-(c_1+c_2)t}). \tag{19}$$

For the effect of each of the y S_2 molecules, we take $f^{(1)}(t) = 0$ and $f^{(2)}(t) = \delta(t)$ and obtain

$$p_2^{(1)}(t) = \frac{c_2}{c_2 + c_1}(1 - e^{-(c_2+c_1)t}),$$

$$p_2^{(2)}(t) = \frac{1}{c_2 + c_1}(c_1 + c_2 e^{-(c_2+c_1)t}). \tag{20}$$

By substituting the probabilities of (19) and (20) in (13), we obtain the distribution for the S_1 molecules.

$$\begin{aligned} P_n^{(1)}(t) = & \begin{pmatrix} x \\ n \end{pmatrix} \left(\frac{1}{c_1 + c_2}(c_2 + c_1 e^{-(c_1+c_2)t}) \right)^n \\ & \times \left(1 - \frac{1}{c_1 + c_2}(c_2 + c_1 e^{-(c_1+c_2)t}) \right)^{x-n} \\ & \otimes \begin{pmatrix} y \\ n \end{pmatrix} \left(\frac{c_2}{c_2 + c_1}(1 - e^{-(c_1+c_2)t}) \right)^n \\ & \times \left(1 - \frac{c_2}{c_2 + c_1}(1 - e^{-(c_1+c_2)t}) \right)^{y-n}, \end{aligned} \tag{21}$$

and similarly for the S_2 molecules.

Note that when only S_1 or only S_2 molecules are present at $t = 0$, the distribution reduces to the first and second convolution factor in (21), respectively. The population distribution for this special case has already been given by McQuarrie.[7]

Also note that deriving the mean and variance for the number of S_1 and S_2 molecules is straightforward, as explained in Sec. III B, and leads to the same results as reported in Refs. 14 and 15.

C. Example 3: A combination of the first-order irreversible and first-order reversible reaction
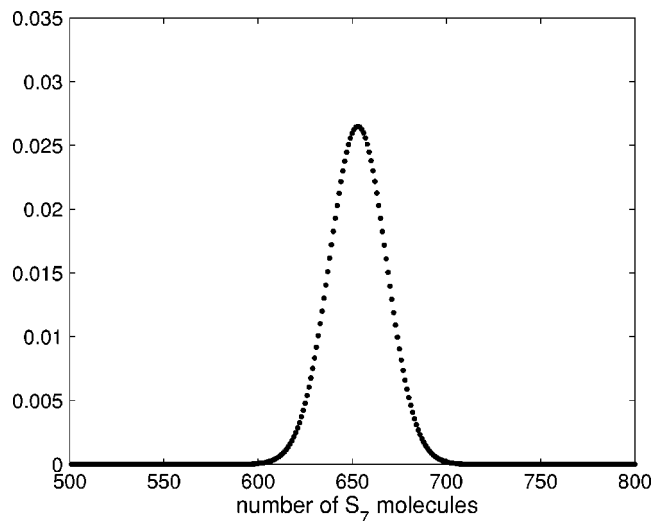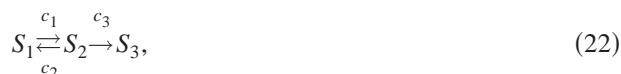
Consider the reaction system

FIG. 1. Direct calculation of the probability mass function of S_7 at time $t = 1.5$ s.

FIG. 3. The probabilities for a single molecule to be in state S_i ($i = 1, 2, \ldots, 7$) at time t.

$$S_1 \underset{c_2}{\overset{c_1}{\rightleftarrows}} S_2 \overset{c_3}{\rightarrow} S_3, \tag{22}$$

and assume there are x S_1 molecules and y S_2 molecules at $t=0$. Again, we first solve (5) with $f^{(1)}(t) = \delta(t)$, $f^{(2)}(t) = f^{(3)}(t) = 0$ to find the probabilities due to one S_1 molecule and obtain

$$p_1^{(1)}(t) = \frac{\sqrt{m} - c_1 + c_2 + c_3}{2\sqrt{m}} e^{-(1/2)(-\sqrt{m}+c_1+c_2+c_3)t}$$
$$+ \frac{\sqrt{m} + c_1 - c_2 - c_3}{2\sqrt{m}} e^{-(1/2)(\sqrt{m}+c_1+c_2+c_3)t},$$

$$p_1^{(2)}(t) = \frac{c_1}{\sqrt{m}} e^{-(1/2)(-\sqrt{m}+c_1+c_2+c_3)t} - \frac{c_1}{\sqrt{m}} e^{-(1/2)(\sqrt{m}+c_1+c_2+c_3)t}, \tag{23}$$
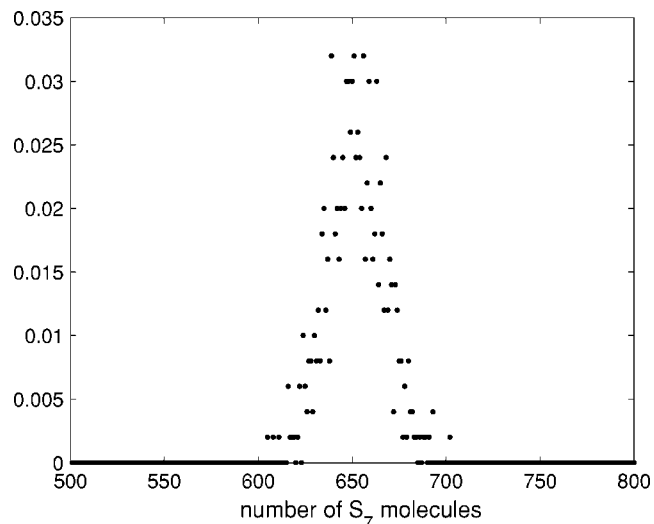
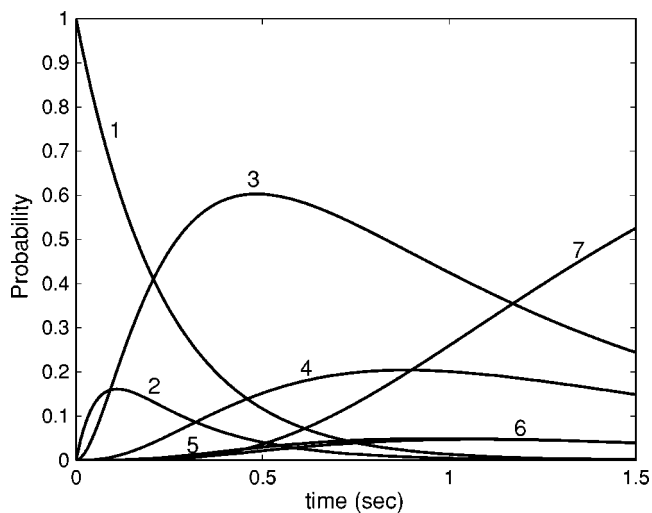FIG. 2. The probability mass of S_7 at time $t=1.5$ s calculated using the SSA, with 500 realizations.

$$p_1^{(3)}(t) = 1 - \frac{\sqrt{m} + c_1 + c_2 + c_3}{2\sqrt{m}} e^{-(1/2)(-\sqrt{m}+c_1+c_2+c_3)t}$$
$$- \frac{\sqrt{m} - c_1 - c_2 - c_3}{2\sqrt{m}} e^{-(1/2)(\sqrt{m}+c_1+c_2+c_3)t},$$

where

$$m = c_1^2 + 2c_1c_2 - 2c_1c_3 + c_2^2 + 2c_2c_3 + c_3^2.$$

Next, we solve (5) with $f^{(1)}(t) = f^{(3)}(t) = 0$, $f^{(2)} = \delta(t)$ to find the probabilities due to one S_2 molecule and find that

$$p_2^{(1)}(t) = \frac{c_2}{\sqrt{m}} e^{-(1/2)(-\sqrt{m}+c_1+c_2+c_3)t} - \frac{c_2}{\sqrt{m}} e^{-(1/2)(\sqrt{m}+c_1+c_2+c_3)t},$$
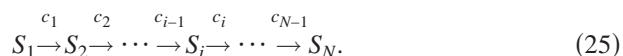
$$p_2^{(2)}(t) = \frac{\sqrt{m} + c_1 - c_2 - c_3}{2\sqrt{m}} e^{-(1/2)(-\sqrt{m}+c_1+c_2+c_3)t}$$
$$+ \frac{\sqrt{m} - c_1 + c_2 + c_3}{2\sqrt{m}} e^{-(1/2)(\sqrt{m}+c_1+c_2+c_3)t}, \tag{24}$$

$$p_2^{(3)}(t) = 1 - \frac{\sqrt{m} + c_1 + c_2 - c_3}{2\sqrt{m}} e^{-(1/2)(-\sqrt{m}+c_1+c_2-c_3)t}$$
$$- \frac{\sqrt{m} - c_1 - c_2 + c_3}{2\sqrt{m}} e^{-(1/2)(\sqrt{m}+c_1+c_2+c_3)t}.$$

The substitution of the probabilities in (13) by the probabilities in (23) and (24) yields the population distribution for the S_1, S_2, and S_3 molecules.

D. Example 4: A first-order reaction chain

Consider a first-order reaction chain in its general form

$$S_1 \overset{c_1}{\rightarrow} S_2 \overset{c_2}{\rightarrow} \cdots \overset{c_{i-1}}{\rightarrow} S_i \overset{c_i}{\rightarrow} \cdots \overset{c_{N-1}}{\rightarrow} S_N. \tag{25}$$

Assume that $c_i \neq c_j$ when $i \neq j$ ($i = 1, \ldots, N$ and $j = 1, \ldots, N$) and that x molecules enter the system in state S_1 at time $t = 0$. The input probability distribution of state S_1 is $f^{(1)}(t)$
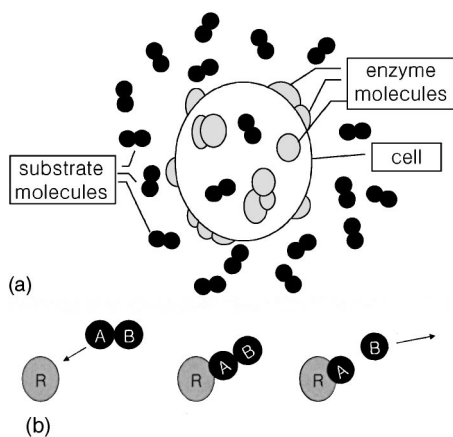
FIG. 4. (a) A cell on which enzyme molecules are immobilized. The substrate molecules float around freely. (b) Illustration of the chemical reaction. The substrate molecule *AB* approaches the enzyme molecule *R* (left). The enzyme recognizes the *A* molecule of the *AB* structure (middle). The enzyme destroys the bond between *A* and *B*, *A* remains attached to *R*, and *B* is released into the solution (right).

$= \delta(t)$. Then, using the method introduced in the previous sections, the state probabilities for one molecule are equal to

$$
p^{(i)}(t) = \begin{cases} \dfrac{1}{c_i}\sum_{j=1}^{i} M_j^{(1,i)} c_j e^{-c_j t}, & c_i > 0 \\ 1 - \sum_{j=1}^{i-1} M_j^{(1,i-1)} e^{-c_j t}, & c_i = 0, i \geq 2, \\ 1, & c_i = 0, i = 1 \end{cases}
$$

where

$$
M_p^{(k,m)} = \begin{cases} \prod_{j=k, j\neq p}^{m} \dfrac{c_j}{c_j - c_p}, & m > k \\ 1, & m = k. \end{cases}
$$

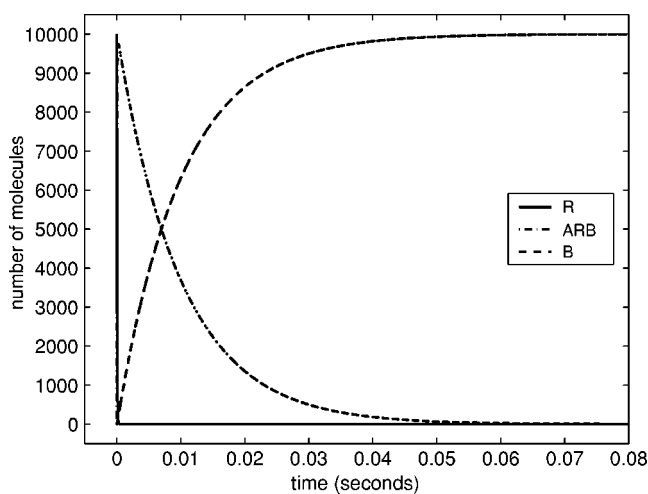The probability that n of the x molecules are in state S_i at time t is given by

FIG. 5. Mean for the population of molecule species *R* (full line), *ARB* (dot-dashed line), and *B* (dashed line).
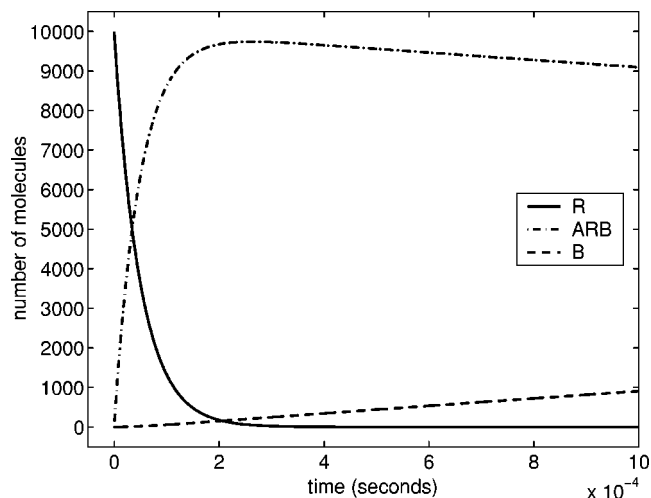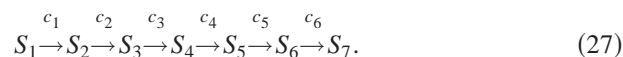

FIG. 6. Zoomed plots of the mean for the population of molecule species *R* (full line), *ARB* (dot-dashed line), and *B* (dashed line).

$$
P_n^{(i)}(t) = \binom{x}{n}(p^{(i)}(t))^n (1 - p^{(i)}(t))^{x-n}. \tag{26}
$$

If there are c_i's equal to c_j's, a closed-form solution is also possible.

E. Example 5: A numerical example for the first-order reaction chain

Consider the first-order reaction chain

$$
S_1 \xrightarrow{c_1} S_2 \xrightarrow{c_2} S_3 \xrightarrow{c_3} S_4 \xrightarrow{c_4} S_5 \xrightarrow{c_5} S_6 \xrightarrow{c_6} S_7. \tag{27}
$$

Suppose there are 1000 S_1 molecules at time $t=0$. The values of the reaction parameters $c_1,\ldots,c_6$ are 4.3, 16.6, 1.2, 2.8, 11.4, 11.9, respectively, with unit s^{-1}. We used two methods to obtain the probability mass of S_7 at time $t=1.5$ s

(1) by direct calculation (DC) using (26), and
(2) by running 500 realizations of the SSA (stochastic simulation algorithm), which is a Monte Carlo method for numerically computing the time evolution of a chemical system.[9]

The resulting probability mass functions are shown in Figs. 1 and 2. From Fig. 2 it is clear that the 500 SSA realizations are not enough for an accurate evaluation of the probability distribution, even though the computation time it took is much longer than the direct calculation using (26). Obviously, methods based on analytic solutions are better than those based on Monte Carlo simulations. However, in cases of more complex reactions where neither analytic solutions nor good approximations of population distributions exist, the SSA is the method of choice.

We have shown in Sec. III A that when $x \to \infty$ and $p^{(i)}(t) \to 0$ [e.g., $p^{(i)}(t) < 0.1$], the binomial distribution (11) can be approximated by the Poisson distribution with mean $\lambda = x p^{(i)}(t)$. The purpose of the approximation is to make the analysis more tractable. In many situations, however, the condition $p^{(i)}(t) \to 0$ is not valid. Using the reaction chain (25) and reaction parameter $c_1,\ldots,c_6$ of this example, the probabilities of a single molecule to be in state S_i (*i*
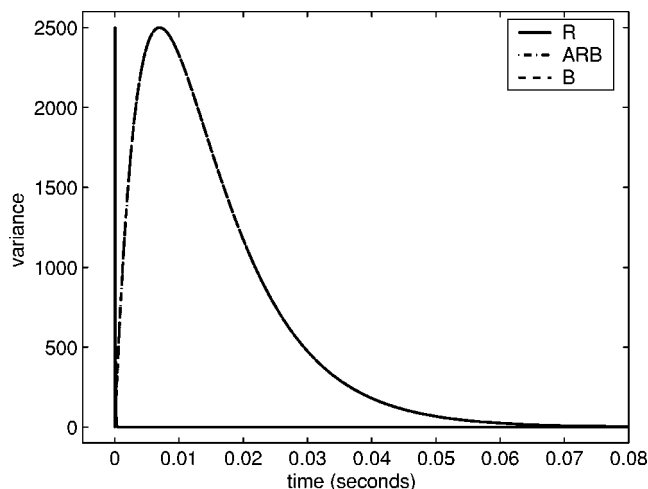
FIG. 7. Variance of the population of molecule species R (full line), ARB (dot-dashed line), and B (dashed line).

$=1, 2, \ldots, 7$) at time t are drawn in Fig. 3, from which it can be seen that during the first 1.5 s, many $p^{(i)}(t)$'s have values greater than 0.1, which implies that the Poisson approximation in that period is not suitable.

F. Example 6: An enzyme–substrate reaction

Let us consider an enzyme reaction where a soluble substrate with an $A-B$ structure reacts with immobilized enzyme molecules located on the surface of cells [see Fig. 4(a)]. The formulation of this chemical reaction is given by

$$AB + R \underset{c_2}{\overset{c_1^{(2)} \quad c_3}{\rightleftarrows}} ARB \rightarrow AR + B, \tag{28}$$

where AB is the substrate molecule, R is the enzyme molecule, and ARB is the enzyme–substrate complex, known as the intermediate product. This reaction is illustrated in Fig. 4(b). The superscript of $c_1^{(2)}$ emphasizes that this specific probability rate constant is for a second-order reaction.

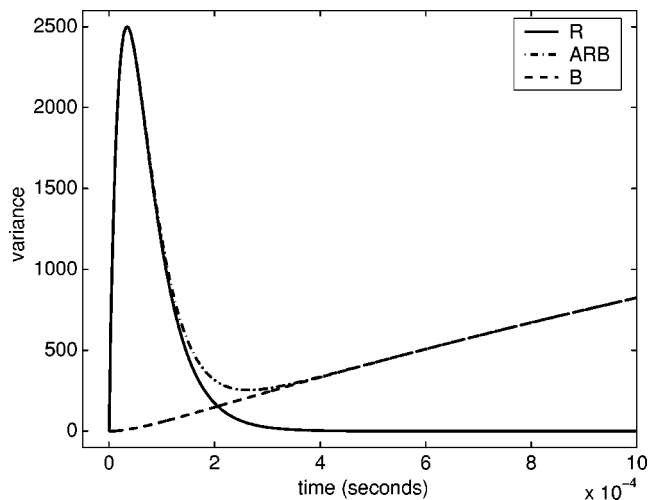Several simplifications can be applied. First, AR can be eliminated because the number of AR molecules has the

same probability distribution as the number of B molecules. Such elimination obviously has no effect on the accuracy of the obtained results. Furthermore, if the number of AB molecules is X_{AB}, the product of X_{AB} and the reaction parameter $c_1^{(2)}$ represents the probability rate for an R molecule to transfer into state ARB. If $c_1 = c_1^{(2)} X_{AB}$, the reaction (28) can be simplified as

$$R \underset{c_2}{\overset{c_1 \quad c_3}{\rightleftarrows}} ARB \rightarrow B.$$
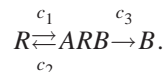
This equation has the same structure as Eq. (22). The only difference is that the value of c_1 varies with time.

A property for many enzyme reactions is that the number of substrate molecules is sufficiently large,[16] and that the number of substrate molecules consumed during the course of the reaction is negligible in comparison to the total number of substrate molecules. Therefore, we can assume that X_{AB} is approximately equal to its initial value, and c_1 can be seen as constant. Thus, after simplification and approximation, the results in (23) can be used here.

As an example, we have calculated the mean and variance of the population of molecule species R, ARB, and B with the following initial values: $X_R(0) = 10\,000$, $X_{AB}(0) = 4.8176 \times 10^{19}$, $X_{AR}(0) = X_{ARB}(0) = X_B(0) = 0$, and reaction parameters $c_2 = 1 \text{ s}^{-1}$, $c_3 = 100 \text{ s}^{-1}$, and $c_1^{(2)} = 4.1930 \times 10^{-16} \text{ s}^{-1}$. The volume in which the reactions are observed is equal to 0.01 l.

The obtained results are shown in Figs. 5–8. The expected population of molecule species are given in Figs. 5 and 6 (zoomed plot). The variance is shown in Figs. 7 and 8.

V. CONCLUSIONS

In this paper we have presented a general approach for studying systems of first-order reactions. By exploiting the independence of the molecules in such systems, we can use this method for studying chemical reactions with any number of molecules and obtain the probability distributions of molecule populations.

With this approach, we have analyzed an enzyme catalyst reaction and have obtained results for both the transient and steady states of the reaction. In addition, we have also derived an analytic solution for the probability distribution of first-order chain reactions. We have found that the Gaussian approximation is more appropriate than the Poisson approximation when the molecule population is large.

[1] A. M. Kierzek, J. Zaim, and P. Zielenkiewicz, J. Biol. Chem. **276**, 8165 (2001).
[2] T. G. Liou and E. J. Campbell, J. Immunol. **157**, 2624 (1996).
[3] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell* (Garland, New York, 2002).
[4] S. Newlands, L. K. Levitt, C. S. Robinson, A. B. C. Karpf, V. R. M. Hodgson, R. P. Wade, and E. C. Hardeman, Genes Dev. **12**, 2748 (1998).
[5] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[6] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
[7] D. A. McQuarrie, J. Appl. Probab. **4**, 413 (1967).
[8] I. J. Laurenzi, J. Chem. Phys. **113**, 3315 (2000).
[9] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[10] J. J. Lukkien, J. P. L. Segers, P. A. J. Hilbers, R. J. Gelten, and A. P. J. Jansen, Phys. Rev. E **58**, 2598 (1998).

FIG. 8. Zoomed plots of the variance for the population of molecule species R (full line), ARB (dot-dashed line), and B (dashed line).

[11]A. B. Stundzia and C. J. Lumsden, J. Comput. Phys. **127**, 196 (1996).

[12]K. De Cock, X. Zhang, M. F. Bugallo, and P. M. Djurić, Proceedings of the 12th European Signal Processing Conference (2004).

[13]G. Grimmett and D. Stirzaker, *Probability and Random Processes* (Oxford University Press, Oxford, 2001).

[14]M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, J. Chem. Phys.

119, 12784 (2003).

[15]K. De Cock, X. Zhang, M. F. Bugallo, and P. M. Djurić, J. Chem. Phys. **121**, 3347 (2004).

[16]X. Zhang, K. De Cock, M. F. Bugallo, and P. M. Djurić, Proceedings of the 2003 IEEE Workshop on Statistical Signal Processing (2003).