Effect of reactant size on discrete stochastic chemical kinetics

Dan T. Gillespie
Dan T. Gillespie Consulting, Castaic, California 91384

Sotiria Lampoudi[a] and Linda R. Petzold
Department of Computer Science, University of California, Santa Barbara, California 93106

This paper is aimed at understanding what happens to the propensity functions (rates) of bimolecular chemical reactions when the volume occupied by the reactant molecules is not negligible compared to the containing volume of the system. For simplicity our analysis focuses on a one-dimensional gas of N hard-rod molecules, each of length l. Assuming these molecules are distributed randomly and uniformly inside the real interval $[0,L]$ in a nonoverlapping way, and that they have Maxwellian distributed velocities, the authors derive an expression for the probability that two rods will collide in the next infinitesimal time dt. This probability controls the rate of any chemical reaction whose occurrence is initiated by such a collision. The result turns out to be a simple generalization of the well-known result for the point molecule case $l=0$: the system volume L in the formula for the propensity function in the point molecule case gets replaced by the "free volume" $L-Nl$. They confirm the result in a series of one-dimensional molecular dynamics simulations. Some possible wider implications of this result are discussed. © *2007 American Institute of Physics.*
[DOI: 10.1063/1.2424461]

I. INTRODUCTION

The mass action formalism for bimolecular chemical reactions, which is reflected in the combinatorial product of the reactant molecular populations in both the continuous-deterministic reaction rate and the discrete-stochastic reaction propensity function, has been a workhorse of chemical kinetics for many decades. This formalism has been shown to break down in practically all but the most "classic," three-dimensional, homogeneous systems. Fractal kinetics was developed to address this breakdown in cases where the molecules move by diffusion.[1] A separate literature explores what happens when the molecules move ballistically, as in a dilute gas.[2]

In both the diffusive and ballistic cases of motion the mass-action formalism's breakdown is thought to occur due to the eventual establishment of correlations in the positions of the molecules (i.e., the deviation from the well-stirred regime), even in cases where the systems are initially prepared as well stirred (for an overview, see Ref. 3). Low dimensionality is thought to exacerbate this effect by making "self-stirring" an insufficient way to rerandomize molecular positions.[4]

Fractal kinetic laws (in which either the phenomenological rate coefficient is made to decrease with time according to a fractional power, or the concentrations of the reactants are raised to a rational power) have been applied to a variety of systems in which reduction and fragmentation of the volume accessible to diffusing reactants are due to the presence of finite-size molecules, formation of depletion zones, and spatial gradients in the molecular concentrations. Fractal ki-

netic laws have a broad range of applicability to realistic problems.[5] The ballistic motion literature, on the other hand, has been more focused on developing an analytical, mechanistic theory of how the distribution of molecular velocities, intermolecular distances, and the resulting reaction kinetics evolve with time, given various initial conditions.[6] Neither literature has concerned itself at any length with the impact of the finite size of the molecules participating in the reactions.

In the present paper we try to answer the following question: What modifications to the mass action law are required to capture the kinetics of a bimolecular reaction in a well-stirred gas-phase system in which volume is excluded only by the reactant molecules? We should note that this scenario ignores several important features of cellular biochemistry, for instance, molecular kinetics in solution and crowding by large nonreacting molecules. However, it has the advantage of being mathematically tractable, at least in the idealized case of a one dimensional, single species system. Therefore, our aim here will be to derive, from first principles, the propensity function for a reaction of the form $A+A \rightarrow products$ in a well-stirred, one-dimensional system that contains only molecules of species A, these molecules having a finite size.

It is known that for the reactions $A+A \rightarrow A$ and $A+A \rightarrow \varnothing$ in *one* dimension, neither ballistic (gas phase) nor diffusive molecular motion can sustain a well-stirred condition over many successive reactions.[4] For such systems, maintaining the well-stirred condition would require either external stirring or a strongly coupled heat bath. However, our goal here is not to explore how the well-stirred condition might be maintained in one dimension; rather, it is to understand the effects of reactant size in three dimensions, where the well-stirred condition *can* often be sustained over many

---
[a]Author to whom correspondence should be addressed.

successive reactions in both the ballistic and diffusive motion cases. The reason we restrict our analysis to a one-dimensional system is that it is practically impossible to mathematically model a uniform, nonoverlapping spatial distribution of finite-size molecules in three (or even two) dimensions. Nevertheless, reactant size effects in a well-stirred one-dimensional system should be *strongly suggestive* of reactant size effects in well-stirred three-dimensional systems, independently of how the well-stirred condition is maintained.

Since bimolecular reactions are always initiated by a collision between two reactant molecules, our main task here will be to compute the probability $p_{col}(dt)$ that a collision will occur, in the next infinitesimal time dt, in a system of N identical hard-rod molecules, each of length l, which are confined to the real interval $[0,L]$ with $L>Nl$. Our analysis will make two important assumptions: First, the *positions* of the rods are randomly uniform in $[0,L]$, subject to the constraint that no two rods overlap. Second, the system is at some constant absolute temperature T, so that the *velocities* of the molecules can be regarded as independent normal (Gaussian) random variables with mean zero and variance,

$$\sigma^2 \equiv k_B T/m. \tag{1}$$

Here, k_B is Boltzmann's constant and m is the mass of a molecule. Equation (1) can be seen most easily from the equipartition theorem, which says that in thermal equilibrium the average kinetic energy per molecule is $\langle \frac{1}{2}mv^2 \rangle = \frac{1}{2}k_B T$. The normality of the velocity distribution follows from the Maxwell-Boltzmann formula. Since each rod velocity v is an independent normal random variable with mean 0 and variance σ^2, it follows that the difference in the velocities of any two rods, i.e., their relative velocity v_{rel}, is a normal random variable with mean 0 and variance $2\sigma^2$. It is not hard to show that the average of the absolute value of v_{rel}, i.e., the average relative speed of a randomly chosen pair of rods, is

$$s_{rel} \equiv \langle |v_{rel}| \rangle = \sqrt{\frac{4}{\pi}}\sigma. \tag{2}$$

Note that, although the initial configuration of the system is chosen randomly, so as to mimic a well-stirred, thermally equilibrated state, the subsequent evolution of the system up until the first collision is entirely deterministic: the molecules move on the interval with their assigned velocities as though the system were isolated. However, in our subsequent analysis we do not continue the evolution of the system beyond the first collision (if we did, we would notice the system "cooling" as fast-moving molecules are depleted, and correlations between the velocities of the molecules would eventually start to form). We make this choice because the focus of this paper is not isolated chemical systems, but rather systems that are in thermal contact with their surroundings, as for instance, a typical cellular system would be.

In Sec. II we review the standard derivation of $p_{col}(dt)$ in the dilute $(l=0)$ case. In Sec. III we give our derivation for the dense $(l>0)$ case, using the mathematical properties of a randomly uniform, nonoverlapping distribution of rods in a fixed interval as derived in Appendices A and B. In Sec. IV

we present some molecular dynamics simulation results that confirm our theoretical results, making use of the algorithm derived in Appendix C for uniformly distributing the rods in $[0,L]$ in a nonoverlapping way. We conclude in Sec. V by discussing some possible implications of these results.

II. THE DILUTE CASE

To compute $p_{col}(dt)$ in the dilute gas limit, where each hard-rod molecule is essentially a point, we begin by choosing from the system a pair of molecules at random, without regard for their positions. In an infinitesimal time dt, one of these chosen molecules will sweep out relative to the other an average distance $s_{rel}dt$. If the second molecule happens to lie inside the length swept out by the first molecule, the two molecules will collide in the next dt. Since the molecules are distributed uniformly inside the interval $[0,L]$, the probability that they will collide in the next dt is simply the ratio of the size of the interval the second molecule needs to be in for a collision to occur to the size of the interval the second molecule is known to be in: $s_{rel}dt/L$. We invoke here the fact that the space occupied by the point molecules themselves is zero, or at least negligible compared to L. Since there are $N(N-1)/2$ distinct ways in which the pair of molecules could have originally been chosen, it follows by the addition law of probability that the probability that some one of those pairs will collide in the next dt is

$$\frac{1}{2}N(N-1)\left(\frac{s_{rel}dt}{L}\right).$$

Thus we conclude that, in the dilute limit $l=0$,

$$p_{col}(dt) = \frac{N(N-1)s_{rel}}{2L}dt. \tag{3}$$

In the foregoing argument, the vanishing smallness of dt ensured that the probability for two or more pairs of molecules to collide in the next dt is negligibly small compared to the probability for only one pair to collide in dt. That in turn implied that pair collisions in time dt can effectively be regarded as mutually exclusive events, and this is what allowed us to invoke the addition law of probability theory to obtain Eq. (3). Equation (3) is thus valid only for infinitesimally small dt.

Some may feel that the foregoing derivation is flawed because it appears to accord the same collision probability to all pairs of molecules, regardless of how far apart the members of the pair might be. But this criticism is unjustified. The single pair collision probability $s_{rel}dt/L$ is expressly conditioned on the premise that the relative locations of the two chosen molecules inside $[0,L]$ are not known—that is what was meant by stipulating that the two molecules be chosen "at random." In a sense, $s_{rel}dt/L$ is the single-pair collision probability averaged over all possible relative positions of the pair members. As has been shown elsewhere,[7] a derivation that explicitly computes the collision probabilities for adjacent point molecules, which are the only ones that can physically collide, yields the same result [Eq. (3)].

III. THE DENSE CASE

If each of the N molecules in the system just considered had a nonzero extent l, how would the result [Eq. (3)] be altered? Naively, we might suppose that since a portion Nl of the interval $[0,L]$ is occupied by the rods themselves, then the probability that the center of one randomly chosen rod will lie inside the infinitesimal collision length $s_{rel}dt$ of the other randomly chosen rod would be the ratio of the collision length $s_{rel}dt$ to the free interval length $(L-Nl)$. This logic would result in simply replacing L in Eq. (3) with $(L-Nl)$. But further reflection might suggest that things should not be that simple: In a situation where the gap between two adjacent rods is less than l, that gap is not available for occupation by another rod, so perhaps it, too, should somehow be counted as part of the interval that is excluded by the rods themselves. We shall show in this section that this last concern is unfounded, and that the seemingly naive replacement of L in Eq. (3) with $(L-Nl)$ does in fact yield the correct generalization of Eq. (3) to the case $l>0$.

We start by choosing at random a pair of adjacent rods, noting that only adjacent rods are capable of colliding with each other. Our selection of this adjacent pair is "random" in the sense that it is made irrespective of the positions and the velocities of the two rods that comprise the pair. All we know about these two rods is that the gap ξ between them is randomly distributed in a way that is consistent with a randomly uniform, nonoverlapping distribution of N rods inside the interval $[0,L]$, and further that their relative velocity v_{rel} is randomly distributed according to the normal (Gaussian) density function $f_G(v_{rel})$ with mean 0 and variance $2\sigma^2$.

The probability density function $P_{gap}(\xi)$ for the inter-rod gap distance is derived in Appendices A and B, and the result is shown in Eq. (B5). Using that function, we assert that the probability that our randomly chosen adjacent pair will collide in the next "nearly" infinitesimal time interval δt is, to first order in δt,

$$p_{col}^{s.p.}(\delta t) = \int_{\xi=0}^{L-Nl} P_{gap}(\xi)d\xi \left(\int_{v_{rel}=-\infty}^{-\xi/\delta t} f_G(v_{rel})dv_{rel} \right) + o(\delta t).$$

(4)

The superscript "s.p." signifies that this is the collision probability for a single pair of adjacent rods, and we note that there are $N-1$ such pairs in the system. The integral in parentheses in Eq. (4) is the probability that the relative velocity of the two rods will have a sufficiently large negative value that the rods will collide in the next δt if their present gap separation is ξ. That probability is multiplied by the probability that the gap separation will be infinitesimally close to ξ, and the result is then summed over all possible values of ξ to obtain the desired collision probability. The term $o(\delta t)$, which goes to zero with δt faster than δt but is otherwise unspecified, accounts for errors in this reasoning caused by interfering collisions of either rod with its other neighbor (or a boundary). The basic assumption being made here is that the probability for two or more collisions to occur in time δt can be made negligibly small compared to the probability of only one collision simply by taking δt close enough to zero. In the end, we will be interested only

in the case in which δt is taken to be infinitesimally small.

Changing the limits in the v_{rel} integral in Eq. (4) from $(-\infty, -\xi/\delta t)$ to $(\xi/\delta t, \infty)$ which is permissible because $f_G(v_{rel})$ is symmetric gives

$$p_{col}^{s.p.}(\delta t) = \int_{\xi=0}^{L-Nl} d\xi \int_{v_{rel}=\xi/\delta t}^{\infty} dv_{rel}[P_{gap}(\xi)f(v_{rel})] + o(\delta t).$$

Next we change the order of integration over ξ and v_{rel},

$$\int_{\xi=0}^{L-Nl} d\xi \int_{v_{rel}=\xi/\delta t}^{\infty} dv_{rel} = \int_{v_{rel}=0}^{(L-Nl)/\delta t} dv_{rel} \int_{\xi=0}^{v_{rel}\delta t} d\xi$$
$$+ \int_{v_{rel}=(L-Nl)/\delta t}^{\infty} dv_{rel} \int_{\xi=0}^{L-Nl} d\xi$$
$$= \int_{v_{rel}=0}^{\infty} dv_{rel} \int_{\xi=0}^{v_{rel}\delta t} d\xi + o(\delta t).$$

The first step here can be seen by graphing the integration region in (ξ, v_{rel}) space. The second step follows by taking δt so small that the limits $(L-Nl)/\delta t$ in the two v_{rel} integrations can be effectively replaced by ∞. Our assumption that doing this introduces errors of higher order in δt than the first is justified by the fact that the Gaussian function $f_G(v_{rel})$ falls off extremely rapidly with increasing v_{rel}, and in practice there will always be some finite upper bound on that variable. So now we have

$$p_{col}^{s.p.}(\delta t) = \int_0^{\infty} dv_{rel} \left(f_G(v_{rel}) \int_0^{v_{rel}\delta t} d\xi[P_{gap}(\xi)] \right) + o(\delta t).$$

Since $P_{gap}(0)>0$, the inner integral over ξ can be easily evaluated to first order in δt,

$$\int_0^{v_{rel}\delta t} d\xi[P_{gap}(\xi)] = P_{gap}(0)(v_{rel}\delta t) + o(\delta t).$$

Therefore,

$$p_{col}^{s.p.}(\delta t) = \int_0^{\infty} dv_{rel}(f_G(v_{rel})P_{gap}(0)v_{rel}\delta t) + o(\delta t)$$

$$= P_{gap}(0)\delta t \left(\frac{1}{2} \int_{-\infty}^{\infty} |v_{rel}|f_G(v_{rel})dv_{rel} \right) + o(\delta t),$$

(5)

$$p_{col}^{s.p.}(\delta t) = \left(\frac{N}{L-Nl} \right) \delta t \left(\frac{1}{2}s_{rel} \right) + o(\delta t).$$

In the second step we have invoked the evenness of the function $f_G(v_{rel})$. In the last step we have invoked the value of the gap density function $P_{gap}(\xi)$ at $\xi=0$, as prescribed by Eq. (B5), and also the definition [Eq. (2)] of s_{rel}.

Recalling now that there were $N-1$ possible ways in which this adjacent pair of rods could have been chosen, we apply the addition law of probability once more to infer that the probability that *some* one of those adjacent pairs will collide in time δt is

$$p_{\text{col}}(\delta t) = (N-1)p_{\text{col}}^{\text{s.p.}}(\delta t) + o(\delta t). \qquad (6)$$

Again, a term $o(\delta t)$ accounts for corrections to the addition law caused by the small possibility that more than one pair of rods might collide in time δt. Substituting Eq. (5) into Eq. (6), and finally taking the limit $\delta t \to dt$ so that the $o(\delta t)$ term can finally be dropped, we obtain the principal result of this paper,

$$p_{\text{col}}(dt) = \frac{N(N-1)s_{\text{rel}}}{2(L-Nl)}dt. \qquad (7)$$

We observe that putting $l=0$ in Eq. (7) indeed gives the expected dilute gas result [Eq. (3)]. More remarkable, though, is the implication that the dilute result [Eq. (3)] can be generalized to the dense case simply by replacing the interval length L with the free interval length, $L-Nl$. This goes against the supposition that, in estimating the available system volume, allowance must somehow be made for gaps between molecules that are smaller than a molecular diameter, since such gaps are not available for occupancy by a molecule. We suggest that the intuitive basis for that concern stems from trying to solve the related but different problem of uniformly scattering the rods inside the interval $[0,L]$, where the nonoverlap condition would indeed preclude placing a rod between two already placed rods that are separated by a distance less than l. But, as is shown in Appendix C, the most efficient algorithm for randomly placing the rods inside $[0,L]$ in a uniform, nonoverlapping way "deals out" the rods from left to right in that interval, and thus never tries to place a rod between two already placed rods.

The result [Eq. (7)] can be viewed from two different perspectives,

$$p_{\text{col}}(dt) = \begin{cases} \frac{1}{2}N(N-1)\dfrac{s_{\text{rel}}dt}{(L-Nl)} \\[2ex] (N-1)\dfrac{Ns_{\text{rel}}dt}{2(L-Nl)}. \end{cases} \qquad (8)$$

The first form in Eq. (8) is the probability that a randomly chosen pair of rods will collide in the next dt, multiplied by the number of such pairs. The second form is the probability that a randomly chosen adjacent pair of rods will collide in the next dt, multiplied by the number of such pairs. This dual perspective on $p_{\text{col}}(dt)$ was pointed out in Ref. 7 for the dilute case $l=0$, with the aim of demonstrating that both of these lines of reasoning lead to the same correct formula for the dt-collision probability. Our present result therefore extends that earlier finding to the case $l>0$.

If every collision between two A molecules produced a chemical reaction, then the propensity function for that reaction would simply be the coefficient of dt in Eq. (7). More generally, the propensity function should contain as a factor the probability $p_{\text{reac|coll}}$ that a colliding pair of molecules will actually undergo the reaction. For simplicity, however, in the following discussion we shall omit this factor and assume that every collision leads to a reaction.

IV. VERIFICATION OF THE EXPONENTIAL DISTRIBUTION

The logic used in discrete stochastic chemical kinetics to deduce the bimolecular propensity function—a logic which forms the foundation for both the chemical master equation and the stochastic simulation algorithm[8]—interprets the result [Eq. (7)] as implying that the time τ_{next} to the next collision/reaction is an *exponential random variable* whose mean is the reciprocal of the coefficient of dt,

$$\langle \tau_{\text{next}} \rangle = \frac{2(L-Nl)}{N(N-1)s_{\text{rel}}}. \qquad (9)$$

But this result requires Eq. (7) to hold at the beginning of *every* infinitesimal time interval prior to the first collision,[9] and hence that at all those times the system be well stirred. Of course, the system is well stirred, by assumption, at the beginning of the first infinitesimal interval $[0,dt)$. But can the same be said for the succeeding intervals $[dt,2dt)$, $[2dt,3dt)$, etc.? A simple heuristic argument giving an affirmative answer to this question goes as follows: So long as no collision has occurred before a given dt interval, we will not have learned anything at all about the positions of the rods beyond what we knew at time $t=0$; thus, for any such dt interval we may regard the system as still being well stirred. To put it in another way, we are suggesting that the Maxwellian ballistic motion of the rods, at least up until the moment of the first collision, does not undo the randomly uniform spatial distribution of those rods.

To see if this expectation is true, and also to check the validity of formula (9), we have performed the following molecular dynamics experiment:

- Using the generating formula [Eq. (C4)], deal out a configuration of the positions of the N rods in $[0,L]$. (This distributes the molecules randomly and uniformly in $[0,L]$.)

- Assign each rod a velocity in the form of an independent sample of the normal random variable with mean 0 and standard deviation $\sigma = \sqrt{\pi/4}s_{\text{rel}}$. (This puts the molecules in thermal equilibrium at temperature T [see Eqs. (1) and (2)].)

- Allow the rods to move accordingly, with rods 1 and N reflecting elastically off the boundaries, until the first collision between a pair of rods occurs, and record the time of that collision.

- Repeat the above steps to obtain many samples of the time to the first collision/reaction, and then make normalized frequency histograms of those times.

We carried out this experiment for a range of values for N and the fraction of the interval occupied by the molecules themselves, κN, where κ is the dimensionless parameter defined by
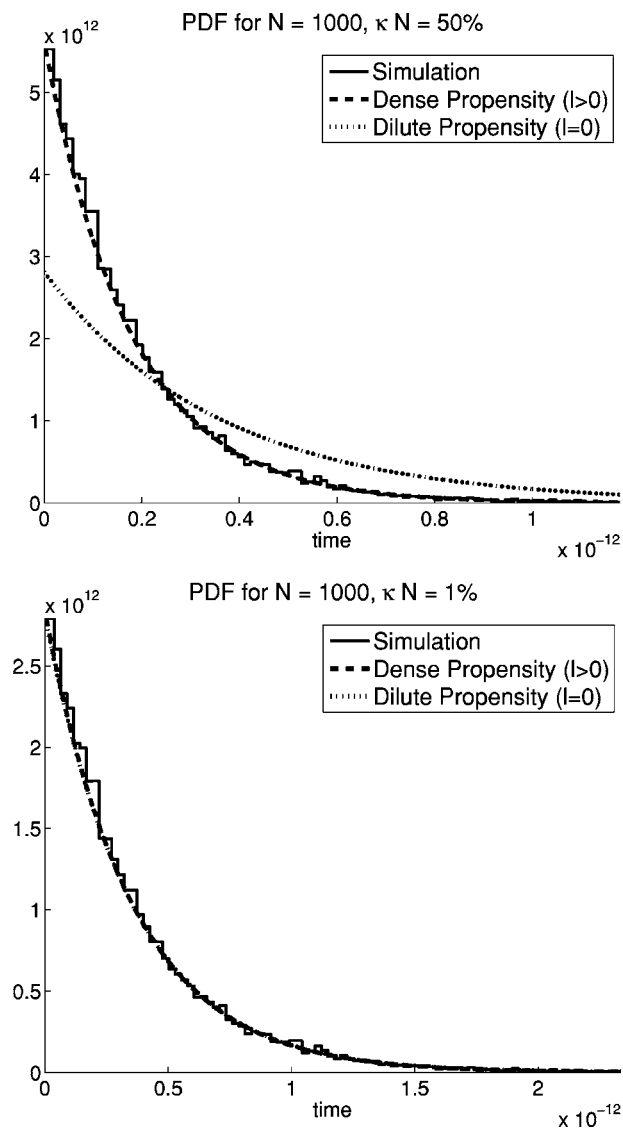
FIG. 1. Histograms of simulation vs analytical pdf's (dense and dilute exponentials) for population $N=1000$ molecules, and $\kappa N=50\%$ and 1%. Here $\kappa = l/L$, where L is the "volume" of the system and l is the "size" of one molecule, so κN is the fraction of the volume of the system occupied by the molecules themselves. Each simulation ensemble contains 20 000 samples. The histograms are computed using 140 equal-sized bins.

$$\kappa \triangleq \frac{l}{L}. \tag{10}$$

We obtained for each case 20 000 samples of the time to the first collision/reaction. The estimated probability density function (pdf) of the time to the first collision is shown as the solid histograms in Figs. 1 and 2. The dashed curves in these figures are the pdf's for the exponential distribution with the dense-case mean [Eq. (9)] and the dotted curves are the pdf's for the exponential distribution for the dilute-case mean with $l=0$.

These results show that the distribution of τ_{next} is, as conjectured, accurately described by an exponential distribution with the dense-case mean [Eq. (9)]. For a very low fraction of volume occupied ($\kappa N=1\%$), the difference between the three curves is practically impossible to detect. As the fraction of volume occupied by the molecules increases
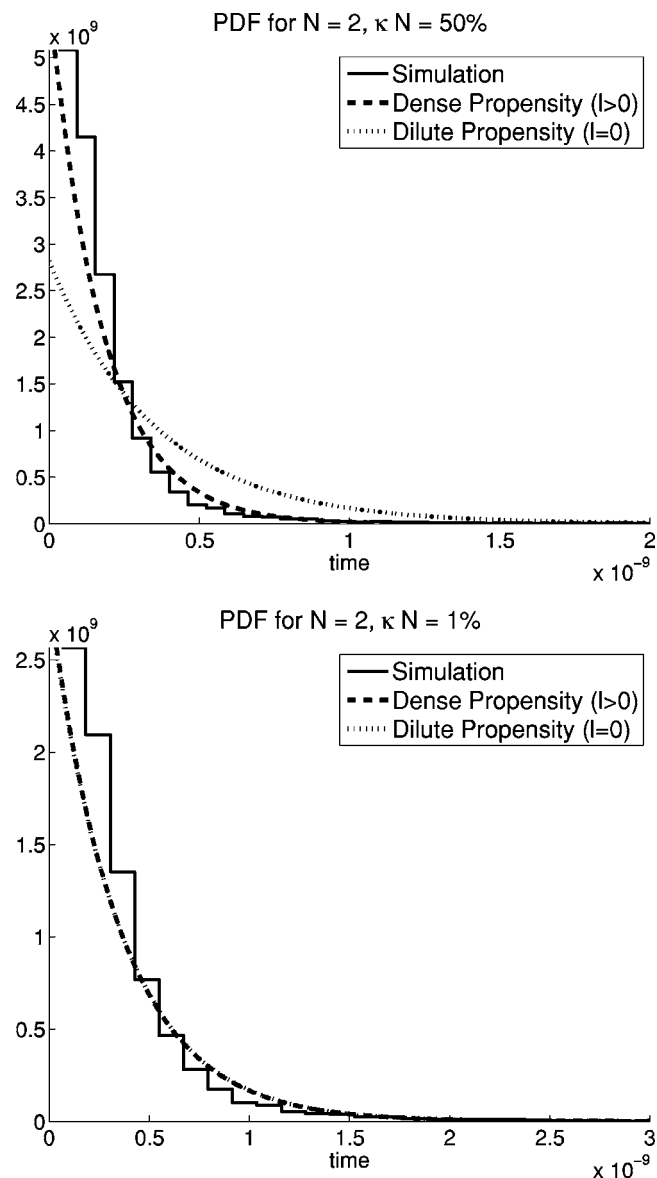

FIG. 2. Histogram of simulation vs analytical pdf's for $N=2$. All other parameters are the same as for Fig. 1.

($\kappa N=50\%$), the difference between the dilute and dense exponential distributions becomes pronounced, and the simulation results in every case match the dense distributions. This happens consistently over reactant populations ranging from $N=1000$ (Fig. 1) to $N=4$ (not shown). For $N=3$ (not shown) and $N=2$ (Fig. 2), a slightly larger difference between the analytical and simulation curves is observed. We attribute this discrepancy to the fact that collisions with the hard boundaries are frequent enough as to be non-negligible when only two or three molecules are present in the system, and it vanishes already at $N=4$. Nevertheless, even for $N=2$ and $\kappa N=50\%$, the dense exponential curve provides a reasonably good approximation to the simulation result.

V. SUMMARY AND DISCUSSION

In this paper, we have derived a formula for the propensity function of the *well-stirred* reaction $A+A \rightarrow products$, assuming that the N molecules of species A are hard rods of

length l that move ballistically with Maxwellian velocities in the one-dimensional interval $[0, L]$. Our formula generalizes the known $l=0$ formula by replacing the system volume L with the "free volume" $(L-Nl)$; perhaps surprisingly, it shows no effect from spaces between adjacent rods that are smaller than l. We also reported results of molecular dynamics experiments that confirm not only the accuracy of our derived formula but also the exponentially distributed nature of the next-collision time that is implicit in both the chemical master equation and the stochastic simulation algorithm.

In the molecular dynamics experiments described in Sec. IV, the initial state of the system was chosen randomly—according to a uniform distribution of molecular positions and a Maxwell-Boltzmann distribution of molecular velocities—but thereafter, until the instant of the first collision/reaction, the system evolved as though it were mechanically isolated. Most real systems of interest are not isolated. A cellular system, for instance, is typically immersed in some form of heat bath with which it interacts, in an essentially random way, to keep its temperature (but not its energy) constant. Since it is not clear how such heat bath interactions should be modeled (see, for example, Kopelman[3] for three alternatives), the following simple but admittedly artificial approach is implicit in our treatment here: During the time between successive collisions the system is treated as though it were isolated. The effects of the random interactions with the heat bath are then assimilated all at once at the moment of each collision, thereby establishing a new spatially homogeneous and thermally equilibrated initial state for the next collision. In the absence of a heat bath, the system would typically "cool" as a result of the removal of usually faster molecules by the reactions. But that thermally isolated case is not of interest to us here, and that is why we did not continue our molecular dynamics simulations beyond the time of the first collision.

We recognize that the well-stirred condition cannot be sustained in this *one*-dimensional system without some sort of external stirring. However, our main interest here is in what our findings suggest for *three-dimensional* systems, where the natural motion of the molecules often *is* sufficient to maintain a well-stirred condition. It seems likely to us that, for such three-dimensional systems, the containing volume in the propensity function formula should similarly be reduced by something like the close-packed volume of the reactant molecules.

It is worth noting the well-behaved nature of our propensity function implied by Eq. (7). As the size of the molecules goes to zero, our dense propensity function limits to the well-known point molecule result. Further, the divergence of our propensity function as the size of an A molecule becomes so large that N of those molecules occupy the entire interval L seems physically quite reasonable, since in that limit the probability of a collision in the next dt should approach something close to 1. More generally, it seems reasonable to expect that, even under diffusing conditions, the N dependence of the propensity function should change as the ratio of the size of an A molecule to the size of the system volume (the ratio κ in our analysis) changes.

Another interesting property of Eq. (7) is revealed by expanding that formula in a power series, using the dimensionless parameter κ defined in Eq. (10),

$$
\begin{aligned}
a(N) &= \left(\frac{s_{\text{rel}}}{L}\right)\frac{\frac{1}{2}N(N-1)}{(1-\kappa N)} \\
&= \left(\frac{s_{\text{rel}}}{L}\right)\frac{1}{2}N(N-1)(1+\kappa N+\kappa^2 N^2+\cdots).
\end{aligned} \tag{11}
$$

Form (11) shows how the usual "combinatorial" dependence of the propensity function on the number of molecules N gets altered by molecular crowding. For a given system volume size L, as the size l of an A molecule goes from zero, where the propensity function behaves quadratically, to its maximum value L/N, where the propensity diverges, the propensity passes successively through stages in which its N dependence can be roughly approximated by a power law, $a(N) \sim kN^X$, where $X=2,3,4\ldots$. As κ is taken larger and larger, the dominant term in parentheses on the right of Eq. (11) will be one with an increasingly higher exponent on N. An amusing instance of this power-law-like fit occurs for $l=0.65$ and $L=1000$: for this value of κ our propensity function mimics rather closely the $dN/dt \sim N^3$ behavior that is expected in a one-dimensional isolated, diffusion-limited system.[1]

There are two directions in which the present work could be extended. First, we would like to see if this result does indeed extend to higher dimensions. However, because there exists no efficient (rejection-free) algorithm for uniformly scattering nonoverlapping molecules in two and three dimensions, numerical simulations of the kind we have done here for those cases would be extremely difficult to carry out. Second, the question begs to be asked: what is the impact of having more than one reacting species, or a mix of reacting and nonreacting species? The challenge in that case is that, in one dimension, the problems of volume segmentation by nonreacting molecules and of reaction front and spatial heterogeneity formation are greatly exacerbated.
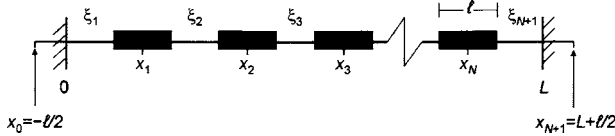
FIG. 3. Schematic of N hard rods, each of length l, distributed inside the x-axis interval $[0,L]$. The variables x_n locate the centers of the rods, and the variables ξ_n measure the gaps between the rods.

APPENDIX A: UNIFORM DISTRIBUTION OF HARD RODS ON A FIXED INTERVAL

Let N rods, each of length l, be randomly and uniformly distributed over the x-axis interval $(0,L)$ subject to the condition that no rods overlap (the rods are thus "hard"). Let the rods be numbered from left to right, and let x_n, for $n = 1, \ldots, N$, locate the center of the nth rod (see Fig. 3). Let $P(x_1, \ldots, x_N)$ be the joint probability density function of these N position variables. We shall prove here that

$$P(x_1, \ldots, x_N) = \prod_{n=1}^{N} P_{n|n-1}(x_n | x_{n-1}), \qquad (A1)$$

where the probability density function of x_n conditioned on x_{n-1} is given by

$$P_{n|n-1}(x_n | x_{n-1}) = \frac{N - n + 1}{(L_{n-1} - x_{n-1})^{N-n+1}} (L_n - x_n)^{N-n}$$

$$(x_{n-1} + l < x_n < L_n; n = 1, \ldots, N). \qquad (A2)$$

Here the parameter x_0 is defined by

$$x_0 \equiv -\tfrac{1}{2} l, \qquad (A3)$$

and the $N+1$ parameters $L_0, \ldots, L_N$ are defined by

$$L_n \equiv L - \tfrac{1}{2} l - (N - n)l \quad (n = 0, \ldots, N), \qquad (A4a)$$

or, equivalently, by the recursion relation

$$L_n \equiv \begin{cases} L - \tfrac{1}{2} l & \text{for } n = N \\ L_{n+1} - l & \text{for } n = N-1, N-2, \ldots, 0. \end{cases} \qquad (A4b)$$

A close examination of Eq. (A4b) in conjunction with Fig. 3 will reveal that, for each $n = 1, \ldots, N$, L_n is the upper limit on x_n, i.e., the value x_n would have if rods n through N were all shoved as far right in the interval $(0,L)$ as possible.

A systematic derivation of the above result was given in Ref. 10 (although there the rods were numbered from right to left instead of from left to right). Since that derivation is lengthy and formal, we shall give here a derivation that is shorter and more transparent. We begin by noting that any joint probability density function $P(x_1, \ldots, x_N)$ can always be written in the "conditioned" form

$$P(x_1, \ldots, x_N) = \prod_{n=1}^{N} P_n(x_n | x_{n-1}, x_{n-2}, \ldots, x_1),$$

where $P_n(x_n | x_{n-1}, \ldots, x_1) dx_n$ gives the probability that the nth variable will have a value in $[x_n, x_n + dx_n)$ given that the $n-1$ lower indexed variables have the respective values $x_{n-1}, \ldots, x_1$. But it is physically clear in our problem (see Fig.

3) that once the position of rod $n-1$ has been fixed, the positions of all the lower indexed rods supply no additional information about the possible position of rod n. Thus, for our problem, the above general conditioning formula reduces to the simpler form [Eq. (A1)].

To establish Eq. (A2), we shall derive an expression for the probability $P_{n|n-1}(x_n | x_{n-1}) dx_n$ that the center of the nth rod from the left will lie in the infinitesimal interval $[x_n, x_n + dx_n)$, given that the center of the $(n-1)$th rod from the left is at x_{n-1}. The "given" here implies that the nth rod from the left, which we label n, will have its center somewhere in the interval $[x_{n-1} + l, L_n)$, and further that the first rod to its right, which we will label $n+1$, will have its center somewhere in the interval $[x_{n-1} + 2l, L_{n+1})$. In general, the kth rod to the right of rod n, which we label $n+k$, will have its center somewhere in the interval $[x_{n-1} + (1+k)l, L_{n+k})$. Now we make the observation that, although the boundaries of these n confining intervals are all different, their lengths are all equal to $L_{n-1} - x_{n-1}$: Thus, the length of the interval containing the center of rod n is

$$L_n - (x_{n-1} + l) = (L_n - l) - x_{n-1} = L_{n-1} - x_{n-1},$$

where the last step invokes the second of relations (A4b), and the length of the interval for the center of rod $n+k$ for any $k = 1, \ldots, N-n$ is

$$L_{n+k} - [x_{n-1} + (1+k)l] = [L_{(n-1)+(1+k)} - (1+k)l] - x_{n-1}.$$

But since the second of relations (A4b) implies that $L_n = L_{n+k} - kl$, the quantity on the right in brackets is L_{n-1}, so the interval length is again $L_{n-1} - x_{n-1}$. So we see that, once the value of x_{n-1} is known, the values of every one of the higher-indexed variables $x_n, \ldots, x_N$ will be restricted to an interval of length $L_{n-1} - x_{n-1}$.

To compute the probability that the center of rod n will lie inside the infinitesimal interval dx_n at x_n, given that rod $n-1$ has its center at x_{n-1}, we first choose at random any one of the $N - n + 1$ rods to the right of rod $n-1$, without regard for its position or ordering. All we know about this randomly chosen rod is that its center must be confined to some interval of length $L_{n-1} - x_{n-1}$, therefore, the probability that its center will actually lie in a particular interval of length dx_n is just the ratio of those two interval lengths,

$$\frac{dx_n}{L_{n-1} - x_{n-1}}.$$

If this dx_n subinterval is located at the value x_n, then our randomly chosen rod will be the leftmost of the rightmost $N - n + 1$ rods if and only if all the $N - n$ remaining rods lie to the right of x_n. That in turn means, by the result developed in the preceding paragraph, that the center of each one of those $N - n$ other rods must be restricted to an interval of length $L_n - x_n$. Since all we know about any one of those other rods is that it is confined to an interval of length $L_{n-1} - x_{n-1}$, the probability that it will actually lie inside a smaller interval of length $L_n - x_n$ is the ratio

$$\frac{L_n - x_n}{L_{n-1} - x_{n-1}}.$$

We conclude, then, that the net probability for the center of the chosen rod to lie in the infinitesimal interval dx_n at x_n and for all the $N-n$ other rods to lie to the right of x_n (thus making the chosen rod the left-most of those $N-n+1$ rods) is

$$\frac{dx_n}{L_{n-1} - x_{n-1}}\left(\frac{L_n - x_n}{L_{n-1} - x_{n-1}}\right)^{N-n}$$

$$= \frac{dx_n}{(L_{n-1} - x_{n-1})^{N-n+1}}(L_n - x_n)^{N-n}.$$

Finally, since there are $N-n+1$ ways in which we could have chosen that first rod, it follows that the probability $P_n(x_n|x_{n-1})dx_n$ must be equal to $N-n+1$ times the above probability. This establishes result (A2).

APPENDIX B: THE GAP DISTRIBUTION

In the problem addressed in Appendix A, number the gaps between adjacent rods so that gap n is bounded on the right by rod n and/or on the left by rod $n-1$. The length ξ_n of gap n will thus be given by (see Fig. 3)

$$\xi_n = x_n - x_{n-1} - l \quad (n = 1, \ldots, N+1). \tag{B1}$$

Here we have recalled definition (A3) of x_0, and additionally defined

$$x_{N+1} \equiv L + \tfrac{1}{2}l. \tag{B2}$$

Counting the two gaps adjacent to the interval boundaries, there are a total of $N+1$ gaps; however, only N of these gaps are algebraically independent, since the sum of all the gaps must be the unoccupied length $L-Nl$.

Let us calculate the probability density function of the length of gap 1, the gap between the origin, and the left edge of rod 1. From Eqs. (B1) and (A3) we see that the size of that gap is related to x_1 by (see Fig. 3)

$$\xi_1 = x_1 - \left(-\tfrac{1}{2}l\right) - l = x_1 - \tfrac{1}{2}l.$$

The density function of x_1 is given by Eq. (A2) with $n=1$,

$$P_{1|0}(x_1|x_0) = \frac{N}{(L_0 - x_0)^N}(L_1 - x_1)^{N-1} \quad (x_0 + l < x_1 < L_1)$$

$$= \frac{N}{(L - Nl)^N}\left(L - Nl + \tfrac{1}{2}l - x_1\right)^{N-1}$$

$$\left(\tfrac{1}{2}l < x_1 < L - Nl + \tfrac{1}{2}l\right),$$

where the last step has used definitions (A3) and (A4). Note that since x_0 is a constant, this is really the unconditioned density function of the variable x_1. The unconditioned density function of the variable ξ_1 can therefore be computed by applying the change-of-variable rule in probability theory,

$$\hat{P}_{1|0}(\xi_1|x_0) = P_{1|0}(x_1|x_0)\left|\frac{dx_1}{d\xi_1}\right|.$$

Here, x_1 on the right is now understood to be the function $x_1(\xi_1) = \xi_1 + \tfrac{1}{2}l$. This immediately yields the result

$$\hat{P}_{1|0}(\xi_1|x_0) = \frac{N}{(L - Nl)^N}(L - Nl - \xi_1)^{N-1}$$

$$(0 < \xi_1 < L - Nl). \tag{B3}$$

It turns out that all the other gaps $\xi_2, \xi_3, \ldots, \xi_N$ have this same unconditioned density function. We shall prove this explicitly for the gap variable ξ_2, and then consider the result for the remaining gaps to be obvious on grounds of symmetry.

To calculate the unconditioned density function for ξ_2, we need to first compute the unconditioned joint density function for x_1 and x_2,

$$P_{1,2|0}(x_1, x_2|x_0) = P_{1|0}(x_1|x_0)P_{2|1}(x_2|x_1)$$

$$= \frac{N(N-1)}{(L_0 - x_0)^N}(L_2 - x_2)^{N-2}$$

$$= \frac{N(N-1)}{(L - Nl)^N}\left(L - Nl + \tfrac{3}{2}l - x_2\right)^{N-2}.$$

The second step above follows from Eq. (A2), and the last step follows from definitions (A3) and (A4). The domain of definition of this joint density function (outside of which the function vanishes) is

$$(x_0 + l < x_1 < L_1; x_1 + l < x_2 < L_2)$$

$$= \left(\tfrac{1}{2}l < x_1 < L - Nl + \tfrac{1}{2}l; x_1 + l < x_2 < L - Nl + \tfrac{3}{2}l\right),$$

where the last step follows from definitions (A3) and (A4). Now we make the transformation of variables $(x_1, x_2) \rightarrow (x_1, \xi_2)$, where in accordance with Eq. (B1), $\xi_2 = x_2 - x_1 - l$. The joint density function of the new pair of variables is

$$\hat{P}_{1,2|0}(x_1, \xi_2|x_0) = P_{1,2|0}(x_1, x_2|x_0)\left|\frac{\partial(x_1, x_2)}{\partial(x_1, \xi_2)}\right|,$$

where x_2 on the right is now regarded as the function $x_2(x_1, \xi_2) = \xi_2 + x_1 + l$. The Jacobian here is easily shown to be unity, so we have

$$\hat{P}_{1,2|0}(x_1, \xi_2|x_0) = \frac{N(N-1)}{(L - Nl)^N}\left(L - Nl + \tfrac{1}{2}l - \xi_2 - x_1\right)^{N-2}.$$

The domain of definition of this function is found by replacing x_2 in the previous domain formula with the function $x_2(x_1, \xi_2)$. That gives

$$\left(\tfrac{1}{2}l < x_1 < L - Nl + \tfrac{1}{2}l; 0 < \xi_2 < L - Nl + \tfrac{1}{2}l - x_1\right)$$

$$= \left(0 < \xi_2 < L - Nl; \tfrac{1}{2}l < x_1 < L - Nl + \tfrac{1}{2}l - \xi_2\right),$$

where the last step follows by interchanging the bounding order of the variables x_1 and ξ_2. Now we can compute the unconditioned density function of ξ_2 by integrating out the variable x_1,

$$\hat{P}_{2|0}(\xi_2|x_0) = \int_{(1/2)l}^{L-Nl+(1/2)l-\xi_2} \hat{P}_{1,2|0}(x_1,\xi_2|x_0)dx_1$$

$$= \frac{N(N-1)}{(L-Nl)^N} \int_{(1/2)l}^{L-Nl+(1/2)l-\xi_2} \left(L - Nl + \tfrac{1}{2}l\right.$$

$$\left. - \xi_2 - x_1\right)^{N-2}dx_1$$

$$= \frac{N(N-1)}{(L-Nl)^N} \int_0^{L-Nl-\xi_2} u^{N-2}du,$$

(B4)

$$\hat{P}_{2|0}(\xi_2|x_0) = \frac{N}{(L-Nl)^N}(L-Nl-\xi_2)^{N-1}$$

$$(0 < \xi_2 < L-Nl).$$

Comparing results (B3) and (B4) for ξ_1 and ξ_2, we conclude by symmetry that the unconditioned density function of any gap length between adjacent rods is

$$P_{\text{gap}}(\xi) = \frac{N}{(L-Nl)^N}(L-Nl-\xi)^{N-1} \quad (0 < \xi < L-Nl).$$

(B5)

APPENDIX C: MONTE CARLO PLACEMENT OF THE RODS

Equations (A1) and (A2) allow us to construct a Monte Carlo algorithm for dealing out the rods from left to right in the interval $0 \leq x \leq L$, successively specifying values for $x_1, x_2, \ldots, x_N$, in such a way that the resulting configuration will be an unbiased sample of the uniform nonoverlapping distribution.[10] The first step in constructing this generating algorithm is to compute the cumulative distribution function $F_{n|n-1}(x_n|x_{n-1})$ corresponding to the probability density function [Eq. (A2)],

$$F_{n|n-1}(x_n|x_{n-1}) \equiv \int_{x_{n-1}+l}^{x_n} P_{n|n-1}(x_n'|x_{n-1})dx_n'$$

$$= \frac{N-n+1}{(L_{n-1}-x_{n-1})^{N-n+1}} \int_{x_{n-1}+l}^{x_n} (L_n - x_n')^{N-n}dx_n'.$$

(C1)

The integration is easily performed using the variable transformation $u = L_n - x_n'$, and recalling from Eq. (A4b) that $L_n - l = L_{n-1}$. The result is

$$F_{n|n-1}(x_n|x_{n-1}) = 1 - \left(\frac{L_n - x_n}{L_{n-1}-x_{n-1}}\right)^{N-n+1}.$$

(C2)

Now we apply the standard Monte Carlo inversion procedure to generate a set of values for $x_1, x_2, \ldots, x_N$ according

to the joint probability density function [Eq. (A2)]: We pick N random numbers $r_1, r_2, \ldots, r_N$ from the unit-interval uniform distribution, and then solve the equation

$$F_{n|n-1}(x_n|x_{n-1}) = r_n,$$

(C3)

for x_n, successively for $n = 1, 2, \ldots, N$. Substituting Eq. (C2) into Eq. (C3), and for convenience replacing r_n by $1 - r_n$ (which is allowed since the latter will also be uniformly distributed in the unit interval), we get

$$\left(\frac{L_n - x_n}{L_{n-1}-x_{n-1}}\right)^{N-n+1} = r_n.$$

Solving this for x_n gives us the generating formula[10]

$$x_n = L_n - (L_{n-1}-x_{n-1})r_n^{1/(N-n+1)} \quad (n = 1, 2, \ldots, N),$$

(C4)

which is to be applied successively for increasing n. As before, $x_0 = -\tfrac{1}{2}l$, and the set of constants $\{L_0, L_1, \ldots, L_N\}$ is as given in Eqs. (A4).

It is shown in Ref. 10 that the alternative generating procedure of independently placing each rod center uniformly at random inside the interval $[\tfrac{1}{2}l, L-\tfrac{1}{2}l]$ and then accepting the final set if and only if no overlap occurs has an associated acceptance probability of $[(L-Nl)/(L-a)]^N$. For $N = 100$ rods of length $l = 1$ on an interval of length $L = 200$, this probability of acceptance is only 1.3×10^{-30}. By contrast, procedure (C4) gives an acceptable configuration every time. The reason is that this procedure knows that the nth rod laid down is to be the nth rod from the left, so it biases the selection appropriately to the left, taking account of the already generated location of the $(n-1)$th rod.

[1] L. W. Anacker and R. Kopelman, J. Chem. Phys. **81**, 6402 (1984); R. Kopelman, J. Stat. Phys. **42**, 185 (1986); K. Lindenberg, P. Argyrakis, and R. Kopelman, J. Phys. Chem. **99**, 7542 (1995).

[2] E. Ben-Naim, P. Krapivsky, F. Leyvraz, and S. Redner, J. Phys. Chem. **98**, 7284 (1994).

[3] J. M. Sancho, A. H. Romero, A. M. Lacasta, and K. Lindenberg, J. Phys.: Condens. Matter (to be published).

[4] P. Argyrakis and R. Kopelman, J. Phys. Chem. **93**, 225 (1989).

[5] M. A. Savageau, J. Theor. Biol. **176**, 115–124 (1995); R. Grima and S. Schnell, ChemPhysChem **7**, 1424 (2006); S. Schnell and T. E. Turner, Prog. Biophys. Mol. Biol. **85**, 235 (2004).

[6] E. Trizac, Phys. Rev. Lett. **88**, 160601 (2002); P. L. Krapivsky and C. Sire, *ibid.* **86**, 2494 (2001); A. Lipowski, D. Lipowska, and A. L. Ferreira, Phys. Rev. E **73**, 032102 (2006).

[7] D. T. Gillespie, J. Stat. Phys. **16**, 311 (1977).

[8] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976); J. Phys. Chem. **81**, 2340 (1977); Physica A **188**, 404 (1992).

[9] Suppose adt is the probability of a collision in *any* infinitesimal time interval dt. The probability of no collision in any finite interval $[0,t)$ can be computed by subdividing $[0,t)$ into n subintervals of size t/n, and then computing the probability of no collision in all of those subintervals. If n is sufficiently large, the latter probability is $(1-a(t/n))^n = e^{-at}$, and this signifies an "exponentially distributed" collision time.

[10] D. T. Gillespie, Naval Weapons Center Technical Publication (NWC TP) No. 5714, 1975 (unpublished).