SIAM J. SCI. COMPUT.
Vol. 28, No. 1, pp. 47–74

INCORPORATING DIFFUSION IN COMPLEX GEOMETRIES INTO STOCHASTIC CHEMICAL KINETICS SIMULATIONS*

SAMUEL A. ISAACSON† AND CHARLES S. PESKIN‡

Abstract. A method is developed for incorporating diffusion of chemicals in complex geometries into stochastic chemical kinetics simulations. Systems are modeled using the reaction-diffusion master equation, with jump rates for diffusive motion between mesh cells calculated from the discretization weights of an embedded boundary method. Since diffusive jumps between cells are treated as first order reactions, individual realizations of the stochastic process can be created by the Gillespie method. Numerical convergence results for the underlying embedded boundary method, and for the stochastic reaction-diffusion method, are presented in two dimensions. A two-dimensional model of transcription, translation, and nuclear membrane transport in eukaryotic cells is presented to demonstrate the feasibility of the method in studying cell-wide biological processes.

Key words. reaction-diffusion, stochastic chemical kinetics, complex geometry, embedded boundary

AMS subject classifications. 65C20, 92C40, 65M06

DOI. 10.1137/040605060

1. Introduction.

Spatially homogeneous, deterministic mass-action kinetics is a standard model for the interactions of proteins, genes, and mRNAs within cellular networks. It assumes that the evolution of any chemical species can be represented by a continuously varying concentration representing the number of that species within the cell divided by the cell volume. Fundamentally, however, the numbers of protein and mRNA molecules within cells are discrete integer variables. The notion of a continuously varying concentration is only well defined for a sufficiently large number of the molecules. Moreover, for systems in which spatial effects are important the population number must not only be large within the cell but also locally within areas of interest. If the population number is sufficiently small, stochastic effects can have a noticeable impact on the behavior of the biological system.

The issue of stochasticity in gene expression was discussed in [3]. The authors subsequently demonstrated numerically that the variation of cell fates within a λ-phage infected population of bacterial cells can be explained as arising from a stochastically driven gene expression switch [4]. The presence of noise in prokaryotic transcription and translation was demonstrated experimentally in both [9] and [23]. More recently, experimental results demonstrating transcriptional noise in eukaryotic cells were presented in [6].

Spatial effects are present in many biological systems, and hence the spatially uniform, or well-mixed, assumption will not always hold. Systems in which spatial effects are known to play a role include Ash1 mRNA localization in budding yeast [1],

*Received by the editors March 11, 2004; accepted for publication (in revised form) August 11, 2005; published electronically March 3, 2006. This work was partially supported by the National Computational Science Alliance under DMS030001N and utilized the NCSA IA–32 Linux Cluster. Partial support was also provided by Defense Advanced Research Projects Agency–Air Force Research Laboratory grant F30602-02-0554.

http://www.siam.org/journals/sisc/28-1/60506.html

†Department of Mathematics, University of Utah, 155 S. 1400 E., Room 233, Salt Lake City, UT 84112 (isaacsas@cims.nyu.edu).

‡Courant Institute of Mathematical Sciences, 251 Mercer St., New York, NY 10012 (peskin@cims.nyu.edu).

morphogen gradients across egg-polarity genes in Drosophila oocyte [1], and the synapse-specificity of long-term facilitation in Aplysia [20]. The exact processes by which proteins and mRNAs move between specific locations, and become localized, are not yet known definitively. Movement is believed to be composed of a mixture of diffusion and various forms of active transport. For example, in [10], mRNA movement within the cytoplasm was found to exhibit diffusive motion, "corralled" diffusive motion, and active, directionally specific, transport. A probabilistic switching between these three types of movement was observed in most cases. Moreover, the fraction of time spent in each of the types of motion was found to differ depending on whether the mRNA was destined to be localized or not.

Diffusion can be incorporated into deterministic mass-action kinetics through the use of reaction-diffusion partial differential equations. Molecular noise can be accounted for through the use of stochastic chemical kinetics, but the usual formulation of stochastic chemical kinetics assumes a spatially homogeneous cellular space. Methods for incorporating diffusion into stochastic chemical kinetics are presented in [25] and [12]. In both methods, space is divided into a collection of equally sized mesh cells. A reaction-diffusion master equation is then derived to give the probability of the system being in a given state. In [12] all transition rates between mesh cells are identical, and formulas for the transition rates are calculated in this case. In [25] a discussion of the continuum-limit for the first several moments of the model is given, but no method is given for calculating the cell-based transition rates.

We present a stochastic reaction-diffusion model for mesh cells of arbitrary volume and varying diffusive transition rates. In section 3, a method is developed for calculating diffusive transition rates between mesh cells within a domain containing multiple complex boundaries. The development of a convergent diffusive approximation within complex geometries is critical for use in individual (biological) cell models where the large number of organelles and membranes could have a significant effect on the overall dynamics of chemical systems. Transition rates are systematically calculated from terms in an embedded boundary discretization of the Fokker–Planck equation for a classically diffusing Brownian particle. We expect the reaction-diffusion master equation to be such a discretization in the case of no reactions and only one particle. The notion of the master equation as an approximation to a continuum equation was used in [26] to obtain approximations of the Fokker–Planck equation for a single particle moving in a one-dimensional potential. The master equation itself has a special structure, requiring a constraint on the weights of the discretization. Our method is found to conserve probability and satisfy detailed balance. The underlying embedded boundary discretization converges between first and second order in practice. As an application of the method, a model of transcription, translation, and nuclear transport within a two-dimensional eukaryotic cell is developed and simulated.

We note that the method we develop is an Eulerian method, using an underlying Cartesian mesh. Other models have been developed using Lagrangian methods that track the movement of each individual molecule within the system. For example, in [2] a Lagrangian method was developed for simulating stochastic reaction-diffusion systems within biological cells. These methods have the benefit of allowing exact local treatments of particle diffusion; however, they introduce other challenges in how to handle reactions between particles, detection of particle collisions, and resolving particle-geometry interactions. Lagrangian methods have been found very useful in the turbulent reactive flow literature. We point the reader interested in numerical methods for convection dominated reactive flows to [22] and [19].

2. Master equation formulation of reaction-diffusion. In section 2.1 the general abstract definition of the master equation is presented. Several properties of the transition rates, $W_{ij}(t)$, are pointed out. These properties constrain the form of the discretization of the Fokker–Planck equation used in generating the diffusive transition rates for the reaction-diffusion master equation. Sections 2.2 and 2.3 review the chemical master equation for a spatially homogeneous system, and the Gillespie method for creating realizations of it. In section 2.4 the reaction-diffusion master equation is presented, along with the motivation for thinking of its diffusive jump rates as weights in a spatial-only discretization of the Fokker–Planck equation.

2.1. General master equation. Let $X(t)$ denote an integer valued, continuous time Markov process. Let $Q_i(t) = \text{Prob}\{X(t) = i | X(0) = x_0\}$, and $W_{ij}(t)dt = \text{Prob}\{X(t+dt) = i | X(t) = j\}$. By definition, $W_{ij}(t)$ is the transition rate at time t to go into state i from state j, for $i \neq j$. Then

$$Q_i(t+dt) = \left(1 - \sum_{j \neq i} W_{ji}(t)dt\right) Q_i(t) + \sum_{j \neq i} W_{ij}(t)Q_j(t)dt + o(dt). \tag{2.1}$$

This says that the probability of being in state i a small time after time t is the probability of being there at time t and staying there added to the probability of being in another state at time t and transitioning to i. The master equation is the coupled set of ODEs obtained by rearranging and taking the limit $dt \to 0$ of (2.1):

$$\begin{aligned} \frac{dQ_i}{dt} &= \sum_{j:j\neq i} \left(W_{ij}(t)Q_j(t) - W_{ji}(t)Q_i(t)\right) \\ &= \left(\sum_{j:j\neq i} W_{ij}(t)Q_j(t)\right) - W_i(t)Q_i(t), \quad W_i(t) = \sum_{j\neq i} W_{ji}(t) \geqslant 0. \end{aligned} \tag{2.2}$$

There are several properties of the structure of (2.2) that we subsequently use. First, $W_{ij}(t) \geq 0$. Second, the term $W_{ij}(t)Q_j(t)$ appears twice with opposite sign: once in the equation for $\partial_t Q_i$ and once in the equation for $\partial_t Q_j$. This implies that the amount of probability that flows per unit time into state i due to transitions from state j to i is exactly the amount of probability that flows per unit time out of state j due to these same transitions. Finally, the only negative term in the equation for $Q_i(t)$ is given by $W_i(t)Q_i(t)$.

2.2. Review of stochastic chemical kinetics. Consider a well-mixed chemical system in a finite closed volume V; that is, the probability of a given particle of the system being in an arbitrary subregion of volume dV is dV/V. Further, let each individual particle's velocity be a Maxwell–Boltzmann distributed random variable. With these assumptions Gillespie showed in [16] that a chemical system can be represented as an integer valued, continuous time Markov process satisfying a master equation.

Let $\boldsymbol{M}(t) = (M^1(t), \ldots, M^L(t))$ denote the state vector of the chemical system. $M^l(t)$ will be the random variable representing the number of molecules of chemical species l at time t. Define $\boldsymbol{m} = (m^1, \ldots, m^L)$ to be a given value of $\boldsymbol{M}(t)$. Denote by S^l the *name* of the lth species. Assume there are K possible reactions, with the function $a^k(\boldsymbol{m})$ giving the probability per unit time of reaction k occurring when $\boldsymbol{M}(t) = \boldsymbol{m}$. For example, letting k label the unimolecular (first order) reaction

$S^i \to S^j$, then $a^k(\boldsymbol{m}) = \alpha\, m^i$, where α is the rate constant in units of number of occurrences of the reaction per molecule of S^i per unit time. Letting k' denote the index of the bimolecular reaction $S^i + S^j \to S^n$, where $i \neq j$, then $a^{k'}(\boldsymbol{m}) = \beta\, m^i m^j$. Here β is the rate constant in units of number of occurrences of the reaction per molecule of S^i and per molecule of S^j, per unit time. Note that if species i and j are the same, then $a^{k'}(\boldsymbol{m}) = (\beta/2)\, m^i(m^i - 1)$. Let $\boldsymbol{\nu}_k = (\nu_k^1, \ldots, \nu_k^L)$ be the change in $\boldsymbol{M}$ that results from one occurrence of reaction k (i.e., $\boldsymbol{M}(t) \to \boldsymbol{M}(t) + \boldsymbol{\nu}_k$). The integers $(\nu_k^1, \ldots, \nu_k^L)$ define the stoichiometry of the kth reaction. In the notation of section 2.1, $a^k(\boldsymbol{m}) = W_{\boldsymbol{m}+\boldsymbol{\nu}_k, \boldsymbol{m}}$, assuming $\boldsymbol{\nu}_i \neq \boldsymbol{\nu}_k$ for all $i \neq k$. (If two or more reactions have identical stoichiometry, their rates are added to obtain the corresponding transition rate W. Also, if there are no reactions leading from state $\boldsymbol{m}'$ to $\boldsymbol{m}$, then $W_{\boldsymbol{m},\boldsymbol{m}'}$ is zero.) Note that the ν_k^l may be positive or negative (or zero). Then if

$$P(\boldsymbol{m}, t) \equiv P(\boldsymbol{m}) \equiv \text{Prob}\{\boldsymbol{M}(t) = \boldsymbol{m} | \boldsymbol{M}(0) = \boldsymbol{m}_0\},$$

the master equation for the time evolution of the probability of the system will be

$$\frac{dP(\boldsymbol{m})}{dt} = \sum_{k=1}^{K} \big(a^k(\boldsymbol{m} - \boldsymbol{\nu}_k) P(\boldsymbol{m} - \boldsymbol{\nu}_k) - a^k(\boldsymbol{m}) P(\boldsymbol{m})\big). \tag{2.3}$$

This is a coupled set of ODEs over all possible integer values of the components of the vector $\boldsymbol{m}$. We shall subsequently refer to (2.3) as the chemical master equation. Assuming all the $\boldsymbol{\nu}_k$ are distinct, in the notation of section 2.1 the master equation is

$$\frac{dP(\boldsymbol{m})}{dt} = \sum_{k=1}^{K} (W_{\boldsymbol{m},\boldsymbol{m}-\boldsymbol{\nu}_k} P(\boldsymbol{m} - \boldsymbol{\nu}_k) - W_{\boldsymbol{m}+\boldsymbol{\nu}_k,\boldsymbol{m}} P(\boldsymbol{m})). \tag{2.4}$$

In both (2.3) and (2.4) the first term in the sum corresponds to all possible transitions into state $\boldsymbol{m}$ from some prestate, $\boldsymbol{m} - \boldsymbol{\nu}_k$, while the second term corresponds to all possible transitions out of state $\boldsymbol{m}$ to $\boldsymbol{m} + \boldsymbol{\nu}_k$.

We now consider the relationship between the traditional macroscopic chemical kinetic equations to the master equation derived above. Define $\boldsymbol{C} \equiv \boldsymbol{M}/V$ to be the vector of random variables for the concentration of each species. Let $\tilde{a}^k$ denote the concentration dependent form of the rate a^k. $\tilde{a}^k$ and a^k are related by $\tilde{a}^k(\boldsymbol{c}) = a^k(V\boldsymbol{c})/V$, and vice versa $a^k(\boldsymbol{m}) = \tilde{a}^k(\boldsymbol{m}/V)V$. Note that $\tilde{a}^k(\boldsymbol{c})$ is the expected number of occurrences of reaction k *per unit volume* per unit time when the vector of concentrations is $\boldsymbol{c}$. From the master equation the mean of C^l, $\langle C^l \rangle$, satisfies

$$\frac{d\langle C^l \rangle}{dt} = \sum_{k=1}^{K} \nu_k^l \langle \tilde{a}^k(\boldsymbol{C}) \rangle.$$

If the averaging operation commutes with the functions $\tilde{a}^k$, then one obtains the equations for classical deterministic chemical kinetics,

$$\frac{d\langle C^l \rangle}{dt} = \sum_{k=1}^{K} \nu_k^l\, \tilde{a}^k(\langle \boldsymbol{C} \rangle);$$

however, in general $\langle \tilde{a}^k(\boldsymbol{C}) \rangle \neq \tilde{a}^k(\langle \boldsymbol{C} \rangle)$, unless $\tilde{a}^k$ is linear. One does expect that asymptotically, as the population sizes become very large, $\langle \tilde{a}^k(\boldsymbol{C}) \rangle \approx \tilde{a}^k(\langle \boldsymbol{C} \rangle)$. For a discussion of the validity of the deterministic approximation see [17].

2.3. Gillespie method. The number of the ODEs comprising the master equation are very large even for simple chemical systems. Direct solution methods become impractical, and thus Monte Carlo methods are used instead. These methods create realizations of the underlying stochastic process, $\boldsymbol{M}(t)$, governed by the master equation. Statistics from many realizations can be used to calculate moments of $\boldsymbol{M}(t)$ and the distribution, $P(\boldsymbol{m}, t)$. In [15] Gillespie presented two equivalent methods that produce exact realizations of $\boldsymbol{M}(t)$.

One of the two methods, the first reaction method, is based on calculating the time of occurrence of each reaction independently, as if no other reactions are present in the system. The reaction with the minimal time to its next occurrence is then executed, and the system state updated. This process is then repeated until the simulation reaches the desired final time. In detail, let

$$p\left(\tau, k|\boldsymbol{m}, t\right) d\tau \equiv \text{ Prob}\{\text{reaction } k \text{ occurs in the interval } [t+\tau, t+\tau+d\tau), \\ \text{ignoring that another reaction could occur first}\}.$$

Then in [15] it is shown that

$$p\left(\tau, k|\boldsymbol{m}, t\right) = a^k\left(\boldsymbol{m}\right) e^{-a^k(\boldsymbol{m})\tau}. \tag{2.5}$$

Knowing $\boldsymbol{m}$ at time t, for each reaction k, a time τ_k at which reaction k would occur can be generated by inverting the probability distribution corresponding to probability density (2.5). The time at which reaction k would occur is given by $t + \tau_k$, where

$$\tau_k = \frac{1}{a^k(\boldsymbol{m})} \ln \frac{1}{r_k},$$

r_k being a uniformly distributed random number in $[0, 1]$. The overall time-evolution algorithm is then the following:

1. Initialize $t = 0$, set initial molecule numbers $\boldsymbol{m} = \boldsymbol{m}_0$
2. Calculate $a^k(\boldsymbol{m})$ for all k
3. For each k, generate $\tau_k = \left(\ln\left(1/r_k\right)\right)/a^k(\boldsymbol{m})$
4. Execute the reaction, k', with the smallest τ_k, update $\boldsymbol{m} := \boldsymbol{m} + \boldsymbol{\nu}_{k'}$
5. Set the time to $t := t + \tau_{k'}$
6. Return to 2.

Optimized versions of Gillespie's algorithms that substantially decrease the computational work were developed by Gibson and Bruck [14]. Their next reaction method is equivalent to the first reaction method but requires only one random number per reaction event. By using a special data structure they reduce finding the reaction with the minimum τ_k to a constant work operation, and updating following a reaction to logarithmic work in the number of reactions, K. For all subsequent Monte Carlo simulations discussed we use the next reaction method.

2.4. Reaction-diffusion master equation. Let the domain of interest be a closed volume V. Divide the domain into a series of computational cells indexed by $i = 1, \ldots, N$. Assume that the size of each cell can be chosen such that within each cell, independently, the master equation formulation of chemical kinetics holds (see section 2.2). Transitions of particles *between* cells are then modeled as first order reactions.

Let $M_i^l(t)$ denote the random variable for the number of particles of species l in cell i, $l = 1, \ldots, L$. Define $\boldsymbol{M}^l(t) = (M_1^l, \ldots, M_N^l)^T$ to be the spatial vector of

species l, and $\boldsymbol{M}(t) = (\boldsymbol{M}^1, \ldots, \boldsymbol{M}^L)$ to be the state matrix of the system. As for the well-mixed case, let

$$P(\boldsymbol{m}, t) \equiv P(\boldsymbol{m}) \equiv \text{Prob}\{\boldsymbol{M}(t) = \boldsymbol{m} | \boldsymbol{M}(0) = \boldsymbol{m}_0\}.$$

Let $\boldsymbol{M}_i(t) = (M_i^1, \ldots, M_i^L)$, $\boldsymbol{e}_i^l$ be the matrix all of whose elements are zero except for the element (i, l), which is one, and let $\boldsymbol{e}_i$ be the column vector which has a 1 in its ith entry and is 0 elsewhere. $\boldsymbol{e}_i\boldsymbol{\nu}_k$ will denote the matrix formed by the product of the column vector $\boldsymbol{e}_i$ with the row vector $\boldsymbol{\nu}_k$. This matrix will be zero everywhere except for the ith row, which will be equal to $\boldsymbol{\nu}_k$. It corresponds to the change in $\boldsymbol{M}(t)$ due to one occurrence of reaction k at location i (i.e., $\boldsymbol{M}(t) \to \boldsymbol{M}(t) + \boldsymbol{e}_i\boldsymbol{\nu}_k$). Let k_{ij}^l denote the jump rate for each individual molecule of the lth chemical species into cell i from cell j, for $i \neq j$. Since diffusion is treated as a first order reaction, and since the molecules are assumed to diffuse independently, the total probability per unit time at time t for one molecule of species l to jump from cell j to cell i is $k_{ij}^l M_j^l(t)$.

The master equation for the time evolution of $P(\boldsymbol{m}, t)$ is then

$$\begin{aligned} \frac{dP(\boldsymbol{m})}{dt} =& \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{L}\left(k_{ij}^l\left(m_j^l + 1\right)P(\boldsymbol{m} + \boldsymbol{e}_j^l - \boldsymbol{e}_i^l) - k_{ji}^l m_i^l P(\boldsymbol{m})\right) \\ &+ \sum_{i=1}^{N}\sum_{k=1}^{K}\left(a_i^k(\boldsymbol{m}_i - \boldsymbol{\nu}_k)P(\boldsymbol{m} - \boldsymbol{e}_i\boldsymbol{\nu}_k) - a_i^k(\boldsymbol{m}_i)P(\boldsymbol{m})\right). \end{aligned} \tag{2.6}$$

This is a coupled set of ODEs over all possible integer values of the matrix $\boldsymbol{m}$. Notice the important point that the reaction probabilities per unit time, $a_i^k(\boldsymbol{m})$, may now depend on spatial location. For example, in an eukaryotic cell some reactions may occur only in the nucleus and others may occur only in the cytoplasm.

Equation (2.6) is separated into two sums. The first term corresponds to diffusive motion between cells i and j of a given species, l. The second is just the components of the chemical master equation, but applied at each individual cell. At this point no specification has been made as to where the rates k_{ij}^l come from, or what their values should be.

For deterministic reaction-diffusion, let $C^l(\boldsymbol{x}, t)$ denote the concentration and D^l the diffusion constant of the lth species. Let $\tilde{a}^k(\boldsymbol{C}(\boldsymbol{x}, t), \boldsymbol{x})$ denote the spatially varying concentration dependent form of a_i^k (see section 2.2). Then

$$\frac{\partial C^l(\boldsymbol{x}, t)}{\partial t} = D^l \Delta C^l + \sum_{k=1}^{K} \nu_k^l \, \tilde{a}^k(\boldsymbol{C}(\boldsymbol{x}, t), \boldsymbol{x}).$$

To see the relation to the stochastic formulation let $C_i^l(t) = M_i^l(t)/V_i$ be the random variable for the chemical concentration of species l, in cell i. Let $\tilde{a}_i^k$ denote the concentration dependent form of a_i^k. $\tilde{a}_i^k$ and a_i^k are related by $\tilde{a}_i^k(\boldsymbol{c}) = a_i^k(V_i\boldsymbol{c})/V_i$, and vice versa $a_i^k(\boldsymbol{m}) = \tilde{a}_i^k(\boldsymbol{m}/V_i)V_i$. From the reaction-diffusion master equation the average of C_i^l then satisfies

$$\frac{d\langle C_i^l\rangle}{dt} = \sum_{j=1}^{N}\left(\frac{V_j}{V_i}k_{ij}^l\langle C_j^l\rangle - k_{ji}^l\langle C_i^l\rangle\right) + \sum_{k=1}^{K}\nu_k^l\langle\tilde{a}_i^k\left(C_i\right)\rangle.$$

Suppose the rates k_{ij}^l could be chosen so that the first term is a discretization of $D^l \Delta C^l$. The only difference between the deterministic and the means of the stochastic formulation in the continuum limit would then be the noncommutativity of the reaction terms with averaging; diffusion would be identical.

Returning to the reaction-diffusion master equation (2.6), we expect that the diffusive jump rates should be independent of the chemical reactions present within a given chemical system. As such, we now consider a chemical system consisting of no reactions. In this case, we expect each particle of each chemical species to independently undergo Brownian motion in the continuum limit as the computational cell size approaches zero. Since, by construction, the diffusive motion of each particle of each chemical species is independent in the reaction-diffusion master equation formulation, it is sufficient to further restrict ourselves to a system consisting of only one chemical species containing just one particle. With these choices, the reaction-diffusion master equation reduces to

$$\frac{dP(\boldsymbol{m})}{dt} = \sum_{i=1}^{N}\sum_{j=1}^{N} k_{ij}\,(m_j+1)\,P(\boldsymbol{m}+\boldsymbol{e}_j-\boldsymbol{e}_i) - k_{ji} m_i P(\boldsymbol{m}). \tag{2.7}$$

Since $\boldsymbol{M}(t)$ is describing one particle, $M_i(t)$ will be zero for all i except the location the particle is currently at. Let $\boldsymbol{X}(t)$ be the Brownian motion process of the particle. Then define

$$Q_i(t) \equiv P(\boldsymbol{e}_i, t) \approx \text{Prob}\{\boldsymbol{X}(t) \in V_i | \boldsymbol{X}(0) = \boldsymbol{x}_0\},$$

where $\boldsymbol{e}_i$ denotes, as before, the unit vector along the ith Cartesian coordinate axis of $\mathbb{R}^N$. With this definition, $Q_i(t)$ is the probability that the particle is in the ith computational cell. Letting $\boldsymbol{m} = \boldsymbol{e}_i$, equation (2.7) implies

$$\frac{dQ_i}{dt} = \sum_{j=1}^{N} k_{ij}\,Q_j - k_{ji}\,Q_i. \tag{2.8}$$

Define a probability density for the particle's position by $p_i(t) \equiv Q_i(t)/V_i$. Then p_i satisfies

$$\frac{dp_i}{dt} = \sum_{j=1}^{N} \frac{V_j}{V_i} k_{ij}\,p_j - k_{ji}\,p_i. \tag{2.9}$$

Again, if the rates form a discretization of $D\Delta p$, for p the continuum single particle density, then the continuum limit as cell size goes to zero would be

$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} = D\Delta p. \tag{2.10}$$

This is the Fokker–Planck equation for a single classical Brownian particle.

With these continuum limits in mind, constructing an operator, L_h, of the form

$$(L_h p)_i = \sum_{j=1}^{N} \frac{V_j}{V_i} k_{ij}\,p_j - k_{ji}\,p_i,$$

where $L_h \to \Delta$ as $h \to 0$, will determine the k_{ij}. (Here h represents the maximum length scale associated with a computational cell.) The diffusion jump rates will then be k_{ij}, provided $k_{ij} \geq 0$ (see section 2.1).

3. Determining the jump rates. From the results of section 2.4, an appropriate spatial discretization of either the classical deterministic formulation of reaction-diffusion, or the Fokker–Planck equation for a single Brownian particle, would determine the jump rates in (2.6). We note that in choosing the jump rates to recover either one of these two formulations, the macroscopic diffusion of the mean concentration from the reaction-diffusion master equation, or the microscopic Brownian motion of independent particles, we will also recover the other formulation. Previous authors [12], [25] have used the requirement of obtaining macroscopic diffusion of the mean concentrations given by the reaction-diffusion master equation to derive diffusive jump rates in the special case of a uniform Cartesian mesh. In this case, it has been shown that the jump rates, with mesh width h, can be given by

$$\frac{D}{h^2}$$

for jumps between neighboring mesh cells, and are zero for jumps between non-neighboring cells. This result has been known from as far back as [11]. In this section, instead of focusing on recovering macroscopic diffusion of the mean concentrations, we focus on the mathematically equivalent requirement of reproducing single particle Brownian motion. We note that if we assume molecules are only moving by diffusion, then the requirement of recovering single particle Brownian motion in the continuum limit is a physical necessity for the reaction-diffusion master equation. With this choice we can derive valid jump rates for diffusion within complex geometries, allowing us to extend the existing methodology to handle arbitrary boundaries. This result is important since many problems of practical interest contain geometrically complex domains, such as the movement of molecules within biological cells where there are many barriers due to organelles and membranes.

To discretize the Laplacian, and hence obtain the diffusive jump rates, boundary conditions must be specified for the domain of interest; this is discussed in section 3.1. Section 3.2 derives the embedded boundary method to discretize the Fokker–Planck equation, and calculates the corresponding diffusive jump rates for the reaction-diffusion master equation. Some consequences of the specific discretization chosen are also presented. In section 3.3 the complete algorithm for solving the reaction-diffusion master equation is summarized. Convergence of the embedded boundary discretization is demonstrated for several examples in section 3.4, while section 3.5 discusses the convergence of the overall algorithm of section 3.3.

3.1. Boundary conditions. Since we are applying the overall method to simulations of biochemical networks in eukaryotic cells, which are cells with a nuclear membrane, we formulate our boundary conditions for this type of domain. Denote the entirety of the cell by Ω and the cell membrane (i.e., exterior boundary) as $\partial\Omega$. A eukaryotic cell will also have a nucleus, $\Omega_n \subset \Omega$, a closed volume contained completely within the interior of the cell. The nuclear membrane, $\partial\Omega_n$, encloses the nucleus, partitioning the cell's interior into two distinct compartments.

Passage through the cellular membrane will not be allowed, so that the exterior boundary will have a no-flux Neumann condition. This can, of course, be changed to a more general flux condition if necessary. The nuclear membrane will be impermeable to certain species, requiring again a no-flux Neumann condition, while allowing a passive flux for others. This second condition will be modeled in this paper by having the nuclear membrane flux proportional to the jump in probability density, or concentration, across the membrane. Note that this boundary condition assumes molecules

cross the nuclear membrane independently of each other. The boundary condition may be derived from a more detailed model of the nuclear membrane as containing channels, called nuclear pores, through which certain species may diffuse. The nuclear membrane is thin relative to the size of a eukaryotic cell, and hence the diffusive motion of molecules within the membrane will equilibrate rapidly. This implies that the diffusive flux within the membrane can be assumed spatially constant. In addition, the number of nuclear pores is large enough that we approximate them as a fixed background density within the membrane. Combining these two approximations, and idealizing the membrane as infinitely thin, implies that the flux through the nuclear membrane is proportional to the difference in probability density, or concentration, across the membrane. Note that in certain situations this boundary condition would need to be modified to a more physically detailed model. One such example would be if the number of molecules crossing the nuclear membrane can become large enough to saturate the nuclear pores.

Consider a single particle moving by classical Brownian motion within the cell. Let $p(\boldsymbol{x},t)$ denote the probability density for the position of the particle. Then within both the nucleus, Ω_n, and the cytoplasm, $\Omega \setminus \Omega_n$, $p(x,t)$ will satisfy (2.10). (Here $\Omega \setminus \Omega_n$ denotes the set of points within Ω that are not also contained in Ω_n.) Let $[p]_n$ denote the jump in $p(\boldsymbol{x},t)$ across the nuclear membrane. Define by ρ the nuclear membrane permeability, and by D the diffusion constant of the particle. Let $\boldsymbol{\eta}$ denote the outward pointing normal to a given surface. Then, incorporating the assumed boundary conditions, $p(\boldsymbol{x},t)$ satisfies

$$\begin{aligned} \frac{\partial p}{\partial t} &= D\Delta p \qquad && \text{in } \Omega, \\ \frac{\partial p}{\partial \boldsymbol{\eta}} &= 0 && \text{on } \partial\Omega, \\ -D\frac{\partial p}{\partial \boldsymbol{\eta}} &= -\rho\,[p]_n && \text{on } \partial\Omega_n, \end{aligned} \tag{3.1}$$

with the initial condition $p(\boldsymbol{x},0) = \delta(\boldsymbol{x}-\boldsymbol{x}_0)$, where $\boldsymbol{x}_0$ is the known initial position of the particle. A discretization of these equations that has the general form of (2.9) will determine the jump rates in the reaction-diffusion master equation.

The corresponding multiple species, deterministic reaction-diffusion model for the assumed boundary conditions follows. Let $[C^l]_n$ denote the jump in C^l across the nuclear membrane, ρ^l the nuclear membrane permeability, and D^l the diffusion constant of the lth species. Then

$$\begin{aligned} \frac{\partial C^l(\boldsymbol{x},t)}{\partial t} &= D^l\Delta C^l + \sum_{k=1}^{K} \nu_k^l\, \tilde{a}^k(\boldsymbol{C}(\boldsymbol{x},t),\boldsymbol{x}) && \text{in } \Omega, \\ \frac{\partial C^l(\boldsymbol{x},t)}{\partial \boldsymbol{\eta}} &= 0 && \text{on } \partial\Omega, \\ -D^l\frac{\partial C^l(\boldsymbol{x},t)}{\partial \boldsymbol{\eta}} &= -\rho^l\left[C^l\right]_n && \text{on } \partial\Omega_n, \end{aligned} \tag{3.2}$$

with the initial condition, $C^l(\boldsymbol{x},0) = C_0(\boldsymbol{x})$.

3.2. Numerical discretization. For simplicity, we discretize (3.1) instead of the deterministic reaction-diffusion formulation (3.2). We may do this because our purpose is not to obtain a numerical scheme for the deterministic equations but instead

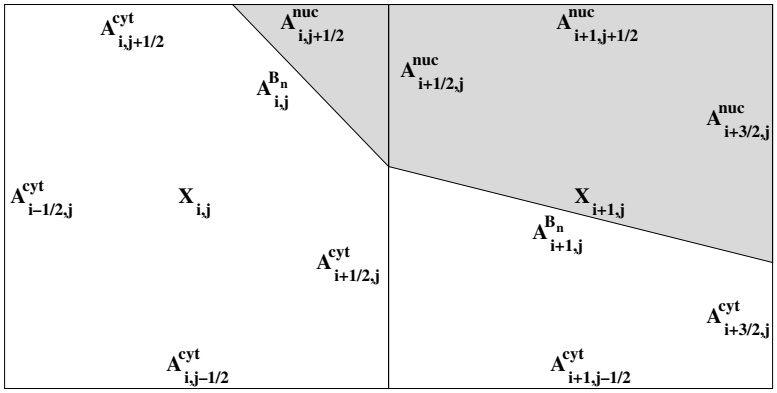


FIG. 1. *Two-dimensional Cartesian mesh cells cut by nuclear membrane. The darker region represents the portion of Ω_n, the nuclear space, within the mesh cells. The lighter region represents the portion of $\Omega \setminus \Omega_n$, the cytoplasm, within the mesh cells. $A^{\text{cyt}}_{i+1/2,j}$ gives the length of the piece of the face connecting cells (i,j) and $(i+1,j)$ that is within the cytoplasmic domain. The remaining $A^{\alpha}_{i,j}$ values, for α a given domain, are defined similarly. $A^{B_n}_{i,j}$ is the length of the piece of the nuclear membrane within the cell (i,j). $X_{i,j}$ gives the center of the Cartesian cell at location (i,j). Also, $V^{\alpha}_{i,j}$ (not shown) gives the area of the portion of cell (i,j) within the domain α.*

to obtain the jump rates for use in our stochastic simulation, and, as was shown in section 2.4, the jump rates are the same whether we use (3.1) or the full deterministic reaction-diffusion system (3.2). For the remainder of this section, we use the word "cell" to refer to a cell of the Cartesian computational mesh, and we specify "biological cell" when that is what we mean. We embed Ω in a Cartesian mesh with cell centers $\boldsymbol{x_i}$ and mesh width h. There are three different domains contained in this description: the nuclear space, Ω_n, the cytoplasm, $\Omega \setminus \Omega_n$, and the space exterior to the biological cell. The total computational domain within which the biological cell is embedded is taken to be a square.

Define three separate solutions to (3.1), corresponding to each of the three domains: p^{nuc}, the value of $p(x,t)$ within Ω_n, p^{cyt}, the value of $p(x,t)$ within $\Omega \setminus \Omega_n$, and p^{ext}, the value of p outside the biological cell. Since the particle is assumed to never leave Ω, p^{ext} will be zero. Denote by p^{α} the value of the solution for domain α, where $\alpha \in \{\text{nuc}, \text{cyt}, \text{ext}\}$. Each of these solutions is assumed to be smoothly extendable across the boundaries of its region of definition. Note that the actual solution $p(x,t)$ will not be smooth across either the nuclear or cellular membrane, but its value on either side of each of these membranes can be smoothly extended to the other side.

For spatial dimension d, let $\boldsymbol{i} \in \mathbb{Z}^d$ denote the index vector for mesh cells. With the smooth extension assumption, a cell-centered solution value, $p^{\alpha}_{\boldsymbol{i}}$, can always be defined within cells containing some portion of the domain α. For a cell completely within this domain only one solution value will be considered, $p^{\alpha}_{\boldsymbol{i}}$. If a cell is split by one of the membranes, then two cell-centered solution values will be stored for the cell. It is assumed that at most one boundary intersects a given cell. For example, in Figure 1, which shows the two-dimensional case for simplicity, cell $\boldsymbol{i} = (i,j)$ is split by the nuclear membrane; hence, there would be two solution values stored for this cell, $p^{\text{nuc}}_{\boldsymbol{i}}$ and $p^{\text{cyt}}_{\boldsymbol{i}}$. These two solution values are subsequently used to approximate the boundary condition on the membrane. They also allow, in the master equation formulation, separate probabilities of the particle being in either portion of the cell.

Again, as the particle cannot cross $\partial\Omega$, $p_{\boldsymbol{i}}^{\text{ext}} = 0$ for all $\boldsymbol{i}$. As we know $p_{\boldsymbol{i}}^{\text{ext}}$, multiple values need not be stored for cells cut by $\partial\Omega$. If the particle could cross the cellular membrane, then $p_{\boldsymbol{i}}^{\text{ext}}$ would not necessarily be zero, and multiple values would be needed in cells split by $\partial\Omega$.

Figure 1 defines several geometric quantities associated with cells that are used in the numerical discretization. Let $\boldsymbol{e}_k$ denote the unit vector along the kth coordinate axis of $\mathbb{R}^d$. Define by

$$\boldsymbol{F}(x,t) = -D\nabla p \tag{3.3}$$

the probability flux vector per unit area. Let $\boldsymbol{F}^\alpha$ be the flux vector of the domain α solution and $(F_k)^\alpha$ the kth component of $\boldsymbol{F}^\alpha$. $\boldsymbol{\eta}^\alpha$ will denote the outward pointing, with respect to domain α, normal to a given surface. Following the conservative discretization method in [21], for all cells, $\boldsymbol{i}$, in which $p_{\boldsymbol{i}}^\alpha$ is defined,

(3.4)

$$\begin{aligned}
\frac{dp_{\boldsymbol{i}}^\alpha}{dt} &= -(\nabla\cdot\boldsymbol{F}^{\boldsymbol{\alpha}})_i \\
&\approx -\frac{1}{V_{\boldsymbol{i}}^\alpha}\int_{V_{\boldsymbol{i}}^\alpha}\nabla\cdot\boldsymbol{F}^{\boldsymbol{\alpha}}\,dV = -\frac{1}{V_{\boldsymbol{i}}^\alpha}\int_{\partial V_{\boldsymbol{i}}^\alpha}\boldsymbol{F}^{\boldsymbol{\alpha}}\cdot\boldsymbol{\eta}^{\boldsymbol{\alpha}}\,dS \\
&= -\frac{1}{V_{\boldsymbol{i}}^\alpha}\left(\sum_{\pm}\sum_{k=1}^{d}\int_{A^\alpha_{\boldsymbol{i}\pm\boldsymbol{e}_k}}\boldsymbol{F}^{\boldsymbol{\alpha}}\cdot\boldsymbol{\eta}^{\boldsymbol{\alpha}},\ dS + \int_{B_n}\boldsymbol{F}^{\boldsymbol{\alpha}}\cdot\boldsymbol{\eta}^{\boldsymbol{\alpha}}\,dS + \int_{B_e}\boldsymbol{F}^{\boldsymbol{\alpha}}\cdot\boldsymbol{\eta}^{\boldsymbol{\alpha}}\,dS\right) \\
&\approx -\frac{1}{V_{\boldsymbol{i}}^\alpha}\left(\sum_{k=1}^{d}\left(A^\alpha_{\boldsymbol{i}+\frac{1}{2}\boldsymbol{e}_k}(F_k)^\alpha_{\boldsymbol{i}+\frac{1}{2}\boldsymbol{e}_k} - A^\alpha_{\boldsymbol{i}-\frac{1}{2}\boldsymbol{e}_k}(F_k)^\alpha_{\boldsymbol{i}-\frac{1}{2}\boldsymbol{e}_k}\right) + A_{\boldsymbol{i}}^{B_n}\boldsymbol{F}_{\boldsymbol{i}}^{B_n}\cdot\boldsymbol{\eta}_{\boldsymbol{i}}^{B_n}\right).
\end{aligned}$$

Here, $B_n = \partial\Omega_n \cap \partial V_{\boldsymbol{i}}^\alpha$ and $B_e = \partial\Omega \cap \partial V_{\boldsymbol{i}}^\alpha$. The surface integral over B_e is identically zero due to the no-flux boundary condition at $\partial\Omega$. If a nonzero flux boundary condition is specified instead, then the integral would contribute another term of the form

$$-\frac{1}{V_{\boldsymbol{i}}^\alpha}A_{\boldsymbol{i}}^{B_e}\boldsymbol{F}_{\boldsymbol{i}}^{B_e}\cdot\boldsymbol{\eta}_{\boldsymbol{i}}^{B_e}$$

in the last line of (3.4).

To give a valid reaction-diffusion master equation the fluxes are chosen as the standard full-face centered difference

$$(F_k)^\alpha_{\boldsymbol{i}\pm\frac{1}{2}\boldsymbol{e}_k} = \mp D\,\frac{p^\alpha_{\boldsymbol{i}\pm\boldsymbol{e}_k} - p^\alpha_{\boldsymbol{i}}}{h}.$$

From the nuclear membrane boundary condition

$$\boldsymbol{F}_{\boldsymbol{i}}^{B_n}\cdot\boldsymbol{\eta}_{\boldsymbol{i}}^{B_n} = -\rho\,[p]_{\boldsymbol{i}}^{B_n}.$$

To enforce this jump condition the cell-centered solution value for the domain on the other side of the nuclear membrane is used. For example, if $\alpha = \text{nuc}$, then $p_{\boldsymbol{i}}^{\alpha'}$ is used, where $\alpha' = \text{cyt}$. Vice versa, if $\alpha = \text{cyt}$, then $\alpha' = \text{nuc}$. Using this definition, the boundary condition is approximated by choosing

$$\rho\,[p]_{\boldsymbol{i}}^{B_n} \equiv \rho(p_{\boldsymbol{i}}^{\alpha'} - p_{\boldsymbol{i}}^\alpha).$$

With all the flux terms specified, the final discretization is then

$$\frac{dp_{\boldsymbol{i}}^{\alpha}}{dt} = \frac{D}{V_{\boldsymbol{i}}^{\alpha} h}\left(\sum_{\pm}\sum_{k=1}^{d} A^{\alpha}_{\boldsymbol{i}\pm\frac{1}{2}\boldsymbol{e}_k}\left(p^{\alpha}_{\boldsymbol{i}\pm\boldsymbol{e}_k} - p^{\alpha}_{\boldsymbol{i}}\right)\right) + \frac{A^{B_n}_{\boldsymbol{i}}\rho}{V^{\alpha}_{\boldsymbol{i}}}(p^{\alpha'}_{\boldsymbol{i}} - p^{\alpha}_{\boldsymbol{i}}). \tag{3.5}$$

For cells that do not intersect any boundary the discretization reduces to the standard five- (seven-) point discrete Laplacian in two dimensions (three dimensions).

Defining the probability of the particle being in the domain α portion of the $\boldsymbol{i}$th cell as

$$Q^{\alpha}_{\boldsymbol{i}}(t) = p^{\alpha}_{\boldsymbol{i}} V^{\alpha}_{\boldsymbol{i}} \approx \text{Prob}\{\boldsymbol{X}(t) \in V^{\alpha}_{\boldsymbol{i}} | \boldsymbol{X}(0) = \boldsymbol{x}_0\},$$

the discretization can be rewritten as

$$\frac{dQ^{\alpha}_{\boldsymbol{i}}}{dt} = \frac{D}{h}\left(\sum_{\pm}\sum_{k=1}^{d} \frac{A^{\alpha}_{\boldsymbol{i}\pm\frac{1}{2}\boldsymbol{e}_k}}{V^{\alpha}_{\boldsymbol{i}\pm\boldsymbol{e}_k}} Q^{\alpha}_{\boldsymbol{i}\pm\boldsymbol{e}_k} - \frac{A^{\alpha}_{\boldsymbol{i}\pm\boldsymbol{e}_k}}{V^{\alpha}_{\boldsymbol{i}}} Q^{\alpha}_{\boldsymbol{i}}\right) + A^{B_n}_{\boldsymbol{i}}\rho\left(\frac{Q^{\alpha'}_{\boldsymbol{i}}}{V^{\alpha'}_{\boldsymbol{i}}} - \frac{Q^{\alpha}_{\boldsymbol{i}}}{V^{\alpha}_{\boldsymbol{i}}}\right). \tag{3.6}$$

The first term in (3.6) represents diffusive motion between computational cells, while the second represents motion across the nuclear membrane within one computational cell. Let the two components of a split computational cell be represented as separate computational cells. In this form, all movement between computational cells within the same domain is by diffusive jumps, while movement between computational cells of different domains is by jumps determined by the membrane fluxes. Despite the appearance of the membrane flux terms, this equation, with an appropriate index relabeling to account for the multiple components of a split computational cell, still has the same form as (2.8). Hence, all motion is still represented as first order reactions, with jump rates given by the coefficients of the $Q^{\{\cdot\}}_{\{\cdot\}}$ terms.

Taking $A^{\alpha}_{\boldsymbol{ij}}$ to be the area of the domain α face shared by cells $\boldsymbol{i}$ and $\boldsymbol{j}$, the jump rate from the domain α component of cell $\boldsymbol{j}$ to the domain α component of cell $\boldsymbol{i}$ is

$$k^{\alpha}_{\boldsymbol{ij}} = \frac{D A^{\alpha}_{\boldsymbol{ij}}}{h V^{\alpha}_{\boldsymbol{j}}}. \tag{3.7}$$

For strictly interior cells this reduces to

$$k^{\alpha}_{\boldsymbol{ij}} = \frac{D}{h^2}.$$

The jump rate across the nuclear membrane in cell $\boldsymbol{i}$ from domain α' to domain α is

$$k^{\alpha\alpha'}_{\boldsymbol{i}} = \frac{A^{B_n}_{\boldsymbol{i}}\rho}{V^{\alpha'}_{\boldsymbol{i}}}. \tag{3.8}$$

A similar equation holds for the jump from domain α to domain α'. Note that $k^{\alpha}_{\boldsymbol{ij}}$ and $k^{\alpha}_{\boldsymbol{ji}}$ are unequal in general because $V^{\alpha}_{\boldsymbol{i}}$ may not be equal to $V^{\alpha}_{\boldsymbol{j}}$. Similarly, $k^{\alpha\alpha'}_{\boldsymbol{i}}$ and $k^{\alpha'\alpha}_{\boldsymbol{i}}$ will also differ as $V^{\alpha}_{\boldsymbol{i}}$ may not be equal to $V^{\alpha'}_{\boldsymbol{i}}$.

There is a fundamental difference in the scaling between (3.7), which scales like $1/h^2$, and (3.8), which scales like $1/h$. This difference arises because the flux for diffusive motion is proportional to the *gradient* of the density, while the flux across the

nuclear membrane is proportional to the *difference* in the density across the membrane.

A consequence of the conservative discretization form is that total probability is conserved:

$$\sum_{\alpha}\sum_{\boldsymbol{i}\in\alpha}\frac{dQ_{\boldsymbol{i}}^{\alpha}}{dt}=\sum_{\alpha}\sum_{\boldsymbol{i}\in\alpha}\frac{dp_{\boldsymbol{i}}^{\alpha}}{dt}V_{\boldsymbol{i}}^{\alpha}=0.$$

The master equation approximation will also satisfy the principle of detailed balance—the statement that at thermodynamic equilibrium the unidirectional probability flux from cell $\boldsymbol{i}$ to cell $\boldsymbol{j}$ is equal to the unidirectional probability flux from cell $\boldsymbol{j}$ to cell $\boldsymbol{i}$. We expect detailed balance to hold, as (3.1) forms a closed and isolated system due to the no-flux cellular membrane condition and passive nuclear membrane flux (no active transport). Letting $p^{eq}(x,t)=p^{eq}$ be the constant equilibrium solution to the continuous problem, $(Q^{eq})_{\boldsymbol{i}}^{\alpha}=p^{eq}V_{\boldsymbol{i}}^{\alpha}$ will be the equilibrium probability of the Brownian particle being in the domain α component of cell $\boldsymbol{i}$. For $A_{\boldsymbol{ij}}^{\alpha}$, the area of the domain α face between cells $\boldsymbol{i}$ and $\boldsymbol{j}$, the detailed balance condition is

$$\begin{aligned}
&k_{\boldsymbol{ji}}^{\alpha}(Q^{eq})_{\boldsymbol{i}}^{\alpha}=k_{\boldsymbol{ij}}^{\alpha}(Q^{eq})_{\boldsymbol{j}}^{\alpha}\\
\iff&\frac{DA_{\boldsymbol{ij}}^{\alpha}}{hV_{\boldsymbol{i}}^{\alpha}}(Q^{eq})_{\boldsymbol{i}}^{\alpha}=\frac{DA_{\boldsymbol{ij}}^{\alpha}}{hV_{\boldsymbol{j}}^{\alpha}}(Q^{eq})_{\boldsymbol{j}}^{\alpha}\\
\iff&\frac{(Q^{eq})_{\boldsymbol{i}}^{\alpha}}{V_{\boldsymbol{i}}^{\alpha}}=\frac{(Q^{eq})_{\boldsymbol{j}}^{\alpha}}{V_{\boldsymbol{j}}^{\alpha}},
\end{aligned}$$

which holds by the definitions of $(Q^{eq})_{\boldsymbol{i}}^{\alpha}$ and $(Q^{eq})_{\boldsymbol{j}}^{\alpha}$. For jumps across a membrane between domains α and α',

$$\begin{aligned}
&k_{\boldsymbol{i}}^{\alpha'\alpha}(Q^{eq})_{\boldsymbol{i}}^{\alpha}=k_{\boldsymbol{i}}^{\alpha\alpha'}(Q^{eq})_{\boldsymbol{i}}^{\alpha'}\\
\iff&\frac{A_{\boldsymbol{i}}^{B_n}\rho}{V_{\boldsymbol{i}}^{\alpha}}(Q^{eq})_{\boldsymbol{i}}^{\alpha}=\frac{A_{\boldsymbol{i}}^{B_n}\rho}{V_{\boldsymbol{i}}^{\alpha'}}(Q^{eq})_{\boldsymbol{i}}^{\alpha'}\\
\iff&\frac{(Q^{eq})_{\boldsymbol{i}}^{\alpha}}{V_{\boldsymbol{i}}^{\alpha}}=\frac{(Q^{eq})_{\boldsymbol{i}}^{\alpha'}}{V_{\boldsymbol{i}}^{\alpha'}},
\end{aligned}$$

which again holds by the definitions of $(Q^{eq})_{\boldsymbol{i}}^{\alpha}$ and $(Q^{eq})_{\boldsymbol{i}}^{\alpha'}$. Note that an active transport boundary condition at the cellular membrane could make the system open, in which case detailed balance would no longer be expected to hold. Moreover, an active transport mechanism at the nuclear membrane would also prevent the system from coming to thermodynamic equilibrium and would result in a steady state in which the above relationships across the nuclear membrane would be violated. This would not affect the rate constants for other (passive) processes, however, which would therefore still satisfy the conditions derived from the principle of detailed balance.

For comparison to our discretization, note that in [21] $(F_k)_{\boldsymbol{i}\pm\frac{1}{2}\boldsymbol{e}_k}$ is chosen to be the standard centered difference approximation when the kth face does not intersect a boundary. This leads to the five- (seven-) point Laplacian as the discretization for strictly interior cells. For faces cut by the boundary, [21] interpolates between the centered difference flux in neighboring cells to approximate the flux at the midpoint of the cut face. This method leads to a second order accurate approximation to the Laplacian, but does not give a valid master equation for the continuous time–discrete space discretization. To see why, note that the general form of the master equation

for the probability, Q_i, of being at site i with transition probability per unit time $W_{ji} \geqslant 0$ is given by (2.2). In general interpolation will introduce negative weights W_{ji} or W_{ij} into the ith equation. Furthermore, interpolation violates the condition that if the ith equation contains a term $W_{ij}Q_j$, then the jth equation should contain a term $-W_{ij}Q_j$.

An alternative discretization method is presented in [13] that accounts for the boundary by using the standard five- (seven-) point Laplacian at all locations and adding an extra forcing term. As the master equation has no forcing terms this method would not give an equation that could be realized by the Gillespie method.

3.3. Overall method. Using the discretization from the previous section, realizations of the stochastic process described by the reaction-diffusion master equation can be created using the Gillespie method. The overall simulation algorithm is as follows:

1. Initialization:
   (a) Given the membrane locations, calculate $A^{\alpha}_{\boldsymbol{i}\pm\boldsymbol{e}_k}$, $V^{\alpha}_{\boldsymbol{i}}$, and $A^{B_n}_{\boldsymbol{i}}$ for each location, $\boldsymbol{i}$, each direction, $\boldsymbol{e}_k$, and each domain, α.
   (b) From (3.7) and (3.8) calculate the jump rates for all species, within all Cartesian cells containing some part of Ω.
   (c) For each piece of a Cartesian cell, calculate the rates of all chemical reactions that can occur there. For reactions with volume dependent rates use $V^{\alpha}_{\boldsymbol{i}}$ to change the rate constants to units of reciprocal time.
2. Time Evolution:
   (a) Simulate individual realizations of the stochastic process described by the reaction-diffusion master equation using the Gillespie method. Diffusive and transmembrane solute motion are represented as first order reactions, using the jump rates calculated in step 1(b). Within each component of a cell, the rates from step 1(c) are used to simulate chemical reactions.
3. Output:
   (a) To estimate moments or distributions, use statistics from many simulations.

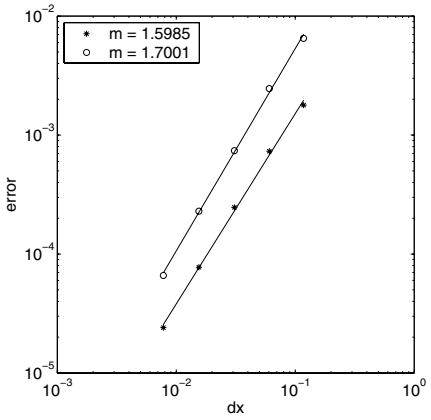
3.4. Convergence of the numerical discretization. The convergence of the discretization provided by (3.5) is illustrated in two dimensions for both the Poisson equation and diffusion equation with circular boundaries. The method was found to converge between first and second order spatially. Let $(\boldsymbol{u}_h)^{\alpha}_{\boldsymbol{i}}$ denote the numerical solution to the Poisson problem, and $\boldsymbol{u}_h$ the vector whose components are given by $(\boldsymbol{u}_h)^{\alpha}_{\boldsymbol{i}}$. Define the maximum norm, $\|\boldsymbol{u}_h\|_{\infty}$, as

$$\|\boldsymbol{u}_h\|_{\infty} = \max_{\alpha,\boldsymbol{i}\in\alpha} |(\boldsymbol{u}_h)^{\alpha}_{\boldsymbol{i}}|.$$

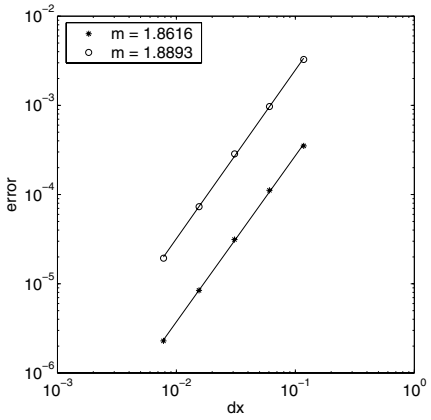
A volume weighted two norm, $\|\boldsymbol{u}_h\|^w_2$, is defined as

$$\|\boldsymbol{u}_h\|^w_2 = \left(\sum_{\alpha} \sum_{\boldsymbol{i}\in\alpha} \left((\boldsymbol{u}_h)^{\alpha}_{\boldsymbol{i}}\right)^2 V^{\alpha}_{\boldsymbol{i}} \right)^{\frac{1}{2}}.$$

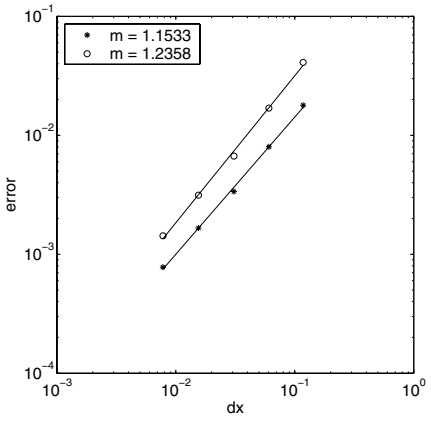
Let $(\boldsymbol{u}^n_h)^{\alpha}_{\boldsymbol{i}}$ denote the numerical solution to the diffusion equation at time t_n, $\boldsymbol{u}^n_h$ the solution vector at time t_n, and $\boldsymbol{u}_h$ the solution over all spatial locations, domains, and times. The maximum norm over all time is defined as $\max_n \|\boldsymbol{u}^n_h\|_{\infty}$. As in section 3.1,



(a) Error for (3.9), with exact solution (3.10).



(b) Error for (3.9), with exact solution (3.11).



(c) Error for (3.12), with exact solution (3.13).

FIG. 2. *Convergence results for Poisson equation solutions; star is two norm error and circle is maximum norm error. These errors are plotted against the mesh width dx on log-log plots. The empirical order of accuracy is m, which is the slope of the best fit straight line in each case.*

$\partial\Omega$ will denote the cell membrane, or outer boundary, while $\partial\Omega_n$ will denote the nuclear membrane, or inner boundary. In all of the examples considered in this section, the two boundaries are concentric circles with radii 1 and 1/2, respectively. The domain into which the circles are embedded is taken, for a given mesh width h, to be the square centered at the origin with sides of length $2 + 4h$. This provides a separation between the boundary of the computational domain and the embedded boundaries.

Figure 2(a) shows the convergence of the discretization for the Poisson equation

with just the outer boundary. The equations are

$$\begin{aligned} \Delta u &= f(\boldsymbol{x}) \quad \text{in } \Omega, \\ \frac{\partial u}{\partial \boldsymbol{\eta}} &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{3.9}$$

where $f(\boldsymbol{x})$ was chosen so that the exact solution was

$$u = \frac{r^4}{4} - \frac{r^3}{3}, \quad r \le 1, \tag{3.10}$$

with $u = 0$ for $r > 1$. The two norm error converged at order 1.6, while the maximum norm error converged at order 1.7.

An angular dependence was added to $f(\boldsymbol{x})$ to give the exact solution

$$u = \left(\frac{r^4}{4} - \frac{r^3}{3}\right)\cos\theta, \quad r \le 1, \tag{3.11}$$

with $u = 0$ for $r > 1$. Figure 2(b) shows the convergence of the errors for this problem. The θ dependence actually improves the convergence, so that the two norm error is order 1.86, while the maximum norm error now converges at order 1.89.

Figure 2(c) shows the convergence of the discretization for the Poisson equation with both boundaries. The equations are

$$\begin{aligned} \Delta u &= f(\boldsymbol{x}) \qquad \text{in } \Omega, \\ \frac{\partial u}{\partial \boldsymbol{\eta}} &= 0 \qquad \text{on } \partial\Omega, \\ -\frac{\partial u}{\partial \boldsymbol{\eta}} &= -\pi\,[u]_n \quad \text{on } \partial\Omega_n, \end{aligned} \tag{3.12}$$

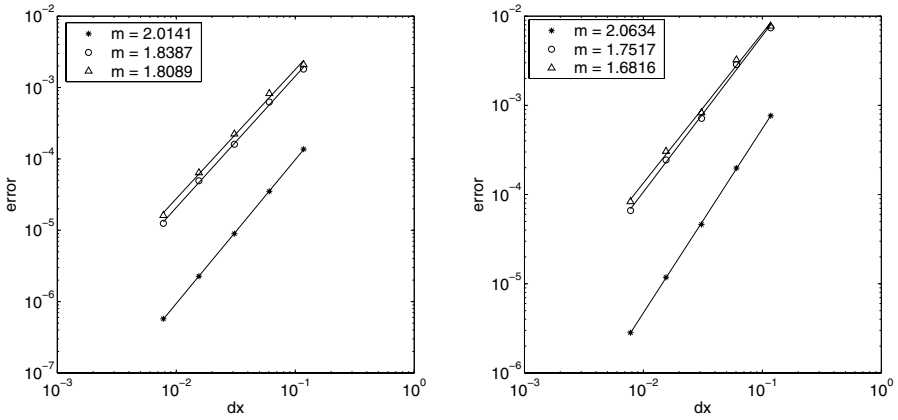
where $f(\boldsymbol{x})$ was chosen so that the exact solution was

$$u = \begin{cases} \cos\pi r + \sin\pi r & \text{in } \Omega_n, \\ \cos\pi r & \text{in } \Omega \setminus \Omega_n, \\ 0 & \text{outside } \Omega. \end{cases} \tag{3.13}$$

The jump condition at the interior boundary reduces the order of convergence to 1.15 in the two norm and 1.23 in the maximum norm.

Figure 3(a) shows the convergence of the discretization for the diffusion equation with just the exterior boundary. The final time, at which the two norm and maximum norm errors were measured, was taken to be $t_f = .025$. At this time the solution was still far from equilibrium. $\Delta t/h$ was fixed at about t_f/L, where L was the diameter of the outer boundary. For the examples in Figure 3, $\Delta t/h \approx .013$. The time integration was performed using the second order, L_0 stable, implicit Runge–Kutta method mentioned in [21]. Similar results were obtained for a Crank–Nicolson discretization. The equations are

$$\begin{aligned} \frac{\partial u}{\partial t} &= \Delta u + f(\boldsymbol{x}, t) \quad \text{in } \Omega, \\ \frac{\partial u}{\partial \boldsymbol{\eta}} &= 0 \qquad\qquad\quad \text{on } \partial\Omega, \end{aligned} \tag{3.14}$$



(a) Error for (3.14), with exact solution (3.15).

(b) Error for (3.16), with exact solution (3.17).

FIG. 3. *Convergence results for diffusion equation solutions; star is two norm error at t_f, circle is the maximum norm error at t_f, and diamond is the maximum of the maximum norm error over all time. These errors are plotted against the mesh width dx on log-log plots. The empirical order of accuracy is m, which is the slope of the best fit straight line in each case.*

where $f(\boldsymbol{x}, t)$ was chosen so that the exact solution was

$$u = \left(\frac{r^4}{4} - \frac{r^3}{3}\right) e^{-20t}, \quad r \leq 1, \tag{3.15}$$

with $u = 0$ for $r > 1$. At t_f, the two norm error converged at about order 2.01, and the maximum norm error converged at about order 1.84. The maximum over all time points of the maximum norm converged at about order 1.81.

Figure 3(b) shows the convergence of the discretization for the diffusion equation with both boundaries. Note that the same time discretization was used as above, again with $\Delta t/h \approx .013$. The equations are

$$\begin{aligned} \frac{\partial u}{\partial t} &= \Delta u + f(\boldsymbol{x}, t) && \text{in } \Omega, \\ \frac{\partial u}{\partial \boldsymbol{\eta}} &= 0 && \text{on } \partial\Omega, \\ -\frac{\partial u}{\partial \boldsymbol{\eta}} &= -\left[u\right]_n && \text{on } \partial\Omega_n, \end{aligned} \tag{3.16}$$

where $f(\boldsymbol{x}, t)$ was chosen so that the exact solution was

$$u = \begin{cases} \left(\frac{br^4}{4} + \frac{ar^3}{3} + \frac{ar^5}{5}\right) e^{-20t} & \text{in } \Omega_n, \\ \left(\frac{r^3}{3} - \frac{r^4}{4}\right) e^{-20t} & \text{in } \Omega \setminus \Omega_n, \\ 0 & \text{outside } \Omega. \end{cases} \tag{3.17}$$

a and b were chosen to satisfy the jump boundary condition at $r = 1/2$. At t_f the two norm error converged at about order 2.06, and the maximum norm error converged at

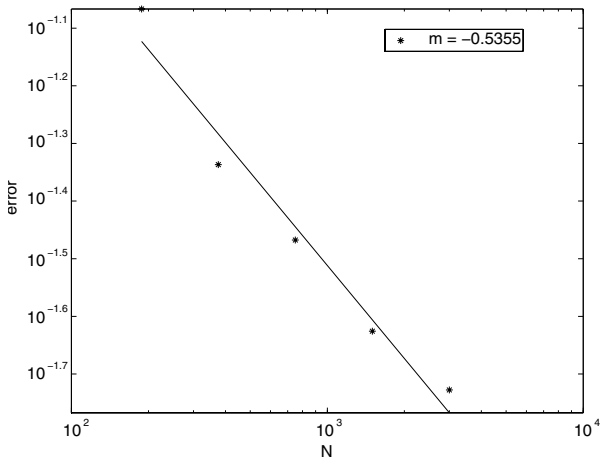


FIG. 4. *Convergence of the mean of Gillespie simulations to the numerical solution to* (3.1). *Here the error is given by the maximum absolute difference over all times and spatial locations between the Gillespie simulation estimate and the numerical solution. A* 37 *by* 37 *mesh was used with mesh width* 2/33, $D = 1$, *and* $\rho = 1$. *The particle was started at the center of the domain for all simulations. m gives the slope of the best fit line to the data.*

about order 1.75. The maximum over all time points of the maximum norm converged at about order 1.68.

3.5. Convergence of the overall method. The overall method presented in section 3.3 has two sources of error associated with it: sampling error and error due to the spatial discretization. The Gillespie method provides exact realizations of the stochastic process defined by the reaction-diffusion master equation, hence the error in any individual realization is due solely to the spatial discretization. From the central limit theorem, for a fixed mesh size, we expect the error between the mean population levels from N simulations and the means of the reaction-diffusion master equation to be approximately $O(1/\sqrt{N})$. Note that the constant in the $O(1/\sqrt{N})$ term will depend on the mesh size. With a sufficient number of samples the joint probability distribution, for a fixed mesh size, can be estimated to any desired accuracy, but the required number of samples might be very large.

As an example, consider a single particle moving within a two-dimensional biological cell. Assume the cell has a circular cellular membrane of radius 1, and a circular nuclear membrane of radius 1/2. The particle's dynamics are assumed to be given by (3.1), with $D = 1$ and $\rho = 1$. Since there is only one particle, the mean number of the particle in a given region is just the probability of the particle being in that region. As the reaction-diffusion master equation is in terms of the population levels of the particle at different locations, its mean should converge to the solution of (3.1) as h goes to zero. Figure 4 shows the maximum error over all times between the mean of N Gillespie simulations and the numerical solution to (3.1), as the number of samples is increased. Here h is fixed at 2/33, for a 37 by 37 mesh. Notice that the error decreases like $1/\sqrt{N}$.

Having obtained the distribution given by the reaction-diffusion master equation, there is still the question of how accurate a representation of the dynamics of the system it gives. It is not clear that the reaction-diffusion master equation has a well-defined continuum limit as h goes to zero. Instead, it is generally considered valid

only for a range of h values, and there are several simultaneous physical conditions which h should satisfy. First, h should be significantly larger than the mean free path, λ, between the elastic collisions driving diffusion. This condition ensures that the system can be considered in local equilibrium within a computational cell because of many nonreactive collisions. Note, however, that to ensure that the system is accurately resolving local behavior, and also that the approximation of diffusion as jumps between cells is reasonable, h should be significantly smaller than the length scale of the entire system, L. Summarizing these constraints,

$$L \gg h \gg \lambda.$$

As we now show, h also needs to be chosen large enough (!) that the system can be considered in local equilibrium and well mixed within each mesh cell on the time scale of the fastest bimolecular reactions. This assumption underlies the use of stochastic chemical kinetics independently within each mesh cell. For this to hold, the time scale for the particle to diffuse throughout a mesh cell should be much faster than the time scale for the fastest bimolecular chemical reaction. The time scale to diffuse throughout a region of length scale h is approximately h^2/D. For a bimolecular reaction, the time scale with rate constant k is h^3/k for a mesh cell of volume h^3. (Recall that the units of the rate constant of a bimolecular reaction are volume per unit time.) Hence, it is necessary that

$$\frac{h^3}{k} \gg \frac{h^2}{D},$$

which implies that

$$h \gg \frac{k}{D}.$$

Thus, the rate constant for a bimolecular reaction places a fundamental lower limit on the spatial size of mesh cells. Fortunately, for cellular processes and the molecules involved in them, k/D is generally not large enough to impose a significant restriction on h. In summary, letting k denote the rate constant of the fastest bimolecular reaction,

$$L \gg h \gg \max\left(\lambda, \frac{k}{D}\right).$$

In these inequalities D is the diffusion coefficient of a particular molecular species and k refers to the fastest bimolecular reaction in which that species participates. Such inequalities must hold simultaneously for all species under consideration. Note that a first order reaction does not restrict the mesh size in any manner since it generally represents an internal molecular event, and is not dependent on the system being well mixed locally.

As of yet, the reaction-diffusion master equation has not been derived from more microscopic physical models. Its validity has been verified numerically in [5] through comparison with microscopic hard sphere models that track individual particle positions and momenta. One might also hope that, for sufficiently fast diffusion, the results from the reaction-diffusion master equation applied to an initially well-mixed system would agree over a range of h values with the spatially homogeneous chemical master equation. For example, consider the simple chemical reaction $A + B \to C$,

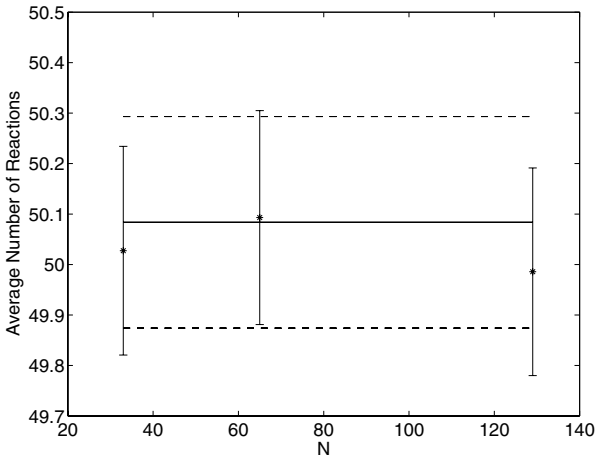


FIG. 5. *Average number of reactions in stochastic diffusion simulations, given by stars, compared to the exact solution for the expected number of reactions of the spatially homogeneous chemical master equation, given by the solid straight line. Error bars give* 99.7 *percent confidence intervals about the data points. The error bars are calculated using the sampled variance. N gives the number of mesh points in each direction. For each data point,* 3000 *sample realizations were used to calculate the average number of reactions. The dashed lines give the predicted* 99.7 *percent confidence interval, for* 3000 *samples, about the exact mean using the exact variance of the well-mixed master equation. Data is from* $t = .1$*, a time at which approximately half the A and half the B chemicals have been converted to C. Note the drastically expanded vertical scale.*

within a 1 unit by 1 unit square. The exact solution, at any time, for the average number of reactions in the spatially homogeneous chemical master equation is given in [18]. This reduces the statistical error in the problem to that in estimating the average number of reactions from the reaction-diffusion master equation. 100 molecules of A and 100 of B are each started randomly distributed throughout the domain. No-flux boundary conditions were assumed for the diffusion of chemicals, but periodic boundary conditions were also tested with no significant impact on the results. With bimolecular rate constant, .0001, and a diffusion constant of 100 for each species, $k/D = .000001$. Hence, by the preceding arguments h should only need to be greater than .000001 for the reaction-diffusion master equation to be valid. Figure 5 shows that as the number of mesh points, N, is increased, the average number of reactions from spatial simulations agrees with the exact number of reactions given by the spatially homogeneous chemical master equation to statistical error.

4. Transcription, translation, transport model. As an application of the method presented in section 3.3 we present a model of transcription, nuclear export, translation, and nuclear import in eukaryotes. While the model is simplified, and uses transcription and translation models that are more appropriate for a prokaryotes, it demonstrates the feasibility of the method in simulating biological networks. Section 4.1 describes the basic molecular cell biology necessary to understand the biochemical reaction model of section 4.2. Section 4.3 shows the time evolution of one realization of the model.

4.1. Background biology. The process by which a functional protein is produced from a segment of DNA that codes for it, called a gene, is a complex series of chemical events. In a coarsest description, the gene is first *transcribed* to produce

a copy of itself called an mRNA. The mRNA is a linear sequence of nucleotides, every three of which comprise an element of a code representing an amino acid. The mRNA is then *translated* by special cellular machinery whereby the linear sequence of nucleotides is read, and the corresponding sequence of amino acids that the mRNA codes for is assembled. The polypeptide chain of amino acids typically undergoes further chemical and structural modifications to then form a functional protein.

For eukaryotes this situation is more complicated: transcription occurs in the nucleus and translation occurs in the cytoplasm of the cell. Between transcription and translation there is then an additional step that is required involving the movement of the mRNA out of the nucleus. This nuclear export process requires a series of chemical events; it cannot occur simply through diffusion of the mRNA. Similarly, those proteins that influence gene expression must be imported into the nucleus to be effective.

Large protein complexes form pores in the nuclear membrane through which all transmembrane traffic flows. These nuclear pore complexes are normally only wide enough for small molecules of less than 9 nanometers to pass through by free diffusion. Large molecules, such as mRNAs and most proteins, require the assistance of nuclear export and import receptors, which allow the pores effectively to dilate so that molecules of up to 26 nanometers can pass [1].

We shall be interested in the export of newly transcribed mRNA from the nucleus, and the import of the resulting protein product back into the nucleus. These two pathways are believed to be distinct in the nuclear receptors used, and in the biasing factors that confer directionality to the process (so that, for example, newly exported mRNA does not get reimported into the nucleus). Since the general mRNA export pathway is still unsettled, we shall use the same pathway for modeling mRNA export and protein import. This pathway is driven by a nuclear membrane RanGTP gradient that confers directionality to the movement of cargo across the membrane. A subset of the full RanGTP cycle model presented in [24] is used to account for the movement of receptors and their cargo. The RanGTP nuclear transport process is sometimes referred to as "active" because there is an energy expenditure in maintaining the gradient of RanGTP across the nuclear membrane. In contrast, the actual movement of cargo across the nuclear membrane is believed to be passive. Note that this pathway is used for the export of some mRNAs, for example, the export of incompletely spliced HIV mRNA encoding viral structural proteins [7].

The basic model presented in the next section is built on the pathway of transcription of one gene to produce mRNA, nuclear export of the mRNA, translation of the mRNA within the cytoplasm, and import of the resulting protein product back into the nucleus. Protein import plays an important role in many transcriptional networks, allowing the protein product of one gene to bind and regulate the expression of another. Several important steps in the transcription-translation cycle for eukaryotes are not modeled, including splicing, the opening of the chromatin to allow physical access to the gene, and the assembly of the transcription initiation complex.

4.2. Chemical model. The eukaryotic cell is modeled in two dimensions as two concentric circles, representing the plasma and nuclear membranes. The plasma membrane radius is taken to be 11.81 μm, while the nuclear membrane radius is 5 μm. The transcription and translation models used are based on those presented in [8]. Specifically, the gene of interest has several states. DNA denotes that the gene is free of RNA polymerase II, hereafter RNAP, the enzyme which reads the nucleotides comprising the gene and then incorporates the corresponding nucleoside triphospate

into the growing mRNA transcript. DNA^0 will denote that the RNAP is bound to the gene's promoter and ready to begin transcription. DNA^l will denote that the first l of the nucleotides forming the gene have been read and incorporated into the mRNA, with M giving the total number of nucleotides in the gene. Finally, the number of mRNA molecules within spatial computational cell i is given by $mRNA_i$. For simplicity, the gene is assumed to be localized in the center of the nucleus, which will have index i_{center}. The concentration of both RNAP and the nucleoside triphospates that are added to the mRNA transcript are assumed fixed so that transitions between the different DNA states are first order reactions. With these definitions, the transcriptional reactions, all defined solely at the cell center, are then

$$\begin{aligned} DNA &\to DNA^0, \\ DNA^{l-1} &\to DNA^l, \quad l = 1, \ldots, M-1, \\ DNA^{L-1} &\to mRNA_{i_{\text{center}}} + DNA. \end{aligned}$$

The rates for these reactions are, respectively, $.05\,\text{s}^{-1}$, $30\,\text{s}^{-1}$, and $30\,\text{s}^{-1}$. The number of nucleotides in the gene, M, is chosen to be 999.

Newly transcribed mRNA is assumed to diffuse freely throughout the nucleus until entering the export pathway. The mRNA export model is based on the RanGTP export system. Denote by NR the nuclear export receptor to which mRNA can bind; by Rt, RanGTP; by Rd, RanGDP; by Rb, RanBP1; and by NR–Rt the nuclear receptor complexed with RanGTP. All five are assumed to be at steady state concentrations, and uniformly distributed throughout the nucleus and cytoplasm. Note that the steady state concentrations are not assumed to be the same within the nucleus and cytoplasm, simply within each individually. Steady state concentrations and subsequent reaction rates were based on the data in [24]. The export process consists of three reactions, beginning with the binding of the export receptor-RanGTP complex to nuclear mRNA. Once this has occurred, the nuclear membrane is assumed to be permeable to the complex. At all locations within the cytoplasm the complex can then bind RanBP1, which subsequently induces the release of the mRNA from the complex. The reactions are

$$\begin{aligned} mRNA_i + NR\text{–}Rt_i &\to mRNA\text{–}NR\text{–}Rt_i && \text{within nucleus,} \\ mRNA\text{–}NR\text{–}Rt_i + Rb_i &\rightleftarrows mRNA\text{–}NR\text{–}Rt\text{–}Rb_i && \text{within cytoplasm,} \\ mRNA\text{–}NR\text{–}Rt\text{–}Rb_i &\to mRNA_i + NR_i + Rb_i + Rd_i && \text{within cytoplasm.} \end{aligned}$$

Using the assumption of constant concentrations, we can reduce these reactions to the following four:

$$\begin{aligned} mRNA_i &\to mRNA\text{–}NR\text{–}Rt_i && \text{within nucleus,} \\ mRNA\text{–}NR\text{–}Rt_i &\rightleftarrows mRNA\text{–}NR\text{–}Rt\text{–}Rb_i && \text{within cytoplasm,} \\ mRNA\text{–}NR\text{–}Rt\text{–}Rb_i &\to mRNA_i && \text{within cytoplasm.} \end{aligned}$$

The rates for these reactions are, respectively, $1184.5\,\text{s}^{-1}$, $298.93\,\text{s}^{-1}$ for the forward, $.5\,\text{s}^{-1}$ for the reverse, and $51.19\,\text{s}^{-1}$.

Once the mRNA is free within the cytoplasm, it can then associate with ribosomes, the molecular machines on which translation occurs. The ribosomes are assumed to be uniformly distributed throughout the cytoplasm as a constant background concentration. The translation process begins with the binding of an mRNA

to a ribosome to form a complex that is ready to translate, denoted by $mRNA_i^0$. As the ribosomes are very large, it is assumed that the translation process is immobile. Translation then consists of a series of elongation steps by which the mRNA is moved through the ribosome and read, and each set of three nucleotides is translated into the appropriate amino acid. This amino acid is then added to the growing polypeptide that will become the protein. Let $mRNA^l$ denote that l amino acids have been incorporated into the polypeptide chain, $l = 1, \ldots, N$, where $N = M/3$. Denote by P protein. Assuming all amino acid concentrations are spatially homogeneous and at steady state, the translation process reduces to the following set of reactions:

$$\begin{aligned} mRNA_i &\to mRNA_i^0 && \text{within cytoplasm,} \\ mRNA_i^{l-1} &\to mRNA_i^l, \quad l = 1, \ldots, N-1, && \text{within cytoplasm,} \\ mRNA_i^{N-1} &\to P_i + mRNA_i && \text{within cytoplasm.} \end{aligned}$$

The rates for these reactions are, respectively, $.5\,\mathrm{s}^{-1}$, $33\,\mathrm{s}^{-1}$, and $33\,\mathrm{s}^{-1}$.

Protein may freely diffuse within both the nucleus and the cytoplasm; however, it may not freely cross the nuclear membrane. Instead, it uses the RanGTP nuclear import process to enter the nucleus. This process consists of two reactions: protein binding to nuclear import receptor within the cytoplasm, and unbinding of protein from the receptor upon binding of RanGTP within the nucleus. The reactions are then

$$\begin{aligned} P_i + NR_i &\to P\text{–}NR_i && \text{within cytoplasm,} \\ P\text{–}NR_i + Rt_i &\to P_i + NR\text{–}Rt_i && \text{within nucleus.} \end{aligned}$$

Using the constant concentration assumption, these reactions can be reduced to the following first order reactions:

$$\begin{aligned} P_i &\to P\text{–}NR_i && \text{within cytoplasm,} \\ P\text{–}NR_i &\to P_i && \text{within nucleus.} \end{aligned}$$

The rates for these reactions are, respectively, $1218.25\,\mathrm{s}^{-1}$ and $1.23\,\mathrm{s}^{-1}$.

A feedback mechanism is incorporated into the model by allowing the protein to repress the expression of its own gene. Denoting by DNA_{rep} the repressed state of the gene, the reactions are

$$DNA + P_{i_{center}} \rightleftarrows DNA_{\mathrm{rep}}.$$

Here the forward reaction rate is $.01\,\mu\mathrm{m}^3\mathrm{s}^{-1}$ and the reverse rate is $.01\,\mathrm{s}^{-1}$.

Finally, mRNA is assumed to be degradable within the cytoplasm with rate $.2\,\mathrm{s}^{-1}$, and protein anywhere with rate $.0025\,\mathrm{s}^{-1}$. Table 4.1 gives the diffusion and nuclear membrane permeability rates for the different species in the model. Note that because of the finite diffusion coefficients, this model will give different statistics than the corresponding well-mixed model. For example, in the nuclear export model the time for an mRNA-nuclear export receptor complex to diffuse throughout the nucleus is approximately

$$\frac{r_{\mathrm{nuc}}^2}{D} = 250 \text{ s}.$$

TABLE 4.1
Diffusion constant, D, and nuclear membrane permeability, ρ, for each species.

Species	D $\mu\text{m}^2\,\text{s}^{-1}$	ρ $\mu\text{m}\,\text{s}^{-1}$
$mRNA$	.1	0
$mRNA^l$	0	0
$mRNA$–NR–Rt	.1	.17
$mRNA$–NR–Rt–Rb	.1	0
P	20.0	0
P–NR	20.0	1.87

In contrast, the time scale for that complex to leave the nucleus in the well-mixed model would be approximately

$$\frac{\frac{4}{3}\pi r_{\text{nuc}}^3}{4\pi r_{\text{nuc}}^2\,\rho} = 9.8 \text{ s}.$$

With the assumed biological constants the two step export process, finding the nuclear membrane and then passing through it, is strongly diffusion limited. Therefore, a well-mixed model with infinitely fast diffusion would ignore the long time scale for an export complex to *find* the nuclear membrane, much less to get through it. This would result in a substantially faster export process than biologically occurs.

4.3. Numerical results. Figures 6 and 7 show the time evolution of one realization of the model described in the previous section, using an underlying 37 by 37 Cartesian mesh with a mesh width of approximately .72 microns. Within the simulations we calculate the number of molecules of each chemical species within each mesh cell. We assume that *within* each individual mesh cell particles are well mixed, and as such the marker for every molecule in each figure is placed randomly within the mesh cell containing that molecule. Initially there are no mRNAs or proteins within the system, and the DNA is in the unbound state. Figure 6 shows the evolution of the system over several minutes. By Figure 6(a) the first mRNA has been fully transcribed and, because of the fast binding rate, is bound to a nuclear export receptor. The DNA is unbound, waiting for the next transcription cycle to begin. After 160 seconds, a second mRNA has been transcribed and bound with a nuclear receptor. The DNA is again unbound. By 310 seconds several proteins have been translated. One is diffusing within the cytoplasm bound to a nuclear receptor, while the other has already undergone the nuclear import process and is diffusing freely within the nucleus. Notice there is only one visible mRNA, as the second is undergoing translation within the cytoplasm and hence not displayed. The DNA is also not visible as it is undergoing transcription. Finally, after 380 seconds, several proteins have accumulated within the nucleus, with one binding to and inhibiting the DNA. A free mRNA is diffusing within the cytoplasm, along with a nuclear receptor coupled protein.

By allowing the protein to feed back and inhibit the transcription of its gene, we create a system in which protein production occurs in bursts. Figure 7 shows that a large amount of protein is built up in the nucleus by 1160 seconds, and that the DNA is in the repressed state. The higher protein levels increase the probability of the gene becoming repressed, and because the on rate for binding is sufficiently fast, and the off rate sufficiently slow, by 1670 seconds the overall protein population has shrunk from 10 to just 2. This is due in part to the fast decay rates of the mRNA. As the DNA is repressed, no mRNAs are transcribed to replace those that quickly decay, and hence the protein population simply decays away over time. Once the protein

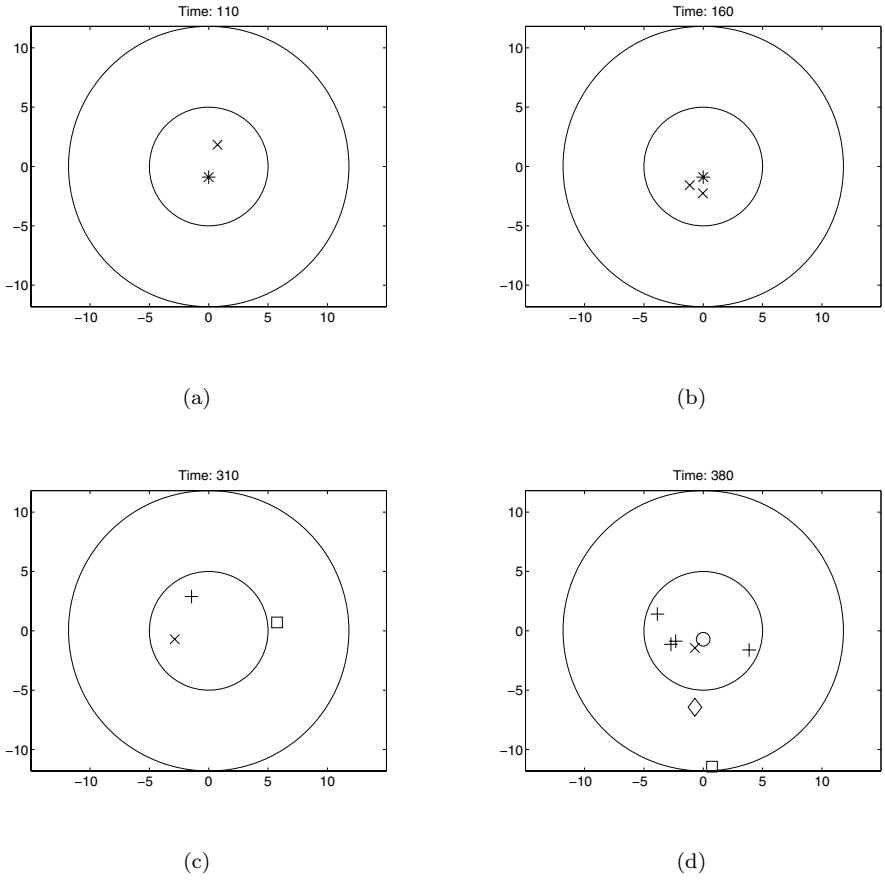


FIG. 6. *Evolution of one realization of the model from section* 4.2 *over several minutes. A star denotes the unbound DNA and a circle that the DNA is repressed. During transcriptional states the DNA is not displayed. A diamond denotes mRNA, and "×'s" denote mRNA bound to nuclear receptor and RanGTP. mRNAs coupled to nuclear receptors, RanGTP, and RanBP*1 *are not present in the images shown. During translation mRNAs are not displayed. A plus denotes a protein, and a box represents protein bound to nuclear receptor. Time is in seconds.*

population becomes sufficiently low, as in Figure 7(b), it becomes more unlikely that protein will find the DNA to bind and repress it, and hence the DNA is able to enter the transcription cycle. By 2010 seconds at least one mRNA has been transcribed and subsequently translated several times, as the nuclear protein level has increased. The production of mRNA followed by its subsequent translation produces new proteins, and consequently the nuclear protein population rises thereafter. By 2080 seconds, 10 protein molecules are within the nucleus.

Figure 8 shows the total number of nuclear proteins in one realization, over 30,000 seconds. The total number is given by the amount of free, receptor-bound, and DNA-bound proteins within the nucleus at a fixed time. Note that the number of proteins tends to quickly jump up to a given amount, and then more slowly decay to either one or zero proteins. When only one protein is present for large periods of time, it is usually bound to the DNA, thereby repressing the gene (data not shown).

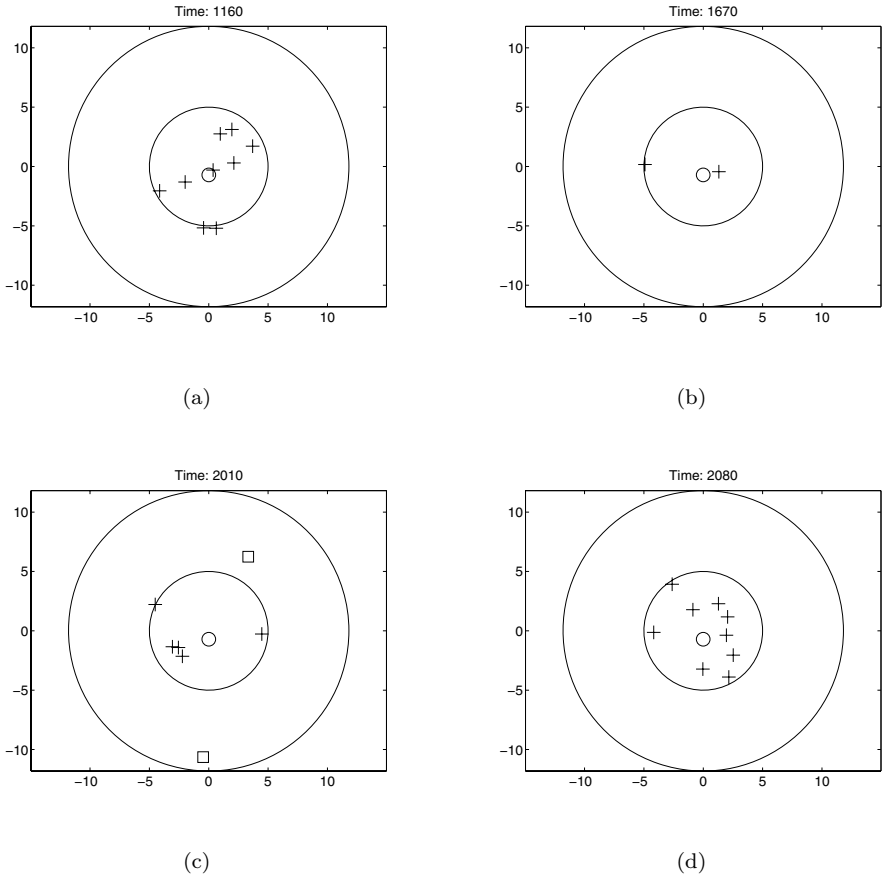


FIG. 7. *Evolution of one realization of the model from section* 4.2 *over a half hour. Symbols have the same codes as in Figure* 6.

5. Conclusions and future work. We have presented a method for calculating diffusive jump rates for the reaction-diffusion master equation within complex geometries. The method allows for multiple domains, and easily handles both jump and Neumann boundary conditions. The underlying discretization on which the method is based is found to converge between first and second order accuracy on both the Poisson equation and the diffusion equation. The discretization also conserves probability, and gives jump rates that satisfy the principle of detailed balance for diffusive motion and for passive boundary fluxes. With only one particle and no chemical reactions the method converges, as the mesh width goes to zero, to the solution to the Fokker–Planck equation for a classical Brownian particle. Further, when the system contains only first order reactions, the mean concentrations given by the method converge to the solution of the reaction-diffusion equations associated with the deterministic model of the chemical system. If the system contains bimolecular or higher order reactions, however, no such agreement with the solution of deterministic reaction-diffusion equations would be expected (except in the limit of large numbers of molecules), since averaging does not commute with multiplication. In such cases,

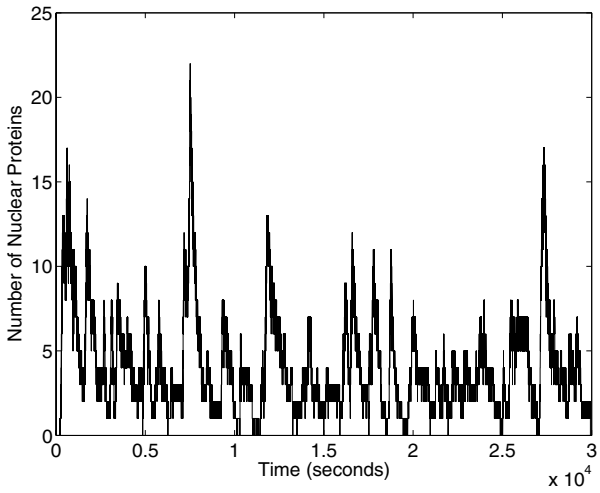


FIG. 8. *Total number of nuclear proteins in one realization of the model from section* 4.2. *Simulation occurred over* 30,000 *seconds, with data points at one second intervals.*

our method should give a more accurate representation of the stochastic dynamics of the spatially distributed chemical system.

As an illustrative application, we developed a model for eukaryotic transcription, nuclear export, translation, nuclear import, and gene regulation, and we simulated this model in a simplified two-dimensional representation of a eukaryotic cell. Spontaneous, but noisy, oscillations in nuclear protein levels were observed in this simple genetic network.

For future work, we plan to extend the current software implementation of the method to three dimensions, with the eventual goal of simulating cellular networks with realistic nuclear membrane and plasma membrane geometry. One interesting question is the interplay between reaction rates and diffusion rates in producing oscillatory nuclear protein levels. For different parameter regimes well-mixed stochastic, well-mixed deterministic, deterministic reaction-diffusion, and stochastic reaction-diffusion models could all show different qualitative behavior. We are also working on developing a more accurate biochemical model for the transcription and translation processes within eukaryotic cells.

Acknowledgments. The authors thank Ian G. Macara for helpful discussions regarding the model in [24]. We would also like to thank the referees for their careful reading of the paper and helpful comments and suggestions.

REFERENCES

[1] B. ALBERTS, A. JOHNSON, J. LEWIS, M. RAFF, K. ROBERTS, AND P. WALTER, *Molecular Biology of the Cell*, 4th ed., Garland Science, New York, 2002.

[2] S. S. ANDREWS AND D. BRAY, *Stochastic simulation of chemical reactions with spatial resolution and single molecule detail*, Physical Biology, 1 (2004), pp. 137–151.

[3] A. ARKIN AND H. H. MCADAMS, *Stochastic mechanisms in gene expression*, Proc. Natl. Acad. Sci. USA, 94 (1997), pp. 814–819.

[4] A. ARKIN, J. ROSS, AND H. H. MCADAMS, *Stochastic kinetic analysis of developmental pathway bifurcation in phage λ–infected escherichia coli cells*, Genetics, 149 (1998), pp. 1633–1648.

[5] F. Baras and M. M. Mansour, *Reaction-diffusion master equation: A comparison with microscopic simulations*, Phys. Rev. E, 54 (1996), pp. 6139–6147.
[6] W. J. Blake, M. Kaern, C. R. Cantor, and J. J. Collins, *Noise in eukaryotic gene expression*, Nature, 422 (2003), pp. 633–637.
[7] B. R. Cullen, *Nuclear mRNA export: Insights from virology*, TRENDS Biochem. Sci., 28 (2003), pp. 419–424.
[8] D. A. Drew, *A mathematical model for prokaryotic protein synthesis*, Bull. Math. Biol., 63 (2001), pp. 329–351.
[9] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Stochastic gene expression in a single cell*, Science, 297 (2002), pp. 1183–1186.
[10] D. Fusco, N. Accornero, B. Lavoie, S. Shenoy, J. Blanchard, R. Singer, and E. Bertrand, *Single mRNA molecules demonstrate probabilistic movement in living mammalian cells*, Curr. Biol., 13 (2003), pp. 161–167.
[11] C. W. Gardiner, K. J. McNeil, D. F. Walls, and I. S. Matheson, *Correlations in stochastic models of chemical reactions*, J. Statist. Phys., 14 (1976), p. 307–331.
[12] C. W. Gardiner, *Handbook of Stochastic Methods: For Physics, Chemistry, and the Natural Sciences*, 2nd ed., Springer Ser. Synergetics 13, Springer-Verlag, New York, 1996.
[13] F. Gibou, R. P. Fedkiw, L. Cheng, and M. Kang, *A second-order-accurate symmetric discretization of the Poisson equation on irregular domains*, J. Comput. Phys., 176 (2002), pp. 205–227.
[14] M. A. Gibson and J. Bruck, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, J. Phys. Chem. A, 104 (2000), pp. 1876–1899.
[15] D. T. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem., 81 (1977), pp. 2340–2361.
[16] D. T. Gillespie, *A rigorous derivation of the chemical master equation*, Phys. A, 188 (1992), pp. 404–425.
[17] D. T. Gillespie, *The chemical Langevin equation*, J. Chem. Phys., 113 (2000), pp. 297–306.
[18] K. Ishida, *Stochastic model for bimolecular reaction*, J. Chem. Phys., 41 (1964), pp. 2472–2478.
[19] P. Jenny, S. B. Pope, M. Muradoglu, and D. A. Caughey, *A hybrid algorithm for the joint pdf equation of turbulent reactive flows*, J. Comput. Phys., 166 (2001), pp. 218–252.
[20] E. R. Kandel, *The molecular biology of memory storage: A dialogue between genes and synapses*, Science, 294 (2001), pp. 1030–1038.
[21] P. McCorquodale, P. Colella, and H. Johansen, *A cartesian grid embedded boundary method for the heat equation on irregular domains*, J. Comput. Phys., 173 (2001), pp. 620–635.
[22] M. Muradoglu, P. Jenny, S. B. Pope, and D. A. Caughey, *A consistent hybrid finite-volume/particle method for the pdf equations of turbulent reactive flows*, J. Comput. Phys., 154 (1999), pp. 342–370.
[23] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, *Regulation of noise in the expression of a single gene*, Nat. Genet., 31 (2002), pp. 69–73.
[24] A. E. Smith, B. M. Slepchenko, J. C. Schaff, L. M. Loew, and I. G. Macara, *Systems analysis of ran transport*, Science, 295 (2002), pp. 488–491.
[25] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, North–Holland, Amsterdam, 2001.
[26] H. Wang, C. S. Peskin, and T. C. Elston, *A robust numerical algorithm for studying biomolecular transport processes*, J. Theoret. Biol., 221 (2003), pp. 491–511.