Dynamics in Genetic Networks

Roderick Edwards and Leon Glass

Abstract. Living organisms contain thousands of genes. These genes contain the genetic code for protein molecules required for the functioning of the organism. During the development of the organism, genes are activated and deactivated at different times and in different places. Molecular biologists have developed remarkable techniques for determining the expression patterns of genes and the mechanisms regulating their expression. Mathematical modeling of these networks proves a formidable task. We show how logical circuits and piecewise linear equations are being used to meet this challenge. The mathematics not only can successfully model gene control in complex organisms but is also posing new questions for mathematical analysis.

1. INTRODUCTION. Different cells of the body are characterized by different sets of protein molecules. Key functions and structures of cells depend on their protein components. For example, proteins are used to build many cellular components, including: ionic channels that regulate the flow of ions into and out of cells; enzymes that catalyze chemical reactions in the body; and muscle tissue, which has tiny protein molecular motors that are the molecular basis of contraction. Proteins are constructed by joining amino acids in a sequence. The particular sequence in any given protein is determined by the sequence of bases in the genetic hereditary material: the DNA. The elucidation of the genetic code and the understanding of mechanisms by which it is controlled and evolves constitute major intellectual triumphs of the second half of the the 20th century.

A key component of this work is the insight provided by Monod and Jacob into the mechanisms of regulation of gene expression [**23**]. Monod and Jacob demonstrated that protein expression in bacteria is partially regulated by the action of special regulatory molecules, which are themselves proteins, called transcription factors. The sequence of amino acids in transcription factors, like any protein, is coded by the DNA, and the production of transcription factors is regulated by other transcription factors that can bind directly to the DNA. Depending on the particular gene, transcription factors can lead to either the activation or repression of the expression of that gene. Monod and Jacob not only discovered the function of transcription factors, but they also hypothesized how control circuits could be built up by hooking together transcription factors. For example, they hypothesized that cell differentiation could be controlled by a circuit comprised of two mutually inhibitory transcription factors. Transcription factor x_1 would bind to DNA inhibiting the production of transcription factor x_2, and in turn, transcription factor x_2 would inhibit the production of x_1. Figure 1 shows this network schematically.

It is instructive to consider from the start a Boolean model for this circuit such as was initially suggested by Kauffman [**17**]. A gene expression pattern can be represented as a binary vector, in which a 1 in the i^{th} place represents that the protein product of the ith gene is expressed (at a high level), and a 0 represents that the protein product is not expressed (or only at a low level). Denoting the logical state of

http://dx.doi.org/10.4169/amer.math.monthly.121.09.793
MSC: Primary 92C42, Secondary 34A36

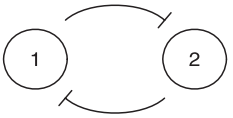

Figure 1. A network of two genes with mutual inhibition: the toggle switch. Each circle represents a gene or the transcription factor it produces. Blunted arrows represent inhibitory regulation.

a variable by a capital letter, we have the Boolean equations $X_1(t+1) = \overline{X_2(t)}$ and $X_2(t+1) = \overline{X_1(t)}$, where the overline represents negation. Thus, if the logical state of variable $X_1(t) = 0$, then $X_2(t+1) = 1$ and if $X_1(t) = 1$, then $X_2(t+1) = 0$. The truth table for this network is given in Table 1.

Table 1. Truth table for the toggle switch.

$X_1(t)$	$X_2(t)$	$X_1(t+1)$	$X_2(t+1)$
0	0	1	1
0	1	0	1
1	0	1	0
1	1	0	0

This network has three asymptotic behaviors. A steady state $X_1 = 0$, $X_2 = 1$, a steady state $X_1 = 1$, $X_2 = 0$, and a cycle between the two state $X_1 = 0$, $X_2 = 0$ and $X_1 = 1$, $X_2 = 1$. However, as Glass and Kauffman [**13**] and Thomas [**25**] recognized, in the absence of synchronous updating, the oscillation between the states 00 and 11 would not exist since two switching events would not be expected to occur simultaneously; in general, one would precede the other and the system would be frozen in either state 01 or 10. One way to represent this would be with the state transition diagram shown in Figure 2.

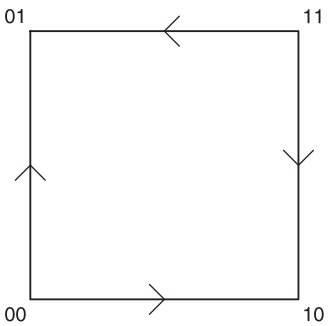

Figure 2. The state transition diagram for the toggle switch.

Monod and Jacob assumed that such a mechanism would lead to the expression of either x_1 or x_2, but not both. Thus, this circuit is a "toggle switch," and an appropriate external signal could push the network into a state where either one (but not both) genes are expressed, in which state it would then remain. More complex versions of this are possible involving more than two genes and more than two stable expression patterns. In this fashion, organisms could have molecular control circuits that could underlie differentiation into different cell types, each of which is characterized by a

particular pattern of gene expression. Monod and Jacob also proposed that hooking together transcription factors in a different fashion could generate circuits that would oscillate, another important type of behavior, necessary, for example, in the control of the cycle of cell division.

Another formulation for the toggle switch topology shown in Figure 2, or more general patterns of behavior, can be developed using differential equations. Such equations can be written in an intuitive fashion:

$$\text{Rate of change of } x_i = \text{ production of } x_i - \text{ destruction of } x_i,$$

where x_i represents the concentration in the cell of the transcription factor produced by the ith gene in the network. The production of x_i may be some nonlinear or piecewise function of the $\{x_j\}$ where $\{x_j\}$ represents some subset of the variables in the system, and we typically assume that the destruction (or *degradation*) term is linear for each x_i. Any molecule as large and complex as a protein can be broken apart by a number of processes that can be considered random, so each transcription factor is naturally degraded at a constant rate. The production of a transcription factor may be regulated by some subset of other transcription factors in the network, according to some logical rule. For example, the production of one transcription factor might be blocked (*inhibited, repressed*) by the presence (in sufficient concentration) of a second, but a third might bind to the second in a way that prevents it from acting as a repressor for the first. Thus, the first transcription factor's production is blocked only if the second is present and the third is absent.

Differential equations for gene control typically involve the Hill function, defined by either

$$f(x) = \frac{x^n}{\theta^n + x^n}, \quad \text{or} \quad \bar{f}(x) = 1 - f(x) = \frac{\theta^n}{\theta^n + x^n}, \tag{1}$$

where x now represents the concentration of a transcription factor in a cell, and f represents activation, while $\bar{f}$ represents repression (Figure 3). These are functions that increase (from 0 to 1) or decrease (from 1 to 0) sigmoidally. The parameter θ is a threshold at which the function value is $\frac{1}{2}$, and n is a positive number (often in the range 2–10) called the Hill coefficient that determines the maximum slope. In the limit

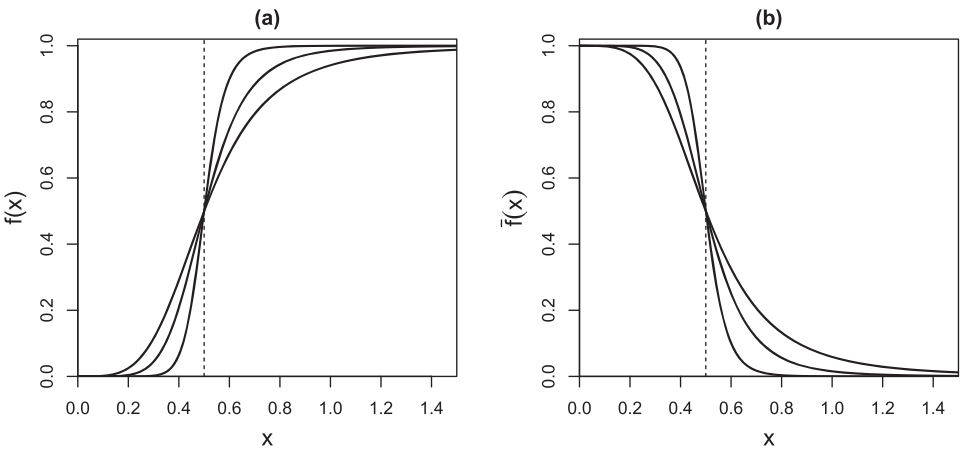

Figure 3. (a) Hill function $f(x)$, representing activation, and (b) $\bar{f}(x)$, representing repression. In both cases, $\theta = 0.5$, and the functions are plotted with $n = 4, 6, 12$ in order of increasing steepness.

$n \to \infty$, the Hill function becomes a step function that switches abruptly from 0 to 1 (or 1 to 0) as x increases through θ.

Now, using the Hill function, we can propose a hypothetical two-dimensional nonlinear ordinary differential equation for the toggle switch of Figure 1 (for example, see [**12**]):

$$\begin{aligned} \dot{x}_1 &= -x_1 + \bar{f}(x_2) \\ \dot{x}_2 &= -x_2 + \bar{f}(x_1). \end{aligned} \tag{2}$$

The qualitative analysis of this equation is a simple textbook problem (for example, see [**14**] or [**16**]). Assuming $\theta = 0.5$, there is a fixed point at $x_1 = x_2 = 0.5$ that is stable if $n < 2$ and is an unstable saddle point for $n > 2$. For $n > 2$, the phase plane is divided into four regions by the x_1 and x_2 nullclines (curves on which $\dot{x}_1 = 0$ and $\dot{x}_2 = 0$, respectively), and there are two stable fixed points: one in which x_1 is high and x_2 is low and another in which x_2 is low and x_1 is high (Figure 4(a)). In fact, there is an isomorphism between the transitions between volume elements defined by the x_1 and x_2 nullclines in Equation (2), as shown in Figure 4(b) and the state transition diagram in Figure 2. A sufficiently large external perturbation can cause a switch from one stable state to the other.

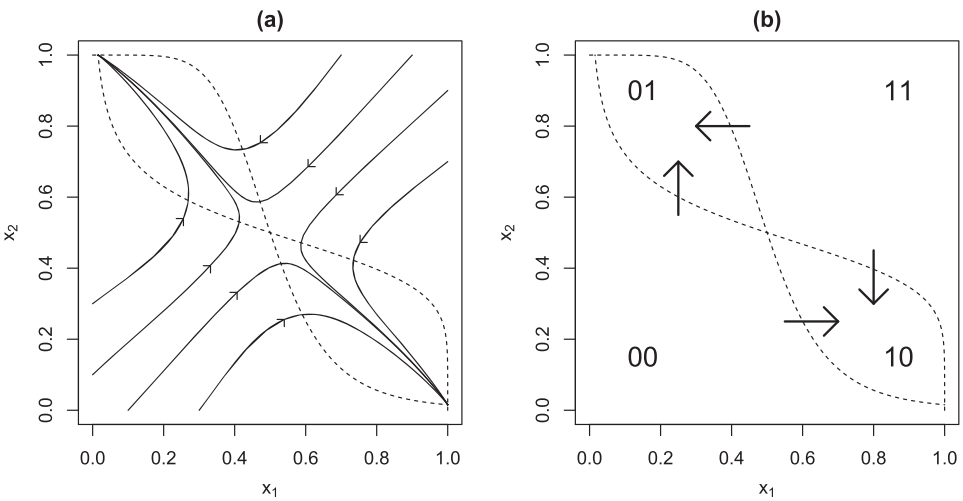

Figure 4. (a) Phase plane for the toggle switch, Equation (2), with several trajectories (solid lines) and the x_1 and x_2 nullclines (dotted lines), and (b) phase plane logic for the toggle switch: The nullclines divide the plane into four regions labelled 00, 01, 10, 11 to correspond to the logical state, and the arrows show the direction of flow between regions. Both graphs were generated using $\theta = 0.5$ and $n = 6$.

Our mathematical analysis of switching networks extends this simple connection between the discrete and continuous switching networks to arbitrarily complex networks that can be represented using state transition diagrams (which are directed graphs on hypercubes of appropriate dimension). A more complete analysis of the equations for the toggle switch can be found, for example, in [**3**].

Parallel to the mathematical analysis have been attempts to analyze natural or manmade genetic control switching networks using similar ideas. Thieffry and Thomas [**24**] made an important contribution when they proposed that the lysogeny-lysogenesis transition in bacteriophage lambda could be modeled using a toggle switch

network. Two transcription factors cro and cI mutually inhibit each other, leading to one or the other being expressed. The mathematical model they developed did not use nonlinear differential equations but rather asynchronous switching networks, with discrete delays needed for the switch. This was followed a few years later by the insertion of the toggle switch topology into the bacterium *E. coli* by Gardner *et al.* [**9**]. This paper used a nonlinear model, slightly more general than the one in Equation (2), to explain some of the technical problems involved in designing a circuit with the appropriate dynamics. The technical problems arise because of the need to have a molecular basis for the parameters, θ and n. The mathematics leads to insight into what values might work, but this leaves hard work to find constructs that have the desired values. The success of Gardner *et al.* (as well as a second paper in the same issue of *Nature* by Elowitz and Leibler that synthesized an oscillator that we describe below) has led to the development of a new field of synthetic biology.

More recently, this early work has been taken a step further by designing circuits that could be synthesized in *E. coli* bacteria that live in the gut. Using the same basic idea, Kotula *et al.* [**18**] were able to design a molecular switch that senses the presence of anhydrotetracycline in the gut, leading to the expression of a molecule that would only be present if anhydrotetracycline was above a critical threshold. Thus, it is possible to engineer a molecular sensor that will sense the concentration of a compound in the gut. This application, though still at a very early stage, shows how the toggle switch network could potentially be used for practical purposes.

This introductory section is meant to provide a motivation for the more mathematical material that follows. The simple ideas here can be greatly expanded to generate a class of fascinating mathematical problems relating to the dynamics of arbitrarily complex nonlinear equations.

More complex networks allow a wide range of dynamical behaviors. The toggle switch is an example with bistability: There are two equilibrium states, both of which are asymptotically stable. In general, networks may exhibit multistability, with two or more asymptotically stable equilibria, representing multiple patterns of gene expression that may become fixed. They can also have limit cycles: periodic solutions to which at least some nearby solutions tend asymptotically. These correspond to repeating oscillations in gene expression patterns, in which the expression of some genes switch on and off in some predictable order. More complex dynamics are also possible, at least in principle, as we will see below.

An underlying motivation for our mathematical analysis is *to determine the possible dynamics given the logical structure of the network*.

We now turn to the formalization of this system and describing some of its basic properties. We show how arbitrary logical patterns of genetic regulation can be modeled by differential equations. We also show how these genetic circuits can lead us into mathematical realms where Fibonacci sequences, hypercubes, and the Perron–Frobenius theorem play a role in understanding their dynamics. Applications of these and related methods to the study of particular biological networks is currently underway. Readers who wish to pursue these directions can look at papers in the recent focus issue of the journal *Chaos* on "Quantitative Approaches to Genetic Networks" [**1**].

2. INHIBITORY LOOPS IN GENETIC CIRCUITS.

One type of genetic control is exerted by transcription factors called repressors. When these proteins are present they block the activity of a particular gene or group of genes. The toggle switch (Figure 1) is an example: The transcription factor coded by gene 1 blocks the activity of gene 2, and in turn, the transcription factor coded by gene 2 blocks the activity of gene 1. As we saw above, the resulting behavior is that either gene 1 dominates (the

concentration, x_1, of its transcription factor stays high, while the concentration, x_2, of gene 2 stays low) or *vice versa*. Figure 5(a) shows another circuit in which there is a loop of three inhibitory elements.

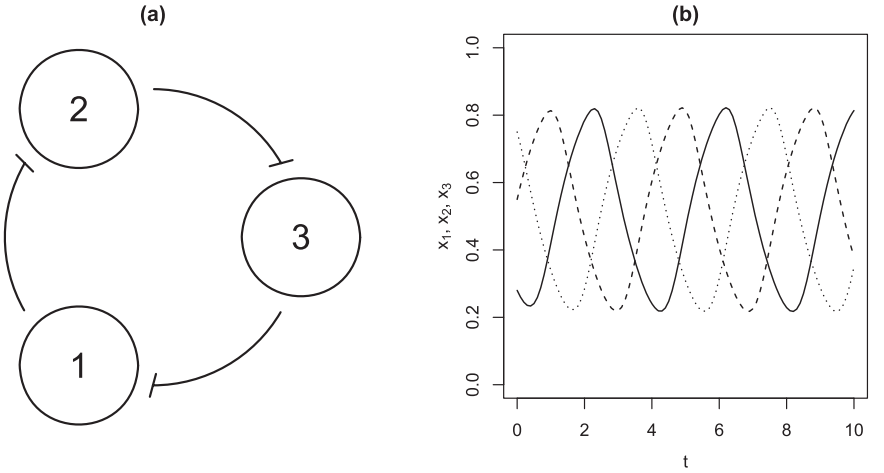

Figure 5. The repressilator. (a) Schematic diagram showing a loop of three inhibitory genes. (b) Time series from numerical integration of 3 with $N = 3,\ \theta = 0.5,\ n = 6$.

To predict the dynamics of such networks we need mathematics.

An appropriate equation to model an inhibitory loop of N elements is

$$\frac{dx_i}{dt} = -x_i + \bar{f}(x_{i-1}), \quad i = 1, \ldots, N, \tag{3}$$

where we identify $x_0 = x_N$, and $\bar{f}$ is the inhibitory Hill function from Equation (1).

Even though (3) is a nonlinear equation and cannot be integrated analytically, a great deal is known about its dynamics. Figure 4(a) shows a phase plane portrait for the case in which $N = 2$, $n = 6$. There are two stable nodes and a saddle point and we have bistability and the possibility of switching. For $N = 3$, we resort to numerical integration of the equation, again taking $n = 6$. Figure 5(b) shows the stable unique limit cycle oscillation that results from this equation, which models the oscillations Elowitz and Leibler found in their synthetic circuit, known as the repressilator [**7**]. If we carry out a local stability analysis of the fixed point at $x_1 = x_2 = x_3 = 0.5$, we find that there is an instability leading to oscillation that occurs as n increases at the value $n = 4$ [**12**].

These results illustrate the well-known observation that feedback loops show bistability if they have an even number of inhibitory steps and oscillation if they have an odd number of inhibitory steps [**26**]. This is a phenomenon known as *frustration*. The intuitive argument runs as follows. If the loop has an even number of steps and at some time the states of the genes alternate between "on" and "off" around the loop, they can all remain in that state forever. However, if the loop has an odd number of steps, this pattern of alternation cannot be achieved on the entire loop, and at least one gene at any time must be "frustrated," *i.e.*, not content to remain in its current state, so it will switch, and this perturbation will continue around the loop, leading to an oscillation. Although we can carry out simulations and stability analysis of simple negative feedback loops in higher dimensions, we are interested in more general techniques that can be used to present a clearer global analysis of the dynamics.

3. PIECEWISE-LINEAR MODELS OF GENETIC NETWORKS. The Hill function has an important property that we now exploit. As $n \to \infty$ it approaches the Heaviside step function: $H(x-\theta) = 0$ for $x < \theta$: $H(x-\theta) = 1$ for $x \geq \theta$. This allows Equation (3) to be numerically integrated in a straightforward and exact manner. Figure 6(a) shows the phase plane for mutual inhibition, and Figure 6(b) shows the time traces for the three-element inhibitory network.

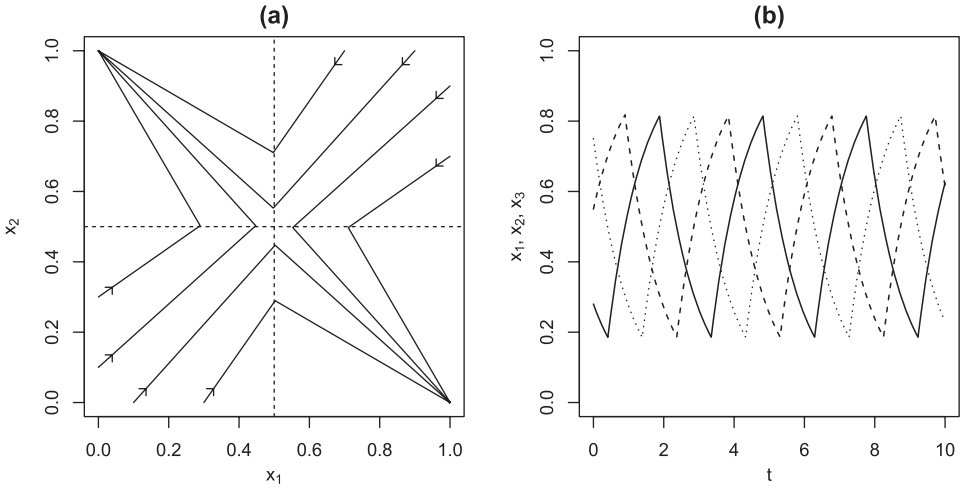

Figure 6. (a) Phase plane for the two-gene mutually inhibitory loop, with $n \to \infty$. (b) Time course of the periodic solution of the three-gene inhibitory loop, with $n \to \infty$.

For these piecewise linear equations, the time traces will be exponentially increasing and decreasing curves that all have the same time constant [**12, 13**]. As a consequence, the trajectories in phase space are straight lines. For example, for mutual inhibition, *i.e.*, (3) with $N = 2$, if we have an initial condition $x_1(0) > 0.5$, $x_2(0) > 0.5$ we find

$$x_1(t) = x_1(0)e^{-t}, \quad x_2(t) = x_2(0)e^{-t}$$

so that the trajectory in phase space is the straight line $\frac{x_2(t)}{x_1(t)} = \frac{x_2(0)}{x_1(0)}$. This is just a straight line between the two points $(x_1(0), x_2(0))$ and $(0, 0)$. However, once the trajectory crosses either $x_1 = 0.5$ or $x_2 = 0.5$, there would be a change in the equations. For example, assume that $x_1(0) < x_2(0)$ so that x_1 would first reach the threshold 0.5 at time t^*. This would lead to a release of the repression of x_2 so that the solution of (3) with the new initial condition at t^* is now

$$x_2(t) = 1 + (x_2(t^*) - 1)e^{-(t-t^*)}, \quad x_1(t) = 0.5e^{-(t-t^*)}.$$

This leads to the equation of the straight line $\frac{x_2(t)-1}{x_1(t)} = \frac{x_2(t^*)-1}{0.5}$, which is the equation for the straight line between the points $(0.5, x_2(t^*))$ and $(0, 1)$.

Clearly, the qualitative features—the numbers and stability of the fixed points—are preserved even though the piecewise linear equation is different from the original nonlinear equation.

For the network with three inhibitory links, there is once again a stable limit cycle oscillation as illustrated in Figure 6(b). Notice that, in this network, there is a symmetry

around the threshold intersection at (0.5, 0.5, 0.5). We can exploit this symmetry and the piecewise linear nature of the network to give a heuristic computation of the points on the limit cycle and the period of the limit cycle [**15**]. Call τ the length of time for any element to transit from its maximum to the threshold (0.5). This is also the length of time for any element to transit from its minimum to the threshold. From Figure 6(b), note that when one variable is at 0.5, another is at an extremum. We find that 2τ is the length of time to transit from 0.5 to either extremum because, in that time, the two other variables make extremum-to-threshold transits in succession. Calling the value of any variable at the maximum $1/2 + \delta$ and the value at the minimum $1/2 - \delta$, we find

$$\left(\frac{1}{2} + \delta\right) e^{-\tau} = \frac{1}{2}, \quad \frac{1}{2} e^{-2\tau} = \frac{1}{2} - \delta.$$

There are two equations and two unknowns, and this can be solved using elementary algebra. Calling $b = e^{-\tau}$, we find $b = \frac{-1+\sqrt{5}}{2}$, $\delta = \frac{\sqrt{5}-1}{4} = .3090\ldots$, and the period of the oscillation is $6\tau = -2\ln(\sqrt{5} - 1) = 2.8873\ldots$.

The extension of this result to higher dimensions leads unexpectedly to generalized Fibonacci series [**15**]. In negative feedback networks in higher dimensions, we find that

$$\left(\frac{1}{2} + \delta\right) e^{-\tau} = \frac{1}{2}, \quad \frac{1}{2} e^{-(N-1)\tau} = \frac{1}{2} - \delta.$$

Once again calling $b = e^{-\tau}$, we find that b is a solution of the polynomial equation $b^{N-1} + b^{N-2} + \cdots - 1 = 0$. We define the generalized N^{th} order Fibonacci series by the recurrence relation $z_k = z_{k-1} + z_{k-2} + \cdots + z_{k-N}$. If we define the asymptotic ratio $\phi_N = \lim_{k\to\infty} z_k/z_{k-1}$ for the N^{th} order series, we find that for the generalized feedback oscillator of N genes, the period is $2N\tau$, where $\tau = -\ln \phi_{N-1}$. It is well known that $1/\phi_2 = \frac{1+\sqrt{5}}{2}$, which is the golden ratio.

Our ability to solve analytically for the period of the oscillation in this class of negative feedback networks is satisfying but leaves open questions of uniqueness and stability. To address these questions, we will place these equations in a much more general framework.

4. LOGICAL NETWORKS AND SYMBOLIC REPRESENTATIONS OF THE DYNAMICS. For these equations, we can separate the phase space into regions defined by the thresholds. Each region can be labeled by a logical state, 0 or 1, depending on whether the variables are below or above their threshold, respectively. Thus, for the two-gene toggle switch network, if $x_1 > \theta$ and $x_2 < \theta$, we would label this state 10. Further, we can label the edges of the resulting graphs showing the directions of the flows between neighboring states. The resulting state transition diagram for the toggle switch is shown in Figure 2. The corresponding diagram for the three-gene repressilator is Figure 7.

These state transitions show a qualitative correspondence with the dynamics in the differential equations. For the switch, there are two vertices that are attracting, corresponding to the two steady states in the differential equation. For the oscillator, there is a stable cycle that corresponds to the circuit in three dimensions in the differential equation. These observations hint at the possibility of a more general approach in which the inhibitory loops we have focused on up until now represent simple paradigmatic examples.

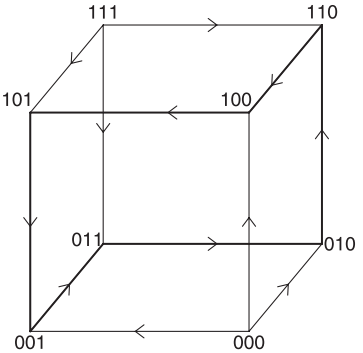

Figure 7. The state transition diagram for the repressilator.

In Table 1, we have already given the logical truth table that defines the dynamics of the toggle switch. We now give the general case. A Boolean switching network with N elements is defined by the equations

$$X_i(t+1) = F_i(X_{i_1}(t), X_{i_2}(t), \ldots, X_{i_K}(t)), \quad i = 1, \ldots, N, \tag{4}$$

where $X_i(t) \in \{0, 1\}$ are logical variables, $F_i(X_{i_1}(t), X_{i_2}(t), \ldots, X_{i_K}(t)) \in \{0, 1\}$, and K is the number of inputs to each element. This is a discrete time and discrete state space system that, therefore, must eventually reach a fixed point or cycle under iteration.

For example, for a negative inhibitory loop of N genes, $F_i = 1 - X_{i-1}$, where X_0 is identified with X_N. Thus, for the repressilator network, the corresponding logical network is given in Table 2. This can then be used to construct the state transition diagram (Figure 7). For each state (node of the state transition diagram or line of the truth table), the outgoing arrows are determined by checking which variables change state at time $t + 1$.

Table 2. Truth table for the repressilator.

$X_1(t)$	$X_2(t)$	$X_3(t)$	$X_1(t+1)$	$X_2(t+1)$	$X_3(t+1)$
0	0	0	1	1	1
0	0	1	0	1	1
0	1	0	1	1	0
0	1	1	0	1	0
1	0	0	1	0	1
1	0	1	0	0	1
1	1	0	1	0	0
1	1	1	0	0	0

For any logical network, we define an analogous differential equation, where for convenience we take all thresholds to be 0. We may think of this as a translation so that each variable x_i now represents the amount by which the concentration of a transcription factor is above its threshold. First, we connect the logical and continuous variables using

$$X_i(t) = 0 \quad \text{if } x_i(t) < 0; \quad \text{otherwise} \quad X_i(t) = 1. \tag{5}$$

Then we have

$$\frac{dx_i}{dt} = -x_i + f_i(X_{i_1}(t), X_{i_2}(t), \ldots, X_{i_K}(t)), \quad \text{for } i = 1, \ldots, N, \tag{6}$$

where $f_i(X_{i_1}(t), X_{i_2}(t), \ldots, X_{i_K}(t))$ takes the value of the corresponding logical variable $F_i(X_{i1}(t), X_{i2}(t), \ldots, X_{i_K}(t))$ [1].

As described for the simplified examples above, for each variable, the temporal evolution is governed by a first-order piecewise linear differential equation. Let $\{t_1, t_2, \ldots, t_k\}$ denote the *switch times* when any variable of the network crosses $\theta_i = 0$. The solution of (6) for each variable x_i for $t_j < t < t_{j+1}$, is

$$x_i(t) = x_i(t_j)\, e^{-(t-t_j)} + f_i(X_{i_1}(t_{j^*}), X_{i_2}(t_{j^*}), \ldots, X_{i_K}(t_{j^*}))(1 - e^{-(t-t_j)}), \tag{7}$$

where t_{j^*} is any time in (t_j, t_{j+1}). This equation has the following property. All trajectories in a given orthant in state space are directed toward a focal point. If the focal point lies in a different orthant from the initial point, then, in general, a threshold hyperplane will eventually be crossed. When the threshold hyperplane is crossed, a new focal point will be selected based on the underlying equations of motion.

Because of their simple mathematical structure, these differential equations admit a simple method for integration. The method consists of setting an initial condition and determining the set of times, $\{\tau_i\}$, when $x_i = 0$ from (7). However, if the system is in an orthant such that none of the variables will ever cross 0, (*i.e.*, the focal point coordinates f_i have the sign of x_i for each i), the system will approach a steady state in that orthant. Otherwise, the system will cross to a new orthant at the time $t_{min} = \min\{\tau_i\}$ at a point that can be explicitly calculated. Once again, using (7), the equations can be analytically integrated, and then the process is iterated.

For this system, the state transition diagram can be generated based solely on the logical structure. The flows between neighboring orthants in state space must follow these prescribed paths. The orientations on edges are determined using the following algorithm. First, define the Hamming distance between two Boolean states to be the number of digits in which they differ. From each Boolean state at time t in the left-hand column of the truth table, determine the Hamming distance to the next state at time $t + 1$ in the right-hand column of the truth table. Now, draw directed edges from each vertex to all vertices that lie on a shortest (undirected) path from the vertex corresponding to the state at time t to the vertex corresponding to the state at time $t + 1$. This will generate a number of directed edges corresponding to the Hamming distance between the two states.

This construction has special properties for the situation in which the differential equations have no self-input, *i.e.*, X_i is **not** one of the inputs of f_i in (6) [**4, 15**]. In this case, the Hamming distance from each column of the truth table at time t to the same column at time $t + 1$ is 2^{N-1} so that the total Hamming distance for the whole truth table is $N \times 2^{N-1}$. This number is identical to the number of edges in the N-cube. Further, from the condition of no self-input, no edge can be directed in two different orientations. Consequently, on the N-cube, each edge is directed in a unique orientation, and there is a 1:1 correspondence between the N-cube, and the truth tables of Boolean networks with no self-input. Further, the N-cube representation that shows the allowed flows between neighboring orthants for any differential equation of the form in (6) can be constructed. Conversely, if thresholds are set, then by observing

[1] For more general formulations of the class of continuous differential equations that correspond to a given Boolean network, see [**4**].

dynamic behavior, it should be possible to determine the underlying logical structure of the network.

We can now be more specific about the objective stated in the introduction. An underlying motivation for our mathematical analysis is to answer the following question: *Once the logical structure in (4) of a network is set, then what are the possible dynamics in the associated differential equation, (6)?* We are only able to answer this question in some cases.

Numerical integration, types of attractors, and decoding the *N*-cube. These equations can display a whole range of qualitatively different types of dynamics, including fixed points (nodes or stable foci), limit cycles, chaos, and quasiperiodicity. In this section, we briefly describe these different types of behavior, and in the next section, we develop analytic methods to analyze particular networks.

Nodes occur when a focal point lies in its own orthant, as described above. Clearly, they are asymptotically stable because all trajectories in that orthant must converge to the focal point. The signature for a stable node in the N-cube graph is a vertex with all N edges directed toward it (e.g., the states 01 and 10 in Figure 2).

Stable foci occur when two or more variables switch in some sequence but approach zero as the number of switchings approaches infinity, while all other variables (if any) converge to some (possibly nonzero) value. Thus, stable foci typically have an associated cyclic sequence of orthants through which an approaching trajectory passes. A necessary condition for a stable focus is a cycle on the N-cube digraph. We say that a cycle has dimension k if k is the dimension of the smallest hypercube on which the given cycle can be drawn.

Limit cycles are closed trajectories toward which at least some nearby trajectories converge, and they, of course, have an associated cyclic sequence of orthants, through which the limiting trajectory passes. A necessary condition for a limit cycle is once again a cycle in the N-cube. Sharper results are available for a class of cycles called cyclic attractors. In a cyclic attractor, each vertex on the cycle is adjacent to $N-2$ vertices not on the cycle, and all the edges from these vertices are directed toward the cycle. For example, there is only one three-dimensional cyclic attractor, and this is the cycle found for the repressilator, Figure 7. When all focal points have coordinates at 0 and 1, (*i.e.*, all $f_i \in \{0, 1\}$ in 6) and all thresholds are at 0.5, the cyclic attractor in two dimensions is associated with a stable focus, whereas all cyclic attractors in higher dimensions are associated with stable limit cycle oscillations (see [**15**] for a proof). However, this class of stable limit cycles represents only a very small fraction of the actual number of possible cycles.

It is nevertheless possible, remarkably, to determine existence and stability of periodic trajectories passing through any particular cyclic sequence of states, even if it is *not* strictly attracting, *i.e.*, if the cycle in the N-cube has some adjacent edges directed away from the cycle. If one computes a trajectory that follows a specified cycle on the N-cube, by starting at an arbitrary initial point on a particular orthant boundary (a transition point, where one variable is at its threshold, *i.e.*, 0) and calculating the sequence of transition points around the cycle given that initial value, one obtains a map from the starting boundary to itself. For the piecewise-linear equations with equal degradation rates and, therefore, trajectories composed of straight line segments, this return map turns out to be a fractional linear map of the form

$$M(\mathbf{x}) = \frac{A\mathbf{x}}{1 + \phi^{\top}\mathbf{x}}, \tag{8}$$

where $\mathbf{x}$ is an $(N-1)$-dimensional vector of gene product concentrations (omitting the one that is at 0, *i.e.*, at threshold), A is an $(N-1)\times(N-1)$ matrix, and ϕ is an $(N-1)$-dimensional vector. Not all initial points on the starting orthant boundary necessarily lead to trajectories that follow the cycle under consideration, but the initial region on the starting boundary from which the cycle *is* followed, called the *returning cone*, can be calculated explicitly. Now, periodic trajectories of the network correspond to fixed points of the return map for the cycle. The denominator of the map is scalar, so any fixed point must satisfy $A\mathbf{x} = (1+\phi^{\top}\mathbf{x})\mathbf{x}$ and, thus, must lie on an eigenvector of the matrix A. If the eigenvector lies in the returning cone for the cycle, then there is actually a periodic trajectory that follows the specified cycle. Its stability can be determined by a straightforward calculation of the Jacobian matrix for the map, and it turns out that it is asymptotically stable if the eigenvalue of A corresponding to the fixed point, $\mathbf{x}$, is the dominant one.

More complex dynamical behaviors are also possible. Quasiperiodicity arises if there are two or more disjoint stable limit cycle attractors in which none of the variables of one of the cycles receive inputs from the variables involved in the other cycles and the periods of all cycles are noncommensurate.

Chaotic dynamics are by definition aperiodic dynamics in deterministic systems with sensitive dependence on initial conditions. Although many refer to chaotic dynamics in (4), this does not conform to standard nonlinear dynamics terminology since such systems have a finite state space and must necessarily cycle. However, in (6), chaotic dynamics are possible and have been demonstrated numerically and analytically in some example networks [**5, 22**]. We have no general way to predict whether any particular logical structure is capable of generating chaotic dynamics for some set of parameters. A necessary condition for chaotic dynamics is that there is a vertex that lies on at least two different cycles. More particular structures with such interacting cycles that allow for chaotic dynamics have been described [**6, 20**].

In the next section, we take a closer look at one such example.

5. CHAOS. Chaotic dynamics arise in networks of a large number of genes typically when there are many network connections, *i.e.*, genes are influenced by many transcription factors. However, it is possible for chaos to arise even in networks of as few as four genes (and even with only three if degradation rates are not all the same [**19**]). An example is the network described by the equations [**5**]:

$$\begin{aligned}
\dot{x}_1 &= -x_1 + 2(\bar{X}_3X_4 + X_2X_3) - 1\\
\dot{x}_2 &= -x_2 + 2(X_1\bar{X}_3X_4 + \bar{X}_1X_3X_4 + \bar{X}_1\bar{X}_3\bar{X}_4) - 1\\
\dot{x}_3 &= -x_3 + 2(\bar{X}_1X_2 + X_1X_4) - 1\\
\dot{x}_4 &= -x_4 + 2(X_2\bar{X}_3 + \bar{X}_1X_3) - 1
\end{aligned} \tag{9}$$

where $\bar{X}_i = 1 - X_i$, and the thresholds have all been translated to 0. The state transition diagram for this network is shown in Figure 8. Here, the most complicated regulation is that of gene 2, but the rule for its expression being switched on and off can nevertheless be understood. It is inherently active but can be deactivated by the presence of any of the three transcription factors 1, 3, or 4, with the exception that transcription factor 4 can prevent the deactivation if it is present in conjunction with either one of transcription factor 1 or 3 (but not if 4 appears by itself, or if all three are present). Such a situation, while complicated, could plausibly occur in a cell.

Solutions to these equations from a large range of initial conditions tend to fall into a dynamic pattern of gene expressions turning on and off according to one of two cycles.

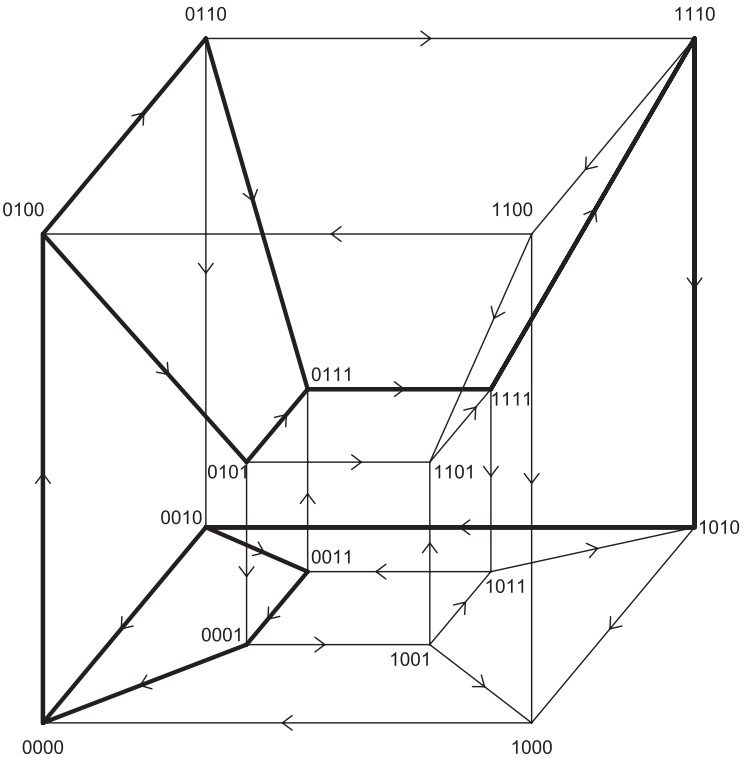

Figure 8. The state transition diagram for the network given by Equations 9. The bold lines indicate the paths taken by the two cycles of the aperiodic attractor.

These two cycles can be represented by their logical states, or hypercube vertices, as follows:

$$A: \quad 1111 \to 1110 \to 1010 \to 0010 \to 0000 \to 0100 \to \mathbf{0110} \to 0111,$$
$$B: \quad 1111 \to 1110 \to 1010 \to 0010 \to \mathbf{0011} \to \mathbf{0001} \to 0000 \to 0100 \to \mathbf{0101} \to 0111.$$

The two cycles are similar, but in a couple of places, the sequence of switchings differs (indicated by bold type). In cycle A, if we start with all genes in the "on" state, they are switched off in sequence, first gene 4, then 2, then 1, and finally 3, to arrive at the 0000 state. Cycle B starts similarly, but before gene 3 is switched off, gene 4 comes on again briefly and then turns off after gene 3 goes off so that we still arrive at 0000. Then, in cycle A, the genes switch on again in the sequence 2, then 3, then 4, and finally 1. In cycle B, the sequence is slightly different, with 4 switching on before 3.

Numerical simulations show that these two cycles follow each other in an apparently random order. Thus, following a circuit of cycle A, the next cycle may be either A or B, though B is always followed by A. It is remarkable that these observations can be proven explicitly by calculating the sets of initial conditions on a Poincaré section from which cycle A is followed, and similarly for cycle B, and then computing the images of these regions. The image of the A region lies in the union of the A and B regions, intersecting them, while the image of the B region lies entirely inside the A region. Thus, A must always be followed by either A or B (no other behavior is possible), and B must always be followed by A (Figure 9).

One might still anticipate that the sequence of cycles A and B might eventually repeat, suggesting that there might be a long and complicated stable periodic orbit.

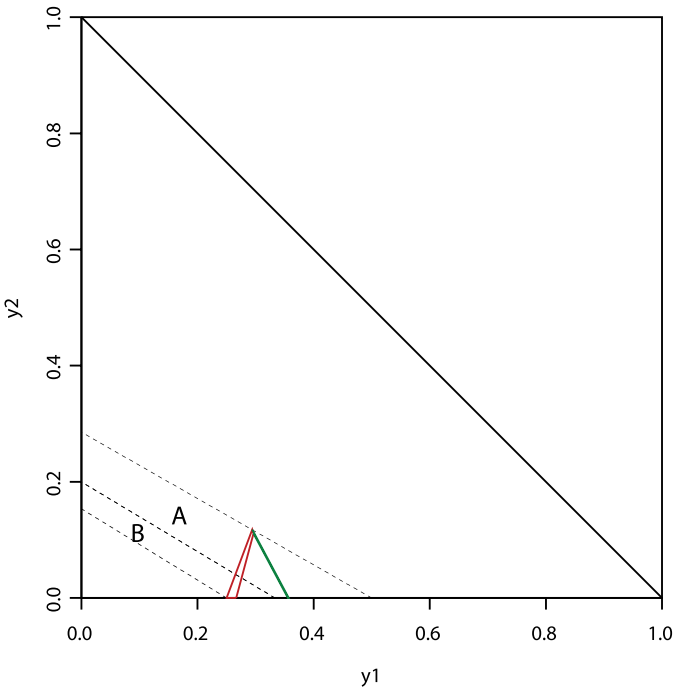

Figure 9. Returning regions for cycles A and B and their images in a projection. The narrow triangular region that crosses over the A and B regions is the image of the A region (under the map for cycle A). The image of the B region (under the map for cycle B) is even narrower and lies entirely in the A region (it is so narrow that it appears as just a line). The boundary on which these maps are defined is $\{y_1 > 0, y_2 > 0, y_3 > 0, y_4 = 0\}$. This has been projected onto the simplex $y_1 + y_2 + y_3 = 1$ and then only the y_1 and y_2 coordinates are plotted. Thus, the triangle in the plot with vertices $(0, 0)$, $(0, 1)$, and $(1, 0)$ represents the simplex, but in fact, the three-dimensional region under consideration is a triangular cone through this simplex.

This, however, can also be shown not to be the case. In other words, it is possible to prove rigorously that there is no stable periodic orbit following any sequence of these two cycles A and B so that whatever attractor is present supports only aperiodic dynamics. This is shown by contradiction using the Perron–Frobenius theorem. In its simplest form, this theorem states that a positive matrix, which maps the positive orthant into itself, has a unique eigenvector (up to scalar multiples) in the positive orthant and it corresponds to the largest, (*i.e.*, the dominant) eigenvalue. The theorem also applies to matrices for which a transformed version of the positive orthant (called a *proper cone*) is mapped into itself [**2**]. Now, if there were a stable periodic orbit for some sequence of As and Bs, then the dominant eigenvector of a matrix appearing in the Poincaré map for this sequence would have to lie somewhere in the intersection of the regions on which the cycles A and B are defined, but it can be shown that the dominant eigenvector of any such sequence is constrained to lie elsewhere [**5**]. This proves only that there is an attractor with aperiodic dynamics. There exist other examples where chaotic dynamics have been proven explicitly [**6**].

6. EXTENSIONS. The modelling framework outlined above enables many of the issues of dynamics of gene regulation to be analyzed, but much of this analysis can be extended to more realistic models where fewer simplifications have been made.

Nonuniform degradation rates. In (6), we made all the degradation rates equal (with value 1 by a scaling), but in general, they will differ:

$$\frac{dx_i}{dt} = -\gamma_i x_i + f_i(X_{i_1}(t), X_{i_2}(t), \ldots, X_{i_k}(t)) . \tag{10}$$

There is no real difficulty in this case with computing solution trajectories from threshold crossing to threshold crossing as before. Trajectories are no longer straight line segments between transitions but are still directed toward a focal point, monotonically in each coordinate. However, the mappings are more complicated and some of the convenient properties of the fractional linear maps of the equal degradation rate case are lost. This makes it more difficult to establish, for example, the existence and stability of periodic orbits, but there are classes of systems for which this can still be done [**8**].

Multiple thresholds. In general, if a particular transcription factor is effective in regulating the expression of more than one gene, it can do this at different threshold concentrations. Thus, enough of a transcription factor may be present to activate (or repress) the expression of one gene but not enough to activate (or repress) the expression of another. In this situation, the phase space of the system is not simply divided in each coordinate into two regions (above and below threshold) but into multiple regions, and the entire phase space is partitioned into "boxes." Nevertheless, the same procedure allows computation of trajectories from threshold crossing to threshold crossing. In fact, it has been shown that the multiple-threshold model, as a class, is really no more difficult than the single-threshold-per-gene model, in that any example of the former class can be reformulated as an example of the latter class, albeit with more variables. The resulting network, though higher dimensional, is inevitably attracted to a lower-dimensional hyperplane equivalent to the phase space of the original network and on which the dynamics are identical to the original [**6**].

Autoregulation, self-input and "singular" domains. If we allow for autoregulation, the possibility that the protein expressed by a gene acts as a transcription factor for its own transcription, then the functions f_i can depend on X_i and we have self-input. This makes it possible for the flow on both sides of a threshold to be directed toward the threshold, or similarly, away from the threshold on both sides. Thus, it becomes possible for the flow to be constrained to such "singular domains" until perhaps another switching event releases it again. It is also possible, when there is self-input, for a pair of genes cycling on and off to converge to their threshold intersection in finite time (after infinitely many switchings). The subsequent flow is then constrained to the region where both these variables are at their threshold values.

Special techniques are required to calculate solution trajectories in these threshold regions because the flow of the differential equation (6) or (10) is not well defined there. Two approaches have been suggested, one using methods from control theory for differential equations with discontinuous vector fields (Filippov methods), the other using singular perturbation methods. There are especially challenging situations in which the former method at least leads to nonuniqueness of solutions so that quite different subsequent trajectories from a particular state are consistent with the equations. Some of these ambiguities can be removed by careful analysis, but some are present because of an intrinsic sensitivity to initial conditions that is revealed by the singular perturbation but masked as nonuniqueness in the step function limit [**21**] .

Most of these mathematically tricky situations are avoided under another assumption: that whenever a transcription factor regulates more than one other gene, it does

so at distinct threshold values. This is the opposite extreme from our initial assumption that there was only one threshold for a given transcription factor, no matter how many other genes it regulated.

Multistep regulation. In a sense, the difficulties associated with flows in threshold domains are artificial in that gene regulation is not in reality a direct action of one transcription factor on another, as implicitly built into the design of the models discussed above. In fact, gene expression consists in the production (transcription) of messenger mRNA molecules that reflect the sequence of nucleotides in the gene. These subsequently control the generation of the gene's specific protein molecule by putting together amino acids in a sequence determined by the mRNA's sequence (this process is called translation). If transcription and translation (and possibly post-translational modifications) are modelled as separate steps, then there is no autoregulation for any variable of the system. However, some of the steps (such as translation) involve linear terms not sigmoidal interactions, so the framework above is not quite sufficient. Models explicitly incorporating multistep regulation are now being studied and, in some cases, predict different behavior than the corresponding single-step models [**10**].

7. CONCLUSIONS. New technologies are generating massive data concerning the structure and regulation of genetic expression in living organisms [**11**]. Because of the huge data sets, it is essential that tools from mathematics and computation be employed to help analyze and interpret the data. In this article, we have presented an introduction to what we believe is the current best candidate method for understanding these networks. The techniques we have described provide a basis for relating the qualitative structure and interactions of biological control networks, with qualitative features of the dynamics. The mathematical techniques utilize standard methods in nonlinear dynamics but also begin to address new and still poorly understood questions involving dynamical systems in huge dimensions and in equations with discontinuous right-hand sides. These mathematical questions are of interest themselves, independent of applications.

We are just at the beginning of understanding mechanisms of genetic control. Developing better methods to understand how these networks function and how they have evolved promises to be a major future intellectual challenge.

ACKNOWLEDGMENT. We thank NSERC for support of our research with grant number 217148.

REFERENCES

1. R. Alberts, J. J. Collins, L. Glass, Introduction to focus issue: Quantitative approaches to genetic networks, *Chaos* **23** (2013) 025001, http://dx.doi.org/10.1063/1.4810923.
2. A. Berman, R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences.* Academic Press, New York, 1994.
3. J. L. Cherry, A. R. Adler, How to make a biological switch, *J. Theoret. Biol.* **203** (2000) 117–133, http://dx.doi.org/10.1006/jtbi.2000.1068.
4. R. Edwards, Analysis of continuous-time switching networks, *Phys. D* **146** (2000) 165–199, http://dx.doi.org/10.1016/s0167-2789(00)00130-5.
5. ———, Chaos in neural and gene networks with hard switching, *Differ. Equ. Dyn. Syst.* **9** (2001) 187–220.
6. R. Edwards, E. Farcot, E. Foxall, Explicit construction of chaotic attractors in Glass networks, *Chaos Solitons Fractals* **45** (2012) 666–680, http://dx.doi.org/10.1016/j.chaos.2012.02.018.
7. M. B. Elowitz, S. Leibler, A synthetic oscillatory network of transcriptional regulators, *Nature* **403** (2000) 335–338.
8. E. Farcot, J.-L. Gouzé, Limit cycles in piecewise-affine gene network models with multiple interaction loops, *Internat. J. Systems Sci.* **41** (2010) 119–130, http://dx.doi.org/10.1080/00207720903144552.

9. T. S. Gardner, C. R. Cantor, J. J. Collins, Construction of a genetic toggle switch in *Escherichia coli*, *Nature* **403** (2000) 339–342.
10. T. Gedeon, G. Cummins, J. J. Heys, Effect of model selection on prediction of periodic behavior in gene regulatory networks, *Bull. Math. Biol.* **74** (2012) 1706–1726, http://dx.doi.org/10.1007/s11538-012-9732-2.
11. M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, and others, Architecture of the human regulatory network derived from ENCODE data, *Nature* **489** (2012) 91–100.
12. L. Glass, Classification of biological networks by their qualitative dynamics, *J. Theoret. Biol.* **54** (1975) 85–107.
13. L. Glass, S. A. Kauffman, The logical analysis of continuous, non-linear biochemical control networks, *J. Theoret. Biol.* **39** (1973) 103–129.
14. L. Glass, M. C. Mackey, *From Clocks to Chaos: The Rhythms of Life.* Princeton Univ. Press, Princeton, NJ, 1988.
15. L. Glass, J. S. Pasternack, Stable oscillations in mathematical models of biological control systems, *J. Math. Biol.* **6** (1978) 207–223.
16. D. Kaplan, L. Glass, *Understanding Nonlinear Dynamics.* Springer, New York, 1995.
17. S. A. Kauffman, Metabolic stability and epigenesis in randomly constructed genetic nets, *J. Theoret. Biol.* **22** (1969) 437–467.
18. J. W. Kotula, S. J. Kerns, L. A. Shaket, L. Siraj, J. J. Collins, J. C. Way, P. A. Silver, Programmable bacteria detect and record an environmental signal in the mammalian gut, *PNAS* (2014) 1321321111, http://dx.doi.org/10.1073/pnas.1321321111.
19. Q. Li, X.-S. Yang, Chaotic dynamics in a class of three dimensional Glass networks, *Chaos* **16** (2006) 033101.
20. L. Lu, R. Edwards, Structural principles for complex dynamics in Glass networks, *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **21** (2011) 237–254, http://dx.doi.org/10.1142/s0218127411028398.
21. A. Machina, R. Edwards, P. van den Driessche, Sensitive dependence on initial conditions in gene networks, *Chaos* **23** (2013) 025101, http://dx.doi.org/10.1063/1.4807480.
22. T. Mestl, C. Lemay, L. Glass, Chaos in high-dimensional neural and gene networks, *Phys. D* **98** (1996) 33–52.
23. J. Monod, F. Jacob, General conclusions: Teleonomic mechanisms in cellular metabolism, growth, and differentiation, *Cold Spring Harb. Symp. Quant. Biol.* **26** (1961) 389–401.
24. D. Thieffry, R. Thomas, Dynamical behaviour of biological regulatory networks. II. Immunity control in bacteriophage lambda, *Bull. Math. Biol.* **57** (1995) 277–297.
25. R. Thomas, Boolean formalization of genetic control circuits, *J. Theoret. Biol.* **42** (1973) 563–585.
26. ———, On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations, in *Numerical Methods in the Study of Critical Phenomena,* Edited by J. Della Dora, J. Demongeot and B. Lacolle, Springer, New York, 1981. 180–193.

RODERICK EDWARDS obtained an M.Sc. degree in applied mathematics from Heriot-Watt University in 1990 and a Ph.D. degree from the University of Victoria in 1994. He was a postdoctoral fellow in the Départment de Kinanthropologie at Université du Québec á Montréal from 1995 to 1998, after which he returned to the University of Victoria, where he is now a professor in the Department of Mathematics and Statistics. His research has involved the analysis of network dynamics arising in biology, including neuronal, genetic, and disease networks, as well as analysis of physiological time series and applications of mathematics in other areas such as applied linguistics.
Department of Mathematics and Statistics, University of Victoria, P.O.Box 1700, STN CSC, Victoria, BC, Canada V8W 2Y2
edwards@uvic.ca

LEON GLASS graduated from Brooklyn College in 1963. He was a graduate student in chemistry at the University of Chicago. He completed postdoctoral fellowships in the Department of Machine Intelligence and Perception, Edinburgh; the Department of Theoretical Biology, Chicago; and the Department of Physics and Astronomy, Rochester. He has been at McGill University in Montreal since 1975 where he is now a professor of physiology and the Isadore Rosenfeld Chair in Cardiology. His research is on mathematical analysis of complex dynamics in biology with focus on cardiac arrhythmias, genetic control, and visual perception.
Department of Physiology, McIntyre Medical Sciences Building, McGill University, 3655 Promenade Sir William Osler, Montréal, Québec, Canada H3G 1Y6
leon.glass@mcgill.ca