Stochasticity in signal transduction pathways

Vidal Rodriguez, J.

Publication date
2009

Chapter 2

Gillespie Multiparticle Method

In this chapter we present a method to simulate large biochemical networks with stochastic and discrete space capabilities: the Gillespie-Multiparticle method (GMP) (Rodriguez et al., 2006). Its underlying model is based on the *Diffusion-Reaction Master Equation* (RDME) described in section 1.4. It is therefore discrete in space, time and in number of particles. Our aim is to produce a method at the mesoscopic scale in order to cope with large biochemical networks. In order to comply with the requirements of a mesoscopic method. This allows us to use a well-known method for stochastic reactions: Gillepie's method (Gillespie, 1977). This discrete-in-space method has the capabilities to capture features such as spatial patterns, gradients and fluctuations created by the interaction of the diffusion and the reaction processes. The resolution of these features will depend ultimately on the granularity, or coarseness, of the lattice. In section 2.2.4 we will discuss how to choose a proper discretisation.

2.1 Background

GMP differentiates from other available methods, discussed in section 1.4, in that it uses an operator-split mechanism for executing the reaction and the diffusion processes, and that integrates the advantages of being able to diffuse particles either one by one or in groups while conserving the normal diffusion properties.

The contents of this chapter are based on the results published in

- Rodriguez, J. V., Kaandorp, J. A., Dobrzyński, M., and Blom, J. G. (2006). Spatial stochastic modelling of the phosphoenolpyruvate-dependent phosphotransferase (pts) pathway in escherichia coli. *Bioinformatics*, 22(15):1895–1901.
- Rodríguez, J. V. and Kaandorp, J. A. (2007). Inferring the distribution of inter-reaction intervals for diffusion-limited reactions on lattices. *Intl. J. Mod. Phys. C*, 18(4):749–757.

Like all methods based on Gillespie's reaction mechanism, dimensionless particles located in the lattice sites are used; diffusion takes place between adjacent lattice sites and reaction occurs locally in each lattice site. In this particular aspect GMP builds on the multiparticle lattice gas automata model developed by Chopard et al. (1994). GMP employs the same diffusion process for particles to distribute them randomly among the nearest lattice site neighbours at certain defined time intervals. Contrast this diffusing mechanism to the individual particle diffusion with exponentially distributed times as in MesoRD by Elf et al. (2003) and by Stundzia and Lumsden (1996).

For reactions, however, instead of Chopard's combinatoric method we use the stochastic simulation algorithm (SSA) of Gillespie (1977).

The GMP method aims at reducing the computational cost of non-split models caused by the extra calculations of individual diffusion events, which is relevant when the system has a mixture of high and low density of particles.

2.2 Description of GMP

Following the typical space discretisations by the RDME-based (Baras and Mansour, 1996 ; Gardiner, 1983) and the Cellular Automata methods (Weimar, 2001 ; Chopard et al., 1994), we discretize the volume of the cell into a cubical lattice with L^3 sites. Each lattice site, of dimension λ, holds a discrete number of uniformly distributed particles. Note that these particles do not have a physical dimension associated because Gillespie's algorithm does not use them explicitly. Fig. 2.1 shows a 2D illustration of a section of a spherical cell. The cytosolic space occupies all the light grey sites, while the darker sites represent the membrane surface that hold the membrane-bound particles (note that this is not the intermembrane space.) As a result of the spatial discretisation, the overall inhomogeneity of the system is limited by the coarseness of the lattice, since within the site no exact spatial information is provided nor required. To simulate the membrane as a impermeable wall we use reflective boundary conditions.

2.2.1 The Operator-Split Reaction-Diffusion

An operator-split Reaction-Diffusion (RD) scheme alternates the execution of each process at predetermined intervals of time (see Algorithm 1). It is therefore an approximation to a Reaction-Diffusion Master Equation solver where, in contrast, these two processes are intertwined. Therefore we are bound to make some error. We will discuss ways to control this error and under what circumstances the method is best used.

In this section we describe the basic algorithm. Let's have a set of different chemical reactants S, where each species can have a different diffusion coefficient D_S. From random walk theory (Chopard and Droz, 1991), we know that the average time required to displace a distance λ is

$$\langle \tau_{D_S} \rangle = \frac{1}{2d} \frac{\lambda^2}{D_S}, \tag{2.1}$$

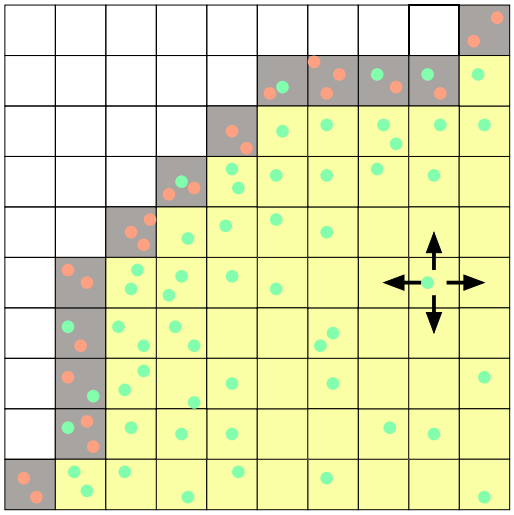

Figure 2.1: *2D slice representation of the spherical cell geometry, including the membrane sites in dark gray and a possible distribution of particles. The length of a site is λ and the neighborhood of a cytosolic site consists of four sites. Reflective boundary conditions are used for sites on the edge of the domain.*

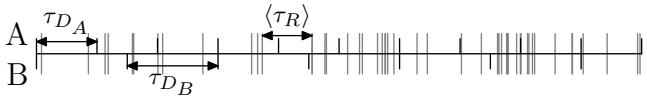

Figure 2.2: *Illustration of the operator-split mechanism for reaction- diffusion.*

with d the dimensionality of the system.

A diffusion step consists in moving randomly the p^S particle of species S to the neighboring sites, which represents a displacement of λ. Therefore, after one diffusion step over all lattice sites, the simulation time is incremented by $\langle\tau_{D_S}\rangle$. We see from this single step that the diffusion process will be executed at times $t_S = n_S\tau_{D_S}$, where n_S is the iteration number, and the diffusion step.

When having different diffusion coefficients for different species, the next diffusion time is the minimum t_S that is larger than the current simulation time t_{sim}. Note that more than one species (even with different D_S) can occasionally have equal t_S. In this case all these species execute a diffusion step. The current simulation time t_{sim} is updated to t_S, and n_S is incremented by one for the diffused species.

We illustrate this process in Fig. 2.2 by having two chemical species A and B. For instance $D_A = 1$ and $D_B = 2/3$, and the reaction process a rate of $k = 1.5$. We observe that the first diffusion event will correspond to A, shortly followed by B, and again A, and then A and B simultaneously. This algorithm is summarized in Alg. 1.

Algorithm 1 Operator-split, Reaction-Diffusion algorithm.

0- Initialize $t_S = \min\{\tau_{D_S}\}$ for all species S
1- $t_{sim} = 0$, $n_S = 1$ for all S
2- while $t_{sim} < t_{end}$ do
3- while $t_{sim} < t_S$ do
4- Reaction μ in τ_R on every lattice site
5- Advance simulation time $t_{sim} = t_{sim} + \tau_R$
6- Diffuse species for which $t_S = t_{sim}$
7- Increment iteration n_S for the diffused species
8- $t_S = \min\{\tau_{D_S} \cdot n_S\}$ for all S

Next, we give further details on the internal mechanism of the diffusion and reaction processes.

2.2.2 The Diffusion Process

We already know that the diffusion process is executed locally at every site. A diffusion step consists in randomly distributing all the particles with equal probability among the six nearest neighbors (Fig. 2.1 shows a 2D neighborhood in black arrows). Chopard et al. (1994) showed that this diffusion rule reproduces the macroscopic diffusion equation in the limit of $\lambda \to 0$, where λ is the lattice site's side length. He also give a method to diffuse particles in bulk, assuming enough particles are present (threshold = 90). This is possible since the stochastic process of moving l particles out of N into two different neighbors is a binomial process. Which in turn, for large N and the theorem of large numbers, the binomial distribution converges to a Gaussian distribution.

The multi-particle method needs to create 6 groups of particles. So first, we create three groups by drawing a normal random number $N(\mu, \sigma)$ such as: the first group contains statistically one third of the total number of particle (l) in the site. Thus, $X_1 \sim N(l/3, l(1/3)(1-1/3))$. The second group contains statistically half of the remaining particles ($X_2 \sim N((l-X_1)/2, (l-X_1)(1/2)(1/2))$), and the third group X_3 contains the remaining from group 2, and no random number is necessary. Now, we only need to split each group following the same procedure to obtain the six groups.

One issue with this type of diffusion is that sharp (or non-smooth) spatial inhomogeneities lead to a check-board effect in diffusion. To illustrate this lets imagine that initially 100 particles are located in one lattice site. After one diffusion step this site will become void of particles (white box contains no particles), while the neighboring ones will contain all the particles (black boxes contain particles). In the next diffusion step the same effect will happen to the neighboring sites that received the particles. Thus, the black ones become white, and the white become black. To overcome this undesired situation, Chopard et al. (1994) uses rest particles, this is, at every diffusion step only a fraction p_r of the particles are diffused. Consequently the average time for a diffusion step

is now

$$\langle \tau_S \rangle \cdot (1 - p_r). \tag{2.2}$$

To obtain the actual number of resting particles we use the same statistical procedure as before, where $X_{p_r} \sim N(lp_r, (lp_r)p_r(1-p_r))$.

2.2.3 The Reaction Process

For the reaction process we use the commonly used Gillespie's method or SSA (Gillespie, 1977), which solves exactly the Chemical Master Equation for well-stirred systems. We note that the use of the more recent improvements (Gibson and Bruck, 2000) of this method do not change the functioning of GMP. There are two important implication by using this method: (i) that the system is reaction limited at the scale of the lattice site, and (ii) that we cannot have more spatial resolution than the lattice site size.

Gillespie's method advances the state of the system by executing chemical reactions sequentially, one reaction at a time. Given a system at time t, the next reaction occurs at time τ_R. The multivariate distribution $P(\tau_R, \mu)\, d\tau$ is the probability that, given the state X_S at time t, the next reaction μ will occur in the infinitesimal time interval $(t+\tau, t+\tau+d\tau)$. Thus, the time of next reaction τ_R that is of type μ is given by:

$$P(\tau_R, \mu) = \begin{cases} a_\mu \exp(-a_0 \tau_R) & , 0 \leq \tau_R < \infty, \\ & \mu = 1, \ldots, M \\ 0 & , \text{otherwise}, \end{cases} \tag{2.3}$$

where a_μ is called propensity function, with M the total number of reaction channels in the system, and a_0 is

$$a_0 = \sum_{\mu=1}^{M} a_\mu.$$

The propensity function quantifies the probability of every reaction channel to occur, when all the reaction in the systems are considered. Hence the normalization factor a_0 in Eqn. 2.3. For a reaction channel μ, that is a bimolecular reaction, this is given by

$$a_\mu = k_\mu \, n^A \, n^B / (V \, N_A), \tag{2.4}$$

where k_μ is the reaction rate, n^A is the number of particles in the site for species A, V is the site volume, and N_A Avogadro's number. A monomolecular reaction, in contrast to a bimolecular, does not depend on the volume of the system. The propensity function for such reaction is then

$$a_\mu = k_\mu \, n^A.$$

The next reaction time τ_R and the reaction channel μ are obtained for a given species S as follows:

$$\tau_R = 1/a_0 \ln(1/r_1) \quad \text{and} \quad \sum_{i=0}^{\mu-1} a_i \leq a_0 r_2 < \sum_{i=\mu}^{M} a_i, \tag{2.5}$$

where, r_1 and r_2 are uniformly distributed random numbers from the interval [0, 1]. Then reaction μ is effectuated and the simulation time t_{sim} is increased with τ_R.

Because of the operator-splitting mechanism, the reaction process executes reactions only for the duration of the current diffusion interval, this is while $\sum \tau_R < \tau_{D_S}$.

2.2.4 Choosing a Lattice Discretisation Size

A critical aspect for the correct functioning of the GMP, and therefore of the family of lattice discretized methods, is the use of a proper lattice size discretization. By adjusting the lattice size, λ, GMP's spatial resolution capabilities range from those offered by spatially resolved particle methods to well-stirred methods. However there is a limitation for the minimum lattice size λ. Despite the convergence of the diffusion process as shown by Chopard et al. (1994), the reaction-diffusion process does not converge for $\lambda \to 0$. In that limit no reactions can occur as the probability of finding two particles in a site is zero. Approaching this limit also implies that the maximum possible reaction rate decreases. Thus, for a pair of reacting particles and a diffusion limited process, there is a minimal lattice size $\lambda_o = 2\sigma$, where σ is the diameter of a particle. Note, however, that under this condition GMP would act almost as Brownian Dynamics with reaction. For complex biochemical systems, in general we do not find purely diffusion limited reactions, which allows us to take larger subvolumes that are considered to be well-stirred.

A reasonable size of the well-stirred subvolume is given by λ_{rmfp}, the reaction mean free path. The reaction mean free path (RMFP) is defined as $\lambda_{rmfp} = (\langle\tau_R\rangle D_{AB})^{1/2}$, where $\langle\tau_R\rangle$ is the mean reaction time given by $1/a_\mu$, and $D_{AB} = D_A + D_B$ is the mutual diffusion coefficient Thus, the λ_{rmfp} is the average distance traveled between two reacting collisions. The validity of the Reaction Diffusion Master Equation is further discussed in Baras and Mansour (1996)).

2.3 Enhancements for Low Numbers of Particles

The multiparticle diffusion method by Chopard et al. (1994) is efficient only for a large number of particles. For two-dimensional systems he uses this method when a site contains more than 40 particles. Based on a comparison of the error between the real binomial distribution and the Gaussian for three dimensional simulations, we set the threshold to 90 (data not shown). Raising the threshold to apply the multiparticle diffusion scheme carries a considerable computational cost penalty, for the one by one particle diffusion is now more likely to occur.

One way to enhance the performance, when the number of molecules is below the threshold, is to apply the multiparticle idea described in section 2.2.2 but this time using the exact distribution, namely the Binomial. This Binomial distribution is the true underlying random process of distributing p particles into n sites with probability $1/2d$, where d is the dimensionality of the system (in our case 3). With this process we are not recovering the exact Reaction

Diffusion Master Equation for diffusion but keeping the approximate operator-split method presented earlier in this chapter.

Unfortunately, the computational cost of drawing random numbers from a Binomial distribution (Press et al., 2007), despite constant, it is rather high (≈ 23 s) and 2.5 times higher than drawing from a Gaussian random number generator (although this time it is not the right distribution) as can be observed in Fig. 2.3 as dash-dot line. Having a closer look at the problem we note that we only need to draw random numbers for a small and finite set of parameters. This small set of parameters are: the number of particles, and the probability of success of the event. The Binomial distribution has two parameters: the number of elements, and the probability of success of choosing an element, which correspond precisely to those we need. The elements to diffuse are the l particles candidate to be shifted to a neighboring site. The probability of success is one over the number of groups, because we follow the same splitting mechanism as before. Because of the reduced parameter space it is easy to generate precomputed tables of values. For our case, these tables would contain distributions up to $90-1$ particles, and the probability to chose each is either $1/2$ or $1/3$ (in order to obtain six groups we first split them probabilistically in half and then each half in three parts). Note that the diffusion coefficient plays no role in these tables as this is taken care of by the timing mechanism described in the Algorithm 1.

Inspecting the inverse of the cumulative distribution function (*cdf*) of a binomial distribution we note that the edges are very steep, while the central area is rather flat. Therefore a pure discretisation of the [0, 1) interval would contain large errors near the edges. We find that the central flat area is approximately comprised in the interval $[1/1024,\ 1 - 1/1024)$.

In searching for the pre-tabulated values, we distinguish two situations: the central region of the distribution and the tails. For the central region it is possible to search by simply scaling the uniform random number (URN), and the cost of this process is $O(1)$. On the other hand, the tails need a different searching approach. Auxiliary tables are precomputed to find the value of $x \in [0,\ 1/1024)$ for the lower tail, that corresponds to a certain y number of particles. A binary search, which has an average cost of $O(\log n)$, of these lower and upper tails will suffice. Because the tails have a low probability of being searched (2/1024), the binary search cost constitutes a very small fraction to the global average searching cost. We see in Fig. 2.3, as a dash-double-dot line, that this method of pre-tabulated tables has a small computational cost, and that it remains basically constant throughout the range of values 10 to 90. If only the binary search was used along the whole precomputed table, then the search time increases logarithmically as shown with the dashed line in Fig. 2.3. Finally, for the sake of completeness, we also show the preformance of the most elementary method of diffusing particle one-by-one, as a solid line. Note that its cost is linear to the number of particles, and that for less than 10 particles it is the most efficient one.

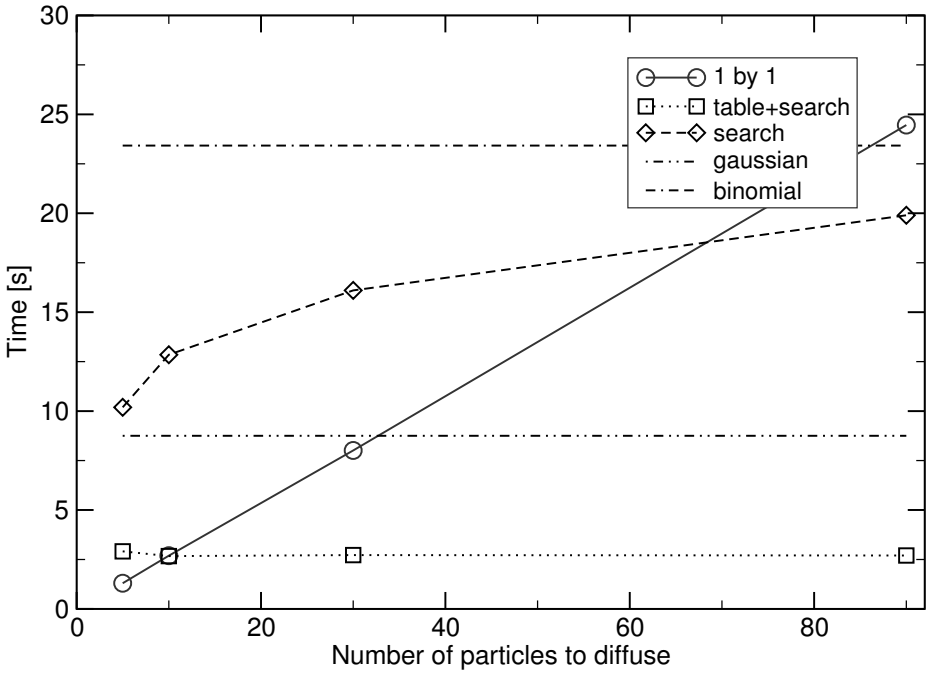

Figure 2.3: *Performance comparison of three dimensional diffusion on a lattice for a low number of particles. The most simple strategy, one-by-one, is the most efficient for less than 10 particles. for a larger number the best is a combination of table and binary search which keeps the global search time low and nearly constant at* ≈ 2.5 *s, nearly 10 times smaller than using an exact Binomial random number generator as in Press et al. (2007).*

2.3.1 Qualitative Computational Cost Comparison

It is difficult to compare quantitatively the performance of methods with a non-homogeneous set of features, such as GMP, against MesoRD, GFRD or Smoldyn. Instead, we find it more instructive to offer a qualitative comparison of the computational cost for the discrete methods that are similar, in terms of features, to GMP. The main factor to account for is how diffusion-limited is the system, if that can be quantified.

GFRD and Smoldyn are both based on the Brownian Dynamics method for diffusion. In contrast to our GMP method, the diffusion process is not executed on a lattice. As in other spatial methods, we can almost find that the diffusion of individual molecules is precisely the most computing-demanding process of the Reaction-Diffusion process. Between these two methods there is an essential difference: the first is event-driven, while the second uses a fix time-step. Because of the few non-reactive collisions (or events), the event-driven methods waste less time in checking for non-reactive events, resulting in better performance. This

advantage is more accentuated as long as the number of molecules is low.

Choosing the right Δt in the latter is not a completely arbitrary procedure since one has to assure that the probability of events per time step is small. In GFRD the maximum simulation time-step during an iteration depends on the distance of the molecules to the target. If the total number of molecules decreases, the inter-particle distances increase, thus allowing for a larger time-step possible.

For the lattice-based methods, GMP and MesoRD, a more direct comparison can be established. In contrast to the previous grid-free methods, the choice of the discretisation size is the limiting factor, because it is coupled to the diffusion. These two methods also present a similar behaviour to the previous lattice-free methods. In the line of GFRD as a event-drien methods, MesoRD method computes only for the next event (as from the system's current state). Then the local state of the site where the event is fired is updated (the essence of the next subvolume method as described in Elf et al. (2003) ; Hattne et al. (2005)), while GMP scans the whole lattice performing all the individuals local updates to the site. Nevertheless MesoRD is not the same as an event-driven method since the evolution of the system is carried out step by step. The main difference is that diffusion events of single particles are treated as reactions, and these are exponentially distributed events just like normal reactions. GMP simplifies this step by using the bulk diffusion described in section 2.2.2.

Another important aspect is that MesoRD only computes the propensity function (a_μ) of the site where the next event occurs, avoiding unnecessary recalculations. However, the cost of maintaining the data structure and inserting sorted events has some penalty. On the other side, for moderate to dense systems (at least in some parts of the system), we find that both perform similarly or GMP is faster due to the advantage of the diffusion process. Notice that just scanning of the lattice and decide that nothing needs to be done results in a considerable time penalty. Such effect was relevant for systems where the reaction only occurred in one and fix site as in the cases studied in section 3.1. We should mention that both methods would also benefit from the performance gain offered by the Gibson-Bruck method or even τ-leap methods for reaction.

The complex dynamics of a Reaction-Diffusion system can vary and given a state that favours one method at one point in time, might evolve to another state that favours the other. Nevertheless, we have seen comparable execution times between MesoRD and GMP for the CheYdiffusion case in section 3.3.

Using similar arguments one can explain differences in the cost of performing diffusion in MesoRD and GMP (the first follows an event-driven scheme, the latter uses a fixed Δt). Obviously the number of molecules of a given species in the sub-volume has to be used instead of the inter-particle distance. Then the average time between diffusive jumps, Equation 2.1 , in MesoRD is inversely proportional to that quantity (see the caption of Table 3.9). Additionally, thanks to the next sub-volume method, MesoRD finds sub-volumes where the next event occurs instead of looping over the whole volume. On the other hand GMP favours higher densities because, contrary to all other methods, particles can be diffused in bulk rather than one by one. The computational cost of the two RDME-level

methods differs also in scaling with the number of reaction channels NR. The usage of the SSA scheme in GMP results in linear scaling with NR; MesoRD achieves approximately $\log NR$ scaling. Note that a diffusion event in the latter method is treated similarly as a reaction, and is also entered into an event queue.

2.4 Detailed Analysis of the GMP Reaction Mechanism

Throughout this thesis we aim at simulating diffusion-limited reactions. Till now we have seen that we can tackle the complexity of this process by means of explicit spatial simulations using, for instance, the GMP method. Applications of this method are presented in Chapter 3. Nonetheless a problem highlighted at the beginning still remains. The problem is that the reaction method used on each lattice site is suited only for reactions which are reaction-limited. Because of the rather restrictive constraint of the local well-mixed condition, we are either coerced to use very small lattice sites—and shifting the simulation towards the micro-scale— or some local error. This type of error is present in the class of methods that uses a lattice discretization with Gillespie's type of method for reactions. Afterall, diffusion-limited reactions are all about space, and therefore the positions of the molecules in space is relevant. It is instructive to show what type of error are introduced in the small timescale, which is that smaller than the diffusion time τ_D as shown in Eqn. 2.1.

2.4.1 The Reversible Diffusion-Limited Reaction of a Pair of Molecules

One the most relevant, simple and ubiquitous reaction in biochemistry for diffusion-limited reaction is the reversible reaction such as

$$A + B \leftrightarrows C. \tag{2.6}$$

A pair of molecules that underwent reaction (A and B producing C) unbind and therefore they have the potential to react again. Because they unbind next to each other their probability of rebind or recombine is high. The unbinding of C produces A and B. In physical terms both molecules are not exactly at contact. Smaller molecules, such as water among others, may lie between them. The separation of molecules can also be due to repellent molecular electrical charges. Thus, a model considers putting the unbound particles at a very small distance from each other (less than the diameter of the particle.) In open systems—which is not the case for cell biology—the distribution of the next reaction time follows a power-law decay for $t \to \infty$ (see the $t^{-3/2}$ line in Fig. 2.4.) But here we deal always with closed systems, and therefore the boundaries play a major role. For closed systems we do not speak of a power-law tail, as the one that has the Smirnov's distribution in Eqn. 2.4, but rather of a power-law head and an exponential tail. These features are captured by Eqn. 2.7 and depicted in

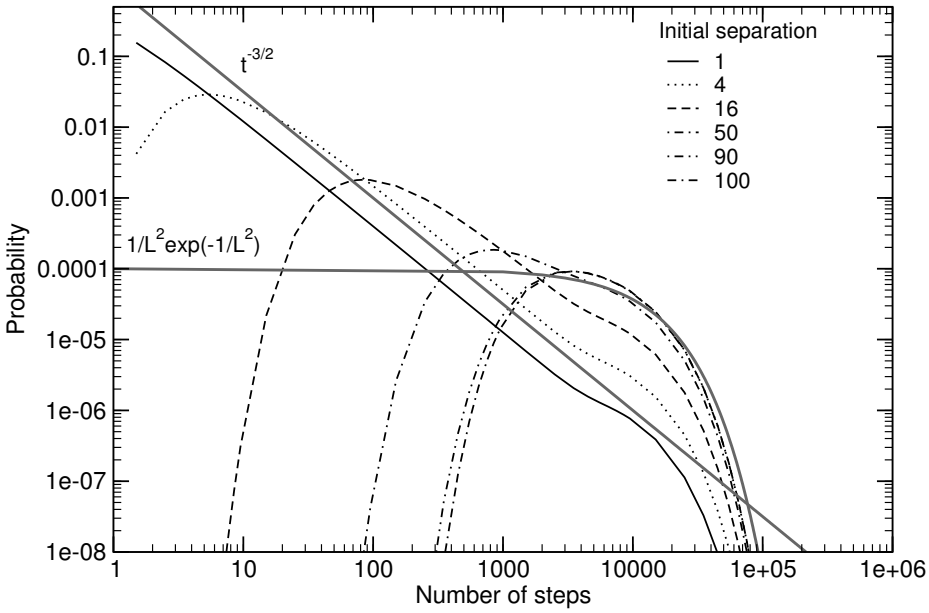

Figure 2.4: *Probability distribution of first passage time in 1D for one random walker starting at a fixed initial separation x_0 from the target at $L = 0$ with reflective boundary condition at $L = 100$.*

Fig. 2.4.

$$f(0,t|x_0,t_0) = \frac{1}{\pi L^2}\sum_{j=0}^{n}\exp\left(-\frac{(1+2j)\pi\sigma\sqrt{t}}{2L}\right)^2(1+2j)\pi^2\sigma^2 \tag{2.7}$$

$$\times\cos\left(\frac{(1+2j)\pi(L-x_0)}{2L}\right)\sin\left(\frac{(1+2j)\pi}{2}\right), \tag{2.8}$$

Strictly speaking the fully detailed *pdf* as shown in Fig. 2.4 has three regions. Before the power-law regime there is an ascension regime due to the fixed initial separation of the two particles. Because this ascension time is very steep, we might refer to this as a cut- off time, since the probability of an event happening before the mode of the distribution is very small.

To further study of the distribution of rebinding times in a more realistic scenario for biochemical reactions and shed light into the error in the small timescale, we use the reversible reaction model shown in Eqn. 2.6. The system is initialized with a single molecule of A and one of B in a cubic volume of 1 μm^3—typical for a bacterial cell. The spatial reversible reaction model is, for example, applicable to gene activation or inhibition, as a part of the gene expression in prokaryotes (van Zon and ten Wolde, 2005 ; Rodriguez et al., 2006 ; Dobrzynski et al., 2007). We require A to be fixed at the center of the

Table 2.1: *Statistics of the inter-reaction times distribution extracted from the duration of the dissociated state for GMP using a lattice size L and GFRD.*

	L				
	30	30 (p=.75)	50	80	GFRD
mean (s)	0.200	0.200	0.200	0.203	0.196
variance	2.904	2.903	5.553	8.224	5.894

domain and B can diffuse with a diffusion coefficient of $1\,\mu m^2 \mathrm{s}^{-1}$. To validate the results, we use the method presented by van Zon and ten Wolde (2005): the Green's Function Reaction Dynamics (GFRD), which is a method specialized in this type of reaction. GFRD is a method which is continuous in space and time at a single particle level such as Brownian Dynamics, and incorporates reactions for pairs of particles drawing from the Smoluchowski theory (Rice, 1985).

With GMP simulations both particles are initially placed in the same lattice site. The forward reaction rate used by the local Gillespie method is $\kappa_f = 3\,10^3\ \mu\mathrm{M}^{-1}\mathrm{s}^{-1}$ and the backward is $\kappa_b = 5\,\mathrm{s}^{-1}$. This configuration results in an average association and disassociation reaction every $5\,\mathrm{s}^{-1}$.

2.4.2 Effects of the Operator-split on the Distribution

In Fig. 2.5 we show the probability density functions of the inter-reaction times (time the particles A and B are in dissociated state.) Each simulation contains more than 2×10^6 data points, of which more than 98.7% lie in the power-law region $t < \tau_S$. Because of the presence of gaps in the distributions, these distributions cannot be directly compared to each other. A more convenient way to compare them is by looking at the mean and variance of the results extracted from the data points (see Table 2.1 for a summary.) As expected, the average inter-reaction time interval is maintained across simulations at $0.2\,s$, corresponding to the forward rate κ_f (Dobrzynski et al., 2007). The variance, instead, differs and only the case $L = 50$ remains in good agreement with the reference result of GFRD, an extended analysis of this observation is provided in Dobrzynski et al. (2007)

In order to gain insight into the details of the distribution of rebinding events, which span over seven orders of magnitude, we use logarithmic histograms in a log-log scale. The plot of the distribution of the inter-reaction time intervals in Fig. 2.5 shows the influence of the lattice size on the variance. Increasing the lattice size reduces the maximum values of the inter-reaction time interval, for example the case $L = 30$, and increases the maximum values of this interval for the case $L = 80$.

In the GFRD reference result shown in Fig. 2.5, we distinguish two regions of the inter-reaction time distribution $\widehat{\Gamma}$: a power-law and an exponential. Note

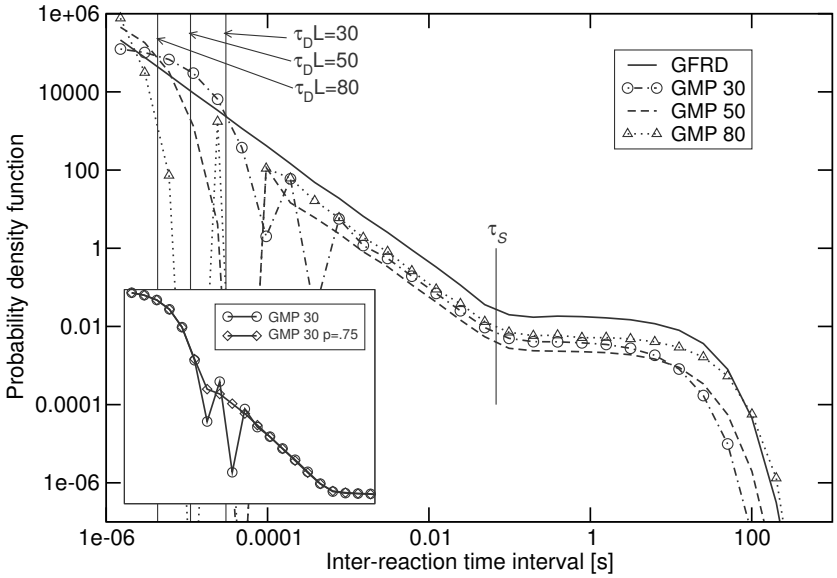

Figure 2.5: *Comparison of $\widehat{\Gamma}$ distributions of inter-reaction times between the split-operator method GMP with lattices $L = 30$, 50 and 80, and the exact solution from GFRD. The inset shows a smoother distribution due to a significant reduction for the gaps when using rest- particle probability $p = 0.75$.*

there is a pronounced transition point between these two regions. This point is system-size dependent, but lattice-size independent, and its value is $\tau_S \approx 0.068\,\mathrm{s}$, corresponding to the average time required for a particle starting from the center of the cube to reach the boundary.

By splitting the diffusion from the reaction process in predetermined intervals, according to section 2.2.1, the method introduces some artifacts in the distribution of the inter-reaction intervals. We note in the power-law regime in Fig. 2.5 that the power-law region differs significantly from that of GFRD. In this distribution we can distinguish three regions.

First, for times in the range of $t < \tau_D$, $\widehat{\Gamma}$ is exponential because of the Gillespie reaction method.

Second, a region with a power-law behaviour in the range $\tau_D < t < \tau_S$. In this region the time for the next reaction events is dominated by the diffusion process across lattice sites, since B has left the site of A and the next reaction interval is larger than τ_D. Additionally, $\widehat{\Gamma}$ has some discontinuities (or gaps) in the lower range of this region, showing up in Fig. 2.5 as large deviations from the expected mean. These fixed width gaps become less relevant in the upper range of the region, where the power-law function in Fig. 2.5 is smooth. These gaps originate from the diffusion process, because the particle can only return to the reactive site after an even number of steps. Reaction times $(2n)\tau_D < t < (2n+1)\,\tau_D$ are impossible and produce gaps in the distribution. We recall from section 2.2.2

that these gaps are common among Lattice Gas Automata-based simulators and they are significant when there are sharp inhomogeneities between adjacent sites. It is possible to reduce the size of these gaps by using rest particles, which means increasing p_r in Eq. (2.2). Simultaneously we allow the particle to stay at the origin site at odd time-steps, and reduce the diffusion time-step. We can get a higher accuracy at the expense of a larger computational time, proportional to $1/(1-p_r)$ due to the smaller diffusion time-step. In the inset of Fig. 2.5 we see that by using a rest-particle probability of $p_r = 0.75$ the power-law distribution appears continuous, yet with smaller gaps than for the distribution with $p_r = 0$.

The third region of $\widehat{\Gamma}$ corresponds to the exponential tail for times $> \tau_S$. This region is equivalent to a reaction-limited situation. In this region, the particle has lost the memory of where it originated, hence the exponential distribution.

2.5 Discussion

We have developed the GMP (Gillespie-Multi-Particle) reaction-diffusion method along the lines of other spatial stochastic kinetic model solvers such as Stundzia and Lumsden (1996) ; Ander et al. (2004) ; Bormann (2001) ; Hattne et al. (2005) ; Bormann (2001). GMP's distinguishing feature is the separation of the reaction and diffusion processes at the scale of a lattice site. The consequent locality of the operations results in a fast algorithm. Since no spatial information of particles at the lattice site scale is used, GMP is suited for parallel computing. Such computational power will be required when simulation studies of whole-cell processes become necessary as experimental biology advances. GMP can cope with both reaction-limited and diffusion-limited problems, as well as with confined distributions of chemicals in biological systems. For complex biochemical networks a trade–off in choosing the lattice size is necessary, since different reaction rates and diffusion coefficients result in different optimal lattice sizes. Consequently, some accuracy is lost in order to gain computational speed when compared to hard-sphere particle models. Nevertheless, the results can be used to guide experimentalists since such loss can be controlled and minimized.

The GMP method takes into account the principal sources of fluctuations, but the class of lattice-discrete methods, to which GMP belongs, smoothes out some noise if the lattice size is too large, because of the use of Gillespie as a reaction method. Note, however, that λ_{rmfp} is a safe lattice size, which is valid for small concentrations. For larger concentrations λ_{rmfp} decreases and in that case the slope of the gradients and the error we allow determine the optimal lattice size. For more complex biochemical systems there is no single optimal lattice size.

We showed the applicability of GMP as an approximate solver for diffusion-limited reactions. We measured the validity in terms of the distribution of a pair of reacting particles as described in section 2.4.1. The coarseness of the lattice has a direct impact on the variance of the reaction process simulated, as seen in Table 2.1. These results show that the proper lattice size lies in the order of the diameter of two particles. The regime used in section 2.4.2 lies in the lower

limits of its applicability, because we are using a lattice site size which is close to what would be the real size of an individual particle (or molecule). The Γ distribution obtained (see Fig. 2.5) shows the effects of using the operator-split and the Gillespie method. Both effects are only noticeable in the lower time scale ($<< \tau_S$). It is possible to reduce the effects of the operator-split by using a rest particle probability larger than 0. However, to remove the exponential head of Γ another reaction method should be used instead of Gillespie.

2.6 Conclusions

In this chapter we have introduced and described some basic properties of the operator-split, lattice-based diffusion-reaction method: Gillespie-Multiparticle Method (GMP), highlighting the side effects at the interaction between two particles in a closed system. Our principal motivation has been to test how computationally simpler and approximate solvers to the RDME perform, and to be able to infer the quality of the final results from elementary bimolecular reactions. The small differences in the resulting probability distribution of a reaction times suggest that the impact in a complex simulation might be small, since the time scale of the events is too fast for biochemical reactions. This method is aimed at addressing complex biochemical networks in which a certain small degree of inaccuracy is tolerable in the diffusion-reaction scheme, unlike other more physical oriented chemical systems that can be described with a more accurate method.

In the following chapter we apply the GMP method to a number of biochemical systems found in bacteria. We pay special emphasis in the spatial and stochastic nature of those systems. These allow us to gain insight into the reliability of GMP as well as to the benefits (or limitations) of lattice-based methods, which usually do not address the specific issue of lattice size.