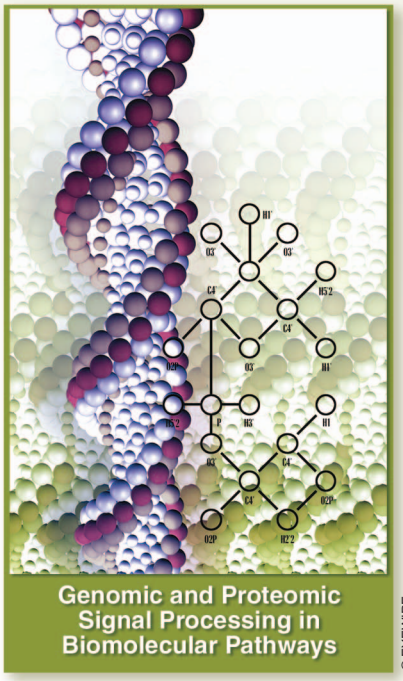
Ranadip Pal, Sonal Bhattacharya, and Mehmet Umut Caglar

# Robust Approaches for Genetic Regulatory Network Modeling and Intervention

## A review of recent advances




© EYEWIRE

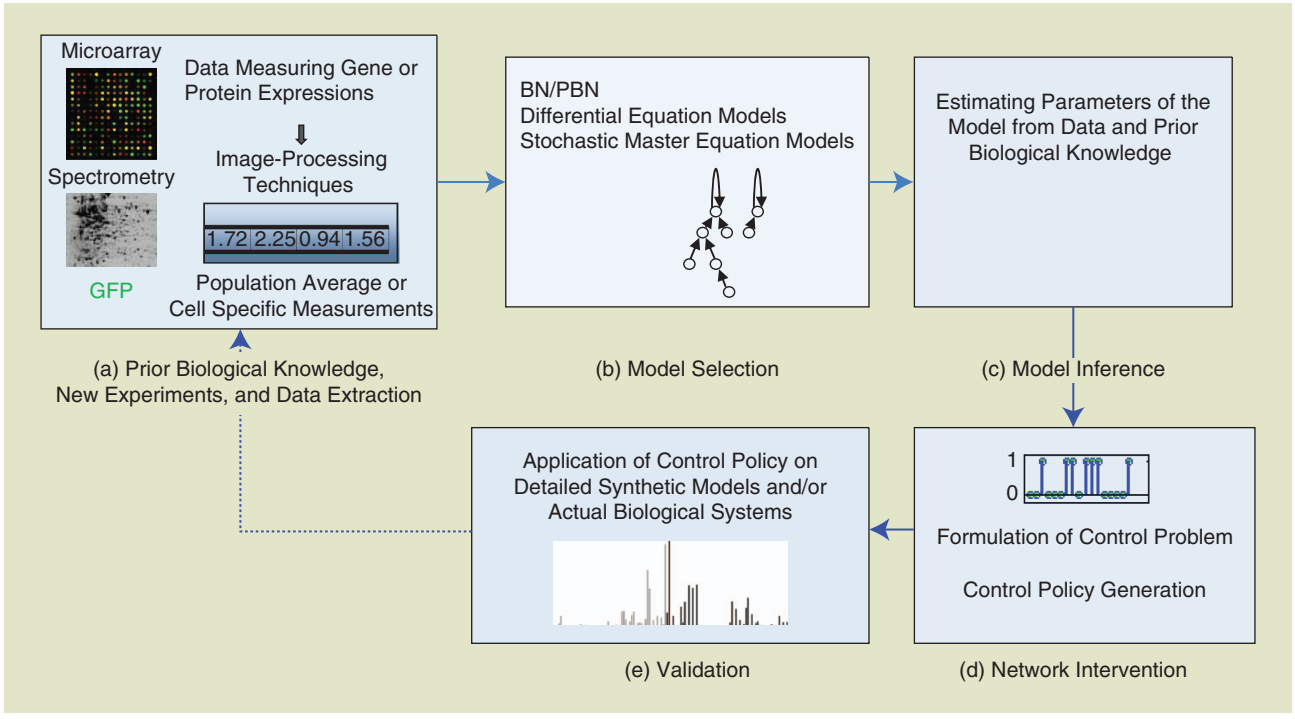
There are two primary objectives for modeling of genetic regulatory networks (GRNs): 1) to better understand the gene interactions and relationships on a holistic level and predict the behavior of biological systems; and 2) to design and analyze control strategies for moving the state of a network from an undesirable location to a desirable one. The control can be for the purpose of systems medicine to move the GRN from a diseased state to a nondiseased state or for synthetic biology purposes to tune the parameters of a synthetic circuit to produce complex novel behaviors. The major steps involved in generating a predictive model and subsequent control policies are illustrated in Figure 1. Step A involves analyzing prior biological knowledge of pathways and conducting new experiments for better understanding of the pathways. Step B consists of selection of a mathematical model to represent the GRN. The next step involves estimating the parameters of the selected model. The last step involves generation of control policies to alter the dynamics of the GRN. Finally, further experiments are conducted to validate the predictions and performance of the designed intervention strategies. Building an accurate mathematical model of the GRN is extremely difficult due to limitations on the experimental data, noise in data extraction, and enormous complexity of the actual biological system. To analyze the robustness of the overall GRN modeling and control approach, we have to examine the effect on predictive and control performance of uncertainties and errors in Steps A–E in Figure 1 independently and in a joint fashion. In this article, we will review approaches to study the robustness of GRN modeling and control strategies with emphasis on the steps of model selection, model inference, and network intervention.

*Digital Object Identifier 10.1109/MSP.2011.942818*
*Date of publication: 8 December 2011*

### GRN MODELING

GRNs represent the interconnections between genomic entities that govern the regulation of gene expression. Since biological regulatory networks are extremely detailed with numerous interactions, a single mathematical model to represent the whole biological regulatory system is generally not


1053-5888/12/$31.00©2012IEEE




**[FIG1] Steps involved in generating a model and intervention strategy for a GRN.**

feasible. Depending on the purpose of modeling, the mathematical model representing the GRN brings in a level of abstraction. The focus of the modeling can be capturing interactions between RNA expressions, protein-protein interactions, or interactions between metabolites. Usually, only parts of the regulome (genes, proteins, and metabolites involved in gene regulation) such as transcription factors, enhancers, and microRNA are made explicit in a mathematical model of a GRN. The modeling can be deterministic capturing the average behavior of a colony of cells or stochastic capturing the inherent noise in biological systems. Furthermore, the models can be fine scale or coarse-scale [1]. This section presents a brief review of the various techniques for modeling GRNs and approaches to study the robustness of the modeling procedure. Relationships between detailed stochastic models and coarse-scale or deterministic models can be used to analyze the predictive power of approximate models.

### *COMMONLY USED MODELS TO REPRESENT A GRN*

Given a set of genes/proteins, the evolution of their expression levels constitutes a dynamical system over time. A large number of approaches have been proposed to model the behavior of GRNs [2]–[4]. Nonlinear ordinary differential equations and piecewise linear differential equations have been proposed as continuous fine-scale deterministic models for GRNs. Differential equation (DE) models assume that species concentration vary continuously and deterministically, which is questionable in case of gene regulation. Thus, stochastic and discrete fine-scale models commonly known as stochastic master equation (SME) models have been proposed for modeling GRNs. To explain an SME model, we will consider a system with $n$ molecular species and $m$ different reaction channels where the state of the system is defined by $\mathbf{x} = [\phi_1, \ldots, \phi_n]$, $\mathbf{x} \in \mathbb{N}^n$ is a vector of integers representing a specific population of each of the $n$ molecular species. For such a system, given the probability density vector $p(\mathbf{x}, t)$ at time $t$, we can derive the DE [5]

$$\dot{p}(\mathbf{x}; t) = -p(\mathbf{x}; t)\sum_{\mu=1}^{m} a_\mu(\mathbf{x}) + \sum_{\mu=1}^{m} p(\mathbf{x} - v_\mu; t)a_\mu(\mathbf{x} - v_\mu), \quad (1)$$

where $a_\mu(\mathbf{x})dt$ denotes the probability that the $\mu$th reaction will happen in a time step of length $dt$ and $v_\mu$ is the stoichiometric transition vector. By considering all the reactions beginning or ending at state $\mathbf{x}$, the time derivative of the probability density of state $\mathbf{x}$ can be written in the form [6]: $\dot{\mathbf{P}}(\mathbf{X}; t) = \mathbf{P}(\mathbf{X}; t)\mathbf{A}$ where $\mathbf{P}(\mathbf{X}; t) = [p(\mathbf{x}_1, t), p(\mathbf{x}_2, t), \ldots]$ is the complete probability density state vector at time $t$ and $\mathbf{A}$ is the state reaction matrix. For the case of finite number of reachable states, the exact solution to the SME can be computed as $\mathbf{P}(\mathbf{X}; t) = \mathbf{P}(\mathbf{X}, 0)e^{\mathbf{A}t}$ [6]. An approximation to the discrete stochastic model in the form of a stochastic DE known as a chemical Langevin equation is often used to make the solution analytically and numerically tractable [4]. The approximation can be represented by the following stochastic DE:

$$\mathbf{x}(t + dt) = bfx(t) + \sum_{\mu=1}^{m} v_\mu a_\mu(\mathbf{x}(t))dt + \sum_{\mu=1}^{m} v_\mu \sqrt{a_\mu(\mathbf{x}(t))}N_\mu(t), \quad (2)$$

where $N_\mu(t)$ for $\mu = 1, \ldots, m$ are independent Gaussian random variables with zero mean and unit variance [7]. The conditions for validity of the approximation require that an

infinitesimal time interval $dt$ is available under which 1) there is no significant change in the propensity functions and 2) the expected number of occurrences of each reaction in $dt$ is significantly greater than one [4].

Probabilistic reasoning based on incomplete prior biological knowledge and current observations has been applied to build models of GRNs commonly known as Bayesian networks or graphical models [8]. Subsequent modification in terms of dynamic Bayesian networks (DBNs) was suggested that allows feedback relations among genes to be modeled [9], [10].

A coarse-scale deterministic model for GRNs has been proposed in the form of the Boolean network (BN) model [11]. Specifically, a BN is composed of a set $V = \{x_1, x_2, \ldots, x_n\}$ consisting of $n$ binary variables, each denoting a gene expression, and a set $F = \{f_1, f_2, \ldots, f_n\}$ of regulatory functions, such that for discrete time, $t = 0, 1, 2, \ldots$, $x_k(t+1) = f_k(x_1(t), \ldots, x_n(t))$. At any time point, the state of the network is given by an expression vector $(x_1(t), \ldots, x_n(t))$, called the gene activity profile (GAP). In the Boolean model, the assumption of a single transition rule for each gene can be problematic with respect to inference: the data are typically noisy, the number of samples is small relative to the number of parameters to be estimated, and there can be unobserved (latent) variables external to the model network and the intrinsic stochasticity of gene expression. Owing to these considerations, the probabilistic BN (PBN) [12] was proposed as a coarse-scale stochastic model for GRNs. The probabilistic structure of the PBN can be modeled as a Markov chain. Thus, PBNs can be considered as part of coarse-scale Markov (CSM) chain models [13].

### *INFERENCE OF GRN MODELS*

Numerous approaches have been proposed for inference of GRN models; in this section, we will discuss few of the recent techniques relevant to robust modeling of GRNs (please refer to [14] and [15] for detailed reviews on inference of GRN models). The term *robust modeling* refers to the ability of the inference and modeling approach to tackle noisy measurements and partial information on biological states. The other form of robustness, related to GRN modeling that is frequently studied, is the robustness of biological processes with respect to generating a reproducible trait under changing conditions. For a current review on GRN robustness, readers are referred to [16], which discusses the current understanding of the role of GRN topology and architecture in the control of transcriptional and phenotypic outputs. Redundancy in GRNs, such as redundant wiring of transcription factors (TFs) converging into a single gene, can be observed in biological processes that are robust to perturbations. On the other hand, stochastic gene expression can generate diversity and provide cells the ability to adapt under adverse conditions. Nodes with low connectivity in GRNs display more stochastic behavior as compared to highly connected nodes [16].

To deal with noise in data extraction, partial observations and stochastic nature of gene expressions, graphical models such as hidden Markov models (HMMs) or generalizations such as DBNs are popular choices to model a GRN. An approach to infer a HMM from time-series data to model biological processes is presented in [17]. Reference [17] uses gene ontology maps as prior biological knowledge for gene-process connections and Kalman filtering for the inference of the HMM parameters. An early method of learning a DBN from time series gene expression microarray data is presented in [18], where an operon map of *E. coli* is used as the prior biological knowledge and expectation maximization algorithm is used to update the conditional probability tables. However, initial application of DBNs such as [18] suffered from high computational cost and low prediction accuracy. Reference [19] tries to address the issue of computational cost by limiting the potential regulator of a gene as the ones having earlier or simultaneous change in expression levels and improve the prediction accuracy by estimating separate transcriptional time lags for different regulator and target combinations.

Estimation of DE models from partial observation of biological states is analyzed in [20]. Adaptive filtering is used in [20] to estimate the parameters of a nonlinear DE model representing the unknown delayed GRN. Another method for inference of nonlinear DE models of GRNs from noisy measurements is considered in [21], where genetic programming is used to infer the model structure and Kalman filtering is applied to estimate the parameters of the model. An alternative technique to increase the reliability of parameter estimates of DE models for GRNs will be model-based experimental design, where the inputs to the biological system is designed to reduce the error in parameter estimation. Reference [22] provides a method of dynamic stimuli design that, when applied to stimulus-response experiments can distinguish among parameterized models with different reaction mechanisms. A mass action kinetics model involving zeroth $(\varnothing \rightarrow x_1)$, first $(x_1 \rightarrow x_2)$- and second $(x_1 + x_2 \rightarrow x_3)$-order reactions are considered that can be represented in matrix form in the following manner:

$$\frac{d\mathbf{x}}{dt} = A_1\mathbf{x} + A_2(\mathbf{x} \otimes \mathbf{x}) + B_1 u + B_2(\mathbf{x} \otimes u) + K$$
$$\mathbf{y} = C\mathbf{x}, \tag{3}$$

where $\otimes$ refers to the Kronecker product and $K$, $[A_1, B_1]$, and $[A_2, B_2]$ refers to zeroth-order, first-order, and second-order reaction matrices respectively. The stimulus design is based on a feedback controller applied to drive the system output $(\mathbf{y}(u, t))$ to a desired response $(\mathbf{y}_r(t))$. The cost function considered is $G(u) = \int_0^T [\mathbf{y}(u, t) - \mathbf{y}_r(t)]^2 dt$. Two approaches are considered to solve the problem: a tangent linear controller based on linear approximations and dynamic optimization controller based on gradient optimization. The suitability of an inferred model is estimated based on the ability of the controller designed using the model to drive the physical system to the desired output. The approach is tested on models of antibody ligand binding, mitogen-activated protein kinase (MAPK) phosphorylation and dephosphorylation, and larger

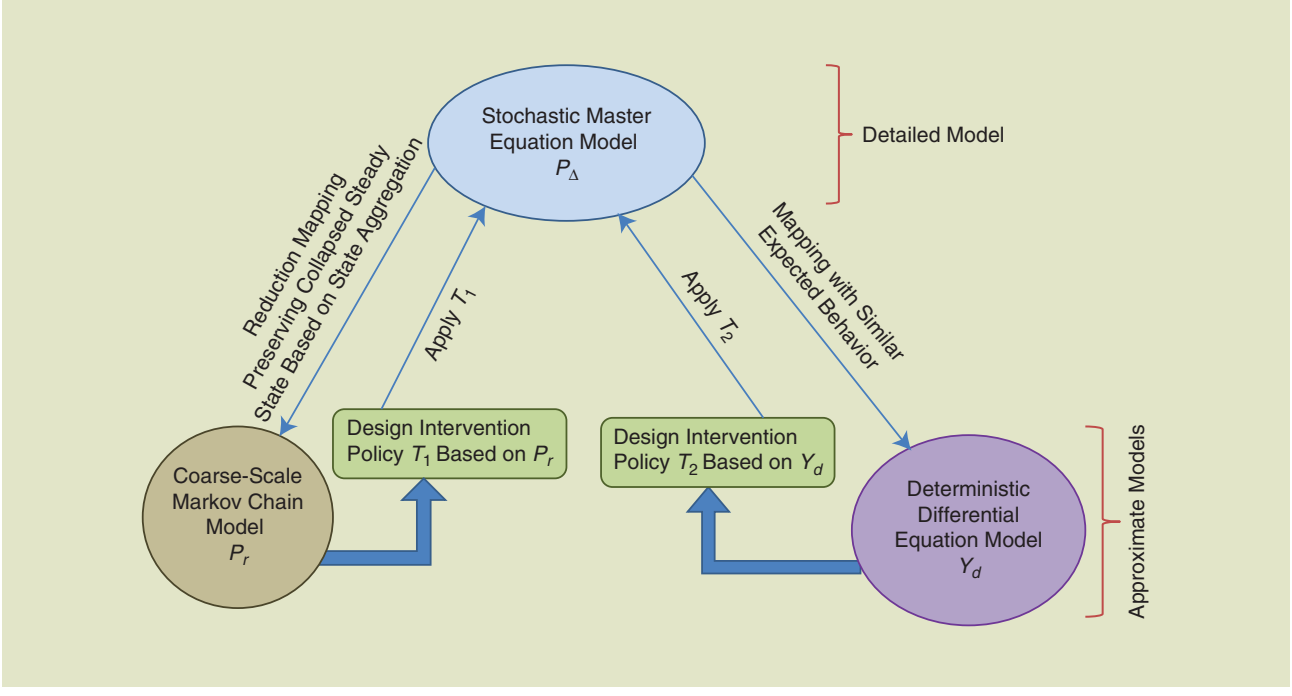
models of the epidermal growth factor receptor (EGFR) pathway. For each of these systems, the correct model produces the controller that minimizes the cost function.

To analyze the effect of uncertainties in molecular concentrations and parameter values on the construction of synthetic gene networks, [23] proposes a robustness analysis approach for gene networks represented by piecewise-multiaffine (PMA) DE models. The state space is partitioned based on the breakpoints of the PMA model for robustness analysis. For a specific set of parameters $P$, the discrete abstraction of the PMA is created and the validity of the dynamic behavior of the system represented by linear temporal logic (LTL) is evaluated by model checking [23]. Another approach on robust parameter estimation for mass action kinetics model using semidefinite programming is considered in [24].

In terms of robust design of BNs, [25] presents an approach based on error correcting codes to design BNs with robustness against perturbations and exhibiting cyclic behavior similar to cell cycle gene regulatory networks. The proposed approach allows cyclic attractors and exhibits checkpoint behavior that tracks and corrects errors due to cellular noise. To avoid information loss through data discretization and retrieve regulatory relations from gene expression data, a fuzzy approach is presented in [26]. The proposed multilayer evolutionary trained neuro-fuzzy recurrent network has the advantages of producing easily interpretable adaptive number of temporal fuzzy rules to represent intergene relationships, avoiding drawbacks of classical neural network training algorithms through the use of particle swarm optimization, and ability to assign scores reflecting the confidence on an inferred relationship.

### *MODEL RELATIONSHIPS*

The selection of a mathematical model to describe the dynamical behavior of a regulatory network is dependent on the available data, estimation techniques, model complexity, and specific purpose of constructing the model. SME models provide the most detailed description of the dynamics of gene expression and imbed, in principle, all the information about the biochemical reactions involved in gene interactions [3], [2]. However, the estimation of the parameters of the fine scale stochastic model requires detailed experiments generating larger data sets and preferably time-series data and cell specific measurements, which is not the case with most biological studies that involves primarily cell average microarrays with limited samples [1]. For a detailed model, if the parameters are poorly estimated from limited data, certain biologically meaningful properties observed in the data could be lost. Furthermore, the use of fine-scale stochastic models is restricted by the inherent computational complexity involved in its simulation [2]. Thus approximate models such as DEs, stochastic DEs, BNs, or CSMs are often used. We would be interested in studying the properties of the detailed model that are still captured by the approximate models. We next provide a brief description of recent work that has been conducted in this area with emphasis on SMEs as the detailed model and ordinary DEs and CSMs as the approximate models. A pictorial representation of the mappings that will be discussed is shown in Figure 2. Other relationships such as between DE and BN models has been explored in [27]. In [27], the BN model is shown as a coarse-grain limit of the DE model and is illustrated on the budding yeast cell cycle network.




**[FIG2] Overview of the model relationships.**

## SME AND ITS DE APPROXIMATION

The relationship between SME and deterministic DE models have been studied analytically recently from the perspective of similarity of the average predictive behavior of the two models [28]–[31]. Taylor series expansion of the expected behavior of the SME model has been used to generate the DE model and its properties studied.

To explain the modeling of the average behavior of a SME model by a DE model, let us consider the SME model shown in (1). Let the number of protein molecules for $\mathbf{x}_j$ be between 0 and $\mathbf{M}_j$. Then the marginal probability of $\mathbf{x}_j = i$ is given by the following formula:

$$p(i;t) = \sum_{l_1=0}^{\mathbf{M}_1} \cdots \sum_{l_{j-1}=0}^{\mathbf{M}_{j-1}} \sum_{l_{j+1}=0}^{\mathbf{M}_{j+1}} \cdots \sum_{l_n=0}^{\mathbf{M}_n} p(l_1, l_2, \ldots, l_j = i, \ldots, l_n; t). \quad (4)$$

Change in expectation of $\mathbf{x}_j(t)$ in the small time duration $dt$ is given by

$$\begin{aligned} E_{x_j}(x_j;t+dt) - E_{x_j}(x_j;t) &= \sum_{l_j=0}^{\mathbf{M}_j} l_j \left( p(x_j = l_j; t+dt) - p(x_j = l_j; t) \right) \\ &= \sum_{\mu=1}^{m} \sum_{u_1=-v_{\mu_1}}^{\mathbf{M}_1 - v_{\mu_1}} \cdots \sum_{u_n=-v_{\mu_n}}^{\mathbf{M}_n - v_{\mu_n}} u_j p(u_1, \ldots, u_n; t) a_\mu(u_1, \ldots, u_n) dt \\ &\quad - \sum_{\mu=1}^{m} \sum_{l_1=0}^{\mathbf{M}_1} \cdots \sum_{l_n=0}^{\mathbf{M}_n} l_j a_\mu(l_1, \ldots, l_n) p(l_1, l_2, \ldots, l_n; t) dt \\ &\quad + \sum_{\mu=1}^{m} v_{\mu_j} \sum_{r_1=-v_{\mu_1}}^{\mathbf{M}_1 - v_{\mu_1}} \cdots \sum_{r_n=-v_{\mu_n}}^{\mathbf{M}_n - v_{\mu_n}} p(r_1, \ldots, r_n; t) a_\mu(r_1, \ldots, r_n) dt. \end{aligned} \quad (5)$$

If we consider $\mathbf{M}_i$ to be high enough, then $p(x_i \geq \mathbf{M}_i - \max_\mu |v_{\mu_i}|; t)$ and $p(x_i \leq \max_\mu |v_{\mu_i}|; t)$ can be assumed to be infinitesimal. This will allow us to ignore the terms corresponding to $x_i \geq \mathbf{M}_i - \max_\mu |v_{\mu_i}|$ and $x_i \leq \max_\mu |v_{\mu_i}|$. If $y_j$ denotes $E_{x_j}(x_j; t)$, (5) can be rewritten as

$$\begin{aligned} \frac{dy_j}{dt} &\approx \sum_{\mu=1}^{m} v_{\mu_j} \sum_{r_1=-v_{\mu_1}}^{\mathbf{M}_1 - v_{\mu_1}} \cdots \sum_{r_n=-v_{\mu_n}}^{\mathbf{M}_n - v_{\mu_n}} p(r_1, \ldots, r_n; t) a_\mu(r_1, \ldots, r_n) \\ &= \sum_{\mu=1}^{m} v_{\mu_j} E_X(a_\mu(X)). \end{aligned} \quad (6)$$

Based on the first-order Taylor series approximation of $a_\mu(X)$ around $\Theta = E_X(X) = [y_1, y_2, \ldots, y_n]$, we have $E_X(a_\mu(X)) \approx a_\mu(\Theta)$. Thus, (6) can be approximated by

$$\frac{dy_j}{dt} \approx \sum_{\mu=1}^{m} v_{\mu_j} a_\mu(y_1, \ldots, y_n) \quad (7)$$

for $j = [1, 2, \ldots, n]$. Equation (7) provides the DE model for tracking the mean of the chemical species represented by the SME. DEs for further tracking of the variances and covariances of the concentration of chemical species have been proposed in [29]. Approximate approaches to calculate the first- and second-order moments for SME models of systems are illustrated in [28]. The DE approximation allows simulating the average behavior of a system such as a colony of cells in a computationally inexpensive manner.

## SME AND ITS CSM APPROXIMATION

Based on the state transitions of a SME model, it can be considered as a continuous time Markov chain with a huge number of states. To explore mappings from fine-scale SME models to CSM models, let us consider $M$ and $N$ to denote the number of states of the SME and CSM model respectively. We consider a sequence $0 = a_0, a_1, \ldots, a_N = M$ such that $a_j - a_{j-1}$ for $j \in \{1, 2, \ldots, 2^n\}$ denote the number of states in the SME model that map to state $j$ in the CSM model. Let $P_\Delta = e^{\mathbf{A}\Delta t}$ represent the $M$-dimensional discrete time SME model ($\Delta t$ is a suitable time period) and $P_r$ represent the reduced $N$-dimensional CSM model. For the SME model, let $\eta$ represent the $M$-dimensional steady-state probability vector and $\zeta$ represent the $N$-dimensional collapsed steady-state vector. Here, collapsing refers to aggregation of states i.e., $\zeta(i) = \sum_{i_2=a_{i-1}+1}^{a_i} \eta(i_2)$ for $i = 1, \ldots, N$.

In [32], a mapping from SME to CSM model is presented where the steady-state probability distribution of the reduced model $P_r$ represented by (8) is equivalent to the collapsed steady-state probability distribution of $P_\Delta$

$$P_r(i,j) = \frac{\sum_{j_1=a_{j-1}+1}^{a_j} \sum_{i_1=a_{i-1}+1}^{a_i} P_\Delta(i_1, j_1) \eta(i_1)}{\sum_{i_2=a_{i-1}+1}^{a_i} \eta(i_2)}. \quad (8)$$

With (8) denoting the state transition probabilities of the CSM model, the steady-state probability distribution vector, $\pi$, of the CSM model is given by [32]: $\pi(i) = \sum_{i_2=a_{i-1}+1}^{a_i} \eta(i_2) = \zeta(i)$ for $i = 1, \ldots, N$. Equation (8) provides a mapping from a fine-scale SME model to a CSM model based on aggregation of states that maintains the collapsed steady-state distribution of the detailed model. If we revisit (8), we notice that $P_r(i,j)$ can be calculated from the transitions of the network once it has reached the steady state. Thus, the mapping described by (8) exists between a network represented by a SME model and a CSM model when the transition probabilities of the CSM are inferred based on state transition data at steady state.

To evaluate the transient behavior of the CSM model, let us consider the difference equation $d_j(t + \Delta t) = |\gamma_j(t + \Delta t) - \tau_j(t + \Delta t)|$ where $\gamma_j(t)$ denote the collapsed probability of states $a_{j-1} + 1$ to $a_j$ of the SME model at any time $t$ and $\tau_j(t)$ denote the probability of state $j$ at any time $t$ for the CSM model. We can bound the difference equation $d_j(t + \Delta t)$ for $\gamma_j(t) = \tau_j(t)$ in the following way:

$$d_j(t + \Delta t) \leq \max_{i_1 \in [1, \ldots, N]} \left( \max_{i_2 \in S_{i_1}} q(i_2, j) - \min_{i_2 \in S_{i_1}} q(i_2, j) \right), \quad (9)$$

where $q(i_2, j) = \sum_{j_1=a_{j-1}+1}^{a_j} P_\Delta(i_2, j_1)$.

## INTERVENTION IN GRNs

Intervention or control applied to GRNs are studied primarily for 1) systems medicine purposes to move the network out of undesirable states, such as those associated with disease, and into desirable ones [33], [34] and 2) synthetic biology purposes for executing new and complex processes: an example will be

alteration of genetic networks of microbes or plants to produce higher efficiency fuels [35]. This section on intervention is organized as follows: The section "Intervention Approaches" introduces the currently available approaches for intervention in GRN for systems medicine and synthetic biology purposes; and the section "Robust Intervention in GRNs" discusses the approaches for robustness analysis and robust design of intervention policies for GRNs. The techniques discussed in the section "Intervention Approaches" are predominantly based on the accurate knowledge of the GRN model whereas the section "Robust Intervention in GRNs" considers the various forms of uncertainties in GRN modeling while analyzing and designing intervention approaches.

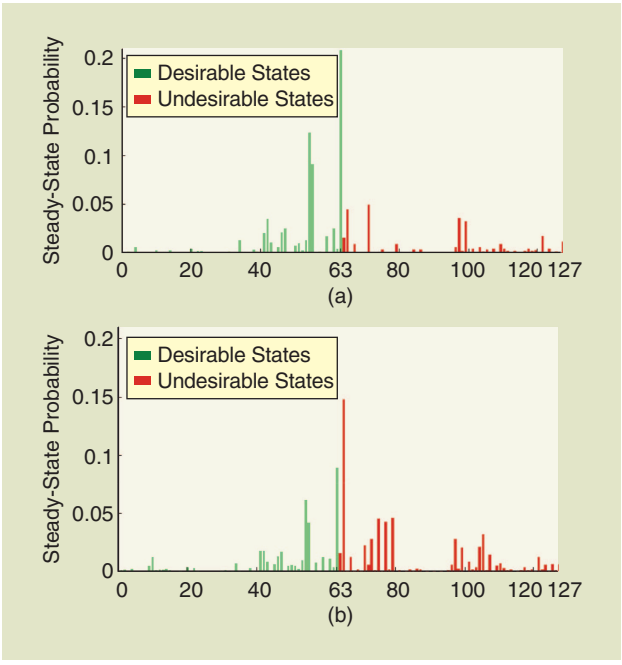
### *INTERVENTION APPROACHES*

Most of the intervention studies related to systems medicine have focused on coarse-scale BN and Markov chain (PBN is one such example) models. The problems with designing and analyzing intervention strategies in fine-scale stochastic models are two-fold:

1) enormous computational complexity involved in simulation of the fine-scale models
2) huge data requirements for inference of the parameters of the fine-scale models.

Consequently, a reasonable approach is constructing intervention strategies for coarse-scale models with the assumption of capturing the overall effects of intervention manifested at the phenotypic (observational) level. We next provide a brief description of the several intervention approaches that have been designed to date for PBNs (a class of CSM models) followed by approaches based on DE and other approximate models.

The motivation behind application of control theory for Markovian models like PBN is to devise optimal policies for manipulating control variables that affect the transition probabilities of the network and can, therefore, be used to desirably affect its dynamic evolution. Since the intervention approaches depends on the Markov chain induced by the PBN, it is extendable to other dynamical systems based on graphical models, such as DBNs. The initial approach consisted of favorably altering the mean first passage times (MFPTs), increasing MFPTs to undesirable states and decreasing MFPTs to desirable states, in the Markov chain associated with a PBN via a one-time perturbation of the expression of a control gene [36]. Shortly thereafter, stochastic control theory was used to alter the transient dynamics over a finite time horizon, both in the presence of full and imperfect information [37], [38]. Subsequently, finite horizon [39] and infinite horizon control [33] for context-sensitive PBNs were addressed using dynamic programming. In practice, intervention will be achieved by [40] 1) targeted small molecule kinase inhibitors (Imatinib, Gefitinib, Erlotinib, Sunitinib, etc.) 2) monoclonal antibodies altering the protein concentrations (Cetuximab, Alemtuzumab, Trastuzumab, etc.) or 3) gene knockdowns. The state desirability is determined by the values of genes/proteins associated



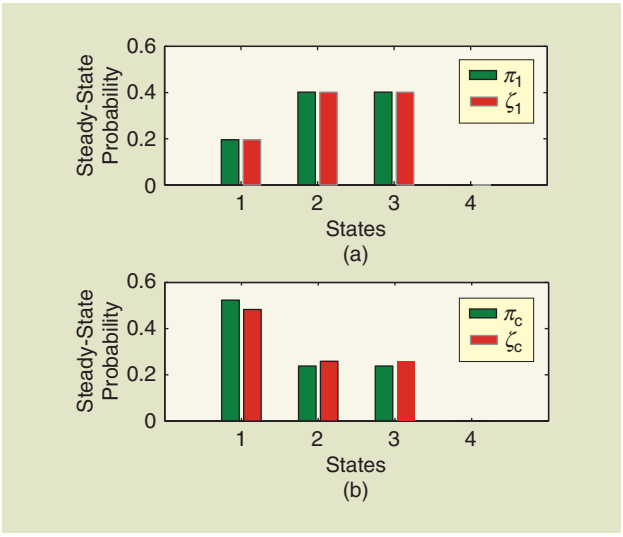

[FIG3] **(a) Steady state distribution after intervention and (b) original steady state [33].**

with the phenotypes of interest. To explain infinite-horizon intervention in PBNs, we next provide a brief mathematical description of the control problem. A PBN with control can be modeled as a finite-state Markov chain [39], [33] described by the control-dependent one-step transition probability $p_{ij}(u) := P(z_{t+1} = j | z_t = i, u_t = u)$ where, for all $t$, the state $z_t$ is an element of a space $S$ and the control input $u_t$ is an element of a space $C$. When the transition probabilities are exactly known, the states make transitions according to $\omega := (P^u)_{u \in C}$. In this case, once a control input is chosen, the resulting controlled transition probability matrix is uniquely determined. Let $\mu = (u_1, u_2, \ldots.)$ represent a generic control policy and $\Pi$ represent the set of all possible $\mu$'s, i.e., the set of all possible control policies. Let $J_{\mu,\omega}$ denote the expected total cost for the discounted cost infinite-horizon problem [33] under control policy $\mu$ and transitions $\omega$

$$J_{\mu,\omega}(z_0) = \lim_{M \to \infty} E_{z_{t+1},\, t=0,\,1,\ldots} \left\{ \sum_{t=0}^{M-1} \alpha^t \, \tilde{g}(z_t, u_t, z_{t+1}) \right\}, \qquad (10)$$

where $0 < \alpha < 1$ denotes the discount factor and $\tilde{g}(z_t, u_t, z_{t+1})$ represents the cost of going from state $z_t$ to $z_{t+1}$ under the control action $u_t$. $\tilde{g}$ is higher for undesirable destination states. For the same destination states, $\tilde{g}$ is higher when the control is active versus when it is not. The control problem here corresponds to minimizing the cost in (10). Consequently, the optimal infinite-horizon discounted cost is given by

$$\Phi(\Pi, \omega, z_0) := \min_{\mu \in \Pi} J_{\mu,\omega}(z_0). \qquad (11)$$




**[FIG4] (a) Steady-state probability distributions $\pi_1$ and $\zeta_1$. (b) Steady-state probability distributions $\pi_c$ and $\zeta_c$ [32].**

The interest in infinite-horizon control is driven by the fact that such policies allow us to optimally alter the steady-state distribution of the network by driving its probability mass into desirable states. In [33], a stationary control policy is derived using dynamic programming principles [41] to shift the steady-state distribution for a PBN derived from gene expression data collected in a metastatic melanoma study, where the abundance of mRNA for the gene WNT5A was found to be highly discriminating between cells with properties typically associated with high versus low metastatic competence, a relationship that suggests an intervention scheme to reduce the WNT5A gene's action. The steady-state distributions shown in Figure 3 corroborate the fact that the intervention policy was able to shift the probability mass to states with lower metastatic competence (down-regulated WNT5A).

In synthetic biology, tuning of the parameters of a system is essential to drive the system towards desirable behavior. A tuning scheme based on PMA models of biological systems and system dynamics represented by LTL is considered in [23] and [42]. The approach is tested on a synthetic gene network of a cascade of genes tetR, lacI, cI, and eyfp. The control input is the input concentration of aTc and the system output is the fluorescence of eyfp protein. The technique was able to generate concentrations of input aTc producing a desired output behavior. An implementation of Boolean logic in biological systems using ribonucleic acid interference (RNAi) is considered in [43]. Small interfering RNAs (siRNA) are given as inputs to the Boolean logic evaluators and the logic circuit consists of multiple mRNAs encoding the same protein but having different noncoding regions.

Some additional approaches that have been proposed for control of GRNs include topological control based on removing negative feedback loops for GRNs represented by a directed graph of gene interconnections [44]; and control of the stochastic hybrid model of the lactose regulation system of *E. coli* based on control designed on a two-state continuous Markov chain model approximation of the system [45].

### *ROBUST INTERVENTION IN GRNs*

For investigating the robustness of intervention approaches for GRNs, two major forms of uncertainties have to be considered: 1) mismatch between the mathematical model of the network used to generate the control policy and the actual biological network and 2) uncertainties in estimating the parameters of the model of the GRN. We next provide some approaches for analyzing the effects of the two forms of uncertainties on control outcome. The section "Mismatch Between the Actual GRN and the Mathematical Model" investigates the effect of control design on CSM or DE model when the actual GRN is represented by a SME model whereas the section "Uncertainties in Estimation of Model Parameters" investigates the effect of uncertainties in model parameter estimation on control outcome and approaches to increase the robustness of the control design.

## MISMATCH BETWEEN THE ACTUAL GRN AND THE MATHEMATICAL MODEL

The mismatch between the actual biological network and the model of the GRN can be studied by assuming that the actual GRN is faithfully represented by an SME model and the model for generation of the control policy is either a CSM model or a deterministic DE model. In this context, an intervention approach designed from a coarse-scale (such as CSM model) or deterministic (such as DE model) model and producing the desired behavior in the coarse-scale or deterministic model will be considered robust if the designed policy when applied to the fine scale model also produces similar coarse-scale or average behavior. If we refer to Figure 2, then we would be interested in investigating 1) the relationship between the coarse-scale behavior of the controlled CSM model ($T_1$ applied to $P_r$) and the controlled SME model ($T_1$ applied to $P_\Delta$), and 2) the similarities in the average behavior of the controlled DE model ($T_2$ applied to $Y_d$) and the controlled SME model ($T_2$ applied to $P_\Delta$).

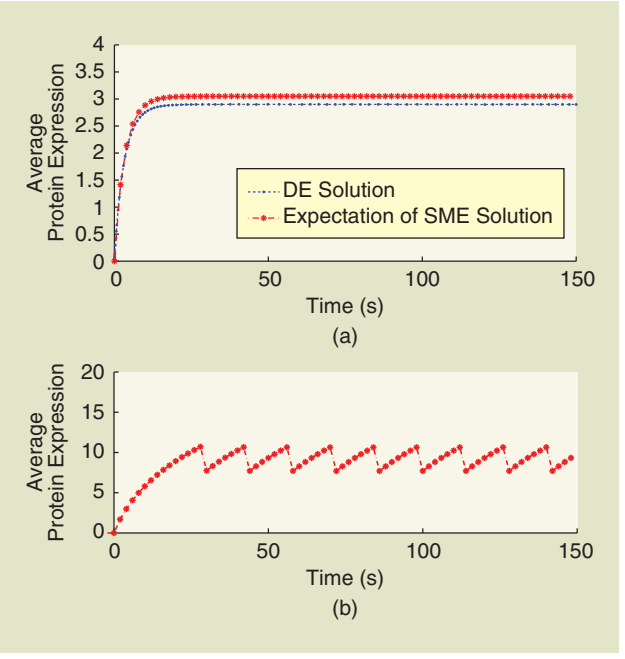
The first relationship is explored in [32]. The control problem is similar to the infinite horizon control considered in (11) to modify the steady-state probabilities from undesirable to desirable states. A parameter $\epsilon$ is formulated whose low value provides a sufficient condition for similarity of the steady-state probability distributions of the controlled CSM model $\pi_d$ and the collapsed steady-state probability distribution of the controlled SME model $\pi_c$. A low value of $\epsilon$ signifies that the probabilities of leaving the state $i$ from the states mapping to state $i$ (i.e., $a_{i-1}+1 \ldots a_i$) are very similar. $\epsilon$ can have a low value for cases such as bimodal gene expressions where thresholds are selected to divide the distinct modes. The infinity norm of the residual for the approximate solution is shown to be upper bounded by $2\epsilon$ [32]. As a biological example for investigating the effect of coarse-scale modeling

on intervention outcome, let us consider a genetic toggle switch where there are two competing proteins $U_1$ and $U_2$, each of which inhibits the transcription of the other [46]. The decay reactions $U_1 \rightarrow \phi$ and $U_2 \rightarrow \phi$ have propensities $c\Psi_1$ and $c\Psi_2$, respectively. The transcription of new copies of the proteins are guided by the reactions $\phi \rightarrow U_1$ and $\phi \rightarrow U_2$ with propensities $b/(b + \Psi_2)$ and $b/(b + \Psi_1)$, respectively. Figure 4(a) shows the equivalence of the steady-state probability distribution $\pi_1$ of $P_r$ and the collapsed steady-state probability distribution $\zeta_1$ of $P_\Delta$ for the system with $b = 0.4$ and $c = 0.05$ . But when a control policy is designed using the reduced network $P_r$ and applied to $P_\Delta$, the collapsed steady-state probability distribution $\zeta_c$ is not exactly the same as the steady-state probability distribution $\pi_c$ of the controlled coarse network as shown in Figure 4(b). The control problem setup is similar to (11). The control objective is to reduce the number of steady state molecules of the two proteins (in other words increase the steady-state probability of State 1) and the control input is assumed to increase the decay of the proteins by changing the value of parameter $c$ to 0.1. The control policy is derived using dynamic programming principles. However $\pi_c$ and $\zeta_c$ are very close.

To investigate the effect of control policy design based on DE models when applied to SME models, let us consider the same genetic toggle switch. Based on the DE approximation provided in (7), Figure 5(a) shows the DE solution (blue dots) and the expectation of the SME solution (red stars) for the model with parameters values $b = 4$, $c = 0.2$ and starting distribution of $p(X) = [1, 0, 0 \ldots, 0]$ i.e., the number of protein molecules for both the protein is zero for the initial state. To study the effect of the control policy designed on the simpler approximate DE model being applied to the original SME, we consider a control problem of maintaining the average expression of Protein 1 to be around the value ten. In this case, the control input is assumed to decrease the decay of the proteins by decreasing the value of the parameter $c$ to 0.01. We design a simple on off controller with deadzone based on the DE model, which produces the following control policy: use System 1 (original system, $b = 4$, $c = 0.2$) if the average protein expression is above 10.5 and use System 2 (system when control is on, $b = 4$, $c = 0.01$) if the average protein expression is below 9.5. The result of this control policy applied to the SME models is shown in Figure 5(b). We notice that the control objective has been almost achieved with the average expression staying between 7.5–10.5. The variation could have been reduced if we apply control in much finer time interval than the current time interval of 2 s. In actual translational medicine, the time interval will be dependent on a number of factors like drug toxicity and technical limitations on the sampling rate of expression measurements.

## UNCERTAINTIES IN ESTIMATION OF MODEL PARAMETERS

The investigation of the robustness of control policy design for GRNs can be approached from the perspective of

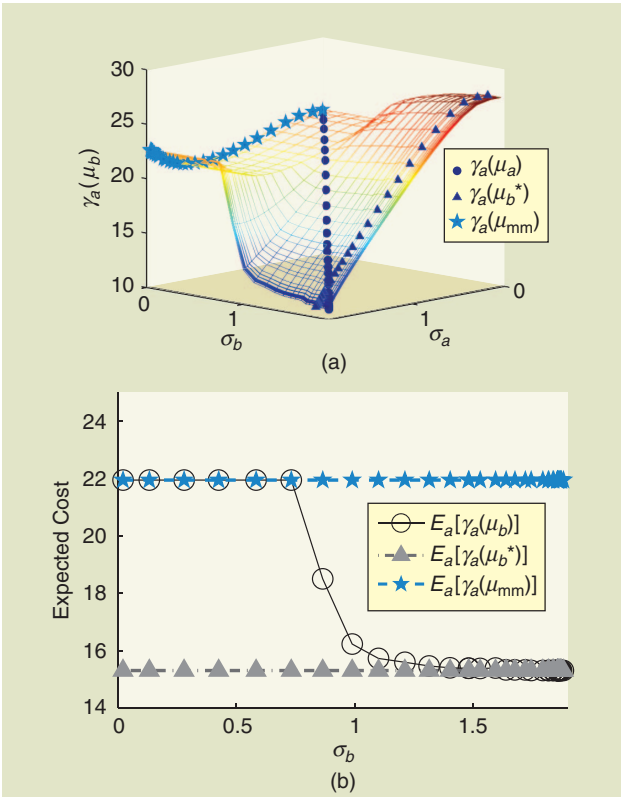



**[FIG5] (a) The average Protein 1 expression simulated using SME (red *) and DE approximation (blue dots) for the original genetic toggle switch with parameters $b = 4$ and $c = 0.2$. (b) The average protein expression simulated using SME after application of the on-off control policy designed using the approximate DE models.**

- studying the robustness of a generated policy in presence of uncertainties
- incorporating the modeling uncertainties in the policy design to arrive at robust control policies.

### Intervention Performance in Presence of Uncertainties

The robustness of a stationary control policy for a Markov chain model similar to (11) is explored in [47] using perturbation theory for Markov chains. When the class of allowed control is restricted to the flipping of genes, the application of a stationary policy converts the uncontrolled transition probability matrix $P$ to a controlled transition probability matrix $P_c$, where $P_c$ and $P$ are related by $P_c = TP$ and $T$ represents a matrix which has only one nonzero entry of one in each row [47]. Let $\pi$ and $\pi_c$ denote the stationary distribution vectors corresponding to the transition matrices $P$ and $P_c$, respectively. Since the probability transition matrix $P$ is estimated from noisy and limited experimental data, there can be errors in estimation. Let $\widetilde{P}$ denote the actual transition matrix of the genetic network and let $\widetilde{P}_c$ denote the controlled transition probability matrix that results from the application of the stationary policy $T$ on $\widetilde{P}$. Let $\widetilde{\pi}$ and $\widetilde{\pi}_c$ denote the stationary distributions of $\widetilde{P}$ and $\widetilde{P}_c$ respectively. For the two Markov chains with transition probabilities $P$ (estimated transition probability matrix) and $\widetilde{P}$ (transition probability matrix of the actual GRN) and sharing a common state space, the difference between the two stationary distributions can be bounded by $|\pi - \widetilde{\pi}|_q \leq k \| E \|_\infty$ where $q = 1$ or $\infty$ and $k > 0$ are some constants and $|\pi - \widetilde{\pi}|_q$ refers to the $q$th norm of the vector




**[FIG6] (a) Cost for policies designed at $\sigma_b$ over the set of all possible $\sigma_a$'s. (b) Expected cost for various policies [51].**

$\pi - \tilde{\pi}$ and $\| E \|_\infty$ denotes the $\infty$ norm of the error matrix $E$, which is equivalent to the maximum absolute row sum of $E$. The constants $K$ are usually referred to as condition numbers [48]. It is shown in [47] that the bound based on the ergodicity coefficient $(K = 1/1 - \tau_1(P))$ is valid for controlled transition probability matrices also i.e., for $\tau_1(P) \neq 1$

$$|\pi_c - \tilde{\pi}_c|_1 \leq \frac{1}{1-\tau_1(P)} \| E \|_\infty. \tag{12}$$

The inequality in (12) implies that if the Markov chain corresponding to an uncontrolled genetic network has a small ergodicity coefficient bound, then the corresponding controlled Markov chain will also have an ergodicity coefficient that is bounded by the same bound. Consequently, if a stationary policy is designed from an estimated Markov chain that is "close" to the actual one for the network, then this policy when applied to the actual network will produce results that are close to the desired outcome, as far as the steady-state behavior is concerned.

Analysis of robustness for PMA models have been considered in [42] by investigating the maximum percentage variations that still keep the system in the range of desired behavior. Another general approach to parameter robustness and tuning based on LTL is considered in [49] and is applied to improve the robustness of a synthetic transcriptional cascade.

### Design of Robust Intervention Strategies

To explain the issue of robust design of control policies for Markovian models, let us revisit the nominal control problem described by (11). In case of uncertainties, we can parameterizing the class of transitions as $\Omega := (P_a^u)_{u \in C,\, a \in F_a}$, where $F_a$ is the noise parameter distribution. One of the ways to approach robust intervention in presence of uncertainties is to consider the worst-case scenario such that therapeutic intervention is not detrimental even for the worst patient. A minimax(worst-case) intervention policy is defined as a policy whose worst performance over the uncertainty class $\Omega$ is best among all admissible policies. The minimax robust policy, denoted by $\mu_{mm}$, is the one that satisfies [47]

$$\Phi(\Pi, \Omega, z_0) := \min_{\mu \in \Pi} \max_{\omega \in \Omega} J_{\mu,\omega}(z_0). \tag{13}$$

There is no guarantee that a minimax robust solution will exist for all distributions. For some uncertainty classes, analytical solutions are feasible. In [47], the minimax robust solution for three different uncertainty classes are worked out. The uncertainties are characterized by 1) transition probabilities in an interval, 2) finite number of transition probability matrices, and 3) upper bound on the $\infty$-norm of the difference in the actual and estimated transition probabilities. However, the condition of rectangularity [50] required for some of the analytical solutions are hardly valid in practice. The worst-case robust policy design approach is typically conservative because it gives too much importance to the scenarios that hardly occur in practice. When the objective is to avoid extremely undesirable results, a minimax design is suitable, but when the objective is to improve the expected chances of success, a Bayesian approach will be preferable. A Bayesian robust policy minimizes the average cost over the uncertainty class $\Omega$. Let $\mu_b$ denote the policy designed to be optimal at point $b$ of the parameter distribution i.e., $\mu_b$ minimizes $J_{\mu,\omega_b}(z_0)$. Here, $\omega_b$ refers to the probability transition matrix at point $b$. Let $\gamma_a(\mu_b) = E_{z_0}[J_{\mu_b,\omega_a}(z_0)]$ denote the expected cost per state at point $a$ of the parameter distribution for the intervention policy $\mu_b$, where the expectation over the states $z_0$ of the network is taken to arrive at a single value for representing the cost of a policy. Consequently, the expected cost of $\gamma_a(\mu_b)$ over the distribution $F_a$ is given by

$$E_a[\gamma_a(\mu_b)] = E_a[E_{z_0}[J_{\mu_b,\omega_a}(z_0)]], \tag{14}$$

where $E_a$ denotes expectation relative to the parameter distribution. A Bayesian control policy, denoted by $\mu_{b^*}$, is one that minimizes $E_a[\gamma_a(\mu_b)]$. The parameter value $b^*$ that achieves the minimum is called a maximally robust parameter value and the corresponding policy is called the Bayesian robust policy [51]. If the actual parameter for the distribution is $b$ and one optimizes at $b$, then there is no gain, and perhaps a loss, in using the Bayesian robust policy because $\gamma_b(\mu_b) \leq \gamma_b(\mu_{b^*})$; however, in the face of uncertainty regarding the actual parameter, the gain in applying the Bayesian robust policy is

$E_a[\gamma_a(\mu_b)] - E_a[\gamma_a(\mu_{b^*})]$, which is always nonnegative and can be substantial for parameter values far from the maximally robust one. The conservativeness of the minimax approach as compared to the Bayesian approach is shown in [51] through simulations using PBNs generated from gene expression data collected in a study of metastatic melanoma. Figure 6(a) shows the cost for policies designed at a particular noise parameter value $\sigma_b$ when applied to all possible values of the noise parameter. Figure 6(b) shows that the expected cost in using the Bayesian policy ($\mu_{b^*}$) is much less than the expected cost in using the minimax policy ($\mu_{\text{mm}}$).

## CONCLUSIONS AND FUTURE DIRECTIONS

In this article, we discussed a number of recent approaches to robust modeling and control of GRNs. The reported results indicate that significant progress has been made in investigating the robust design of intervention polices of GRNs; however, a large number of open issues still remain. We next discuss some of the open issues in this area.

- *Explore the robustness of models that encompasses transcriptome, proteome and metabolome:* Robust analysis and intervention approaches designed so far in genomics are primarily focused on one level such as transcriptome, proteome, or metabolome. With the advent of combined data sets and improved approaches to connect different experimental data sets, integrated modeling of the regulome is gaining momentum. We require robustness analysis of such integrated models along with robust control policy design of hybrid integrated models containing a combination of different time scales.
- *Robust adaptive control:* Most of the currently available robust modeling and control approaches for GRNs consider the actual biological model to be fixed and the uncertainties are assumed to be in the estimation of the parameters of the fixed biological model. However, the model connectivity and logical relationships can keep changing with time—especially in the case of genetic diseases—and we require control approaches that will be robust and also adaptively fine tune the parameters of the model to new experimental data.
- *Approximate control policy design:* Control policy design especially robust policy design encompasses a huge computational complexity and thus suboptimal approaches to control policy design have to be explored. The tuning technique based on discrete abstractions of the state space proposed in [23] is one such approach. Further approaches to control policy design for higher-order detailed nonlinear models needs to be explored.
- *Explore model free control:* Ideas for the design of control policies for translational genomics based on network measurements without the inference of the parameters of the model needs to be explored [52]. When the ultimate goal is intervention and there is huge uncertainty in estimating the actual model of the biological system, function approximators based on polynomials or neural networks can be used to design the control policy.
- *Issues in component integration and noise propagation for synthetic biology:* Combining multiple biological blocks to produce complex behaviors is being extensively studied [53], [54]. One of the many open issues in this area is the analysis and fault detection of synthetic systems that do not produce the desired behavior as predicted by the given descriptions of standardized individual parts [53]. To produce a desired output response based on the integration of simpler component devices, it is beneficial to have standardized biological blocks with properly defined input and output signals where the interactions between multiple blocks are through the input and output signals only. Thus, further research efforts are required in the direction of standardization and understanding cellular context dependent modularity in biological components [54]. The other important issue in synthetic biology is to understand and characterize the noise (intrinsic and extrinsic) in artificial gene networks. Some initial analysis of the effect of negative feedback on noise propagation in transcriptional networks is considered in [55] using simulation studies on a stochastic model of single gene and multigene cascade networks. It is shown that negative auto-regulation for single or multistage cascades do not attenuate noise, whereas oscillatory behavior can be observed for delayed negative feedback.

## ACKNOWLEDGMENT

This research was supported by NSF grant CCF0953366.

## AUTHORS

***Ranadip Pal*** (ranadip.pal@ttu.edu) received the B.Tech. degree in electronics and electrical and computer engineering from the Indian Institute of Technology, Kharagpur, in 2002 and the M.S. and Ph.D. degrees in electrical engineering from Texas A&M University, College Station, in 2004 and 2007, respectively. While at Texas A&M, he was the recipient of the Ebensberger/Fouraker Fellowship, a Distinguished Graduate Student Masters Research Award, and a National Instruments Fellowship. 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, and control of genetic regulatory networks. He is the recipient of the 2010 NSF CAREER Award and is an Indian National Math Olympiad Awardee.

***Sonal Bhattacharya*** (sonal.bhattacharya@ttu.edu) received the B.Sc. and M.Sc. degrees in physics from Calcutta University, India, in 1998 and 2001, respectively, and the post M.S. diploma in physics from Saha Institute of Nuclear Physics, Calcutta, in 2002. She is currently pursuing the Ph.D. degree in electrical and computer engineering at Texas Tech University, Lubbock. Her research areas are genomic signal processing, control of genetic regulatory networks, image processing, solid-state physics, and fluid mechanics.

***Mehmet Umut Caglar*** (umut.caglar@ttu.edu) received his B.S. degree from the Middle East Technical University Physics Department Advanced Physics program, Ankara, Turkey, in

2006. He received his M.S. degree in physics from Texas Tech University, Lubbock, in 2010. He is currently a Ph.D. student in the physics department and an M.S. student in the electrical and computer engineering department at Texas Tech University. His research areas are genetic regulatory networks, systems biology, synthetic biology, network theory, bioinformatics, and data mining.

## REFERENCES


[1] G. Karlebach and R. Shamir, "Modelling and analysis of gene regulatory networks," *Nat. Rev. Mol. Cell Biol.*, vol. 9, no. 10, pp. 770–780, Sept. 2008.

[2] H. De Jong, "Modeling and simulation of genetic regulatory systems: A literature review," *J. Comput. Biol.*, vol. 9, no. 1, pp. 67–103, 2001.

[3] Z. Szallasi, J. Stelling, and V. Periwal, *System Modeling in Cell Biology from Concepts to Nuts and Bolts*. Cambridge, MA: MIT Press, 2006.

[4] X. Cai and X. Wang, "Stochastic modeling and simulation of gene networks: A review of the state-of-the-art research on stochastic simulations," *IEEE Signal Processing Mag.*, vol. 24, no. 1, pp. 27–36, 2007.

[5] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Phys. A*, vol. 188, no. 1–3, pp. 404–425, 1992.

[6] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," *J. Chem. Phys.*, vol. 124, no. 4, 13 pages, 2006.

[7] D. T. Gillespie, "The chemical Langevin equation," *J. Chem. Phys.*, vol. 113, no. 1, pp. 297–306, 2000.

[8] M. Jordan, *Learning in Graphical Models*. Cambridge, MA: MIT Press, 1999.

[9] K. Murphy and S. Mian. (1999). "Modelling gene expression data using dynamic Bayesian networks." *Comput. Sci. Div.*, Univ. of California, Berkeley, CA, Tech. Rep. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.30.9391&rep=rep1&type=pdf

[10] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using Bayesian networks to analyze expression data," *J. Comput. Biol.*, vol. 7, nos. 3 and 4, pp. 601–620, 2000.

[11] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford Univ. Press, 1993.

[12] 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.

[13] S. Kim, H. Li, E. R. Dougherty, N. Chao, Y. Chen, M. L. Bittner, and E. B. Suh, "Can Markov chain models mimic biological regulation?," *Biol. Syst.*, vol. 10, no. 4, pp. 337–357, 2002.

[14] M. Hecker, S. Lambeck, S. Toepfer, E. van Someren, and R. Guthke, "Gene regulatory network inference: Data integration in dynamic models. A review," *Biosystems*, vol. 96, no. 1, pp. 86–103, 2009.

[15] C. Sima, J. Hua, and S. Jung, "Inference of gene regulatory networks using time-series data: A survey," *Curr. Genomics.*, vol. 10, no. 6, pp. 416–429, 2009.

[16] L. T. Macneil and A. J. Walhout, "Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression," *Genome Res.*, vol. 21, no. 5, pp. 645–657, 2011

[17] Y. Shi, M. Klustein, I. Simon, T. Mitchell, and Z. Bar-Joseph, "Continuous hidden process model for time series expression experiments," *Bioinformatics*, vol. 23, no. 13, pp. i459–i467, 2007.

[18] I. M. Ong and D. Page. (2001). "Inferring regulatory pathways in E. coli using dynamic Bayesian networks." in *Proc. Annu. Int. Conf. Intelligent Systems for Molecular Biology*. [Online]. Available: http://ismb01.cbs.dtu.dk/poster_abstracts.html#A158

[19] M. Zou and S. D. Conzen, "A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data," *Bioinformatics*, vol. 21, no. 1, pp. 71–79, 2005.

[20] W. Yu, J. Lu, G. Chen, Z. Duan, and Q. Zhou, "Estimating uncertain delayed genetic regulatory networks: An adaptive filtering approach," *IEEE Trans. Automat. Contr.*, vol. 54, no. 4, pp. 892–897, 2009.

[21] L. Qian, H. Wang, and E. Dougherty, "Inference of noisy nonlinear differential equation models for gene regulatory networks using genetic programming and Kalman filtering," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 3327–3339, 2008.

[22] J. F. Apgar, J. E. Toettcher, D. Endy, F. M. White, and B. Tidor, "Stimulus design for model selection and validation in cell signaling," *PLoS Comput. Biol*, vol. 4, no. 2, p. e30, Feb. 2008.

[23] G. Batt, B. Yordanov, R. Weiss, and C. Belta, "Robustness analysis and tuning of synthetic gene networks," *Bioinformatics*, vol. 23, pp. 2415–2422, Sept. 2007.

[24] L. Kuepfer, U. Sauer, and P. A. Parrilo, "Efficient classification of complete parameter regions based on semidefinite programming," *BMC Bioinf.*, vol. 8, pp. 12, 2007.

[25] B. Vasic, V. Ravanmehr, and A. R. Krishnan, "An information theoretic approach to constructing robust Boolean gene regulatory networks," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. PP, no. 99, 2011.

[26] I. A. Maraziotis, A. Dragomir, and D. Thanos, "Gene regulatory networks modelling using a dynamic evolutionary hybrid," *BMC Bioinf.*, vol. 11, p. 140, 2010.

[27] M. Davidich and S. Bornholdt, "The transition from differential equations to Boolean networks: A case study in simplifying a regulatory network model," *J. Theor. Biol.*, vol. 255, no. 3, pp. 269–277, 2008.

[28] J. Goutsias, "Classical versus stochastic kinetics modeling of biochemical reaction systems," *Biophys. J.*, vol. 92, no. 7, pp. 2350–2365, Apr. 2007.

[29] C. A. Gomez-Uribe and G. C. Verghese, "Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations," *J. Chem. Phys.*, vol. 126, no. 2, 12 pages, 2007.

[30] C. H. Lee, K. H. Kim, and P. Kim, "A moment closure method for stochastic reaction networks," *J. Chem. Phys.*, vol. 130, no. 13, 15 pages, 2009.

[31] R. Pal and M. U. Caglar, "Control of stochastic master equation models of genetic regulatory networks by approximating their average behavior," in *Proc. IEEE Int. Workshop Genomic Signal Processing and Statistics*, 2010.

[32] R. Pal and S. Bhattacharya, "Characterizing the effect of coarse-scale PBN modeling on dynamics and intervention performance of genetic regulatory networks represented by stochastic master equation models," *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3341–3351, 2010.

[33] R. Pal, A. Datta, and E. R. Dougherty, "Optimal infinite horizon control for probabilistic Boolean networks," *IEEE Trans. Signal Processing*, vol. 54, no. 6, pp. 2375–2387, 2006.

[34] T. Liu and D. A. Lauffenburger, *Systems Biomedicine: Concepts and Perspectives*, 1st ed. New York: Academic Press, 2009.

[35] A. Mukhopadhyay, A. M. Redding, B. J. Rutherford, and J. D. Keasling, "Importance of systems biology in engineering microbes for biofuel production," *Curr. Opin. Biotechnol.*, vol. 19, no. 3, pp. 228–234, 2008.

[36] I. Shmulevich, E. R. Dougherty, and W. Zhang, "Gene perturbation and intervention in probabilistic Boolean networks," *Bioinformatics*, vol. 18, no. 10, pp. 1319–1331, 2002.

[37] A. Datta, A. Choudhary, M. L. Bittner, and E. R. Dougherty, "External control in Markovian genetic regulatory networks," *Mach. Learn.*, vol. 52, no. 1–2, pp. 169–191, 2003.

[38] A. Datta, A. Choudhary, M. L. Bittner, and E. R. Dougherty, "External control in Markovian genetic regulatory networks: The imperfect information case," *Bioinformatics,* vol. 20, no. 6, pp. 924–930, 2004.

[39] R. Pal, A. Datta, M. L. Bittner, and E. R. Dougherty, "Intervention in context-sensitive probabilistic Boolean networks," *Bioinformatics*, vol. 21, no. 7, pp. 1211–1218, 2005.

[40] J. Clark, "Molecular targeted drugs," in *Harrison's Manual of Oncology*. B. A. Chabner, D. L. Longo, and T. J. Lynch, eds., New York: McGraw-Hill, 2007, ch. 1.10, pp. 67–75.

[41] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed., Nashua, NH: Athena Scientific, 2001.

[42] G. Batt, C. Belta, and R. Weiss, "Temporal logic analysis of gene networks under parameter uncertainty," *IEEE Trans. Automat. Contr.* (Special Issue on Systems Biology), vol. 53, pp. 215–229, 2008.

[43] K. Rinaudo, L. Bleris, R. Maddamsetti, S. Subramanian, R. Weiss, and Y. Benenson, "A universal RNAi-based logic evaluator that operates in mammalian cells," *Nat. Biotechnol.*, vol. 25, pp. 795–801, May 2007.

[44] A. Aswani and C. Tomlin, "Topology based control of biological genetic networks," in *Proc. 47th IEEE Conf. Decision and Control*, 2008, pp. 3205–3210.

[45] A. A. Julius, A. Halasz, M. S. Sakar, H. Rubin, V. Kumar, and G. J. Pappas, "Stochastic modeling and control of biological systems: The lactose regulation system of Escherichia coli," *IEEE Trans. Automat. Contr.* (Special Issue on Systems Biology), vol. 53, pp. 51–65, 2008.

[46] M. Hegland, C. Burden, L. Santoso, S. MacNamara, and H. Booth, "A solver for the stochastic master equation applied to gene regulatory networks," *J. Comput. Appl. Math.*, vol. 205, no. 2, pp. 708–724, 2007.

[47] R. Pal, A. Datta, and E. R. Dougherty, "Robust intervention in probabilistic Boolean networks," *IEEE Trans. Signal Processing*, vol. 56, no. 3, pp. 1280–1294, 2008.

[48] G. E. Cho and C. D. Meyer, "Comparison of perturbation bounds for the stationary distribution of a Markov chain," *Linear Algeb. Applicat.*, vol. 335, no. 1–3, pp. 137–150, 2001.

[49] A. Rizk, G. Batt, F. Fages, and S. Soliman, "A general computational method for robustness analysis with applications to synthetic gene networks," *Bioinformatics*, vol. 25, no. 12, pp. i169–i178, 2009.

[50] G. Iyengar, "Robust dynamic programming," *Math. Oper. Res.*, vol. 30, no. 2, pp. 257–280, 2005.

[51] R. Pal, A. Datta, and E. R. Dougherty, "Bayesian robustness in the control of gene regulatory networks," *IEEE Trans. Signal Processing*, vol. 57, no. 9, pp. 3667–3678, 2009.

[52] J. C. Spall and J. A. Cristion, "Model-free control of nonlinear stochastic systems with discrete-time measurements," *IEEE Trans. Automat. Contr.*, vol. 43, no. 9, pp. 1198–1210, 1998.

[53] B. Canton, A. Labno, and D. Endy, "Refinement and standardization of synthetic biological parts and devices," *Nat. Biotechnol.*, vol. 26, no. 7, pp. 787–793, 2008.

[54] P. Purnick and R. Weiss, "The second wave of synthetic biology: From modules to systems," *Nat. Rev. Mol. Cell Biol.*, vol. 10, no. 6, pp. 410–422, June 2009.

[55] S. Hooshangi and R. Weiss, "The effect of negative feedback on noise propagation in transcriptional gene networks," *Chaos: Interdiscip. J. Nonlinear Sci.*, vol. 16, no. 2, 10 pages, 2006.


[SP]