INTERNATIONAL JOURNAL OF ROBUST AND NONLINEAR CONTROL
Int. J. Robust Nonlinear Control 2005; **15**:691–711
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/rnc.1018

Stochastic modelling of gene regulatory networks

Hana El Samad[1], Mustafa Khammash[1,*,†], Linda Petzold[2] and Dan Gillespie[3]

[1] *Mechanical Engineering, University of California at Santa Barbara, U.S.A.*
[2] *Computer Science, University of California at Santa Barbara, U.S.A.*
[3] *Dan T Gillespie Consulting, Castaic, California, U.S.A.*

SUMMARY

Gene regulatory networks are dynamic and stochastic in nature, and exhibit exquisite feedback and feedforward control loops that regulate their biological function at different levels. Modelling of such networks poses new challenges due, in part, to the small number of molecules involved and the stochastic nature of their interactions. In this article, we motivate the stochastic modelling of genetic networks and demonstrate the approach using several examples. We discuss the mathematics of molecular noise models including the chemical master equation, the chemical Langevin equation, and the reaction rate equation. We then discuss numerical simulation approaches using the stochastic simulation algorithm (SSA) and its variants. Finally, we present some recent advances for dealing with stochastic stiffness, which is the key challenge in efficiently simulating stochastic chemical kinetics. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: gene regulatory networks; stochasticity; molecular noise

1. INTRODUCTION: THE ROLE OF MATHEMATICS AND SYSTEMS THEORY IN MODELLING BIOLOGICAL DYNAMICS

The large amounts of data being generated by high-throughput technologies have motivated the emergence of systems biology, a discipline which emphasizes a system's characterization of biological networks. Such a system's view of biological organization is aimed to draw on mathematical methods developed in the context of dynamical systems and computational theories in order to create powerful simulation and analysis tools to decipher existing data and

*Correspondence to: M. Khammash, Mechanical Engineering, University of California at Santa Barbara, U.S.A.
†E-mail: khammash@engineering.ucsb.edu

Contract/grant sponsor: U.S. Air Force Office of Scientific Research
Contract/grant sponsor: California Institute of Technology; contract/grant number: F30602-01-2-0558
Contract/grant sponsor: U.S. Department of Energy; contract/grant number: DE-FG02-04ER25621
Contract/grant sponsor: National Science Foundation; contract/grant numbers: CCF-0326576, ACI00-86061
Contract/grant sponsor: Institute for Collaborative Biotechnologies; contract/grant number: DAAD19-03-D-0004
Contract/grant sponsor: U.S. Army Research Office
Contract/grant sponsor: Molecular Sciences Institute; contract/grant number: 244725
Contract/grant sponsor: Department of Energy's Genomes to Life Program

Received 31 October 2004
Revised 1 April 2005
Accepted 1 May 2005

devise new experiments. One goal is to generate an integrated knowledge of biological complexity that unravels functional properties and sources of health and disease in cells, organs, and organisms. This is commonly known as the 'reverse engineering' problem. Another goal is to successfully interface naturally occurring genetic circuits with *de novo* designed and implemented systems that interfere with the sources of disease or malfunction and revert their effects. This is known as the 'forward engineering' problem. Although reverse and forward engineering of even the simplest biological systems has proven to be a daunting task, mathematical approaches coupled to rounds of iteration with experimental approaches greatly facilitate the road to biological discovery in a manner that is otherwise difficult. This, however, necessitates a serious effort devoted to the characterization of salient biological features, followed by the successful extension of engineering/mathematical tools, and the creation of new tools and theories to accommodate and/or exploit them.

Much of the mathematical modelling of genetic networks represents gene expression and regulation as deterministic processes [1]. There is now, however, considerable experimental evidence indicating that significant stochastic fluctuations are present in these processes, both in prokaryotic and eukaryotic cells [2–6]. Furthermore, studies of engineered genetic circuits designed to act as toggle switches or oscillators have revealed large stochastic effects [7–9]. Stochasticity is therefore an inherent feature of biological dynamics, and as such, should be the subject of in depth investigation and analysis. A study of stochastic properties in genetic systems is a challenging task. It involves the formulation of a correct representation of molecular noise, followed by the formulation of mathematically sound approximations for these representations. It also involves devising efficient computational algorithms capable of tackling the complexity of the dynamics involved.

2. ELEMENTS OF GENE REGULATION

Computation in the cellular environment is carried out by proteins, whose abundance and activity is often tightly regulated. Proteins are produced through an elaborate process of gene expression, which is often referred to as the 'Central Dogma' of molecular biology.

2.1. Gene regulatory networks and the central dogma of molecular biology

The synthesis of cellular protein is a multi-step process that involves the use of various cellular machines. One very important machine in bacteria is the so called RNA polymerase (RNAP). RNAP is an enzyme that can be recruited to transcribe any given gene. However, RNAP bound to regulatory sigma factors recognizes specific sequences in the DNA, referred to as the *promoter*. Whereas the role of RNAP is to transcribe genes, the main role of σ factors is to recognize the promoter sequence and signal to RNAP in order to initiate the transcription of the appropriate genes. The transcription process itself consists of synthesizing a messenger RNA (mRNA) molecule that carries the information encoded by the gene. Here, RNAP acts as a 'reading head' transcribing DNA sequences into mRNA. Once a few nucleotides on the DNA have been transcribed, the σ-factor molecule dissociates from RNAP, while RNAP continues transcribing the genes until it recognizes a particular sequence called a *terminator sequence*. At this point, the mRNA is complete and RNAP disengages from the DNA. During the transcription process, ribosomes bind to the nascent mRNA and initiate translation of the

message. The process of translation consists of sequentially assembling amino acids in an order that corresponds to the mRNA sequence, with each set of three nucleotides corresponding to a single unique amino acid. This combined process of gene transcription and mRNA translation constitutes gene expression, and is often referred to as the *central dogma* of molecular biology.

Gene regulatory networks can be broadly defined as groups of genes that are activated or deactivated by particular signals and stimuli, and as such produce or halt the production of certain proteins. Through various combinatorial logic at the gene or the end-product protein level, these networks orchestrate their operation to regulate certain biological functions such as metabolism, development, or the cellular clocks. Regulation schemes in gene regulatory networks often involve positive and negative feedback loops. A simple scheme consists, for example, of a protein that binds to the promoter of its own gene and shields it from the RNAP-σ complex, thereby auto-regulating its own production. When interfaced and connected together according to a certain logic, a network of such building blocks (and possibly others possessing different architectures and components) generates intricate systems that possess a wide range of dynamical behaviours and functionalities.

3. MODELLING GENETIC NETWORKS

The mathematical approaches used to model gene regulatory network differ in the level of resolution they achieve and their underlying assumptions. A broad classification of these methods separates the resulting models into deterministic and stochastic, each class embodying various subclasses with their different mathematical formalisms.

3.1. Deterministic rate equations modelling

Cellular processes, such as transcription and translation, are often perceived to be systems of distinct chemical reactions that can be described using the laws of mass-action, yielding a set of differential equations (linear or nonlinear) that give the succession of states (usually concentration of species) adopted by the network over time. The equations are usually of the form

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = f_i(x), \quad 1 \leqslant i \leqslant n \tag{1}$$

where $x = [x_1, \ldots, x_n]^{\mathrm{T}}$ is a vector of non-negative real numbers describing concentrations and $f_i : \mathbb{R}^n \to \mathbb{R}^n$ is a function of the concentrations. Ordinary differential equations are arguably the most widespread formalism for modelling gene regulatory networks and their use goes back to the 'operon' model of Jacob and Monod [10] and the early work of Goodwin [11]. The main rationale of deterministic chemical kinetics is that at constant temperature, elementary chemical reaction rates vary with reactant concentration in a simple manner. These rates are proportional to the frequency at which the reacting molecules collide, which is again dependent on the molecular concentrations in these reactions.

3.2. Stochastic modelling

Although the time evolution of a well-stirred chemically reacting system is traditionally described by a set of coupled, ordinary differential equations that characterize the evolution of

the molecular populations as a continuous, deterministic process. Chemically reacting systems in general, and genetic networks in particular, actually possesses neither of those attributes: molecular populations are whole numbers, and when they change they always do so by discrete, integer amounts. Furthermore, a knowledge of the system's current molecular populations is not by itself sufficient to predict with certainty the future molecular populations. Just as rolled dice are essentially random or 'stochastic' when we do not precisely track their positions and velocities and all the forces acting on them, so is the time evolution of a well-stirred chemically reacting system for all practical purposes stochastic. If discreteness and stochasticity are not noticeable, for example in chemical systems of 'test-tube' size or larger, then the traditional continuous deterministic description seems to be adequate. But if the molecular populations of some reactant species are very small, or if the dynamic structure of the system makes it susceptible to noise amplification, as is often the case in cellular systems, discreteness and stochasticity can play an important role. Whenever that happens, the ordinary differential equations approach does not accurately describe the true behaviour of the system. Alternatively, one should resort to an overtly discrete and stochastic description evolving in real (continuous) time that accurately reflects how chemical reactions physically occur at the molecular level. We first give a number of motivating examples that illustrate the importance of such stochastic methods, then present the details of stochastic chemical kinetics.

3.3. Deterministic versus stochastic: some examples

Here, we present a number of examples that illustrate situations where accurate accounting of noise dynamics is crucial. The first example depicts a situation where the deterministic description of a genetic network does not correctly represent the evolution of the mean of the inherently stochastic system. The second example illustrates the effect of noise on a system exhibiting bistability, making the convergence to any one of the equilibria a probabilistic event, and even causing random switching between these different equilibria. The third example illustrates the effect of noise in inducing oscillations in an otherwise stable system. Finally, the fourth example, which will also motivate the theoretical treatment in the rest of the paper, illustrates the situation where fluctuations in the concentration of key cellular regulators are of special interest even in the absence of a noise-induced change in the dynamical behaviour.

3.3.1. A monostable system. Molecular noise acting upon dynamical structures can generate disorderly behaviour in homeostatic systems. Consider a simple example where protein molecules X and Y are synthesized from the reservoirs A and B at an equal rate k. X and Y are assumed to associate irreversibly with association rate constant k_a in the formation of a heterodimer C. Molecules of X and Y can also decay with first-order rate constant α_1 and α_2, respectively, as described by the scheme

$$A \xrightarrow{k} X \xrightarrow{\alpha_1} \phi, \quad B \xrightarrow{k} Y \xrightarrow{\alpha_2} \phi, \quad X + Y \xrightarrow{k_a} C \tag{2}$$

In the deterministic setting, the reactions in (2) can be described by the rate equations

$$\begin{aligned} \frac{\mathrm{d}\phi_1}{\mathrm{d}t} &= k - \alpha_1\phi_1 - k_a\phi_1\phi_2 \\ \frac{\mathrm{d}\phi_2}{\mathrm{d}t} &= k - \alpha_2\phi_2 - k_a\phi_1\phi_2 \end{aligned} \tag{3}$$

where ϕ_1 and ϕ_2 are the concentrations of X and Y, respectively. The system in Equation (3) can only have one stable equilibrium point in the positive quadrant whose region of attraction of this equilibrium is the entire positive quadrant. For two sets of values the system's parameters given by

$$\begin{aligned} &k = 10, \quad \alpha_1 = 10^{-6}, \quad \alpha_2 = 10^{-5}, \quad k_a = 10^{-5} \\ &k = 10^3, \quad \alpha_1 = 10^{-4}, \quad \alpha_2 = 10^{-3}, \quad k_a = 10^{-3} \end{aligned} \tag{4}$$

the steady-state values for X and Y are equal and given by $\phi_1^{ss} \simeq \sqrt{k/k_a} = 1000$ and $\phi_2^{ss} \simeq 100$. Sample stochastic trajectories of the system are given in Figure 1. A plot for Y, for example, suggests that the mean follows closely the deterministic trajectory for the first set of parameters. Furthermore, fluctuations around the mean are relatively small. For the second set of parameters, however, there is a noticeable discrepancy between the behaviour of the mean and that of the deterministic trajectory. Stochastic excursions reach up to four-fold the deterministic trajectory, indicating a severe effect of the noise on the system. Such an effect indicates that a deterministic approach to the analysis of such a system can be misleading and calls for a thorough stochastic treatment.

3.3.2. A genetic switch. Multistable biological systems are abundant in nature [12,8]. Stochastic effects in these systems can be substantial as noise can influence the convergence to equilibria or even cause switching from one equilibrium to another. One of the best studied examples of multistability in genetic systems is the bacteriophage λ system [6]. A simplified model for the bacteriophage λ was proposed by Hasty and coworkers [13]. In their model, the gene *cI*

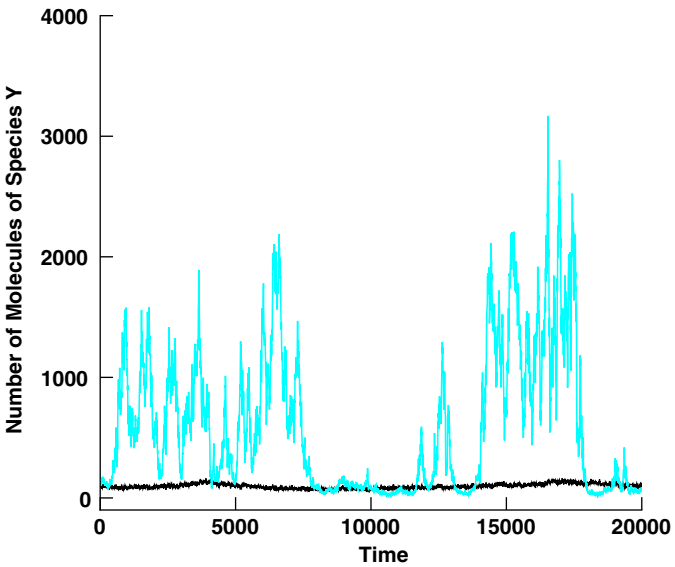

Figure 1. Stochastic simulation of the system in (2). The first set of parameters in (4) yields the plot in black while the second set of parameters yields the plot in gray.

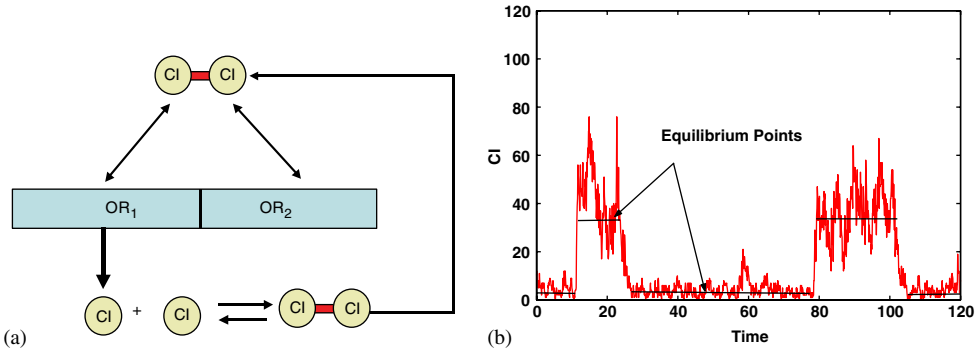

Figure 2. (a) A simple model of the lambda bacteriophage; and (b) stochastic time trajectory of CI. Noise causes switching between the two equilibrium points of the system.

expresses the λ repressor CI which dimerises and binds to DNA as a transcription factor at either of two binding sites, OR1 or OR2. Binding of this transcription factor to OR1 enhances transcription of CI (positive feedback), while binding to OR2 represses transcription of CI (negative feedback) (see Figure 2(a)). The molecular reactions in this system proceed as follows:

$$2\mathrm{CI} \overset{K_1}{\rightleftharpoons} \mathrm{CI}_2, \quad \mathrm{CI}_2 + \mathrm{D} \overset{K_2}{\rightleftharpoons} \mathrm{DCI}_2, \quad \mathrm{CI}_2 + \mathrm{D} \overset{K_s}{\rightleftharpoons} \mathrm{DCI}_2^*$$

$$\mathrm{DCI}_2 + \mathrm{CI}_2 \overset{K_4}{\rightleftharpoons} \mathrm{DCI}_2\mathrm{CI}_2, \quad \mathrm{DCI}_2 + P \overset{k_t}{\rightarrow} \mathrm{DCI}_2 + P + n\mathrm{CI}, \quad \mathrm{CI} \overset{k_d}{\rightarrow} \phi$$

where the DCI_2 and DCI_2^* complexes denote the binding to OR_1 and OR_2, respectively, and $\mathrm{DCI}_2\mathrm{CI}_2$ denotes binding to both sites. K_i are forward equilibrium constants, k_t is protein synthesis rate, and k_d is degradation rate. P is the concentration of the RNA polymerase assumed here to be constant, and n is the number of proteins per mRNA transcript, taken here to be 2. Ordinary differential equations that describe these chemical reactions implement a bistable system. In the deterministic setting, and for any given set of initial conditions, the system's trajectories converge to one or the other of the equilibria and stay there for all times. However, as we incorporate the effect of molecular noise in this description, we notice that switching between the two stable equilibria is possible if the noise amplitude is sufficient to drive the trajectories occasionally out of the basin of attraction of one equilibrium into the basin of attraction of the other equilibrium. This effect is shown in Figure 2(b).

3.3.3. A genetic oscillator. To adapt to natural periodicity, such as the alternation of day and night, most living organisms have developed the capability of generating oscillating expressions of proteins in their cells with a period close to 24 h. The molecular mechanisms that generate these oscillations, known as the circadian rhythm, have been the subject of extensive experimental and mathematical investigation in various organisms. The Vilar–Kueh–Barkai–Leibler (VKBL in short) description of the circadian oscillator incorporates an abstraction of a minimal set of essential, experimentally determined mechanisms for the circadian system [14]. More specifically, the VKBL model involves two genes, an activator A and a repressor R, which

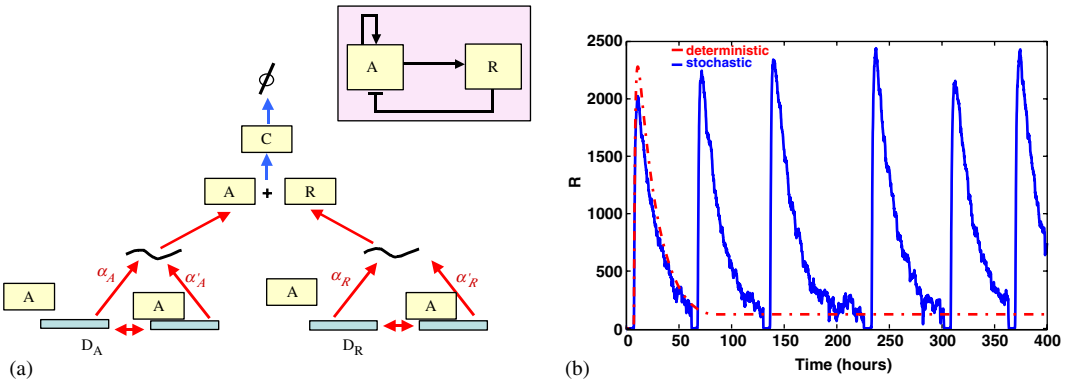

Figure 3. (a) The molecular components of the VKBL model of the circadian oscillator; and (b) noise induced oscillations.

are transcribed into mRNA and subsequently translated into proteins. The activator A binds to the A and R promoters and increases their expression rate. Therefore, A implements a positive loop acting on its own transcription. At the same time, R sequesters A to form a complex C, therefore inhibiting it from binding to the gene promoter and acting as a negative feedback loop. These interactions are depicted in Figure 3. For the parameter values given in Reference [14], a differential equations model for the dynamics of Figure 3 exhibits autonomous oscillations with an approximate period of 24 h. These oscillations, however, promptly disappear from the deterministic model as the degradation rate of the repressor δ_R is decreased to the quarter of its value. Actually, a bifurcation diagram shows that the system undergoes a supercritical Hopf bifurcation in the neighbourhood of this value. The unique deterministic equilibrium of the system becomes stable (see Figure 3(b)). However, as the effects of molecular noise are incorporated into the description of the system, it is observed that oscillations in the stochastic system pertain (see Figure 3(b)). In fact, the regularity of these noise-induced oscillations, in an otherwise stable deterministic system, can be manipulated by tuning the level of noise in the network. This can be done, for example, by changing the number of molecules or speed of molecular reactions [15]. This phenomenon is a manifestation of *coherence resonance*, and illustrates the crucial interplay between noise and dynamics.

3.3.4. Accounting for molecular fluctuations. Major regulators such as σ-factors in prokaryotic cells are present in small numbers and hence, removal or addition of a single regulator molecule can have substantial effects. An example of such a situation is present in the heat shock response of the bacterium *Escherichia coli*, which refers to the mechanism by which organisms react to a sudden increase in the ambient temperature. The consequence at the cellular level is the unfolding of cell proteins, which threatens the life of the cell. Under these circumstances, the enzyme RNA polymerase (RNAP) bound to the regulatory sigma factor, σ^{32}, recognizes the HS gene promoters and transcribes specific HS genes. The HS genes encode predominantly molecular chaperones that are involved in refolding denatured proteins and proteases that degrade unfolded proteins. In addition to their binding to unfolded proteins, chaperones can also bind σ^{32}. Chaperone-bound σ^{32} is incapable of binding to RNAP. This has the effect of a negative feedback loop that modulates the activity of σ^{32}. Other feedback and feedforward loops

modulate the stability and synthesis of σ^{32}. A detailed mechanistic model for the heat shock response was developed in Reference [16]. Here, we give an abstraction that illustrates the challenges encountered in such systems, while keeping the presentation simple. The example consists of the following molecular reactions:

$$S_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_2 \xrightarrow{c_3} S_3$$

For parameter values $c_1 = 10$, $c_2 = 4 \times 10^4$, and $c_3 = 2$, the average population of species S_2 is 0.5. Most of the time, the S_2 population is 0, sometimes it is 1, occasionally it is 2 or 3, and only rarely is anything more. An S_2 molecule here has a very short lifetime, usually turning into an S_1 molecule. In our analogy to the heat shock model, σ^{32} molecules would be sequestered by the chaperones as soon as they become available. An S_2 molecule occasionally turns into an S_3 molecule, corresponding to an important event of gene expression. It is therefore of interest to track precisely when this event happens, motivating a look at the statistics of the S_2 molecule, rather than an averaged behaviour of this quantity as depicted in a deterministic description of the system. We will re-examine this example later in the context of the computational advances and challenges in stochastic chemical kinetics.

4. MATHEMATICS OF MOLECULAR NOISE MODELS

4.1. Foundations of stochastic chemical kinetics

We consider a well-stirred system of molecules of N chemical species $\{S_1, \ldots, S_N\}$ interacting through M chemical reaction channels $\{R_1, \ldots, R_M\}$. The system is assumed to be confined to a constant volume Ω, and to be in thermal (but not chemical) equilibrium at some constant temperature. With $X_i(t)$ denoting the number of molecules of species S_i in the system at time t, we wish to study the evolution of the state vector $X(t) = (X_1(t), \ldots, X_N(t))$, given that the system is initially in some state $X(t_0) = x_0$. Each reaction channel R_j is assumed to be 'elemental' in the sense that it describes a distinct physical event which happens essentially instantaneously. Reaction channel R_j is characterized mathematically by two quantities. The first is its state-change vector $v_j = (v_{1j}, \ldots, v_{Nj})$, where v_{ij} is defined to be the change in the S_i molecular population caused by one R_j reaction; thus, if the system is in state x and an R_j reaction occurs, the system immediately jumps to state $x + v_j$. The array $\{v_{ij}\}$ is commonly known as the stoichiometric matrix. The other characterizing quantity for reaction channel R_j is its propensity function a_j. It is defined so that $a_j(x)\,\mathrm{d}t$ gives the probability, given $X(t) = x$, that one R_j reaction will occur somewhere inside Ω in the next infinitesimal time interval $[t, t + \mathrm{d}t)$. If R_j is the monomolecular reaction $S_i \rightarrow$ products, the underlying physics is quantum mechanical, and implies the existence of some constant c_j such that $a_j(x) = c_j x_i$. If R_j is the bimolecular reaction $S_i + S_{i'} \rightarrow$ products, the underlying physics implies a different constant c_j, and a propensity function $a_j(x)$ of the form $c_j x_i x_{i'}$ if $i \neq i'$, or $c_j \frac{1}{2} x_i(x_i - 1)$ if $i = i'$ [17,18]. The stochasticity of a bimolecular reaction stems from the fact that we do not know the precise positions and velocities of all the molecules in the system; thus we can predict only the probability that an S_i molecule and an $S_{i'}$ molecule will collide in the next $\mathrm{d}t$, and only the probability that such a collision will result in an R_j reaction. It turns out that c_j for a monomolecular reaction is numerically equal to the reaction rate constant k_j of conventional deterministic chemical

kinetics, while c_j for a bimolecular reaction is equal to k_j/Ω if the reactants are different species, or $2k_j/\Omega$ if they are the same [17–19].

4.2. The chemical master equation

Due to the probabilistic nature of the dynamics described above, we need to compute the probability $P(x, t \mid x_0, t_0)$ that $X(t)$ will equal x, given that $X(t_0) = x_0$. We can deduce a time-evolution equation for this function by using the laws of probability to write

$$P(x, t + \mathrm{d}t \mid x_0, t_0) = P(x, t \mid x_0, t_0) \times \left[1 - \sum_{j=1}^{M} a_j(x)\,\mathrm{d}t\right] + \sum_{j=1}^{M} P(x - v_j, t \mid x_0, t_0) \times a_j(x - v_j)\,\mathrm{d}t$$

The first term on the right is the probability that the system is already in state x at time t and no reaction of any kind occurs in $[t, t + \mathrm{d}t)$, and the generic second term is the probability that the system is one R_j reaction removed from state x at time t and one R_j reaction occurs in $[t, t + \mathrm{d}t)$. That these $M + 1$ routes from time t to state x at time $t + \mathrm{d}t$ are mutually exclusive and collectively exhaustive is ensured by taking $\mathrm{d}t$ so small that no more than one reaction of any kind can occur in $[t, t + \mathrm{d}t)$. Subtracting $P(x, t \mid x_0, t_0)$ from both sides, dividing through by $\mathrm{d}t$, and taking the limit $\mathrm{d}t \to 0$, we obtain [17,20]

$$\frac{\partial P(x, t \mid x_0, t_0)}{\partial t} = \sum_{j=1}^{M} [a_j(x - v_j)P(x - v_j, t \mid x_0, t_0) - a_j(x)P(x, t \mid x_0, t_0)] \tag{5}$$

This is the *chemical master equation* (CME). In principle, it completely determines the function $P(x, t \mid x_0, t_0)$. But the CME is really a set of nearly as many coupled ordinary differential equations as there are combinations of molecules that can exist in the system. So it is not surprising that the CME can be solved analytically for only a very few very simple systems, and numerical solutions are usually prohibitively difficult. One might hope to learn something from the CME about the behaviour of averages like $\langle f(X(t))\rangle \equiv \sum_x f(x)P(x, t \mid x_0, t_0)$. For example, it can be proved from Equation (5) that

$$\frac{\mathrm{d}\langle X_i(t)\rangle}{\mathrm{d}t} = \sum_{j=1}^{M} v_{ij}\,\langle a_j(X(t))\rangle \quad (i = 1, \ldots, N)$$

If all the reactions were monomolecular, the propensity functions would all be linear in the state variables, and we would have $\langle a_j(X(t))\rangle = a_j(\langle X(t)\rangle)$. The above equation would then become a closed ordinary differential equation for the first moments, $\langle X_i(t)\rangle$. But if any reaction is bimolecular, the right-hand side will contain at least one quadratic moment of the form $\langle X_i(t)X_{i'}(t)\rangle$, and the equation then becomes merely the first of an infinite, open-ended set of equations for all the moments. In the hypothetical case that there are *no fluctuations*, we would have $\langle f(X(t))\rangle = f(X(t))$ for all functions f. The above equation for $\langle X_i(t)\rangle$ would then reduce to

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = \sum_{j=1}^{M} v_{ij}\,a_j(X(t)) \quad (i = 1, \ldots, N) \tag{6}$$

This is the *reaction rate equation* (RRE) of traditional deterministic chemical kinetics—a set of N coupled first-order ordinary differential equations for the $X_i(t)$, which are now continuous (real) variables. The RRE is more commonly written in terms of the concentration variables $X_i(t)/\Omega$ as in Equation (1), but that scalar transformation is inconsequential for our purposes here. We shall later see how the RRE follows more deductively from a series of physically transparent approximating assumptions to the stochastic theory.

4.3. The stochastic simulation algorithm

Since the CME (5) is rarely of much use in computing $P(x, t \mid x_0, t_0)$ of $X(t)$, we need another computational approach. One approach that has proven fruitful is to construct *numerical realizations* of $X(t)$, i.e. simulated trajectories of $X(t)$-versus-t. This is *not* the same as solving the CME numerically; however, much the same effect can be achieved by either histogramming or averaging the results of many realizations. The key to generating simulated trajectories of $X(t)$ is a new function, $p(\tau, j \mid x, t)$ [19]. It is defined so that $p(\tau, j \mid x, t)\,\mathrm{d}\tau$ is the probability, given $X(t) = x$, that the *next* reaction in the system will occur in the infinitesimal time interval $[t + \tau,\ t + \tau + \mathrm{d}\tau)$, *and* will be an R_j reaction. Formally, this function is the joint probability density function of the two random variables 'time to the next reaction' (τ) and 'index of the next reaction' (j). To derive an analytical expression for $p(\tau, j \mid x, t)$, we begin by noting that if $P_0(\tau \mid x, t)$ is the probability, given $X(t) = x$, that no reaction of any kind occurs in the time interval $[t, t + \tau)$, then the laws of probability imply the relations

$$p(\tau, j \mid x, t)\,\mathrm{d}\tau = P_0(\tau \mid x, t) \times a_j(x)\,\mathrm{d}\tau$$

$$P_0(\tau + \mathrm{d}\tau \mid x, t) = P_0(\tau \mid x, t) \times \left[1 - \sum_{j'=1}^{M} a_{j'}(x)\,\mathrm{d}\tau\right]$$

An algebraic rearrangement of this last equation and passage to the limit $\mathrm{d}\tau \to 0$ results in a differential equation whose solution is easily found to be $P_0(\tau \mid x, t) = \exp(-a_0(x)\tau)$, where $a_0(x) \equiv \sum_{j'=1}^{M} a_{j'}(x)$. When we insert this result into the previous equation, we get

$$p(\tau, j \mid x, t) = a_j(x) \exp(-a_0(x)\tau) \tag{7}$$

Equation (7) is the mathematical basis for the simulation approach. It implies that the joint density function of τ and j can be written as the product of the τ-density function, $a_0(x) \exp(-a_0(x)\tau)$, and the j-density function, $a_j(\mathbf{x})/a_0(\mathbf{x})$. We can easily generate random samples from these two density functions by using the inversion method of Monte Carlo theory [19]: Draw two random numbers r_1 and r_2 from the uniform distribution in the unit-interval, and select τ and j according to

$$\tau = \frac{1}{a_0(x)} \ln\left(\frac{1}{r_1}\right) \tag{8a}$$

$$\sum_{j'=1}^{j-1} a_{j'}(x) \leqslant r_2\, a_0(x) < \sum_{j'=1}^{j} a_{j'}(x) \tag{8b}$$

Thus we arrive at the following version of the *stochastic simulation algorithm* (SSA) [19,21]:

1. Initialize the time $t = t_0$ and the system's state $x = x_0$.
2. With the system in state x at time t, evaluate all the $a_j(x)$ and their sum $a_0(x)$.
3. Generate values for τ and j according to Equations (8).
4. Effect the next reaction by replacing $t \leftarrow t + \tau$ and $x \leftarrow x + v_j$.
5. Record (x, t) as desired. Return to Step 2, or else end the simulation.

The $X(t)$ trajectory that is produced by the SSA might be thought of as a 'stochastic version' of the trajectory that would be obtained by solving the RRE (6). But note that the time step τ in the SSA is exact, and is not a finite approximation to some infinitesimal $\mathrm{d}t$, as is the time step in most numerical solvers for the RRE. The SSA and the CME are logically equivalent to each other; yet even when the CME is completely intractable, the SSA is straightforward to implement. The problem with the SSA is that it is often very slow. The source of this slowness can be traced to the factor $1/a_0(x)$ in Equation (8a); this factor will be very small if the population of any reactant species is sufficiently large, and that is often the case in practice. There are several variations to the above method for implementing the SSA, some of which are more efficient than others [22,23]. But *any* procedure that simulates *every* reaction event one at a time will inevitably be too slow for most practical applications. This prompts us to look for ways of giving up some of the exactness of the SSA in return for greater simulation speed.

4.4. Tau-leaping

One approximate accelerated simulation strategy is *tau-leaping* [24]. It advances the system by a *pre-selected* time τ which encompasses more than one reaction event. In its simplest form, tau-leaping requires that τ be chosen small enough that the following *Leap Condition* is satisfied: The expected state change induced by the leap must be sufficiently small that no propensity function changes its value by a significant amount. Therefore, if $X(t) = x$, and if we choose τ small enough to satisfy the Leap Condition, so that the propensity functions stay approximately constant, then reaction R_j should fire approximately $\mathscr{P}_j(a_j(x), \tau)$ times in $[t, t + \tau)$. So to the degree that the Leap Condition is satisfied, we can leap by a time τ from state x at time t by taking [24]

$$X(t + \tau) \doteq x + \sum_{j=1}^{M} v_j \, \mathscr{P}_j(a_j(x), \tau) \tag{9}$$

where the Poisson random variable $\mathscr{P}(a, \tau)$ is by definition the number of events that will occur in time τ given that $a\,\mathrm{d}t$ is the probability that an event will occur in any infinitesimal time $\mathrm{d}t$, where a can be any positive constant. Doing this evidently requires generating M Poisson random numbers for each leap. It will result in a faster simulation than the SSA to the degree that the total number of reactions leapt over, $\sum_{j=1}^{M} \mathscr{P}_j(a_j(x), \tau)$, is large compared to M. To apply this simulation technique, we must be able to estimate in advance the largest value of τ that is compatible with the Leap Condition. At least one reasonably efficient way of doing that has been developed [25]. In the limit that $\tau \to 0$, tau-leaping becomes mathematically equivalent to the SSA. But tau-leaping also becomes very inefficient in that limit because all the random numbers in (9) will approach zero, giving a very small time step with usually no reactions firing. As a practical matter, tau-leaping should not be used if the largest value of τ that satisfies the

Leap Condition is less than a few multiples of $1/a_0(x)$, the expected time to the next reaction in the SSA, since it would then be more efficient to use the SSA. Tau-leaping has been shown to significantly speed up the simulation of *some* systems [24,25]. But it is not as foolproof as the SSA. If one takes leaps that are too large, problems can arise; e.g. some species populations might be driven negative [26–28]. If the system is 'stiff', meaning that it has widely varying time scales with the fastest mode being stable, the Leap Condition will generally limit the size of τ to the time scale of the fastest mode, with the result that large leaps cannot be taken. Stiffness is very common in cellular chemical systems, and is discussed in more detail in Section 5, along with several possible remedies.

It is tempting to try to formulate a 'higher-order' tau-leaping formula by extending higher-order ODE methods in a straightforward manner for discrete stochastic simulation. However, to do this correctly is very challenging. Most such extensions give state updates whose stochastic parts are not even correct to first order in τ. A theory of consistency, order and convergence for tau-leaping methods is given in Reference [29], where it is shown that the tau-leaping method defined above, and the implicit tau-leaping method given in Section 5.1, are correct to first order in τ as $\tau \to 0$.

4.5. Transitioning to the macroscale: the chemical Langevin equation and the reaction rate equation

Suppose we can choose τ small enough to satisfy the Leap Condition, so that approximation (9) is good, but nevertheless large enough that

$$a_j(x)\tau \gg 1 \quad \text{for all } j = 1, \ldots, M \tag{10}$$

Since $a_j(x)\tau$ is the mean of the random variable $\mathscr{P}_j(a_j(x), \tau)$, the physical significance of condition (10) is that each reaction channel is expected to fire many more times than once in the next τ. It will not always be possible to find a τ that satisfies both the Leap Condition and condition (10), but it usually will be if the populations of all the reactant species are sufficiently large. When condition (10) does hold, we can make a useful approximation to the tau-leaping formula (9). This approximation stems from the purely mathematical fact that the Poisson random variable $\mathscr{P}(a, \tau)$ can be well approximated when $a\tau \gg 1$ by a *normal* random variable with the same mean and variance. Denoting the normal random variable with mean m and variance σ^2 by $\mathscr{N}(m, \sigma^2)$, it thus follows that when condition (10) holds,

$$\mathscr{P}_j(a_j(x), \tau) \doteq \mathscr{N}_j(a_j(x)\tau,\ a_j(x)\tau) = a_j(x)\tau + (a_j(x)\tau)^{1/2}\, \mathscr{N}_j(0, 1)$$

The last step follows from the fact that $\mathscr{N}(m, \sigma^2) = m + \sigma\mathscr{N}(0, 1)$. Inserting this approximation into Equation (9) gives the formula for leaping by τ from state x at time t [30,31]

$$X(t + \tau) \doteq x + \sum_{j=1}^{M} v_j a_j(x)\tau + \sum_{j=1}^{M} v_j \sqrt{a_j(x)}\mathscr{N}_j(0, 1)\sqrt{\tau} \tag{11}$$

This is called the *Langevin leaping formula*. It evidently expresses the state increment $X(t + \tau) - x$ as the sum of two terms: a deterministic 'drift' term proportional to τ, and a fluctuating 'diffusion' term proportional to $\sqrt{\tau}$. It is important to keep in mind that Equation (11) is an approximation which is valid only to the extent that τ is (i) *small* enough that no

propensity function changes its value significantly during τ, yet (ii) *large* enough that every reaction fires many more times than once during τ. The approximate nature of Equation (11) is underscored by the fact that $X(t)$ therein is now a continuous (real-valued) random variable instead of discrete (integer-valued) random variable; we lost discreteness when we replaced the integer-valued Poisson random variable with a real-valued normal random variable. The 'small-but-large' character of τ in Equation (11) marks that variable as a 'macroscopic infinitesimal'. If we subtract x from both sides and then divide through by τ, the result can be shown to be the following (approximate) stochastic differential equation, which we call the *chemical Langevin equation* (CLE) [30–32]:

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} \doteq \sum_{j=1}^{M} v_j a_j(X(t)) + \sum_{j=1}^{M} v_j \sqrt{a_j(X(t))}\, \Gamma_j(t) \tag{12}$$

The $\Gamma_j(t)$ here are statistically independent 'Gaussian white noise' processes satisfying $\langle \Gamma_j(t)\, \Gamma_{j'}(t') \rangle = \delta_{jj'}\, \delta(t - t')$, where the first delta function is Kronecker's and the second is Dirac's. The CLE (12) is mathematically equivalent to Equation (11), and is subject to the same conditions for validity. Langevin equations arise in many areas of physics, but the usual way of obtaining them is to start with a macroscopically inspired drift term (the first term on the right side of the CLE) and then *assume* a form for the diffusion term (the second term on the right side of the CLE) with an eye to obtaining some pre-conceived outcome. So it is noteworthy that our derivation here inferred the forms of *both* the drift and diffusion terms from the premises underlying the CME/SSA.

Molecular systems become 'macroscopic' in the *thermodynamic limit*. This limit is formally defined when system volume Ω and the species populations X_i all approach ∞ in such a way that the species concentrations X_i/Ω all remain constant. The large populations in chemical systems near the thermodynamic limit generally mean that such systems will be well described by the Langevin formulas (11) and (12). To discern the implications of those formulas in the thermodynamic limit, we evidently need to know the behaviour of the propensity functions in that limit. It turns out that all propensity functions grow *linearly* with the system size as the thermodynamic limit is approached. For a monomolecular propensity function of the form $c_j x_i$ this behaviour is obvious, since c_j will be independent of the system size. For a bimolecular propensity function of the form $c_j x_i x_{i'}$ this behaviour is a consequence of the fact that bimolecular c_j's are always inversely proportional to Ω, reflecting the fact that two reactant molecules have a harder time finding each other in larger volumes. It follows that as the thermodynamic limit is approached, the deterministic drift term in Equation (11) grows like the size of the system, while the fluctuating diffusion term grows like the *square root* of the size of the system. And likewise for the CLE (12). This establishes the well known rule-of-thumb in chemical kinetics that *relative fluctuation effects in chemical systems typically scale as the inverse square root of the size of the system*. In the *full thermodynamic limit*, the size of the second term on the right-hand side of Equation (12) will usually be negligibly small compared to the size of the first term, in which case the CLE reduces to the RRE (6). Thus we have *derived* the RRE as a series of limiting approximations to the stochastic theory that underlies the CME and the SSA. The tau-leaping and Langevin-leaping formulas evidently form a conceptual bridge between stochastic chemical kinetics (the CME and SSA) and conventional deterministic chemical kinetics (the RRE), enabling us to see how the latter emerges as a limiting approximation of the former.

5. ADVANCES IN THE SIMULATION OF MOLECULAR NOISE MODELS

5.1. SSA, tau-leaping, and connections to stiffness

Despite recent improvements in its implementation [22,23], as a procedure that simulates *every* reaction event, the SSA is necessarily inefficient for most realistic problems. There are two main reasons for this, both arising from the multiscale nature of the underlying problem: (1) *stiffness*, i.e. the presence of multiple timescales, the fastest of which are stable; and (2) the need to include in the simulation both species that are present in relatively small quantities and should be modelled by a discrete stochastic process, and species that are present in larger quantities and are more efficiently modelled by a deterministic differential equation.

We have already discussed tau-leaping, which accelerates discrete stochastic simulation for systems involving moderate to large populations of all chemical species. There are two limitations to the tau-leaping method as it has been described so far. The first is that it is not appropriate or accurate for describing the evolution of chemical species that are present in *very small* populations. In an adaptive algorithm, these can be handled by SSA. The second limitation is that the stepsize τ that one can take in the original tau-leaping method turns out to be limited not only by the rate of change of the solution variables, but by the timescales of the fastest reactions, even if those reactions are having only a slow net effect on the species concentrations. This limitation is due to the explicit nature of the tau-leaping formula, and is very closely related to a phenomenon called *stiffness* which has been studied extensively in the context of simulation of deterministic (ODE) chemical kinetics. The second, and more serious limitation can be handled by slow scale SSA, which will be discussed in Section 5.2, or by implicit tau-leaping which will be discussed next.

In deterministic systems of ODEs, stiffness generally manifests itself when there are well separated 'fast' and 'slow' time scales present, and the 'fast modes' are stable. Because of the fast stable modes, all initial conditions result in trajectories which, after a short and rapid transient, lead to the 'stable manifold' where the 'slow modes' determine the dynamics and the fast modes have decayed. In general, a given trajectory of such a system will exhibit rapid change for a short duration (corresponding to the fast time scales) called the 'transient', and then evolve slowly (corresponding to the slow time scales). During the initial transient, the problem is said to be non-stiff, whereas while the solution is evolving slowly it is said to be stiff. One would expect that a reasonable numerical scheme should be able to take bigger time steps once the trajectory has come sufficiently close to the slow manifold, without compromising the accuracy of the computed trajectory. That this is not always the case is well known to numerical analysts and to many practitioners, as is the fact that in general explicit methods must continue to take time steps that are of the order of the fastest time scale. Implicit methods, on the other hand, avoid the above described instability, but at the expense of having to solve a nonlinear system of equations for the unknown point at each time step. In fact, implicit methods often damp the perturbations off the slow manifold. Once the solution has reached the stable manifold, this damping keeps the solution on the manifold, and is desirable. Although implicit methods are clearly more expensive per step than explicit methods, this cost is more than compensated for in solving stiff systems by the fact that the implicit methods can take much larger timesteps [33].

When stochasticity is introduced into a system with fast and slow time scales, with fast modes being stable as before, one may still expect a slow manifold corresponding to the equilibrium of the fast scales. But the picture changes in a fundamental way. After an initial rapid transient,

while the mean trajectory is almost on the slow manifold, any sample trajectory will still be fluctuating on the fast time scale in a direction transverse to the slow manifold. In some cases the size of the fluctuations about the slow manifold will be practically negligible, and an implicit scheme may take large steps, corresponding to the time scale of the slow mode. But in other cases, the fluctuations about the slow manifold will *not* be negligible in size. In those instances, an implicit scheme that takes time steps much larger than the time scale of the fast dynamics will dampen these fluctuations, and will consequently fail to capture the variance correctly. This can be corrected at relatively low expense via a procedure called *down-shifting* that we will describe shortly.

The original (explicit) tau-leaping method is an explicit method because the propensity functions a_j are evaluated at the current known state, so the future unknown random state $X(t+\tau)$ is given as an explicit function of $X(t)$. It is the explicit construction of this formula that leads to the problem with stiffness, just as in the case of explicit methods for ODEs. An implicit tau-leaping method can be defined by [34]

$$X(t+\tau) = X(t) + \sum_{j=1}^{M} v_j a_j(X(t+\tau))\tau + \sum_{j=1}^{M} v_j(\mathscr{P}_j(a_j(X(t)),\tau) - a_j(X(t))\tau) \tag{13}$$

In the implementation of the method (13), the random variables $\mathscr{P}_j(a_j(X(t)),\tau)$ can be generated without knowing $X(t+\tau)$. Also, once the $\mathscr{P}_j(a_j(X(t)),\tau)$ have been generated, the unknown state $X(t+\tau)$ depends on $\mathscr{P}_j(a_j(X(t)),\tau)$ in a deterministic way, even though this dependence is given by an implicit equation. As is done in the case of deterministic ODE solution by implicit methods, $X(t+\tau)$ can be computed by applying Newton's method for the solution of nonlinear systems of equations to (13) where the $\mathscr{P}_j(a_j(X(t)),\tau)$ are all known values. Just as the explicit-tau method segues to the explicit Euler methods for SDEs and ODEs, the implicit-tau method segues to the implicit Euler methods for SDEs and ODEs. However, while the implicit tau-leaping method computes the *slow* variables with their correct distributions, it computes the *fast* variables with the correct means but with distributions about those means that are too narrow. A time-stepping strategy called *down-shifting* [34] can restore the overly-damped fluctuations in the fast variables. The idea is to interlace implicit tau-leaps, each of which is on the order of the time scale of the slow variables and hence 'large', with a sequence of much smaller time steps on the order of the time scale of the fast variables. The smaller time steps are taken over a duration that is comparable to the 'relaxation/decorrelation' time of the fast variables. These small timesteps may be taken with either the explicit-tau method or the SSA. This sequence of small steps is intended to 'regenerate' the correct statistical distributions of the fast variables. The fact that the underlying kinetics is Markovian or 'past-forgetting' is key to being able to apply this procedure.

5.2. The slow-scale SSA

When reactions take place on vastly different time scales, with 'fast' reaction channels firing very much more frequently than 'slow' ones, a simulation with the SSA will spend most of its time on the more numerous fast reaction events. This is an inefficient allocation of computational effort, especially when fast reaction events are much less important than slow ones. One possibility for speeding up the computation is the use of hybrid methods [35,36]. These methods combine the RRE used to simulate fast reactions involving large populations, with the SSA used to simulate slow reactions with small populations. However, these algorithms do not provide any

efficient way to simulate fast reactions involving a species with small population (for example the heat shock problem discussed in Section 3.3.4). Another possibility is to find a way to skip over the fast reactions and explicitly simulate only the slow reactions. The recently developed [37] *slow-scale SSA* (ssSSA) provides a way to do this. This algorithm is closely related to the stochastic quasi steady-state assumption [38]. The first step in setting up the ssSSA is to divide (and re-index) the M reaction channels $\{R_1, \ldots, R_M\}$ into fast and slow subsets, $\{R_1^{\mathrm{f}}, \ldots, R_{M_{\mathrm{f}}}^{\mathrm{f}}\}$ and $\{R_1^{\mathrm{s}}, \ldots, R_{M_{\mathrm{s}}}^{\mathrm{s}}\}$, where $M_{\mathrm{f}} + M_{\mathrm{s}} = M$. We initially do this provisionally (subject to possible later change) according to the following criterion: the propensity functions of the fast reactions, $a_1^{\mathrm{f}}, \ldots, a_{M_{\mathrm{f}}}^{\mathrm{f}}$, should usually be very much larger than the propensity functions of the slow reactions, $a_1^{\mathrm{s}}, \ldots, a_{M_{\mathrm{s}}}^{\mathrm{s}}$. Due to this partitioning, the time to the occurrence of the next fast reaction will usually be very much smaller than the time to the occurrence of the next slow reaction.

Next we divide (and re-index) the N species $\{S_1, \ldots, S_N\}$ into fast and slow subsets, $\{S_1^{\mathrm{f}}, \ldots, S_{N_{\mathrm{f}}}^{\mathrm{f}}\}$ and $\{S_1^{\mathrm{s}}, \ldots, S_{N_{\mathrm{s}}}^{\mathrm{s}}\}$, where $N_{\mathrm{f}} + N_{\mathrm{s}} = N$. This gives rise to a partitioning of the state vector $X(t) = (X^{\mathrm{f}}(t), X^{\mathrm{s}}(t))$, and also the generic state space variable $x = (x^{\mathrm{f}}, x^{\mathrm{s}})$, into fast and slow parts. A fast species is defined to be any species whose population *gets changed* by some fast reaction; all the other species are slow. Note the asymmetry in this definition: a slow species cannot get changed by a fast reaction, but a fast species can get changed by a slow reaction. Note also that a_j^{f} and a_j^{s} can both depend on both fast and slow variables. The state-change vectors can now be re-indexed

$$v_j^{\mathrm{f}} \equiv (v_{1j}^{\mathrm{ff}}, \ldots, v_{N_{\mathrm{f}}j}^{\mathrm{ff}}), \quad j = 1, \ldots, M_{\mathrm{f}}$$

$$v_j^{\mathrm{s}} \equiv (v_{1j}^{\mathrm{fs}}, \ldots, v_{N_{\mathrm{f}}j}^{\mathrm{fs}}, v_{1j}^{\mathrm{ss}}, \ldots, v_{N_{\mathrm{s}}j}^{\mathrm{ss}}), \quad j = 1, \ldots, M_{\mathrm{s}}$$

where $v_{ij}^{\sigma\rho}$ denotes the change in the number of molecules of species S_i^{σ} ($\sigma = \mathrm{f}, \mathrm{s}$) induced by one reaction R_j^{ρ} ($\rho = \mathrm{f}, \mathrm{s}$). We can regard v_j^{f} as a vector with the same dimensionality (N_{f}) as X^{f}, because $v_{ij}^{\mathrm{sf}} \equiv 0$ (slow species do not get changed by fast reactions).

The next step in setting up the ssSSA is to introduce the *virtual fast process* $\hat{X}^{\mathrm{f}}(t)$. It is composed of the same fast species state variables as the real fast process $X^{\mathrm{f}}(t)$, but evolving *only* through the fast reactions; i.e. $\hat{X}^{\mathrm{f}}(t)$ is $X^{\mathrm{f}}(t)$ with all the slow reactions switched off. To the extent that the slow reactions do not occur very often, we may expect that $\hat{X}^{\mathrm{f}}(t)$ will provide a good approximation to $X^{\mathrm{f}}(t)$. But from a mathematical standpoint there is an profound difference: $X^{\mathrm{f}}(t)$ by itself is not a Markov process, whereas $\hat{X}^{\mathrm{f}}(t)$ is. Since the evolution of $X^{\mathrm{f}}(t)$ depends on the evolving slow process $X^{\mathrm{s}}(t)$, $X^{\mathrm{f}}(t)$ will not be governed by a master equation of the simple Markovian form. But for the *virtual* fast process $\hat{X}^{\mathrm{f}}(t)$, the slow process $X^{\mathrm{s}}(t)$ stays fixed at some constant initial value x_0^{s}; therefore, $\hat{X}^{\mathrm{f}}(t)$ evolves according to the master equation,

$$\frac{\partial \hat{P}(x^{\mathrm{f}}, t \mid x_0, t_0)}{\partial t} = \sum_{j=1}^{M_{\mathrm{f}}} [a_j^{\mathrm{f}}(x^{\mathrm{f}} - v_j^{\mathrm{f}}, x_0^{\mathrm{s}}) \hat{P}(x^{\mathrm{f}} - v_j^{\mathrm{f}}, t \mid x_0, t_0) - a_j^{\mathrm{f}}(x^{\mathrm{f}}, x_0^{\mathrm{s}}) \hat{P}(x^{\mathrm{f}}, t \mid x_0, t_0)]$$

where $\hat{P}(x^{\mathrm{f}}, t \mid x_0, t_0)$ is the probability that $\hat{X}^{\mathrm{f}}(t) = x^{\mathrm{f}}$, given that $X(t_0) = x_0$. Note that the j-sum here runs over only the fast reactions. In order to apply the ssSSA, *two conditions* must be satisfied. The first is that the virtual fast process $\hat{X}^{\mathrm{f}}(t)$ be *stable*, in the sense that it approaches a well defined, time-independent random variable $\hat{X}^{\mathrm{f}}(\infty)$ as $t \to \infty$; i.e.

$$\lim_{t \to \infty} \hat{P}(x^{\mathrm{f}}, t \mid x_0, t_0) \equiv \hat{P}(x^{\mathrm{f}}, \infty \mid x_0)$$

$\hat{P}(x^{\mathrm{f}}, \infty \mid x_0)$ can be calculated from the stationary form of the master equation,

$$0 = \sum_{j=1}^{M_{\mathrm{f}}} [a_j^{\mathrm{f}}(x^{\mathrm{f}} - v_j^{\mathrm{f}}, x_0^{\mathrm{s}})\hat{P}(x^{\mathrm{f}} - v_j^{\mathrm{f}}, \infty \mid x_0) - a_j^{\mathrm{f}}(x^{\mathrm{f}}, x_0^{\mathrm{s}})\hat{P}(x^{\mathrm{f}}, \infty \mid x_0)]$$

which will be easier to solve since it is purely algebraic. The second condition is that the relaxation of $\hat{X}^{\mathrm{f}}(t)$ to its stationary asymptotic form $\hat{X}^{\mathrm{f}}(\infty)$ happen *very quickly* on the time scale of the *slow* reactions. More precisely, we require that the relaxation time of the virtual fast process be very much less than the expected time to the next slow reaction. These two conditions will usually be satisfied if the system is *stiff*. If these conditions cannot be satisfied for any partitioning of the reactions, we shall take that as a sign that the fast reactions are no less important than the slow ones, so it is not a good idea to skip over them.

Given these definitions and conditions, it is possible to prove the *slow-scale approximation* [37]: If the system is in state $(x^{\mathrm{f}}, x^{\mathrm{s}})$ at time t, and if Δ_{s} is a time increment that is *very large* compared to relaxation time of $\hat{X}^{\mathrm{f}}(t)$ but *very small* compared to the expected time to the next slow reaction, then the probability that one R_j^{s} reaction will occur in the time interval $[t, t + \Delta_{\mathrm{s}})$ can be well approximated by $\bar{a}_j^{\mathrm{s}}(x^{\mathrm{s}}; x^{\mathrm{f}})\,\Delta_{\mathrm{s}}$ where

$$\bar{a}_j^{\mathrm{s}}(x^{\mathrm{s}}; x^{\mathrm{f}}) \triangleq \sum_{x^{\mathrm{f}\prime}} \hat{P}(x^{\mathrm{f}\prime}, \infty \mid x^{\mathrm{f}}, x^{\mathrm{s}})\, a_j^{\mathrm{s}}(x^{\mathrm{f}\prime}, x^{\mathrm{s}}) \tag{14}$$

We call $\bar{a}_j^{\mathrm{s}}(x^{\mathrm{s}}; x^{\mathrm{f}})$ the *slow-scale propensity* function for reaction channel R_j^{s}. It is the average of the regular R_j^{s} propensity function over the fast variables, treated as though they were distributed according to the asymptotic virtual fast process $\hat{X}^{\mathrm{f}}(\infty)$. The ssSSA is an immediate consequence of this slow-scale approximation as we shall see in the next section.

5.3. Example: a toy model of a heat shock subsystem

To illustrate the ssSSA, we consider again the reaction set discussed in Section 3.3.4. depicting a simplified model of the heat shock response

$$S_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_2 \overset{c_3}{\rightarrow} S_3 \tag{15}$$

under the condition $c_2 \gg c_3$. An SSA simulation of this model will be mostly occupied with simulating occurrences of reactions R_1 and R_2. These reactions are 'uninteresting' since they just keep undoing each other. Only rarely will the SSA produce an important transcription-initiating R_3 reaction. We take the fast reactions to be R_1 and R_2, and the slow reaction to be R_3. The fast species will then be S_1 and S_2, and the slow species S_3. The virtual fast process $\hat{X}^{\mathrm{f}}(t)$ will be the S_1 and S_2 populations undergoing *only* the fast reactions R_1 and R_2. Unlike the *real* fast process, the *virtual* fast process obeys the conservation relation

$$\hat{X}_1(t) + \hat{X}_2(t) = x_{\mathrm{T}} \quad \text{(constant)} \tag{16}$$

This relation greatly simplifies the analysis of the virtual fast process, since it reduces the problem to a single independent state variable. Eliminating $\hat{X}_2(t)$ in favour of $\hat{X}_1(t)$ by means of Equation (16), we see that given $\hat{X}_1(t) = x_1'$, $\hat{X}_1(t + \mathrm{d}t)$ will equal $x_1' - 1$ with probability $c_1 x_1'\,\mathrm{d}t$, and $x_1' + 1$ with probability $c_2(x_{\mathrm{T}} - x_1')\,\mathrm{d}t$. $\hat{X}_1(t)$ is therefore what is known mathematically as a 'bounded birth–death' Markov process [18]. It can be shown [31] that this process has, for any

initial value $x_1 \in [0, x_T]$, the asymptotic stationary distribution

$$\hat{P}(x_1', \infty \mid x_T) = \frac{x_T!}{x_1'!(x_T - x_1')!} q^{x_1'}(1-q)^{x_T - x_1'} \quad (x_1' = 0, 1, \ldots, x_T) \tag{17}$$

where $q \equiv c_2/(c_1 + c_2)$. Thus, $\hat{X}_1(\infty)$ is the binomial random variable $\mathscr{B}(q, x_T)$, whose mean and variance are

$$\langle \hat{X}_1(\infty) \rangle = x_T q = \frac{c_2 x_T}{c_1 + c_2} \tag{18a}$$

$$\mathrm{var}\{\hat{X}_1(\infty)\} = x_T q(1-q) = \frac{c_1 c_2 x_T}{(c_1 + c_2)^2} \tag{18b}$$

It can also be shown [37] that $\hat{X}_1(t)$ relaxes to $\hat{X}_1(\infty)$ in a time of order $(c_1 + c_2)^{-1}$.

The slow scale propensity function for the slow reaction R_3 is, according to the result (5.2), the average of $a_3(x) = c_3 x_2$ with respect to $\hat{X}^f(\infty)$. Therefore, using Equations (16) and (18a),

$$\bar{a}_3(x_3; x_1, x_2) = c_3 \langle \hat{X}_2(\infty) \rangle = \frac{c_3 c_1 (x_1 + x_2)}{c_1 + c_2} \tag{19}$$

Since the reciprocal of $\bar{a}_3(x_3; x_1, x_2)$ estimates the average time to the next R_3 reaction, the condition that the relaxation time of the virtual fast process be very much smaller than the mean time to the next slow reaction is

$$c_1 + c_2 \gg \frac{c_3 c_1 (x_1 + x_2)}{c_1 + c_2} \tag{20}$$

This condition will evidently be satisfied if the inequality $c_2 \gg c_3$ is sufficiently strong. When it is satisfied, the ssSSA for reactions (15) goes as follows:

1. Given $X(t_0) = (x_{10}, x_{20}, x_{30})$, set $t \leftarrow t_0$ and $x_i \leftarrow x_{i0}$ $(i = 1, 2, 3)$.
2. In state (x_1, x_2, x_3) at time t, compute $\bar{a}_3(x_3; x_1, x_2)$ from Equation (19).
3. Draw a unit-interval uniform random number r, and compute
$$\tau = \frac{1}{\bar{a}_3(x_3; x_1, x_2)} \ln\left(\frac{1}{r}\right)$$
4. Advance to the next R_3 reaction by replacing $t \leftarrow t + \tau$ and
$$x_3 \leftarrow x_3 + 1, \quad x_2 \leftarrow x_2 - 1$$
Then, with $x_T = x_1 + x_2$,
$x_1 \leftarrow$ sample of $\mathscr{B}(\frac{c_2}{c_1 + c_2}, x_T)$, and $x_2 \leftarrow x_T - x_1$.
5. Record (t, x_1, x_2, x_3) if desired. Then return to Step 2, or else stop.

In Step 4, the first pair of updates actualizes the R_3 reaction while the second pair of updates 'relaxes' the fast variables in a manner consistent with the stationary distribution (17) and the new value of x_T.

Figure 4(a) shows the results of an exact SSA run of reactions (15) for the parameter values

$$c_1 = 10, \quad c_2 = 4 \times 10^4, \quad c_3 = 2; \quad x_{10} = 2000, \quad x_{20} = x_{30} = 0 \tag{21}$$

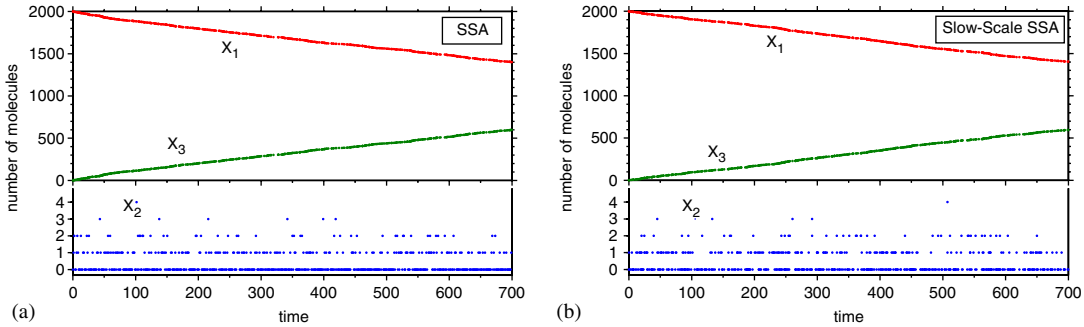

Figure 4. Two simulations of reactions (15) using the parameter values (21): (a) shows the trajectories produced in an exact SSA run, with the populations being plotted out essentially after each R_3 reaction (see text for details); and (b) shows the trajectories produced in a ssSSA run in which only the R_3 reactions, which totaled 587, were directly simulated, and the populations are plotted out after each of those. The ssSSA simulation ran over 1000 times faster than the SSA simulation.

The S_1 and S_3 populations here are plotted out immediately after each R_3 reaction. There were over 23 million reactions simulated here, but only about 600 of those were R_3 reactions. The S_2 population, which is shown on a separate scale, is plotted out at the same number of equally spaced time intervals. This plotting gives a more typical picture of the behaviour of the S_2 population than would be obtained by plotting it immediately after each R_3 reaction, because R_3 reactions are more likely to occur when the S_2 population is larger.

For the parameter values (21), condition (20) is satisfied by 4 orders of magnitude initially, and even more so as the total population of S_1 and S_2 declines; therefore, this reaction set should be amenable to simulation using the ssSSA procedure. Figure 4(b) shows the results of such a simulation, plotted after each R_3 reaction. We note that all the species trajectories in this approximate ssSSA run agree very well with those in the exact SSA run of Figure 4(a); even the behaviour of the sparsely populated species S_2 is accurately replicated by the ssSSA. But whereas the construction of the SSA trajectories in Figure 4(a) required simulating over 23 million reactions, the construction of the slow-scale SSA trajectories in Figure 4(b) entailed simulating only the 587 slow (R_3) reactions. The slow-scale SSA run was over 1000 times faster than the SSA run, with no perceptible loss of accuracy.

6. CONCLUSION

The dynamic and stochastic nature of gene networks offers new research opportunities in the efficient modelling, simulation, and analysis of such networks. In this article, we motivated a stochastic approach to modelling genetic networks and outlined some of the key mathematical approaches to molecular noise modelling. We described numerical simulation approaches based on the SSA and outlined recent advances in the simulation of molecular noise models.

ACKNOWLEDGEMENTS

The authors would like to acknowledge John Doyle, Caltech for many insightful discussions. This work was supported in part by the U.S. Air Force Office of Scientific Research and the California Institute of

Technology under DARPA Award No. F30602-01-2-0558, by the U.S. Department of Energy under DOE award No. DE-FG02-04ER25621, by the National Science Foundation under NSF awards CCF-0326576 and ACI00-86061, and by the Institute for Collaborative Biotechnologies through grant DAAD19-03-D-0004 from the U.S. Army Research Office. Dan Gillespie received additional support from the Molecular Sciences Institute under contract No. 244725 with Sandia National Laboratories and the Department of Energy's Genomes to Life Program.

REFERENCES

1. Hasty J, McMillen D, Isaacs F, Collins JJ. Computational studies of gene regulatory networks: in numero molecular biology. *Nature Reviews Genetics* 2001; **2**(4):269–268.
2. Bennett DC. Differentiation in mouse melanoma cells: initial reversibility and an in-off stochastic model. *Cell* 1983; **34**(2):445–453.
3. Dingemanse MA, De Boer PA, Moorman AF, Charles R, Lamers WH. The expression of liver specific genes within rat embryonic hepatocytes is a discontinuous process. *Differentiation* 1994; **56**(3):153–162.
4. Walters MC, Fiering S, Eidemiller J, Magis W, Groudine M, Martin DIK. Enhancers increase the probability but not the leven of gene expression. *Proceedings of the National Academy of Sciences USA* 1995; **92**(15): 7125–7129.
5. Ko M, Nakauchi SH, Takahashi N. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *EMBO Journal* 1990; **9**(9):2835–2842.
6. Arkin A, Ross J, McAdams HH. Stochastic kinetic analysis of the developmental pathway bifurcation in phase λ-infected *Escherichia coli* cells. *Genetics* 1998; **149**(4):1633–1648.
7. Elowitz M, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature* 2000; **403**(6767):335–338.
8. Besckei A, Seraphin B, Serrano L. Positive feedback in eukaryotic gene networks: cell differentiation by graded binary response. *EMBO Journal* 2001; **20**(10):2528–2535.
9. Gardner TS, Cantor CR, Collings JJ. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 2000; **403**(6767):339–342.
10. Jacob F, Monod J. On the regulation of gene activity. *Cold Spring Harbor Symposium on Quantitative Biology*, vol. 26, 1961; 193–211, 389–401.
11. Goodwin BC. *Temporal Organization in Cells*. Academic Press: New York, 1963.
12. Ferrell JE, Machleder EM. The biochemical basis for an all-or-none cell fate in xenopus oocytes. *Science* 1998; **280**(1):895–898.
13. Hasty J, Pradines J, Dolnik J, Collins JJ. Noise-based switches and amplifiers for gene expression. *Proceedings of the National Academy of Science USA* 2000; **97**(5):2075–2080.
14. Vilar J, Kueh HY, Barkai N, Leibler S. Mechanisms of noise resistance in genetic oscillators. *Proceedings of the National Academy of Sciences USA* 2002; **99**(9):5988–5992.
15. El Samad H. Mechanisms of noise exploitation in gene regulatory networks. Biological design principles for robustness, performance, and selective interactions with noise. *Ph.D. Dissertation*, University of California at Santa Barbara, 2004.
16. El-Samad H, Kurata H, Doyle J, Gross C, Khammash M. Surviving heat shock: control strategies for robustness and performance. *Proceedings of the National Academy of Sciences USA* 2005; **102**(8):2736–2741.
17. Gillespie D. A rigorous derivation of the chemical master equation. *Physica A* 1992; **188**:404–425.
18. Gillespie D. *Markov Processes*: *An Introduction for Physical Scientists*. Academic Press: San Diego, 1992.
19. Gillespie D. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* 1976; **22**(4):403–434.
20. McQuarrie D. Stochastic approach to chemical kinetics. *Journal of Applied Probability* 1967; **4**:413–478.
21. Gillespie D. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry* 1977; **81**: 2340–2361.
22. Gibson M, Bruck J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry* 2000; **104**:1876–1889.
23. Cao Y, Li H, Petzold L. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *Journal of Chemical Physics* 2004; **121**(9):4059–4067.
24. Gillespie D. Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Chemical Physics* 2001; **115**:1716–1733.
25. Gillespie D, Petzold L. Improved leap-size selection for accelerated stochastic simulation. *Journal of Chemical Physics* 2003; **119**:8229–8234.
26. Cao Y, Gillespie DT, Petzold LR. Avoiding negative populations in explicit Poisson tau-leaping. *Journal of Chemical Physics* 2005; **123**(5).

27. Tian T, Burrage K. Binomial leap methods for simulating stochastic chemical kinetics. *Journal of Chemical Physics* 2004; **121**:10356–10364.
28. Chatterjee A, Vlachos D, Katsoulakis M. Time accelerated Monte Carlo simulations of biological networks using the binomial tau-leap method. *Journal of Chemical Physics* 2005; **122**:014112.
29. Rathinam M, Petzold LR, Cao Y, Gillespie D. Consistency and stability of tau-leaping schemes for chemical reaction systems. *Multiscale Modeling and Simulation* (SIAM) 2005, to appear.
30. Gillespie D. The chemical Langevin equation. *Journal of Chemical Physics* 2000; **113**:297–306.
31. Gillespie D. The chemical Langevin and Fokker–Planck equations for the reversible isomerization reaction. *Journal of Physical Chemistry* 2002; **106**:5063–5071.
32. Kurtz T. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications* 1978; **6**:223–240.
33. Ascher UM, Petzold LR. *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM: Philadelphia, 1998.
34. Rathinam M, Petzold L, Cao Y, Gillespie D. Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method. *Journal of Chemical Physics* 2003; **119**:12784–12794.
35. Haseltine E, Rawlings J. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *Journal of Chemical Physics* 2002; **117**:6959–6969.
36. Kiehl TR, Mattheyses T, Simmons M. Hybrid simulation of cellular behavior. *Bioinformatics* 2004; **20**:316322.
37. Cao Y, Gillespie D, Petzold L. The slow-scale stochastic simulation algorithm. *Journal of Chemical Physics* 2005; **122**:014116(18 pp.).
38. Rao C, Arkin A. Stochastic chemical kinetics and the quasi steady-state assumption: application to the Gillespie algorithm. *Journal of Chemical Physics* 118 + 4999-5010, 2003.

