Published in final edited form as:

Curr Opin Chem Eng. 2014 August 1; 5: 90–95. doi:10.1016/j.coche.2014.07.001.

Solution of Chemical Master Equations for Nonlinear Stochastic Reaction Networks

Patrick Smadbeck and Yiannis N. Kaznessis*

Department of Chemical Engineering and Materials Science, University of Minnesota,
Minneapolis, MN 55455, USA

Abstract

Stochasticity in the dynamics of small reacting systems requires discrete-probabilistic models of reaction kinetics instead of traditional continuous-deterministic ones. The master probability equation is a complete model of randomly evolving molecular populations. Because of its ambitious character, the master equation remained unsolved for all but the simplest of molecular interaction networks. With the first solution of chemical master equations, a wide range of experimental observations of small-system interactions may be mathematically conceptualized.

Keywords

Stochastic Reaction Networks; Chemical Master Equation; Closure Scheme

Introduction

Ramkrishna and Amundson discussed in a compelling 2004 review how mathematical modeling of chemical reaction phenomena propelled chemical engineering to its unique, successful identity [1]. For example, the work by Billous and Amundson set the foundation for chemical reactor stability and sensitivity analysis [2, 3]. Aris and Amundson then built on this foundation, developing the chemical reactor control theories still taught in chemical engineering programs worldwide [4, 5].

An argument can be made, in an analogous way, that mathematical models of biological systems can become an important component for progress in biological engineering. Of course, although the principles of kinetics apply to biological systems, these systems differ from industrial-scale chemical systems in numerous fundamental ways, including this one: biomolecular systems are often far from the thermodynamic limit.

The thermodynamic limit is theoretically attained when the number of molecules of chemical species in the system increases towards infinity. If reactants of a reaction are near

© Â© 2XXX Elsevier Ltd. All rights reserved.

*Corresponding author's yiannis@umn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

the thermodynamic limit, continuous-deterministic modeling formalisms, i.e., using ordinary differential equations, are accurate representations of reality.

Biomolecular systems cannot be assumed to be near the thermodynamic limit. For example, if DNA is one of the interacting components, as is often the case in transcription kinetics, the number of DNA molecule copies in a cellular organism may be as small as 1. In such cases, using ordinary differential equations for simulating the reaction kinetics is bound to result into erroneous results. Stochastic models that account for the randomness of small reacting systems are better suited to model biomolecular systems.

The importance of modeling formalisms appropriate for systems away from the thermodynamic limit was recognized more than 50 years ago by McQuarrie, Moyal and Oppenheim, among others [6–10]. These physical chemists developed the chemical master equation (CME) that follows the time changes of the probability the state is at any point in the available state space.

The common form of the CME is equivalent to the more general Chapman-Kolmogorov equation as applied to Markov processes [7]. In particular, reaction events are modeled with a Markov chain with a discrete set of possible states, or “state space”, occurring in continuous time. Here, states refer to numbers of molecules present in the system. In general, for time t , this will be a vector $X(t) = [X_1 \dots X_N]$, where X_i is number of molecules of the i -th chemical species with $i = 1, \dots, N$.

Transitions between states of the Markov chain occur when a chemical reaction occurs. Reactions in biological systems may include covalent reactions, bindings, conformational changes, transcriptional elongation events, etc.

The general form of the CME is:

$$\frac{\partial P(\underline{X};t)}{\partial t} = \sum_{\underline{X}'} [T(\underline{X}|\underline{X}')P(\underline{X}';t) - T(\underline{X}'|\underline{X})P(\underline{X};t)] \quad (1)$$

where $P(\underline{X};t)$ is the probability of being in state $\underline{X}$ at time t . The transition probability $T(\underline{X}|\underline{X}')$ of going from state $\underline{X}'$ to state $\underline{X}$ per unit time is completely determined by the reaction event kinetics.

The CME describes the dynamics of stochastic systems exactly, but has been until recently mathematically intractable for all but the simplest of linear systems [6]. The reason analytical solutions to the CME remained elusive becomes clear when the master equation is recast in equivalent terms of probability moments – the probability distribution average, the variance, and so on:

$$\frac{\partial \underline{\mu}}{\partial t} = A \underline{\mu} + A' \underline{\mu}' \quad (2)$$

where $\underline{\mu}$ is the vector of moments up to order M and A is the matrix describing the linear portion of the moment equations. On the right, $\underline{\mu}'$ is the vector of higher-order moments, and the corresponding matrix A' .

Generating the matrices in Equation 2 can be performed either analytically [11, 12] or numerically [13]. For linear systems with only 0th or 1st-order reactions, A' is empty. For other systems, A' is not empty and the set of ODEs becomes infinite, and thus intractable.

Without the ability to close the set of equations in (2), or otherwise solve the CME, scientists turned to approximations. In 1976, Daniel Gillespie presented a computer algorithm to sample the master probability distribution with numerical simulations of networks of reactions [14, 15]. Although Gillespie's methods were not widely recognized for almost 20 years, his algorithms found fertile ground for development in efforts to model biological systems. Nowadays, a community of scientists and engineers is continually working to improve the computational efficiency and accuracy of algorithms that simulate chemical reacting systems [16–25].

We joined the efforts of this community, working to improve the computational efficiency and accuracy of algorithms that simulate stochastic chemical reacting systems. We developed a hybrid stochastic-discrete and stochastic-continuous modeling formalism for treating reacting systems that span multiple time scales [26, 27]. We worked on probabilistic steady-state approximations [28], stochastic model reduction techniques [29, 30, 31], and adaptive time-stepping algorithms for stochastic differential equations [32]. We made our algorithms publicly available on sourceforge.net [33–35] and used them to simulate gene regulatory networks [36–41].

Zero-information closure of the master chemical equation

It was our work on probability moments [12, 13] that led to the solution of the master equation. We modeled the master probability with its moments, i.e., the mean, variance, skewness, etc. (Equation 2). As explained, the challenge has always been that for nonlinear reaction networks, lower order moments depend on higher order ones (the mean depends on the variance, the variance depends on the skewness, and so on so forth). There is then no closure scheme.

However, we quickly realized that although the numerical value of higher order moments is too significant to be neglected, higher order moments contribute little information in reconstructing the master probability. Our zero information closure (ZI-closure) scheme then finds the lower order moments by maximizing the information entropy of any reaction network [42].

This section briefly explains the zero-information moment closure method. It concludes with the example of the Schlögl model in order to clarify the calculation of several important equations [43]. This section expands upon a previously described ODE solving scheme and steady-state determination method [42].

For simplicity's sake we limit the discussion to one-dimensional systems. We begin by defining the information entropy for a single component system with a probability distribution $p(x)$:

$$H = - \sum_{x=0}^{\infty} p(x) \ln p(x) \quad (3)$$

For an unconstrained system the resulting maximum-entropy distribution is simply uniform. However, values of the lower-order moments, $\underline{\mu}$, act as constraints on the system. The result is best solved using a Lagrange multiplier method (here assuming a simple component with M known lower order moments):

$$\Lambda = H - \lambda_0 g_0 - \lambda_1 g_1 - \dots - \lambda_M g_M \quad g_0 = \sum_{x=0}^{\infty} p(x) - 1 \quad g_1 = \sum_{x=0}^{\infty} x p(x) - \langle x \rangle \quad \vdots \quad g_M = \sum_{x=0}^{\infty} x^M p(x) - \langle x^M \rangle \quad (4)$$

The maximum is found by differentiating by $p(x)$ and setting the result to zero:

$$\frac{\partial \Lambda}{\partial p(x)} = -\ln p(x) - 1 - \lambda_0 - \lambda_1 x - \dots - \lambda_M x^M = 0 \quad (5)$$

or, trivially:

$$p_H(x) = \exp(-1 - \lambda_0 - \lambda_1 x - \dots - \lambda_M x^M) \quad (6)$$

An analytical expression for the maximum entropy distribution is determined with the same number of parameters as the number of known lower-order moments.

The maximum-entropy moments, denoted with a subscript H, can now be determined. Take, for example, the determination of $\langle x^m \rangle_H$ in a single component system:

$$\langle x^m \rangle_H = \sum_{x=0}^{\infty} x^m p_H(x) \quad (7)$$

Zero-information closure uses maximum-entropy moments as approximations for higher-order moments in simulation:

$$\underline{\mu}' = \underline{\mu}'_H \quad (14)$$

Note that the order of closure, M, is chosen with trial-and-error since it is not a priori clear what order closure scheme will be accurate enough for any nonlinear reaction network.

We demonstrate the accuracy and utility of ZI-closure with the Schlögl model. The structure of this model along with baseline parameter values are presented in Table 1. The Schlögl reaction network has served as a canonical model of one-dimensional stochastic processes with a non-Gaussian probability distribution. Because it has one degree of freedom, the Schlögl model is easily tractable. Because the underlying probability distribution is bimodal for a wide range of parameter values, the Schlögl model affords the study of nonlinear effects on stochastic dynamics.

When the Schlögl model exhibits two main peaks in its probability distribution, up to 12 probability distribution moments may be required to accurately reproduce the underlying distribution. Figure 1 demonstrates how accuracy improves as the closure-order is increased. Several maximum entropy optimizations are executed using an increasing closure order (from 2nd to 10th-order closure). The comparison is made with the steady state probability distribution obtained from averaging over 10⁷ stochastic simulation algorithm trajectories.

An algorithm equivalent to the next-reaction SSA developed by Gillespie and improved by Gibson and Bruck [16] was used to generate all stochastic trajectories. The algorithm was implemented in our own Hy3S algorithm [33]. As described elsewhere, this method samples the underlying CME solution accurately [26, 33], allowing for comparisons with the ZI-closure scheme.

With moment closure it is now possible to obtain steady-states quickly, in stark contrast to stochastic simulations. Typically, the user of stochastic simulations has to simulate the dynamic behavior of an ensemble of reaction networks from an initial condition to the steady state. With ZI-closure, by setting the left side of Equation 2 to zero, a steady-state distribution is immediately available.

With fast calculations of steady states, sensitivity analysis of stochastic, nonlinear chemical reaction networks is possible. Figure 2 shows an example of the profile obtained by solving for steady-states across a range of key values. The parameter range (k_4 from 2 to 5 s⁻¹) was chosen such that the left and right sides of the graph have a single peaked distribution, whereas the blue shaded region in the middle is where the steady-state distribution has two peaks. This is explicitly shown in Figure 3 as well, where three example parameter values show how the Schlögl model flows from normally distributed to bimodal and back again.

These examples show two important features of ZI-closure. First, there are 300 steady-state values obtained using ZI-closure in Figure 2. While this number of simulations would be impractical for SSA simulations (indeed, hundreds of hours of simulation time were used to obtain the 31 SSA points for comparison), it takes a matter of minutes with ZI-closure scheme. With ZI-closure, steady-state optimization can be near immediate for some small systems. Second, the line itself is, for all practical purposes, a sensitivity analysis for parameter k_4 for the Schlögl system. Sensitivity analysis is an important tool in chemical reaction design because it can identify the parameters that most influence a specific behavior in a system. In this case it is clear that in the bimodal region the distribution mean is very sensitive to the value of k_4 . This type of analysis could have otherwise taken hours of computation time with a direct SSA simulation.

Conclusions

We anticipate that ZI-closure will aid scientists in modeling small chemical reaction networks in a way fit for analysis and design. In reference 42, we discuss numerous, example reaction networks, such as the the Schlögl model and the Michaelis-Menten model, demonstrating that ZI-closure is applicable on arbitrary small nonlinear reaction networks.

In particular, the calculation of steady states with ZI-closure affords a parameter sensitivity analysis and stability analysis for stochastic reaction networks. Modeling ranges of kinetic and thermodynamic parameters may rationalize the design of control systems for complex, nonlinear networks of reactions, away from the thermodynamic limit.

Of particular utility may be the application of ZI-closure in identifying components and interactions for synthetic biological systems. Modeling may then play an increasingly important role in biological engineering. Admittedly, the field of computational synthetic biology is still nascent. A long distance separates theory from practical implementation, but with advances in modeling techniques, from Gillespie's SSA, to hybrid simulation algorithms, to CME closure schemes, this distance is being shortened.

Acknowledgments

This work was supported by a grant from the National Institutes of Health (American Recovery and Reinvestment Act grant GM086865) and a grant from the National Science Foundation (CBET-0644792) with computational support from the Minnesota Supercomputing Institute (MSI). Support from the University of Minnesota Digital Technology Center and the University of Minnesota Biotechnology Institute is also acknowledged.

References

1. Ramkrishna D, Amundson NR. Mathematics in Chemical Engineering: A 50 Year Introspection. *AIChE J.* 2004; 50:7–23. ** This is a tour de force review of the importance of mathematical modeling in chemical engineering
2. Bilous O, Amundson NR. Chemical Reactor Stability and Sensitivity. *AIChE J.* 1955; 1:513. ** This is the seminal paper that ushered the rational analysis of chemical reactors.
3. Bilous O, Amundson NR. Chemical Reactor Stability and Sensitivity II. Effect of Parameters on Sensitivity of Empty Tubular Reactors. *AIChE J.* 1956; 2:117.
4. Aris R, Amundson NR. An Analysis of Chemical Reactor Stability and Control: I. The Possibility of Local Control, with Perfect or Imperfect Control Mechanisms, ii. The Evolution of Proportional Control, iii. The Principles of Programming Reactor Calculations. Some Extensions, *Chem. Eng. Sci.* 1958; 5:121.
5. Aris, R. Introduction to the Analysis of Chemical Reactors. Englewood Cliffs, NJ: Prentice Hall; 1965.
6. McQuarrie DA. Stochastic approach to chemical kinetics. *J Appl Probab.* 1967; 4:413–478. ** This manuscript defines chemical master equations with clarity.
7. Moyal JE. Stochastic Processes and Statistical Physics. *Journal of the Royal Statistical Society Series B (Methodological).* 1949; 11:150–210.
8. Oppenheim I, Shuler KE. Master equations and markov processes. *Phys Rev.* 1965; 138:1007–1011.
9. Oppenheim I, Shuler KE, Weiss GH. Stochastic theory of multistate relaxation processes. *Adv Mol Relax Processes.* 1967; 1:13–68.
10. Oppenheim, I.; Shuler, KE.; Weiss, GH. Stochastic processes in chemical physics: The master equation. Cambridge, MA: The MIT Press; 1977.
11. Gillespie CS. Moment closure approximations for mass-action models. *IET systems biology.* 2009; 3:52–58. [PubMed: 19154084]
12. Sotiropoulos V, Kaznessis YN. Analytical Derivation of Moment Equations in Stochastic Chemical Kinetics. *Chem Eng Sci.* 2011; 66:268–277. [PubMed: 21949443]
13. Smadbeck P, Kaznessis YN. Efficient Moment Matrix Generation for Arbitrary Chemical Networks. *Chem Eng Sci.* 2012; 84:612–618. [PubMed: 23175571]
14. Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comp Phys.* 1976; 22:403–434. ** Seminal contribution setting the foundation for stochastic kinetic simulations.

15. Gillespie DT. Exact Stochastic Simulation of Coupled Chemical Reactions. *J Phys Chem.* 1977; 81:2340–2361.
16. Gibson MA, Bruck J. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *J Phys Chem.* 2000; 104:1876–1889. * An important contribution that renders stochastic simulations significantly more efficient than the original algorithm by Gillespie.
17. Haseltine EL, Rawlings JB. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J Chem Phys.* 2002; 117:6959–6969. * A smart way to move from discrete-stochastic to continuous-stochastic algorithms.
18. Cao Y, Li H, Petzold L. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J Chem Phys.* 2004; 121:4059–4067. [PubMed: 15332951]
19. Chatterjee A, Mayawala K, Edwards JS, Vlachos DG. Time accelerated monte carlo simulations of biological networks using the binomial {tau}-leap method. *Bioinformatics.* 2005; 21:2136–2137. [PubMed: 15699024]
20. Tian T, Burrage K. Binomial leap methods for simulating stochastic chemical kinetics. *J Chem Phys.* 2004; 121:10356–10364. [PubMed: 15549913]
21. E W, Liu D, Vanden-Eijnden E. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *J Chem Phys.* 2005; 123:194107. [PubMed: 16321076] * One of the first successful attempts to simulate multiscale reaction networks
22. Munsky B, Khammash M. The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys.* 2006; 124:044104. [PubMed: 16460146] ** A very important contribution in efforts to reduce stochastic reaction models.
23. Cao Y, Gillespie DT, Petzold LR. Avoiding negative populations in explicit poisson tau-leaping. *J Chem Phys.* 2005; 123:054104. [PubMed: 16108628]
24. Chatterjee A, Vlachos DG. Temporal acceleration of spatially distributed kinetic monte carlo simulations. *J Comput Phys.* 2006; 211:596–615.
25. Harris LA, Clancy PA. A partitioned leaping" approach for multiscale modeling of chemical reaction dynamics. *J Chem Phys.* 2006; 125:144107. [PubMed: 17042579]
26. Salis H, Kaznessis YK. Accurate Hybrid Stochastic Simulation of a System of Coupled Chemical or Biochemical Reactions. *J. Chem. Phys.* 2005; 122:1–13. * A computationally efficient hybrid method for multiscale simulations.
27. Salis H, Kaznessis YN. Numerical simulation of stochastic gene circuits. *Comp Chem Eng.* 2005; 29:577–588.
28. Salis H, Kaznessis YN. An equation-free probabilistic steady state approximation: Dynamic application to the stochastic simulation of biochemical reaction networks. *J Chem Phys.* 2005; 123:214106. [PubMed: 16356038]
29. Sotiropoulos V, Daoutidis P, Kaznessis YN. Model Reduction of Multiscale Chemical Langevin Equations: A Numerical Case Study. *IEEE Trans Comp Bio Bioinf.* 2009; 6:470–482.
30. Contou-Carrere MN, Sotiropoulos V, Kaznessis YN, Daoutidis P. Model Reduction of Multi-Scale Chemical Langevin Equations. *Syst. Contr. Lett.* 2011; 60:75–86.
31. Smadbeck P, Kaznessis YN. Stochastic model reduction using a modified Hill-type kinetic law. *J. Chem. Phys.* 2012; 137:23. 234109.
32. Sotiropoulos V, Kaznessis YN. An adaptive time step scheme for a system of SDEs with multiple multiplicative noise. Chemical Langevin equation, a proof of concept. *J Chem Phys.* 2008; 128:014103. [PubMed: 18190181]
33. Salis H, Sotiropoulos V, Kaznessis YN. Multiscale Hy3S: Hybrid Stochastic Simulations for Supercomputers. *BMC Bioinformatics.* 2006; 7:93. 2006. [PubMed: 16504125]
34. Hill A, Tomshine J, Wedding E, Sotiropoulos V, Kaznessis YK. SynBioSS: the Synthetic Biology Modeling Suite. *Bioinformatics.* 2008; 24:2551–2553. [PubMed: 18757873]
35. Weeding E, Houle J, Kaznessis YN. SynBioSS designer: a web-based tool for the automated generation of kinetic models for synthetic biological constructs. *Brief Bioinform.* 2010; 11(4): 394–402. [PubMed: 20639523]
36. Tuttle L, Salis H, Tomshine J, Kaznessis YN. Model-Driven Design Principles of Gene Networks: the Oscillator. *Biophys J.* 2005; 89:3873–3883. [PubMed: 16183880]

37. Tomshine J, Kaznessis YN. Optimization of a stochastically simulated gene network model via simulated annealing. *Biophys J*. 2006; 91:3196–3205. [PubMed: 16920827]
38. Kaznessis Y. Multi-Scale Models for Gene Network Engineering. *Chem Eng Sci*. 2006; 61:940–953.
39. Kaznessis Y. Models for Synthetic Biology. *BMC Systems Biology*. 2007; 1:47. [PubMed: 17986347]
40. Ramalingam K, Maynard J, Kaznessis YN. Forward engineering of synthetic biological AND gates. *Biochemical Engineering Journal*. 2009; 47:38–47. * Models explain and guide designs of synthetic biological systems
41. Volzing K, Biliouris K, Kaznessis YN. proTeOn and proTeOff, New Protein Devices That Inducibly Activate Bacterial Gene Expression. *ACS Chem Biol*. 2011; 6(10):1107–1116. [PubMed: 21819083]
42. Smadbeck P, Kaznessis YN. A closure scheme for chemical master equations. *Proc Natl Acad Sci U S A*. 2013; 110(35):14261–14265. [PubMed: 23940327] ** Numerical solution to chemical master equations for nonlinear stochastic reaction networks.
43. Schlögl F. Chemical reaction models for non-equilibrium phase transition. *Z. Physik*. 1972; 253:147–161.

Highlights

- Randomness is a defining feature of small reaction networks
- The most complete model of random reaction networks is the master probability equation
- ZI-closure is an accurate numerical solution to the master equation through moment closure
- ZI-closure offers an alternative to Monte Carlo based stochastic simulations
- Stability analysis may lead to control and optimization theories for stochastic systems

Figure 1. Comparison between ZI-Closure and SSA

The Schlögl model parameters are $k_1 \cdot A = 0.15$ ($\text{molec.}^{-1}\text{s}^{-1}$), $k_2 = 0.0015$ ($\text{molec.}^{-2}\text{s}^{-1}$), $k_3 \cdot B = 20$ $\text{molec.}^{-1}\text{s}^{-1}$, and $k_4 = 3.5$ s^{-1} . The stochastic simulation algorithm results (black line) are an excellent approximation of the chemical master equation solution, assuming sufficient trajectories are simulated (here the results are from 10^6 trajectories). As more moments are added to the maximum entropy optimization program (from 2 to 12 moments), the ZI-closure optimization results become more representative of the exact solution distribution.

Figure 2. Schlögl Model Sensitivity Analysis

The average number of molecules of X is shown at steady states obtained with varying values of kinetic constant k_4 . The other parameter values are $k_1 \cdot A = 0.15 \text{ molec.}^{-1}\text{s}^{-1}$, $k_2 = 0.0015 \text{ molec.}^{-2}\text{s}^{-1}$, and $k_3 \cdot B = 20 \text{ molec.}^{-1}\text{s}^{-1}$. The range for k_4 was chosen such that at low values (left side) the distribution has a single peak, for middle range values (shaded-blue region) the distribution has two peaks, and at high values (right side) it returns to a single peaked distribution. In this figure the red line represents the ZI-Closure solution with 12-order closure, with each square representing a corresponding SSA solution using 10^6 trajectories.

Figure 3. Schlögl Model Probability Distributions

Distributions obtained from both SSA and ZI-Closure are shown for three k_4 values: a) $k_4 = 2 \text{ s}^{-1}$, b) $k_4 = 3.5 \text{ s}^{-1}$, and c) $k_4 = 5 \text{ s}^{-1}$. The parameters are $k_1 \cdot A = 0.15 \text{ molec.}^{-1} \text{ s}^{-1}$, $k_2 = 0.0015 \text{ molec.}^{-2} \text{ s}^{-1}$, and $k_3 \cdot B = 20 \text{ molec.}^{-1} \text{ s}^{-1}$. The distribution shifts from a single peak to two peaks and then back to one. The SSA results are represented as lines, while the ZI-Closure results (12th order closure) are circles, demonstrating the versatility of the method.

Table 1

Schlögl model network description. Note that the reservoirs A and B are assumed constant and defined as part of the first and third kinetic constants.

Model	Schlögl
Reactions	$2X + A \xrightarrow{k_1} 3X$ $3X \xrightarrow{k_2} 2X + A$ $B \xrightarrow{k_3} X$ $X \xrightarrow{k_4} B$
Kinetic parameters	$k_1 A = 0.15 \text{ (1/molecules-s)}$ $k_2 = 0.0015 \text{ (1/molecules}^2\text{-s)}$ $k_3 B = 20 \text{ (molecules/s)}$ $k_4 = 3.5 \text{ (1/s)}$ <i>unless otherwise specified</i>