Accurate particle-based reaction algorithms for fixed timestep simulators

Stuart T. Johnston, Christopher N. Angstmann, Satya N.V. Arjunan, Casper H.L. Beentjes, Adrien Coulier, Samuel A. Isaacson, Ash A. Khan, Karen Lipkow, and Steven S. Andrews*

Abstract Particle-based simulators are widely used to study biochemical systems involving spatial transport and chemical reactions on sub-cellular length scales. Fixed time step methods can often offer good performance even when simulating complex many-particle systems. However, current reaction algorithms approximate more detailed molecular dynamics models either inaccurately or slowly. Here,

Stuart T. Johnston
Department of Biomedical Engineering, University of Melbourne, Melbourne, Victoria, Australia e-mail: stuart.johnston@unimelb.edu.au

Christopher N. Angstmann
School of Mathematics and Statistics, University of New South Wales, Sydney, NSW, Australia e-mail: c.angstmann@unsw.edu.au

Satya N.V. Arjunan
RIKEN Center for Biosystems Dynamics Research, Osaka, Japan e-mail: satya.arjunan@gmail.com

Casper H.L. Beentjes
Mathematical Institute, University of Oxford, Oxford, UK e-mail: casper.beentjes@new.ox.ac.uk

Adrien Coulier
Department of Information Technology, Division of Scientific Computing, Uppsala University, Uppsala, Sweden e-mail: adrien.coulier@it.uu.se

Samuel A. Isaacson
Department of Mathematics and Statistics, Boston University, Boston, MA, USA e-mail: isaacsas@bu.edu

Ash A. Khan
RMIT University, Melbourne, Australia e-mail: ash.khan@rmit.edu.au

Karen Lipkow
Babraham Institute and Cambridge Systems Biology Centre, University of Cambridge, Cambridge, UK e-mail: karen.lipkow@babraham.ac.uk

Steven S. Andrews *Corresponding author
Division of Basic Sciences, Fred Hutchinson Cancer Research Center, and Department of Physics, Seattle University, Seattle, WA, USA e-mail: steven.s.andrews@gmail.com

we present new reaction algorithms that better approximate microscopic molecular dynamics models while maintaining good computational efficiency. A "Brownian bridge" algorithm samples reactions using reactant positions both before and after each diffusive step; its simulated dynamics exactly match those of appropriate underlying idealised models. Simpler but less accurate "RDF-matching" algorithms sample reactions by only using reactant positions after diffusive steps; they accurately reproduce the steady-state radial distribution function of the underlying idealised model. These algorithms can accurately approximate both commonly used reaction models and more realistic models that account for intermolecular potentials.

1 Introduction

Nearly all biochemical systems within cells, from metabolism to signalling, rely in part on chemical reactions between pairs of diffusing molecules. Algorithms for simulating these reactions are thus essential to biochemical modelling software. The approaches that have been developed over the past several decades work at varying levels of detail [16, 23]. Some treat molecular species as well-mixed chemical concentrations [2, 31], others capture spatial heterogeneity or the stochastic effects that arise from the statistics of single molecules [8, 10, 18, 20, 30], and yet others explicitly model the movement of every individual atom or molecule in response to applied forces [21]. In general, increasing the level of simulated detail yields more accurate results but at the expense of increasing the time needed to run simulations.

This raises the question of how one can mitigate the trade-off between simulation accuracy and computational efficiency. Hybrid simulation techniques address this by representing the simulated system at multiple levels of detail at once, using more detail for portions of the system that are of particular interest and less detail for portions that are of less interest [17, 25, 26]. Another approach focuses computational effort on simulation *times* that are of greatest interest, either by using adaptive time steps [3, 22, 32] or by selectively repeating simulations of interesting events [15]. Yet a third approach is to develop new reaction algorithms that offer improved accuracy but with minimal sacrifice of computational efficiency [9, 19, 29]. Here, we follow this last approach. We began developing these methods at the 2018 MATRIX workshop on "Spatio-temporal stochastic systems in biology", held in Creswick, Victoria, Australia.

This work focuses on particle-based simulation methods that use fixed simulation time steps [6, 27]. Particle-based methods represent every individual molecule of interest as either a point-like or spherical particle in three-dimensional continuous space. These molecules diffuse, may interact with membranes or other surfaces, and undergo chemical reactions when two reactant molecules collide with each other, often at some reaction rate. Representing simulated time with fixed length steps introduces approximations because it approximates physically continuous processes with discrete intervals. However, these methods can be reasonably efficient and, if

designed well, can always be made more accurate by reducing the time step size. The algorithms described here would be straightforward to implement in software, such as Smoldyn [7, 4, 5].

2 Assumptions and Definitions

Consider the generic irreversible bimolecular reaction A + B $\rightarrow$ C, where all three chemical species are composed of spherical molecules that undergo ideal diffusion and only interact with each other through this reaction. We ignore any potential volume exclusion interactions between pairs of A molecules or pairs of B molecules.

Assume that there are many more B molecules than A molecules so that we can work in the reference frame of a single A molecule. In this reference frame, an A molecule is at the origin and effectively stationary, surrounded by many diffusing B molecules. These B molecules diffuse with the sum of the physical A and B diffusion coefficients, which is called the mutual diffusion coefficient and denoted by D [24]. As in the classical Smoluchowski diffusion-limited reaction model and the more physically realistic Collins-Kimball model (described below), the A molecule is not removed from the system upon reaction, but acts as a permanent absorbing sink for B molecules [13, 24, 28]. In effect, we observe a single A molecule until it reacts, transfer the coordinates to some other A molecule and observe it until it reacts, and so on. We define the "binding radius", σ_b, as the centre-to-centre distance between A and B molecules where they start to interact with each other. It is easiest to assign this entire binding radius to the A molecule, effectively making it a sphere of radius σ_b and the B molecules simple points. The left panel of Figure 1 illustrates this model.

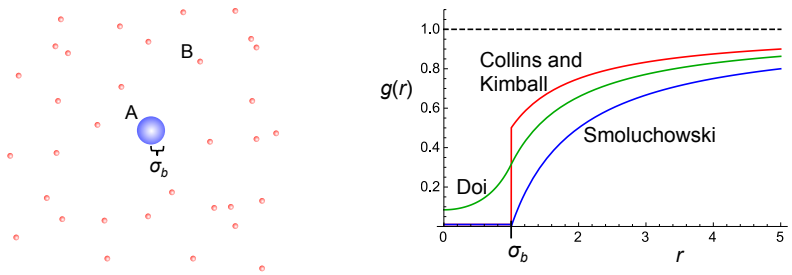

Fig. 1 (Left) Cartoon of the reaction model assumptions. (Right) Example steady-state radial distribution functions for the Smoluchowski, Collins and Kimball, and Doi reaction models.

The radial distribution function (RDF), $g(r,t)$, characterises the mean spatial distribution of B molecules about A molecules and is normalised so that $\lim_{r\rightarrow\infty} g(r,t) = 1$. Because we assumed ideal diffusion, the RDF evolves according to the diffusion equation; under the further assumption of three-dimensional rotational symmetry, it

is given by

$$\frac{\partial g(r,t)}{\partial t} = \frac{D}{r^2}\frac{\partial}{\partial r}\left[r^2\frac{\partial g(r,t)}{\partial r}\right]. \tag{1}$$

The new algorithms described here were designed to work with a variety of different reaction models so that the most appropriate one can be chosen for any particular system. Nevertheless, a few standard models are of particular interest. The *Smoluchowski model* assumes that reactants react immediately upon contact [24, 28], implying a Dirichlet boundary condition,

$$g(\sigma_b,t) = 0 \qquad \text{(Smoluchowski)} \tag{2}$$

(blue curve in the right panel of Figure 1). The *Collins and Kimball model* replaces the assumption of immediate reactions with the more realistic assumption that reactants can either react or reflect off of each other [13, 24]. This is modelled through the use of a radiation or Robin boundary condition,

$$\left.\frac{\partial g(r,t)}{\partial r}\right|_{r=\sigma_b^+} = \frac{g(\sigma_b,t)}{\gamma} \qquad \text{(Collins and Kimball)} \tag{3}$$

(red curve in right panel of Figure 1). Here, γ is a reactivity term that approaches 0 in the Smoluchowski model limit and infinity in the non-reactive limit. Finally, the *Doi model* allows reactants to physically overlap each other, meaning that their centre-to-centre separations are less than the binding radius, at which point they react with an "intrinsic reaction" rate constant, λ [1, 14]. Both the RDF and its first spatial derivative are continuous at σ_b, because there is no abrupt change of behaviour there, so the resulting RDF is smooth and increases monotonically with r (green curve in right panel of Figure 1). B molecules are only consumed at the binding radius for the first two models, so the net inward flux of their RDFs at the binding radius represents the chemical reaction rate coefficient,

$$k(t) = -4\pi r^2 D\left.\frac{\partial g(r,t)}{\partial r}\right|_{r=\sigma_b^+}. \tag{4}$$

This is the time-dependent chemical reaction rate per molecule of A and per unit concentration of B. This coefficient asymptotically approaches a constant value as the system evolves toward its steady state, where the constant value is then called the reaction rate constant and written as k. The Doi model is a little more complicated because B molecules can cross the binding radius without reacting, but equation 4 still approaches the steady-state reaction rate constant and is often a close approximation for transient behaviour.

Typically, a modeler would specify a reaction rate constant, the reaction model, and any necessary reaction model parameters, such as the sum of the physical molecule radii, the Collins and Kimball γ value, or the Doi model λ value. They would also specify the reactant diffusion coefficients, which combine to give D, and the simulation time step, Δt. From these, the simulation software needs to compute

simulation algorithm parameters, such as the binding radius and any reaction probabilities. When it then simulates the reaction system using those parameters, it should replicate the model results as closely as possible.

When comparing simulation and model results, it helps to think of time as progressing continuously in the simulation but with the algorithms only reporting their results as discrete snapshots at the end of each time step. We now discuss several new numerical methods that approximate these continuous-time models at fixed time steps reasonably well. They are called "exact" if their results are statistically identical with solutions of the appropriate continuous-time model.

3 Results

3.1 Brownian bridge method, two-step algorithm

The first new algorithm, which we call a Brownian bridge method, considers molecule positions both before and after diffusing. Working in the A molecule frame of reference, this algorithm is based on the idea that any of the continuous-time models can be solved exactly for any initial B molecule position, $\mathbf{r}_i$, to give both the probability of reaction during the time step, $P_{model}(\text{react}|\mathbf{r}_i)$ and the remaining probability density of the B molecule position at the end of the time step, $p_{model}(\mathbf{r}_f|\mathbf{r}_i)$. This probability density does not integrate to 1, but integrates to the probability that the molecule survives the time step; in other words, it integrates to $1-P_{model}(\text{react}|\mathbf{r}_i)$. Assume for now that both functions are known. The goal of the algorithm is to replicate this behaviour.

The algorithm involves two steps. The first is to diffuse all B molecules using random Gaussian-distributed displacements, as in current algorithms [9, 19]. Here, an individual B molecule diffuses from its initial location, $\mathbf{r}_i$, to a nearby point, $\mathbf{r}_f$, which is chosen from a Gaussian probability density, $p_{diff}(\mathbf{r}_f|\mathbf{r}_i)$. Assuming this density is everywhere larger than the model probability density, meaning $p_{diff}(\mathbf{r}_f|\mathbf{r}_i) \geq p_{model}(\mathbf{r}_f|\mathbf{r}_i)$, then the second step is that the simulator absorbs, meaning reacts, B molecules where there is excess probability density. B molecules that don't react are not moved. Quantitatively, the desired probability density, $p_{model}(\mathbf{r}_f|\mathbf{r}_i)$, is achieved if the simulator retains molecules with probability $p_{model}(\mathbf{r}_f|\mathbf{r}_i)/p_{diff}(\mathbf{r}_f|\mathbf{r}_i)$. This means that molecules should react with probability

$$P(\text{react}|\mathbf{r}_i,\mathbf{r}_f) = 1 - \frac{p_{model}(\mathbf{r}_f|\mathbf{r}_i)}{p_{diff}(\mathbf{r}_f|\mathbf{r}_i)}. \tag{5}$$

As an example, this Brownian bridge algorithm has been solved for a one-dimensional version of this system that has an absorbing boundary at $x = 0$ [9, 12]. Solving the diffusion equation for a molecule that starts at x_i (assuming $x_i > 0$) and that is subject to the absorbing boundary at $x = 0$ shows that its probability density after the end of one time step is equal to

$$p_{model}(x_f|x_i) = \left\{ \begin{array}{cc} 0 & x_f \leq 0 \\ G_s(x_f - x_i) - G_s(x_f + x_i) & x_f > 0 \end{array} \right\} \tag{6}$$

where $G_s(x)$ is a normalised Gaussian with mean 0 and standard deviation s, s is the rms step length which is equal to $s = \sqrt{2D\Delta t}$, and Δt is the time step length (Figure 2). Again, this probability density does not integrate to 1 due to the probability of the molecule being absorbed at the boundary. In the first step of the algorithm, the simulator diffuses the molecule from its initial location, x_i, using a Gaussian-distributed displacement. This causes its probability density after diffusion to be

$$p_{diff}(x_f|x_i) = G_s(x_f - x_i). \tag{7}$$

Comparing these two probability densities shows that $p_{diff}(x_f|x_i) \geq p_{model}(x_f|x_i)$ for all x_f values, so it is possible to recover the desired probability density by simply absorbing the molecule with the appropriate probability. Substituting these $p_{model}(x_f|x_i)$ and $p_{diff}(x_f|x_i)$ solutions into eq. 5 and simplifying gives

$$P(\text{react}|x_i, x_f) = \left\{ \begin{array}{cc} 1 & x \leq 0 \\ \exp\left(-\frac{2x_i x_f}{s^2}\right) & x > 0 \end{array} \right\}. \tag{8}$$

Thus, the algorithm for this one-dimensional case is that diffusion occurs with standard Gaussian-distributed displacements, going from x_i to x_f, and then absorption occurs with probability 1 if $x_f \leq 0$ or with probability $\exp(-\frac{2x_i x_f}{s^2})$ if $x_f > 0$. The results from this algorithm are statistically identical to those for the continuous-time model, making it an exact algorithm.

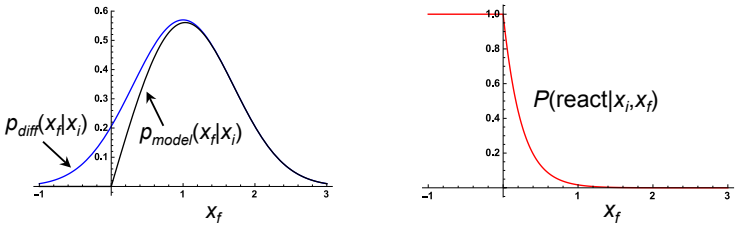

Fig. 2 Illustration of the Brownian bridge algorithm in 1 dimension for a Smoluchowski boundary condition at $x = 0$, for a molecule starting at $x_i = 1$ with an rms step length of $s = 0.7$. The left panel shows the model (black) and diffused (blue) probability densities; the region between the two represents the probability that molecules should react. The right panel shows the reaction probability as a function of position for the same parameters.

The formalism is the same when extending this algorithm to three dimensions. The numerator and denominator of eq. 5 are Green's functions with and without the model boundary conditions. The denominator is just a three-dimensional Gaussian. The numerator would be straightforward as well if radial symmetry could be assumed (the Green's function equations are in [9, 11]) but it cannot. To see this,

consider an A molecule at the origin and a B molecule that starts along the positive z-axis above the origin and ends at either the same location or equally far along the negative z-axis, below the origin. Both possibilities have the same initial and final distances from the origin, but the latter trajectory is much more likely to go close to the origin at some point along its path, and hence to undergo a reaction.

When accounting for the fact that the reference frame can be translated, rotated, and scaled as needed, the two vectors that are needed for computing the Brownian bridge absorption probability, $\mathbf{r}_i$ and $\mathbf{r}_f$, simplify to three values: the reduced initial separation between the two molecules, $\tilde{r}_i$, the reduced final separation, $\tilde{r}_f$, and the internal angle between the initial and final B molecule vectors, θ. The two reduced parameters were defined by dividing by the binding radius, so $\tilde{r}_i = |\mathbf{r}_i|/\sigma_b$ and $\tilde{r}_f = |\mathbf{r}_f|/\sigma_b$. Additional inputs to the probability calculation are constant over the course of the simulation; they are the reduced rms step length, $\tilde{s}$, which equals s/σ_b, and any model parameters, such as γ for the Collins and Kimball model and λ for the Doi model.

In the special case of the Smoluchowski model, the Green's function necessary for the numerator of eq. 5 has been solved (section 14.16 part III in ref. [11]). It is

$$p_{model}(\mathbf{r}_f|\mathbf{r}_i) = \frac{1}{4\pi\sqrt{\tilde{r}_i\tilde{r}_f}} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\theta) \int_0^{\infty} \frac{C_{n+\frac{1}{2}}(u\tilde{r}_i) C_{n+\frac{1}{2}}(u\tilde{r}_f)}{J^2_{n+\frac{1}{2}}(u) + Y^2_{n+\frac{1}{2}}(u)} e^{-Du^2 t} u\,du \tag{9}$$

where

$$C_{n+\frac{1}{2}}(z) = J_{n+\frac{1}{2}}(z) Y_{n+\frac{1}{2}}(u) - Y_{n+\frac{1}{2}}(z) J_{n+\frac{1}{2}}(u),$$

and P_n functions are Legendre polynomials, $J_{n+\frac{1}{2}}$ functions are Bessel functions, and $Y_{n+\frac{1}{2}}$ functions are spherical harmonics. Evaluating this Green's function during the simulation for every possible individual molecule interaction in a simulation would be impractical. A better approach would be to tabulate these results in a four dimensional table ($\tilde{r}_i$, $\tilde{r}_f$, θ, and $\tilde{s}$), enabling much quicker lookup. However, this would still incur significant computational costs, especially when a separate lookup, ideally with interpolation, would be needed for every possible molecule interaction at every time step.

The Doi model could also be simulated exactly with this Brownian bridge method. The sole change is that the $p_{model}(\mathbf{r}_f|\mathbf{r}_i)$ term in eq. 5 would need to be the Doi model Green's function. It is undoubtedly even more complicated than the one in eq. 9 but could be computed numerically. The resulting lookup table would be as computationally intensive as the one for the Smoluchowski model.

3.2 Brownian bridge method, three-step algorithm

The two-step Brownian bridge method would not work for the Collins and Kimball model because the requirement that $p_{model}(\mathbf{r}_f|\mathbf{r}_i) \leq p_{diff}(\mathbf{r}_f|\mathbf{r}_i)$ for all $\mathbf{r}_f$ values does not hold for some parameter values. To see this, consider the limit of low

reactivity ($\gamma \to \infty$), in which B molecules simply reflect off of A molecules and never react. In the model, the B molecule probability density would be relatively high just outside of the surface of the A molecules due to the reflecting boundary. However, the B molecule probability density would have a substantially lower value here after the algorithm's diffusion step because this step ignores the A molecule boundary, letting A and B molecules overlap instead. As a result, the model probability density could not be attained by simply removing particular B molecules.

A simple solution to this problem is to add a reflection step to the middle of the previous two-step algorithm structure to yield the following three-step algorithm:

1. Diffuse molecules with Gaussian distributed displacements with rms step length s.
2. Reflect A and B molecules off of each other to account for their excluded volumes. The previously described "overlap algorithm" and "reflection algorithm" would suffice, of which the latter agrees very closely with exact results [5].
3. Evaluate pairs of molecules to see whether they should react; if so, remove them from the simulation and replace them with products, as before. Simple generalisation of eq. 5 shows that the reaction probability should be

$$P(\text{react}|\mathbf{r}_i, \mathbf{r}_f) = 1 - \frac{p_{model}(\mathbf{r}_f|\mathbf{r}_i)}{p_{d,r}(\mathbf{r}_f|\mathbf{r}_i)} \tag{10}$$

   where $p_{d,r}(\mathbf{r}_f|\mathbf{r}_i)$ is the probability density that a molecule ends at $\mathbf{r}_f$ after being moved by both the algorithm's diffusion and reflection steps.

The Green's function equations in eq. 10 are complicated (see section 14.16 part IV of ref. [11] for the denominator) but, again, could be calculated numerically. The resulting reaction probability lookup table would have the same input parameters as before, $\tilde{r}_i$, $\tilde{r}_f$, θ, and $\tilde{s}$, plus any additional model-specific parameters, such as γ for the Collins and Kimball model.

This three-step version of the Brownian bridge algorithm would work for the Smoluchowski or Collins and Kimball models, but not for models that allow molecules to overlap, such as the Doi model. This is because $p_{d,r}(\mathbf{r}_f|\mathbf{r}_i) = 0$ for $|\mathbf{r}_f| < \sigma_b$, again violating the requirement that the simulated probability density must be at least as large as the model probability density.

3.3 RDF-matching, two-step

A different set of new algorithms, which we call RDF-matching methods, only considers molecule positions after diffusion (or after diffusion and reflection), $\mathbf{r}_f$. Ignoring the initial molecule positions, $\mathbf{r}_i$, necessarily makes these algorithms inexact, but they are still more accurate than existing algorithms; in particular, their simulated RDFs exactly match those of the models when at steady state. In essence, the RDF-matching algorithms react or reflect molecules as needed so that their simulated steady-state RDFs match the appropriate model RDFs.

Radial symmetry can be assumed now because only one B molecule position is used. Define $g_{model}(r)$ as the steady-state model RDF and $g_{diff}(r)$ as the RDF after one round of the algorithm's diffusion step after starting from the steady-state model RDF (we dropped the f subscript from r_f for simplicity because only final positions are considered here). This latter RDF is

$$g_{diff}(r) = \int_0^\infty 4\pi r'^2 \frac{1}{4\pi r r'} [G_s(r-r') - G_s(r+r')] g_{model}(r') dr' \tag{11}$$

where this uses the Green's function for radially symmetric diffusion in three dimensions. Assuming that $g_{diff}(r) \geq g_{model}(r)$ for all r, which is always true for the Smoluchowski and Doi models, then the system returns to $g_{model}(r)$ after one cycle if a molecule survives with probability $g_{model}(r)/g_{diff}(r)$. From this, the absorption probability for a molecule that diffuses to r is

$$P(\text{react}|r) = 1 - \frac{g_{model}(r)}{g_{diff}(r)}. \tag{12}$$

Note the similarity to eq. 5. Thus, the algorithm, which we call the two-step RDF-matching approach, is to first diffuse all molecules with Gaussian distributed displacements, and to then react pairs of molecules based on their separations using the probability given in eq. 12 (see Figure 3). This algorithm produces an RDF that exactly matches that of the model when at steady state. However, it's inexact when away from steady state; additionally, the fact that it ignores initial molecule positions makes it inexact when considering the reaction probabilities of individual molecules.

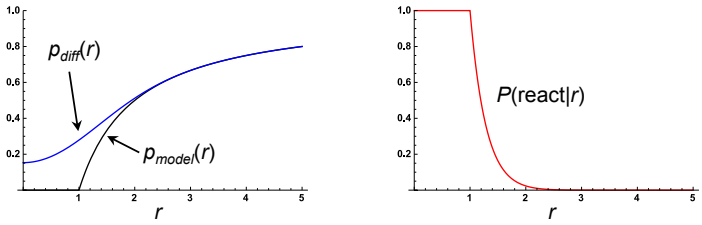

Fig. 3 Illustration of 2-step RDF matching algorithm for Smoluchowski boundary condition, for a binding radius of $\sigma_b = 1$ and an rms step length of $s = 0.7$. The left panel shows the model and simulated RDFs; the region between the two represents the probability that molecules should react. The right panel shows the reaction probability as a function of radius for the same parameters.

The probability function in eq. 12 can be computed explicitly in simple cases. For the Smoluchowski model, the steady-state model RDF is

$$g_{model}(r) = 1 - \frac{\sigma_b}{r}. \qquad \text{(Smoluchowski)} \tag{13}$$

Convolving this with the Green's function for radially symmetric 3-dimensional diffusion from eq. 11 leads to [9]

$$g_{diff}(r) = \frac{s^2}{r}\left[G_s(r-\sigma_b) - G_s(r+\sigma_b)\right] + \frac{1}{2}(e_+ + e_-) + \frac{\sigma_b}{2r}(e_+ - e_-) \tag{14}$$

where

$$e_\pm = \operatorname{erfc}\frac{\sigma_b \pm r}{s\sqrt{2}}$$

Combining these results with eq. 12 then gives the reaction probability as

$$P(\text{react}|r) = 1 - \frac{r-\sigma_b}{s^2\left[G_s(r-\sigma_b) - G_s(r+\sigma_b)\right] + \frac{1}{2}(e_+ + e_-) + \frac{\sigma_b}{2r}(e_+ - e_-)} \tag{15}$$

This equation is relatively simple to evaluate but still might be more efficient in a simulation if expressed in a lookup table (inputs are $\tilde{r}$ and $\tilde{s}$).

Other models yield more complicated reaction probabilities, but could again be expressed in lookup tables.

3.4 RDF-matching, three-step

The two-step algorithm does not work if $p_{diff}(r) < p_{model}(r)$ for some r value, which again can arise for the Collins and Kimball model. As before, a simple solution is to move to a three-step simulation algorithm with a middle reflection step. Extending the prior results gives the reaction probability for the three-step RDF matching algorithm as

$$P(\text{react}|r) = 1 - \frac{g_{model}(r)}{g_{d,r}(r)} \tag{16}$$

where $g_{d,r}(r)$ is the RDF in the simulation, after the model RDF undergoes both diffusion and reflection steps.

Although conceptually simple, this makes the equations much difficult to evaluate (in the Smoluchowski case, $g_{d,r}(r)$ can be found by integrating the product of $g_{model}(r)$ from eq. 13 and the Green's function for diffusion near a reflecting spherical boundary, which is given in eq. 16 of section 14.7 of ref. [11]). Again, numerical computation and lookup tables are practical solutions.

As in the two-step case, this three-step algorithm reproduces the model RDF when at steady state, but is inexact away from steady state and when considering individual molecules. In this case though, the incorrect reaction probabilities for individual molecules are worse than they might seem. During the diffusion step of the algorithm, suppose one B molecules diffused to the centre of the A molecule and another to just outside the edge of the A molecule. During the reflection step, the first B molecule would get reflected far away from the A molecule while the second wouldn't be moved at all. Because the first B molecule ends up far away, it

would have the lower reaction probability. However, this contradicts the fact that the continuous-time trajectory of the first B molecule probably spent much more time overlapping the A molecule, so it should have had the higher reaction probability. Thus, this algorithm causes the molecules that have the greatest overlap at the end of the diffusion step to have the lowest reaction probability, which is unphysical.

3.5 RDF-matching, two-step with remapping

A solution to this problem can be found by returning to the two-step RDF-matching method, but with a new remapping option. Now, the reaction step is that molecules react with probability $P(\text{react}|r)$, or get moved to a new position with probability $P(\text{move}|r)$ and relocation mapping $r \to r_{\text{out}}$, or stay put with probability $P(\text{stay}|r)$. Clearly,

$$P(\text{react}|r) + P(\text{move}|r) + P(\text{stay}|r) = 1 \tag{17}$$

Introducing this new option of moving molecules during the reaction step creates much more flexibility. We describe some options for this algorithm, focusing on models in which molecules are not permitted to overlap, such as the Collins and Kimball model.

The total amount, or mass, of the RDF after diffusion that needs to be reacted in order to return it to the steady-state RDF, $g_{model}(r)$, is

$$m_{react} = 4\pi \int_0^\infty [g_{diff}(r) - g_{model}(r)] r^2 \, dr \tag{18}$$

The next question is which portion of the RDF after diffusion should be reacted. One reasonable choice is to react the B molecules that have the greatest overlap with the A molecule. In particular, we can define an absorption radius, σ_a, for which all molecules inside are reacted and others are not. In other words, $P(\text{react}|r) = 1$ for $r < \sigma_a$ and $P(\text{react}|r) = 0$ for $r > \sigma_a$. The value of this absorption radius can be found by integrating the RDF after diffusion until the correct mass has been reached,

$$m_{react} = 4\pi \int_0^{\sigma_a} g_{diff}(r) r^2 \, dr, \tag{19}$$

An alternative choice would be to react B molecules with uniform probability, p_{react}, up to the binding radius and not beyond it, meaning that $P(\text{react}|r) = p_{react}$ for $r < \sigma_b$ and $P(\text{react}|r) = 0$ for $r > \sigma_b$. For this choice, p_{react} can be found from

$$m_{react} = 4\pi p_{react} \int_0^{\sigma_b} g_{diff}(r) r^2 \, dr. \tag{20}$$

With either choice, reactions then convert the RDF after diffusion, $g_{diff}(r)$, to the RDF after diffusion and reaction, $g_{d,rxn}(r)$,

$$g_{d,rxn}(r) = g_{diff}(r)[1 - P(\text{react}|r)] \tag{21}$$

Next, this modified RDF, $g_{d,rxn}(r)$, needs to be remapped to make it equal to the model RDF. Because the correct number of molecules have been reacted at this point, the mass of the RDF is conserved during this mapping step. Using the assumption that molecules are not allowed to overlap in the model, all RDF mass within the binding radius is excess and needs to be remapped to points outside of the binding radius. For convenience, also assume that the $g_{d,rxn}(r)$ function is smaller than $g_{model}(r)$ at all values that are outside of the binding radius (it is necessarily smaller on average, but things get complicated if it's not also smaller at all individual points). The mass of the RDF that needs to be remapped is

$$m_{map} = 4\pi \int_0^{\sigma_b} g_{d,rxn}(r) r^2 \, dr = 4\pi \int_{\sigma_b}^{\infty} [g_{model}(r) - g_{d,rxn}(r)] r^2 \, dr. \tag{22}$$

Again, there are multiple options, now for how to map these molecules from $[0, \sigma_b)$ to their new locations in (σ_b, ∞) in order to recover the steady-state profile. Simple choices are to maintain or invert radial ordering, where molecules near the origin are moved just outside σ_b in the former option and further out toward ∞ in the latter option. It is not intuitively clear which would be more accurate. For the first approach, in which radial ordering is maintained, consider the cumulative function for the molecules that need to be moved,

$$C_{in}(r_{in}) = 4\pi \int_0^{r_{in}} g_{d,rxn}(r') r'^2 \, dr', \tag{23}$$

This is defined on $0 < r_{in} < \sigma_b$ and returns a $C_{in}(r_{in})$ value that increases monotonically from 0 to the mass of molecules that need to undergo mapping, m_{map}. As this mass is conserved upon mapping, there is an equivalent value in the cumulative function for the locations where the molecules get mapped to,

$$C_{out}(r_{out}) = 4\pi \int_{\sigma_b}^{r_{out}} [g_{model}(r') - g_{d,rxn}(r')] r'^2 \, dr'. \tag{24}$$

This cumulative function is defined on $\sigma_b < r_{out} < \infty$ and also increases monotonically from 0 to the mass of molecules that need to undergo mapping, but now represents the spaces available for those molecules. The process of mapping a molecule from $r_{in} \in [0, \sigma_b)$ to its new location at $r_{out} \in (\sigma_b, \infty)$ is then:

1. Calculate $C_{in}(r_{in})$ for the initial location of the molecule from eq. 23.
2. Calculate the value of r_{out} such that $C_{out}(r_{out}) = C_{in}(r_{in})$ from eq. 24.
3. Move the molecule from r_{in} to r_{out}.

The second case, in which radial ordering is reversed, is identical but with the exception that the outer cumulative function runs in reverse order, being defined as

$$C_{out}(r_{out}) = 4\pi \int_{r_{out}}^{\infty} [g_{model}(r') - g_{d,rxn}(r')] r'^2 \, dr', \tag{25}$$

The same mapping process defined above works here as well.

3.6 Example of RDF-matching with remapping

Consider a 3D reaction process where the steady-state distribution follows the Collins and Kimball RDF. Using reduced variables for simplicity, it is

$$g_{model}(\tilde{r}) = \left\{ \begin{array}{cc} 0 & \tilde{r} < 1 \\ 1 - \frac{1}{\tilde{r}(1+\tilde{\gamma})} & 1 \leq \tilde{r} \end{array} \right\} \tag{26}$$

where $\tilde{\gamma}$ is the reduced boundary coefficient (equal to γ/σ_b). After the diffusion step, the distribution becomes, from eq. 11 and ref. [9]

$$g_{diff}(\tilde{r}) = \frac{\tilde{s}^2}{r}\left[G_{\tilde{s}}(\tilde{r}-1) - G_{\tilde{s}}(\tilde{r}+1)\right] + \frac{1}{2}(e_+ + e_-) + \frac{1}{2\tilde{r}(\tilde{\gamma}+1)}(e_+ - e_-) \tag{27}$$

The mass to be absorbed is given by substituting these two RDFs into eq. 18, yielding

$$m_{react} = \frac{2\pi\tilde{s}^2}{1+\tilde{\gamma}} \tag{28}$$

Next, we described two possibilities for the reaction step, of which one is to react all molecules up to some radius σ_a and the other was to react molecules up to the radius σ_b with probability p_{react}. The former is more difficult to solve analytically, so we consider the latter in this example. From eq. 20, the reaction probability is

$$p_{react} = \frac{m_{react}}{4\pi \int_0^1 g_{diff}(\tilde{r})\tilde{r}^2\, d\tilde{r}} \tag{29}$$

$$= \frac{6\tilde{s}^2}{\frac{2\tilde{s}\sqrt{2}}{\sqrt{\pi}}\{(\tilde{s}^2-1)(\tilde{\gamma}+1)e^{\frac{-2}{\tilde{s}^2}} + [\tilde{s}^2(\tilde{\gamma}+1) - 3\tilde{\gamma}]\} + [3\tilde{s}^2 - 4(\tilde{\gamma}+1)]\,\mathrm{erf}\frac{\sqrt{2}}{\tilde{s}} + 4(\tilde{\gamma}+1)} \tag{30}$$

Applying this to the RDF after diffusion yields

$$g_{d,rxn}(\tilde{r}) = \left\{ \begin{array}{cc} p_{react} g_{diff}(\tilde{r}) & \tilde{r} < 1 \\ g_{diff}(\tilde{r}) & 1 < \tilde{r} \end{array} \right\} \tag{31}$$

with substitutions from eqs. 27 and 30. This RDF is lengthy but can still be expressed in closed form. However, the next step is to compute the cumulative masses both inside and outside of the binding radii with eqs. 23 and 24, which can only be done numerically. Once those numerical integrals are computed, they are equated to each other and then solved for r_{out} as a function of r_{in}. This solution gives the required mapping.

4 Discussion

This work presents two new algorithms for simulating bimolecular chemical reactions with particle-based simulators that use fixed time steps. Both are more accurate than existing methods but do not incur substantial computational penalties.

In the Brownian bridge approach, the simulator considers both the initial and final separation vectors between potential reactants and computes the probability of a reaction occurring for those values. All simulated results exactly match those of the underlying model for isolated pairs of molecules, making it exact at this level of detail (interactions among 3 or more molecules are still approximate). A 2-step version of this algorithm, in which the algorithm only diffuses and reacts molecules, is sufficient for the Smoluchowski and Doi models, whereas a 3-step version, adding an intermediate reflection step, is necessary for the Collins and Kimball model. The primary disadvantage of the Brownian bridge method is that it requires looking up reaction probabilities for each reaction in lookup tables that have a minimum of three dimensions (initial separation, final separation, and interior angle) and often more dimensions. This may create an undesirable computational cost.

In the RDF-matching method, the simulator only considers final separations between potential reactants, while effectively assuming that the initial separations are randomly chosen from the steady-state distribution for the model. This algorithm enables the simulator to match the model radial distribution function exactly, but only when at steady state. Moreover, the precise dynamics of single molecules do not quite statistically match those of the underlying particle reaction-diffusion model. We again developed 2-step and 3-step algorithm versions. Additionally, we developed a remapping method that resamples the position of unreacted molecules after a diffusion step to correctly reproduce the steady-state radial distribution function. This may aid in reducing the error introduced in the algorithm by only considering molecular separations. The underlying reaction probabilities that need to be sampled can be computed from closed form equations in simple cases, or can be stored in a one dimensional lookup table.

One way in which the proposed algorithms could be optimised is by replacing the lookup tables with suitable approximating functions. For example, many functions can be closely approximated by rational functions or continued fractions, for which there are simple and efficient evaluation methods.

These new algorithms are not just two of many possible improvements on existing algorithms, but are particularly accurate methods for particle-based simulations that use fixed time steps. Simulating diffusion using Gaussian distributed molecule displacements is a sensible approach as it simulates diffusion exactly for non-interacting molecules in free space. If separating diffusion and reaction into separate steps, as is common, then sampling reaction probabilities with the Brownian bridge method is the unique solution that produces exact agreement with the underlying model. Also, the RDF-matching approach is the simplest option for producing exact agreement with the steady-state model RDF.

Absent from this work was any consideration of reversible reactions. Accounting for them would minimally affect the Brownian bridge algorithm because it doesn't

make any assumptions about molecule starting locations. On the other hand, they would affect the RDF-matching algorithm because reversibility changes the steady-state radial distribution functions.

An intriguing aspect of this work is that these algorithms can be set up to simulate realistic intermolecular potentials, such as a Lennard-Jones potential, nearly as easily as they can simulate the Smoluchowski or other simple models. Doing so could enable much more efficient simulation of these reaction dynamics than is currently possible.

Acknowledgements We thank Mark Flegg, Kevin Burrage, Ruth Baker, Samuel Isaacson, and Hans Othmer for organising the 2018 MATRIX workshop on "Spatio-temporal stochastic systems in biology", where we began this work. SNVA was partially supported by JSPS KAKENHI Challenging Research (Pioneering) Grant No. 18H05371. SAI was partially supported by National Science Foundation award DMS-1255408.

References

1. Agbanusi, I.C., Isaacson, S.A.: A comparison of bimolecular reaction models for stochastic reaction–diffusion systems. Bulletin of Mathematical Biology **76**(4), 922–946 (2014)
2. Aldridge, B.B., Burke, J.M., Lauffenburger, D.A., Sorger, P.K.: Physicochemical modelling of cell signalling pathways. Nature Cell Biology **8**(11), 1195 (2006)
3. Andrews, S.S.: Serial rebinding of ligands to clustered receptors as exemplified by bacterial chemotaxis. Phys. Biol. **2**, 111–122 (2005)
4. Andrews, S.S.: Spatial and stochastic cellular modeling with the Smoldyn simulator. In: Bacterial Molecular Networks, pp. 519–542. Springer (2012)
5. Andrews, S.S.: Smoldyn: particle-based simulation with rule-based modeling, improved molecular interaction and a library interface. Bioinformatics **33**(5), 710–717 (2017)
6. Andrews, S.S.: Particle-based stochastic simulators. Encyclopedia of Computational Neuroscience (2018)
7. Andrews, S.S., Addy, N.J., Brent, R., Arkin, A.P.: Detailed simulations of cell biology with Smoldyn 2.1. PLoS Comput. Biol. **6**, e1000,705 (2010)
8. Andrews, S.S., Arkin, A.P.: Simulating cell biology. Current Biology **16**(14), R523–R527 (2006)
9. Andrews, S.S., Bray, D.: Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. Physical Biology **1**(3), 137 (2004)
10. Andrews, S.S., Dinh, T., Arkin, A.P.: Stochastic models of biological processes. In: Encyclopedia of Complexity and Systems Science, pp. 8730–8749. Springer (2009)
11. Carslaw, H., Jaeger, J.: Conduction of Heat in Solids. Oxford University Press, Oxford, England (1959)
12. Clifford, P., Green, N.: On the simulation of the Smoluchowski boundary condition and the interpolation of Brownian paths. Molecular Physics **57**(1), 123–128 (1986)
13. Collins, F.C., Kimball, G.E.: Diffusion-controlled reaction rates. Journal of Colloid Science **4**(4), 425–437 (1949)
14. Doi, M.: Stochastic theory of diffusion-controlled reaction. Journal of Physics A: Mathematical and General **9**(9), 1479 (1976)
15. Donovan, R.M., Tapia, J.J., Sullivan, D.P., Faeder, J.R., Murphy, R.F., Dittrich, M., Zuckerman, D.M.: Unbiased rare event sampling in spatial stochastic systems biology models using a weighted ensemble of trajectories. PLOS Comput Biol **12**(2), e1004,611 (2016)

16. ElKalaawy, N., Amr, W.: Methodologies for the modeling and simulation of biochemical networks, illustrated for signal transduction pathways: A primer. Biosystems **129**, 1–18 (2015)
17. Erban, R.: From molecular dynamics to brownian dynamics. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **470**(2167), 20140,036 (2014)
18. Erban, R., Chapman, J., Maini, P.: A practical guide to stochastic simulations of reaction-diffusion processes. arXiv preprint arXiv:0704.1908 (2007)
19. Erban, R., Chapman, S.J.: Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions. Physical Biology **6**(4), 046,001 (2009)
20. Grima, R., Schnell, S.: Modelling reaction kinetics inside cells. Essays in Biochemistry **45**, 41–56 (2008)
21. Karplus, M., McCammon, J.A.: Molecular dynamics simulations of biomolecules. Nature Structural and Molecular Biology **9**(9), 646 (2002)
22. Kerr, R.A., Bartol, T.M., Kaminsky, B., Dittrich, M., Chang, J.C.J., Baden, S.B., Sejnowski, T.J., Stiles, J.R.: Fast Monte Carlo simulation methods for biological reaction-diffusion systems in solution and on surfaces. SIAM Journal on Scientific Computing **30**(6), 3126–3149 (2008)
23. Mogilner, A., Allard, J., Wollman, R.: Cell polarity: quantitative modeling as a tool in cell biology. Science **336**(6078), 175–179 (2012)
24. Rice, S.A.: Diffusion-Limited Reactions. Elsevier (1985)
25. Robinson, M., Andrews, S.S., Erban, R.: Multiscale reaction-diffusion simulations with Smoldyn. Bioinformatics **31**, 2406–2408 (2015)
26. Robinson, M., Flegg, M., Erban, R.: Adaptive two-regime method: application to front propagation. The Journal of Chemical Physics **140**(12), 124,109 (2014)
27. Schöneberg, J., Ullrich, A., Noé, F.: Simulation tools for particle-based reaction-diffusion dynamics in continuous space. BMC Biophysics **7**(1), 11 (2014)
28. Smoluchowski, M.v.: Versuch einer mathematischen theorie der koagulationskinetik kolloider lösungen. Z. Phys. Chem **92**(129-168), 9 (1917)
29. Tournier, A.L., Fitzjohn, P.W., Bates, P.A.: Probability-based model of protein-protein interactions on biological timescales. Algorithms for Molecular Biology **1**(1), 25 (2006)
30. Turner, T.E., Schnell, S., Burrage, K.: Stochastic approaches for modelling in vivo reactions. Computational Biology and Chemistry **28**(3), 165–178 (2004)
31. Tyson, J.J., Chen, K.C., Novak, B.: Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Current Opinion in Cell Biology **15**(2), 221–231 (2003)
32. van Zon, J.S., Ten Wolde, P.R.: Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. Physical Review Letters **94**(12), 128,103 (2005)