A Stochastic Particle-Based Biological System Simulator

Laurier Boulianne[1] Michel Dumontier[2] and Warren J. Gross[1]
[1]**Department of Electrical and Computer Engineering, McGill University, Montreal, QC, H3A 2A7 Canada**
[2]**Department of Biology, Carleton University, Ottawa, ON, K1S 5B6 Canada**
laurier.boulianne@mail.mcgill.ca, michel_dumontier@carleton.ca, warren.gross@mcgill.ca

Keywords: Stochastic simulation, Particle systems, Biological systems

Abstract

The simulation and visualization of biological systems is expected to enhance our understanding of biological processes towards the development of effective therapeutic treatments. Biological systems are inherently stochastic at the molecular level, exhibit modified behavior under crowded conditions and may be affected by spatial locality. Common simulation approaches fail to account for these important aspects of biological systems, in part because they are computationally expensive. Here, we describe a stochastic, particle-based simulator that takes spatial locality into account. Each particle in the system is represented explicitly on a 3D grid where only one particle can occupy a grid location. The grid structure and stochastic approach removes the need for distance calculation and particle search. We demonstrate the effect of molecular crowding and spatial locality for a simple biological system. We anticipate that this system will be useful in examining more complex systems. Finally, this system is expected to be suitable for acceleration with parallel customizable hardware, a necessary requirement towards the simulation of an entire cell.

1. INTRODUCTION

Computational cell biology is currently one of the most exciting cross-disciplinary areas of research [1]. An ambitious, long term goal of this field is to simulate a biological cell in which researchers are focused on producing accurate simulations of biological systems with molecular resolution. This task has two major challenges: 1) the construction of accurate biological models and 2) the development of scalable simulation architectures. Efforts towards realizing these objectives are expected to support therapeutic drug development against human diseases by increasing our understanding of biological systems.

Molecular systems are inherently stochastic and spatially dependent such that they require compatible simulation methods. It is well known that molecules behave with Brownian dynamics [2]. The functionality of certain proteins known as enzymes are limited by the rate of diffusion in the solution medium. An important aspect to this functionality has to do with the mobility of particles in what is increasingly believed to be a crowded environment [3]. Importantly, certain cellular responses occur as a result of single or few particle fluctuations, and this precludes the use of modeling the systems with continuum dynamics. As well, the effect of spatial localization is expected to play an important role in the behavior of the system [1]. The idealization of a "well mixed" system is unlikely to reflect biological reality where molecular complexes form scaffolds for recruitment for cellular signaling and metabolism. Indeed, stochastic and spatial considerations are necessary for the in silico simulation of biological cells.

The simulation of discrete, stochastic, spatially-dependent molecular systems is however extremely computationally expensive and most current simulators do not support all of these functionalities. A brief overview of some of the current popular simulator follows.

The stochastic simulation algorithm (SSA) from Gillespie [4] is a stochastic and dimensionless algorithm that has been used to simulate many different systems. The position of the particle is not tracked and a "well mixed" system is assumed. Virtual Cell [5] is a deterministic simulator that solves partial differential equations (PDE) to compute the concentration and location of every species. The simulation space is divided into compartments that have no specific geometry. StochSim [6] is a stochastic simulator that tracks individual molecules on a discretized 2D grid. Simple 2D structure can be created where nearest-neighbour interactions of molecules can be simulated. MCell [7] is a stochastic simulator that tracks individual particles in a continuous 3D space. The diffusing particles move independently with Brownian dynamics. 2D membrane surfaces and sites of chemical reactions are mapped in the 3D volume. A ray-tracing algorithm is used to detect collisions between particles and the surface, resulting in chemical reactions. SmartCell [8] is a stochastic simulator that divides the 3D space into smaller cells that can contain several molecules. Inside each of these small cells the well mixed assumption holds and an algorithm similar to the SSA is used. Particles can diffuse to neighboring cells. ChemCell [9] induces stochastic behavior for individual particles with a dependency on diffusion coefficient in a 3D volume, and evaluates chemical events by converting rate laws into probabilities. A binning algorithm is used to find the neighbor particles that are susceptible to enter in a reaction. The binning is an $O(N)$ algorithm in the number of particles N. While many of these methods are effective in simulating biologi-

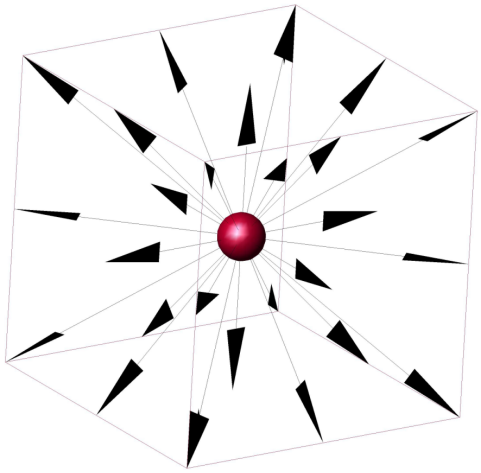

Figure 1. The 27 possible directions of a D3Q27 grid.

cal processes, they do not address the issue of parallelization, performance and scalability which are necessary for whole cell simulations.

We have developed a stochastic simulator that handles spatial locality, very low particle concentrations and collision between particles using a discrete 3D grid. This paper is structured as follow. Section 2 describes the simulation model. Section 3 gives preliminary results of our model. We discuss current challenges and future directions in Section 4. Conclusions are offered in Section 5.

2. THE SIMULATION MODEL

We propose a simplified model for molecular movement and interaction. Molecules are represented as particles that move within discrete volumes in discrete time steps. An integer-addressed 3D grid avoids floating-point computation and distance calculations. These choices were made with the future goal of enabling highly parallel, large-scale simulations using custom hardware. The model components including the grid, particle movement and interaction, and cellular geometry are described below.

2.1. Grid-Based Stochastic Model

The simulation volume is divided into a three-dimensional grid of *voxels* on the nanometer scale. The grid follows the D3Q27 model shown in Figure 1 meaning that a given voxel has access to its 27 neighbors (including itself) for movement and reaction and anything outside this immediate neighborhood is ignored. Each voxel may contain at most one particle. Together, these remove the need for distance calculation while enabling a tremendous amount of parallelism that can be exploited with parallel architectures. The voxel dimensions can be selected to obtain a trade off between space resolution and total model volume.

2.2. Particle Movement

A particle represents macro-molecules such as protein, small molecules such as ions, inert particles that contribute to molecular crowding or complex structures such as membranes. Particles can move in the grid and interact with each other. A particle can only move in one of 27 surrounding locations, including staying in the current location. A particle can move at most once per time step or turn. A "moving ratio" between 0 and 1 representing the probability of movement at every turn is set for each different species. Particles with two different "moving ratios" have different diffusion speeds. The random selection of the movement direction results in the particle following a Brownian random walk. The particle data structure contains the type of the particle and flags indicating whether it has moved or reacted in the current time step (turn).

2.3. Probability of Interaction

Particles may interact with each other when in spatially adjacent grid cells. Common interactions include aggregation events such as molecular complex formation/dissolution or conversion events such as chemical reactions. The probability of reaction per time step is derived from the rate of reaction. Particles may react only with their immediate neighbors and only once per turn. Complex reactions involving more than three particles are decomposed into several elementary reactions of up to three particles. Three different reactions involve three or less particles: one reactant and one product, one reactant and two products and finally, two reactants and one product. Let's consider the two reactions that involve a single reactant.

$$A \rightarrow B, \tag{1}$$

$$A \rightarrow B + C. \tag{2}$$

Both reactions have a forward rate of reaction k in units of time^{-1}. The time step which is the elapsed time between 2 successive iterations of the algorithm is t seconds. Assuming N particles of type A are in the system, then in both cases the expected number of reactions per turn is given by Nkt. Considering each particle individually, a particle reacts at each turn with probability equal to kt. In our stochastic model, a uniform random number R_n ranging between 0 and 1 is generated for each particle and the reaction takes place if and only if $R_n < kt$. For a given k, it is possible to have a product kt larger than one. Should that happen, a smaller time step t is required in order to maintain a coherent simulation.

In a reaction with only one reactant and one product, the reactant is replaced by the product if $R_n < kt$. In a reaction with one reactant and two products, a search is first conducted in the surrounding area of the reactant if $R_n < kt$. If the search detects at least one free voxel in the cube surrounding of the

particle, the reaction takes place and the second product is positioned in that free location while the first product is placed at the position of the initial reactant. The reaction is blocked if no free position is found. Consider the following reaction with two reactants:

$$A + B \rightarrow C, \tag{3}$$

with a rate constant k in units of (molarity*time)$^{-1}$ and a time step between each iteration of t second. The total number of reactions N_r is given by

$$N_r = \frac{kN_aN_bt}{A_vV}, \tag{4}$$

assuming N_a particles of type A, N_b particles of type B, a volume V and Avogadro's number A_v. On average, the desired number of reactions in our system should be equivalent to the result of the above equation. In a well-mixed system, the number of (A,B) pairs that are close enough to each other to generate a reaction is given by $N = N_aN_bV_c/V$ where V_c is the volume of the cube containing the 26 neighboring voxels and V is the total volume of the simulation. If each of those pairs react with probability P, then $N_r = NP$. Setting the two equations $N_r = NP = kN_aN_bt/(A_vV)$ gives the equation

$$P = \frac{kt}{A_vV_c}. \tag{5}$$

The formula is independent of V, N_a and N_b as expected. Similar to the previous case, for a given rate constant k, it is possible to have a set of parameter t/V_c such that P is greater than 1. If that is the case a smaller time step or larger voxels (proportional to V_c) have to be selected. A smaller time step reduces proportionally the number of reactions taking place during a turn. Similarly, a greater V_c increases the search area which increases the number of (A,B) pairs that can react together. Since the number of reactions must remain the same, the probability of reaction for each of those pairs is decreased. Each turn, a random number R_n between 0 and 1 is generated for the first of the two reactants. If $R_n < P$, then the first reactant will search its surrounding area for the second reactant. If one is found, the reaction takes place and the product is placed at the location of the first reactant. If no reactant is found, the reaction is aborted.

2.4. Complex reactions

More complex reactions are implemented by cascading several elementary equations. Complex reactions are broken down into a series of simpler reactions by introducing "temporary" species. Consider the following reaction with 1 reactant and 5 products,

$$A \rightarrow B + C + D + E + F, \tag{6}$$

where k is the rate of reaction in units of time^{-1}. For each additional product exceeding two products, a "temporary" species will be created. In this case, three temporary species are created. It follows that the reaction will be broken down into:

$$T_1 \rightarrow B + C, \tag{7}$$

$$T_2 \rightarrow D + E, \tag{8}$$

$$T_3 \rightarrow F + T_1, \tag{9}$$

$$A \rightarrow T_2 + T_3, \tag{10}$$

where T_1, T_2 and T_3 are the first, second and third temporary species respectively. By setting the rate of reaction of Equation 10 equal to k and the probability of reaction of equations containing any temporary species on the reactant side equal to one, we reduce the artifacts due to the creation of the temporary species to a minimum. The temporary species disappear from the system as quickly as possible and the overall rate of reaction is identical. Shown below is the case where more than two reactants merge into a single product:

$$A + B + C + D + E \rightarrow F. \tag{11}$$

The procedure is similar to the previous case and for more than two reactants such that one temporary species will be created for each additional reactant.

$$A + B \rightarrow T_1, \tag{12}$$

$$C + D \rightarrow T_2, \tag{13}$$

$$E + T_1 \rightarrow T_3, \tag{14}$$

$$T_2 + T_3 \rightarrow F. \tag{15}$$

where T_1, T_2 and T_3 are the first, second and third temporary species respectively. In order to obtain the same overall probability of reaction and to reduce the impact of the temporary species on the system to a minimum, the probability of reaction of any reaction containing temporary species on the reactant side (Equations 13 and 14) is set to 1. Assuming that P is the probability of reaction of the reaction presented in 11 and P_1 and P_2 are the probability of the first and second simple reaction $A + B \rightarrow T_1$ and $C + D \rightarrow T_2$ then, we set $P = P_1P_2$. We also set $P_1 = P_2$ and equating the two equations gives $P_1 = P_2 = \sqrt{P}$. In general, the probability of the simple reactions P_n containing no temporary species is equal to

$$P_n = P^{\lfloor \frac{2}{N_{reactants}} \rfloor}, \tag{16}$$

where P is the probability of reaction and $N_{reactants}$ is the number of reactants of the initial reaction.

Each temporary particle has a parameter $lifetime$ which indicates the number of turns the particle has to live in the system before reverting back to its previous state. The short lifetime of "temporary" particles is important for two reasons.

First, it makes sure that "temporary" particles are effectively temporary and do not stay in the system for a long period of time. It also makes sure that all the reactants are to be close to each other in order for the reaction to complete. A lifetime of 2 or 3 turns is reasonable since it gives enough time to react with the neighboring particles while making sure temporary particles do not constitute the bulk of the system.

2.5. Reversible reactions

Reversible reactions are handled by creating two separate reactions, one for the forward reaction with the forward reaction rate and one for the backward reaction with the corresponding backward reaction rate. Assuming the following reaction

$$A+B+C+D+E \leftrightarrow F, \tag{17}$$

with forward reaction rate k_f and backward reaction rate k_b. This reversible reaction is then split into

$$A+B+C+D+E \rightarrow F, \tag{18}$$

with a reaction rate k_f and

$$F \rightarrow A+B+C+D+E, \tag{19}$$

with reaction rate k_b. Temporary particles involved in a reversible reaction must also remember if they are participating in a forward or backward reaction such that they revert back to the proper reactants when their lifetime reaches zero.

3. RESULTS

Several molecular models were simulated. The first two models consist of only a few different reactions and a limited number of species. The first model is a simple reversible reaction $A+B \leftrightarrow C$. The second one is a Michaelis-Menten system, which describes the kinetics of many enzymes. It is important to validate our approach by comparing the results with an already well known and proven approach, the SSA from Gillespie [4]. It is a very good candidate to validate our model as both approaches should provide the equivalent results for well mixed systems. The third example will demonstrate the effects of crowding by adding inert particles to a Michaelis-Menten system and the last example shows how the structure of a system can play in important role.

3.1. Simple reaction

This system is a simple reversible reaction involving three different species A, B and C in the following manner: $A+B \leftrightarrow C$. The forward reaction $A+B \rightarrow C$ has a rate of reaction k_f of 10^{10} per mole per second. The reverse reaction $C \rightarrow A+B$ has a rate of reaction k_b of 1 per second. The simulation space is a cube with a volume of 10^{-11} liters and the time step is 10^{-4} seconds. The initial number of particles is 3000 A particles, 1000 B particles and 0 C particles. The system reaches steady-state fairly quickly and after 1 second of simulated time, the species reach equilibrium with the exception of some small stochastic noise. The SSA simulator has been set with similar parameters: a simulation length of 1 second and a time step of 10^{-3} second. The results are shown in Figure 2. Both simulators produce the same results with small, but expected stochastic fluctuations. This model has been simulated by the ChemCell software with similar results [9]. These results support the idea that the discretization of the volume into a grid does not affect system behavior under these conditions.

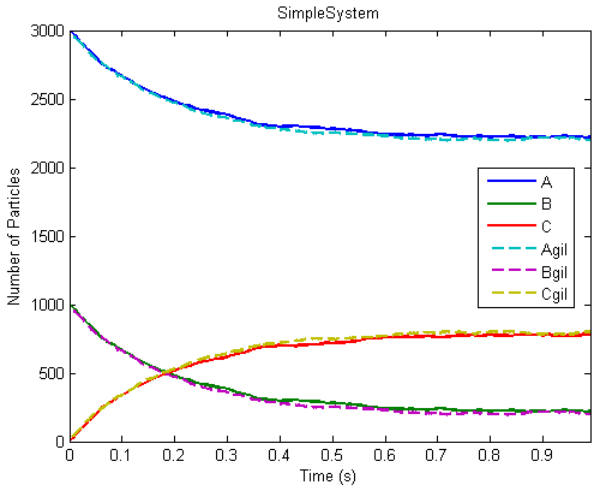

Figure 2. Comparison between our model and the Gillespie approach for the reaction $A+B \leftrightarrow C$.

3.2. Michaelis-Menten reaction

The Michaelis-Menten equations are used to describe most enzymatic reactions. Michaelis-Menten kinetics are described by the following equation:

$$E+S \leftrightarrow ES \rightarrow E+P. \tag{20}$$

The species E is an enzyme which can react with the substrate S to form the complex enzyme-substrate ES. The ES complex can revert back to its original dissociated form $E+S$ or create the product P, liberating at the same time the original enzyme E. As in the previous case, the simulation takes place in a cube of 10^{-11} liters, the number of enzymes E is 1000 particles and the initial amount of substrate S is 3000 particles. The forward rate of reaction k_1 of $E+S \leftrightarrow ES$ is 10^{10} per mole per second and the reverse rate of reaction is 1 per second. The forward rate of reaction of $ES \rightarrow E+P$ is also 1 per second. The simulation runs for 10 seconds, the time step in both cases has been set to 10^{-3}. The results produced our simulators is compared with the SSA algorithm and presented in Figure 3. Similar to the previous case, both approaches produce the same results.

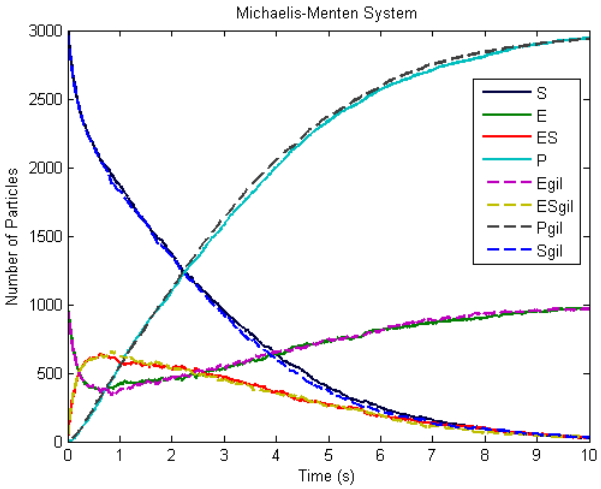

Figure 3. Comparison GridCell and the dimensionless SSA approach for a Michaelis-Menten system.

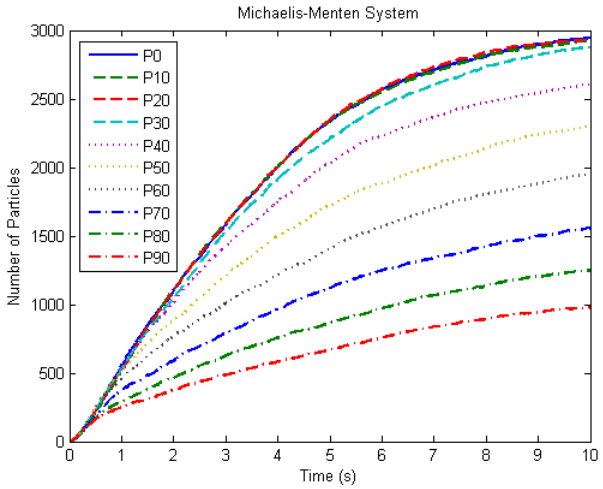

Figure 4. Effect on crowding on the creation of products in a Michaelis-Menten system. The inert particles occupy 0% to 90% of the voxels.

3.3. Crowding

Molecular crowding occurs when density of particles reduces their movement and hence affects their reactivity. Crowding is typically ignored in most models since the kinetics are often based on controlled, in vitro conditions that are not crowded. As well, simulators do not typically accommodate for this feature since it is computationally expensive to keep track of all particle positions, their excluded volume and the implementation of collision detection algorithms. Here, we demonstrate the effect of crowding by adding inert particles to the system. Inert particles do not react with other molecules but reduce movement in the grid and affect the overall number of reactions. The effect of crowding has been tested with the Michaelis-Menten system described in the previous subsection. Figure 4 shows the number of products over time for a wide range of varying concentrations of inert particles. The number after the P in the legend signifies the percent of the voxels occupied by inert particles. The inert particles have a minimal impact on the system when they occupy less than 30% of the available space. However, above 30% the reaction slows down linearly as more and more inert particles are added. Interestingly, no matter the concentration of inert particles, product formation is similar in the first 0.5 seconds. After 0.5s, the slope of curve begins to deviate from the non-crowded case. The likely explanation for this is that there is sufficient substrate S near every enzyme particle (E) to react at full speed. As time passes, nearby particles (S) are converted into product particles (P) and the enzymes and reactants must move to other positions to contact each other. At this point the reaction slows down dramatically, as noted by an abrupt change in the slope of graph.

3.4. Localization

Localization of particles, either by recruitment to a specific location or by anchoring them in relatively structured environments is expected to affect cellular processes. Here, we examine the effect of localization on reaction rates when a system is not well-mixed. Localization to cellular structures such as membranes may influence the overall behavior of the system by fixing position, reducing diffusion and hence affecting the rate of collision between interacting particles. The biochemical model is a Michaelis-Menten reaction where enzymes are localized to regions of a semi-porous membrane made of immobile inert particles. The substrate particles are also all placed on one side of the membrane. This example is similar to the one presented in [10]. The top view of the structure is shown in Figure 5. Substrate particles initially located on the left side slowly migrate to the right side as shown in Figure 6. The S concentration is still much higher on the left side than the right after 10s of simulation time. Concentration of the S is lowest at the two enzyme sites. This is because S are converted to product when interacting with enzymes embedded in the membrane. Figure 7 shows the evolution of the species P.

Figure 8 shows the difference in the overall reaction rate between a well-mixed system and a system with the structure described by Figure 5. Both simulations have the same number of particles, the same volume and the same reaction rate. However, the overall speed of reaction is substantially different between the two systems. Due to the presence of the semi-porous membrane and only two specific areas where the reaction can take place, the non well-mixed system exhibits a much slower reaction rate than the ideal well-mixed case. This demonstrates that the structure can have a significant

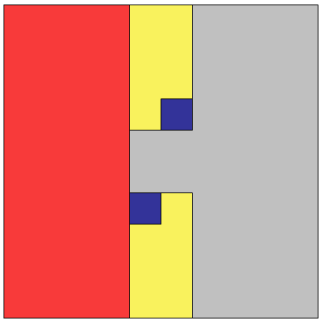

Figure 5. Top view of the simulation structure. The red area indicates the location of the substrate, blue areas indicate the location of the enzymes and the yellow areas indicate the inert particles forming the porous membrane.

impact on the behavior of a biological pathway and that the well-mixed assumption can induce a large amount of error.

4. FUTURE DIRECTIONS

4.1. Compartments

Compartments are important for many biological processes because they play a role in regulating the concentration of particles required for normal cellular function and may be involved in the recruitment of macromolecular complexes involved in cell signaling. Models that include compartments are aiming to understand the role of species partition, and this necessarily abandons the assumption of a well mixed system.

Compartments are not yet not fully supported. Support for these entities is currently being implemented and should be available for the next version of our simulator. Briefly, these will be treated as follows. Each voxel in the model will be assigned a compartment number. A particle moving from one compartment to another has to check a compartment table that contains a probability indicating how often that particle will go through that new compartment. Particles may not interact with particles located in different compartments.

4.2. FPGA Acceleration

While tracking every single particle in the system may provide additional biological insight it comes at a steep computational cost over conventional approaches. Simulations that could have been performed in a matter of seconds with differential equations may take several hours to complete. One way to minimize this additional computation is to parallelize the algorithm.

A major motivation in applying a discrete 3D grid with a simple neighbor finding system for ”collision” detecting is

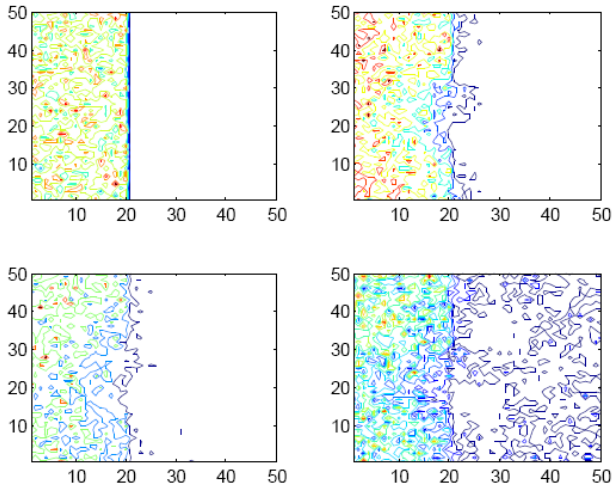

Figure 6. Top view contour plot of the concentration of substrate at $t = 0$, $t = 2$, $t = 6$ and $t = 10$.

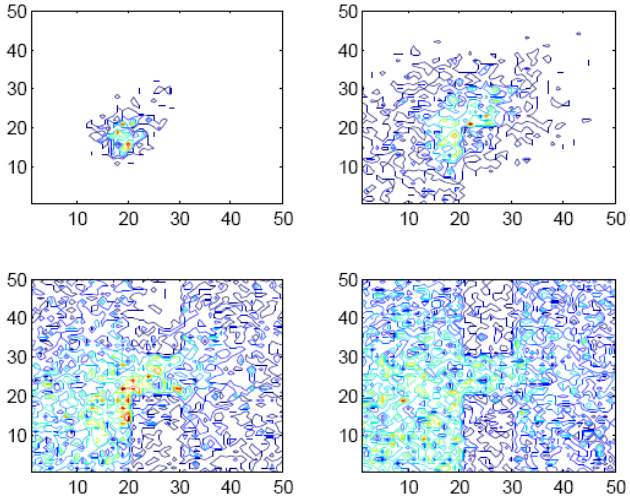

Figure 7. Top view contour plot of the concentration of product at $t = 0.1$, $t = 2$, $t = 6$ and $t = 10$.

that it avoids expensive searches and makes possible the implementation to a parallel architecture. By keeping the algorithm simple and regular, it is possible to design a simple custom pipelined architecture which could operate on many particles at the same time. In the current system all particles apart by more than two locations in every direction are completely independent from each other and can be processed at the same time, exposing a high degree of parallelism. Fine-grained parallel devices such as Field-Programmable Gate Arrays (FPGAs) are prime candidates for hardware acceleration.

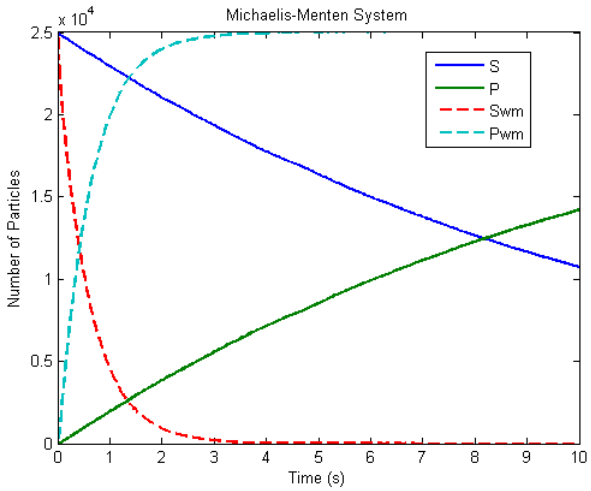

Figure 8. Concentration plot of the same Michaelis-Menten system under two different conditions. The first one considers localization effect, the second one assumes a well-mixed system.

5. CONCLUSIONS

We describe a stochastic simulator that handles locality, very low species concentration and collisions using a discrete 3D grid. Simulation results were found to be comparable to those obtained with the SSA algorithm. We demonstrate the negative impact that 1) crowding and 2) locality may have on reaction rates. Hence, biological simulations should no longer assume well mixed systems as these may lead to significant error. The adoption of a discrete 3D grid facilitates nearest-neighbour determination for interactions and reduce the requirement for computionally expensive distance calculations that depend on floating-point arithmetic. The regularity and simplicity of the algorithm makes it a good candidate for acceleration with a parallel architecture.

REFERENCES

[1] C. Lemerle, B. D. Ventura, and L. Serrano, "Space as the final frontier in stochastic simulations of biological systems," *FEBS Letters*, vol. 578, no. 8, pp. 1789–1794, 2005.

[2] M. Kerker, "Brownian movements and molecular reality prior to 1900," *Journal of Chemical Education*, no. 51, pp. 764–768, 1974.

[3] A. Minton, "The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences," *Mol Cell Biochem*, no. 55, pp. 119–140, 1983.

[4] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, no. 4, pp. 403–434, December 1976.

[5] J. Schaff *et al.*, "A general computational framework for modeling cellular structure and function," *Journal of Biophys.*, no. 73, pp. 1135–1146, November 1997.

[6] N. L. Novere and T. Shimizu, "Stochsim: modelling of stochastic biomolecular processes," *Bioinformatics*, no. 17, pp. 575–576, 2001.

[7] J. R. Stiles and T. M. Bartol, "Monte carlo methods for simulating realistic synaptic microphysiology using mcell," in *Computational Neuroscience: Realistic Modeling for Experimentalists*, E. D. Schutter, Ed. CRC Press, Boca Raton, 2001, pp. 87–127.

[8] M. Ander *et al.*, "Smartcell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks," *Syst. Biol.*, no. 1, pp. 129–138, 2004.

[9] S. Plimpton and A. Slepoy, "Chemcell: A particle-based model of protein chemistry and diffusion in microbial cells," *Sandia Technical Report SAND2003-4509*, December 2003.

[10] R. Grima and S. Schnell, "A mesoscopic simulation approach for modeling intracellular reactions," *Journal of Statistical Physics*, in press (DOI:10.1007/s10955-006-9202-z) 2007.

BIOGRAPHIES

Laurier Boulianne received the B.Eng. in electrical engineering in 2005 from McGill University. Currently, he is pursuing a PHD in electrical engineering at McGill University. His research interests include reconfigurable computing, computer simulations, bioinformatics and systems biology.

Michel Dumontier is a faculty member at Carleton University with a growing record of interdisciplinary research that leverages the disciplines of molecular and cell biology, biochemistry, bioinformatics, organic chemistry, computer science and engineering. Dr. Dumontier's long-term research objective is to merge these disciplines to develop innovative solutions that merge knowledge management with simulation towards realizing the promise of personalized medicine.

Warren J. Gross received the B.A.Sc. degree in electrical engineering from the University of Waterloo, Waterloo, Ontario, Canada, in 1996, and the M.A.Sc. and Ph.D. degrees from the University of Toronto, Toronto, Ontario, Canada, in 1999 and 2003, respectively. Currently, he is an Assistant Professor with the Department of Electrical and Computer

Engineering, McGill University, Montreal, Quebec, Canada. During the summers of 2004 and 2005 he was a Visiting Professor at the Universit de Bretagne-Sud, Lorient, France. His research interests are in the design and applications of signal processing microsystems and custom computer architectures.

Dr. Gross served on the Program Committees of the 2006 IEEE Workshop on Signal Processing Systems and the 2006 IEEE Symposium on Field-Programmable Custom Computing Machines. He is a member of the Design and Implementation of Signal Processing Systems Technical Committee of the IEEE Signal Processing Society.