Stochastic Simulation of Biological Cellular Processes Using VHDL-AMS

Gregory D. Peterson and Joseph M. Lancaster
Electrical and Computer Engineering
The University of Tennessee
gdp@utk.edu

Abstract

Accurate, predictive models of biological cellular processes are a key component of the quest to transform the engineering of genetic controls within organisms. Biological systems include a dizzying variety of interacting biochemical pathways, each consisting of numerous reactions and chemical species. Moreover, such classical assumptions of differential equations models as equilibrium (including well-stirred species) and a large numbers of reactants do not hold for mesoscalar, intra-cellular modeling. In fact, these systems require the seamless, accurate modeling of interacting discrete and continuous behaviors. Previous research has demonstrated the potential of using portions of biochemical pathways as switches to control behavior, with emergent behaviors such as logic gates created. Accurate, efficient models require the ability to mix discrete and continuous descriptions while providing for the high-fidelity interaction between these domains. Although a number of previous researchers have discussed the use of analog and mixed signal hardware description languages (AMS HDLs) such as VHDL-AMS and Verilog-AMS for representing the behavior of MEMS, microfluidic, optical, and thermal systems, similar benefits can result from the use of AMS HDLs for biological systems modeling. We present preliminary research into the VHDL-AMS representation of biological systems at different levels of abstraction and the capability to support multi-resolution modeling. We discuss the role of this research within the context of the DARPA BioSPICE research program, our modeling and simulation approach, and future plans.

Introduction

The understanding of biological systems remains one of our primary scientific activities. In one of the most important accomplishments in the history of science, the sequence of the human genome was recently completed [Pen00, Mar00]. This DNA sequence information is quite important because DNA (deoxyribonucleic acid) contains a biochemical code for the specific characteristics and functions of living organisms. These characteristics, or traits, are referred to as genes. The complete set of genes for an organism is its genome.

The role of genes in biological processes is very important, with genes controlling cell differentiation in developmental processes though proteins. Genes produce messenger RNA, or mRNA, to direct the operation of ribosomes, which translate mRNA information into proteins, which are strings of amino acids. The genes, mRNA, ribosomes, and proteins interact in biochemical *pathways*, or series of reactions that perform some cellular function. Each chemical step in the pathway is catalyzed by enzymes, and enzyme amounts are controlled by the expression of specific genes. Specific genes may promote or inhibit the production of specific proteins. Pathways can be very complex, with dozens or even hundreds of reactions that can occur simultaneously. The genes and proteins act as elements in a complex feedback control system for the organism [FS00].

The completion of the human genome mapping promises to change the way we approach biological research by allowing scientists to consider whole-organism approaches to the mechanisms controlling cell biochemistry. Control of biochemical cellular processes will provide exciting new capabilities for treating and even preventing

disease. The complexity of the biological systems and biochemical pathways of interest renders it impractical to analytically model them. Similarly, the cost and difficulty in performing laboratory with organisms significantly hampers scientific advance. The role of computer modeling and simulation cannot be overstated in facilitating data interpretation and understanding of complex biochemical pathways and mechanisms.

Traditionally, biochemical pathway research has employed experimentation, but this approach will require scientists to test every possible case of gene expression under all environmental conditions, which is practically impossible. Hence, research is addressing experimental and computational techniques to study the function of every gene in an organism, measuring mRNA concentrations, determining the function and structure of protein function, and modeling the interconnected regulatory network of the whole cell [FS00]. As Fitch and Sokhansanj point out, "Pathway simulation would obviate the need for an impractical number of biochemical experiments, and, ultimately, such simulation can be used to design control systems in organisms to produce useful proteins in a regulated fashion, in other words, an artificial genetic control system."

Biological Application

A primary goal of biological research is the development of accurate models that can be used to explain biological processes, with predictive models particularly promising for drug development, epidemiology, bio-engineering, and genetic applications. In this paper we focus on the use of VHDL-AMS for developing highly predictive, accurate models of cellular processes.

The potential impact of appropriate modeling tools or environments on biological research may be as far-reaching as that of circuit simulation on electrical engineering. Genetic bioengineering will benefit from the ability to employ simulations of complex biochemical networks to reduce the amount of experimentation required to understand biochemical pathways or to influence their function. The ability to understand, and eventually to control, these biochemical pathways will profoundly impact our ability to develop new bioengineered systems, crops, and medical treatments [FS00].

Rigorous, fundamental models are necessary for many potential applications of prediction and control of cell behavior. Furthermore, classical differential equation approaches that treat the molecular concentrations as continuous discrete functions are rendered invalid by the small population levels of specific regulatory molecules within individual cells; therefore, a stochastic approach is warranted [Gil77]. However, application of exact stochastic simulation of the fundamental equations describing a genetic circuit to communities containing thousands of cells presents a massive computational problem. Ongoing research activities at The University of Tennessee and Oak Ridge National Laboratory seek to address this problem through an integrated modeling approach that employs: (a) fundamental and rigorous lower-tier models of simple genetic circuits that employ stochastic simulations of the population dynamics of individual molecules within the cell as affected by reaction and diffusion mechanisms; and (b) upper-tier models consisting of discrete stochastic abstractions of the lower-tier models that are less computationally intensive and able to simulate diffusion of signaling molecules between randomly distributed cells and the resulting group behavior. These modeling approaches have very different requirements for modeling languages or tools, but can be modeled together within the VHDL-AMS environment.

Related Modeling Efforts

In most models of biological processes, a differential equations-based chemical master equation is used to describe the time-varying populations of the various chemical species. Such models are used to describe macro level intercellular biological

behaviors [Gil77, Gil92]. The chemical master equation approach does not maintain its accuracy for intracellular modeling, because the population of chemical species is too small for the differential equation approach. To accurately model the lower populations, an exact stochastic reaction diffusion simulation can be used [Stu96, Gib00]. Performance issues preclude the application of such an exact simulation approach for large numbers of chemical species.

The biochemical genetic function of organisms can be described at a macroscopic level by the use of deterministic, continuous differential equations representing chemical species concentrations and the chemical reaction rate laws that concern the concentrations of the various molecules within the cell. This approach is not always useful in practice for several reasons. First of all, the chemical rates given in the master equation may include a number of intermediate steps that are not explicitly modeled in this manner. Secondly, in practice the master equation includes a large number of biochemical reaction rates that are difficult to find. Thirdly, these rates are associated with an equilibrium state with respect to thermodynamics and chemical concentrations. Finally, the reactions described by the continuous differential equations of the chemical master equation occur in individual cells, where low populations of each of the chemical species are located. This renders many of the assumptions inherent in the macroscopic chemical reaction rate laws invalid, and thus makes the deterministic solutions invalid as well.

For intracellular modeling, we use behavioral modeling techniques with VHDL-AMS for stochastic simulation to describe genetic regulatory dynamics. Stochastic simulation explicitly considers the effect of small population numbers of regulatory molecules within the cell. This facilitates generalization of computer models because fewer types of basic reactions must be considered and the rate constants for elementary reaction steps (such as the elongation of mNRA by a single nucleotide during transcription) may be more accessible than the macroscopic rate constants that embody dozens of elementary steps.

Application of VHDL-AMS

Multiresolution simulation can be applied to improve the performance of simulation for intercellular interactions of biological processes modeled at the intracellular level. The VHDL-AMS language provides modeling support for the master equation modeling with differential equations, the stochastic simulation approach, as well as multiresolution modeling using both approaches together [APT02].

To support this modeling application, we developed VHDL-AMS packages, natures, types, and components for general modeling of biochemical pathways. Some preliminary validation work of these modeling activities addresses the behavior of quorum sensing for bioluminescence control with Vibrio fischeri [Jam00] and iron transport with Pseudomonas auruginosa [Vas99]. Both the quorum sensing and iron transport processes are known to impact the virulence of bacterial infections, so accurate modeling capabilities can potentially help in limiting the effectiveness of bacterial infections.

Quorum sensing is a process commonly used by bacterial populations to coordinate their efforts for specific functions such as infection onset, antibiotic production, or biofilm formation. For example, a biofilm may be created by bacterial populations to inhibit immune defenses once a critical mass of bacteria is achieved. Biofilm production before a sufficient population density exists expends energy without achieving associated benefits, so quorum sensing is employed to establish intercellular communications concerning bacterial density. The physical mechanism employed is the release of diffusible signal molecules called autoinducers, with a threshold level used to allow the individual bacterium to coordinate their activation or repression of

specific gene expression. This process involves biochemical behavior at both the intracellular microscopic level (e.g., the production of autoinducers and gene expression) as well as intercellular macroscopic behavior (e.g., the diffusion and density of autoinducers among bacterial populations). Figure 1 illustrates the quorum sensing process for V*ibrio fischeri*. The VHDL-AMS language supports this multiresolution modeling task well.

Conclusions

The VHDL-AMS language was developed to support the modeling of mixed-signal and mixed-technology systems. The expressive power of VHDL-AMS makes it very well suited to modeling biochemical pathways as well. The development of standard VHDL-AMS natures, quantities, and biochemical components for modeling biological systems was presented and shown to be an interesting new application for behavioral modeling techniques. Future work will address the most efficient means of modeling biological systems, how to make the models most helpful to support biological research, and to consider interactions between biological systems and those of other domains.

Acknowledgements

The authors would like to acknowledge the help from our colleagues at The University of Tennessee, particularly Michael L. Simpson, Chris D. Cox, Michael J. Roberts, and Paul Frymier in understanding the biological systems of interest and the current state of the art in biological system modeling.

References

[ARM98] A. Arkin, J. Ross, and H. H. McAdams, "Stochastic kinetic analysis of developmental pathway bifurcation in phage-infected Escherichia coli cells," *Genetic*s, vol. 149, pp. 1633–1648, Aug. 1998.

[APT02] P. J. Ashenden, G. D. Peterson, and D. Teegarden, *The System Designer's Guide to VHDL-AMS: Analog, Mixed-Signal, and Mixed-Technology Modeling*. Morgan-Kaufmann Publishers, San Francisco, CA, 2002.

[EL00] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Natur*e, vol. 403, pp. 335–338, Jan. 2000.

[EKY97] D. Endy, D. Kong, and J. Yin, "Intracellular kinetics of a growing virus: A genetically structured simulation for bacteriophage T7," *Biotechnology Bioengineering*, vol. 55, pp. 375–389, July 1997.

[FS00] J. Patrick Fitch and Bahrad Sokhansanj. "Genomic Engineering: Moving Beyond DNA Sequence to Function." *Proceedings of the IEEE*, 88(12): 1949-1971, December 2000.

[Gib00] Michael A. Gibson and Jehoshua Bruck. "Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels." *Journal of Physical Chemistry*, 104:1876-1889, 2000.

[Gil76] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Phys*. vol. 22, pp. 403–434, Dec. 1976.

[Gil77] Daniel T. Gillespie. "Exact Stochastic Simulation of Coupled Chemical Reactions." *The Journal of Physical Chem*. 81(25):2340-2361, 1977.

[Gil92] Daniel T. Gillespie. "A Rigorous Derivation of the Chemical Master Equation." *Physica A*. 188:404-425, 1992.

[Ham93] B. J. Hammond, "Quantitative study of the control of HIV-1 gene expression," *J. Theoretical Biol*., vol. 163, pp. 199–221, July 1993.

[Jam00] Sally James, Patric Nilsson, Geoffrey James, Staffan Kjelleberg, and Torbjorn Fagerstrom. "Luminescence Control in the Marine Bacterium Vibrio fischeri: An Analysis of the Dynamics of lux Regulation." *Journal of Molecular Biology*, 296:1127-1137, 2000.

[Luk98] J. J. Lukkien, J. P. L. Segers, P. A. J. Hilbers, R. J. Gellen, and A.P. J. Jansen, "Efficient Monte Carlo methods for the simulation of catalytic surface reactions," *Phys. Rev.* E, vol. 58, pp. 2598–2610, Aug. 1998.

[MAA97] H. H. McAdams and A. Arkin, "Stochastic mechanisms in gene expression," in *Proc. Natl. Acad. Sci. US*A, vol. 94, Feb. 1997, pp. 814–819.

[MAA98] H. H. McAdams and A. Arkin, "Simulation of prokaryotic genetic circuits," *Annu. Rev. Biophys. Biomolecular Structur*e, vol. 27, pp. 199–224, 1998.

[MAA99] H. H. McAdams and A. Arkin, "It's a noisy business! Genetic regulation at the nanomolar scale," *Trends in Genetic*s, vol. 15, pp. 65–69, Feb. 1999.
[Stu96] Audrius B. Stundzia and Charles J. Lumsden. "Stochastic Simulation of Coupled Reaction-Diffusion Processes." *Journal of Computational Physics*, 127:196-207, 1996.
[Vas99] M.L. Vasil and U.A. Ochsner. "The Response of Pseudomonas aeruginosa to Iron: Genetics, Biochemistry, and Virulence." *Molecular Microbiology*, 34:399-413, 1999.

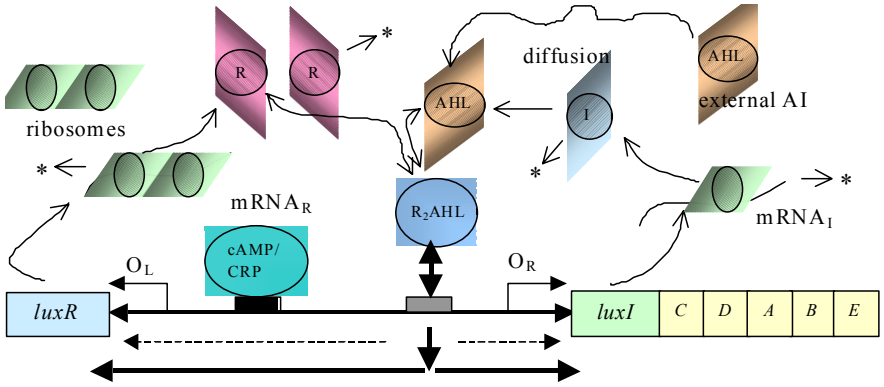

Figure 1. Simplified model of Lux I/R quorum sensing system of V*ibrio fischeri.*