

# Heuristic Continuous-time Associative Memories

Truong Quang Dang Khoa, Masahiro Nakagawa

**Abstract**—In this paper, a novel associative memory model will be proposed and applied to memory retrievals based on the conventional continuous time model. The conventional model presents memory capacity is very low and retrieval process easily converges to an equilibrium state which is very different from the stored patterns. Genetic Algorithms is well-known with the capability of global optimal search escaping local optimum on progress to reach a global optimum. Based on the well-known idea of Genetic Algorithms, this work proposes a heuristic rule to make a mutation when the state of the network is trapped in a spurious memory. The proposal heuristic associative memory show the stored capacity does not depend on the number of stored patterns and the retrieval ability is up to  $\sim 1$ .

**Keywords**—Artificial Intelligent, Soft Computing, Neural Networks, Genetic Algorithms, Hopfield Neural Networks, and Associative Memories.

## I. INTRODUCTION

**D**YNAMIC associative neural memory architectures and their learning algorithms are considered as nonlinear dynamical systems that information retrieval is performed as an evolution of the system's state in a high-dimensional state space. The retrieval is implemented by first initializing state with noisy or partial input pattern and then allowing the memory to perform a progressing search to find out the closest associative stored memory.

During the past quarter century, the numerous autoassociative models have been extensively investigated on the basis of the autocorrelation dynamics. Since the proposals of the retrieval models by Anderson in [2], and Hopfield in [3], it has been well appreciated that the storage capacity of the autocorrelation model, or the number of stored pattern vectors,  $L$ , to be completely associated vs the number of neurons,  $N$ , which is called the relative storage capacity or loading rate and denoted as  $L/N$ , is estimated as  $\sim 0.14$  at most for the autocorrelation learning model with the activation function as the signum one.

For the limited capacity as above mention, many of researches were proposed to improve it. One of the considered

approaches is changes of active functions. Morita [10] replaced the monotonous active functions by a neuro-dynamic model with nonmonotonous mapping. He reported that the nonmonotonous mapping in a neuro-dynamics possesses a remarkable advantage in storage capacity,  $\sim 0.27$ , superior than the conventional association models with monotonous mappings. Nakagawa [12] [13] proposed sinusoidal mappings or skew tent mappings. The chaotic behavior of the network based on the nonmonotonous chaos dynamics and relaxation of the monotonous decreasing of the energy was found. The convergence to the spurious memories may be avoided through the parameter control from the monotonously increasing function of the activation function to the monotonously decreasing one. Therefore, high association ability was shown up to  $\sim 0.5$ .

Another approach is improvement of energy functions. The energy function of networks always evolves in the direction that leads to lower network energy. Updating rules that based on the quadratic form of energy functions were investigated. Nakagawa [14] defined in terms of the entropy function instead of the conventional quadratic functions. The author realized the twice larger storage capacity in comparison with conventional model.

The correlation weight matrix was also considered as improved models. Venkatesh [2] proposed updating rules of weight matrix modified by a diagonal matrix of positive eigenvalues and individual orthogonal matrix for increasing the radii of attraction-spheres around memories and specifying increased attraction in certain directions around each of memories.

Yanai and Amari in [7] studied the probability of a state of networks belonging to the basin of attraction. The authors proposed the statistical neurodynamical method to analyze the one-step recall dynamics of the two-stage dynamic model and obtained the capacity of memory  $\sim N/\sqrt{\log N}$

Although many improved models have been proposed, major problems of autocorrelation model have still presented. Memory capacity is very low. Basins of attraction of stored patterns are small. There are many spurious memories, equilibrium states different from stored patterns.

To improve the conventional associative memory model, this work proposes a novel approach based on the idea of well-known Genetic Algorithms (GAs). GAs is known with the capability of global optimal search through series of procedures as selection, crossover and mutation. GAs always

Manuscript received July 4<sup>th</sup>, 2007.

This work was supported in part by The 21<sup>st</sup> Century COE (Center of Excellence) Program "Global Renaissance by Green Energy Revolution" from Ministry of Education, Culture, Sports, Science and Technology of Japan.

Authors are with the Chaos and Fractals Informatics Laboratory, Nagaoka University of Technology, 1603-1 Kamitomiokamachi, Nagaoka, Niigata 940-2188, JAPAN (e-mail: khoa@ieee.org).

escapes local optimum on progress to reach a global optimum. This advantage is applied for associative memory models that almost easily trap at local spurious memories. Furthermore, this approach still keeps the conventional active function as well as the autocorrelation weight matrix because this matrix is really includes itself much more information capacity. The advantage of improvement is the stored capacity does not depend on the number of stored patterns and the retrieval ability is up to  $\sim 1$ . The remainder of this paper is organized as follows. In section 2, the mathematic models of associative memories and heuristic rule are introduced. In section 3, some illustrations and discussion is described. Finally, section 4 is conclusion.

## II. MATHEMATICS MODELS

### A. Continuous-time Associative Memory

The stored binary vector:

$$e_i^{(r)} = \pm 1 \quad (1 \leq i \leq N, 1 \leq r \leq L) \quad (1)$$

where  $N$  and  $L$  are the number of neurons and the number of stored pattern vectors.

The states of the neural network are characterized in terms of the output vector

$$S_i \quad (1 \leq i \leq N)$$

and the internal states

$$\sigma_i \quad (1 \leq i \leq N)$$

which are related each other in terms of

$$S_i = f(\sigma_i) \quad (2)$$

where  $f(\cdot)$  is the activation function of the neuron. To define a continuous time model, it is usually chosen a nonlinear function such *Tanh*

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

A matrix of weights  $W$  is defined by Hebbian learning rule as

$$W = \frac{1}{N} \sum_{i=1}^N e_i e_i^T \quad (4)$$

The continuous associative memory is described by the following differential equation

$$\frac{d\sigma(t)}{dt} = W.f[\sigma(t)] \quad (5)$$

Updating rule

$$\sigma(t + \delta) = \int \frac{d\sigma(t)}{dt} dt \approx \sigma(t) + \delta \cdot \frac{d\sigma(t)}{dt} \quad (6)$$

where  $\delta$  is distance between two simulated computing steps.

The overlaps  $m^{(r)}$  are defined by

$$m^{(r)} = \sum_{i=1}^N e_i^{+(r)} s_i \quad (7)$$

Here the covariant vector  $e_i^{+(s)}$  is defined in the terms of the following orthogonal relation,

$$\sum_{i=1}^N e_i^{+(r)} e_i^{+(s)}(t) = \delta_{rs} \quad (1 \leq r, s \leq L) \quad (8)$$

$$e_i^{+(r)} = \sum_{r'=1}^L W_{rr'}^{-1} e_i^{r'} \quad (9)$$

Then the overlaps are calculated in terms of orthogonal relation:

$$|m^{(r)}| = \delta_{rs} \quad (10)$$

and

$$\sum_{r=1}^L |m^{(r)}| = 1 \quad (11)$$

Initial vector

$$s_i(0) = \begin{cases} -e_i^s & (1 \leq i \leq H_d) \\ +e_i^s & (H_d + 1 \leq i \leq N) \end{cases} \quad (12)$$

where  $e^{(s)}$  is a target pattern that needs to be retrieved and  $H_d$  is the Hamming distance between the initial vector  $s_i(0)$  and the target vector  $e^{(s)}$ .

The retrieval process is succeeded if

$$m^{(s)}(t) = \sum_{i=1}^N e_i^{+(s)} s_i(t) \quad (13)$$

results in  $\pm 1$  for  $t \geq 1$ , in which the system will be in steady state such that

$$\begin{aligned} s_i(t+1) &= s_i(t) \\ \sigma_i(t+1) &= \sigma_i(t) \end{aligned} \quad (14)$$

### B. Mutative Procedure of Genetic Algorithms

Mutation is a genetic operator that alters one or more gene values in a chromosome from its initial state. This can result in entirely new gene values being added to the gene pool. With these new gene values, the genetic algorithm may be able to arrive at better solution than was previously possible. Mutation is an important part of the genetic search as help helps to prevent the population from stagnating at any local optima. Mutation occurs during evolution according to a user-definable mutation probability.

### C. Improvement of Continuous-time Associative Memory by Mutation

There are two phases to the operation of autoassociative memory, namely the storage phase and the retrieval phase. The first one according to the outer product rule, Hebbian learning rule, defines a symmetric weight matrix, representing the patterns to be memorized by the network. During the retrieval phase, an  $N$ -dimensional vector called a probe, representing the state of network, typically contents an incomplete or noisy version of stored memories. The asynchronous update described above is continued until there are no further changes to report. Unfortunately, the spurious states of networks increase exponential with  $N$ . The associative memory has difficultly reached limited stored memories as shown in Fig. 1.

This paper defines a mutative error to detect retrieval process and proposed a heuristic rule to make a mutation when the state of the network is trapped in a spurious memory. The term, mutative error, is the difference of overlaps between present state and previous state. When the mutative error is small enough to determine the state of network is being trapped, the mutation phase is excited to try a better solution.

III. ILLUSTRATION AND DISCUSSION

A. 4-pattern problem

A simple illustration is used to examine the continuous-time model. This example has 4 stored patterns which are encoded by a 10x10 binary pixel grids, are shown in Fig. 2. In this test, we present the high noisy stored patterns as input patterns to test how the model behaved upon pattern association.

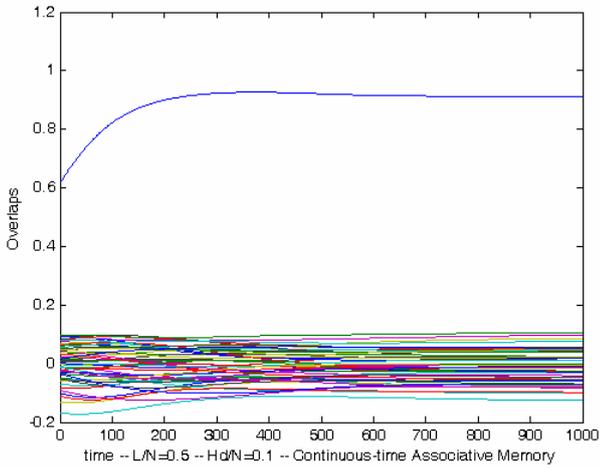


Fig. 1 Overlaps of unsuccessful retrieval process.

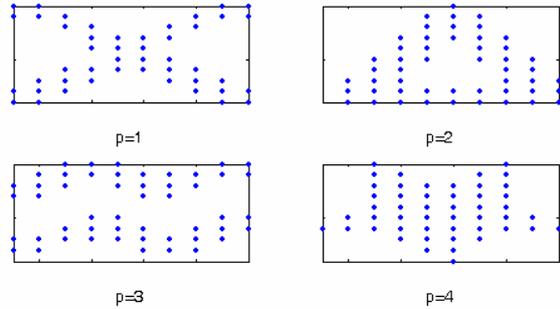


Fig. 2 The four stored patterns.

As shown in Fig. 3. , the continuous time model can be successfully recalled the correct stored pattern within around 30 steps, and it will be stabilize in a steady state. All of 4 patterns are included 40% noisy as initial patterns. The

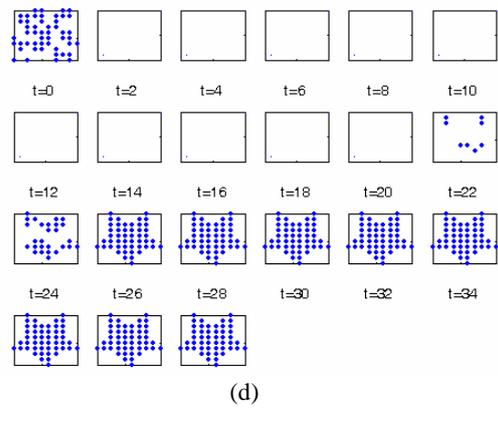
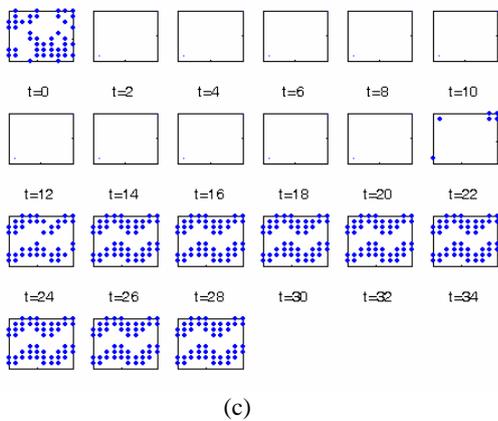
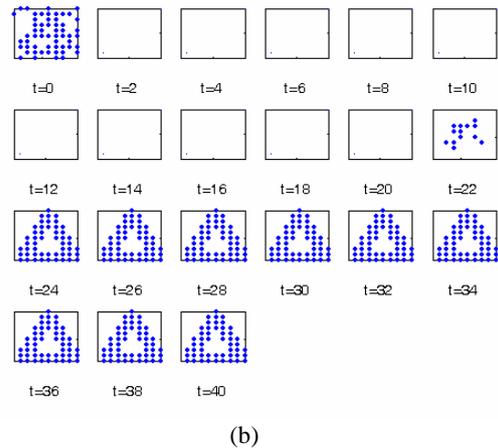
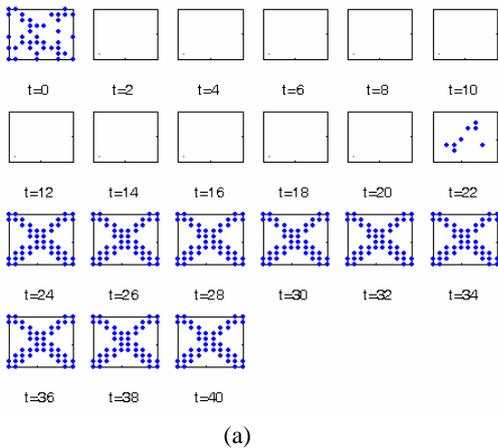


Fig. 3 (a)-(d). The sequence of 4 patterns with 40% noisy.

computing time of convergence depends on the value of simulation computing step. This example can be convergence in a shorter time with the larger computing step.

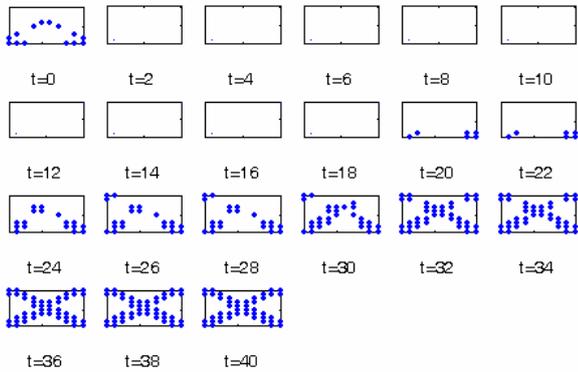
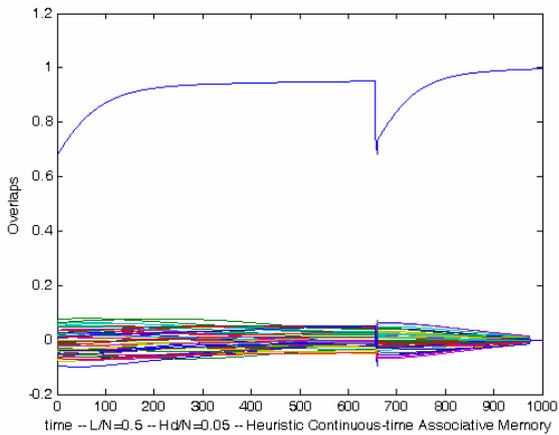


Fig. 4. Retrieval to the closest associative stored memory.

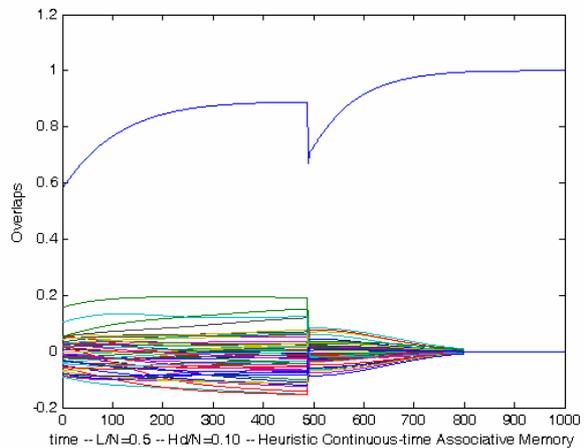
The second test is shown on Fig. 4. to illustrate that the successful retrieval pattern is the nearest one among stored patterns. Hamming distances of from the initial pattern as shown in Fig. 4. to every stored pattern are correspondent to [0.3800; 0.4200; 0.5600; 0.5500]. The final target pattern converges to the first one. Comparing to Lee model in [6], the second one was retrieved.

*B. Enlarge capacity of memory up to ~1*

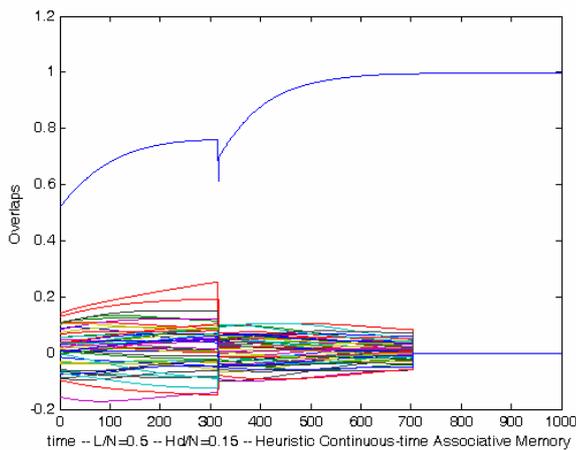
We first present an example of the dynamics of the overlaps in Fig. 5(a)-(d). In this case, we choose the number of neurons,  $N=100$ , and the stored rate,  $L/N=0.5$ , is set as const. Fig. 5. (a), (b), (c) and (d) correspond to the different initial values of Hamming distances,  $Hd/N=\{0.05; 0.1; 0.15; 0.2\}$ . It is clear to see the Heuristic rule is shown in this illustration. When the retrieval process traps in a local optimal called a spurious, the Heuristic rule generates a new state and the final results can be found out. The time that one of the overlaps increases up to 1 and the others go to zero represents the



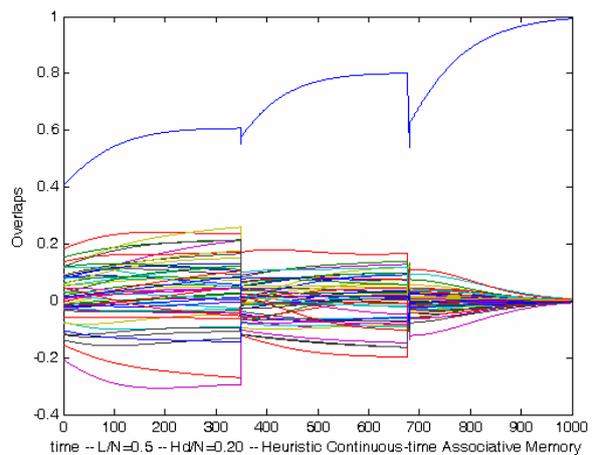
(a)



(b)



(c)



(d)

Fig. 5(a)-(d). The time dependence of overlaps of the proposed model.

success of retrievals. In Fig. 5(d) the retrieval process is successful after 2 trials.

To see the retrieval ability of the present model, the success rate  $S_r$  is defined as the rate of the success of 10 trials with the different stored patterns  $e_i^{(r)}$  ( $1 \leq i \leq N, 1 \leq r \leq L$ ). Every trial the maximum of iterations is set 1,000. Then we present the dependence of the success rate  $S_r$  on the stored capacity  $L/N$  are depicted in Fig. 6.(a)-(d), corresponding to the different initial values of Hamming distances,  $H_d/N=0.05; 0.1; 0.15; 0.2$ .

For the comparison, the corresponding results of conventional models are shown in Fig. 7(a)-(d). It is found that the present approach may achieve larger memory capacity than the conventional autocorrelation strategy. The storage capacity beyond the conventional one with the depression of the success rate does not depend on the number of stored patterns.

It is easy to see that when the memory capacity rate  $L/N$  reaches to 1, the Heuristic model mostly can retrieve successfully. The retrieval process does not depend on the

number of stored patterns,  $L$ . It has only depended on the limited time of iterated process. Increasing number of iterations up to 10,000 the result is shown in Fig. 8.

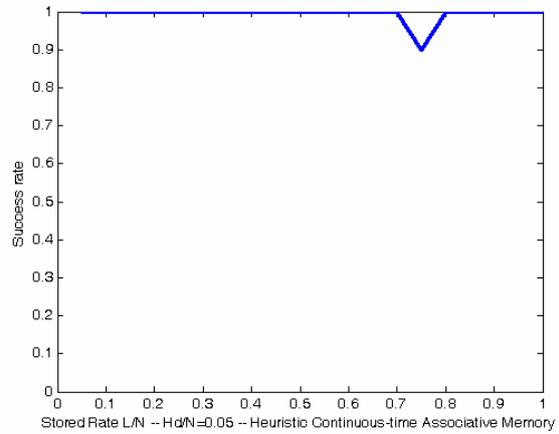
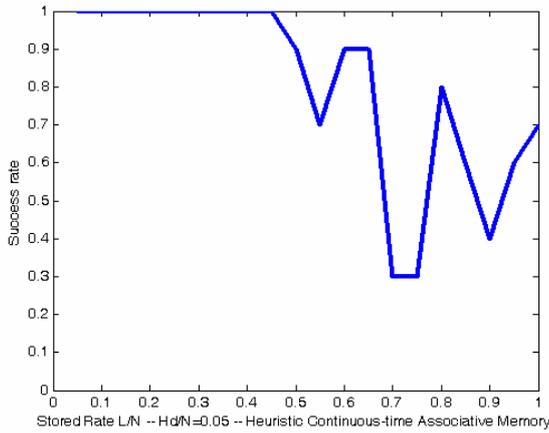
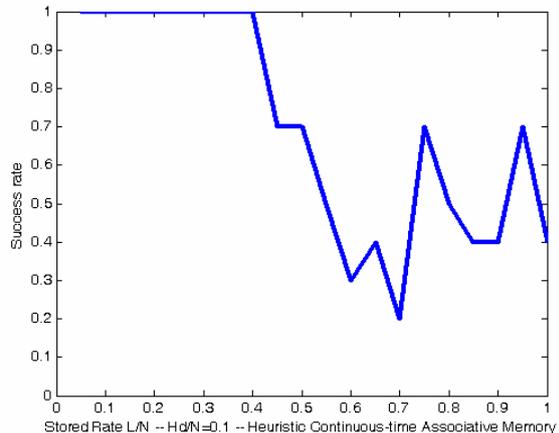


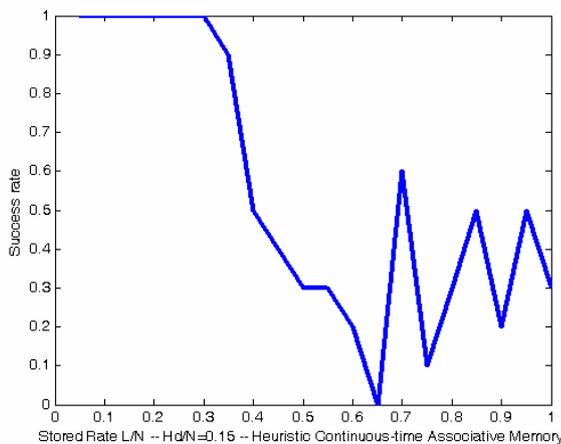
Fig. 8 The independence of the memory capacity  $L/N$  on number of stored patterns,  $L$ , of Heuristic Continuous-time Associative



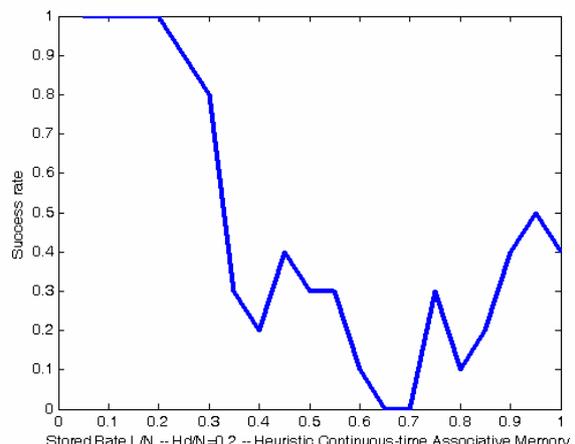
(a)



(b)



(c)



(d)

Fig. 6(a)-(d). The dependence of the success rate on the memory capacity  $L/N$  of Heuristic Continuous-time Associative Memories.

Memories with number of iterations up to 10,000.

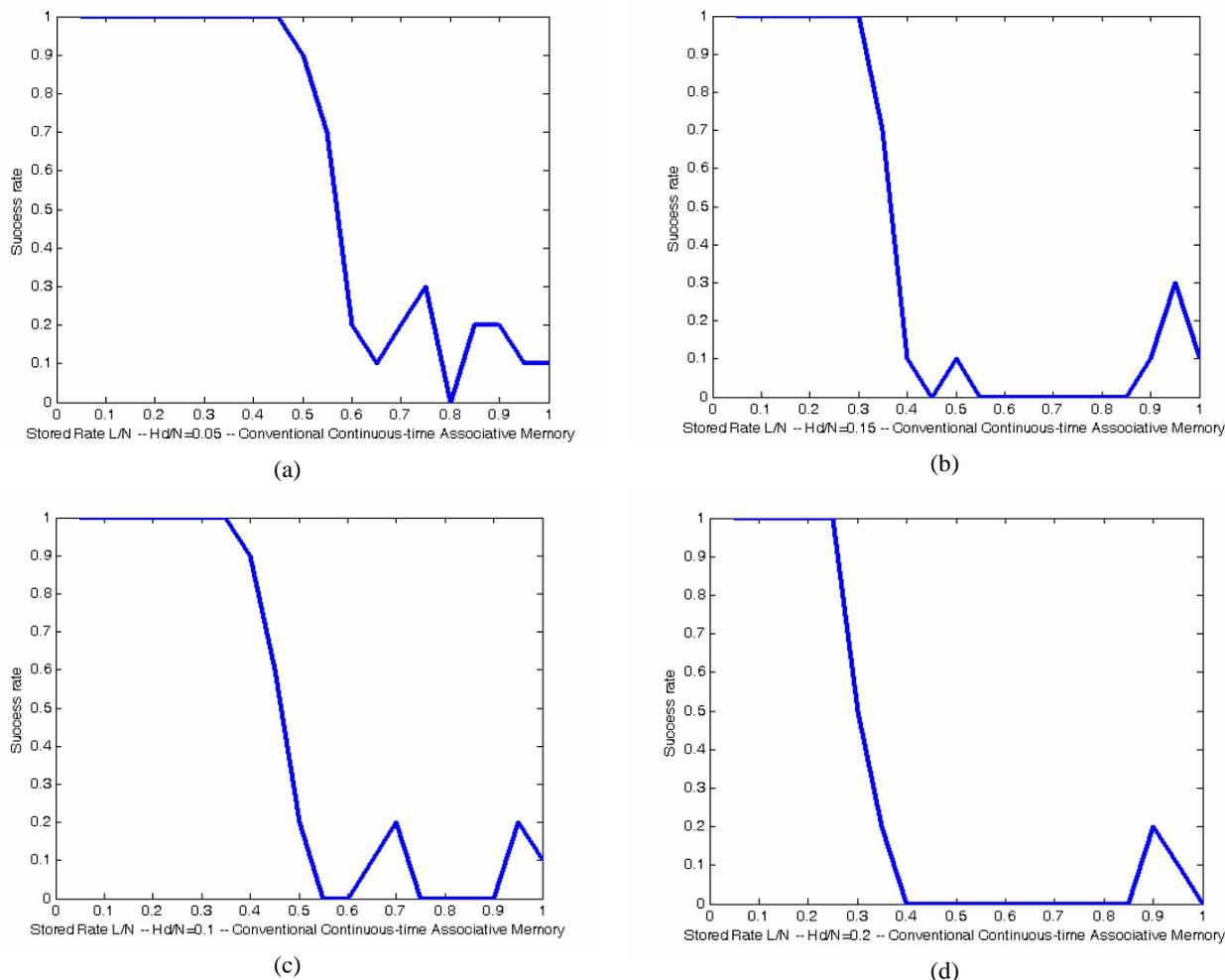


Fig. 7(a)-(d). The dependence of the success rate on the memory capacity  $L/N$  of Conventional Continuous-time Associative Memories.

#### IV. CONCLUSION

In the present paper, we have proposed a novel associative memory based on the hybrid algorithm between conventional model and the mutative idea of Genetic Algorithms. From computer simulation results, it has been found that the large storage capacity, characterized by the success rate curve, does not depend on the number of stored memories. As a future work, controlling the mutative process can obtain better results for associative memory model around the high sensitive area.

#### ACKNOWLEDGMENT

The authors would like to thank the WASET Editors, and the referees for interesting and valuable discussions and publications.

#### REFERENCES

- [1] Masahiro Nakagawa, *Chaos and Fractals in Engineering*, World Scientific, Singapore, 1999.
- [2] Mohamad H. Hassoun, *Associative Neural Memories: Theory and Implementation*, Oxford University Press, USA 1993.
- [3] S. Haykin, *Neural Networks: A Comprehensive Foundation, Second Edition*, Macmillan, New York, 1999.
- [4] C. M. Marcus and R. M. Westervelt, "Stability and convergence of analog neural networks with multiple-time-step parallel dynamics", *Phys. Rev. A* 42, 1990, pp.2410
- [5] F. R. Waugh, and R. M. Westervelt, "Associative memory in an analog iterated-map neural network C. M. Marcus", *Phys. Rev. A* 41, 1990, pp.3355
- [6] Lee, R.S.T., "A transient-chaotic autoassociative network (TCAN) based on Lee oscillators", *IEEE Trans. Neural Networks*, Vol. 15, No. 5, Sep. 2004, pp.1228 - 1243
- [7] Yanai, H.-F., and Amari, S.-I., "Auto-associative memory with two-stage dynamics of nonmonotonic neurons", *IEEE Trans. Neural Networks*, Vol. 7, No. 4, July 1996, pp.803 - 815
- [8] Seiji Miyoshi, and Masato Okada, "Storage Capacity Diverges With Synaptic Efficiency in an Associative Memory Model With Synaptic

- Delay and Pruning”, *IEEE Trans. Neural Networks*, Vol. 15, No. 5, Sep. 2004, pp.1215
- [9] Masaharu Adachi and Kazuyuki Aihara, “Associative Dynamics in a Chaotic Neural Network”, *J. Neural Networks*, Vol. 10, No. 1, Jan. 1997, pp. 83-98
- [10] Masahiko Morita, “Associative memory with nonmonotone dynamics”, *J. Neural Networks*, Vol. 6, No. 1, 1993, pp. 115-126
- [11] Masahiro Nakagawa, “Statistical Properties of Chaos Associative Memory”, *J. of the Physical Society of Japan*, Vol. 71, No. 9, Sep. 2002, pp. 2316-2325
- [12] Masahiro Nakagawa, “A Chaos Associative Memory with a Skew-Tent Activation Function”, *J. of the Physical Society of Japan*, Vol.69, No.3, Mar. 2000, pp. 972-980
- [13] Masahiro Nakagawa, “Chaos Associative Memory with a Periodic Activation Function”, *J. of the Physical Society of Japan*, Vol. 67, No. 7, July 1998, pp. 2281-2293
- [14] Masahiro Nakagawa, “Entropy Based Associative Memory”, *IEICE Trans. Fundamentals of Electronics, Communications and Computer*, Vol.E89-A, No.4, 2006, pp.895-901
- [15] Truong Quang Dang Khoa, Masahiro Nakagawa, Neural Network Learning based on Chaos, *WASET Inter. J. Computer, Information, and Systems Science, and Engineering*, Vol. 1, No. 2, 2007, pp.97-102

**Truong Quang Dang Khoa** was born in the South of Vietnam in 1979. He graduated in 2002, received the Master Degree and became a lecturer of the Hochiminh University of Technology in 2004. Now he is pursuing PhD study and working as a research assistant in Nagaoka University of Technology under the support of The 21<sup>st</sup> Century COE (Center of Excellence) Program “Global Renaissance by Green Energy Revolution” from Ministry of Education, Culture, Sports, Science and Technology of Japan. His research interests are Artificial Intelligent Computation, and Power System Analysis. He is a member of IEEE.

**Masahiro Nakagawa** was born in Kyoto, Japan in 1958. He received the B.Sc. and M. Sc. degrees from Nagaoka University of Technology, in 1980 and 1982, respectively. He also received Ph.D degree Nagoya University in 1988. He has been a Professor at the Nagaoka University of Technology since 2001. His recent activity is concerned with the fractals and chaos in science and engineering, e.g. the chaos in EEG and vocal sounds, the chaos neural networks with a periodic activation function, the iterated function system for image coding etc. Also he is interested in the physics of liquid crystals involving the molecular theory and the viscoelastic continuum theory. He is a member of the Japanese Neural Network Society, the Physical Society of Japan, and Japanese Society of Applied Physics.