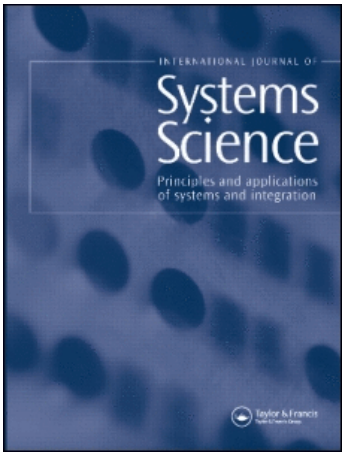


International Journal of Systems Science
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713697751

Stationary and structural control in gene regulatory networks: basic concepts

Edward R. Dougherty [ab]; Ranadip Pal [c]; Xiaoning Qian [d]; Michael L. Bittner [b]; Aniruddha Datta [a]
[a] Department of Electrical and Computer Engineering, Texas A&M University,, College Station, TX, 77843, USA [b] Computational Biology Division, Translational Genomics Research Institute, Phoenix, AZ, 85004, USA [c] Department of Electrical and Computer Engineering, Texas Tech. University, Lubbock, TX, 79409, USA [d] Department of Electrical and Computer Engineering, The Department of Statistics, Texas A&M University College Station, TX, 77843, USA

Online publication date: 08 January 2010

To cite this Article Dougherty, Edward R., Pal, Ranadip, Qian, Xiaoning, Bittner, Michael L. and Datta, Aniruddha(2010) 'Stationary and structural control in gene regulatory networks: basic concepts', International Journal of Systems Science, 41: 1, 5 — 16
To link to this Article: DOI: 10.1080/00207720903144560
URL: http://dx.doi.org/10.1080/00207720903144560

International Journal of Systems Science
Vol. 41, No. 1, January 2010, 5–16

Taylor & Francis
Taylor & Francis Group

Stationary and structural control in gene regulatory networks: basic concepts

Edward R. Dougherty[ab]*, Ranadip Pal[c], Xiaoning Qian[d], Michael L. Bittner[b] and Aniruddha Datta[a]

[a]*Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77843, USA;*
[b]*Computational Biology Division, Translational Genomics Research Institude, Phoenix, AZ, 85004, USA;*
[c]*Department of Electrical and Computer Engineering, Texas Tech. University, Lubbock, TX, 79409, USA;*
[d]*Department of Electrical and Computer Engineering, The Department of Statistics, Texas A&M University College Station, TX, 77843, USA*

(*Received 13 April 2009; final version received 5 June 2009*)

A major reason for constructing gene regulatory networks is to use them as models for determining therapeutic intervention strategies by deriving ways of altering their long-run dynamics in such a way as to reduce the likelihood of entering undesirable states. In general, two paradigms have been taken for gene network intervention: (1) stationary external control is based on optimally altering the status of a control gene (or genes) over time to drive network dynamics; and (2) structural intervention involves an optimal one-time change of the network structure (wiring) to beneficially alter the long-run behaviour of the network. These intervention approaches have mainly been developed within the context of the probabilistic Boolean network model for gene regulation. This article reviews both types of intervention and applies them to reducing the metastatic competence of cells via intervention in a melanoma-related network.

Keywords: control; gene regulatory network; genomic signal processing; Markov chains; perturbation theory

1. Introduction

The modelling of gene regulatory networks is necessary to describe the manner in which cells execute and control normal function and how abnormal function results from a breakdown in regulation. Hence, gene regulatory networks are critical to translational genomics, whose aim is to develop therapies based on the disruption or mitigation of aberrant gene function contributing to the pathology of a disease. Mitigation would be accomplished by the use of drugs to act on gene products. Engineering therapeutic tools involves synthesising dynamical networks, analysing these networks to characterise gene regulation and developing intervention strategies to modify dynamical behaviour (Dougherty and Datta 2005; Datta and Dougherty 2007; Shmulevich and Dougherty 2007). Two basic intervention approaches have been considered for gene regulatory networks in the context of probabilistic Boolean networks (PBNs) (Shmulevich, Dougherty, and Zhang 2002a; Shmulevich, Dougherty, Kim, and Zhang 2002), external control and structural intervention.

External control takes advantage of the fact that the dynamic behaviour of a PBN can be modelled by a Markov chain, thereby making intervention in PBNs amenable to the theory of Markov decision processes. Control is generally based on flipping (or not flipping) the value of a control gene. The first intervention approach of this kind involved the determination of an optimal single-gene flip to the network to drive it away from an undesirable state to a desirable state, optimality being based on mean first passage time (Shmulevich, Dougherty, and Zhang 2002b). Following this simple, one-time intervention, attention turned to dynamic-programming-based finite-horizon external control (Datta, Choudhary, Bittner, and Dougherty 2003), in which the effects are transient and the steady-state distribution is not changed, and then to infinite-horizon stationary external control (Pal, Datta, and Dougherty 2006) in which the steady-state distribution is altered. In this article we consider infinite-horizon stationary control.

Structural intervention involves a one-time change of the network structure (wiring) to beneficially alter the long-run behaviour (steady state) of the network (Shmulevich, Dougherty, and Zhang 2002c; Xiao and Dougherty 2007; Qian and Dougherty 2008). Given a class of potential structural changes, the problem is to find the optimal structural intervention resulting in a desired alteration of the steady-state distribution. Ignoring random effects, PBN state transitions are determined by regulatory rules among the genes and a structural intervention results from perturbing one of the rules. Two issues are involved: (1) characterise the

*Corresponding author. Email: edward@ece.tamu.edu

ISSN 0020–7721 print/ISSN 1464–5319 online

DOI: 10.1080/00207720903144560
http://www.informaworld.com

effects of a rule perturbation and (2) determine the optimal perturbation(s) to achieve a desirable alteration in the steady-state distribution, one in which the steady-state mass is moved in the direction of desirable states and away from undesirable states.

2. Probabilistic Boolean networks

Before discussing intervention, we provide a brief review of PBNs, one sufficient to make the present article self-contained. We refer the reader to (Shmulevich and Dougherty 2007) for a comprehensive review. Probabilistic Boolean networks extend the classical Boolean networks (Kauffman 1969, 1993) by introducing uncertainty in the rule structure. This uncertainty is motivated by randomness in the inference procedure, inherent biological randomness, and model stochasticity owing to latent variables outside the model that are involved in regulation. We begin by defining a Boolean network and then introduce the PBN model as being composed of a set of Boolean networks.

A Boolean network $G(V,F)$ is defined by a set $V=\{x_1,x_2,\ldots,x_n\}$ of binary variables, $x_i\in\{0,1\}$, $i=1,\ldots,n$, and a list of Boolean functions $F=(f_1,f_2,\ldots,f_n)$. The value of x_i at time $t+1$ is completely determined by a subset $\{x_{i1},x_{i2},\ldots,x_{ik_i}\}\subset V$ at time t via a Boolean function $f_i\colon\{0,1\}^{k_i}\mapsto\{0,1\}$. Transitions are homogeneous in time and we have the update

$$x_i(t+1)=f_i(x_{i1}(t),x_{i2}(t),\ldots,x_{ik_i}(t)). \qquad (1)$$

Each x_i represents the state (expression) of gene i, where $x_i=1$ and $x_i=0$ represent gene i being expressed and not expressed, respectively. It is commonplace to refer to x_i as the i-th gene. The list F of Boolean functions represents the rules of regulatory interactions between genes. That is, any given gene transforms its inputs (regulatory factors that bind to it) into an output, which is the state or expression of the gene itself. All genes are assumed to update synchronously in accordance with the functions assigned to them and this process is then repeated. At any time t, the state of the network is defined by a state vector $\mathbf{x}(t)=(x_1(t),x_2(t),\ldots,x_n(t))$, called a *gene activity profile* (GAP). Given an initial state, a BN will eventually reach a set of states, called an *attractor cycle*, through which it will cycle endlessly. Each initial state corresponds to a unique attractor cycle and the set of states leading to a specific attractor cycle is known as the *basin of attraction* (BOA) of the attractor cycle.

A *Boolean network with perturbation* (BNp) is defined by allowing each gene to possess the possibility of randomly flipping its value with a positive probability p. Implicitly, we assume that there is an i.i.d. random perturbation vector $\gamma=(\gamma_1,\gamma_2,\ldots,\gamma_n)$, where $\gamma_i\in\{0,1\}$, the i-th gene flips if and only if $\gamma_i=1$, and $p=P(\gamma_i=1)$ for $i=1,2,\ldots,n$. If $\mathbf{x}(t)$ is the GAP at time t, then the next state $\mathbf{x}(t+1)$ is either $\mathbf{f}(\mathbf{x}(t))$ with probability $(1-p)^n$ or $\mathbf{x}(t)\oplus\boldsymbol{\gamma}$ with probability $1-(1-p)^n$, where $\mathbf{f}$ is the multi-output function from the truth table and $\oplus$ is component-wise addition modulo 2. Perturbation makes the corresponding Markov chain of a BNp irreducible and ergodic. Hence, the network possesses a steady-state distribution π describing its long-run behaviour. A BNp inherits the attractor structure from the original Boolean network without perturbation, the difference being that a random perturbation can cause a BNp to jump out of an attractor cycle, perhaps then transitioning to a different attractor cycle prior to another perturbation. If p is sufficiently small, then π will reflect the attractor structure within the original network. We can derive the transition probability matrix P if we know the truth table and the perturbation probability p for a BNp.

A binary (PBN) is composed of a family $\{B_1,B_2,\ldots,B_m\}$ of BNps together with probabilities governing the selection of a BNp at each time. The m constituent BNps are characterised by m network functions, $\{\mathbf{f}_1,\mathbf{f}_2,\ldots,\mathbf{f}_m\}$. At any time point there is a positive probability q of switching from the current governing constituent BNp to another, with the selection probabilities for transitioning to $B_1,B_2,\ldots,B_m$ given by $c_1,c_2,\ldots,c_m$, respectively. Note that the probability of switching to constituent network B_r is independent of the current network; indeed, when a switch is called for, the current network may 'switch' to itself. By definition, a PBN inherits the attractor cycles of its constituent BNps. There are two modeling interpretations regarding q. If $q<1$, the interpretation is that there are latent variables outside the network whose changes cause the model network to behave stochastically, the PBN is said to be *context-sensitive* (Brun, Dougherty and Shmulevich 2005), and q is usually assumed to be small, reflecting the assumption that the BNp governing network dynamics is only changed infrequently. If $q=1$, as in the original formulation of PBNs (Shmulevich et al. 2002), then the interpretation is that the uncertainty in the BNp arises from uncertainty in model inference and the PBN is said to be *instantaneously random*. Although we have defined PBNs as having binary gene values, there is nothing inherent in this restriction and the general definition assumes that each gene can take a finite number of values, say in the set $\{0,1,\ldots,d\}$. Both binary and ternary PBNs have been applied, where in the ternary case, the values -1, 0, and 1 correspond to a gene being down-regulated invariant, and up-regulated, respectively.

For the purposes of intervention we will be focusing on the Markov chain associated with a PBN. Hence, we are concerned with the transitions constituting the probability transition matrix. For a BNp, the probability of transitioning from GAP $\mathbf{y}$ to GAP $\mathbf{x}$ is given by

$$P_{\mathbf{y}}(\mathbf{x}) = \mathbf{1}_{[\mathbf{f}(\mathbf{y})=\mathbf{x}]}(1-p)^n + \mathbf{1}_{[\mathbf{x}\neq\mathbf{y}]}p^{\eta(\mathbf{x},\mathbf{y})}(1-p)^{n-\eta(\mathbf{x},\mathbf{y})}, \tag{2}$$

where $\eta(\mathbf{x}, \mathbf{y})$ is the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$, and $\mathbf{1}_{[\mathbf{f}(\mathbf{y})=\mathbf{x}]}$ is the indicator function that takes value 1 if $\mathbf{f}(\mathbf{y}) = \mathbf{x}$ according to the truth table and is equal to 0 otherwise. For instantaneously random PBNs,

$$P_{\mathbf{y}}(\mathbf{x}) = \sum_{j=1}^{m} c_j \mathbf{1}_{[\mathbf{f}_j(\mathbf{y})=\mathbf{x}]}(1-p)^n + \mathbf{1}_{[\mathbf{x}\neq\mathbf{y}]}p^{\eta(\mathbf{x},\mathbf{y})}(1-p)^{n-\eta(\mathbf{x},\mathbf{y})}. \tag{3}$$

For context-sensitive PBNs, the matter is more complicated. Here the states of the Markov chain consist of (context, GAP) pairs and the probability of transitioning from state (s, $\mathbf{y}$) to state (r, $\mathbf{x}$) is given by

$$\begin{aligned} P_{s,\mathbf{y}}(r,\mathbf{x}) = {} & \mathbf{1}_{[r=s]}(1-q+qc_s)\Big\{\mathbf{1}_{[\mathbf{f}_s(\mathbf{y})=\mathbf{x}]}(1-p)^n \\ & + \mathbf{1}_{[\mathbf{x}\neq\mathbf{y}]}p^{\eta(\mathbf{x},\mathbf{y})}(1-p)^{n-\eta(\mathbf{x},\mathbf{y})}\Big\} \qquad (4) \\ & + \mathbf{1}_{[r\neq s]}qc_r\Big\{\mathbf{1}_{[\mathbf{f}_r(\mathbf{y})=\mathbf{x}]}(1-p)^n \\ & + \mathbf{1}_{[\mathbf{x}\neq\mathbf{y}]}p^{\eta(\mathbf{x},\mathbf{y})}(1-p)^{n-\eta(\mathbf{x},\mathbf{y})}\Big\}, \qquad (5) \end{aligned}$$

where r, s denote the rth and sth BNp, which are the BNps at time $t+1$ and t (see (Faryabi, Vahedi, Chamberland, Datta, and Dougherty 2009) for derivations).

3. Stationary control

A PBN with control can be modelled as a stationary discrete-time dynamic system

$$z_{t+1} = f(z_t, u_t, w_t),\ t = 0, 1, \ldots, \tag{6}$$

where for all t, z_t is an element of a state space S, the control input u_t is an element of a control space C, the disturbance w_t is an element of a space D and $f: S \times C \times D \mapsto S$. In the particular case of a context-sensitive PBN with n genes composed of m Boolean networks with perturbation probability p and network switching probability q, the state space has $m2^n$ elements of the form (r, $\mathbf{x}$). By numbering these from 0 to $m2^n - 1$, we can equivalently take the state space to be $S = \{0, 1, 2, \ldots, m2^n - 1\}$, which is what we will do. In the special case of an instantaneously random PBN, the state space reduces to $S = \{0, 1, 2, \ldots, 2^n - 1\}$. We assume that there are k binary control inputs, so that, upon numbering these from 0 to $2^k - 1$, the control input u_t is constrained to take values in the space $C = \{0, 1, \ldots, 2^k - 1\}$. The disturbance w_t is manifested in terms of change of network based on the network transition probability q or change of state due to perturbation probability p. w_t is independent of prior disturbances $w_0, w_1, \ldots, w_{t-1}$.

Our objective is to come up with a sequence of control inputs, referred to as a *control strategy*, such that an appropriate cost function is minimised over the entire class of allowable control strategies. In doing so, we follow Pal et al. (2006). To arrive at a useful solution, a cost function must capture the costs and the benefits of using any control. Thus, we define a cost per stage, $\tilde{g}(i, u, j)$, depending on the origin state i, the destination state j, and the applied control input u. The actual design of a 'good' cost function is application dependent and is likely to require considerable expert knowledge. In finite-horizon control one can sum the costs over the number of time points constituting the horizon and take the expectation; however, this cannot safely be done with infinite horizon because there is a possibility that the summation of the one-stage costs may go to infinity (for all controls), thereby leading to an ill-posed optimisation problem.

One standard way of defining the cost function is to assume that the cost per stage $\tilde{g}(i, u, j)$ is *bounded* for any $i, j \in S$ and $u \in C$ and to introduce a *discount factor* $\alpha \in (0, 1)$ in the cost to guarantee that the limit of the finite sums converges as the horizon length goes to infinity. Specifically, our objective is to find a policy $\pi = \{\mu_0, \mu_1, \ldots\}$, where $\mu_t : S \to C, t = 0, 1, \ldots$, that minimises the cost function

$$J_\pi(z_0) = \lim_{M\to\infty} E\left\{\sum_{t=0}^{M-1} \alpha^t \tilde{g}(z_t, \boldsymbol{\mu}_t(z_t), z_{t+1})\right\}, \tag{7}$$

where the cost per stage $\tilde{g} : S \times C \times D \to \Re$ is given and the expectation is with respect to the state and the disturbance. In the general formulation, the inclusion of α in the cost captures the fact that costs incurred at a later time are less significant. In the case of cancer treatment, $\alpha < 1$ signifies that the condition of the patient in the initial stages of treatment is more important than the condition at a later stage, or in other words, the reward for improving the condition of the patient in the present is more significant than the reward obtained from similar improvement at a later stage.

Since in Equation (7) the cost is obtained by taking the expectation with respect to the origin and destination states, it is possible to replace $\tilde{g}(z_t, u_t, z_{t+1})$ by an equivalent cost per stage that does not depend on the destination state. This is accomplished by taking the expectation with respect to the destination state and leaving only the expectation with respect to the

original state. More specifically, we use as cost per stage the expected cost $g(i, u)$ given by (Bertsekas 2001):

$$g(i,u) = \sum_{j=0}^{m2^n-1} p_{ij}(u)\tilde{g}(i,u,j), \tag{8}$$

where $p_{ij}(u)$ is the transition probability under control u. In general, the cost $\tilde{g}(i,u,j)$ of moving from state i to state j under control u may depend on the starting state i; however, in the case of PBNs, we have no obvious basis for assigning different costs based on different initial states. Accordingly, we assume that the penalty $\tilde{g}(i,u,j)$ is independent of the starting state i and its value is based on the control effort and the terminal state j. The penalty is high if the end state is a bad state regardless of the starting state, and vice-versa. Hence $\tilde{g}(i,u,j) = \tilde{g}(u,j)$ and Equation (8) becomes

$$g(i,u) = \sum_{j=0}^{m2^n-1} p_{ij}(u)\tilde{g}(u,j). \tag{9}$$

If we denote by Ω the set of all *admissible* policies ω, i.e. the set of all function sequences $\omega = \{\mu_0, \mu_1, \dots\}$ with $\mu_t(z): S \to C, t = 0, 1, \dots,$ then the optimal cost function J^* is defined by

$$J^*(z) = \min_{\omega\in\Omega} J_\omega(z), z \in S. \tag{10}$$

A *stationary policy* is an admissible policy of the form $\omega = \mu, \mu, \dots,$ and its corresponding cost function is denoted by J_μ. The stationary policy $\omega = \mu, \mu \dots$ is optimal if $J_\mu(z) = J^*(z)$ for all states z.

We minimise the cost function of Equation (7) under the assumption that the cost per stage is bounded, an assumption that holds for PBNs because the expected cost, $g(i,u)$, for state i is given by Equation (9), $\sum_{j=0}^{2^n-1} p_{ij}(u) = 1$, and $\tilde{g}(u,j)$ is bounded since the control and state spaces are finite. This is accomplished by introducing two mappings. For any cost function $J: S \to \Re$, define the mapping $TJ: S \to \Re$ by

$$(TJ)(i) = \min_{u\in C}[\, g(i,u) + \alpha \sum_{j=0}^{m2^n-1} p_{ij}(u)J(j)], \; i \in S. \tag{11}$$

TJ is the optimal cost function for the one-stage (finite-horizon) problem that has stage cost g and terminal cost αJ. Similarly for any cost function $J: S \to \Re$ and control function $\mu: S \to C$, define the mapping $T_\mu J: S \to \Re$ by

$$(T_\mu J)(i) = g(i,\mu(i)) + \alpha \sum_{j=0}^{m2^n-1} p_{ij}(\mu(i))J(j), \; i \in S. \tag{12}$$

$T_\mu J$ can be viewed as the cost function associated with the policy μ for the one-stage problem that has stage cost function g and terminal cost αJ. Since the mappings T and T_μ map functions $J: S \to \Re$ into new functions mapping S to $\Re$, one can define the composition of T with itself and T_μ with itself as follows:

$$(T^k J)(i) = (T(T^{k-1}J))(i), i \in S, \; k = 1, 2, \dots, \tag{13}$$

$$(T^0 J)(i) = J(i), \; i \in S, \tag{14}$$

and

$$(T_\mu^k J)(i) = (T_\mu(T_\mu^{k-1}J))(i), i \in S, \; k = 1, 2, \dots, \tag{15}$$

$$(T_\mu^0 J)(i) = J(i), \; i \in S. \tag{16}$$

The mappings T and T_μ play an important role in the solution of the optimal control problem of this section. Specifically, it can be shown that: (i) the optimal cost function J^* is the unique fixed point of the map T; (ii) the iteration $J_{t+1} = TJ_t$ converges to J^* as $t \to \infty$; and (iii) the mapping T_μ can be used to characterise the conditions under which a given stationary policy μ is optimal. These ideas are formalised in the following three theorems adapted from Bertsekas (2001) (see (Pal et al. 2006) for proofs).

Theorem (Convergence of the discounted-cost algorithm)**:** *For any bounded cost function $J: S \to \Re$, the optimal cost function J^* satisfies*

$$J^*(i) = \lim_{M\to\infty} (T^M J)(i), \quad \text{for all } i \in S. \tag{17}$$

Theorem (Bellman's equation)**:** *The optimal cost function J^* satisfies*

$$J^*(i) = \min_{u\in C}[\, g(i,u) + \alpha \sum_{j=0}^{m2^n-1} p_{ij}(u)J^*(j))], \quad \text{for all } i \in S, \tag{18}$$

or, equivalently, $J^* = TJ^*$. Furthermore, J^* is the unique solution of this equation within the class of bounded functions.

Theorem (Necessary and sufficient condition for optimality)**:** *A stationary policy μ is optimal if and only if $\mu(i)$ attains the minimum in Bellman's equation* (18) *for each $i \in S$; i.e.*

$$TJ^* = T_\mu J^*. \tag{19}$$

The three theorems provide a basis for computational algorithms for determining the optimal policy. The second theorem asserts that the optimal cost

function satisfies Bellman's equation while the first theorem states that the optimal cost function can be iteratively determined by running the recursion

$$J_{t+1} = TJ_t,\ t = 0, 1, 2, \ldots \tag{20}$$

for any bounded initial cost function $J_0 : S \to \Re$. Since this iteration is guaranteed to converge to J^*, one can keep on running this iteration until some stopping criterion is reached. The resulting policy is a stationary one, which by the third theorem, must be optimal. The iteration described in (20) is referred to as the *Value Iteration* procedure since, at every stage we are iterating on the values of the cost function and the optimal policy simply falls out as a by-product when the iteration converges to the optimal value of the cost function (see (Pal et al. 2006) for solution by *Policy Iteration*).

4. Stationary control in a melanoma network

To pragmatically discuss optimal intervention, one must work with a network for which the goal of intervention can be quantitatively specified. Here, we consider a gene regulatory network developed from data collected in a study of metastatic melanoma in which the abundance of messenger RNA for the gene WNT5A was found to be a highly discriminating difference between cells with properties typically associated with high metastatic competence versus those with low metastatic competence (Bittner, et al. 2000). These findings were validated and expanded in a second study (Weeraratna, et al. 2002). In this study, experimentally increasing the levels of the Wnt5a protein secreted by a melanoma cell line via genetic engineering methods directly altered the metastatic competence of that cell as measured by the standard *in vitro* assays for metastasis. A further finding of interest in the current study was that an intervention that blocked the Wnt5a protein from activating its receptor, the use of an antibody that binds the Wnt5a protein, could substantially reduce Wnt5a's ability to induce a metastatic phenotype. This suggests a study of control based on interventions that alter the contribution of the WNT5A gene's action to biological regulation, since the available data suggest that disruption of this influence could reduce the chance of a melanoma metastasizing, a desirable outcome.

We consider a 7-gene network containing the genes WNT5A, pirin, S100P, RET1, MART1, HADHB and STC2. To obtain the PBN, we have used the algorithms described in Pal, Ivanov, Datta, Bittner, and Dougherty (2005) to construct four highly probable Boolean networks to use as the constituent networks. There are $4 \times 2^7 = 512$ states in the Markov chain, 4 contexts and 128 GAPs. The last two binary digits of the state number represent the constituent network and the first seven binary digits constitute the GAP. The genes are ordered as WNT5A, pirin, S100P, RET1, MART1, HADHB and STC2, with WNT5A as the most significant bit (MSB) of the GAP profile and STC2 as the least significant bit (LSB) of the GAP profile.

The discounted-cost control strategy has been applied to the designed PBN with pirin chosen as the control gene ($u = 1$ signifying the state of pirin is reversed and $u = 0$ signifying no intervention), $p = q = 0.01$ being the probabilistic parameters of the PBN, and $\alpha = 0.9$ being the discount factor. The cost of control is assumed to be 1 and the states are assigned penalties as follows:

$$\tilde{g}(u, j) = \begin{cases} 5 & \text{if } u = 0 \text{ and WNT5A is 1 for state } j \\ 6 & \text{if } u = 1 \text{ and WNT5A is 1 for state } j \\ 1 & \text{if } u = 1 \text{ and WNT5A is 0 for state } j \\ 0 & \text{if } u = 0 \text{ and WNT5A is 0 for state } j \end{cases}$$

Since our objective is to down-regulate the WNT5A gene, a higher penalty is assigned for destination states having WNT5A up-regulated. Also for a given WNT5A status for the destination state, a higher penalty is assigned when the control is active versus when it is not.

The control objective is to down-regulate the WNT5A gene. Parts (a) and (b) of Figure 1 show the steady-state distributions of the original and controlled PBNs, respectively. States 0–255 have WNT5A = 0 (desirable) and states 256–511 have WNT5A = 1 (undesirable). The different steady-state distributions show that the stationary policy has shifted the probability mass from states with higher to lower metastatic competence. The probability mass of undesirable states in the original steady-state distribution is 0.4939, while in the controlled steady-state distribution it has been reduced to 0.3061. To numerically quantify the change, we multiply the steady-state distribution by the cost vector. For the original PBN the cost vector is 0 for states 0–255 and 5 for states 256–511. For the stationary policy, the cost vector is $\tilde{g}(\mu(z), z),\ z \in [0, 1, 2, \ldots, 511]$. The value for the stationary policy is 1.5551 as compared to 2.4695 for no control.

5. Structural intervention

In the present framework, a function perturbation is a change to one of the Boolean functions and

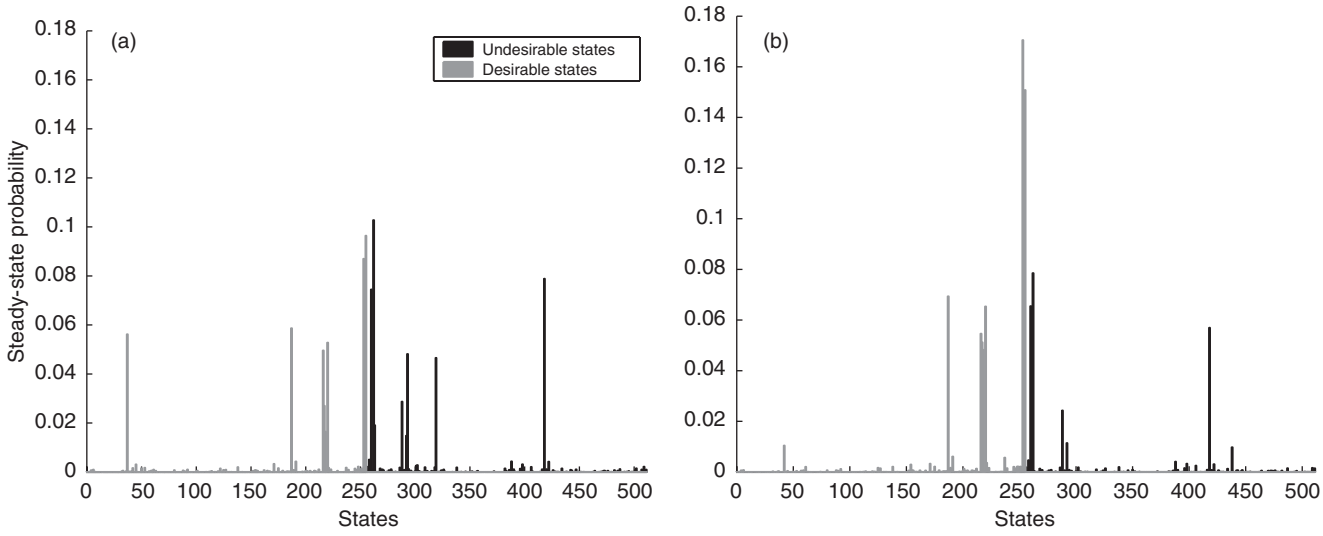


Figure 1. Steady-state distribution shifts: (a) original steady state distribution; (b) steady-state distribution using discounted cost stationary policy.

a structural intervention is a change to the truth table governing network transitions via a function perturbation (the term 'perturbation' being used in two different senses in the context of network intervention). To find an optimal structural intervention we need to determine the long-run effect of such a perturbation. Following function perturbation the original transition matrix P and steady-state distribution π are changed to $\tilde{P}$ and $\tilde{\pi}$, respectively. Our desire is to represent $\tilde{\pi}$ in terms of π and P so that an optimal perturbation can be determined directly from π and P.

The basic idea is to derive $\tilde{\pi}$ for PBNs based on the results for general Markov chains. For the original and perturbed networks, $\pi^T P = \pi^T$ and $\tilde{\pi}^T \tilde{P} = \tilde{\pi}^T$, where T denotes transpose. Since the Markov chain is irreducible, the time-average transition matrix is defined by

$$P^{\infty} = \lim_{j\to\infty}[P + P^2 + \cdots + P^j]/j = e\pi^T, \tag{21}$$

where e is a column vector whose components are all unity. Letting $\tilde{P} = P + E$, the steady-state distribution change is

$$(\tilde{\pi} - \pi)^T(I - P) = \tilde{\pi}^T E. \tag{22}$$

Analytical expressions for the steady-state distribution change can be obtained using generalised inverses (Hunter 1986), g-inverses (Cho and Meyer 2001) and the fundamental matrix (Schweitzer 1968). In general, a *g-inverse* of a matrix A is any matrix A^- such that $AA^-A = A$. If P and π are the transition matrix and the stationary distribution, respectively, of a finite irreducible Markov chain, and t and u are any vectors such that $\pi^T t \neq 0$ and $u^T e \neq 0$, then (1) $I - P + tu^T$ is nonsingular; (2) $[I - P + tu^T]^{-1}$ is a g-inverse of $I - P$; and (3) all g-inverses of $I - P$ are of the form $[I - P + tu^T]^{-1} + ef^T + g\pi^T$ for arbitrary f and g. The fundamental matrix is given by $Z = [I - P + e\pi^T]^{-1}$ and exists in our case since it exists for any ergodic chain (Kemeny and Snell 1960). While various results can be obtained in the general setting, we are interested in special cases that apply to PBNs; in particular, we want to express $\tilde{\pi} - \pi$ without reference to $\tilde{\pi}$.

For a *rank-one perturbation*, the perturbed Markov chain has the transition matrix $\tilde{P} = P + ab^T$, where a, b are two arbitrary vectors satisfying $b^T e = 0$, and ab^T represents a rank-one perturbation to the original Markov chain P. We now state the key result for this kind of perturbation (Hunter 2005; Qian and Dougherty 2008) (see (Qian and Dougherty 2008) for the details leading up to the theorem).

Theorem: *The steady-state distribution, fundamental matrix, and a g-inverse for $I - \tilde{P}$ for the rank-one perturbed network are given by*

$$\tilde{\pi}^T = \pi^T + \frac{\pi^T a}{1 - b^T Z a} b^T Z, \tag{23}$$

$$\tilde{Z} = \left[I - \frac{(\pi^T a) e b^T Z}{1 - b^T Z a}\right]\left[Z + \frac{Z a b^T Z}{1 - b^T Z a}\right], \tag{24}$$

$$[I - \tilde{P} + a(\pi + b)^T]^{-1} = Z\left[I + (e - a)\frac{\pi^T}{\pi^T a}\right]. \tag{25}$$

An important special case occurs when the transition mechanisms before and after perturbation differ only in one state, say the k-th state. Then $E = e_k b^T$ has

non-zero values only in its k-th row, where e_k is the elementary vector with a 1 in the k-th position and 0s elsewhere. Substituting this into Equation (23) yields

$$\tilde{\pi}^T = \pi^T + \frac{\pi^T e_k}{1 - b^T Z e_k} b^T Z = \pi^T + \frac{\pi_k}{1 - \beta_k} \beta^T \quad (26)$$

with $\beta^T = b^T Z$ (Schweitzer 1968). For the i-th state,

$$\tilde{\pi}_i = \pi_i + \frac{\pi_k \beta_i}{1 - \beta_k}. \quad (27)$$

The results for these special cases can be extended to arbitrary types of perturbations so that it is possible to compute the steady-state distributions of arbitrarily perturbed Markov chains in an iterative fashion (Qian and Dougherty 2008). For complicated perturbations, the computation for the iterative algorithm is in the form of vector-matrix multiplication and the complexity of the procedure increases linearly with the increasing complexity of perturbations.

To accomplish optimal intervention in PBNs, these general Markov chain results must be adapted to PBNs. First consider a single perturbation on a BNp, which is a flip for one target gene at the response side for the predictor state $\mathbf{k}$ in the truth table. Ignoring perturbation, each row in the transition matrix corresponds to a single input state $\mathbf{i}$ and the row consists of all 0s except for a 1 in the column of the output state $\mathbf{f}(\mathbf{i})$. A one-bit perturbation for input state $\mathbf{k}$ means that there is a network function transformation $\mathbf{f} \to \tilde{\mathbf{f}}$ with $\mathbf{v} = \tilde{\mathbf{f}}(\mathbf{k}) \neq \mathbf{f}(\mathbf{k}) = \mathbf{u}$ and $\tilde{\mathbf{f}}(\mathbf{i}) = \mathbf{f}(\mathbf{i})$ for all $\mathbf{i} \neq \mathbf{k}$. Hence, there is now (with $\tilde{\mathbf{f}}$) a 1 in the column of the output state $\mathbf{v}$ and a 0 in the column of state $\mathbf{u}$, whereas originally (with $\mathbf{f}$) there was a 0 in the column of the output state $\mathbf{v}$ and a 1 in the column of state $\mathbf{u}$, and all other entries in the matrix are unchanged. Relative to the BNp, this changes the transition matrix P only at p_{ku} and p_{kv} (note that we write states in italic when they are subscripts of transition probabilities – for instance, k in place of $\mathbf{k}$). Thus, $\tilde{p}_{ku} = p_{ku} - \tau$ and $\tilde{p}_{kv} = p_{kv} + \tau$, where τ is decided by the change to the truth table and the perturbation probability p: $\tau = (1-p)^n$. With Equation (27), we have

$$\tilde{\pi}_i = \pi_i + \frac{(1-p)^n \pi_k (z_{vi} - z_{ui})}{1 - (1-p)^n (z_{vk} - z_{uk})}. \quad (28)$$

This scheme can be extended to more sophisticated function perturbations by applying the results for either rank-one perturbations in Equation (23) or the iterative updating scheme for multi-row perturbations. For instance, for a function perturbation to a single regulatory function f_i for gene i having K_i predictor genes there will be changes to 2^{n-K_i} rows of the transition matrix P.

Now consider an instantaneously random PBN. Ignoring perturbation, each row in the transition matrix corresponds to a single input state $\mathbf{i}$ and that row contains a number of positive values summing to 1, with the remaining values being 0. The value in column j is the sum of all context-selection probabilities c_r such that $\mathbf{f}_r(\mathbf{i}) = \mathbf{j}$, where $\mathbf{f}_r$ is the network function for the rth constituent BNp. A one-bit perturbation for input state $\mathbf{k}$ corresponds to a change in one network function $\mathbf{f}_r$, so that exactly two columns are changed. This means that there is a network function transformation $\mathbf{f}_r \to \tilde{\mathbf{f}}_r$ with $\mathbf{v} = \tilde{\mathbf{f}}_r(\mathbf{k}) \neq \mathbf{f}_r(\mathbf{k}) = \mathbf{u}$ and $\tilde{\mathbf{f}}_r(\mathbf{i}) = \mathbf{f}_r(\mathbf{i})$ for all $\mathbf{i} \neq \mathbf{k}$. There is a gain of c_r in the column of the output state $\mathbf{v}$ and a loss of c_r in the column of state $\mathbf{u}$, and all other entries in the matrix are unchanged. Hence, we can proceed in the same manner as with a one-bit perturbation for a BNp using Equation (27). If we are not limited to changing a single constituent BNp network function for input state $\mathbf{k}$, then the situation becomes more complicated because there could be several changes in the transition matrix in the row corresponding to state $\mathbf{k}$, but this too can be handled (Qian and Dougherty 2008).

For a context-sensitive PBN, each state of the Markov chain is of the form $(r, \mathbf{i})$, where r denotes the r-th constituent BNp and $\mathbf{i}$ the gene state vector. If there are m contexts and n genes, then the size of the transition matrix is $m2^n \times m2^n$. A one-bit perturbation for input state $(s, \mathbf{k})$ corresponds to a change in network function $\mathbf{f}_s$, so that exactly two columns are changed, both being columns of the matrix corresponding to the sth constituent BNp. Hence, we can proceed in the same manner as with a rank-one perturbation for a BNp using Equation (23). To change d constituent BNp network functions for input gene state $\mathbf{k}$ means one-bit perturbations to d states, $(s_1, \mathbf{k}), (s_2, \mathbf{k}), \ldots, (s_d, \mathbf{k})$, in the Markov chain. Relative to the matrix this means multiple rank-one perturbations caused by d perturbations. This can be solved analogously as in the multi-row situation with BNps.

The objective of intervention is to keep cells away from certain states (for example, metastatic cancerous phenotypes). The goal of structural intervention is to alter the long-run dynamics of the network by altering the rule structure. Given a set $\mathscr{U}$ of undesired states, we want to minimise $\sum_{\mathbf{i} \in \mathscr{U}} \hat{\pi}_i$ over all structural changes within a given class. This can be done analytically using the fact that $\tilde{\pi}_i$ can be expressed in terms of π_i, the difference vector for the k-th rows of the transition matrices, and the fundamental matrix (Equation (27)). One can constrain the optimisation in many ways, for instance, by limiting the amount of mass in any given state. One might also put limits on the change in mass

$|\hat{\pi}_i - \pi_i|$ of some states, so that there is no over concentration of mass in certain states of the altered network.

Before proceeding to an illustration of structural intervention in a biological network, we note that the perturbation theory for PBNs can be used to characterise the long-run sensitivity of a PBN to structural perturbations and that this long-run sensitivity is related to the inference and controllability of PBNs (Qian and Dougherty 2009).

6. Structural intervention in a melanoma network

Here, we consider a 7-gene Boolean network constructed from the data of (Bittner et al. 2000) using the Bayesian connectivity-based approach of (Zhou, et al. 2004). The network contains the genes WNT5A, pirin, S100P, RET1, MART1, HADHB and STC2, which we label as $x_1, x_2, \ldots, x_7$, respectively. The regulatory rules inferred in (Pal, Datta, Bittner, and Dougherty 2005) are given in Table 1, where the i-th bit of the output binary string represents the output value for the i-th input predictor in binary representation. For example, in the last row of Table 1, 1101 means that, for the input predictors 00, 01, 10, 11, f_7 has the output 1, 1, 0 and 1. We let the gene perturbation probability be $p = 0.001$.

If we are only interested in steady-state mass, then the objective function is chosen to minimise that mass for the states with $x_1 = 1$ (WNT5A expressed), namely, we minimise $\sum_{x_1=1} \tilde{\pi}(\mathbf{x})$. This actually leads to an obvious intervention strategy: perturb the output from 1 to 0 for the predictor $x_6 = 0$ of f_1. For this perturbation, $\sum_{x_1=1} \tilde{\pi}(\mathbf{x}) = 0.0010$. Figure 2(a) shows the original steady-state distribution and Figure 2(b) shows the steady-state distribution resulting from the structural intervention.

While the preceding strategy appears obvious, it has the effect of significantly altering the attractor structure. The original network has four attractors possessing the following steady-state probabilities: $\pi(0101111) = 0.3645$, $\pi(0110110) = 0.0260$, $\pi(0111110) = 0.1301$ and $\pi(1000001) = 0.4662$. Since we desire the down-regulation of WNT5A, the attractor 1000001 is undesirable. This is accomplished by the intervention, with $\tilde{\pi}(1000001) = 0.0005$. But the intervention introduces a new attractor, 0000001, with $\tilde{\pi}(0000001) = 0.4677$. If we lack knowledge of this phenotype, although the overall result is beneficial from the standpoint of WNT5A, we may not want such a large mass associated with a little known phenotype.

Suppose we take a different approach. Since we want to down-regulate WNT5A, the attractor 1000001 is undesirable and we want to reduce its steady-state probability but, in addition, suppose we do not want to introduce other attractors that might lead to unexpected cellular behaviour. Then we can maximise the following objective function $\tilde{\pi}(0101111) + \tilde{\pi}(0110110) + \tilde{\pi}(0111110)$. The optimal function perturbation in this case is to perturb the output from 0 to 1 for the predictor $x_4x_6x_7 = 011$ of f_4. The perturbed steady-state probabilities are $\tilde{\pi}(0101111) = 0.6921$, $\tilde{\pi}(0110110) = 0.0495$, $\tilde{\pi}(0111110) = 0.2472$ and $\tilde{\pi}(1000001) = 0.0007$. We can see that the steady-state mass for the rest of states is equal to 0.0105, which appears to be a safe intervening strategy. Figure 2(c) illustrates that the steady-state distribution shifts from undesirable to desirable states after this one-bit function perturbation. This shift is permanent since the perturbation changes the underlying structure of the BNp.

Table 1. Definitions of Boolean functions for the WNT5A BNp.

Function	Input variables	Output
f_1	x_6	10
f_2	x_2, x_4, x_6	00010111
f_3	x_3, x_4, x_7	10101010
f_4	x_4, x_6, x_7	00001111
f_5	x_2, x_5, x_7	10101111
f_6	x_2, x_3, x_4	01110111
f_7	x_2, x_7	1101

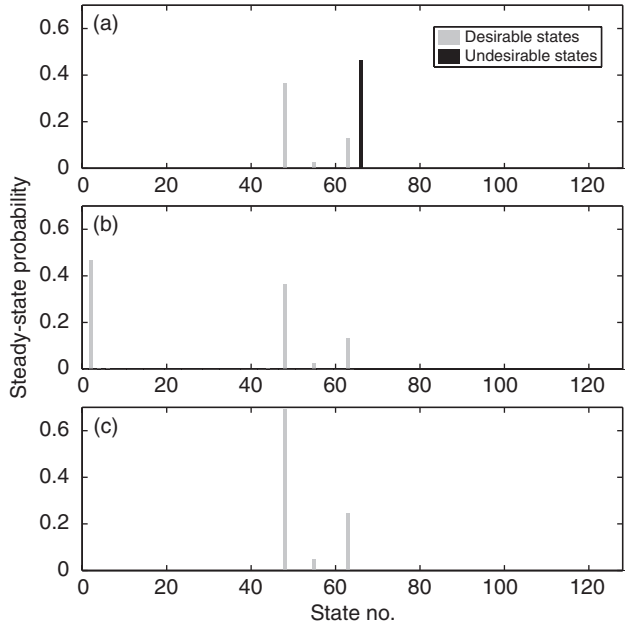


Figure 2. Steady-state distribution shifts for BNps with $p = 0.001$: (a) original steady-state distributions; (b) after optimal perturbation to f_1; (c) after optimal perturbation to f_4.

We can consider many potential intervention strategies because the analytic results lead to an efficient search for the optimal function perturbation. In addition, it is easy to find optimal practical strategies by looking into the subset of biologically admissible intervening strategies. Another point to notice is that the entire process of finding an optimal function perturbation requires computation of a single matrix inverse – the fundamental matrix of the original network. Thus, the efficiency of finding an optimal strategy in one-bit function perturbation is determined by the capability of computing the matrix inverse.

7. Conclusion

In this article, we have reviewed the fundamental aspects of stationary and structural intervention in Markovian gene regulatory networks, in particular, probablistic Boolean networks. As noted at the outset, various issues regarding regulatory intervention have been addressed in a number of papers. Many challenging issues need to be addressed before mathematically derived intervention strategies become a therapeutic reality. Three major areas of concern are model inference, computational complexity and implementation.

Here we have assumed the model, in this case, the PBN, or at least its associated Markov chain, is known. Although various approaches have been taken for model inference, these remain problematic owing to the lack of appropriate time-course data to infer the regulatory rules and probabilistic structure of a PBN, or even the state-to-state transition probabilities. Not only does satisfactory inference of a PBN require a large amount of data (Marshall, Yu, Xiao, and Dougherty 2007), but current technology is incapable of obtaining the necessary time-course measurements. This has resulted in methods that infer PBNs from steady-state data, a severely ill-posed inverse problem, under the assumption that a single, or family of, inferred networks will facilitate derivation of an intervention strategy that is beneficial for the actual network, say via a robust control policy (Pal, Datta, and Dougherty 2008, 2009).

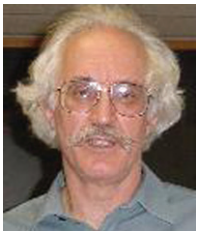
Regarding complexity, the computational burden of dynamic programming severely limits the size of the networks considered, as well as any effort to use finer quantisation (Akutsu, Hayashida, Ching, and Ng 2007). Full-model optimisation has been limited to 10 genes using binary quantisation. A number of approaches have been considered that in one way or another approximate explicit, full optimisation. These include a linear optimisation model to reduce computational complexity (Ng, Zhang, Ching, and Akutsu 2006), control when the states are not observable, thereby resulting in imperfect information for state feedback (Datta, Choudhary, Bittner, and Dougherty 2004), design of the control policy on an instantaneously random PBN obtained from a context-sensitive PBN via state reduction (Pal et al. 2005; Faryabi et al. 2009), and approximate policies designed via reinforcement learning (Faryabi, Datta, and Dougherty 2007). A promising approach to reduce both computational complexity and data requirements for inference is to forego the full model and use a greedy procedure only dependent on certain network statistics, for instance, mean first passage times (Vahedi, Faryabi, Chamberland, Datta, and Dougherty 2008). If one should have the full model, with too many genes for full optimizsation, one can reduce the size of the network (state space) by eliminating genes or states (Ivanov and Dougherty 2004; Ivanov, Pal, and Dougherty 2007). Active research involves linking this reduction to the control objectives (Ghafarri, Ivanov, and Dougherty 2008; Qian and Dougherty 2009).

Lastly, let us note that there are practical issues related to implementation. One must have some method to control the expression of a gene (or genes) to implement the desired intervention strategy. For instance, one might block a gene product. Such intervention will not mirror the simple modelling assumption that a gene's expression can be precisely controlled, either in amount or duration, as assumed in the derivation of stationary policies herein, nor will it necessarily be the case that the effects of a drug will be limited to the control gene. Modelling these practical complexities significantly increases the mathematical difficulties involved in obtaining intervention strategies, but they must be treated. Two directions in which practical therapeutic issues have been addressed is with regard to constraining the intervention so as to limit the expected number of interventions (Ching, et al. 2009; Faryabi, Vahedi, Chamberland, Datta, and Dougherty 2008) or to require a recovery period following intervention (Vahedi, Faryabi, Chamberland, Datta, and Dougherty 2009), and to treat intervention in asynchronous gene regulatory networks (Faryabi, Chamberland, Vahedi, Datta and Dougherty 2008).

Ultimately, the successful application of mathematical network intervention strategies depends upon the investigative procedure utilised for translational science (Dougherty 2009). The mathematical analysis must try to get as close as possible to the physical processes and therapeutic methods and, just as importantly, experiments need to be guided by our mathematical understanding of dynamical systems and their control. If data are not gathered in a way that can be used to infer useful models or if models are not

developed with an eye towards potential experimental protocols, then the full potential for translational science will not be achieved.

Notes on contributors



Edward R. Dougherty is a Professor in the Department of Electrical and Computer Engineering at Texas A&M University in College Station, TX, where he holds the Robert M. Kennedy 26 Chair in Electrical Engineering and is Director of the Genomic Signal Processing Laboratory. He is also the Director of the Computational Biology Division of the Translational Genomics Research Institute in Phoenix, AZ, and an Adjunct Professor in the Department of Pathology of the University of Texas MD Anderson Cancer Center in Houston, TX. He holds a PhD in mathematics from Rutgers University and an MS in Computer Science from Stevens Institute of Technology, and has been awarded the *Doctor Honoris Causa* by the Tampere University of Technology in Finland. He is a fellow of SPIE, has received the SPIE President's Award and served as the editor of the SPIE/IS&T Journal of Electronic Imaging. At Texas A&M he has received the Association of Former Students Distinguished Achievement Award in Research, been named Fellow of the Texas Engineering Experiment Station and named Halliburton Professor of the Dwight Look College of Engineering. Prof. Dougherty is author of 14 books, editor of 5 others, and author of more than 200 journal papers.

Ranadip Pal received the B Tech. degree in Electronics and Electrical Communication Engineering from the Indian Institute of Technology, Kharagpur, India in 2002 and the MS and PhD degrees in Electrical Engineering from Texas A&M University, College Station in 2004 and 2007, respectively. In August 2007, he joined the Department of Electrical and Computer Engineering at Texas Tech University, Lubbock as an Assistant Professor. His research areas are computational biology, genomic signal processing, control of stochastic networks. While at Texas A&M, Pal was the recipient of the Ebensberger/Fouraker Fellowship, a Distinguished Graduate Student Masters Research Award and a National Instruments Fellowship. Pal was also an Indian National Math Olympiad Awardee and was ranked 1st in the Regional Math Olympiad, West Bengal, India.

Xiaoning Qian received the PhD degree in electrical engineering from Yale University, New Haven, CT, in 2005. Currently, he is in the Bioinformatics Training Program, Texas A&M University, College Station. His current research interests include biomedical image analysis, genomic signal processing and computational biology.

Michael M. Bittner was initially trained as a biochemical geneticist, studying phage replication and bacterial transposition with a variety of biochemical and bacterial genetic methods at Princeton University (PhD), the Washington University School of Medicine and the Population and Molecular Genetics Department of the University of Georgia (postdoctoral). Efforts since that time have concentrated on the practical application of knowledge about the control systems operating in prokaryotes and eukaryotes. At Monsanto Corp., methods for the biologic production of peptides and proteins useful in human medicine and agriculture were developed. At Amoco Corp. methods for producing, in yeast, small molecule precursors of vitamins of human and veterinary pharmacologic interest were developed. During this time, efforts to devise cytogenetic molecular diagnostics based on *in situ* hybridization produced a series of biological, biochemical and biophysical technologies that led to the founding of the spin-off molecular diagnostics company, Vysis. Recent efforts at the National Institutes of Health and the Translational Genomics Research Institute focus on developing ways of making accurate measures of the transcriptional and genomic status of cells and analytic tools that allows inferences to be drawn from these measures that provide insight into the cellular processes operating in healthy and diseased cells.

Aniruddha Datta received the B Tech degree in Electrical Engineering from the Indian Institute of Technology, Kharagpur in 1985, the MSEE degree from Southern Illinois University, Carbondale in 1987 and the MS (Applied Mathematics) and PhD degrees from the University of Southern California in 1991. In August 1991, he joined the Department of Electrical and Computer Engineering at Texas A&M University where he is currently a Professor and holder of the J. W. Runyon, Jr '35 Professorship II. His areas of interest include adaptive control, robust control, PID control and Genomic Signal Processing. He has authored or coauthored 5 books and over 100 journal and conference papers on these topics. He is a Fellow of IEEE, has served as an Associate Editor of the *IEEE Transactions on Automatic Control* from 2001 to 2003, the *IEEE Transactions on Systems, Man and Cybernetics-Part B* from 2005 to 2006 and is currently serving as an Associate Editor of the *EURASIP Journal on Bioinformatics and Systems Biology*.

References

Akutsu, T., Hayashida, M., Ching, W.K., and Ng, M.K. (2007), 'Control of Boolean Networks: Hardness Results and Algorithms for Tree Structured Networks', *Theoretical Biology*, 244(4), 670–679.

Bertsekas, D.P. (2001), *Dynamic Programming and Optimal Control* (Vols 1 & 2), Belmont, MA: Athena Scientific.

Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., and Sondak, V. (2000), 'Molecular Classification of Cutaneous Malignant Melanoma by Gene Expression Profiling', *Nature*, 406(6795), 536–540.

Brun, M., Dougherty, E.R., and Shmulevich, I. (2005), 'Steady-state Probabilities for Attractors in Probabilistic Boolean Networks', *EURASIP Journal on Signal Processing*, 85(10), 1993–2013.

Ching, W.-K., Zhang, S.-Q., Jiao, Y., Akutsu, T., Tsing, N.-K., and Wong, A.S. (2009), 'Optimal Control Policy for Probabilistic Boolean Networks with Hard Constraints', *IET Systems Biology*, 3(2), 90–99.

Cho, G.E., and Meyer, C.D. (2001), 'Comparison of Perturbation Bounds for the Stationary Distribution of Markov chains', *Linear Algebra and its Applications*, 335, 137–150.

Datta, A., and Dougherty, E.R. (2007), *Introduction to Genomic Signal Processing with Control*, New York: CRC Press.

Datta, A., Choudhary, A., Bittner, M.L., and Dougherty, E.R. (2003), 'External Control in Markovian Genetic Regulatory Networks', *Machine Learning*, 52, 169–191.

Datta, A., Choudhary, A., Bittner, M.L., and Dougherty, E.R. (2004), 'External Control in Markovian Genetic Regulatory Networks: The Imperfect Information Vase', *Bioinformatics*, 20(6), 924–930.

Dougherty, E.R. (2009), 'Translational Science: Epistemology and the Investigative Process', *Current Genomics*, 10(2), 102–109.

Dougherty, E.R., and Datta, A. (2005), 'Genomic Signal Processing: Diagnosis and Therapy', *IEEE Signal Processing Magazine*, 22(1), 107–112.

Faryabi, B., Datta, A., and Dougherty, E.R. (2007), 'On Approximate Stochastic Control in Genetic Regulatory Networks', *IET Systems Biology*, 1(6), 361–368.

Faryabi, B., Chamberland, J.-F., Vahedi, G., Datta, A., and Dougherty, E.R. (2008), 'Optimal Intervention in Asynchronous Genetic Regulatory Networks', *IEEE Journal of Selected Topics in Signal Processing*, 2(3), 412–423.

Faryabi, B., Vahedi, G., Chamberland, J.-F., Datta, A., and Dougherty, E.R. (2008), 'Optimal Constrained Intervention in Genetic Regulatory Networks', *EURASIP Journal on Bioinformatics and Systems Biology*, Article ID 620767, May 2008.

Faryabi, B., Vahedi, G., Chamberland, J-F., Datta, A., and Dougherty, E.R. (2009), 'Intervention in Context-sensitive Probabilistic Boolean Networks Revisited', *EURASIP Journal of Bioinformatics and Systems Biology*, Article ID 360864, 13 pages.

Ghafarri, N., Ivanov, I., and Dougherty, E.R. (2008), 'Reduction Mappings and Control Policies for Intervention in Boolean Networks', in *IEEE International Workshop on Genomic Signal Processing and Statistics*, Phoenix.

Hunter, J.J. (1986), 'Stationary Distributions of Perturbed Markov Chains', *Linear Algebra and its Applications*, 82, 201–214.

Hunter, J.J. (2005), 'Stationary Distributions and Mean First Passage Times of Perturbed Markov Chains', *Linear Algebra and its Applications*, 410, 217–243.

Ivanov, I., and Dougherty, E.R. (2004), 'Reduction Mappings Between Probabilistic Boolean Networks', *EURASIP Journal on Applied Signal Processing*, 4(1), 125–131.

Ivanov, I., Pal, R., and Dougherty, E.R. (2007), 'Dynamics Preserving Size Reduction Mappings for Probabilistic Boolean Networks', *IEEE Transactions on Signal Processing*, 55(5), 2310–2322.

Kauffman, S.A. (1969), 'Metabolic Stability and Epigenesis in Randomly Constructed Genetic Nets', *Theoretical Biology*, 22, 437–467.

Kauffman, S.A. (1993), *The Origins of Order*, New York: Oxford University Press.

Kemeny, J.G., and Snell, J.L. (1960), *Finite Markov Chains*, New York: Van Nostrand.

Marshall, S., Yu, L., Xiao, Y., and Dougherty, E.R. (2007), 'Inference of Probabilistic Boolean Networks from a Single Observed Temporal Sequence', *EURASIP Journal on Bioinformatics and Systems Biology*, Article ID 32454, 15 pages.

Ng, M.K., Zhang, S.Q., Ching, W.K., and Akutsu, T. (2006), 'A Control Model for Markovian Genetic Regulatory Networks', *Transactions on Computational and System Biology*, LNBI 4070, 36–48.

Pal, R., Datta, A., Bittner, M.L., and Dougherty, E.R. (2005), 'Intervention in Context-sensitive Probabilistic Boolean Networks', *Bioinformatics*, 21, 1211–1218.

Pal, R., Ivanov, I., Datta, A., Bittner, M.L., and Dougherty, E.R. (2005), 'Generating Boolean Networks with a Prescribed Attractor Structure', *Bioinformatics*, 21, 4021–4025.

Pal, R., Datta, A., and Dougherty, E.R. (2006), 'Optimal Infinite horizon Control for Probabilistic Boolean Networks', *IEEE Transactions and Signal Processing*, 54(6-2), 2375–2387.

Pal, R., Datta, A., and Dougherty, E.R. (2008), 'Robust Intervention in Probabilistic Boolean Networks', *IEEE Transactions and Signal Processing*, 56(3), 1280–1294.

Pal, R., Datta, A., and Dougherty, E.R. (2009), 'Bayesian Robustness in the Control of Gene Regulatory Networks', *IEEE Transactions Signal Processing*, in press.

Qian, X., and Dougherty, E.R. (2008), 'Effect of Function Perturbation on the Steady-state Distribution of Genetic regulatory Networks: Optimal Structural Intervention', *IEEE Transactions and Signal Processing*, 56(10-1), 4966–4975.

Qian, X., and Dougherty, E.R. (2009), 'On the Long-run Sensitivity of Probabilistic Boolean Networks', *Journal of Theoretical Biology*, 257(4), 560–577.

Qian, X., and Dougherty, E.R. (2009), 'Control-compatible State Reduction for Boolean Networks', in *IEEE*

International Workshop on Genomic Signal Processing and Statistics, Minneapolis.

Schweitzer, P.J. (1968), 'Perturbation Theory and Finite Markov Chains', *Journal of Applied Probability*, 5, 401–413.

Shmulevich, I., and Dougherty, E.R. (2007), *Genomic Signal Processing*, Princeton: Princeton University Press.

Shmulevich, I., Dougherty, E.R., Kim, S., and Zhang, W. (2002), 'Probabilistic Booleannetworks: A Rule-based Uncertainty Model for Gene Regulatory Networks', *Bioinformatics*, 18, 261–274.

Shmulevich, I., Dougherty, E.R., and Zhang, W. (2002a), 'From Boolean to Probabilistic Boolean Networks as Models of Genetic Regulatory Networks', *Proceedings of the IEEE*, 90, 1778–1792.

Shmulevich, I., Dougherty, E.R., and Zhang, W. (2002b), 'Gene Perturbation and Intervention in Probabilistic Boolean Networks', *Bioinformatics*, 18, 1319–1331.

Shmulevich, I., Dougherty, E.R., and Zhang, W. (2002c), 'Control of Stationary Behaviour in Probabilistic Boolean Networks by Means of Structural Intervention', *Biological Systems*, 10(4), 431–446.

Vahedi, G., Faryabi, B., Chamberland, J-F., Datta, A, and Dougherty, E.R. (2008), 'Intervention in Gene Regulatory Networks Via a Stationary Mean-first-passage-time Control Policy', *IEEE Transactions on Biomedical Engineering*, 15(10), 2319–2331.

Vahedi, G., Faryabi, B., Chamberland, J.-F., Datta, A., and Dougherty, E.R. (2009), 'Optimal Intervention Strategies for Cyclic Therapeutic Methods', *IEEE Transactions on Biomedical Engineering*, 56(2), 281–291.

Weeraratna, A.T., Jiang, Y., Hostetter, G., Rosenblatt, K., Duray, P., Bittner, M., and Trent, J.M. (2002), 'Wnt5a Signalling Directly Affects Cell Motility and Invasion of Metastatic Melanoma', *Cancer Cell*, 1, 279–288.

Xiao, Y., and Dougherty, E.R. (2007), 'The impact of Function Perturbations in Boolean Networks', *Bioinformatics*, 23(10), 1265–1273.

Zhou, X., Wang, X., Pal, R., Ivanov, I., Bittner, M.L., and Dougherty, E.R. (2004), 'A Bayesian Connectivity-based Approach to Constructing Probabilistic Gene Regulatory Networks', *Bioinformatics*, 20, 2918–2927.