Simulating Biochemical Networks at the Particle Level and in Time and Space: Green's Function Reaction Dynamics

Jeroen S. van Zon[1] and Pieter Rein ten Wolde[2,1,*]

[1]*Division of Physics and Astronomy, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands*
[2]*FOM Institute for Atomic and Molecular Physics, Kruislaan 407, 1098 SJ Amsterdam, The Netherlands*

(Received 9 June 2004; published 1 April 2005)

We present a technique, called Green's function reaction dynamics (GFRD), for particle-based simulations of reaction-diffusion systems. GFRD uses a maximum time step such that only single particles or pairs of particles have to be considered. For these particles, the Smoluchowski equations are solved analytically using Green's functions, which are used to set up an event-driven algorithm. We apply the technique to a model of gene expression. Under biologically relevant conditions, GFRD is up to 5 orders of magnitude faster than conventional particle-based schemes.

DOI: 10.1103/PhysRevLett.94.128103 PACS numbers: 87.16.Ac, 02.70.−c, 05.40.−a, 87.16.Yc

In many networks of interacting components, the spatial distribution of the reactants and the stochastic character of their interactions is of crucial importance for the macroscopic behavior of the system. Examples are to be found in biological evolution and population dynamics [1], but arguably the best examples are biochemical networks [1,2]. Biochemical networks are the computational devices of living cells. They allow the living cell to detect, amplify, and integrate signals, as well as transmit signals from one place to another. Importantly, the concentrations of the components are often low and, as a result, biochemical networks can be highly stochastic. Indeed, an important question is how the ability to resist noise constrains the design of the network [3].

In principle, computer simulations are ideally suited for elucidating the design principles that allow biochemical networks to process information reliably in time and space. However, the current techniques to study biochemical networks are of limited use. The commonly used reaction-diffusion equations based upon the macroscopic chemical rate equations ignore the discrete nature of the reactants and the stochastic character of their interactions, while techniques based upon the (zero-dimensional) chemical master equation, such as the Gillespie algorithm [4], assume that at each instant the particles are uniformly distributed in space. In order to take into account both the full spatial distribution of the components and the stochastic character of their interactions, it would seem natural to use a technique based upon Brownian dynamics. However, such a technique, while correct, would be highly inefficient, because the reactant concentrations are usually low and, as a consequence, much CPU time would be wasted in propagating the particles toward one another. We have developed an event-driven algorithm, named Green's function reaction dynamics (GFRD), which uses Green's functions to combine in one step the propagation of the particles in space with the reactions between them.

Here, we apply GFRD to a model of gene expression. The calculations reveal that the event-driven nature of GFRD makes it up to 5 orders of magnitude more efficient than schemes based upon Brownian dynamics. The algorithm is generic and not limited to biochemical networks. GFRD can be applied to a wide variety of reaction-diffusion problems, including those in population dynamics, evolution, and condensed-matter physics.

Overview of the algorithm.—Figure 1 shows a typical configuration of reactants. The particles move diffusively; the circles indicate the maximum distance each particle can travel in a time step. The essence of GFRD is to determine a maximum time step, $\Delta t_{\max}$, such that only single particles or pairs of particles have to be considered

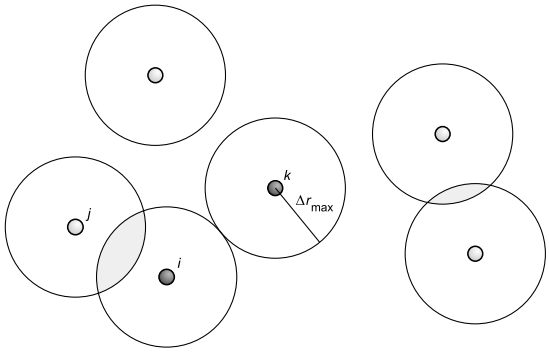

FIG. 1. Determination of the maximum time step, $\Delta t_{\max}$. The maximum size of the time step is set by the requirement that each particle can interact with at most one other particle during that time step; it can thus travel a distance of at most $\Delta r_{\max,i}$ during a time step, as indicated by the circles. We have used the operational criterion $\Delta r_{\max,i} = H\sqrt{6D_i\Delta t_{\max,i}}$, with D_i being the diffusion constant of particle i. A value of $H = 3$ was found to yield a good conservation of the spatial distribution of particles. In this example, each particle is assumed to have the same diffusion constant; the time step is limited by the constraint that particles i and k should not interact as particle i can already interact with particle j. Note that with this maximum time step the many-body problem of propagating the N particles is reduced to that of propagating single particles and pairs of particles.

(see Fig. 1). For these cases, the Smoluchowski equation [5,6] can be solved *analytically* using Green's functions. The analytical solutions can then be used to set up an event-driven algorithm [7]. In contrast to the event-driven schemes of [4,8], GFRD follows all the reactants in both time and space.

Monomolecular reactions: the Green's function for a single particle.—We first consider a single, spherical particle A that moves diffusively and can "decay" according to

$$A \xrightarrow{k_d} B + C \ldots. \tag{1}$$

We assume that reaction (1) is a Poisson process and that if a reaction event happens, it happens *instantaneously.* This means that the reaction event can be decoupled from the diffusive motion of the particle. If $k_d dt$ is the probability that a reaction event occurs in an infinitesimal time interval dt, then the probability that the *next* reaction occurs between t and $t + dt$ is given by

$$q_d(t|t_0)dt = k_d \exp[-k_d(t - t_0)]dt. \tag{2}$$

The diffusive motion of the particle is described by the Einstein diffusion equation, the solution of which is given by

$$p_1(\mathbf{r}, t|\mathbf{r}_0, t_0) = \frac{1}{[4\pi D(t - t_0)]^{3/2}} \exp\left[-\frac{|\mathbf{r} - \mathbf{r}_0|^2}{4D(t - t_0)}\right]. \tag{3}$$

Here, D is the diffusion constant and $p_1(\mathbf{r}, t|\mathbf{r}_0, t_0)$ is the probability that the particle is at position $\mathbf{r}$ at time t, given that it was at $\mathbf{r}_0$ at time t_0.

Bimolecular reactions: the Green's function for a pair of particles.—We next consider one pair of particles A and B that can react according to

$$A + B \xrightarrow{k_a} C + D \ldots. \tag{4}$$

We assume that A and B are spherical and move by diffusion with diffusion constants D_A and D_B, respectively. Furthermore, the particles react with an intrinsic rate constant k_a when they have approached each other within the reaction distance $\sigma = (d_A + d_B)/2$, where d_A and d_B are the diameters of A and B. The particles could interact via a potential $U(\mathbf{r})$ that depends upon the interparticle vector $\mathbf{r}$, although here we restrict ourselves to $U(\mathbf{r}) = 0$ for $|\mathbf{r}| > \sigma$.

The diffusive motion of such a pair of particles is described by the Smoluchowski equation [5,6]. By making the coordinate transformation $\mathbf{R} = \sqrt{D_B/D_A}\mathbf{r}_A + \sqrt{D_A/D_B}\mathbf{r}_B$ and $\mathbf{r} = \mathbf{r}_B - \mathbf{r}_A$, it can be shown that the Smoluchowski equation describes two independent random processes—free diffusion in the coordinate $\mathbf{R}$ and diffusion (with a drift) in the coordinate $\mathbf{r}$ [9]. The former process is described by

$$p_2^{\mathbf{R}}(\mathbf{R}, t|\mathbf{R}_0, t_0) = \frac{\exp[-\frac{|\mathbf{R} - \mathbf{R}_0|^2}{4(D_A + D_B)(t - t_0)}]}{[4\pi(D_A + D_B)(t - t_0)]^{3/2}}. \tag{5}$$

The Green's function for the interparticle vector $\mathbf{r}$, $p_2^{\mathbf{r}}(\mathbf{r}, t|\mathbf{r}_0, t_0)$, takes into account the reaction between A and B and is obtained via a *radiation* boundary condition on the solution of the Smoluchoswki equation for the interparticle vector $\mathbf{r}$ [9–11]. Two important quantities can be derived from this Green's function. The first is the *survival probability*

$$S_a(t|\mathbf{r}_0, t_0) = \int_{|\mathbf{r}|>\sigma} d\mathbf{r}\, p_2^{\mathbf{r}}(\mathbf{r}, t|\mathbf{r}_0, t_0). \tag{6}$$

The second quantity is the propensity function $q_a(t|\mathbf{r}_0, t_0)$, which is the probability per unit time that the *next* reaction of a pair of particles, initially separated by $\mathbf{r}_0$, occurs at time t. It is related to $S_a(t|\mathbf{r}_0, t_0)$ via

$$q_a(t|\mathbf{r}_0, t_0) \equiv -\frac{\partial S_a(t|\mathbf{r}_0, t_0)}{\partial t}. \tag{7}$$

Outline of the algorithm.—The GFRD algorithm consists of iterating the following steps.

(i) Determine maximum time step $\Delta t_{\max}$. The maximum possible time step is determined such that only single particles or pairs of particles have to be considered (see Fig. 1).

(ii) Determine next reaction and next reaction time. For each reaction R_ν, we draw a tentative next reaction time $\Delta t_\nu = t_\nu - t_0$ from the distribution $q_\nu(t - t_0) = -\partial S_\nu(t - t_0)/\partial t$ [see Eqs. (2) and (7) for the monomolecular and bimolecular reactions, respectively]. The system will then be propagated through a time Δt as given by

$$\Delta t = \min(\{\Delta t_\nu\}, \Delta t_{\max}). \tag{8}$$

Note that if there is no reaction with $\Delta t_\nu < \Delta t_{\max}$, then no reaction will occur within the time step.

(iii) Propagate particles. Single particles are propagated according to $p_1(\mathbf{r}, t|\mathbf{r}_0, t_0)$ in Eq. (3); if a particle decays, the products are placed next to each other at $\mathbf{r}$. For each pair of particles, the following two substeps are executed: (1) a new position for the coordinate $\mathbf{R}$ is obtained from Eq. (5); (2) if the pair has not reacted, a new interparticle vector $\mathbf{r}$ is obtained from $p_2^{\mathbf{r}}(\mathbf{r}, t|\mathbf{r}_0, t_0)$; otherwise, if it has reacted, the products are placed adjacent to each other at positions around $\mathbf{R}$.

(iv) Update particles. The identities of the particles are updated according to the executed reaction.

During the simulations we use a two-dimensional look-up table for $q_a(t - t_0|\mathbf{r}_0)$ and a four-dimensional table for the full solution of $p_2^{\mathbf{r}}(\mathbf{r}, t - t_0|\mathbf{r}_0)$; they are 1 Kb and 10 Mb, respectively. This procedure is more efficient than that of [9], where $p_2^{\mathbf{r}}(\mathbf{r}, t - t_0|\mathbf{r}_0)$ was constructed on the fly.

The event-driven nature makes GFRD particularly useful for networks in which the times between events are

distributed over a wide range of length and time scales, as, e.g., in biochemical networks: GFRD takes small steps when the particles are close to one another, while it takes large jumps in time and space when the particles are far apart from each other; this cannot be accomplished with methods that use a fixed-time step [12]. Further, in GFRD it never happens that more than two interaction partners are within a reaction zone, in contrast to fixed-time step schemes [12]. Most importantly, the event-driven nature allows the simulation of arbitrary complex networks: branching pathways—where a component can undergo a number of competing reactions [see Eqs. (9) and (10)]—can be handled, because only single and binary Green's functions of reactants are required to propagate the system till the next event; in contrast, fixed-time step methods [12] would here require intractable many-body Green's functions involving both reactants and products.

As a maximum time step is chosen such that the reactions occur *independently* of each other, it can be shown that GFRD yields the correct system dynamics [9]. More complicated reactions, such as trimolecular reactions, can be handled since they can be decomposed into monomolecular and bimolecular reactions. In addition, an event-driven algorithm of this type could be set up for, e.g., interacting particles [with $U(\mathbf{r}) \neq 0$ for $|\mathbf{r}| > \sigma$]—$\Delta t_{\max}$ is then determined by the *range* of $U(\mathbf{r})$—and/or those moving by other mechanisms than diffusion such as active transport. If necessary, the required Green's functions and propensity functions could be obtained numerically.

Application.—We demonstrate GFRD using the following model of gene expression:

$$A + B \underset{k_d}{\overset{k_a}{\rightleftarrows}} C, \tag{9}$$

$$C \underset{k_{\text{prod}}}{\longrightarrow} P + A + B, \tag{10}$$

$$P \underset{k_{\text{dec}}}{\longrightarrow} \varnothing \tag{11}$$

In Eqs. (9)–(11), A represents a promoter site on the DNA and B a RNA polymerase (RNAP) molecule that moves by free diffusion and that can bind with a forward rate k_a to the promoter site to form the RNAP-DNA complex C. This complex can dissociate with a rate constant k_d. Alternatively, it can produce a protein P at a production rate k_{prod}. Proteins degrade with decay rate k_{dec}.

In the simulations, we fix the promoter site A in the center of a spherical container of volume $V = 1\ \mu\text{m}^3$, which is comparable to that of the *Escherichia coli* cell. The concentration of RNAP is 30 nM [13] and its diffusion constant is $D = 1\ \mu\text{m}^2\,\text{s}^{-1}$ [14]. At contact, the RNAP associates with the promoter site at a rate determined by the Maxwell-Boltzmann velocity distribution, leading to $k_a = 3 \times 10^9\ \text{M}^{-1}\,\text{s}^{-1}$ [15]. The dissociation rate is $k_d = 21.5\ \text{s}^{-1}$, corresponding to the equilibrium constant $K = k_a/k_d$ reported in [13]. The diameters of the promoter site and the RNAP molecules are $\sigma = 5$ nm.

In Fig. 2 we show the average number of proteins $\overline{N}_P$ as a function of k_{prod}, keeping k_{decay} fixed at $0.04\ \text{s}^{-1}$. As the concentration of RNAP is low and spatial correlations are negligible, the GFRD results for $\overline{N}_P$ follow the mean-field prediction

$$\overline{N}_P = K_1 K_2 \frac{N_B}{V + K_1 N_B}, \tag{12}$$

where $K_1 = k_a/(k_d + k_{\text{prod}})$ and $K_2 = k_{\text{prod}}/k_{\text{dec}}$. However, in contrast to the mean-field analysis, GFRD also allows us to quantify the effect of the *spatial fluctuations* of the RNAP molecules on the *noise* in gene expression.

To assess the importance of spatial fluctuations we compare the GFRD results to those obtained using the zero-dimensional chemical master equation, which does take into account the stochastic character of chemical reactions, but, in contrast to GFRD, assumes that at each instant the particles are uniformly distributed in space [4]. The noise η_P in the protein concentration, defined as the standard variation divided by the mean, can be obtained analytically from the master equation [6]:

$$\eta_P^2 = \frac{1}{\overline{N}_P} - \frac{k_{\text{prod}} k_a N_B}{k_{\text{prod}} k_a N_B + \overline{N}_P (k_a N_B + k_d + k_{\text{prod}})^2}. \tag{13}$$

The second term in Eq. (13) goes through a minimum at $k_{\text{prod}} = k_a N_B + k_d$ and vanishes for both small and large k_{prod}. In these regimes, the master equation predicts that gene expression reduces to a linear birth-and-death process. In Fig. 3, we compare this prediction to that of GFRD. For small k_{prod} both approaches yield identical results. In this regime, protein synthesis is the rate-limiting step. On

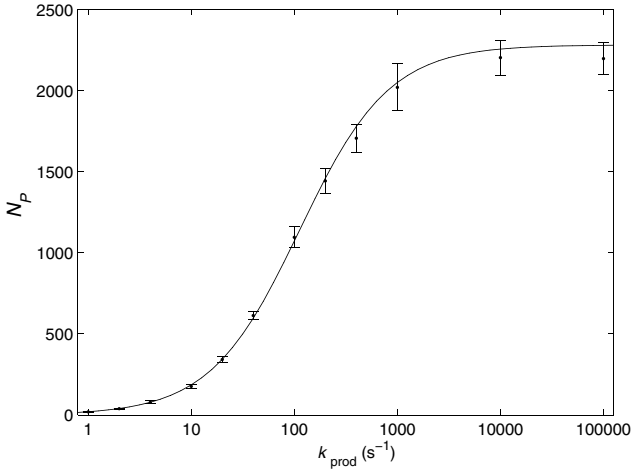

FIG. 2. The mean protein number $\overline{N}_P$ as a function of the protein production rate k_{prod} as obtained from the GFRD simulations for the reaction scheme shown in Eqs. (9)–(11). The solid line denotes the mean-field prediction given by Eq. (12).

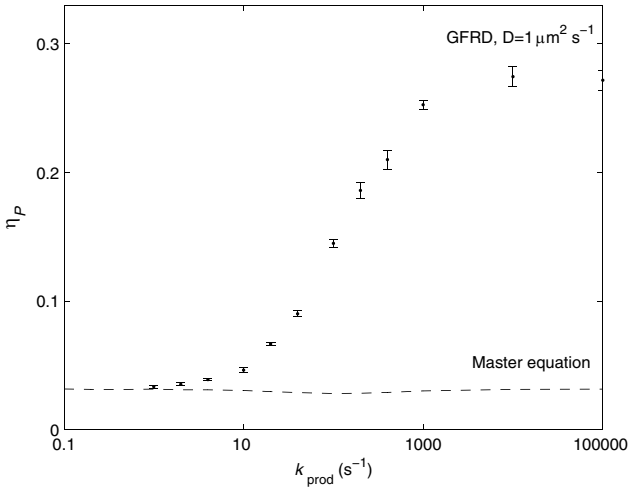

FIG. 3. The noise in protein level η_P as a function of synthesis rate k_{prod} for the reaction scheme shown in Eqs. (9)–(11). To elucidate the effect of spatial fluctuations, we fix the mean number of proteins at $\overline{N}_P = 1000$ by changing the protein degradation rate k_{decay}.

the time scale of gene expression the RNAP molecules have sufficient time to become well mixed and gene expression, indeed, reduces to a linear birth-and-death process. For $k_{\text{prod}} \gtrsim 1\ \text{s}^{-1}$, however, the noise of the spatially resolved model is larger than that of the ''well-stirred reactor'' model and grows with increasing k_{prod}. The larger noise is due to the broad distribution of arrival times of RNAP molecules at the promoter site, which is much broader than the corresponding Poisson distribution for the system without spatial fluctuations. These calculations show that spatial fluctuations could be an important source of noise in gene expression.

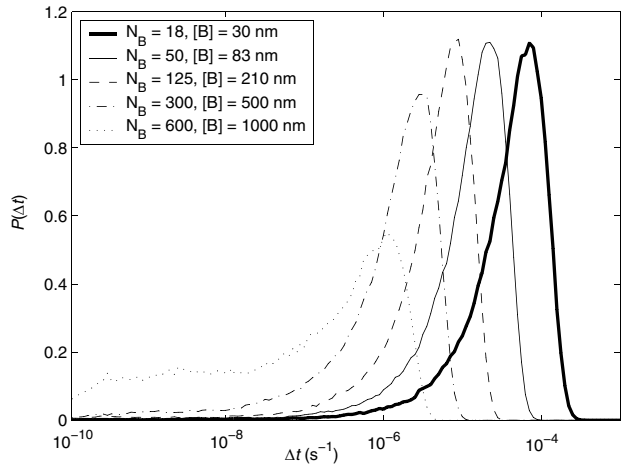

FIG. 4. The distribution of propagation times Δt in GFRD for a system consisting of a single particle A and N_B particles B that can react according to the scheme in Eq. (9). In brute-force Brownian dynamics, the time step is $\Delta t \approx 1 \times 10^{-10}$ s.

Outlook.—In biochemical networks, the reactant concentrations are often very low, ranging from nMs for gene networks to μMs for signal transduction pathways. Figure 4 suggests that with GFRD it should be possible to reach time steps of at least 10^{-6}–10^{-4} s for such networks. In contrast, with conventional Brownian dynamics [16] we typically cannot use time steps larger than 10^{-10}–10^{-9} s $[(10^{-5}$–$10^{-4})\sigma^2/D]$. Hence, even though in GFRD the required CPU time to execute one step is about a factor of 10 larger than in Brownian dynamics, the overall computational efficiency of GFRD is 2 to 5 orders of magnitude higher than that of Brownian Dynamics under biologically relevant conditions. GFRD thus brings particle-based simulations of biochemical networks on biological time scales of seconds to hours within reach.

We thank Daan Frenkel, Fred MacKintosh, Rosalind Allen, and Bela Mulder for useful discussions. The work is part of the research program of the ''Stichting voor Fundamenteel Onderzoek der Materie (FOM),'' which is financially supported by the ''Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO).''

*Electronic address: tenwolde@amolf.nl

[1] N. M. Shnerb, Y. Louzoun, E. Bettelheim, and S. Solomon, Proc. Natl. Acad. Sci. U.S.A. **97**, 10 322 (2000).
[2] Y. Togashi and K. Kaneko, Phys. Rev. Lett. **86**, 2459 (2001); Phys. Rev. E **70**, 020901(R) (2004).
[3] H. H. McAdams and A. Arkin, Proc. Natl. Acad. Sci. U.S.A. **94**, 814 (1997); M. B. Elowitz and S. Leibler, Nature (London) **403**, 335 (2000).
[4] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[5] M. Smoluchowski, Z. Phys. Chem. **92**, 129 (1917); S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[6] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).
[7] A. B. Bortz *et al.*, J. Comput. Phys. **17**, 10 (1975).
[8] E. M. Sevick and R. C. Ball, J. Colloid Interface Sci. **144**, 561 (1991).
[9] J. S. van Zon and P. R. ten Wolde, q-bio.MN/0404002.
[10] H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids* (Oxford University Press, New York, 1959).
[11] N. Agmon and A. Szabo, J. Chem. Phys. **92**, 5270 (1990).
[12] H. Kim, M. Yang, and K. J. Shin, J. Chem. Phys. **111**, 1068 (1999); A. V. Popov and N. Agmon, J. Chem. Phys. **115**, 8921 (2001).
[13] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Proc. Natl. Acad. Sci. U.S.A. **99**, 12 795 (2002).
[14] M. B. Elowitz *et al.*, J. Bacteriol. **181**, 197 (1999).
[15] H. X. Zhou and A. Szabo, J. Chem. Phys. **95**, 5948 (1991); J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, San Diego, 1986), 2nd ed.
[16] P. Meakin, Phys. Rev. Lett. **51**, 1119 (1983); M. Kolb *et al.*, Phys. Rev. Lett. **51**, 1123 (1983); M. D. Haw *et al.*, Phys. Rev. E **56**, 1918 (1997).