AN APPLICATION OF SYSTEM THEORY TO STOCHASTIC MODELS FOR FIRST ORDER CHEMICAL REACTIONS

Xueying Zhang, Katrien De Cock, Mónica F. Bugallo, Petar M. Djurić

Department of Electrical & Computer Engineering,
Stony Brook University,
Stony Brook, New York 11794-2350, USA
sherry,decock,monica,djuric@ece.sunysb.edu

ABSTRACT

A new approach for the computation of probability distributions for coupled first order chemical reactions is introduced. The approach is based on system theory, where the system states are chemical species and the signals are probabilities. We derive the transfer functions of the so defined systems and show that they can be applied to various reaction environments. The use of block diagrams offers a clear, visual, and convenient way to decompose a complicated reaction system into simpler sub-systems and vice versa. Since the state of the system is defined as a molecule species instead of molecule population, with this method one can study chemical reactions involving any number of molecules.

1. INTRODUCTION

In the last 50 years there have been some very important discoveries in the field of biology. The Genome Project and advances in experimental techniques will lead to full knowledge of DNA sequences, identification of most genes, and temporal gene expression patterns for many organisms. Complete understanding of how groupings of genes and related protein reactions interact to orchestrate cellular functions is very important and possible. These interactions are often formulated by coupled chemical reactions. As an example, ribosomal translation of mRNA, the process of ribosome moving along mRNA and translating the mRNA into protein, can be modeled as the following first order reaction chain:

$$\begin{aligned} &Ribosome \cdot mRNA_0 \xrightarrow{k_0} Ribosome \cdot mRNA_1 \xrightarrow{k_1} \dots \\ &\xrightarrow{k_{i-1}} Ribosome \cdot mRNA_i \xrightarrow{k_i} \dots \\ &\xrightarrow{k_{MAX-1}} Ribosome \cdot mRNA_{MAX} \\ &\xrightarrow{k_{MAX}} Ribosome_{free} + mRNA_{free} + protein \end{aligned} \tag{1}$$

Chemical reactions are traditionally studied through the deterministic approach, which assumes that the number of molecules involved in reactions is large, and that the time evolution of the molecule population is deterministic. However, the number of molecules in biological systems can be very small [1, 2]. As an example, it has been estimated that there are about 450 proteins, 30 ribosomes, 340 tRNA molecules, some mRNA molecules, and $30,000$ small organic molecules (e.g. amino acids, sugar, nucleotides, ATP, NAD and others). Many studies have shown that biological processes are actually stochastic [3, 4, 5]. Therefore, in small biological systems, it is often more appropriate to model the chemical reactions in a stochastic way.

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], and the second relies on Monte Carlo simulation methods [7]. In this paper, we consider the first approach.

The objectives here are (i) to introduce a system with probability signals, present a theory of represention of first order reaction networks with system block diagrams and derive solutions from them, (ii) to study the distribution of molecule populations in the presence of multiple sources.

The structure of the paper is as follows. In Section 2 we introduce the elementary system functions. Rules are given for translation of chemical equations into block diagrams. In Section 3, the process of solving probability distribution functions through block diagrams is presented. Subsequently, the case of combining multiple sources is discussed. Finally, in Section 4 we give some concluding remarks.

2. REPRESENTATION OF CHEMICAL REACTIONS BY SYSTEM BLOCK DIAGRAMS

In this section we introduce the representation of first order chemical reactions by system block diagrams. The state of the system is the chemical species involved in the observed chemical reactions. The signal of the system is the probability that a single molecule is in state S_i at time t. This probability is denoted by $p^{(i)}(t)$, and such systems are referred to as probability systems. First, we define the elementary system, and later we use it as a foundation for building more complicated systems. We then discuss a special form of the elementary system function. Additional reactions can be incorporated in the studied biochemical reaction by using rules for expanding system block diagrams.

2.1. Elementary system function

We give the definition of an elementary system as follows:

Definition 1. *An elementary system is defined by the state of interest, a parent state (a previous state), the probability rate constant for a molecule to transfer from the parent state to the state of interest, and the probability rate constant for a molecule to leave the state of interest.*

For a first order reaction μ, the specific probability rate constant c_μ is such that

$$c_\mu \Delta t \equiv \text{the probability per molecule for the first order reaction } \mu \text{ to occur in the very small time interval } \Delta t.$$

Now, consider a single molecule in a first order reaction chain

$$\cdots S_{i-1} \xrightarrow{c_{i-1}} S_i \xrightarrow{c_i} \cdots \tag{2}$$

and suppose that we are interested in the state S_i. Then, the elementary system is composed of the state S_i, its parent state S_{i-1}, the incoming arrow $\xrightarrow{c_{i-1}}$, and the outgoing arrow $\xrightarrow{c_i}$.

Let the probability of an observed molecule in state S_i at time t by denoted by $p^{(i)}(t)$, i.e.,

$$p^{(i)}(t) \equiv \text{the probability that an observed molecule is in state } S_i \text{ at time } t.$$

Following the definition of the specific rate constant c, for a single molecule, it can be derived that

$$p^{(i)}(t+\Delta t) = c_{i-1}\Delta t p^{(i-1)}(t) + (1 - c_i \Delta t) p^{(i)}(t).$$

The rearrangement of this expression with $\Delta t \to 0$, yields a differential equation. By applying the unilateral Laplace transformation, we obtain

$$p^{(i)}(s) = \frac{c_{i-1}}{c_i + s} p^{(i-1)}(s) + \frac{1}{s + c_i} p^{(i)}(t=0). \tag{3}$$

where s is a complex variable used in the definition of the Laplace transform. We assume throughout the rest of this paper that the probability for the molecule to be in any state at time zero is zero, that is, $p^{(i)}(t=0) = 0$. For a species that has molecules at time zero, a source is attached to the corresponding state in the block diagram. This is discussed in more detail in Section 2.2. Then (3) becomes

$$p^{(i)}(s) = \frac{c_{i-1}}{c_i + s} p^{(i-1)}(s)\ . \tag{4}$$

The factor $\frac{c_{i-1}}{c_i+s}$ in (4) is the transfer function of the elementary system. The block representing the system connects the states S_{i-1} and S_i, and is illustrated in Figure 1. The input of the system is the probability that a molecule is in state S_{i-1} at time t. The output is the probability that a molecule is in state S_i, also at time t. From (4), it is easy to see that the probability system is homogeneous, i.e., the output of the system scales linearly with the input.

In summary, the impulse response and the transfer function of the elementary system in (2) are

$$h(t) = c_{i-1} e^{-c_i t}\ ,\ \ H(s) = \frac{c_{i-1}}{s + c_i}. \tag{5}$$

2.2. A special form of the elementary system function

A special form of the elementary system function is the transfer function for the system connected to a source. The source reflects the probability that the observed molecule is introduced to a certain state from the exterior. For example, if $S_i^{(in)}$ is a source connected to state S_i, then we denote it according to

$$S_i^{(in)} \Rightarrow S_i \xrightarrow{c_i} \cdots \tag{6}$$

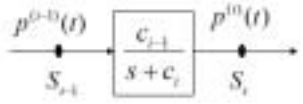

Fig. 1. The system block diagram of reaction (2). The block represents the transfer function of the elementary system. It connects the state S_{i-1} and state S_i. The input of this block is the probability as a function of time t that a molecule is in state S_{i-1}. The output is the probability as a function of t that a molecule is in state S_i.

The detailed balance yields

$$\begin{aligned} p^{(i)}(t+\Delta t) &= p^{(i)}(t)(1 - c_i \Delta t) + p^{(i,in)}(t) \\ &= p^{(i)}(t)(1 - c_i \Delta t) + f^{(i)}(t)\Delta t \end{aligned}$$

where $p^{(i,in)}(t)$ is the probability that a molecule from the source is injected into state S_i in the interval $[t\ ,\ t + \Delta t]$, and $f^{(i)}(t)$ is the source's probability density function of injection into state S_i. Therefore,

$$p^{(i)}(s) = \frac{f^{(i)}(s)}{s + c_i}.$$

The impulse response and the transfer function of the system which connects the source to state S_i is

$$h(t) = e^{-c_i t}\ ,\ H(s) = \frac{1}{s + c_i}\ .$$

Note that the input signal of this system is the probability density function of injection.

For a system where the source injects the observed molecule into state S_i at time $t = 0$, the source's probability density function is $\delta(t)$ whose unilateral Laplace transform is 1. For open systems, possible source functions are $ae^{-a(t-\tau)}$, $a\delta(t-\tau_1)+(1-a)\delta(t-\tau_2)$, and $\sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!}\delta(t - n\tau)$.

2.3. Expansion of an existing system block diagram

When new reactions are included to a reaction network, the system block diagram can be modified using the following rules:

2.3.1. Addition of a parent

Consider the reaction network

$$\begin{array}{l} \cdots S_{i-1} \xrightarrow{c_{i-1}} \quad S_i \quad \xrightarrow{c_i} \cdots \\ \qquad\qquad\quad \uparrow c_j \\ \qquad\qquad\quad S_j \\ \qquad\qquad\quad \vdots \end{array} \tag{7}$$

This can be seen as adding another parent to the state S_i of the reaction chain (2). The modified block diagram of Figure 1 is plotted in Figure 2. Note that $p^{(i-1)}(t)$ is the probability for a molecule to be in state S_{i-1} and $p^{(j)}(t)$ is the probability for the same molecule to be in state S_j. The probability for this molecule to be in state S_i is the sum of the two outputs. This rule reflects the additivity of the probability system.

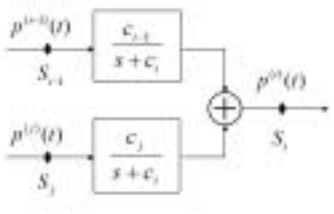

Fig. 2. The system block diagram for reaction network (7). This can be achieved by adding another parent to state S_i of the block diagram in Figure 1.

2.3.2. Addition of a child

In the following reaction network

$$\begin{array}{ccccc} \cdots S_{i-1} & \xrightarrow{c_{i-1}} & S_i & \xrightarrow{c_i^{(1)}} & \cdots \\ & & \downarrow c_i^{(2)} & & \\ & & \vdots & & \end{array} \tag{8}$$

the species S_i has two children. As a result, the modified transfer function has one more term $c_i^{(2)}$ in its denominator compared to that of (2), since there is one more outgoing edge with parameter $c_i^{(2)}$. The modified block diagram is shown in Figure 3.

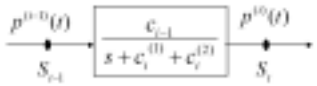

Fig. 3. The system block diagram for reaction network (8). Compared to Figure 1, the transfer function of this system has one more term $c_i^{(2)}$ in its denominator.

Example 1:
The system block diagram for the coupled first order reactions (1) is shown in Figure 4, with $Ribosome \cdot mRNA_i$ $(i = 0, \dots)$ denoted as S_i, $Ribosome{\cdot}mRNA_{MAX}$ as S_{N-1} and $Ribosome_{free}$ as S_N. The possible source is S_0.

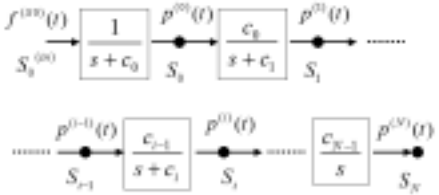

Fig. 4. Block diagram for the first order reaction network (1).

3. FROM SYSTEM BLOCK DIAGRAM TO THE SOLUTION

3.1. From system block diagram to the solution of $p^{(i)}(t)$

In Section 2, we discussed that the probability system has the properties of homogeneity and additivity. Therefore, the probability system is a linear system and the properties of linear systems can be applied.

Example 2:
Assume that $c_i \neq c_j$ when $i \neq j$ ($i = 1, \dots, N$ and $j = 1, \dots, N$), and that the input probability distribution of state S_1 is $f^{(1)}(t) = \delta(t)$. Then, using system theory, the block diagram in Figure 4 leads to the state probabilities for one molecule

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

where

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

Note that if there are c_is equal to c_js, a closed form solution is also possible.

3.2. From $p^{(i)}(t)$ to the probability distribution of the molecule population

Suppose there are x_0 molecules of species A injected into the system from source $A^{(in)}$ with the same probability density $f(t)$, and no molecules of other species. Denote the number of molecules at state S_i by $N^{(i)}$ and let $K_j^{(i)} = 1$ if molecule j is at state S_i and $K_j^{(i)} = 0$ if it is not. Thus $N^{(i)} = \sum_j K_j^{(i)}$. Note that $K_1^{(i)}, K_2^{(i)}, \cdots, K_{x_0}^{(i)}$ are independent Bernoulli random variables with the same probability $p^{(i)}(t)$. Therefore, N_i has a binomial distribution. The probability that there are n molecules in state S_i at time t is

$$P_n^{(i)}(t) = \binom{x_0}{n} \left(p_A^{(i)}(t)\right)^n \left(1 - p_A^{(i)}(t)\right)^{x_0 - n} \tag{9}$$

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

$$\begin{aligned} \mu_i(t) &= x_0 p_A^{(i)}(t) \\ \sigma_i^2(t) &= x_0 p_A^{(i)}(t)\left(1 - p_A^{(i)}(t)\right). \end{aligned}$$

Example 3: A numerical example for the first order reaction chain

Consider the first order reaction chain

$$S_0 \xrightarrow{c_0} S_1 \xrightarrow{c_1} S_2 \xrightarrow{c_2} S_3 \xrightarrow{c_3} S_4 \xrightarrow{c_4} S_5 \xrightarrow{c_5} S_6 \tag{10}$$

Suppose there are 1000 S_0 molecules at time $t = 0$. The values of the reaction parameters $c_0 \dots c_5$ are 4.3, 16.6, 1.2, 2.8, 11.4, 11.9, respectively, with unit s^{-1}. The probability mass of S_6 at time $t = 1.5$ seconds, is shown in Figure 5.

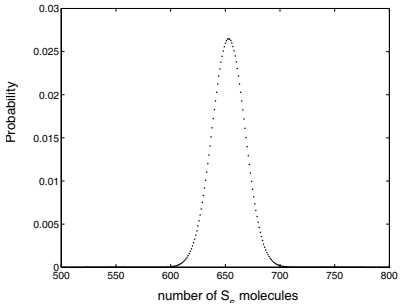

Fig. 5. Direct calculation of the probability mass function of S_6 at time $t = 1.5$ seconds.

When $x_0 \to \infty$ and $p^{(i)}(t) \to 0$ (e.g. $p^{(i)}(t) < 0.1$), the binomial distribution (9) can be approximated by the Poisson distribution with mean $\lambda = x_0 \cdot p^{(i)}(t)$. The purpose of the approximation is to make the analysis more tractable. In many cases, however, the condition $p^{(i)}(t) \to 0$ is not valid. Using the reaction chain (10) and reaction parameter $c_0, \ldots, c_5$ of Example 3, the probabilities of a single molecule to be in state S_i ($i = 0, 1, \ldots, 6$) at time t are drawn in Figure 6, from which it can be seen that during the first 1.5 seconds, many $p^{(i)}(t)$'s have values greater than 0.1. Therefore, during that period the Poisson approximation is not suitable.

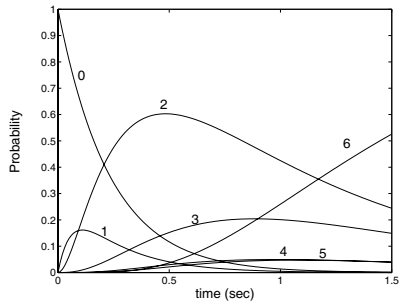

Fig. 6. The probabilities for a single molecule to be in state S_i ($i = 0, 1, \ldots, 6$) at time t.

However, if there are multiple sources which are independent from each other, the probability mass function for state S_i is the convolution of binomial distributions. For example, if there are y_0 molecules injected from source $B^{(in)}$ besides the x_0 molecules injected from source $A^{(in)}$, then

$$\begin{aligned} P_n^{(i)}(t) = \quad & \tbinom{x_0}{n}\big(p_A^{(i)}(t)\big)^n\big(1-p_A^{(i)}(t)\big)^{x_0-n} \\ & \otimes \tbinom{y_0}{n}\big(p_B^{(i)}(t)\big)^n\big(1-p_B^{(i)}(t)\big)^{y_0-n} \end{aligned}$$

Note that "$\otimes$" represents the convolution along the n axis. We refer to each of these binomial distributions as components. The mean $\mu_i(t)$ and variance $\sigma_i^2(t)$ of $P_n^{(i)}(t)$ are

$$\begin{aligned} \mu_i(t) &= x_0 p_A^{(i)}(t) + y_0 p_B^{(i)}(t) \\ \sigma_i^2(t) &= x_0 p_A^{(i)}(t)\big(1-p_A^{(i)}(t)\big) + y_0 p_B^{(i)}(t)\big(1-p_B^{(i)}(t)\big). \end{aligned}$$

When the values of x_0 and y_0 are large, the convolution becomes difficult to compute. To approximate the convolution result, a discretized Gaussian distribution is then employed, for example, as follows:

$$P_n^{(i)}(t) = \frac{1}{D}\frac{1}{\sqrt{2\pi\sigma_i^2(t)}} e^{-\frac{\left(n-\mu_i(t)\right)^2}{2\sigma_i^2(t)}} \sum_{j=0}^{x_0+y_0} \delta(n-j)\,, \quad (11)$$

where

$$D = \sum_{j=0}^{x_0+y_0} \frac{1}{\sqrt{2\pi\sigma_i^2(t)}} e^{-\frac{\left(n-\mu_i(t)\right)^2}{2\sigma_i^2(t)}}.$$

4. CONCLUSION

In this paper we have presented a new approach for studying first order reaction networks. By defining the states of a probability signal system as molecule species instead of molecule populations, we can use the method for studying chemical reactions with any number of molecules. The system block diagrams that are employed in our approach can provide visual decomposition of complicated networks into simpler networks and vice versa. We have also shown how to obtain the probability distributions of molecule populations. With this approach, we have derived an analytic solution for the probability distribution of the first order chain reaction. Under the situation where the molecule population is large, we have found that the Gaussian approximation is more appropriate than the Poisson approximation.

5. REFERENCES

[1] A. M. Kierzek, J. Zaim, and P. Zielenkiewicz, "The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression," *Journal of Biological Chemistry*, vol. 276(11), pp. 8165–8172, 2001.

[2] V. A. Kuznetsov, G. D. Knott, and R. F. Bonner, "General statistics of stochastic process of gene expression in eukaryotic cells," *Genetics*, vol. 161, no. 3, pp. 1321 – 1332, July 2002.

[3] S. Newlands, L. K. Levitt, C. S. Robinson, A.B. C. Karpf, V. R.M. Hodgson, R. P. Wade, and E. C. Hardeman, "Transcription occurs in pulses in muscle fibers," *Gene Development*, vol. 12(17), pp. 2748–2758, 1998.

[4] I. L. Ross, C. M. Browne, and D.A. Hume, "Transcription of individual genes in eukaryotic cells occurs randomly and infrequently," *Immunology and cell biology*, vol. 72, pp. 177–185, 1994.

[5] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, no. 5584, pp. 1183–1186, August 2002.

[6] I. J. Laurenzi, "An analytical solution of the stochastic master equation for reversible bimolecular reaction kinetics," *Journal of Chemical Physics*, vol. 113, no. 8, pp. 3315, August 2000.

[7] J. J. Lukkien, J. P. L. Segers, P. A. J. Hilbers, R. J. Gelten, and A. P. J. Jansen, "Efficient Monte Carlo methods for the simulation of catalytic surface reactions," *Physical Review E*, vol. 58, pp. 2598–2610, 1998.