Molecular Communication and Signaling in Human Cells

I. Habibi[1] and A. Abdi[1,2]
[1]Electrical and Computer Engineering Department
and [2]Biological Sciences Department
New Jersey Institute of Technology, Newark, NJ, USA
ih26@njit.edu, ali.abdi@njit.edu

E. S. Emamian
Advanced Technologies for Novel Therapeutics
Enterprise Development Center,
Newark, NJ, USA
emame@atnt-usa.com

***Abstract*— Signaling networks in human cells convey signals from the cell membrane to specific target molecules via biochemical interactions, to control a variety of cellular functions. We have modeled signaling networks as communication channels where molecules communicate with each other to transfer signals. We have defined and computed the fundamental parameters of transmission error probability and signaling capacity in signaling networks. This systematic approach can be used to understand how cell signaling errors and malfunctioning molecules may contribute to the development of complex human disorders with unknown molecular bases.**

Keywords—systems biology; signal transduction; molecular networks; human disease; signaling error; communication channels; transmission error probability; capacity.

I. INTRODUCTION

Over the past few decades, molecular biologists have discovered numerous human proteins and their molecular interactions. The human proteome map [1] includes thousands of proteins, many of them interacting with each other via different post-translational activatory and inhibitory modifications such as phosphorylation, methylation and ubiquitination. It is now well established that the dysfunction of one or more molecules within the molecular networks can contribute to the development of different diseases. The major challenge is to figure out which molecules in such large interconnected networks are more important for the development of a certain disease, and how much each molecule contributes to the pathology. For example, schizophrenia is a complex trait disorder and in recent years several signaling proteins, such as DISC1 (disrupted in schizophrenia 1) and NRG1 (neuregulin 1) and AKT1 have been consistently reported to be associated with increased risk of schizophrenia in several different populations around the world [2]. However we do not know how much each one of these molecules contribute to the disease development and which molecules are the most critical ones.

In this paper, a complex disorder is viewed as the outcome of miscommunication in signaling networks with dysfunctional molecules. So, we propose to use communication engineering concepts, models and methods to model complex trait disorders, by quantifying the function of different molecules that are known to be associated with the disorder, to find out which molecules are the most critical ones for the development of pathology. Our approach can also identify previously unknown critical molecules and their contributions to the disease development. In this model, a molecule is considered critically important if its dysfunction disrupts the expected function of the intracellular communication network.

In order to proceed with the molecular communication approach, we need to choose a modeling framework for molecular networks. In large networks with so many molecules, we often do not have enough information on mechanistic details and kinetic parameters, therefore, differential equation-based models are not feasible for the analyses of such large networks. Instead, logical network models such as binary (Boolean) are appropriate as they can be constructed with minimal amount of information about the molecules and their interactions, and yet they can provide useful biologically-relevant predictions that can be verified experimentally [3]. The efficacy of such logical models are described with details in some recent review and research articles [4]-[13].

In this paper we use a ternary logic framework to study communication among molecules in signaling networks, which complements our prior study using a binary framework [14].

The rest of this paper is organized as follows: In Section II we summarize the approach that we developed in [14] to model signaling networks as communication channels, where molecules were considered to follow a binary activity model. Section III is devoted to a ternary activity model for molecules, to study how a more complex activity model may affect signaling network communication parameters such as transmission error probability and signaling capacity. Transmission error probability and signaling capacity are defined and computed in Section IV, and are used to find critical molecule in signaling networks. Some concluding remarks are provided in Section V.

II. SIGNALING NETWORKS AS COMMUNICATION CHANNELS

Here we summarize the model that we developed in our prior study [14], to lay the foundation for a more complex model we have introduced in the present study.

To model a signaling network as a communication channel, the first step is to specify input(s) and output(s), as well as intermediate molecules in the network. After that we need to specify the types of interactions among the molecules within the network, using molecular biology data. In the next step we should determine binary (active/inactive) logic equations that show how each molecule is regulated by its upstream and the other molecules in the network. These equations allow to calculate transition probability channel matrix **M** for the network, where each element of the matrix specifies the probability of the outputs to be 0 or 1, inactive or active, conditioned on the 0 or 1 states of the inputs. As an example of this modeling approach, consider caspase3 signaling network in Fig. 1, where there are three inputs molecules, i.e., insulin, EGF and TNF, 17 intermediate molecules, and one output

molecule, i.e., caspase3. Communication channel model for the normal caspase3 signaling network is shown in Fig. 2, where all the molecules in the network are functioning normally.

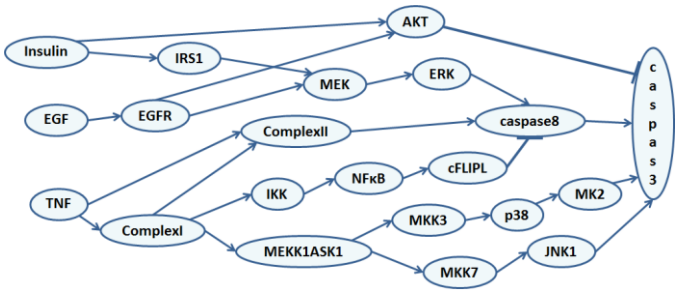

Fig. 1. The Caspase3 signaling network.

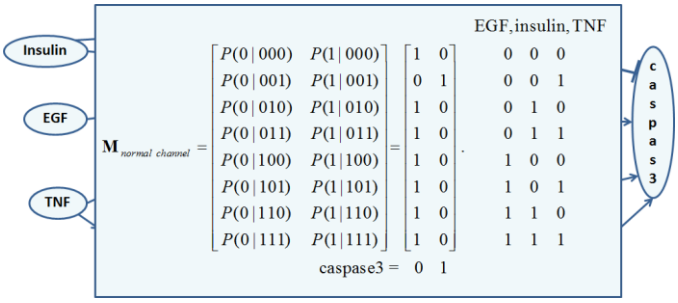

Fig. 2. Transition probability channel matrix of a binary communication channel model for the normal caspase3 signaling network.

To model a pathological (abnormal) signaling network, two assumptions are made [14]. First, a dysfunctional molecule remains active, 1, or inactive, 0, with an equal probability of 1/2, regardless of the activity states of the molecules that regulate the dysfunctional molecule. Second, the probability of each molecule to be dysfunctional is β, except for a dominant molecule whose dysfunctionality probability is $k\beta$, $k \geq 1$, where k is the dominance factor. This means the dysfunctionality probability of the dominant molecule is k times greater than other molecules. For $k \to \infty$, this model simplifies to the pathological network model introduced in our previous publication [3].

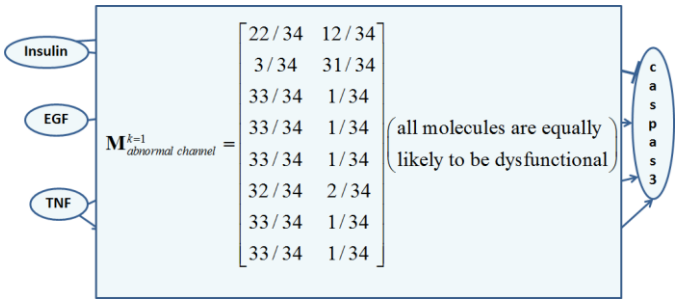

Fig. 3. Transition probability channel matrix of a binary communication channel model for a pathological caspase3 signaling network, where all molecules are equally likely to be dysfunctional.

Based on these assumptions and using the network binary logic equations, transition probability channel matrices for the pathological signaling network model can be derived as functions of k, where matrix entries depend on the molecule which is considered to be dominant [14]. Transition probability channel matrix **M** for a pathological caspase3 signaling network where all molecules are equally likely to be dysfunctional, $k = 1$, is shown in Fig. 3.

III. Communication Channel Representataion of a Ternary Signaling Network Model

Here we expand the binary model of the previous section to a ternary model. In a ternary model each molecule can be either active, 1, inactive, 0, or partially active, 1/2. One possible definition for ternary logic functions AND, OR and NOT are provided in Table 1 of [15]. Ternary logic equations for the caspase3 network are derived and presented in TABLE I, where "max" and "min" stand for maximum and minimum, respectively. Using TABLE I we derive the transition probability channel matrix M for the normal caspase3 network, which is shown in Fig. 4.

TABLE I. Ternary Logic Equations for the Caspase3 Signaling Network

Molecule	Ternary logic equation
AKT	AKT= max(EGFR,insulin)
caspase8	caspase8= min((1-cFLIPL) ,max(ComplexII,ERK))
cFLIPL	cFLIPL=NFκB
ComplexI	ComplexI=TNF
ComplexII	ComplexII= max(TNF,ComplexI)
EGFR	EGFR=EGF
ERK	ERK=MEK
IKK	IKK=ComplexI
IRS1	IRS1=Insulin
JNK1	JNK1=MKK7
MEK	MEK= max(EGFR,IRS1)
MEKK1ASK1	MEKK1ASK1=ComplexI
MK2	MK2=p38
MKK3	MKK3=MEKK1ASK1
MKK7	MKK7=MEKK1ASK1
NFκB	NFκB=IKK
p38	p38=MKK3
caspase3	caspase3= min((1-AKT),max(caspase8,JNK1,MK2))

To model a pathological signaling network, we assume that a dysfunctional molecule remains active, 1, partially active, 1/2, or inactive, 0, with equal probability of 1/3, regardless of its input signals. We also assume that the probability of each molecule to be dysfunctional is β, except for a dominant molecule whose dysfunctionality probability is $k\beta$, $k \geq 1$, where k is the dominance factor. This means the dysfunctionality probability of the dominant molecule is k times greater than other molecules. A pathological network model based on different assumptions is introduced in [16].

Using the total probability theorem, each element of transition probability channels matrix **M**, P(output(s) |

input(s)), for a pathological network model can be calculated as follows:

$$\begin{aligned} &P(\text{output(s)}\,|\,\text{input(s)}) = \\ &\sum_{i=1}^{n}\{P(\text{output(s)}\,|\,\text{input(s)},\mathcal{D}_i^0)P(\mathcal{D}_i^0) \\ &+\ P(\text{output(s)}\,|\,\text{input(s)},\mathcal{D}_i^{1/2})P(\mathcal{D}_i^{1/2}) \\ &+\ P(\text{output(s)}\,|\,\text{input(s)},\mathcal{D}_i^1)P(\mathcal{D}_i^1)\}. \end{aligned} \tag{1}$$

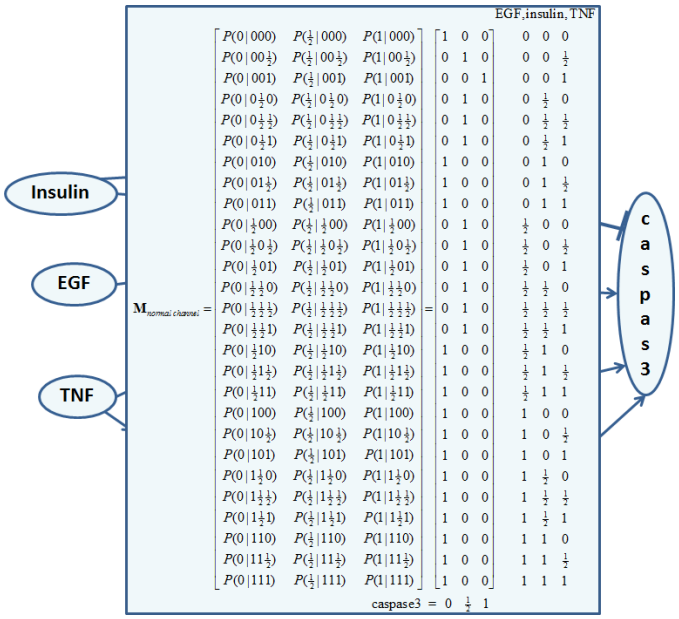

Fig. 4. Transition probability channel matrix of a ternary communication channel model for the normal caspase3 signaling network.

Here $\mathcal{D}_i^0$ represents the event that X_i is dysfunctional such that its activity level is locked at $X_i = 0$, where X_i is the i-th intermediate molecule in the network and n is the total number of intermediate molecules. Similarly, $\mathcal{D}_i^{1/2}$ and $\mathcal{D}_i^1$ denote the events that X_i is dysfunctional and locked at $X_i = 1/2$ and $X_i = 1$, respectively. Moreover we have:

$$\begin{aligned} P(\mathcal{D}_i^0) &= P(X_i = 0\,|\,X_i \text{ is dysfunc})P(X_i \text{ is dysfunc}) \\ &= (1/3)\,p_i, \end{aligned} \tag{2}$$

$$\begin{aligned} P(\mathcal{D}_i^{1/2}) &= P(X_i = 1/2\,|\,X_i \text{ is dysfunc})P(X_i \text{ is dysfunc}) \\ &= (1/3)\,p_i, \end{aligned} \tag{3}$$

$$\begin{aligned} P(\mathcal{D}_i^1) &= P(X_i = 1\,|\,X_i \text{ is dysfunc})P(X_i \text{ is dysfunc}) \\ &= (1/3)\,p_i, \end{aligned} \tag{4}$$

where $p_i = P(X_i \text{ is dysfunc})$.

To study the pathological behavior of the caspase3 network in Fig. 1, elements of the associated transition probability channels matrix **M**, $P(\text{caspase3}\,|\,\text{EGF},\text{insulin},\text{TNF})$, can be calculated by combining Equations (1)-(4):

$$\begin{aligned} &P(\text{caspase3}\,|\,\text{EGF},\text{insulin},\text{TNF}) \\ &= (1/3)\sum_{i=1}^{17} p_i\{P(\text{caspase3}\,|\,\text{EGF},\text{insulin},\text{TNF},\mathcal{D}_i^0) \\ &+\ P(\text{caspase3}\,|\,\text{EGF},\text{insulin},\text{TNF},\mathcal{D}_i^{1/2}) \\ &+\ P(\text{caspase3}\,|\,\text{EGF},\text{insulin},\text{TNF},\mathcal{D}_i^1)\}. \end{aligned} \tag{5}$$

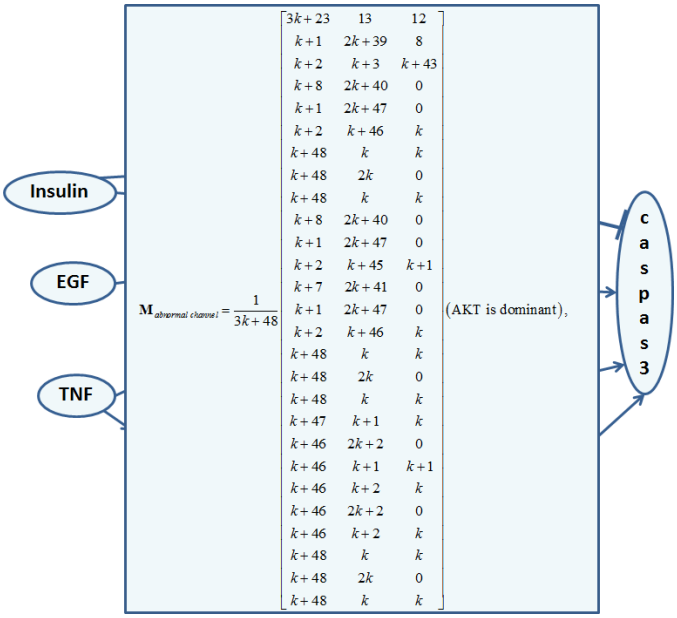

Fig. 5. Transition probability channel matrix of a ternary communication channel model for the pathological caspase3 signaling network, when AKT is the dominant dysfunctional molecule.

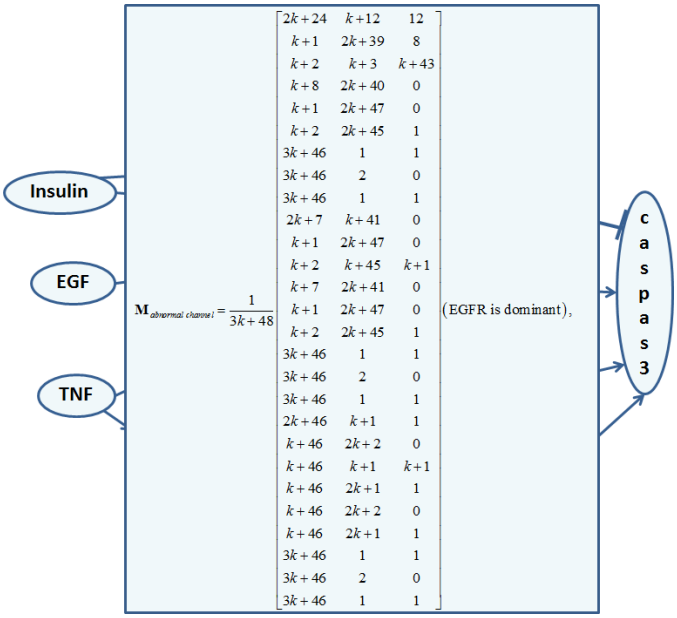

Fig. 6. Transition probability channel matrix of a ternary communication channel model for the pathological caspase3 signaling network, when EGFR is the dominant dysfunctional molecule.

Using Equation (5) we have calculated $P(\text{caspase3}\,|\,\text{EGF},\text{insulin},\text{TNF})$ depending on which molecule is dominant, for all 27 possible inputs in the ternary model. The **M** matrix of the pathological caspase3 signaling network when

AKT is the dominant dysfunctional molecule is given in Fig. 5, whereas Fig. 6 shows the **M** matrix when EGFR is the dominant dysfunctional molecule in the pathological network. Distinct **M** matrices for several other dominant dysfunctional molecules are not provided due to space limitation.

IV. Transmission Error Probability and Signaling Capacity of Pathological Networks

Here we use transmission error probability and signaling capacity as two communication-related metrics to find critical molecules in signaling networks.

The transmission error probability can be calculated using the total probability theorem as follow:

$$P_e = \sum P(\text{incorrect output(s)}|\text{input(s)})P(\text{input(s)}) . \quad (6)$$

For the caspase3 network we have:

$$P_e = \sum (P(\text{incorrect caspase3}|\text{EGF, insulin, TNF}) \times P(\text{EGF, insulin, TNF})). \quad (7)$$

Using the transition probability channel matrix in Fig. 4 , and since in our model all the 27 input combinations are assumed to be equally probable, Equation (7) can be written as:

$$\begin{aligned} P_e = \{&P(\tfrac{1}{2}|000)+P(1|000)+P(0|00\tfrac{1}{2})+P(1|00\tfrac{1}{2}) \\ &+P(0|001)+P(\tfrac{1}{2}|001)+P(0|0\tfrac{1}{2}0)+P(1|0\tfrac{1}{2}0) \\ &+P(0|0\tfrac{1}{2}\tfrac{1}{2})+P(1|0\tfrac{1}{2}\tfrac{1}{2})+P(0|0\tfrac{1}{2}1)+P(1|0\tfrac{1}{2}1) \\ &+P(\tfrac{1}{2}|010)+P(1|010)+P(\tfrac{1}{2}|01\tfrac{1}{2})+P(1|01\tfrac{1}{2}) \\ &+P(\tfrac{1}{2}|011)+P(1|011)+P(0|\tfrac{1}{2}00)+P(1|\tfrac{1}{2}00) \\ &+P(0|\tfrac{1}{2}0\tfrac{1}{2})+P(1|\tfrac{1}{2}0\tfrac{1}{2})+P(0|\tfrac{1}{2}01)+P(1|\tfrac{1}{2}01) \\ &+P(0|\tfrac{1}{2}\tfrac{1}{2}0)+P(1|\tfrac{1}{2}\tfrac{1}{2}0)+P(0|\tfrac{1}{2}\tfrac{1}{2}\tfrac{1}{2})+P(1|\tfrac{1}{2}\tfrac{1}{2}\tfrac{1}{2}) \\ &+P(0|\tfrac{1}{2}\tfrac{1}{2}1)+P(1|\tfrac{1}{2}\tfrac{1}{2}1)+P(\tfrac{1}{2}|\tfrac{1}{2}10)+P(1|\tfrac{1}{2}10) \\ &+P(\tfrac{1}{2}|\tfrac{1}{2}1\tfrac{1}{2})+P(1|\tfrac{1}{2}1\tfrac{1}{2})+P(\tfrac{1}{2}|\tfrac{1}{2}11)+P(1|\tfrac{1}{2}11) \\ &+P(\tfrac{1}{2}|100)+P(1|100)+P(\tfrac{1}{2}|10\tfrac{1}{2})+P(1|10\tfrac{1}{2}) \\ &+P(\tfrac{1}{2}|101)+P(1|101)+P(\tfrac{1}{2}|1\tfrac{1}{2}0)+P(1|1\tfrac{1}{2}0) \\ &+P(\tfrac{1}{2}|1\tfrac{1}{2}\tfrac{1}{2})+P(1|1\tfrac{1}{2}\tfrac{1}{2})+P(\tfrac{1}{2}|1\tfrac{1}{2}1)+P(1|1\tfrac{1}{2}1) \\ &+P(\tfrac{1}{2}|110)+P(1|110)+P(\tfrac{1}{2}|11\tfrac{1}{2})+P(1|11\tfrac{1}{2}) \\ &+P(\tfrac{1}{2}|111)+P(1|111)\}/27. \end{aligned} \quad (8)$$

Now we can use Equation (8) to find transmission error probability P_e depending on which molecule is dominant in the pathological network, by replacing the conditional probabilities in (8) with the corresponding elements of the **M** matrix. For example, using some elements of the **M** matrix in Fig. 5, P_e in (8) can be simplified to Equation (9), when AKT is the dominant dysfunctional molecule in the pathological network. Expressions for P_e versus k are given in Equations (10)-(16), when each of the rest of the molecules is dominant.

$$P_e = \frac{5}{9} - \frac{637}{81(k+16)} \text{ (AKT is dominant)}, \quad (9)$$

$$P_e = \frac{26}{81} - \frac{314}{81(k+16)} \text{ (EGFR is dominant)}, \quad (10)$$

$$P_e = \frac{8}{81} - \frac{8}{81(k+16)} \text{ (MEKK1ASK1 is dominant)}, \quad (11)$$

$$P_e = \frac{2}{27} + \frac{26}{81(k+16)} \text{ (caspase8 is dominant)}, \quad (12)$$

$$P_e = \frac{5}{81} + \frac{43}{81(k+16)} \text{ (ERK or MEK is dominant)}, \quad (13)$$

$$P_e = \frac{4}{81} + \frac{20}{27(k+16)} \text{ (ComplexI is dominant)}, \quad (14)$$

$$P_e = \frac{1}{27} + \frac{77}{81(k+16)} \text{ ($cFLIP_L$ or IKK or NFκB or IRS1 or JNK1 or MK2 or MKK3 or MKK7 or p38 is dominant)}, \quad (15)$$

$$P_e = \frac{2}{81} + \frac{94}{81(k+16)} \text{ (ComplexII is dominant)}. \quad (16)$$

The above transmission error probabilities are graphed in Fig. 7 versus k, and will be discussed at the end of this section.

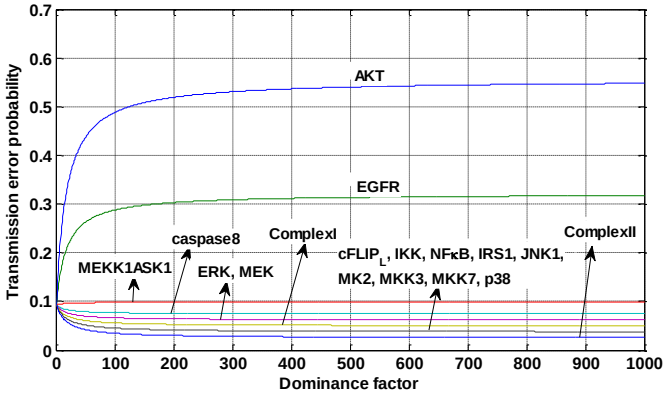

Fig. 7. Transmission error probability versus the dominance factor k in a ternary model for the pathological caspase3 network.

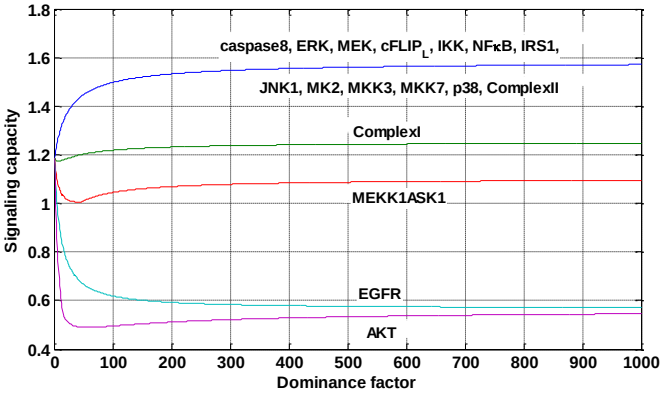

Fig. 8. Signaling capacity versus the dominance factor k in a ternary model for the pathological caspase3 network.

In addition to the transmission error probability, another important communication-related metric that we calculate for pathological signaling networks modeled as communication channels is the signaling capacities C. Depending on which

molecule in the channel is the dominant dysfunctional molecule, we can calculate the associated signaling capacity using the transition probability channel matrix **M**.

We have used the Arimoto algorithm [17] to numerically calculate the signaling capacities of the pathological caspase3 signaling network, for various dominant dysfunctional molecules in the network. The results are graphed in Fig. 8 versus the dominance factor *k*.

Comparison of Fig. 7 and Fig. 8 reveal that AKT is a very important molecule in the caspase3 signaling network. According to Fig. 7, when AKT is the dominant dysfunctional molecule, transmission error probability rapidly increases with the dominance factor, and moreover, AKT exhibits the highest transmission error probability. Additionally, in Fig. 8 we observe when AKT is the dominant dysfunctional molecule, signaling capacity rapidly decreases with the dominance factor, and remains at the lowest level, compared to other molecules. These observations indicate the critical role of AKT in the network. On the other hand, few molecules play a less critical role in the network function. For example when MEKK1ASK1 is dominant, transmission error probability is low and signaling capacity is relatively high, and both metrics are less sensitive to the changes in *k*. Interestingly, most of the molecules are not critical, i.e., when they are dominantly dysfunctional, transmission error probability is very low and signaling capacity is very high. Examples of such molecules include $cFLIP_L$, IKK, NFκB, IRS1, JNK1, MK2, MKK3, MKK7, p38 and ComplexII.

With regard to the binary model [14] summarized in Section II and the ternary model developed in Section III, we observe the increased resolution power of the ternary model in specifying critical molecules, at the cost of being a more complex model than the binary model. More specifically, in the binary model we have 4 different groups of molecules exhibiting 4 different transmission error probability curves (see Fig. 2a in [14]). However, according to Fig. 7 for the ternary model in the present paper, there are 8 different groups of molecules with 8 different transmission error probability curves. This indicates more precision in identifying the roles of different molecules in the network, at the cost of more complex model and method. For example EGFR and MEKK1ASK1 have the same transmission error probability in the binary model (see Fig 2a in [14]), whereas they exhibit different transmission error probabilities in the ternary model (see Fig. 7 in the present paper). We have made the same overall observation regarding the signaling capacity of binary and ternary network models (see Fig. 2b in [14] and Fig. 8 in the present paper).

V. Conclution

In this paper we have shown how signaling networks in human cells can be modeled as communication channels in which inputs and outputs are certain molecules such as ligands and transcription factors, respectively. We have shown how using binary and ternary activity models for molecules, one can study pathological/abnormal network behaviors, as well as signal transmission errors, when there are dysfunctional molecules in the network. We have observed that a ternary model is more complex to build and analyze. The ternary model, however, provides a higher resolution power especially for separation of molecules with similar error probabilities.

The proposed approach has the potential to pinpoint those critical molecules that have causative effects in some complex human disorders. This is an important step for understanding the molecular basis of these disorders and finding therapeutics that target the activity of such critical molecules [18].

References

[1] H. Pearson, "Biologists initiate plan to map human proteome," *Nature*, vol. 452, no. 7190, pp. 920-921, 2008.

[2] E.S. Emamian, "AKT/GSK3 signaling pathway and schizophrenia," *Frontiers in Molecular Neuroscience*, vol. 5, 2012.

[3] A. Abdi, M. B. Tahoori and E. S. Emamian, "Fault diagnosis engineering of digital circuits can identify vulnerable molecules in complex cellular pathways," *Science Signaling*, vol. 1, no. 42, pp. 48-61, 2008.

[4] R.S. Wang, A. Saadatpour and R. Albert, "Boolean modeling in systems biology: an overview of methodology and applications," *Phys. Biol*., vol. 9, no. 5, 055001, 2012.

[5] C. Chaouiya and E. Remy, "Logical Modelling of Regulatory Networks, Methods and Applications," *Bulletin of Mathematical Biology*, vol. 75, no. 6, pp. 891-895, 2013.

[6] C. Chaouiya, A. Naldi , D. Thieffry, "Logical modelling of gene regulatory networks with GINsim," *Methods Mol Biol*, vol. 804, pp. 463-479, 2012.

[7] M.I. Davidich and S. Bornholdt, "Boolean Network Model Predicts Knockout Mutant Phenotypes of Fission Yeast," *PLoS ONE*, vol. 8, no. 9, e71786, 2013.

[8] R. Schlatter, et al., "Integration of Boolean models exemplified on hepatocyte signal transduction," *Briefings in bioinformatics*, vol. 13, pp. 365-376, 2012.

[9] G. Karlebach, "Inferring Boolean network states from partial information," *EURASIP J Bioinform Syst Biol*, vol. 2013, no. 1, pp. 1-13, 2013.

[10] K. Kobayashi and K. Hiraishi, "Optimal Control of Gene Regulatory Networks with Effectiveness of Multiple Drugs: A Boolean Network Approach," *BioMed Research International*, vol. 2013, 2013.

[11] 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, p. 1323-1337, 2012.

[12] 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, 2010.

[13] R. Samaga and S. Klamt, "Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks," *Cell Communication and Signaling*, vol. 11, no. 1, p. 43, 2013.

[14] I. Habibi, E.S. Emamian and A. Abdi, "Quantitative analysis of intracellular communication and signaling errors in signaling networks," *BMC Systems Biology*, vol. 8, no. 1, 2014.

[15] M. Mukaidono, "Regular ternary logic functions—Ternary logic functions suitable for treating ambiguity," *IEEE Trans. Comput*., vol. C-35, no. 2, pp. 179–183, Feb. 1986.

[16] I. Habibi, E.S. Emamian and A. Abdi, "Advanced fault diagnosis methods in molecular networks," *PLoS One*, vol. 9, e108830, 2014.

[17] S. Arimoto, "An algorithm for calculating the capacity of an arbitrary discrete memoryless channel," *IEEE Trans. Inform. Theory*, vol. 18, pp. 14-20, Jan. 1972.

[18] A. Abdi and E.S. Emamian, "Fault diagnosis engineering in molecular signaling networks: An overview and applications in target discovery," *Chemistry and Biodiversity*, vol. 7, no. 5, pp. 1111-1123, 2010 (Summit on Systems Biology special issue).