Hybrid simulations of stochastic reaction-diffusion processes for modeling intracellular signaling pathways

K.-H. Chiam,* Chee Meng Tan,† Vipul Bhargava,‡ and Gunaretnam Rajagopal
Bioinformatics Institute, 30 Biopolis Street, Singapore 138671, Singapore
(Received 16 May 2006; published 14 November 2006)

In the intracellular environment, signaling takes place in a nonideal environment that is spatially heterogeneous and that is noisy, with the noise arising from the low copy numbers of the signaling molecules involved. In this paper, we model intracellular signaling pathways as stochastic reaction-diffusion processes and adapt techniques commonly used by physicists to solve for the spatiotemporal evolution of the signaling pathways. We then apply it to study two problems of relevance to the modeling of intracellular signaling pathways. First, we show that, in the limit of small protein diffusion which is typically the case for proteins in the cytosol crowded by other macromolecules, the extent of diffusion control, in the transient regime, on reactions is greater than previous predictions. Second, we show that the presence of scaffold proteins can modify the phosphorylation activity of a mitogen-activated protein kinase cascade, and explain how this activity is modulated by the scaffold protein concentration.

DOI: 10.1103/PhysRevE.74.051910 PACS number(s): 82.39.−k, 87.15.Aa, 87.15.Vv, 87.15.Ya

I. INTRODUCTION

At the dawn of the twenty-first century, dramatic advances in experimental methodology in genomics and molecular biology have enabled us to probe the structure and function of living organisms to unprecedented levels of resolution. This has, however, led to a deluge of data. Analysis of this data to understand how genes and proteins work collectively to control cellular function has led to intensive use of mathematical and computational techniques both for modeling and extracting information and knowledge from the raw data resulting from high-throughput experiments. To this end, there is increasing interest in adopting a systems-level approach to model (and ultimately predict) the behavior of biological systems. The goal is to use a quantitative and integrative approach to model the complex biochemical reactions and dynamic influence of cellular networks to understand cellular behavior in its normal and diseased states. Quantitative models are essential for understanding the dynamics of intracellular signaling pathways, given the inherent complexity of the components and their network architecture. However, existing models of intracellular signaling pathways are rather restrictive. To a large extent, the focus has been on using deterministic rate laws which has at its core the law of mass action, an empirical law relating reaction rates and molecular concentrations that are treated as continuous functions of time. The chemical components are assumed to be located in a homogeneous environment (i.e., a well-stirred reactor) thereby neglecting spatial aspects. This is of course very far from reality especially when one considers the structural organization of the intracellular environment. A high degree of macromolecular crowding as well as the presence of endogenous obstacles in cellular media have important thermodynamic and kinetic consequences especially affecting diffusion processes [1]. For example, the diffusion of macromolecules in the cytoplasm can be 5 to 20 times lower than in saline solutions [2,3]. Furthermore, many important intracellular processes depend on spatial heterogeneity, among which is morphogenesis, chemotaxis, and local signal processing in neurons [4]. This therefore makes the validity of a continuous, deterministic approach in modeling intracellular signaling pathways that neglects spatial heterogeneity even more tenuous.

Incorporating heterogeneity into the modeling process highlights another key assumption inherent in the deterministic approach. This assumption states that every biomolecular species is assumed to be present in such copious quantities that their temporal fluctuations can be considered negligible, and thus a deterministic treatment is appropriate. Under *in vivo* conditions, the low copy numbers of some of the components that make up the pathway give rise to nonnegligible fluctuations, resulting in dynamics that is inherently stochastic. (Indeed, these intrinsic fluctuations have recently been measured using fluorescent probes [5–7].) This stochasticity manifests itself as a source of noise in experimental measurements of the species concentrations or other measurements derived from them [8,9]. The implications of stochasticity resulting from the small number of participating molecules in cellular processes have been well-studied [10]. For example, there have been theoretical studies [11–13] of how the presence of noise modifies the switching properties of ultrasensitive signaling cascades (such as enhanced sensitivity, reduction in the sharpness of thresholding responses, etc.), and proposals of novel phenomena such as stochastic focusing in signaling [14,15] that have eluded deterministic models. In principle, an exact way to quantitatively incorporate stochasticity in models of intracellular signaling pathways that takes into account the discrete nature of the molecules involved is to use chemical master equations [16]. They describe the dynamical changes in the probabilities of

*Electronic address: chiamkh@mailaps.org
†Present address: Biomedical Engineering Department, Duke University, Durham, North Carolina 27708, USA.
‡Present address: Bioinformatics Program, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA.

each species having a particular number of molecules at a particular time, with the transition rates dependent on the biochemical kinetics. Numerically evaluating trajectories of the chemical master equation is in principle a computationally intensive procedure. However, there are Monte Carlo techniques to speed up the process [17]. Recently, approximate methods have also been developed to accelerate the Monte Carlo techniques, making them even more efficient, though at the price of sacrificing accuracy [18–21]. (Accuracy in this case refers to the lack of discrepancy between the probability of generating the solution by the approximate techniques and the probability assigned by the chemical master equation.)

There have also been attempts to include spatial aspects of cellular signaling such as incorporating the effects of molecular diffusion and other spatial variations in species concentration into stochastic biomolecular reactions. A common approach is to subdivide the cytosolic volume into a spatial mesh, and then model molecular diffusion as random jump processes of the molecular species from one mesh point to an adjacent one [22,23]. The resulting set of stochastic reaction and jump (i.e., diffusion) processes will then form the quantitative framework that overcomes the limitations posed by the deterministic approach. A number of recently published investigations attempt to address these issues with bacterial chemotaxis being the main focus. Levin *et al.* [24] and Shimizu *et al.* [25] have investigated both the diffusive properties of signaling molecules and the implications for the spatial organization of receptor arrays. Bhalla [26] examined how diffusion and subcellular compartmentalization influences the underlying signaling processes and gives rise to a diversity of signaling outcomes which may include washout of the signals, signal amplification, and conversion of steady responses to transients. Andrews and Bray [27] have developed techniques for simulating biochemical reaction networks with a spatial resolution that is accurate to nearly the size scale of individual molecules and highlighted the capabilities of this approach by simulating simple biomolecular reactions. Elf and Ehrenberg [28,29] have adapted the Monte Carlo techniques described above to efficiently sample trajectories of reaction-diffusion master equations and applied it to understand the separation of bistable biochemical systems into spatial domains of opposite phases. Finally, Metzler [30] looked at how spatial fluctuations play a non-negligible part in cellular genetic switching processes. We note, however, that whatever the choice of algorithm used to model the stochastic spatiotemporal variations of the concentrations of signaling components in the intracellular environment, simulating the dynamics of this large number of stochastic reaction-diffusion processes is computationally very demanding even on massively parallel clusters that are widely available today.

In this paper, we first describe how intracellular signaling pathways can be modeled as a set of stochastic reaction-diffusion processes that reflect *in vivo* conditions. We then describe an algorithm to simulate these processes efficiently. The details of this algorithm will be discussed in Sec. II, where we also describe tests to check the accuracy and efficiency of the algorithm. In Sec. III, we apply the algorithm to investigate two problems relevant to intracellular signal transduction. First, we calculate the extent of diffusion control on biomolecular reactions in the limit of small biomolecular diffusion, which is typically the case for proteins in the cytosol crowded by other macromolecules. Second, we calculate quantitatively how the presence of scaffold proteins modify the phosphorylation activity of a mitogen-activated protein kinase signaling cascade. Finally, in Sec. IV, we present our conclusions.

II. METHODS

A. Definitions

We first introduce the notations that will be used throughout this paper. Let the intracellular signaling pathway we want to model consist of N species. The concentrations of these species are denoted collectively by the vector

$$\vec{X}(\vec{r},t) = \{X_1(\vec{r},t), \ldots, X_N(\vec{r},t)\}. \tag{1}$$

The coordinates denote spatial location $\vec{r} \in \Omega$ and time $0 \leqslant t \leqslant T$. Here, Ω is the modeling spatial domain, which could denote the three-dimensional cytosolic volume or the two-dimensional membrane surface, and T is the modeling period, i.e., the time up to which the pathway is to be modeled. For modeling convenience, we work with dimensionless quantities by an appropriate choice of units, i.e., length in units of 1 μm and time in units of 1 min. The components X_i's are therefore measured in units of $(A\|\Omega\|)^{-1}$, where A is Avogadro's number and $\|\Omega\|$ the volume of Ω. This choice ensures that a value for X_i of, say, 10, can be taken to mean either that the concentration of the ith species is 10 or that there are 10 molecules of the ith species present in Ω. Hereafter, throughout this paper, we will use the terms concentration and number of molecules interchangeably.

In addition, molecules of the ith species are assumed to be diffusing in the cytosolic volume with a diffusion coefficient of D_i, $i=1,\ldots,N$, measured in units of 1 $\mu m^2/\text{min}$. For the purpose of this paper, the D_i's are assumed to be constant in space and time although, in general, they need not be. Recent progress in imaging techniques has made it possible to measure the diffusion coefficients of proteins in the cytosol experimentally [2,3] with values in the order of 10 to 10^3.

Members of each reacting species may interact with one another via M elementary reactions, denoted by the vector

$$\vec{R} = \{R_1, \ldots, R_M\}. \tag{2}$$

These reactions are completely specified by two pieces of information, the stoichiometric matrix ν_{ji} that specifies the change in the number of X_i molecules per jth reaction, $j=1,\ldots,M$, $i=1,\ldots,N$, and the rate constants k_j, $j=1,\ldots,M$, that specify how fast the jth reaction proceeds.

Thus given an initial value for the concentrations, $X_i(\vec{r},t=0)$, $i=1,\ldots,N$, $\vec{r} \in \Omega$, the task of quantitatively modeling the intracellular signaling pathway can be formulated as solving for the spatiotemporal dynamics of $\vec{X}$ subject to the reactions $\vec{R}$ and molecular diffusion specified by the D_i's.

B. Algorithm

In our algorithm, the modeling domain Ω is discretized into a Cartesian spatial mesh. In this paper, for simplicity, we will let Ω be the two-dimensional unit square ($0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1$), and discretize it into $N_x \times N_y$ rectangular mesh points, each of dimensions $N_x^{-1} \times N_y^{-1}$. We will also impose periodic boundary conditions. Of course, to realistically model the dynamics of intracellular signaling pathways, we will have to consider the three-dimensional cytosolic volume with the appropriate boundary conditions at the two-dimensional cellular membrane. However, preliminary work on a simple three-dimensional system leads us to the view that we should not expect the conclusions in this paper to change if a three-dimensional domain is used, nor if more realistic boundary conditions are used. The mesh points are then populated with initial concentrations of the reacting species whose spatial distribution will reflect the spatial heterogeneity of the problem at hand. In each mesh point, the set of M reactions, $\vec{R}$, is carried out independently. Because the reactions in each mesh point involve only species residing in that mesh point, we can assume that there are $N \times N_x \times N_y$ distinct species, and rewrite Eq. (1) notationally as

$$\vec{X}(t) = \{X_i(x,y,t)\},$$

$$i = 1, \ldots, N, \quad x = 1, \ldots, N_x, \quad y = 1, \ldots, N_y. \tag{3}$$

Molecular diffusion is modeled as a set of random jump processes from mesh points to adjacent ones [22,23,28,29]. The rate constant $\tilde{k}_i$ for one of these jump processes to happen for a molecule of the ith species is given by

$$\tilde{k}_i = \frac{D_i}{(2d)l^2}, \tag{4}$$

where l is either N_x^{-1} for jumps in the x-direction or N_y^{-1} for jumps in the y-direction, and d is the dimensionality of the modeling domain Ω ($d=2$ for the unit square). Equation (4) follows from a discretized version of Fick's law, which states that the local flux of a diffusing species (i.e., flow per unit time per unit area) is proportional to the local gradient of the concentration of that species and the diffusion coefficient.

There are now a total of $(M+2dN) \times N_x \times N_y$ processes, corresponding to $M \times N_x \times N_y$ reactions and $2dN \times N_x \times N_y$ jump processes. (The factor of $2d$ arises because each mesh point has $2d$ nearest neighbors.) We can rewrite Eq. (2) notationally as

$$\vec{R} = \{R_j(x,y); \tilde{R}_i(x,y)\},$$

$$i = 1, \ldots, 2dN, \quad j = 1, \ldots, M,$$

$$x = 1, \ldots, N_x, \quad y = 1, \ldots, N_y, \tag{5}$$

where $\tilde{R}$ denotes random jump processes emulating diffusion described above.

With this formalism, the task posed above can be reformulated as solving for the evolution of $\vec{X}$ of Eq. (3), subject to the set of reaction and jump (i.e., diffusion) processes, $\vec{R}$ of Eq. (5). If this evolution can be performed stochastically, i.e., by using an approach that incorporates the small number of molecules involved, such as Gillespie's algorithm [17–19,28,29], then the two limitations of the deterministic approach raised in the Introduction will be addressed.

In this paper, we propose a hybrid approach that will allow for more efficient simulation of the large set of reaction-diffusion processes. By hybrid, we mean that either a stochastic or a deterministic treatment will be applied subject to criteria determined dynamically in the course of the spatiotemporal evolution of the system. First, we compute the propensity functions

$$\vec{a}(t) = \{a_j[\vec{X}(t)]\}, \quad j = 1, \ldots, (M + 2dN) \times N_x \times N_y \tag{6}$$

for every process in Eq. (5). The quantity $a_j[\vec{X}(t)] \times dt$, for small dt, is the probability that the jth process will occur in the time interval from t to $t+dt$. For example, consider the simple process of two monomers, X_1 and X_2, reacting at a rate k to form a dimer X_3. The propensity function for this process is

$$a[X_1(t), X_2(t)] = kX_1(t)X_2(t). \tag{7}$$

For fixed k, the propensity a is directly proportional to the product of the number of molecules of X_1 and X_2. When there are only a few molecules of X_1 and X_2 present, dimerization is less likely to occur. This likelihood increases with more molecules of X_1 and X_2 present. Similarly, for a fixed number of molecules $X_{1,2}$, the propensity a is directly proportional to the rate k. When the rate of the reaction is small, dimerization is less likely to occur. This likelihood increases with a higher rate of dimerization. Thus processes with small propensities occur more rarely than processes with large propensities. It is then natural to partition the set of processes $\vec{R}$ into two subsets—those with small propensities and those with large propensities. Exactly what is meant by small or large and how this partitioning is implemented will be discussed below. Those processes with small propensities can then be solved accurately using a stochastic approach such as Gillespie's algorithm [17–19]. We can then take advantage of those processes with large propensities and solve them using a deterministic approach. These kinds of "hybrid" algorithms have been proposed [21,31,32] and implemented [33] previously. We note, however, that only reactive processes within a homogeneous well-stirred environment in conformance with the assumptions on which Gillespie's algorithms is based on were considered, i.e., the "hybridization" was applied only to temporal updates of the species concentrations. The main contribution of our algorithm is to consider the addition of diffusion to this hybridization, thereby allowing for hybridization in both temporal and spatial updates.

The partitioning proceeds as follows. The jth process is sorted into the subset for processes with small propensities (the "small-propensity-subset") when

$$a_j \leqslant \xi \max(\vec{a}). \tag{8}$$

The parameter $0 \leqslant \xi \leqslant 1$ is a user-specified parameter of our hybrid algorithm. A value of $\xi = 1$ sorts all processes into the small-propensity-subset and hence makes the algorithm

solve all processes using a stochastic approach. On the other hand, a value of $\xi=0$ sorts all processes into the large-propensity-subset and hence makes the algorithm solve all processes using a deterministic approach. Thus the parameter ξ can be thought of as a tuning parameter that continuously tunes the species concentrations from being solved fully deterministically (i.e., which is efficient but not necessarily accurate), to being solved fully stochastically (i.e., which is accurate but not necessarily efficient). The exact amount of this tradeoff between efficiency and accuracy will depend on the specific problem and has to be specified by the user. For example, consider the example of a system of reactions involving the turnover of proteins and mRNAs. The turnover rates of the protein species are of the time scale of hours and those of the mRNA species are of the time scale of minutes. However, the protein species are present in copy numbers on the order of, say, hundreds, whereas the mRNA species are present in copy numbers on the order of unity. Thus the propensities for the mRNA turnovers are approximately of the same order of magnitude as those for protein turnovers. Hence it is possible that any choice of $\xi>0$ will result in some of the mRNA turnovers being partitioned into the large-propensity-subset, resulting in some mRNA turnovers being solved deterministically even though their copy numbers are of the order of unity. To prevent this from occurring in the algorithm, one solution is to "flag" particular reactions and explicitly "force" them to be solved deterministically or stochastically at all times. For example, the mRNA turnovers may be flagged to be partitioned into the small-propensity-subset to be solved stochastically at all times. However, the proper partitioning of systems exhibiting "stiffness," i.e., systems whose components exhibit vastly different concentrations or vastly different time scales, is an active area of research. [21] The effect of choosing a value of $\xi\neq 0,1$ on the species concentrations is shown in Fig. 1. Some time intervals of the dynamics are solutions of a deterministic approach; these intervals show smooth concentration changes and exhibit small fluctuations. Other intervals are solutions via a stochastic approach that show large fluctuations. It is also apparent from the figure that those time intervals corresponding to deterministic solutions being used are characterized by steep concentration gradients which can, intuitively, be thought of as being proportional to the propensities (though not strictly valid in multispecies and nonlinear reactions). In this sense, our hybrid algorithm can also be thought of as implementing partitioning based on the steepness of the concentration gradients, with the parameter ξ specifying the steepness threshold. In the appendixes, a pseudocode outline of the algorithm is presented, as well as a discussion on the possible sources of error.

C. Example: A diffusing-decaying-dimerizing system

We now apply our hybrid algorithm to solve an example system in order to illustrate its working properties. The example system is a set consisting of particles undergoing diffusing-decaying-dimerizing processes. (The same system but without diffusion was used as a test case in previous studies of approximate stochastic algorithms [18].) In this example system, an unstable monomer X_1 is simultaneously undergoing three competing processes:

(1) decaying at a rate k_1: $X_1 \xrightarrow{k_1} \varnothing$,

(2) dimerizing to form an unstable dimer X_2 at a forward rate k_2 and a backward rate k_3, $X_1+X_1 \underset{k_3}{\overset{k_2}{\rightleftharpoons}} X_2$, and

(3) diffusing with diffusion coefficient D_1.

In addition, the unstable dimer X_2 stablizes to the stable conformation X_3 at a rate k_4, $X_2 \xrightarrow{k_4} X_3$, and both dimers X_2 and X_3 are also diffusing with diffusion coefficients D_2 and D_3. The rates and diffusion coefficients used in this example are summarized in Table I.

Initially, the X_1 monomers are localized in a central region in the mesh, and there are no X_2 nor X_3 dimers. In Fig. 2, the time series of the total concentration of the dimer X_2 in the whole spatial mesh is plotted. It accumulates from zero, reaches a maxima, and then decays (to form the stable X_3). Several solutions are shown: the fully deterministic solution ($\xi=0$), the fully stochastic solution ($\xi=1$), and hybrid solutions at $\xi=10^{-3}$, 10^{-2}, and 10^{-1}. One peculiar property of this

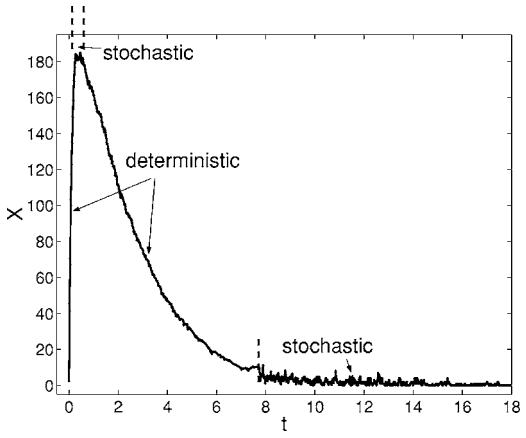

FIG. 1. A schematic illustrating how a hypothetical species concentration X varies with time t using our hybrid algorithm. Those time intervals indicated by the words "stochastic" mean that, in those intervals, the propensity for the reaction that involves species X satisfies Eq. (8) in the text, i.e., a stochastic approach is used, resulting in the characteristic large fluctuations. Conversely, those time intervals indicated by the words "deterministic" mean that, in those intervals, the propensity for the reaction that involve species X violates Eq. (8) in the text. In this case, the concentration X is solved by a deterministic approach, resulting in the smooth evolution without fluctuations.

TABLE I. Rate constants and diffusion coefficients in dimensionless units used in simulating the diffusing-decaying-dimerizing system.

k_1	k_2	k_3	k_4	D_1	D_2	D_3	N_x	N_y	Δt
1	0.002	0.5	0.04	0.1	0.1	0.1	100	100	10^{-3}

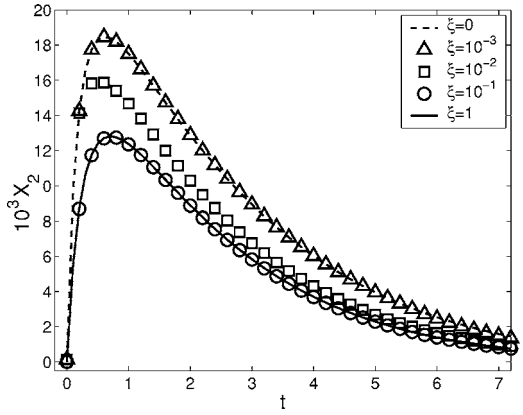

FIG. 2. Dynamics of the total concentration of the dimer X_2 summed over the domain vs time t for several values of ξ, namely $\xi=0$ (corresponding to a fully deterministic solution, dashed line), $\xi=10^{-3}$, 10^{-2}, and 10^{-1} (corresponding to hybrid solutions), and $\xi=1$ (corresponding to a fully stochastic solution, solid line). The concentration obtained by the hybrid algorithm is bounded by both the concentrations obtained deterministically and stochastically. All realizations use $N_x=N_y=50$ and $\Delta t=10^{-3}$. In addition, for $\xi\neq 0$, the realizations have been averaged over 100 sets of simulations started with different random seeds.

example system is that concentrations of the dimers X_2 and X_3 obtained fully deterministically differ from that obtained fully stochastically [18]. Thus this example system offers an opportunity to allow us to observe how this error varies in the hybrid regime with intermediate values of ξ. The figure shows that the concentration obtained by the hybrid algorithm is bounded by both the concentrations obtained deterministically and stochastically.

In Fig. 3, spatial snapshots of the concentrations fields for species X_1 and X_3 are shown at several times. The monomer X_1 is seen to diffuse outwards from the central region. However, it is evident from panel (b) that, in addition to outward spreading, localized stochastic processes are taking place (resulting in isolated dark spots in the snapshot, denoting low concentrations, scattered around the white background, denoting high concentrations). The dimer X_3 is also seen to be spreading outwards, but in a very inhomogeneous manner. This spatial inhomogeneity is the result of stochastic processes taking place. In general, the presence of deterministiclike diffusion and stochasticlike fluctuations in the same dynamics is a hallmark of our hybrid algorithm.

D. Accuracy and efficiency

We now discuss how the accuracy of the algorithm scales with the parameter ξ. We use the diffusing-decaying-dimerizing system as a test case using two metrics to quantify accuracy. First, the relative error of the mean of the ith species concentration is computed with respect to the concentration obtained from a fully deterministic approach,

$$\epsilon_{i,\mathrm{mean}} = \left| \frac{\langle \Sigma X_i(x,y,t;\xi)\rangle - \Sigma X_i(x,y,t;\xi=0)}{\Sigma X_i(x,y,t;\xi=0)} \right|. \qquad (9)$$

The Σ denotes sum over all mesh points, i.e., $\Sigma_{x=1}^{N_x}\Sigma_{y=1}^{N_y}$. The angle brackets $\langle\cdot\rangle$ denote averaging over a large set of realizations obtained with different random seeds. Note that the error is, by definition, zero at $\xi=0$. In Fig. 4, the value of ϵ_{mean} for the dimer X_3 is shown at time $t=20$ when the system has reached steady state. It remains at approximately

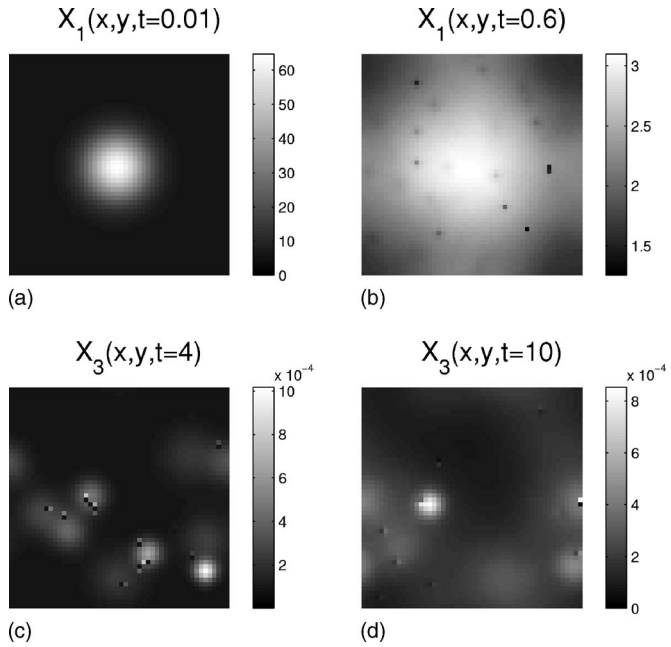

FIG. 3. Snapshots illustrating the spatiotemporal dynamics obtained from our hybrid algorithm. The concentration fields of the monomer $X_1(x,y,t)$ and dimer $X_3(x,y,t)$ are plotted for times (a) $t=0.1$, (b) $t=0.6$, and (c) $t=4$, (d) $t=10$, respectively. The reactant X_1 is initially localized at the center of the domain and is seen to diffuse outwards until it becomes more or less homogeneous, with sporadic localized fluctuations, resulting in isolated dark spots in the snapshot (denoting low concentrations) scattered around the white background (denoting high concentrations). The dimer X_3, initially at zero concentration everywhere, grows in a random and inhomogeneous manner, also with sporadic localized fluctuations. The simulation was performed with $N_x=N_y=50$, $\Delta t=10^{-4}$, and $\xi=10^{-2}$.

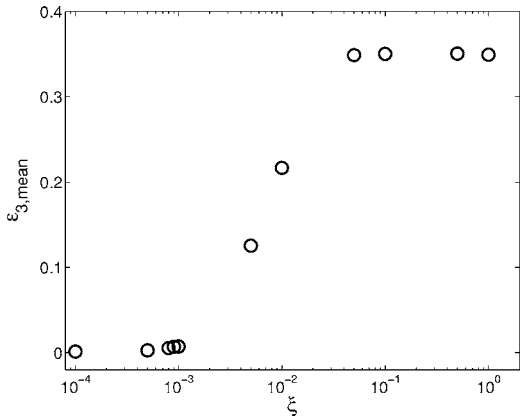

FIG. 4. The relative error $\epsilon_{3,\mathrm{mean}}$ of the mean concentration of the dimer X_3 vs the parameter ξ, calculated using Eq. (9) in the text. It increases with increasing ξ. The mean at each ξ is calculated over 100 realizations started with different random seeds.

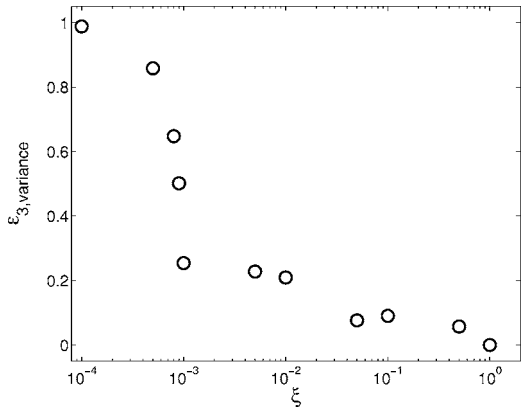

FIG. 5. The relative error $\epsilon_{3,\text{variance}}$ of the variance of the concentration of the dimer X_3 vs the parameter ξ, calculated using Eq. (10) in the text. It increases with decreasing ξ. The variance at each ξ is calculated over 100 realizations started with different random seeds.

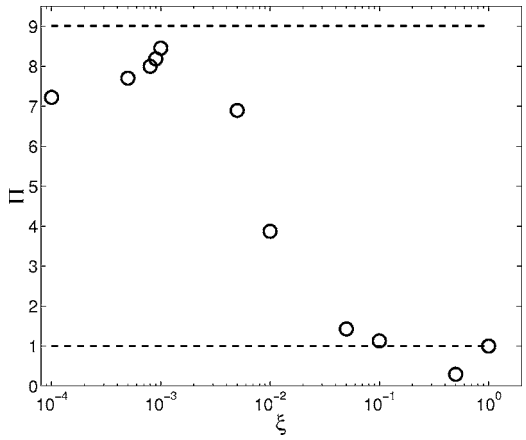

FIG. 6. Speedup Π measured by the ratio of the wall clock time taken to complete the fully stochastic dynamics of the diffusing-decaying-dimerizing system to the wall clock time taken to complete the hybrid dynamics of the same system, vs the parameter ξ. The speedups at each ξ are averaged over 100 simulations started with different random seeds. The bottom dashed line is the reference $\Pi=1$ line that indicates performance on par with a fully stochastic dynamics (i.e., no speedup). The top dashed line indicates performance that is on par with a fully deterministic dynamics (i.e., maximum speedup possible).

zero up to about $\xi \approx 10^{-3}$, at which point it increases monotonically to its maximal value at $\xi=10^{-1}$.

A second metric used is the relative error of the variance of the ith species concentration computed with respect to the variance obtained from a fully stochastic approach,

$$\epsilon_{i,\text{variance}} = \left| \frac{\sigma^2[X_i(x,y,t;\xi)] - \sigma^2[X_i(x,y,t;\xi=1)]}{\sigma^2[X_i(x,y,t;\xi=1)]} \right|, \tag{10}$$

where the variance function is

$$\sigma^2[X_i(x,y,t;\xi)] = \langle [\Sigma X_i(x,y,t;\xi)]^2 \rangle - \langle \Sigma X_i(x,y,t;\xi) \rangle^2. \tag{11}$$

In Fig. 5, the relative error for the variance of the concentration of the dimer X_3 is shown at time $t=20$ when the evolution has reached steady state. The error is, by definition, zero at $\xi=1$. It then increases as ξ decreases, up to unity at $\xi=0$ (since there, the concentration is obtained deterministically and hence does not exhibit any fluctuations).

Thus the hybrid algorithm exhibits a tradeoff between accumulating errors in the mean of the concentrations and in the variance of the concentrations. The parameter ξ can then be interpreted as quantifying this tradeoff. In this sense, there is no "best" or "optimal" choice for the value of the parameter ξ. All values of ξ between zero and unity yield valid solutions to the hybrid algorithm subject to this tradeoff.

Finally, we quantify the efficiency of the hybrid algorithm using the speedup number Π, which we define to be the wall clock time taken to complete the fully stochastic evolution divided by the wall clock time taken to complete the hybrid evolution. A value of $\Pi=1$ indicates no speedup. A value of $\Pi \gg 1$ is desirable. In Fig. 6, the speedup Π is plotted for simulations performed on the diffusing-decaying-dimerizing system. A speedup of $\Pi \approx 8.5$ is attained at $\xi=10^{-3}$. This is very close to the maximal speedup possible, $\Pi=9$ (i.e., the speed of solving the system fully deterministically). For $\xi < 10^{-3}$, the speedup actually decreases, presumably owing to computational overheads incurred in the partitioning of the processes in a situation when most if not all of them will be sorted into the large-propensity-subset to be solved deterministically anyway. It is also interesting to observe that a portion of the curve, $10^{-1} \leqslant \xi < 1$, exhibits "speeding-down" with $\Pi < 1$. This indicates that the hybrid algorithm in this regime is less efficient than the fully stochastic one. This can be attributed again to computational overheads incurred in the partitioning in a situation when most if not all of the processes will be sorted into the small-propensity-subset to be solved stochastically anyway.

The speedup curve, together with the error curves for the concentration mean and concentration variance, indicate that a value of ξ lying between 10^{-3} and 10^{-1} is ideal for carrying out hybridization for this particular system. For values of ξ less than 10^{-3}, no changes to the mean (and the variance, to a lesser extent) are observed, indicating that $\xi \approx 10^{-3}$ is the "most deterministic" extreme of the hybridization. For values of ξ greater than 10^{-1}, no changes to the variance and the mean are observed, indicating that $\xi \approx 10^{-1}$ is the "most stochastic" extreme of the hybridization.

III. RESULTS

A. Diffusion-controlled reactions

In this section, we study a simple diffusion-controlled reaction that will highlight the utility of our hybrid algorithm. Biochemical reactions that are modulated by the diffusion of the reactants are important in several aspects of signaling, such as ligand-receptor interactions [34], receptor-kinase interactions [35], catalysis by localized enzymes [36], as well as protein localization within scaffold complexes (to be discussed in the following section). We start with a brief overview of the theory of diffusion-controlled reactions [37,38].

TABLE II. Rate constants, diffusion coefficients, and other parameters in dimensionless units used in simulating Eq. (19).

r_f	r_b	D_1	D_2	R	ρ	N_x	N_y	Δt
1	0.1	2.5×10^{-6}	2.5×10^{-6}	0.01	0.1	100	100	10^{-3}

Consider, for simplicity, the formation of a dimer X_3 from its two constituent reactants, X_1 and X_2,

$$X_1 + X_2 \underset{k_b}{\overset{k_f}{\rightleftharpoons}} X_3, \tag{12}$$

where k_f and k_b are the forward and reverse rate constants. It is implicitly assumed that Eq. (12) is an approximation to a more complex underlying process involving two (or more) steps,

$$X_1 + X_2 \underset{d_b}{\overset{d_f}{\rightleftharpoons}} X_4 \underset{r_b}{\overset{r_f}{\rightleftharpoons}} X_3. \tag{13}$$

Here, d_f and d_b are the rates of formation and dissolution of an intermediate encounter complex X_4, and r_f and r_b are the forward and reverse rate constants for the encounter complex to dimerize. The encounter complex is formed temporarily whenever a molecule of X_1 wanders by random walk to within an encounter distance R of a molecule of X_2, and vice versa. (In most cases neglecting interactions, the encounter distance is simply the sum of the radii of the molecules of X_1 and X_2, i.e., the closest distance that a molecule of X_1 and a molecule of X_2 can get together before dimerization occurs.) Therefore the rates d_f and d_b are related to the diffusion coefficients D_1 and D_2 for X_1 and X_2, respectively [37,38],

$$d_f = 2(D_1 + D_2), \tag{14}$$

$$d_b = 2(D_1 + D_2)R^{-2}. \tag{15}$$

By assuming that the concentration of the encounter complex X_4 is small compared to those of the reactants and product, i.e., that it is at quasi-steady-state, then it can be shown that Eq. (12) is identical to Eq. (13) when [37,38]

$$k_f = \frac{d_f r_f}{d_b + r_f}, \tag{16}$$

$$k_b = \frac{d_b r_b}{d_b + r_f}. \tag{17}$$

If in addition

$$\rho \equiv \frac{d_b}{r_f} < 1, \tag{18}$$

then $k_f \approx d_f$ and the dimerization is said to be diffusion-controlled (i.e., where reactant diffusion is so slow that it completely determines the reaction rate). Conversely, if $\rho > 1$, the dimerization is said to be reaction-or activation-controlled (i.e., the reaction rate is fully determined by an activation energy barrier making reactant diffusion irrelevant).

It is common to use Eqs. (12), (16), and (17) as a theoretical model for reaction-diffusion processes [35] as it is computationally less expensive to solve. However, in this section, we show that this theoretical model does not yield the correct concentrations in the transient regime by explicitly solving for the stochastic reaction-diffusion processes using our hybrid algorithm. We are interested in the diffusion-controlled regime, i.e., at small values of ρ. One scenario where this occurs is in scaffold protein complexes, which typically have large molecular weights and thus small diffusion coefficients. Another scenario involves the diffusion of proteins in the cytosol which is crowded with other large molecules such as microtubules, actin filaments, and organelles [1]. In this scenario, the mean free paths of the proteins are reduced, resulting in a smaller diffusion coefficient.

We implement the dimerization

$$X_1 + X_2 \underset{r_b}{\overset{r_f}{\rightleftharpoons}} X_3 \tag{19}$$

with forward and reverse rate constants r_f and r_b, respectively, on the unit square with a spatial mesh of $N_x = N_y = 100$ points in each direction. We let the species diffuse with diffusion coefficients D_1, D_2, and D_3, respectively. The encounter distance is the dimension of one mesh point, $R = N_x^{-1} = N_y^{-1}$, i.e., a molecule of X_1 will only react with a molecule of X_2 when they are both in the same mesh point. The rate constants and diffusion coefficients used are summarized in Table II. Initially, X_1 and X_2 molecules are randomly distributed in the spatial mesh, and there are no X_3 molecules.

In Fig. 7, snapshots of the spatiotemporal evolution of the dimer concentration field X_3 at several instances of time t averaged over 100 sets of simulations started with different random seeds are shown. The "coarsening" observed is representative of the effects of reactant diffusion.

The time evolution of the total concentration of the dimer X_3 in the whole mesh is then plotted vs time t and compared with the result obtained from solving Eqs. (12), (16), and (17). In Fig. 8, the concentration profiles from our hybrid simulations and from the theoretical model are plotted. In addition, another set of profiles for $\rho = 1$ is also shown for reference.

We find that the concentration profile for $\rho = 0.1$ from our hybrid simulations does not agree with the theoretical model at early times $t < 200$. We note that this time is approximately on the order of $\sim R^2/D_{1,2}$. Thus Eqs. (12), (16), and (17) are invalid at early times $t < R^2/D_{1,2}$ and when the reactants' diffusion coefficients are small, i.e., in the diffusion-controlled regime. (No such discrepancies were found for the concentration profiles for $\rho = 1$.) Of course, the theoretical model still correctly predicts the long-time steady-state value

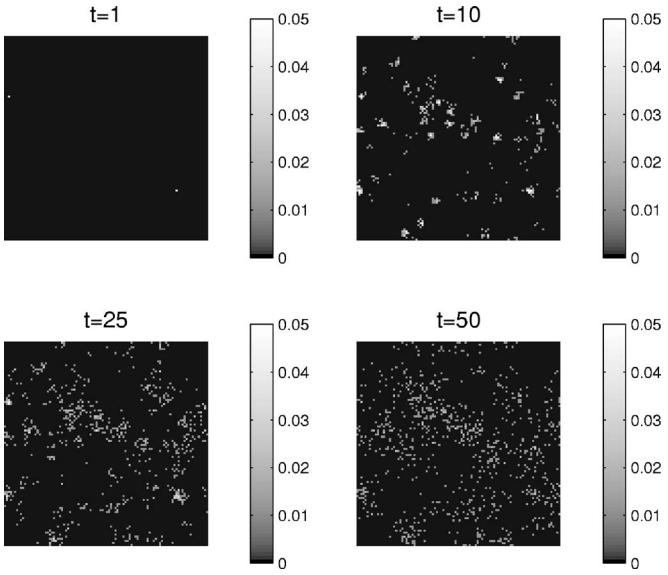

FIG. 7. Snapshots of the mean spatial distribution of the dimer $\langle X_3(x,y,t)\rangle$ for $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$ and at times $t=1$, 10, 25, and 50 and subject to the reaction-diffusion process of Eq. (19) in the text with parameters listed in Table II. The angle brackets $\langle\cdot\rangle$ denote averaging over 100 sets of simulations started with different random seeds.

for the concentrations. Nevertheless, transient behaviors may have kinetic consequences, especially when the reactions are cascaded, as is frequently the case in intracellular signaling pathways. Thus whatever happens at early times upstream may be propagated downstream at later times. In scenarios in which these effects are anticipated to be important, it is suggested that our hybrid algorithm explicitly includes diffusive processes be used.

B. Scaffold proteins in intracellular signaling pathways

In this section, we use the hybrid algorithm to study the role of scaffold proteins in signal transduction. In particular,

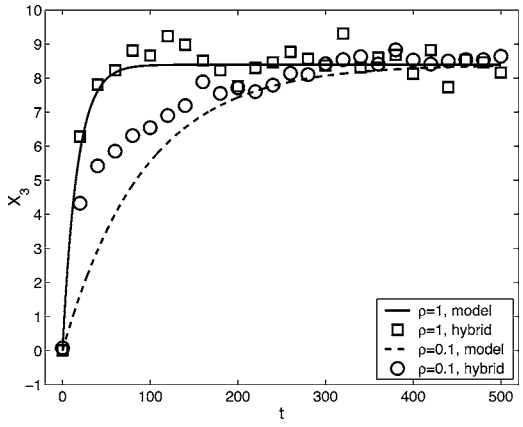

FIG. 8. Concentration of species X_3 vs time t generated from our hybrid simulations, Eq. (19), and from the theoretical model, Eqs. (12), (16), and (17), for two values of $\rho=0.1$ and $\rho=1$. The parameters used are shown in Table II. The profiles generated from our model are averaged over 100 sets of simulations started with different random seeds.

we focus on a model of the mitogen-activated protein kinases (MAPK), which are activated by a protein kinase cascade mechanism in which the MAPK is phosphorylated and activated by a MAP-kinase-kinase (MAPKK) that is, in turn, phosphorylated and activated by a MAP-kinase-kinase-kinase (MAPKKK) [39–41]. The cascade reactions occur in the cytosol resulting in the activated MAPK phosphorylating various targets in the cytosol and the nucleus. The protein kinases are evolutionary conserved features present in a variety of receptor-mediated signaling pathways in eukaryotic organisms from yeast to humans. In yeast (*S. cerevisiae*) the MAPK cascade acts in three distinct signaling pathways [39,41]: a pheromone sensing pathway, a pathway that monitors extracellular osmotic conditions, and a nutrient-sensitive pathway that converts individual yeast cells into a connected filamentous form. While these pathways may share some identical members, nevertheless specificity is normally faithfully maintained. It has been established experimentally that pathway specificity can be maintained by the presence of scaffold proteins. For example, seminal experiments on the yeast mating pathway have established that scaffold proteins (such as Ste5) are indeed physiologically relevant to the regulation of MAPK modules [42]. When the Ste5 scaffold protein is present, the Fus3 kinase (MAPK), the Ste7 kinase (MAPKK), and the Ste11 kinase (MAPKKK) are recruited to the scaffold. Signals then propagate down the MAPK cascade triggering pheromones resulting in a mating response. When Ste5 is absent, signal propagation is inhibited and consequently pheromones cannot induce mating. In addition, it has also been shown through the creation of artificial scaffold proteins that the latter can enable signal transmission through a MAPK module [43].

One of the consequences of scaffold protein maintaining pathway specificity is the precise regulation of MAPK signaling, i.e., scaffold proteins prevent the activation of MAPK modules by irrelevant stimuli as well as providing spatial and temporal control of MAPK signaling. However, it is unclear what other roles, if any scaffold proteins play. As such, exploring the roles of scaffold proteins in addition to maintaining pathway specificity is still an active area of research [40]. In addition to experimental investigations, there have been several computational studies on scaffold proteins [44,45] in recent years to try to uncover possible new roles in the signal transduction process. These computational studies have yielded two important insights concerning scaffolds. First, scaffolds amplify the signaling, i.e., it results in a larger fraction of the MAPK being activated, but only for a range of scaffold concentrations. Second, scaffolds make the cascade less sensitive to fluctuations within the cellular environment, in the sense that the dosage-response (or input stimulus vs output activity) curve is more graded and less sharp.

We use the hybrid algorithm to explore the first of these two issues in more detail, namely, the amplification of signaling activity by scaffold proteins. Our contribution lies in the fact that our model explicitly takes into account species diffusion and stochasticity. Diffusion is especially crucial for accurate modeling because the formation of scaffold complexes is likely to be diffusion-controlled, given the large masses typical of scaffold proteins and hence their low diffusion coefficient. However, rather than study a specific sys-

tem, we choose to model a generic scaffold complex in a generic MAPK phosphorylation cycle, namely the model proposed by Goldbeter and Koshland [46]. This model comprises a kinase species K and a kinase-kinase species E. The kinase-kinase molecules catalyze the phosphorylation of the kinase molecules K into their active configuration K^*. There are also phosphatases P that catalyze the dephosphorylation of the active kinase molecules into their inactive form. The scaffold proteins F can recruit and bind to both the kinase K (of the inactive form) and the kinase-kinase E, when either or both of the latter happen to have diffused to within an encounter distance. We assume that trimolecular encounters are rare, and are represented by two bimolecular encounters, i.e., the scaffold F can bind to K (or E) to form an intermediate scaffold complex, and then this intermediate scaffold complex can bind to E (or K). Once both the kinase K and kinase-kinase E molecule are recruited to the scaffold, the catalysis of the kinase K to its active form K^* occurs. Upon activation, the scaffold complex breaks up and K^* and E diffuse away.

In summary, our model comprises the following reactions:

$$K + E \underset{k_{12}}{\overset{k_{11}}{\rightleftharpoons}} I_1 \xrightarrow{k_{13}} K^* + E, \tag{20}$$

$$K^* + P \underset{k_{22}}{\overset{k_{21}}{\rightleftharpoons}} I_2 \xrightarrow{k_{23}} K + P, \tag{21}$$

$$F + K \underset{k_{32}}{\overset{k_{31}}{\rightleftharpoons}} C_1, \tag{22}$$

$$C_1 + E \underset{k_{42}}{\overset{k_{41}}{\rightleftharpoons}} C_2 \xrightarrow{k_{43}} F + K^* + E, \tag{23}$$

$$F + E \underset{k_{52}}{\overset{k_{51}}{\rightleftharpoons}} C_3, \tag{24}$$

$$K + C_3 \underset{k_{62}}{\overset{k_{61}}{\rightleftharpoons}} C_2 \xrightarrow{k_{63}} F + K^* + E. \tag{25}$$

The kinase molecules K and K^* are diffusing with a diffusion coefficient of D_K. Similarly, the kinase-kinase molecules and the phosphatases have diffusion coefficients D_E and D_P, respectively. The intermediate complexes I_1 and I_2 have diffusion coefficients D_{I_1} and D_{I_2}, respectively. (Note that all diffusion coefficients are assumed to be constant in this paper.) We assume that the scaffold proteins have a molecular weight that is very large compared to the other species, and so, do not diffuse. Thus the scaffold proteins F and the scaffold complexes C_1 to C_3 have diffusion coefficients of zero. The rate constants and diffusion coefficients used in simulating the model are summarized in Table III. We have found that our results are still qualitatively valid for a wide range of rate constants and diffusion coefficients.

Initially, the kinases, kinase-kinases, phosphatases, and scaffolds are randomly distributed on the spatial mesh. In Fig. 9, we plot the time course of the total kinase phosphorylation activity $\alpha(t)$ over the whole mesh, which is the fraction of kinase molecules in the mesh that are in the phosphorylated form at time t, for several values of the scaffold-to-kinase concentration ratio f. We see that the addition of scaffolds, $f>0$, can either increase or decrease the steady-state activity.

TABLE III. Rate constants and diffusion coefficients in dimensionless units used in simulating Eqs. (20)–(25).

i	k_{i1}	k_{i2}	k_{i3}
1	0.4167	150	600
2	0.4167	150	600
3	4.167	60	
4	2.0835	150	600
5	4.167	60	
6	2.0835	150	600
D_k	D_E, D_P	D_{I_1}, D_{I_2}	
1	0.33	0.28	
N_x	N_y	Δt	ξ
10	10	10^{-3}	10^{-3}

This is shown more clearly in Fig. 10, which plots the relative change of the activity,

$$\Delta\alpha(f) = \frac{\alpha(f) - \alpha(f=0)}{\alpha(f=0)}, \tag{26}$$

as a function of the scaffold-to-kinase concentration ratio f. It is seen that the steady-state activity can be amplified by up to approximately 20% when there are as many scaffolds as there are kinases ($f=1$). However, the further increase of scaffolds causes the steady-state activity to eventually decrease. Finally, when there are approximately 50 times more scaffolds than kinases, the steady-state activity is almost zero, i.e., there are almost no phosphorylation events.

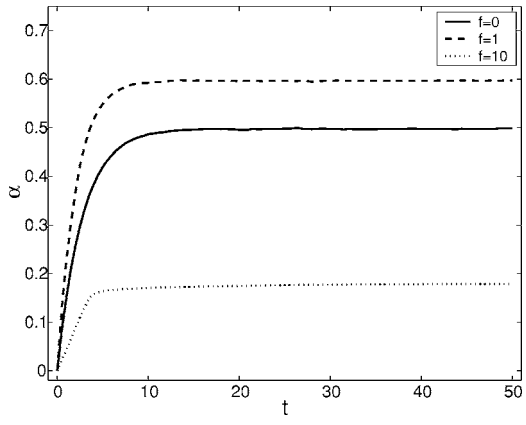

FIG. 9. Kinase phosphorylation activity α summed over the whole spatial mesh vs time t for several values of the scaffold-kinase concentration ratio f. In the absence of scaffold proteins ($f=0$), the activity achieves a steady-state value of $\alpha=0.5$. However, the addition of scaffold proteins can either increase (dashed line) or decrease (dotted line) the steady-state activity.

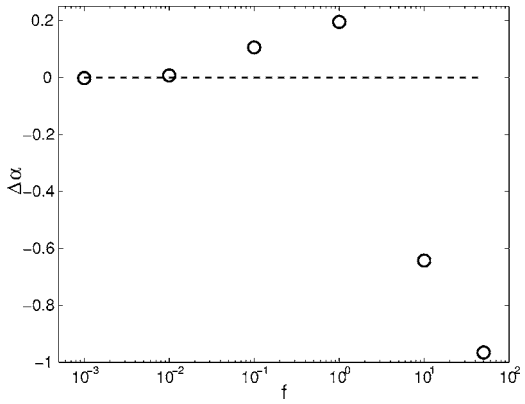

FIG. 10. Relative change $\Delta\alpha$ in the steady-state activity vs the scaffold-to-kinase concentration ratio f. The activity is amplified for an optimal range of scaffold concentrations $10^{-2} \leqslant f \leqslant 5$.

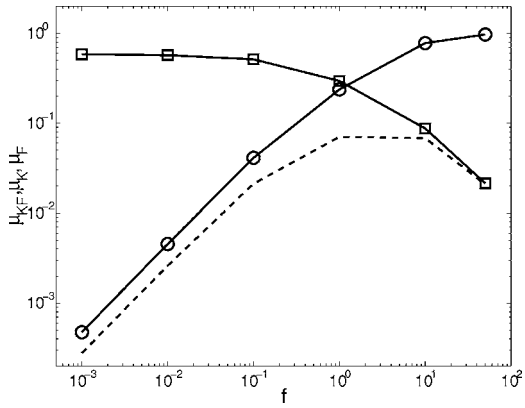

FIG. 11. "Utilization ratio" of scaffolds μ_F (denoted by circles), kinases μ_K (square symbols), and the product of these two ratios μ_{KF} (dashed line) vs the scaffold-to-kinase concentration ratio f. The product peaks when both utilization ratios are equal, suggesting that steady-state activity can be amplified by the presence of scaffolds only when the ratio of bound-to-free kinases and the ratio of bound-to-free scaffolds are equal.

The existence of an optimal range of scaffold concentrations $10^{-2} \leqslant f \leqslant 5$ for which the steady-state kinase activity is enhanced was first observed by Levchenko *et al.* [45]. However, their model, though quantitative, does not account for spatial diffusion and is deterministic. Here, with information on the spatial distribution of the proteins available, we explain how this optimal range of scaffold concentration arises. At low scaffold concentrations, the number of scaffold-bound kinases is small compared to the number of freely diffusing kinases. As the scaffold concentration increases, this ratio of scaffold-bound-to-free kinases increases, and eventually, all kinases will be scaffold-bound. The same is also true of the ratio of scaffold-bound-to-free kinase-kinases. Thus we can compute a "utilization ratio," μ_K, which measures the ratio of bound-to-free kinases, and it increases monotonically with scaffold concentration. However, at low scaffold concentrations, the number of bound scaffolds is large compared to the number of free scaffolds. As the scaffold concentration increases, this ratio of bound-to-free scaffolds will decrease, simply because there will not be enough kinases to recruit. Thus we can compute another "utilization ratio," μ_F, which measures the ratio of bound-to-free scaffolds. It decreases monotonically with scaffold concentration. These two ratios are plotted in Fig. 11. Their product, μ_{KF}, then exhibits a peak centered at around $f \approx 1$. Thus the optimal range of scaffold concentration for amplifying steady-state activity is attained, not when either the kinases or scaffolds are separately optimally "utilized," but when both the kinases and scaffolds are equally "utilized," i.e., when the ratio of bound-to-free kinases and the ratio of bound-to-free scaffolds are equal.

The results in this section highlight how scaffold proteins in conjunction with environmental factors such as stochasticity and diffusion can influence signal transduction by modifying the activity levels of the signaling process.

IV. CONCLUSIONS

In this paper, we have argued why models for intracellular signaling pathways based on a deterministic approach that has as its basis, a well-stirred homogeneous environment, are inadequate for *in vivo* conditions. The main limitations are fluctuations that arise due to the small number of some of the signaling components and the heterogeneity of the structurally organized intracellular environment. We then discussed how stochastic reaction-diffusion processes may be used to overcome these limitations so as to model realistically biochemical processes that couple reaction with diffusion. We described a hybrid algorithm that can be used to solve these processes efficiently that is very easy to implement. A key feature of the hybrid algorithm is that the user chooses the degree of hybridization to use by specifying a threshold parameter for the set of propensities. Those processes whose propensities fall below this threshold will be solved by a stochastic approach (accurate but not necessarily efficient) whereas those processes whose propensities are above this threshold will be solved by a deterministic approach (efficient but not necessarily accurate). We also quantified the accuracy and efficiency of our hybrid algorithm by showing how the errors in the mean and variance of species concentrations scale vary with the threshold parameter. We then applied the hybrid algorithm to study two problems. First, we showed that the theoretical model of diffusion-controlled reactions is not valid at early times when the reactants' diffusion coefficients are small. Second, we showed how scaffold proteins can modify the steady-state activity of kinase phosphorylation, and quantified this modification by studying scaffold and kinase utilizations. We are currently applying the computational framework developed and described in this paper to model signal transduction in the Wnt signaling pathway [47] to study the role of stochasticity and heterogeneity in modulating signaling in this pathway which is critical in development biology and tumorgenesis.

ACKNOWLEDGMENTS

We acknowledge discussions with Sydney Brenner, Louis Lim, Stefan Lim, and Jerry Solomon. This work was sup-

ported by the Biomedical Research Council of the Agency for Science, Technology, and Research of Singapore.

APPENDIX A: IMPLEMENTATION OF ALGORITHM

A pseudocode outline of the algorithm is as follows. The set of concentrations $\vec{X}(t)$ was defined in Eq. (3). The set of reaction-diffusion process $\vec{R}$ and the set of propensities $\vec{a}(t)$ were defined in Eqs. (5) and (6), respectively. The set of processes belonging to the small-propensity-subset will be denoted by $\vec{R}_S(t)$ and their propensities $\vec{a}_S(t)$; the set of processes belonging to the large-propensity-subset will be denoted by $\vec{R}_L(t)$ and their propensities $\vec{a}_L(t)$. (Note that $\vec{R}_S$ and $\vec{R}_L$ depend on time t, but their union, $\vec{R}$, does not.)

(1) `initialize` and set $t=0$
(2) while $(t<T)$ do
(3) $\vec{a}(t)=$`calculatePropensities`$[\vec{R},\vec{X}(t)]$
(4) $[\vec{R}_S(t),\vec{a}_S(t),\vec{R}_L(t),\vec{a}_L(t)]=$`partition`$(\vec{R},\vec{a}(t),\xi)$
(5) $\tau=$`calculateTau`$[\vec{a}_S(t)]$
(6) if $(\Delta t\leqslant\tau)$ do
(7) $t_0=t$
(8) while $(t\leqslant t_0+\tau)$ do
(9) $\vec{X}(t+\Delta t)=$`solveDeterministic`$[\vec{R}_L(t),\vec{X}(t)]$
(10) $\vec{a}(t)=$`calculatePropensities`$[\vec{R}_S(t),\vec{X}(t)]$
(11) $\tau=$`calculateTau`$[\vec{a}_S(t)]$
(12) $t=t+\Delta t$
(13) end while
(14) $\vec{X}(t+\tau)=$`solveStochastic`$[\vec{R}_S(t),\vec{X}(t)]$
(15) else
(16) $\vec{X}(t+\tau)=$`solveDeterministic`$[\vec{R}(t),\vec{X}(t)]$
(17) $t=t+\tau$
(18) end if
(19) end while

The `initialize` subroutine sets up the algorithm by reading the input specifications of the pathway to be modeled, which is given in the SBML format [48]. In addition, the other parameters such as N_x, N_y, and ξ, as well as initial concentrations for the species, are specified. The discretization of the spatial mesh and the population of species are also carried out.

The `calculatePropensities` subroutine computes the propensity functions as described above for the set of processes passed into it. The `partition` subroutine carries out the partitioning of $\vec{R}$ into $\vec{R}_S(t)$ and $\vec{R}_L(t)$ as described above. The `calculateTau` subroutine takes the propensities for the processes in the small-propensity-subset and selects a time value τ randomly which is the time that one of these processes will occur [17]. As can be seen from the pseudocode, τ is calculated within the `while` loop of lines (8)–(13), i.e., it is updated after every deterministic time step. This is necessary because some of the species that participate in reactions partitioned into the large-propensity-subset may also be involved in the small-propensity-subset of reactions and hence contribute to the calculation of τ. To make the recalculation of τ efficient, we follow the method of Gibson and Bruck [49] and have stored, in the `initialize` subroutine, a dependency graph that specifies which reactions will modify which propensity. Note that this dependency graph is not modified as the code runs. Thus in the `calculatePropensities` subroutine, only those propensities that need to be modified are updated. This results in an efficient way to evaluate τ. From time t up to time $t+\tau$, no other events from the small-propensity-subset will occur. However, in this same time interval, the processes in the large-propensity-subset will have to be updated. This is handled by the `solveDeterministic` subroutine, which solves the processes in the large-propensity-subset in time steps of $\Delta t<\tau$. At the end of each time step Δt, the propensities for the processes in the small-propensity-subset are reevaluated and the value of τ is updated with these new propensities. This update helps to reduce errors that will be discussed later.

The `solveDeterministic` subroutine updates the processes in the large-propensity-subset deterministically. It rewrites the processes as differential equations, which is most generally the reaction-diffusion equations,

$$\frac{\partial X_i(\vec{r},t)}{\partial t}=\sum_{j=1}^{J}\nu_{ji}a_j(X_1,\ldots,X_N)+D_i\vec{\nabla}^2X_i, \tag{A1}$$

where $j=1,\ldots,J$ with J the number of processes in the large-propensity-subset. There are many techniques to solve Eq. (A1) numerically, and we have chosen to use a simple technique, the forward-time and centered-space discretization method. It discretizes time using the forward Euler scheme and discretizes space using the central difference scheme [50]. Using this scheme, Eq. (A1) becomes

$$\begin{aligned}\frac{X_i(x,y,t+\Delta t)-X_i(x,y,t)}{\Delta t}&=\sum_{j=1}^{J}\nu_{ji}a_j[X_1(x,y,t),\ldots,X_N(x,y,t)]\\&+D_i\frac{X_i(x+1,y,t)-2X_i(x,y,t)+X_i(x-1,y,t)}{N_x^{-2}}\\&+D_i\frac{X_i(x,y+1,t)-2X_i(x,y,t)+X_i(x,y-1,t)}{N_y^{-2}}.\end{aligned} \tag{A2}$$

Thus $\vec{X}(x,y,t+\Delta t)$ is completely determined by $\vec{X}(x,y,t)$. The choice of the time step Δt will have to guarantee stability and convergence [50].

The `solveStochastic` subroutine simply chooses one of the processes in the small-propensity-subset randomly [17], and updates the concentration $\vec{X}$ at time $t+\tau$ based on the values of the stoichiometric matrix.

APPENDIX B: SOURCES OF ERROR

Since our hybrid algorithm is, after all, an approximate one, the species concentrations will contain errors (except of

course when $\xi=0$ or 1, in which case the algorithm reproduces the exact deterministic or stochastic solutions, respectively).

One source of error occurs when a jump (i.e., diffusion) process is sorted into the large-propensity-subset to be solved deterministically. Recall that a jump process is a local one involving only two mesh points. However, solving diffusion deterministically is, by definition, a global procedure involving every mesh point in the modeling domain Ω. For example, consider the following scenario: At time t, a jump process for a particular species at mesh point (x_1,y_1) is being sorted into the large-propensity-subset, while a jump process for the same species at another mesh point (x_2,y_2) is being sorted into the small-propensity-subset. Thus from time t to $t+\tau$, diffusion of the species will be solved deterministically for all mesh points. At time $t+\tau$, the jump process at (x_2,y_2) is chosen as the stochastic event to be updated. The species concentration at (x_2,y_2) will thus be "doubly updated" at time $t+\tau$.

Another source of error occurs when a particular species is involved in reactions that are being sorted into both the large-propensity-subset and the small-propensity-subset. It may happen that, by the time the reaction in the small-propensity-subset is to be updated stochastically, the species concentration will have already been modified by the reaction in the large-propensity-subset to such an extent that the time at which the stochastic process will take place [τ generated in line (5) of the pseudocode] will be inaccurate. This is overcome by recalculating τ after every deterministic update in lines (10) and (11) of the pseudocode.

Finally, a third source of error occurs when the concentration of a particular species, previously solved by a deterministic approach, now has to be solved by a stochastic approach. This can happen when the propensity for the process involving this species changes to an extent that it becomes repartitioned from the large-propensity-subset to the small-propensity-subset. Thus one has to "connect" a deterministic solution to a stochastic solution. This is a problem when the `solveStochastic` subroutine takes as input, say, a noninteger concentration value of, say, 3.14 for the species that is to be updated. One solution to overcome this error is to simply round off the concentrations (so that 3.14 becomes 3 and 2.72 becomes 3) prior to calling `solveStochastic`. Thus line (14) of the pseudocode can be replaced by

$$(14)\ \vec{X}(t) = \texttt{solveStochastic}(\vec{R}_S(t), \texttt{NINT}[\vec{X}(t)])$$

with $\texttt{NINT}[\cdot]$ denoting the rounding-off function. However, this solution is not ideal because it is possible that the errors due to the rounding-off may accumulate and grow with time [20]. An alternate solution is to adopt a "random-rounding" approach such as that used by Bhalla [26] whereby a noninteger concentration value is randomly rounded up or down to the nearest integer. In our algorithm, we did not implement any rounding operations. We decided to preserve the law of mass conservation and speak of, say, 3.14 molecules even though the statement is meaningless. Nevertheless, we view this as a side effect of a hybrid algorithm that is approximate.

[1] R. J. Ellis and A. P. Minton, Nature (London) **425**, 27 (2003).
[2] A. S. Verkman, Trends Biochem. Sci. **27**, 27 (2002).
[3] D. Gerlich and J. Ellenberg, Nat. Cell Biol. **5**, S14 (2003).
[4] E. A. Finch and G. J. Augustine, Nature (London) **396**, 753 (1998).
[5] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[6] W. J. Blake, M. Kærn, C. R. Cantor, and J. J. Collins, Nature (London) **422**, 633 (2003).
[7] Mads Kærn, T. C. Elston, W. J. Blake, and J. J. Collins, Nat. Rev. Genet. **6**, 451 (2005).
[8] M. Sasai and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **100**, 2374 (2003).
[9] J. Paulsson, Nature (London) **427**, 415 (2004).
[10] C. V. Rao, D. M. Wolf, and A. P. Arkin, Nature (London) **420**, 231 (2002).
[11] O. G. Berg, J. Paulsson, and M. Ehrenberg, Biophys. J. **79**, 1228 (2000).
[12] R. Metzler and P. G. Wolynes, Chem. Phys. **284**, 469 (2002).
[13] M. Thattai and A. van Oudenaarden, Biophys. J. **82**, 2943 (2002).
[14] J. Paulsson, O. G. Berg, and M. Ehrenberg, Proc. Natl. Acad. Sci. U.S.A. **97**, 7148 (2000).
[15] J. Paulsson and M. Ehrenberg, Phys. Rev. Lett. **84**, 5447 (2000).
[16] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1983).
[17] D. T. Gillespie, J. Phys. Chem. A **81**, 2340 (1977).
[18] D. T. Gillespie, J. Chem. Phys. **115**, 1716 (2001).
[19] D. T. Gillespie and L. R. Petzold, J. Chem. Phys. **119**, 8229 (2003).
[20] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, J. Chem. Phys. **119**, 12784 (2003).
[21] Y. Cao, D. T. Gillespie, and L. R. Petzold, J. Chem. Phys. **122**, 014116 (2005).
[22] F. Baras and M. M. Mansour, Phys. Rev. E **54**, 6139 (1996).
[23] A. B. Stundzia and C. J. Lumsden, J. Comput. Phys. **127**, 196 (1996).
[24] M. D. Levin, T. S. Shimizu, and D. Bray, Biophys. J. **82**, 1809 (2002).
[25] T. S. Shimizu, S. V. Aksenov, and D. Bray, J. Mol. Biol. **329**, 291 (2003).
[26] U. S. Bhalla, Biophys. J. **87**, 733 (2004).
[27] S. S. Andrews and D. Bray, Phys. Biol. **1**, 137 (2004).
[28] J. Elf, A. Dončić, and M. Ehrenberg, Proc. SPIE **5110**, 114 (2003).
[29] J. Elf and M. Ehrenberg, Systems Biology **1**, 230 (2004).
[30] R. Metzler, Phys. Rev. Lett. **87**, 068103 (2001).
[31] E. L. Haseltine and J. B. Rawlings, J. Chem. Phys. **117**, 6959 (2002).
[32] C. V. Rao and A. P. Arkin, J. Chem. Phys. **118**, 4999 (2003).
[33] J. Puchałka and A. M. Kierzek, Biophys. J. **86**, 1357 (2004).

[34] D. A. Lauffenburger and J. J. Linderman, *Receptors: Models for Binding, Trafficking and Signaling* (Oxford University Press, Oxford, 1993).
[35] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Müller, Nat. Biotechnol. **20**, 370 (2002).
[36] G. L. Nelsestuen and M. B. Martinez, Biochemistry **36**, 9081 (1997).
[37] M. Eigen, *Quantum Statistical Mechanics In The Natural Sciences* (Plenum, New York, 1974), Chap. Diffusion control in biochemical reactions, pp. 37–61.
[38] G. I. Bell, Science **200**, 618 (1978).
[39] W. R. Burack and A. S. Shaw, Curr. Opin. Cell Biol. **12**, 211 (2000).
[40] J. E. Ferrell, Jr., Science Signal Transduction Knowledge Environment **2000**, PE1 (2000).
[41] D. K. Morrison and R. J. Davis, Annu. Rev. Cell Dev. Biol. **19**, 91 (2003).
[42] E. A. Elion, J. Cell. Sci. **114**, 3967 (2001).
[43] S.-H. Park, A. Zarrinpar, and W. A. Lim, Science **299**, 1061 (2003).
[44] D. Bray and S. Lay, Proc. Natl. Acad. Sci. U.S.A. **94**, 13493 (1997).
[45] A. Levchenko, J. Bruck, and P. W. Sternberg, Proc. Natl. Acad. Sci. U.S.A. **97**, 5818 (2000).
[46] A. Goldbeter and D. E. Koshland, Jr., Proc. Natl. Acad. Sci. U.S.A. **78**, 6840 (1981).
[47] E. Lee, A. Salic, R. Krüger, R. Heinrich, and M. W. Kirschner, PLoS Biol. **1**, 116 (2004).
[48] M. Hucka *et al.*, Bioinformatics **19**, 524 (2003).
[49] M. A. Gibson and J. Bruck, J. Phys. Chem. A **104**, 1876 (2000).
[50] W. H. Press, B. P. Glannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 1989).