

# The small-voxel tracking algorithm for simulating chemical reactions among diffusing molecules

Daniel T. Gillespie,<sup>1,a)</sup> Effrosyni Seitaridou,<sup>2,b)</sup> and Carol A. Gillespie<sup>1,c)</sup>

<sup>1</sup>Dan T Gillespie Consulting, 30504 Cordoba Pl., Castaic, California 91384, USA

<sup>2</sup>Department of Physics, Oxford College of Emory University, 100 Hamill St., Oxford, Georgia 30054, USA

(Received 6 September 2014; accepted 1 December 2014; published online 19 December 2014)

Simulating the evolution of a chemically reacting system using the bimolecular propensity function, as is done by the stochastic simulation algorithm and its reaction-diffusion extension, entails making statistically inspired guesses as to where the reactant molecules are at any given time. Those guesses will be physically justified if the system is dilute and well-mixed in the reactant molecules. Otherwise, an accurate simulation will require the extra effort and expense of keeping track of the positions of the reactant molecules as the system evolves. One molecule-tracking algorithm that pays careful attention to the physics of molecular diffusion is the enhanced Green's function reaction dynamics (eGFRD) of Takahashi, Tănase-Nicola, and ten Wolde [Proc. Natl. Acad. Sci. U.S.A. **107**, 2473 (2010)]. We introduce here a molecule-tracking algorithm that has the same theoretical underpinnings and strategic aims as eGFRD, but a different implementation procedure. Called the small-voxel tracking algorithm (SVTA), it combines the well known voxel-hopping method for simulating molecular diffusion with a novel procedure for rectifying the unphysical predictions of the diffusion equation on the small spatiotemporal scale of molecular collisions. Indications are that the SVTA might be more computationally efficient than eGFRD for the problematic class of non-dilute systems. A widely applicable, user-friendly software implementation of the SVTA has yet to be developed, but we exhibit some simple examples which show that the algorithm is computationally feasible and gives plausible results. © 2014 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4903962>]

## I. INTRODUCTION

Chemically reacting systems at the cellular level, where the molecular population of one or more reactant species can be relatively low, are often analyzed using the chemical master equation (CME) or the logically equivalent stochastic simulation algorithm (SSA). But the CME and SSA require some rather restrictive conditions to be satisfied. The best known of those is the requirement that the reactant molecules be *well-mixed* throughout the system volume  $\Omega$ . However, it has recently been argued<sup>1</sup> that a more critical requirement is that the reactant molecules be *dilute* in  $\Omega$ . Here, dilute means that the average separation between the reactant molecules is very large compared to their diameters, or equivalently, that the reactant molecules occlude only a negligibly small fraction of  $\Omega$ .<sup>2</sup> Two observations support the view that diluteness is a necessary (but not sufficient) condition for using the CME and the SSA: First, the derivation from molecular physics of the bimolecular propensity function, which plays a key role in both the CME and the SSA, requires that the reactant molecules be not only well-mixed but also dilute in  $\Omega$ ,<sup>1</sup> e.g., analyzing a bimolecular reaction as though it involves only two molecules, without the complications of a third molecule, will be justified in the CME and SSA only if the system is so dilute that three co-located molecules will be a rarity. Second, only if the reactant molecules are dilute in  $\Omega$  can ordi-

nary diffusion keep them well-mixed.<sup>3</sup> The underlying reason for that is easy to understand: only if the reactant molecules are separated by distances comfortably larger than their diameters will they be able to wander around without colliding with each other for a long enough time to become well-mixed between collisions. It follows that diluteness is also a prerequisite for using the ordinary differential equations of traditional chemical kinetics, since those equations arise as an approximation of the CME/SSA when the molecular populations of the species and the system volume are all very large.

Diluteness is also a necessary (but not sufficient) condition for the reaction-diffusion CME/SSA. That formalism seeks to circumvent the globally well-mixed requirement of the CME and the SSA by subdividing the system into imaginary subvolumes or *voxels* which individually are well-mixed—a condition that generally requires the voxels to be “small.” Chemical reactions are assumed to occur inside individual voxels, and reactant molecules are moved from voxel to voxel by special “hopping reactions” that replicate the diffusion equation when the voxels are, again, “small.” But we cannot legitimately use a bimolecular propensity function *inside a voxel* unless the reactant molecules are *dilute in that voxel*. That requires, at a minimum, that the voxel volume be much larger than the volume occluded by two reactant molecules, and that requirement might not always be compatible with the requirement that the voxels be “small.”

Overcoming the diluteness limitation of the SSA and its voxel extension would seem to require a simulation

<sup>a)</sup>Electronic mail: gillespiedt@mailaps.org

<sup>b)</sup>Electronic mail: eseitar@emory.edu

<sup>c)</sup>Electronic mail: gillespiec@acm.org

strategy that tracks the positions of the reactant molecules, thereby obviating unwarranted assumptions about where those molecules are and are not. The gold standard of such strategies is molecular dynamics. It tracks both the positions and the velocities of not only all the reactant (solute) molecules, but also all the more numerous solvent molecules. That approach, however, is rarely feasible. A much less ambitious, somewhat less accurate, but still computationally challenging approach is to track only the *positions* of only the *solute* molecules. An algorithm of that kind which pays careful attention to the physics of diffusion is the enhanced Green’s function reaction dynamics (eGFRD) of Takahashi, Tănase-Nicola, and ten Wolde.<sup>4</sup> An important feature of eGFRD is its use of imaginary absorbing surfaces to segregate or corral the reactant molecules into small groups, a strategy originally devised by Opplestrup *et al.*<sup>5</sup> The time evolutions of the molecules inside individual corrals are then simulated independently and in parallel until the absorption of some molecule at a corral boundary signals that a corral has been breached. When such a breach occurs, the simulations are temporarily halted to allow any necessary adjustments to be made to the corrals, and any necessary synchronizations to be made in the corral times.

Of the several tasks that comprise the eGFRD simulation strategy, we will be concerned here *only* with the task of accurately simulating the time evolution of the reactant molecules inside a single corral until either a chemical reaction or a breach of the corral occurs. Such “halting events” are simulated in the current formulation of eGFRD by using analytical Green’s function techniques to compute the times of those events and the locations of the reactant molecules at those times.<sup>4</sup> Here we will present another way of performing those tasks, which we have dubbed the small-voxel tracking algorithm (SVTA), which might offer some computational advantages in certain situations. The implementation procedure of the SVTA is quite different from that of eGFRD, but as we will show, the SVTA and eGFRD share the same theoretical underpinnings. We will *not* address in this paper the problem of determining the best way to lay out the corrals, nor the problem of how to most efficiently coordinate and synchronize their local times when a halting event occurs.

After reviewing in Sec. II some results in the theory of diffusion and diffusion-induced chemical reactions that we will need later, we introduce the SVTA in Sec. III in the context of carrying out the core calculations for “single particle events” as defined in eGFRD.<sup>4</sup> We illustrate that with a simple numerical example in Sec. IV. In Sec. V we develop the SVTA’s procedure for carrying out the core calculations in eGFRD’s “two particle events.” After showing in Sec. VI that the SVTA and eGFRD share the same theoretical foundation, we give in Sec. VII some simple numerical examples of two-molecule events which show that the SVTA is computationally feasible and gives plausible results. In Sec. VIII we show how the SVTA can deal with three-molecule events, and in Sec. IX we show how it can take account of forces between reactant molecules when such forces are prescribed. We conclude in Sec. X with a summary of our work and a forward-looking assessment of the SVTA.

## II. BACKGROUND THEORY

In this section, we will review some results in the theory of diffusion and diffusion-induced chemical reactions which we will need later to derive the SVTA, and also to make clear its relation to other simulation strategies such as the SSA, the reaction-diffusion SSA, and eGFRD. Derivations of most of these results can be found in Ref. 6.

### A. Modeling assumptions

We begin by stating the physical assumptions that underlie our analysis. All reactant molecules are assumed to be *solute* molecules that are in solution with very many, much smaller, chemically inert *solvent* molecules. All molecules are confined to a region  $\Omega$  whose boundary  $\partial\Omega$  is perfectly reflecting, and kept at absolute temperature  $T$ . Reactant molecule  $i$  is assumed to be a hard sphere with diameter  $\sigma_i$ , mass  $m_i$ , diffusion coefficient  $D_i$ , and *unimolecular* reaction constant  $k_i$ ; the last is defined so that  $k_i dt$  is the probability that molecule  $i$  will so react in the next infinitesimal time  $dt$ . And finally, two molecules  $i$  and  $j$  will react bimolecularly *only if they collide*, and then only with probability  $q_{ij}$ , where  $0 \leq q_{ij} \leq 1$ . The collision-conditioned reaction probability  $q_{ij}$  can, at least in principle, be calculated from a detailed knowledge of the physical properties of the two molecules; e.g., if the reaction occurs only if the two molecules have a collisional kinetic energy above some threshold value  $E_{\text{thr}}$ , then  $q_{ij}$  will be the Arrhenius factor  $\exp(-E_{\text{thr}}/k_B T)$ , where  $k_B$  is Boltzmann’s constant.

### B. Molecular diffusion and its simulation

On sufficiently large space and time scales, the position  $\mathbf{R}(t)$  of the center of a diffusing reactant molecule at time  $t$  can be regarded as a random variable whose probability density function (PDF)  $p(\mathbf{r}, t | \mathbf{r}_0, t_0)$ , conditioned on  $\mathbf{R}(t_0) = \mathbf{r}_0$  for  $t_0 \leq t$ , obeys the Einstein diffusion equation,

$$\frac{\partial p(\mathbf{r}, t | \mathbf{r}_0, t_0)}{\partial t} = D \nabla_{\mathbf{r}}^2 p(\mathbf{r}, t | \mathbf{r}_0, t_0). \quad (1)$$

The Einstein diffusion equation differs from the traditional diffusion equation in that the latter’s solution is the average density of many non-interacting solute molecules, instead of the PDF of a single solute molecule. The restriction of Eq. (1) to “sufficiently large space and time scales” can be spelled out quantitatively as follows:<sup>6</sup> Equation (1) accurately describes the physical motion of a diffusing solute molecule only if the time scale on which that motion is being observed is large enough that the time

$$\tau \equiv \frac{mD}{k_B T} \quad (2)$$

is infinitesimally small, and only if the length scale on which that motion is being observed is large enough that the length

$$\lambda \equiv D \sqrt{\frac{m}{k_B T}} \quad (3)$$

is infinitesimally small. If instead we observe the solute molecule on scales where  $\tau$  or  $\lambda$  is very large, then the

motion of the solute molecule will appear to be *ballistic*, as in an ideal gas. That is because during such small times and over such small distances, the solute molecule will not experience enough collisions with the much smaller and lighter solvent molecules to noticeably alter its velocity. Constant velocity motion is ballistic motion, and ballistic motion cannot be described by Eq. (1). Indeed, on very small space-time scales the motion described by Eqs. (1) is not merely inaccurate; it is bizarrely unphysical.<sup>6</sup>

A mathematically exact way of simulating the unrestricted motion of a molecule that moves according to Eq. (1) is provided by the Einstein stepping formulas, which allow us to compute the location of the center of the molecule at time  $t + \Delta t$  in terms of its location at earlier time  $t$ :

$$\left. \begin{aligned} x(t + \Delta t) &= x(t) + \sqrt{2D \cdot \Delta t} \cdot n_x \\ y(t + \Delta t) &= y(t) + \sqrt{2D \cdot \Delta t} \cdot n_y \\ z(t + \Delta t) &= z(t) + \sqrt{2D \cdot \Delta t} \cdot n_z \end{aligned} \right\}. \quad (4)$$

Here  $n_x$ ,  $n_y$ , and  $n_z$  are independent samples of the normal random variable with mean 0 and variance 1. Equation (4) is exact with respect to Eq. (1) for all  $\Delta t \geq 0$ . But since Eq. (1) itself is *physically* accurate only on time scales much larger than  $\tau$ , formulas (4) will be physically accurate only if  $\Delta t \gg \tau$ . A generalization of formulas (4) in which  $\Delta t$  is not subject to the restriction  $\Delta t \gg \tau$  exists,<sup>7</sup> but we will not be needing it in this paper. But even in the great majority of situations in which Eq. (4) is accurate, they have a serious drawback from a simulation standpoint: they cannot tell us where the molecule went *between* times  $t$  and  $t + \Delta t$ .

A way of simulating the movement prescribed by the diffusion equation which does not have this drawback, and which is asymptotically exact, is the voxel-hopping algorithm used in the reaction-diffusion CME and SSA. With space subdivided into cubic voxels of edge length  $h$ , so that each interior voxel shares a voxel face with its six adjacent voxels, this algorithm requires the center of the diffusing molecule to hop according to the following rule:

$$(D/h^2) \cdot dt = \text{probability that the center of the molecule will hop from its present voxel to any particular accessible adjacent voxel in the next infinitesimal time } dt. \quad (5)$$

Voxels that are *not* accessible to the center of the molecule are those whose occupancy would put some part of the molecule either outside the system volume or inside another reactant molecule.

Rule (5) is “asymptotically exact” with respect to the diffusion equation (1) in the following sense: The master equation which describes the stochastic time evolution of a random walker that hops from voxel to voxel according to rule (5) becomes Eq. (1) in the limit  $h \rightarrow 0$ .<sup>8</sup> We know of no other rule for jumping on a lattice that models the diffusion equation with such clearly established mathematical fidelity, provided the lattice spacing  $h$  is “small.” However, since the *physical* validity of rule (5) is contingent on Eq. (1) being physically accurate on the scale of an individual voxel, then the limitation of Eq. (1) to lengths much larger than  $\lambda$  in Eq. (3) implies that rule (5) will be *physically* accurate only if  $h$  is large

enough that

$$h \gg D \sqrt{\frac{m}{k_B T}}. \quad (6)$$

Detailed studies<sup>9</sup> have shown that the voxel-hopping rule (5) most accurately describes the first-passage of the molecule’s center *from* a point that is *randomly uniform* inside its current voxel *to* a point that is *randomly uniform* inside its destination voxel. This result is corroborated by the observation that if the center of the molecule were more likely closer to one voxel face than another, then the hopping rule (5) would be incorrect in treating the six adjacent voxels as equally likely destinations for the next jump. We conclude that the voxel-hopping strategy (5) carries the additional implication that *between hops, the center of the molecule will be randomly and uniformly distributed inside its current voxel*. This result will be critical to our derivation of the SVTA.

### C. Diffusion-induced chemical reactions in a dilute, well-mixed system

If the system is *dilute and well-mixed*—a circumstance that is *not* assumed by the SVTA—then two molecules in  $\Omega$  chosen at random, call them molecules 1 and 2, will undergo a bimolecular reaction in the next infinitesimally small time  $dt$  with probability  $k|\Omega|^{-1} \cdot dt$ , where  $|\Omega|$  is the volume of  $\Omega$  and<sup>10</sup>

$$k = \frac{4\pi\sigma_{12}^2 D_{12} \bar{v}_{12} q_{12}}{4D_{12} + \sigma_{12} \bar{v}_{12} q_{12}}. \quad (7)$$

In formula (7),

$$\sigma_{12} \equiv \frac{1}{2}(\sigma_1 + \sigma_2), \quad (8)$$

$$D_{12} \equiv D_1 + D_2, \quad (9)$$

and

$$\bar{v}_{12} = \sqrt{8k_B T / (\pi m_{12})} \quad (10)$$

is the average speed of either molecule relative to the other if they were in thermal equilibrium in a dilute gas, with  $m_{12} \equiv m_1 m_2 / (m_1 + m_2)$  being the so-called reduced mass of the two molecules. The derivation of Eq. (7) is summarized in Appendix A; we will be needing some intermediate results of that derivation later in this paper.

If molecules 1 and 2 are of different species, and if there are  $x_1$  molecules of species 1 and  $x_2$  molecules of species 2, all dilute and well-mixed inside  $\Omega$ , then  $k|\Omega|^{-1} x_1 x_2$  is called the *propensity function* of the reaction in question. Its product with  $dt$  gives the probability that a bimolecular reaction between some species 1 molecule and some species 2 molecule will occur somewhere inside  $\Omega$  in the next  $dt$ . If molecules 1 and 2 are of the same species, then the propensity function will instead be  $k|\Omega|^{-1} \frac{1}{2} x_1 (x_1 - 1)$ . The existence of a propensity function for a bimolecular reaction—i.e., the existence of some function whose product with  $dt$  gives the probability that the reaction will occur in the next infinitesimal time  $dt$ —is essential for the system to be described by a CME and simulated by the SSA.

Two important limiting cases of formula (7) are

$$4D_{12} \gg \sigma_{12}\bar{v}_{12}q_{12} \Rightarrow k \approx \pi\sigma_{12}^2\bar{v}_{12}q_{12} \equiv k_{\text{bal}} \quad (11a)$$

and

$$4D_{12} \ll \sigma_{12}\bar{v}_{12}q_{12} \Rightarrow k \approx 4\pi\sigma_{12}D_{12} \equiv k_{\text{dif}}. \quad (11b)$$

The regime defined by the condition in (11a) is called the *ballistic* regime, because the formula for the reaction probability  $k_{\text{bal}}|\Omega|^{-1}dt$  implied by Eq. (11a) is exactly what one obtains using simple kinetic theory arguments when the two molecules are moving ballistically, as in an ideal gas.<sup>11,1</sup> The regime defined by the condition in (11b) is called the *diffusion-limited* regime; it is the regime that characterizes most cellular reactions.

A better understanding of what diffusion-limited means can be gained by noting that the formulas for  $k_{\text{bal}}$  and  $k_{\text{dif}}$  in Eqs. (11a) and (11b) allow us to rewrite Eq. (7) as

$$\frac{1}{k} = \frac{1}{k_{\text{dif}}} + \frac{1}{k_{\text{bal}}}. \quad (12)$$

Since  $k$  is a measure of how quickly the reaction occurs, then  $1/k$  is a measure of how slowly the reaction occurs. In an electrical circuit analogy,  $k$  is the “conductance” of the reaction, while  $1/k$  is the “resistance” of the reaction. Equation (12) evidently expresses the resistance of the reaction as a sum of a “diffusional resistance” and a “ballistic resistance.” That way of combining these resistances implies that they are “connected in series,” and that mirrors rather well the actual physical situation: In order for the reaction to occur, molecules 1 and 2 must first overcome the diffusional resistance associated with their approach to each other via classical diffusion over a relatively large distance. Then they must overcome the ballistic resistance associated with their approach to each other immediately before they collide. If  $k_{\text{dif}} \ll k_{\text{bal}}$ , so that the diffusional resistance  $1/k_{\text{dif}}$  is much greater than the ballistic resistance  $1/k_{\text{bal}}$ , then it is diffusion that is the bottleneck to the reaction: the ballistic resistance term in Eq. (12) will then be negligibly small, and we will have  $k \approx k_{\text{dif}}$ .

There are several surprises in Eqs. (11a) and (11b). First, in spite of the fact that the exact equation (7) shows that  $k \rightarrow 0$  as  $q_{12} \rightarrow 0$ , the diffusion-limited regime’s approximation for  $k$  in Eq. (11b) has  $k$  independent of  $q_{12}$ . What that means is that if  $q_{12}$ , whose value is confined to the interval  $0 \leq q_{12} \leq 1$ , is non-zero enough that the condition  $4D_{12} \ll \sigma_{12}\bar{v}_{12}q_{12}$  is satisfied, then  $k$  will be insensitive to the exact value of  $q_{12}$ . But notice that even if  $q_{12}$  were equal to 1, so that every collision produced a reaction, it would *not* necessarily follow that the reaction will be diffusion-limited; in order for that to be so it would also have to be true that  $4D_{12} \ll \sigma_{12}\bar{v}_{12}$ . At the other extreme, we can always imagine  $q_{12}$  to be small enough (yet not zero) that condition (11a) will be satisfied; in that case the reaction will be in the ballistic regime and we would have  $k \approx k_{\text{bal}}$ , even though the system is obviously a solution and not a gas.

## D. Connection with some other approaches

If the numerator and denominator on the right of Eq. (7) are multiplied by  $\pi\sigma_{12}$ , that formula becomes, after invoking the definition in (11a),

$$k \equiv \frac{4\pi\sigma_{12}^2D_{12}\bar{v}_{12}q_{12}}{4D_{12} + \sigma_{12}\bar{v}_{12}q_{12}} = \frac{4\pi\sigma_{12}D_{12}k_{\text{bal}}}{4\pi\sigma_{12}D_{12} + k_{\text{bal}}}. \quad (13)$$

The form on the right appears in earlier literature more frequently than the form on the left, although with  $k_{\text{bal}}$  usually denoted differently—for example, as  $k_a$  by Takahashi *et al.*,<sup>4</sup> or as  $k_r$  by Hellander *et al.*<sup>12</sup>—and referred to as either the “intrinsic” rate constant or the “microscopic” rate constant. In either guise,  $k_{\text{bal}}$  in Eq. (13) is usually regarded as an independent parameter, with no reference made to the formula for it in Eq. (11a) in terms of  $\sigma_{12}$ ,  $\bar{v}_{12}$ , and  $q_{12}$ . Indeed, the parameter  $q_{12}$  rarely appears in existing works on chemical reaction simulation, although it is a well established parameter in chemical kinetics more generally. If  $k_{\text{bal}}$  is considered to be the given quantity instead of  $q_{12}$ , then  $q_{12}$  can easily be deduced from Eq. (11a) as

$$q_{12} = \frac{k_{\text{bal}}}{\pi\sigma_{12}^2\bar{v}_{12}}. \quad (14)$$

When this expression for  $q_{12}$  is substituted into Eq. (11b), the condition for being diffusion-limited becomes  $4\pi\sigma_{12}D_{12} \ll k_{\text{bal}}$ .

## III. SVTA’S ONE-MOLECULE CORRAL

eGFRD<sup>4</sup> considers two different corraling scenarios: *one* molecule in a corral, and *two* molecules in a corral. We will first describe how the SVTA deals with the simpler one-molecule corral. That situation is illustrated (in two dimensions) in Fig. 1: molecule 1 at initial time  $t_0$  is cordoned off from the other molecules in  $\Omega$  by an imaginary absorbing surface  $\Sigma$ , either all by itself as in Fig. 1(a), or with the help of part of the system boundary  $\partial\Omega$  as in Fig. 1(b). The surfaces  $\Sigma$  and  $\partial\Omega$  corral the entire molecule, but we will be locating the molecule by its center. So we introduce the two surfaces  $\Sigma_1$  and  $\partial\Omega_1$  (dashed lines in Fig. 1) which are set in from the respective surfaces  $\Sigma$  and  $\partial\Omega$  by the radius  $\sigma_1/2$  of the molecule;  $\Sigma_1$  and  $\partial\Omega_1$  corral the *center* of molecule 1. If the center of molecule 1 comes into contact with  $\partial\Omega_1$ , the molecule will be reflected. If the center of molecule 1 comes into contact with  $\Sigma_1$ , we will halt the simulation because of a corral breach. This one-molecule scenario is a little more general than the single particle event scenario in eGFRD;<sup>4</sup> there the corraling surface is assumed to be a sphere that is centered on the initial location of molecule 1 and lies completely inside  $\Omega$ .

The SVTA will simulate the diffusional motion of molecule 1 using the voxel-hopping rule (5). That requires us to partition the space accessible to the center of molecule 1 into voxels in the form of *cubes* of edge length  $h$ . This “voxelation” of space transforms the situation illustrated in Fig. 1(b) to that illustrated in Fig. 2. The voxel containing the center of molecule 2 is designated  $VC_1$ . In Fig. 2 we have colored that voxel a uniform gray to emphasize the fact,



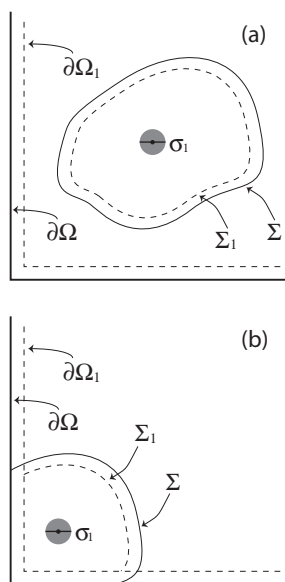


FIG. 1. A single solute molecule of diameter  $\sigma_1$ , one of the possibly many solute molecules inside a containing region  $\Omega$ , is at time  $t_0$  separated from the other molecules by an imaginary surface  $\Sigma$ . That surface “corrals” molecule 1, either entirely as in (a), or with the help of the system boundary  $\partial\Omega$  as in (b). In the 2-dimensional representation of these figures, the surfaces  $\Sigma$  and  $\partial\Omega$  are shown as solid lines. The surfaces  $\Sigma_1$  and  $\partial\Omega_1$  (dashed lines) are formed by tracing the position of the molecule’s center when the molecule is rolled around on the insides of  $\Sigma$  and  $\partial\Omega$ , respectively. So if the center of molecule 1 contacts any point on  $\Sigma_1$  or  $\partial\Omega_1$ , then molecule 1 will be in contact with the boundary of its corral.

discussed at the end of Sec. II B, that between hops the center of molecule 1 is distributed randomly and uniformly inside  $VC_1$ . When molecule 1 hops to an adjacent voxel, it is  $VC_1$  that does the hopping. Also, we replace the dashed line surfaces  $\Sigma_1$  and  $\partial\Omega_1$  in Fig. 1(b) with the sets of voxels  $V\Sigma_1$  and  $V\partial\Omega_1$  which those respective surfaces intersect. We will *not* allow  $VC_1$  to hop into any voxel belonging to  $V\partial\Omega_1$  (the red voxel set in Fig. 2) because that would place some part of molecule 1 outside of  $\Omega$ . We *will* allow  $VC_1$  to hop into any voxel belonging to  $V\Sigma_1$  (the yellow voxel set in Fig. 2), but that will constitute a corral breach which halts the simulation.

At any instant  $t$  during the simulation, two potential “events” compete with each other to occur first: a *unimolecular reaction*, and a *hop*. The definition of the unimolecular reaction rate constant  $k_1$  (see Sec. II A) can be shown to imply that the time until that reaction fires is the exponential random variable  $\mathcal{E}(k_1)$  with decay constant  $k_1$  (and mean  $k_1^{-1}$ ). Therefore, if a corral breach does not occur earlier, molecule 1 will undergo its unimolecular reaction at time

$$t_{\text{uni}} = t_0 + \tau_{\text{uni}}, \quad (15a)$$

where

$$\tau_{\text{uni}} = \mathcal{E}(k_1). \quad (15b)$$

The right side of Eq. (15b) may be read “a sample value of  $\mathcal{E}(k_1)$ .”

Competing with the unimolecular reaction event is a hop of  $VC_1$  to one of its  $n_1^{\text{open}}$  accessible adjacent voxels. In the one-molecule scenario, the only voxels that are *not* accessible to  $VC_1$  are voxels belonging to  $V\partial\Omega_1$ . For example, if  $\Omega$  were

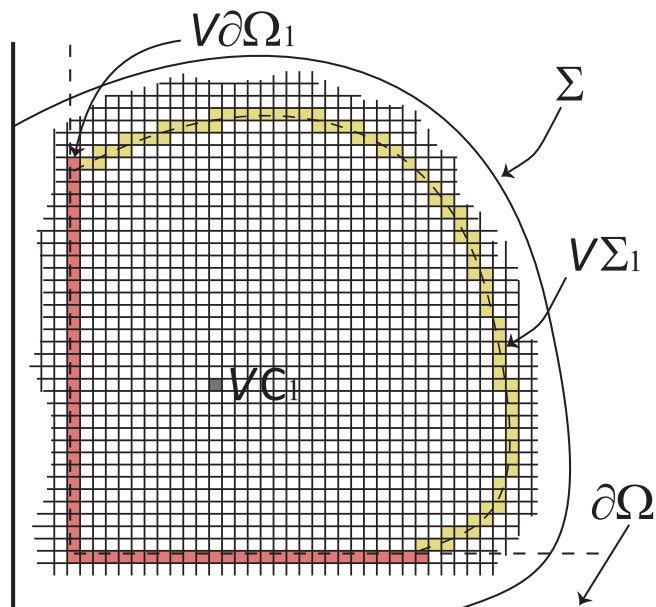


FIG. 2. The configuration in Fig. 1(b) as represented by the SVTA. The corraling region for the molecule’s center, namely the region enclosed by  $\Sigma_1$  and portions of  $\partial\Omega_1$ , has been subdivided into cubic voxels of edge length  $h < \sigma_1$ . In this voxelated space, the center of molecule 1 is assumed to be distributed randomly and uniformly inside the particular voxel  $VC_1$  that contains that point. The corraling surfaces  $\Sigma_1$  and  $\partial\Omega_1$  have been replaced by the respective sets of voxels that those surfaces intersect; those “voxelated surfaces”  $V\Sigma_1$  and  $V\partial\Omega_1$  are shown in yellow and red, respectively. If voxel  $VC_1$  were to coincide with any voxel belonging to  $V\Sigma_1$  or  $V\partial\Omega_1$ , then some part of molecule 1 would lie outside the corraling region formed by  $\Sigma$  and  $\Omega$ . So  $VC_1$  will *not* be allowed to jump to a voxel belonging to  $V\partial\Omega_1$ .  $VC_1$  will be allowed to jump to a voxel belonging to  $V\Sigma_1$ , but that will immediately suspend the simulation.

a *cube*, then  $n_1^{\text{open}}$  could be as large as 6 if  $VC_1$  were well inside  $\Omega$ , or as small as 3 if  $VC_1$  were in a corner of  $\Omega$ . In any case, by applying the addition law of probability to (5) we may infer that the probability that  $VC_1$  will hop in the next  $dt$  to any one of its  $n_1^{\text{open}}$  accessible adjacent voxels is  $n_1^{\text{open}} \cdot (D_1/h^2)dt$ . That implies that the time to the next hop will be an exponential random variable with decay constant  $n_1^{\text{open}} D_1/h^2$ . So at the current time  $t$ , unless molecule 1 reacts unimolecularly earlier,  $VC_1$  will hop to an adjacent voxel at time

$$t_{\text{hop}} = t + \tau_{\text{hop}}, \quad (16a)$$

where

$$\tau_{\text{hop}} = \mathcal{E}(n_1^{\text{open}} D_1/h^2). \quad (16b)$$

The particular voxel to which  $VC_1$  hops at this time is obtained by selecting one of the  $n_1^{\text{open}}$  accessible adjacent voxels “at random.” If those voxels are indexed in any order from 1 to  $n_1^{\text{open}}$ , then the index  $k_{\text{dest}}$  of the destination voxel can be computed as

$$k_{\text{dest}} = [1 + n_1^{\text{open}} \cdot \mathcal{U}(0, 1)], \quad (17)$$

where  $[x]$  denotes “the greatest integer in  $x$ ” and  $\mathcal{U}(0, 1)$  denotes a sample value of the uniform random variable in the unit interval.

The SVTA imposes some stringent bounds on the voxel edge length  $h$ . Since a minimal requirement for a molecule

tracking algorithm is that the resolution of the tracking should be smaller than the molecule, the SVTA imposes the upper bound  $h < \sigma_1$ . Thus, a voxel cannot contain the center of more than one molecule, and one molecule generally occludes many voxels. The SVTA also imposes a lower bound on  $h$ . For reasons explained in connection with Eq. (3), the diffusion equation (1), and hence also the voxel-hopping rule (5), will be *physically* accurate on the scale of a voxel *only* if  $h \gg D_1 \sqrt{m_1 / (k_B T)}$ . Thus, in the SVTA the voxel edge length  $h$  must satisfy

$$D_1 \sqrt{\frac{m_1}{k_B T}} \ll h < \sigma_1. \quad (18)$$

To get some idea of what condition (18) implies for a typical cellular molecule, let us put in some order-of-magnitude values for its parameters. If our spherical molecule 1 has the same mass ( $m_1 = 1.66 \times 10^{-19}$  g) and density ( $1.37$  g/cm<sup>3</sup>) as a typical protein molecule,<sup>13</sup> then simple geometry gives for its diameter  $\sigma_1 = 6.1 \times 10^{-7}$  cm. If this molecule is diffusing in water (viscosity  $\eta = 0.001$  kgm<sup>-1</sup> s<sup>-1</sup>) at room temperature ( $T = 293$  K), then the Stokes formula  $\gamma_1 = 3\pi\eta\sigma_1$  for the molecule's drag coefficient  $\gamma_1$ , together with the Einstein relation  $D_1\gamma_1 = k_B T$ , imply a diffusion coefficient of  $D_1 = 0.7 \times 10^{-10}$  m<sup>2</sup>/s. That value for  $D_1$  is probably an overestimate;<sup>14</sup> nevertheless, with all these values, condition (18) becomes

$$0.014 \times 10^{-7} \text{ cm} \ll h < 6.1 \times 10^{-7} \text{ cm}. \quad (19)$$

Since the right side of (19) is over 400 times larger than the left side, and since also the figure on the left side is probably an overestimate, we conclude that condition (18) on  $h$  should be feasible for cellular systems.

The SVTA's strategy for simulating a single-molecule corral is quite straightforward: It hops  $VC_1$  over the voxels repeatedly until *either*  $t$  exceeds  $t_{\text{uni}}$ , *or*  $VC_1$  arrives in a voxel belonging to  $V\Sigma_1$ . The following steps describe in more detail how this strategy is actually carried out.

1. Divide the space accessible to the center of molecule 1 into cubic voxels of edge length  $h$  which satisfies condition (18). Label each voxel with three (signed) integers  $i_x, i_y, i_z$ , so that the center of voxel  $(i_x, i_y, i_z)$  is at point  $(x, y, z) = (i_x h, i_y h, i_z h)$ .
2. Identify the voxels that belong to the voxelated corraling surface  $V\Sigma_1$ , and also the voxels that belong to the voxelated boundary  $V\partial\Omega_1$ .
3. Set the current time  $t$  to the given initial time  $t_0$ . Take voxel  $VC_1 \equiv (i_{1x}, i_{1y}, i_{1z})$  to be the voxel that contains the given initial location of the center of molecule 1.
4. Using Eqs. (15a) and (15b), compute the tentative time  $t_{\text{uni}}$  at which molecule 1 will undergo its unimolecular reaction. If no unimolecular reaction is possible, set  $t_{\text{uni}} = \infty$ .
5. (*Begin simulation loop*) If  $VC_1$  is a voxel belonging to  $V\Sigma_1$ , set  $t_{\text{halt}} = t$  and halt.
6. Determine which if any of the six voxels adjacent to  $VC_1$ —namely voxels  $(i_{1x} \pm 1, i_{1y}, i_{1z})$ ,  $(i_{1x}, i_{1y} \pm 1, i_{1z})$  and  $(i_{1x}, i_{1y}, i_{1z} \pm 1)$ —belong to  $V\partial\Omega_1$ , and hence are not accessible to  $VC_1$  on its next jump. Num-

ber the remaining  $n_1^{\text{open}}$  accessible adjacent voxels with the index  $k = 1, \dots, n_1^{\text{open}}$ . Then compute the tentative time  $t_{\text{hop}}$  of the next jump of  $VC_1$  using Eqs. (16a) and (16b).

7. Branch on which of the two times  $t_{\text{hop}}$  and  $t_{\text{uni}}$  is the *earlier*:<sup>15</sup>
  - a. If  $t_{\text{hop}} < t_{\text{uni}}$ : Set  $t = t_{\text{hop}}$ . Compute the index  $k = k_{\text{dest}}$  of the destination voxel using Eq. (17). Execute the hop by moving  $VC_1 \equiv (i_{1x}, i_{1y}, i_{1z})$  to that destination voxel. Then return to 5.
  - b. If  $t_{\text{uni}} \leq t_{\text{hop}}$ : Set  $t_{\text{halt}} = t_{\text{uni}}$ . Execute the unimolecular reaction, and then halt.

The foregoing procedure is totally focused on what is happening inside this one corral, and is oblivious to what is happening inside other corrals in  $\Omega$ . If it is subsequently determined that a solute molecule from outside this corral has diffused to  $\Sigma$  at a time  $t_{\text{encroach}}$  that is earlier than the time  $t_{\text{halt}}$  determined above, then  $t_{\text{encroach}}$  becomes the actual halting time for our corral. We will then have to backtrack the simulation, and restore molecule 1 to its voxel location at time  $t_{\text{encroach}}$ . The voxel-hopping procedure makes that relatively easy to do: we need only keep a log of the times and the destinations of each hop. More specifically, if the journey of molecule 1 from its initial voxel  $v_0$  at time  $t_0$  to time  $t_{\text{halt}}$  contained a total of  $J$  jumps, then we will keep a record of the time  $t_j$  of the  $j$ th jump, and the voxel  $v_j$  which molecule 1 jumped *to* on that jump, for all  $j = 1, \dots, J$ . Since during each time interval  $[t_{i-1}, t_i]$  molecule 1 was always in voxel  $v_{i-1}$ , the two data sets  $\{t_0, t_1, \dots, t_J\}$  and  $\{v_0, v_1, \dots, v_J\}$  allow us to immediately determine where molecule 1 was at any time  $t' \in [t_0, t_{\text{halt}}]$ . These two data sets can be discarded once it is certain that no encroachment occurred prior to time  $t_{\text{halt}}$ .

#### IV. EXAMPLE SIMULATION OF A ONE-MOLECULE CORRAL EVENT USING THE SVTA

We will illustrate the SVTA's single-molecule simulation procedure with an example that will show that the voxel-hopping rule (5) is indeed a very accurate way to simulate the diffusion equation if we take  $h$  "sufficiently small." To that end, we take the corraling surface  $\Sigma$  to be a sphere that is wholly inside  $\partial\Omega$  and centered on the center of molecule 1 at time 0. And we set  $k_1 = 0$  so that the halting event will always be the arrival of molecule 1 at the surface of  $\Sigma$ . The halting time will therefore be a first-passage time random variable, which we will denote by  $T$ . We take  $D_1 = 1$ , and we let the radius of  $\Sigma$  be equal to 100 units of length plus the radius of molecule 1; thus, the sphere  $\Sigma_1$ , which corrals the center of molecule 1, is a sphere of radius 100 length units.

To generate a sample value of  $T$  using the SVTA, we use the procedure described in Eqs. (16) and (17) with  $n_1^{\text{open}} = 6$  to hop  $VC_1$  from its initial location at the center of  $\Sigma_1$  until it enters a voxel whose center is  $\geq 100$  length units from the center of  $\Sigma_1$ . The final value of  $t$  is then the desired sample of  $T$ . We carried out this procedure for five values of  $h$  ranging from 1 to 20. For each value of  $h$  we made  $10^6$  runs, and from the  $10^6$  sample values of  $T$  thus obtained we constructed a normalized frequency histogram by plotting the number of

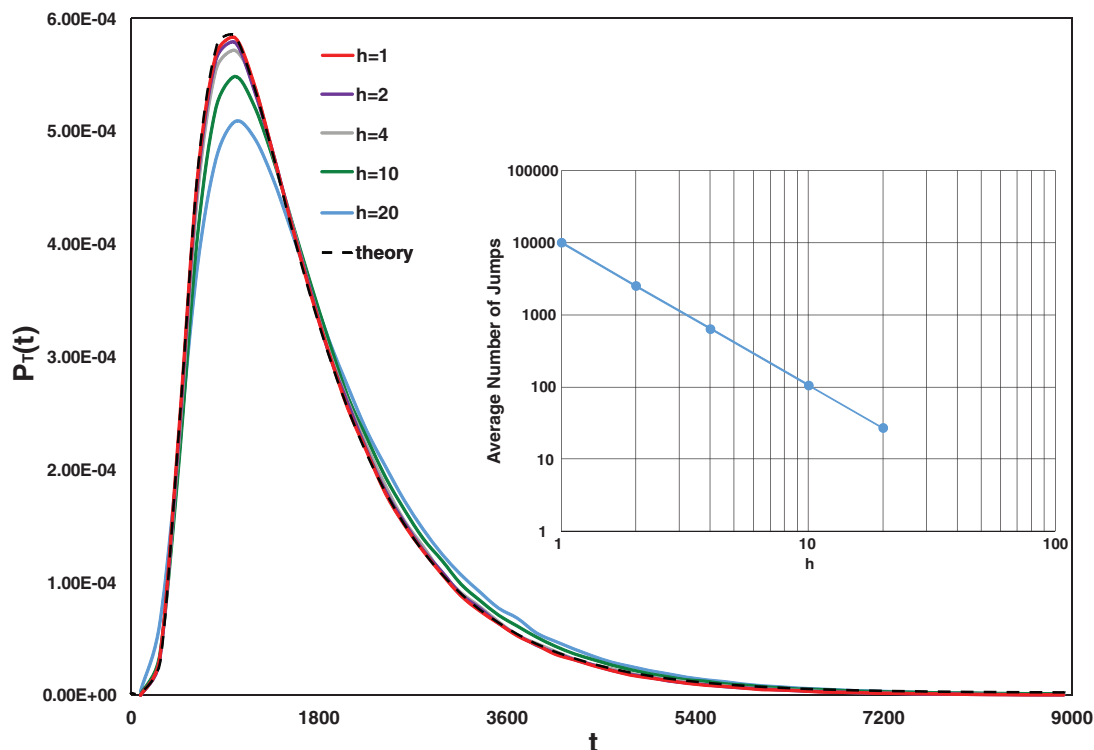


FIG. 3. Example of a first-passage time calculation in the single-molecule SVTA. The dashed black curve is the theoretically predicted PDF  $P_T(t)$  of the first passage time  $T$  for the case in which the corraling surface is a sphere of radius 100 around the initial location of the center of a molecule with diffusion coefficient 1. The colored curves are the normalized frequency histograms provided by SVTA for five  $h$  values,  $h = 1, 2, 4, 10,$  and  $20$ . The theoretical PDF was computed as described in the supplemental material.<sup>16</sup> It can be seen that as  $h \rightarrow 0$ , the SVTA histograms converge to  $P_T(t)$ . The excellent agreement for the three smallest  $h$  values indicates that the SVTA is an accurate algorithm for simulating molecular diffusion. The inset shows a log-log plot of the average number of jumps per run as a function of  $h$ . The slope of the fitted linear curve is  $-1.97 \approx -2$ , which shows that the average number of jumps is proportional to  $h^{-2}$ . Since halving  $h$  increases the number of jumps by a factor of 4, the computation time will increase accordingly. Thus, there is a tradeoff between accuracy and computation time.

sample values that fell into time bin  $j$  divided by the product of the bin width (180) times  $10^6$ . We then compared these histograms not only with each other, but also with a plot of the theoretically predicted PDF of  $T$ ,  $P_T(t)$ , which we computed using essentially the same theory used by eGFRD;<sup>4</sup> for details of how we computed  $P_T(t)$ , see Ref. 16. Our results are shown in Fig. 3. The dashed black curve is the theoretically predicted PDF, and the five colored curves are the normalized frequency histograms for  $h = 1, 2, 4, 10,$  and  $20$ . Each histogram curve is a smooth interpolation of the values found in 50 bins over the time range shown in the figure. The large number ( $10^6$ ) of sample values of  $T$  that we used for each  $h$ -value had the salutary effect of making the one-standard deviation uncertainty range for each bin-value approximately equal to the line width of the curves in the figure.

The expected convergence of the histograms to  $P_T(t)$  as  $h \rightarrow 0$  is clear from the figure. The  $h = 1$  and  $h = 2$  curves are excellent approximations to the theoretically predicted curve, and the  $h = 4$  curve is arguably “good enough.” Accuracy drops off as  $h$  is increased to 10 and then 20. But since for  $h = 20$  the radius of  $\Sigma_1$  is only 5 voxels, it is perhaps more surprising that the  $h = 20$  histogram is as good as it is. The inset figure shows that the average number of jumps per run decreases with increasing  $h$  in proportion to  $h^{-2}$ , falling from 10 066 jumps per run at  $h = 1$  to 27.3 jumps per run at  $h = 20$ . Since the simulation time is roughly proportional to the num-

ber of jumps per run, this implies that the average simulation time for a run also increases with decreasing  $h$  in proportion to  $h^{-2}$ . The trade-off between accuracy and computational effort is illustrated by the findings here that the “excellent”  $h = 2$  curve requires on average 2533 jumps for a first passage, and the “very good”  $h = 4$  curve requires on average 639 jumps for a first passage. The implication seems to be that on the order of a thousand hops need to be made in order for the voxel-hopping method (5) to give an accurate modeling of diffusion.

This example should not be construed as suggesting that the usual goal of the SVTA is to estimate the form of the PDF  $P_T(t)$  of the first-passage time  $T$ . However, estimating  $P_T(t)$  is the goal of the first step of eGFRD,<sup>4</sup> because eGFRD uses that function in its second step to compute a sample value of  $T$ . The third step of eGFRD is to compute the molecule’s arrival point on  $\Sigma_1$ . That task is easy when  $\Sigma_1$  is a sphere, but not if  $\Sigma_1$  has some other shape, or if  $\partial\Omega_1$  forms part of the corral. In contrast, the SVTA accomplishes *all* of these tasks at the end of a *single* run. But of course, that single SVTA run usually entails very many hops.

## V. SVTA’S TWO-MOLECULE CORRAL

Next we consider two solute molecules 1 and 2 inside a corral. As in the one-molecule case just discussed, these

molecules will usually be part of a larger collection of reactant molecules that are all contained within a common boundary  $\partial\Omega$ . But we assume that at an initial time  $t_0$  molecules 1 and 2 have been cordoned off from the other molecules by an imaginary absorbing surface  $\Sigma$ . We define the *four* surfaces  $\Sigma_j$  and  $\partial\Omega_j$  ( $j = 1, 2$ ) as the surfaces generated by the center of molecule  $j$  when molecule  $j$  is rolled around on the insides of  $\Sigma$  and  $\partial\Omega$ , respectively. See Fig. 4(a). In eGFRD<sup>4</sup> the specification of  $\Sigma$  for two molecules is more complicated: it involves two spheres associated with two auxiliary variables which are linear transforms of the position variables of the centers of the two molecules, and no provision is made for portions of  $\partial\Omega$  to form a part of the corral for the molecules.

We subdivide the region accessible to the centers of the molecules into cubic voxels of edge length  $h$ , and we denote by  $VC_1$  and  $VC_2$  the voxels that currently contain the centers of molecules 1 and 2, respectively. See Fig. 4(b). A hop by molecule  $j$  will be implemented as a hop by the voxel  $VC_j$ . Between hops, the center of molecule  $j$  is distributed randomly and uniformly inside  $VC_j$ . We replace the surfaces  $\Sigma_1$ ,  $\Sigma_2$ ,  $\partial\Omega_1$ , and  $\partial\Omega_2$  by the sets of all voxels which those surfaces intersect, denoting those respective voxel sets by  $V\Sigma_1$ ,  $V\Sigma_2$ ,  $V\partial\Omega_1$ , and  $V\partial\Omega_2$ . If  $VC_j$  hops into a voxel belonging to  $V\Sigma_j$  we will halt the simulation. We will not allow  $VC_j$  to hop into a voxel belonging to  $V\partial\Omega_j$  since that would put some part of molecule  $j$  physically outside of  $\Omega$ .

In a straightforward generalization of (18), the SVTA for a two-molecule corral imposes the following conditions on the voxel size  $h$ :

$$\max_{j=1,2} \left( D_i \sqrt{\frac{m_j}{k_B T}} \right) \ll h < \min_{j=1,2} (\sigma_j). \quad (20)$$

Notice that the upper bound on  $h$  here *contradicts* the requirement  $h \gg (\sigma_1, \sigma_2)$  that is imposed by the reaction-diffusion SSA (RD-SSA) on its voxels. The latter requirement arises because the RD-SSA simulates bimolecular reactions inside individual voxels using the bimolecular propensity function, and that will not be physically justified unless the two reacting molecules occlude no more than a negligibly small fraction of a voxel. This difference has two important consequences. First, the SVTA cannot use the bimolecular propensity function to model bimolecular reactions. Second, the much smaller  $h$  values that are used in the SVTA make the simulation of diffusion using the voxel-hopping rule (5) much more accurate in the SVTA than in the RD-SSA. These observations should make it clear that, even though the SVTA and the RD-SSA both use the voxel-hopping strategy (5), they are really *very different algorithms*: the RD-SSA requires its voxels to be large enough that many reactant molecules can fit inside a single voxel, whereas the SVTA requires its voxels to be so small that they cannot contain the center of more than one molecule at a time; the RD-SSA uses the bimolecular propensity function, whereas the SVTA does not; and the SVTA *tracks* individual reactant molecules, whereas the RD-SSA does not.

The plan of the SVTA for a two-molecule corral is to hop  $VC_1$  and  $VC_2$  repeatedly, using the hopping rule (5),

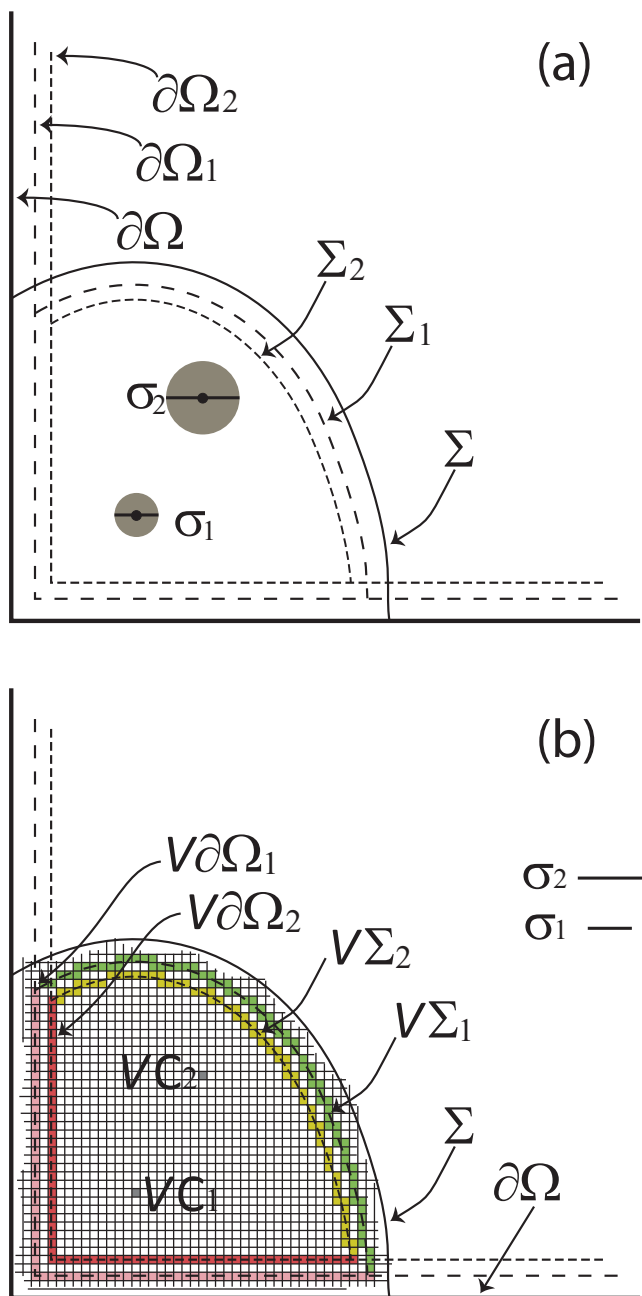


FIG. 4. (a) Two solute molecules of diameters  $\sigma_1$  and  $\sigma_2$ , part of a larger system of solute molecules enclosed in some containing region  $\Omega$ , are at time  $t_0$  separated from the other molecules in  $\Omega$  by an imaginary surface  $\Sigma$  and possibly a portion of the system boundary  $\partial\Omega$ . The surfaces  $\Sigma_1$  and  $\partial\Omega_1$  (dashed lines) are formed by tracing the positions of the center of molecule 1 when that molecule is rolled around on the insides of  $\Sigma$  and  $\partial\Omega$ , respectively. The surfaces  $\Sigma_2$  and  $\partial\Omega_2$  are formed in a similar way by the center of molecule 2. (b) After subdividing the corraling region enclosed by the larger of the  $\Sigma_1$  and  $\Sigma_2$  surfaces ( $\Sigma_1$  in this case) into cubic voxels of edge length  $h < \min(\sigma_1, \sigma_2)$ , the SVTA assumes that the *centers* of molecules 1 and 2 are distributed randomly and uniformly inside the particular voxels in which they respectively reside; those two voxels are shown here in gray, and are labeled  $VC_1$  and  $VC_2$ , respectively. The surfaces  $\Sigma_1$ ,  $\partial\Omega_1$ ,  $\Sigma_2$ , and  $\partial\Omega_2$  are replaced by their voxelated counterparts  $V\Sigma_1$  (in green),  $V\partial\Omega_1$  (in pink),  $V\Sigma_2$  (in yellow), and  $V\partial\Omega_2$  (in red). Although in this pictorial representation  $V\Sigma_1$  and  $V\Sigma_2$  do not have any voxels in common, they will if  $\sigma_1 \approx \sigma_2$ , and in fact they will be identical if  $\sigma_1 = \sigma_2$ . Likewise for  $V\partial\Omega_1$  and  $V\partial\Omega_2$ .

until *either* molecule 1 or molecule 2 reacts unimolecularly, *or*  $VC_1$  hops into a voxel belonging to  $V\Sigma_1$  or  $VC_2$  hops into a voxel belonging to  $V\Sigma_2$ , *or* molecules 1 and 2 react



bimolecularly. The task of determining the tentative time until either molecule reacts unimolecularly is a straightforward generalization of what was done in the one-molecule corral simulation. But hopping  $VC_1$  and  $VC_2$  is more complicated, because now we have to take care that a hop does not result in the molecules overlapping each other. And determining the tentative time when the molecules react bimolecularly is even more complicated; because, in the SVTA two molecules will react with each other only if they *collide*, and collisions occur on the smallest of space-time scales where the diffusion equation (1), and hence also the voxel-hopping rule (5), do not accurately describe the motion of the molecules. We will now describe in turn how the SVTA accomplishes the foregoing three tasks.

### A. Simulating a unimolecular reaction

For the unimolecular reactions, it follows from the definition of  $k_i$  (see Sec. II A) and the addition law of probability that the probability that *either* molecule will react unimolecularly in the next  $dt$  is

$$k_1 dt + k_2 dt = (k_1 + k_2) dt.$$

So, barring interfering eventualities, the time to the next unimolecular reaction will be exponentially distributed with decay constant  $k_1 + k_2$ . The tentative time to a unimolecular reaction can therefore be computed as

$$t_{\text{uni}} = t_0 + \tau_{\text{uni}}, \quad (21a)$$

where

$$\tau_{\text{uni}} = \mathcal{E}(k_1 + k_2). \quad (21b)$$

The molecule that reacts at this time will be

$$\text{molecule 1 if } \mathcal{U}(0, 1) \leq \frac{k_1}{k_1 + k_2}; \text{ otherwise molecule 2.} \quad (22)$$

### B. Simulating a hop

A new challenge in the two-molecule corral is that we must make sure that a hop by  $VC_1$  or  $VC_2$  does not result in the two molecules overlapping each other. This is another significant difference between the SVTA and the RD-SSA: The latter never checks to see if the destination voxel in a hop actually has room for the hopping molecule; the tacit assumption that there *always is* room can be justified, at least approximately, only if the system is dilute in the reactant molecules. Thus, the RD-SSA does not model *molecular crowding*, whereas the SVTA does.

The SVTA's procedure for doing this is based on the simple rule that molecules 1 and 2 will be in collisional contact with each other if and only if the center of each lies on an imaginary sphere of radius  $\sigma_{12}$  that is concentric about the other. These spheres are called "action spheres." To adapt this condition for collisional contact to our voxelated space, we must construct *voxel approximations* of the action spheres:  $VAS_{ij}$  ( $i \neq j$ ) will denote the set of voxels that approximates the action sphere *about* molecule  $i$  *relative to* molecule  $j$ . One way to define  $VAS_{ij}$  would be as the set of all voxels that are intersected by a sphere of radius  $\sigma_{12}$  which is centered

on voxel  $VC_i$ . But analytically determining if a voxel is intersected by a sphere is complicated. So we adopt instead the following nearly equivalent but more computationally convenient definition:  $VAS_{ij}$  is the set of all voxels whose centers are a distance  $\leq \sigma_{12}$  from the center of voxel  $VC_i$  and which have at least one adjacent voxel whose center is a distance  $> \sigma_{12}$  from the center of  $VC_i$ . Appendix B describes a computer-oriented procedure for constructing the set of voxels  $VAS_{ij}$  according to this definition for given values of  $h$  and  $\sigma_{12}$ .

Figure 5 attempts to clarify our definition of the voxelated action spheres in a simpler two-dimensional context. Figure 5(a) reminds us what molecules 1 and 2 are assumed to look like in the "real" world; the dashed circles are the imaginary action spheres around each molecule. Figure 5(b) shows molecules 1 and 2 in a voxelated world. Here the voxels  $VC_1$  and  $VC_2$  that contain the centers of the respective molecules are colored flat gray to emphasize that the center of each molecule is uniformly distributed inside its voxel. Spheres of radius  $\sigma_{12}$  are shown ascribed about the centers of voxels  $VC_1$  and  $VC_2$ . The voxels colored dark blue have their centers inside the sphere about  $VC_2$ , and also have at least one adjacent voxel whose center is not inside that sphere; they comprise the voxelated action sphere  $VAS_{21}$ . The voxels colored light blue similarly comprise the voxelated action sphere  $VAS_{12}$ . The two voxelated action spheres are of course identical structures.  $VAS_{21}$  surrounds and hops with  $VC_2$ , and  $VAS_{12}$  surrounds and hops with  $VC_1$ .

Molecular crowding is modeled in the SVTA simply by making the voxels comprising  $VAS_{ij}$  *inaccessible* to  $VC_j$ , in the same sense that the voxels comprising  $V\partial\Omega_j$  are inaccessible to  $VC_j$ ; i.e.,  $VC_j$  will not be allowed to hop into any voxel that belongs to  $VAS_{ij}$ . This means that before each hop of  $VC_j$ , we must determine which if any of its six adjacent voxels belong to  $VAS_{ij}$ . Since voxelated action spheres may contain hundreds or even thousand of voxels, this task can be very time consuming in two ways: (i) many voxel comparisons have to be made prior to each hop of  $VC_j$ , and (ii) the voxels that belong to  $VAS_{ij}$ , unlike the voxels that belong to  $V\partial\Omega_j$ , move every time  $VC_i$  hops.

Although we have not found a way to alleviate problem (i), we have find a way to alleviate problem (ii): we can determine which adjacent voxels to  $VC_j$  belong to  $VAS_{ij}$  *without having to explicitly hop*  $VAS_{ij}$  every time  $VC_i$  hops. This is accomplished with the help of the *virtual* voxelated action sphere  $vVAS_{ij}$ , which is a copy of  $VAS_{ij}$  that is centered on the origin. Unlike  $VAS_{ij}$ ,  $vVAS_{ij}$  does not hop with  $VC_i$ , and it does not inhibit the motion of  $VC_j$ . The utility of  $vVAS_{ij}$  owes to the following lemma.

*Lemma 1:* The  $\pm x$ -adjacent voxel to  $VC_j$  will belong to  $VAS_{ij}$  if and only if the voxel obtained by subtracting from the coordinates of that voxel the coordinates of  $VC_i$  belongs to  $vVAS_{ij}$ ; and likewise for the  $\pm y$ -adjacent voxel to  $VC_j$  and the  $\pm z$ -adjacent voxel to  $VC_j$ .

Lemma 1 allows us to determine all the voxels adjacent to  $VC_j$  that belong to  $VAS_{ij}$  without directly consulting  $VAS_{ij}$ . Once that is done, the next lemma allows us to immediately determine which voxels adjacent to  $VC_i$  belong to  $VAS_{ji}$ .

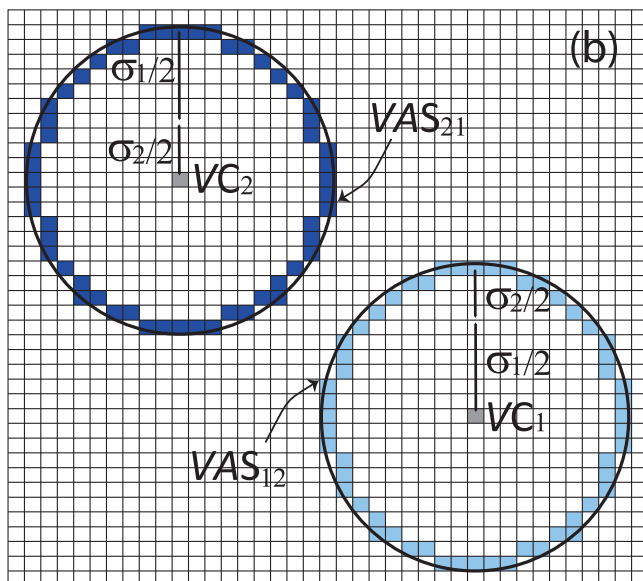
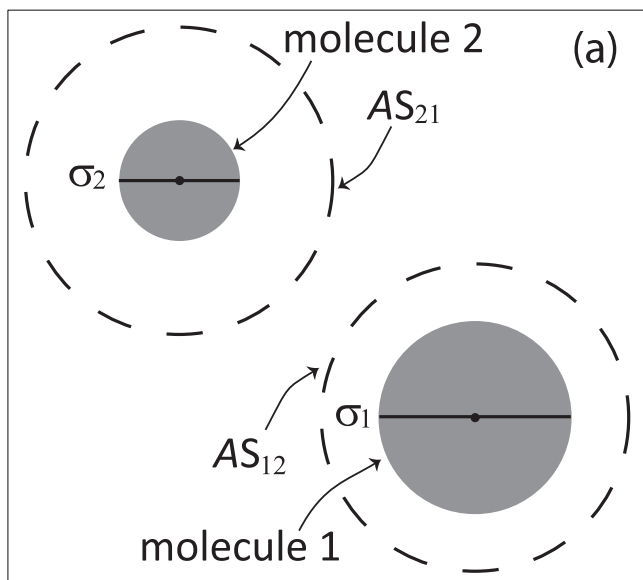


FIG. 5. (a) Molecules 1 and 2 with their respective diameters  $\sigma_1$  and  $\sigma_2$ . We imagine each molecule to be concentrically surrounded by an “action sphere” of radius  $\frac{1}{2}(\sigma_1 + \sigma_2) \equiv \sigma_{12}$ . When the two molecules *collide*, the center of each molecule will be in contact with the action sphere about the other molecule. (b) In the SVTA, the centers of molecules 1 and 2 are assumed to be distributed randomly and uniformly inside voxels  $VC_1$  and  $VC_2$ , respectively. The action sphere about molecule  $j$  is now represented by the set of all voxels whose centers are  $\leq \sigma_{12}$  from the center of voxel  $VC_j$  and which have at least one *adjacent* voxel that does *not* satisfy that condition. Here the *voxelated action sphere*  $VAS_{21}$  about molecule 2 relative to molecule 1 is the set of dark blue voxels, and the voxelated action sphere  $VAS_{12}$  about molecule 1 relative to molecule 2 is the set of light blue voxels. In the SVTA, the two molecules *collide* with each other whenever the *center* of either molecule comes into contact with the *boundary* of the voxelated action sphere about the other molecule. That will be possible only if either  $VC_1$  has hopped to a voxel that is adjacent to  $VAS_{21}$ , or  $VC_2$  has hopped to a voxel that is adjacent to  $VAS_{12}$ .  $VAS_{12}$  hops rigidly with  $VC_1$ , and  $VAS_{21}$  hops rigidly with  $VC_2$ ; however, in an actual simulation we do not need to explicitly execute the hops of either voxelated action sphere.

*Lemma 2:* If the  $\pm x$ -face of  $VC_j$  abuts  $VAS_{ij}$ , then the  $\mp x$ -face of  $VC_i$  will abut  $VAS_{ji}$ ; likewise for the  $\pm y$ -faces and the  $\pm z$ -faces of  $VC_j$ .

The foregoing is the procedure that the SVTA uses before each hop by either  $VC_1$  or  $VC_2$  to determine the number  $n_{12}^{VAS}$  of voxels adjacent to  $VC_1$  that belong to  $VAS_{21}$ —a number that is also equal to the number of voxels adjacent to  $VC_2$  that belong to  $VAS_{12}$ . The number  $n_j^{\text{open}}$  of adjacent voxels to  $VC_j$  that are *accessible* to  $VC_j$  on its next jump will then be 6 minus  $n_{12}^{VAS}$  minus the number of voxels adjacent to  $VC_j$  that belong to  $V\partial\Omega_j$ . (The latter number might not be the same for  $j = 1$  and 2, so  $n_1^{\text{open}}$  will not always be equal to  $n_2^{\text{open}}$ .) Since all this information is obtained without ever explicitly referencing the two ever-moving voxelated action spheres  $VAS_{21}$  and  $VAS_{12}$ , a hop will always involve *explicitly* relocating *only one* voxel—namely  $VC_1$  or  $VC_2$ .

Using the foregoing procedure to determine before each hop which voxels adjacent to  $VC_1$  and  $VC_2$  are open, the procedure for executing the next hop is a fairly straightforward extension of the procedure used in the one-molecule corral. Applying the addition law of probability to the hopping rule (5), we see that the probability that *either* molecule will hop in the next  $dt$  is

$$n_1^{\text{open}} \frac{D_1}{h^2} dt + n_2^{\text{open}} \frac{D_2}{h^2} dt = \frac{n_1^{\text{open}} D_1 + n_2^{\text{open}} D_2}{h^2} dt.$$

This implies that if  $t$  is the current time, then the tentative time of the next hop is

$$t_{\text{hop}} = t + \tau_{\text{hop}}, \quad (23a)$$

where

$$\tau_{\text{hop}} = \mathcal{E} \left( \frac{n_1^{\text{open}} D_1 + n_2^{\text{open}} D_2}{h^2} \right). \quad (23b)$$

Since  $n_j^{\text{open}} D_j / (n_1^{\text{open}} D_1 + n_2^{\text{open}} D_2)$  is the probability that it will be molecule  $j$  that hops at this time, then the molecule that makes the hop can be computed as

$$\begin{aligned} &\text{molecule 1 if } \mathcal{U}(0, 1) \leq \frac{n_1^{\text{open}} D_1}{n_1^{\text{open}} D_1 + n_2^{\text{open}} D_2}; \\ &\text{otherwise molecule 2.} \end{aligned} \quad (24)$$

The destination of the hopping molecule  $j^*$  is then obtained by selecting one of that molecule’s  $n_{j^*}^{\text{open}}$  accessible adjacent voxels “at random.” One way to do that is to number those voxels in any order from 1 to  $n_{j^*}^{\text{open}}$  and then take the number of the destination voxel to be

$$k_{\text{dest}} = [1 + n_{j^*}^{\text{open}} \cdot \mathcal{U}(0, 1)], \quad (25)$$

where  $[x]$  denotes “the greatest integer in  $x$ .” We implement the hop by moving  $VC_{j^*}$  accordingly.

### C. Simulating a collision

The third and final part of the SVTA’s two-molecule corral plan is to determine the tentative time when molecules 1 and 2 will react *bimolecularly*. But here we face a problem: molecules 1 and 2 will react with each other only if they collide, and collisions occur on small space-time scales where neither the diffusion equation (1) nor the voxel-hopping rule (5) is physically valid. In devising a rule for

when two molecules *collide* in voxelated space, we cannot just make something up; we must be guided by the physics that describes the motion of solute molecules. To that end, we turn to a result that is used in the derivation of the bimolecular propensity function (which ironically is *not* used by the SVTA). As stated in Appendix A, the fact that solute molecules move *ballistically* on small space-time scales implies the result (A4). That result says that if the PDF of a solute molecule near a small surface element of area  $a$  has the value  $p_0$ , then the probability that the molecule will *impinge on* or *collide with* that surface element in the next infinitesimal time  $dt$  will be equal to  $\frac{1}{4}p_0\bar{v} \cdot a \cdot dt$ , where  $\bar{v}$  is the average speed the molecule would have if it were in an ideal gas at the same temperature. As a preliminary test of this result in a voxel context, let us use it to calculate the probability that the center of our molecule 1, which is uniformly distributed inside voxel  $VC_1$ , will impinge on (or collide with) a particular one of  $VC_1$ 's six faces. The value of the constant PDF of the center of molecule 1 inside  $VC_1$  must be equal to the reciprocal of the voxel volume  $h^3$  for normalization, and the area of a voxel face is  $h^2$ . So, according to (A4), the probability that the center of molecule 1 will impinge on a given face of  $VC_1$  in the next  $dt$  is

$$\frac{1}{4}p_0\bar{v} \cdot a \cdot dt = \frac{1}{4}h^{-3}\bar{v}_1 \cdot h^2 \cdot dt = \frac{\bar{v}_1}{4h} \cdot dt. \quad (26)$$

Here,  $\bar{v}_1 = \sqrt{8k_B T / (\pi m_1)}$  is the average speed of molecule 1 in a dilute gas at temperature  $T$ . Is this result plausible? It evidently implies that the reciprocal of the coefficient of  $dt$  on the right of Eq. (26) is the *average time* to the impingement. Physically, we expect that {the average time until the center of molecule 1 *impinges on a voxel face*} should be much *smaller* than {the average time until the center of molecule 1 *hops across that voxel face* and uniformly populates the voxel on the other side}. The latter time is of course the reciprocal of the hopping probability rate  $D_1/h^2$ . It is not hard to show that the requirement  $h \gg D_1\sqrt{m_1/k_B T}$  in (20) implies exactly the expected result:

$$4h/\bar{v}_1 \ll h^2/D_1. \quad (27)$$

So we conclude that the result (26) is indeed plausible, and hence that (A4) should enable us to correctly describe a collision between two molecules in voxelated space.

Returning to that problem, we note that the two “adjacency” situations depicted in Figs. 6(a) and 6(b) could have come about as a result of several possible hops. In Fig. 6(a), for example,  $VC_1$  and  $VAS_{12}$  could have just jumped one voxel up or down or left; or  $VC_2$  and  $VAS_{21}$  could have just jumped one voxel down or up or right. Regardless of which of these six hops led to the situation depicted in Fig. 6(a), *molecules 1 and 2 have not yet collided*. But a collision is now *possible*: molecules 1 and 2 will *collide* whenever the *center* of either molecule *impinges* on a voxel face that is shared with the voxelated action sphere about the other molecule. However, when we work on the small length scale of an impingement, we must abandon some voxel-inspired approximations that were acceptable on larger length scales: On small length scales, the action sphere about each molecule can no longer be regarded as a stationary structure that is centered on the *voxel*

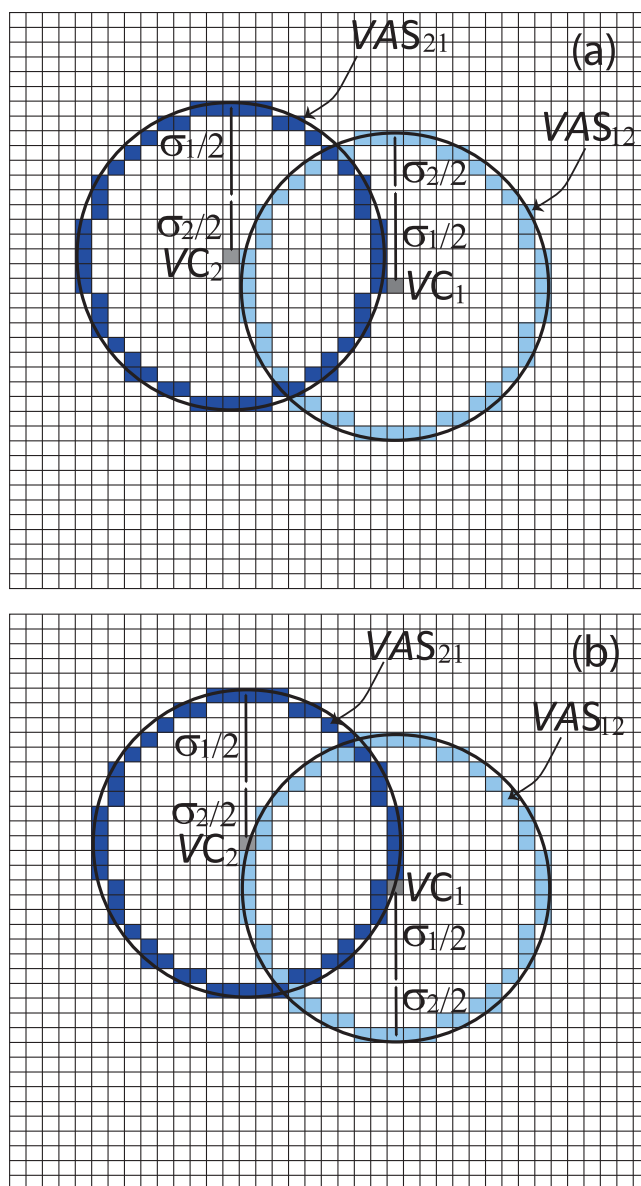


FIG. 6. (a)  $VC_1$  shares one voxel face with  $VAS_{21}$ , and  $VC_2$  shares one voxel face with  $VAS_{12}$ . (b) Two voxel faces are being shared; the number of shared faces will always be the same. A collision between molecules 1 and 2 requires either the center of molecule 1, which is uniformly distributed inside voxel  $VC_1$ , to impinge on the face of a voxel that belongs to  $VAS_{21}$ , or the center of molecule 2, which is uniformly distributed inside voxel  $VC_2$ , to impinge on the face of a voxel that belongs to  $VAS_{12}$ . But on the small spatiotemporal scale of a collision, those two events will always occur *together*; also, since the action sphere about molecule  $j$  is really centered on molecule  $j$  rather than voxel  $VC_j$ , then each voxelated action sphere will see the center of the other molecule moving with the *relative* speed of the two molecules.

containing the molecule's center; it must instead be regarded as a structure that is centered on, and rigidly moving with, the *molecule* which it surrounds. That implies that the average speed of the impingement will be the average *relative* speed of the two molecules. It also implies that an impingement of the center of molecule 1 on the action sphere about molecule 2 will always be *accompanied* by the impingement of the center of molecule 2 on the action sphere about molecule 1; i.e., although molecules 1 and 2 *hop independently*, they *collide*

together. Therefore, a *collision* between molecules 1 and 2 is the *impingement of the center of molecule 1 on any face of its containing voxel*  $VC_1$  that is shared with  $VAS_{21}$ , and the impingement occurs at the average relative speed  $\bar{v}_{12}$  of the two molecules.

With this perspective, we can now use (A4) to compute the probability of a *collision* between molecules 1 and 2 in the next  $dt$  when  $VC_1$  shares  $n_{12}^{VAS}$  faces with  $VAS_{21}$  (in which case  $VC_2$  will share  $n_{12}^{VAS}$  faces with  $VAS_{12}$ ). Since the PDF of the center of molecule 1 everywhere inside  $VC_1$  is  $p_0 = h^{-3}$ , and since the total surface area shared by  $VC_1$  and  $VAS_{21}$  is  $a = n_{12}^{VAS} \cdot h^2$ , then the impingement probability formula yields

$$\begin{aligned} \frac{1}{4} \cdot p_0 \cdot \bar{v} \cdot a \cdot dt &= \frac{1}{4} \cdot h^{-3} \cdot \bar{v}_{12} \cdot (n_{12}^{VAS} h^2) \cdot dt \\ &= \left( \frac{n_{12}^{VAS} \bar{v}_{12}}{4h} \right) \cdot dt, \end{aligned} \quad (28)$$

where  $\bar{v}_{12}$  is given by Eq. (10). Therefore, defining

$$\beta_{12} \equiv \frac{n_{12}^{VAS} \bar{v}_{12}}{4h}, \quad (29)$$

we conclude that

$$\beta_{12} \cdot dt = \text{the probability that molecules 1 and 2 will collide in the next } dt. \quad (30)$$

This result implies that if, at the current time  $t$ ,  $VC_1$  shares  $n_{12}^{VAS}$  faces with  $VAS_{21}$ , then unless a unimolecular reaction or a hop occurs first, molecules 1 and 2 will *collide* at time

$$t_{\text{col}} = t + \tau_{\text{col}}, \quad (31a)$$

where

$$\tau_{\text{col}} = \mathcal{E}(\beta_{12}). \quad (31b)$$

#### D. Simulating a bimolecular reaction

The collision time  $\tau_{\text{col}}$  would not be needed if  $q_{12} = 0$ , because then it would not matter when or even if molecules 1 and 2 collide. All that is needed when  $q_{12} = 0$  is the rule that  $VC_j$  is not allowed to hop into any voxel belonging to  $VAS_{ij}$ . But if  $q_{12} > 0$ , we need the result (31) to determine if and when a *reaction* between the two molecules occurs. We do that as follows: Let  $n_{\text{col}}$  be the number of collisions between molecules 1 and 2 that are required to initiate a reaction. Since  $q_{12}$  is the probability that one collision will produce a reaction, then

$$\Pr(n_{\text{col}} = n) = (1 - q_{12})^{n-1} q_{12} \quad (n = 1, 2, \dots). \quad (32)$$

This functional form implies that  $n_{\text{col}}$  is the *geometric random variable* with mean  $1/q_{12}$ :

$$n_{\text{col}} = \text{Geo}(q_{12}). \quad (33)$$

So, at the initial time  $t_0$ , we generate a random value for  $n_{\text{col}}$  according to Eq. (33). Then at each subsequent collision between molecules 1 and 2 we reduce  $n_{\text{col}}$  by 1, and we execute the bimolecular reaction when  $n_{\text{col}}$  reaches zero.<sup>17</sup> Note that the collisions might be interspersed with hops, and thus

be distributed over several separate  $VC_1$ - $VAS_{21}$  adjacencies, whose values for  $\beta_{12}$  cannot be predicted in advance.

To summarize: In a two-molecule corral, the next *event* will always be either a *unimolecular reaction* of either molecule, or a *diffusional hop* by either molecule, or a *collision* between the two molecules. We have computed in Eqs. (21), (23), and (31) the tentative times  $t_{\text{uni}}$ ,  $t_{\text{hop}}$ , and  $t_{\text{col}}$  for these possible next events. If  $t_{\text{col}} < \min(t_{\text{hop}}, t_{\text{uni}})$ , we will execute the collision by advancing  $t$  to  $t_{\text{col}}$  and simultaneously decrementing  $n_{\text{col}}$  (which was generated according to Eq. (33) at time  $t_0$ ) by 1. If that decrement makes  $n_{\text{col}} = 0$ , we execute the bimolecular reaction and halt. Otherwise, since the collision did not change the voxel location of either molecule, we will generate a new tentative time of the next collision according to Eqs. (31a) and (31b) and continue the simulation. But if instead  $t_{\text{col}} \geq \min(t_{\text{hop}}, t_{\text{uni}})$ , we will proceed as follows: if  $t_{\text{hop}} < t_{\text{uni}}$ , we will advance  $t$  to  $t_{\text{hop}}$ , execute a hop, and then continue the simulation; otherwise, we will advance  $t$  to  $t_{\text{uni}}$ , execute a unimolecular reaction, and halt. The full simulation procedure for the two-molecule SVTA is outlined in Fig. 7.

#### VI. EQUIVALENCE OF THE SVTA AND eGFRD

We will now show that the SVTA and eGFRD<sup>4</sup> are equivalent, in that they can be derived from the same line of reasoning. That line of reasoning is the derivation in Appendix A of the bimolecular propensity function and its rate constant  $k$  in Eq. (7), but with one important difference: the bimolecular propensity function requires the system to be *dilute* in the solute molecules, but the SVTA and eGFRD do *not*. This difference arises because any analysis of the interaction between two molecules makes the tacit assumption that their relative motion will not be affected by the presence of a third reactant molecule. The SVTA and eGFRD satisfy that assumption by putting a corral around the two molecules. But the bimolecular propensity function, which underlies both the CME and the SSA, does not know where any of the reactant molecules really are, so the only way it can satisfy the assumption is to require the reactant molecules be so *dilute* in  $\Omega$  that clusterings of three or more molecules will be so infrequent that they can be ignored. We will now show that, except for this difference, the SVTA and eGFRD follow from the same reasoning used to derive Eq. (7).

As noted in Appendix A, the derivation of Eq. (7) proceeds in two phases. The first phase uses the Einstein diffusion equation (1) to derive Eq. (A3). The second phase takes account of the fact that on the small length scales where molecular collisions occur, the *physical* motion of a solute molecule is actually ballistic, and hence not described by the diffusion equation. The second phase combines the ballistic motion result (A4) with the definition of  $q_{12}$  to obtain Eq. (A6). When Eq. (A6) is combined with Eq. (A3) from the first phase of the derivation, the result is the explicit formula for  $k$  in Eq. (7).

That the SVTA's procedure for simulating a bimolecular reaction follows this same line of reasoning can be seen as follows. Just as the first phase of the derivation of  $k$  used the *diffusion equation* in the region outside the action sphere around molecule 2 to derive Eq. (A3), the SVTA begins by using the



### Initialization

1. Divide the space that is accessible to the centers of molecules 1 and 2 into cubic voxels of edge length  $h$ , subject to condition (20). Label each voxel with three signed integers so that the center of voxel  $(i_x, i_y, i_z)$  is at point  $(x, y, z) = (i_x h, i_y h, i_z h)$ .
2. Identify the voxels that belong to the voxelated corraling surfaces  $V\Sigma_1$  and  $V\Sigma_2$ , and also the voxels that belong to the voxelated boundaries  $V\partial\Omega_1$  and  $V\partial\Omega_2$ .
3. From the given values for  $\sigma_{12}$  and  $h$ , construct the virtual voxelated action sphere  $vVAS_{21}$ , centered on  $(0,0,0)$ , according to the procedure outlined in Appendix B.
4.  $t \leftarrow t_0$ . Take  $VC_1 \equiv (i_{1x}, i_{1y}, i_{1z})$  and  $VC_2 \equiv (i_{2x}, i_{2y}, i_{2z})$  to be the voxels that contain the given initial positions of the centers of molecules 1 and 2.
5. Compute the time  $t_{\text{uni}}$  at which a unimolecular reaction will occur, using Eqs. (21). If no unimolecular reaction is possible,  $t_{\text{uni}} \leftarrow \infty$ .
6. If  $q_{12} > 0$ , compute the number  $n_{\text{col}}$  of collisions needed to produce a bimolecular reaction using Eq. (33).

### Main Simulation Loop

7. If  $VC_1 \in V\Sigma_1$  or  $VC_2 \in V\Sigma_2$ , then  $t_{\text{halt}} \leftarrow t$ . Record  $t_{\text{halt}}$  and the molecular positions, and then halt.
8. For  $j=1$  and 2, determine which of the six voxels adjacent to  $VC_j$  are accessible to  $VC_j$  on its next hop, and collect those voxels in a set indexed from 1 to  $n_j^{\text{open}}$ ; also determine the number  $n_{12}^{\text{VAS}}$  of voxels adjacent to  $VC_1$  that belong to  $VAS_{21}$ . The procedure for doing all that is this:

$$n_{12}^{\text{VAS}} \leftarrow 0.$$

$$\{\text{Set of open adjacent voxels for } VC_1\} \leftarrow \emptyset.$$

$$\{\text{Set of open adjacent voxels for } VC_2\} \leftarrow \emptyset.$$

- \* If  $(i_{1x}+1-i_{2x}, i_{1y}-i_{2y}, i_{1z}-i_{2z}) \in vVAS_{21}$ , then it must be that  $(i_{1x}+1, i_{1y}, i_{1z}) \in VAS_{21}$ , so  $n_{12}^{\text{VAS}} \leftarrow n_{12}^{\text{VAS}} + 1$ .

Otherwise:

If  $(i_{1x}+1, i_{1y}, i_{1z}) \notin V\partial\Omega_1$ , then add  $(i_{1x}+1, i_{1y}, i_{1z})$  to the set of open voxels for  $VC_1$ .

If  $(i_{2x}-1, i_{2y}, i_{2z}) \notin V\partial\Omega_2$ , then add  $(i_{2x}-1, i_{2y}, i_{2z})$  to the set of open voxels for  $VC_2$ .

Repeat \* for the other adjacent voxels to  $VC_1$ , namely  $(i_{1x}-1, i_{1y}, i_{1z})$ ,  $(i_{1x}, i_{1y}\pm 1, i_{1z})$ , and  $(i_{1x}, i_{1y}, i_{1z}\pm 1)$ .

Index the set of open voxels for  $VC_1$  from 1 to  $n_1^{\text{open}}$ .

Index the set of open voxels for  $VC_2$  from 1 to  $n_2^{\text{open}}$ .

9. Generate the time  $t_{\text{hop}}$  of the next hop per Eqs. (23).
10. Generate the time  $t_{\text{col}}$  of the next potentially reacting collision:

If  $(q_{12} > 0$  and  $n_{12}^{\text{VAS}} \geq 1)$ , then  $\tau_{\text{col}} \leftarrow \mathcal{E}(\beta_{12})$ .

Otherwise,  $\tau_{\text{col}} \leftarrow \infty$ .

$$t_{\text{col}} \leftarrow t + \tau_{\text{col}}.$$

11. Now that we know  $t_{\text{uni}}$ ,  $t_{\text{hop}}$  and  $t_{\text{col}}$ , we can determine what happens next, and make it happen:

If  $t_{\text{col}} \leq \min(t_{\text{hop}}, t_{\text{uni}})$ , the next event is a collision:

$$t \leftarrow t_{\text{col}}.$$

$$n_{\text{col}} \leftarrow n_{\text{col}} - 1.$$

If  $n_{\text{col}} = 0$ :

Execute the bimolecular reaction.

$$t_{\text{halt}} \leftarrow t.$$

Record  $t_{\text{halt}}$  and the molecular positions, and then halt.

Otherwise, return to step 10.

Otherwise the next event is either a unimolecular reaction or a hop:

If  $t_{\text{hop}} < t_{\text{uni}}$ , the next event is a hop:

$$t \leftarrow t_{\text{hop}}.$$

Execute a hop of either  $VC_1$  or  $VC_2$  according to Eqs. (24) and (25).

Return to step 7.

Otherwise, the next event is a unimolecular reaction:

$$t_{\text{halt}} \leftarrow t_{\text{uni}}.$$

Execute the unimolecular reaction per Eq. (22).

Record  $t_{\text{halt}}$  and the molecular positions, and then halt.

If another solute molecule from outside the corral diffuses to  $\Sigma$  at a time  $t_{\text{encroach}}$  which is earlier than  $t_{\text{halt}}$ , then we will have to backtrack the simulation and restore molecules 1 and 2 to their voxel locations at time  $t_{\text{encroach}}$ , which becomes the actual halting time. To do that, we keep a log of the times and the destinations of each jump: if the journey of  $VC_1$  and  $VC_2$  from their initial voxels  $v_{1,0}$  and  $v_{2,0}$  at time  $t_0$  to time  $t_{\text{halt}}$  contained a total of  $J$  jumps, then we log the time  $t_i$  of the  $i^{\text{th}}$  jump, along with the voxels  $v_{1,i}$  and  $v_{2,i}$  that  $VC_1$  and  $VC_2$  were respectively in immediately *after* the  $i^{\text{th}}$  jump, for all  $i=1, \dots, J$ . Since  $VC_1$  and  $VC_2$  were in the respective voxels  $v_{1,i-1}$  and  $v_{2,i-1}$  during  $[t_{i-1}, t_i)$ , then the three sequences  $\{t_0, \dots, t_J\}$ ,  $\{v_{1,0}, \dots, v_{1,J}\}$ , and  $\{v_{2,0}, \dots, v_{2,J}\}$  enable us to determine where both molecules were at any instant  $t' \in [t_0, t_{\text{halt}})$ . Those three data sets can be discarded once it is certain that no encroachment occurred prior to time  $t_{\text{halt}}$ .

FIG. 7. Outline of the two-molecule SVTA procedure.

*voxel-hopping procedure* (5) to bring molecules 1 and 2 to within *almost* contact distance of each other. The smallness of  $h$  demanded by the SVTA ensures that the voxel-hopping procedure will give an accurate rendering of the motion prescribed by the diffusion equation.<sup>8</sup> Next, just as the second phase of the derivation of  $k$  combined the ballistic result (A4) with the definition of  $q_{12}$  to obtain Eq. (A6), the SVTA uses the ballistic result (A4) to obtain Eqs. (29) through (31), and then invokes the definition of  $q_{12}$  to obtain Eqs. (32) and (33). Thus we see that the SVTA's procedure for simulating a bimolecular reaction emerges from the same logic that gave us the bimolecular reaction rate constant  $k$  in Eq. (7), although without the latter's requirement that the system be dilute in the reactant molecules.

To show that eGFRD's procedure for simulating a bimolecular reaction likewise finds its justification in the derivation of Eq. (7), we begin by noting that both that derivation and eGFRD use the Einstein diffusion equation (1), with  $D$  replaced by  $D_{12} \equiv D_1 + D_2$ , to describe the motion of molecule 1 relative to molecule 2 in the region outside the action sphere around molecule 2. Indeed, in Appendix A, the PDF  $p_1(\mathbf{r}, t | \mathbf{r}_0, t_0)$  of the position of the center of molecule 1 relative to the center of molecule 2 is none other than the "Green's function" of eGFRD. But let us focus now on Eqs. (A1) and (A5) in Appendix A. In both of those equations,  $da$  is an infinitesimal area element on the surface of the action sphere at point  $\mathbf{r}' = \sigma_{12} \hat{\mathbf{r}}'$ . Equation (A1) encapsulates the *first* phase of the derivation of  $k$ : it describes the initial diffusional approach of the two molecules over center-separation distances that are large compared to  $\sigma_{12}$ . Equation (A5) encapsulates the *second* phase of the derivation of  $k$ ; it describes the final ballistic approach of the two molecules when the distance between their centers is nearly equal to  $\sigma_{12}$ . But since the *right* sides of Eqs. (A1) and (A5) describe the same probability, then their *left* sides must be equal:

$$\begin{aligned} D_{12} \nabla_{\mathbf{r}} p_1(\mathbf{r}', t | \mathbf{r}_0, t_0) \cdot \hat{\mathbf{r}}' \cdot da \cdot dt \\ = \frac{1}{4} p_1(\mathbf{r}', t | \mathbf{r}_0, t_0) \bar{v}_{12} q_{12} \cdot da \cdot dt. \end{aligned}$$

Simplifying this gives

$$\nabla_{\mathbf{r}} p_1(\mathbf{r}', t | \mathbf{r}_0, t_0) \cdot \hat{\mathbf{r}}' = \left( \frac{\bar{v}_{12} q_{12}}{4D_{12}} \right) p_1(\mathbf{r}', t | \mathbf{r}_0, t_0). \quad (34a)$$

Equation (34a) says that the gradient of  $p_1$  at any point on the action sphere must at all times be equal to a certain constant times the value of  $p_1$  at that point. Equation (34a) can therefore be used as a *boundary condition* for the diffusion equation at the action sphere; indeed, it is a kind of boundary condition that is known generically as a "radiation" boundary condition. Since the derivation we just gave of Eq. (34a) made use of the ballistic relation (A4), then solving the diffusion equation subject to boundary condition (34a) should yield a solution that takes proper account of the ballistic motion of molecule 1 relative to molecule 2 in the immediate vicinity of the action sphere. And that is exactly what eGFRD<sup>4</sup> does, although it eliminates the parameter  $q_{12}$  in Eq. (34a) in favor of the parameter  $k_{\text{bal}}$  via Eq. (14) to obtain the equivalent

formula

$$4\pi\sigma_{12}^2 D_{12} \nabla_{\mathbf{r}} p_1(\mathbf{r}', t | \mathbf{r}_0, t_0) \cdot \hat{\mathbf{r}}' = k_{\text{bal}} p_1(\mathbf{r}', t | \mathbf{r}_0, t_0). \quad (34b)$$

Equation (34b) is eGFRD's radiation boundary condition as written in Eq. (S20) of Ref. 4, allowing for the fact, mentioned earlier in connection with Eq. (13), that the "intrinsic rate constant"  $k_a$  in Ref. 4 is the same as our  $k_{\text{bal}}$ . (We note as an aside that Eq. (A7) is a more explicit rendering of the radiation boundary condition when the system is dilute and well-mixed.) The fact that eGFRD's radiation boundary condition (34b) can be derived from the same physics that underlies the SVTA establishes the logical equivalence of these two simulation procedures.

Those who have long known, e.g., from the work of Bicoût *et al.*,<sup>18</sup> that imposing the radiation boundary condition (34b) on the diffusion equation yields a physically correct modeling of a bimolecular reaction, might choose to adopt the following perspective: the SVTA's procedures at Eqs. (29)–(33) amount to a way of imposing the radiation boundary condition on the *voxel-hopping method of simulating* the diffusion equation.

## VII. EXAMPLE SIMULATIONS WITH THE TWO-MOLECULE SVTA

We will illustrate the two-molecule SVTA simulation procedure of Sec. V by using a simplified version of it in which molecules 1 and 2 are the only reactant molecules in  $\Omega$ , there is no corraling surface  $\Sigma$ , and the molecules cannot react unimolecularly. Thus, the only halting event will be the reaction of the two molecules with each other. We will assume that molecules 1 and 2 have the same diameters, so that the surfaces  $\partial\Omega_1$  and  $\partial\Omega_2$  which confine their centers will coincide. To lend some element of realism to our example, we will assign values to the physical parameters that are roughly typical of real cellular systems [see the discussion leading to Eq. (19)]:

$$\sigma_1 = \sigma_2 = \sigma_{12} = 6.1 \times 10^{-9} \text{ m}, \quad (35a)$$

$$D_1 = D_2 = \frac{1}{2} D_{12} = 0.7 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}, \quad (35b)$$

$$m_1 = m_2 = 2m_{12} = 1.66 \times 10^{-22} \text{ kg}. \quad (35c)$$

With  $T = 293\text{K}$ , Eq. (10) gives for the average relative speed of the two molecules

$$\bar{v}_{12} = 11.14 \text{ m s}^{-1}. \quad (35d)$$

Lacking guidance for a realistic choice for the collision-conditioned reaction probability  $q_{12}$ , we will arbitrarily take

$$q_{12} = 0.2, \quad (35e)$$

so that on average 5 collisions will be needed to induce a reaction. Finally, we will take the containing volume  $\Omega$  to be a cube of edge length  $H$ . For our first three examples, we will assign  $H$  the value

$$H = 128.71 \times 10^{-9} \text{ m}, \quad (36)$$

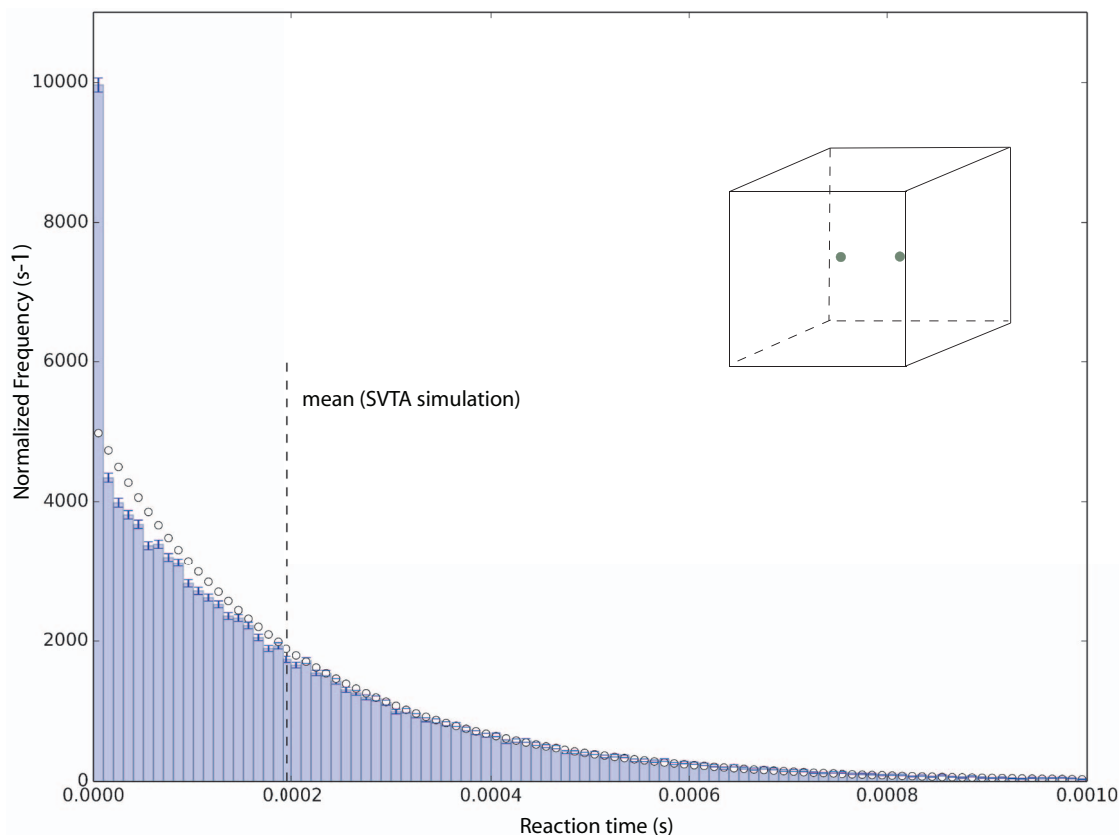


FIG. 8. A normalized frequency histogram of  $10^5$  times to reaction of two molecules in a cubic container, as simulated by the SVTA. The parameter values are as given in Eqs. (35)–(37). The inset figure shows the size of the two molecules in relation to the size of  $\Omega$ , as well as the initial positions of the centers of the two molecules at points  $(-3\sigma_{12}, 0, 0)$  and  $(3\sigma_{12}, 0, 0)$ . The vertical bar around each histogram bin level denotes the one-standard deviation statistical uncertainty interval for that bin. The vertical dashed line is the mean of the data, and the open-circle curve is the exponential distribution corresponding to that mean. Owing to the discrepancies between the histogram and the open-circle curve below the mean, especially in the first histogram bin, the distribution of the time to reaction is not exponential. That implies that these reaction times cannot be characterized by a reaction rate constant; i.e., there is no constant  $c$  such that  $c \cdot dt$  gives the probability that molecules 1 and 2 will react with each other in the next infinitesimal time  $dt$ .

which is slightly over 21 times larger than the diameter of each molecule. And for the voxel edge length  $h$ , we will choose the value

$$h = 0.61 \times 10^{-9} \text{ m}, \quad (37)$$

so that there are exactly 201 voxels that are accessible to  $VC_1$  and  $VC_2$  along each axis of the cube  $\Omega$ . The value of the left side of condition (20) in this case is  $0.014 \times 10^{-9}$  m, so the lower bound condition on  $h$  is well satisfied.

In our first example, molecules 1 and 2 are initially placed near the middle of the cube  $\Omega$ , with their edges separated by five molecular diameters. This initial condition is depicted in the inset to Fig. 8, in which the sizes of the molecules and  $\Omega$  are drawn to scale. With this initial condition, we made  $10^5$  runs of the STVA. A normalized frequency histogram of the times to reaction thus obtained is shown in Fig. 8. The vertical bar over each histogram bin shows the one-standard deviation uncertainty limits for that bin. The dashed vertical line locates the mean of the data. The open-circle curve shows what the histogram would look like if the data were *exponentially* distributed with that mean. We see that while the match between that exponential distribution and the histogram is good above the mean, it is not good below the mean; in particular, the large spike in the first bin makes it clear that these

reaction times are *not* exponentially distributed. That fact has an interesting consequence: there is *no reaction rate constant* which characterizes these reaction times; i.e., there is no constant  $c$  such that  $c \cdot dt$  gives the probability that molecules 1 and 2 will react with each other in the next infinitesimal time  $dt$ .

Our second example is the same system but with a different initial condition. Now we start with molecules 1 and 2 in opposite corners of  $\Omega$ , as shown in the inset to Fig. 9. The accompanying histogram is of the reaction times obtained in  $10^5$  runs of the STVA for this initial condition. Compared to the histogram in Fig. 8, the mean here is 18% larger, and the spike in the first bin in Fig. 8 has been replaced by a deep valley. But once again, the histogram is not exponential, so there is no reaction rate constant.

The results in Figs. 8 and 9 are not surprising: it is perfectly reasonable that if the molecules are initially close together their time to reaction will tend to be smaller than if they were initially far apart. But the result that neither case can be described by a reaction rate constant may come as a surprise to many. Some might counter that this simply means that the reaction rate in these cases is not a constant, but rather a function of time. However, the facts that (i) the form of that function of time apparently can be deduced only through a

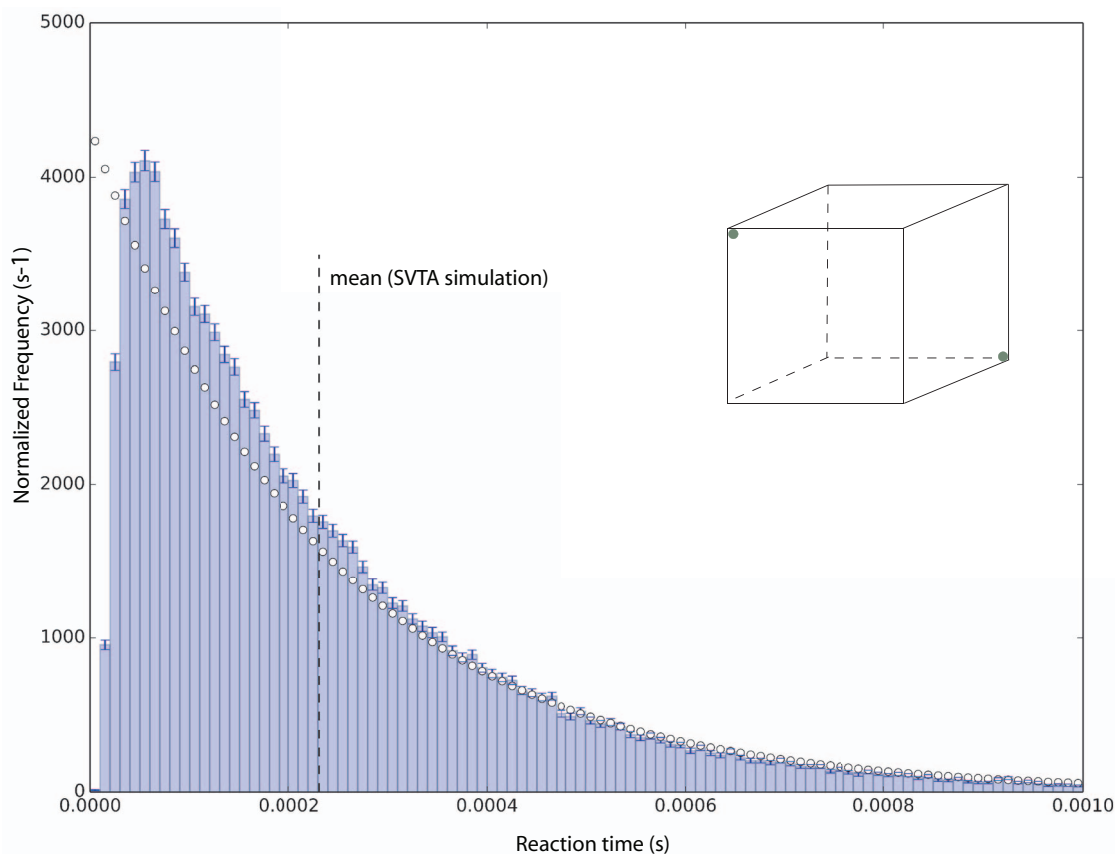


FIG. 9. A repeat of Fig. 8, but now with molecules 1 and 2 initially located in opposite corners of  $\Omega$ , as illustrated in the inset. Again the time to reaction is not exponentially distributed. The implication of Figs. 8 and 9 is that the PDF of the time to reaction for a delta-function initial condition will generally be non-exponential, and moreover will be different for different initial positions of the two molecules. As a consequence, the notion of a “reaction rate” loses practical utility. But this poses no problem for the SVTA.

*post hoc* analysis, and (ii) there will be a *different* reaction rate function for every different initial condition, seem rather to suggest that the concept of “reaction rate” is simply *not useful* when the initial positions of the molecules are sharply specified. Although that might be a problem if one were intent on describing these reactions with a differential equation, it is not a problem for the SVTA, nor for eGFRD either.

The foregoing results naturally make one wonder how the SVTA connects with situations in which reaction rates *do* exist. Those situations are described by the *bimolecular propensity function*. As discussed in Sec. II C and Appendix A, for a system containing two reactant molecules the bimolecular propensity function is the constant  $k|\Omega|^{-1}$ , where  $k$  is given by Eq. (7). By definition,  $k|\Omega|^{-1}dt$  is the probability that molecules 1 and 2 will react with each other in the next infinitesimal time  $dt$ . That implies that the time to reaction will be the exponential random variable with decay constant  $k|\Omega|$  and mean  $|\Omega|/k$ . However, all this requires the system to be *dilute and well-mixed*. Satisfying those two conditions turns out to be more complicated than one might imagine.

*Diluteness* in this case requires  $\sigma_{12}$  to be small enough, or  $\Omega$  to be large enough, that the volume occluded by molecules 1 and 2 is negligibly small compare to  $|\Omega|$ . That means that diluteness is a *limiting ideal* that cannot be attained by any finite system; any finite system can be, at best, only *approx-*

*imately* dilute. However, there is one aspect of the small occluded volume requirement that we can fold into the theory, at least approximately, rather easily: An inspection of the derivation of the bimolecular propensity function<sup>10</sup> reveals that  $|\Omega|$  in the formula  $k/|\Omega|$  for the bimolecular propensity function is actually supposed to be the volume that is *accessible to the center of molecule 1*. For a truly dilute system in which  $\sigma_{12}/H \approx 0$ , that volume would be, to a satisfactory approximation,  $H^3$ . But if  $\sigma_{12}/H$  is not infinitesimally small, we should recognize that the center of molecule 2 actually has access only to a cube  $\tilde{\Omega}$  whose walls are set in from those of  $\Omega$  by  $\sigma_{12}/2$ , and whose edge length is therefore  $H - \sigma_{12}$ . Furthermore, for most (but not all) of the time, the action sphere around molecule 2 will occlude an additional volume  $\frac{4}{3}\pi\sigma_{12}^3$ . It should therefore be more accurate to take the bimolecular propensity function to be

$$\frac{k}{|\Omega|} \approx \frac{k}{(H - \sigma_{12})^3 - \frac{4}{3}\pi\sigma_{12}^3}. \quad (38)$$

This expression obviously reduces to  $k/H^3$  in the dilute limit  $\sigma_{12}/H \rightarrow 0$ . In the system depicted in the insets to Figs. 8 and 9, a system which at first sight might seem to be “reasonably dilute,” this correction to the bimolecular propensity function amounts to a surprisingly large 16% increase in the value of the bimolecular propensity function.



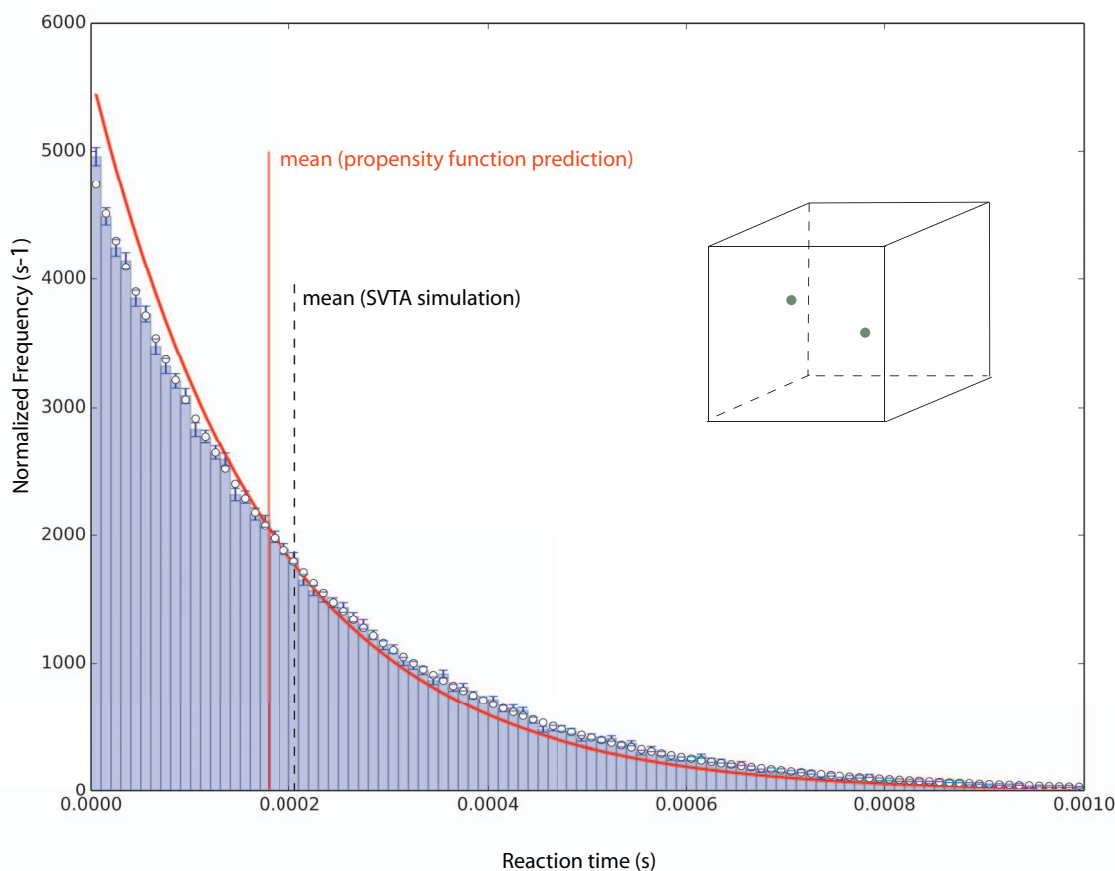


FIG. 10. As in Figs. 8 and 9, but now the initial positions of the molecules are *not* as shown in the inset figure; instead, the initial positions were generated from the probability distribution that is assumed in the derivation of the bimolecular propensity function (see text). The histogram of the generated reaction times now approximates very closely the exponential curve (open-circles) corresponding to the data mean (vertical dashed black line); there is only a small discrepancy in the very first bin. Therefore, to a very good approximation, the reaction times for this distribution of initial positions *can* be described by a reaction rate constant—namely, the reciprocal of the data mean. The solid vertical red line locates the mean reaction time predicted by the bimolecular propensity function in formula (38), which takes account of the fact that its derivation considers  $|\Omega|$  to be the volume that is actually accessible to the *center* of either molecule. The solid red curve is the exponential distribution corresponding to that mean. A perspective on the mismatch between the propensity function prediction and the SVTA data can be gained by comparing with the results in Fig. 11.

As for the propensity function’s *well-mixed* requirement, one might think that it could be satisfied simply by choosing the initial positions  $\mathbf{r}_1$  and  $\mathbf{r}_2$  of the centers of molecules 1 and 2 in a randomly uniform way inside the cube  $\tilde{\Omega}$ , discarding pairs for which  $|\mathbf{r}_1 - \mathbf{r}_2| < \sigma_{12}$ . However, doing that would *not* establish the initial distribution that is actually assumed in the physics derivation of the bimolecular propensity function and its constant  $k$  in Eq. (7). As is shown in Appendix A, that derivation assumes that the PDF of the initial vector  $\mathbf{r}_1 - \mathbf{r}_2 \equiv \mathbf{r}$  is the solution of the steady-state, isotropic diffusion equation

$$0 = \frac{d^2}{dr^2}(rP_1(r)) \quad (\sigma_{12} \leq r) \quad (39)$$

which satisfies the two boundary conditions

$$P_1(r \gg \sigma_{12}) = |\Omega|^{-1} \quad (40a)$$

and

$$P_1(r = \sigma_{12}) = \frac{4D_{12}}{4D_{12} + \sigma_{12}\bar{v}_{12}q_{12}} |\Omega|^{-1}. \quad (40b)$$

If molecules 1 and 2 were non-reacting, so that  $q_{12} = 0$ , then the boundary condition (40b) on the action sphere would be

the same as the far-field boundary condition (40a), and the solution to Eq. (39) would then be  $P_1(r) = |\Omega|^{-1}$  for all accessible  $r$ ; that is the distribution that would be generated by the uniform seeding procedure just described. But in the present case we have  $q_{12} = 0.2$ , and the prefactor in Eq. (40b) is 0.04 instead of 1. In Ref. 16, we derive a Monte Carlo procedure for generating the initial positions  $\mathbf{r}_1$  and  $\mathbf{r}_2$  of the centers of molecules 1 and 2 in such a way that  $r$  will distributed according to the solution of Eq. (39) for the boundary conditions (40a) and (40b) for *any* value of  $q_{12}$  in  $[0,1]$ .

Using this special random seeding procedure to generate the initial positions of molecules 1 and 2, we again made  $10^5$  runs of the system in Figs. 8 and 9. The histogram of the reaction times thus obtained is shown in Fig. 10. Again the vertical dashed black line locates the mean of the data, and the open-circle curve is the exponential distribution corresponding to that mean. We see that except for a very slight excess of events in the first histogram bin, the reaction times *can* now be considered exponentially distributed. So now we have a reaction rate constant, namely, the reciprocal of the data average. The red vertical line in Fig. 10 locates the average reaction time corresponding to the “corrected” bimolecular propensity function (38)—i.e., the reciprocal of the

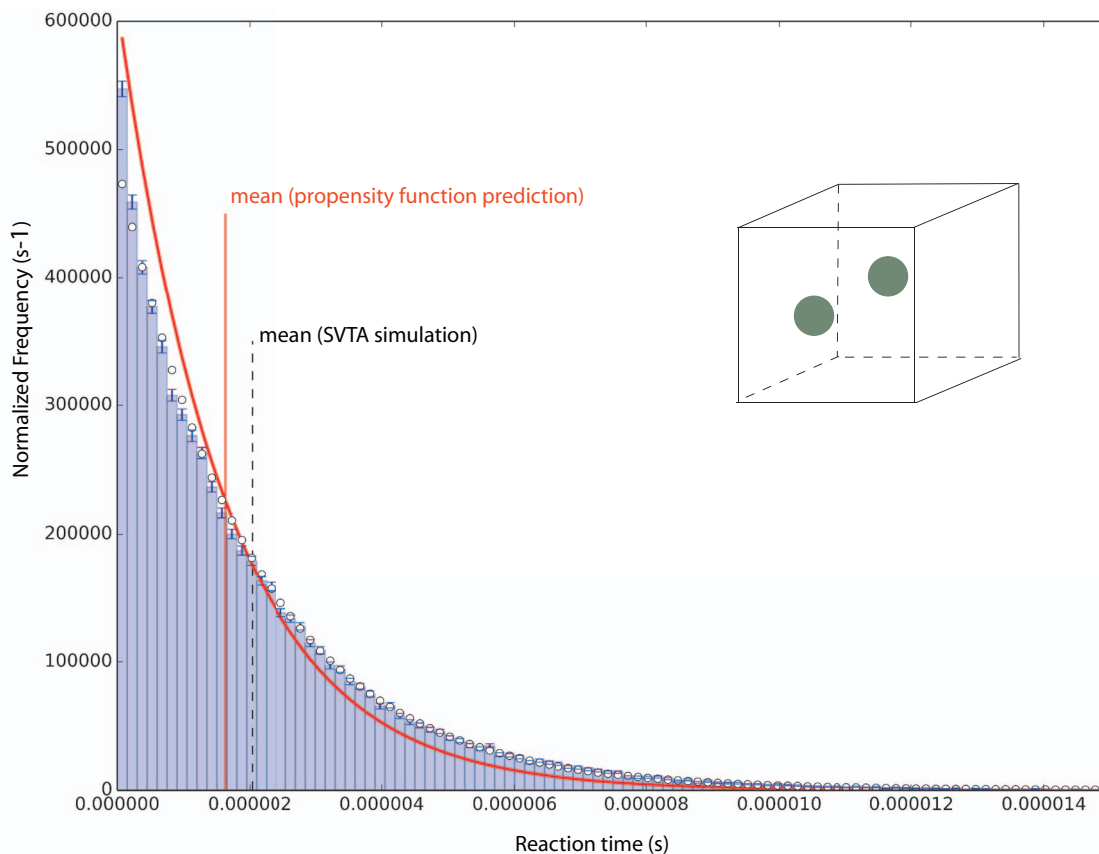


FIG. 11. A re-do of the run in Fig. 10 after reducing the edge length  $H$  of  $\Omega$  by a factor of *one-fourth*, and making a corresponding reduction in  $h$ . Again, the inset figure illustrates only the relative sizes of the molecules and  $\Omega$ ; the initial positions of the molecules were generated from the probability distribution that is assumed in the derivation of the bimolecular propensity function. Notice that the time scale on the horizontal axis here is nearly two orders of magnitude smaller than in Fig. 10. In contemplating the differences between the SVTA results and the bimolecular propensity function predictions in Figs. 10 and 11, several points should be kept in mind: (i) the SVTA never takes notice of the bimolecular propensity function's rate constant  $k$  in Eq. (7), which determines the locations of the red lines in these figures; (ii) neither of these systems is *perfectly* dilute, as is assumed by the bimolecular propensity function; and (iii) the predictions of the bimolecular propensity function may not be exact, because its derivation (see Appendix A) makes approximating assumptions. What these two figures do show is that the differences between the SVTA results and the predictions of the bimolecular propensity function are *smaller* in the *more dilute* system of Fig. 10 than in the *less dilute* system of Fig. 11, as theory predicts.

right side of (38)—and the red curve shows the exponential distribution corresponding to that average. There is evidently a noticeable difference between the SVTA simulation results and the predictions of the (corrected) bimolecular propensity function theory: the mean reaction time of the data is 15% larger than the mean reaction time predicted by the theory underlying the bimolecular propensity function.

Any attempt to understand the apparent discrepancy in Fig. 10 must begin by noting that, if we had *not* made the correction to the propensity function formula in (38), which reduced the mean predicted by the bimolecular propensity function by 16%, we would be looking here at only a 1% discrepancy, and we might be feeling quite satisfied. However, the logic motivating the correction in (38) is very compelling; moreover, the results obtained in an additional simulation suggest that correction is in fact warranted. That additional simulation is of the same two molecules as in the preceding three simulations, but now enclosed in a *smaller cube* whose edge length,  $H = 32.177 \times 10^{-9}$  m, is *one-fourth* of the value in Eq. (36). The relative sizes of the two molecules and their new  $\Omega$  are illustrated to scale in the inset to Fig. 11. In simulating this system, we reduced the voxel

edge length to  $h = 0.1297 \times 10^{-9}$  m in order to maintain the condition that there will be exactly 201 voxels accessible to  $VC_1$  and  $VC_2$  along each axis of  $\Omega$ . With the initial locations of the two molecules again chosen according to the probability distribution assumed in deriving the bimolecular propensity function (see Ref. 16), we made another series of  $10^5$  SVTA simulations. The histogram of the reaction times thus obtained is shown in Fig. 11.

A careful comparison of Figs. 10 and 11 reveals two significant facts. First, exponential distribution that corresponds to the mean of the data (the open-circle curve) fits the histogram of the data *better* in Fig. 10 than in Fig. 11; this can be seen by comparing the fits in the first two histogram bins of each figure. Second, the fractional difference between the mean of the data (vertical dashed black line) and the mean predicted by the bimolecular propensity function theory (vertical red line) is *smaller* in Fig. 10 (15%) than in Fig. 11 (25%). Qualitatively, these results are just what we expect: neither system is *perfectly* dilute, as is required by the derivation of the bimolecular propensity function, but the system that is *more dilute* (Fig. 10) agrees with the prediction of the bimolecular propensity function *better* than the *less*

*dilute* system (Fig. 11). On top of all that is the fact that we really have no warrant for regarding the theory underlying the bimolecular propensity function as being *exact*. That is because the derivation of formula (7) for  $k$  (see Appendix A) made some *assumptions*: it assumed a steady-state, isotropic initial condition, which provided us with Eq. (39); and it assumed a large reservoir at infinity, in order to sustain the outer boundary condition (40a) in the presence of the lower inner boundary condition (40b). Although those assumptions are physically reasonable for a system that is dilute and well-mixed, they ultimately amount to *approximations*. Therefore, their consequences cannot be presumed to be exact. For example, formula (7) for the bimolecular propensity function is completely independent of the *shape* of  $\Omega$ , yet it is easy to imagine extreme shapes that would invalidate that formula. So the red lines in Figs. 10 and 11 cannot confidently be regarded as exact benchmarks. Indeed, it might turn out that the approximations made in deriving the SVTA are more accurate than the approximations that were made in deriving the bimolecular propensity function, in which case Figs. 10 and 11 would be revealing errors in the propensity function theory. Finally, in further defense of the excluded volume correction (38), the highly plausible similarity of Figs. 10 and 11 would be spoiled without that correction.

With all this in mind, and also taking into account that (i) the time scales on the horizontal axes in Figs. 10 and 11 differ by nearly two orders of magnitude, and (ii) the SVTA never takes notice of the propensity function's formula (7) for  $k$  or its formula (38), it could be argued that the level of agreement between the SVTA results and the bimolecular propensity function theory in Figs. 10 and 11 is actually rather good. We conclude that the simulations in Figs. 10 and 11 show that the SVTA gives results that appear to be as compatible with the bimolecular propensity function as we are presently entitled to expect.

### VIII. THREE-MOLECULE EVENTS IN THE SVTA

Situations will often arise in which three solute molecules will be situated so equitably relative to each other that putting an absorbing surface  $\Sigma$  around two of them and a second absorbing surface  $\Sigma'$  around the third would seem arbitrary. While doing so anyway would certainly not be incorrect, it might be more efficient to put a single absorbing surface  $\Sigma$  around all three molecules. This would appear to be difficult to do using the current method for implementing eGFRD.<sup>4</sup> But the SVTA can treat the three-molecule case by making straightforward extensions of the two-molecule procedure described in Sec. V. Here we outline those extensions in just enough detail that the interested reader should be able to generalize the two-molecule procedure in Fig. 7 to the three-molecule case.

The voxel  $VC_j$  containing the center of molecule  $j$  for  $j = 1, 2, 3$ , and the voxelated corraling surfaces  $V\Sigma_j$  and  $V\partial\Omega_j$  for those centers, are all defined in the same way as in a two-molecule corral. The bounds on  $h$  are the obvious generalization of those in (20). There will now be six voxelated action spheres:  $VAS_{12}$  and  $VAS_{13}$  surround and hop with  $VC_1$ ;  $VAS_{21}$  and  $VAS_{23}$  surround and hop with  $VC_2$ ;

and  $VAS_{31}$  and  $VAS_{32}$  surround and hop with  $VC_3$ . They are defined in the same way as in the two-molecule case (see Appendix B). But as discussed in Sec. V, we will not need to explicitly track any of them. We need only construct one *virtual* voxelated action sphere for each pair—e.g.,  $vVAS_{12}$ ,  $vVAS_{23}$ , and  $vVAS_{31}$ —each of which will remain centered on voxel (0, 0, 0) throughout the simulation.

The tentative time of a unimolecular reaction will be [cf. Eqs. (21a) and (21b)]

$$t_{\text{uni}} = t_0 + \mathcal{E}(k_1 + k_2 + k_3). \quad (41)$$

The molecule that so reacts at that time will be molecule  $j$  with probability  $k_j/(k_1 + k_2 + k_3)$ ; thus, the index of the reacting molecule can be generated by an obvious extension of Eq. (22).

In order to determine the tentative times to the next hop and the next collision, we must first determine the status of the six voxels adjacent to each of  $VC_1$ ,  $VC_2$ , and  $VC_3$ . This is done in the same way as in the two-molecule case. In that way we learn the value of  $n_1^{\text{open}}$ , the number of voxels adjacent to  $VC_1$  which do *not* belong to  $V\partial\Omega_1$  or  $VAS_{21}$  or  $VAS_{31}$ , and similarly the values of  $n_2^{\text{open}}$  and  $n_3^{\text{open}}$ . We also learn the values of  $n_{jj'}^{\text{VAS}}$ , the number of voxels adjacent to  $VC_j$  that belong to  $VAS_{j'j}$  (which is the same as the number of voxels adjacent to  $VC_{j'}$  that belong to  $VAS_{jj'}$ ) for all  $j' \neq j$ .

The tentative time of the next hop will be [cf. Eqs. (23a) and (23b)]

$$t_{\text{hop}} = t + \mathcal{E}\left(\frac{n_1^{\text{open}}D_1 + n_2^{\text{open}}D_2 + n_3^{\text{open}}D_3}{h^2}\right), \quad (42)$$

where  $t$  is the current time. The molecule that hops at that time will be molecule  $j$  with probability  $n_j^{\text{open}}D_j/(n_1^{\text{open}}D_1 + n_2^{\text{open}}D_2 + n_3^{\text{open}}D_3)$ ; thus, the index of the hopping molecule can be generated by an obvious extension of Eq. (24).

As for bimolecular reactions, if none or only one of the  $q_{jj'}$ 's are  $>0$ , the simulation procedure will be an obvious extension of the two-molecule case. To illustrate what happens if more than one pair of molecules can react bimolecularly, let us consider the case in which pairs (1, 2) and (2, 3) can react, but the pair (3, 1) cannot:

$$q_{12} > 0, \quad q_{23} > 0, \quad q_{31} = 0. \quad (43a)$$

Before entering the main simulation loop, we will generate, for each pair that can react, the number of collisions that will produce a reaction of that pair [cf. Eqs. (32) and (33)]:

$$n_{12}^{\text{col}} = \text{Geo}(q_{12}), \quad n_{23}^{\text{col}} = \text{Geo}(q_{23}). \quad (43b)$$

Then in the main simulation loop, as part of the decision process to determine whether the next event will be a *unimolecular reaction* or a *hop* or a *collision between two reactable molecules*, we will generate a tentative time of the last as [cf. Eqs. (29) and (31)]

$$t_{\text{col}} = t + \mathcal{E}(\beta_{12} + \beta_{23}), \quad (43c)$$

where  $t$  is the current time and

$$\beta_{jj'} \equiv \frac{n_{jj'}^{\text{VAS}}\bar{v}_{jj'}}{4h}. \quad (43d)$$

Note that we do not care when or even if the non-reactable pair (3, 1) collides, because that collision has no consequence other than the blocked hopping destinations of molecules 1 and 3, and those we account for simply by obeying the blocks. If  $t_{\text{col}} \leq \min(t_{\text{uni}}, t_{\text{hop}})$ , we will select the pair that actually collides by generating  $u = \mathcal{U}(0, 1)$  and then choosing

$$\text{pair } (1, 2) \text{ if } u \leq \frac{\beta_{12}}{\beta_{12} + \beta_{23}}; \quad \text{otherwise pair } (2, 3). \quad (43e)$$

We will then execute the *collision* of the chosen pair  $(j, j')$  by reducing  $n_{jj'}^{\text{col}}$  by 1. If that reduction brings  $n_{jj'}^{\text{col}}$  to zero, we will immediately execute the *reaction* of that pair.

The three-molecule SVTA procedure can obviously be extended to more molecules, and indeed, at least in principle, to *all* the molecules in the system *without* any corralling surfaces. However, the increasing combinatorial complexity associated with checking to see which molecular pairs are imminently colliding, coupled with the concomitant loss of the computational parallelism that is provided by the corrals, will inevitably result in some optimal (and probably small) number of molecules to corral. This is a matter that requires future investigation.

## IX. ACCOMMODATING NON-COLLISIONAL FORCES BETWEEN REACTANT MOLECULES

Practically all reactant molecules in a cell are electrically neutral, in that they contain the same number of protons and electrons. But those positive and negative charges are distributed differently over the molecules, and that gives rise to electrical forces between the molecules. Such forces will generally be of much shorter range than the  $1/r^2$  force between two point charges. Examples of such forces in cell biology are the van der Waals interactions that hold together the nonpolar regions of fat molecules to form the lipid bilayer in cells, and the forces between proteins as they fold and bind with each other to form bigger structures (e.g., red blood cells).<sup>19,20</sup>

When a solute molecule  $i$  is subject to a force  $\mathbf{F}$  that does not arise from collisions with the surrounding solvent molecules, the motion of the molecule on sufficiently large spatial and temporal scales is described by the Smoluchowski equation:<sup>21</sup>

$$\frac{\partial p(\mathbf{r}_i, t)}{\partial t} = D_i \nabla_{\mathbf{r}_i}^2 p(\mathbf{r}_i, t) - \nabla_{\mathbf{r}_i} \cdot \left( \frac{D_i}{k_B T} \mathbf{F}(\mathbf{r}_i) p(\mathbf{r}_i, t) \right). \quad (44)$$

The Smoluchowski equation reduces to the standard diffusion equation (1) when  $\mathbf{F} \equiv 0$ . It turns out that the voxel-hopping procedure (5) for simulating the motion of a molecule according to the diffusion equation (1) can be extended to the Smoluchowski equation (44), once again in a way that becomes exact in the limit  $h \rightarrow 0$ , by a fairly simple modification of the hopping probability rate  $D_i/h^2$ .<sup>22</sup> This allows the SVTA to take account of forces of attraction or repulsion between reactant molecules. We will now describe how this is procedurally implemented.

For definiteness, let us suppose that two solute molecules 1 and 2 *attract* each other with a force whose magnitude  $F$  depends only on the distance between their centers, and

whose direction is along their line of centers. To adapt this situation to the voxelated space of the SVTA, we let  $\mathbf{r}_1$  and  $\mathbf{r}_2$  be the position vectors, relative to some fixed point  $O$ , of the *centers of the voxels*  $VC_1$  and  $VC_2$  that contain the respective molecules. Then with  $\mathbf{r}_{21} \equiv \mathbf{r}_2 - \mathbf{r}_1$ ,  $r_{21} \equiv |\mathbf{r}_{21}|$ , and  $\hat{\mathbf{r}}_{21} \equiv \mathbf{r}_{21}/r_{21}$ , molecule 1 feels a force  $F(r_{21})\hat{\mathbf{r}}_{21}$  due to molecule 2, and molecule 2 feels a force  $-F(r_{21})\hat{\mathbf{r}}_{21}$  due to molecule 1. With the point  $O$  as origin, we set up a Cartesian reference frame in the space of the voxel mesh with the mutually orthogonal unit vectors  $\hat{\mathbf{x}}$ ,  $\hat{\mathbf{y}}$ , and  $\hat{\mathbf{z}}$  aligned with the edges of the voxels. Then we have

$$\mathbf{r}_j = x_j \hat{\mathbf{x}} + y_j \hat{\mathbf{y}} + z_j \hat{\mathbf{z}} \quad (j = 1, 2) \quad (45a)$$

and

$$\begin{aligned} \mathbf{r}_{21} &= (x_2 - x_1)\hat{\mathbf{x}} + (y_2 - y_1)\hat{\mathbf{y}} + (z_2 - z_1)\hat{\mathbf{z}} \\ &\equiv x_{21}\hat{\mathbf{x}} + y_{21}\hat{\mathbf{y}} + z_{21}\hat{\mathbf{z}}. \end{aligned} \quad (45b)$$

We resolve the force exerted on molecule 1 by molecule 2 into its Cartesian components by invoking the definition  $\hat{\mathbf{r}}_{21} \equiv \mathbf{r}_{21}/r_{21}$  and Eq. (45b):

$$F(r_{21})\hat{\mathbf{r}}_{21} = F(r_{21})\frac{x_{21}}{r_{21}}\hat{\mathbf{x}} + F(r_{21})\frac{y_{21}}{r_{21}}\hat{\mathbf{y}} + F(r_{21})\frac{z_{21}}{r_{21}}\hat{\mathbf{z}}. \quad (46)$$

We call the voxel whose center is at  $\pm h\hat{\mathbf{x}}$  relative to the center of voxel  $VC_1$  “adjacent voxel  $x_{\pm}$ ,” and we similarly define adjacent voxels  $y_{\pm}$ , and  $z_{\pm}$ . Then it can be shown (see Ref. 16) that the jump probability rate for molecule 1 will be

$$\frac{D_1}{h^2} \left( 1 \pm h \frac{F(r_{21}) x_{21}}{2k_B T r_{21}} \right) \text{ for a hop to adjacent voxel } x_{\pm}, \quad (47)$$

and similarly for hops to adjacent voxels  $y_{\pm}$  and  $z_{\pm}$ . This result of course assumes that the adjacent voxel in question is accessible to molecule 1. It also assumes that  $h$  is “small,” and in particular small enough that all of the jump probability rates are non-negative.

If we label the six adjacent voxels to  $VC_2$  in the same way (i.e.,  $x_{\pm}$ ,  $y_{\pm}$ ,  $z_{\pm}$ ), then the jump probability rates for molecule 2 will be given by Eq. (47) *except* that the “ $\pm$ ” signs inside the large parentheses must be replaced by “ $\mp$ .” That same sign replacement must be made to *both* molecule jump probability rates if the force is repulsive instead of attractive.

## X. SUMMARY AND CONCLUSIONS

This paper has introduced the small-voxel tracking algorithm (SVTA) for simulating chemical reactions among diffusing hard-sphere molecules. At first sight, the SVTA might seem to be merely a modified version of the reaction-diffusion SSA (RD-SSA), since both use the voxel-hopping procedure (5) to move molecules around on a discrete spatial lattice in continuous time. But the differences between the two algorithms run much deeper than those similarities. The fact that the SVTA requires its voxels to be *smaller* than the reactant molecules while the RD-SSA requires its voxels to be *much larger* than the reactant molecules has major ramifications: First, it means that voxel-hopping in the SVTA simulates the diffusion equation more accurately than does voxel-hopping



in the RD-SSA, since (5) is exact only in the limit  $h \rightarrow 0$ . Second, it enables the SVTA to *track* individual reactant molecules (at least to resolution  $h$ ), which the RD-SSA does not do. Third, it enables the SVTA to prevent a molecule from hopping into a voxel that is occluded by another molecule, and to thereby model the effects of *molecular crowding*. The RD-SSA does not do that either, because it allows a molecule to hop into a voxel without checking to see whether the voxel actually has room for the molecule. Many of the limitations of the RD-SSA can be traced to the fact that it requires the system to be *dilute* in the reactant (solute) molecules. That requirement stems from the fact that non-diluteness invalidates the physics derivation of the bimolecular propensity function, which is central to the RD-SSA. But the SVTA does not use the bimolecular propensity function. And more importantly, the SVTA does not require the system to be dilute in the reactant molecules.

The real inspiration for the SVTA is eGFRD.<sup>2</sup> The SVTA and eGFRD have the same strategic aims, but very different implementation procedures. Nevertheless, we showed in Sec. VI that the SVTA and eGFRD are equivalent in that both find their theoretical justification in the same physical theory, namely, the theory used to derive the bimolecular propensity function (see Appendix A). This equivalence reflects yet another significant difference between the SVTA and the RD-SSA: In addition to using the probability (5) for hopping to an adjacent voxel, the SVTA also uses the probability (26) for impinging on a voxel face. More specifically, the SVTA supplements the *hopping* probability rule (5) with the *collision* probability rule (30); when the latter is combined with the collision-condition reaction probability  $q_{12}$ , the result is a modeling of a collision-induced reaction that is equivalent to what eGFRD obtains when it supplements the standard diffusion equation with the radiation boundary condition (34b).

Both the SVTA and eGFRD need additional research to determine (i) an efficient strategy for erecting the corrals, (ii) the optimal number of solute molecules to enclose in a corral, (iii) the best way to synchronize the corrals when a breach occurs, and (iv) how to place multiple product molecules of a reaction in relation to the prior location of the parent molecule(s). None of these important issues has been addressed in this paper. In addition, the SVTA needs additional research on how to optimally choose  $h$ .<sup>23</sup>

In closing, we list what we perceive to be the strengths and weaknesses of the SVTA.

#### *Weaknesses of the SVTA:*

- (a) It requires us to impose a voxel mesh on the region enclosed by the corraling surface  $\Sigma$ , and to identify the voxels belonging to  $V\Sigma_i$  and  $V\partial\Omega_i$  for each corralled molecule. In multi-molecule scenarios, we also have to stay continually aware of all abutments between the voxel containing a molecule's center and the associated voxelated action spheres around the other molecules.
- (b) Since  $h$  is "small" there will be many voxels inside  $\Sigma$ , so many hops will usually have to be executed before something interesting happens. And the more dilute the system is, the more acute this problem will be.
- (c) We must determine an optimal value for  $h$ ; i.e., we must find the best compromise between the greater accuracy that comes with smaller values of  $h$ , subject to condition (20), and the faster simulation speed that comes with larger values of  $h$ .

#### *Strengths of the SVTA:*

- (a) Like eGFRD, it has a carefully considered basis in molecular physics.
- (b) It works for non-dilute systems.
- (c) It easily models the effects of molecular crowding, simply by not allowing a molecule to hop into a voxel that is barred by the presence of another molecule or the system boundary.
- (d) It does not require solving any partial differential equations.
- (e) It allows arbitrary shapes and positions for the corraling surface  $\Sigma$ .
- (f) It is straightforwardly extendable to a three-molecule corral.
- (g) It can accommodate non-collisional forces between reactant molecules.
- (h) Once a reaction or a corral breach occurs, we immediately know what happened, when it happened, and where the molecules were.
- (i) By logging the times and destinations of the hops, we can straightforwardly backtrack if an outside molecule breaches  $\Sigma$  before the computed halting time.
- (j) Specific areas on the boundary  $\partial\Omega$ , which we have thus far assumed to be reflecting, could easily be changed to reaction sites or absorption sites.
- (k) Active transport of reactant molecules along physically confined pathways, which is often very important in cellular systems, could be modeled as non-diffusional hops along connected voxel strings.

Finally, we note that the fact that the SVTA appears to be more efficient for less dilute systems, while eGFRD appears to be more efficient for more dilute systems, suggests that it would be fruitful to explore the possibility of an eGFRD-SVTA hybrid algorithm.

## ACKNOWLEDGMENTS

The authors thank Linda Petzold and her group at UCSB, especially Brian Drawert, Andreas Hellander, and Stefan Hellander, for several enlightening and helpful discussions. D.G. was funded by the University of California, Santa Barbara under Professional Services Agreement No. 130401A40, pursuant to NIH under Award No. R01-EB014877-01. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of these agencies.

## APPENDIX A: OUTLINE OF THE DERIVATION OF EQ. (7)

The derivation of Eq. (7) as given in Ref. 9 proceeds in two phases. The first phase is based solely on the Einstein dif-

fusion equation (1). Working in the rest frame of molecule 2, we take the center of that molecule to be the origin of coordinates for the position vector  $\mathbf{R}(t) \equiv \mathbf{R}_1(t) - \mathbf{R}_2(t)$  of the center of molecule 1 relative to the center of molecule 2. It can be shown that the conditional PDF  $p_1(\mathbf{r}, t | \mathbf{r}_0, t_0)$  of  $\mathbf{R}(t)$  obeys Eq. (1) with  $D$  replaced by  $D_{12}$ . In that equation, the hard sphere assumption restricts the variable  $\mathbf{r}$  to the region outside the “action sphere” of radius  $\sigma_{12}$  about the origin. If the center of molecule 1 lies anywhere *on* the action sphere, then the two molecules will be *colliding* with each other. In contrast, a *reaction* between the two molecules is mathematically represented as the *absorption* of the center of molecule 1 on the action sphere. From the fact that the right side of Eq. (1) can be written as minus the divergence of  $-D_{12}\nabla_{\mathbf{r}}p_1(\mathbf{r}, t | \mathbf{r}_0, t_0)$ , it follows that the latter is the probability flux for the position of the center of molecule 1 relative to molecule 2. That implies in particular that, if  $da$  is an infinitesimal area element on the action sphere at point  $\mathbf{r}' = \sigma_{12}\hat{\mathbf{r}}'$ , where the circumflex denotes a unit vector, then

$$\begin{aligned} D_{12}\nabla_{\mathbf{r}}p_1(\mathbf{r}', t | \mathbf{r}_0, t_0) \cdot \hat{\mathbf{r}}' da \cdot dt \\ = \text{the probability that molecule 1 will be} \\ \text{absorbed in } [t, t + dt) \text{ by the area element} \\ \text{da on the action sphere at point } \mathbf{r}'. \end{aligned} \quad (\text{A1})$$

Therefore, the probability that molecules 1 and 2 will *react* in the next  $dt$ , which by definition is  $k|\Omega|^{-1} \cdot dt$ , can be computed as the sum of the absorption probabilities (A1) over all the infinitesimal area elements  $da$  on the action sphere:

$$k|\Omega|^{-1} \cdot dt = \left( \oint_{\text{act sph}} D_{12}\nabla_{\mathbf{r}}p_1(\mathbf{r}', t | \mathbf{r}_0, t_0) \cdot \hat{\mathbf{r}}' da \right) \cdot dt. \quad (\text{A2})$$

To evaluate Eq. (A2), we must first solve Eq. (1) for  $p_1$  subject to *physically appropriate* initial and boundary conditions. The physically appropriate initial condition for  $p_1$  is taken to be neither the function  $\delta(\mathbf{r} - \mathbf{r}_0)$  that is suggested by the conditioning on  $p_1$  in Eq. (A2), nor the uniform distribution inside  $\Omega$ , but rather the *steady state solution* of the diffusion equation for *spherically symmetric* boundary conditions. The function  $p_1$  therefore does not vary with time, and it depends only on the magnitude  $r$  of  $\mathbf{r}$ . The physically appropriate “outer” boundary condition is taken to be  $p_1 = 1/|\Omega|$  everywhere on and outside a sphere centered on the origin whose radius  $\rho$  satisfies  $\sigma_{12} \ll \rho \ll |\Omega|^{1/3}$ . The physically appropriate “inner” boundary condition is taken to be some as yet unspecified value  $p_1(\sigma_{12})$ . When the stationary solution to Eq. (1) which satisfies all these conditions is computed and then substituted into Eq. (A2), the integration is straightforwardly performed and yields

$$k|\Omega|^{-1} \cdot dt = 4\pi D_{12}\sigma_{12}(|\Omega|^{-1} - p_1(\sigma_{12})) \cdot dt. \quad (\text{A3})$$

The result (A3) is a formula for  $k$  that is incomplete since it involves the as yet unknown value of  $p_1(\sigma_{12})$ . The fact that the Einstein model of diffusion is unable to specify the value of  $p_1(\sigma_{12})$  is a consequence of the fact that the *collision* between the two molecules which initiates the reaction takes

place on a vanishingly small length scale where Eq. (1) is not physically valid. The resolution of this difficulty is to recognize that on very small length scales, a solute molecule moves *ballistically*, as in an ideal gas. That brings us to the second phase of the derivation.

A well known result in the kinetic theory of the ideal gas is that if  $\rho$  is the average number of gas molecules per unit volume and  $\bar{v}$  is the average speed of those molecules, then the average number of collisions of the molecules with the boundary of  $\Omega$ , per unit area and per unit time, is  $\frac{1}{4}\rho\bar{v}$ .<sup>24</sup> A stochastic reformulation of the derivation of this result focusing on the small-scale motion of a *single solute* molecule establishes the following general result: If  $p_0$  is the value of the PDF of the solute molecule in the immediate neighborhood of an infinitesimally small stationary area element  $da$ , then<sup>25</sup>

$$\begin{aligned} \frac{1}{4}p_0\bar{v} \cdot da \cdot dt = \text{the probability that the solute molecule} \\ \text{will collide with } da \text{ in the next} \\ \text{infinitesimally small time } dt. \end{aligned} \quad (\text{A4})$$

Here,  $\bar{v}$  is the average speed of the solute molecule if it were in a dilute gas at the system temperature  $T$ . Applying (A4) to our present problem, where the diffusing molecule is molecule 1 and  $da$  is an infinitesimal area element on the action sphere at point  $\mathbf{r}'$ , and noting that multiplying a *collision* probability by the *collision-conditioned reaction* probability  $q_{12}$  gives us a *reaction* probability, we obtain

$$\begin{aligned} \frac{1}{4}p_1(\mathbf{r}', t | \mathbf{r}_0, t_0)\bar{v}_{12}q_{12} \cdot da \cdot dt \\ = \text{the probability that molecule 1 will react with} \\ \text{molecule 2 in } [t, t + dt) \text{ via a collision with the area} \\ \text{element } da \text{ on the action sphere at point } \mathbf{r}'. \end{aligned} \quad (\text{A5})$$

Since in our problem  $p_1(\mathbf{r}', t | \mathbf{r}_0, t_0) = p(\sigma_{12})$  everywhere on the action sphere, then summing (integrating) the probability (A5) over all area elements  $da$  on the action sphere just replaces  $da$  with  $4\pi\sigma_{12}^2$ . Thus we obtain the following formula for the probability that molecules 1 and 2 will *react* in the next  $dt$ :

$$k|\Omega|^{-1} \cdot dt = p_1(\sigma_{12})\pi\sigma_{12}^2\bar{v}_{12}q_{12} \cdot dt. \quad (\text{A6})$$

In Eqs. (A3) and (A6), we evidently have two formulas for the two unknowns  $k$  and  $p_1(\sigma_{12})$ . Formula (A3) was derived from the Einstein diffusion equation, and formula (A6) was derived from the equations for ballistic motion. Upon solving Eqs. (A6) and (A3) simultaneously, we obtain for  $k$  the formula in Eq. (7), and also the result

$$p_1(\sigma_{12}) = \frac{4D_{12}}{4D_{12} + \sigma_{12}\bar{v}_{12}q_{12}}|\Omega|^{-1}. \quad (\text{A7})$$

## APPENDIX B: THE VOXELATED ACTION SPHERE

We consider space to be subdivided into cubic voxels whose edges, of length  $h$ , are aligned with a Cartesian reference frame. We label each voxel with three signed integers  $i_x, i_y$ , and  $i_z$ , and we define the voxel  $(i_x, i_y, i_z)$  to be the voxel whose center is at the Cartesian point  $(x, y, z) = (i_x h, i_y h, i_z h)$ . Here we describe a computer-oriented procedure for identifying all the voxels belonging to the *virtual voxelated action sphere*  $vVAS_{21}$  for given values of the action sphere radius

$\sigma_{12}$  and  $h$ .  $vVAS_{21}$  is just  $VAS_{21}$ , except it is centered on voxel (0,0,0) instead of on voxel  $VC_2$ .

We begin by defining  $S_{\text{ball}}$  to be the set of all voxels whose centers are located on or inside a sphere of radius  $\sigma_{12}$  about the center of voxel (0,0,0). Then we define  $vVAS_{21}$  to be the set of all voxels belonging to  $S_{\text{ball}}$  that have at least one *adjacent* voxel that does *not* belong to  $S_{\text{ball}}$ ; i.e.,  $vVAS_{21}$  consists of all the exterior boundary voxels of  $S_{\text{ball}}$ .

The number of voxels belonging to the set  $S_{\text{ball}}$  that have their centers on the positive  $x$ -axis is

$$v \equiv \left\lceil \frac{\sigma_{12}}{h} \right\rceil, \quad (\text{B1})$$

where  $\lceil a \rceil$  denotes “the greatest integer in  $a$ .” It follows that the set of voxels

$$S_{\text{cube}} \equiv \{(i_x, i_y, i_z) \mid -v \leq i_x \leq v, -v \leq i_y \leq v, -v \leq i_z \leq v\} \quad (\text{B2})$$

comprises a cube which contains the voxel set  $S_{\text{ball}}$  as a subset. We now construct the voxel set  $vVAS_{21}$  by carrying out the following two steps:

*Step 1.* Select from the set  $S_{\text{cube}}$  all voxels whose centers are a distance  $\leq \sigma_{12}$  from the origin. These selected voxels comprise the set  $S_{\text{ball}}$ .

*Step 2.* Select from the set  $S_{\text{ball}}$  all voxels  $(i_x, i_y, i_z)$  for which at *least one* of its adjacent voxels  $(i_x \pm 1, i_y, i_z)$ ,  $(i_x, i_y \pm 1, i_z)$ ,  $(i_x, i_y, i_z \pm 1)$  does *not* belong to  $S_{\text{ball}}$ .

The final set of selected voxels is  $vVAS_{21}$ .

What the SVTA requires is the list of the indices  $(i_x, i_y, i_z)$  of the voxels that make up  $vVAS_{21}$ . It is fortunate that only the computer, and not the human modeler, needs to interact with that list, because the number of voxels that comprise  $vVAS_{21}$  increases rapidly with increasing  $\sigma_{12}/h$ , and visualizing the structure of  $vVAS_{21}$  is extremely difficult for all but the smallest values of  $\sigma_{12}/h$ . Two examples illustrate this point: In the simplest (and least accurate) case where  $h$  is taken to be as large as possible, so that  $\sigma_{12}/h$  is slightly larger than 1,  $vVAS_{21}$  consists of the 6 voxels that are adjacent to  $VC_2$ :  $(\pm 1, 0, 0)$ ,  $(0, \pm 1, 0)$ , and  $(0, 0, \pm 1)$ . In this obviously crude approximation to a sphere, there are a total of 18 voxels exteriorly adjacent to  $vVAS_{21}$ , of which 6 share one voxel face with  $vVAS_{21}$  and 12 share two. In the more accurate case in which  $\sigma_{12}/h = 10$ , which is the case illustrated in 2D in Figs. 5 and 6,  $vVAS_{21}$  consists of 978 voxels; those voxels have a total of 1118 exteriorly adjacent voxels, of which 558 share one face with  $vVAS_{21}$ , 336 share two faces, and 224 share three faces.

There is one artifact of the voxelated action sphere that deserves mentioning. As  $h$  is taken smaller and smaller, the voxelated action sphere provides an increasingly better approximation to the true action sphere as regards the latter’s *diameter*, but *not* as regards its *surface area*. Whereas the minimum and maximum diameters of  $vVAS_{21}$  both converge to  $2\sigma_{12}$  as  $h \rightarrow 0$ , the ratio of the exterior surface area of  $vVAS_{21}$  to  $4\pi\sigma_{12}^2$  fluctuates around 1.5, with no indication of approaching 1 as  $h \rightarrow 0$ . Is this a problem? We do not think so. We suspect that this increase in the surface area of a voxelated action sphere over the area of the true action sphere compensates for the restriction in the hopping rule (5) to six

mutually orthogonal hopping directions. And it seems clear, not only from theory<sup>6</sup> but also from the simulation results in Fig. 3, that the hopping rule (5) does simulate the diffusion equation more and more accurately as  $h \rightarrow 0$ .

<sup>1</sup>D. Gillespie, L. Petzold, and E. Seitaridou, *J. Chem. Phys.* **140**, 054111 (2014).

<sup>2</sup>By this definition of diluteness, a collection of molecules at any given concentration (number of molecules per unit volume) could be *dilute* if the molecules are sufficiently small, and *not dilute* if the molecules are sufficiently large. Diluteness therefore cannot be quantified solely by the concentration; account must also be taken of the sizes of the molecules. We are unaware of any widely accepted measure of diluteness, but candidates might be (a) the ratio of {the average separation distance between the reactant molecules} to {the average size of those molecules}, or (b) the ratio of  $|\Omega|$  to {the volume occluded by the reactant molecules}, or (c) the logarithm of either (a) or (b).

<sup>3</sup>This result was derived analytically in Sec. III of Ref. 1 where, unfortunately, the section title incorrectly asserted that diluteness is a necessary *and sufficient* condition for the system to stay well-mixed. In fact, the analysis there demonstrates only necessity, not sufficiency; i.e., the molecules must be dilute to stay well-mixed, but being dilute does not guarantee that they will *at all times* be well-mixed. For example, the close proximity of the two product molecules of a reaction immediately after the reaction occurs will temporarily violate the well-mixed condition, even if the system is “on average” well-mixed. That such a temporary small-scale violation of the well-mixed condition can cause qualitative changes in the behavior of a cellular system, even if the system is dilute, has been demonstrated in the paper by Takahashi *et al.* which is cited in Ref. 4.

<sup>4</sup>K. Takahashi, S. Tănase-Nicola and P. R. ten Wolde, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2473 (2010); see especially the supplementary material for that paper. The enhancement in eGFRD consists of major modifications prompted by Ref. 5 that resolve several shortcomings of the original GFRD of J. S. van Zon and P. R. ten Wolde, *J. Chem. Phys.* **123**, 234910 (2005).

<sup>5</sup>T. Opplestrup, W. Bulatov, G. Gilmer, M. Kalos, and B. Sadigh, *Phys. Rev. Lett.* **97**, 230602 (2006).

<sup>6</sup>D. Gillespie and E. Seitaridou, *Simple Brownian Diffusion: An Introduction to the Standard Theories* (Oxford University Press, 2012). A downloadable Errata for this book, which features a heavily revised Section 5.6, is available at <http://ukcatalogue.oup.com/product/9780199664504.do#>.

<sup>7</sup>The transition of a solute molecule’s motion from diffusional to ballistic as the spatiotemporal scale is decreased below the values in Eqs. (2) and (3) is clear in Langevin’s theory of molecular diffusion, which is more comprehensive than Einstein’s theory. The transition is especially vivid in the Langevin counterpart of the Einstein stepping formulas (4), in which not only the position but also the velocity of a diffusing molecule gets updated. Unencumbered by any restrictions on the size of  $\Delta t$ , the Langevin stepping formulas reduce smoothly to the Einstein formulas (4) for  $\Delta t \gg \tau$ , and to the equations of ballistic motion for  $\Delta t \ll \tau$ . For details, see Ref. 6, Chap. 9.

<sup>8</sup>A concise proof of the fact that the hopping rule (5) becomes exactly equivalent to the diffusion equation (1) in the limit  $h \rightarrow 0$  can be found in the Appendix of Ref. 1. A version of that proof which gives the more general form of the hopping rule when an external force is acting on the diffusing molecule can be found in Ref. 16.

<sup>9</sup>See Sections 11.7–11.9 of Ref. 6.

<sup>10</sup>The original derivation of Eq. (7) was given in D. Gillespie, *J. Chem. Phys.* **131**, 164109 (2009). An improved derivation, which is the one summarized in Appendix A, is given in Ref. 6, Sections 3.7 and 4.8.

<sup>11</sup>D. Gillespie, *J. Comput. Phys.* **22**, 403 (1976).

<sup>12</sup>S. Hellander, A. Hellander, and L. Petzold, *Phys. Rev. E* **85**, 042901 (2012).

<sup>13</sup>H. P. Erickson, *Biol. Proced. Online* **11**, 32 (2009).

<sup>14</sup>As noted in Ref. 13, a real protein molecule is not a smooth sphere, and it carries around with it a hydrated layer of water molecules. Those complicating features increase the drag coefficient  $\gamma_1$  over the value given by the Stokes formula, and that in turn is expected to produce, through the Einstein relation, a smaller value for  $D_1$ .

<sup>15</sup>If two “events”  $E_1$  and  $E_2$  can occur in the next infinitesimal time interval  $dt$  with respective probabilities  $a_1 dt$  and  $a_2 dt$ , then the time to and identity of the next occurring event can be computed in either of two ways. The *direct method* is to take the time to the next event as  $\tau = \mathcal{E}(a_1 + a_2)$  and the index of the next event as 1 if  $a_1 \geq (a_1 + a_2) \cdot \mathcal{U}(0, 1)$ , else 2. The *first*

reaction method is to compute two tentative reaction times  $\tau_1 = \mathcal{E}(a_1)$  and  $\tau_2 = \mathcal{E}(a_2)$ , and then take  $\tau = \min(\tau_1, \tau_2)$ , and the event as the one whose tentative time is smaller. That these two methods are equivalent is proved in Ref. 11, Section 5b. The SVTA employs both methods.

<sup>16</sup>See supplementary material at <http://dx.doi.org/10.1063/1.4903962> for three additional appendices to this article.

<sup>17</sup>If  $q_{12} \ll 1$ ,  $n_{\text{col}}$  in Eq. (33) might be so large that many successive  $VC_1-VAS_{21}$  adjacencies will usually occur before a bimolecular reaction occurs. In that case, the following approximate procedure should be faster than the exact procedure described in the text: At each  $VC_1-VAS_{21}$  adjacency, advance  $t$  by  $\tau' = \min(\tau_{\text{hop}}, \tau_{\text{uni}})$  and subtract from  $n_{\text{col}}$  a sample value of the Poisson random variable with mean  $\beta_{12}t_{\text{min}}$ , viewing that sample as the number of collisions that occur during the current  $VC_1-VAS_{21}$  adjacency. If that subtraction leaves  $n_{\text{col}}$  positive, continue the simulation by executing either the hop or the unimolecular reaction as is called for. But if the subtraction reduces  $n_{\text{col}}$  from a value  $n_b > 0$  to some value  $n_a \leq 0$ , then estimate the bimolecular reaction to have occurred at time  $t + [n_a / (n_b - n_a)] \cdot \tau'$ . The actual time of the bimolecular reaction will usually be some unknown time between  $t - \tau'$  and  $t$ , but the consequent relative error in the total time to the reaction will be small if many  $VC_1-VAS_{21}$  adjacencies have occurred. Note that  $\beta_{12}$  will in general take different values in successive  $VC_1-VAS_{21}$  adjacencies.

<sup>18</sup>D. Bicout, A. Berezhkovskii, and A. Szabo, *J. Chem. Phys.* **114**, 2293 (2001); see in particular Sec. II A.

<sup>19</sup>K. A. Dill and S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology* (Garland Science, 2003).

<sup>20</sup>R. Phillips, J. Kondev, J. Theriot, and H. Garcia, *Physical Biology of the Cell* (Garland Science, 2008).

<sup>21</sup>Two different derivations of the Smoluchowski equation can be found in Chapter 10 of Ref. 6.

<sup>22</sup>We do not know how widely known this generalized hopping result is, but we first learned of it from R. Erban, LMS/EPSC Short Instructional Course: Stochastic Modelling in Biological Systems (March 18–23, Oxford University, 2012), lecture #5. A derivation of Erban's result from the perspective of the present paper is given in Ref. 16.

<sup>23</sup>Future work on the SVTA should include a study of how best to choose the voxel size  $h$ . The usual expectation, that smaller values of  $h$  give more accurate results, is nicely illustrated by our one-molecule simulations in Fig. 3. But it is not always true in the SVTA, because of the fuzzy lower bounds on  $h$  in conditions (18) and (20). A test made during our two-molecule simulations for Fig. 10 indicated that decreasing  $h$  by 50% resulted in a 3% increase in the mean time to reaction, but no other change in the character of the histogram. Since that decrease in  $h$  left  $h$  21 times larger than the value on the left side of condition (20), its results probably were indeed 3% more accurate. But that decrease in  $h$  also caused the average run time to increase by 400%.

<sup>24</sup>R. Present, *Kinetic Theory of Gases* (McGraw-Hill, 1958).

<sup>25</sup>For details of the derivation of Eq. (A4), see Secs. 4.6 and 4.7 of Ref. 6.