Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Engineering

Solution of chemical master equations for nonlinear stochastic reaction networks

Patrick Smadbeck and Yiannis N Kaznessis

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 (CMEs), a wide range of experimental observations of small-system interactions may be mathematically conceptualized.

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

Corresponding author: Kaznessis, Yiannis N (yiannis@umn.edu, Yiannis@cems.umn.edu)

Current Opinion in Chemical Engineering 2014, **5**:90–95

This review comes from a themed issue on **Reaction engineering and catalysis**

Edited by **Marc-Olivier Coppens** and **Theodore T Tsotsis**

http://dx.doi.org/10.1016/j.coche.2014.07.001

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 Bilous 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 toward infinity. If reactants of a reaction are near the thermodynamic limit, continuous-deterministic modeling formalisms, that is, 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••,7–10]. 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 $\underline{X}(t) = [X_1, \ldots, X_N]$, where X_i is number of molecules of the ith chemical species with $i = 1, \ldots, N$.

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:

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

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:

$$\frac{\partial \underline{\mu}}{\partial t} = A\underline{\mu} + A'\underline{\mu}' \tag{2}$$

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 Eqn 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, 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•,17•,18–20,21•,22••,23–25].

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 [29–31], 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 [36–39,40•,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, that is, the mean, variance, skewness, etc. (Eqn 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 and 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 $p(x)$:

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

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):

$$\begin{aligned} \Lambda &= H - \lambda_0 g_0 - \lambda_1 g_1 - \cdots - \lambda_M g_M \\ g_0 &= \sum_{x=0}^{\infty} p(x) - 1 \\ g_1 &= \sum_{x=0}^{\infty} x\,p(x) - \langle x \rangle \\ &\vdots \\ g_M &= \sum_{x=0}^{\infty} x^M p(x) - \langle x^M \rangle \end{aligned} \tag{4}$$

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

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

or, trivially:

$$p_H(x) = \exp(-1 - \lambda_0 - \lambda_1 x - \cdots - \lambda_M x^M) \tag{6}$$

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