Mesoscopic-microscopic spatial stochastic simulation with automatic system partitioning

Stefan Hellander, Andreas Hellander, and Linda Petzold

Citation: The Journal of Chemical Physics **147**, 234101 (2017);
View online: https://doi.org/10.1063/1.5002773
View Table of Contents: http://aip.scitation.org/toc/jcp/147/23
Published by the American Institute of Physics

Articles you may be interested in

From properties to materials: An efficient and simple approach
The Journal of Chemical Physics **147**, 234105 (2017); 10.1063/1.5009548

Reliability assessment for large-scale molecular dynamics approximations
The Journal of Chemical Physics **147**, 234106 (2017); 10.1063/1.5009431

Parallel replica dynamics method for bistable stochastic reaction networks: Simulation and sensitivity analysis
The Journal of Chemical Physics **147**, 234110 (2017); 10.1063/1.5017955

Coupling discrete and continuum concentration particle models for multiscale and hybrid molecular-continuum simulations
The Journal of Chemical Physics **147**, 234112 (2017); 10.1063/1.5001703

Communication: Spectroscopic characterization of a strongly interacting $C_{(2)}H$ group on the $EMIM^+$ cation in the $(EMIM^+)_2X^-$ (X = BF_4, Cl, Br, and I) ternary building blocks of ionic liquids
The Journal of Chemical Physics **147**, 231101 (2017); 10.1063/1.5009009

Principal component analysis on a torus: Theory and application to protein dynamics
The Journal of Chemical Physics **147**, 244101 (2017); 10.1063/1.4998259

Mesoscopic-microscopic spatial stochastic simulation with automatic system partitioning

Stefan Hellander,[1] Andreas Hellander,[1] and Linda Petzold[2]
[1]*Department of Information Technology, Uppsala University, P.O.Box 337, SE-75105 Uppsala, Sweden*
[2]*Department of Computer Science, University of California, Santa Barbara, California 93106-5070, USA*

(Received 1 September 2017; accepted 27 November 2017; published online 15 December 2017)

The reaction-diffusion master equation (RDME) is a model that allows for efficient on-lattice simulation of spatially resolved stochastic chemical kinetics. Compared to off-lattice hard-sphere simulations with Brownian dynamics or Green's function reaction dynamics, the RDME can be orders of magnitude faster if the lattice spacing can be chosen coarse enough. However, strongly diffusion-controlled reactions mandate a very fine mesh resolution for acceptable accuracy. It is common that reactions in the same model differ in their degree of diffusion control and therefore require different degrees of mesh resolution. This renders mesoscopic simulation inefficient for systems with multiscale properties. Mesoscopic-microscopic hybrid methods address this problem by resolving the most challenging reactions with a microscale, off-lattice simulation. However, all methods to date require manual partitioning of a system, effectively limiting their usefulness as "black-box" simulation codes. In this paper, we propose a hybrid simulation algorithm with automatic system partitioning based on indirect *a priori* error estimates. We demonstrate the accuracy and efficiency of the method on models of diffusion-controlled networks in 3D. *Published by AIP Publishing.* https://doi.org/10.1063/1.5002773

I. INTRODUCTION

Cellular control systems are inherently spatial, as reactions in a network involve macromolecules confined to certain locations or sub-compartments. For example, MAPK pathways involve sensing a stimulus by receptors localized to the cell membrane, propagating the signal in the cytoplasm via a phosphorylation cascade that modifies transcription factors, which are eventually imported into the nucleus where they bind to promoter sites on DNA and affect the expression of downstream genes. On an even finer level, spatial correlations on very short length scales impact the dynamics of various biochemical systems.[1] The low copy numbers of key molecules introduce stochasticity in gene regulatory networks (GRNs). This is an important factor to account for when studying the regulatory properties of GRNs. Examples where both spatial and stochastic effects are predicted to be important[1] include spatial gene regulation of Hes1,[2] polarization in budding yeast,[3] and the MinD-system in *E. Coli*.[4]

As a consequence, spatio-temporal simulation of reaction-diffusion systems is an important tool to analyze GRNs. In particular, two modeling frameworks have attracted considerable attention in the systems' biology community: the mesoscopic, discrete stochastic reaction-diffusion master equation (RDME) in which point-like molecules are tracked on a grid and Brownian Dynamics (BD) in which hard-sphere particles are tracked in continuous space. Many capable software packages have been created to support such spatial modeling, including MCell,[5] Smoldyn,[6] E-Cell,[7] MesoRD,[8] VCell,[9] STEPS,[10] NeuroRD,[11] ReaDDy,[12] URDME,[13] PyURDME,[14] and StochSS,[15] the latter integrates spatial capabilities via PyURDME.

Mesoscopic simulators are efficient if a reasonably coarse mesh can be used. However, for some diffusion-limited systems, it is critical to capture short-range, short-time scale interactions between the molecules.[1,16,17] In this case a microscopic, particle-scale resolution is needed for accurate simulation. This raises the question of how well the mesoscopic model can capture the microscale dynamics as the mesh size approaches the molecule size. Unless measures are taken, the RDME approximation quality degrades as the mesh size becomes increasingly fine.[18–20]

The point-particle mesoscopic model can be made to approximate a microscopic model by deriving scale-dependent reaction rates[17,19,21,22] down to a critical mesh size after which no one-neighbor stencil can provide increased accuracy.[21] After this critical limit, it is possible to improve the approximation even further by considering a wider stencil.[17,20] This approach results in a lattice method with a lattice spacing on the order of the size of the molecules. By considering the Doi model[23] rather than the Smoluchowski model on the microscopic scale, it is also possible to arrive at a non-local, convergent mesoscopic model by directly discretizing the microscopic model.[24] But even if these solutions allow for accurate simulation of diffusion-limited systems down to a mesh size close to individual particles, the simulation cost increases dramatically as the mesh becomes finer: the number of diffusion events (and so the simulation time) grows proportionally to h^{-2}, where h is a measure of the mesh size.

A promising approach to efficiently simulate systems with multiscale properties is hybrid methods in which the reaction network is partitioned and parts of it are simulated on different modeling levels and with different solvers. Examples include mesoscopic-macroscopic methods,[25–28] macroscopic-mesoscopic methods,[29] macroscopic-microscopic methods,[30]

and mesoscopic-microscopic methods, in which parts of a system are simulated with the RDME and parts of it are simulated with a particle-tracking algorithm.[31–33] If a good partitioning can be found, the cost savings of hybridization can be substantial by keeping the number of particles handled on the microscale small, yet maintaining a reasonable mesh size for the RDME simulator. However, one key problem with hybrid methods is that prior knowledge about the system is needed in order to partition it correctly. Also, the system dynamics may change over the course of the simulation or in different regions in the spatial domain, making the initial system partitioning invalid or suboptimal. These issues make hybrid solvers hard to use without expert knowledge, which limits their usefulness as black-box simulation tools. Another problem is that they are complex and challenging to implement, thus there is a lack of software capable of hybrid simulation.

In this paper, we develop theory and a practical method that, given a user-supplied error tolerance, is capable of automatic selection of the appropriate modeling level for each species in a spatial stochastic model. We show that the hybrid method converges with decreasing time step and demonstrate numerically that it accurately reproduces the fine-grained reaction dynamics of microscopic methods. Finally, we show that the hybrid method can simulate systems that are intractable with pure mesoscopic methods, without having to simulate the whole system microscopically.

The rest of the paper is organized as follows. In Sec. II, we briefly introduce the mesoscopic and microscopic models and review how they are related. In Sec. III, we describe the proposed hybrid algorithm, we show how to split a general system, and we develop the condition to be used for adaptive system partitioning. In Sec. IV, we discuss the practical implementation of the method and demonstrate the accuracy and efficiency of the developed method on two challenging test problems in Sec. V. Section VI concludes the paper.

II. BACKGROUND

Two modeling frameworks that are popular for simulating reaction-diffusion kinetics in systems biology are the microscopic Smoluchowski model and the mesoscopic reaction-diffusion master equation (RDME). In the former, particles are modeled as hard spheres and their positions are tracked continuously in space, whereas in the latter, particles are point-particles and their positions are tracked up to the resolution of a structured or unstructured grid approximating the domain.

A. Mesoscopic model

On the mesoscopic scale, we model molecules as point particles and treat diffusion as jumps between adjacent voxels on a mesh. The state of the system is the discrete number s_i of each chemical species S_i, $i = 1 \dots M$, in the voxels of the grid, where the voxels are denoted by $\mathcal{V}_j$, $j = 1 \dots N$.

A diffusive jump is a linear event

$$S_{1i} \xrightarrow{D_1} S_{1j}, \tag{1}$$

taking a molecule of species S_1 from voxel $\mathcal{V}_i$ to an adjacent voxel $\mathcal{V}_j$, where D_1 is a rate constant that depends on the diffusion constant of species S_1 and the shape and size of the voxels.[34]

Chemical reactions occur between molecules residing in the same voxel. For example, a bimolecular reaction between species S_1 and S_2 producing S_3 in voxel $\mathcal{V}_j$ can be written as

$$S_{1j} + S_{2j} \xrightarrow{k_a} S_{3j}, \tag{2}$$

where k_a is a mesoscopic reaction rate. The mathematical formalism is the continuous-time discrete-space Markov process. In this framework, the propensity for the reaction, $a(s_{1j}, s_{2j}) = k_a s_{1j} s_{2j}/V_j$, is the probability of the reaction occurring in an infinitesimal interval $(t, t + \delta t)$, where V_j is the volume of voxel $\mathcal{V}_j$. With this assumption, realizations of the process can be efficiently simulated using versions of the stochastic simulation algorithm (SSA) optimized for reaction-diffusion systems such as the Next Subvolume Method (NSM).[35]

B. Microscopic model

Consider two molecules M_1 and M_2, of species S_1 and S_2, respectively. The molecules can react irreversibly according to $S_1 + S_2 \rightarrow S_3$. In the microscopic Smoluchowski model, molecules are modeled as hard spheres with a finite reaction radius, diffusing according to Brownian motion. We denote the molecules' reaction radii by σ_1 and σ_2 and their diffusion constants by D_1 and D_2. Their positions in $\mathbb{R}^3$ at time t_0 are denoted by $\mathbf{r}_{10}$ and $\mathbf{r}_{20}$. The probability that the molecules are at positions $\mathbf{r}_1$ and $\mathbf{r}_2$ at time t is described by the probability density function (pdf) $p(\mathbf{r}_1, \mathbf{r}_2, t|\mathbf{r}_{10}, \mathbf{r}_{20}, t_0)$. Let $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ and $\mathbf{R} = \sqrt{\frac{D_2}{D_1}}\mathbf{r}_1 + \sqrt{\frac{D_1}{D_2}}\mathbf{r}_2$. We can now rewrite the pdf as

$$p(\mathbf{r}_1, \mathbf{r}_2, t|\mathbf{r}_{10}, \mathbf{r}_{20}, t_0) = p_{\mathbf{R}}(\mathbf{R}, t|\mathbf{r}_0, t_0) p_{\mathbf{r}}(\mathbf{r}, t|\mathbf{r}_0, t_0), \tag{3}$$

and it can be shown that the dynamics of the two molecules in $\mathbb{R}^3$ is governed by the following system of equations:[36]

$$\frac{\partial p_{\mathbf{R}}}{\partial t} = D\Delta_{\mathbf{R}} p_{\mathbf{R}}, \tag{4}$$

$$\frac{\partial p_{\mathbf{r}}}{\partial t} = D\Delta_{\mathbf{r}} p_{\mathbf{r}}, \tag{5}$$

where $D = D_1 + D_2$. The initial and boundary conditions are given by

$$4\pi\sigma^2 D \frac{\partial p_{\mathbf{r}}}{\partial n}\bigg|_{\|\mathbf{r}\|=\sigma} = k_a p_{\mathbf{r}}(\|\mathbf{r}\| = \sigma, t|\mathbf{r}_0, t_0), \tag{6}$$

$$p_{\mathbf{r}}(\|\mathbf{r}\| \rightarrow \infty, t|\mathbf{r}_0, t_0) = 0, \tag{7}$$

$$p_{\mathbf{r}}(\mathbf{r}, t_0|\mathbf{r}_0, t_0) = \delta(\mathbf{r} - \mathbf{r}_0). \tag{8}$$

This system of equations can be solved exactly[36,37] or using an operator split approach.[38] We will call k_a the microscopic reaction rate. The probability that the two molecules react between times t_0 and t is given by

$$p_{\mathbf{r}}(*, t|\mathbf{r}_0, t_0) = 1 - \int_{t_0}^{t} p_{\mathbf{r}}(\mathbf{r}, t|\mathbf{r}_0, t_0)\, d\mathbf{r}. \tag{9}$$

The time t_d until a molecule undergoes a unimolecular reaction event with reaction rate k_d is given by sampling t_d from an exponential distribution with mean k_d. If the dissociation event produces two molecules, then they are placed at a distance of σ apart. We will outline in Sec. IV B how to use $p_{\mathbf{r}}(*, t|\mathbf{r}_0, t_0)$ and $p(\mathbf{r}_1, \mathbf{r}_2, t|\mathbf{r}_{10}, \mathbf{r}_{20}, t_0)$ to simulate a more complex system within a bounded domain.

C. Mesoscopic parameters

We now ask how the mesoscopic and microscopic models are related. Specifically we need to relate the mesoscopic parameters to the microscopic parameters. The diffusion jump rates on the mesoscopic scale are obtained by discretizing the diffusion equation. On a structured mesh, this is straightforward; for details on how to do it on an unstructured mesh, see Ref. 13. Below we outline an approach to relating the mesoscopic reaction rates to the microscopic reaction rates.

It is well known that if the reaction volume V in 3D is much larger than the molecules, i.e., $V \gg \sigma$, then the mesoscopic reaction rate, k_{CK}, relates to the microscopic reaction rate as

$$k_{\mathrm{CK}} = \frac{1}{V}\frac{4\pi\sigma D k_a}{4\pi\sigma D + k_a}, \tag{10}$$

where D is the sum of the molecules' diffusion constants and σ is the sum of the reaction radii. This expression was first derived by Collins and Kimball[39] and later re-derived by Gillespie.[40] It is easy to see that for a spatially discretized well-mixed system, or if the voxels are large enough ($h \gg \sigma$, where h^3 is the volume of a voxel), the mesoscopic reaction rate, k_a^{meso}, will be given by

$$k_a^{\mathrm{meso}} = \frac{1}{h^3}\frac{4\pi\sigma D k_a}{4\pi\sigma D + k_a}. \tag{11}$$

In models with multiscale properties, we often need to resolve part of the system to very high accuracy, which requires a highly resolved mesh. This implies that the condition $h \gg \sigma$ might not be satisfied. We thus need to derive reaction rates for this case.

Following the analysis in Refs. 19 and 21, we start by considering a single irreversible reaction

$$\mathrm{S}_1 + \mathrm{S}_2 \xrightarrow{k_a} \mathrm{S}_3 \tag{12}$$

in a cube discretized by a Cartesian mesh. Additionally, we will assume that the S_1 molecule is fixed inside a voxel close to the center of the domain, while the S_2 molecule diffuses freely with diffusion rate D.

To study the relationship between the RDME and Smoluchowski models for small h, we compare the expected time until the molecules first react on the microscopic scale to the time on the mesoscopic scale.

First, consider the microscopic scale. Let the S_2 molecule have an initial position sampled from a uniform distribution and denote the mean binding time, or the average time until the molecules react given that the S_1 molecule is uniformly distributed, by τ_{micro}. Following the approach in Ref. 21, we split τ_{micro} into two parts,

$$\tau_{\mathrm{micro}} = \tau_{\mathrm{diff}}^{\mathrm{micro}} + \tau_{\mathrm{react}}^{\mathrm{micro}}, \tag{13}$$

where $\tau_{\mathrm{diff}}^{\mathrm{micro}}$ is the average time until the S_1 molecule is in contact with the S_2 molecule for the first time and $\tau_{\mathrm{react}}^{\mathrm{micro}}$ is the average time until the molecules react given that they are in contact.

We know that[17,19]

$$\tau_{\mathrm{diff}}^{\mathrm{micro}} \approx \begin{cases} \dfrac{V}{4\pi\sigma D} & (3D) \\ \dfrac{V\left\{\log\left(\pi^{-1}\frac{V^{1/2}}{\sigma}\right)\right\}}{2\pi D} & (2D) \end{cases} \tag{14}$$

and that

$$\tau_{\mathrm{react}}^{\mathrm{micro}} = \frac{V}{k_a} \quad (1\mathrm{D},\ 2\mathrm{D},\ 3\mathrm{D}). \tag{15}$$

We now consider system (12) on the mesoscopic scale. The S_2 molecule is fixed in a voxel close to the origin so that it is far from the boundaries. The S_1 molecule is sampled uniformly on the mesh, and τ_{meso} denotes the average time until the two molecules react for the first time. Let $\tau_{\mathrm{diff}}^{\mathrm{meso}}$ be the average time until the molecules are in the same voxel for the first time, and let $\tau_{\mathrm{react}}^{\mathrm{meso}}$ denote the average time until they react given that the molecules start in the same voxel.

Again we split the average binding time into two parts

$$\tau_{\mathrm{meso}} = \tau_{\mathrm{diff}}^{\mathrm{meso}} + \tau_{\mathrm{react}}^{\mathrm{meso}}, \tag{16}$$

where (with $C_2 = 0.1951$ and $C_3 = 1.5164$)[19,21,41,42]

$$\tau_{\mathrm{diff}}^{\mathrm{meso}} = \begin{cases} \dfrac{C_3 V}{6Dh} + O\left(N^{\frac{1}{2}}\right) & (3D) \\ \dfrac{V}{4\pi D}\log(N) + \dfrac{C_2 V}{4D} + O\left(N^{-1}\right) & (2D) \end{cases} \tag{17}$$

and

$$\tau_{\mathrm{react}}^{\mathrm{meso}} = \frac{N}{k_a^{\mathrm{meso}}}. \tag{18}$$

We make the ansatz that the mean reaction times on the mesoscopic and microscopic scales are equal, to obtain

$$\tau_{\mathrm{meso}} = \tau_{\mathrm{micro}} \tag{19}$$

$$\iff \tau_{\mathrm{diff}}^{\mathrm{meso}} + \tau_{\mathrm{react}}^{\mathrm{meso}} = \tau_{\mathrm{micro}} \tag{20}$$

$$\iff k_a^{\mathrm{meso}} = \frac{N}{\tau_{\mathrm{micro}} - \tau_{\mathrm{diff}}^{\mathrm{meso}}}, \tag{21}$$

where the last equality follows from (18). Note that τ_{micro} and $\tau_{\mathrm{diff}}^{\mathrm{meso}}$ are both known so that we can compute k_a^{meso} using (21). We showed in Ref. 19 that k_a^{meso} can be rewritten as

$$k_a^{\mathrm{meso}} = \frac{k_a}{h^3}\left(1 + \frac{k_a}{D}G(h,\sigma)\right)^{-1}, \tag{22}$$

where G in 3D is given by

$$G(h,\sigma) = \frac{1}{4\pi\sigma} - \frac{C_3}{6h}. \tag{23}$$

From the analysis follows the existence of an (for accuracy) optimal mesh size. To see this, consider that

$$\tau_{\mathrm{diff}}^{\mathrm{meso}} < \tau_{\mathrm{diff}}^{\mathrm{micro}} \implies \tau_{\mathrm{react}}^{\mathrm{meso}} > \tau_{\mathrm{react}}^{\mathrm{micro}}, \tag{24}$$

$$\tau_{\mathrm{diff}}^{\mathrm{meso}} = \tau_{\mathrm{diff}}^{\mathrm{micro}} \implies \tau_{\mathrm{react}}^{\mathrm{meso}} = \tau_{\mathrm{react}}^{\mathrm{micro}}, \tag{25}$$

$$\tau_{\mathrm{diff}}^{\mathrm{meso}} > \tau_{\mathrm{diff}}^{\mathrm{micro}} \implies \tau_{\mathrm{react}}^{\mathrm{meso}} < \tau_{\mathrm{react}}^{\mathrm{micro}}. \tag{26}$$

The reaction dynamics is better resolved on the microscopic scale than on the mesoscopic scale, and so we expect the mesoscopic accuracy to increase as $\tau_{\mathrm{react}}^{\mathrm{meso}}$ approaches $\tau_{\mathrm{react}}^{\mathrm{micro}}$ from above. This was shown to be true in Ref. 19. However, as $\tau_{\mathrm{react}}^{\mathrm{meso}}$ decreases further, the accuracy also worsens. This was also shown in Ref. 19.

Thus, since $\tau_{\mathrm{diff}}^{\mathrm{meso}}$ increases with decreasing h, we note that in general we expect the most accurate mesoscopic simulations by selecting h such that (25) holds. Solving $\tau_{\mathrm{diff}}^{\mathrm{meso}} = \tau_{\mathrm{diff}}^{\mathrm{micro}}$ for h yields the optimal mesh size h^*,[19]

$$h^* = \begin{cases} \dfrac{2C_3}{3}\pi\sigma \approx 3.2\sigma & (3D) \\ \sqrt{\pi}e^{\frac{3+2C_2\pi}{4}}\sigma \approx 5.1\sigma & (2D) \end{cases}, \tag{27}$$

where

$$C_3 \approx 1.5164, \tag{28}$$

$$C_2 \approx 0.1951. \tag{29}$$

D. Hybrid methods

We previously developed a hybrid method[31] that allowed a given system to be split into two parts: a mesoscopic part and a microscopic part. Species are divided into one subsystem simulated on the microscopic scale and one subsystem simulated on the mesoscopic scale. The division could depend on spatial constraints so that a species would be simulated as part of the microscopic subset only in certain parts of space, but not in others. With the system split into two subsets, we would proceed to simulate the system in sequence:

1. Initialize and set $t = 0$. Let the final time be T. Select a splitting time step Δt.
2. Simulate the mesoscopic subset for Δt seconds, while keeping the microscopic subset fixed.
3. Simulate the microscopic subset for Δt seconds, while keeping the mesoscopic subset fixed. However, we allow microscopic molecules to react bimolecularly with mesoscopic molecules.
4. Synchronize and assign all newly created molecules to their respective scale.
5. Add Δt to t. Repeat 2-4 until $t = T$.

A crucial and counter-intuitive aspect of this algorithm is that it is necessary to select a time step Δt that is neither too small nor too large. The method does not, in general, converge with $\Delta t \to 0$. A rule of thumb is that Δt should be selected such that molecules diffuse on the length scale of individual voxels in between synchronization. However, it is straightforward to design a system that would require a smaller Δt for accurate simulation and for which the above hybrid method does not work.

To see why the simple scheme outlined above leads to the existence of an optimal time step Δt, consider the following model system:

$$\mathrm{S}_1 \xrightarrow{k_1} \mathrm{S}_{11} + \mathrm{S}_{12} \xrightarrow{k_2} \mathrm{S}_2, \tag{30}$$

with S_1 microscopic and S_{11}, S_{12}, and S_2 mesoscopic. When S_1 dissociates, S_{11} and S_{12} are placed at contact and they might therefore rebind quickly to form S_2. If Δt is large, then it is likely that this fast interaction will be captured on the microscopic scale during that time step, and the accuracy will consequently be high. If Δt is small on the other hand, then S_{11} and S_{12} will become mesoscopic quickly after S_1 dissociates, and information about the spatial correlation of S_{11} and S_{12} will be lost.

Below we propose a way to improve the algorithm to address this problem.

III. A CONVERGENT HYBRID METHOD

Here we propose a hybrid method which builds on the algorithm[31] outlined in Sec. II D but improves it in two critical ways: First, we propose a new scheme to make the simulation convergent as the splitting time step $\Delta t \to 0$ and second, we use the theory in Sec. III C below to enable automatic system partitioning.

A. Algorithm

To make the method converge monotonically as its time step decreases, we here generalize the splitting over species to allow for *dynamic splitting*, in which a *time-dependent* function maps molecules to either scale. Let t_j denote the time elapsed since the molecule with index j was created, and let $F(S_j, t_j)$ denote the function mapping a molecule of species S of age t_j to either the mesoscopic subset or the microscopic subset. Then, for system (30), we let

$$F(S_1, t) = \text{microscale}, \quad \text{for all } t, \tag{31}$$

$$F(S_{11}, t) = F(S_{12}, t) = \begin{cases} \text{microscale}, & t \le t_m, \\ \text{mesoscale}, & t > t_m, \end{cases} \tag{32}$$

$$F(S_2, t) = \text{mesoscale}, \quad \text{for all } t, \tag{33}$$

where t_m is chosen sufficiently large (see Sec. III D for how to choose t_m) and where t is the time since the molecule was created.

The algorithm in Ref. 31 now becomes a special case ($t_m = 0$) of the algorithm proposed here:

B. How to split a system

For a given system, we will need to determine a suitable splitting in order to achieve high accuracy as well as efficient simulations. Again, consider the system in Eq. (30). Taking symmetry into account and by observing that for this particular system the species S_2 can be safely simulated on the mesoscopic scale as it only diffuses, we can split the system in five different ways,

$$\mathcal{X}_1 : S_1 \text{ micro}; S_{11}, S_{12}, S_2 \text{ meso}, \tag{34}$$

$$\mathcal{X}_2 : S_1, S_{11} \text{ micro}; S_{12}, S_2 \text{ meso}, \tag{35}$$

$$\mathcal{X}_3 : S_1, S_{11}, S_{12} \text{ micro}; S_2 \text{ meso}, \tag{36}$$

$$\mathcal{X}_4 : S_{11}, S_{12} \text{ micro}; S_1, S_2 \text{ meso}, \tag{37}$$

$$\mathcal{X}_5 : S_{11} \text{ micro}; S_1, S_{12}, S_2 \text{ meso}. \tag{38}$$

We will now consider the accuracy and convergence of each of the splittings $\mathcal{X}_1$–$\mathcal{X}_5$.

$\mathcal{X}_1$: In Algorithm 1, the S_{11} and S_{12} molecules will be simulated on the mesoscale for t_m seconds, and therefore, if t_m is chosen large enough such that the molecules either rebind or can be considered well-mixed inside their respective voxels at the end of t_m, the system will be accurately simulated. Importantly, this is also true for $\Delta t \to 0$.

$\mathcal{X}_2$: The accuracy is the same as for the splitting in $\mathcal{X}_1$, since S_{12} is mesoscopic. Association events between microscale and mesoscale molecules have the same spatial resolution as a pure mesoscopic association event.

$\mathcal{X}_3$: All molecules of interest are simulated on the microscopic scale; the accuracy is therefore the same as for a pure microscale simulation, but with no efficiency gained.

$\mathcal{X}_4$: The S_1 molecule dissociates on the mesoscopic scale, so all spatial correlation is lost (up to the size of the voxel) upon dissociation. Even though we proceed to simulate the S_{11} and S_{12} molecules on the microscopic scale, we still get the accuracy of a mesoscopic simulation.

Algorithm 1. Hybrid method.

1. Initialize the system. Set the time $t = 0$. Let T be the length of the simulation.
2. Assign molecules to the mesoscopic and microscopic subsets according to $F(S, t)$.
3. Simulate the mesoscopic molecules for Δt seconds. Mesoscopic molecules cannot interact with microscopic molecules during the time step. Any molecules produced will, for the remainder of the time step, be simulated on the mesoscopic scale.
4. Simulate the microscopic molecules for Δt seconds, while freezing the mesoscopic molecules. Microscopic molecules can react mesoscopically with mesoscopic molecules. Any molecules produced will, for the remainder of the time step, be simulated on the microscopic scale.
5. Add Δt to t.
6. Repeat 2-5 until $t = T$.

$\underline{\mathcal{X}_5}$: The argument from $\mathcal{X}_4$ holds here as well. The accuracy will be the same as for a mesoscopic simulation.

In conclusion, for this model problem, the *only* viable splitting of system (30) (apart from the trivial pure microscopic simulation) is $\mathcal{X}_1$, assuming that we want to simulate as few species as possible on the microscopic scale.

Note that we may, in some cases, be able to simulate the system described by Eq. (30) accurately with the method in Ref. 31. However, with that algorithm, we cannot take the splitting time step arbitrarily small. The reason for this is described in detail in Ref. 31. While this is acceptable for some systems, it will lead to inaccuracies for many others.

As an example, consider the following system:

$$\mathrm{S}_1 \xrightarrow{k_1^1} \mathrm{S}_{11} + \mathrm{S}_{12} \xrightarrow{k_2^1} \mathrm{S}_2, \tag{39}$$

$$\mathrm{S}_2 \xrightarrow{k_1^2} \mathrm{S}_{21} + \mathrm{S}_{22} \xrightarrow{k_2^2} \mathrm{S}_3, \tag{40}$$

where k_2^1 and k_2^2 are large so that both association reactions are diffusion limited.

First consider the case of $t_m^1 = 0$ and $t_m^2 = 0$. The method now reduces to the method in Ref. 31. We know that Δt has to be large enough. S_1 dissociates on the microscale, so both S_{11} and S_{12} will be microscale until the end of the time step. However, if they are created near the end of the time step, they are likely to survive until the end of the time step and then turn mesoscopic. If they react on the mesoscale, then the product S_2 will be mesoscopic until the end of the time step. If $1/k_1^2 \ll \Delta t$, the S_2 molecule is likely to dissociate before the end of the time step, in which case S_{21} and S_{22} are initially mesoscopic. Information about the spatial correlation between S_{21} and S_{22} is lost, up to the size of the voxel, and the rest of the simulation may therefore be inaccurate.

With Algorithm 1, we can guarantee high accuracy, by fixing t_m^1 and t_m^2 large enough and choosing Δt small enough. We then ensure that S_{11} and S_{12} exist on the microscale long enough to either react quickly or become well-mixed insider their respective voxels, while also ensuring that S_2 is mesoscopic only on a time scale much shorter than k_1^2 (by selecting $\Delta t \ll 1/k_1^2$).

C. Criteria for selecting the modeling scale of different species

Based on the work outlined in Sec. II C, it is clear that choosing a mesh size $h = h^*$ in general leads to the most accurate mesoscopic simulations. In fact, it is possible to push h^* almost to the size of the molecules if the model is extended to allow for reactions between molecules in adjacent voxels.[17,20] The problem is that h^* is small and this makes the mesoscopic simulations expensive, sometimes significantly more expensive than microscopic simulations.[20] Another problem is that h^* is a function of the reaction radius, and thus different reactions may require different mesh resolutions to be resolved. This can make it impossible to simulate a system accurately with the RDME.[20] We therefore want to perform mesoscopic simulations for the majority of the system with $h \gg h^*$ and handle reactions that require a very fine mesh with a microscopic solver.

For any given $h \geq h^*$, the RDME solver will match the mean binding time of the microscopic model if the mesoscale propensity functions from Ref. 19 are used, but if $h > h^*$, it will not perfectly match the fine-grained reaction dynamics. Here, we use the relative error in $\tau_{\text{react}}^{\text{meso}}$ to estimate this error. We let

$$W(h) = \frac{\left|\tau_{\text{react}}^{\text{meso}} - \tau_{\text{react}}^{\text{micro}}\right|}{\tau_{\text{react}}^{\text{micro}}} \tag{41}$$

$$= \frac{\frac{N}{k_a^{\text{meso}}} - \frac{V}{k_a}}{\frac{V}{k_a}} \tag{42}$$

$$= \frac{k_a}{h^d k_a^{\text{meso}}} - 1, \tag{43}$$

where we have used that $\tau_{\text{react}}^{\text{meso}} > \tau_{\text{react}}^{\text{micro}}$ for $h > h^*$. Now using (22) in place of k_a^{meso}, we obtain

$$W = \frac{k_a}{h^d \frac{k_a}{h^d}\left(1 + \frac{k_a}{D} G(h, \sigma)\right)^{-1}} - 1 \tag{44}$$

$$= \frac{k_a}{D} G(h, \sigma). \tag{45}$$

We assume that for $W(h) < \epsilon$, for some small enough ϵ, a reaction is sufficiently resolved on the mesoscopic scale. We can hence use $W(h)$ to decide which species need to be handled on the microscopic scale in order to resolve the reaction dynamics to high enough accuracy. The assumption $W < \epsilon$ with (22) and (45) holds if and only if

$$\frac{k_a}{h^3}(1 + \epsilon)^{-1} < k_a^{\text{meso}} \leq \frac{k_a}{h^3}. \tag{46}$$

In other words, a reaction is well resolved mesoscopically when the mesoscopic reaction rate is sufficiently close to the microscopic reaction rate (scaled by the volume of the voxel).

In Sec. V A, we suggest a reasonable default value for ϵ based on numerical experiments.

D. How to select t_m

We need to select t_m in Eq. (32) so that the hybrid method accurately simulates the system in Eq. (30). In particular we want the relative error

$$E_{hybrid} = \frac{|\tau_{\text{react}}^{\text{hybrid}} - \tau_{\text{react}}^{\text{micro}}|}{\tau_{\text{react}}^{\text{micro}}} \tag{47}$$

to be small for t_m large enough. Following a dissociation of S_1, the average time until the molecules react is given by

$$\tau_{\text{react}}^{\text{hybrid}} = \tau_{\text{react}}^{\text{micro}}|_{t\le t_m} + \tau_{\text{react}}^{\text{meso}}|_{t>t_m}, \tag{48}$$

where $\tau_{\text{react}}^{\text{micro}}|_{t\le t_m}$ is the average time until the molecules react on the microscopic scale, given that they react before t_m, and $\tau_{\text{react}}^{\text{meso}}|_{t>t_m}$ is the average time until the molecules react on the mesoscopic scale, given that they react after t_m.

Let $S(t)$ denote the probability that the molecules do not react before time t. Then

$$\tau_{\text{react}}^{\text{meso}}|_{t>t_m} = S(t_m)(\tau_{\text{diff}}^{\text{meso}}|_{t>t_m} + \tau_{\text{react}}^{\text{meso}}), \tag{49}$$

where $\tau_{\text{diff}}^{\text{meso}}|_{t>t_m}$ is the average time that it takes for the molecule to diffuse back to the origin voxel, given that the molecules did not react before time t_m. By a similar argument, we can write

$$\tau_{\text{react}}^{\text{micro}} = \tau_{\text{react}}^{\text{micro}}|_{t\le t_m} + \tau_{\text{react}}^{\text{micro}}|_{t>t_m} \tag{50}$$

$$= \tau_{\text{react}}^{\text{micro}}|_{t\le t_m} + S(t_m)(\tau_{\text{diff}}^{\text{micro}}|_{t>t_m} + \tau_{\text{react}}^{\text{micro}}). \tag{51}$$

Now, by (48)–(51)

$$E_{hybrid} = \frac{\tau_{\text{react}}^{\text{hybrid}} - \tau_{\text{react}}^{\text{micro}}}{\tau_{\text{react}}^{\text{micro}}} \tag{52}$$

$$= S(t_m)\frac{\tau_{\text{diff}}^{\text{meso}}|_{t>t_m} - \tau_{\text{diff}}^{\text{micro}}|_{t>t_m} + \tau_{\text{react}}^{\text{meso}} - \tau_{\text{react}}^{\text{micro}}}{\tau_{\text{react}}^{\text{micro}}} \tag{53}$$

$$= S(t_m)\left(Q + \frac{k_a}{D}G(h,\sigma)\right) \tag{54}$$

$$\le S(t_m)(Q+\epsilon), \tag{55}$$

with

$$Q = \frac{\tau_{\text{diff}}^{\text{meso}}|_{t>t_m} - \tau_{\text{diff}}^{\text{micro}}|_{t>t_m}}{\tau_{\text{react}}^{\text{micro}}}. \tag{56}$$

It is now easy to see that since we are considering a bounded domain, where $\tau_{\text{diff}}^{\text{meso}}|_{t>t_m} \to \tau_{\text{diff}}^{\text{meso}}$, $\tau_{\text{diff}}^{\text{micro}}|_{t>t_m} \to \tau_{\text{diff}}^{\text{micro}}$, then as $t_m \to \infty$

$$S(t_m) \to 0, \;\; Q \to 0, \;\; \text{as } t_m \to \infty. \tag{57}$$

The method thus converges as $t_m \to \infty$, which is also easy to see intuitively, as the simulation in practice becomes purely microscopic for t_m large enough.

We have seen numerically that the mesoscopic method incurs an error in average rebind time that is on the order of the time that it takes for a molecule to become well-mixed inside a voxel.[19] We therefore propose to select $t_m = K^2 V_{\text{vox}}^{2/3}/(6D)$, that is, a time proportional to the time that it takes for a molecule to diffuse a distance proportional to the length scale of a voxel. Here K is a constant, and we have found $K = 6$ to be sufficiently large.

We note that $S(t)$ is not known in general since we are considering a bounded domain. However, assuming that t_m will be small enough, and the molecules are some distance from the boundary, we can approximate $S(t)$ by the survival probability for an unbounded domain. This quantity is known analytically and in 3D is given by[43]

$$S(t) = 1 - \frac{k_a}{4\pi\sigma D + k_a}\left(1 - \exp(\alpha^2 t)\text{erfc}(\alpha\sqrt{t}\right), \tag{58}$$

where

$$\alpha = \left(1 + \frac{k_a}{4\pi\sigma D}\right)\frac{\sqrt{D}}{\sigma}. \tag{59}$$

We thus know all terms of E_{hybrid} except for Q. By selecting t_m large enough, we are ensuring that either $Q \approx -k_a/DG(h,\sigma)$ or $S(t_m) \approx 0$, and hence $E_{hybrid} \approx 0$. We show numerically in Fig. 1 that our choice of t_m is sufficiently large to accurately reproduce the microscopic distribution of rebind times for two different diffusion-limited reaction rates.

E. Splitting a more complex system

To further illuminate how to practically implement the automatic splitting of a system, we will consider some more

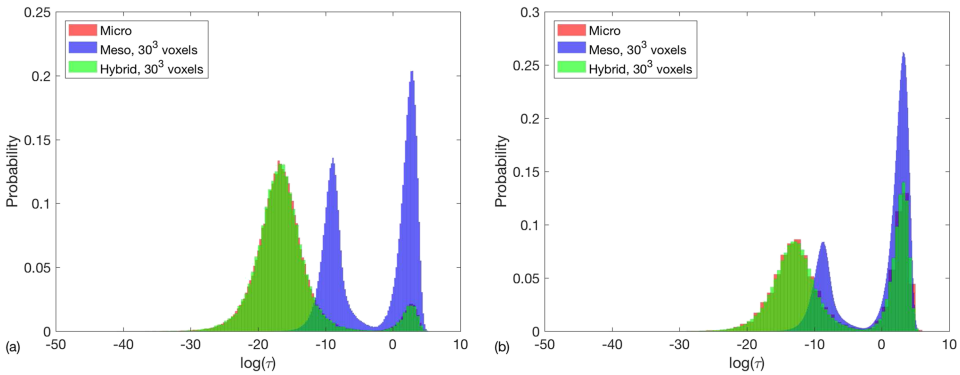

FIG. 1. We consider two molecules, one fixed and one diffusing at diffusion rate 1.0, in a cube of volume 1.0. The molecules react irreversibly with reaction rate k_a. Let τ denote the time until the molecules react, given that they start in contact (or in the same voxel on the mesoscopic scale), with a total reaction radius of 0.005. Above we plot the distribution of the logarithm of the rebind time τ. In (a), $k_a = 1.0$, and in (b), $k_a = 0.1$. We can see that the hybrid method reproduces the microscopic distribution closely (note that green and red overlap in the figures above), while the mesoscopic RDME does not reproduce the same distribution for diffusion-limited reactions. In particular, the RDME is unable to accurately resolve reaction events occurring on a spatial scale of one voxel or less.

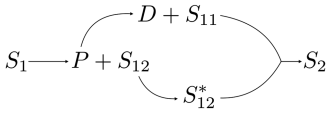

FIG. 2. Schematic illustration of system (64)–(67).

complex cases. The simple sequence

$$\mathrm{S}_1 \longrightarrow \mathrm{S}_{11} + \mathrm{S}_{12} \longrightarrow \mathrm{S}_2 \tag{60}$$

serves as the base case. A single molecule produces two new molecules, spatially correlated, that can then react to form a new molecule. However, we can consider more complex variants of this simple case.

For example, consider the following system:

$$S_1 \longrightarrow S_{11} + S_{12}, \tag{61}$$
$$S_{11} \longrightarrow S_{11}^*, \tag{62}$$
$$S_{11}^* + S_{12} \longrightarrow S_2. \tag{63}$$

In this case S_1 produces two molecules, S_{11} and S_{12}, that do not directly react. However, if the reaction $\mathrm{S}_{11} \longrightarrow \mathrm{S}_{11}^*$ is fast enough, the above system behaves similarly to the simple system in (60). We might therefore need to simulate S_1 on the microscopic scale.

Another example is the system

$$S_1 \longrightarrow P + S_{12}, \tag{64}$$
$$P \longrightarrow D + S_{11}, \tag{65}$$
$$S_{12} \longrightarrow S_{12}^*, \tag{66}$$
$$S_{11} + S_{12}^* \longrightarrow S_2. \tag{67}$$

We first find that the only bimolecular reaction is $S_{11} + S_{12} \rightarrow S_2$. Assume that we have $W > \epsilon$. We find no dissociation reaction producing S_{11} and S_{12}. We now look for reactions producing either of the molecules or both. No reactions produce both, but we find that S_{12} produces S_{12}^* and that P produces S_{11} and D. We proceed to look for any reactions producing either S_{12} or P or both. We now find only one such reaction, $S_1 \rightarrow P + S_{12}$. Since this is a dissociation reaction, S_1 will be simulated on the microscopic scale.

This is schematically illustrated in Fig. 2.

The production of S_2 originates from a dissociation, with an intermediate association reaction that is unresolved at fine length scales on the mesoscopic scale. The dissociating molecule is therefore a candidate for the microscopic scale.

We employ the following general strategy to identify molecules that should be simulated on the microscopic scale:

For system (64)–(67), the algorithm would thus first compute W for the association reaction $S_{11} + S_{12}^* \rightarrow S_2$. If $W > \epsilon$, then the algorithm proceeds to find the sequence of reactions $S_1 \rightarrow P + S_{12}$, $P \rightarrow D + S_{11}$ producing S_{11}, and the sequence of reactions $S_1 \rightarrow P + S_{12}$, $S_{12} \rightarrow S_{12}^*$ producing S_{12}^*. Since the two sequences both start with S_1 dissociating, we will simulate S_1 on the microscopic scale.

IV. IMPLEMENTATION

We have implemented the hybrid method as an extension to the high-level PyURDME[14] Python API. This allows for specification of microscopic/hybrid systems and execution via a simple, object-oriented Python modeling interface. In Secs. IV A–IV C, we describe the different components of the solver and discuss computational complexity and performance aspects of hybrid simulation.

A. Mesoscopic solver: Next-particle method (NPM)

In the NSM, reaction and diffusion events in each voxel are grouped, and the heap is organized so that each leaf corresponds to a voxel. In each iteration, the next reaction or diffusion event is executed and the next event time is updated along with the heap for each affected voxel. For a fine mesh, the vast majority of events are diffusion events and the simulation cost is dominated by the time to execute diffusion events. Ignoring reactions, the simulation cost, C_{NSM}, on a uniform grid with N voxels and M molecules of a single diffusing species can be written as

$$C_{NSM}(N, M) = C_1 N^2 M \log N, \tag{68}$$

where C_1 is an implementation- and architecture-dependent constant. Here we instead propose a particle-centric mesoscopic algorithm, the Next-Particle Method (NPM), that tracks individual particles on the grid. We simulate a mesoscopic system as follows:

The main advantage of the NPM in the context of the hybrid method is that it minimizes the overhead of switching between the mesoscopic and the microscopic solvers, since the two solvers can share one data structure for the particle list. The microscopic solver needs to know the position of individual molecules, so maintaining one particle list simplifies the mapping between discrete positions on a grid and continuous

Algorithm 2. Identifying microscale particles.

1. Identify all bimolecular reactions for which $W > \epsilon$, for some sufficiently small ϵ, where W is the relative error in the mesoscopic mean binding time, as defined in (41).
2. For each such reaction, determine if the product originates from a dissociation reaction. If it does, the dissociating molecule should be simulated on the microscopic scale.
   (a) If the two reactants are produced by a dissociating molecule we are done. The product of the association reaction then originates from a dissociation [an example of this is (60)].
   (b) If they are not, find all sequences of reactions producing reactant 1 of the association reaction and all sequences of reactions producing reactant 2.
   (c) For each sequence producing reactant 1, and starting with a dissociation, search for a sequence producing reactant 2 that starts with the same dissociation. If we find two such sequences, then the initially dissociating molecule will be simulated on the microscopic scale.

Algorithm 3. Next-particle method.

1. Particles are stored in a list, with information about species type and which voxel they currently occupy.
2. For each particle, we generate the time to and the destination voxel of its next diffusion event. Add each diffusion event to the heap. For N particles, the size of the heap will be N.
3. For two reactive particles occupying the same voxel, we generate the time until the next tentative event and add that event to the heap.
4. Execute the next event.
5. Update all dependent events. If a molecule diffused, add its next diffusion and reaction events to the heap. If molecules reacted, add new diffusion and reaction events to the heap for all molecules that were affected.
6. Repeat 4-5 until the end of the simulation.

positions in space. The cost for the NPM for the example above can be written as

$$C_{NPM}(N, M) = C_2 N^2 M \log M, \tag{69}$$

where C_2 is a constant. This highlights another potential advantage of the NPM in the context of hybrid simulation: it can be more efficient than NSM for highly resolved meshes if the number of voxels is larger than the number of particles. This will often be the case for highly resolved geometries.

Note, however, that the hybrid framework proposed here does not depend on the particular implementation of the mesoscopic method and that it would be possible to alternate algorithms depending on the particular values of N and M.

B. Microscopic solver: Green's function reaction dynamics (GFRD)

On the microscopic scale, the system is simulated with the Smoluchowski model, as described in Sec. II B. For a system of more than one or two molecules, we have an intractable many-body problem. To deal with this, we employ a strategy conceptually similar to the GFRD algorithm.[36]

The first step of the algorithm is to divide the system into subsets of one or two molecules and to select a time step Δt, such that to high accuracy we can update the subsets independently during Δt. Molecules that are each others' nearest neighbors are updated in pairs, while all other molecules are updated as single molecules. The time step Δt is chosen as large as possible, with the constraint that the probability of interactions between the separate subsets is small. We then simulate each subset for Δt seconds.

For each subset, we look for the next reaction: if two molecules react bimolecularly, we can sample the time until they react from $p_{\mathbf{r}}(*, t|\mathbf{r}_0, t_0)$ [defined in (9)], if either or both molecules can react unimolecularly, we can sample the next reaction time from exponential distributions, and finally we will look for possible interactions with the boundary. The reaction that occurs first is executed. We then repeat the procedure until the subsystem has been advanced to time $t_0 + \Delta t$.

Instead of solving Eq. (4) with boundary conditions given by Eqs. (6)–(8) exactly, we solve it using the operator split approach described in Ref. 38. Furthermore, we only sample from $p_{\mathbf{r}}(*, t|\mathbf{r}_0, t_0)$ and $p_{\mathbf{r}}(\mathbf{r}, t|\mathbf{r}_0, t_0)$ if the distance between the molecules is small and the probability of a reaction is fairly large. If the probability of reaction during the time step is small, it can be more efficient to simulate the pair using small, purely diffusive, steps of Brownian dynamics until the molecules are close. The cutoff is typically at a distance of around a few reaction radii.

C. Hybrid solver

We can now couple the mesoscopic NPM with the microscopic solver in a simple loop. Since the NPM keeps track of individual molecules, it is straightforward to map each molecule to either scale according to the splitting function F. Both solvers can keep track of how long a molecule has existed, thus making it easy to determine whether a microscopic molecule can be mapped to the mesoscopic scale.

When a molecule switches from the mesoscopic scale to the microscopic scale, we need to know its position in continuous space. We sample its position from a uniform distribution on its voxel. Similarly, when a molecule switches from the microscopic scale to the mesoscopic scale, we need to know which voxel the molecule occupies. This is straightforward, since we track which voxel a molecule occupies to accurately simulate its interaction with the boundary. This process is described in detail in Ref. 31. We summarize some of the practical details of the hybrid method in Algorithm 4.

The overhead from switching between the scales is inversely proportional to the splitting time step, $C_{coupling} = C_3(\Delta t_s)^{-1}$, where we have assumed that the number of particles that switch in each time step is small, compared to the total number of particles on both scales.

With M_1 the average number of mesoscopic particles and M_2 the average number of microscopic particles during the course of a simulation, the complexity of the overall hybrid method can be described by

$$C_{hybrid}(N, M_1, M_2) = \frac{C_2 M_1 \log M_1}{N^{-2}} + C_{gfrd}(M_2) + C_3(\Delta t_s)^{-1}. \tag{70}$$

Since the number of particles handled on the different scales depends on the mesh resolution, i.e., M_1 and M_2 are functions of N, the cost of the solver is complicated to estimate *a priori*, and it implies the existence of an optimal choice of N for performance. This will be illustrated in Sec. V B.

Note that each solver could be optimized depending on the system. If we were mainly interested in simulating systems in a cube, the microscopic solver could be significantly optimized by simplifying the process of keeping track of the boundary. For a system with many more particles than voxels, we could

Algorithm 4. Hybrid method.

1. Initialize the system. Assign molecules to the mesoscopic and microscopic subsets according to $F(S, t)$. Set the time $t = 0$. Let T be the length of the simulation.
2. Mesoscopic molecules are simulated with the NPM for Δt seconds. Microscopic molecules are frozen.
3. Simulate the microscopic molecules for Δt seconds with GFRD as implemented in Sec. IV B. The mesoscopic molecules do not diffuse. Microscopic molecules can react with mesoscopic molecules that occupy the voxel that the microscopic molecule currently resides in. The microscopic and mesoscopic molecules thus react mesoscopically by sampling a next reaction time from an exponential distribution with the correct mesoscopic reaction rate.
4. Determine the simulation scale of each molecule by evaluating $F(S, t)$.
5. For each mesoscopic molecule that transitioned to the microscopic scale:
   (a) Set its scale to microscopic.
   (b) Set the continuous position by sampling uniformly on the voxel that it currently occupies. For details on how to sample the position, see Ref. 31.
6. For each microscopic molecule that transitioned to the mesoscopic scale:
   (a) Set its scale to mesoscopic.
   (b) Determine which voxel contains its current continuous position. This will be the molecule's new mesoscopic voxel.
7. Add Δt to t.
8. Repeat 2-7 until $t = T$.

choose to simulate the mesoscopic part of the system with the NSM rather than with the NPM.

In Sec. V B, we show how the contribution to the total execution time of each solver depends on the mesh size and the system. The total cost of a simulation depends non-linearly on the mesh size, since we need to balance the trade-off between a coarse mesh and fast mesoscopic simulations but expensive microscopic simulations, with a fine mesh on which the mesoscopic simulations are more expensive while the microscopic simulations will be faster (due to the fact that we will simulate fewer molecules on the microscopic scale on a fine mesh).

V. NUMERICAL EXAMPLES

While the theory above is derived under the assumption of a Cartesian mesh, we have shown that in most cases it can be applied also to the case of unstructured meshes,[44] by substituting the voxel width h for $V_{vox}^{1/3}$, where V_{vox} is the volume of a voxel in a mesh. In particular, we show in Sec. V A that we can accurately split and simulate a system on an unstructured mesh.

Furthermore, we demonstrate that we can accurately simulate a problem previously shown to be intractable with the standard RDME model,[44] and finally we show the existence of an optimal mesh size, from an efficiency perspective, in between the coarsest and finest possible mesh sizes.

A. Splitting species: Accuracy and efficiency

In this example, we demonstrate that for a given system, we can split the species into a microscopic subset and a mesoscopic subset using $W(h)$ defined in Eq. (41). The resulting splitting of species should yield accurate and efficient simulations on a given unstructured mesh.

First we want to determine a suitable ϵ such that $W < \epsilon$ indicates that the reaction is sufficiently resolved on the mesoscopic scale. We again consider the simple system

$$\mathrm{S}_1 \xrightarrow{k_1} \mathrm{S}_{11} + \mathrm{S}_{12} \xrightarrow{k_2} S_2. \tag{71}$$

In Sec. III C, we found that

$$W = \frac{k_2}{D} G\left(V_{\mathrm{vox}}^{\frac{1}{3}}, \sigma\right) \tag{72}$$

is a measure of how well the rebinding time of a pair of molecules is resolved [where k_2^{meso} is given by (22)]. Note that we substituted h for $V_{\mathrm{vox}}^{\frac{1}{3}}$ in (72), where, on a Cartesian mesh, we would have $h^3 = V_{\mathrm{vox}}$.

While we have no theory relating W directly to the error in the mesoscopic simulation of the system, we can use it as an indirect measure of the error. Via numerical simulations we can find an ϵ such that $W < \epsilon$ implies that the simulations will be accurate.

For simplicity we consider system (71) in a cube. We let the microscopic parameters be given by

$$\begin{cases} \sigma_1 = \sigma_{11} = \sigma_{12} = \sigma_2 = 0.0025, \\ D_1 = D_{11} = D_{12} = D_2 = 1.0, \\ k_1 = 10.0, \\ V = (50 h^* \sigma)^3, \end{cases} \tag{73}$$

and we sample k_2 from [0.001, 1.0]. By design we expect the best agreement between mesoscale and microscale simulations for a mesh of 50^3 voxels. Note that these parameters are chosen arbitrarily, but we will proceed to show that the results can be applied successfully to a numerical example with different parameters.

First we compare pure mesoscopic and microscopic simulations. Let $\mathbf{y}_{\mathrm{me}} = (y_{\mathrm{me}}^1, \ldots, y_{\mathrm{me}}^N)$ be the average number of S_2 molecules, computed from M_{me} mesoscale trajectories sampled at the time points $t_1, \ldots, t_L$, and let $\mathbf{y}_{\mathrm{mi}} = (y_{\mathrm{mi}}^1, \ldots, y_{\mathrm{mi}}^L)$ be the average number of S_2 molecules computed as the average of M_{mi} microscale trajectories sampled at the time points

$t_1, \ldots, t_L$. We consider the max-norm error E, defined as

$$E = \max_{1 \le i \le L} \left| y^i_{\text{me}} - y^i_{\text{mi}} \right|. \tag{74}$$

For small values of k_2, we expect the mesoscopic simulations to be accurate also for coarse meshes, while for large k_2, we expect the error to be large unless the spatial resolution is near the maximum resolution of 50^3 voxels. In Fig. 3, we show how W correlates with the error E for different k_2 and that $\epsilon = 0.025$, although arbitrary, is a reasonable choice yielding an error of the order of 1.

We now apply the choice of $\epsilon = 0.025$ to an expanded system of three bimolecular reactions,

$$\mathrm{S}_1 \xrightarrow{k^1_1} \mathrm{S}_{11} + \mathrm{S}_{12} \xrightarrow{k^1_2} \mathrm{S}_2, \tag{75}$$

$$\mathrm{S}_2 \xrightarrow{k^2_1} \mathrm{S}_{21} + \mathrm{S}_{22} \xrightarrow{k^2_2} \mathrm{S}_3, \tag{76}$$

$$\mathrm{S}_3 \xrightarrow{k^3_1} \mathrm{S}_{31} + \mathrm{S}_{32} \xrightarrow{k^3_2} \mathrm{S}_4, \tag{77}$$

with parameters different from the simple system above. The system is simulated inside a sphere of radius 0.5, discretized with an unstructured mesh consisting of 6395 voxels.

Depending on the values of k^i_2, $i = 1, 2, 3$, we will simulate some combination of S_1, S_2, and S_3 on the microscopic scale. For $W_i > \epsilon$, S_i is simulated on the microscopic scale. The minimum time that a molecule has to exist on the microscopic scale before it becomes mesoscopic is given by $t_m = \frac{K^2 V_{\text{vox}}^{2/3}}{6D_i}$, with $K = 6$, cf. Sec. III D.

We consider six different combinations of reaction rates; see Table I. In each case, we will have a different combination of S_1, S_2, and S_3 on the microscopic scale. For $k^*_2 > 0.1$, $W > \epsilon$, while for $k^*_2 = 0.001$, we have $W < \epsilon$. Thus, for case 6 the hybrid method will simulate all molecules on the mesoscopic scale, and we therefore expect the mesoscopic simulation to agree well with the microscopic simulation.

In Fig. 4, we show that the hybrid method agrees well with the microscopic simulations. In case 6, the RDME agrees well with the microscopic model, as expected. In addition, we show that the accuracy increases with a decreasing splitting time step. For a relatively large splitting time step of 0.1, the method produces results with a fairly large error, but as we refine the splitting time step, the results approach that of a pure microscopic simulation. We have tabulated the errors of the hybrid method errors in Table II with the errors of pure mesoscopic simulations. Even with a fairly large splitting time step, the error in the hybrid method is smaller. For case 6, in which the reactions are slow compared to diffusion, both the hybrid method and the RDME produce accurate results.

TABLE I. Association rates for the six different cases.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
k^1_2	0.1	0.001	0.1	0.1	0.001	0.001
k^2_2	0.1	0.3	0.3	0.001	0.001	0.001
k^3_2	0.1	0.001	0.001	0.2	0.2	0.001

While one reason to use a hybrid method is to gain efficiency over a very fine-grained RDME simulation, another is that some systems cannot be simulated accurately with a standard RDME model. In Ref. 44, we considered the following system:

$$\mathrm{S}_1 \xrightarrow{k_d} \mathrm{S}_{11} + \mathrm{S}_{12} \xrightarrow{k_r} \mathrm{S}_2, \tag{78}$$

$$\mathrm{S}_2 \xrightarrow{k_d} \mathrm{S}_{21} + \mathrm{S}_{22} \xrightarrow{k_r} \mathrm{S}_3, \tag{79}$$

where $k_d = 10.0$ and $k_r = 0.1$. If σ_i is the reaction radius of species S_i and σ_{ij} the reaction radius of species S_{ij}, then $\sigma_1 = 10^{-3}$, $\sigma_{11} = 0.8 \times 10^{-3}$, $\sigma_{12} = 0.8 \times 10^{-3}$, $\sigma_2 = 2.0 \times 10^{-3}$, $\sigma_{21} = 1.8 \times 10^{-3}$, $\sigma_{22} = 1.8 \times 10^{-3}$, and $\sigma_3 = 2.5 \times 10^{-3}$. For simplicity, all molecules diffuse with diffusion rate 1.0. The domain is a cube of volume 1.0.

To resolve the first association, we need a mesh size of around $h^*_1 = \frac{2}{3} C_3 \pi (\sigma_{11} + \sigma_{12}) \approx 5.0 \times 10^{-3}$, and to resolve the second association, we need a mesh size of around $h^*_1 = \frac{2}{3} C_3 \pi (\sigma_{21} + \sigma_{22}) \approx 1.14 \times 10^{-2}$. We showed in Ref. 44 that we cannot resolve both reactions simultaneously with the standard local RDME; we could simulate the system by allowing reactions between neighboring voxels. However, these simulations become expensive as the mesh needs to be highly refined, and they cannot be trivially extended to unstructured meshes.

We show here that another viable approach is to simulate the system with a hybrid method. The system is simulated for 2 s, with 201 uniform time samples including $t = 0$. In Fig. 5, we plot the error E as a function of the mesh size, where E is defined as in Ref. 44. Let $\mathcal{S} = \{S_1, S_{11}, S_{12}, S_2, S_{21}, S_{22}, S_3\}$.

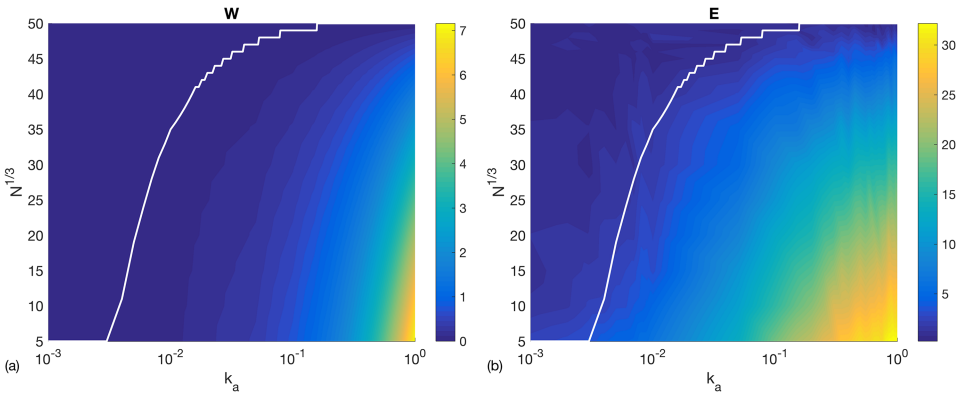

FIG. 3. In (a) we plot W as a function of $N^{\frac{1}{3}}$ and k_a. For points above the white solid line, we have $W < 0.025$. In (b) we plot the error E as a function of $N^{\frac{1}{3}}$ and k_a. Again, for points above the white solid line, $W < 0.025$. For small k_a, the error is small for all mesh sizes, while as k_a increases, we need a large N to keep the error small. We can see that for $W < 0.025$, the error is roughly on the order of 1.

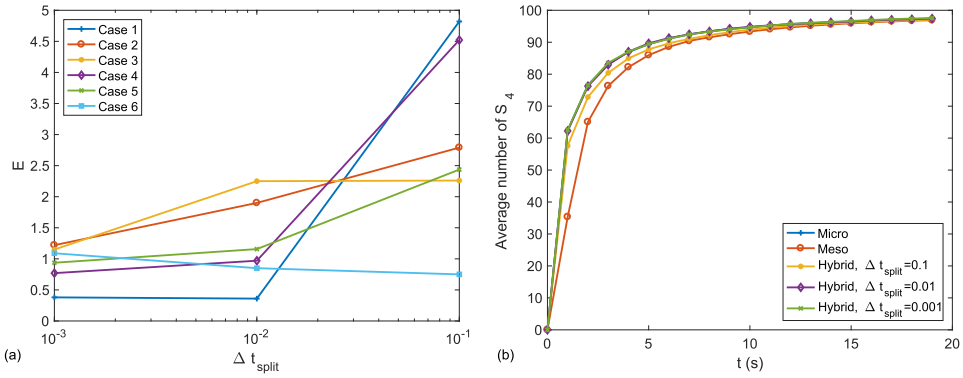

FIG. 4. (a) We plot the error E, as defined in Eq. (74) as a function of the splitting time step Δt_{split}. For the smallest time step, $\Delta t_{\text{split}} = 0.001$, the error remains small for case 6 but is larger for cases 1 and 4, in which some or all of the reactions are diffusion limited. (b) The average number of S_4 molecules over time in case 1. We see that the hybrid method with $\Delta t_{\text{split}} = 0.1$ underestimates the average number of S_4 molecules but still produces better results than with a pure mesoscopic simulation. The hybrid method matches the microscopic results closely for $\Delta t_{\text{split}} \leq 0.01$.

TABLE II. Max-norm error. We see that the hybrid method, even for a large splitting time step, produces results that are more accurate than pure mesoscopic simulations. In the case where all reactions are slow compared to diffusion, the RDME produces accurate results, as does the hybrid method.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
RDME	27.08	3.78	7.19	9.87	4.597	0.81
Hybrid, $\Delta t_{split} = 0.1$	4.82	2.79	2.26	4.52	2.437	0.75
Hybrid, $\Delta t_{split} = 0.01$	0.36	1.90	2.25	0.97	1.157	0.85
Hybrid, $\Delta t_{split} = 0.001$	0.38	1.22	1.15	0.77	0.937	1.09

Then

$$E(h) = \frac{1}{201}\sum_{i=1}^{201}\sum_{S\in\mathcal{S}} |S^*_{h,i} - S^{micro}_i|, \tag{80}$$

where S^{micro}_i is the average population of species S at time t_i, obtained with the microscopic algorithm and where $S^*_{h,i}$ is the average population of species S at time t_i obtained with either the hybrid algorithm or with the NPM, simulated on a mesh with a voxel width of h.

B. Efficiency: Non-linear dependence on the mesh size

As already discussed in Sec. IV C, the total execution time is the sum of the time spent on the mesoscopic scale, T_{meso}, the microscopic scale, T_{micro}, and the overhead incurred from the coupling of the scales. The time spent on the microscopic scale depends on how many of the species are microscopic, which in turn depends on the resolution of the mesh. On a fine mesh, we will simulate fewer molecules on the microscopic scale, and for a shorter time, but we pay the price of a more costly mesoscopic simulation.

In this numerical example, we show that the total execution time T is a non-linear function of T_{micro} and T_{meso} and that to optimize T we need to balance T_{micro} and T_{meso} in a non-trivial way.

We consider the system

$$\mathrm{S}_1 \xrightarrow{k_1^1} \mathrm{S}_{11} + \mathrm{S}_{12} \xrightarrow{k_2^1} \mathrm{S}_2,$$
$$\mathrm{S}_2 \xrightarrow{k_1^2} \mathrm{S}_{21} + \mathrm{S}_{22} \xrightarrow{k_2^2} \mathrm{S}_3,$$
$$\mathrm{S}_3 \xrightarrow{k_1^3} \mathrm{S}_{31} + \mathrm{S}_{32} \xrightarrow{k_2^3} \mathrm{S}_4,$$

where $k_i^1 = 20.0$, $i = 1, 2, 3$, $k_2^1 = 0.0016$, $k_2^2 = 0.001\,45$, and $k_2^3 = 0.0014$. We initialize the system with 200 S_1 molecules and 200 S_2 molecules, all with uniformly sampled positions on the domain. The domain is a sphere with radius 0.5, and we consider a sequence of meshes, ranging from coarse to fine. In Fig. 6, we show that there exists an optimum, with respect to total execution time, between the coarsest and the most resolved mesh. Note that this particular system can be accurately simulated on any mesh resolution with the hybrid method, and therefore the error remains small for all mesh sizes, with the only thing changing being the number of molecules simulated on either scale, and for how long the microscopic molecules remain microscopic.

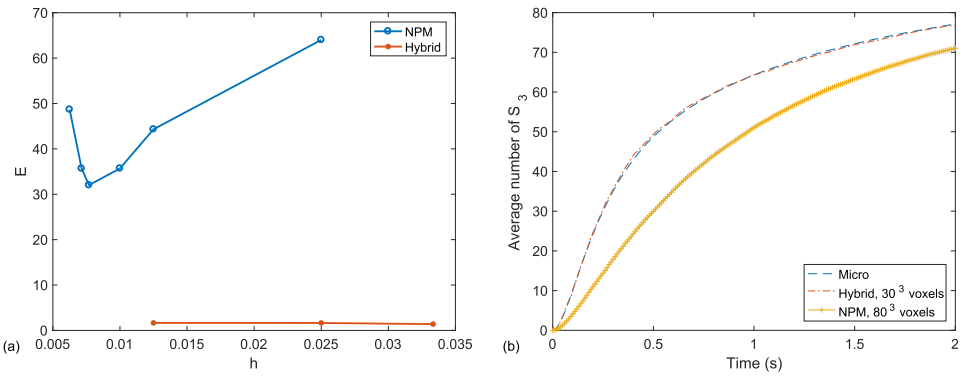

FIG. 5. (a) The error E, as defined by Eq. (80). The RDME does not match the microscopic dynamics for any mesh size. The hybrid method is able to capture the dynamics of the system by simulating the S_1 and S_2 species on the microscopic scale and all other species on the mesoscopic scale. (b) Example trajectory of the average population of S_3 molecules. The hybrid method agrees well with the microscopic results, while the NPM on a mesh of 80^3 voxels does not agree with the microscopic simulations.

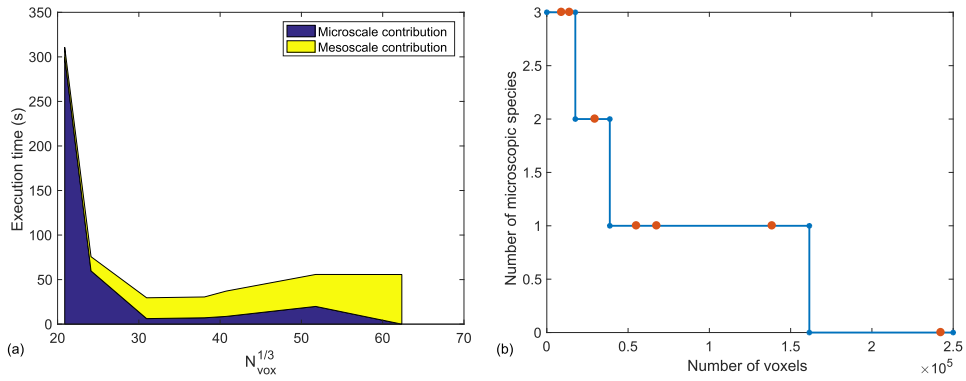

FIG. 6. (a) The total execution time is a nonlinear function of the mesh size. For a coarse mesh, most of the simulation time will be spent on the microscopic scale. As the mesh is successively refined, the time spent on the mesoscopic scale starts to dominate. The shortest total simulation time is obtained for a mesh of around 30 000 voxels. (b) Number of particles on the microscopic scale as a function of the mesh resolution. Red dots indicate the sample points used to generate the plot in (a). As we can see, for the two left-most points, we simulate S_1, S_2, and S_3 on the microscopic scale. As we move to the right, we will simulate two, one, and finally, for the right-most point, no species on the microscopic scale. We can see in (a) that a pure mesoscopic simulation is slower than a simulation with one microscopic particle, but with a much coarser mesh.

In general there is no way to *a priori* determine the optimal mesh size, as it will depend on the system under study as well as the initial condition. It will also depend on the size of the simulation; if we are planning on running many or very long trajectories, making the total simulation time substantial, we can afford an expensive preprocessing step. On the other hand, if the total simulation time is short to moderate, an expensive preprocessing step will not be worthwhile. We therefore propose a heuristic approach to selecting the mesh size.

If we can afford an expensive preprocessing step, we can simulate either full trajectories or shortened trajectories on a sequence of mesh resolutions to find a mesh resolution that appears to minimize the total execution time (there is of course no guarantee that we have found an actual minima). To speed-up this process we could, if the system allows it, perform the simulations on a structured Cartesian grid on a regular domain. That way we avoid the costly process of generating a sequence of unstructured meshes. While there is no guarantee that the system behaves the same way on a structured Cartesian mesh as on the actual domain of interest, we can still expect to get an approximation of the relative cost of simulations on different mesh sizes.

Also note that in many cases the mesh size will be constrained by the geometry of the problem. Internal structures could require a certain minimum mesh resolution, meaning that we cannot select the mesh size that optimizes the execution time but that we instead are constrained to a certain mesh and have to choose the best splitting given the mesh resolution.

VI. CONCLUSIONS

We have developed a hybrid method coupling simulation of the mesoscopic and microscopic modeling scales. The method can, for a certain class of systems, automatically propose a splitting of species based on how diffusion-limited the reactions are. Furthermore, we show that the new method converges with decreasing splitting time step for a larger class of systems than a previously developed method.[31]

We apply the method to a numerical example, showing that it accurately, and with increased efficiency compared to microscopic simulations, splits the system. We also show how the optimal splitting can be found for a mesh between the coarsest and the finest possible resolutions. It is therefore necessary to find a balance between how many molecules to simulate on the microscopic scale and how fine the mesh should be.

The approach described in this paper can, in general, be applied to systems where molecules are created in spatial proximity through some sequence of unimolecular and bimolecular reactions. Another possibility is that molecules are created in spatial proximity due to more complex interactions with internal membranes or fibers, processes not necessarily captured by the scheme outlined above. It is also plausible that microscale resolution could be needed for other reasons, such as for processes where molecules in 3D react with complex membranes or move due to active transport. Automatic splitting of such systems would require a different analysis.

ACKNOWLEDGMENTS

This work has been funded by National Institutes of Health (NIH) NIBIB Award No. R01-EB014877, Department of Energy (DOE) Award No. DE-SC0008975, the Swedish Research Council under Award No. 2015-03964, and the eSSENCE strategic collaboration on eScience.

APPENDIX: SOFTWARE AND DATA AVAILABILITY

The source code used to generate the data and figures in this article is available for download at http://shellander.bitbucket.io/rdsim.

[1] A. Mahmutovic, D. Fange, O. G. Berg, and J. Elf, Nat. Methods **9**, 1163 (2012).
[2] M. Sturrock, A. Hellander, S. Aldakheel, L. Petzold, and M. Chaplain, Bull. Math. Biol. **76**, 766 (2013).
[3] M. J. Lawson, B. Drawert, M. Khammash, and L. Petzold, PLoS Comput. Biol. **9**, e1003139 (2013).
[4] D. Fange and J. Elf, PLoS Comput. Biol. **2**, e80 (2006).
[5] R. A. Kerr, T. M. Bartol, B. Kaminsky, M. Dittrich, J.-C. J. Chang, S. B. Baden, T. J. Sejnowski, and J. R. Stiles, SIAM J. Sci. Comput. **30**, 3126 (2008).
[6] S. S. Andrews and D. Bray, Phys. Biol. **1**, 137 (2004).

[7]M. Tomita, K. Hashimoto, K. Takahashi, T. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J. Craig, V. Clyde, and A. Hutchison, Bioinformatics **15**, 72 (1999).
[8]J. Hattne, D. Fange, and J. Elf, Bioinformatics **21**, 2923 (2005).
[9]J. Schaff, C. C. Fink, B. Slepchenko, J. H. Carson, and L. M. Loew, Biophys. J. **73**, 1135 (1997).
[10]I. Hepburn, W. Chen, S. Wils, and E. D. Schutter, BMC Syst. Biol. **6**, 425 (2012).
[11]R. F. Oliveira, A. Terrin, G. D. Benedetto, R. C. Cannon, W. Koh, M. Kim, M. Zaccolo, and K. T. Blackwell, PLoS One **4**, e11725 (2010).
[12]J. Schöneberg and F. Noé, PLoS One **8**, e74261 (2013).
[13]B. Drawert, S. Engblom, and A. Hellander, BMC Syst. Biol. **6**, 76 (2012).
[14]B. Drawert, M. Trogdon, S. Toor, L. Petzold, and A. Hellander, SIAM J. Sci. Comput. **38**, C179 (2016).
[15]B. Drawert, A. Hellander, B. Bales, D. Banerjee, G. Bellesia, B. J. Daigle, Jr., G. Douglas, M. Gu, A. Gupta, S. Hellander, C. Horuk, D. Nath, A. Takkar, S. Wu, P. Lötstedt, C. Krintz, and L. R. Petzold, PLoS Comput. Biol. **12**, 1 (2016).
[16]K. Takahashi, S. Tănase-Nicola, and P. R. ten Wolde, Proc. Natl. Acad. Sci. U. S. A. **107**, 2473 (2010).
[17]D. Fange, O. G. Berg, P. Sjöberg, and J. Elf, Proc. Natl. Acad. Sci. U. S. A. **107**, 19820 (2010).
[18]S. A. Isaacson, SIAM J. Appl. Math. **70**, 77 (2009).
[19]S. Hellander, A. Hellander, and L. R. Petzold, Phys. Rev. E **91**, 023312 (2015).
[20]S. Hellander and L. R. Petzold, Phys. Rev. E **93**, 013307 (2016).
[21]S. Hellander, A. Hellander, and L. R. Petzold, Phys. Rev. E **85**, 042901 (2012).
[22]R. Erban and J. Chapman, Phys. Biol. **6**, 046001 (2009).
[23]M. Doi, J. Phys. A: Math. Gen. **9**, 1465 (1976).
[24]S. Isaacson, J. Chem. Phys. **139**, 054101 (2013).
[25]L. Ferm, A. Hellander, and P. Lötstedt, J. Comput. Phys. **229**, 343 (2010).
[26]M. Flegg, J. Chapman, and R. Erban, J. R. Soc., Interface **9**, 859 (2012).
[27]F. Spill, P. Guerrero, T. Alarcon, P. K. Maini, and H. Byrne, J. Comput. Phys. **299**, 429 (2015).
[28]W.-C. Lo, L. Zheng, and Q. Nie, R. Soc. Open Sci. **3**, 160485 (2016).
[29]C. A. Yates and M. B. Flegg, J. R. Soc., Interface **12** (2015).
[30]B. Franz, M. B. Flegg, S. J. Chapman, and R. Erban, SIAM J. Appl. Math. **73**, 1224 (2013).
[31]A. Hellander, S. Hellander, and P. Lötstedt, Multiscale Model. Simul. **10**, 585 (2012).
[32]M. B. Flegg, S. Hellander, and R. Erban, J. Comput. Phys. **289**, 1 (2015).
[33]M. Robinson, S. S. Andrews, and R. Erban, Bioinformatics **31**, 2406 (2015).
[34]S. Engblom, L. Ferm, A. Hellander, and P. Lötstedt, J. Sci. Comput. **31**, 1774 (2009).
[35]J. Elf and M. Ehrenberg, Syst. Biol. **1**, 230 (2004).
[36]J. S. van Zon and P. R. ten Wolde, J. Chem. Phys. **123**, 234910 (2005).
[37]H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids* (Oxford University Press, 1959).
[38]S. Hellander and P. Lötstedt, J. Comput. Phys. **230**, 3948 (2011).
[39]F. C. Collins and G. E. Kimball, J. Colloid Sci. **4**, 425 (1949).
[40]D. T. Gillespie, J. Chem. Phys. **131**, 164109 (2009).
[41]E. W. Montroll and G. H. Weiss, J. Math. Phys. **6**, 167 (1965).
[42]E. W. Montroll, J. Math. Phys. **10**, 753 (1969).
[43]H. Kim and K. J. Shin, Phys. Rev. Lett. **82**, 1578 (1999).
[44]S. Hellander and L. Petzold, J. Chem. Phys. **146**, 064101 (2017).