FLEXIBLE SINGLE MOLECULE SIMULATION OF REACTION-DIFFUSION PROCESSES

STEFAN HELLANDER[1], PER LÖTSTEDT[1] *

May 14, 2010

[1]*Division of Scientific Computing, Department of Information Technology*
Uppsala University, P. O. Box 337, SE-75105 Uppsala, Sweden
emails: `stefan.hellander@it.uu.se, perl@it.uu.se`

Abstract

An algorithm is developed for simulation of the motion and reactions of single molecules at a microscopic level. The molecules diffuse in a solvent and react with each other or a polymer and molecules can dissociate. Such simulations are of interest e.g. in molecular biology. The algorithm is similar to the Green's function reaction dynamics (GFRD) algorithm by van Zon and ten Wolde where longer time steps can be taken by computing the probability density functions (PDFs) and then sample from its distribution function. Our computation of the PDFs is much less complicated than GFRD and more flexible. The solution of the partial differential equation for the PDF is split into two steps to simplify the calculations. The sampling is without splitting error in two of the coordinate directions for a pair of molecules and a molecule-polymer interaction and is approximate in the third direction. The PDF is obtained either from an analytical solution or a numerical discretization. The errors due to the operator splitting, the partitioning of the system, and the numerical approximations are analyzed. The method is applied to three different systems involving up to four reactions. Comparisons with other mesoscopic and macroscopic models show excellent agreement.

Keywords: single molecule, chemical reactions, diffusion, operator splitting.

AMS subject classification: 65C05, 65C35, 82C80, 92C40.
PACS subject classification: 02.50.Ey, 02.70.Uu, 83.10.Rs, 87.10.Rt.

Abbreviations

CDF	cumulative distribution function
CME	chemical master equation
GFRD	Green's function reaction dynamics
ODE	ordinary differential equation
PDE	partial differential equation
PDF	probability density function
RDME	reaction-diffusion master equation
SDE	stochastic differential equation
SSA	stochastic simulation algorithm

1   Introduction

It is well established and generally accepted that the discrete nature of chemical reactions requires stochastic modeling when low numbers of molecules are involved in biological cells (25). The governing equation at a *mesoscopic* level is the chemical master equation (CME) for the probability density function (PDF) of the state of the system (17; 21). The underlying assumption is then that the system is well stirred without spatial gradients of the concentrations of the chemical species. This assumption is not always valid and then spatial effects have to be taken into account (12; 16) and the model equation is the reaction-diffusion master equation (RDME) (17; 21). The domain is partitioned into voxels or compartments and the state of the system is the number of molecules of every species in each one of the voxels.

In a mesoscopic description, there is no upper bound on the copy number of the molecules in a voxel and the path of separate molecules is not followed. Molecules diffuse between voxels but only the copy numbers of the different molecules in the voxels are recorded. Trajectories of these systems are usually generated by the stochastic simulation algorithm (SSA) by Gillespie (18) and improvements of it. The *macroscopic* mean field equations in a well stirred domain are the reaction rate equations, a system of nonlinear ordinary differential equations (ODEs), and in a case with spatial variation the reaction-diffusion equations, a system of partial differential equations (PDEs).

Sometimes it is necessary to follow the motion and reactions of single molecules. One reason for a more expensive *microscopic* modeling can be that a detailed model of a complex behavior is required. For example, no accurate mesoscopic description is known for crowding and a particle model is the only alternative (4; 29). Another reason is that the mesoscopic SSA is valid only when its voxel sizes are of the same order as the average distance traveled by a particle between two reactions or the reactive mean free path (3). Hence, sharp gradients in the

concentrations may not be sufficiently resolved on a mesoscopic mesh (20; 26). Furthermore, the assumption that the molecules are well stirred in a voxel is not satisfied for low concentrations with few molecules. Rare events are then not well modeled by the RDME. An example is gene expression with time and length scales differing by orders of magnitude (40). Corrections to the RDME are possible (15; 31) to improve the modeling for small voxel volumes but there are cases where a microscopic model is the only alternative.

In a microscopic model, each molecule is tracked and when they are in the vicinity of each other they may react and form a new compound. Either the molecules diffuse as in Brownian motion or jump with some probability between discrete points or vertices in a lattice. The diffusion is caused by collisions with small solvent molecules. The discretized space in the RDME can be viewed as a lattice with the center of the voxel at the lattice vertex. The difference between RDME and a lattice model is that only one molecule may occupy one lattice point thus allowing for crowding effects. In Brownian dynamics, the molecules move according to a time discretization of a stochastic differential equation (SDE). The simulation is then usually *time-driven* in the sense that the state is advanced one small time step Δt and then possible reactions are processed at the end of the step. These two approaches are analyzed with different boundary conditions in (14).

There is plenty of software implementing these algorithms for applications in molecular biology. Freely moving molecules react with membranes in MCell (6). In Smoldyn (2), molecules diffuse by Brownian motion as in MCell and reactions occur when they are sufficiently close to each other. Large molecules are tracked in space and small and abundant molecules move in a spatial lattice in Cell++ (30). The molecules jump to adjacent vertices in a lattice and react with each other in GridCell (4) including crowding effects. Space is discretized in voxels in a generalization of the SSA algorithm (18) to include diffusion on a Cartesian mesh in (35), in SmartCell (1), and in an efficient implementation MesoRD (19). An unstructured mesh (or a generalized lattice) is used in URDME (8; 13). Some of the methods and the implementations are compared in (9) and reviewed in (36).

The simulation based on SSA is *event-driven*. The molecular state of the system is constant in time intervals Δt and is updated after an event such as a reaction or a jump to a neighboring voxel. These time steps are in general much longer than the steps in a time-driven simulation. A general review of time-driven and event-driven molecular simulations is found in (10).

An event-driven algorithm for diffusing and reacting molecules is Green's function reaction dynamics (GFRD) by van Zon and ten Wolde (41; 42). Weak interaction such as electrostatic forces and the shape of the molecules are ignored. The full multi-body problem is decomposed into a collection of single- and two-body problems. The molecules collide and react and dissociate with a PDF following Smoluchowski's theory (33; 41). Free molecules diffuse according to Brownian motion and an SDE. A new position after a predetermined time step

3

is sampled from the cumulative distribution function (CDF) corresponding to the PDF. Much longer time steps are allowed at the cost of computing the PDF and the CDF.

The PDF solution satisfies a Fokker-Planck equation or a Kolmogorov forward equation with boundary conditions for colliding molecules by Collins and Kimball (7). Analytical solutions are known in special cases but can be difficult to evaluate since they often involve infinite summation and integration of special functions. An algorithm similar to GFRD is developed in (11; 27) where single molecules are surrounded by protective regions and the exit time from these regions determine the time steps in the method. Pairs of molecules are simulated using small time steps and Brownian motion.

In this paper, we are interested in the dynamics of biological systems with relatively few molecules. A mesoscopic model is much more efficient for large copy numbers. We adopt the GFRD approach of determining a PDF for the new position of the molecule. The computation is simplified here by splitting the spatial operator of the governing Smoluchowski (or Fokker-Planck) equation. The dimension of the subproblems is then further reduced after splitting with simpler PDEs to solve. More flexibility in the modeling is another advantage of the operator splitting. Either analytical solutions are used or the equations are solved numerically such that reaction models can be simulated where analytical solutions are unknown. Hard spherical molecules react with each other or with a long molecular chain or polymer modeled as a hard cylinder representing e.g. a DNA strand. The reactions are irreversible or reversible. The subsystems consisting of a single molecule or a pair of molecules are advanced in time asynchronously. Between time points where the system is synchronized, one subsystem may use a different sequence of time steps than another subsystem.

By computing the PDF, longer time steps are possible where a number of association-dissociation events may have occurred. Large tables are avoided making the algorithm compute-intense and memory-lean and more suitable for modern processor architectures. The simplification of the PDF of the system, the operator splitting, the evaluation of the analytical solutions, and the numerical solution of the Smoluchowski equations all introduce approximation errors. These errors depend on the discretization parameters and are analyzed in the paper. By adjusting the numerical parameters, the errors can be controled to be of the same order as or smaller than the modeling errors (e.g. proteins are not hard spheres or cylinders). We are willing to accept larger numerical errors and gain more flexibility in the modeling and implementation of the method. The influence of those errors may not be so important. This is the case in molecular dynamics simulations. Even though the modeling and numerical errors measured in a traditional way are large, statistical quantities such as expectations are determined accurately (32).

The contents of the sections are organized as follows. The Smoluchowski equations are presented in Section 2. The algorithm is described in Section 3. The

induced approximation errors are analyzed in Section 4. Systems with irreversible and reversible reactions and up to five reacting species are simulated in Section 5 and compared to solutions obtained with the macroscopic reaction rate equations and mesoscopic SSA simulations. The agreement between different models is very good.

2 Stochastic model

We consider a system of N molecules initially. The molecules move in space by diffusion, react with each other, and dissociate into two products. There is a PDF for the positions of the molecules at time t given the positions at time t_n but the PDE satisfied by the PDF is impossible to solve analytically for $N > 2$. Instead, the system is decomposed into subsystems as in (11; 41) with one or two molecules. Each subsystem can be advanced in time independently of the other subsystems if the single molecules or pairs of molecules are sufficiently well separated from each other. Much longer time steps can be taken compared to time-driven Brownian dynamics. The PDF satisfies a Smoluchowski equation. By sampling the distribution associated with the PDF, the next locations of the molecules at $t > t_n$ are determined.

2.1 A single molecule

Assume that a molecule is at position $\mathbf{x}_n$ at time t_n. Then the PDF $p_1(\mathbf{x}, t | \mathbf{x}_n, t_n)$ for its position $\mathbf{x}$ at time $t > t_n$ satisfies the diffusion equation

$$\partial_t p_1 = D \Delta_{\mathbf{x}} p_1 \tag{2.1}$$

with the diffusion coefficient D. The time derivative is denoted by ∂_t. The Laplace operator $\Delta_{\mathbf{x}}$ operates in the $\mathbf{x}$-coordinates. The initial condition is the Dirac delta function and the solution vanishes at infinity

$$p_1(\mathbf{x}, t_n | \mathbf{x}_n, t_n) = \delta(\mathbf{x} - \mathbf{x}_n), \ \lim_{\mathbf{x} \to \infty} p_1(\mathbf{x}, t | \mathbf{x}_n, t_n) = 0. \tag{2.2}$$

The solution is the well known Green's function for the heat equation

$$p_1(\mathbf{x}, t | \mathbf{x}_n, t_n) = \frac{1}{(4\pi D(t - t_n))^{3/2}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_n\|^2}{4D(t - t_n)}\right), \tag{2.3}$$

where $\| \cdot \|$ is the Euclidean vector norm.

2.2 A pair of molecules

The Smoluchowski equation for the PDF $p_2(\mathbf{x}_1, \mathbf{x}_2, t | \mathbf{x}_{1n}, \mathbf{x}_{2n}, t_n)$ of two molecules at $\mathbf{x}_1$ and $\mathbf{x}_2$ at t which were at $\mathbf{x}_{1n}$ and $\mathbf{x}_{2n}$ at t_n is

$$\partial_t p_2 = D_1 \Delta_{\mathbf{x}_1} p_2 + D_2 \Delta_{\mathbf{x}_2} p_2. \tag{2.4}$$

The molecules diffuse with different coefficients D_1 and D_2. Two new coordinates are introduced in (11; 41)

$$\mathbf{Y} = \sqrt{D_2/D_1}\mathbf{x}_1 + \sqrt{D_1/D_2}\mathbf{x}_2, \ \mathbf{y} = \mathbf{x}_2 - \mathbf{x}_1. \tag{2.5}$$

The diffusion in $\mathbf{Y}$ and $\mathbf{y}$ is independent of each other and p_2 in (2.4) can be written

$$p_2(\mathbf{x}_1, \mathbf{x}_2, t | \mathbf{x}_{1n}, \mathbf{x}_{2n}, t_n) = p_{\mathbf{Y}}(\mathbf{Y}, t | \mathbf{Y}_n, t_n) p_{\mathbf{y}}(\mathbf{y}, t | \mathbf{y}_n, t_n), \tag{2.6}$$

where $\mathbf{Y}_n$ and $\mathbf{y}_n$ are defined by $\mathbf{x}_{1n}, \mathbf{x}_{2n}$, and (2.5). With the diffusion coefficient $D_{12} = D_1 + D_2$, the PDFs $p_{\mathbf{Y}}(\mathbf{Y}, t | \mathbf{Y}_n, t_n)$ and $p_{\mathbf{y}}(\mathbf{y}, t | \mathbf{y}_n, t_n)$ solve the PDEs

$$\partial_t p_{\mathbf{Y}} = D_{12} \Delta_{\mathbf{Y}} p_{\mathbf{Y}}, \quad \partial_t p_{\mathbf{y}} = D_{12} \Delta_{\mathbf{y}} p_{\mathbf{y}}. \tag{2.7}$$

The motion in the $\mathbf{Y}$-coordinate is in free space and the initial condition and condition at ∞ are as in (2.2) with the solution in (2.3).

In an irreversible reaction $A_1 + A_2 \to B$ between species A_1 and A_2 producing B, there is an additional boundary condition at the reaction radius $r = \|\mathbf{y}\| = \sigma$. The molecules are spherical with diameters d_1 and d_2 and $\sigma = (d_1 + d_2)/2$. When the distance between the molecules is σ, they may react with each other or not depending on the association coefficient k_a. In a spherical coordinate system $\mathbf{r}^T = (r, \theta, \phi)$, we have the diffusion equation for $p_{\mathbf{r}}(\mathbf{r}, t | \mathbf{r}_n, t_n)$

$$\partial_t p_{\mathbf{r}} = D_{12} \left(\frac{\partial^2 p_{\mathbf{r}}}{\partial r^2} + \frac{2}{r} \frac{\partial p_{\mathbf{r}}}{\partial r} + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial p_{\mathbf{r}}}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 p_{\mathbf{r}}}{\partial \phi^2} \right), \tag{2.8}$$

with the boundary condition at $r = \sigma$

$$4\pi \sigma^2 D_{12} \frac{\partial p_{\mathbf{r}}}{\partial r}\Big|_{r=\sigma} = k_a p_{\mathbf{r}}(r = \sigma, t | \mathbf{r}_n, t_n). \tag{2.9}$$

In the far field,

$$\lim_{r \to \infty} p_{\mathbf{r}}(r, \theta, \phi, t | \mathbf{r}_n, t_n) = 0, \tag{2.10}$$

and the initial condition is $p_{\mathbf{r}}(\mathbf{r}, t | \mathbf{r}_n, t_n) = \delta(\mathbf{r} - \mathbf{r}_n)$ as in (2.2).

When k_a vanishes then we have a reflection or Neumann boundary condition, $\partial p_{\mathbf{r}}/\partial r|_{r=\sigma} = 0$, and the molecules do not react with each other. When $k_a \to \infty$ the boundary is absorbing, $p_{\mathbf{r}}(r = \sigma, t | \mathbf{r}_n, t_n) = 0$, and a new chemical species B is always formed.

The analytical solution of (2.8), (2.9), and (2.10) is an infinite sum of terms consisting of a product of a Legendre function and an infinite integral of a complicated expression in Bessel functions of the first and second kind and their derivatives (5, p. 382), (41).

If the reaction is reversible, $A_1 + A_2 \rightleftarrows B$, then there is a probability for the B-molecule to dissociate into A_1 and A_2 depending on the dissociation rate k_d. The boundary condition at $r = \sigma$ in (2.9) for a reversible geminate dissociation is modified to

$$4\pi\sigma^2 D_{12} \frac{\partial p_{\mathbf{r}}}{\partial r}|_{r=\sigma} = k_a p_{\mathbf{r}}(r = \sigma, t|\mathbf{r}_n, t_n) - k_d(1 - S(t|\mathbf{r}_n, t_n)), \qquad (2.11)$$

see (22; 23). The survival probability S of the compound molecule B given that A_1 and A_2 are separated by $\mathbf{r}_n$ at t_n in (2.11) is

$$S(t|r_n, t_n) = 1 - \int_{t_n}^{t} 4\pi\sigma^2 D_{12} \frac{\partial p_{\mathbf{r}}}{\partial r}|_{r=\sigma} \, d\tau. \qquad (2.12)$$

A solution $p(r, t|r_n, t_n)$ of (2.8), (2.11), and (2.12) independent of θ and ϕ is derived in (22). A system of three nonlinear equations is solved for coefficients in the arguments of the exponential and the complementary error function erfc in a sum with five terms. The analytical solution of (2.8), (2.9), and (2.10) is simplified considerably if the solution is independent of θ and ϕ and $k_d = 0$, see (22, (17)) and (5, p. 368). Other models for irreversible reactions are found in (28).

2.3   A molecule and a polymer

The interaction between a molecule A and a stationary polymer P can be modeled in the same manner. Introduce a cylindrical coordinate system $\mathbf{r}^T = (r, \phi, z)$. The polymer is situated at $r = 0$ along the z-axis. An irreversible reaction $A + P \to P$ where A is absorbed by P takes place with a probability depending on k_c at the reaction radius $r = \sigma$. Then the PDF $p_{\mathbf{r}}$ of a free molecule A in space $r > \sigma$ diffusing with coefficient D satisfies

$$\partial_t p_{\mathbf{r}} = D \left(\frac{\partial^2 p_{\mathbf{r}}}{\partial r^2} + \frac{1}{r} \frac{\partial p_{\mathbf{r}}}{\partial r} + \frac{1}{r^2} \frac{\partial^2 p_{\mathbf{r}}}{\partial \phi^2} + \frac{\partial^2 p_{\mathbf{r}}}{\partial z^2} \right), \qquad (2.13)$$

with the boundary condition at $r = \sigma$

$$4\pi\sigma^2 D \frac{\partial p_{\mathbf{r}}}{\partial r}|_{r=\sigma} = k_c p_{\mathbf{r}}(r = \sigma, t|\mathbf{r}_n, t_n). \qquad (2.14)$$

The analytical solution of (2.13) and (2.14) is not known to us but with $p_{\mathbf{r}}$ independent of ϕ and z a solution is derived in (5, p. 370). It is still fairly intricate. Let J_0 and J_1 be Bessel functions of the first kind and Y_0 and Y_1 Bessel functions of the second kind. The definition of the function C is

$$C(u, r, k, h) = \frac{J_0(ur)(kuY_1(\sigma u) + hY_0(\sigma u)) - Y_0(ur)(kuJ_1(\sigma u) + hJ_0(\sigma u))}{((kuY_1(\sigma u) + hY_0(\sigma u))^2 + (kuJ_1(\sigma u) + hJ_0(\sigma u))^2)^{1/2}}.$$

$$(2.15)$$

Then the PDF is

$$p_r(r, t|r_0, t_n) = \frac{1}{2\pi} \int_0^\infty \exp(-Du^2(t - t_n))C(u, r, k, k_c)C(u, r_0, k, k_c)u \, du. \quad (2.16)$$

with $k = 4\pi\sigma^2 D$.

3   Algorithm

A trajectory of the system is advanced in time by time steps $\Delta\tau$ from t_n to $t_n + \Delta\tau$. We start each time step by dividing the system into subsystems consisting of a single molecule or a pair of molecules. The partitioning is done in the following way:

- To each molecule we assign a nearest neighbor and a second nearest neighbor.

- Two molecules that are nearest neighbors of each other will constitute a subsystem consisting of a pair of two molecules.

- Molecules that cannot be partitioned into pairs under the above category will be subsystems of single molecules.

For each subsystem i we determine a time interval of length $\Delta\tau_i$ during which the probability for interaction with any other subsystem is small. This $\Delta\tau_i$ is computed using the distances to the nearest neighbors of the subsystems. Then let $\Delta\tau = \min_i \Delta\tau_i$. In Section 4.3, we specify the details of how $\Delta\tau_i$ is computed.

The GFRD algorithm (41) is based on an event-driven time stepping. A time step is chosen as the minimum of the time to the next reaction and the $\Delta\tau$ computed above.

We take on a slightly different approach. Each subsystem is simulated independently in $[t_n, t_n + \Delta\tau]$ but they are synchronized at $t_{n+1} = t_n + \Delta t$. In this interval, many reversible reactions can occur in each subsystem. Furthermore, we do not have to necessarily simulate each reaction explicitly. Instead, certain fast associations and dissociations can be skipped by including both the association and the dissociation in the governing equations for the PDF (2.8) with boundary condition (2.11). An example is a system i with a reversible reaction

$$A + B \rightleftharpoons C. \quad (3.1)$$
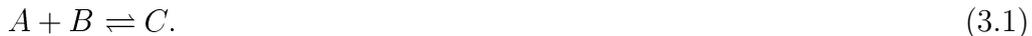
Many associative and dissociative reactions can occur during a time step $\Delta\tau_i$ and there is no need to let $\Delta\tau_i$ be the time to the first reaction.

The presence of an outer boundary must be taken into account. The computational domain is a cube with reflecting boundary conditions in the examples. Since no analytical results are available for pairs of particles and a boundary, we must choose time steps for each subsystem such that the probability for a particle to react with both the boundary and another particle during one time step is small.

The polymer is modeled by a cylinder along the z-axis and with its center at the origin of the xy-plane. The cylinder is implemented as an inner boundary with the boundary condition given by (2.14).

3.1 Simulating a subsystem with asynchronous time stepping

A subsystem will consist of either one or two molecules. It would be possible to start by partitioning the system into larger subsystems and then apply the algorithm recursively to each subsystem. This could be more efficient for large systems.

3.1.1 Subsystem of one molecule

There are two cases depending on whether the molecule can dissociate or not.

- If it cannot dissociate, then update the position according to free diffusion. If it is outside the domain, then reflect it back into the domain.

- If it can dissociate, then there are two possibilities. If it is involved in a single reversible reaction (3.1), update the position and the state according to Section 3.2.1. If the reactions are more complicated we may have to sample a time for the next reaction and update the subsystem to this time. The next step is to apply the algorithm to the subsystem with the new state.

3.1.2 Subsystem with two molecules

The first step is to check the distance to the inner boundary at the cylinder and the outer boundary of the computational domain. Since pairs of molecules close to a boundary cannot be updated using an easily computed PDF, a time step $\Delta\tau_b$ is determined such that the molecules can react only with each other or with the boundary but not both. The molecules may also dissociate. If either one or both of them can, a time $\Delta\tau_d$ is sampled to the next dissociation. We take the minimum of $\Delta\tau$, $\Delta\tau_b$ and $\Delta\tau_d$ and update the subsystem accordingly.

As an example, take the chemical system

$$
\begin{aligned}
C &\rightarrow A + B, \\
C + D &\rightarrow E.
\end{aligned}
\tag{3.2}
$$

The molecule C in a pair with molecule D may dissociate or react with D to form E at a time $t < t_n + \Delta\tau$. If the reactions are as in (3.1) in a pair with one A and one B molecule, then the state and location of the molecules are given at $t_n + \Delta\tau$ by the PDF in (2.8) and (2.11).

If the minimum is $\Delta\tau_b$ there are two alternatives.

- If the molecules are closer to the boundary than to each other, the position of each molecule is updated using the boundary condition for the outer boundary or the cylinder. In the latter case the coordinates are determined as described in Section 3.2.2.

- If the molecules are closer to each other than to the boundary, their positions are updated as in Section 3.2.1.

If the minimum is $\Delta\tau_d$ and if the molecules can react as in (3.2), then a time $\Delta\tau_a$ is generated to the next association. Again we have two cases.

- If $\Delta\tau_a$ is smaller than $\Delta\tau_d$, the subsystem is updated to $t_n + \Delta\tau_a$. We generate a position for the new molecule according to $p_\mathbf{Y}$ with $\mathbf{Y}$ defined as in (2.5). With a $C - D$ pair in (3.2), the reaction $C + D \rightarrow E$ fires at $t_n + \Delta\tau_a$ before $C \rightarrow A + B$ at $t_n + \Delta\tau_d$.

- If $\Delta\tau_d$ is smaller than $\Delta\tau_a$ the pair is updated as in Section 3.2.1 but now given that the particles do not react before $t_n + \Delta\tau_d$. At $t_n + \Delta\tau_d$ a dissociation will occur and in (3.2) C will split into A and B, $C \rightarrow A + B$.

The algorithm is applied repeatedly until time $t_n + \Delta\tau$ is reached. In practice, there is a small minimum time step to guarantee that we do not get stuck simulating subsystems where many molecules are close to each other or the boundaries. Since τ_b, τ_d, and τ_a are different for different pairs of molecules, the subsystems use different asynchronous time stepping to $t_n + \Delta\tau$.

3.2   Operator splitting

The analytical solution of (2.8), (2.9), and (2.10) is difficult to compute and if the reaction is reversible then the solution of (2.8), (2.10), and (2.11) is not known. The CDF of the solution in the irreversible case is tabulated in (41) in a large four-dimensional array in the parameters r, θ, r_n, and $t - t_n$. There is one table for every type of irreversible reaction in the system defined by the parameters σ, k_a, and D_{12}.

An often used technique to solve complicated time-dependent PDEs numerically is the operator splitting or fractional step method (24; 34). It is suitable when the original equation can be split into a number of substeps whose solution is very much simpler than the solution of the full equation. Firstly, a time step

Δt is chosen. Secondly, the diffusion terms on the right hand side of (2.8), and (2.13) are integrated one after another Δt forward in time in a sequence. Then the solution is advanced in substeps in time again by Δt until the final time is reached. The relation between $\Delta \tau$ and Δt is $\Delta \tau = m \Delta t$ with an integer m. The numerical results in Section 5 are obtained with $\Delta \tau = \Delta t$. A similar idea is the Alternating Direction Implicit (ADI) method for iterative solution of steady state equations (37).

A second order accurate technique is Strang splitting (34). If the diffusion consists of two terms, then the solution is advanced half a time step $\Delta t/2$ with the first term. Then in the second part, the solution is integrated a full time step Δt with the second term. Finally, half a time step is taken again with the first term.

3.2.1 A pair of molecules

Apply the operator splitting to (2.8) and rotate the coordinate system such that the north pole is at $\mathbf{r}_n^T = (r_n, \theta_n, \phi_n)$ and let $t_{n+1} = t_n + \Delta t$. The procedure starting at $\mathbf{r}_n$ and time t_n is:

1. Solve in the radial direction for $p_r(r, t | r_n, t_n), r \geq \sigma, t \geq t_n$, in

$$\partial_t p_r = D_{12} \left(\frac{\partial^2 p_r}{\partial r^2} + \frac{2}{r} \frac{\partial p_r}{\partial r} \right) \equiv \mathcal{A}_{r2}(r) p_r, \tag{3.3}$$

with initial condition $p_r(r, t_n | r_n, t_n) = \delta(r - r_n)$ and boundary condition

$$4\pi \sigma^2 D_{12} \frac{\partial p_r}{\partial r} |_{r=\sigma} = k_a p_r(\sigma, t | r_n, t_n) - k_d(1 - S(t | r_n, t_n)). \tag{3.4}$$

Sample the distribution at t_{n+1} with the PDF p_r to obtain r_{n+1}.

2. Solve in the polar and azimuthal directions for $p_\theta(\theta, t | r, 0, \phi_n, t_n), \theta \in [0, \pi], \phi \in [0, 2\pi)$, in

$$\partial_t p_\theta = \frac{D_{12}}{r^2} \left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial p_\theta}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 p_\theta}{\partial^2 \phi} \right) \equiv \mathcal{A}_\theta(r) p_\theta, \tag{3.5}$$

with initial condition $p_\theta(\theta, \phi, t_n | r, 0, \phi_n, t_n) = \delta(\theta)$. The solution of (3.5) is independent of ϕ

$$p_\theta(\theta, t | r, 0, \phi_n, t_n) = \sum_{\ell=0}^{\infty} \sqrt{\frac{2\ell + 1}{4\pi}} \exp(-\ell(\ell + 1) D_{12}(t - t_n)/r^2) P_\ell(\cos \theta),$$

$$\tag{3.6}$$

where P_ℓ denotes the Legendre function. The solution in (3.6) is independent of ϕ and ϕ'_{n+1} is sampled from the uniform distribution in $[0, 2\pi)$.

Sample the distribution at t_{n+1} and r_{n+1} with the PDF p_θ to obtain θ'_{n+1}. Then rotate the sphere such that the north pole coincides with θ_n to obtain θ_{n+1} and ϕ_{n+1}.

The new relative position of the molecules at t_{n+1} is $(r_{n+1}, \theta_{n+1}, \phi_{n+1})$.

The exact solution $p_\mathbf{r}$ of (2.8) at t_{n+1} can be written

$$
\begin{aligned}
&p_\mathbf{r}(r_{n+1}, \theta_{n+1}, \phi_{n+1}, t_{n+1} | r_n, \theta_n, \phi_n, t_n) \\
&\quad = p_{\mathbf{r}|r}(\theta_{n+1}, \phi_{n+1}, t_{n+1} | r_{n+1}, r_n, \theta_n, \phi_n, t_n) p_r(r_{n+1}, t_{n+1} | r_n, t_n),
\end{aligned}
\tag{3.7}
$$

where p_r solves (3.3). The operator splitting approximation $p_{\mathbf{r}, ops}$ of $p_\mathbf{r}$ is

$$
\begin{aligned}
&p_{\mathbf{r}, ops}(r_{n+1}, \theta_{n+1}, \phi_{n+1}, t_{n+1} | r_n, \theta_n, \phi_n, t_n) \\
&\quad = p_\theta(\theta_{n+1}, t_{n+1} | r_{n+1}, \theta_n, \phi_n, t_n) p_r(r_{n+1}, t_{n+1} | r_n, t_n),
\end{aligned}
\tag{3.8}
$$

where p_θ is the solution of (3.5) and p_r is the same as in (3.7). The sampling of r and ϕ is exact. Instead of sampling the CDF of $p_{\mathbf{r}|r}$ for θ_{n+1}, we sample the CDF of the approximation p_θ. The error due to the operator splitting is analyzed in the next section.

For the reversible reaction, there is an analytical solution of (3.3) and (3.4) derived in (23). These solutions are integrated exactly, and thus there is an analytical expression for the necessary CDF. These expressions involve the complex-valued error function which is computed according to the algorithm in (39). The CDF can be computed fairly efficiently and instead of using tables the CDF is computed in each step. This allows us to have large systems and change values of the parameters often without running into memory issues related to the tables.

The solution (3.6) depends on two parameters: the shifted and scaled time $\hat{t} = D_{12}(t - t_n)/r^2$ and the polar angle θ. A two dimensional table can be precomputed for the CDF in θ and $\hat{t}$. This table is the same for all reactions.

With Strang splitting (34), a reduced splitting error is achieved. The procedure is:

1. Solve (3.3) in the radial direction for $p_r(r, t | r_n, t_n)$ with the same initial and boundary conditions as in (3.3) and (3.4) and sample the distribution defined by p_r to obtain r_{n+1} at t_{n+1}.

2. Solve (3.5) with $\mathcal{A}_\theta(r_n)$ for $p_\theta(\theta, t | r_n, 0, \phi_n, t_n)$ with the same initial condition as in (3.5). Sample the distribution at $t_{n+1/2} = t_n + \Delta t/2$ for $\theta'_{n+1/2}, \phi'_{n+1/2}$ and update $\theta_{n+1/2}, \phi_{n+1/2}$.

3. Solve (3.5) with $\mathcal{A}_\theta(r_{n+1})$ for $p_\theta(\theta, t | r_{n+1}, 0, \phi_n, t_n)$ with the same initial condition as in (3.5). Sample the distribution at $t_{n+1/2}$ for $\theta'_{n+1}, \phi'_{n+1}$ and update θ_{n+1}, ϕ_{n+1}.

12

It is shown in Section 4.1 that the Strang approximation $p_{\mathbf{r},Str}$ of $p_{\mathbf{r}}$ in (2.8) at t_{n+1} can be written

$$
\begin{aligned}
&p_{\mathbf{r},Str}(r_{n+1}, \theta_{n+1}, \phi_{n+1}, t_{n+1} | r_n, \theta_n, \phi_n, t_n) \\
&\quad = p_\Theta(\theta_{n+1}, t_{n+1} | r_{n+1}, r_n, \theta_n, \phi_n, t_n) p_r(r_{n+1}, t_{n+1} | r_n, t_n).
\end{aligned}
\tag{3.9}
$$

There are at least three advantages of the above splitting of the spatial operator compared to solving (2.8) directly. The equation in the radial direction can include dissociation. The equations in the angular directions are simple with analytical solutions and the modeling is more flexible. The disadvantage is that an additional error due to the splitting is introduced. The local error is of $\mathcal{O}\left(\Delta t^2\right)$ for plain operator splitting and of $\mathcal{O}\left(\Delta t^3\right)$ for Strang splitting, see Section 4.1, and it is very small in the numerical experiments in Section 5.

3.2.2 A molecule and a polymer

Similarly, the diffusion equation in the cylindrical coordinate system (2.13) is split into two parts. The solution in the r-direction for $r \geq \sigma$ is first determined from

$$
\partial_t p_r = D \left(\frac{\partial^2 p_r}{\partial r^2} + \frac{1}{r} \frac{\partial p_r}{\partial r} \right) \equiv \mathcal{A}_{r1}(r) p_r,
\tag{3.10}
$$

with the boundary condition (2.14) at $r = \sigma$. With a sampled r_{n+1} from the distribution with the PDF p_r and after rotation of the coordinate system such that $\phi_n = 0$, we solve

$$
\partial_t p_{\phi z} = D \left(\frac{1}{r^2} \frac{\partial^2 p_{\phi z}}{\partial \phi^2} + \frac{\partial^2 p_{\phi z}}{\partial z^2} \right) \equiv \mathcal{A}_\phi(r) p_{\phi z},
\tag{3.11}
$$

with a periodic boundary condition in the ϕ-direction $p_{\phi z}(-\pi, z, t | 0, z_n, t_n) = p_{\phi z}(\pi, z, t | 0, z_n, t_n)$, a vanishing solution as $|z| \to \infty$, and the initial condition $p_{\phi z}(\phi, z, t_n | 0, z_n, t_n) = \delta(\phi)\delta(z - z_n)$. The PDF of (3.11) with $\mathcal{A}_\phi(r_{n+1})$ is a Gaussian in each one of the coordinate directions z and $r_{n+1}\phi$ with the angle determined modulo 2π. This PDF defines the CDF to be sampled for ϕ_{n+1} and z_{n+1}. The sampling of r and z is exact.

The Strang splitting follows the same scheme as in the previous section. Firstly, p_r is determined at t_{n+1} by (3.10). Secondly, $p_{\phi z}$ is computed at $t_{n+1/2}$ using $\mathcal{A}_\phi(r_n)$ and thirdly, $p_{\phi z}$ is integrated from $t_{n+1/2}$ to t_{n+1} with the diffusion $\mathcal{A}_\phi(r_{n+1})$. The sampling of r and z is here exact but there is an error in the sampling of ϕ.

The evaluation of the solution (2.16) of (3.10) requires an upper bound $u_{\max}$ on the integration interval and then approximation of the integral by a quadrature rule. The integrand is complicated to calculate and it is non-trivial to estimate a reliable $u_{\max}$. Interpolation in a table with the coordinates $r, r_0, \Delta t$ and depending

on the parameters D, σ, and k_c would be necessary. Therefore, p_r is computed numerically by a finite difference approximation of (3.10). We have used a second-order centered finite difference approximation in space and Euler backwards in time.

4 Error analysis

Three sources of error in the PDF and the corresponding CDF used for sampling of the next relative position of the molecules in a pair are analyzed. One error in the PDF is caused by the operator splitting. A second error in the PDF is due to the safety margin for the new position of a single molecule. There is a small probability that a collision with another molecule is ignored. A third source is the discretization errors in the accumulation of the CDFs from the PDFs, the linear interpolation for computing the samples, the numerical solution of the PDEs for the PDFs, and in the approximation of infinite sums.

4.1 Operator splitting

The standard splitting error analysis (34) is not directly applicable because the Green's functions in Section 3 have Dirac functions as initial data but the errors in the PDFs can be derived indirectly.

Let $\mathcal{A}_1(x_1)$ and $\mathcal{A}_2(x_1, \mathbf{x}_2)$ be two differential operators with derivatives in the x_1 and $\mathbf{x}_2$ coordinates, respectively. Consider the PDE

$$u_t = \mathcal{A}_1(x_1)u + \mathcal{A}_2(x_1, \mathbf{x}_2)u, \quad (x_1, \mathbf{x}_2) \in [\sigma, a) \times \Omega, \tag{4.1}$$

with homogeneous boundary conditions at $x_1 = \sigma$ and $x_1 = a$ independent of $\mathbf{x}_2$ and at the boundary of a rectangular domain Ω independent of x_1 and with initial condition at t_n

$$u(x_1, \mathbf{x}_2, t_n) = u_{10}(x_1)u_{20}(\mathbf{x}_2). \tag{4.2}$$

The solution of (4.1) is split into two parts with initial conditions:

$$u_{1t} = \mathcal{A}_1(x_1)u_1, \qquad u_1(x_1, t_n) = u_{10}(x_1)u_{20}(\mathbf{x}_2), \qquad \text{(4.3a)}$$
$$u_{2t} = \mathcal{A}_2(x_1, \mathbf{x}_2)u_2, \qquad u_2(x_1, \mathbf{x}_2, t_n) = u_1(x_1, \mathbf{x}_2, t_{n+1}), \qquad \text{(4.3b)}$$

with u_1 and u_2 satisfying the conditions at the boundaries in the x_1 and $\mathbf{x}_2$-directions, respectively. We will compare the solutions u of (4.1) and u_2 of (4.3b) to estimate the splitting error in the Green's functions p_θ and $p_{\phi z}$ in Section 3.2.

The Green's function $G_1(x_1, t | \xi, t_n)$ of (4.3a) solves

$$G_{1t} = \mathcal{A}_1(x_1)G_1, \ G_1(x_1, t_n | \xi, t_n) = \delta(x_1 - \xi), \tag{4.4}$$

with the same boundary conditions as u_1. Then u_1 can be written

$$u_1(x_1, \mathbf{x}_2, t_{n+1}) = u_{20}(\mathbf{x}_2) \int_\sigma^a G_1(x_1, t_{n+1}|\xi, t_n) u_{10}(\xi)\, d\xi. \qquad (4.5)$$

In the same manner for the second equation in (4.3) there is a Green's function $G_2(x_1, \mathbf{x}_2, t_{n+1}|\boldsymbol{\eta}, t_n)$ such that

$$u_2(x_1, \mathbf{x}_2, t_{n+1}) = \int_{\boldsymbol{\eta} \in \Omega} G_2(x_1, \mathbf{x}_2, t_{n+1}|\boldsymbol{\eta}, t_n) u_1(x_1, \boldsymbol{\eta}, t_{n+1})\, d\boldsymbol{\eta}. \qquad (4.6)$$

The operator splitting approximation of the solution of (4.1) and (4.2) is by (4.5) and (4.6)

$$u_2(x_1, \mathbf{x}_2, t_{n+1}) = \int_\Omega \int_\sigma^a G_2(x_1, \mathbf{x}_2, t_{n+1}|\boldsymbol{\eta}, t_n) G_1(x_1, t_{n+1}|\xi, t_n) u_{10}(\xi) u_{20}(\boldsymbol{\eta})\, d\xi d\boldsymbol{\eta}. \qquad (4.7)$$

With the Green's function $G(x_1, \mathbf{x}_2, t|\xi, \boldsymbol{\eta}, t_n)$ of the unsplit problem (4.1) the solution is

$$u(x_1, \mathbf{x}_2, t_{n+1}) = \int_\Omega \int_\sigma^a G(x_1, \mathbf{x}_2, t_{n+1}|\xi, \boldsymbol{\eta}, t_n) u_{10}(\xi) u_{20}(\boldsymbol{\eta})\, d\xi d\boldsymbol{\eta}. \qquad (4.8)$$

Hence, the difference Δu between u and u_2 is

$$\Delta u(x_1, \mathbf{x}_2, t_{n+1}) = u - u_2 = \int_\Omega \int_\sigma^a \Delta G(x_1, \mathbf{x}_2, t_{n+1}|\xi, \boldsymbol{\eta}, t_n) u_{10}(\xi) u_{20}(\boldsymbol{\eta})\, d\xi d\boldsymbol{\eta}, \qquad (4.9)$$

where $\Delta G = G - G_2 G_1$.

After Taylor expansion in t of u_1, u_2, and u the difference between u and u_2 is

$$\Delta u(x_1, \mathbf{x}_2, t_{n+1}) = \frac{1}{2}\Delta t^2 (\mathcal{A}_1 \mathcal{A}_2 - \mathcal{A}_2 \mathcal{A}_1) u_{10}(x_1) u_{20}(\mathbf{x}_2) + \mathcal{O}\left(\Delta t^3\right). \qquad (4.10)$$

Comparing (4.9) and (4.10) we find that

$$\int_\Omega \int_\sigma^a \Delta G(x_1, \mathbf{x}_2, t_{n+1}|\xi, \boldsymbol{\eta}, t_0) u_{10}(\xi) u_{20}(\boldsymbol{\eta})\, d\xi d\boldsymbol{\eta} \\ = \tfrac{1}{2}\Delta t^2 (\mathcal{A}_1 \mathcal{A}_2 - \mathcal{A}_2 \mathcal{A}_1) u_{10}(x_1) u_{20}(\mathbf{x}_2) + \mathcal{O}\left(\Delta t^3\right), \qquad (4.11)$$

for *any smooth* u_{10} and u_{20} fulfilling the boundary conditions. Thus, the difference between the Green's functions for the unsplit problem G and the operator split problem $G_2 G_1$ is of $\mathcal{O}\left(\Delta t^2\right)$.

With $x_1 = r$ and $\mathbf{x}_2 = (\theta, \phi)$, $a \to \infty$ and $\Omega = [0, \pi] \times [-\pi, \pi]$, in Section 3.2.1, the commutator in the right hand side of (4.11) is

$$[\mathcal{A}_1, \mathcal{A}_2] = \mathcal{A}_1 \mathcal{A}_2 - \mathcal{A}_2 \mathcal{A}_1 = \mathcal{A}_{r2} \mathcal{A}_\theta - \mathcal{A}_\theta \mathcal{A}_{r2} = \frac{2D_{12}^2}{r^3} \left(\frac{1}{r} - 2 \frac{\partial}{\partial r} \right) \mathcal{A}_\vartheta, \quad (4.12)$$

where $\mathcal{A}_\vartheta = r^2 D_{12}^{-1} \mathcal{A}_\theta$. The solutions p_r and p_θ to (3.3) and (3.5) are the Green's functions G_1 and G_2 above. The difference ΔG in (4.9) is obtained from (3.7) and (3.8)

$$\Delta G = \; (p_{\mathbf{r}|r}(\theta_{n+1}, \phi_{n+1}, t_{n+1}|r_{n+1}, r_n, \theta_n, \phi_n, t_n) - p_\theta(\theta_{n+1}, t_{n+1}|r_{n+1}, \theta_n, \phi_n, t_n))$$
$$\cdot p_r(r_{n+1}, t_{n+1}|r_n, t_n).$$
$$(4.13)$$

The conclusion from (4.13) and (4.11) is that the local error due to operator splitting appears in the PDF for the sampling of θ and is proportional to Δt^2.

The analysis of the case in Section 3.2.2 is similar. Let $x_1 = r$ and $\mathbf{x}_2 = (\phi, z)$ in the interaction between a molecule and a polymer. The solutions p_r and $p_{\phi z}$ of (3.3) and (3.5) are the Green's functions G_1 and G_2 in (4.5) and (4.6). The difference ΔG is now as in (4.13)

$$\Delta G = \; (p_{\mathbf{r}|r}(\phi_{n+1}, z_{n+1}, t_{n+1}|r_{n+1}, r_n, \theta_n, \phi_n, t_n)$$
$$- p_{\phi z}(\phi_{n+1}, z_{n+1}, t_{n+1}|r_{n+1}, \theta_n, \phi_n, t_n)) p_r(r_{n+1}, t_{n+1}|r_n, \theta_n, \phi_n, t_n).$$
$$(4.14)$$

The commutator in the right hand side of (4.9) is

$$[\mathcal{A}_1, \mathcal{A}_2] = \mathcal{A}_{r1} \mathcal{A}_\phi - \mathcal{A}_\phi \mathcal{A}_{r1} = \frac{4D^2}{r^3} \left(\frac{1}{r} - \frac{\partial}{\partial r} \right) \frac{\partial^2}{\partial \phi^2}. \quad (4.15)$$

There is no contribution from the z-direction since $\mathcal{A}_{r1}$ and $\partial^2 / \partial z^2$ commute. The local error in the sampling of ϕ is of $\mathcal{O}(\Delta t^2)$.

4.2 Strang splitting

The solution of (4.1) is split into three parts in Strang splitting:

$$u_{2t} = \mathcal{A}_2(x_1, \mathbf{x}_2) u_2, \qquad u_2(x_1, \mathbf{x}_2, t_n) = u_{10}(x_1) u_{20}(\mathbf{x}_2), \qquad (4.16a)$$
$$u_{1t} = \mathcal{A}_1(x_1) u_1, \qquad u_1(x_1, \mathbf{x}_2, t_n) = u_2(x_1, \mathbf{x}_2, t_{n+1/2}), \qquad (4.16b)$$
$$u_{2t} = \mathcal{A}_2(x_1, \mathbf{x}_2) u_2, \qquad u_2(x_1, \mathbf{x}_2, t_{n+1/2}) = u_1(x_1, \mathbf{x}_2, t_{n+1}). \qquad (4.16c)$$

The solution of (4.1) at t_{n+1} can be written using the Green's functions

$$
\begin{aligned}
u_2(x_1, \mathbf{x}_2, t_{n+1}) &= \int_{\boldsymbol{\zeta} \in \Omega} G_2(x_1, \mathbf{x}_2, t_{n+1}|\boldsymbol{\zeta}, t_{n+1/2}) u_1(x_1, \boldsymbol{\zeta}, t_{n+1}) \, d\boldsymbol{\zeta} \\
&= \int_{\boldsymbol{\zeta} \in \Omega} G_2(x_1, \mathbf{x}_2, t_{n+1}|\boldsymbol{\zeta}, t_{n+1/2}) \int_\sigma^a G_1(x_1, t_{n+1}|\xi, t_n) u_2(\xi, \boldsymbol{\zeta}, t_{n+1/2}) \, d\xi d\boldsymbol{\zeta} \\
&= \int_{\boldsymbol{\zeta} \in \Omega} G_2(x_1, \mathbf{x}_2, t_{n+1}|\boldsymbol{\zeta}, t_{n+1/2}) \int_\sigma^a G_1(x_1, t_{n+1}|\xi, t_n) \\
&\quad \cdot \int_{\boldsymbol{\eta} \in \Omega} G_2(\xi, \boldsymbol{\zeta}, t_{n+1/2}|\boldsymbol{\eta}, t_n) u_{10}(\xi) u_{20}(\boldsymbol{\eta}) \, d\boldsymbol{\eta} d\xi d\boldsymbol{\zeta} \\
&= \int_{\boldsymbol{\eta} \in \Omega} \int_\sigma^a G_3(x_1, \mathbf{x}_2, t_{n+1}|\xi, \boldsymbol{\eta}, t_n) G_1(x_1, t_{n+1}|\xi, t_n) u_{10}(\xi) u_{20}(\boldsymbol{\eta}) \, d\xi d\boldsymbol{\eta}, \\
G_3(x_1, \mathbf{x}_2, t_{n+1}|\xi, \boldsymbol{\eta}, t_n) &= \int_{\boldsymbol{\zeta} \in \Omega} G_2(x_1, \mathbf{x}_2, t_{n+1}|\boldsymbol{\zeta}, t_{n+1/2}) G_2(\xi, \boldsymbol{\zeta}, t_{n+1/2}|\boldsymbol{\eta}, t_n) \, d\boldsymbol{\zeta}.
\end{aligned}
$$

(4.17)

The difference Δu between the solution (4.8) determined by the unsplit equation (4.1) and the solution by the Strang splitting (4.17) is as in (4.9) with $\Delta G = G - G_3 G_1$ from (4.17). By Taylor expansion in t, it is shown in (34) that Δu is of $\mathcal{O}(\Delta t^3)$ and consequently, $\Delta G = \mathcal{O}(\Delta t^3)$ locally.

The solutions p_r and p_θ to (3.3) and (3.5) are the Green's functions G_1 and G_2 above. The difference ΔG in (4.9) is obtained from (3.7) and (3.9)

$$
\begin{aligned}
\Delta G &= (p_{\mathbf{r}|r}(\theta_{n+1}, \phi_{n+1}, t_{n+1}|r_{n+1}, r_n, \theta_n, \phi_n, t_n) \\
&\quad - p_\Theta(\theta_{n+1}, t_{n+1}|r_{n+1}, r_n, \theta_n, \phi_n, t_n)) p_r(r_{n+1}, t_{n+1}|r_n, t_n), \\
p_\Theta(\theta_{n+1}, t_{n+1}|r_{n+1}, r_n, \theta_n, \phi_n, t_n) & \\
&= \int_{\boldsymbol{\zeta} \in \Omega} p_\theta(\theta_{n+1}, t_{n+1}|r_{n+1}, \boldsymbol{\zeta}, \phi_n, t_{n+1/2}) p_\theta(\boldsymbol{\zeta}, t_{n+1/2}|r_n, \theta_n, \phi_n, t_n) \, d\boldsymbol{\zeta}.
\end{aligned}
$$

(4.18)

The conclusion from (4.18) is that the local error due to operator splitting appears in the PDF for the sampling of θ and is proportional to Δt^3.

The analysis of the Strang splitting for the molecule and the polymer in Section 3.2.2 can be carried out in the same manner to show that the error in the PDF for sampling of ϕ is also of $\mathcal{O}(\Delta t^3)$.

4.3   Collisions between two molecules

The algorithm is based on dividing the system of molecules into subsystems which are advanced independently. In order to do so we compute a time step $\Delta\tau$ during which interaction between the different subsystems is unlikely.

Let us consider the simplest case: two single freely diffusing molecules. They are assumed to have an initial distance of r_n and now we want to compute a $\Delta\tau$ such that the probability that they collide in $[t_n, t_n + \Delta\tau]$ is smaller than ϵ.

The probability that they do not collide is the same as the probability that they survive during $[t_n, t_n + \Delta\tau]$ with absorbing boundary conditions in (2.11) with $k_a \to \infty$. Thus, if $S_{k_a}(t|r_n, t_n)$ denotes the probability to survive at time t, $t > t_n$ for two particles with initial distance r_n as in (2.12), then $\Delta\tau$ is chosen such that

$$S_{k_a \to \infty}(t_n + \Delta\tau | r_n, t_n) > 1 - \epsilon. \tag{4.19}$$

An expression for this probability is found in (22) for a finite k_a and by letting $k_a \to \infty$ we obtain

$$S_{k_a \to \infty}(t_n + \Delta\tau | t_n, r_n) = 1 - \left(\frac{\sigma}{r_n}\right) \operatorname{erfc}\left(\frac{r_n - \sigma}{\sqrt{4D\Delta\tau}}\right).$$

In (41) they suggest, inspired by the average distance a molecule travels in a time interval, that a reasonable choice of $\Delta\tau$ is

$$\Delta\tau = \frac{(r_n - \sigma)^2}{6H^2 D} \tag{4.20}$$

with $H = 3$. This H is motivated by saying that it is sufficient to preserve the correct steady state distribution.

Instead, we propose that $\Delta\tau$ is taken to be

$$\Delta\tau(r_n) = \frac{(r_n - \sigma)^2}{4H(r_n)^2 D}, \tag{4.21}$$

where $\Delta\tau$ and H now are functions of r_n. Inserting (4.21) into (4.19) we conclude that H should satisfy

$$\left(\frac{\sigma}{r_n}\right) \operatorname{erfc}\left(H(r_n)\right) < \epsilon. \tag{4.22}$$

To see the difference between (4.20) and (4.21), let ϵ be such that the optimal choice of $H(r_n)$ in (4.22) is $H = 3\sqrt{3/2}$ at $r_n = \sigma$. Then $\Delta\tau = \Delta\tau(r_n)$ when $r_n = \sigma$ and $\epsilon \approx 2.035 \cdot 10^{-7}$.
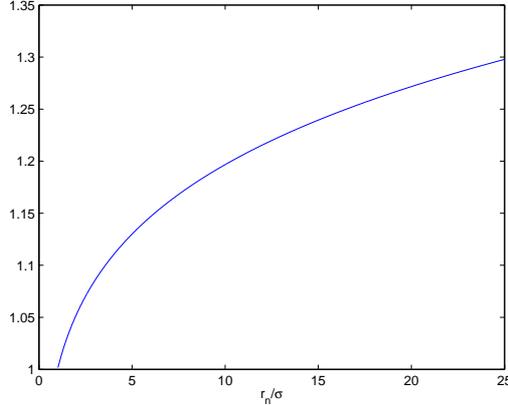
Figure 4.1: The quotient between τ in (4.21) and (4.20) for an optimal H in (4.22) with r_n between σ and 25σ.

Although the difference is not too great, longer time steps can be taken with $\Delta\tau$ in (4.21) also for short distances between the molecules and still the condition in (4.19) is fulfilled.

Note that for a system of particles, what we should consider is really the total probability for a particle in a pair to interact with *any* other particle. To obtain that probability would indeed be very difficult since all possible collisions are dependent events. However, if the concentration is sufficiently small, this dependency will also be small and our simplification is justifiable.

4.4 Discretization errors

The position of a single molecule, ϕ in a pair of molecules, and ϕ and z in a molecule-polymer pair are sampled from the normal distribution. The PDFs in (3.3), (3.5), (3.10), and (3.11) are evaluated by an analytical formula or by solving a numerical discretization of the PDE in each time step Δt for each pair of molecules or molecule-polymer pair. Then the CDF is obtained by numerical integration of the PDF or by an analytical formula. Finally, the sampled value for r or θ is computed by solving the inverse problem where the CDF is equal to the uniformly sampled $\chi \in [0, 1]$.

An analytical formula for the CDF P_r corresponding to p_r in (3.3) is derived in (23). Intervals $[r_{i,\text{low}}, r_{i,\text{up}}]$, $i = 0, 1, \ldots$ are determined here such that $P_r(r_{i,\text{low}}) < \chi < P_r(r_{i,\text{up}})$. The interval is halved as i increases and $r_{i,\text{low}}$ and $r_{i,\text{up}}$ are chosen to keep the inequalities with P_r fulfilled until the length of the interval Δr_i is sufficiently small. Then r is computed by linear interpolation in the interval. The error is proportional to Δr_i^2.

The CDF P_θ with the PDF in (3.5) is computed by numerical integration with the trapezoidal method applied to (3.6) starting at $\theta_0 = 0$ with step $\Delta\theta$

until $P_\theta(\theta_{i+1}) > \chi$. The upper bound on the summation $l_{\max}$ in (3.6) is chosen such that the next term in the series is less than some bound. Then the sampled value θ is determined by linear interpolation between θ_i and θ_{i+1}. The error in the integration and interpolation is proportional to $\Delta\theta^2$. In the same manner, P_r is integrated from the numerical solution of p_r in (3.10) using the step Δr between σ and $r_{\max}$. The time step is $\Delta t/n_{\mathrm{tim}}$ where n_{tim} is a fixed integer and $r_{\max}$ is a few standard deviations of the PDF of a free molecule. The discretization error in r is of $\mathcal{O}\left(\Delta r^2 + \Delta t/n_{\mathrm{tim}}\right)$.

An alternative is to precompute the inverse of the CDFs in tables using analytical or numerical PDFs for a domain in parameter space. Then the sampled values are calculated by multidimensional interpolation. These tables may be quite large for good accuracy and large domains. The interpolation errors are of the same kind as above. This procedure may be faster for parameter values in the domain but requires much more memory and for extreme parameter combinations the only possibility is to sample r and θ as above.

5   Numerical results

Our algorithm is applied to the simulation of three different molecular systems. In the first system, there are one reversible reaction and one irreversible reaction between four species and a decay of one species. The second system consists of two reversible reactions for five species. The third system has one reversible reaction and one irreversible reaction between one polymer and three molecular species. The results with our microscopic method are compared to the concentrations obtained with the macroscopic reaction rate equations and well stirred mesoscopic simulations with the SSA algorithm (18). For comparison with the macroscopic solution, the average is taken over many realizations of the stochastic microscopic and mesoscopic processes. Strang splitting is used in all examples. The difference between straightforward operator splitting and Strang splitting is evaluated for the second system. The computational domain is a cube with reflecting boundaries and the molecules are initially placed randomly inside it.

5.1   One reversible and two irreversible reactions

Consider the following reactions:

$$A + B \underset{k_d}{\overset{k_a}{\rightleftharpoons}} C$$

$$C \xrightarrow{k_{\mathrm{prod}}} A + B + P$$

$$P \xrightarrow{k_{\mathrm{dec}}} \emptyset.$$
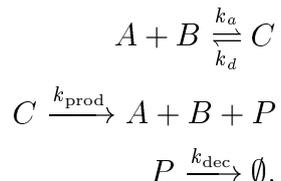
This example was also studied in (41), and here we show that we can reproduce their results. The values of the parameters in the microscopic model in (2.11) are

$k_a = 150$, $k_d = 1$, $k_{\mathrm{prod}} = 2$, $k_{\mathrm{dec}} = 0.2$, $\sigma = 0.04$ and the volume $V = (250\sigma)^3$. The diffusion constant is $D = 4$ for all molecules.
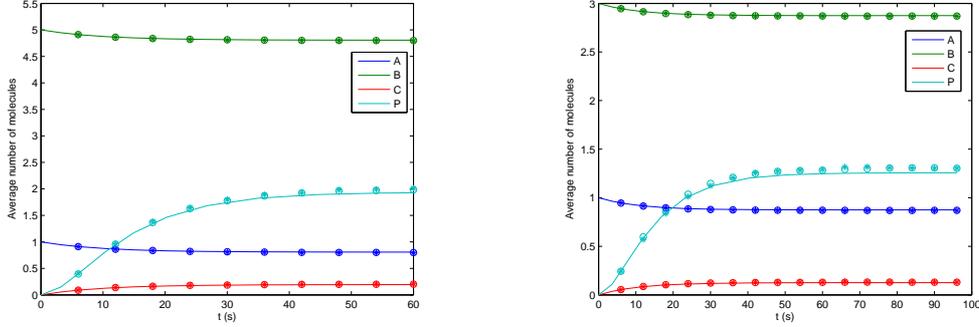

Figure 5.1: The time evolution of the average number of molecules of each species in the first example in Section 5.1. The solid lines represent the macroscopic ODE solution. The data computed with SSA are marked by $\circ$ and with the microscopic simulation by $*$.
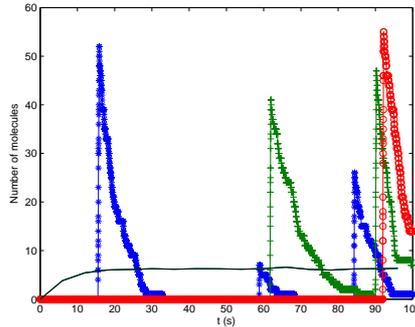

Figure 5.2: The number of P molecules in three realizations $(*, +, \circ)$ of the first example. The solid line shows the average number of P molecules in about 1000 trajectories.

The molecules A, B, and C are treated spatially, but the positions of the P molecules are ignored. Since C can dissociate in two different ways, for a pair of one A molecule and one B molecule the time to the next reaction will be generated. The agreement is excellent with averaged values obtained from simulating the system with SSA using about 1000 realizations in Figure 5.1. The initial data for the first system is five A molecules and one B molecule and for the second system three A molecules and one B molecule. A small discrepancy with the macroscopic model is noted, but this is as expected for small copy numbers.

The variation of the number of P molecules in three different trajectories is displayed in Figure 5.2. As we can see, the variance is quite large. The number of molecules can be zero for some time before an A and a B molecule react and

there is a sudden explosion of the number of P molecules. Then the A and B molecules eventually dissociate and diffuse away from each other and the number of P molecules will decay exponentially.
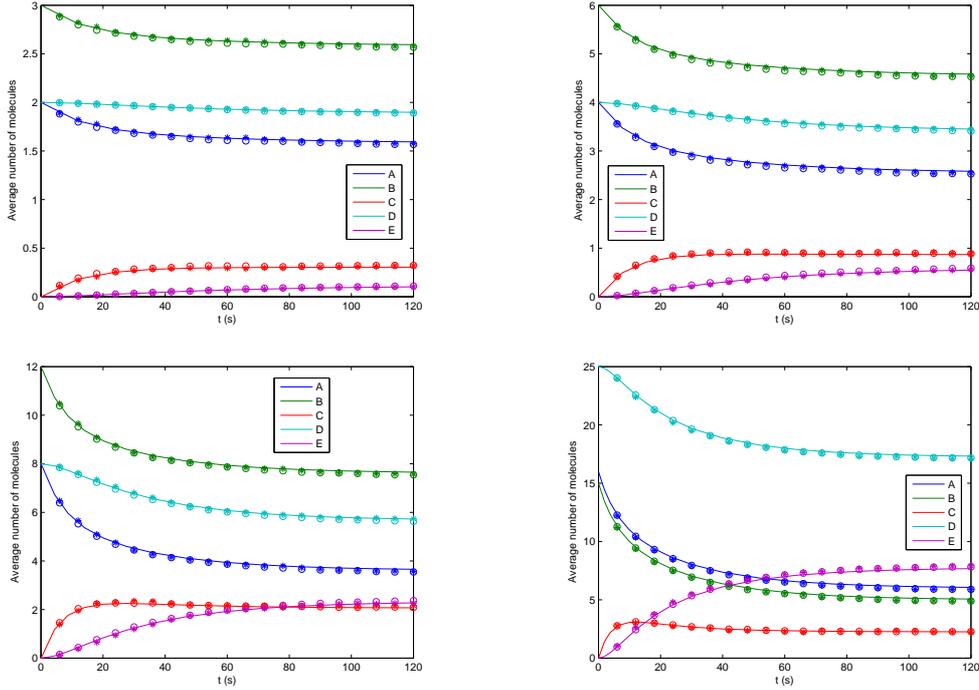
5.2 Two reversible reactions

Figure 5.3: The number of initial molecules is doubled between the panels starting at the upper left panel and ending at the lower right panel while keeping all parameters constant for the second example in Section 5.2. The solid lines represent the macroscopic solution. The data computed with SSA are marked by ∘ and with the microscopic simulation by ∗.

In this example we consider the following system of two coupled reactions:

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C$$

$$C + D \underset{k_4}{\overset{k_3}{\rightleftharpoons}} E.$$
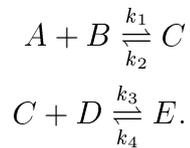
The values of the parameters are $k_1 = 150$, $k_2 = 2$, $k_3 = 200$, $k_4 = 1$, corresponding to k_a and k_d in (2.11), the diffusion $D_i = 4$ for all species, $\sigma = 0.04$ and the volume $V = (250\sigma)^3$.

The initial system has seven molecules, see the upper left panel of Figure 5.3, two A, three B, and two D molecules. There is a decline in the number of A,

B, and D molecules, while the expected number of C and E molecules increases in time. The number of molecules is successively doubled in the subfigures in Figure 5.3 while keeping the parameters constant. The averages in the microscopic simulations with about 1000 trajectories agree well with those obtained by simulating this system with SSA. The relative discrepancy between the stochastic simulations and the deterministic solution of the ODE system is the same in the figures when the total number of molecules increases. The standard deviations computed mesoscopically or microscopically are almost the same in Figure 5.4 with seven molecules initially. The results for the standard deviations in the other three cases in Figure 5.3 are similar. In Figure 5.5, we show the time series of a single realization of the system.
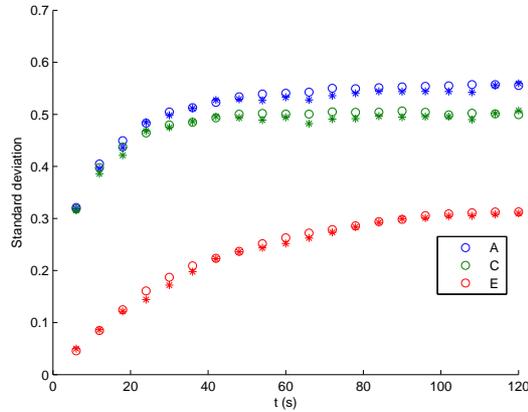
Figure 5.4: The standard deviation in the second example with seven molecules initially computed with SSA (o) and microscopically (∗). The standard deviation for the number of A- and B-molecules is the same as well as for the number of C- and D-molecules.
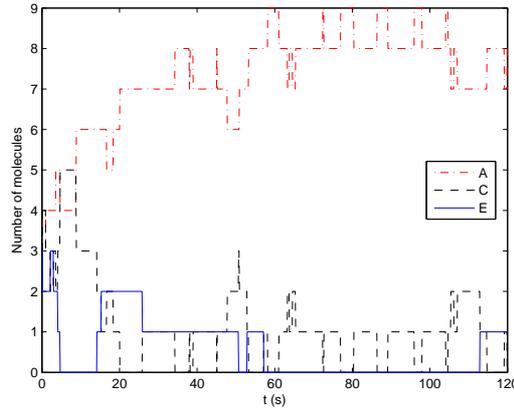
Figure 5.5: The result of one simulation of the system in Section 5.2. The values of the parameters in this example are $k_1 = 300$, $k_2 = 10$, $k_3 = 300$, $k_4 = 10$, $\sigma = 0.04$, $V = (250\sigma)^3$ and $D = 4$ for all species.

In Figure 5.6, the average computational work for one realization is measured in our present implementation on a Nehalem 2.26 Ghz processor when the total number of initial molecules N increases. The average CPU-time depends quadratically on N in Figure 5.6. When doubling the number of molecules while keeping the concentration constant by expanding V, the CPU-time per trajectory increases linearly.
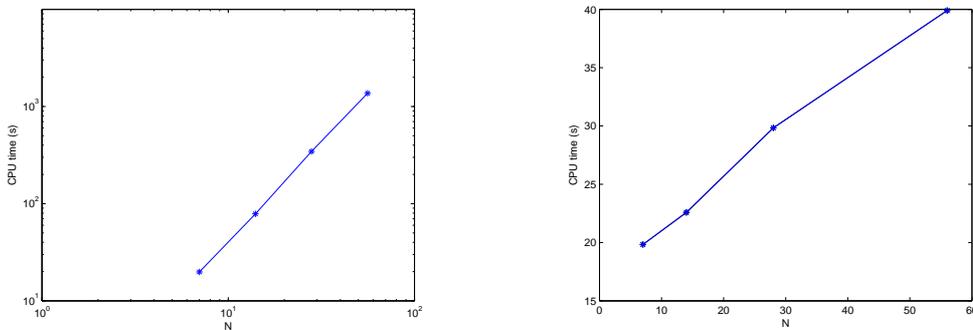
Figure 5.6: The CPU-time in seconds as a function of the number of initial molecules N in a fixed volume (left) and with an expanding volume to keep the concentration constant (right).

In Figure 5.7, we compare the results when simulating 10000 trajectories of the above example with operator splitting as well as with Strang splitting. There is no noticeable difference in the average number of molecules when using operator splitting. One reason for this could be that there is no bias in the errors and they are averaged out when choosing initial positions randomly and simulating

24

the system longer periods of time. Another possible reason is that the time steps $\Delta\tau$ from (4.21) are so short that the difference is not visible with the resolution in the figure. A conclusion is that the operator splitting is *not* a major source of error in the simulations.
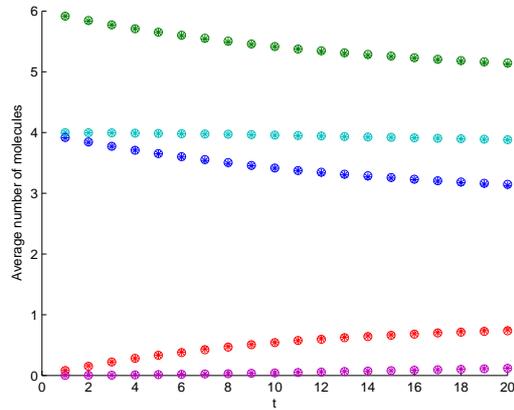

Figure 5.7: Results obtained using operator splitting (o) compared to results using Strang splitting ($*$).
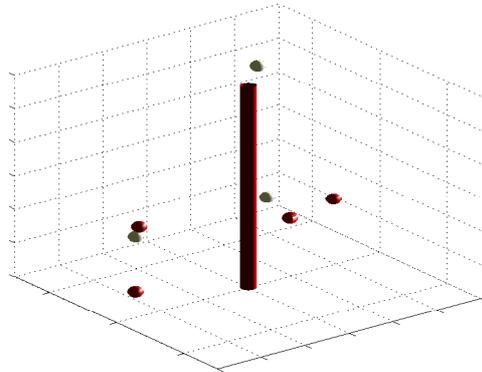
5.3   Reactions with a polymer

Figure 5.8: A snapshot of a molecule-polymer simulation. A- and B-molecules are diffusing around the polymer. They may react and form C-molecules, which in turn can be absorbed by the polymer.

Finally we consider a single reversible reaction involving three species, where one of the molecules can react with a polymer (which could for instance be DNA

or RNA)

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C$$

$$C + \text{Polymer} \overset{k_{\text{cyl}}}{\rightarrow} \emptyset.$$
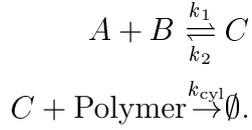
As described in Section 3, the polymer is modeled by a fixed immobile cylinder along the z-axis, see Figure 5.8. Several C molecules may react with the polymer at the same time. The values of the parameters are $k_1 = 200$, $k_2 = 1$, $k_{\text{cyl}} = 75$, $V = (200\sigma)^3$, $\sigma_{\text{cyl}} = 0.05$, $\sigma = 0.04$ and $D = 4$ for all species.
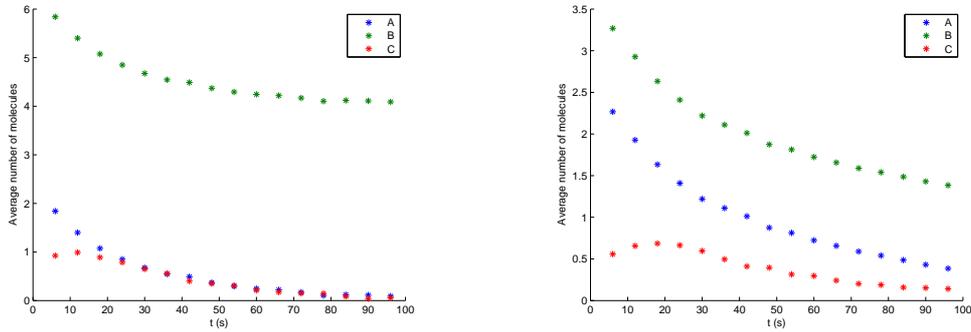

Figure 5.9: The averages of the species over 1000 trajectories in the third example in Section 5.3 determined by the microscopic model and with different initial data.

The results with the microscopic model are as expected in Figure 5.9. After an initial increase in the number of C molecules they will gradually be absorbed by the cylinder and eventually there will be only A or B molecules left. Thus, if we look at the standard deviation of the number of molecules in Figure 5.10, we find that it tends to zero.
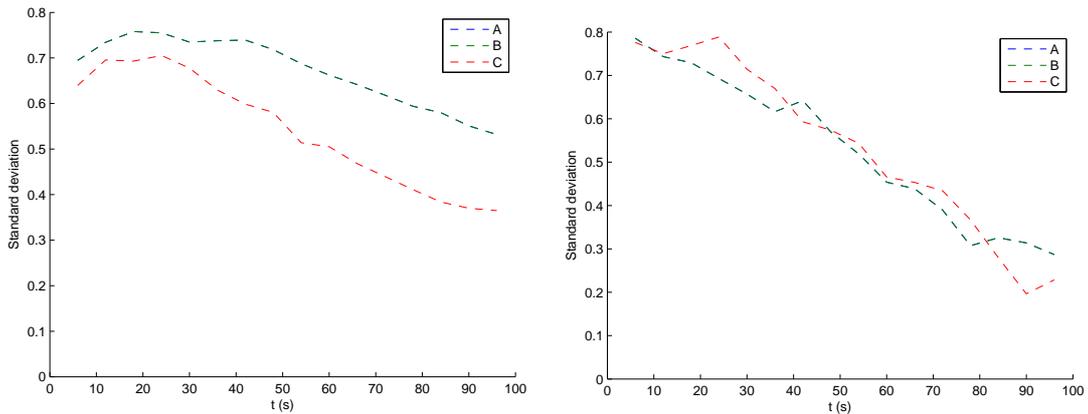

Figure 5.10: The standard deviation decays as the average number of molecules decays for the same initial data as in Figure 5.9. The values for molecules A and B overlap.

26

The agreement between the macroscopic, mesoscopic, and microscopic modeling of the chemical systems in this section is very good. Then the question arises: Why use a complicated single molecule method if the ODE model is as accurate? The answer is that the examples here are chosen for verification of the algorithm for single molecules and for validation of the microscopic model. Many other systems are known where simulations with macroscopic and mesoscopic models differ e.g. in (38) and in the Introduction problems are listed where a microscopic model is the only alternative.

6 Conclusions

An algorithm is proposed for Monte Carlo simulation of the motion and reactions of molecules with low copy numbers, particularly suitable for problems in molecular biology. Our work is based on the GFRD algorithm (41). It is an event-driven algorithm for taking longer time steps than is possible in methods based on Brownian dynamics. In this way, orders of magnitude in computing time can be saved.

The new contribution is an operator splitting technique introducing an additional error in one of the coordinate directions but simplifying the implementation of the method considerably and increasing the applicability and flexibility. Furthermore, we can include dissociation in the boundary condition of the governing PDE, enabling the possibility of leaping over fast reactions. The molecules are partitioned into subsystems and each subsystem is advanced in time independently from the other subsystems between synchronization time points. We suggest a way of choosing the time steps based on the probability for two particles to collide, in order to ensure that the error made is small when partitioning the molecules into subsystems.

The errors in every time step of the algorithm caused by the discretization in time and space and in the parameter space are analyzed. The errors due to two versions of the operator splitting are quantified in a numerical example and they are small. The accuracy of the method is demonstrated in three examples: a system with four species and four reactions, a system with five species and four reactions and a system with three molecular species, one polymer, and three reactions.

Acknowledgment

References

[1] M. Ander, P. Beltrao, B. Di Ventura, J. Ferkinghoff-Borg, M. Foglierini, A. Kaplan, C. Lemerle, I. Tomás-Oliveira, and L. Serrano. SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. *Syst. Biol.*, 1:129–138, 2004.

[2] S. S. Andrews and D. Bray. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys. Biol.*, 1:137–151, 2004.

[3] F. Baras and M. Malek Mansour. Reaction-diffusion master equation: A comparison with microscopic simulations. *Phys. Rev. E*, 54:6139–6148, 1996.

[4] L. Boulianne, S. Al Assaad, M. Dumontier, and W. J. Gross. GridCell: a stochastic particle-based biological system simulator. *BMC Syst. Biol.*, 2:66, 2008.

[5] H. S. Carslaw and J. C. Jaeger. *Conduction of Heat in Solids*. Oxford University Press, London, 2nd edition, 1959.

[6] J. S. Coggan, T. M. Bartol, E. Esquenazi, J. R. Stiles, S. Lamont, M. E. Martone, D. K. Berg, M. H. Ellisman, and T. J. Sejnowski. Evidence for ectopic neurotransmission at a neuronal synapse. *Science*, 309:446–451, 2005.

[7] F. C. Collins and G. E. Kimball. Diffusion-controlled reaction rates. *J. Colloid. Sci.*, 4:425–437, 1949.

[8] J. Cullhed, S. Engblom, and A. Hellander. The URDME manual version 1.0. Technical Report 2008-022, Dept. of Information Technology, Uppsala University, Uppsala, Sweden, 2008. Available at `http://www.it.uu.se/research`.

[9] M. Dobrzyński, J. V. Rodríguez, J. A. Kaandorp, and J. G. Blom. Computational methods for diffusion-influenced biochemical reactions. *Bioinformatics*, 23:1969–1977, 2007.

[10] A. Donev. Asynchronous event-driven particle algorithms. *Simulation*, 85(4):229–242, 2009.

[11] A. Donev, V. V. Bulatov, T. Oppelstrup, G. H. Gilmer, B. Sadigh, and M. H. Kalos. A first-passage kinetic Monte Carlo algorithm for complex diffusion-reaction systems. *J. Comput. Phys.*, 229:3214–3236, 2010.

[12] K. Doubrovinski and M. Howard. Stochastic model for Soj relocation dynamics in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*, 102:9808–9813, 2005.

[13] S. Engblom, L. Ferm, A. Hellander, and P. Lötstedt. Simulation of stochastic reaction-diffusion processes on unstructured meshes. *SIAM J. Sci. Comput.*, 31:1774–1797, 2009.

[14] R. Erban and J. Chapman. Reactive boundary conditions for stochastic simulation of reaction-diffusion processes. *Phys. Biol.*, 4:16–28, 2007.

[15] R. Erban and J. Chapman. Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions. *Phys. Biol.*, 6:046001, 2009.

[16] D. Fange and J. Elf. Noise induced Min phenotypes in *E. coli. PLoS Comput. Biol.*, 2(6):e80, 2006.

[17] C. W. Gardiner. *Handbook of Stochastic Methods.* Springer Series in Synergetics. Springer-Verlag, Berlin, 3rd edition, 2004.

[18] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22(4):403–434, 1976.

[19] J. Hattne, D. Fange, and J. Elf. Stochastic reaction-diffusion simulation with MesoRD. *Bioinformatics*, 21:2923–2924, 2005.

[20] J. M. Haugh. A unified model for signal transduction reactions in cellular membranes. *Biophys. J.*, 82:591–604, 2002.

[21] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry.* Elsevier, Amsterdam, 5th edition, 2004.

[22] H. Kim and K. J. Shin. Exact solution of the reversible diffusion-influenced reaction for an isolated pair in three dimensions. *Phys. Rev. Lett.*, 82(7):1578–1581, 1999.

[23] H. Kim, M. Yang, and K. J. Shin. Dynamic correlation effect in reversible diffusion-influenced reactions: Brownian dynamics simulation in three dimensions. *J. Chem. Phys.*, 111(3):1068–1075, 1999.

[24] G. I. Marchuk. Splitting and alternating direction methods. In P. G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis*, pages 197–462, Amsterdam, 1990. North-Holland.

[25] H. H. McAdams and A. Arkin. It's a noisy business. Genetic regulation at the nanomolar scale. *Trends Gen.*, 15:65–69, 1999.

[26] M. I. Monine and J. M. Haugh. Reactions on cell membranes: Comparison of continuum theory and Brownian dynamics simulations. *J. Chem. Phys.*, 123:074908, 2005.

[27] T. Oppelstrup, V. V. Bulatov, A. Donev, M. H. Kalos, G. H. Gilmer, and B. Sadigh. First-passage kinetic Monte Carlo method. *Phys. Rev. E*, 80:066701, 2009.

[28] Alexander V. Popov and Noam Agmon. Three-dimensional simulations of reversible bimolecular reactions: The simple target problem. *J. Chem. Phys.*, 115(19):8921–8932, Nov. 2001.

[29] D. Ridgway, G. Broderick, A. Lopez-Campistrous, M.Ru'aini, P. Winter, M. Hamilton, P. Boulanger, A. Kovalenko, and M. J. Ellison. Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm. *Biophys. J.*, 94:3748–3759, 2008.

[30] C. Sanford, M. L. K. Yip, C. White, and J. Parkinson. Cell++ – simulating biochemical pathways. *Bioinformatics*, 22:2918–2925, 2006.

[31] P. Sjöberg, O. G. Berg, and J. Elf. Taking the reaction-diffusion master equation to the microscopic limit. Technical report, Dept. of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden, 2009.

[32] R. D. Skeel. What makes molecular dynamics work? *SIAM J. Sci. Comput.*, 31:1363–1378, 2009.

[33] M. v. Smoluchowski. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Z. phys. Chemie*, 92:129–168, 1917.

[34] G. Strang. On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.*, 5:506–517, 1968.

[35] A. B. Stundzia and C. L. Lumsden. Stochastic simulation of coupled reaction-diffusion processes. *J. Comput. Phys.*, 127:196–207, 1996.

[36] K. Takahashi, S. N. V. Arjunan, and M. Tomita. Space in systems biology of signaling pathways–towards intracellular molecular crowding in silico. *FEBS Lett.*, 579:1782–1788, 2005.

[37] R. S. Varga. *Matrix Iterative Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1962.

[38] J. M. G. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler. Mechanism of noise-resistance in genetic oscillators. *Proc. Nat. Acad. Sci.*, 99:5988–5992, 2002.

[39] J. A. C. Weideman. Computation of the complex error function. *SIAM J. Numer. Anal.*, 31(5):1497–1518, 1994.

[40] J. S. van Zon and M. J. Morelli and S. Tănase-Nicola and P. R. ten Wolde. Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophys. J.*, 91:4350–4367, 2006.

[41] J. S. van Zon and P. R. ten Wolde. Green's-function reaction dynamics: A particle-based approach for simulating biochemical networks in time and space. *J. Chem. Phys.*, 123:234910, 2005.

[42] J. S. van Zon and P. R. ten Wolde. Simulating biochemical networks at the particle level and in time and space: Green's-function reaction dynamics. *Phys. Rev. Lett.*, 94:128103, 2005.