34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

A game-of-life like simulator for design-oriented modeling of BioBricks in synthetic biology

Morgan Madec, Yves Gendrault, Christophe Lallement, *Member, IEEE,* Jacques Haiech

Abstract— **This paper deals with the development of a new simulator that will be very helpful to establish new accurate and predictive design-oriented models for the BioBricks used in synthetic biology. The simulator uses the principle of the game-of-life: molecules can move on a grid and, at every iteration, binding and dissociation rules are applied when two molecules are on same node. The principle is elementary but it can highlight interesting biological phenomenon. Those can be modeled by mathematical equations to achieve design-oriented models. In this case, the simulator also helps to make to link between mathematical parameters and the microscopic parameters. A first version of the software has been implemented in MATLAB. It permits to retrieve very interesting results, such as the Hill's equation and the properties of Hill's coefficient.**

I. Introduction

Synthetic biology is a new extension of biotechnologies that aims at designing new complex biological systems based on the assembly of biological elements decoupled from their environment [1]. Its main application fields are health care [2] and environment [3].

By definition, the design approach used in synthetic biology is similar to the one used in the engineering of systems (*eg*. microelectronics). At mid-term, the design of artificial biological functions is expected to be a kind of LEGO™ game, which consists in assembling a set of elementary biological standardized parts, called BioBricks [4]. In microelectronics, such approach makes possible the design of powerful circuits with more than 2 billion of transistors, like for example the last generation of microprocessors (*eg*. Intel Xeon Nehalem-EX).

Obviously, the current designed "biological circuits" are not as complex as a microprocessor in term of involved function number. Nevertheless, this complexity is expected to increase as quickly as it was the case for microelectronics [5]. As a consequence, the need of a powerful Genetic Design Automation tool should become more and more concrete in a near future [6].

M. Madec, Y. Gendrault and C. Lallement are with the Institut d'Electronique du Solide et des Systèmes *(InESS Institute of Electronics for Solids and Systems)*, a research unit from the Université de Strasbourg (Unistra) and the Centre National de Recherches Scientifiques (CNRS), UMR7163, Strasbourg, France.

J. Haiech is with Laboratoire d'Innovation Thérapeutique *(LIT, Therapeutic Innovation Laboratory)*, a research unit from the Université de Strasbourg (Unistra) and the Centre National de Recherches Scientifiques (CNRS), UMR7200, Strasbourg, France.

Corresponding author is Morgan MADEC, (+33)388106227, e-mail: madec@iness.c-strasbourg.fr.

One of the main revolutions in systems design, which occurs at the end of last century, is the development of virtual prototyping to speed up design process and reduce design costs. Nowadays, complex and heterogeneous systems, such as cars or aircrafts [7], can be virtually designed, by assembling models for each involved components instead of assembling actual components in an actual prototype. Obviously, the efficiency and the reliability of a virtual prototype are directly linked with the accuracy and the predictivity of the involved models. As a consequence, a big effort has been made over the past decade in order to develop accurate and fast-simulating models as well as powerful languages and tools to support the virtual prototyping process. Unfortunately, current models used in biology are not accurate enough to make virtual prototyping possible.

One of the goals of our research team is to develop an equivalent computer-aided design environment for synthetic biology [8]. The development of design-oriented models for elementary biological part should be a first breakthrough in this way. In biotechnology, the measurement of species concentration inside a cell is not as obvious as the measurement of physical quantities (voltage, torque, temperature…). To get round this difficulty, we develop a low abstraction level simulator, based on the principle of a game-of-life. This simulator should permit to accurately reproduce elementary biological mechanisms and BioBricks behavior in order to give access to missing data required for the development of efficient design-oriented models (fast-simulating model with a highest abstraction used and assembled by the biosystem designers).

This point is clarified in the first part of the paper, taking as an example the development of the model of an electronic diode. Then, our simulator is presented in the second part of the paper. Finally, some examples are given in order to show the interest of our approach.

II. Design-Oriented Modeling in Microelectronics

The development of design-oriented models can be carried out thru three ways: i) the theoretical study of the system to model, ii) the exploitation of experimental results obtained on actual systems or iii) the use of complex adapted simulation tools (*eg*. COMSOL for the simulation of physical systems, SILVACO for microelectronic devices…). Nevertheless, a combination of these three approaches is often required to establish the most suitable model.

To illustrate this, let us take the example of an electronic diode. The theoretical study of this device leads to a well-known *I-V* model of the diode [9]:

$$I(V) = I_S \cdot \left(\exp \frac{V}{V_T} - 1 \right) \quad (1)$$

with I_S the saturation current and V_T the thermodynamic potential. Nevertheless, this theoretical model does not fit with actual I-V characteristic. How can I be improved?

First, we observe that the slope of the log(I)-V characteristic is not equal 1, which is the theoretical expected value. This is due to generation-recombination inside the depletion region, which varies with process and cannot be described easily with physical analytic equations. As a consequence, to take this effect into account, an ideality factor, n, has been introduced.

This new model fits the actual characteristic in the elbow but a shift between characteristics still remains for high current bias. To understand this phenomenon, the diode is modeled in a dedicated semiconductor simulator (SILVACO in this case), which solves the space-dependent differential equations of electrons and holes diffusion in the semiconductor. Simulation results shows some voltage gradients in areas of the device in which the electrical potential is assumed to be constant (this assumption is necessary to establish the equation (1)). This observation suggests the addition of a series resistance R_S leading to the following implicit equation:

$$I_S \cdot \left(1 - \exp \frac{V - R_S \cdot I}{V_T} \right) + I = 0 \quad (2)$$

This model is now acceptable and can be implemented with dedicated hardware description languages such as VHDL-AMS or Verilog-A [10]. As a consequence, it can be used to simulate a diode in a complete circuit.

Up to now, there is nothing very revolutionary. But what is it about applying the same methodology for BioBricks? In fact, it is not obvious because of two main difficulties. First, the theory of semiconductor device seems to be more domesticated than the theory of biological mechanisms involved in BioBricks. Second, it is easier to access to voltage or current measurements in a silicon device than to measure species concentrations or flows in a living cell. As a consequence, the use of a dedicated simulator stands out as our best ally to tackle this challenging task.

III. Description of the Simulator

We can found in the literature or on the Internet some existing simulation tools for synthetic biology. Nevertheless, most of them are already based on high-level models and are not suitable for our purpose. As far as we know there is no simulator in biology that can play the same role as SILVACO in microelectronics. As a consequence, our idea is to develop our own software with a dedicated simulation engine. The concept is very easy and looks like a kind of game-of-life. First, the cell in divided into a 2-D or 3-D mesh, on which the different involved species may randomly evolve: they travel randomly from node to node and, when they meet each other on a given node, they have a probability to interact.

With this first version of the simulator, our goal is to show the feasibility and interest of our approach. To simplify implementation and spare computation time, we use an NxN 2-D rectangular mesh. The software is implemented in MATLAB, a dedicated tool for matrix computation, which is very worthwhile for this application.

A. Data structure

The first task was to choose the data structure. There are two natural approaches. The first one consists in defining, for each involved species, a vector giving the position of each molecule in the mesh. Modeling of species displacement with this approach is obvious but the modeling of interaction becomes challenging. It requires a cross search in the species vectors to find the nodes shared by two molecules and where an interaction may occur. In the other hand, another approach consists in using a matrix for each species. Each element of the matrix corresponds to one node of the mesh and is equal to the number of species in the node. The modeling of displacement is more complex than with vector but the modeling of interactions is optimized.

A comparative evaluation of both methods leads to the conclusion that the first one is more interesting when the mesh is fine (N increase) whereas the second one is more interesting when the number of involved species M increase. The memory space required to store data is in $o(N)$ and in $O(M)$ for the first approach, whereas it is in $O(N^2)$ and in $O(M)$ for the second one. Conversely, the computation time varies in $o(N)$ and in $O(M^2)$ for the first approach and in $O(N^2)$ and in $O(M)$ for the second method. In this version of the software, the second approach is used. For each involves species P_k, an NxN matrix, $\mathbf{A_k}$, is defined

B. Modeling of species displacement

The movement of a given molecule inside the cell depends on its shape and size. We define for each kind of species a mobility parameter μ_k. μ_k is equal to 0 for a fast molecule that can change node at each time step. Conversely, for a fixed molecule μ_k tends to 1.

The displacement is computed through two NxN random matrices $\mathbf{R_X}$ and $\mathbf{R_Y}$ that are generated according to a uniform distribution between -1 and 1. First, a point-by-point multiplication is carried out between $\mathbf{A_k}$ and $\mathbf{R_X}$. The elements of the resulting matrix are compared with the value of the mobility μ_k. Then, an NxN x-displacement matrix $\mathbf{DX_k}$ is computed according to the following rules:

If $\mathbf{A_k}(i,j) \cdot \mathbf{R_X}(i,j) < -\mu_k$ and i > 1,
$\mathbf{DX_k}(i,j) = -1$ and $\mathbf{DX_k}(i-1,j) = +1$
Else if $\mathbf{A_k}(i,j) \cdot \mathbf{R_X}(i,j) > \mu_k$ and i < N,
$\mathbf{DX_k}(i,j) = -1$ and $\mathbf{DX_k}(i+1,j) = +1$
Else $|\mathbf{A_k}(i,j) \cdot \mathbf{R_X}(i,j)| < \mu_k$,
$\mathbf{DX_k}(i,j) = 0$

The new value of $\mathbf{A_k}$ is $\mathbf{A_k}+\mathbf{DX_k}$. The same operation is carried out for the y-displacement thanks to a $\mathbf{DY_k}$.

C. Modeling of species interactions

Modeling of interactions is more complex but the use of a matrix instead of vector to represent the positions of the molecules facilitates the task. Each potential interaction is treated case by case, but all of them depend on two *N*x*N* random matrices, one for the binding **B** and one for the dissociation **D**.

Let us take the example of the following binding equilibrium characterized by k_{ON} and k_{OFF} parameters:

$$P_T + P_U \underset{k_{OFF}}{\overset{k_{ON}}{\longleftrightarrow}} P_V \tag{3}$$

In our simulator, those coefficient cannot be directly implemented and are replaced by two probabilities $PB_{T,U,V}$ and $PD_{T,U,V}$ which represent respectively the probability that species P_T and P_U bind themselves to form P_V when they share the same node, and the probability that the species P_V dissociates to give back P_T and P_U.

For the dissociation, we first search for non-zero values in the matrix $\mathbf{A_V}$. This gives the position where the interactions may take place. If, on this node, the element in the **D** matrix is lower than $PD_{T,U,V}$, the dissociation occurs (the corresponding value is decremented in $\mathbf{A_V}$ and incremented in $\mathbf{A_T}$ and $\mathbf{A_U}$).

The process is the same for the binding reaction except that the position of the potential interaction are computed with the matrix $\mathbf{A_T}$*$\mathbf{A_U}$ (* is the point-by-point multiplication). The values in **B** is compared with $PB_{T,U,V}$ and, if the binding occurs, the value corresponding value is incremented in $\mathbf{A_V}$ and decremented in $\mathbf{A_T}$ and $\mathbf{A_U}$. In practice, binding is computed before dissociation.

IV. Examples

In the following, we try to show the potential of our approach on some standard examples.

A. Binding reaction with two ligands

We first consider the following biological scheme:

$$A + B \underset{k_{OFF}}{\overset{k_{ON}}{\longleftrightarrow}} AB + B \underset{k'_{OFF}}{\overset{k'_{ON}}{\longleftrightarrow}} AB_2 \tag{4}$$

The species A and B can bind each other to give the complex AB, which can itself bind again with B to give AB_2. The mobility μ_A μ_{AB} and μ_{AB2} are set to 0.9 whereas the mobility μ_B is fixed to 0.1. The binding and dissociation probability are fixed resp. to 0.7 and 0.003 for both reactions. The cell is divided into a 50x50 mesh. Simulation results are given on Figs. 1 and 2.

At the initial step (Fig. 1a), 10 A molecules (red dots) are randomly dispatched in the cell, and 50 B molecules (blue dots) are injected from the top-left corner. After a few iterations (Fig. 1b), a B molecule meets an A molecule, which leads to the apparition of the AB complex (pink). The molecules continue to spread in the cell and reach the other A. In the same time, B binds with AB to form AB_2 complex (magenta in Fig. 2c). After about 1,000 iterations (Fig. 1c), equilibrium occurs: B molecules are randomly dispatched in the space and the four species A, B, AB and AB_2 coexist.

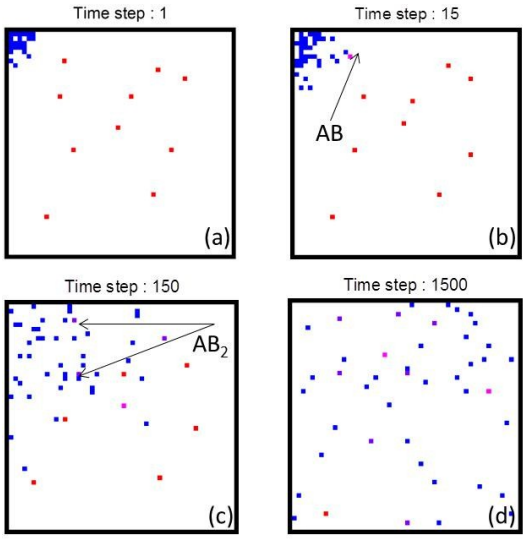

Figure 1. Simulation of a double binding reaction. Blue, red pink and magenta dots corresponds to the position of A, B, AB and AB_2 molecules.

Due to the small number of molecules, the time-evolution of each species concentration is very noisy. To get exploitable curves, an averaging is required (Fig. 2).

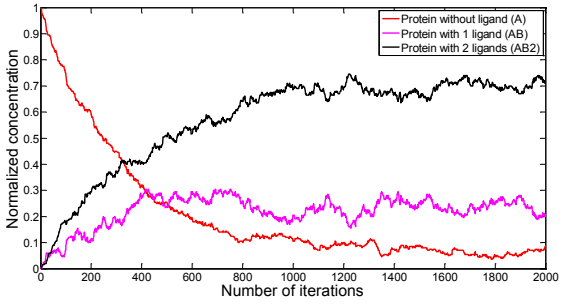

Figure 2. Transient simulation of the concentration of A, AB and AB_2. Results are obtained after an averaging over 20 simulations.

B. About Hill's equation

For this part, we go one step further by considering a tetramer A which is able to bind up to 4 ligands L. In most cases, in design-oriented models, such systems are modeled thanks to the Hill's equation [11]:

$$\theta_A = \frac{1}{1 + \left(K_A / [B] \right)^n} \tag{5}$$

In such an equation, θ_A is the ratio of occupied sites on the tetramers A, K_A is the microscopic dissociation constant which corresponds to the concentration of B for which half of the sites are occupied and n is the Hill's coefficient which depends on the number of available sites on A as well as cooperativity between sites. Our simulator is now used to retrieve Hill's equation parameters (esp. n) and properties as a function of microscopic considerations.

First, we simulate this system with our game-of-life like simulator. The mobility of A and L are respectively set to 0.95 and 0.05. At the initial step, 20 A and a given quantity of B are dispatched randomly inside the cell. We observe the behavior of the signal θ_A at the equilibrium as a function of the ligand concentration [L]. Simulation results are given in Fig. 3 for five different cases described in Tab. I: the non-cooperative binding, a slightly positive cooperative binding, a slightly negative cooperative binding and two case where the tetramer can only bind two ligands, one without cooperativity and one with a strong one. Let $PB_{0,1}$ and $PU_{0,1}$ be the binding and unbinding probability of the first ligand. Without cooperativity, the first ligand has 4 available sites on A whereas the *k*-th has only 4-*k* free sites to bind. As a consequence, $PB_{k\text{-}1,k} = PB_{0,1}/k$. By the same way, the probability that AL_k unbinds is equal to *k* time the probability that AL unbinds ($PU_{k\text{-}1,k} = k.PB_{0,1}$). With cooperativity, the $PB_{k\text{-}1,k}$ and $PU_{k\text{-}1,k}$ increase or decrease faster as in the uncooperative case. The simulated θ_A-[L] curves are fitted with Hill's equations and the extracted Hill's coefficient is also given in Tab. I.

TABLE I. MODEL PARAMETERS AND EXTRACTED HILL'S COEFFICIENT

Case	Parameters				Extracted value for n
	$PB_{0,1}$ $PU_{1,0}$	$PB_{1,2}$ $PU_{2,1}$	$PB_{2,3}$ $PU_{3,2}$	$PB_{3,4}$ $PU_{4,3}$	
Case 1: 4 ligands w/o cooperativy	0.3 0.003	0.15 0.006	0.1 0.009	0.075 0.012	1.046
Case 2: 4 ligands Pos. cooperativity	0.3 0.003	0.3 0.003	0.3 0.003	0.3 0.003	1.89
Case 3: 4 ligands Neg. cooperativity	0.3 0.003	0.1 0.009	0.044 0.020	0.022 0.041	0.92
Case 4: 2 ligands w/o cooperativy	0.3 0.003	0.15 0.006	0 0	0 0	1.011
Case 5: 2 ligands Strong coop.	0.1 0.009	0.9 0.001	0 0	0 0	1.785

$PB_{k,k+1}$ is the probability of binding and additional B when the tetramer is in configuration AB_k
$PU_{k+1,k}$ is the probability of unbinding one B when the tetramer is in configuration AB_{k+1}

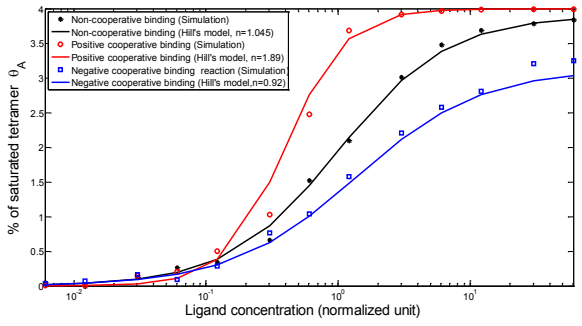

Figure 3. Percentage of saturated tetramer at the equilibrium as a function of the ligand initial concentration. Markers corresponds to the simulation results and lines represent the corresponding Hill's model. K_A in the case of a non-cooperative binding has been chosen as the reference concentration in the x-axis.

This study permits to confirm properties about Hill's coefficient that have been already observed: i) *n* is about 1 for a non-cooperative case (case 1 and 4); ii) *n* increase with positive cooperativity (case 2) and decrease with negative one (case 3); iii) for a strong cooperativity, *n* is about equal of the number of operating sites.

V. CONCLUSION

The work presented in this paper is a first version of a new simulator which should help to the development of design-oriented models for synthetic biology. It should facilitate the link between mathematical parameters of macro-models (*eg.* Hill's equation) and biochemical parameters (molecule inertia, binding probability of two species, degradation…). The software proves itself on some examples. Nevertheless, some improvements should be carried out for the second version of the software : comparison between 2-D and 3-D mesh, the use of more complex mesh, the reduction of the parameter set, the introduction of more sophisticated cell shapes and more realistic displacements process… They are currently under investigation. Up to now, the major issue encountered in the development and the use of this simulation engine is the computation time (about 1 minute to simulate the system presented in Sec. IV.B with 2000 iterations). This point should also be improved for the next version of the simulator.

ACKNOWLEDGMENT

The authors would like to thank Paul Bonnet, Romain Cherpantier and Hélène Moulet, students from Telecom Physique Strasbourg, an engineering school of the University of Strasbourg, for their contribution in the development of the presented software.

REFERENCES

[1] D. Endy, "Foundations for engineering biology" in *Nature*, vol. 438 (24), pp. 449-453, 2005.

[2] V. Martin, D. Pitera, S. Withers, J. Newman, J. Keasling, "Engineering a mevalonate pathway in Escherichia coli for production of terpenoids" in *Nature Biotechnol.,* vol. 21, pp. 796–802, 2003.

[3] E. Rubin.,"Genomics of cellulosic biofuels" in *Nature*, vol. 454, pp. 841–845, 2008.

[4] B. Canton, A. Labno, and D. Endy, "Refinement and standardiza-tion of synthetic biological parts and devices," *Nat. Biotechnol.*, vol. 26, p. 787793, 2008.

[5] R. Carlson, "*Biology Is Technology: The Promise, Peril, and New Business of Engineering Life*", Harvard University Press ed., 2011.

[6] M. W. Lux, B. W. Bramlett, D. A. Ball, and J. Peccoud, "Genetic design automation: engineering fantasy or scientific renewal?", in *Trends in Biotechnology*, vol. 30, pp. 120–126, 2011.

[7] Y. Bennet, "The application of virtual prototyping in the development of complex aerospace products" in *Aircraft Engineering and Aerospace Technology,* Vol. 69(1), pp.19 - 25. 1997

[8] Y. Gendrault, M. Madec, C. Lallement, F. Pecheux, J. Haiech, "Computer-aided design in synthetic biology: a system designer approach", *4th IEEE International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2011),* Barcelone (Spain), October 24-28, 2011.

[9] S.M. Sze, "*Physics of semiconductor devices"*, John Wiley and Sons ed. 2007.

[10] F. Pecheux, C. Lallement, A. Vachoux, "VHDL-AMS and Verilog-AMS as alternative hardware description languages for efficient modeling of multidiscipline systems" in *IEEE Trans. on Computer-Aided. Design.*, vol. 24, pp. 204–225, 2005.

[11] J. Weiss, "The Hill equation revisited: uses and misuses" in *FASEB Journal*, vol. 11(11), pp. 835-841, 1997.