The Third International Symposium on Optimization and Systems Biology (OSB'09)
Zhangjiajie, China, September 20–22, 2009

Nonlinear Coupling for Improved Stochastic Network Model: A Study of Schnakenberg Model*

Youfang Cao[1,2] Jie Liang[1,2]

[1]Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Department of Bioengineering, University of Illinois at Chicago, Chicago 60607, USA

Abstract Langevin equation is widely used to study the stochastic effects in molecular networks, as it often approximates well the underlying chemical master equation. However, frequently it is not clear when such an approximation is applicable and when it breaks down. Here we study the simple Schnakenberg model consisting of two molecular species whose concentrations vary and three reversible reactions. To reduce residual errors from the conventional formulation of the Langevin equation, we propose to explicitly model the effective coupling between macroscopic concentrations of different molecular species. Our preliminary results show that this formulation is effective in correcting residual errors from the original uncoupled Langevin equation and can approximate the underlying chemical master equation very accurately.

Keywords Schnakenberg model; Langevin Equation; Master Equation; Noise

1 Introduction

Chemical master equation (CME) provides a fundamental theoretical framework for modeling biochemical reaction systems. However, master equation is difficult to solve analytically, except in a few simple special cases. Currently, there are two approaches for obtaining solutions to the master equation. The first is the widely used stochastic simulation algorithm [3]. This is an Monte Carlo sampling method, but has serious limitations. For example, it is difficult to know *a priori* whether the amount of sampling is sufficient, especially when facing the difficult task of estimating small probabilities of biological events that occur rarely but are critically important. Stochastic simulations can be very inefficient in capturing these rare events of interests.

The second is to solve the chemical master equation computationally. Methods in reducing the state space with controlled error bounds have been developed for achieving this goal [7]. Recently, progress has also been made in the optimal enumeration of the state space of the underlying Markov process of the CME, and in obtaining the exact

*This study was supported by a phase II 985 Project (Sub-Project: T226208001 at Shanghai Jiao Tong University, Shanghai, China), NIH grants GM079804-01A1, GM081682, NSF grants DBI-0646035 and DMS-0800257, and ONR N00014-09-1-0028.

stationary probability distribution over these states [1]. It was shown that these techniques can be applied to realistic biological problems, including the well-known system of the epigenetic switch of phage lambda [2].

However, to go beyond biochemical networks of small or moderate size, whose chemical master equations cannot be solved directly, one needs to approximate the stochastic biochemical system. One widely used approach is that of the Langevin equation. van Kampen summarized the "Langevin approach" as one consisting of two terms: The *drift* term represent the macroscopic deterministic part of the system, and the *diffusion* term represents the intrinsic stochasticity or noise in the system. The basic form of Langevin equation is:

$$\frac{d\mathbf{X}}{dt} = \mu(\mathbf{X}) + \sigma(\mathbf{X})\mathcal{N}(0, \frac{1}{dt}). \tag{1}$$

Here $\mathbf{X}$ is the vector of concentrations of molecular species in the reaction system, $\mu(\mathbf{X})$ the deterministic component of the equation, or the drift term. The second term is the diffusion term. Here $\mathcal{N}(0, 1/dt)$ is a vector of 1D Gaussians, with zero mean and $1/dt$ variance. The coefficient $\sigma^2(\mathbf{X})$ controls the amplitude of the Gaussian noise. It can be either a function of $\mathbf{X}$ or a constant. The key issue in developing Langevin models for biochemical networks is to determine $\mu(\mathbf{X})$ and $\sigma^2(\mathbf{X})$. When $\sigma^2(\mathbf{X})$ is a vector of constants, one adjusts its values so the variance of the Gaussian noise produce the correct fluctuations in the system [12].

In this study, we explore a novel approach for developing stochastic models of biochemical networks extending the Langevin formulation. Using the well-studied Schnakenberg's model as the example, we show that by explicitly incorporating the coupling of macroscopic concentrations of different molecular species, a more accurate model whose solution approaches the original chemical master equation can be obtained. This paper is organized as follows: we describe the development of the coupling Langevin equation in the next section. We then present results of numerical solutions and analysis. This is followed by discussion and conclusion.

2 Models

Schnakenberg model

The Schnakenberg model was originally developed for studying the limit cycle behavior in a simple chemical reaction system [11]. This system consists of only two reacting components. However, such a simple system can already produce complex behavior such as oscillation. There have been many theoretical studies on the Schnakenberg model (see [10, 9] for recent examples).

The Schnakenberg model used in this study consists of three reversible reactions:

$$X \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} A, \qquad B \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} Y, \qquad 2X + Y \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} 3X, \tag{2}$$

where X and Y are reacting species of the system, and A and B are external reactants whose copy numbers or concentrations are fixed constants. Each reaction has a corresponding microscopic reaction rate. The fixed copy numbers or concentrations of A and B can be adjusted, which lead to different behavior of the system.

Chemical master equation (CME)

The chemical master equation for the Schnakenberg model can be written as:

$$\begin{aligned}\frac{dp(X,Y,t)}{dt} &= k_{-1}Ap(X-1,Y,t)+k_2Bp(X,Y-1,t)+\frac{k_3}{V^2}(X-1)(X-2)(Y+1)p(X-1,Y+1,t)\\ &+k_1(X+1)p(X+1,Y,t)+k_{-2}(Y+1)p(X,Y+1,t)+\frac{k_{-3}}{V^2}(X+1)X(X-1)p(X+1,Y-1)\\ &-(k_{-1}a+k_2b+\frac{k_3}{V^2}X(X-1)Y+k_1X+k_{-2}Y+\frac{k_{-3}}{V^2}X(X-1)(X-2))p(X,Y,t),\end{aligned} \tag{3}$$

where X, Y, A and B are copy numbers of molecular species X, Y, A and B. $p(X,Y,t)$ is the probability that the system has the state of X copies of X and Y copies of Y at time t. V represents the volume of the system.

Macroscopic deterministic model (ODE)

The macroscopic deterministic model or the ODE model provides the first degree approximation to the master equation. It does not model the stochasticity in the system as it contains no noise-related diffusion term. The deterministic equations of the concentration of X and Y can be written as:

$$\frac{dx}{dt}=k_{-1}a-k_1x+k_3x^2y-k_{-3}x^3, \quad \text{and} \quad \frac{dy}{dt}=k_2b-k_{-2}y-k_3x^2y+k_{-3}x^3, \tag{4}$$

in which x, y, a and b are concentrations of molecular species X, Y, A and B, respectively. Concentrations can be calculated from copy numbers of corresponding molecular species after division by the system volume V.

Coupling Langevin Equation

The drift term. The noise-related diffusion term is added to the deterministic ODE model to account for the stochasticity in the system. We first derive the differential form of the mean of X and Y from the chemical master equation. After multiplying both sides of Eqn. (3) with X and sum over X and Y from 0 to ∞, we combine similar terms, and re-index the summations. We obtain:

$$\begin{aligned}\frac{d\sum_{X=0}^{\infty}\sum_{Y=0}^{\infty}Xp(X,Y,t)}{dt} &= k_{-1}A\sum_{X=0}^{\infty}\sum_{Y=0}^{\infty}p(X,Y,t)-k_1\sum_{X=0}^{\infty}\sum_{Y=0}^{\infty}Xp(X,Y,t)\\ &+\frac{k_3}{V^2}\sum_{X=0}^{\infty}\sum_{Y=0}^{\infty}X(X-1)Yp(X,Y,t)-\frac{k_{-3}}{V^2}\sum_{X=0}^{\infty}\sum_{Y=0}^{\infty}X(X-1)(X-2)p(X,Y,t).\end{aligned} \tag{5}$$

That is,

$$\frac{d\langle X\rangle}{dt}=k_{-1}A-k_1\langle X\rangle+\frac{k_3}{V^2}\langle X(X-1)Y\rangle-\frac{k_{-3}}{V^2}\langle X(X-1)(X-2)\rangle, \tag{6}$$

where $\langle\cdots\rangle$ stands for the mean of stochastic variable "$\cdots$". Similarly, we have for $\langle Y\rangle$:

$$\frac{d\langle Y\rangle}{dt}=k_2B-k_{-2}\langle Y\rangle-\frac{k_3}{V^2}\langle X(X-1)Y\rangle+\frac{k_{-3}}{V^2}\langle X(X-1)(X-2)\rangle \tag{7}$$

It is easy to see that Eqn (6) and Eqn (7) are equivalent to the macroscopic deterministic equations Eqn (4), after both sides are divided by the system volume V and $\langle X\rangle$ and $\langle Y\rangle$ are transformed into concentration (Eqn (8)). We will use these as the drift term of the Langevin equation:

$$\mu_x(x,y)=k_{-1}a-k_1x+k_3x^2y-k_{-3}x^3, \quad \text{and} \quad \mu_y(x,y)=k_2b-k_{-2}y-k_3x^2y+k_{-3}x^3 \tag{8}$$

The diffusion term. As in Eqn 1, we use Gaussian noise to model the diffusion term. The coefficient of the diffusion term can be derived from the variance of the copy numbers of X and Y. By the definition of $\sigma^2(\mathbf{X}) = \langle \mathbf{X}^2 \rangle - \langle \mathbf{X} \rangle^2$, we have:

$$\frac{d\sigma^2(\mathbf{X})}{dt} = \frac{d\langle \mathbf{X}^2 \rangle}{dt} - 2\langle \mathbf{X} \rangle \frac{d\langle \mathbf{X} \rangle}{dt}. \tag{9}$$

Here, $\frac{d\langle \mathbf{X} \rangle}{dt}$ has already been derived (Eqn 6). $\frac{d\langle \mathbf{X}^2 \rangle}{dt}$ can be obtained using the same approach as that of Eqn (6) and Eqn (7). After multiplying both sides of Eqn (3) with X^2 and sum over X and Y from 0 to ∞, we combine the terms, and re-index them. We obtain:

$$\begin{aligned} \frac{d\langle X^2 \rangle}{dt} &= 2k_{-1}A\langle X \rangle + k_{-1}A - 2k_1\langle X^2 \rangle + k_1\langle X \rangle + \frac{2k_3}{V^2}\langle X^2(X-1)Y \rangle \\ &+ \frac{k_3}{V^2}\langle X(X-1)Y \rangle - \frac{2k_{-3}}{V^2}\langle X^2(X-1)(X-2) \rangle + \frac{k_{-3}}{V^2}\langle X(X-1)(X-2) \rangle \end{aligned} \tag{10}$$

Substituting Eqn. (10) and (6) into Eqn. (9), we now have obtained $\frac{d\sigma^2(X)}{dt}$ as:

$$\frac{d\sigma_X^2(X,Y)}{dt} = \frac{d\langle X^2 \rangle}{dt} - 2\langle X \rangle \frac{d\langle X \rangle}{dt} = k_{-1}A + k_1\langle X \rangle + \frac{k_3}{V^2}\langle X(X-1)Y \rangle + \frac{k_{-3}}{V^2}\langle X(X-1)(X-2) \rangle \tag{11}$$

In order to convert the copy numbers X and Y into concentrations, Eqn (11) is divided by V^2. We arrive at the expression for the variance of the concentration of molecular species X as:

$$\boldsymbol{\sigma^2}(\mathbf{x}) = \frac{d\sigma_x^2(x,y)}{dt} = \frac{1}{V}(k_{-1}a + k_1x + k_3x^2y + k_{-3}x^3) \tag{12}$$

This is also the variance in Eqn (1). Similarly, we have:

$$\boldsymbol{\sigma^2}(\mathbf{y}) = \frac{1}{V}(k_2b + k_{-2}y + k_3x^2y + k_{-3}x^3) \tag{13}$$

We can now obtain the Langevin equation for the Schnakenberg model from its underlying chemical master equation:

$$\begin{aligned} \frac{dx}{dt} &= (k_{-1}a - k_1x + k_3x^2y - k_{-3}x^3) + \sqrt{\frac{1}{V}(k_{-1}a + k_1x + k_3x^2y + k_{-3}x^3)}\mathcal{N}(0, \frac{1}{dt}) \\ \frac{dy}{dt} &= (k_2b - k_{-2}y - k_3x^2y + k_{-3}x^3) + \sqrt{\frac{1}{V}(k_2b + k_{-2}y + k_3x^2y + k_{-3}x^3)}\mathcal{N}(0, \frac{1}{dt}) \end{aligned} \tag{14}$$

We call this the uncoupled Langevin equation (ULE).

The coupling terms between molecular species. Although the Langevin equations obtained thus far can approximate the chemical master equation well, effects of higher order coupling intrinsic in complex biochemical systems are not accounted for in this model. For example, the molecular species X and Y in the Schnakenberg model are strongly coupled, but they are treated as independent random variables in the present form of the Langevin equation. In order to obtain better approximation to the chemical

master equation, we now extend the conventional Langevin formulation and add explicitly a coupling term.

To account for the coupling effect between X and Y, we derive the $\frac{d\langle XY\rangle}{dt}$ term from the chemical master equation Eqn (3) using similar approach as before:

$$\frac{d\langle XY\rangle}{dt} = k_{-1}A\langle Y\rangle + k_2B\langle X\rangle - (k_1+k_{-2})\langle XY\rangle - \frac{k_3}{V^2}\langle XY(X-1)(X-Y+1)\rangle + \frac{k_{-3}}{V^2}\langle X(X-1)(X-2)(X-Y-1)\rangle. \tag{15}$$

In the form of concentrations, we have:

$$\frac{dxy}{dt} = k_{-1}ay + k_2bx - (k_1+k_{-2})xy - k_3x^2y(x-y) + k_{-3}x^3(x-y) \tag{16}$$

To add this coupling term to the Langevin equation with proper dimension, we decompose this term by dividing it with y and x, respectively, to obtain $\frac{1}{y}\frac{dxy}{dt}$ and $\frac{1}{x}\frac{dxy}{dt}$. These are then added to Eqn (14) with a scaling coefficient α:

$$\begin{aligned}\frac{dx}{dt} &= (k_{-1}a - k_1x + k_3x^2y - k_{-3}x^3) + \sqrt{\frac{1}{V}(k_{-1}a + k_1x + k_3x^2y + k_{-3}x^3)}\mathcal{N}(0,\frac{1}{dt}) + \alpha\frac{dxy}{ydt}\\ \frac{dy}{dt} &= (k_2b - k_{-2}y - k_3x^2y + k_{-3}x^3) + \sqrt{\frac{1}{V}(k_2b + k_{-2}y + k_3x^2y + k_{-3}x^3)}\mathcal{N}(0,\frac{1}{dt}) + \alpha\frac{dxy}{xdt}\end{aligned} \tag{17}$$

The determination of the scaling coefficient α, analysis of the computed results, and comparisons to other Langevin equation-based models are discussed below.

3 Numerical Simulations and Analysis

We have computed the exact steady state probability distributions of the system at different concentrations of X and Y by solving the chemical master equation using the method described in [1] (Fig 1a and 1b, CME for chemical master equation). The reaction rates are set at fixed value of $k_1 = k_2 = k_3 = k_{-1} = 1.0$ and $k_{-2} = k_{-3} = 0.01$. The system volume is set to a constant of $V = 100$. We use two sets of copy numbers of (A,B) of $(10,50)$ and $(20,40)$ for the fixed parameters A and B in the Schnakenberg model in our calculations. We have also reconstructed the probability distributions based on $200,000$ simulations of the improved coupling Langevin equation (Fig 1c and 1d, CLE for coupling Langevin Equation), and of the original uncoupled Langevin equation (Fig 1e and 1f, ULE for uncoupled Langevin Equation). For numerical simulation of the Langevin Equations, we follow Ito's formula [8] and use the Eular-Maruyama method [5]. In addition, we have computed the trajectories of the concentrations of X and Y evolving at different time according to the deterministic model (Fig 1g and 1h, ODE for Ordinary Differential Equation). Here the fixed constants a and b are set at values of $(0.1,0.5)$ and $(0.2,0.4)$, which are equivalent to the values of copy numbers used for the stochastic models.

Comparison of results from models of chemical master equations, Langevin equation, and ODE models

At the parameter values of $A = 10$ and $B = 50$, the well-known oscillating limit cycle behavior of the Schnakenberg model can be seen in Fig 1g. At $A = 20$ and $B = 40$,

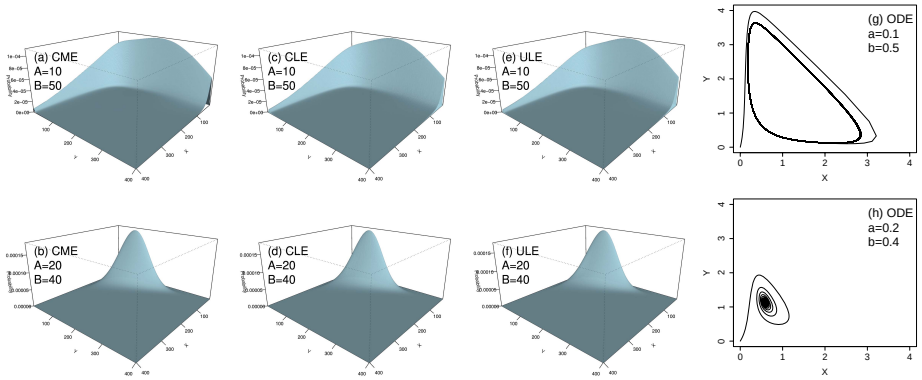

Figure 1: Calculated probability distributions over different copy numbers of X and Y and the trajectories of evolving concentrations of X and Y. (a) and (b): Exact probability distributions over copy numbers X and Y obtained by solving the chemical master equation (CME). Two sets of copy numbers of (A,B) at $(10,50)$ and $(20,40)$ are used for the fixed parameters A and B; (c) and (d): Reconstructed probability distributions over X and Y obtained from $200,000$ simulations of the coupling Langevin equation (CLE); (e) and (f): Reconstructed probability distributions over X and Y obtained from $200,000$ simulations of the original uncoupled Langevin equation (ULE); and (g) and (h): Trajectories of evolving concentrations of X and Y according to the deterministic ordinary differential equation (ODE). Here (g) shows the well-known oscillating limit cycle behavior of the Schnakenberg model, and (h) shows the convergence towards a fixed point. The concentrations of A and B are set at values equivalent to the copy numbers used in stochastic models.

the behavior of the system converges towards a fixed point (Fig 1h). The landscapes of the steady state probability distributions obtained from solving both the chemical master equation and the two forms of the Langevin equation all show a crater surrounded by a mountainous ridge for the parameter set of $A = 10$ and $B = 50$ (detailed figures not shown). This corresponds well with the limit cycle behavior observed in the ODE model. At $A = 20$ and $B = 40$, the landscapes show a single peak, which again corresponds well with the fixed-point behavior observed in the ODE model.

We find that the model of Langevin equation generally approximates the probability landscape obtained directly from chemical mater equation well. From the comparisons of results from 25 different parameter sets (detailed data not shown), we find that overall the location of the maximum probabilities in copy numbers of X and Y from the Langevin equation is very close to that obtained from the chemical master equation (Fig 2c, red circles). This indicates that the noise term derived from the chemical master equation works very well for the Schnakenberg model.

In contrast, we find the location of the fixed point can deviate significantly from that of the maximum probability derived from the CME model, especially when the landscape is relatively flat and the maximum probability is less pronounced (Fig 2c, black squares).

Overall, our results show that the noise term plays essential role in determining the behavior of the Schnakenberg system, as the distance of maxima between solutions from the Langevin equation and the chemical master equation is usually less than the distance

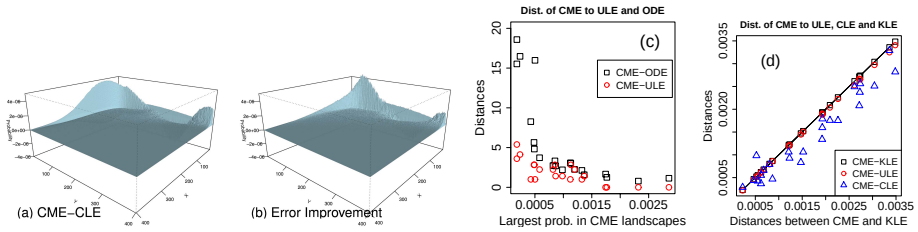

Figure 2: Comparison of errors between different models. (a): Difference in probability landscape of the uncoupled Langevin equation (ULE) and that of the chemical master equation (CME). This represents errors in ULE. (b): The amount of the errors in (a) that are corrected by the Coupling Langevin Equation (CLE) model. This shows the improvement due to the coupling term. (c): Distances between the locations of the maxima of the probability distribution landscapes obtained from the chemical master equation (CME), the uncoupled Langevin equation (ULE), and the locations of the fixed points derived from the ODE model. Here the X-axis is the maximum probability value of the landscape derived from the CME, and Y-axis is the Euclidean distance between the locations of the maximum points by CME and ULE, and the locations of the maximum point by CME and the fixed point derived from the ODE model. Here locations have coordinates in copy numbers of X and Y. Black squares are comparisons between the CME and the ODE models, and the red circles are comparisons between the CME and the ULE models. (d): Comparisons of distances from the CME landscape to the landscape from the Kurtz-Langevin equation (KLE), to the landscape from our uncoupled Langevin equation (ULE), and to that of the coupling Langevin equation (CLE). Here the distance between CME and KLE are plotted on the x-axis, and the distances between CME and ULE (red circle) as well as between CME and CLE (blue triangle) are plotted on the y-axis. The scaling coefficient for the coupling terms is $\alpha = 0.02$.

of maxima and fixed point between the solutions from the chemical master equation and the ODE. In Fig. 2c, most of the red circles are below the black squares.

Effects of the coupling between molecular species

For comparison, we have also performed simulations using a different formulation of the Langevin equation for the same Schnakenberg model following Kurtz and Gillespie [6, 4]. We compare the difference between two landscapes a and b by integrating the overall differences in probabilities: $d_{a,b} = [\sum_{(X,Y)}[p_a(X,Y) - p_b(X,Y)]^2]^{1/2}$. For each of the 25 parameter sets, we calculate the difference $d_{CME,KLE}$ between the landscapes from CME and from the Kurtz-Langevin equation formulation, the difference of $d_{CME,ULE}$ between CME and the uncoupled Langevin equation (ULE), as well as the difference $d_{CME,CLE}$ between CME and the coupling Langevin equation (CLE). The results are plotted in Fig 2d. Both the uncoupled and coupling Langevin equations we have obtained in this study have smaller differences from the exact landscape probabilities obtained from the chemical master equation than the solution obtain from the Kurtz-Langevin equation. We find that our model following the original uncouple Langevin equation performs as good as or slightly better than the Kurtz-Langevin model (Fig. 2d, red circle).

However, it is the coupling Langevin equation that performs significantly better than the Kurtz-Langevin equation as well as the uncoupled Langevin equation, as can be seen in Fig. 2 a-d, where the errors in ULE and the amount of correction by CLE are shown.

These results suggest that the introduction of the coupling effects of the macroscopic concentrations of the molecular species can significantly improve the accuracy of the stochastic model of biochemical networks.

4 Discussions

In this study, we have developed a novel formulation to account for the stochasticity in the biochemical networks. We introduce the noise term to the Langevin equation from the master equation formulation of the Schnakenberg model. An important development is the additional incorporation of the nonlinear coupling effect between molecular species. We find with simulations that for the Schnakenberg model, our coupling Langevin equation model provides the best approximation to the probability landscape of the underlying chemical master equation. Our approach is general and can be applied to other biochemical networks as well. Our work suggest that the coupling effects are not negligible and should be accounted for accurately.

Acknowledgements

We thank Profs Ping Ao and Hong Qian for helpful discussions.

References

[1] Youfang Cao and Jie Liang. Optimal enumeration of state space of finitely buffered stochastic molecular networks and exact computation of steady state landscape probability. *BMC Systems Biology*, 2:30, 2008.

[2] Youfang Cao, Hsiao-Mei Lu, and Jie Liang. Stochastic probability landscape model for switching efficiency, robustness, and differential threshold for induction of genetic circuit in phage lambda. *Conf Proc IEEE Eng Med Biol Soc.*, 1:611–614, 2008.

[3] Daniel. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81:2340–2361, 1977.

[4] Daniel. T. Gillespie. The chemical langevin equation. *Journal of Chemical Physics*, 113:297–306, 2000.

[5] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin, 1999.

[6] Thomas G. Kurtz. Strong approximation theorems for density dependent markov chains. *Stochastic Processes and their Applications*, 6:233–240, 1978.

[7] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics*, 124(4):044104, 2006.

[8] Bernt Oksendal. *Stochastic Differential Equations*. Springer, 2003.

[9] Hong Qian. Open-system nonequilibrium steady state: Statistical thermodynamics, fluctuations, and chemical oscillations. *The Journal of Physical Chemistry B*, 110:15063–15074, 2006.

[10] Hong Qian, Saveez Saffarian, and Elliot L. Elson. Concentration fluctuations in a mesoscopic oscillating chemical reaction system. *PNAS*, 99:10376–10381, 2002.

[11] J. Schnakenberg. Simple chemical reaction systems with limit cycle behaviour. *Journal of Theoretical Biology*, 81:389–400, 1979.

[12] N.G. Van Kampen. *Stochastic processes in physics and chemistry, 3rd Edition*. Elsevier Science and Technology books, 2007.