Optimal Intervention Strategies for Cyclic Therapeutic Methods

Golnaz Vahedi*, *Student Member, IEEE*, Babak Faryabi, *Student Member, IEEE*, Jean-Francois Chamberland, *Member, IEEE*, Aniruddha Datta, *Senior Member, IEEE*, and Edward R. Dougherty, *Member, IEEE*

***Abstract*—External control of a genetic regulatory network is used for the purpose of avoiding undesirable states such as those associated with a disease. Certain types of cancer therapies, such as chemotherapy, are given in cycles with each treatment being followed by a recovery period. During the recovery period, the side effects tend to gradually subside. In this paper, it is shown how an optimal cyclic intervention strategy can be devised for any Markovian genetic regulatory network. The effectiveness of optimal cyclic therapies is demonstrated through numerical studies for random networks. Furthermore, an optimal cyclic policy is derived to control the behavior of a regulatory model of the mammalian cell-cycle network.**

***Index Terms*—Cyclic therapy, dynamic programming, genetic regulatory networks, probabilistic Boolean networks (PBNs), stochastic optimal control.**

I. Introduction

SUCCESSFUL treatment of bacterial infections is largely a result of our ability to exploit the biochemical differences between bacteria and human cells so as to achieve toxic drug concentrations in the former while sparing the latter. Unfortunately, such high selectivity is at present elusive in the chemotherapy of human cancers. Hence, great efforts are required to determine dose schedules that maximize the benefit to toxicity ratio in cancer treatment [1]. To this end, chemotherapy is generally given in *cycles*: each treatment is followed by a *recovery phase*. During a recovery period, the side effects tend to gradually subside. Dose intensity is a measure of chemotherapy delivery that looks at the amount of drug delivered per unit of time. A higher drug dose intensity can be delivered by increasing the dose per cycle (dose escalation) or by reducing the interval between cycles (dose density).

For a given integrated drug effect, the chance of eradicating the tumor is maximized by delivering the most effective dose level of drug over as short a time as possible. Tumors given less time to grow between treatments are more likely to be eradicated. Administering high quantities of drugs at the beginning of a chemotherapy cycle might fail for two reasons. First, levels higher than a certain concentration may not increase the killing rate of cancer cells. Second, even if they did, the toxicity could be intolerable to the patient. In practice, optimizing the schedule means determining a way to give the maximum integrated effect over as short a time as possible, consistent with reasonable quality of life [1].

A prime objective of modeling genetic regulatory networks is to develop therapies based on gene regulation, in particular, the disruption or mitigation of aberrant gene function contributing to the pathology of a disease. Engineering therapeutic tools involve synthesizing nonlinear dynamical networks, analyzing these networks to characterize gene regulation, and developing intervention strategies to modify dynamical behaviors [2]. In this paper, we derive an optimal cyclic intervention strategy for gene regulation in the context of probabilistic Boolean networks (PBNs).

PBNs, a class of stochastic models for gene regulatory networks [3], have recently received considerable attention in the literature. The dynamic behavior of PBNs can be studied within the context of Markov chains [3]. Each state of the Markov chain basically represents the concatenation of all the quantized gene values usually referred to as the gene activity profile (GAP). To date, effective intervention strategies have been studied for PBNs. Initial efforts have focused on manipulating external (control) variables to desirably affect system evolution over a finite-time horizon [4]. These short-term policies have been shown to possess the ability to change the dynamical behavior of regulatory networks over a small number of stages. However, they are not effective in changing the long-run behaviors. To address this issue, stochastic control policies have been designed to affect the long-run network behaviors of regulatory networks [5]. For a predefined cost of intervention and cost of undesirable states at each stage, the objective is to find a control strategy that minimizes the expected total discounted cost in the long run. The optimal control policy with regard to the cost values can be found via dynamic programming. It is worthwhile to point out that solving the stochastic control problem fails when the number of genes present in the

Manuscript received February 1, 2008; revised June 14, 2008. First published August 15, 2008; current version published March 25, 2009. This work was supported in part by the National Science Foundation under Grant ECS-0355227, Grant CCF-0514644, and Grant ECS-0701531, in part by the National Cancer Institute under Grant R01 CA-104620, and in part by the Translational Genomics Research Institute. *Asterisk indicates corresponding author.*

*G. Vahedi is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: golnaz@tamu.edu).

B. Faryabi, J.-F. Chamberland, and A. Datta are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: bfariabi@tamu.edu; chmbrlnd@tamu.edu; datta@ece.tamu.edu).

E. R. Dougherty is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA, and also with the Translational Genomics Research Institute, Phoenix, AZ 85004 USA (e-mail: edward@ece.tamu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2008.2003092

network goes beyond relatively modest numbers. To mitigate this problem and bypass the impediment of model estimation, model-free methods based on reinforcement learning [6] and mean first-passage times [7] have also been introduced.

For hitherto intervention strategies [4]–[7], at every state transition of the system, the intervention strategy dictates whether to apply treatment or not. In this paper, our objective is to devise an effective intervention strategy under the constraint that intervention is permitted only every W transitions, where $W \in \mathbb{N}$ denotes the length of the recovery period. An intervention strategy that is optimal for the case where intervention is permitted at every transition is not necessarily optimal (i.e., may not minimize the expected total discounted cost) if one is only permitted to apply treatment every W transitions. We will refer to a policy that is optimal when intervention is permitted every transition as an *optimal one-transition policy*. Similarly, we refer to the policy that is optimal when intervention is permitted every W transitions as an *optimal W-transition policy*.

We define a *treatment window* to be every W transitions of the system. Intervention is permitted at the beginning of a treatment window. Thereafter, the system transitions $W-1$ steps without intervention. To incorporate the cyclic constraint on interventions, we construct a Markov chain with an augmented state space based on the original Markov chain. An optimal cyclic intervention policy, i.e., optimal W-transition policy, can be found by solving the stochastic control problem for the Markov chain with the augmented state space via dynamic programming algorithms. However, this procedure maybe prohibitive due to the size of the augmented state space. We show that the augmented state space can be collapsed resulting in a compressed space of size equal to the original state space. We accomplish this reduction in the size of the state space by accumulating the expected cost of the system progressing during a period. The new cost function is used to select the proper action when intervention is permitted. We establish the convergence of the dynamic programming algorithm and show how the optimal W-transition intervention strategy can be found. Furthermore, we compare the performance of an optimal W-transition policy to that of an optimal one-transition policy when intervention is applied every W transitions. We show that although this may not be true in general, in our intervention framework, optimal one-transition policy can be used as an approximation of optimal W-transition policy.

Our focus in this paper is on binary PBNs. However, the proposed procedure for devising an optimal W-transition strategy applies without change to a PBN having any discrete range of values. The difficulty with PBNs possessing more than two values per gene is that the size of the state space increases dramatically. Conceptually, networks with finer quantization can be analyzed using the same tools; indeed, the original application of automatic control considered a ternary network arising from cDNA-microarray data quantized into three values: -1 (down-regulated), $+1$ (up-regulated), and 0 (invariant) [4]. More generally, the proposed procedure can be applied to any Markovian regulatory network and should be mathematically viewed in this manner. For instance, one can devise an optimal W-transition intervention strategy for dynamic Bayesian networks (DBNs) [8]. The proposed method can actually be viewed for DBNs in the same manner as with PBNs because any DBN can be represented by a probabilistically equivalent PBN [9].

This paper is structured in the following manner. Necessary definitions are provided in Section II. We show how to derive an optimal W-transition control policy in Section III. In Section IV, we investigate the performance of optimal one-transition and W-transition policies on synthetic networks with various properties. Furthermore, we apply the proposed intervention method to control a regulatory model of the mammalian cell-cycle network.

II. Background

A. Probabilistic Boolean Network

An instantaneous PBN consists of a sequence $V = \{x_i\}_{i=1}^{n}$ of n nodes, where $x_i \in \{0, \ldots, d-1\}$, together with a sequence $\{\mathbf{f}_l\}_{l=1}^{k}$ of vector-valued functions called *predictor functions*. In the framework of gene regulation, each element x_i represents the expression level of a gene. It is common to mix the terminology by referring to x_i as the ith gene. Each vector-valued function $\mathbf{f}_l = (f_{l1}, \ldots, f_{ln})$ determines a constituent network of the PBN. The function $f_{li} : \{0, \ldots, d-1\}^n \rightarrow \{0, \ldots, d-1\}$ is the predictor of gene i, whenever network l is selected. The number of quantization levels is denoted by d. At each step, a predictor function is randomly selected according to probability distribution $\{p_l\}_{l=1}^{k}$. After selecting the predictor function $\mathbf{f}_l$, the values of the genes are updated accordingly; that is, in conformity with the network determined by $\mathbf{f}_l$. We consider PBNs with perturbation in which each gene may change its value with a small perturbation probability p at each time unit. The dynamics of a PBN can be represented via a Markov chain and, as a consequence of the perturbation, this Markov chain is ergodic and possesses a steady-state distribution.

Two quantization levels have thus far been used in practice. If $d = 2$ (binary), then the constituent networks are Boolean networks with 0 or 1 meaning OFF or ON, respectively. The case $d = 3$ (ternary) arises when we consider a gene to be -1 (down-regulated), $+1$ (up-regulated), and 0 (invariant). This situation commonly occurs with cDNA microarrays, where a ratio is taken between the expression values on the test channel (red) and the base channel (green). In this paper, we will develop the methodology for $d = 2$, so that gene values are either 0 or 1. However, the methodology is applicable to any finite number of levels.

The GAP is an n-digit binary vector $\boldsymbol{x}(t) = (x_1(t), \ldots, x_n(t))$ giving the expression values of the genes at time t, where $x_i(t) \in \{0, \ldots, d-1\}$. We note that there is a natural bijection between the GAP $\boldsymbol{x}(t)$ and its decimal representation, which takes values in $\mathcal{W} = \{0, 1, \ldots, d^n - 1\}$.

In the presence of external controls, we suppose that the PBN has m binary control inputs, $\{c_i(t)\}_{i=1}^{m}$, which specify the interventions on control genes $g_1, \ldots, g_m$. A control $c_i(t)$, which can take values 0 or 1 at each updating epoch t, specifies the action on the control gene g_i. The decimal bijection of the control vector, $u_{g_1,\ldots,g_m}(t) \in \mathcal{C} = \{0, 1, \ldots, 2^m - 1\}$, describes

the complete status of all the control inputs. As in previous applications, we focus on a single control gene, $u_g(t) \in \mathcal{C} = \{0, 1\}$. Treatment alters the status of the control gene g, which can be selected among all the genes in the network. If the control at updating epoch t is on, $u_g(t) = 1$, then the state of the control gene g is toggled; if $u_g(t) = 0$, then the state of the control gene g remains unchanged. We assume that the control gene g is given, and we refer to $u_g(t)$ as $u(t)$ in the rest of this paper.

Brun *et al.* showed that the dynamic behavior of a PBN can be modeled by a Markov chain [10]. In this case, system evolution is represented by a stationary discrete-time equation

$$z(t+1) = f(z(t), u(t), w(t)), \qquad \text{for } t = 0, 1, \ldots$$

where state $z(t)$ is an element of the state-space $\mathcal{S} = \{(c, s) : c \in \{1, \ldots, k\} \wedge s \in \{0, 1, \ldots, d^n - 1\}\}$. It should be noted that k is the total number of predictor functions. In this paper, we denote the cardinality of $\mathcal{S}$ by N.

The disturbance $w(t)$ is the manifestation of uncertainties in the PBN. It is assumed that both the gene perturbation distribution and the network switching distribution are independent and identical for all time steps t. Originating from a state i, the successor state j is selected randomly within set $\mathcal{S}$ according to the transition probability $p_{ij}(u)$

$$p_{ij}(u) \stackrel{\triangle}{=} P(z(t+1) = j \,|\, z(t) = i, u(t) = u)$$

for all i and j in $\mathcal{S}$, and for all u in $\mathcal{C}$. Gene perturbation insures that all the states in the Markov chain communicate with one another. Hence, the finite-state Markov chain has a unique steady-state distribution [11].

B. Optimal Intervention for PBNs

In cancer, one can consider the correlation between metastasis and the abundances of mRNA for certain genes [12]. A gene, which associates with metastasis, is called a *target gene*. Such a gene can be used to partition the state space into subsets of desirable and undesirable states, $\mathcal{D}$ and $\mathcal{U}$, respectively.

Pal *et al.* employed Markov decision processes to devise an optimal intervention strategy that can alter the likelihood of undesirable states in the long run [5]. They assumed that intervention is allowed at every transition. In the following, we summarize their approach to derive an optimal one-transition policy. A cost-per-stage $g(i, j, u)$ is associated to each transition in the system. In general, the cost-per-stage may depend on the origin state i, the successor state j, and the control input u. We assume that the cost-per-stage is stationary and bounded for all , in $\mathcal{S}$, and u in $\mathcal{C}$. We define the expected immediate cost in state i, when control u is selected, by

$$\overline{g}(i, u) = \sum_{j \in \mathcal{S}} p_{i,j}(u)\, g(i, j, u).$$

We consider a discounted formulation of the expected total cost. The discounting factor, $\lambda \in (0, 1)$, ensures the convergence of the expected total cost over the long run [13]. In the case of cancer therapy, the discounting factor emphasizes that obtaining treatment at an earlier stage is favored over later stages. The expected total discounted cost given policy π and initial state i is denoted by

$$J_\pi(i) = \lim_{N \to \infty} E\left\{ \sum_{t=0}^{N-1} \lambda^t\, g\left(z(t), z(t+1), \mu(z(t))\right) | z(0) = i \right\} \tag{1}$$

where state $z(t)$ is an element of the state space $\mathcal{S}$ at time t. A policy $\pi = \{\mu(0), \mu(1), \ldots\}$ is a sequence of decision rules $\mu(t) : \mathcal{S} \to \mathcal{C}$, for each time step t. The vector $\mathbf{J}_\pi$ of the expected total discounted costs is called the value function. In this stochastic control problem, we seek an intervention strategy π^* among all the admissible intervention strategies Π that minimizes the value function for each state i

$$\pi^*(i) = \arg \min_{\pi \in \Pi} J_\pi(i), \qquad \forall\, i \in \mathcal{S}. \tag{2}$$

As we mentioned earlier, this is the optimal one-transition policy. For a finite-time horizon, the dynamic programming algorithm describes how the optimal cost J_{k+1} propagates backward in time to the optimal cost J_k

$$J_k(i) = \min_{u \in \mathcal{C}} \left[\overline{g}(i, u) + \lambda \sum_{j=0}^{N-1} p_{i,j}(u) J_{k+1}(j) \right]. \tag{3}$$

The previous equation motivates the introduction of the mapping $T : \mathcal{S} \mapsto \Re$ defined by

$$TJ(i) = \min_{u \in \mathcal{C}} \left[\overline{g}(i, u) + \lambda \sum_{j=0}^{N-1} p_{i,j}(u) J(j) \right] \qquad \forall i \in \mathcal{S} \tag{4}$$

for any value function $J : \mathcal{S} \mapsto \Re$. Given the mapping of (4), the following propositions summarize how one can devise an optimal one-transition policy. Proofs of these statements can be found in [13].

Proposition 1 (Convergence of discounted cost algorithm): For any $x \in \mathcal{S}$ and any bounded function $J : \mathcal{S} \mapsto \Re$, the optimal cost function satisfies

$$J^*(x) = \lim_{M \longrightarrow \infty} (T^M J)(x) \qquad \forall\, x \in \mathcal{S}.$$

Proposition 2 (Bellman's optimality equation): The optimal cost function J^* satisfies

$$J^* = TJ^*. \tag{5}$$

Furthermore, J^* is the unique solution of this equation within the class of bounded functions.

Proposition 3 (Necessary and sufficient condition for optimality): A stationary policy μ is optimal if and only if it attains the minimum in Bellman's optimality equation of (5).

The three aforementioned propositions provide the basis for a method for determining an optimal one-transition policy. Proposition 2 asserts that the optimal cost function satisfies Bellman's optimality equation, while Proposition 1 states that the optimal cost function can be iteratively determined by running the recursion equation

$$J_{k+1} = TJ_k, \qquad k = 0, 1, 2, \ldots \tag{6}$$

for any bounded initial cost function $J_0 : \mathcal{S} \mapsto \Re$. Since this iteration is guaranteed to converge to J^*, one can continue

the iteration until some stopping criterion is reached. By Proposition 3, the resulting optimal policy is also stationary. The procedure described in (6) is referred to as the value iteration algorithm since, at every stage, we are iterating on the value function. The optimal one-transition policy is obtained as the argument of the minimization step once the iterative procedure has converged.

III. Optimal Control Strategy for Cyclic Therapeutic Methods

Our objective is to find an effective intervention policy when we are allowed to apply treatment only every W transitions, in other words, at times $t = 0, W, 2W, \ldots$. To incorporate this cyclic constraint in our mathematical framework, we construct a Markov chain with an augmented state space based on the original Markov chain. The new (augmented) state space is defined as

$$\tilde{\mathcal{S}} = \{(i,j) \mid i \in \{0,\ldots,N-1\}, j \in \{0,\ldots,W-1\}\}$$

where N is the size of the original state space $\mathcal{S}$. There are two types of states in the augmented state space: state (i,j) with $j = 0$, represented as $(i,0)$, where intervention is permitted, and state (i,j) with $j \neq 0$, where intervention is not permitted. In the augmented state space, the control u is constrained to take values in $U(i,j)$, a given nonempty subset of $\mathcal{C}$. For the first type of states, $(i,0)$, we have $U(i,0) = \{0,1\}$, while for the second type of states, (i,j) where $j \neq 0$, we have $U(i,j) = \{0\}$.

The transition probabilities in the augmented state space are defined as a function of control u. For state $(i,0)$, we define the probability of transitioning to state (i',j') given control u as

$$p_{(i,0)(i',j')}(u) = \begin{cases} p_{i,i'}(u), & \text{if } j' = 1 \\ 0, & \text{otherwise} \end{cases}$$

where $p_{i,i'}(u)$ denotes the probability of transitioning from state i to state i' under control u. On the other hand, for states (i,j), where $j \neq 0$, control u only admits one value, $u \in \{0\}$. For these states, the transition probability is defined as

$$p_{(i,j)(i',j')}(u=0) = \begin{cases} p_{i,i'}(u=0), & \text{if } j' = (j+1) \bmod W \\ 0, & \text{otherwise} \end{cases}$$

where $p_{i,i'}(u=0) = p_{i,i'}$ denotes the uncontrolled probability of transitioning from state i to state i'. It should be noted that $(j' = (j+1) \bmod W)$ is true if either $(j' = j+1)$ or $(j = W-1$ and $j' = 0)$ is true. Considering that $u \in U(i,j)$, the probability of transitioning from state (i,j) to state (i',j') can be compactly defined as

$$p_{(i,j)(i',j')}(u) = \begin{cases} p_{i,i'}(u), & \text{if } j' = (j+1) \bmod W \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Let us now consider an example to explain how the aforementioned definition simulates the cyclic intervention scenario. Assume that at time $t = 0$, we observe state i. At this time, we are allowed to apply control $u \in \{0,1\}$. The augmented state corresponding to state i at time $t = 0$ is $(i,0)$. From augmented state $(i,0)$, under control u, the system transitions to the augmented state $(i',1)$ with probability $p_{i,i'}(u)$, where $u \in \{0,1\}$. The probability of transitioning to any other state (i',j), where $j \neq 1$, is zero. At time $t = 1$ and from augmented state $(i',1)$, the system transitions to state $(i'',2)$ with probability $p_{i',i''}(0)$ since $u \in \{0\}$. Likewise, one can consider transitions for $t = 2,\ldots,W-2$. Similarly, assume that we observe state k at time $t = W-1$. The probability of transitioning to the augmented state $(k',0)$ is $p_{k,k'}(0)$. The probability of transitioning to any other state (k',j), where $j \neq 0$, is zero.

The cost-per-stage for transitioning from augmented state (i,j) to augmented state (i',j'), given control u, is defined as

$$g(i,j,i',j',u) = \begin{cases} C+c & i' \in \mathcal{U} \text{ and } \{j=0 \text{ and } j'=1\} \text{ and } u=1 \\ C & i' \in \mathcal{U} \text{ and } \{j' = (j+1) \bmod W\} \text{ and } u=0 \\ c & i' \in \mathcal{D} \text{ and } \{j=0 \text{ and } j'=1\} \text{ and } u=1 \\ 0 & i' \in \mathcal{D} \text{ and } \{j' = (j+1) \bmod W\} \text{ and } u=0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where C and c represent the cost of undesirable states and the cost of treatment (control), respectively. Given $u = 1$, we assign a cost to a transition from state (i,j) to state (i',j') only when $j = 0$ and $j' = 1$. In this case, if i' is an undesirable state, the corresponding cost is $C + c$; if i' is a desirable state, the only cost incurred is c. When $u = 0$, it is possible to transition to (i',j') if $j' = (j+1) \bmod W$ is true. In this case, if i' is an undesirable state, the corresponding cost is C; if i' is a desirable state, no cost is incurred. For all the other cases, no cost is assigned.

Based on (8), we define the expected immediate cost at state (i,j) when control u is selected by

$$\begin{aligned} \overline{g}(i,j,u) &= \sum_{i'=0}^{N-1} \sum_{j'=0}^{W-1} p_{(i,j)(i',j')}(u) g(i,j,i',j',u) \\ &= \sum_{i'=0}^{N-1} [p_{(i,j)(i',0)}(u) g(i,j,i',0,u) \\ &\quad + \cdots + p_{(i,j)(i',W-1)}(u) g(i,j,i',W-1,u)]. \end{aligned}$$

In this equation, for each value of i', only one term inside the brackets is nonzero [based on the definition of the transition probabilities in (7)]. Hence

$$\overline{g}(i,j,u) = \sum_{i'=0}^{N-1} p_{(i,i')}(u) g(i,j,i',j',u)$$

where $j' = (j+1) \bmod W$ is true. Using the definition of $g(i,j,i',j',u)$ in (8), we have

$$\overline{g}(i,j,u) = \begin{cases} C \sum_{i' \in \mathcal{U}} p_{i,i'}(u=1) + c, & \text{if } u = 1 \\ C \sum_{i' \in \mathcal{U}} p_{i,i'}(u=0), & \text{if } u = 0. \end{cases} \tag{9}$$

From (9), it is clear that $\overline{g}(i,j,u)$ does not depend on j, i.e., $\overline{g}(i,j,u) = \overline{g}(i,u)$.

As we explained in Section II-B, the dynamic programming algorithm captures how the optimal cost at J_{k+1} propagates backward in time to the optimal cost J_k. For the augmented

state space, we have

$$J_k(i,j) = \min_{u\in U(i,j)}\left[\overline{g}(i,j,u)+\lambda\sum_{i'=0}^{N-1}\sum_{j'=0}^{W-1} p_{(i,j)(i',j')}(u)\ J_{k+1}(i',j')\right] \quad \forall (i,j)\in\tilde{\mathcal{S}}. \tag{10}$$

Since $\overline{g}(i,j,u)=\overline{g}(i,u)$, we can rewrite (10) as

$$J_k(i,j) = \min_{u\in U(i,j)}\left[\overline{g}(i,u)+\lambda\sum_{i'=0}^{N-1}\sum_{j'=0}^{W-1} p_{(i,j)(i',j')}(u)\ J_{k+1}(i',j')\right] \quad \forall (i,j)\in\tilde{\mathcal{S}}. \tag{11}$$

Our goal is to derive the value functions for the original state space, i.e., $\mathcal{S}$, based on (11). To this end, for every treatment window starting with $i\in\mathcal{S}$, we accumulate the total discounted cost of all states in the window where no control can be applied and add it to the average cost of state i. We then show how the accumulated cost at the beginning of the $(s+1)$th window affects the accumulated cost at the beginning of the sth window, where $s=0,1,2,\ldots$. This approach is in accord with the dynamic programming technique that ranks decisions based on the sum of the present cost and the expected future cost, assuming optimal decision making for subsequent stages. This manipulation of the value function lets us collapse the augmented state space $\tilde{\mathcal{S}}$ to the much smaller space $\mathcal{S}$. We prove the convergence of the discounted cost algorithm in this framework and show how an optimal W-transition control policy can be found using standard dynamic programming algorithms.

Assume that $\mathbf{P}$ is the transition probability matrix of the uncontrolled Markov chain. For $i,j\in\mathcal{S}$, let $p_{i,j}^{(r)}$ be the probability of going from state i to state j in r steps, i.e., the (i,j)th entry of the matrix $\mathbf{P}^{(r)}$. The objective is to compute the recursive relation of the value function starting at time $t=sW$, given the cost value at time $t=(s+1)W$.

Without loss of generality, we assume that $s=0$. In the augmented state space $\tilde{\mathcal{S}}$, we are not allowed to apply any control at state $(i,W-1)$, hence from (7) and (11)

$$\begin{aligned} J_{W-1}(i,W-1) &= \min_{u\in U(i,W-1)}\left\{\overline{g}(i,u)+\lambda\sum_{j=0}^{N-1} p_{i,j}(u)J_W(j,0)\right\}\\ &= \overline{g}(i,0)+\lambda\sum_{j=0}^{N-1} p_{i,j}J_W(j,0). \end{aligned} \tag{12}$$

Given J_{W-1}, one can compute J_{W-2} as

$$\begin{aligned} J_{W-2}(i,W-2) &= \min_{u\in U(i,W-2)}\left\{\overline{g}(i,u)+\lambda\sum_{j=0}^{N-1} p_{i,j}(u)J_{W-1}(j,W-1)\right\}\\ &= \overline{g}(i,0)+\lambda\sum_{j=0}^{N-1} p_{i,j}J_{W-1}(j,W-1). \end{aligned}$$

Replacing J_{W-1} from (12), we have $J_{W-2}(i,W-2)$ as a function of $J_W(k,0)$ for all $k\in\mathcal{S}$,

$$\begin{aligned} J_{W-2}(i,W-2) &= \overline{g}(i,0)+\lambda\sum_{j=0}^{N-1} p_{i,j}\left(\overline{g}(j,0)+\lambda\sum_{k=0}^{N-1} p_{j,k}J_W(k,0)\right)\\ &= \overline{g}(i,0)+\lambda\sum_{j=0}^{N-1} p_{i,j}\overline{g}(j,0)+\lambda^2\sum_{k=0}^{N-1} p_{i,k}^{(2)}J_W(k,0). \end{aligned}$$

Similarly, we can compute J_{W-3} as

$$\begin{aligned} J_{W-3}(i,W-3) &= \sum_{j=0}^{N-1}\left(p_{i,j}^{(0)}+\lambda p_{i,j}^{(1)}+\lambda^2 p_{i,j}^{(2)}\right)\overline{g}(j,0)\\ &\quad+\lambda^3\sum_{k=0}^{N-1} p_{i,k}^{(3)}J_W(k,0). \end{aligned}$$

One can recursively evaluate the value function for the last state in a treatment window where no control is allowed, i.e., $J_1(i,1)$, as follow:

$$\begin{aligned} J_1(i,1) &= \sum_{j=0}^{N-1}\left(\sum_{r=0}^{W-2}\lambda^r p_{i,j}^{(r)}\right)\overline{g}(j,0)\\ &\quad+\ \lambda^{W-1}\sum_{k=0}^{N-1} p_{i,k}^{(W-1)}J_W(k,0). \end{aligned} \tag{13}$$

Finally, at time 0, intervention is allowed and the following minimization problem leads to $J_0(i,0)$:

$$J_0(i,0) = \min_{u\in U(i,0)}\left\{\overline{g}(i,u)+\lambda\sum_{j=0}^{N-1} p_{i,j}(u)J_1(j,1)\right\}. \tag{14}$$

Using (13) and (14), we obtain

$$\begin{aligned} J_0(i,0) &= \min_{u\in U(i,0)}\left\{\overline{g}(i,u)+\lambda\sum_{j=0}^{N-1} p_{i,j}(u)\left(\sum_{k=0}^{N-1}\left(\sum_{r=0}^{W-2}\lambda^r p_{j,k}^{(r)}\right)\overline{g}(k,0)\right.\right.\\ &\quad\left.\left.+\lambda^{W-1}\sum_{k=0}^{N-1} p_{j,k}^{(W-1)}J_W(k,0)\right)\right\}. \end{aligned} \tag{15}$$

We can rewrite (15) as

$$\begin{aligned} J_{sW}(i,0) &= \min_{u\in U(i,0)}\left\{\overline{g}(i,u)+\lambda\sum_{j=0}^{N-1} p_{i,j}(u)\left(\sum_{k=0}^{N-1}\left(\sum_{r=0}^{W-2}\lambda^r p_{j,k}^{(r)}\right)\overline{g}(k,0)\right.\right.\\ &\quad\left.\left.+\lambda^{W-1}\sum_{k=0}^{N-1} p_{j,k}^{(W-1)}J_{(s+1)W}(k,0)\right)\right\} \end{aligned} \tag{16}$$

for $s=0,1,2,\ldots$. This equation reveals how the cost at the beginning of the $(s+1)$th window affects the cost at the beginning

of the sth window. This equation ranks decisions based on the sum of the present cost and the expected future cost considering the cost of all the states where no control is allowed between two treatment times. This manipulation of the value function lets us collapse the state space from $\tilde{\mathcal{S}}$ to $\mathcal{S}$ and leads to

$$J_s(i) = \min_{u\in U(i)} \left\{ \overline{g}(i,u) + \lambda \sum_{j=0}^{N-1} p_{i,j}(u) \left(\sum_{k=0}^{N-1} \left(\sum_{r=0}^{W-2} \lambda^r p_{j,k}^{(r)} \right) \overline{g}(k,0) + \lambda^{W-1} \sum_{k=0}^{N-1} p_{j,k}^{(W-1)} J_{(s+1)}(k) \right) \right\} \quad (17)$$

for $s = 0, 1, 2, \ldots$. It should be noted that the aforementioned backward propagation of costs applies to every W transitions of the Markov chain in which we are permitted to apply intervention and $U(i) = \{0, 1\}$.

Similar to Section II-B, the following proposition discusses how an optimal W-transition stationary control policy can be devised. In Proposition 4, we prove the convergence of the discounted cost algorithm as it is defined in this paper. The proof of the proposition can be found in the supplementary materials. Propositions 2 and 3 can be restated for the following operator T.

Proposition 4 (Convergence of discounted cost algorithm): For any $i \in \mathcal{S}$, bounded function $J : \mathcal{S} \mapsto \Re$, and $T : \mathcal{S} \mapsto \Re$, where

$$TJ(i) = \min_{u\in U(i)} \left\{ \overline{g}(i,u) + \lambda \sum_{j=0}^{N-1} p_{i,j}(u) \left(\sum_{k=0}^{N-1} \left(\sum_{r=0}^{W-2} \lambda^r p_{j,k}^{(r)} \right) \overline{g}(k,0) + \lambda^{W-1} \sum_{k=0}^{N-1} p_{j,k}^{(W-1)} J(k) \right) \right\} \quad (18)$$

the optimal cost function satisfies

$$J^*(x) = \lim_{j' \longrightarrow \infty} \left(T^M J \right)(x) \qquad \forall\, x \in \mathcal{S}.$$

Propositions 2–4 provide the basis for computational algorithms to determine an optimal W-transition policy. Proposition 2 asserts that the optimal cost function satisfies Bellman's optimality equation, while Proposition 4 states that the optimal cost function can be iteratively determined by running the recursion

$$J_{s+1} = TJ_s, \qquad s = 0, 1, 2, \ldots \quad (19)$$

for any bounded initial cost function $J_0 : \mathcal{S} \mapsto \Re$, an optimal W-transition policy is found when the iteration converges to the optimal value of the cost function.

IV. Results and Discussion

As we mentioned in the Introduction, an optimal one-transition policy is no longer optimal, i.e., does not necessarily minimize the expected total discounted cost, if one is restricted to apply treatment only every W transitions. Nevertheless, we can apply an optimal one-transition policy every W transitions and compare the effect and cost of such a policy to the ones of an optimal W-transition policy, which truly minimizes the expected total discounted cost.

We anticipate an effective control policy to reduce the likelihood of visiting undesirable states as compared to a network without intervention by modifying the long-run behavior of the network. The effectiveness of a control policy can be measured by the amount of change (*shift*) in the aggregated probability of undesirable states before and after intervention. We should emphasize that an optimal policy does not necessarily result in a maximum shift in the steady-state distribution, as explained earlier, since we are minimizing the expected total discounted cost. The amount of shift in the aggregated probability of undesirable states before and after intervention can be computed as

$$\Delta P^W = \frac{\sum_{i\in\mathcal{U}} \pi_i - \sum_{i\in\mathcal{U}} \tilde{\pi}_i^W}{\sum_{i\in\mathcal{U}} \pi_i}. \quad (20)$$

In (20), $\tilde{\pi}_i^W$ is the probability of being in undesirable state i in the long run using a policy that is applied every W transitions. In this equation, π_i is the probability of being in undesirable state i in the long run when there is no intervention. In other words, given a Markovian gene regulatory network, one can shift the aggregated probability of undesirable states to desirable ones through appropriately altering the expression of the control gene every W time instants.

Formulation of ΔP^W requires the computation of $\tilde{\pi}^W$, i.e., the steady-state distribution of the Markov chain under a W-transition policy μ_W, a policy that is applied every W transitions. To this end, we derive the transition probability matrix of the system when a W-transition policy μ_W is applied. In general, W possible cases can happen for the transition of state i to state j in W steps under a cyclic policy depending on the instants in which states i and j are observed with respect to the treatment times. Let us denote the transition probability matrix under the W-transition policy μ_W by $\mathbf{P}_{\mu_W}$. In the first case, there are $W-1$ uncontrolled transitions and the corresponding transition probability matrix is $\mathbf{P}^{W-1}$. Afterward, in Wth transition, policy μ_W decides whether to apply control or not. The system transitions to state j and the corresponding transition probability matrix is $\mathbf{P}_{\mu_W}$. Consequently, the transition probability matrix corresponding to the first case is $\mathbf{P}^{(W-1)}\mathbf{P}_{\mu_W}$. In the second case, starting from state i, there are $W-2$ uncontrolled transitions and the corresponding transition probability matrix is $\mathbf{P}^{W-2}$. At next transition, policy μ_W decides whether to apply control or not and the system transitions according to the transition probability matrix $\mathbf{P}_{\mu_W}$. Thereafter, the system transitions to state j according to the original transition probability matrix $\mathbf{P}$. The transition probability matrix corresponding to the second case is $\mathbf{P}^{(W-2)}\mathbf{P}_{\mu_W}\mathbf{P}$. Likewise, the transition probability matrix for $W-2$ other cases can be derived. Fig. 1 demonstrates an example for $W = 4$. As this figure suggests, four possible cases can happen depending on where state i is observed with respect to treatment times.

To find the transition probability matrix of the Markov chain under optimal W-transition policy, one should consider the possibility of these cases. Since each of these cases are equally

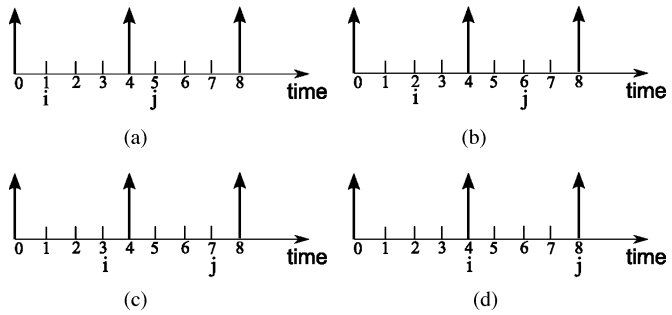


Fig. 1. Example of cyclic intervention strategy for $W = 4$. Arrows represent treatment times. (a)–(d) show the four possible cases that can happen depending on the instants in which states i and j are observed with respect to treatment times.

probable, the following transition probability matrix represents the probabilities of transitions among states when the W-transition intervention policy μ_W is applied:

$$\hat{\mathbf{P}}_{\mu_W} = \frac{1}{W}\sum_{w=1}^{W} \mathbf{P}^{(W-w)}\mathbf{P}_{\mu_W}\mathbf{P}^{(w-1)}. \tag{21}$$

The steady-state distribution $\tilde{\pi}^W$ is the invariant distribution of $\hat{\mathbf{P}}_{\mu_W}$.

In the following sections, we first derive optimal one-transition and W-transition policies for synthetic networks. We generate random PBNs with various properties. We vary the values of bias and connectivity of the PBNs. The bias of a PBN is the probability that each constituent Boolean function takes on the value 1 and the connectivity corresponds to the maximum number of predictors for each Boolean function. Since the bias and connectivity affect the dynamical properties of randomly generated Boolean networks (BNs) [14], we take them as parameters in our simulations. Whenever not specified, the connectivity of the PBN is 3. Furthermore, we investigate the effect of the cost of control on each type of policy. The set of simulations is presented in full in the supplementary materials. We provide some of these in the sequel. We then present a similar investigation for the network obtained from the mammalian cell-cycle network proposed in [15].

A. Synthetic Networks

We generate random PBNs with seven genes. Each PBN consists of four constituent BNs. For each PBN, the probability transition matrix of the corresponding Markov chain is computed [3]. Without loss of generality, the target gene is chosen to be the most significant gene in the states. We assume that the up-regulation of the target gene is undesirable. Consequently, the state space is partitioned into desirable states, $\mathcal{D} = \{0, \ldots, N/2-1\}$, and undesirable states, $\mathcal{U} = \{N/2, \ldots, N\}$, where N represents the total number of states. Since our objective is to down-regulate the target gene, a higher cost is assigned to destination states having an up-regulated target gene. We postulate the following cost-per-stage:

$$g(u,j) = \begin{cases} 0, & \text{if } u = 0 \text{ and } j \in \mathcal{D} \\ 5, & \text{if } u = 0 \text{ and } j \in \mathcal{U} \\ c, & \text{if } u = 1 \text{ and } j \in \mathcal{D} \\ 5 + c, & \text{if } u = 1 \text{ and } j \in \mathcal{U} \end{cases} \tag{22}$$

where c represents the cost of control. Whenever it is not specified, the cost of control is selected to be zero. Results for various costs of control are provided in the supplementary materials. Note that the cost scheme reflects our objective. In practice, the actual values would have to be assigned by a physician according to his or her understanding of the disease.

For each PBN, we vary the value of W from 1 to 10. For each W, the optimal W-transition policy is derived and the corresponding ΔP^W is computed from (21). Given the optimal W-transition policy, we estimate the average total discounted cost induced by this policy. To this end, we generate synthetic time-course data for 1000 time-steps from each PBN model while the optimal W-transition policy is applied. Using this synthetic time-course data, we estimate the discounted cost by accumulating the discounted cost of each state given the policy at that state. This procedure is repeated 10 000 times for random initial states and the average of the induced discounted cost is computed. Furthermore, the optimal one-transition policy is applied every W transitions and the corresponding ΔP^W is computed from (21). To compute the average discounted cost of the optimal one-transition policy when it is applied every W transitions, we generate synthetic time-course data as explained before and the average total discounted cost of the optimal one-transition policy is similarly computed. In sum, for each PBN model, we have the following: ($\bar{C}^W$) average total discounted cost resulting from the optimal W-transition policy; (ΔP^W) the value of ΔP^W resulting from the optimal W-transition policy; ($\bar{C}^{W,1}$) the average total discounted cost of the optimal one-transition policy when it is applied every W transitions; and ($\Delta P^{W,1}$) the value of ΔP^W resulting from the optimal one-transition policy applied every W transitions. The preceding procedure is repeated for 1000 random PBNs, thereby yielding 1000 values for each statistic: $\bar{C}_1^W, \ldots, \bar{C}_{1000}^W$; $\Delta P_1^W, \ldots, \Delta P_{1000}^W$; $\bar{C}_1^{W,1}, \ldots, \bar{C}_{1000}^{W,1}$; $\Delta P_1^{W,1}, \ldots, \Delta P_{1000}^{W,1}$. Using these, we compare the optimal W-transition and one-transition polices via the empirical averages $\mathbf{M}[C^W]$ of $\bar{C}_1^W, \ldots, \bar{C}_{1000}^W$; $\mathbf{M}[C^{W,1}]$ of $\bar{C}_1^{W,1}, \ldots, \bar{C}_{1000}^{W,1}$; $\mathbf{M}[\Delta P^W]$ of $\Delta P_1^W, \ldots, \Delta P_{1000}^W$; and $\mathbf{M}[\Delta P^{W,1}]$ of $\Delta P_1^{W,1}, \ldots, \Delta P_{1000}^{W,1}$. In addition, for each value of W, the histograms of the differences $\bar{C}_i^{W,1} - \bar{C}_i^W$ and $\Delta P_i^W - \Delta P_i^{W,1}$, $i = 1, \ldots, 1000$, are also found. We will see that the means tend to be close, $\mathbf{M}[C^W] \approx \mathbf{M}[C^{W,1}]$ and $\mathbf{M}[\Delta P^W] \approx \mathbf{M}[\Delta P^{W,1}]$, but that the histograms of the differences have long tails to the right, indicating that there are cases for which using the optimal one-transition policy can have strongly detrimental effects.

In the first set of experiments, each constituent BN is randomly generated with a bias, the bias being the probability that a Boolean function takes on the value 1. We randomly select the bias b of a BN from a beta distribution. The mean of the beta distribution is chosen to be 0.3, 0.5, or 0.7. The variance of

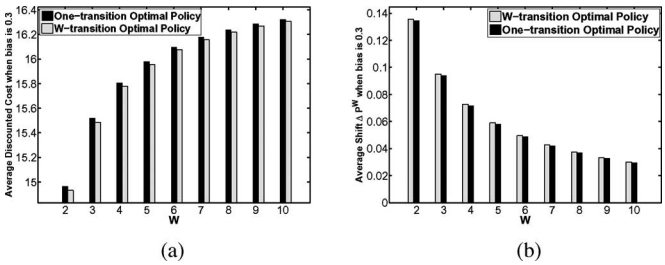


Fig. 2. Comparison of optimal W-transition and one-transition policies based on the average values of ΔP^W and average total discounted cost for $W \in \{1, \ldots, 10\}$ for random PBNs with bias mean $= 0.3$. (a) Average of ΔP^W. (b) Average of discounted cost.

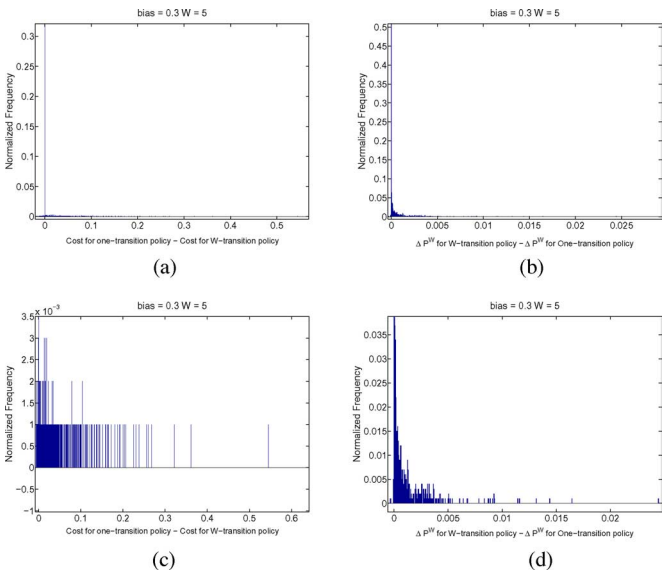


Fig. 3. Comparison of optimal W-transition and one-transition policies based on the histogram of difference of W-transition and optimal one-transition policies for $W = 5$ on random PBNs with bias mean $= 0.3$. (a) Histogram of ΔP^W associated to optimal W-transition policy minus ΔP^W associated to optimal one-transition policy. (b) Histogram of the average discounted cost associated to optimal one-transition policy minus the average discounted cost associated to optimal W-transition policy. (c) Enlarged view of (a). (d) Enlarged view of (b).

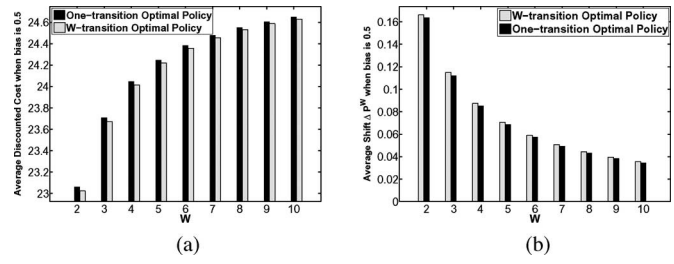


Fig. 4. Comparison of optimal W-transition and one-transition policies based on the average values of ΔP^W and average total discounted cost for $W \in \{1, \ldots, 10\}$ for random PBNs with bias mean $= 0.5$. (a) Average of ΔP^W. (b) Average of discounted cost.

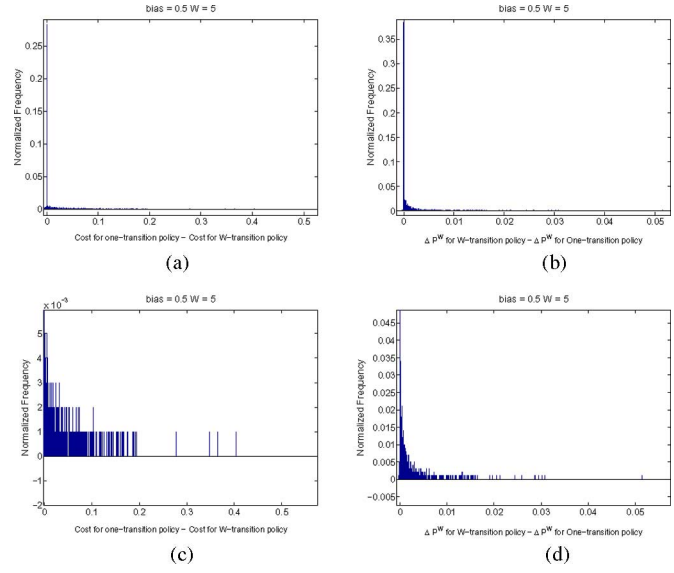
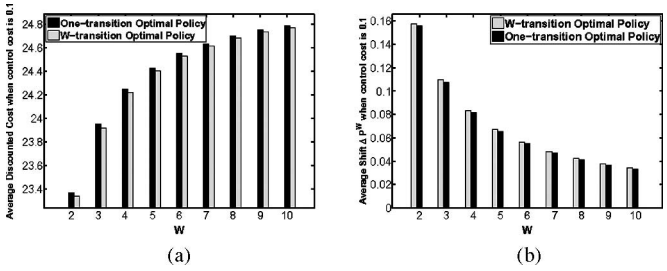


Fig. 5. Comparison of optimal W-transition and one-transition policies based on the histogram of difference of W-transition and optimal one-transition policies for $W = 5$ on random PBNs with bias mean $= 0.5$. (a) Histogram of ΔP^W associated to optimal W-transition policy minus ΔP^W associated to optimal one-transition policy. (b) Histogram of the average discounted cost associated to optimal one-transition policy minus the average discounted cost associated to optimal W-transition policy. (c) Enlarged view of (a). (d) Enlarged view of (b).

the beta distribution, σ^2, is set to a constant value 0.0001. The average values of ΔP^W and the average total discounted costs for both one-transition and W-transition policies are shown in Figs. 2 and 4 for bias values of 0.3 and 0.5, respectively. Similarly, the histograms of the differences of the two policies in terms of ΔP^W and the average total discounted costs are shown in Figs. 3 and 5. Similar studies for bias value of 0.7 can be found in the supplementary materials.

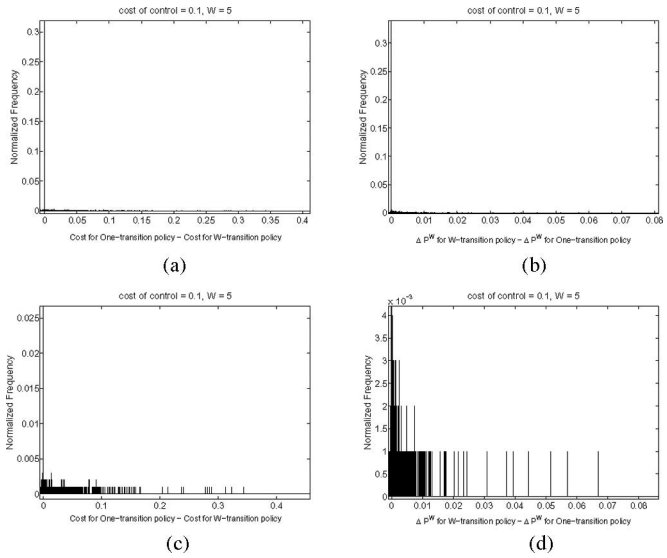
We observe that the average of ΔP^W for both policies decreases as W increases. This behavior is in accordance with the intuition that treatments that are further apart in time are less effective. As we stated in the Introduction, tumors given less time to grow between treatments are more likely to be eradicated [1]. In the long run, less treatment is applied for a larger W, and consequently, more cost is induced. Hence, for a fixed bias, the average discounted costs of both one-transition and W-transition policies increase as W increases. On average, the optimal W-transition policy results in lower discounted cost and higher ΔP^W as compared to the optimal one-transition policy. The histograms of the differences of the two policies in terms of ΔP^W and average discounted show how often they generate similar outcomes and how often the effect of the two policies differ. Note that the differences are not positive for all PBNs. This is because the optimal policies minimize the "expected" total discounted cost. Hence, the W-transition control policy can induce a larger average discounted cost compared to the one-transition control policy, but rarely.

In the second set of experiments, we generate constituent BNs with connectivities 2, 3, and 4. For each connectivity, predictors and Boolean functions are randomly generated with a bias b, randomly selected from a beta distribution with mean 0.5. Results of this experiment can be found in the supplementary materials. Similar to the previous experiment, we observe that the optimal W-transition policy results in lower average discounted cost and higher ΔP^W as compared to the optimal one-transition policy. In the third set of experiments, we repeat the simulations for the cost of control being 0, 0.1, and 0.5.



(a) (b)

Fig. 6. Comparison of optimal W-transition and one-transition policies based on the average values of ΔP^W and average total discounted cost for $W \in \{1, \ldots, 10\}$ for random PBNs with control cost $= 0.1$. (a) Average of ΔP^W. (b) Average of discounted cost.



(a) (b) (c) (d)

Fig. 7. Comparison of optimal W-transition and one-transition policies based on the histogram of difference of optimal W-transition and one-transition policies for $W = 5$ on random PBNs when cost of control is 0.1. (a) Histogram of ΔP^W associated to optimal W-transition policy minus ΔP^W associated to optimal one-transition policy. (b) Histogram of the average discounted cost associated to optimal one-transition policy minus the average discounted cost associated to optimal W-transition policy. (c) Enlarged view of (a). (d) Enlarged view of (b).

Results of this experiment for control cost of 0.1 are shown in Figs. 6 and 7. Further experiments for control cost of 0 and 0.5 can be found in the supplementary materials.

B. Mammalian Cell-Cycle Network

In this section, we construct a PBN that is a probabilistic version of the Boolean model for the mammalian cell cycle regulation proposed in [15]. This PBN postulates the mammalian cell cycle with a mutated phenotype. The proposed intervention method is then applied to hinder the cell growth in the absence of growth factors when treatment is allowed every W transitions.

During the late 1970s and early 1980s, yeast geneticists identified the cell-cycle genes encoding for new classes of molecules, including the cyclins (so-called because of their cyclic pattern of activation) and their cyclin dependent kinases (cdk) partners [15]. Our model is rooted in the work of Faure *et al.*, who have recently derived and analyzed the Boolean functions of the mammalian cell cycle [15]. The authors have been able to quantitatively reproduce the main known features of the wild-type biological system, as well as the consequences of several types of mutations.

TABLE I
BOOLEAN FUNCTIONS OF NORMAL MAMMALIAN CELL CYCLE

Product	Predictors
$CycD$	Input
Rb	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{CycD} \wedge \overline{CycB})$
$E2F$	$(\overline{Rb} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{Rb} \wedge \overline{CycB})$
$CycE$	$(E2F \wedge \overline{Rb})$
$CycA$	$(E2F \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge (\overline{Cdh1 \wedge UbcH10})) \vee (CycA \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge (\overline{Cdh1 \wedge UbcH10}))$
$p27$	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge (\overline{CycE} \wedge \overline{CycA}) \wedge \overline{CycB} \wedge \overline{CycD})$
$Cdc20$	$CycB$
$Cdh1$	$(\overline{CycA} \wedge \overline{CycB}) \vee (Cdc20)$
$UbcH10$	$(\overline{Cdh1}) \vee (Cdh1 \wedge UbcH10 \wedge (Cdc20 \vee CycA \vee CycB))$
$CycB$	$(\overline{Cdc20} \wedge \overline{Cdh1})$

TABLE II
MUTATED BOOLEAN FUNCTIONS OF MAMMALIAN CELL CYCLE

Product	Predictors
$CycD$	Input
Rb	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB})$
$E2F$	$(\overline{Rb} \wedge \overline{CycA} \wedge \overline{CycB})$
$CycE$	$(E2F \wedge \overline{Rb})$
$CycA$	$(E2F \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge (\overline{Cdh1 \wedge UbcH10})) \vee (CycA \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge (\overline{Cdh1 \wedge UbcH10}))$
$Cdc20$	$CycB$
$Cdh1$	$(\overline{CycA} \wedge \overline{CycB}) \vee (Cdc20)$
$UbcH10$	$(\overline{Cdh1}) \vee (Cdh1 \wedge UbcH10 \wedge (Cdc20 \vee CycA \vee CycB))$
$CycB$	$(\overline{Cdc20} \wedge \overline{Cdh1})$

Mammalian cell division is tightly controlled. In a growing mammal, the cell division should coordinate with the overall growth of the organism. This coordination is controlled via extracellular signals. These signals indicate whether a cell should divide or remain in a resting state. The positive signals, or growth factors, instigate the activation of Cyclin D (CycD) in the cell.

The key genes in this model are CycD, retinoblastoma (Rb), and p27. Rb is a tumor-suppressor gene. This gene is expressed in the absence of the cyclins, which inhibit the Rb by phosphorylation. Whenever p27 is present, Rb can be expressed even in the presence of CycE or CycA. Gene p27 is active in the absence of the cyclins. Whenever p27 is present, it blocks the action of CycE or CycA. Hence, it stops the cell cycle. Table I summarizes the Boolean functions of the wild-type cell cycle network.

The preceding explanation represents the wild-type cell-cycle model. Following one of the proposed mutations in [15], we assume p27 is mutated and its logical rule is always zero (OFF). In this cancerous scenario, p27 can never be activated. As we mentioned earlier, whenever p27 is present, Rb can be expressed even in the presence of CycE or CycA. For the mutated cell cycle network, p27 is always zero and Rb cannot be expressed in a case where CycD is not present but CycE or CycA are active [15]. This mutation introduces a situation where both CycD and Rb might be inactive. As a result, in this mutated phenotype, the cell cycles in the absence of any growth factor. In other words, we consider the logical states in which both Rb and CycD are down-regulated as "undesirable states," when p27 is mutated. Table II summarizes the mutated Boolean functions.

The Boolean functions in Table II are used to construct the PBN model for the cell cycle. To this end, we assume that the extracellular signal to the cell-cycle model is a latent variable. The growth factor is not part of the cell and its value is determined

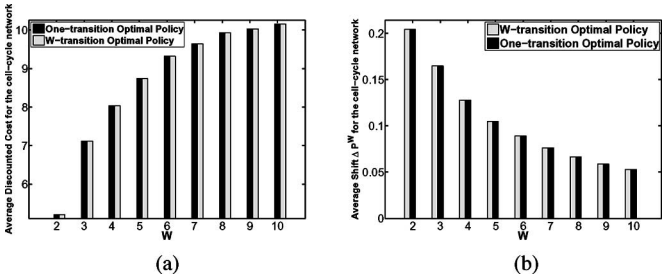


Fig. 8. Comparison of optimal W-transition and one-transition policies based on the values of ΔP^W and average total discounted cost for $W \in \{1, \ldots, 10\}$ for the mammalian cell-cycle network. (a) ΔP^W. (b) Average of total discounted cost.

by the surrounding cells. The expression of CycD changes independently of the cell's content and reflects the state of the growth factor. Depending on the expression status of CycD, we obtain two constituent Boolean networks for the PBN. The first constituent Boolean network is determined from Table II when the value of CycD is equal to zero. Similarly, the second constituent Boolean network is determined by setting the variable of CycD to one. To completely define the PBN, the switching probability, the perturbation probability, and the probability of selecting each constituent Boolean network have to be specified. We assume that these are known. Here, we set the switching probability and the perturbation probabilities equal to 10^{-3}, and the two constituent Boolean networks are equally likely.

According to Table II, the cell-cycle PBN consists of nine genes: CycD, Rb, E2F, CycE, CycA, Cdc20, Cdh1, UbcH10, and CycB. The aforementioned order of genes is used in the binary representation of the logical states, with CycD as the most significant bit and CycB as the least significant bit. This order of genes in the logical states facilitates the presentation of our results and does not affect the computed control policies.

Having CycD and Rb as the most significant genes, we assume that the down-regulations of the CycD and Rb, i.e., the cell growth in the absence of growth factors, is undesirable. Consequently, the state space is partitioned into undesirable states and desirable states. Application of the proposed method, or any of the other methods developed for control of gene regulatory networks, requires the designation of desirable and undesirable states, and this depends upon the existence of relevant biological knowledge. In the cell-cycle example when p27 is mutated, we consider the logical states in which both Rb and CycD are down-regulated as undesirable states. We assume that the cost of the logical states with down-regulated Rb and CycD is higher than that for the states in which these two genes are not simultaneously down-regulated. The costs of undesirable states and control are defined as in (22).

The value of ΔP^W and the average total discounted cost for both optimal one-transition and W-transition policies derived for the cell-cycle network are shown in Fig. 8. In the long run, less treatment is applied for a larger W, and consequently, more cost is induced. Hence, the average discounted costs of both optimal one-transition and W-transition policies increase as W increases. It should be noted that the previous experiments show the average behavior of 1000 random PBNs, while this experiment considers the behavior of one network, i.e., the mammalian cell-cycle network. In this instance, the optimal one-transition and W-transition policies are close to parity for the cell-cycle network.

V. Conclusion

In this paper, our objective has been to devise an effective intervention strategy under the constraint that intervention is permitted only after a recovery period. To incorporate the cyclic constraint on interventions, we have constructed a Markov chain with an augmented state space based on the original Markov chain associated with a gene regulatory network. We have shown how to derive an optimal W-transition policy for the Markov chain with the augmented state space via dynamic programming algorithms. The dynamic programming approach can be computationally prohibitive due to the size of the augmented state space. To mitigate this computational burden, we can collapse the augmented state space into the original state space of the unconstrained Markov chain. Despite this reduction in the size of the state space, the application of our proposed method is still restricted to small number of genes since the complexity of the dynamic programming algorithm increases exponentially with the number of genes. We point out, however, that in our application of interest, intervention in gene regulatory networks, the goal is not to model fine-grained molecular interactions among a host of genes, but rather to model a limited number of genes, typically with very coarse quantization, whose regulatory activities are significantly related to a particular aspect of a specific disease.

In the result section, we have compared the average total discounted cost and ΔP^W of optimal one-transition and W-transition policies for random PBNs with various properties. We have also considered the histogram of the difference of ΔP^W generated by the two policies along with the histogram of the difference between the average discounted costs of the two policies. It is evident from Figs. 2, 4, and 6 that the optimal W-transition policy, on average, induces lower cost and higher ΔP^W as compared to the optimal one-transition policy. However, the discounted cost and ΔP^W corresponding to the optimal one-transition policy is very close to the correspondents of the optimal W-transition policy. The histograms of the difference between the two policies in terms of ΔP^W and the average discounted cost are shown in Figs. 3, 5, and 7. These figures suggest that the performance of these two policies are, on average, close to parity (substantial mass on 0, where the two policies are identical), but that there are PBNs for which the two policies are distinct, on occasion, significantly so. These observations apply across PBNs possessing various properties—bias, connectivity, and cost of control.

References

[1] R. Simon and L. Norton, "The norton-simon hypothesis: Designing more effective and less toxic chemotherapeutic regimens," *Nature Clinical Practice Oncology*, vol. 3, no. 8, pp. 406–407, 2006.

[2] E. R. Dougherty and A. Datta, "Genomic signal processing: Diagnosis and therapy," *IEEE Signal Process. Mag.*, vol. 22, no. 1, pp. 107–112, Jan. 2005.

[3] I. Shmulevich, E. R. Dougherty, S. Kim, and W. Zhang, "Probabilistic Boolean networks: A rule-based uncertainty model for gene regulatory networks," *Bioinformatics*, vol. 18, no. 2, pp. 261–274, 2002.
[4] A. Datta, A. Choudhary, M. Bittner, and E. R. Dougherty, "External control in Markovian genetic regulatory networks," *Mach. Learning*, vol. 52, no. 1/2, pp. 169–191, 2003.
[5] R. Pal, A. Datta, and E. R. Dougherty, "Optimal infinite-horizon control for probabilistic Boolean networks," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2375–2387, Jun. 2006.
[6] B. Faryabi, A. Datta, and E. Dougherty, "On approximate stochastic control in genetic regulatory networks," *IET Syst. Biol.*, vol. 1, no. 6, pp. 361–368, 2007.
[7] G. Vahedi, B. Faryabi, J. Chamberland, A. Datta, and E. R. Dougherty, "Intervention in gene regulatory networks via a stationary mean-first-passage-time control policy," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 10, pp. 2319–2331, 2008.
[8] N. Friedman, K. Murphy, and S. Russell, "Learning the structure of dynamic probabilistic networks," in *Proc. 15th Nat. Conf. Artif. Intell. AAAI-98*, Madison, WI, pp. 139–147.
[9] H. Lahdesmaki, S. Hautaniemi, I. Shmulevich, and O. Yli-Harja, "Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks," *Signal Process.*, vol. 86, no. 4, pp. 814–834, 2006.
[10] M. Brun, E. R. Dougherty, and I. Shmulevich, "Steady-state probabilities for attractors in probabilistic boolean networks," *Signal Process.*, vol. 85, no. 10, pp. 1993–2013, 2005.
[11] J. R. Norris, *Markov Chains*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[12] M. Bittner, P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, N. Sampas, E. Dougherty, E. Wang, F. Marincola, C. Gooden, J. Lueders, A. Glatfelter, P. Pollock, J. Carpten, E. Gillanders, D. Leja, K. Dietrich, C. Beaudry, M. Berens, D. Alberts, V. Sondak, N. Hayward, and J. Trent, "Molecular classification of cutaneous malignant melanoma by gene expression profiling," *Nature*, vol. 406, no. 6795, pp. 536–450, 2000.
[13] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA: Athena Scientific, 2005.
[14] I. Shmulevich and E. R. Dougherty, *Genomic Signal Processing*. Princeton, NJ: Princeton Series in Applied Mathematics, 2007.
[15] A. Faure, A. Naldi, C. Chaouiya, and D. Theiffry, "Dynamical analysis of a genetic boolean model for the control of the mammalian cell cycle," *Bioinformatics*, vol. 22, no. 14, pp. e124–e131, 2006.

Golnaz Vahedi (S'02) received the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, and the M.Sc. degree in electrical and computer engineering from the University of Alberta, Edmonton, AB, Canada, in 2001 and 2004, respectively. She is currently working toward the Ph.D. degree in electrical and computer engineering at Texas A&M University, College Station.

Her current research interests include systems biology, genomic signal processing, and inference and control of genetic regulatory networks.

Babak Faryabi (S'03) received the M.Sc. degree in electrical and computer engineering from Sharif University of Technology, Tehran, Iran, in 1997. He is currently working toward the Ph.D. degree at the Texas A&M University, College Station.

He was a Research Assistant at the University of Toronto, ON, Canada. His current research interests include the applications of Markovian decision processes and their approximations in the modeling and control of gene regulatory networks.

Mr. Faryabi is a member of the International Society for Computational Biology. From 2003 to 2005, he received the Edward S. Rogers Scholarship from the Department of Electrical and Computer Engineering, University of Toronto.

Jean-Francois Chamberland (S'98–M'04) received the B.Eng. degree from McGill University, Montreal, PQ, Canada, the M.Sc. degree from Cornell University, Ithaca, NY, and the Ph.D. degree from the University of Illinois at Urbana-Champaign, in 1998, 2000, and 2004, respectively, all in electrical engineering.

Since 2004, he has been with Texas A&M University, College Station, where he is currently an Assistant Professor in the Department of Electrical and Computer Engineering. His current research interests include communication systems, queueing theory, detection and estimation, and genomic signal processing.

Dr. Chamberland was the recipient of the Young Author Best Paper Award in 2006 from the IEEE Signal Processing Society and the Faculty Early Career (CAREER) Award in 2008 from the National Science Foundation.

Aniruddha Datta (S'87–M'91–SM'97) received the B.Tech degree in electrical engineering from the Indian Institute of Technology, Kharagpur, India, in 1985, the M.S.E.E. degree from the Southern Illinois University, Carbondale, in 1987, and the M.S. (applied mathematics) and Ph.D. degrees from the University of Southern California, Los Angeles, in 1991.

In August 1991, he joined the Department of Electrical and Computer Engineering, Texas A&M University, College Station, where he is currently a Professor and the holder of the J. W. Runyon, Jr. '35 Professorship II. From 2001 to 2003, he worked on cancer research as a Postdoctoral Trainee under a National Cancer Institute (NCI) Training Grant. He is the author or coauthor of *Adaptive Internal Model Control* (Springer-Verlag, 1998), *Structure and Synthesis of PID Controllers* (Springer-Verlag, 2000), *PID Controllers for Time Delay Systems* (Birkhauser, 2004), and *Introduction to Genomic Signal Processing With Control* (CRC, 2007). His current research interests include adaptive control, robust control, proportional–integral–derivative (PID) control, and genomic signal processing.

Dr. Datta was an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL from 2001 to 2003 and the IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS–PART B from 2005 to 2006.

Edward R. Dougherty (M'05) received the M.Sc. degree in computer science from Stevens Institute of Technology, Hoboken, NJ, and the Ph.D. degree in mathematics from Rutgers University, Newark, NJ.

He is currently a Professor in the Department of Electrical and Computer Engineering, Texas A&M University, College Station, where he holds the Robert M. Kennedy Chair and is the Director of the Genomic Signal Processing Laboratory. He is also the Director of the Computational Biology Division, Translational Genomics Research Institute, Phoenix, AZ. He is the author or coauthor of more than 200 papers and 14 books, and editor of five books. He has contributed extensively to the statistical design of nonlinear operators for image processing and the consequent application of pattern recognition theory to nonlinear image processing. His current research interests include genomic signal processing.

Prof. Dougherty was the recipient of the Doctor Honoris Causa by the Tampere University of Technology, Finland, and the International Society for Optical Engineering (SPIE) President's Award, and the Association of Former Students Distinguished Achievement Award in Research at Texas A&M University. He is a Fellow of the SPIE. He was the Editor of the *SPIE/IS&T Journal of Electronic Imaging*. He has been named Fellow of the Texas Engineering Experiment Station and Halliburton Professor of the Dwight Look College of Engineering.