Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT

Signal Processing 83 (2003) 789–798

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Dynamics in high-dimensional model gene networks

K. Kappler[a], R. Edwards[a,b,*], L. Glass[b]

[a] *Department of Mathematics and Statistics, University of Victoria, P.O. Box 3045, STN CSC, Victoria, BC, Canada V8W 3P4*
[b] *Centre for Nonlinear Dynamics in Physiology and Medicine, Department of Physiology, McIntyre Medical Sciences Building, McGill University, 3655 Drummond St., Montréal, Qué., Canada H3G 1Y6*

Received 18 June 2002; received in revised form 3 September 2002

Abstract

We consider dynamics in a class of ordinary differential equations representing a simplified model of a genetic network. In this network, the model genes control the production rates of products of other genes by a logical function. Thus, the concentration of a particular gene's product is driven either up or down depending on the particular combination of activity states (active or inactive, i.e., gene product above or below threshold) of a set of relevant 'input' genes. The interactions are based on binary functions but the evolution of gene product concentrations is continuous in time, unlike the discrete-time Boolean networks of Kauffman and others. Numerical methods allow rapid and accurate integration of the model equations. Also, theory now allows analytic confirmation of numerically observed attractors. This enables us to determine changes in the distribution of dynamical properties (attractor types) of random networks as the size of the networks increase and as the number of inputs to each gene increases.

Keywords: Gene networks; Gene regulation; Qualitative dynamics; Limit cycles; Chaos

1. Introduction

A gene is a sequence of DNA bases that codes for a sequence of amino acids that constitute a protein. According to current estimates, there are more than 30,000 different genes in human beings. Although it is an idealization, it is common for molecular biologists to think of genes as being turned "on" or "off". If a gene is "on", the associated protein would be synthesized (through a sequence of intermediate steps that are themselves subject to various controls), whereas if the gene is "off" the associated protein would not be synthesized. In any particular cell of the body, only a subset of these genes will be expressed at any one time. The expression of the genes in a cell is controlled by various diffusible factors, called transcription factors. A transcription factor is a protein whose sequence is in turn coded by a gene and whose expression is controlled by transcription factors. Thus, since the proteins in a cell play a major role in determining the cell properties, the differential gene expression in different cells leads to differences in structure and function in different cell types.

Various attempts have been made to develop mathematical models of genetic control networks.

* Corresponding author. Department of Mathematics and Statistics, University of Victoria, P.O. Box 3045, STN CSC, Victoria, BC, Canada V8W 3P4.
E-mail addresses: kappler@uvic.ca (K. Kappler), edwards@math.uvic.ca (R. Edwards), glass@cnd.mcgill.ca (L. Glass).

0165-1684/03/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0165-1684(02)00479-6

Kauffman proposed that genetic networks could be modeled by random Boolean networks in which time is discrete and each element computes a Boolean function based on the values of inputs to that element [22,23]. The Boolean formalization was developed further by Thomas and collaborators [32,33]. However, since gene networks do not act in discrete time and gene product concentrations are continuous variables, we believe that the discrete networks above, or even asynchronous versions of them, are less suitable as model gene networks than ordinary differential equations in which gene interactions are incorporated as logical functions [10,12,18–21,28,29]. Differential equations and logical networks have been proposed to model a variety of different specific gene networks [25–27,30,31].

Kauffman and others have studied the properties of discrete Boolean switching networks in discrete time as the size of the network increases [1,22]. Calling N the number of model genes, for networks with $K = 2$ inputs, there is generally a well-behaved dynamics in which there are comparatively few (Kauffman suggests $\approx \sqrt{N}$) attractors, and the average length of attractors also scales as $\approx \sqrt{N}$. However, more recent work shows that there are difficulties in counting the number of attractors, and suggest strongly that it actually grows linearly with N [1,6]. Discrete switching networks, such as those proposed by Kauffman, have discrete state spaces and so must necessarily always cycle. The possible cycle lengths can potentially be of the order 2^N; thus, the finding that the cycle lengths are much shorter than the longest possible lengths has attracted a good deal of theoretical interest [2–5,7–9,15,24,34]. Discrete switching networks with $K = 1$ inputs are more likely to reach fixed points and isolated very short cycles than $K = 2$ networks: they are more likely to become 'frozen'. Networks with $K > 2$ have a larger probability of very long cycles (increasing with K). Although the discrete networks with very long cycle lengths are sometimes called "chaotic", this terminology is at odds with the standard definition in nonlinear dynamics of "chaos" which demands an aperiodic dynamics and must necessarily arise in infinite state spaces. However, in the continuous differential equations that capture the logical structure of the switching networks, chaotic dynamics is possible in four and more dimensions [10,11,28,29].

The point of the current communication is to explore the dynamics in differential equation models of genetic networks as the number of elements in the network increases, and for various numbers of inputs.

2. A differential equation

In this section we briefly present a mathematical model of gene networks. Since many aspects of the model have recently been reviewed [10,12], we refer the reader to these earlier publications for further mathematical details. Our exposition follows closely an earlier publication [13].

A Boolean switching network with N elements is represented by

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

where $F_i(X_{i_1}(j), X_{i_2}(j), \ldots, X_{i_K}(j)) \in \{0, 1\}$ and K is the number of inputs. This is a discrete time and discrete state space system which, therefore, must eventually reach a fixed point or cycle under iteration.

Since biological systems are not believed to have clocking devices that simultaneously update the network, a differential equation would be a more suitable class of mathematical model. The logical structure of Eq. (1) can be captured by a differential equation [17,19,21]. To a continuous variable $x_i(t)$, we associate a discrete variable $X_i(t)$:

$$X_i(t) = 0 \text{ if } x_i(t) < 0, \quad \text{otherwise } X_i(t) = 1. \tag{2}$$

For any logical network, we define an analogous differential equation

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_i + f_i(X_{i_1}(t), X_{i_2}(t), \ldots, X_{i_K}(t)), \quad i = 1, \ldots, N, \tag{3}$$

where $f_i(X_{i_1}(t), X_{i_2}(t), \ldots, X_{i_K}(t))$ is a scalar whose sign is negative (positive) if the corresponding logical variable $F_i(X_{i1}(t), X_{i2}(t), \ldots, X_{i_K}(t))$ is 0 (1).

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 0. The

solution of Eq. (3) for each variable x_i for $t_j < t < t_{j+1}$ is

$$x_i(t) = x_i(t_j)\,\mathrm{e}^{-(t-t_j)} + f_i(X_{i_1}(t_{j^*}), X_{i_2}(t_{j^*}), \ldots, X_{i_K}(t_{j^*}))(1 - \mathrm{e}^{-(t-t_j)}), \quad (4)$$

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 towards a focal point. If the focal point lies in a different orthant from the initial condition, 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.

Even though Eq. (3) is more realistic than Eq. (1) as a model for biological systems, this equation still is highly oversimplified. Yet, this equation has remarkable mathematical properties that facilitate theoretical analysis. Moreover, there is an expectation, demonstrated in some simple examples, that the qualitative dynamics in the model system will be preserved in more realistic versions, for example, when the discontinuous step functions are replaced by continuous sigmoidal functions [20]. Also, synthetic gene networks have been created that show some of the simple types of dynamical behaviour found in our class of networks—in particular, bistability (two fixed points) [16] and oscillation in an inhibitory loop [14].

In the differential equation with no self-input, in general, only one variable will cross its threshold at a given time. Therefore, the dynamics in the differential equation can be mapped onto an N-cube where directed edges represent allowed transitions between logical states. The allowed transitions are also equivalent to the allowed transitions in an asynchronous switching network with the same logical structure [17,21].

3. Numerical integration and types of attractors

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 Eq. (4). 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 $\tau_{\min} = \min\{\tau_i\}$. Once again, using Eq. (4), the equations can be analytically integrated for a time interval $\tau_{\min}$, and then the process is iterated.

These equations can display several qualititatively different types of dynamics, including fixed points (nodes or spiral points), limit cycles, chaos, and quasiperiodicity.

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. Stable spiral points 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 non-zero) value. Thus, spiral nodes typically have an associated cyclic sequence of orthants through which an approaching trajectory passes.

Limit cycles are closed trajectories towards which at least some nearby trajectories converge and they of course have an associated cyclic sequence of orthants, through which the limiting trajectory passes. When the focal points have coordinates that are all ± 1, as in the model we consider here, it is possible for a limit cycle to involve simultaneously switching variables. An approaching trajectory will have a well-defined cyclic sequence of orthants with one variable switching at a time, but in the limit as $t \to \infty$ two or more switchings will approach simultaneity. This is still clearly a limit cycle, even though our network definition may leave the subsequent trajectory from an exact simultaneous switching ambiguous.

Chaotic dynamics has been demonstrated analytically in some example networks [11,28]. Here, where we have to rely on numerical identification of fixed points or limit cycles, we will classify any network in which we find none of these simple attractors as ‘chaotic’. While we expect most of these to be truly chaotic as in the examples cited above, it is possible that we could occasionally confuse very long transients or quasiperiodicity with chaotic behaviour. Quasiperiodicity can occur when there are cycles independent of each other and with non-commensurate periods. This type of behaviour we do not anticipate to be common.

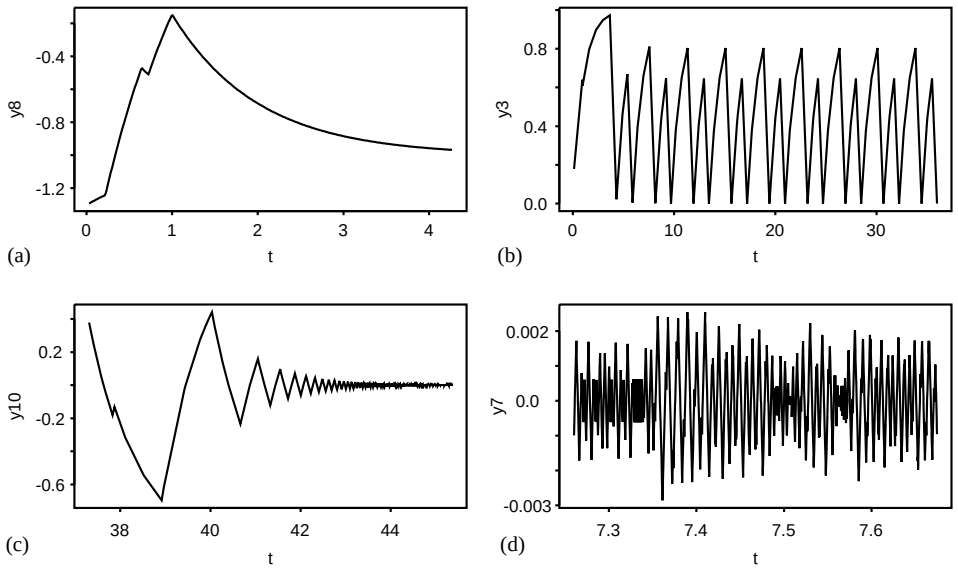

Fig. 1. Time series showing the four classes of behaviour: node, spiral point, limit cycle, chaos in continuous-time *NK* networks: (a) $N = 10$, $K = 3$; (b) $N = 8$, $K = 2$; (c) $N = 10$, $K = 3$; and (d) $N = 8$, $K = 2$.

Numerical integration can give strong evidence for the existence of an attractor of a particular type in a network. Fig. 1 shows the simulated time evolution of one variable in each of four networks, typifying the four classes of behaviour.

However, there are also analytic methods for confirming the existence of attractors in this class of networks. Nodes found by integration need no further verification: if we find a focal point lying in its own orthant (i.e., with the same sign structure), then that focal point is a stable node. It is also possible to prove the existence and stability of limit cycles analytically [10,21,29]. The return map for a given cycle of orthants starting on a particular orthant boundary crossing (Poincaré section) can be calculated explicitly and has the form of a fractional linear map:

$$M(\mathbf{y}) = \frac{A\mathbf{y}}{1 + \phi^t \mathbf{y}},$$

where $\mathbf{y} \in \mathbf{R}^{(n-1)}$ is a point on the initial orthant boundary (in an n-dimensional network), A is an $(n-1) \times (n-1)$ matrix, and $\phi \in \mathbf{R}^{(n-1)}$. This mapping is defined on a region (called a 'returning cone') which may be only part of the orthant boundary, from which trajectories follow the cycle of orthants under consideration. Letting λ_1 and $\mathbf{v}_1$ be the dominant eigenvalue of A and its corresponding eigenvector, stable periodic orbits through a given cycle of orthants correspond exactly to the following conditions being met:

1. $\lambda_1 > 1$ (and real),
2. $\mathbf{v}_1$ lies in the returning cone (or on its boundary).

If $\lambda_1 = |\lambda_2|$ then stability is only neutral. If $\mathbf{v}_1$ lies on the boundary of the returning cone, then the cycle is degenerate in the sense that somewhere around the cycle there occurs a simultaneous switching of two or more variables in the limit.

Thus, if numerical integration suggests convergence to a limit cycle, this can be confirmed or disproved by selecting any orthant boundary crossing along that cycle and calculating the return map and the returning cone, as well as the eigenvalues and eigenvectors of the map. If the above conditions are met, then it is indeed a stable limit cycle. If not then integration should be continued as the trajectory must eventually deviate from the purported cycle. Of course, these calculations to verify the existence of a stable limit cycle must also generally be done by computer and so accuracy must be considered. In the class of networks considered here, with all focal point coordinates at ± 1,

the matrix A and the vector ϕ in the return map have integer components and $\det(A) = 1$.

If a repeating sequence of orthants is checked by the above procedure and the dominant eigenvalue, λ_1, is exactly 1, then the cycle corresponds to a spiral point [10]. There must still be an eigenvector in the returning cone, but $\lambda_1 = 1$ and $\det(A) = 1$ imply $|\lambda_j| = 1$ for all j, so it will often happen that λ_1 has an eigenspace of dimension greater than 1, which makes it more difficult in practice to check whether the eigenspace intersects the returning cone.

4. Survey of random networks

The number of different structural classes of these networks (Eq. (3)), taking symmetries into account, grows extremely rapidly with N [12]. For $N=4$, $K=3$, there are already over 11 million classes. The $N = 4$, $K=2$ networks are clearly a subset of the $N=4$, $K=3$ networks since some choices of the Boolean functions do not actually depend on all their inputs.

In order to assess the effects on dynamics of different values of N, the number of genes, and K, the number of inputs per gene, we generated many networks randomly and integrated them from random initial conditions. Initially, one of us (Glass) conducted such a survey of $K = 2$ networks basing the results only on the output of the integrations. Then we (Edwards and Kappler) conducted another survey, covering fewer values of N but more values of K, using independently written code and employing the analytic theory to confirm the results of the integrations. Both approaches are non-trivial to carry out in practice. We point out some of the potential pitfalls below.

The numerics-only approach has the advantage of simplicity and speed, but conclusions (except for nodes) are inevitably based on appearances and not on proof. For example, it is possible in principle for a cycle of orthants to be followed many times, but not correspond to a stable periodic orbit. This could happen if the dominant eigenvector for the return map lay very close to but outside the returning cone. Then, trajectories could come back to the returning cone repeatedly but eventually must fall outside and go off on some other path. In practice, conclusions based on the numerical integration are probably fairly reliable.

The numerics-plus-analysis approach has the advantage of analytic confirmation of existence of limit cycles and spiral points, but slows the process down, particularly because precision of the calculations required are critical, especially for long cycles. Even in small networks ($N = 4$), it has been shown that stable cycles involving hundreds of switchings exist, though rare [12], and the matrix A in the return map for such a cycle has very large integer entries. Even small integer matrices for shorter cycles can have high condition number and therefore make accurate eigenvalue and eigenvector calculations difficult.

The procedure for random generation of networks requires some comment. We take a straightforward approach, essentially that used by Kauffman [23, p. 192]. First, for each gene we randomly choose the K genes that provide its inputs from amongst the other $N - 1$ genes, each possible set being equally likely. We do not allow self-input. Then we randomly assign a Boolean function of K inputs to each gene so that each set of binary values is equally likely. One implication of this procedure is that some network structures are selected more often than others. Consider the task of selecting a network with $N = 3$ and $K = 1$. For each gene, there is a choice of two other genes for an input. If the Boolean function we end up selecting is the one that does not depend on the input (for example, if the function value is 0, whether the input is 0 or 1), then which gene was selected as the input gene makes no difference to the structure of the network. In essence, the *effective* number of inputs to a given gene is less than or equal to K, but if less than K then more than one choice of input genes gives the same network. While it would be possible to design a sampling procedure to select each network *structure* with equal likelihood, it is not clear whether this is a more appropriate procedure for modelling random gene networks, and the straightforward approach has the added advantage of allowing comparison with work on discrete-time Boolean NK networks. However, the implication of the selection procedure should be noted, and can be significant. For example, it is possible to calculate explicitly the probabilities of finding nodes, spiral points and limit cycles in $N = 3$, $K = 1$ networks by combinatorial arguments. These calculations are relatively easy in this case because coexistence of two types of attractor in one network is not possible and only two nodes or two spiral points are

possible. Using our straightforward random selection procedure,

$$Pr(\text{node}) = \frac{456}{512} = 0.890625,$$

$$Pr(\text{spiral point}) = \frac{48}{512} = 0.09375,$$

$$Pr(\text{limit cycle}) = \frac{8}{512} = 0.015625,$$

whereas, if we select each network structure with equal likelihood,

$$Pr(\text{node}) = \frac{172}{216} = 0.7962963,$$

$$Pr(\text{spiral point}) = \frac{36}{216} = 0.1666667,$$

$$Pr(\text{limit cycle}) = \frac{8}{216} = 0.037037.$$

The former set of values can be used as a check on our numerical results below. We also know that for $N = 3$, $K = 2$ networks, there is only one type of limit cycle that can occur and it occurs in 8 of the 4096 equally likely networks, so that $Pr(\text{limit cycle}) = 0.001953125$. While it is hard to calculate exact probabilities for general N and K even for nodes, it is clear that the probability of a random network having *no* nodes increases with K. For $K = 0$, this probability is 0; for $K = N$, this probability is $(1 - 2^{-N})^{(2^N)}$, which rapidly converges to e^{-1} as $N \to \infty$. Thus, the probability that a random NK network has at least one node approaches a number greater than or equal to $1 - e^{-1} \approx 0.632$ as N increases, the specific number depending on K. The presence of nodes in a network does not, of course, guarantee that all trajectories converge to one of them, but it is clearly possible for large N that they may be reached after very long transients.

For the first survey, numerical integrations were carried out for 9000 transition steps (switchings) per network, and the sequence of switching variables was searched for periodicities (usually after a transient) of length up to 500 transition steps. Nodes are found when there are no more transitions, as described above. Network behaviour was classed as 'spiral point' if there were no periodic orbits, and some subset of the variables rapidly cycled around 0. Network behaviour was classed as 'limit cycle' if there was a periodic oscillation of some subset of the variables with the other variables approaching a fixed value. Network behaviour was classed as 'chaotic' if they did not fit into any of the other classes.

This classification is subject to several inaccuracies, aside from numerical error and the possibility of deceptive transients as mentioned above. In particular, networks in which there are uncoupled subsets of genes (i.e., subsets of genes with no inputs from one set to the other or vice versa) behaving differently can cause classification difficulties. This is not generally expected to be common, but may occur especially in networks with $K = 1$, as is the case for discrete-time Boolean networks [23, p. 203]. If one subset of genes considered as a smaller independent network converges to a spiral point, while another subset not coupled to the first converges to a limit cycle, then our procedure would classify it as a spiral point. From a mathematical point of view, such a network considered as a whole converges to a limit cycle on which some of the variables are simply fixed. If two or more subsets converge to limit cycles with periods that are not commensurate (i.e., are irrationally related), then the network behaviour as a whole is quasiperiodic, but will be classified as chaotic here. This appears to be very rare even though it is clearly possible. Other combinations of behaviours in uncoupled subsets of genes will be classified 'correctly' in the sense of the attractor type of the network as a whole. A chaotic subset and a periodic subset will be classified as chaos, for example.

The results of this survey are shown in Table 1 and Fig. 2.

For the second survey the numerical integration was carried out by a C program which generated random Boolean function values for each orthant encountered (rather than generating an entire network structure since state space is enormous for large N). Each network was first integrated for 3000 transition steps. If no attractor was identified at that point, integration was continued for another 30,000 transition steps. There were two conditions under which integration was stopped before the full set of 3000 or 30,000 steps was complete. If a stable node (focal point) was reached, no further switchings could occur. Secondly, if the time between subsequent transitions dropped below 10^{-15}, integration was stopped to avoid problems with roundoff error near simultaneous switchings. In

Table 1
Numerical survey of dynamics of $K = 2$ networks

N	Node	Spiral	Cycle	Chaos
3	7272	2710	18	0
4	6909	2583	489	19
5	6859	2430	684	27
6	7035	2440	522	3
7	6988	2434	575	3
8	334	135	30	1
9	333	120	47	0
10	332	127	41	0
15	334	102	61	3
20	321	123	56	0
25	291	120	84	5
30	302	108	83	7
35	299	101	97	3
40	280	118	95	7
50	284	104	108	4
60	261	122	108	9
70	269	109	106	16
80	257	107	117	19
90	252	114	119	15
100	264	101	115	20
125	251	112	108	29
150	235	114	131	20
175	246	91	120	43
200	240	109	111	40

Numbers of attractors found in randomly generated networks with random initial conditions. In dimensions 3–7 the simulations were carried out in 10,000 different networks and in higher dimensions the simulations were carried out in 500 networks.

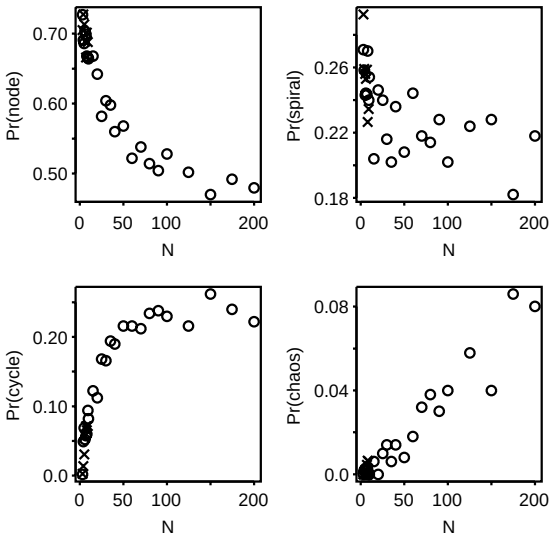

Fig. 2. Size of network, N, versus probability of each type of attractor on $K=2$ networks. Circles indicate data from the numerics-only survey; crosses indicate data from the numerical/analytic survey.

all such cases, a limit cycle or a spiral point was identified, so no misclassification occurred as a result of stopping the integration.

After each integration run, an attempt was made to classify the network behaviour by the following procedure. First, the sequence of switching variables was searched for periodic subsequences (after a transient) in which every variable switched an even number of times (so that it ended up in the same orthant it started in). If a subsequence had repeated over at least 3000 steps or the last half of the integration run, whichever was less, and if these repetitions occurred at least three times (so that cycles of length up to 1000 might be found), it was identified as a possible cycle.

If no possible cycle was found and the integrations had been stopped at a stable node, then the classification was complete. If no possible cycle was found after the full 3000 steps, integration was continued. If no possible cycle was found after the full 30,000 steps in the second integration run, the network behaviour was classed as 'chaos'.

If a possible cycle was found, it was passed to an analysis routine to check the conditions noted above for existence of stable periodic orbits or spiral points. This analysis was carried out mostly in Maple, where eigenvalues and eigenvectors can be calculated with high precision (we used 30 digits, which gave eigenvalues and eigenvectors of sufficient accuracy), and where the large integers in the mapping matrix, A, for long cycles could be handled reliably. If the dominant eigenvalue of A was greater than 1 and the corresponding eigenvector lay in the returning cone for the cycle map, then the network behaviour was classed as 'limit cycle'. If the dominant eigenvalue was exactly 1 (identified with a small tolerance in Maple), the network behaviour was classed as 'spiral point' (and no check on the returning cone was made). In any other case (such as the dominant eigenvector not lying in the returning cone, in which case the purported limit cycle does not exist), integration was continued, or if the second run of 30,000 steps had been done, the network behaviour was classed as 'chaos'.

Finally, as a check on the problem of transients, we repeated some of the simulations allowing the number of integration steps to go to 300,000. The results were very close to those allowing only 30,000 steps and are not reported here.

While some of the potential problems of the numerics-only approach are avoided here and we have more confidence in the distributions of attractors found, there are still potential inaccuracies. Quasiperiodic attractors will still be classed as chaotic. Other combinations of uncoupled subnetworks could also be misclassified as chaotic, such as a limit cycle and a spiral or two spirals with irregularly interspersed switchings. If a spiral point involved a complicated sequence of switchings, we relied on the numerical repetition to assert that it would continue indefinitely, rather than showing that an eigenvector lay in the returning cone as we did for limit cycles. For the vast majority of spiralling cycles, this should be sufficient because if the cycle was eventually to be lost, a different variable would eventually have to switch but would be very unlikely to wait longer than our integration time.

The results of this survey are shown in Table 2. The numbers of occurrences of each behaviour type are out of 2000 network simulations. Fig. 3 gives one graphical view of these results, showing the dependence of probabilities of each type of attractor on K for each value of $N \leqslant 9$.

Although there are some differences in the results from the two surveys, which presumably reflect a small percentage of misclassifications, the overall trend is similar for the $K = 2$ networks, which helps to verify the results of the numerics-only method. The crosses in Fig. 2 indicate results from this survey. For $K = 2$ networks, the probabilities of stable nodes decreased exponentially with increasing N, perhaps approaching a non-zero value. The same is probably true for spiral points, though there appeared to be more variability in these. The probability of limit cycles increased in what looks like exponential approach to some maximum value (from the available evidence, this seems to be about $\frac{1}{4}$ but larger N would be required to draw a firm conclusion).

The effect of increasing K was to decrease the probability of finding nodes but to increase the probability of spiral points, at least for small N. The probability of finding a cycle peaked at $K = 2$ (or 3 for $N = 9$) decreasing for larger K. The probability for $K = 1$ was less dependent on N and for $N \geqslant 5$ there was an increase in the probability of finding a cycle between $K = 1$ and 2. Thus, in general $K = 2$ networks seemed to favour limit cycles more than any other K. The

Table 2
Numerical/analytic survey of dynamics of NK networks

N	K	Node	Spiral	Cycle	Chaos
3	1	1776	188	36	0
3	2	1410	585	5	0
4	1	1784	171	43	2
4	2	1455	518	27	0
4	3	1103	876	19	2
5	1	1795	152	49	4
5	2	1426	512	61	1
5	3	1118	837	42	2
5	4	869	1094	34	3
6	1	1818	141	39	2
6	2	1377	506	115	2
6	3	1063	856	77	4
6	4	833	1113	46	8
6	5	651	1299	42	8
7	1	1789	150	58	3
7	2	1332	517	142	9
7	3	1074	822	97	7
7	4	812	1114	69	5
7	5	605	1332	51	12
7	6	493	1475	25	7
8	1	1803	136	58	3
8	2	1403	453	131	13
8	3	1034	828	119	19
8	4	752	1143	98	7
8	5	583	1349	58	10
8	6	497	1462	35	6
8	7	419	1551	24	6
9	1	1806	142	49	3
9	2	1376	469	143	12
9	3	967	850	156	27
9	4	728	1149	98	25
9	5	555	1365	65	15
9	6	415	1536	39	10
9	7	392	1585	18	5
9	8	340	1645	9	6

Numbers of attractors found in randomly generated networks with random initial conditions. Simulations were carried out in 2000 different networks for each combination of N and K.

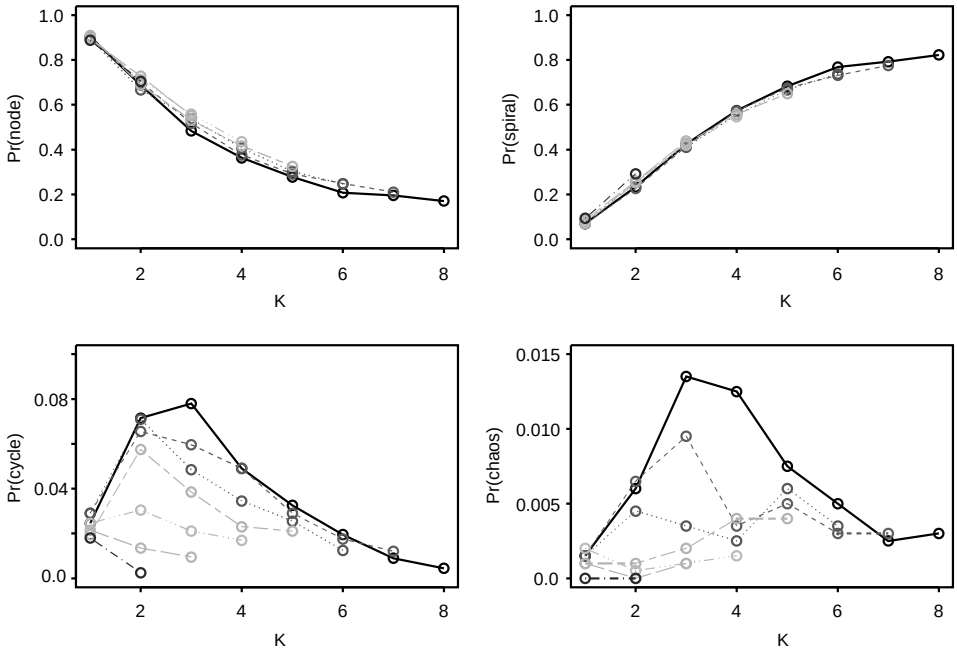

Fig. 3. Number of inputs, K, versus probability of each type of attractor for small NK networks ($N = 3$–9). Each value of N has a different line type which can be determined by observing that the maximum value of K simulated was $N - 1$ in each case.

probability of finding chaos followed a less clear pattern as K varied. In our simulations it increased with $K \geqslant 2$ in the cases $N = 4, 5, 6$ but for $N = 8$ and 9, it appeared to reach a maximum at $K = 3$ and decrease for larger K. From another point of view, we find that for fixed $K > 2$, the number of networks reaching nodes generally decreases with increasing N and the number of chaotic networks generally increases.

5. Conclusions

The extent to which the methods here are applicable to real genetic systems is not known. The current equations are not realistic for many reasons: (i) control of gene expression is not on or off but is graded; (ii) there may be time delays associated with synthesis or degradation of gene products not accounted for here; (iii) decay rates of different gene products are different; (iv) although a single gene product might control expression of many different genes, the threshold levels for activation and/or inhibition may be different for different targets. However, all these represent quantitative changes in the equations, and the extent to which these changes influence the qualitative properties of the equations is still largely unknown. Further, it is intriguing that a recent analysis of gene expression in sea urchin concludes that the gene control can be approximated by a logical function based on the presence or absence of relevant factors that control the gene expression [35].

Our work shows a number of consistent patterns in probabilities of reaching different attractor types as the size of the network, N, and the number of inputs per gene, K, are changed. Increasing N for fixed K generally decreases the probability of reaching a fixed point, but increases the probability of reaching a cycle or a chaotic attractor (at least for $K = 2$). Increasing K for fixed N decreases the probability of reaching a node, but correspondingly increases the probability of converging to a spiral point, at least for small N, while the probabilities of cycles or chaos peaks at about $K = 2$ or 3. It would be interesting to extend these results to the case of larger N and $K > 2$, where some preliminary tests we have conducted suggest that chaos becomes very likely as K increases.

Acknowledgements

This research has been supported by grants from NSERC, and the National Research Resource for Complex Physiologic Signals. Leon Glass thanks Dr. J. Collins, Boston University, and Dr. A. Goldberger, Beth Israel Deaconess Hospital, for hospitality during the 2001–2002 academic year, when some of the research described here was carried out.

References

[1] R.J. Bagley, L. Glass, Counting and classifying attractors in high dimensional dynamical systems, J. Theoret. Biol. 183 (1996) 269–284.
[2] U. Bastolla, G. Parisi, Closing probabilities in the Kauffman model: an annealed computation, Physica D 98 (1996) 1–25.
[3] U. Bastolla, G. Parisi, A numerical study of the critical line of Kauffmann networks, J. Theoret. Biol. 187 (1997) 117–133.
[4] A. Bhattacharjya, S. Liang, Power-law distributions in some random Boolean networks, Phys. Rev. Lett. 77 (1996) 1644–1647.
[5] A. Bhattacharjya, S. Liang, Median attractor and transients in random Boolean nets, Physica D 95 (1997) 29–34.
[6] S. Bilke, F. Sjunnesson, Stability of the Kauffman model, Phys. Rev. E 65 (2001) 016129.
[7] B. Derrida, Valleys and overlaps in Kauffman's model, Philos. Mag. B56 (1987) 917–923.
[8] B. Derrida, Y. Pomeau, Random networks of automata: a simple annealed approximation, Europhys. Lett. 1 (1986) 45–49.
[9] B. Derrida, D. Stauffer, Phase transitions in two dimensional Kauffman cellular automata, Europhys. Lett. 2 (1986) 739–745.
[10] R. Edwards, Analysis of continuous-time switching networks, Physica D 146 (2000) 165–199.
[11] R. Edwards, Chaos in neural and gene networks with hard switching, Differential Equations Dyn. Systems 9 (2001) 187–220.
[12] R. Edwards, L. Glass, Combinatorial explosion in model gene networks, Chaos 10 (2000) 691–704.
[13] R. Edwards, H.T. Siegelmann, K. Aziza, L. Glass, Symbolic dynamics and computation in model gene networks, Chaos 11 (2001) 160–169.
[14] M.B. Elowitz, S. Leibler, A synthetic oscillatory network of transcriptional regulators, Nature 403 (2000) 335–338.
[15] H. Flyvbjerg, An order parameter for networks of automata, J. Phys. A 21 (1988) L955–L960.
[16] T.S. Gardner, C.R. Cantor, J.J. Collins, Construction of a genetic toggle switch in *Escherichia coli*, Nature 403 (2000) 339–342.
[17] L. Glass, Combinatorial and topological methods in nonlinear chemical kinetics, J. Chem. Phys. 63 (1975) 1325–1335.
[18] L. Glass, S.A. Kauffman, Cooperative components, spatial localization and oscillatory cellular dynamics, J. Theoret. Biol. 34 (1972) 219–237.
[19] L. Glass, S.A. Kauffman, The logical analysis of continuous, nonlinear biochemical control networks, J. Theoret. Biol. 39 (1973) l03–l29.
[20] L. Glass, J.S. Pasternack, Prediction of limit cycles in mathematical models of biological oscillations, Bull. Math. Biol. 40 (1978) 27–44.
[21] L. Glass, J.S. Pasternack, Stable oscillations in mathematical models of biological control systems, J. Math. Biol. 6 (1978) 207–223.
[22] S.A. Kauffman, Metabolic stability and epigenesis in randomly constructed genetic networks, J. Theoret. Biol. 22 (1969) 437–467.
[23] S.A. Kauffman, Origins of Order: Self-Organization and Selection in Evolution, Oxford University Press, Oxford, 1993.
[24] B. Luque, R.V. Solé, Phase transitions in random networks: simple analytic determination of critical points, Phys. Rev. E 55 (1997) 257–260.
[25] G. Marnellos, E. Mjolsness, A gene network approach to modelling early neurogeneisis in drosophila, in: R.B. Altman, A.K. Dunker, L. Hunter, T.E. Klein (Eds.), Pacific Symposium on Biocomputing '98, World Scientific, Singapore, 1998, pp. 30–41.
[26] H.H. McAdams, L. Shapiro, Circuit simulation of genetic networks, Science 269 (1995) 650–656.
[27] L. Mendoza, E. Alvarez-Buylla, Dynamics of the genetic regulatory network for *Arabidopsis thaliana* flower morphogenesis, J. Theoret. Biol. 193 (1998) 307–319.
[28] T. Mestl, C. Lemay, L. Glass, Chaos in high dimensional neural and gene networks, Physica D 98 (1996) 33–52.
[29] T. Mestl, E. Plahte, S.W. Omholt, Periodic solutions in systems of piecewise-linear differential equations, Dyn. Stability Systems 10 (1995) 179–193.
[30] J. Reinitz, D.H. Sharp, Mechanism of *eve* stripe formation, Mech. Devel. 49 (1995) 133–158.
[31] D. Thieffry, R. Thomas, Dynamical behavior of biological regulatory networks. II. Immunity control in bacteriophage lambda, Bull. Math. Biol. 57 (1995) 277–295.
[32] R. Thomas, Boolean formalization of genetic control circuits, J. Theoret. Biol. 42 (1973) 563–585.
[33] R. Thomas, R. D'Ari, Biological Feedback, CRC Press, Boca Raton, FL, 1990.
[34] G. Weisbuch, Complex Systems Dynamics, Addison-Wesley, Redwood City, 1991.
[35] C.-H. Yuh, H. Bolouri, E.H. Davidson, Genomic Cis-regulatory logic: experimental and computational analysis of a sea urchin gene, Science 279 (1998) 1896–1902.