

# A Study on the Design of a Biologizing Control System

\*Byung-Jae Choi and \*\*Paul P. Wang

\*: School of Electronic Engineering, Daegu University  
15 Naeri Jinryang Gyungsan Gyungbuk, 712-714 Korea

\*\* : Department of Electrical & Computer Engineering, Duke University  
Durham NC, 27708 USA

## Abstract

According to the progress of an information-oriented society, more human friendly systems are required. The systems can be implemented by a kind of intelligent algorithms. In this paper we propose the possibility of the implementation of an intelligent algorithm from gene behavior of human beings, which has some properties such as self organization and self regulation. The regulation of gene behavior is widely analyzed by Boolean network. Also the SORE (Self Organizable and Regulating Engine) is one of those algorithms. This paper does not report detailed research results; rather, it studies the feasibility of gene behavior in biocontrol systems based upon computer simulations.

**Key Words** : Boolean network, gene behavior, intelligent algorithms, SORE, cellular automata, self organization, self regulation

## 1. Introduction

Many complex processes are difficult to control using existing techniques because they are highly interconnected nonlinear systems that operate over a wide range of conditions. Some novel control techniques are required in order to cope with increasing demands on convenience, comfort, and high performance. Some pioneering researchers such as James Albus have even undertaken the task of constructing a road map for the engineering of the mind [1-2]. A biocontrol system is recently considered as one of most important issues in control fields. This can also play an important role in the design of a kind of human friendly systems.

The development of control theory has a long history due to fairly intensive research efforts for at least half a century. Some issues were thoroughly investigated and have reached a very mature status, while others were left nearly untouched. In an article entitled "*Biologizing control theory: How to make a control system come alive*" [3], John L. Casti coined the word "Biologizing" to reflect the recognition of so called "reliability and survivability" as topics of primary concern for engineers. A correct value judgment for a control system was thus finally determined. A blue-ribbon panel of 52 experts has nearly created a large set of research problems that lie ahead for control research engineers [4].

The concept of "feedback" in control theory is so far only known as a special trivial case of "homeostasis" - the function of an organism to regulate and to keep a constant "internal environment". These topics on biocontrol systems nevertheless bring much excitement and vision [5-6] to the research community.

There is no doubt that the structures of biocontrol systems are very complicated hierarchical structures. But what kinds of system components are necessary to make biocontrol systems work? Without any specific examples, everything remains vague, unclear and uncertain. This paper provides such an example.

Some models for a gene regulatory network were reported: The Boolean network (BN) model has received a considerable amount of attention, which was originally introduced by Stuart A. Kauffman [7]. This is also called by NK-network. In a genetic network, the total number of genes is represented by  $N$  and  $K$  is the largest number of genes which regulates any one of the  $N$  genes in the genetic network.

SORE (Self Organizable and Regulating Engine) was first discovered as a classifier [8]. More interesting and robust properties subsequently emerged [9-12]. SORE is by far the most general mathematical structure of a family of automata theories listed as follows:  $\text{SORE} \supset \text{Boolean Network} \supset \text{Cellular Automata} \supset \text{Linear Automata}$ , with the exception of the condition  $K < 5$  where the cellular automata could be more general than the Boolean network. Based upon the theory of Stuart A. Kauffman, a biological genetic network usually has a much smaller  $K$  to more realistically model a biological setting. However, it is the condition of  $K = N$  in which the network assumes its full strength for the best

접수일자 : 2004년 3월 31일

완료일자 : 2004년 5월 28일

This work was supported by the Post-doctoral Fellowship Program of Korea Science & Engineering Foundation (KOSEF).

possible performance in a "biologizing" control system. This  $K=N$  condition is what distinguishes SORE from the standard Boolean network [9].

We discuss Boolean network and SORE in modern control languages and a concept of colonization in Section 2 and 3, respectively. In Section 4 we describe some properties for a biologizing control system. Discussions and conclusions are presented in Section 5.

## II. Gene Regulatory Networks in Modern Control Language

The roots of the Boolean network lie in the automata theory which was a subject of Turing and von Neumann work. A Boolean network is a system of  $N$  interconnected binary elements (nodes). It is widely used to model gene regulatory networks and was originally introduced by Stuart A. Kauffman in 1969. It is frequently called by NK-network, where  $N$  and  $K$  are the number of genes and the maximum connectivity of a Boolean network, respectively. The number of inputs  $K$  may vary or be the same for all nodes in the network.

A Boolean network  $G(V, F)$  is defined by a set of nodes (genes)  $V = \{x_1, x_2, \dots, x_n\}$  and a set of Boolean functions  $F = \{f_1, f_2, \dots, f_n\}$ .  $x_i \in \{0, 1\}$  is a Boolean variable, where  $i = 1, 2, \dots, n$ . We write simply  $x_i = 1$  to denote that the  $i$ -th node (gene) is expressed ( $x_i = 0$  denotes that the  $i$ -th gene is not expressed). Each Boolean function  $f_i(x_{i1}, \dots, x_{iK})$  with  $K$  specific input nodes is assigned to node  $x_i$  and is used to update its value. The values of all the nodes in  $V$  are then updated synchronously. In general, there are  $2^{2^N}$  possible Boolean activation functions for a node with  $K$  inputs.

Regulation of nodes is defined by the set  $F$  of Boolean functions. In detail, given the value of the nodes  $V$  at time  $t$ , the Boolean functions are used to update the value of the nodes at time  $t+1$ . In the NK network model for gene regulation, a symbolic logic function is used to describe the self-organization and self-regulation of genes' expressions. The Boolean functions are made up of the logical connectors AND, OR, and NOT, forming a complete logic set. The HIGH's and LOW's of gene expressions for the next state in the Boolean network are controlled by these Boolean functions. In other words, these logic functions become the rules that govern the HIGH's or LOW's of the genes' expressions in the next instant of time. The synchronous update process is then repeated making the network dynamic. In order to capture the dynamic nature of the network, it may be useful to consider the wiring diagram which gives an explicit way of implementing the updating procedure.

During the operation of the network, a sequence of states will converge to a limit cycle or an attractor (a

limit cycle with a length of one). Each specific initial condition will converge to a specific attractor or limit cycle. For a genetic network of  $N$  genes, there are exactly  $2^N$  possible initial states.

For an example, consider the lambda bacteriophage. It is a virus that invades E. coli bacteria and 2 distinct modes of operation: 1) it can become integrated into the host cell DNA, and be replicated automatically each time the bacterium divides. 2) it can multiply in the cytoplasm of bacterium, eventually killing its host. There are 2 proteins: lambda repressor and cro protein. The lambda repressor blocks the expression of the gene for the cro protein, and vice versa. So, it can be modelled by a one-input Boolean network with 2 nodes (lambda repressor and cro protein) as shown in Fig. 1. It shows that both of the nodes are INVERSE properties because cro protein blocks the expression of the lambda repressor and vice versa. Two of four states are 10 and 01, and they correspond to the exclusive expression of each protein. There are another cycle of period 2 (11  $\rightarrow$  00  $\rightarrow$  11) which is not a behavior observed in the lambda bacteriophage. This shows that the simple model is not a complete description of the lambda bacteriophage system.

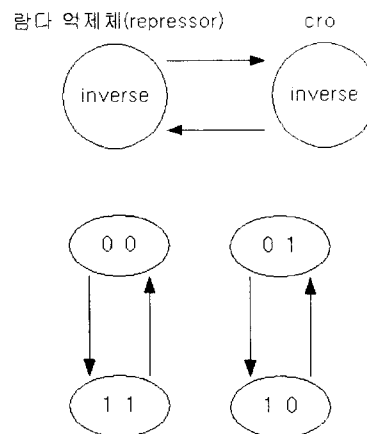


Fig. 1 The Boolean network model for the lambda bacteriophage

SORE is the most general Boolean network of which Stuart A. Kauffman's NK-network is a special case [9]. The NK-network imposes the severe restriction of  $K < N$ , which closely approximate the biological reality of  $K < N$  or  $K \ll N$ . Since SORE uses the most general NN-network, it can be used in applications other than genetic networks alone. SORE is the most general mathematical structure used to represent self-organizing and self-replicating systems.

Theoretically speaking, the whole family of the Boolean network and the cellular automata can best be explained by using the modern control theoretical language of state space. The  $N$  genes' expression levels naturally constitute  $N \times 1$  state vectors in a vector space. This discrete state space will consist of precisely  $2^N$

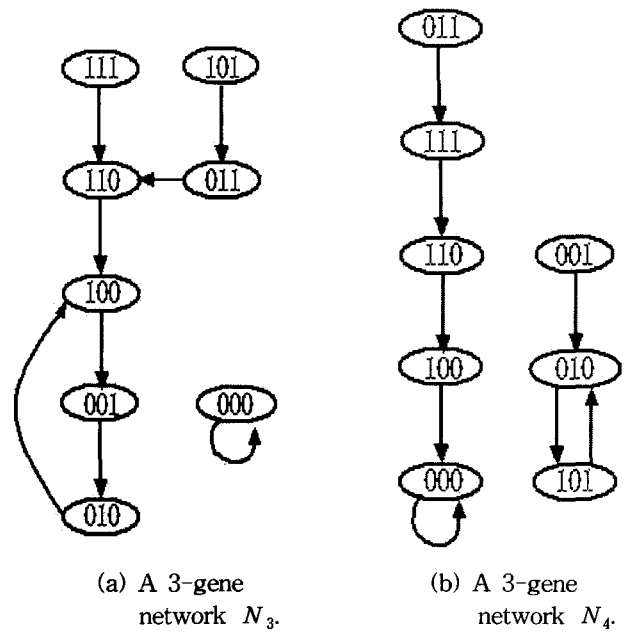
state vectors. Unlike continuous differential dynamic systems, the discrete space can be completely displayed as long as  $N$  is not too large. For example, we can visualize a three-dimensional cube for  $N=3$ . An autonomous solution subject to an initial state vector is of primary concern because it gives the sign of life in a cell. In other words, a zero-state response does not even exist, and only a zero-input response has to be dealt with. For all NK-network and all SORE, the whole state vector space can always be partitioned into  $I$  independent subspaces  $\{S_1, S_2, \dots, S_I\}$ , where  $I$  is the total number of attractors or limit cycles. If  $\alpha_{S_j}$  and  $\alpha_{S_k}$  represent the elements of subspace  $S_j$  and  $S_k$  respectively, then  $\alpha_{S_j} \cap \alpha_{S_k} = \emptyset$  and  $S_j \cap S_k = \emptyset$  for all  $j \neq k$ . All isolated islands (subgraphs that describe the subspaces) are called "basins of attractors".

### III. Concept of Colonization

Usually the complexity of a network is designated by the number of genes. Even with a few genes, the "colonizing action" of a network can create a much more complex colony in a short amount of time. As the "biologizing" control system becomes more complex, so do the fundamental issues of reachability, controllability, and observability.

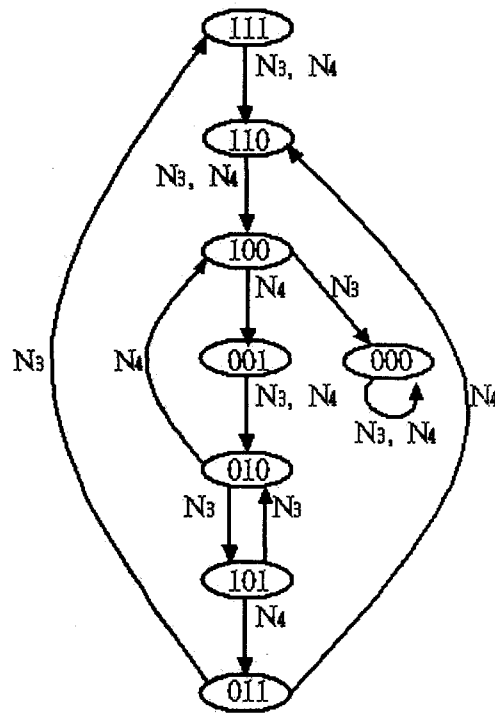
To illustrate only one example, let us assume that there are two known genetic networks and that both are made up of three genes (Fig. 2 (a) and (b)). There is only one attractor for each network where the existence of two three-gene networks would act autonomously. However, once the states of the two networks are connected (we call this a serial cascade), the behavior changes dramatically. Fig. 2 (c) shows that no less than five limit cycles now exist. The existence of these limit cycles carries special meaning when considered in reference to living cells. It is precisely this constant cyclic behavior that is prevalent in many forms of life. The number of limit cycles observed is also significant. A large number of them will exhibit the biological characteristics which Stuart A. Kauffman called "the edge of chaos." This chaotic behavior is triggered by the time-varying rules of Boolean logic.

The complexity in chaotic orders is due to the time-varying nature of the rules of Boolean logic. However, all of the complexity is due to autonomous activities. So far, there exists no zero-state response. The equivalence of such a response does exist, but it takes the form of dynamic systems expansion and enlargement; one may call it "colonization". Each gene is made up of a combination of the same molecules used in molecular genomics G, A, T, and C. When a new gene is introduced into genetic networks, a larger network emerges. For example, if a new gene is added to a four-gene network  $N_4$  and a five-gene network  $N_5$ , a



(a) A 3-gene network  $N_3$ .

(b) A 3-gene network  $N_4$ .



(c) Serial cascade of  $N_3$  and  $N_4$ .

Fig. 2 Two 3-gene networks and their serial cascade network

single much larger ten-gene network  $N_{10}$  emerges. The new genetic network will have  $2^{10}$  total state vectors, which is  $2^6$  or 64 times larger than the  $N_4$  genetic network and  $2^5$  or 32 times larger than the  $N_5$  genetic network. Why does one gene possess so much power? This single new gene may be viewed as a combination of big molecules, or it may represent a change in the

chemical environment, such as temperature, pressure, or other changes.

The modeling of various biological phenomena within the general framework of automata has been documented for quite some time. However, the renewed interest in modeling gene regulations in molecular biology in recent years has caught the attention of a much larger set of researchers. By far, SORE is the most generalized mathematical structure among all automata theories used in modeling gene regulations. Since a generalized example of ten genes demonstrating colonization requires much more space to document, its salient features will simply be mentioned without the presentation of its computer simulations.

#### IV. Other Properties for Biologizing Control System

The "biologizing" control theory states that reliability and survivability are the top priorities for a biocontrol system. How does one achieve this goal? One example is a system component like the Boolean network or SORE which can deliver this requirement of reliability and survivability. SORE exhibits redundancy characteristics as described in reference [11]. Error correction code itself already exhibits redundancy characteristics. This allows SORE to have the reliability and survivability necessary for a biocontrol system. Of course, much more research is needed to have a deeper understanding of the issue.

What method of control should be exercised in an autonomous manner? Redundancy with a simple switching action after a decision is not a good design with respect to control and economics. Self organization, regulation, and control are basic requirements for a "biologizing" control system. Smooth control operation only be expected from a self organizable system. Self organization and regulation is a main capability of a gene behavior.

Classification lies in the heart of any intelligent and autonomous control system. For a "biologizing" control system to be the most effective, some classification must be made before the best choice can be selected. The main thesis of reference [12] is that SORE could be the most powerful known classifier. One of the most important issues of classification and pattern recognition theory is what classifier which yields highest percentage of correct recognition with data in any kind of structure. The emphasis of this issue is that the data points in problems will not necessarily be clustered nicely together. If this performance criterion is adopted, one can say with certainty that the best known classifiers such as Support Vector Machine (SVM) and Artificial Neural Network (ANN) will never be able to compete with SORE. In reference [10], the simplest case of SORE, the two gene network, was investigated and proven to be the best possible classifier suitable for any kind of data

structure. Unfortunately, for more complicated networks, more time and research is needed because the problem becomes a synthesis problem for discovering Boolean logic functions. SORE's abilities as a classifier may allow it to become a generic system component for a future "biologizing" control system.

With advances in MEMS and nanotechnology, all types of physical, chemical and biological sensors will soon be readily available to the public. This will allow future control systems to make more informed decisions based on their surroundings. Decision making capability under uncertainty, namely the approximation reasoning of fuzzy logic, is only a special case of SORE. The use of Boolean functions in Boolean networks allows each node or each gene to take any combination of logical connectors. Most of the methodology of fuzzy logic employed so far is a simple set of "IF ... THEN ..." rules, which is less general than SORE.

Brain wave bio potentials have been used in experiments for mobile robot control [13]. This is just one of the many possible brain-inspired applications. Intelligent computational analysis of the human genome will drive medicine for at least the next half century [14]. Will research into bioinformatics result in useful information well beyond what is currently known? Perhaps the manipulation of brain information may be governed by the device such as SORE. Obviously these ideas are not certain, but there is a definite chance that SORE may be used in future brain-inspired applications.

#### V. Concluding Remarks

We discussed that a model for gene regulatory network can be a generic building block for a "biologizing" control system. There has been a lot of research into control systems for many years. Everything from the relationships between intelligent systems and the fundamental rules of biology to system configurations and road maps for the future of control systems have been studied. On the other hand, we present a generic building block for control systems, because without effective system blocks, there can be no successful system.

The evolutionary processes of Boolean networks are also very important. SORE demonstrates evolutionary development through its adaptation and learning capabilities. In colonization, the input of one gene (logically connected) to two genetic networks of  $N_4$  and  $N_5$  will produce a colony of  $N_{10}$  which is many times larger. The biological phenomena of self-reproduction can also be explained.

As a model for a gene regulatory network can be considered as a generic building block for a biologizing control system, a gene behavior can be studied as a new idea for the systemization of an intelligent algorithm.

## References

- [1] J. S. Albus, "Outline for a theory of intelligence", IEEE Trans. SMC, vol.21, no.3, pp.473-509, 1991
- [2] J. S. Albus, "The engineering of mind", Information Sciences, vol.117, pp.1-18, 1999
- [3] J. L. Casti, "Biologizing" control theory: How to make a control system come alive", Complexity, vol.7, no.4, pp.10-12, 2002
- [4] Report of the workshop Held at the University of Santa Clara on September 18-19, 1986, "Challenges to control: A collective view", IEEE Trans. Automatic Control, vol.32, no.4, pp.275-285, 1987
- [5] R. Lathrop, "Intelligent systems in biology: Why the excitement?", IEEE Intelligent Systems, pp.8-13, Nov./Dec. 2001
- [6] J. Leith *et al.*, "Toward more intelligent annotation tools: A Prototype", IEEE Intelligent Systems, pp.42-51, Nov./Dec. 2001
- [7] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets", J. of Theoretical Biology, vol.22, pp.437-467, 1969
- [8] P. P. Wang and H. D. Cheng, "SORE: Self organizable and regulating engine" A research proposal submitted to US NSF, Feb. 20, 2003, File No.: 03248-27, Fastlane
- [9] P. P. Wang and J. Robinson, "What is SORE?", The 7th Proceedings of the Joint Conferences on Information Sciences, Sep. 2003
- [10] P. P. Wang *et al.*, "A study of the Two-Gene Network - The simplest special case of SORE", The 7th Proceedings of the Joint Conferences on Information Sciences, Sep. 2003
- [11] P. P. Wang and H. Tao, "A Novel method of error correcting code generation based upon SORE", The 7th Proceedings of the Joint Conferences on Information Sciences, Sep. 2003
- [12] P. P. Wang and J. Yu. "SORE - A powerful classifier", The 7th Proceedings of the Joint Conferences on Information Sciences, Sep. 2003
- [13] K. H. Choi and M. Sasaki, "Brain-wave bio potentials based mobile robot control: Wavelet-neural network pattern recognition approach", IEEE Int. Conference on SMC, vol.1, pp.322-328, 2001
- [14] R. B. Altman, "Challenges for intelligent systems

in biology", IEEE Intelligent Systems, pp.14-18, Nov./Dec. 2001

- [15] J. M. Evans *et al.*, "Knowledge engineering for real time intelligent control", Proceedings of the 2002 IEEE Int. Symposium on Intelligent Control, Oct. 2002

## 저 자 소 개



### Byung-Jae Choi

He received the B.S. degree in Department of Electronic Engineering from Gyungbuk National University, Korea, in 1987, and the M.S. and Ph. D. degrees in Department of Electrical Engineering from KAIST, Korea, in 1987 and 1998 respectively. Since 1999,

he has been a faculty member of the School of Electronic Engineering at the Daegu University, where he is currently an Assistant Professor. His research interests are intelligent control theory and applications. He is a member of IEEE, KFIS, KITE, KIEE, and ICASE.

Phone : +82-53-850-6633  
Fax : +82-53-850-6619  
E-mail : bjchoi@daegu.ac.kr

### Paul P. Wang

He received the B.S. degree in Department of Electrical Engineering from Taiwan University, Taiwan, in 1958, and the Ph. D degree in Department of Electrical Engineering from Ohio State University, USA. Since 1968, he has been a faculty member of the Department of Electrical & Computer Engineering at the Duke University, USA, where he is currently a Professor. His research interests are intelligent control theory, bioinformatics, and bio-engineering.

Phone : +1-919-660-5259  
Fax : +1-919-660-5293  
E-mail : ppw@ee.duke.edu