INFERRING THE DISTRIBUTION OF INTER-REACTION INTERVALS FOR DIFFUSION-LIMITED REACTIONS ON LATTICES

JORDI VIDAL RODRÍGUEZ* and JAAP A. KAANDORP†

*Section of Computational Science, University of Amsterdam, Kruislaan 403
 1098 SJ Amsterdam, The Netherlands*

**jrodrigu@science.uva.nl*

†jaapk@science.uva.nl

We study the operator-split method Gillespie-MultiParticle method (GMP) for a diffusion-limited reaction in a finite medium by looking at the detailed distribution of reaction times for a broad range of time scales. The diffusion-limited reaction lies at the lower limits of the applicability of the GMP method. We propose an empirical distribution for the inter-reaction times (duration of the dissociated state) that accounts for the power-law and exponential regimes. The knowledge of this distribution can be used to change the reactive method of GMP and enable it to use coarser discretization, with consequent improvement in performance.

Keywords: Reaction-diffusion; finite systems; power-law; gene transcription.

1. Introduction

Recent technological advances allow to study individual reactions in single cells, and have provided new detailed dynamical data of biochemical reactions.⁵ At such small scale, the discrete and stochastic nature of reactions should be modeled, since the system may exhibit states which cannot be predicted by deterministic models.^{3,8}

Under circumstances of low number of molecules and slow diffusion due to crowding or viscosity and rapid reaction rates, the reactions become diffusion-limited and spatial inhomogeneities arise which induce larger fluctuations of the system.^{4,14} The diffusion-limited effects, in contrast to the reaction-limited, modify the distribution of the reactions. There are two main causes for such alteration: i) Due to the random initial separation of reactants and the diffusion-limited conditions mentioned above, the probability of reacting at $t = 0$ is close to zero. The actual distribution has a maximum at a non-zero time, instead of at $t = 0$ as is the case of the exponential distribution used in the reaction-limited case. ii) For reversible reactions, the separation of the reactants after dissociation is fixed. This type of reaction is typical of enzymatic reactions and of gene transcription.^{16,9} In this paper we focus on this type of reversible reaction

using explicit spatial stochastic simulations. Where A , B and C are molecules with

forward and backward reaction constant κ_f and κ_b respectively.

Theory for diffusion-limited reactions shows that for infinite media, the distribution of the reactive events follows a power-law $t^{-d/2}$, which is d -dimension dependent.^{1,12} In contrast to infinite media where the particle can diffuse infinitely far away from the reactive particle, in finite media the particle is bound within a domain. In the later case, the power-law distribution remains unaffected for times smaller than the typical time for the particle to reach the boundary (τ_S). For longer times, the system behaves as if it were reaction-limited, thus the distribution remains exponential. Unfortunately, analytical expressions for finite media are difficult to obtain and have complex terms.^{11,10}

We can use some approximations of finite and infinite systems to infer an approximate distribution for our finite system. Redner studied analytically the first-passage reaction time, $F(X, t)$, in a finite domain of a particle initially placed very close to the absorbing boundary.¹¹ Since an analytical form is not possible in the time domain, he suggests that F has an approximated distribution like:

$$F(X, t) = t^{-\alpha} \exp(t/\tau_S), \quad (2)$$

where X is the state of the particles: associated or dissociated, t time, $\alpha = 3/2$ the characteristic power-law exponent for a three dimensional system, and τ_S a system size related parameter. Note that Eq. 2 does not reflect the initial separation mentioned before in case i). Its main property is that it shows the typical power-law behavior for small t and an exponential tail, dependant on the system size. In this paper we will focus on the analysis of a pair of molecules, with zero separation of dissociated particles for simplicity.

Explicit spatio-temporal simulations are commonly used for simulating complex biochemical networks at the cell level, that cannot easily be solved analytically.^{15,13} They rely on space discretization into a lattice, and running a local reaction and diffusion method. Discretization allows us to solve a smaller spatial and temporal domain which has a simpler solution. Despite system inhomogeneities, the smaller lattice sites should be considered nearly homogeneous, which is a simpler case to solve. Additionally, separating the reaction from the diffusion process simplifies the handling of both processes.^{13,2} A proper discretization is an important feature to obtain better approximations. When the discretization is too small, simulations of complex networks cannot be simulated efficiently. In reaction-limited systems efficient solvers use knowledge of the reaction distribution, knowledge of an diffusion-limited reaction distribution may improve explicit simulations.

In this paper we use the operator-split, or time-splitting², method GMP (Gillespie-MultiParticle) to study its suitability for simulating diffusion-limited reactions.¹³ We first look at the detailed distribution of inter-reaction intervals for a broad range of time scales to infer which regimes are properly modelled. The test case is an irreversible reaction of a pair of particles. It helps us to validate the method and to infer the behavior for this basic case. Finally, based on the empirical observations, we propose a distribution for the inter-reaction time intervals.

2. Gillespie-MultiParticle (GMP): an Operator-Split Method

We use the Gillespie-MultiParticle (GMP) operator-split method¹³ to solve the reversible reaction in (1). This method is based on the multiparticle lattice gas automata and it allows to handle complex reaction networks as well as it allows for an arbitrary number of dimensionless particles to be in the same site.⁶ As the name indicates, the reaction and diffusion processes are split (or executed independently). First a diffusion step of all particles is executed, followed by a reaction step in every lattice site.

To simulate a biochemical process within a cell, we approximate the cell's cytosol by a discrete cubic lattice of side length l which contains L^3 lattice sites. The impermeable cell membrane, that encloses the cytosol is represented by a reflective boundary condition. If a particle were to diffuse outside the domain after a diffusion step, we force that particle to remain at the same position of origin. The execution of a diffusion step in all lattice sites and for all particles, represents a simulation-time increase of τ_D given by:

$$\tau_D = \frac{1}{6D} \left(\frac{l}{L} \right)^2 (1 - p), \quad (3)$$

where l is the length of the system, L the number of sites (or divisions) along one dimension, D the diffusion coefficient and p is the rest particle probability. This last parameter reduces the diffusion time step and avoids the situation that a particle can only return to its original site in a even number of steps. Consequently it increases the resolution of the diffusion process in time and in some cases also in space. Because we assume a local well-mixed domain, we model the reaction by using the Gillespie method with the macroscopic (under well-mixed conditions) reaction rate κ .⁷ This assumption requires the lattice site to be small enough for the diffusion not to have an effect. The interval between reactive events is τ_R and its distribution at a smaller time scale than the diffusion time step (τ_D) is, consequently, exponential.

3. The Reversible Diffusion-Limited Reaction of a Pair of Molecules

To study the distribution of inter-reaction times we use the reversible reaction (1) of a pair of molecules, A and B , in a cubic volume of $1\mu\text{m}^3$. We can use this reaction as a simplified model of gene transcription. We assume that A (the gene) is immobile, as well as the gene RNA-polymerase complex C , whereas B is a RNA-polymerase molecule. When C dissociates into A and B , it is assumed that A and B are at contact. Only B can diffuse by executing a random walk on the lattice.

We require A to be fixed at the center of the domain and let B diffuse with a diffusion coefficient of $1\mu\text{m}^2\text{s}^{-1}$. To validate the results, we use Green's Function Reaction Dynamics (GFRD).¹⁶ 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 described by the Smoluchowski theory.¹²

Table 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

In the following simulations, both particles are initially placed in the same lattice site, and they are in dissociated state. The forward reaction rate used by the local Gillespie method is $\kappa_f = 3 \cdot 10^3 \mu\text{M}^{-1}\text{s}^{-1}$ and the backward is $\kappa_b = 5 \text{ s}^{-1}$. This configuration results in an average association and disassociation reaction every 5 s^{-1} . In the following sections, we will focus on the distribution of the duration of the dissociated state intervals, the inter-reaction time.

3.1. Effects of the Operator-split on the Distribution

In Fig. 1 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 $2 \cdot 10^6$ reactive events, of which more than 98.7% lie in the power-law region $t < \tau_S$. Because there gaps in the distributions, the 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 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 . The variance, instead, varies and only the case $L = 50$ is in good agreement with the reference result of GFRD.

In order to get insight to the details of the distribution of inter-reaction events, which spans 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. 1 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 $L = 80$. Note that the average time for a particle to return to its origin in a finite cubical space is L^3 steps. If each step takes τ_D , then the average time is proportional to L . The duration of the intervals between returns increase and hence the variance as seen in Table 1. These effects have not been considered in previous works, since they mainly contrast average results against deterministic solutions.¹⁵

In the GFRD reference result shown in Fig. 1, we distinguish two regions of the inter-reaction time distribution $\hat{\Gamma}$: a power-law and an exponential. Note there is a prominent transition point between these two regions. This point is system size dependent, but lattice size independent, and its value is $\tau_S \approx 0.068 \text{ s}$, corresponding to the average time required for a particle starting from the center of the cube to reach the boundary.

Fig. 1. Comparison of $\hat{\Gamma}$ distributions of inter-reaction times between the split-operator method GMP with lattices $L = 30, 50$ and 80 , and the exact solution from GFRD. Note that the exponential head of $\hat{\Gamma}$ extends slightly beyond τ_D given a lattice L . The inset shows a smoother distribution due to a significant reduction for the gaps when using rest-particle probability $p = 0.75$.

By splitting the diffusion from the reaction process in predetermined intervals, the method introduces some artifacts in the distribution of the inter-reaction intervals. We note in the lower time scale in Fig. 1, that the power-law region differs significantly from that of GFRD. We can distinguish three regions. First, for times in the range of $t < \tau_D$, $\hat{\Gamma}$ is exponential because of the Gillespie reaction method. Second, a region with a power-law behavior 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, $\hat{\Gamma}$ has some discontinuities (or gaps) in the lower range of this region, showing up in Fig. 1 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. 1 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. These gaps are common among Lattice Gas Automata simulators and they are significant when there are sharp inhomogeneities between adjacent sites.⁶ It is possible to reduce the size of these gaps by using rest particles, which means to increase p in Eq. (3). Simultaneously we allow the particle to stay at the origin site in 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

$1/(1-p)$ due to the smaller diffusion time step. In the inset of Fig. 1, we see that by using a rest-particle probability of $p = 0.75$ the power-law distribution appears continuous, yet smaller gaps than for the distribution with $p = 0$ still exist. The third region of $\hat{\Gamma}$ corresponds to the exponential tail for times $> \tau_S$. This region is equivalent to a reaction-limited situation.

3.2. Compound Distribution

In the previous section we have discussed the dual composition of the probability density function of the inter-reaction times for a diffusion-limited reversible reaction in a finite system: a part characterized by a power-law and another by an exponential.

We argue that diffusion-limited reversible reactions with a partially absorbing boundary condition can be approximated by a linear combination of an absorbing boundary condition with a diffusion-limited reaction and a reaction-limited reaction. We can describe this process by the probability distribution function:

$$\Gamma(X, t) = \beta t^{-\alpha} \exp(-t/\tau_S) + \lambda \exp(-t\lambda), \quad (4)$$

where the first term comes from the general result shown in Eq. (2) of a diffusion-limited reaction with absorbing boundary condition in a finite system.¹¹ The second term corresponds to the overall reaction rate $\lambda^{-1} = 1/(4\pi D\sigma) + 1/\kappa_f$, where D is the diffusion coefficient of B , σ is the radius of the particles (5 nm), and κ_f is the intrinsic reaction rate.¹⁶ Next we determine the β parameter of the first term. Applying the scaling $\Gamma^* = \Gamma/\lambda$, we notice in Fig. 2 that the power-law part of the distribution remains constant for the different reaction rates. If we consider the intersection point of $\beta t^{-3/2} = 1$ at $t = \tau_S$, we find

$$\beta \propto \lambda \tau_S^{3/2}. \quad (5)$$

From Fig. 3 it can be observed that the power-law factor should be set to a slightly larger value, such as $t = \omega \tau_S$. By using $\omega = 2$ we obtain the theoretical distribution for Γ shown in Fig. 3.

When we make the reaction system in (1) more reaction-limited, by decreasing κ_f , the power-law component becomes less dominant as β decreases linearly in λ . Conversely, as the system becomes more diffusion-limited, by increasing κ_f , GMP reaches a point where the measured distribution $\hat{\Gamma}$ does not change significantly. This is illustrated in Fig. 2 where the result for $\kappa_f = 3 \cdot 10^5$ overlaps with the one for $\kappa_f = 3 \cdot 10^3$. This limiting situation approximates the absorbing boundary condition as the one used by Redner to derive the approximation of Eq. (2).¹¹ However our results shown in Fig. 2 show a stronger influence of the exponential term. This exponential tail is a consequence of the diffusion process used in the GMP method, which coerces the particle to leave the site of the A particle.

Using the proposed Γ distribution in Eq. (4), Fig. 3 shows a fit of the result obtained with GFRD. For clarity, we also plot each component of Eq. (4) to illustrate

Fig. 2. Scaled distribution function Γ^* showing the constant powerlaw part. Γ is scaled by λ^{-1} .

Fig. 3. Fit of the probability distribution function of reaction intervals (using $\hat{\beta}$, $\hat{\tau}_S$ and $\hat{\lambda}$), the analytical Γ distribution using β , τ_S and λ and the GFRD distribution for comparison. For clarity of the Γ distribution construction, we also show its components.

the contribution of each term of the proposed distribution with the corresponding fitted parameters with values $\hat{\beta} = 0.002$, $\hat{\tau}_S = 0.057$ and $\hat{\lambda} = 0.064$. These fitted results, shown in Fig. 3, agree with the analytical results for Γ with $\tau_S = 0.068$ and $\lambda = 0.062$ obtained previously.

4. Conclusions

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 3. The coarseness of the lattice has a direct impact on the variance of the reaction process simulated, as seen in Table 1. These results show that the proper lattice size lies in the order of the diameter of two particles. This regime lies in the lower limits of its applicability, because we are using a lattice site size which is close to size of an individual particle. The Γ distribution obtained (see Fig. 1) shows the effects of using the operator-split and the Gillespie method. Both are only noticeable in the lower time scale ($\ll \tau_S$). 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 in place of Gillespie.

The reaction studied in this paper is typical for gene expression models, where the particle representing the gene is immobile.¹⁶ We have found that the distribution Γ in Eq. 4 can be approximated well by combining two simple distributions: a power-law for short-times and an exponential one for larger times. A good approximation of Γ can be calculated from τ_S and λ . However, for exact simulations ω needs to be determined. Another improvement of the gene expression model is the addition of a separation of the dissociated gene and a RNA-polymerase molecule (A and B .)

We have seen that the regime that cannot be approximated well by using GMP is the short time scale. Since the distribution Γ in the vicinity of this regime is approximated well by Eq. 2, we might improve GMP accuracy by introducing knowledge of this distribution in a modified reaction method. Such modification would enable GMP to use a coarser discretization, with a probable improvement in performance and accuracy.

Acknowledgments

We thank Joke G. Blom and Maciej Dobrzyński for the discussions and providing the GFRD data, and Alfons G. Hoekstra for comments on the manuscript. This work is funded by The Netherlands Organization for Scientific Research in Computational Life Sciences NWO-CLS-635.100.007.

References

1. N. Agmon and A. Szabo, *J. Chem. Phys.* **92**, 5270 (1990).

2. D. Alemani, B. Chopard, J. Galceran, and J. Buffle. LBGK method coupled to time splitting technique for solving reaction-diffusion processes in complex systems. *Phys. Chem. Chem. Phys.* **7**, 3331 (2005).
3. A. Arkin, J. Ross, and H.H. McAdams. *Genetics* **149**, 1633 (1998).
4. A. Becskei, B. B. Haufmann, and A. van Oudenaarden. *Nature Genetics* **37**, 937 (2005).
5. L. Cai, N. Friedman, and X. Sunney Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358 (2006).
6. B. Chopard, L. Frachebourg, and M. Droz. *Int. J. Mod. Phys. C* **5**, 47 (1994).
7. D.T. Gillespie. *J. Phys. Chem.* **81**, 2340 (1977).
8. Thomas B. Kepler and Timonthy C. Elston. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys. J.* **81**, 3116 (2001).
9. S. Krishna, B. Banerjee, T.V. Ramakrishnan, and G.V. Shivashankar. *Proc. Natl. Acad. Sci.* **102**, 4771 (2005).
10. W. Nadler and D.L. Stein. Reaction-diffusion description of biological transport processes in general dimension. *J. Chem. Phys.* **104**(5), 1918 (1996).
11. S. Redner. *A Guide to First-Passage Processes*. (Cambridge University Press, 2001).
12. S. A. Rice. *Diffusion-Limited Reactions*, vol. 25 of *Comprehensive Chemical Kinetics*. (Elsevier, 1985).
13. J. Vidal Rodríguez, J.A. Kaandorp, M. Dobrzyński, and J.G. Blom. Spatial stochastic modelling of the phosphoenolpyruvate-dependent phosphotransferase (PTS) pathway in *Escherichia coli*. *Bioinformatics* **22**, 1895 (2006).
14. S. Schnell and T.E. Turner. *Prog. Biophys. Mol. Biol.* **85**, 235 (2004).
15. A.B. Stundzia and C.J. Lumsden. *J. Comput. Phys.* **127**, 196 (1996).
16. J. S. van Zon and P. R. ten Wolde. *Phys. Rev. Lett.* **94**, 128103 (2005).