Natural Information Networks

A Universal Paradigm

by Jehoshua (Shuki) Bruck

Claude Shannon established the field of Information Theory in 1948. His groundbreaking work provided a quantitative characterization of information (the notions of bits and entropy), quantitative limits on information transmission (the notion of channel capacity) and methods to achieve the capacity through coding. With this new language to describe information, including, compression, and coding, the technologies for information communications and storage have flourished. The past fifteen years have presented new challenges for reasoning about information and networks, including (i) the marriage of information storage and communications, namely, the emergence of the World Wide Web and (ii) continuous connectivity between people, namely, the emergence of wireless networks. In a sense, these years saw the emergence of a connected civilization—or a natural information network. These natural information networks are not limited to the macro level, and it is also intriguing to consider micro level networks, such as biological networks, including neural networks and gene regulatory networks. Also, the emergence of natural information networks has led to studies in social science and econonomics.

My research program strongly benefited from the generous support of the Lee Center and allowed me to explore topics relating to natural networks that are not yet supported by traditional funding agencies. Examples of topics include: wireless networks and percolation theory, computation with biological networks, networks of relation and capacity of storage, and most recently, information representation in flash memories. In addition, it provided a flexible framework to attract and train the best graduate students and postdocs. In fact, two of my students supported by the Lee Center won the Wilts prize for the best PhD thesis in Electrical Engineering at Caltech (M. Franceschetti in 2003 and M. Riedel in 2004).

The Physics of Wireless Networks

Continuum percolation relates to the study of points that are distributed uniformly on the plane, where each point is the center of a disc of given radius. A natural question is: what is the minimal radius that results in an emergence of a large area of connected discs? This critical event, the emergence of a large connected area, is called a phase transition. This question has been extensively studied; in fact, in 1982 Kenneth Wilson (Caltech PhD, 1961) won the Nobel Prize for using these concepts to make calculations and predictions related to solid-state physics. Motivated by the challenges of wireless networks, especially, the notion of multi-hop routing, we considered a generalization in which a deterministic algorithm places the discs on the plane, but in such a way that each disc covers at least one point. We study the percolation properties of this new model, showing that, in the most general case, an unbounded connected component of discs does not exist. However, we identify some large families of covering algorithms, for which such an unbounded component does arise. This work introduced the tools from percolation theory to the domain of wireless networks and resulted in new methods for computing capacity, routing efficiency and reliability in wireless networks. This work was mainly done by M. Franceschetti (Caltech PhD, 2003), who is currently a tenured associate professor at UCSD.

Computation in Stochastic Biological Networks

Stochastic chemical reaction networks are among the most fundamental models used in chemistry, biochemistry, and most recently, computational biology. Traditionally, analysis has focused on mass action kinetics, where reactions are assumed to involve sufficiently many molecules that the state of the system can be accurately represented by continuous molecular concentrations with the dynamics given by deterministic differential equations. However, analyzing the kinetics of small-scale chemical processes involving a finite number of molecules, such as occurs within cells, requires stochastic dynamics that explicitly track the exact number of each molecular species. For example, over 80% of the genes in the *E. coli* chromosome are expressed at fewer than a hundred copies per cell. Further, observations and computer simulations have shown that stochastic effects resulting from these small numbers may be physiologically significant.

The power of different systems to do computa-

tion can vary greatly. It has previously been assumed that systems such as genetic regulatory networks and chemical reaction networks are much weaker than the gold standard computational systems such as Turing machines. On the other hand, we have become accustomed to proofs that even some of the simplest systems are capable of universal computation, meaning that they are in some sense equivalent in power to Turing machines. As a result, predicting their eventual behavior is impossible even in theory. Chemical reaction networks have been shown to be computationally universal when combined with some form of memory such as a polymer memory or membrane separated compartmental memory; however, researchers have previously assumed that, on their own, a finite number of species in a well-mixed medium can only perform bounded computations.

In contrast with this historical intuition, we have shown that, in fact, such "plain" chemical reaction networks can indeed perform unbounded computation, using the concentration (number of molecules) of each species as the storage medium. Both the power and the weakness of chemical reaction network computation were also pinpointed by showing that they are as fast as a Turing machine, but require exponentially more space. This universality of chemical reaction networks derives from their probabilistic nature. If the possible reactions in a chemical system could be prioritized so that the next reaction at each step is always the one with highest priority, then universal behavior would be attainable; but, of course, chemistry does not behave in this way. However, since the reaction rates in a chemical system are influenced by the concentrations, they are somewhat under the control of the system itself and this weak form of prioritization is sufficient to let the system perform universal computation with a high probability of success. If we require that the chemical system guarantee the correct answer without fail, then the system is effectively deprived of the opportunity to use its reaction rates and in this situation, it is incapable of universal computation. Thus, the stochastic reaction rate foundation turns out to be the source of the computational power in chemical reaction networks.

This work was performed in collaboration with M. Cook (Caltech PhD, 2005), D. Soloveichik (Caltech PhD, 2008) and Professor Erik Winfree. My collaboration with Erik Winfree led to continued collaboration on the Molecular Programming Project, funded by the NSF's 2008 Expeditions in Computing Program.

Current and Future Work

Why do natural systems seem miraculous? The key reason is because we still do not know how to design systems that do what biological cells do. However, while we do not know how to design complex molecular systems, we do know how to design computers that are highly complex information systems. For example, we are capable of designing and manufacturing computer chips with billions of transistors. How did we get there? The breakthrough in digital circuit design came when we discovered a way to systematically express our ideas using a language of symbols (calculus) that in turn could be mapped to working physical systems. Historically, two key milestones in logic design are: (i) George Boole in 1854 showing how to express syllogism (logic) with algebra (Boolean algebra), thereby, creating a calculus for logic; and (ii) Claude Shannon in 1938 showing how to implement arbitrary Boolean functions by series-parallel relay networks, hence, connecting Boolean calculus with physics.

"The stochastic reaction rate foundation turns out to be the source of the computational power of chemical reaction networks."

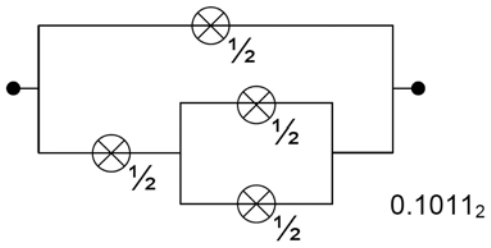

Figure 1: A stochastic network that implements the probability 11/16 with 4 probabilistic switches that are closed with probability 1/2.

Inspired by Shannon's work, we are now consid-

ering stochastic switching networks. These are relay networks that consist of probabilistic switches, and we study the expressive power and the reliability of these networks. In addition, we are experimenting with molecular systems, based on DNA strands, that implement stochastic networks. LEE

Jehoshua (Shuki) Bruck is the Gordon and Betty Moore Professor of Computation and Neural Systems and Electrical Engineering.

Read more at: http://paradise.caltech.edu

References

[1] A. Jiang, R. Mateescu, M. Schwartz and J. Bruck, "Rank Modulation for Flash Memories," *IEEE Trans. on Inform. Theory*, Vol. **55**, No. 6, pp. 2659-2673, June 2009.

[2] M. Schwartz and J. Bruck, "Constrained Codes as Networks of Relations," *IEEE Trans. on Inform. Theory*, Vol. **54**, No. 5, pp. 2179-2195, May 2008.

[3] D. Soloveichik, M. Cook, E. Winfree and J. Bruck, "Computation with Finite Stochastic Chemical Reaction Networks," *Natural Computing*, 2008.

[4] M. Franceschetti, J. Bruck and L. Schulman, "A Random Walk Model of Wave Propagation," *IEEE Trans. on Antennas and Propagation*, Vol. **52**, No. 5, pp. 1304-1317, May 2004. (Winner of the Schelkunoff prize for the best paper in the IEEE Transactions on Antennas and Propagation for 2005.)

[5] L. Booth, J. Bruck, M. Franceschetti and R. Meester, "Covering Algorithms, Continuum Percolation, and the Geometry of Wireless Networks," *Annals of Applied Probability*, **13**(2), pp. 722-741, May 2003.