Stochastic Models of Biological Processes

Steven S. Andrews[1], Tuan Dinh[1], Adam P. Arkin[1,2]
[1] Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, USA
[2] Department of Bioengineering and Howard Hughes Medical Institute, University of California, Berkeley, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Non-Spatial Stochastic Modeling
Spatial Stochastic Modeling
Future Directions
Box 1: The *E. Coli* Min System
Acknowledgments
Bibliography

Glossary

Brownian dynamics A level of detail in which each molecule is represented by a point-like particle and molecules move in response to diffusion and collisions

Chemical Fokker–Planck equation (CFPE) Master equation for well-mixed systems that corresponds to the chemical Langevin equation

Chemical master equation (CME) Master equation for the probability that the system has specific integer copy numbers for each type of chemical species; it is exact for a well-mixed system

Chemical Langevin equation (CLE) Approximate stochastic differential equation for well-mixed systems which is based on continuous Gaussian statistics

Direct method An implementation of the Gillespie algorithm

Extrinsic noise In genetic noise studies, expression fluctuations of a gene that arise from upstream genes or global fluctuations

First-reaction method An implementation of the Gillespie algorithm

Gillespie algorithm Exact algorithm for simulating individual trajectories of the CME

Hybrid algorithms Algorithms that are designed to efficiently simulate systems that have multiple timescales

Individual-based spatial models Models that track individual molecules as they diffuse or react

Intrinsic noise Expression fluctuations of a gene that arise from that particular gene

Jump process A process in which the system abruptly changes from one state to another

Optimized direct method A computationally efficient implementation of the Gillespie algorithm

Population-based spatial models Models that track how many molecules of each chemical species are in various spatial compartments

Reaction channel A possible reaction between specified reactant and product chemical species (the terminology distinguishes this meaning from an individual reaction event between single molecules)

Reaction-diffusion equation Deterministic partial differential equation that combines mass action reaction kinetics and normal chemical diffusion

Reaction-diffusion master equation (RDME) Chemical master equation that accounts for diffusion as well as reactions

Reaction rate equation (RRE) Deterministic ordinary differential equation for the net production rate of each chemical species from chemical reactions

Spatial chemical Langevin equation Chemical Langevin equation that accounts for diffusion as well as reactions

Stochastic simulation algorithm (SSA) Alternative term for the Gillespie algorithm

Stoichiometric matrix (ν) Matrix that gives the net production of each chemical species, for each chemical reaction

Tau-leaping method Approximate simulation method for well-mixed systems in which molecule numbers are updated using discrete Poisson statistics

Well-mixed hypothesis Assumption that mixing processes occur faster than the relevant reaction processes

Definition of the Subject

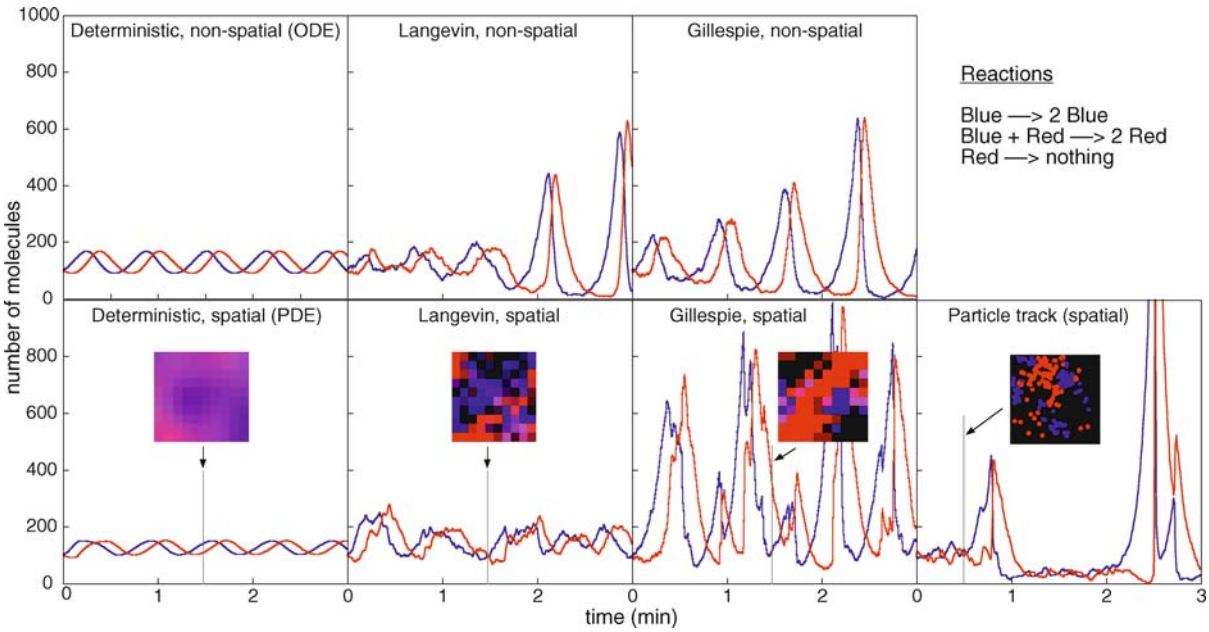
Many processes in cell biology, such as those that carry out metabolism, the cell cycle, and various types of signaling, are comprised of biochemical reaction networks. It has proven useful to study these networks using computer simulations because they allow us to quantitatively investigate hypotheses about the networks. Deterministic simulations are sufficient to predict average behaviors at the population level, but they cannot address questions about noise, random switching between stable states of the system, or the behaviors of systems with very few molecules of key species. These topics are investigated with stochastic simulations. In this article, we review the dominant types of stochastic simulation methods that are used to investigate biochemical reaction networks, as well as some of the results that have been found with them. As new biological experiments continue to reveal more detail about biological systems, and as computers continue to become more powerful, researchers will increasingly turn to simulation methods that can address stochastic and spatial details.

Introduction

Random events are ubiquitous throughout biology. Diffusion, chemical reactions, gene expression, homologous recombination, and most other fundamental biological processes are governed to a large extent by the inherently discrete and stochastic interactions of molecules [1]. In many cases, the random events that occur on very small length and time scales become averaged out when one focuses on larger length or time scales. However, there also exist many examples in which stochastic fluctuations at small scales propagate up to and then influence the system behavior at large scales. Examples range from the swimming trajectories of individual bacteria all the way up to the genetic diversity on which evolution depends.

Research on stochasticity in biochemical systems has received a great deal of attention lately, leading to many recent reviews [2,3,4,5,6,7,8,9]. One reason for its popularity is that the basic designs of many biochemical systems, such as metabolism, cell division, and chemotaxis, are becoming reasonably well understood. Starting from this understanding, researchers are delving deeper to examine the quantitative behaviors of these systems, including the roles of stochastic influences. Also, there is an increasing awareness of the importance of stochasticity in biological systems. For example, it has become clear that noisy gene expression is the rule rather than the exception [9]; this leads to important questions about non-genetic individuality and about biological robustness to gene expression noise. Thirdly, stochasticity is often investigated using computationally demanding simulations. Cheaper and faster computers, as well as improved simulation algorithms, are making it feasible for researchers to investigate more complex biochemical systems, at increasingly realistic levels of detail. Finally, and perhaps most importantly, the last ten years have witnessed incredible progress in experimental biochemical methods, some of which allow direct measurements of stochasticity on microscopic size scales. These methods include gene expression measurements [10,11], flow cytometry [12,13,14,15,16], single molecule detection methods [17,18,19,20], and applications of synthetic biology [21].

At the most fundamental level, the quantum mechanics that describe the dynamics of all physical systems are well-understood and completely deterministic [22]. It is



Stochastic Models of Biological Processes, Figure 1

Simulation results for a simple chemical oscillator using different simulation methods. The Lotka–Volterra system is shown, which shares key features with cellular oscillators such as circadian rhythms. *Insets* show the spatial distributions of molecules at the indicated times. In the *top panels*, note that stochasticity allows the system to drift to large amplitude oscillations and that the Langevin and Gillespie methods yield similar results. In the *bottom panels*, all of which were started with nearly homogeneous initial states, differences arise from the approximations: the PDE simulation has predictable oscillations due to the minimal stochasticity (which is only in the initial state); the Gillespie simulation has larger peaks than the Langevin one because it only allows integer numbers of molecules in each bin; and the particle tracking simulation shows larger and fewer bursts than does the Gillespie simulation because it accurately treats diffusion at all length scales (this difference was reduced with a spatial Gillespie simulation that used smaller subvolumes). Parameters: rate constants are 10 min^{-1}, 8000 nm^3 $molec^{-1}$ min^{-1}, and 10 min^{-1}, for the respective reactions shown in the *top-right corner*, systems start with 100 of each *blue* and *red* molecules, their diffusion coefficients are 100 nm^2 min^{-1}, the volume is 100 nm high and wide by 10 nm deep, and the first three spatial simulations divide this volume into cubic subvolumes that are 10 nm on a side. This figure is reproduced from [8]

only when systems are observed that there arises unavoidable randomness, although these aspects of quantum mechanics remain murky and as close to philosophy as science. More importantly, essentially any system that is governed by nonlinear dynamics, including nearly all physical systems, rapidly becomes chaotic as the system size is increased beyond a few molecules, and thus becomes effectively unpredictable [23]. For all intents and purposes, the diffusive trajectories of individual molecules, and the probabilities of chemical reactions occurring between neighboring reactants at specific times, are fundamentally stochastic.

In the laboratory, partly by design, this stochasticity usually averages out. For systems comprised of many particles, diffusion is observed to be described well by Fick's law of diffusion and chemical reaction kinetics are described well by the deterministic reaction rate equations [24]. The situation is often different within biological cells for several reasons: (i) the program for cellular behavior, the genome, is present at low copy number and yet each gene governs the expression of possibly thousands of proteins, (ii) the low copy numbers of many proteins and mRNA transcripts within cells make random variation of their numbers a relatively large fraction of the total, (iii) stochasticity is often amplified during sequential biochemical steps, of which an especially important case is the sequence of DNA transcription followed by mRNA translation, and (iv) because of spatial organization within cells, it often takes very few proteins in specific locations to achieve large effects on the entire cell dynamics.

This review provides a broad account of stochastic modeling of biological processes. The emphasis is placed on stochastic processes at the cellular level although much of the work presented here also applies to other scales and systems. Our goal is to briefly familiarize the reader with the mathematical forms of the most important equations,

the tools for analyzing and simulating them, and some applications for which they have been particularly successful. As shown below, the mathematics, the software implementation, and even the applications of modeling methods are often closely linked.

There are several ways to categorize stochastic biochemical modeling methods. A key designation is whether a method is spatial or non-spatial: spatial models treat spatial organization of proteins and membranes explicitly, whereas non-spatial models include an implicit assumption that mixing processes occur faster than the relevant reaction processes, which is called the *well-mixed hypothesis*. No single modeling method can efficiently capture stochastic dynamics over wide ranges of time scales, so separate methods have been developed that operate at levels of temporal detail that range from nanoseconds to hours. Hybrid simulators combine methods that operate at different timescales to allow the efficient simulation of systems that include both fast and slow processes. Whereas many modeling methods are designed solely to address the reactions in a biochemical reaction network, others also consider system boundaries, mechanics, or the multiple states that proteins can be in. Here, we present non-spatial modeling methods first, followed by spatial methods, with diversions along the way to touch on as many of the other topics as possible. Simulation results from many of the methods that are discussed are compared in Fig. 1, where it is seen that the differences can be quite significant.

Non-Spatial Stochastic Modeling

Deterministic Modeling and Notation

Before focusing on stochastic modeling, it is helpful to introduce the notation and some terminology by summarizing a few aspects of deterministic modeling. As applied to chemical reaction networks, deterministic modeling is based upon ordinary differential equations (ODEs) for the individual chemical reactions. Consider the generic elementary reversible reaction

$$A + B \underset{k_r}{\overset{k_f}{\rightleftharpoons}} C \,, \tag{1}$$

where k_f and k_r are the forward and reverse reaction rates, respectively. We assume that the system is kept well-mixed so that diffusion effects can be ignored. The reaction rate equations for components A, B, and C are the ODEs

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -k_f[A][B] + k_r[C] \,, \tag{2a}$$

$$\frac{d[C]}{dt} = k_f[A][B] - k_r[C] \,. \tag{2b}$$

More complex reaction networks are expressed analogously, with one equation for each chemical species and with terms in the equations that represent chemical reactions. From these equations, the reactions can be simulated to show how the concentrations change over time. Or, after setting the left sides of the equations to zero, they can be solved to yield the steady-state chemical concentrations. One can also investigate the dynamic or steady-state behaviors as the reaction rate parameters (k_f and k_r), or initial concentrations, are varied [25]; this can yield phase diagrams for the reactions and additional insight.

It is helpful to generalize the rate equations given above to make them more convenient for computational or analytical work and to show their forms more clearly. First, each chemical concentration is replaced by the variable $Z_i(t)$, where i is an index for A, B, or C and the time-dependence is written out explicitly. Next, the product terms in the equations are replaced by the functions $\tilde{a}_f(\mathbf{Z}(t))$ and $\tilde{a}_r(\mathbf{Z}(t))$ for the forward and reverse reactions, respectively; the tildes indicate that molecule quantities are given as concentrations rather than as molecule numbers. These functions are called the *reaction propensities*. Finally, the '+' or '–' signs show the reaction stoichiometry. They are replaced by ν_{fi} and ν_{ri} for the forward and reverse reactions, respectively, which are elements of the so-called *stoichiometric matrix* (throughout this review, we follow Gillespie's notation [6]). In this example, one unit of each A and B are lost in a unit amount of forward reaction ($\nu_{fA} = \nu_{fB} = -1$) as one unit of C is formed ($\nu_{fC} = 1$); the ν_{ri} values have the opposite signs. With these substitutions, the rate equations become

$$\frac{dZ_i(t)}{dt} = \nu_{fi}\tilde{a}_f(\mathbf{Z}(t)) + \nu_{ri}\tilde{a}_r(\mathbf{Z}(t)) \,, \tag{3a}$$

$$\tilde{a}_f(\mathbf{Z}(t)) = k_f Z_A Z_B \,, \tag{3b}$$

$$\tilde{a}_r(\mathbf{Z}(t)) = k_r Z_C \,. \tag{3c}$$

These equations are trivially extended to arbitrarily large reaction networks. Consider a system with N chemical species that can react via M different *reaction channels* (a "reaction channel" is simply unambiguous terminology for a reaction between specific reactant and product chemical species). The dynamics of this system are given with the *reaction rate equation* (RRE):

$$\frac{dZ_i(t)}{dt} = \sum_{j=1}^{M} \nu_{ji}\tilde{a}_j(\mathbf{Z}(t)) \,. \tag{4}$$

The reaction propensity equations are typically the products of reaction rate constants and the appropriate chemical concentrations, as shown above (Eq. (3b) and (3c)),

but they may also describe non-elementary processes such as Michaelis–Menten kinetics. It is worth noting that the state of the system, at any point in time, is fully expressed with the vector $\mathbf{Z}(t)$. This means that the entire trajectory of the system can be deterministically calculated from the RRE and any single $\mathbf{Z}(t)$ snapshot.

The RRE is at the heart of many branches of quantitative biochemistry and systems biology. A great deal of metabolic theory is based on either the steady-state solutions of the RRE, or the set of steady-state solutions that are possible, given only knowledge of the stoichiometric matrix [26,27,28,29]. Studies of biochemical oscillations [25], including the cell cycle [30,31], circadian rhythms [32,33], and certain spatial patterns [34,35] are usually based on the dynamics of the deterministic RRE. Research on biochemical switches, as are found in prion diseases [36], developmental processes [37], and some protein kinase cascades [38], often focuses on the multiple steady-state solutions of the RRE [39]. Deterministic modeling has been, and still is, the conventional modeling method for most biological systems.

The Chemical Master Equation

Although the RRE is tremendously useful, it cannot address the stochastic processes that are inherent to biochemical systems. This is because the RRE arises from a series of approximations to a more physically rigorous stochastic model of chemical reactions [40,41].

As above, we assume that diffusive processes are much faster than reactive ones [8], which allows us to ignore spatial organization (this assumption is usually valid for genetic and metabolic networks, but often invalid for signaling networks). Nevertheless, the stochasticity of diffusion plays an essential role because it makes the precise timing of individual reactions effectively random. These reactions occur in abrupt transitions, in which reactants are converted effectively instantaneously into products, making this a type of *jump process*. Also, random diffusion causes the system to rapidly lose any memory of its prior states, and thus of the sequence of reactions that led up to the current state. This independence of the system dynamics on its history, called the Markov property, implies that the probability that a specific reaction occurs depends only on the state of the system at that time [41].

Because of the random reaction timing, reactant concentrations do not follow the deterministic trajectory that is predicted by the RRE. Instead, many concentration trajectories are possible, of which a single effectively random one actually occurs. There are two primary ways to investigate the possible trajectories with computational methods. One can simultaneously track the probability of every possible outcome or one can simulate many independent stochastic trajectories and then analyze them as one would with several repetitions of an experiment. These methods are described in this and subsequent sections, respectively.

To mathematically track the probability that the system is in each possible state, it is helpful to first replace the vector of chemical concentrations that was introduced earlier, $\mathbf{Z}(t)$, with a vector of integer-valued molecule numbers, $\mathbf{X}(t)$. These are related to each other, within round-off error, through the volume of the system, which is given as Ω,

$$\mathbf{X}(t) \simeq \Omega \mathbf{Z}(t) \,. \tag{5}$$

The state of the stochastic system is fully captured by $\mathbf{X}(t)$. The probability that the system is in state $\mathbf{x}$ at time t, given that it started in state $\mathbf{x}_0$ at time t_0, is written as $P(\mathbf{x}, t \mid \mathbf{x}_0, t_0)$. This probability changes over time because chemical reactions can transfer the system either into this state from other ones, or out of this state and into others. These possible transitions are combined to yield the *chemical master equation* (CME) [6,42]:

$$\frac{\partial P(\mathbf{x}, t \mid \mathbf{x}_0, t_0)}{\partial t} = \sum_{j=1}^{M} \Big[a_j(\mathbf{x} - \underline{\boldsymbol{\nu}}_j)\, P(\mathbf{x} - \boldsymbol{\nu}_j, t \mid \mathbf{x}_0, t_0) - a_j(\mathbf{x})\, P(\mathbf{x}, t \mid \mathbf{x}_0, t_0) \Big] \,. \tag{6}$$

The sum is over the reaction channels that can occur in the system. The two terms within brackets give the rate at which the probability of being in state $\mathbf{x}$ increases or decreases over time because of reactions into or out of state $\mathbf{x}$, respectively. These are proportional to the reaction propensities for the respective reactions. They are also proportional to the probability that the system was in the starting state, because the system can only leave a state if it was there in the first place.

The reaction propensity $a_j(\mathbf{x})$, given here without a tilde because it is for molecule numbers rather than concentrations, is a probability density: $a_j(\mathbf{x})\mathrm{d}t$ is the probability that exactly one reaction of type j will occur in a system in state $\mathbf{x}$ within the next dt amount of time. This microscopic propensity function is as central to stochastic chemical kinetics as its macroscopic analog is to the reaction rate equation. However, the microscopic propensity rests on a solid microphysical basis, and has in fact been shown to have an exact solution for a well-stirred thermally-equilibrated gas-phase system [43]. For such a system, the propensity function is

$$a_j = h_j c_j \,, \tag{7}$$

where h_j is the number of distinct combinations of individual reactants for reaction j and c_j is the probability density for one of those reactions to occur. That is, $c_j(t)\mathrm{d}t$ is the probability that a randomly selected set of reactants for reaction j will collide and react in the next infinitesimal time interval $\mathrm{d}t$. For a variety of reaction mechanisms, the c_j values can be calculated quite accurately from only the physical properties of the system [43].

The primary approximation made for the CME is that the reactive system is well-mixed. This implies both that there is no spatial organization and that there are no significant correlations between successive reactions (the Markov property). Examples of correlated reactions include metabolite channeling, which is the transfer of a metabolite from one enzyme to the next before it has a chance to equilibrate into the cytoplasm [44], and geminate recombinations, which are multiple bindings between molecules that bind reversibly [45]. The well-mixed statement also encompasses the assumption that the system is isothermal, which is typically the case in biological systems.

Because the CME becomes computationally intractable with any but the simplest systems, some recent work has focused on efficient solution methods. An algorithm called the finite state projection method accomplishes this by projecting a matrix form of the CME onto a smaller space [46,47]. By choosing the size of the projection space, the accuracy can be adapted to any level of precision. Less formal methods for state-space reduction of the CME have been proposed as well [48]. Another method uses a sparse grid, which can work efficiently for up to 10 proteins [49]. Work has also gone into separations of the CME into fast and slow components, as described in the section on hybrid methods [42,50,51].

Applications of the Chemical Master Equation

The solution of the CME suffers from the so-called "curse of dimensionality" as the size of the state space, and hence the number of equations, increases exponentially with the number of chemical species involved. Except for very small and simple systems, it is extremely difficult to obtain solutions of the CME, either analytically or numerically. However, a few papers do report quite interesting results from direct simulations of the CME.

The master equation was used to investigate the dynamics of transiently denatured segments of double stranded DNA [52]. The authors derived the dynamics, formation rates, and lifetimes of these "bubbles", which can be compared to fluorescence correlation microscopy experiments of fluorescently tagged base pairs [53]. In another use of the CME, studies on molecular motors [54,55] demonstrate how the load-velocity curve, including rectified motion, arises from nucleotide triphosphate binding free energies. These works more fully investigate ideas on thermal ratchets that were presented previously [56]. A particularly intriguing study on the copy number control system for bacterial plasmids [57] showed that stochasticity in a regulatory portion of a system can actually decrease the stochasticity elsewhere in the system. This "stochastic-focusing" changes the behaviors of gradual-response systems towards those of threshold systems [58,59] in a manner that is analogous to the oscillation enhancement that stochastic resonance can create in oscillating systems [60]. These results contradict the widely held belief that an increase in stochasticity in one portion of a system will necessarily increase the stochasticity everywhere downstream of it. Studies of simple signal transduction motifs have shown how the predictions of the RRE can be qualitatively wrong compared to the CME treatment in that the stochastic systems might be bistable or oscillate when the deterministic system has one stable state [40,61]. Many of these studies that used the CME also used other theoretical techniques as well, which allow fruitful comparisons between the methods.

The Gillespie Algorithm

Because of the challenges in working with the CME, it is most often investigated using a Monte Carlo approach in which individual sample trajectories are simulated. These simulations can be exact or approximate. In this context, "exact" means that if the simulation were run many times, the distribution of simulated trajectories would agree exactly with that which would be predicted by an analytical solution, were it obtainable, of the chemical master equation. Exactness implies nothing about the validity of the CME or about the limitations of the computational accuracy (such as round-off errors and imperfect pseudo-random number generators), but only that no further approximations are made beyond those that are assumed by the CME.

In 1976, Gillespie introduced an exact algorithm for simulating the CME [6,62,63] which is called the *stochastic simulation algorithm* (SSA) in his papers, but is better known as the *Gillespie algorithm*. This algorithm cycles through three portions: (i) generate the time step until the next reaction, (ii) determine which reaction that will be, and (iii) execute the reaction by advancing the time and molecule counts to reflect it. The Gillespie algorithm was introduced with two varieties, called the *direct method* and the *first-reaction method*. In the former, the time step to

the next reaction, τ, and the reaction number, j, are chosen from the following probability distributions:

$$P(\tau) = a e^{-a\tau} , \tag{8a}$$

$$P(j) = \frac{a_j(\mathbf{X})}{a} . \tag{8b}$$

The variable a represents the summed reaction propensity,

$$a \equiv \sum_j a_j(\mathbf{X}) . \tag{8c}$$

In the first-reaction method, a time step, τ_j, is generated for each possible reaction channel. Again, these are exponentially distributed random numbers,

$$\tau_j = a_j(\mathbf{X})\, e^{-a_j(\mathbf{X})\tau} . \tag{9}$$

The smallest of these time steps is chosen as the next simulation time step, while its subscript dictates the reaction channel that is executed at that time. The direct method is usually preferred because it is a little easier to program and runs slightly faster with a simple implementation. However, the latter has been favored as a basis for improvements on computational efficiency.

The exactness of the Gillespie algorithm comes at the cost of its being computationally demanding. Even if one simulation can be performed reasonably quickly, such as in a few minutes, this can still be too slow to investigate the behaviors of hundreds of mutant cells or to explore different regions of parameter space. Thus, much effort has been devoted to improving the efficiency of the Gillespie algorithm, while still maintaining exactness. These methods are all based upon either the direct method or the first-reaction method, but are carefully designed, usually with priority queues or other indexing methods, so that internal variables are recalculated as infrequently as possible [6,64,65,66,67]. Of these, it appears that the *optimized direct method* is probably best for most practical problems [66]. Because the faster methods are significantly more difficult to program than the original ones, both sets of methods are still commonly used.

The computational intensity of the Gillespie algorithm, even with more efficient implementations, makes it difficult to perform sensitivity analyzes. In these analyzes, one investigates the extent to which the results depend upon input parameters, which can helpful for determining which parameters need additional experimental investigation or which are particularly important for system control. An algorithm for stochastic sensitivity analysis was recently developed and applied to biochemical reaction networks [68]. It involves the addition of just two steps to the basic loop of the Gillespie algorithm.

Another difficulty of the Gillespie algorithm, which also applies to simulations of the RRE and other algorithms presented below, is called *combinatorial explosion*. Suppose a scaffold protein has several sites with which it can bind other proteins, and suppose that each of those proteins can bind to none, one, or two phosphate groups (this is the situation for the Ste5 protein in the yeast pheromone response pathway [69]). There are clearly a tremendous number of possible binding states that the scaffold protein can be in, each of which has to be treated as a separate chemical species. Just listing all of these states is tedious, and simulating their reaction dynamics with the Gillespie algorithm is very slow. One solution is to not list every possibility when the simulation starts, but to create states when they are needed and to destroy them when they are no longer required [65]. Another option is to use an algorithm that is implemented in a program called StochSim [70,71]. Unlike the Gillespie algorithm, this one does not stochastically choose reactions to execute, but it instead chooses reactant pairs from the pool of existing molecules. A probabilistic scheme is used to determine if these reactants should be made to react with each other.

What if the system volume changes as a function of time? This might seem like an unusual concern, but it occurs during cell growth (and cell division) and it affects the reaction propensities. The necessary modifications to the Gillespie algorithm were recently derived, which are likely to be particularly useful for relatively slow processes, such as protein production from infrequently expressed genes [72].

Applications of the Gillespie Algorithm

Models based on the Gillespie algorithm have provided critical insights into the stochastic nature of gene expression [3,73,74]. In particular, fluctuations in the rates of gene transcription are amplified at the translation stage to yield highly erratic time patterns of protein production [75]. When multiple regulatory proteins act together, or compete with each other, this randomness is amplified further because of the random sequence of protein bursts [75]. These effects were shown to stochastically switch a model of phage-λ between the bistable lysis and lysogeny states [76], with results that are consistent with experimental ones. Stochastic gene expression is also used by many pathogenic organisms to randomly switch their surface features so they can evade host responses [77], may be used by the HIV virus to stochastically delay viral expression long enough for transformation of its activated T-cell host to a memory cell and thereby trap HIV

as a latent phage [16], can establish asymmetries that determine cell differentiation [78], and can cause circadian clocks to lose synchrony [21,79]. In fluctuating environments, stochastic gene expression can permit an isogenic bacterial population to grow faster than it would if all individuals were phenotypically homogeneous [77,80,81].

From combined modeling and experimental approaches, the dominant noise sources in the stochastic expression of a specific gene are: (i) the expression fluctuations of that particular gene, which is called the *intrinsic noise*, (ii) noise that is transmitted to it from upstream genes, and (iii) global noise that affects all genes. The latter two sources are often combined and called *extrinsic noise*, which is the total noise source that is extrinsic to that specific gene [12,74,82,83,84]. Noise arises at both the transcription and translation stages, for which the relative importance depends on the strength of the promoter and on whether prokaryotic or eukaryotic transcriptional machinery is used [14,15,75,85,86]. Direct measurements of gene expression have generally confirmed the predictions made by stochastic simulations [11,12,13,15,17,18,21,87].

Gene expression appears to be unavoidably stochastic, and this randomness is usually amplified at each stage, so how does biology function reliably amidst all the noise? This is a central topic of many papers on biological robustness [5,9,88,89,90,91], several of which use the Gillespie algorithm or other types of stochastic modeling. One answer is that many reaction network structures are inherently less susceptible to noise than others. These include ones in which the reaction rates do not depend on the number of mRNA transcripts [92], certain scale-free reaction networks [93], and networks that are designed to function near saturation [94] (analogous to binary logic). Secondly, there are several mechanisms for biological robustness to noise, including negative feedback, integral feedback [95], checkpoints, and redundancy [9]. Because gene expression noise is usually detrimental to biological function, it has been suggested that there is active selection for robustness mechanisms [96,97].

Approximate Stochastic Methods

Because every reaction is simulated individually in the Gillespie algorithm, it is unavoidably computationally demanding, even with the algorithmic methods that have been developed to speed it up. To address this, several approximate methods have been developed.

The most accurate of these approximate methods is called the *tau-leaping* method [6,98,99,100]. In contrast to the Gillespie algorithm, the τ-leaping method uses a simulation time step which is long enough that many individual reactions are likely to occur during the time interval. The reaction propensities ought to change slightly as each reaction occurs to reflect the new chemical populations, although this algorithm uses the assumption that changes within a single time step are negligible. This is the sole approximation made for the τ-leaping method. Each reaction is considered to be an independent event (a consequence of both the well-mixed hypothesis and the constant reaction propensities), so the number of reactions that occur during time step τ for reaction channel j is a Poisson-distributed random variable; it is denoted k_j and has a mean value equal to the reaction propensity $a_j(\mathbf{X}(t))$. The formula used by τ-leaping that updates the system state over one time step is

$$\mathbf{X}(t+\tau) = \mathbf{X}(t) + \sum_{j=1}^{M} k_j \boldsymbol{\nu}_j \,. \tag{10}$$

The algorithm alternates steps in which the state of the system is updated and those in which new reaction propensities are calculated.

As the time step is reduced to zero, the τ-leaping simulation method approaches that of the Gillespie algorithm, although with more computational overhead. In the other direction, increasing the time step makes the simulation becomes more and more approximate. It has been suggested that τ should be chosen by first predicting the change in molecular populations over time using deterministic methods; then, the time step is chosen so that no molecular species is likely to change its population during this time step by more than some pre-determined fraction of its total population [98]. A difficulty that can occur with τ-leaping (which is also an issue with ODE integration), is that it is possible for a molecular species to be assigned a negative population. Several methods have been proposed to avoid this problem, some of which are also able to improve the performance of the algorithm in other ways as well [101,102,103]. Despite several papers on the development of τ-leaping, this method has yet to be applied to novel biological problems.

Two additional approximations allow the application of many more theoretical methods. First, the vector that defines the state of the system, $\mathbf{X}(t)$, is allowed to take on real values as well as integer values. As one would expect, this is usually a reasonable assumption for large chemical populations and a poor assumption for low copy numbers. In particular, it can a very poor approximation in cases where there might become no copies of a chemical species at all; an approximate value of, say, 0.01 protein copies can lead a system to entirely different outcomes than one would find with exactly 0 copies. The second

approximation is to replace the Poisson-distributed random variables that were used in the τ-leaping algorithm with Gaussian-distributed random variables [104]. The effect of this change again decreases as the copy numbers of chemical species are increased. It also decreases as longer time steps are used because that leads to more individual chemical reactions per time step. These approximations allow the updating equation of the τ-leaping algorithm to be replaced by a stochastic differential equation called the *chemical Langevin equation* [105,106,107] (CLE),

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = \sum_{j=1}^{M} \nu_{ji}\, a_j(\mathbf{X}(t)) + \sum_{j=1}^{M} \Gamma_j(t)\, \nu_{ji} \sqrt{a_j(\mathbf{X}(t))}. \quad (11)$$

The first term is simply the reaction rate equation that is given above (Eq. (4)), but for molecule counts rather than concentrations. The second term adds Gaussian noise to the deterministic result, where $\Gamma_j(t)$ represents a temporally uncorrelated, statistically independent Gaussian white noise with mean 0 and variance 1. In other words, the integral of $\Gamma_j(t)$ is a one-dimensional continuous random walk. The chemical Langevin equation describes a continuous Markov process which is an approximation of the jump Markov process that underlies the chemical master equation. Simulations with the CLE, which are based on an equation that is quite similar to Eq. (11) [106], yield single stochastic trajectories of the system state, much like simulations with the Gillespie algorithm or the τ-leaping method.

Alternatively, instead of following a single system as it moves along one of its many possible stochastic trajectories, it is possible to focus on a single portion of the state space to see how likely it is that the system will be in this region of state space as a function of time. This latter picture is described by *chemical Fokker–Planck equation* [41,105,107] (CFPE),

$$\begin{aligned}\frac{\partial P(\mathbf{x}, t \mid \mathbf{x}_0, t_0)}{\partial t} = & \\ -\sum_{i=1}^{N} \frac{\partial}{\partial x_i} & \left[\left(\sum_{j=1}^{M} \nu_{ji} a_j(\mathbf{x}) \right) P(\mathbf{x}, t \mid \mathbf{x}_0, t_0) \right] \\ + \frac{1}{2} \sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2} & \left[\left(\sum_{j=1}^{M} \nu_{ji}^2 a_j(\mathbf{x}) \right) P(\mathbf{x}, t \mid \mathbf{x}_0, t_0) \right] \\ + \sum_{\substack{i,i'=1 \\ i<i'}}^{N} \frac{\partial^2}{\partial x_i \partial x_{i'}} & \left[\left(\sum_{j=1}^{M} \nu_{ji} \nu_{ji'} a_j(\mathbf{x}) \right) P(\mathbf{x}, t \mid \mathbf{x}_0, t_0) \right]. \end{aligned} \quad (12)$$

The first term, often called the drift term, arises from the deterministic behavior of the system. The latter two terms, collectively called the diffusion term, represent the stochastic deviations away from deterministic behavior. Because the CFPE represents continuous processes, it is significantly more analytically tractable than the chemical master equation.

To compute the probabilities of possible system behaviors using the CFPE, state space is usually discretized into a grid and then the CFPE is integrated using standard numerical methods [108,109]. This analysis method is similar to that employed for the CME, but is usually less computationally demanding because the discretized state space is typically significantly coarser. Nevertheless, the dimensionality of state space still increases exponentially with additional chemical species, so the CFPE still suffers from the curse of dimensionality.

Applications of Approximate Stochastic Methods

Perhaps because stochastic simulations are still a relatively new field of study, many studies with the CLE and CFPE focus more on the mathematical techniques than on the biological applications [109,110,111,112,113]. The approximate CLE and CFPE have been shown to yield results that are in good agreement with exact simulations for a reversible isomerization reaction, even with very few molecules [107].

The CLE was used to analytically investigate the role of noise-induced phenomena in enzymatic futile cycles, which is a motif that is common to many biochemical networks [61]. The analysis indicated that the presence of external noise is sufficient to induce switching bistability in the system, a phenomenon that is often attributed to feedback loops [25]. In combination with experimental data, the CLE was also used to show that translational efficiency is the predominant source of intracellular noise for a single-gene system [15]. The Fokker–Planck equation has been used to model cell growth [111,112] and cell migration [113,114]. Of particular interest, the Fokker–Planck equation has provided a convenient framework to describe the behaviors of molecular motors [109]. A motor protein is approximated as a diffusion particle in a periodic asymmetric free-energy surface. Under the input of chemical energy, the motor switches stochastically between different potentials that describe distinct biochemical states of the motor. The model has been used to explain key experimental observations for molecular motors, most notably for the F_1F_0-ATPase system [115] and a bacterial flagellar motor [109,116].

Hybrid Algorithms

Systems that involve multiple time scales provide major simulation challenges. If the fast time scale is simulated with high precision, then the simulation takes too long for the dynamics of the slow one to be observed with any reasonable efficiency. On the other hand, if the simulation time steps are optimized for the slow timescale, then they are too long for the fast reactions and numerical errors become problematic. In the language of differential equations, these are stiff systems which require special solution techniques. For stochastic simulations with multiple timescales, several methods have been developed recently.

One class of hybrid methods focuses on new mathematics to allow approximations of the Gillespie algorithm, or related algorithms, to function with reasonable accuracy over a wide range of timescales [42,50,51,117,118,119,120,121]. The other class generally involves the coupling of multiple simulators, usually including ODE, Langevin, and/or Gillespie; the high-population molecular species are simulated with less stochastic detail and the low-population species are simulated with more stochastic detail [122,123].

Spatial Stochastic Modeling

Most biological systems are highly organized. For example, *Escherichia coli* bacteria have helical cytoskeletons, polar-localized proteins, centrally positioned chromosomes, and elaborate flagellar motor complexes. Eukaryotes are even more organized, with elaborate organelles, microtubules and other complex cytoskeletal elements, motor proteins that shuttle back and forth, and carefully controlled cell shapes. Even phages display remarkable order in the way the DNA is packed into the outer shell. Where does this order come from? And how does this order influence the biochemical reaction network? These questions are being investigated with new imaging experiments [124,212] and with new computer simulation methods that can account for spatial heterogeneity. These spatial simulation methods are the focus of this section.

Spatial simulations have been used to investigate a wide variety of topics. These include: morphogen gradients across *Drosophila* and *Xenopus* oocytes [125,126,127], the *Escherichia coli* cell division plane localization system [128,129,130,131,132,133,134] (see Sect. "Box 1: The *E. Coli* Min System"), intracellular signaling [135,136,137,138], and rebinding of ligands to receptor complexes [139,140,141].

As described above, many successful biochemical models do not account for spatial heterogeneity; in fact, non-spatial models are in the vast majority. Typically, non-spatial models get away with ignoring space because they model dynamics that occur more slowly than the time it takes for a molecule to diffuse across a cell, because they investigate processes that are not intrinsically spatial, and because they do not demand high quantitative accuracy. As the tools are becoming available, including both fast computers and new software algorithms [45,142,143,144], the interest in including spatial detail is increasing. These spatial models can be either deterministic or stochastic, of which our primary focus is on the latter ones.

As with the non-spatial methods that are described above, stochastic effects in spatial models arise from the discreteness of molecules. This leads to fluctuations in the numbers of molecules, which are typically on the order of the square root of the number of molecules in the appropriate characteristic volume (near steady-state and equilibrium points, but frequently greater near critical points). In spatial models, the characteristic length scale is no longer the size of the entire system but is dictated by the length scale of the spatial heterogeneity. With the shorter length scale, the characteristic volume size is reduced, fewer molecules are in these volumes, and stochastic effects increase. Thus, stochastic simulations can be required for spatial models, even if they were not needed for the corresponding non-spatial model. There are also other good reasons to model stochastic effects in spatial simulations. Many spatial phenomena, such as noise that arises from ligand rebinding [141], cannot be adequately treated without considering the detailed molecular interactions. Finally, a model is only as good as its weakest aspect. If one increases the accuracy of a model in one way, such as by accurately treating either space or stochastics, then the benefits may not be realized until the other aspect is addressed as well.

Schemes for investigating a chemical system with spatial and stochastic detail can be classified by whether they consider molecules within populations or as individuals. In the former case, space is divided it into small subvolumes, whereas in the latter, space is continuous. These classes are described in detail below. Another approach is lattice-based methods [145,146,147,148,149]. However, we do not discuss them here because they are rarely used for quantitative modeling. Furthermore, the underlying lattice geometry usually affects the results, thus making them less realistic.

Population-Based Spatial Models

In a top-down approach towards spatial modeling, one starts with a simple, deterministic, macroscopic description and then adds successive layers of detail. In this case,

the natural starting point is with the standard textbook descriptions of chemical reactions and diffusion [24]. Reactions are described with mass action reaction kinetics expressed with the reaction rate equation that was discussed above (Eq. (4)). Diffusion is described with the diffusion equation [24], also called Fick's second law of diffusion, which is

$$\frac{\partial Z_i(\mathbf{r}, t)}{\partial t} = D_i \nabla^2 Z_i(\mathbf{r}, t) \,. \tag{13}$$

In an extension of the definition given before, $Z_i(\mathbf{r}, t)$ is the concentration of component i at the 3-dimensional position $\mathbf{r}$ and time t. D_i is the diffusion coefficient for component i.

Because reactions and diffusion occur simultaneously, the respective equations are combined to express the simultaneous effects of both processes to yield the *reaction-diffusion equation*,

$$\frac{\partial Z_i(\mathbf{r}, t)}{\partial t} = \sum_{j=1}^{M} \nu_{ji} \tilde{a}_j(\mathbf{Z}(\mathbf{r}, t)) + D_i \nabla^2 Z_i(\mathbf{r}, t) \,. \tag{14}$$

This partial differential equation (PDE) underlies a great deal of theory on chemical and biological pattern formation [126,136,150,151]. The Virtual Cell computer program [152] is general-purpose software that simulates the reaction-diffusion equation. It has been used primarily to explore spatial effects in intracellular signaling [153,154,155].

The reaction-diffusion equation is deterministic, so it captures neither the discreteness of reaction events nor the Brownian motion processes that underlie diffusion. It is possible to add this stochasticity directly into the deterministic theory but that would create a set of coupled stochastic scalar field equations, which would be extraordinarily complicated. Neither the deterministic nor the stochastic PDEs are tractable to work with analytically for any but the very simplest systems except, perhaps, in steady-state. Thus, most analysis is either computational or approximate.

In most such analyses, the equations are first simplified by dividing the system volume into an array of small cubic subvolumes, each with width l. This spatial discretization changes the diffusion portion of the reaction-diffusion equation into a discrete form:

$$\frac{\mathrm{d}Z_{i,k}(t)}{\mathrm{d}t} = \sum_{j=1}^{M} \nu_{ji}\, \tilde{a}_j(\mathbf{Z}_k(t)) + \frac{D_i}{l^2} \sum_{k'} \left[Z_{i,k'}(t) - Z_{i,k}(t) \right] \,. \tag{15}$$

The index k denotes the subvolume number, much as $\mathbf{r}$ represented the spatial location. The latter summation in this discrete reaction-diffusion equation extends over all nearest neighbors of subvolume k, denoted by k'. Because the description of space was changed from continuous states to discrete states, much like the discrete kinds of molecules that are labeled by the index i, diffusion is now formally identical to reactions. The "reaction rate constant" for diffusion [156] between one subvolume and its neighbor is D_i/l^2. Because of this mathematical equivalence, much of the following discussion on the stochastic simulation of the reaction-diffusion equation parallels the discussion presented earlier on non-spatial stochastic simulations.

The first spatial stochastic equation that we present is the one that accounts for the least detail. It is the *spatial chemical Langevin equation*, which results from adding white Gaussian noise to the discrete reaction-diffusion equation. It is

$$\begin{aligned} \frac{\mathrm{d}Z_{i,k}(t)}{\mathrm{d}t} = & \sum_{j=1}^{M} \nu_{ji} \left[\tilde{a}_j(\mathbf{Z}_k(t)) + \Gamma_j(t) \sqrt{\tilde{a}_j(\mathbf{Z}_k(t))} \right] \\ & + \sum_{k'} \left\{ \frac{D_i}{l^2} \left[Z_{i,k'}(t) - Z_{i,k}(t) \right] \right. \\ & \left. + \Gamma_{k'}(t) \sqrt{\frac{D_i}{l^2}} \left[\sqrt{Z_{i,k'}(t)} - \sqrt{Z_{i,k}(t)} \right] \right\} \,. \end{aligned} \tag{16}$$

In an extension to what was presented before, $\Gamma_j(t)$ and $\Gamma_{k'}(t)$ represent temporally uncorrelated, statistically independent Gaussian white noises [106]. This is a specific example of the more general multivariate Langevin equation; it, and the multivariate Fokker–Planck equation, have been explored in depth [41,105]. However, the more specific spatial chemical Langevin equation has essentially never been used, investigated mathematically, or simulated. The sole exception that we are aware of was its simulation for a figure for a tutorial article [8] (those results are reproduced in Fig. 1).

The spatial chemical Langevin equation captures stochasticity reasonably accurately for systems in which there are many molecules per subvolume but not for those with few molecules per subvolume. Errors arise both because Gaussian white noise is the incorrect fluctuation distribution [100,104] and because it treats molecule amounts as continuously variable quantities. These are addressed by moving to the next level of detail in which the continuous molecular concentrations are replaced by discrete numbers of molecules. This changes the temporally

continuous reaction and diffusion processes to stochastic jump processes. The *reaction-diffusion master equation* [156,157,158,159] (RDME) describes the time dependence of the system at this level of description. It is

$$\frac{\mathrm{d}P(\mathbf{x},t)}{\mathrm{d}t} = \sum_{j=1}^{M}\left[a_j(\mathbf{x}-\underline{\boldsymbol{\nu}}_j)P(\mathbf{x}-\boldsymbol{\nu}_j,t) - a_j(\mathbf{x})P(\mathbf{x},t)\right] + \sum_{i=1}^{N}\frac{D_i}{l^2}\sum_{k,k'}\Big[(X_{i,k}+1)\,P(\ldots,X_{i,k}+1, X_{i,k'}-1,\ldots,t) - X_{i,k}P(\mathbf{x},t)\Big]. \tag{17}$$

$P(\mathbf{x}, t)$ is the probability that the system is in state $\mathbf{X} = \mathbf{x}$ at time t, $X_{i,k}$ is the number of molecules of type i in subvolume k, $\mathbf{X}$ is the vector of all $X_{i,k}$ values, and a_j is the propensity of reaction j.

The RDME expresses as much detail as is possible through these successive improvements of the reaction-diffusion equation. It is tempting to think of it as the fundamental equation for reactions and diffusion, and thus the basis for a statistical theory of chemistry. In fact, it is sufficiently accurate for most systems, but it nevertheless involves approximations that can be important in some situations. Firstly, neither of the starting equations, which are mass action kinetics and Fickian diffusion, are completely accurate even for very large systems. Mass action kinetics does not address the increased reaction rates that occur on extremely short time scales, which arise from reduced spatial correlations [160,161]. Nor does it address the geminate recombinations that can occur between the products of a dissociation reaction [162,163]. The diffusion equation is usually quite accurate for dilute solutions but fails for highly crowded ones [164,165], including most biological systems [166]. Secondly, the discretization of space into small subvolumes can also lead to inaccuracies, or exacerbate the inaccuracies just mentioned. The subvolume sizes must not be so small that they impinge on the microscopic details of the reaction or diffusion processes. This means that they need to be significantly larger than single molecules and larger than the mean free path lengths of diffusion [156,167]. Conversely, the subvolumes must not be so large that there would be appreciable concentration gradients across them. This means that the subvolume width needs to be less than the reactant correlation length. The correlation length is hard to predict but is at least as large as the average distance that a reactant travels before it reacts, called the reactive mean free path [156,167].

The RDME is even more intractable than the nonspatial chemical master equation because of the addition of spatial states and the many transitions that can occur between the spatial states. These additional states and transitions also make stochastic simulations of the RDME with Gillespie's direct algorithm extremely slow [157]. Several faster algorithms have been developed to address this problem. The "next reaction method" of Gibson and Bruck [64] was adapted to spatial simulations [168], and then further improved, to yield the "next subvolume method" [132,142]. Also, a fast version of the direct method [169] has been developed for spatial simulations [167]. All of these methods yield exactly the same results as Gillespie's original methods [62,63] but use carefully optimized data structures to minimize the number of computations.

A separate challenge with simulating the RDME concerns the cubical subvolumes into which space was discretized. Biological systems rarely have square corners, so the basic theory requires adaptation to account for realistic boundaries. In one approach, the mathematics was developed for dividing boundary subvolumes into two separate portions [170]. Using another approach, the theory was developed for curved surfaces, which was implemented in the MesoRD program [132,171]. Although it has not been developed yet, it has been proposed that automatic mesh refinement could simultaneously account for complex boundaries and lead to significant computational efficiencies [167].

Along with simulations of the *E. coli* Min system, presented in Sect. "Box 1: The E. Coli Min System", population-based spatial stochastic models have been used for a variety of test systems. In the first implementation of a spatial Gillespie algorithm, Stundzia and Lumsden used a one-dimensional simulation to demonstrate stochastic calcium wave propagation [157]. Elf and Ehrenberg showed that spatial and stochastic effects can cause an intrinsically bistable system to lose its global hysteresis through the formation of spatial domains [142]. In a third study, Isaacson and Peskin demonstrated their method for simulating porous boundaries with a model that includes transcription, translation, and nuclear membrane transport [170].

Individual-Based Spatial Models

In a bottom-up approach to spatial modeling, one starts with a very detailed consideration and then makes successive approximations. A convenient place to start is by considering every individual molecule in the system, along with some of the molecular structures. The motions of

these molecules are governed by physical forces including steric repulsion, bond mechanics, and electrostatics. The simulation of the motions that result from these forces is called molecular dynamics [172]. Molecular dynamics can yield very accurate results but is so computationally intensive that it is rarely used for more than hundreds of cubic nanometers of volume or more than tens of nanoseconds of time. These size and time scales are too confining for studying biochemical reaction networks, so approximations are made.

At the Smoluchowski level of detail, all solvent molecules are ignored, solute molecules are treated as spheres, diffusion proceeds stochastically, and molecular rotation, molecular momentum, and long-range intermolecular forces are all ignored. This is a vast simplification, but is often valid. It is usually reasonably accurate for size scales that are larger than a few nanometers and for timescales that are longer than a few nanoseconds, constraints that are acceptable for an enormous range of chemical and biological phenomena.

For diffusion at the Smoluchowski level of detail, the effects of solvent-solute interactions on the solute motion are approximated by assuming that solute molecules diffuse with mathematically ideal Brownian motion [173,174]. This is a key approximation that replaces the deterministic molecular motions that result from solvent collisions with stochastic trajectories. It is often the only source of stochasticity in the theory, or in simulations that derive from this individual-based approach. More precisely, the position of molecule i at time t is given with the probability density $p_i(\mathbf{r}, t)$, which evolves over time according to the master equation

$$\frac{\partial p_i(\mathbf{r}, t)}{\partial t} = D_i \nabla^2 p_i(\mathbf{r}, t) \,. \tag{18}$$

This equation is nearly identical to the diffusion equation, given above (Eq. (13)), differing only in the definitions of the variables and the interpretation. Now, it is not a population of molecules that diffuse, but the positional probability density for a single molecule.

Because it is so simple, the diffusion master equation is analytically tractable, in contrast to the other master equations that were discussed. One result is an entire body of analytical theory on diffusion-influenced reactions [160,175]. Nevertheless, it too becomes unmanageable for systems that have several interacting molecules, so it is simulated with a technique called *Brownian dynamics* [176,177,178,179,180]. In this method, molecules have well-defined point-like positions which are updated at each simulation time step using random displacements. The displacements are chosen by solving the diffusion master equation for molecule i, which is taken to be at the well-defined position $\mathbf{r}_0$ at time t_0. One simulation time step later, at time $t_0 + \Delta t$, the probability density for the molecule's position is found to be a 3-dimensional Gaussian density that is centered at $\mathbf{r}_0$,

$$p_i(r, \Delta t) = \frac{1}{s_i^3 (2\pi)^{3/2}} \exp\left[-\frac{(\mathbf{r} - \mathbf{r}_0)^2}{2 s_i^2} \right] . \tag{19a}$$

The standard deviation of this Gaussian, called the root mean square step length, is

$$s_i = \sqrt{2 D_i \Delta t} \,. \tag{19b}$$

Brownian dynamics simulations provide accuracy that is below that of molecular dynamics, but still captures single molecule behavior.

Brownian dynamics has been used extensively for examining the rates of diffusion-influenced chemical reactions in solution [139,177,179,180,181,182] and for the rates of binding between ligands and receptor arrays [139,141,178,183,184]. In these studies, simulated molecules diffuse in solution; at the moment that a reactant pair, or a ligand and its cognate receptor, come into contact, they undergo a chemical reaction. While diffusing, intermolecular forces are often ignored, although some studies account for these interactions as well [185,186].

To achieve the necessary level of detail, Brownian dynamics simulations usually use very short simulation time steps, often on the order of picoseconds [139]. Adaptive time steps, such that time steps are long when reactants are widely separated and short when they are close, can speed simulations up by several orders of magnitude, but are easy to implement only if there is just one diffusing particle present in the simulation volume [141]. A more sophisticated method that has the same general goal of computational efficiency is called Green's function reaction dynamics [143,187,188] (GFRD). In GFRD, which works with any number of molecules, the system is inspected to see how soon the next molecular collision or reaction could occur. The system is then advanced to that time using a single simulation time step, the event is executed if appropriate, and the cycle repeats. Yet another method, used in a program written by one of us (SSA) called Smoldyn, achieves computational efficiency by modifying the effective radii of simulated molecules so that the same reaction rate is achieved with long simulation time steps as with short ones [45,189,190]. This method does not achieve the same spatial or temporal precision as classical Brownian dynamics or GFRD, but the level of detail is still more than adequate for most biological applications and has been

shown to be indistinguishable from more accurate simulations in many cases [191].

Technically, all of these algorithms execute Brownian dynamics. However, the term "Brownian dynamics" is typically used to describe highly detailed studies in which reaction rates, rebinding dynamics, or similar phenomena are found from fundamental molecular properties such as molecular radii and intermolecular forces. In contrast, GFRD and the methods used in Smoldyn are more often used to determine system-level behaviors from known or estimated reaction rates. These are more often called particle-based stochastic simulation methods [192].

MCell is another program that performs particle-based stochastic simulations [144]. Unlike the others, it cannot simulate reactions that occur in free solution, but instead only treats reactions at surfaces. Despite the decrease of versatility, it is still useful for studying a wide variety of biological phenomena [193,194,195]; in particular, it was developed to investigate the neuromuscular junction [196,197,198]. In MCell, surface-bound receptors are not modeled as single molecules as they would be in Smoldyn or GFRD methods, but as a uniform binding probability that applies to an entire surface tile. This decreases the spatial resolution some, but increases the computational efficiency.

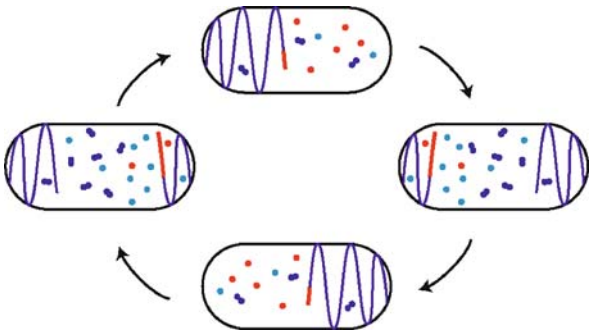
Future Directions

The classic advice of using the right tool for the job is as true in biochemical modeling as it is elsewhere. Several modeling tools have been presented here. Deterministic ordinary differential equation models are simple, easy to use, and can be analyzed with many powerful theoretical and analytical methods. They are the right tool for systems that can be treated as being well-mixed and that are both large enough and sufficiently far from critical points that stochastic effects are unimportant. In contrast, systems that include low copy numbers of important components, and/or that can be triggered by random events, require stochastic modeling methods for their investigation. These include integration of the chemical master equation and random sampling of the stochastic trajectories using the Gillespie algorithm, both of which are exact methods. Approximate methods include τ-leaping stochastic simulations, integration of the chemical Fokker–Planck equation, and sampling with the chemical Langevin equation. Of these, the Gillespie algorithm has proven to be the most popular. Finally, if the system cannot be considered to be well-mixed, then yet different tools are needed. These include spatial variants of the same list of simulation methods, including partial differential equations for deterministic simulations and a spatial Gillespie algorithm for stochastic simulations. Particle-tracking simulation methods allow an even greater degree of detail.

In general, more detailed simulation methods yield more accurate results and are based more closely on underlying processes and less on phenomenological descriptions. However, they are also more computationally intensive and require more model parameters. This parametrization poses a significant problem for current models because the necessary quantitative experimental data are typically only marginally adequate or are completely non-existent. For an ODE model, it is sometimes possible to address this problem by exploring model behaviors over wide ranges of parameter space, from which one can draw phase diagrams that graphically depict how the model behaves for different parameter choices. From this, one can sometimes constrain parameters or gain additional insight into the model; for example, Tyson showed how two enzyme concentrations can be used to regulate the cell cycle, bringing an oocyte from metaphase arrest to autonomous oscillations, and on to growth-controlled cell division [30]. Because of the computational demands of spatial and stochastic models, as well as the richer behavior possibilities, it is much more difficult to explore parameter space with these more complex models. Thus, much work is needed on this topic.

More generally, the mathematical infrastructure for designing and interpreting stochastic models lags far behind that for non-spatial deterministic models. This poses some challenges for theorists. For example, what new theories and graphical tools will help scientists gain intuition into the dynamics of stochastic systems? and what are the controlling elements of stochastic systems? The theory is even more unexplored when spatial organization is considered as well. Nevertheless, spatial considerations are essential because no biological life has been found that is well-mixed; instead, a tremendous amount of biochemical activity involves membranes, polymers, protein scaffolds, large multimeric complexes, and other spatial structures. Theories that address these topics will not be as elegant as those that focus on the chemical master equation, but biology is not always elegant either.

Although research on stochastic modeling of biochemistry grew slowly from the 1950s to the 1990s, the pace has accelerated dramatically during the last 10 to 15 years. This acceleration will likely continue for many more years, in response to the faster computers that become available every year and to the ever-increasing complexity of biochemical data. With this growth, stochastic modeling may open up entire new ways to understand cell biology.



Stochastic Models of Biological Processes, Figure 2
Diagram of the *E. coli* Min system, which is used to position the cell division plane at the cell center. *Dots* represent cytoplasmic proteins, while *curved lines* represent helical membrane-bound protein polymers. *Colors* identify the proteins: *light blue* for MinD bound to ADP, *dark blue* for MinD protein bound to ATP, and *red* for MinE. The system dynamics are summarized in the text of Sect. "Box 1: The E. Coli Min System"

Box 1: The *E. Coli* Min System

The *E. coli* Min system has served as a proving ground for spatial stochastic simulation methods. The Min system is used by *E. coli*, in conjunction with other systems, to position the cell division plane accurately at the cell center [131]. The system is comprised of the proteins MinC, MinD and MinE, which oscillate back and forth across the cell, from one pole to the other, with a period of about 40s (Fig. 2). Of these, only MinD and MinE are required for the oscillation, making this a relatively simple system that exhibits remarkably interesting dynamics. Cytoplasmic MinD proteins bind ATP, dimerize, and polymerize on the inside of the cell membrane to form long helical structures that extend outwards from one of the two cell poles. When MinE binds to the cell-center end of a MinD polymer, it activates ATP hydrolysis which depolymerizes the terminal subunit. As MinE progressively disassembles a MinD polymer at one end of the cell, it reassembles again from the opposite pole to start the next oscillation cycle. The oscillating Min proteins continually inhibit cell division plane formation near the poles using MinC, which colocalizes with MinD, thus only permitting cell division at the cell center.

This system was explored for several years with deterministic reaction-diffusion models [128,199,200]. One of these models, by Howard, Rutenberg, and de Vet [199], was also explored by the same group using a one-dimensional population-based stochastic method [129] (it uses discrete particle numbers and fixed time steps, thus conceptually placing it between the spatial Langevin and spatial Gillespie methods). The authors found that stochastic effects were essential for generating oscillations in some parameter regimes, in a spatial version of stochastic resonance [60,201]. These models helped direct new experiments [202,203,204,205] that clarified the processes of the system.

Building on the prior models and the new experimental data, Huang, Meir, and Wingreen [130] developed a new reaction-diffusion model that was more closely connected with the biology than were previous models and that accounted for several mutant phenotypes. This model became the basis of several stochastic simulations. The spatial Gillespie method was employed by Fange and Elf [132] using their MesoRD program. They showed that a stochastic model can account for a "spotty" phenotype and for oscillations in spherical mutant cells, neither of which can be explained by the deterministic model. The MCell particle-tracking program was used by Kerr and coworkers [133] to show that the Min system alone is insufficient to center the cell division plane with high accuracy.

Yet unexplained with these simulations were convincing experimental results that MinD forms polymers on the cell membrane [205,206,207]. These were explored with another particle-tracking model [208], using a method based on Smoluchowski dynamics [45]. Although this group did simulate spontaneous polymer formation, they observed many randomly oriented short filaments, in contrast to the few helical polymers that are observed experimentally. This inherent difficulty with the reaction-diffusion model [209], whether deterministic or stochastic, has led to several studies that have focused specifically on the polymer dynamics and shapes [134,210,211].

The *E. coli* Min system is already well on its way to becoming the prototypical system for studying spatial biochemical dynamics, much as *E. coli* chemotaxis has become the prototypical system for investigating bacterial signaling.

Acknowledgments

This work was funded by the US Department of Energy.

Bibliography

1. McQuarrie DA (1967) Stochastic approach to chemical kinetics. J Appl Probab 4:413–478
2. Turner TE, Schnell S, Burrage K (2004) Stochastic approaches for modelling in vivo reactions. Comp Biol Chem 28:165–178
3. Raser JM, O'Shea EK (2005) Noise in gene expression: Origins, consequences, and control. Science 309:2010–2013
4. Samoilov MS, Price G, Arkin AP (2006) From fluctuations to phenotypes: The physiology of noise. Sci STKE 2006:re17

5. Rao CV, Wolf DM, Arkin AP (2002) Control, exploitation, and tolerance of intracellular noise. Nature 420:231–237
6. Gillespie DT (2007) Stochastic simulation of chemical kinetics. Ann Rev Phys Chem 58:35–55
7. Wolf DM, Arkin AP (2003) Motifs, modules and games in bacteria. Curr Opin Microbiol 6:125–134
8. Andrews SS, Arkin AP (2006) Simulating cell biology. Curr Biol 16:R523–R527
9. McAdams HH, Arkin A (1999) It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet 15:65–69
10. Singer RH, Lawrence DS, Ovryn B, Condeelis J (2005) Imaging of gene expression in living cells and tissues. Biomed J Optics 10:051406
11. Levsky JM, Shenoy SM, Pezo RC, Singer RH (2002) Single-cell gene expression profiling. Science 297:836–840
12. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297:1183–1186
13. Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. Science 304:1811–1814
14. Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. Nature 422:633–637
15. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. Nature Genet 31:69–73
16. Weinberger LS, Burnett JC, Toettcher JE, Arkin AP, Schaffer DV (2005) Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell 122:169–182
17. Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. Nature 440:358–362
18. Yu J, Xiao J, Ren X, Lao K, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. Science 311:1600–1603
19. Golding I, Cox EC (2006) Protein synthesis molecule by molecule. Genome Biol 7:212
20. Fusco D, Accornero N, Lavoie B, Shenoy SM, Blanchard J-M, Singer RH, Bertrand E (2003) Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. Curr Biol 13:161–167
21. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403:335–338
22. Sakurai JJ (1994) Modern Quantum Mechanics. Addison-Wesley, Boston
23. Strogatz SH (1994) Nonlinear Dynamics and Chaos. Westview Press, Cambridge
24. Atkins PW (1986) Physical Chemistry. Freeman, New York
25. Tyson JJ, Chen KC, Novak B (2003) Sniffers, buzzers, toggles, and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr Opin Cell Biol 15:221–231
26. Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, Palsson BO (2001) Metabolic modeling of microbial strains in silico. Trends Biochem Sci 26:179–186
27. Varma A, Palsson BO (1994) Metabolic flux balancing: Basic concepts, scientific and practical use. Nature Biotech 12:994–998
28. Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. Curr Opin Biotech 14:491–496
29. Fell D (1997) Understanding the Control of Metabolism. Portland Press, London
30. Tyson JJ (1991) Modeling the cell division cycle: cdc2 and cyclin interactions. Proc Natl Acad Sci USA 88:7328–7332
31. Chen KC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B, Tyson JJ (2004) Integrative analysis of cell cycle control in budding yeast. Mol Biol Cell 15:3841–3862
32. Goldbeter A (2002) Computational approaches to cellular rhythms. Nature 420:238–245
33. van Zon JS, Lubensky DK, Altena PRH, ten Wolde PR (2007) An allosteric model of circadian KaiC phosphorylation. Proc Natl Acad Sci USA 104:7420–7425
34. Reinitz J, Mjolsness E, Sharp DH (1995) Model for cooperative control of positional information in Drosophila by bicoid and maternal hunchback. Exp J Zool 271:47–56
35. von Dassow G, Meir E, Munro EM, Odell GM (2000) The segment polarity network is a robust developmental module. Nature 406:188–192
36. Kellershohn N, Laurent M (2001) Prion diseases: dynamics of the infection and properties of the bistable transition. Biophys J 81:2517–2529
37. Ferrell JEJ, Machleder EM (1998) The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. Science 280:895–898
38. Huang C-YF, Ferrell JEJ (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc Natl Acad Sci USA 93:10078–10083
39. Laurent M, Kellershohn N (1999) Multistability: a major means of differentiation and evolution in biochemical systems. Trends Biochem Sci 24:418–422
40. Samoilov MS, Arkin AP (2006) Deviant effects in molecular reaction pathways. Nature Biotech 24:1235–1240
41. van Kampen NG (1992) Stochastic Processes in Physics and Chemistry. Elsevier, Amsterdam
42. Haseltine EL, Rawlings JB (2005) On the origins of approximations for stochastic chemical kinetics. Chem J Phys 123:164115
43. Gillespie DT (1992) A rigorous derivation of the chemical master equation. Physica A 188:404–425
44. Rohwer JM, Postma PW, Kholodenko BN, Westerhoff HV (1998) Implications of macromolecular crowding for signal transduction and metabolite channeling. Proc Natl Acad Sci USA 95:10547–10552
45. Andrews SS, Bray D (2004) Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. Phys Biol 1:137–151
46. Munsky B, Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. Chem J Phys 124:044104
47. Peles S, Munsky B, Khammash M (2006) Reduction and solution of the chemical master equation using time scale separation and finite state projection. Chem J Phys 125:204104
48. Kuwahara H, Myers CJ, Samoilov MS, Barker NA, Arkin AP (2006) Automated abstraction methodology for genetic regulatory networks. Trans Comput Syst Biol 6:150–175
49. Hegland M, Burden C, Santoso L, MacNamara S, Booth H (2007) A solver for the stochastic master equation applied to gene regulatory networks. Comp J Appl Math 205:708–724
50. Nedea SV, Jansen APJ, Lukkien JJ, Hilbers PAJ (2003) Infinitely fast diffusion in single-file systems. Phys Rev E 67:046707
51. Chatterjee A, Vlachos DG (2006) Multiscale spatial Monte Carlo simulations: Multigriding, computational singular per-

turbation, and hierarchical stochastic closures. Chem J Phys 124:064110
52. Ambjörnsson T, Banik SK, Lomholt MA, Metzler R (2007) Master equation approach to DNA breathing in heteropolymer DNA. Phys Rev E 75:021908
53. Altan-Bonnet G, Libchaber A, Krichevsky O (2003) Bubble dynamics in double-stranded DNA. Phys Rev Lett 90:138101
54. Lattanzi G, Maritan A (2001) Master equation approach to molecular motors. Phys Rev E 64:061905
55. Wang H-Y, Elston T, Mogilner A, Oster G (1998) Force generation in RNA polymerase. Biophys J 74:1186–1202
56. Peskin CS, Odell GM, Oster GF (1993) Cellular motions and thermal fluctuations: the Brownian ratchet. Biophys J 65:316–324
57. Paulsson J, Ehrenberg M (2000) Random signal fluctuations can reduce random fluctuations in regulated component of chemical regulatory networks. Phys Rev Lett 84:5447–5450
58. Paulsson J, Berg OG, Ehrenberg M (2000) Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. Proc Natl Acad Sci USA 97:7148–7153
59. Berg OG, Paulsson J, Ehrenberg M (2000) Fluctuations in repressor control: thermodynamic constraints on stochastic focusing. Biophys J 79:2944–2953
60. Li H, Hou Z, Xin H (2005) Internal noise stochastic resonance for intracellular calcium oscillations in a cell system. Phys Rev E 71:061916
61. Samoilov M, Plyasunov S, Arkin AP (2005) Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. Proc Natl Acad Sci USA 102:2310–2315
62. Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Comp J Phys 22:435–450
63. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. Phys J Chem 81:2340–2361
64. Gibson MA, Bruck J (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. Phys J Chem A 104:1876–1889
65. Lok L, Brent R (2005) Automatic generation of cellular reaction networks with Molecularizer 1.0. Nature Biotech 23:131–136
66. Cao Y, Li H, Petzold L (2004) Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. Chem J Phys 121:4059–4067
67. McCollum JM, Peterson GD, Cox CD, Simpson ML, Samatova NF (2006) The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior. Comp Biol Chem 30:39–49
68. Plyasunov S, Arkin AP (2007) Efficient stochastic sensitivity analysis of discrete event systems. Comput J Phys 221:724–738
69. Bardwell L (2004) A walk-through of the yeast mating pheromone response pathway. Peptides 25:1465–1476
70. Morton-Firth CJ, Bray D (1998) Predicting temporal fluctuations in an intracellular signalling pathway. Theor J Biol 192:117–128
71. LeNovère N, Shimizu TS (2001) StochSim: modelling of stochastic biomolecular processes. Bioinformatics 17:575–576
72. Lu T, Volfson D, Tsimring L, Hasty J (2004) Cellular growth and division in the Gillespie algorithm. Syst Biol 1:121–128
73. McAdams H, Arkin A (1998) Simulation of prokaryotic genetic circuits. Annu Rev Biophys Biomol Struct 27:199–224
74. Paulsson J (2004) Summing up the noise in gene networks. Nature 427:415–418
75. McAdams HH, Arkin A (1997) Stochastic mechanisms in gene expression. Proc Natl Acad Sci USA 94:814–819
76. Arkin A, Ross J, McAdams HH (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. Genetics 149:1633–1648
77. Wolf DM, Vazirani VV, Arkin AP (2005) Diversity in times of adversity: probabilistic strategies in microbial survival games. Theor J Biol 234:227–253
78. Fiering S, Whitelaw E, Martin DIK (2000) To be or not to be active: the stochastic nature of enhancer action. BioEssays 22:381–387
79. Barkai N, Leibler S (2000) Biological rhythms: Circadian clocks limited by noise. Nature 403:267–268
80. Thattai M, van Oudenaarden A (2004) Stochastic gene expression in fluctuating environments. Genetics 167:523–530
81. Kussell E, Leibler S (2005) Phenotypic diversity, population growth, and information in fluctuating environments. Science 309:2075–2078
82. Pedraza JM, van Oudenaarden A (2005) Noise propagation in gene networks. Science 307:1965–1969
83. Swain PS, Elowittz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci USA 99:12795–12800
84. Mettetal JT, Muzzey D, Pedraza JM, Ozbudak EM, van Oudenaarden A (2006) Predicting stochastic gene expression dynamics in single cells. Proc Natl Acad Sci USA 103:7304–7309
85. Kierzek AM, Zaim J, Zielenkiewicz P (2001) The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. Biol J Chem 276:8165–8172
86. Peccoud J, Ycart B (1995) Markovian modeling of gene-product synthesis. Theor Popul Biol 48:222–234
87. Rosenfeld N, Young JW, Alon U, Swain PS, Elowittz MB (2005) Gene regulation at the single-cell level. Science 307:1962–1965
88. Kitano H (2004) Biological robustness. Nature Rev Genet 5:826–837
89. Alon U, Surette MG, Barkai N, Leibler S (1999) Robustness in bacterial chemotaxis. Nature 397:168–171
90. Barkai N, Leibler S (1997) Robustness in simple biochemical networks. Nature 387:913–917
91. Stelling J, Sauer U, Szallasi Z, Doyle FJI, Doyle J (2004) Robustness of cellular functions. Cell 118:675–685
92. Vilar JMG, Kueh HY, Barkai N, Leibler S (2002) Mechanisms of noise-resistance in genetic oscillators. Proc Natl Acad Sci USA 99:5988–5992
93. Aldana M, Cluzel P (2003) A natural class of robust networks. Proc Natl Acad Sci USA 100:8710–8714
94. Thattai M, van Oudenaarden A (2002) Attenuation of noise in ultrasensitive signaling cascades. Biophys J 82:2943–2950
95. Yi T-M, Huang Y, Simon MI, Doyle J (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. Proc Natl Acad Sci USA 97:4649–4653
96. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB (2004) Noise minimization in eukaryotic gene expression. PLoBiol S 2:1–5
97. Voigt CA, Wolf DM, Arkin AP (2005) The Bacillus subtilis sin operon: an evolvable network motif. Genetics 169:1187–1202

98. Cao Y, Gillespie DT, Petzold LR (2006) Efficient step size selection for the tau-leaping simulation method. Chem J Phys 124:044109
99. Gillespie DT, Petzold LR (2003) Improved leap-size selection for accelerated stochastic simulation. Chem J Phys 119:8229–8234
100. Gillespie DT (2001) Approximate accelerated stochastic simulation of chemically reacting systems. Chem J Phys 115:1716–1733
101. Cao Y, Gillespie DT, Petzold LR (2005) Avoiding negative populations in explicit Poisson tau-leaping. Chem J Phys 123:054104
102. Chatterjee A, Mayawala K, Edwards JS, Vlachos DG (2005) Time accelerated Monte Carlo simulations of biological networks using the binomial tau-leap method. Bioinformatics 21:2136–2137
103. Pettigrew MF, Resat H (2007) Multinomial tau-leaping method for stochastic kinetic simulations. Chem J Phys 126:084101
104. Zwanzig R (2001) A chemical Langevin equation with non-Gaussian noise. Phys J Chem B 105:6472–6473
105. Gillespie DT (1996) The multivariate Langevin and Fokker–Planck equations. Am Phys J 64:1246–1257
106. Gillespie DT (2000) The chemical Langevin equation. Chem J Phys 113:297–306
107. Gillespie DT (2002) The chemical Langevin and Fokker–Planck equations for the reversible isomerization reaction. Phys J Chem A 106:5063–5071
108. Wang H, Peskin CS, Elston TC (2003) A robust numerical algorithm for studying biomolecular transport processes. Theor J Biol 221:491–511
109. Xing J, Wang H, Oster G (2005) From continuum Fokker–Planck models to discrete kinetic models. Biophys J 89:1551–1563
110. Tao Y (2004) Intrinsic noise, gene regulation and steady-state statistics in a two-gene network. Theor J Biol 231:563–568
111. van der Mee CVM, Zweifel PF (1987) A Fokker–Planck equation for growing cell populations. Math J Biol 25:61–72
112. Sato K, Kaneko K (2006) On the distribution of state values of reproducing cells. Phys Biol 3:74–82
113. Hill NA, Häder D-P (1997) A biased random walk model for the trajectories of swimming micro-organisms. Theor J Biol 186:503–526
114. Schienbein M, Gruler H (1993) Langevin equation, Fokker–Planck equation and cell migration. Bull Math Biol 55:585–608
115. Xing J, Liao J-C, Oster G (2005) Making ATP. Proc Natl Acad Sci USA 102:16539–16546
116. Elston TC, Oster G (1997) Protein turbines I: the bacterial flagellar motor. Biophys J 73:703–721
117. Allen RJ, Frenkel D, ten Wolde PR (2006) Simulating rare events in equilibrium or nonequilibrium stochastic systems. Chem J Phys 124:024102
118. Rathinam M, Petzold LR, Cao Y, Gillespie DT (2003) Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. Chem J Phys 119:12784
119. Cao Y, Gillespie DT, Petzold LR (2005) The slow-scale stochastic simulation algorithm. Chem J Phys 122:014116
120. Cao Y, Gillespie DT, Petzold LR (2005) Accelerated stochastic simulation of the stiff enzyme-substrate reaction. Chem J Phys 123:144917
121. Rao CV, Arkin AP (2003) Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. Chem J Phys 118:4999–5010
122. Adalsteinsson D, McMillen D, Elston TC (2004) Biochemical network stochastic simulator (BioNetS): software for stochastic modeling of biochemical networks. Bioinformatics BMC 5:24
123. Vasudeva K, Bhalla US (2004) Adaptive stochastic-deterministic chemical kinetic simulations. Bioinformatics 20:78–84
124. Baumeister W (2002) Electron tomography: towards visualizing the molecular organization of the cytoplasm. Curr Opin Struct Biol 12:679–684
125. Gierer A, Meinhardt H (1972) A theory of biological pattern formation. Biol Cyber 12:30–39
126. Maini PK, Painter KJ, Chau HNP (1997) Spatial pattern formation in chemical and biological systems. Chem J Soc Faraday Trans 93:3601–3610
127. Gurdon JB, Bourillot P-Y (2001) Morphogen gradient interpretation. Nature 413:797–803
128. Meinhardt H, de Boer PAJ (2001) Pattern formation in *Escherichia coli*: A model for the pole-to-pole oscillations of Min proteins and the localization of the division site. Proc Natl Acad Sci USA 98:14202–14207
129. Howard M, Rutenberg AD (2003) Pattern formation inside bacteria: fluctuations due to the low copy number of proteins. Phys Rev Lett 90:128102
130. Huang KC, Meir Y, Wingreen NS (2003) Dynamic structures in *Escherichia coli*: spontaneous formation of MinE rings and MinD polar zones. Proc Natl Acad Sci USA 100:12724–12728
131. Lutkenhaus J (2007) Assembly and dynamics of the bacterial MinCDE system and spatial regulation of the Z ring. Ann Rev Biochem 76:14.11–14.24
132. Fange D, Elf J (2006) Noise-induced Min phenotypes in *E coli*. PLoComp S Biol 2:637–648
133. Kerr RA, Levine H, Sejnowski TJ, Rappel W-J (2006) Division accuracy in a stochastic model of Min oscillations in *Escherichia coli*. Proc Natl Acad Sci USA 103:347–352
134. Cytrynbaum E, Marshall BDL (2007) A multi-stranded polymer model explains MinDE dynamics in *E coli* cell division. Biophys J 93:1134–1150
135. Bray D (1998) Signaling complexes: biophysical constraints on intracellular communication. Annu Rev Biophys Biomol Struct 27:59–75
136. Slepchenko BM, Schaff JC, Carson JH, Loew LM (2002) Computational cell biology: Spatiotemporal simulation of cellular events. Annu Rev Biophys Biomol Struct 31:423–441
137. Meyers J, Craig J, Odde DJ (2006) Potential for control of signaling pathways via cell size and shape. Curr Biol 16:1685–1693
138. Rao CV, Kirby JR, Arkin AP (2005) Phosphatase localization in bacterial chemotaxis: divergent mechanisms, convergent principles. Phys Biol 2:148–158
139. Agmon N, Edelstein AL (1997) Collective binding properties of receptor arrays. Biophys J 72:1582–1594
140. Lagerholm BC, Thompson NL (1998) Theory for ligand rebinding at cell membrane surfaces. Biophys J 74:1215–1228
141. Andrews SS (2005) Serial rebinding of ligands to clustered receptors as exemplified by bacterial chemotaxis. Phys Biol 2:111–122

142. Elf J, Ehrenberg M (2004) Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. Syst Biol 1:230–236
143. van Zon JS, ten Wolde PR (2005) Green's function reaction dynamics: A particle-based approach for simulating biochemical networks in time and space. Chem J Phys 123:234910
144. Stiles JR, Bartol TM (2001) Monte Carlo methods for simulating realistic synaptic microphysiology using MCell. In: De Schutter E (ed) Computational Neuroscience: Realistic Modeling for Experimentalists. Press CRC, Boca Raton
145. Dab D, Boon J-P, Li Y-X (1991) Lattice-gas automata for coupled reaction-diffusion equation. Phys Rev Lett 66:2535–2539
146. Ermentrout GB, Edelstein-Keshet L (1993) Cellular automata approaches to biological modeling. Theor J Biol 160:97–133
147. Duke TAJ, LeNovère N, Bray D (2001) Conformational spread in a ring of proteins: a stochastic approach to allostery. Mol J Biol 308:541–553
148. Goldman J, Andrews SS, Bray D (2004) Size and composition of membrane protein clusters predicted by Monte Carlo analysis. Eur Biophys J 33:506–512
149. Grima R, Schnell S (2006) A systematic investigation of the rate laws valid in intracellular environments. Biophys Chem 124:1–10
150. Turing AM (1990) The chemical basis of morphogenesis. Bull Math Biol 52:153–197
151. Cross MC, Hohenberg PC (1993) Pattern formation outside of equilibrium. Rev Mod Phys 65:851–1123
152. Slepchenko B, Schaff J, Macara I, Loew LM (2003) Quantitative cell biology with the Virtual Cell. Cell TRENDS Biol 13:570–576
153. Fink CC, Slepchenko B, Moraru II, Watras J, Schaff JC, Loew LM (2000) An image-based model of calcium waves in differentiated neuroblastoma cells. Biophys J 79:163–183
154. Fink CC, Slepchenko B, Moraru II, Schaff J, Watras J, Loew LM (1999) Morphological control of inositol-1,4,5-triphosphate-dependent signals. Cell J Biol 147:929–935
155. Hernjak N, Slepchenko B, Fernald K, Fink CC, Fortin D, Moraru II, Watras J, Loew LM (2005) Modeling and analysis of calcium signaling events leading to long-term depression in cerebellar Purkinje cells. Biophys J 89:3790–3806
156. Baras F, Malek-Mansour M (1996) Reaction-diffusion master equation: A comparison with microscopic simulations. Phys Rev E 54:6139–6148
157. Stundzia AB, Lumsden CJ (1996) Stochastic simulation of coupled reaction-diffusion processes. Comput J Phys 127:196–207
158. Nicolis G, Malek-Mansour M (1980) Systematic analysis of the multivariate master equation for a reaction-diffusion system. Stat J Phys 22:495–512
159. Kruse K, Elf J (2006) Kinetics in spatially extended systems. In: Szallasi Z, Stelling J, Periwal V (eds) System Modeling in Cell Biology From Concepts to Nuts and Bolts. Press MIT, Cambridge, pp 177–198
160. Hynes JT (1985) The theory of reactions in solution. In: Baer M (ed) Theory of Chemical Reaction Dynamics. Press CRC, Boca Raton, pp 171–234
161. Cohen B, Huppert D, Agmon N (2000) Non-exponential Smoluchowski dynamics in fast acid-base reaction. Am J Chem Soc 122:9838–9839
162. Noyes RM (1955) Kinetics of competitive processes when reactive fragments are produced in pairs. Am J Chem Soc 77:2042–2045
163. Pines E, Huppert D (1988) Geminate recombination in excited-state proton transfer reactions: Numerical solution of the Debye–Smoluchowski equation with backreaction and comparison with experimental results. Chem J Phys 88:5620–5630
164. Verkman AS (2002) Solute and macromolecule diffusion in cellular aqueous compartments. Trends Biochem Sci 27:27–33
165. Schnell S, Turner TE (2004) Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. Prog Biophys Mol Biol 85:235–260
166. Fulton AB (1982) How crowded is the cytoplasm? Cell 30:345–347
167. Bernstein D (2005) Simulating mesoscopic reaction-diffusion systems using the Gillespie algorithm. Phys Rev E 71:041103
168. Ander M, Beltrao P, Di Ventura B, Ferkinghoff-Borg J, Foglierini M, Kaplan A, Lemerle C, Tomás-Oliveira I, Serrano L (2004) SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. Syst Biol 1: 129–138
169. Fricke T, Schnakenberg J (1990) Monte-Carlo simulation of an inhomogeneous reaction-diffusion system in the biophysics of receptor cells. Z Phys B 83:277–284
170. Isaacson SA, Peskin CS (2006) Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations. Sci SIAMJ Comput 28:47–74
171. Hattne J, Fange D, Elf J (2005) Stochastic reaction-diffusion simulation with MesoRD. Bioinformatics 21:2923–2924
172. Frenkel D, Smit B (2002) Understanding molecular simulation: from algorithms to applications. Academic, San Diego
173. Berg HC (1993) Random Walks in Biology. Princeton Univ Press, Princeton
174. Gillespie DT (1996) The mathematics of Brownian motion and Johnson noise. Am Phys J 64:225–240
175. Rice SA (1985) Diffusion Limited Reactions. Elsevier, Amsterdam
176. Ermak DL, McCammon JA (1978) Brownian dynamics with hydrodynamic interactions. Chem J Phys 69:1352–1360
177. Northrup SH, Allison SA, McCammon JA (1984) Brownian dynamics simulation of diffusion-influenced bimolecular reactions. Chem J Phys 80:1517–1524
178. Northrup SH (1988) Diffusion-controlled ligand binding to multiple competing cell-bound receptors. Phys J Chem 92:5847–5850
179. Northrup SH, Erickson HP (1992) Kinetics of protein-protein association explained by Brownian dynamics computer simulation. Proc Natl Acad Sci USA 89:3338–3342
180. Edelstein AL, Agmon N (1993) Brownian dynamics simulations of reversible reactions in one dimension. Chem J Phys 99:5396–5404
181. Oh C, Kim H, Shin KJ (2002) Excited-state diffusion-influenced reversible association-dissociation reaction: Brownian dynamics simulation in three dimensions. Chem J Phys 117:3269–3277
182. Kim H, Yang M, Shin KJ (1999) Dynamic correlation effect in reversible diffusion-influenced reactions: Brownian dynamics simulation in three dimensions. Chem J Phys 111:1068–1075
183. Agmon N, Edelstein A (1995) Geometric and many-particle aspects of transmitter binding. Biophys J 68:815–825

184. Edelstein AL, Agmon N (1997) Brownian simulation of many-particle binding to a reversible receptor array. Comput J Phys 132:260–275
185. Agmon N, Szabo A (1990) Theory of reversible diffusion-influenced reactions. Chem J Phys 92:5270–5284
186. Kim H, Shin KJ (1999) Exact solution of the reversible diffusion-influenced reaction for an isolated pair in three dimensions. Phys Rev Lett 82:1578–1581
187. van Zon JS, ten Wolde PR (2005) Simulating biochemical networks at the particle level in time and space: Green's function reaction dynamics. Phys Rev Lett 94:128103
188. van Zon JS, Morelli MJ, Tanase-Nicola S, ten Wolde PR (2006) Diffusion of transcription factors can drastically enhance the noise in gene expression. Biophys J 91:4350–4367
189. Lipkow K (2006) Changing cellular location of CheZ predicted by molecular simulations. Comp PLOS Biol 2:301–310
190. Lipkow K, Andrews SS, Bray D (2004) Simulated diffusion of CheYp through the cytoplasm of *E coli*. J Bact 187:45–53
191. Tournier AL, Fitzjohn PW, Bates PA (2006) Probability-based model of protein-protein interactions on biological timescales. Algorithms Molec Biol 1:25
192. Tolle DP, Le Novère N (2006) Particle-based stochastic simulation in systems biology. Curr Bioinformatics 1:1–6
193. Franks KM, Bartol TM, Sejnowski TJ (2002) A Monte Carlo model reveals independent signaling at central glutametergic synapses. Biophys J 83:2333–2348
194. Coggan JS, Bartol TM, Esquenazi E, Stiles JR, Lamont S, Martone ME, Berg DK, Ellisman MH, Sejnowski TJ (2005) Evidence for ectopic neurotransmission at a neuronal synapse. Science 309:446–451
195. Koh X, Srinivasan B, Ching HS, Levchenko A (2006) A 3D Monte Carlo analysis of the role of dyadic space geometry in spark generation. Biophys J 90:1999–2014
196. Stiles JR, van Helden D, Thomas J, Bartol M, Salpeter EE, Salpeter MM (1996) Miniature endplate current rise times < 100 microseconds from improved dual recordings can be modeled with passive acetylcholine diffusion from a synaptic vesicle. Proc Natl Acad Sci USA 93:5747–5752
197. Stiles JR, Kovyazina IV, Salpeter EE, Salpeter MM (1999) The temperature sensitivity of miniature endplate currents is mostly governed by channel gating: Evidence from optimized recordings and Monte Carlo simulations. Biophys J 77:1177–1187
198. Bartol TMJ, Land BR, Salpeter EE, Salpeter MM (1991) Monte Carlo simulation of miniature endplate current generation in the vertebrate neuromuscular junction. Biophys J 59:1290–1307
199. Howard M, Rutenberg AD, de Vet S (2001) Dynamic compartmentalization of bacteria: accurate division in *E coli*. Phys Rev Lett 87:278102
200. Kruse K (2002) A dynamic model for determining the middle of *Escherichia coli*. Biophys J 82:618–627
201. Wio HS (1996) Stochastic resonance in a spatially extended system. Phys Rev E 54:R3075–R3078
202. Hu Z, Gogol EP, Lutkenhaus J (2002) Dynamic assembly of MinD on phospholipid vesicles regulated by ATP and MinE. Proc Natl Acad Sci USA 99:6761–6766
203. Hu Z, Saez C, Lutkenhaus J (2003) Recruitment of MinC, an inhibitor of Z-ring formation, to the membrane in *Escherichia coli*: role of MinD and MinE. J Bact 185:196–203
204. Shih Y-L, Fu X, King GF, Le T, Rothfield L (2002) Division site placement in *E coli*: mutations that prevent formation of the MinE ring lead to loss of the normal midcell arrest of growth of polar MinD membrane domains. EMBOJ 21:3347–3357
205. Shih Y-L, Le T, Rothfield L (2003) Division site selection in *Escherichia coli* involves dynamic redistribution of Min proteins within coiled structures that extend between the two cell poles. Proc Natl Acad Sci USA 100:7865–7870
206. Shih Y-L, Kawagishi I, Rothfield L (2005) The MreB and Min cytoskeletal-like systems play independent roles in prokaryotic polar differentiation. Mol Microbiol 58:917–928
207. Suefuji K, Valluzzi R, RayChaudhuri D (2002) Dynamic assembly of MinD into filament bundles modulated by ATP, phospholipids, and MinE. Proc Natl Acad Sci USA 99:16776–16781
208. Pavin N, Paljetak C, Krstic V (2006) Min-protein oscillations in *Escherichia coli* with spontaneous formation of two-stranded filaments in a three-dimensional stochastic reaction-diffusion model. Phys Rev E 73:021904
209. Adelman JL, Andrews SS (2004) Intracellular pattern formation: A spatial stochastic model of bacterial division site selection proteins MinProc CDE. Complex Systems Summer School Final Project Papers, Santa Fe Institute, Santa Fe
210. Drew DA, Osborn MJ, Rothfield LI (2005) A polymerization-depolymerization model that accurately generates the self-sustained oscillatory system involved in bacterial division site placement. Proc Natl Acad Sci USA 102:6114–6118
211. Andrews SS, Arkin AP (2007) A mechanical explanation for cytoskeletal rings and helices in bacteria. Biophys J 93:1872–1884
212. Lippincott-Schwartz J, Snapp E, Kenworthy A (2001) Studying protein dynamics in living cells. Nat Rev Mol Cell Biol 2:444–456

Stochastic Noises, Observation, Identification and Realization with

Giorgio Picci
Department of Information Engineering,
University of Padua, Padua, Italy

Article Outline

Glossary
Introduction
Stochastic Realization
Wide-Sense Stochastic Realization
Geometric Stochastic Realization
Dynamical System Identification
Future Directions
Bibliography

Glossary

Wiener process A *Wiener process* is a continuous time stochastic process $w = \{w(t)\,;\, t \in \mathbb{R}\}$ with continuous sample paths and stationary independent incre-