IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 29, 2019, accepted January 2, 2020, date of publication January 15, 2020, date of current version January 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2966229

Identifying Desirable Function Perturbations in Signaling Pathways Through Stochastic Analysis

JUN HU[1], HUI-JIA LI[2], JIE GAO[3], ZEBO ZHOU[4], AND CHENGYI XIA[5]

[1]School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100080, China
[2]School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
[3]College of Science, Southwest Petroleum University, Chengdu 610500, China
[4]School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China
[5]School of Computer and Communication Engineering, Tianjin University of Technology, Tianjin 300191, China

Corresponding authors: Hui-jia Li (hjli@mass.ac.cn) and Zebo Zhou (klinsmann.zhou@gmail.com)

This work was supported in part by the National Science Foundation of China under Grant 71871233, Grant 61601371, and Grant 41604025, in part by the Beijing Natural Science Foundation under Grant 9182015, and in part by the Fundamental Research Funds for the Central Universities under Grant 2672018ZYGX2018J082.

ABSTRACT The logical representation and analyses of biological systems provide us with valuable insights. Inherent fluctuations play an important role in the long-term behaviors of biological systems. In this work, we mainly investigate the effects of function perturbations that might contribute to the development of effective therapies. We propose a stochastic model for one-bit perturbations and perform effective analyses of biological systems subjected to function perturbations. In addition, we also consider the scenario of a multibit perturbation (i.e., a one-bit perturbation occurring in multiple columns); thereby, we can determine the most effective column combination for the purpose of maximizing or minimizing the desired signal probability. Through stochastic analysis of the caspase3 signaling pathway, the corresponding practicability and effectiveness are demonstrated. Consequently, appropriate therapies can be determined to maximize or minimize the probability of caspase3 signaling.

INDEX TERMS Stochastic analysis, signaling pathway, function perturbation, potential target or target combination.

I. INTRODUCTION

Biological functions are widely used to indicate the interactions among genes, proteins, and other intracellular molecules. Biological systems (either signaling pathways or genetic networks) are used to mediate the signal transduction inside cells and have been extensively studied. The chains of interactions have been actively investigated to better understand biological processes at the system level. Thus far, to gain biological insights, various approaches have been proposed, including logical models [1]–[3] and continuous models using differential equations [4]–[7]. Among those models, logic models, especially Boolean networks (BNs), have been thoroughly studied and have been adopted as an informative and effective means of qualitatively modeling biological systems [2], [3], [8], [9].

For biological systems, considerable research has focused on investigating dynamics or performing robustness analysis of situations with fluctuations. State perturbations and function perturbations are predominantly investigated. In [10], the effects of gene perturbations were thoroughly investigated with the adoption ofprobabilistic BNs (PBNs). Moreover, the effects of stochasticity and variability were further investigated in [11]. Furthermore, state perturbations in multivalued scenarios, usually referred to as probabilistic multiple-valued networks (PMNs), were studied in [12], [13]. To improve the efficiency of analyzing systems withstate perturbations, stochastic models have been proposed for PBNs and PMNs in [14] and [15], respectively. In addition, the authors in [16] consider the effects of state perturbations in context-sensitive genetic networks. Sometimes, the dysfunction of a certain gene node might lead to an undesirable transition from a normal network to a defective one. Better understanding of the importance of molecules can provide valuable insights for the identification of potential drug targets [17], [18]. Besides, the authors in [19] thoroughly studied the node vulnerability in signaling pathways where a certain gene node was deliberately suppressed or mutated [19]. However, these methods are

The associate editor coordinating the review of this manuscript and approving it for publication was Baoping Cai.

usually very time consuming and have a high computational complexity, which hinder the further applications in the real worlds.

Consequently, we mainly focus on varying the rule-based structure, i.e., function perturbation, which also provides valuable insights into biological systems. Function perturbation is used to model the influences of uncertainty and latent variables that can be formulated by varying the updating functions. Structural intervention has the potential to alter the dynamic behavior permanently while enabling the system to evolve in a desirable direction. In [20], the authors thoroughly investigated the effects of function perturbation on the attractors of BNs. To perform an analytic study of function perturbation, perturbation theory was used to analyze finite Markov chains in [21]. Moreover, the impacts of function perturbations were also investigated and applied to a *D. melanogastergene network* [22]. In [23], algorithms were proposed to identify the optimal one-bit perturbation strategy for BNs. For simplicity, the function perturbation, including one-bit perturbations and multi-bit perturbations, in this work mainly refers to the alteration of updating rules. One-bit perturbation refers to a one-bit flip (i.e., changing from 1 to 0, or vice versa) of the truth table. However, these methods are usually not able to obtain the explicit solutions, which can't be use to real applications directly.

To obtain therapeutic benefits, we might determine desirable strategies by applying function perturbations to certain molecules. Thus, attractive targets can be determined to facilitate the design of intervention strategies aiming to alter the direction of a system's evolution or to maximize (or minimize) the signal probabilities. As discussed above, stochastic analysis can improve the efficiency of investigating biological systems. Thereby, we try to perform a stochastic analysis of systems with function perturbations in this work. Furthermore, as in [24], a pair of molecules with high vulnerability typically includes a highly vulnerable molecule. Inspired by that finding, we also consider multi-bit perturbation for the purpose of maximizing or minimizing signal probabilities.

Overall, this paper makes the following three folds contributions:

1. first, stochastic models for biological systems are proposed with relationships that are implemented by logic gates.

2. Then, we perform the corresponding stochastic analysis in order to study the effects of function perturbations on signaling pathways effectively. To identify desirable strategies, we also study the impacts of applying function perturbations (either single column or column combinations) to different molecules.

3. To save space without losing generality, in the present study, we mainly focus on the investigation of signaling pathways. Moreover, with modifications, the analyses in this work are also applicable to multistate scenarios.

The remainder of the paper is organized as follows. Section II first reviews the fundamentals related to the stochastic computational approach. Section III presents an introduction to one-bit perturbations and the corresponding stochastic architecture. In Section IV, the model is generalized subsequently to consider multi-bit perturbation. Subsequently, analyses of several benchmarks are performed in Section V; moreover, we also study the effects of incorporating function perturbation into the caspase3 signaling pathway through stochastic analysis. Finally, Section VI concludes the manuscript.

II. STOCHASTIC COMPUTATION

A. PGMs FOR RELIABILITY EVALUATION

A probabilistic gate model (PGM) relates the output probability of a gate to its input and error probabilities; this is accomplished according to the function and malfunction (such as in the presence of an error) of the gate [30]. In general, the output probability of a gate can be calculated by the following equation,

$$Z = P(output\text{“1”}|gatefaulty) \cdot P(gatefaulty) + P(output\text{“1”}|gatenotfaulty) \cdot P(gatenotfaulty). \quad (1)$$

Consider a von Neumann fault, i.e., a fault that flips the correct output of a gate and resembles the behavior of a soft error. Let'sdenotethe errorrate, i.e., s = P(gate faulty), and p, the fault-free output probability, i.e., p = P(output "1"| gate not faulty). The following equation is then applicable to any logic gate/function for the calculationof its output probability,

$$Z_v = (1 - p) \cdot \varepsilon + p \cdot (1 - \varepsilon). \quad (2)$$

Stuck-at faults can also be modeled in a PGM. For a stuck-at-1 fault, (6) becomes

$$Z_{SA1} = \varepsilon + p \cdot (1 - \varepsilon) \quad (3)$$

For a stuck-at-0 fault, this is given by

$$Z_{SA1} = p \cdot (1 - \varepsilon). \quad (4)$$

An accurate algorithm using the PGMs accounts for signal dependencies in a circuit [30]. If all inputs are mutually independent, reconvergentfanouts are the only topological structures that introduce signal correlations in a circuit with no feedback. Signalcorrelationscanbe eliminatedby decomposing a circuit into twosub-circuits for each reconvergent fanout. When all reconvergent fanouts are eliminated by this fanout decomposition, the gate PGMs can then be applied to obtain the reliability of the original circuit. As the required computation almost doubles for each reconvergentfanout, however, the PGM algorithm has a computational complexity that increases exponentially with the number of dependent reconvergent fanouts [30]. As applicable to any analytical approach, the accurate analysis of large circuits is therefore likely to be intractable due to its very large computational overhead.

B. STOCHASTIC COMPUTATIONAL MODELS (SCMs)

For reliable circuit design, a stochastic computational approach was initially presented in the 1960s [25]. For

stochastic analysis, signal probabilities are indicated by random bit streams (also referred to as stochastic sequences) by setting a proportional number of bits to a specific value [26].

This computational capability of stochastic logic allows the numerical evaluation of circuit reliability using stochastic computational models(SCMs).SCMs are based on the operation of stochastic logic and the notions of PGMs. As discussed previously, any gate affected by a von Neumann fault can be modeled by (2). In fact, (2) can be implemented by the stochastic logic of an XOR gate [31], as follows:

$$XORcto(p, \varepsilon) = p(1-\varepsilon) + (1-p)\varepsilon, \tag{5}$$

where p is the fault-free output probability and ε is the gate error rate. The special case of a stochastic XOR is used to compute (2) because gate errors are assumed to occur independently. The general model must be used if there is a correlation between the gate error and the input signals. (5) shows that regardless of the type of logic gate modeled by PGM, (2) can be implemented by a stochastic XOR logic. Therefore, an SCM can be obtained by adding an XOR gate to an unreliable gate and using an input of XOR to implement the gate error rate. In this case,

$$p = P(X1 = 1 \text{ and } X2 = 1), \tag{6}$$

and the XOR gate computes (5).

In addition to the von Neumann fault that was originally modeled in [31], the stuck-at faults can also be modeled by SCMs. For (3) considering a stuck-at-1 fault, an SCM can be constructed by adding an OR gate to the unreliable gate and using an input of the OR to implement the gate error rate, as

$$ORcto(p, \varepsilon) = p + \varepsilon - p \cdot \varepsilon = \varepsilon + p \cdot (1-\varepsilon). \tag{7}$$

For a stuck-at-0 fault, AND and NOT gates are used to implement the function of (8):

$$ANDcto(p, \varepsilon) = p \cdot (1-\varepsilon). \tag{8}$$

As indicated in (6), (7) and (8), an SCM is universal, because it can be constructed for an arbitrary logic gate. Moreover, it also masks errors through the function of a logic gate; so logic masking is explicitly considered in an SER analysis. Signal correlations are inherently accounted, so the useofSCMssignificantlyreducesthecomputational complexity of a probabilistic analysis by using redundancy in the time domain and stochasticlogicfor processingand calculating the gate error rate.

Fig. 1 illustrates the logic gates adopted for stochastic computation in this manuscript. As shown in Fig. 1(a), an inverter can efficiently compute the complement of a probability, whereas the multiplication of two independent probabilities is implemented with the adoption of an AND gate (Fig. 1(c)). Moreover, the weighted sum can be efficiently implemented by the 2-to-1 multiplexer shown in Fig. 1(e), and the output is affected by the distributions of zeros and ones in the control sequence. For a simple example, a sequence

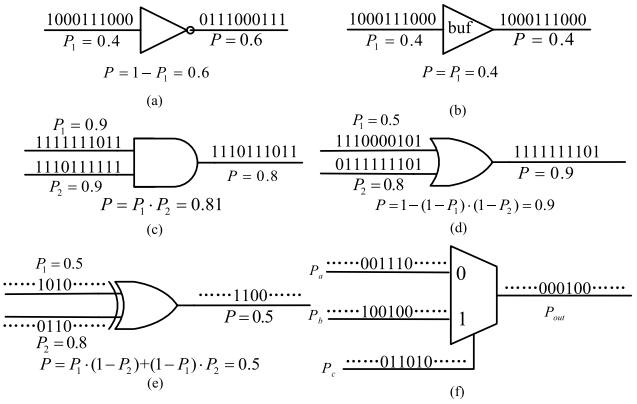


FIGURE 1. Boolean logic gates for stochastic computation. (a) An inverter; (b) buffer; (c) AND; (d) OR; (e) XOR; (f) 2-to-1 multiplexer.

of 10 bits is adopted to encode the probabilities, as shown in Figs. 1(a)–(d). To achieve a significantly higher accuracy, adopting a longer sequence length is a reasonable method, as shown in Figs. 1(e)–(f).

By propagating the binary sequences through logic gates, a probabilistic computation can be efficiently conducted. However, the obtained results are not deterministic but are probabilistic because of the inevitable stochastic fluctuations. With the adoption of non-Bernoulli sequences, the fluctuations can be reduced to a large extent [26]. That is, if a reasonable sequence length is adopted, the inaccuracy can be significantly reduced or can even be negligible.Moreover, signal correlations are also inherently handled by the bitwise dependencies.

III. ONE-BIT PERTURBATION

For biological systems, we usually assume that each molecule possesses two activity states (i.e., active or inactive, usually indicated by 1 and 0, respectively). For a simple example, a BN consisting of n nodes (denoted $x_1, x_2, \cdots, x_n$) is investigated here, and the network is described as:

$$\begin{cases} x_1(t+1) = f_1(x_1(t), x_2(t), \cdots, x_n(t)) \\ x_2(t+1) = f_2(x_1(t), x_2(t), \cdots, x_n(t)) \\ \qquad\qquad \vdots \\ x_n(t+1) = f_n(x_1(t), x_2(t), \cdots, x_n(t)) \end{cases} \tag{9}$$

where f_i is a Boolean updating function for x_i. Similarly, signaling pathways can also be described by representing relationships among molecules with Boolean functions. Given a molecule x_i with k_i parent molecules, i.e., $x_{i1}, x_{i2}, \cdots, x_{ik_i}$ ($k_i \leq n$), the state of x_i at time $t+1$, i.e., $x_i(t+1)$, is represented as:

$$x_i(t+1) = f_i(x_{i1}(t), x_{i2}(t), \cdots, x_{ik_i}(t)) \tag{10}$$

where k_i represents the indegree for molecule x_i and f_i denotes the corresponding Boolean function. According to the definition of a one-bit perturbation in Section 1, if the perturbation occurs at the j_{th} entry (i.e., flipping the value at the j_{th} entry), the perturbation is denoted $f_i \to f_i^{(j)}$.

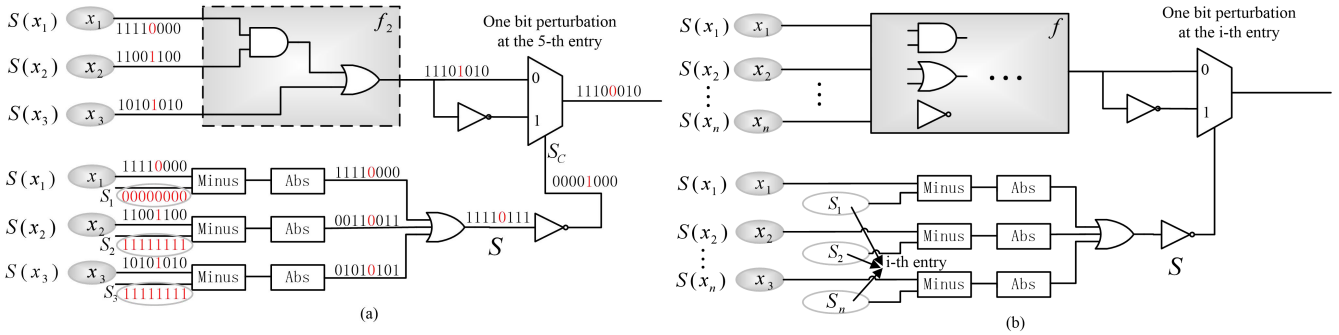


FIGURE 2. (a) Stochastic model for $f_2 \rightarrow f_2^{(5)}$; (b) a general stochastic model.

TABLE 1. Truth table for f_2 (x_1, x_2, x_3).

Label	1	2	3	4	5	6	7	8
$x_1x_2x_3$	111	110	101	100	011	010	001	000
f_2	1	1	1	0	1	0	1	0

TABLE 2. Signal probability distribution of x_1, x_2, x_3.

$p(x_1) = 0.4, p(x_2) = 0.3, p(x_3) = 0.8$								
$x_1x_2x_3$	111	110	101	100	**011**	010	001	000
f_2	1	1	1	0	**1**	0	1	0
$p(\times 10^{-3})$	96	24	224	56	**144**	36	336	84
$\tilde{f}_2$	1	1	1	0	**0**	0	1	0

Given the occurrence of a function perturbation, the corresponding output signal probability distribution is likely affected. These changes might be helpful in diagnosing changes caused by a disease, radiationtherapy, drug treatment, etc.Nevertheless, function perturbation is studied in this manuscript from another perspective (specifically, not as an identification issue but a utilization issue). We focus on developing function perturbation strategies to encourage the system to evolve in a desirable direction. We know that BNs and signaling pathways can both be described by Boolean functions, as indicated by (9) and (10), respectively; hence, the function perturbation models for BNs are also applicable to signaling pathways.

Example 1: Here, we consider a BN consisting of three molecules; f_2 represents the updating function for x_2 and is denoted $f_2\,(x_1, x_2, x_3) = OR(AND(x_1, x_2), x_3)$, where x_2 is a molecule with an indegree of three. According to the analysis of f_2, the corresponding truth tableis listed inTable 1.

With the provided signal probabilities for x_1, x_2 and x_3 (i.e., $p(x_1) = 0.4$, $p(x_2) = 0.3$, $p(x_3) = 0.8$), the probabilities for different state combinations are illustrated in Table 2. Assuming a one-bit perturbation occurs at the 5_{th} column (i.e., $f_2 \rightarrow f_2^{(5)}$), as we see in Table 2, the value at the 5_{th} entry $f_2\,(0, 1, 1) = 1$ will switch to $\tilde{f}_2\,(0, 1, 1) = 0$.

According to Table 2, the probabilities for x_2 being 1 and 0 before and after the one-bit perturbation are calculated as

$$p(1) = 0.096 + 0.024 + 0.224 + 0.144 + 0.336 = 0.824$$

and

$$p(0) = 0.056 + 0.036 + 0.084 = 0.176,$$

respectively. However, given the occurrence of $f_2 \rightarrow f_2^{(5)}$, the corresponding signal probabilities are calculated as

$$p'\,(1) = 0.096 + 0.024 + 0.224 + 0.336 = 0.68$$

and

$$p'\,(0) = 0.056 + 0.036 + 0.144 + 0.084 = 0.32,$$

respectively. Similar analysis can be performed if $f_2 \rightarrow f_2^{(i)}$ occurs ($j \in \{1,2, \cdots, 8\}$); the corresponding results obtained through accurate analysis are presented in Table 3.

For a simple example, a one-bit perturbation at $f_2\,(x_1, x_2, x_3) = OR(AND(x_1, x_2), x_3)$ is used here. A stochastic architecture for $f_2 \rightarrow f_2^{(5)}$ is presented in Fig. 2(a). Initially, the signal probabilities for x_1, x_2, x_3 are encoded as stochastic sequences (i.e., $S\,(x_1)$, $S\,(x_2)$, and $S\,(x_3)$, respectively (a longer sequence is usually required to improve the accuracy for stochastic analysis)). By propagating sequences through the stochastic architecture as shown in Fig. 2(a), the output sequence can be efficiently derived and further used to determine the signal probability.

For simplicity, 8 bits from the sequences (corresponding to the number of possible combinations of x_1, x_2, x_3 as shown in Table 2) are adopted for an illustration here. For the input combination 011, the output equals 1, which is anticipated to be switched to 0 given the occurrence of a function perturbation at the 5_{th} entry (marked in red). This switch can be easily implemented with the adoption of a 2-to-1 multiplexer (with a control sequence of S_C). For S_C, if the function perturbation is occurring at the i_{th} entry, then the entry corresponding to the anticipated state combination should be 1. For instance, in Fig. 2(a), we are supposed to perturb the 5_{th} entry; thus, the entries with an input combination of 011 equal 1. According to the mechanism of the multiplexer, if $(S_C)_i = 1$, the corresponding bit in the output sequence for f_2 is flipped, which is efficiently implemented by an inverter; otherwise, the corresponding bit is maintained. Hence, a control sequence of 00001000 is necessary for $f_2 \rightarrow f_2^{(5)}$.

TABLE 3. Signal probabilities for x_2 with the occurrence of $f_2 \rightarrow f_2^{(f)}$. For stochastic analysis, the adopted sequence length is 10K.

$f_2 \rightarrow f_2^{(j)}$	Accurate analysis		Stochastic analysis	
	$p'(1)$	$p'(0)$	$p'(1)$	$p'(0)$
at 1_{st} entry	0.728	0.272	0.7302	0.2698
at 2_{nd} entry	0.8	0.2	0.8	0.2
at 3_{rd} entry	0.6	0.4	0.5991	0.4009
at 4_{th} entry	0.88	0.12	0.8810	0.1190
at 5_{th} entry	0.68	0.32	0.6828	0.3172
at 6_{th} entry	0.86	0.14	0.8616	0.1384
at 7_{th} entry	0.488	0.512	0.4874	0.5126
at 8_{th} entry	0.908	0.092	0.9040	0.0960

Thus, the remaining problem is proposing a stochastic architecture to derive S_C from $S(x_1)$, $S(x_2)$, and $S(x_3)$. As indicated by the example in Fig. 2(a), we can construct the architecture with the adoption of minus and abs gates, where the minus gate outputs the difference between two inputs and the abs gate gives the absolute value of the input. For the 5_{th} entry, S_1, S_2 and S_3 can be adopted to encode the anticipated entry. By feeding the sequences through the stochastic architecture, we can obtain $S'(x_1)$, $S'(x_2)$, and $S'(x_3)$. As we see, the bits at the 5_{th} entries equal 0; then, an OR gate can be adopted. Hence, a sequence S is obtained; the bit corresponding to the combination 011 equals 1, while the other bits equal 0. That is, the previously mentioned control sequence for the 2-to-1 multiplexer is derived with the proposed stochastic architecture.

Accordingly, one-bit perturbation at any entry ($f_2 \rightarrow f_2^{(i)}$) can be easily implemented by varying S_1, S_2 and S_3. Moreover, for a general function f, the corresponding stochastic model can be easily derived with the adoption of combinations of logic gates, as presented in Fig. 2(b).

The results for perturbations at different entries are presented in Table 3 and are obtained through different approaches. As shown in Table 3, various output probabilities are obtained by applying one-bit perturbations at different entries. Thus, different perturbation strategies can be adopted according to our expectations. Moreover, as we see in Table 3, stochastic analysis can predict the output probability approximately with an acceptable inaccuracy. The disparity incurred by the inevitable stochastic fluctuations can be reduced greatly by increasing the sequence length.

If $f_2 \rightarrow f_2^{(2)}$ occurs, the results obtained by stochastic analysis are the same as those obtained by an accurate analysis. For $f_2 \rightarrow f_2^{(2)}$, the truth table is further presented in Table 4. Here, the signal probability after $f_2 \rightarrow f_2^{(2)}$ is calculated as $p'(1) = p(x_3) = 0.8$ as f_2 is changed to $\tilde{f}_2 = x_3$, which holds regardless of the states of x_1 and x_2. Thereby, as in Table 3, the results obtained through stochastic analysis with the occurrence of $f_2 \rightarrow f_2^{(2)}$ are accurate, as $p(x_3)$ can be accurately encoded.

For x_2, the state of 1 is anticipated to be desirable. The results of applying a one-bit perturbation at different entries are presented Fig. 3. As we see in Fig. 3, we can apply a one-bit perturbation at the 4_{th}, 6_{th}, or 8_{th} entry for the purpose of increasing the probability of being expressed (i.e., $p(1)$). Moreover, applying a one-bit perturbation at the 8_{th} entry results in the highest probability, whereas the perturbation occurring at the 7_{th} entry is the most undesirable, as the lowest probability is obtained.

TABLE 4. Signal probability distribution of x_1, x_2, x_3 when $f_2 \rightarrow f_2^{(2)}$.

$p(x_1) = 0.4, p(x_2) = 0.3, p(x_3) = 0.8$								
$x_1x_2x_3$	111	110	101	100	011	010	001	000
f_2	1	1	1	0	1	0	1	0
$\tilde{f}_2$	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**
$p(\times 10^{-3})$	96	24	224	56	144	36	336	84

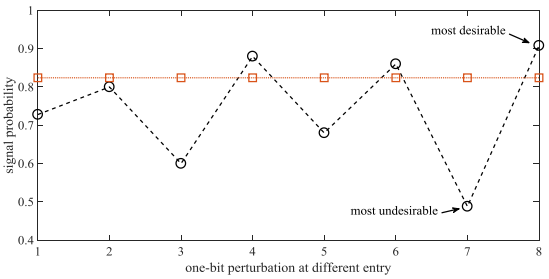


FIGURE 3. Signal probabilities obtained through stochastic analysis (the sequence length is 10k).

IV. MULTIBIT PERTURBATIONS

Furthermore, the previously discussed perturbations might also occur at multiple entries, and the output probability is likely to be affected to a larger extent than that for the incorporation of only a one-bit perturbation. The stochastic model of f_2 with a two-bit perturbation (occurring at the 3_{rd} and 5_{th} entries) is presented in Fig. 4(a) (for simplicity, the sequence length is 8 bits here, corresponding to the possible combinations of x_1, x_2 and x_3). As shown in Fig. 4(a), after propagating $S(x_1)$, $S(x_2)$, and $S(x_3)$ through the stochastic architecture in the lower box, two stochastic sequences S_1 and S_2 are obtained. Please refer to the explanation for the architecture in Fig. 2 in Section III for details. If $S_{1,i} = 1$, then the outputs of the entries corresponding to $(S(x_1), S(x_2)S(x_3)) = 011$ are anticipated to be flipped. Similarly, the outputs for entries corresponding to 101 will also be changed. Here, in order to incorporate two-bit perturbation, an OR gate is used to combine S_1 and S_2. Hence, the control sequence for the 2-to-1 multiplexer with which the anticipated column combination is to be perturbed can be obtained. Similarly, any possible combination of two-bit perturbation can be easily implemented. Moreover, a general stochastic model can be constructed for a general function f, as illustrated in Fig. 4(b). The simulation results are presented in Fig. 5 for two-bit perturbation (3_{rd} and 5_{th}) via stochastic analysis Here, for simplicity, we also assume that the state of 1 for x_2 is desirable. According to the analysis in Section III, a one-bit perturbation at the 4_{th}, 6_{th}, or 8_{th} entry can increase the likelihood of x_2 being expressed. As we see in Fig. 5, two-bit perturbations can have a greater effect than one-bit perturbations in increasing the expressing or suppressing the output. Among

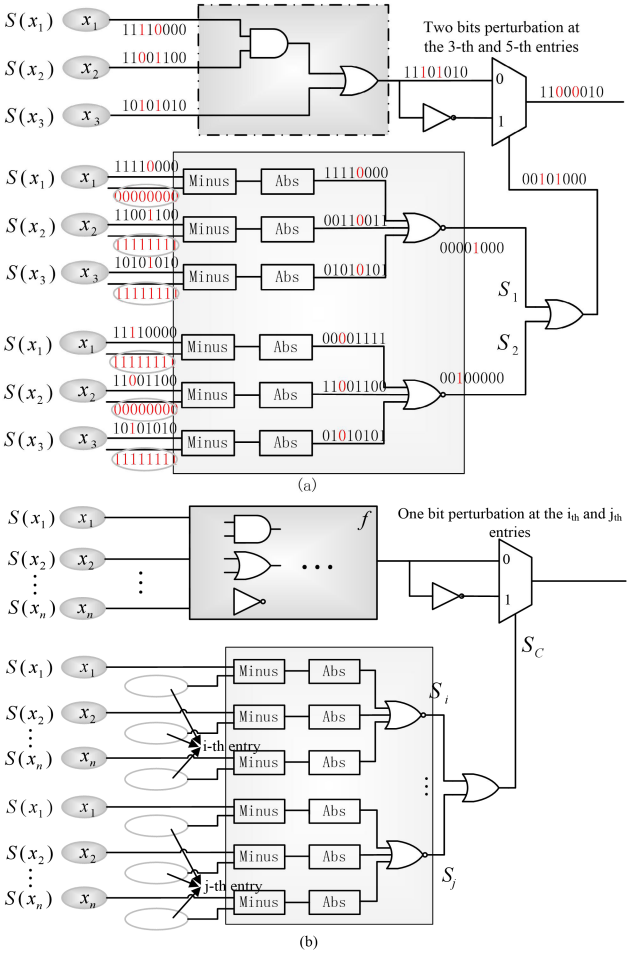


FIGURE 4. **(a) Stochastic model for performing perturbations at the 3rd and 5th entries in f_2; (b) a stochastic model for a general function f with multibit perturbations.**

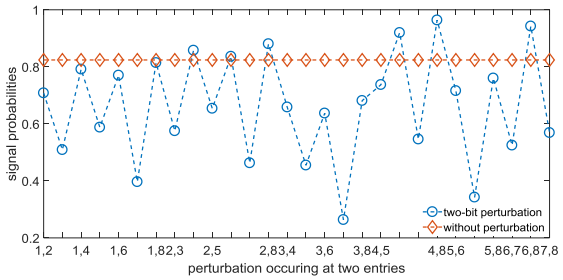


FIGURE 5. **Signal probabilities for two-bit perturbations at the 3rd and 5th entries through stochastic analysis (here, the sequence length is 10k).**

the possible combinations in Fig. 5, the entry combinations of (2,4), (2,6), (2,8), (4,6), (4,8), and (6,8) are options to be perturbed. As in [24], we know that the pair of molecules with a higher vulnerability usually includes a highly vulnerable molecule identified by the single fault analysis. Here, we can draw a similar conclusion that a desirable two-bit or multibit perturbation is usually composed of combinations including a desirable one-bit perturbation. Furthermore, the highest probability is obtained by performing perturbations at the 4_{th} and 8_{th} entries. Through comparison with the results in Fig. 3,

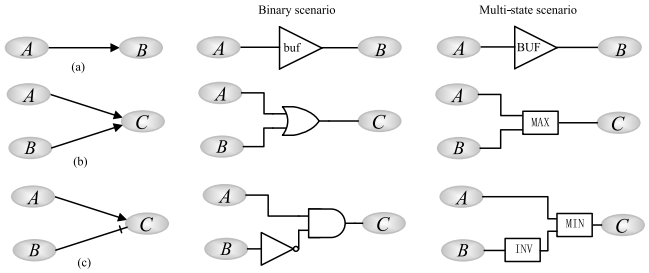


FIGURE 6. **(a) A activates B directl; (b) scenario without inhibition; (c) scenario with inhibition.**

we can see that for multiple-bit perturbations, the incorporation of a desirable/undesirable one-bit perturbation will result in a higher/lower output probability. For the entry combination of (3,7), an even lower probability is obtained compared with that of the other strategies.

V. APPLICATION OF THE PROPOSED MODEL

A. CONSTRUCTION OF THE STOCHASTIC PATHWAY

For a signaling pathway, the complex relationships among molecules can be represented by combinations of logic gates. As in [17], the following assumptions are also adopted: if the state of a molecule is affected by those of multiple inputs and no inhibitors exist, then the molecule will be active if any one of the inputs is active. Such relationship can be easily modeled by an OR gate, as shown in Fig. 6(a). Moreover, if at least one of those inputs is inhibited, then the investigated molecule will be inactive as long as at least one of the inhibitors is active, as indicated in Fig. 6(b).

Usually, an increased level of granularity indicated by multivalued variables is introduced for accurate representation of the states. Without loss of generality, a ternary pathway is analyzed in [15], where a molecule could be active, partially active, or inactive, denoted by 2, 1, and 0, respectively. This representation is biologically meaningful, as the activities or protein levels of molecules are proved to be partially affected. For a ternary signaling pathway, the corresponding input-output relationships can also be represented with the adoption of ternary logic gates, as shown in Fig. 6. For instance, for a binary pathway, if B is directly activated by A, then the relationship is indicated by $B = buf(A)$. In contrast, for a ternary pathway, the relationship is denoted $B = BUF(A)$. This relationship is also applicable to multistate pathways. Hence, a stochastic model for multiple-valued scenarios with function perturbations can be constructed with minor modifications (replacing the binary gates with multivalued ones). To perform stochastic analysis for multivalued scenarios, randomly permuted sequences consisting of fixed numbers of multiple values can be adopted, as in [15].

B. ANALYSIS OF THE CASPASE3 NETWORK

Caspase3, which plays an important role in cell death and survival, is a well-characterized molecule. The widely utilized signaling pathway shown in Fig. 7 is adopted here to regulate the activity of *caspase*3. As we see in Fig. 7, the investigated

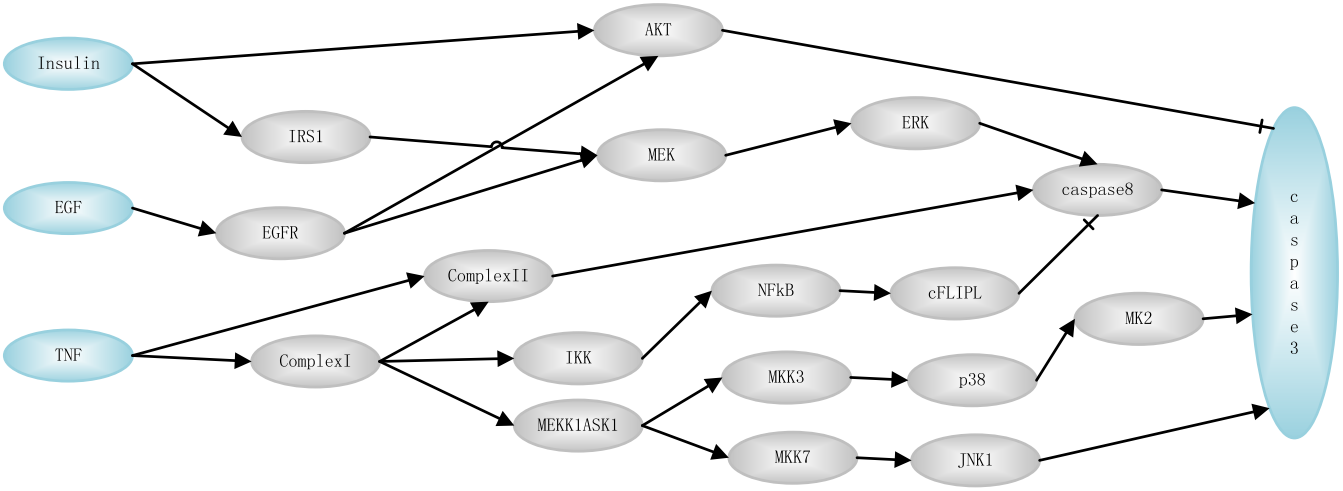


FIGURE 7. Illustration of the caspase3 network. An arrow indicates a positive regulatory interaction or activation between two genes, while a blunt arrow represents an inhibitory relationship [17].

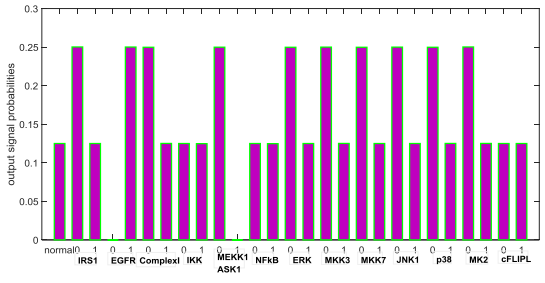


FIGURE 8. Signal probabilities for molecules with 1 indegree subjected to a one-bit perturbation.

caspase3 pathway typically originates from *EGF*, insulin, and *TNF*, and the pathway is composed of 21 molecules (including three input molecules, 17 intermediate molecules, and one output molecule). Such a pathway for caspase3, as shown in Fig. 10, was experimentally verified in [27], [28]. For an illustration, the input signal probabilities are set to $p(Insulin) = 0.5$, $p(EGF) = 0.5$, and $p(TNF) = 0.5$. Moreover, to investigate the relationship between the input probability and desirable strategies, different input probability distributions are also considered later.

First, a one-bit perturbation is applied to the molecules with one indegree. According to the analysis in Section V, the corresponding stochastic architecture can be constructed for the caspase3 pathway, as shown in Fig. 7. If a one-bit perturbation is considered, the stochastic architecture in Fig. 2(b) can be adopted. The results obtained through stochastic analysis are presented in Fig. 8. With the incorporation of a one-bit perturbation, the mean value of the probability for caspase3 is determined to be 0.1250 through stochastic analysis. Here, for each simulation, the result is probabilistic and approximate, and inaccuracy occurs due to the inevitable stochastic fluctuations, which can be reduced greatly with the adoption of a reasonable sequence length.

Nevertheless, as shown in Fig. 8, regardless of whether one-bit perturbations occur at the 1_{st} or 2_{nd} entry for *IKK*, *NFkB* and *cFLIPL*, the corresponding output probability is not changed. This result occurs because in the investigated pathway, if *ComplexI* is activated (i.e., 1), then the subsequent molecules *ComplexII*, *IKK* and *MEKK*1*ASK*1 will also be activated. Hence, caspase8 will be suppressed. Thus, the resulting value of the expression OR(*MK*2, *caspase*8, *JNK*1) is one. In such a case, *MEKK*1*ASK*1 remains activated even if a one-bit perturbation occurs at the 2_{nd} entry for *IKK*. However, if *ComplexI* is inactivated (i.e., 0), then the state of *caspase*8 is determined by that of *ERK*. The updating function for *AKT* is the same as that for *ERK*, i.e., OR(*Insulin*, *EGF*). Thus, the value of the expression OR(*MK*2, *caspase*8, *JNK*1) is the same as that of *ERK*. The expression for caspase3 is represented as AND(not(*AKT*), OR(*MK*2, *caspase*8, *JNK*1)). Hence, caspase3 is suppressed if *ComplexI* is inactivated. Even if the 1_{st} entry for *IKK* is perturbed, *caspase*3 remains suppressed. Similar analysis can be performed for the one-bit perturbation of *NFkB* and *cFLIPL*.

We see that the probability increases to approximately 0.25 when the one-bit perturbation occurs at the 1st entry of a molecule from {*IRS*1, *ComplexI*, *ERK*, *MKK*3, *MKK*7, *JNK*1, *p*38, *MK*2}, whereas the corresponding probabilities remain the same for perturbations at the 2_{nd} entry. Moreover, if the probability of caspase3 is anticipated to be small, the perturbation strategy of affecting the 1_{st} entry of *EGFR* or the 2_{nd} entry of *MEKK*1*ASK*1 can be adopted. According to the results presented in Fig. 8, strategies to maximize/minimize the probability of caspase3 can be determined easily.

Moreover, we also consider the scenario of function perturbation occurring at molecules with an indegree of 2, e.g., *AKT*, *ComplexII*, and *MEK*. First, the effects of one-bit perturbations are incorporated, with the corresponding probabilities being presented in Fig. 9. If a one-bit perturbation is applied to the 1st entry of *ComplexII* or *MEK*, the probability increases to approximately 0.25; otherwise, the

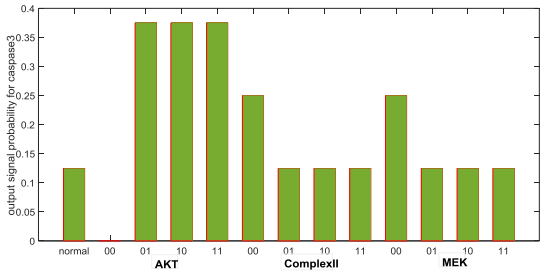


FIGURE 9. Signal probabilities for molecules with an indegree of 2 subjected to a one-bit perturbation.

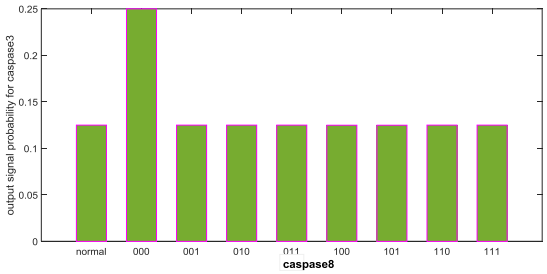


FIGURE 10. Signal probabilities obtained when caspase8 is subjected to a one-bit perturbation.

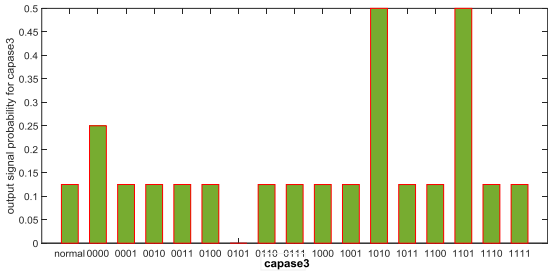


FIGURE 11. Signal probabilities if caspase3 is subjected to a one-bit perturbation.

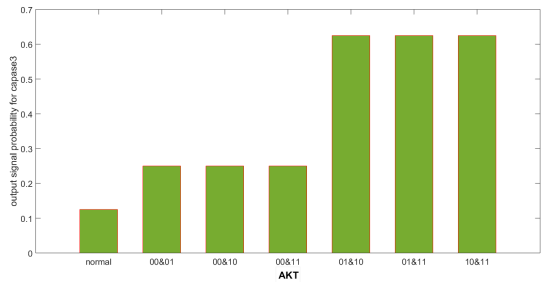


FIGURE 12. Signal probabilities for caspase3 obtained if a two-bit perturbation is applied to AKT.

corresponding value remains the same. In contrast, for *AKT*, the probability equals approximately 0.375 if the 2_{nd}, 3_{rd}, or 4_{th} entry is affected by a one-bit perturbation, whereas the result decreases to zero if the 1_{st} entry is perturbed. Thus, we can conclude that molecules with an indegree of two are usually better options than those with an indegree of one to maximize the output probability. As indicated in Fig. 9, compared to other molecules, *AKT* is the best candidate for the purpose of minimizing the probability. This result is because of the updating function for *AKT*, i.e., *AKT* = OR(*Insulin*, *EGFR*). Given a one-bit perturbation on the 1st entry, *AKT* is activated if (*Insulin*, *EGFR*) is 00. That is, caspase3 will be suppressed in such a case.

For *caspase*8 and *caspase*3, the indegree values are 3 and 4, respectively. Fig. 10 illustrates the results of applying a one-bit perturbation at different entries of the updating function for *caspase*8, and the corresponding results for caspase3 are presented in Fig. 11. Similar to Fig. 10, the probability is doubled if a one-bit perturbation is applied to the entry corresponding to 000; otherwise, the value remains unchanged. Although *caspase*8 is also directly connected to *caspase*3, the activation is less sensitive than the inhibition. Thus, determination of a good target to maximize the signal probability is not only affected by the distance from the output molecule but also by the corresponding relationship.

Compared with that for the scenario without perturbation, the probability increases to 0.5 if the entry corresponding to 1010 or 1101 is chosen for perturbation. If a one-bit perturbation is applied to the entry corresponding to 0000, the probability is 0.25, whereas the value decreases to zero if the entry corresponding to 0101 is perturbed. Otherwise, the probability remains the same. By comparison with the previous results in Figs. 8-10, we can see that for a one-bit perturbation, the output molecule itself is always the best option (here, a maximum value of 0.5 is obtained). Overall, by introducing a one-bit perturbation, several strategies can be adopted, including perturbing 0101 for caspase3, 00 for *AKT*, or 0 for *EGFR*, to derive an output probability of zero.

For molecules with an indegree greater than one, the effect of multibit perturbation was further investigated. Other than caspase3 itself, *AKT* is also a good option, as indicated by Fig. 9, which was obtained according to the analyses in [17]–[29]. Here, two-bit perturbation is conducted on *AKT* for an illustration. The stochastic model in Fig. 6(b) can be used for stochastic analysis, and the obtained results are presented in Fig. 12.

As shown in Fig. 12, the output probability increases to 0.6250 if a one-bit perturbation is applied to 01&10, 01&11, or 10&11 for *AKT*. In contrast, as indicated in Fig. 9, the corresponding probability is 0.375 if a one-bit perturbation is applied to 01, 10, or 11. We see that the combination of any two elements in {01, 10, 11} can yield a higher signal probability. This finding is consistent with our previous conclusion that a desirable strategy for multibit perturbation always includes the desirable strategy for the one-bit perturbation.

Furthermore, the effects of varying the input probability distributions are also considered. The corresponding output signal probabilities are presented in Fig. 13. For simplicity, only a one-bit perturbation for molecules with an indegree of one is incorporated. A similar analysis can be conducted accordingly.

We can see that in Fig. 13, the corresponding output probability is substantially affected by the initial input probability distributions. To minimize the output probability, a one-bit perturbation at the 1_{st} entry for *EGFR* or the 2_{nd} entry for *MEKK*1*ASK*1 is a good option. For *EGFR*, if the

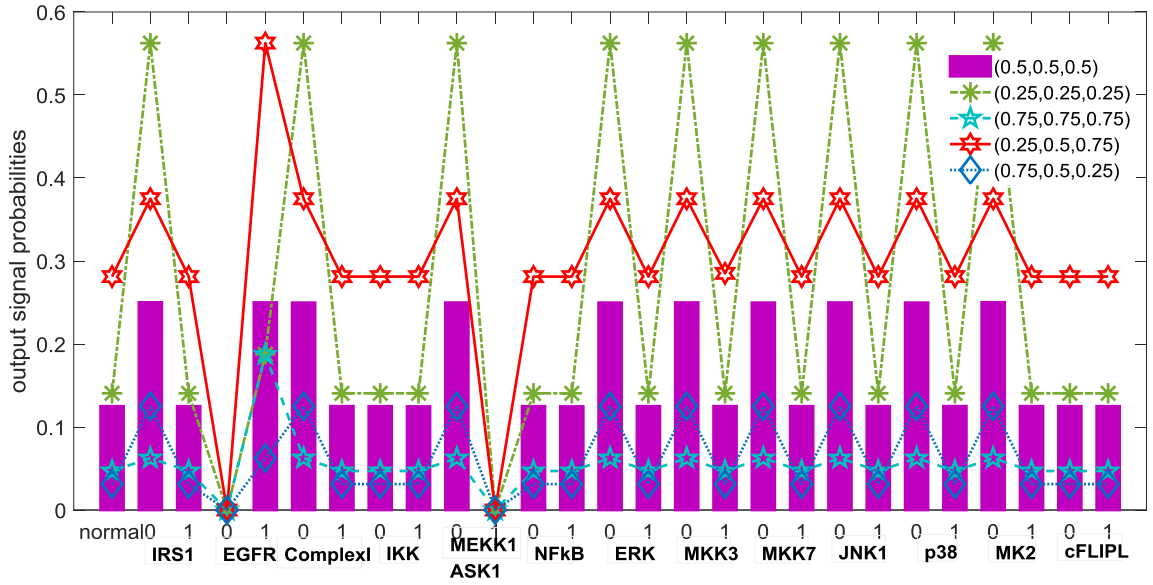


FIGURE 13. Signal probabilities of caspase3 for different input probability distributions given the occurrence of a one-bit perturbation.

1_{st} entry is perturbed, *EGFR* is always activated. Then, *AKT* is also activated because of the updating function, i.e., $AKT = \mathrm{OR}(Insulin, EGFR)$. This result means that caspase3 is always inactive. Similarly, in response to a one-bit perturbation at the 2_{nd} entry of $MEKK1ASK1$, the corresponding molecule will always be inactive, making caspase3 inactive.

If a one-bit perturbation is applied to a molecule in $\{IRS1, ComplexI, ERK, MKK3, MKK7, JNK1, p38, MK2\}$, the obtained probability when perturbing the 1_{st} entry is always larger than that of affecting the 2_{nd} one. The effects of perturbing molecules from $\{MKK3, MKK7, JNK1, p38, MK2\}$ is similar; this result occurs because these molecules originate directly from $MEKK1ASK1$. For instance, because of the relationship between $MKK7$ and $JNK1$, the effect of a one-bit perturbation on $MKK7$ is equivalent to that on $JNK1$. Similar analyses can be performed for $MKK3$, $p38$, and $MK2$.

Nevertheless, for different input probability distributions, the same perturbation strategy is capable of yielding the maximum output probability. For example, for (0.25, 0.5, 0.75), the maximum probability of 0.5626 is obtained in response to the application of a one-bit perturbation at the 2_{nd} entry of *EFGR*, and the maximum probability is also obtained by adopting same strategy for (0.75, 0.75, 0.75). However, this strategy is unable to yield the maximum value for the other input distributions here. Thus, we can conclude that the best strategy is determined by the input probability distributions, while the pathway topology also plays an important role in determining the most desirable strategy.

For the sake of simplicity, only a binary pathway is investigated here; nevertheless, stochastic analysis can certainly be performed if a multi-state scenario is considered. For instance, similar to the situation in [24], a ternary signaling pathway can be constructed according to the multistate relationships between molecules. For more information, please refer to [15]. That is, the proposed stochastic model for a one-bit perturbation can be generalized with modifications to the analysis of a multistate scenario to determine appropriate perturbation strategies.

VI. CONCLUSION

Logical representation and analyses of biological systems provide us with valuable insights. In this work, we focused on efficient analysis of function perturbation in pathways. We find that function perturbations of different molecules contribute to the development of effective therapies. To perform effective analysis, stochastic architectures are proposed to investigate the effects of function perturbations on long-term behaviors in biological systems. Then, stochastic analysis of a *Caspase*3 pathway subjected to function perturbations is performed while the practicability and effectiveness of the proposed stochastic approach are determined. Moreover, the effects of multibit perturbation are also considered. We see that the effects are largely determined by the relationships among molecules or the network topology. The simulation results show that in addition to directly affecting caspase3, *AKT* is a better candidate than other molecules for maximizing the signal probability of caspase3. Furthermore, the effect of applying a multibit perturbation is likely to be better than that of adopting a one-bit perturbation. In contrast, we also find that even for the same perturbation strategy, the corresponding effect is largely affected by the input probability distributions.

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers for their valuable suggestions.

REFERENCES

[1] S. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *J. Theor. Biol.*, vol. 22, no. 3, pp. 437–467, Mar. 1969.

[2] R. Layek, A. Datta, M. Bittner, and E. R. Dougherty, "Cancer therapy design based on pathway logic," *Bioinformatics*, vol. 27, no. 4, pp. 548–555, Feb. 2011.

[3] M. K. Morris, J. Saez-Rodriguez, P. K. Sorger, and D. A. Lauffenburger, "Logic-based models for the analysis of cell signaling networks," *Biochemistry*, vol. 49, no. 15, pp. 3216–3224, Apr. 2010.

[4] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Boca Raton, FL, USA: CRC Press, 2006.

[5] P. Wang, J. Lu, and X. Yu, "Colored noise induced bistable switch in the genetic toggle switch systems," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 12, no. 3, pp. 579–589, May 2015.

[6] P. Wang, Y. Zhang, J. Lü, and X. Yu, "Functional characteristics of additional positive feedback in genetic circuits," *Nonlinear Dyn.*, vol. 79, no. 1, pp. 397–408, Jan. 2015.

[7] P. Wang, J. Lü, and M. J. Ogorzalek, "Global relative parameter sensitivities of the feed-forward loops in genetic networks," *Neurocomputing*, vol. 78, no. 1, pp. 155–165, Feb. 2012.

[8] S. Watterson, S. Marshall, and P. Ghazal, "Logic models of pathway biology," *Drug Discovery Today*, vol. 13, nos. 9–10, pp. 447–456, May 2008.

[9] M. L. Wynn, N. Consul, S. D. Merajver, and S. Schnell, "Logic-based models in systems biology: A predictive and parameter-free network analysis method," *Integr. Biol.*, vol. 4, no. 11, pp. 1323–1337, 2012.

[10] 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, Feb. 2002.

[11] D. Murrugarra, "Modeling stochasticity and variability in gene regulatory networks," *EURASIP J. Bioinform. Syst. Biol.*, vol. 1, Jun. 2012, Art. no. 5.

[12] S. Kim, H. Li, E. R. Dougherty, N. Cao, Y. Chen, M. Bittner, and E. B. Suh, "Can Markov chain models mimic biological regulation?" *J. Biol. Syst.*, vol. 10, no. 04, pp. 337–357, Dec. 2002.

[13] A. Garg, L. Mendoza, I. Xenarios, and G. Demicheli, "Modeling of multiple valued gene regulatory networks," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2007, pp. 1398–1404.

[14] J. Liang and J. Han, "Stochastic Boolean networks: An efficient approach to modeling gene regulatory networks," *BMC Syst. Biol.*, vol. 6, no. 1, p. 113, 2012.

[15] P. Zhu and J. Han, "Stochastic multiple-valued gene networks," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 1, pp. 42–53, Feb. 2014.

[16] P. Zhu, J. Liang, and J. Han, "Gene perturbation and intervention in context-sensitive stochastic Boolean networks," *BMC Syst. Biolo.*, vol. 8, no. 1, p. 60, 2014.

[17] A. Abdi, M. B. Tahoori, and E. S. Emamian, "Fault diagnosis engineering of digital circuits can identify vulnerable molecules in complex cellular pathways," *Sci. Signaling*, vol. 1, no. 42, p. Ra10, Oct. 2008.

[18] W. Xiong, L. Xu, T. Huang, X. Yu, and Y. Liu, "Finite-iteration tracking of singular coupled systems based on learning control with packet losses," *IEEE Trans. Syst. Man Cybern, Syst.*, vol. 50, no. 1, pp. 245–255, Jan. 2020.

[19] P. Zhu, H. M. Aliabadi, H. Uludag, and J. Han, "Identification of potential drug targets in cancer signaling pathways using stochastic logical models," *Sci. Rep.*, vol. 6, Mar. 2016, Art. no. 23078.

[20] Y. Xiao and E. R. Dougherty, "The impact of function perturbations in Boolean networks," *Bioinformatics*, vol. 23, no. 10, pp. 1265–1273, May 2007.

[21] W. Xiong, L. Xu, D. W. C. Ho, J. Cao, and T. Huang, "Synchronous and asynchronous iterative learning strategies of T–S fuzzy systems with measurable and unmeasurable state information," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3042–3053, Oct. 2018.

[22] Z. Zhou, "Optimal batch distributed asynchronous multisensor fusion with feedback," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 1, pp. 46–56, Feb. 2019.

[23] B. Cai, Y. Zhao, H. Liu, and M. Xie, "A data-driven fault diagnosis methodology in three-phase inverters for PMSM drive systems," *IEEE Trans. Power Electron.*, vol. 32, no. 7, pp. 5590–5600, Jul. 2017.

[24] B. Cai, H. Liu, and M. Xie, "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks," *Mech. Syst. Signal Process.*, vol. 80, pp. 31–44, Dec. 2016.

[25] B. R. Gaines, "Stochastic computing systems," *Adv. Inf. Syst.*, vol. 2, pp. 37–172. Mar. 1969.

[26] J. Han, H. Chen, J. Liang, P. Zhu, Z. Yang, and F. Lombardi, "A stochastic computational approach for accurate and efficient reliability evaluation," *IEEE Trans. Comput.*, vol. 63, no. 6, pp. 1336–1350, Jun. 2014.

[27] K. A. Janes, S. Gaudet, J. G. Albeck, U. B. Nielsen, D. A. Lauffenburger, and P. K. Sorger, "The response of human epithelial cells to TNF involves an inducible autocrine cascade," *Cell*, vol. 124, pp. 1225–1239, Mar. 2006.

[28] P. Zhu, X. Wang, S. Li, Y. Guo, and Z. Wang, "Investigation of epidemic spreading process on multiplex networks by incorporating fatal properties," *Appl. Math. Comput.*, vol. 359, pp. 512–524, Oct. 2019.

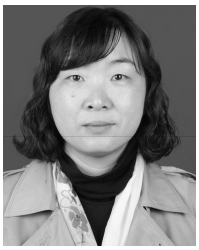
[29] J. Luo, B. D. Manning, and L. C. Cantley, "Targeting the PI3K-Akt pathway in humancancer: Rationale and promise," *Cancer Cell*, vol. 4, pp. 257–262, Oct. 2003.

[30] J. Han, H. Chen, E. Boykin, and J. Fortes, "Reliability evaluation of logic circuits using probabilistic gate models," *Microelectron. Rel.*, vol. 51, no. 2, pp. 468–476, Feb. 2011.

[31] H. Chen and J. Han, "Stochastic computational models for accurate reliability evaluation of logic circuits," in *Proc. 20th Symp. Great lakes Symp. VLSI*, Providence, RI, USA, 2010, pp. 61–66.

JUN HU received the B.S. degree in mathematics and applied mathematics from Southwest Petroleum University, in 2017. He is currently pursuing the master's degree in project management with the Central University of Financial and Economics, Beijing, China. His research interests involve complex networks, cluster, social networks, and control theory.

HUI-JIA LI received the Ph.D. degree in operational research and control theory from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, in 2013. He is currently a Distinguished Research Fellow with the School of Science, Beijing University of Posts and Telecommunications. His research interests include data mining, pattern recognition, complex networks, and control theory.



JIE GAO received the B.S. degree in information and computing science from Xiangtan University, Xiangtan, China, in 2004, the M.S. degree in computational mathematics from Wuhan University, Wuhan, China, in 2006, and the Ph.D. degree in applied mathematics from the University of Electronic Science and Technology of China, Chengdu, China, in 2018. Her research interests include synchronization and control of complex networks, and consensus of multiagent systems.

ZEBO ZHOU received the B.Sc. and M.Sc. degrees from the School of Geodesy and Geomatics, Wuhan University, Wuhan, China, in 2004 and 2006, respectively, and the Ph.D. degree from the College of Surveying and Geo-informatics, Tongji University, Shanghai, China, in 2009, all in geodesy and surveying engineering. He was a Visiting Fellow with the Surveying and Geospatial Engineering Group, School of Civil and Environmental Engineering, University of New South Wales, Kensington, Australia, from 2009 to 2015. He is currently an Associate Professor with the School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu, China. His research interests include global navigation satellite systems (GNSS) navigation and positioning, GNSS/INS integrated navigation, and multisensor fusion.

CHENGYI XIA was born in Hefei, Anhui, China, in 1976. He received the B.S. degree in mechanical engineering from the Hefei University of Technology, Hefei, China, in 1998, the M.S. degree in nuclear energy science and engineering from the Institute of Plasma Physics, Chinese Academy of Sciences, Hefei, in 2001, and the Ph.D. degree in control theory and control engineering from Nankai University, Tianjin, China, in 2008. From 2001 to 2013, he was a Lecturer, an Assistant Professor, and an Associate Professor with the Tianjin University of Technology, Tianjin, China, where he has been a Professor, since 2013. He has coauthored more than 70 peer-reviewed journal or conference papers. His research interests include complex system modeling and analysis, risk analysis and management, complex networks, epidemic spreading, and evolutionary game theory.