Performance Limits on Biochemical Computation

George E. Homsy

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: Recent work addresses the possibility of designing biochemical systems that compute, using concentrations of distinct chemical species as logic signals. Some researchers claim the performance of such biochemical computers will eventually rival or exceed electronic computers for some tasks. We critically evaluate this claim from a theoretical perspective, by deriving bounds on the performance of biochemical computers and comparing these bounds to the performance of current electronic computers.

Motivation: Current models of chemical computation differ from the electronic model: signals are not carried on wires, but are co-distributed in solution and diffuse freely. In fact signal molecule diffusion is an integral part of chemical computer action.

In electronic computation, the sequesterment of signals to wires allows distinction of one signal from another. In chemical computation, since signals are spatially mixed, they must be distinguished by some other mechanism. Typically a different chemical species is chosen for each signal, and the differences in chemical activity distinguish the signals.

The physics of this chemical distinction mechanism, combined with some reasonable assumptions about the classes of molecules being used as signals, some information theoretic tools, and simple diffusion kinetics, can give us fundamental bounds on the (complexity, speed, error) envelope within which a chemical computer must operate.

Previous Work: Monod and Jacob suggested long ago [7] that organisms regulate gene expression using logic-like functions. Many have since noted [2, 3, 8, 10, 11] that living systems implement logic functions using biochemical reaction networks. Smith and Schweitzer [9] argue that in fact, many biological functions performed by cells are Turing complete.

Not all work in this field has been analytic. Engineers and computer scientists have long dreamed of a synthetic approach to designing chemical computers. In the early 1990's Hjelmfelt *et al* [4] put the field on a firm theoretical footing, by giving explicit design procedures for chemical gates, amplifiers, perceptrons, *etc*. In 1994 Adleman [1] computed a solution to the Hamiltonian Path problem *in vitro* using DNA hybridization. In 1997 Knight and Sussman [5] gave an example of how logic gates might be implemented using synthetic genetic regulatory networks.

In 1997 Magnasco [6] offered a proof that chemical kinetics is Turing universal. Magnasco treats chemistry as abstract: He chooses an abstract set of reactions and rate constants and then assumes a set of chemical species can be found to meet those specifications. It is the primary contribution of this work to demonstrate that at least for biopolymers, this assumption is unfounded, and that any single-vessel chemical reaction system based on biopolymers has bounded computational capability.

Approach: The approach combines results from the areas of chemical physics, information theory, transport phenomena, and statistical mechanics. We assume the design variables we may choose for a chemical computer are: The operating density, the length of the signaling molecules, and the number of signals. From these we derive bounds on the operating frequency and error rate of such a computer, in terms of the independent variables and some constants related to the chemistry being used.

The operating frequency is derived from the mean separation distance and the signal molecule diffusion constant, which is in turn derived by applying a scaling relation to the diffusion constant of a monomer, and the signal molecule length.

Now considering the signal molecules as codewords in a space of strings of length L, we can obtain a lower bound on the error rate per gate operation. The number of signals combined with the code length and the base of the code determines an upper bound on the minimum separation distance between codewords. We apply an interaction energy scaling relation to this minimum separation distance to obtain an upper bound on the difference in Gibbs free energy between binding of the correct signal molecule and an incorrect signal molecule. We then apply the Maxwell-Boltzmann distribution to this energy separation to find the fractional occupancy of the receptor site by incorrect signal molecules.

Taking this fractional occupancy as an error rate, we can then compute a figure of merit for the chemical computer in terms of the number of signals, the operating frequency, and the error rate.

Difficulty: The difficulty in this work lies mainly in pinning the theory to reasonable assumptions about the underlying physics of the system. While the results may be intuitively believable, much attention must still be devoted to clarifying and substantiating the assumptions, and delineating the classes of systems for which they are valid. This is an ongoing process.

Impact: The most immediate impact of this result will be to demonstrate what is not possible with traditional chemical computation. But positive things might also come of a better understanding of the limitations: This might inform directions for future chemical computation technology by indicating which aspects of traditional chemical computation are the most performance-limiting.

It is also possible that evolutionary biologists and researchers into the origins of life would find this work interesting and informative.

Future Work: We wish to extend this result to more general chemistries (*i.e.* non-polymers). In this case it is not so simple to put a lower bound on the free energy of association of nonmatched molecules. This problem might be addressed with either an information-theoretic approach; or computationally, for instance by using a lattice model of binding sites.

Testing by real physical experiment is also an area we wish to explore. This might be done for instance using an *in vitro* RNA transcription model of computation.

Marginal sensitivity analysis might also be interesting to predict the error rate of computational subsystems *added* to an extant chemical computer. Such analyses might be useful in predicting error rates of *in vivo* logic systems carried on plasmids.

Research Support: This work was supported partially by a National Science Foundation Graduate Research Fellowship, and partially by a Merck Corporation Graduate Research Fellowship.

References:

[1] L. M. Adleman. Molecular Computation of Solutions to Combinatorial Problems. *Science* 266 (5187) pp 1021. Nov 11, 1994

[2] A. Arkin and J. Ross. Computational Functions in Biochemical Reaction Networks. *Biophys. J.* 67 pp 560–578, Aug 1994

[3] Dennis Bray. Protein molecules as computational elements in living cells. *Nature* 376 pp 307–312, Jul 1995

[4] A. Hjelmfelt, E. D. Weinberger and J. Ross. Chemical Implementation of Neural Networks and Turing Machines. *Proc. Natl. Acad. Sci.* 88 pp 10983–10987, Dec 1991

[5] Thomas F. Knight Jr. and Gerald Jay Sussman. Cellular Gate Technology. in *Unconventional Models of Computation* pp 257–272. Springer, New York, 1998

[6] Marcelo O. Magnasco Chemical Kinetics is Turing Universal *Phys. Rev. Lett.* 68(6) pp 1190–1193, 1997

[7] J. Monod and F. Jacob. in *Cellular Regulatory Mechanisms* pp 389–401. Cold Spring Harbor press, New York, 1961

[8] Frederick C. Neidhardt and Michael A. Savageau. Regulation Beyond the Operon. in *Escherichia Coli and Salmonella* pp 1310–1324. ASM Press, Washington, D.C., 1992

[9] Warren D. Smith and Allan Schweitzer. DNA computers in vitro and vivo. NECI Tech Report number 983 Princeton, NJ, Apr 1995

[10] M. Sugita *J. Theor. Biol.* 4 pp 179–192, 1963

[11] R. Thomas Boolean Formalization of Genetic Control Circuits. *J. Theor. Biol.* 42 pp 563–585, 1973