Structure and timescale analysis in genetic regulatory networks

Madalena Chaves, Eduardo D. Sontag and Réka Albert

***Abstract*— Regulation of gene expression is achieved through networks of interactions among genes and gene products. Genetic networks are sometimes described in a qualitative way, for instance by means of discrete or even Boolean models. Even when such models accurately reflect the basic structure of interactions, they are in general not suitable for robustness analysis, as one needs to study the effect of biologically-relevant perturbations on the dynamics of the system.**

This work is concerned with the study of the robustness and fragility of gene regulation networks to variability in the timescales of the distinct biological processes involved. It explores and compares two methods: introducing asynchronous updates in a Boolean model, or integrating the Boolean rules in a continuous, piecewise linear model. As an example, the segment polarity network of the fruit fly is analyzed. A theoretical characterization is given of the model's ability to predict the correct development of the segmented embryo, in terms of the specific timescales of the various regulation interactions.

I. Introduction

Biological systems are known to exhibit several robustness properties, allowing them to respond appropriately to perturbations in their environment. Such properties ensure that the right physiological and regulatory functions are accomplished at the right time, despite local variations in an organism's state [2]. The large amounts of experimental data available on biological networks mainly provide information on their components and interactions [9], [11]. In the absence of quantitative and kinetic information, regulatory networks are often described by discrete models, where the interactions are given by a set of logical rules [1], [12].

In this context, we will say that a Boolean model of a genetic network is *robust*, if it predicts the appropriate gene expression patterns with a fairly high frequency (eg, $\geq 85\%$ of experiments), when subject to random perturbations in its time dynamics. Otherwise, the model will be considered *fragile*, or vulnerable to such perturbations. In this paper, a methodology for analyzing the robustness of Boolean models is developed, which couples random perturbations in parameters with ideas from asynchronous parallel processing, or from simple hybrid systems (piecewise linear equations). The resulting stochastic methods provide a way to systematically investigate the space of parameters (timescales and effective concentrations), and identify regions of biologically meaningful dynamics. An application to a model of the segment polarity network of *Drosophila* uncovers points of vulnerability of the network, as well as regions of robustness to timescale perturbations.

M. Chaves is with the Institute for Systems Theory and Automatic Control, University of Stuttgart, Pfaffenwaldring 9, 70550 Stuttgart, Germany chaves@ist.uni-stuttgart.de.

E.D. Sontag is with the Department of Mathematics, Rutgers University, Piscataway, NJ 08854 USA sontag@math.rutgers.edu.

R. Albert is with the Department of Physics and Huck Institutes for the Life Sciences, Pennsylvania State University, University Park, PA 16802 USA ralbert@phys.psu.edu.

II. Asynchronous algorithms

The issues encountered in dynamic analysis of genetic regulatory networks are often similar to features of computation in large networks of processors (nodes). Consider a Boolean model for a network X_1, X_2, ..., X_N of genes and gene products. Synchronous updates mean that, at each step $k+1$, the states of all the nodes in the network are updated simultaneously, according to

$$X_i^{k+1} = F_i(X_1^k, X_2^k, \ldots, X_N^k), \quad i = 1, \ldots, N$$

where F_i is the regulating function for node X_i (in our example, mRNAs or proteins).

In processor networks performing parallel computation usually every processor is assigned a set of tasks to be performed at appointed times, the result of which is then communicated to the other nodes in the network [3]. The data communication and update among nodes should take into account that each of them has its particular processing rate. A perfect synchronization is often not possible, and asynchronous methods have been developed (as in the work of R. Thomas [13], [14]). We have also previously developed asynchronous methods which allow different timescales for the different processes within the network [4]. Here we introduce an alternative asynchronous algorithm where each node is updated according to its own specific time unit. The updating times of i-th node are pre-specified as:

$$T_i^{k+1} = T_i^k + \gamma_i = k\gamma_i, \quad k \in \mathbb{N}, \tag{1}$$

where γ_i, $i = 1, \ldots, N$, are fixed constants denoting the time unit for each node. The γ_i are randomly chosen from a uniform distribution in an interval $[1-\varepsilon, 1+\varepsilon]$, where $\varepsilon \in (0,1)$. The constant ε defines the magnitude of the perturbations to the synchronous time unit. Note that the case $\varepsilon = 0$ reduces to the synchronous model, where every node is updated simultaneously ($\gamma_1 = \cdots = \gamma_N$), at the same time instants: $T_i^k = k$, for all $i = 1, \ldots, N$. Other distributions could be used, but in the present study we chose a uniform distribution to avoid imposing extra assumptions on the various biological timescales at this early stage in the analysis - as will be seen, some assumptions will arise naturally later on. At any given time t, the next node(s) to be updated is(are) i such that $T_i^k = \min_{j,\ell}\{T_j^\ell \geq t\}$, for some k. The variables X_i are updated according to:

$$X_i(T_i^k) = F_i(\, X_1(\tau_{1i}^k), \ldots, X_N(\tau_{Ni}^k)\,), \tag{2}$$

where τ_{ji}^k defines the most recent instant when node j was updated, that is

$$\tau_{ji}^k = \max_\ell \{T_j^\ell : T_j^\ell < T_i^k\}.$$

This algorithm is well adapted to the analysis of the genetic regulatory networks, as it is well known that the timescales of transcription, translation, and degradation processes can vary widely from gene to gene and can be anywhere from minutes to hours. For instance, in the example (Section IV), $\gamma_{wG_i} < \gamma_{wg_i}$ means that Wingless protein in cell i is translated at a faster rate (shorter time intervals) than *wingless* mRNA is produced.

A. Random order algorithm

For comparison purposes, we briefly describe an algorithm developed in [4], which guarantees that every node is updated exactly once during each unit time interval. A random order of updates for the N nodes is generated as a permutation ϕ^k of $\{1, \dots N\}$. This permutation is randomly chosen out of a uniform distribution over the set of all $N!$ possible permutations, at the beginning of the time unit k. The updating times for each node are now written as

$$T_i^k = N(k-1) + \phi^k(i), \quad k \in \mathbb{N},$$

so that $\phi^k(j) < \phi^k(i)$ implies $T_j^k < T_i^k$, and node j is updated before node i at the k-th iteration.

III. Glass-type networks

An alternative method for analysis of varying timescales in a genetic network, based in the work of L. Glass [10], is now described. In that paper, Glass introduced a class of piecewise linear differential equations that combine logical rules for the synthesis of products with linear (free) decay. In this method, each node is represented by two variables, one discrete and one continuous. The interactions among nodes are still modeled by Boolean functions [10], [7], [6]. The notation is as follows: $\widehat{X}_i$ denotes the continuous variable associated with node i, X_i its discrete variable, and the discrete variable's Boolean rule is F_i. The Glass-type model is then

$$\frac{d\widehat{X}_i}{dt} = \alpha_i(\ -\widehat{X}_i + F_i(X_1, X_2, \dots, X_N)\ ), \tag{3}$$

with $\alpha_i \geq \varepsilon$ for some fixed $\varepsilon > 0$, for $i = 1, \dots, N$. At each instant t, the discrete variable X_i is defined as a function of the continuous variable according to a threshold value:

$$X_i(t) = \begin{cases} 0, & \widehat{X}_i(t) \leq \theta_i \\ 1, & \widehat{X}_i(t) > \theta_i\ , \end{cases} \tag{4}$$

where $\theta_i \in (0,1)$.[1] The discrete variables X_i represent the ON and OFF levels of the nodes in the Boolean model. Since the initial conditions coincide (i.e., $X(0) = \widehat{X}(0)$) and $F_i \in \{0,1\}$, it is easy to see that solutions of (3) evolve in the hypercube $[0,1]^N$. Under these conditions, the limiting values "0" and "1" of the continuous variable $\widehat{X}_i$ represent, respectively, "absence of species i" and "maximal concentration of species i" – thus we can view the $\widehat{X}_i$ as dimensionless variables, scaled to attain their maximal values at 1. The continuous dynamics is translated into a Boolean ON/OFF response, according to θ_i: as soon as $\widehat{X}_i$ increases above θ_i, species i is considered to be in the ON state; otherwise it remains in the OFF state (see also [6]). Thus the parameter θ_i defines the fraction of "maximal concentration" necessary for a protein or mRNA to effectively perform its biological function.

The α_i represent different timescales for the different processes, and their inverses may be interpreted as half-lives of mRNA or proteins. In fact, they are naturally related to the individual time units γ_i: using Euler's method to discretize system (3) obtains

$$\widehat{X}_i(t + \Delta t) = \widehat{X}_i(t) + \alpha_i \Delta t(-\widehat{X}_i(t) + F_i(X(t))).$$

Now notice that choosing the integrating time interval to be such that $\alpha_i \Delta t = 1$ recovers the discrete asynchronous algorithm with specific time units

$$\gamma_i = \Delta t = \alpha_i^{-1}.$$

With this method the structure and relative timescales of the network may be analyzed, by randomly varying the half-lives α_i, and the effective ON concentrations θ_i. It is easy to see that the steady states of the piecewise linear equations (3) are still those of the Boolean model, since:

$$\frac{d\widehat{X}_i}{dt} = 0 \ \Leftrightarrow\ \widehat{X}_i = X_i = F_i(X_1, X_2, \dots, X_N),$$

for $i = 1, \dots, N$, independently of θ_i.

IV. The segment polarity genes in the fruit fly

The methods above will be applied to the robustness analysis of a Boolean model of *Drosophila melanogaster* segment polarity network. This gene network is responsible for defining the segmentation of the embryo of the fly. The best characterized segment polarity genes include *engrailed* (*en*), *wingless* (*wg*), *hedgehog* (*hh*), *patched* (*ptc*), *cubitus interruptus* (*ci*) and *sloppy paired* (*slp*), coding for the corresponding proteins, which we will represent by capital letters (*EN*, *WG*, *HH*, *PTC*, *CI* and *SLP*). Two additional proteins, *CIA*, and *CIR*, resulting from transformations of the protein *CI*, also play important roles.

The Boolean model to be analyzed is depicted in Table I, and was introduced in [1]. (Further robustness analysis was also developed in [4], [5].) In this model, a parasegment of four cells is considered: the variables are the expression levels of the segment polarity genes and proteins (listed above) in each of the four cells. The model successfully describes the transition from the initial expression pattern (5) to a final pattern two or three developmental stages later, when the embryo has been clearly divided into parasegments

[1] We will analyze the behaviors of trajectories of systems of the form (3), assuming that trajectories are well-defined. Since the right-hand sides of equations of these type are discontinuous, it is very difficult to give general existence and uniqueness theorems for solutions of inital-value problems. One must impose additional assumptions, insuring that only a finite number of switches can take place on any finite time interval, and often tools from the theory of differential inclusions must be applied, see for instance [8], [6].

TABLE I

BOOLEAN OF SEGMENT POLARITY GENE PRODUCTS IN THE MODEL. SUBSCRIPTS $i = 1, 2, 3, 4$ DENOTE CELL NUMBER.

Node	Boolean updating function (synchronous algorithm)
SLP_i	$SLP_i(k+1) = \begin{cases} 0 & \text{if} \quad i \in \{1,2\} \\ 1 & \text{if} \quad i \in \{3,4\} \end{cases}$
wg_i	$wg_i(k+1) = (CIA_i(k)$ and $SLP_i(k)$ and not $CIR_i(k))$ or $[wg_i(k)$ and $(CIA_i(k)$ or $SLP_i(k))$ and not $CIR_i(k)]$
WG_i	$WG_i(k+1) = wg_i(k)$
en_i	$en_i(k+1) = (WG_{i-1}(k)$ or $WG_{i+1}(k))$ and not $SLP_i(k)$
EN_i	$EN_i(k+1) = en_i(k)$
hh_i	$hh_i(k+1) = EN_i(k)$ and not $CIR_i(k)$
HH_i	$HH_i(k+1) = hh_i(k)$
ptc_i	$ptc_i(k+1) = CIA_i(k)$ and not $EN_i(k)$ and not $CIR_i(k)$
PTC_i	$PTC_i(k+1) = ptc_i(k)$ or $(PTC_i(k)$ and not $HH_{i-1}(k)$ and not $HH_{i+1}(k))$
ci_i	$ci_i(k+1) =$ not $EN_i(k)$
CI_i	$CI_i(k+1) = ci_i(k)$
CIA_i	$CIA_i(k+1) = CI_i(k)$ and [not $PTC_i(k)$ or $HH_{i-1}(k)$ or $HH_{i+1}(k)$ or $hh_{i-1}(k)$ or $hh_{i+1}(k)$]
CIR_i	$CIR_i(k+1) = CI_i(k)$ and $PTC_i(k)$ and not $HH_{i-1}(k)$ and not $HH_{i+1}(k)$ and not $hh_{i-1}(k)$ and not $hh_{i+1}(k)$

of about four cells each (see first entry of Table II). We adopt the notation "wg_1^k" or "$wg_1(k)$" to represent the state of *wingless* mRNA in the first cell of the parasegment at time k. Similar notations apply for other mRNAs and proteins. Periodic boundary conditions are assumed, meaning that: $node_{4+1} = node_1$ and $node_{1-1} = node_4$. The wild type initial pattern corresponds to:

$$wg_4^0 = 1,\ en_1^0 = 1,\ hh_1^0 = 1,\ ptc_{2,3,4}^0 = 1,\ ci_{2,3,4}^0 = 1, \quad (5)$$

with the remaining nodes zero.

A complete analysis of the steady states for the Boolean model (Table I) is found in [1]. Table II summarizes these results, indicating the expressed nodes in each of the six steady-states. We note that three of the steady states agree perfectly with experimentally observed states corresponding to wild type, to *ptc* knockout mutant (broad striped) and to *en*, *wg* or *hh* knockout mutant (non-segmented) embryonic patterns, the latter two corresponding to embryonic lethal phenotypes (see [1] for appropriate references).

TABLE II

COMPLETE CHARACTERIZATION OF THE MODEL'S STEADY STATES. ONLY ON (EXPRESSED) NODES ARE INDICATED.

Steady state	Expressed nodes
wild type (WT)	wg_4, WG_4, en_1, EN_1, hh_1, HH_1, $PTC_{2,3,4}$, $ptc_{2,4}$, $ci_{2,3,4}$, $CI_{2,3,4}$, $CIA_{2,4}$, CIR_3
broad stripes (BS)	$wg_{3,4}$, $WG_{3,4}$, $en_{1,2}$, $EN_{1,2}$, $hh_{1,2}$, $HH_{1,2}$, $ptc_{3,4}$, $PTC_{3,4}$, $ci_{3,4}$, $CI_{3,4}$, $CIA_{3,4}$
no segmentation (NS)	$ci_{1,2,3,4}$, $CI_{1,2,3,4}$, $PTC_{1,2,3,4}$, $CIR_{1,2,3,4}$
wild type variant (WTV)	wg_4, WG_4, en_1, EN_1, hh_1, HH_1, $PTC_{1,2,3,4}$, $ptc_{2,4}$, $ci_{2,3,4}$, $CI_{2,3,4}$, $CIA_{2,4}$, CIR_3
ectopic (EC)	wg_3, WG_3, en_2, EN_2, hh_2, HH_2, $PTC_{1,3,4}$, $ptc_{1,3}$, $ci_{1,3,4}$, $CI_{1,3,4}$, $CIA_{1,3}$, CIR_4
ectopic variant (ECV)	wg_3, WG_3, en_2, EN_2, hh_2, HH_2, $PTC_{1,2,3,4}$, $ptc_{1,3}$, $ci_{1,3,4}$, $CI_{1,3,4}$, $CIA_{1,3}$, CIR_4

V. RESULTS

Applying the asynchronous algorithm to the segment polarity network shows that, when started from the initial wild type state (5), any of the steady states of the model (Table II) may occur with a certain probability. The probability of occurrence of each pattern depends on the range over which the individual time units γ_i are allowed to vary (see Fig. 1). For $\varepsilon = 0$, the wild type steady state is attained with probability 100% (corresponding to the synchronous Boolean model). As ε increases to 0.01 (resp. 0.1) this value decreases to 60% (resp. 44%). However, further increase in ε (hence larger time intervals) unexpectedly leads to an increase in the occurrence of the wild type state, up to 51% for $\varepsilon = 0.9$. Other final states observed are the broad-striped pattern ($25\% - 38\%$) observed in heat-shock experiments and *ptc* mutants and the pattern with no segmentation ($12\% - 15\%$) observed in *en*, *hh* or *wg* mutants. Each of the other three steady states occurs with frequencies less than 5%. (These values were obtained from 10000 numerical experiments.)

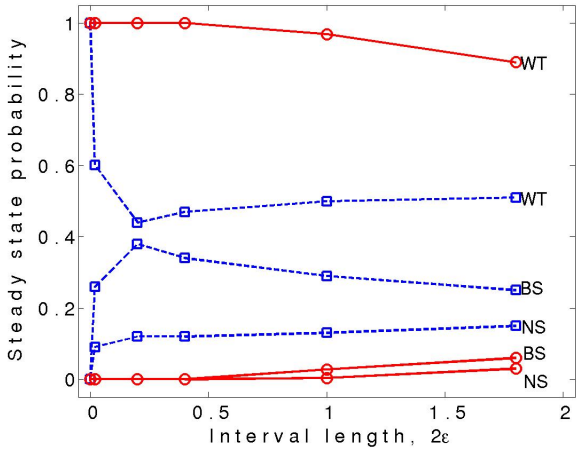

Fig. 1. Probability of occurrence of the three most frequent patterns under variable range of timescales (see Table II for notation). Dashed lines/squares represent asynchronous algorithm results, while solid lines/circles represent Glass-type model results (out of 1000 runs, with $\theta_i = 0.5$, for all i.

With the random order algorithm, the WT steady state is reached with a probability of 56%, followed by BS (24%) and NS (15%). Less frequent are WTV (4.2%) and EC, ECV ($< 1\%$). To apply the Glass type model in a comparable way, the scale factors α_i^{-1} are randomly chosen from a uniform distribution in intervals of the form $[1-\varepsilon, 1+\varepsilon]$, $\varepsilon \in (0,1)$. Also, to separately study the effect of varying ON thresholds, in Fig. 1 all θ_i are fixed at 0.5. With this method, in contrast to the asynchronous Boolean model, the wild type pattern occurs with frequencies that decrease monotonically with ε, down to 89% for $\varepsilon = 0.9$ (Fig. 1). The next more frequently achieved patterns are BS (around 6%), NS (3%), and WTV (1%). To analyze the effect of varying θ_i, we next randomly chose α_i from the interval $[0.5, 1.5]$, and also chose θ_i randomly from uniform distributions in the intervals $(0,1)$, $(0, 0.5]$ and [0.5,1). The results are summarized in Table III and indicate that higher thresholds are not realistic. On the

contrary, small fractions of the maximal concentration (below 50%) are already sufficient for a gene or protein to be active. (For instance, in [15] a threshold of 10% was used.)

VI. Robustness under timescale separation

It is well known that post-translational processes such as protein conformational changes or complex formation, usually have shorter durations than transcription, translation or mRNA decay. In this section, a timescale separation among processes is introduced, equivalent to updating proteins first and mRNAs later. Timescale separation is straightforwardly implemented in the random order algorithm presented in Section II; at the k-th updating step we generate two random permutations, ϕ^k_{Prot} and ϕ^k_{mRNA}, within the set of proteins and mRNAs, respectively. Then the N nodes are updated in the order given by

$$\phi^k = (\phi^k_{Prot}, \phi^k_{mRNA}).$$

This method shows that the Boolean model is very robust, in the sense that when started from the wild type initial condition, the wild type pattern occurs with a frequency of 87.5% and only one other steady state is observed, the broad striped pattern, with a frequency of 12.5%. Furthermore, these frequencies are exact, as follows from a complete characterization of the model resulting from incorporation of a protein/mRNA timescale separation into the random order algorithm. We summarize the results in the next theorem, stated without proof, and refer to [4] for more details.

Theorem 1: In the random order algorithm with timescale separation, let $wg_3^0 = 0$, $ptc_3^0 = 1$, $hh_{2,4}^0 = 0$ and $ci_3^0 = 1$ (as satisfied by initial condition (5)). Then system diverges from the wild type pattern if and only if the permutation ϕ^1 satisfies the following sequence among the proteins *CI*, *CIA*, *CIR* and *PTC*:

$$\begin{array}{lllllll} CIR_3 & CI_3 & & CIA_3 & & PTC_3, \\ & CI_3 & CIR_3 & CIA_3 & & PTC_3, \\ & CI_3 & & CIA_3 & CIR_3 & PTC_3. \end{array} \qquad (6)$$

The other proteins may appear in any of the remaining slots. ∎

Thus we can compute the exact probability with which the random order algorithm (with timescale separation) leads to either the wild type or broad stripes pattern: the latter is simply the fraction of sequences of the form (6) [4].

For the Glass-type and asynchronous algorithms, time separation among processes is implemented by using two non-overlapping intervals for the scaling factors:

$$\begin{array}{ll} \gamma_i^{-1}, \alpha_i \in A_{mRNA}, & \text{if } \ X_i \in \{wg, en, hh, ptc, ci\} \\ \gamma_i^{-1}, \alpha_i \in A_{Prot}, & \text{otherwise} \end{array}$$

with, for instance, $A_{mRNA} = [0.2, 0.6]$ and $A_{Prot} = [1.4, 1.8]$. Under these conditions, choosing the factors α_i from a uniform distribution in these intervals, numerical experiments indicate that the two methods respond in mostly similar ways, with only the wild type and broad stripes patterns occurring at steady state when the systems start from (wild type) initial condition (5).

For the asynchronous algorithm, the probabilities of convergence to each of the steady states clearly depend on the distance between the two intervals (see [4]). Convergence to wild type is between 93% and 100%.

For the Glass-type model, two cases can be distinguished. For $\theta_i \leq 0.5$, numerical simulations show that the model reaches wild type pattern with probability near 100%, even when there is some overlap between A_{mRNA} and A_{Prot}. In fact, we next theoretically prove that *the wild type pattern is indeed the unique possible steady state* of the hybrid system (3) and initial condition (5), when there is a suitable distance between the intervals, $\theta_i = \theta$ for all i, and a lower bound on θ (Theorem 2). For $\theta_i > 0.5$, we have found no condition that guarantees convergence to the wild type steady state, and indeed numerical simulations show that, even for large interval separation, the system may converge to one of the mutant patterns.

Theorem 2: Consider system (3), with $\theta_i = \theta$ for all $i = 1, \ldots, N$, and initial condition (5). Assume that the scaling factors α_i are chosen from intervals A_{mRNA} and A_{Prot} that satisfy:

$$\text{For all} \quad a \in A_{mRNA} \ \text{ and } \ b \in A_{Prot}: \quad 0 < 2a < b. \qquad (7)$$

Assume also that one of the following conditions holds:

(a) $\theta \leq 1/2$ and $(1-\theta)^2 \leq \theta$ or equivalently $0.382 \approx (3-\sqrt{5})/2 \leq \theta \leq 1/2$;
(b) $\theta \leq 1/2$ and $\alpha_{PTC_3} > \alpha_{CI_3}$;

then $wg_3(t) = 0$ for all t.

The differences and similarities between discrete and continuous models are illustrated by Theorems 1 and 2. The second (sufficient) condition of Theorem 2 guarantees convergence to the wild type steady state for all $0 < \theta \leq 0.5$, but assumes that $\alpha_{PTC_3} > \alpha_{CI_3}$. This is an analog to Theorem 1: if $\alpha_{PTC_3} > \alpha_{CI_3}$, then (starting from $PTC_3(0) = CI_3(0) = 0$ and assuming $F_{PTC_3} = F_{CI_3} = 1$) $\widehat{PTC}_3$ increases faster than $\widehat{CI}_3$, implying that PTC_3 becomes ON faster than CI_3. Such response prevents the events listed in Theorem 1, which would lead to a mutant state. Thus, both discrete and piecewise linear model predict that the sequence of *PTC*, *CI* expression in the third cell is one of the fundamental pieces in establishing the correct development of embryo segmentation.

Condition (a) (Theorem 2) applies only for $0.382 \leq \theta \leq 0.5$, but does not require any extra conditions to prevent the single "jump" event described by Theorem 1.

Theorem 2 identifies essentially three distinct regions of behavior for the case $\theta_i = \theta$: $(0, (3-\sqrt{5})/2]$, $[(3-\sqrt{5})/2, 1/2]$, and $[1/2, 1)$. To test the performance of the system and compare it to previous results, we considered two timescale situations: $\alpha_i \in [0.5, 1.5]$ for all i, or the timescale separation $A_{mRNA} = [0.2, 0.6]$, $A_{Prot} = [1.4, 1.8]$. In each case, we randomly assigned values to θ_i from uniform distributions in the intervals $(0, 1)$, $(0, 0.5)$ and $(0.4, 0.5)$. Table III summarizes the results for all combinations of θ_i and α_i regions. The most general case, allowing a large degree of freedom in both timescales and concentration thresholds, indicates

TABLE III

PROBABILITIES OF CONVERGENCE TO A GIVEN STEADY STATE, WITH DISTINCT CONCENTRATION THRESHOLDS θ_i AND DISTINCT TIMESCALES α_i, IN THE GLASS-TYPE MODEL. (PROBABILITIES COMPUTED OUT OF 1000 SIMULATIONS FOR EACH CASE.)

Steady state pattern	$(0,1)$	$(0.5,1)$	$(0,0.5]$	$(0,0.5]$ $\alpha_{PTC_3} > \alpha_{CI_3}$	$[0.4,0.5]$
$A_{mRNA} = A_{Prot} = [0.5, 1.5]$					
WT	45.6%	57.1%	84.1%	90%	92.6%
BS	27.8%	15.1%	12%	6.2%	7.3%
NS	24.4%	25.8%	0.9%	0.9%	0.05%
WTV	2.1%	1.9%	2.9%	2.9%	0%
$A_{mRNA} = [0.2, 0.6]$, $A_{Prot} = [1.4, 1.8]$					
WT	74.1%	52.7%	96.6%	97.1%	100%
BS	10.8%	3.3%	3.3%	2.8%	0%
NS	14.1%	43.9%	0%	0%	0%
WTV	1.0%	0%	0.1%	0.1%	0%

the fragility of the network, with a very high incidence on mutant patterns ($\theta_i \in (0,1)$, $\alpha_i \in [0.5, 1.5]$). Comparing the reasonable results obtained for $\theta_i \leq 0.5$ with the bad performance for $0.5 < \theta_i < 1$, we conclude that the optimal ON concentratin for proteins or mRNA is below 50% of maximal concentration. Restricting θ_i even further to one of the conditions given in Theorem 2 dramatically increases the probability that the system develops in the correct way.

VII. GLASS-TYPE MODEL PROVIDES EXACT CONVERGENCE TO WILD TYPE PATTERN

In this section we will require that the intervals A_{mRNA} and A_{Prot} do not overlap, by satisfying assumption (7). A second assumption is that the effective maximal concentration is equal for all nodes and satisfies $\theta \leq 1/2$, which is equivalent to: $\leq \ln \frac{1}{\theta}$. Theorem 2 shows that the steady state representing the broad stripes pattern cannot ever be reached in system (3) from the initial condition (5), when $\theta \leq 1/2$ and either of the extra conditions holds.

Theorem 3: Consider system (3), with $\theta_i = \theta$ for all $i = 1, \ldots, N$, and initial condition (5). Assume that the scaling factors α_i satisfy (7), and that $\theta \leq 1/2$. Then $wg_4(t) = 1$ and $PTC_1(t) = 0$ for all t.

This shows that the steady states NS, WTV, EC and ECV cannot ever be reached in system (3) with (5). From Theorems 2 and 3 we conclude that, under the timescale separation assumption, the Glass-type model (3) (with $\theta_i = \theta$ for all $i = 1, \ldots, N$) can only converge to the wild type pattern, when starting from (5), and for appropriate θ values (Table III).

The proof of Theorem 2 is given next, and the reader is referred to [5] for a proof of Theorem 3. We first summarize some useful observations. Let X denote any of the nodes in the network, and α its time rate. Since equations (3) are either of the form $d\widehat{X}/dt = \alpha(-\widehat{X} + 1)$ or $d\widehat{X}/dt = -\alpha\widehat{X}$, their solutions are continous functions, piecewise combinations of:

$$\widehat{X}^1(t) = 1 - (1 - \widehat{X}^1(t_0))\, e^{-\alpha(t-t_0)} \quad (8)$$

$$\widehat{X}^0(t) = \widehat{X}^0(t_0)\, e^{-\alpha(t-t_0)} \quad (9)$$

$\widehat{X}^1(t)$ (resp. $\widehat{X}^0(t)$) is monotonically increasing (resp. decreasing). In addition, note that discrete variables X can only switch between 0 and 1 at instants such that $\widehat{X}(t_{\text{switch}}) = \theta$:

$$t^1_{\text{switch}} = t_0 + \frac{1}{\alpha} \ln \frac{(1 - \widehat{X}(t_0))}{1 - \theta} \quad (10)$$

$$t^0_{\text{switch}} = t_0 + \frac{1}{\alpha} \ln \frac{\widehat{X}(t_0)}{\theta} \quad (11)$$

From the initial conditions, together with the constant values of SLP_i ($i = 1, 2, 3, 4$), we can immediately conclude:

$$\widehat{wg}_{1,2}(t) = WG_{1,2}(t) = 0, \quad (12)$$

$$\widehat{en}_{3,4}(t) = \widehat{EN}_{3,4}(t) = 0,$$

$$\widehat{hh}_{3,4}(t) = \widehat{HH}_{3,4}(t) = 0, \quad (13)$$

for all $t \geq 0$. Then, because $ci_{3,4}(0) = 1$ and $F_{ci_{3,4}} =$ not $EN_{3,4}$,

$$\widehat{ci}_{3,4}(t) = 1 \text{ and } \widehat{CI}_{3,4}(t) = 1 - e^{-\alpha_{CI_{3,4}}\, t}. \quad (14)$$

Lemma 7.1: Let $0 \leq t_0 < t_3 \leq t_1$ and $0 \leq t_2 < t_3$. Define $\delta = \ln \frac{1}{1-\theta} / \max_{1,\ldots,N} \alpha_i$. Assume $CIA_3(t) = 0$ for $t \in (t_2, t_3)$, and $wg_3(t) = 0$ for $t \in [0, t_3)$. Then

(a) $wg_3(t) = 0$ for $t \in [0, t_3 + \delta)$;
(b) $WG_3(t) = 0$ for $t \in [0, t_3 + \delta)$;
(c) $en_2(t) = EN_2(t) = 0$ for $t \in [0, t_3 + \delta)$;
(d) $hh_2(t) = HH_2(t) = 0$ for $t \in [0, t_3 + \delta)$.

Assume further that $PTC_3(t) = 1$ for $t \in (t_0, t_1)$. Then

(e) $PTC_3(t) = 1$ for all $t \in (t_0, t_3 + \delta)$.
(f) $CIA_3(t) = 0$ for all $t \in (t_2, t_3 + \delta)$.

Proof: Part (a) follows directly from the fact that $F_{wg_3}(t) = 0$ on $[0, t_3)$, and from (10).

To prove parts (b), (c), and (d), first note that initial conditions together with $wg_3(t) = 0$ for $t \in [0, t_3)$ imply $\widehat{WG}_3(t) = 0$, $\widehat{en}_2(t) = \widehat{EN}_2(t) = 0$, and $\widehat{hh}_2(t) = \widehat{HH}_2(t) = 0$, for $t \in [0, t_3]$. Then, from equations (8) to (11) we conclude that the corresponding discrete variables cannot switch from 0 to 1 during an interval of the form $[0, t_3 + \frac{1}{\alpha_j} \ln \frac{1}{1-\theta})$. Taking the largest common interval yields the desired results.

To prove parts (e) and (f), assume also that $PTC_3(t) = 1$ for $t \in (t_0, t_1)$. From (13) and part (d), it follows that function F_{PTC_3} does not switch in the interval $(t_0, t_3 + \delta)$ and in fact $PTC_3(t) = 1$ for all t in this interval. This, together with (13) and part (d) yield $F_{CIA_3}(t) = 0$ for $(t_0, t_3 + \delta)$, so that $\widehat{CIA}_3$ cannot increase in this interval and the discrete level satisfies $CIA_3(t) = 0$ for all $t \in (t_2, t_3 + \delta)$, as we wanted to show. ∎

Corollary 7.2: Let $0 \leq t_0 < t_3 \leq t_1$ and $0 \leq t_2 < t_3$. If $PTC_3(t) = 1$ for $t \in (t_0, t_1)$, $CIA_3(t) = 0$ for $t \in (t_2, t_3)$, and $wg_3(t) = 0$ for $t \in [0, t_3)$, then $wg_3(t) = 0$ for all t.

The proof follows by induction on k, setting $\tilde{t}_3 = t_3 + k\delta$ (see [4]).

Proof of Theorem 2: The rule for CIA_3 may be simplified to (by (13)) $F_{CIA_3} = CI_3$ and [notPTC_3 or hh_2 or HH_2]. From equation (14), we have that

$$CI_3(t) = 1, \quad \text{for all } t > \frac{1}{\alpha_{CI_3}} \ln \frac{1}{1 - \theta}. \quad (15)$$

On the other hand, since $ptc_3(0) = 1$, by continuity of solutions $ptc_3(t) = 1$ for all $t < \frac{1}{\alpha_{ptc_3}} \ln \frac{1}{\theta}$. This implies that the Patched protein satisfies

$$\widehat{PTC}_3(t) = 1 - e^{-\alpha_{PTC_3}\, t}, \quad 0 \le t \le \frac{1}{\alpha_{ptc_3}} \ln \frac{1}{\theta}$$

and therefore

$$PTC_3(t) = \begin{cases} 0, & 0 \le t \le \frac{1}{\alpha_{PTC_3}} \ln \frac{1}{1-\theta} \\ 1, & \frac{1}{\alpha_{PTC_3}} \ln \frac{1}{1-\theta} < t < \frac{1}{\alpha_{ptc_3}} \ln \frac{1}{\theta}. \end{cases} \tag{16}$$

By assumption, $\alpha_{PTC_3} > \alpha_{ptc_3}$ and also $\ln \frac{1}{1-\theta} \le \ln \frac{1}{\theta}$, defining a nonempty interval where PTC_3 is expressed. Now let $t_c = \frac{1}{\alpha_{CI_3}} \ln \frac{1}{1-\theta}$ and $t_p = \frac{1}{\alpha_{PTC_3}} \ln \frac{1}{1-\theta}$. $\widehat{CIA}_3(t)$ starts at zero and must remain so while $CI_3 = 0$, so that $CIA_3(t) = 0$ for $0 < t < t_c$. In the case $t_c > t_p$, letting $t_0 = t_p$, $t_1 = \frac{1}{\alpha_{ptc_3}} \ln \frac{1}{\theta}$, $t_2 = 0$, and $t_3 = t_c$ in Corollary 7.2, obtains $wg_3(t) = 0$ for all t. This proves item (b) of the theorem, and part of (a).

To finish the proof of item (a), we assume that $(1-\theta)^2 < \theta$ and must now consider the case $t_c \le t_p$. Then

$$\widehat{CIA}_3(t) = \begin{cases} 0, & 0 \le t \le t_c \\ 1 - e^{-\alpha_{CIA_3}\,(t-t_c)}, & t_c < t \le t_p \\ \widehat{CIA}_3(t_p)\, e^{-\alpha_{CIA_3}\,(t-t_p)}, & t_p < t \le \frac{1}{\alpha_{ptc_3}} \ln \frac{1}{\theta}, \end{cases}$$

Following equation (10) with $t_0 = t_c$ and $\widehat{CIA}_3(t_0) = 0$, CIA_3 might become expressed at time $t_c < t_a < t_p$:

$$t_a = t_c + \frac{1}{\alpha_{CIA_3}} \ln \frac{1}{1-\theta},$$

but it would then become zero again at (equation (11) with $t_0 = t_p$)

$$t_b = t_p + \frac{1}{\alpha_{CIA_3}} \ln \frac{\widehat{CIA}_3(t_p)}{\theta}.$$

Finally, we show that, even if $CIA_3(t) = 1$ for $t \in (t_a, t_b)$, wg_3 cannot become expressed in this interval. In this interval, $\widehat{wg}_3$ evolves according to $\widehat{wg}_3(t) = 1 - e^{-\alpha_{wg_3}\,(t-t_a)}$, and wg_3 can switch to 1 at time

$$t_w = t_a + \frac{1}{\alpha_{wg_3}} \ln \frac{1}{1-\theta}.$$

We will show that $t_w > t_b$, so $wg_3(t) = 0$ in the interval $[0, t_b)$. Writing

$$\begin{aligned} \ln \frac{\widehat{CIA}_3(t_p)}{\theta} &= \ln \frac{\widehat{CIA}_3(t_p)}{1-\theta} \frac{1-\theta}{\theta} \\ &= \ln \frac{\widehat{CIA}_3(t_p)}{1-\theta} + \ln \frac{1-\theta}{\theta} \\ &\le \ln \frac{1}{1-\theta} + \ln \frac{1}{1-\theta} \end{aligned}$$

where we have used $\widehat{CIA}_3(t_p) \le 1$ and the assumption on θ: $\frac{1-\theta}{\theta} \le \frac{1}{1-\theta}$. Therefore

$$\begin{aligned} t_b &\le t_p + \frac{2}{\alpha_{CIA_3}} \ln \frac{1}{1-\theta} \\ &< \frac{1}{\alpha_{wg_3}} \ln \frac{1}{1-\theta} + \frac{1}{\alpha_{CIA_3}} \ln \frac{1}{1-\theta} < t_w \end{aligned}$$

where we have used the timescale separation assumption (7). Letting $t_0 = t_p$, $t_2 = 0$, and $t_1 = t_3 = \min\{t_b, \alpha_{ptc_3}^{-1} \ln \frac{1}{\theta}\}$ in the Corollary, obtains $wg_3(t) = 0$ for all t. ∎

VIII. CONCLUSIONS

Two alternative methods were discussed for modeling gene expression networks: asynchronous Boolean methods and piecewise linear hybrid systems. Unrestricted variability in timescales or ON levels may lead to significant deviations from the experimentally observed dynamical behavior thus suggesting the fragility of the developmental process under severe perturbations. Both models agree in predicting the fundamental sequences of gene expression that irreversibly lead to a deviation from the wild type behavior and to a convergence to a mutant state. However, a single biologically justified restriction, that is, separating the timescales of post-translational and transcription/translation processes, unveils a remarkable robustness of the Boolean model in predicting the correct gene expression pattern. The Glass-type system with time separation, as a model of the segment polarity gene network, indicates that the topology of the network is more important than the fine-tuning of the kinetic parameters [15], since its results are robust for a large region of parameter (scaling factor, activation threshold) space. This result underscores the applicability of qualitative modeling when detailed kinetic information is unavailable.

REFERENCES

[1] R. Albert and H. G. Othmer. The topology of the regulatory interactions predicts the expression pattern of the *Drosophila* segment polarity genes. *J. Theor. Biol.*, vol 223, 2003, pp 1-18.

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular biology of the cell (4th ed.)*. Garland Publishing, New York, 2002.

[3] D.P. Bertsekas and J.N. Tsitsiklis. *Parallel and Distributed Computation, Numerical Method*. Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[4] M. Chaves, R. Albert, and E.D. Sontag. Robustness and fragility of boolean models for genetic regulatory networks. *J. Theor. Biol.*, vol 235, 2005, pp 431-449.

[5] M. Chaves, E.D. Sontag, and R. Albert. Methods of robustness analysis for boolean models of gene control networks. *IEE Proc. Syst. Biol.*, to appear, 2006.

[6] H. de Jong, J.L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Qualitative simulation of genetic regulatory networks using piecewise linear models. *Bull. Math. Biol.*, vol 66, 2004, pp 301-340.

[7] R. Edwards and L. Glass. Combinatorial explosion in model gene networks. *Chaos*, vol 10, 2000, pp 691-704.

[8] T. Gedeon. Attractors in continuous-time switching networks. *Communications on Pure and Applied Analysis*, vol 2, 2003, pp 187-209.

[9] L. Giot, J. S. Bader, C. Brouwer, A. Chaudhuri, and B. Kuang et al. A protein interaction map of *Drosophila melanogaster*. *Science*, vol 302, 2003, pp 1727-1736.

[10] L. Glass. Classification of biological networks by their qualitative dynamics. *J. Theor. Biol.*, vol 54, 1975, pp 85-107.

[11] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, and et al. Z. Bar-Joseph. Transcriptional regulatory networks in saccharomyces cerevisiae. *Science*, vol 298, 2002, pp 799-804.

[12] L. Sánchez and D. Thieffry. A logical analysis of the *Drosophila* gap-gene system. *J. Theor. Biol.*, vol 211, 2001, pp 115-141.

[13] R. Thomas. Boolean formalization of genetic control circuits. *J. Theor. Biol.*, vol 42, 1973, pp 563-585.

[14] R. Thomas. Regulatory networks seen as asynchronous automata: a logical description. *J. Theor. Biol.*, vol 153, 1991, pp 1-23.

[15] G. von Dassow, E. Meir, E.M. Munro, and G.M. Odell. The segment polarity network is a robust developmental module. *Nature*, vol 406, 2000, pp 188-192.