Evolutionary and Hebbian analysis of hierarchically coupled associative memories

ROGÉRIO MARTINS GOMES[1]
HENRIQUE ELIAS BORGES[1]

Centro Federal de Educação Tecnológica de Minas Gerais - CEFET-MG
Intelligent Systems Laboratory - LSI
Av. Amazonas 7675, CEP 30.510-000, Belo Horizonte, MG, Brazil
[1] (rogerio,henrique)@lsi.cefetmg.br

Abstract. Inspired by The Theory of Neuronal Group Selection (TNGS) we have conducted a comparative storage and retrieval analysis of a multi-level or hierarchically coupled associative memory through evolutionary computation and Hebbian learning. The TNGS establishes that memory processes can be described as being organized, functionally, in hierarchical levels, where higher levels coordinate sets of functions of the lower levels. The most basic units in the cortical area of the brain are formed during epigenesis and are called neuronal groups, which are defined as a set of localized tightly coupled neurons constituting what we call our first-level blocks of memories. On the other hand the higher levels are formed during our lives, or ontogeny, through selective strengthening or weakening of the neural connections amongst the neuronal groups. In this sense, this paper describes and compares a method of acquiring the inter- group synapses for the proposed coupled system using both evolutionary computation and Hebbian learning. The results show that evolutionary computation, more specifically genetic algorithms, is more suitable for network acquisition than Hebbian learning because it allows for the emergence of complex behaviours which are potentially excluded due to the well known crossover effect constraints presented in Hebbian learning. Simulations have been carried out considering a wide range of the system parameters.

Keywords: Associative memory, Evolutionary Computation, Generalized-brain-state-in-a-box (GBSB) model, Theory of neuronal group selection (TNGS).

(Received August 20th, 2012 / Accepted October 14th, 2012)

1 Introduction

Theories and mathematical models have been studied over the years in an attempt to capture particular features of the brain mainly of the neurons and their interactions [2], [7], [8], [9],[11], [14], [20], [27], [28], [30], [35], and [37]. It is reasonable to assume that many of the general principles upon which the nervous system is conceived may be described through a much simplified model that takes into account the most relevant features of each neuron and the interactions amongst them. Although incomplete, some models are suitable for theoretical investigations and can enable us to understand some brain features through the subdivision of the neural networks into smaller structures [7].

In general, these studies treat the neocortex is regarded as an associative memory in which some of the long and short-range cortical connections are responsible for the storage and retrieval of global patterns. From this perspective, the neocortex could be divided into various discrete modular units where the short-range connections correspond to the synapses between neurons of the same module while the long-range connections can be represented as synapses between neurons of different modules [7], [9], and [20].

Consenquently, Gomes et al. [16] addressed in their work the pattern storage and retrieval problem and they also considered the implication of developmental organisation of neurobiological structure. This development may result in pre-existing synaptic connections as well as the hierarchically coupled organisation of the network. Thus, based on the same principles of multi-module organisation, the Theory of Neuronal Group Selection (TNGS), proposed by Edelman [11], establishes that memory processes can be described as being organized, functionally, in hierarchical levels, where higher levels coordinate sets of functions of the lower levels. In Edelman's theory, synapses of the localized neural cells in the cortical area of the brain generate a hierarchy of cluster units denoted as: neuronal groups (clusters of tightly coupled neural cells), local maps (reentrant clusters of coupled neuronal groups) and global maps (reentrant clusters of coupled neural maps). Edelman argues that the neuronal group is the most basic unit in the cortical area of the brain and, it is the basic constructor of memories. These neuronal groups are a set of tightly coupled neurons, firing and oscillating synchronically, developed in the embryo as well as during the beginning of a child's life, *i.e.*, they are structured during phylogeny and are responsible for the most primitive functions in human beings. In other words, the neuronal groups are hardwired, meaning that they are difficult to change. Considering these principles, these neuronal groups would be, equivalently, the first-level memories of our artificial model.

Immediately after birth, the human brain rapidly starts creating and modifying synaptic connections between neuronal groups. According to this proposition, Edelman proposed an analogy based on the Darwin's Theory of Natural Selection and Darwinian theories of population dynamics. The term Neural Darwinism could be used to describe a physical process observed in neurodevelopment in which synapses, which were used amongst different clusters (neuronal groups) are strengthened, while unused ones are weakened, giving rise to a second level physical structure regarded in the TNGS as a local map. Each of these arrangements of connections amongst clusters within a given local map results in a certain inter-neuronal group activity which yield a second-level memory. In other words, the second-level memory could be viewed as a correlation amongst the first-level memories. This process of coupling smaller structures through synaptic interconnections between neurons of different neuronal groups in order to generate larger ones could be repeated recursively. Consequently, new hierarchical levels of memories emerge through selected correlations of the lower level memories [11].

Based on these arguments, Gomes et al. [13], [15], and [16] presented a model of multi-level associative memory having its first level structure formed by generalised-Brain-State-in-a-Box (GBSB) neural network. They also proposed an energy function for the coupled model and they proved that the inter-group coupling that enabled the emergence of second-level memories do not destroy the first-level memory structures.

The strategy adopted by Gomes et al. [16] is basically to build a predefined non-symmetric weight matrix for these neuronal groups, i.e. synthesised, by means of an algebraic calculation proposed by Lillo and collaborators [23], to meet the main characteristics of the TNGS theory. Thus, the matrix was carefully designed to store all of the desired patterns while minimising the number of the spurious states (undesired asymptotically stable equilibrium points). In addition, the second level was designed to follow the generalised Hebb rule or Outer Product Method algorithm through the selected patterns extracted from the first-level memories. This strategy for building a network where the first level is designed with tightly synaptic weights as the coupling amongst the other levels exhibit synaptic plasticity (Hebbian Learning) is closer to the model of neuronal organisation proposed by Edelman. Hence, the local maps or our second level memories, are not synthesized, instead, the correlations would emerge through a learning or adaptive mechanism.

As a result, this paper describes and compares a method of acquiring the inter-group synapses for the proposed coupled system using both evolutionary computation and Hebbian learning and is organized as follows. In section 2 we present the model of hierarchically coupled GBSB neural networks and show how multi-level memories may emerge from it. Section 3 illustrates the analysis made through a sequence of experiments, showing the behaviour of the global network and its capacity of convergence to global patterns for orthogonal and linearly independent (LI) vectors for both Hebbian learning and evolutionary computation. Finally, Section 4 concludes the paper.

2 Multi-level memories

In order to develop this new model, Gomes [13] uses an extension of the original ***BSB - Brain-State-in-a-Box*** [4] called ***GBSB*** (*Generalized-Brain-State-in-Box*) [18] which can be applied in the implementation of associative memories, where each stored pattern, *i.e.*, a memory, is an asymptotically stable equilibrium point [34]. The design of artificial neural network associative memories have been explored in the last two decades and some methods have been proposed in [10], [17], [21], [22], [24], [25], and [31]

Associative memories have also been studied, particularly in the cases where they are a part of a hierarchical or coupled system. Some authors regard the neocortex as being a kind of associative memory in which some of the long and short-range cortico-cortical connections implement the storage and retrieval of global patterns [29], [32], and [36]. Thus, the cortex could be divided into various discrete modular elements where the short-range connections will be those synapses between neurons of the same module while the long-range connections would be synapses between neurons of different modules. In addition, these authors have considered in their models, symmetric connections, asynchronous updating, local and global features formed by Hebbian learning [30], [27], [28], and [35]. Notwithstanding, these synapses are expected to mimic some important characteristics inherent to biological systems [11], which have not been considered: parallelism amongst synapses in different regions of the brain, re-entrant and asymmetric connections, synchronous activation, different *bias* as well as different maximum and minimum fire rates, redundancy, nonlinear dynamics and self-connection for each neuron. For this reason, taking as inspiration the Theory of Neuronal Group Selection (TNGS) proposed by Edelman [8], and [11], a multi-level or hierarchically coupled associative memory based on coupled Generalized-Brain-State-in-a-Box (GBSB) neural networks was proposed and analyzed in [13], [15], and [33].

The Generalized-Brain-State-in-a-Box (GBSB) model [18] can be described by the following equation:

$$\mathbf{x}^{k+1} = \varphi((\mathbf{I}_n + \beta\mathbf{W})\mathbf{x}^k + \beta\mathbf{f}), \qquad (1)$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\beta > 0$ is a small and positive gain factor, $\mathbf{W} \in \Re^{n\times n}$ is the weight matrix, which need not be symmetrical, and $\mathbf{f} \in \Re^n$ is the bias field allowing us to better control the extent of the basins of attraction of the fixed points of the system. It is worth mentioning that when the weight matrix $\mathbf{W}$ is symmetric and $\mathbf{f} = \mathbf{0}$, the original model discussed in [39] will be recovered.

The activation function φ is a linear saturating function whose i^{th} component is defined as:

$$\begin{aligned} x_i^{k+1} &= \varphi(y_i^k) \\ \varphi(y_i^k) &= \begin{cases} +1 & \text{if } y_i^k > +1 \\ y_i^k & \text{if } -1 \le y_i^k \le +1 \\ -1 & \text{if } y_i^k < -1, \end{cases} \end{aligned} \qquad (2)$$

where y_i^k is the argument of the function φ in (1).

In our multi-level memories, each GBSB neural network plays the role of our first-level memory inspired by Neuronal Groups of the TNGS. In order to build a second-level memory we can couple any number of GBSB networks by means of bidirectional synapses. These new structures will play the role of our second-level memory memories analogous to the local maps of the TNGS. Hence, some global patterns could emerge as selected couplings of the first-level stored patterns.

Figure 1 illustrates a two-level hierarchical memory via coupled GBSB model where each one of the neural networks A, B and C, represents a GBSB network. In a given network, each single neuron has synaptic connections with all neurons of the same network, *i.e.*, the GBSB is a fully connected non-symmetric neural network. Besides, some selected neurons in a given network are bidirectionally connected with some selected neurons in the other networks [2], [3], [7], [27], [28], and [35]. These inter-network connections, named in this paper inter-group connections, can be represented by a weight inter-group matrix $\mathbf{W}_{cor}$, which accounts for the interconnections of the networks due to coupling. An analogous procedure could be followed in order to build higher levels in the proposed aforementioned hierarchy [1], and [11].

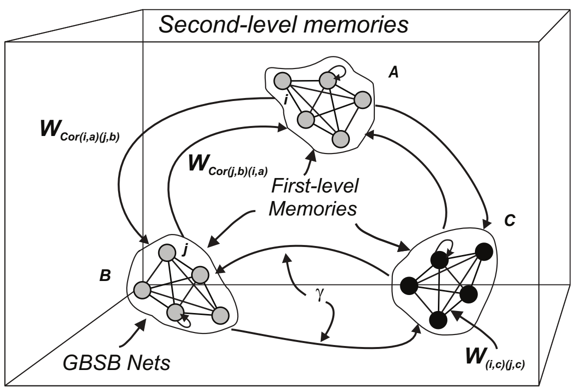

Figure 1: Coupled neural network design

In order to observe the results of coupling of a given

GBSB network with the remaining GBSB networks, one should extend (1) by adding to it a term which represents the inter-group coupling. Consequently, our general version of the multi-level associative memory model can be defined by:

$$x_{(i,a)}^{k+1} = \varphi\left(x_{(i,a)}^{k} + \sum_{j=1}^{N_a} \beta_a w_{(i,a)(j,a)} x_{(j,a)}^{k} + \beta_a f_{(i,a)} + \mu \sum_{\substack{b=1 \\ b\neq a}}^{N_r} \sum_{j=1}^{N_q} \gamma_{(a,b)} w_{cor(i,a)(j,b)} x_{(j,b)}^{k}\right), \quad (3)$$

where $x_{(i,a)}^{k}$ is the state of the i^{th} neuron of the a^{th} network at time k, $\beta_a > 0$ is a small and positive constant referred to as intra-group gain of the a^{th} network and $f_{(i,a)}$ is the bias field of the i^{th} neuron of the a^{th} network, $w_{(i,a)(j,a)}$ is the synaptic weight between the i^{th} and the j^{th} neuron of the a^{th} network, N_a is the number of neurons of the a^{th} network N_r is the number of networks, N_q is the number of neurons of the b^{th} network, *i.e.*, the number of neurons of the b^{th} network that are coupled to i^{th} neuron of the a^{th} network, μ is the coupling density amongst the networks, $w_{cor(i,a)(j,b)}$ is the synaptic weight between the i_{th} of the a_{th} network and the $j_t h$ neuron of the b_{th} network, and $\gamma_{(a,b)}$ is a positive constant referred to as inter-group gain between the a^{th} and b^{th} network, and $x_{(j,b)}^{k}$ is the state of the j^{th} neuron of the b^{th} network at time k. To sum it up, the first three terms represent the a_{th} single GBSB networks. The fourth term of (3), the sum over j, labels the N_q neurons in the b^{th} network that are connected to neuron i in the a^{th} network being the inter-group gain and the coupling density parameterised by $\gamma_{(a,b)}$ and μ respectively.

It is important to note that, in this general model, different β_a and $\gamma_{(a,b)}$ values could be assigned to each network as well as to pairs of them, respectively. However, without loss of generality we will be analyzing a particular case of this general version of the multi-level associative memory model in which the intra-group and inter-group gains are constant, *i.e.*:

$$\left.\begin{array}{lcl} \beta_a & \equiv & \beta \\ \gamma_{(a,b)} & \equiv & \gamma \end{array}\right\} \forall\, a, b \quad (4)$$

Equation (3) can be rewritten, in vectorial notation, as:

$$\mathbf{x}_a^{k+1} = \varphi\left(\left(\mathbf{I}_n + \beta\mathbf{W}_a\right)\mathbf{x}_a^k + \beta\mathbf{f}_a + \mu\gamma \sum_{b=1, b\neq a}^{N_r} \mathbf{W}_{cor}\mathbf{x}_b^k\right) \quad (5)$$

being $N_q = N_a = N_n$, that is, the networks have the same number of neurons.

In our model, it is necessary to consider that the weight matrix $\mathbf{W}_a$ is designed by following the algorithm proposed in [23]. Such algorithm ensures that the patterns which are symmetrical to the desired ones are not automatically stored as asymptotically stable equilibrium points of the network, therefore causing a minimisation in the number of spurious states as a result.

3 Simulation results

Up to this point, we have presented a model of multi-level associative memories and its associated equations that allow the system to evolve dynamically towards a global pattern when one of the networks is initialized in one of the previously stored patterns as a first-level memory [16].

In their work, [16] presented some simulations that validated their claims. The computational experiments consisting of three or more GBSB networks connected as in Figure 1 were conducted and each network was designed to present the same number of neurons and patterns stored as first-level memories. The weight matrix of each individual network was designed according to the algorithm proposed in [23], which ensures that those patterns symmetrical to desired ones are not automatically stored as asymptotically stable equilibrium points of the network, minimizing the number of spurious states as a result. The second-level memories, or global emergent patterns, were built by randomly selecting a set of patterns, which were stored as first-level memories considering linearly independent (LI) or orthogonal vectors. Assuming that each network contains *m* stored patterns or memories, a vector state in the μ^{th} memory configuration could be written as $\mathbf{P}_\mu$, $\mu = 1, \ldots, m$. In addition to this, the number and values of the stored patterns can be different in each network.

The selected patterns extracted from the first-level memories used to form a global pattern determine the inter-group weight matrix $\mathbf{W}_{cor(a,b)}$ by observing two analyses, one based on the generalized Hebb rule or Outer Product Method proposed in [16] and the other based on evolutionary algorithms.

3.1 Hebbian analysis

The inter-group weight matrix $\mathbf{W}_{cor(a,b)}$ considering the generalized Hebb rule was calculated as follows:

$$\mathbf{W}_{cor(a,b)} = \frac{1}{\sqrt{N_a}\sqrt{N_b}} \sum_{\mu=1}^{p} \mathbf{P}_{(\mu,a)} \mathbf{P}'_{(\mu,b)} \qquad (6)$$

where, $\mathbf{W}_{cor(a,b)}$ is the inter-group weight matrix between the a^{th} network and the b^{th} network, N_a is the number of neurons of the a^{th} network, N_b is the number of neurons of the b^{th} network and p is the number of stored patterns chosen as first-level memories to be second-level memories.

In our simulations each network contains 12 neurons producing 4096 possible patterns from which 6 were selected to be stored as our first-level memories. The selected set of 6 patterns stored as first-level memories were chosen randomly considering LI or orthogonal vectors. In addition, in the first experiment 3 amongst the $6^3 = 216$ possible combinations of the 3 sets of first-level memories have been chosen randomly to be our second level memories.

The system was initialized at time $k = 0$; randomly in one of the networks A, B or C, and in one of its first-level memories which compose a second level memory considering a typical value of β ($\beta = 0.2878$) [16].

The two other networks, in their turn, were initialized in one of the 4096 possible combination of patterns, also at random. Then, we measured the number of times that a system consisting of three coupled networks converged to a configuration of triplets. A triplet is one of the global emergent patterns (vector length 36) which constitutes a second-level memory when three networks are coupled. In the experiment, we considered a density of coupling amongst the inter-group neurons of 100%. The neurons which took part in the inter-group connections were chosen randomly and the rate of memory recovery in our experiments were averaged over 1000 trials for each value of γ. The results for LI and orthogonal vectors can be seen in Figure 2 which shows that our model presented a mean recovery rate of global patterns higher than 80% for LI vectors and a rate near 100% for orthogonal vectors considering a specific optimal value of γ (Table 1).

Table 1: Maximum mean rate of memory recovery and gamma values for orthogonal and LI vectors considering 3, 4 and 5 coupled networks

	3		4		5	
	ORT	LI	**ORT**	**LI**	**ORT**	**LI**
CONV. (%)	98.4	82.7	95.5	81.8	90.2	70.6
optimal gamma	0.9	1.4	0.7	1.2	0.6	1

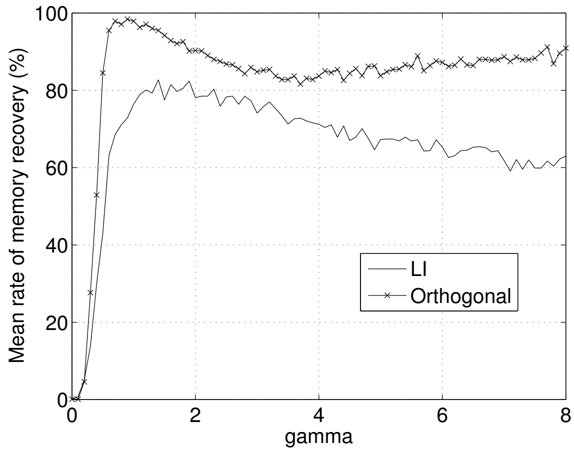

Figure 2: Triplets measured for LI and orthogonal vectors.

In the second experiment, we analyzed the capacity of convergence to a global pattern in systems when three, four and five networks are Coupled and three patterns of each network (first-level memories) were chosen at random to be second-level memories.

For example, considering a system with three coupled networks as shown in Figure 3, we assume that the stored patterns $\mathbf{P}_{(1,A)}$, $\mathbf{P}_{(4,A)}$ and $\mathbf{P}_{(6,A)}$ from network **A**, $\mathbf{P}_{(2,B)}$, $\mathbf{P}_{(5,B)}$ and $\mathbf{P}_{(6,B)}$ from network B and that $\mathbf{P}_{(1,C)}$, $\mathbf{P}_{(3,C)}$ and $\mathbf{P}_{(5,C)}$ from network C were chosen as first-level memories of each network to be the second-level memories. Therefore, our second-level memories will be a combination of these first-level memories, which are:

- second-level Memory 1: [$\mathbf{P}_{(1,A)}$ $\mathbf{P}_{(2,B)}$ $\mathbf{P}_{(1,C)}$];
- second-level Memory 2: [$\mathbf{P}_{(4,A)}$ $\mathbf{P}_{(5,B)}$ $\mathbf{P}_{(3,C)}$];
- second-level Memory 3: [$\mathbf{P}_{(6,A)}$ $\mathbf{P}_{(6,B)}$ $\mathbf{P}_{(5,C)}$].

The proceeder for four, five or more coupled networks is a straightforward extension of the previous one.

A comparison between all these different couplings can be seen in Figures 4 and 5. It can be observed that, for both LI and orthogonal vectors, the capacity of convergence to a global pattern decreases when more networks are coupled. In the experiment the system presented a better performance in relation to its capacity of convergence when orthogonal vectors were used (table 1).

In the experiments carried out so far, we stored 6 patterns (first-level memories) in each network. However, only 3 of these 6 stored patterns were chosen

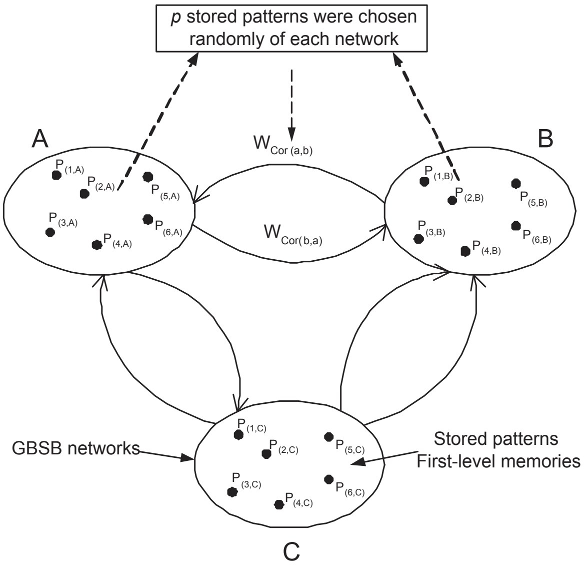

Figure 3: Coupled neural network design

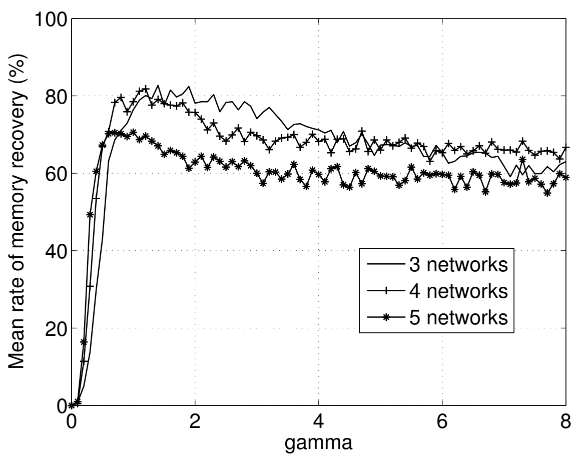

Figure 4: Mean rate of memory recovery for 3 to 5 coupled networks - LI vectors.

to compose the second-level memories. In the following experiment considering 3 coupled networks we will choose from 1 to 6 of these first-level memories to compose our second level-memories. Therefore we will have up to 6 different sets of triplets or global memories. In addition to it simulations considering $\beta = 0.2878$ and LI and orthogonal vectors will be performed. In Figures 6 and 7 we drew the convergence graph of the system. It can be observed that the system looses its capacity when a larger set of triplets is chosen to perform a second-level memory. This happens because our inter-group weight matrix ($w_{cor(i,a)(j,b)}$) has been determined by the generalized Hebb rule whose

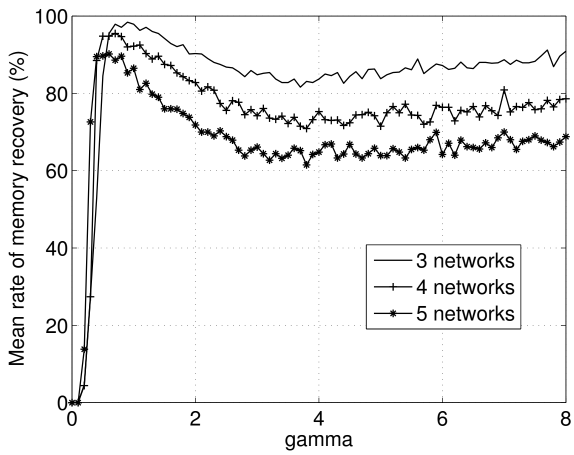

Figure 5: Mean rate of memory recovery for 3 to 5 coupled networks - Orthogonal vectors.

absolute capacity is equal $\frac{N}{2\ln(N)}$, where N is the total number of neurons. Furthermore, the differences in the rate of memory recovery between LI and orthogonal vectors owe to the term called *Cross Talk* or *Interference Term* which appears interfering with the recovery capacity. The aforesaid term is extremely dependent on the number and representation of the input vectors. In this way, when LI vectors are used to be our patterns this error term will represent an important value affecting the recovery rate of the system. On the other hand, when orthogonal vectors are used, this term will be equal to zero, thus decreasing the error rate of the system when retrieving the stored patterns. Table 2 shows the maximum mean rate of memory recovery and gamma values for orthogonal and LI vectors considering from 1 to 6 patterns chosen as first-level memories.

Table 2: Maximum mean rate of memory recovery and gamma values for orthogonal and LI vectors considering from 1 to 6 patterns chosen as first-level memories

Patterns	Type	Conv (%)	gamma
1	ORT	100	1
	LI	100	1
2	ORT	99.4	1
	LI	98	1.2
3	ORT	98.4	0.9
	LI	82.7	1.4
4	ORT	76.6	0.9
	LI	60	1
5	ORT	64.5	0.6
	LI	40.3	0.5
6	ORT	53.9	0.6
	LI	38.2	0.7

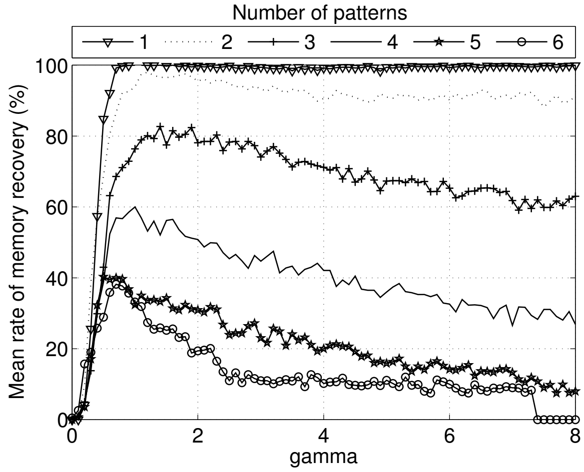

Figure 6: Mean rate of memory recovery obtained for 3 coupled networks considering from 1 to 6 patterns chosen as first-level memories - LI vectors.

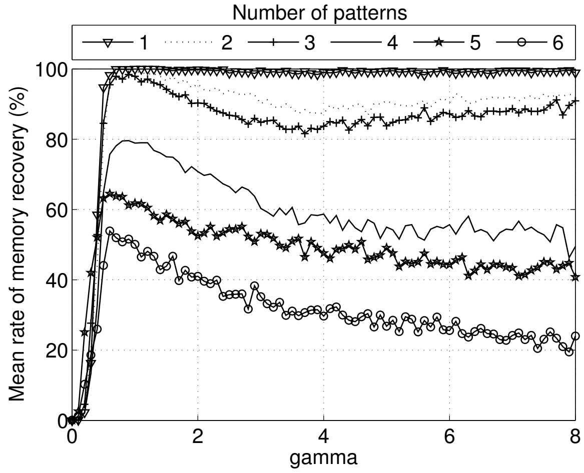

Figure 7: Mean rate of memory recovery obtained for 3 coupled networks considering from 1 to 6 patterns chosen as first-level memories - orthogonal vectors.

3.2 Evolutionary analysis

The convergence and capacity of the system was measured using the inter-group value (γ) and the inter-group weight matrix $\mathbf{W}_{cor(a,b)}$ which was calculated in accordance with a genetic algorithm strategy. In our simulations, the characteristics of the networks were as used in section 3.1 considering three to five GBSB networks connected as shown in Figure 3.

Firstly, the representation of each individual chosen was composed of real-valued variables, or genes. The aforementioned individual variables account for the γ values and the components w_{ij} of the inter-group weight matrix $\mathbf{W}_{cor(a,b)}$. This representation acts as the genotypes (chromosome values) and is uniquely mapped onto the decision-variable (phenotypic) domain.

The next step is to create an initial population consisting of 50 individuals whose first variable of each single one is the γ value. The remaining variables of each individual represent each one of the w_{ij} elements of the inter-group weight matrix $\mathbf{W}_{cor(a,b)}$. γ is a random real number uniformly distributed and ranges from 1 to 2 and w_{ij} is a random real number uniformly distributed which ranges from -0.5 to 0.5 (Figure 8) Moreover, one individual of the initial population has been seeded with the inter-group matrix developed in section 3.1. This technique aims to guarantee that the solutions produced by the Genetic Algorithm (GA) will not be less effective than the one generated by the Hebbian analysis.

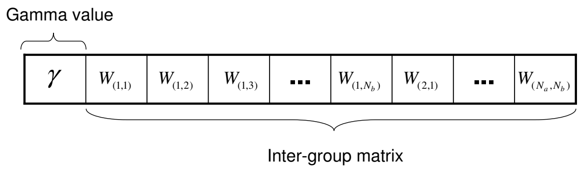

Figure 8: Individuals - Chromosome values.

The objective function used to measure how individuals have performed a convergence to a global pattern was settled at {-10, -5, -2 0}, being -10 the payoff for a complete recovery ($N_r \longrightarrow$ Number of networks), -5 and -2 for a partial recovery ($N_r - 1 \longrightarrow$ Number of networks minus 1 and $N_r - 2 \longrightarrow$ Number of networks minus 2, respectively) and 0 for no recovery.

The fitness function used to transform the objective function value into a measure of relative fitness was developed through a linear ranking method. The selective pressure was chosen equal 2 and individuals were assigned a fitness value according to their rank in the population rather than their raw performance. This fitness function suggests that by limiting the reproductive range, no individual generates an excessive number of offspring in order to prevent premature convergence [5].

The next phase, a number of individuals is chosen for reproduction, such individuals will determine the number of offspring a population will produce. The selection method used was the Stochastic Universal Sampling (SUS) [6] with a generation gap of 0.7 (70%).

Once the individuals to be reproduced are chosen, a recombination operation takes place. The type of crossover developed in this paper was intermediate recombination, considering that we are using real-valued encoding of the chromosome structure. Intermediate recombination is a method of producing new phenotypes around and between the values of the parents phenotypes [26]. Offspring are produced according to the rule

$$O_1 = P_1 + \alpha(P_2 - P_1), \quad (7)$$

where α is a scaling factor chosen uniformly at random over some interval, typically [-0.25, 1.25] and P_1 and P_2 are the parent chromosomes [26]. Each variable in the offspring is the result of the combination of the variables in the parents' genes according to the above expression with a new α chosen for each pair of parent genes.

Now as in natural evolution, it is necessary to establish a mutation process [12]. For real-valued populations, mutation processes are achieved by either perturbing the gene values or by doing a random selection of new values within the allowed range [38, 19]. A real-valued mutation was carried out at a mutation rate of $1/N_{var}$, where N_{var} is the number of variables of each individual.

Given the fact that by means of recombination, the new population becomes smaller than the original one by 30% (generation gap of 70%), it becomes necessary, in order to maintain the size of the original population, we decided to reinsert 90% of the new individuals into the old population in order to replace its least fitted members.

The system was initialized randomly at time $k = 0$ in one of the networks, and in one of its first-level memories which compose a second level memory. The other networks, in their turn, were initialized in one of the 4096 possible combination of patterns, also at random. In the experiment, we considered a density of coupling amongst the inter-group neurons of 100%. The GA was run in 5 trials being the algorithm terminated after a number of 100 generations. After all, the quality of the best members of the population was tested against the definition of the problem.

In the first experiment a typical value of β was chosen ($\beta = 0.2878$) and we measured the number of times that a system consisting of three coupled networks converged to a configuration of triplets. The rate of memory recovery in our experiments were averaged over 5 trials of 1000 iterations of the algorithm proposed in section 2 for each population.

The capacity of convergence of the global system can be seen in Figures 9 and 10 which show that our model presented a maximum mean rate of memory recovery higher than 90% for LI vectors and a rate of nearly 100% for orthogonal vectors (Table 3 - 3 coupled networks). The upper and lower limit, which represent the mean curve of the maximum and minimum convergence in all trials were close to the mean score of the system. Fig 11 and 12 depict the standard deviation of the population whilst Figures 13 and 14 show the evolution of the mean error of the system. The highest score achieved was 97.3% and 92.2% for orthogonal and LI vectors, respectively (Table 3). Moreover, the system developed through genetic algorithm presented a slightly better recovery performance than it did in Hebbian analysis when 3 networks were coupled and LI vectors were used (Tables 1 and 3).

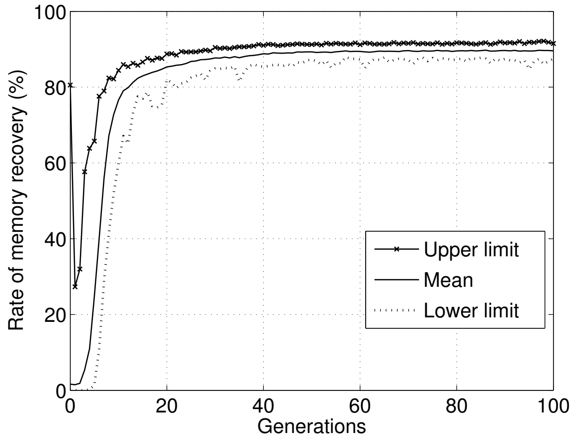

Figure 9: Score of triplets in the population as a function of the number of generations averaged across all 5 trials for LI vectors.

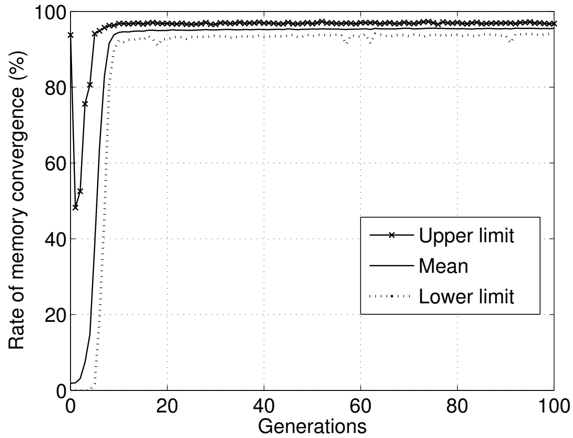

Figure 10: Score of triplets in the population as a function of the number of generations averaged across all 5 trials for orthogonal vectors.

In the second experiment, we analyzed the capacity of convergence into a global pattern in systems when three, four and five networks are coupled. Three patterns of each network (first-level memories) were chosen at random to be the second-level memories as shown in the example of the subsection 3.1.

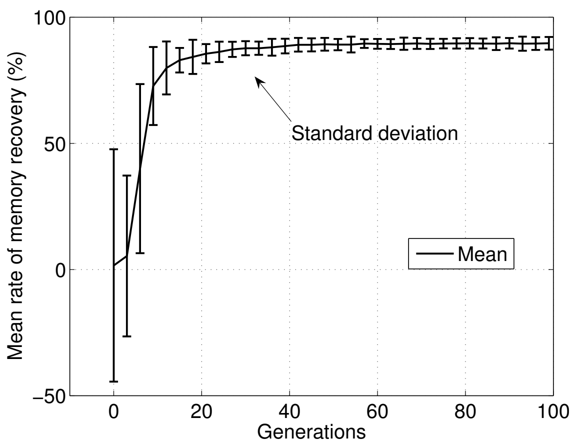

Figure 11: Mean and standard deviation of the triplets in the population as a function of the number of generations averaged across all 5 trials for LI vectors.

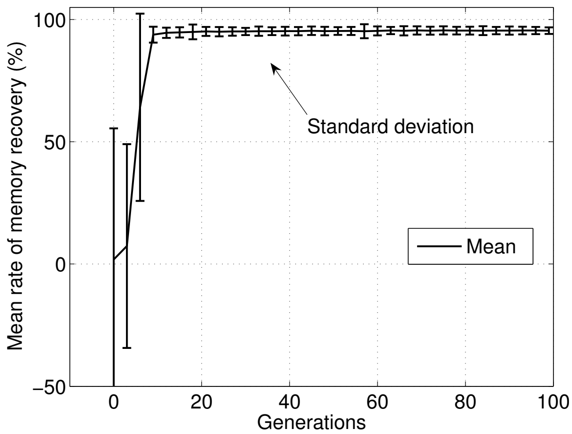

Figure 12: Mean and standard deviation of the triplets in the population as a function of the number of generations averaged across all 5 trials for orthogonal vectors.

Table 3: Maximum rate of memory recovery and gamma values for orthogonal and LI vectors considering 3, 4 and 5 coupled networks

	3		**4**		**5**	
	ORT	LI	**ORT**	**LI**	**ORT**	**LI**
CONV. (%)	97.3	92.2	91.4	83.9	85.18	70.9
gamma	1.42	1.55	1.53	1.55	1.64	1.55

A comparison between all these different couplings can be seen in Figures 15 and 16. It can be observed that the memory recovery into a global pattern decreases when more networks are coupled. Comparing the results depicted in the tables 1 and 3 it is possible to infer

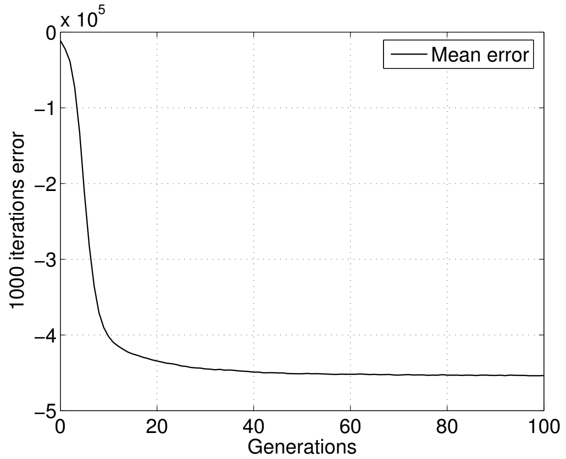

Figure 13: Error Evolution as function of the number of generations for LI vectors.

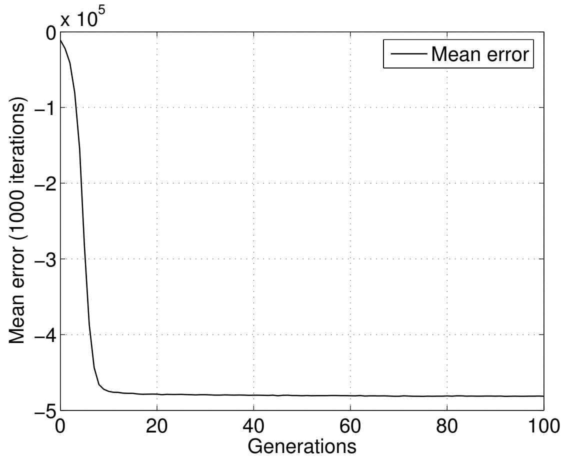

Figure 14: Error Evolution as function of the number of generations for orthogonal vectors.

that the system did not show considerable discrepancies amongst the methods (GA and Hebbian). However, the system recovery rate for LI vectors proves to be better when GA is used. Likewise as seen in Hebbian analysis (subsection 3.1), the system presented a better performance regarding its capacity of memory recovery when orthogonal vectors were used.

Finally, repeating the last experiments carried out in section 3.1, considering 3 coupled networks we will choose from 1 to 6 of these first-level memories to compose our second level-memories. Therefore we will have up to 6 different sets of triplets or global memories. In Figures 17 and 18 we plot the recovery capacity of the system to the chosen global patterns (Table 4). It can be noticed that the system loses its capacity of recovering when a larger set of triplets are chosen to per-

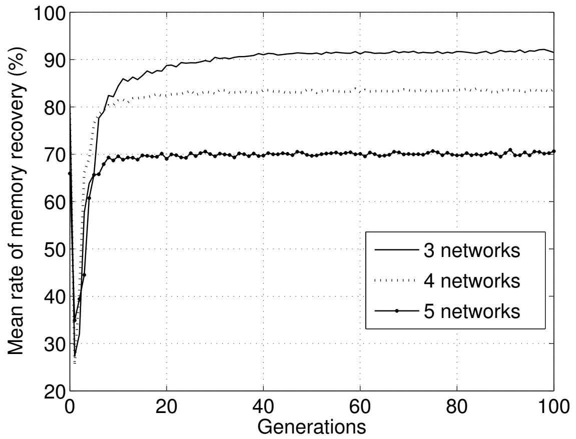

Figure 15: Mean score of memory recovery for 3 to 5 coupled networks - LI vectors.

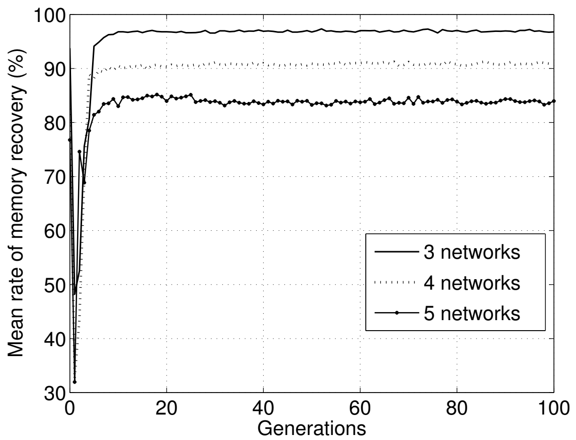

Figure 16: Mean score of memory recovery for 3 to 5 coupled networks - Orthogonal vectors.

form a second-level memory, however as can be seen in Table 5 the system, performing GA method, presents a better performance mainly when the number of patterns increase for both LI and orthogonal vectors. Besides that, despite a decrease in recovery capacity for all cases, the difference between LI and orthogonal vectors remains almost level or present a variation of around 12% for genetic algorithm whereas a more intense deterioration in its recovery capacity of global patterns, especially for LI vectors, occurs in the Hebbian learning method. This happens, as exposed in subsection 3.1, due to the term *Cross Talk* or *Interference Term* which appears interfering with the recovery capacity.

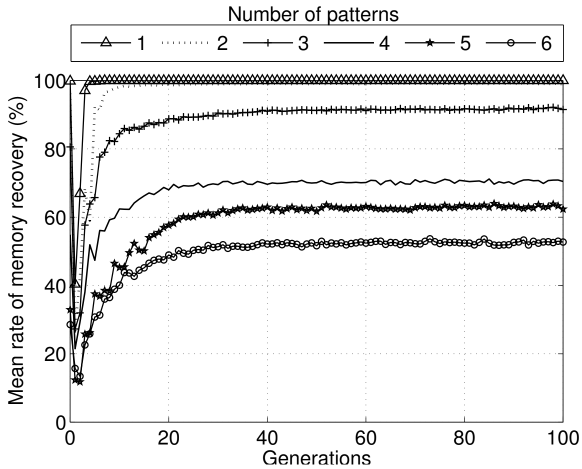

Figure 17: Mean score of triplets in the population as a function of the number of generations averaged across all 5 trials for LI vectors considering from 1 to 6 patterns chosen as first-level memories.

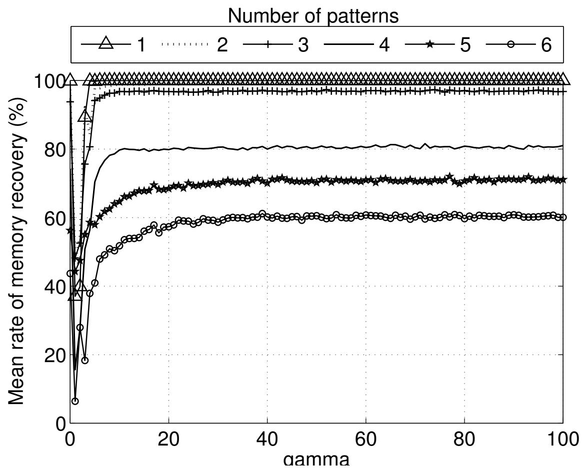

Figure 18: Mean score of triplets in the population as a function of the number of generations averaged across all 5 trials for orthogonal vectors considering from 1 to 6 patterns chosen as first-level memories.

4 Conclusions

comparative information storage and retrieval analysis of a multi-level or hierarchically coupled associative memory based on coupled Generalized-Brain-State-in-a-Box (GBSB) neural networks through evolutionary computation and Hebbian learning

In this paper, we have perfomed a comparative information storage and retrieval analysis of multi-level associative memories using a set of coupled GBSB neural networks as basic building blocks. We performed numerical computations for a two-level memory system by following two strategies, Hebbian and GA analysis.

It was verified that the capacity of convergence to a

Table 4: Maximum rate of memory recovery and gamma values for orthogonal and LI vectors considering from 1 to 6 patterns chosen as first-level memories

Patterns	Type	Conv (%)	gamma
1	ORT	100	1.49
	LI	100	1.43
2	ORT	99.4	1.44
	LI	99.3	1.49
3	ORT	97.3	1.42
	LI	92.16	1.55
4	ORT	81.6	1.49
	LI	71.2	1.42
5	ORT	72.0	1.48
	LI	64.0	1.52
6	ORT	61.2	1.63
	LI	53.7	1.39

Table 5: Maximum rate of comparison of memory recovery between Genetic and Hebbian algorithms for orthogonal and LI vectors considering from 4 to 6 patterns chosen as first-level memories

Algorithms →		Genetic	Hebbian
Patterns	Type	Conv (%)	Conv (%)
4	ORT	81.6	76.6
	LI	71.2	60
5	ORT	72.0	64.5
	LI	64.0	40.3
6	ORT	61.2	53.9
	LI	53.7	38.2

global pattern proved to be significant for both LI and orthogonal vector, even though the percentage of convergence achieved for orthogonal vectors exceeded that of LI vectors in both cases, Hebbian and GA analysis, as expected.

However, when GA was used, our experiments showed that the performance of the system was better than it was when Hebbian learning was applied. The recovery of global patterns were even more evident as the number of first-level memories that compose the repertoire of the second-level memories were increased. In fact, GA performs a compensation, reducing the effect of the *Cross Talk* or *Interference Term* which appears in Hebbian learning, suggesting that in those cases one should be using GA and orthogonal vectors.

Our experiments showed that it is possible to build multi-level memories and that higher levels could present higher performance when built using GA. In particular, we are interested in comparing multi-level memories with two-level memories that have the same number of first-level memories in further works. We expect that the simulations presented in this paper can be used to design further experiments that may lead to a better understanding of the behavior and capacity of hierarchical memory systems.

References

[1] Aleksander, I. What is thought? *NATURE*, 429(6993):701–702, 2004.

[2] Amit, D. J. *Modeling Brain Function*. Cambridge University Press, Cambridge, UK, 1989.

[3] Anderson, J. A. *An introduction to neural network*. MIT Press, Cambridge, Massachusetts, 1995.

[4] Anderson, J. A., Silverstein, J. W., Ritz, S. A., and Jones, R. S. *Distinctive features, categorical perception, probability learning: some applications of a neural model*, chapter 22, pages 283–325. MIT Press, Cambridge, Massachusetts, 1985.

[5] Baker, J. E. Adaptive selection methods for genetic algorithms. In Grefenstette, J. J., editor, *Proceedings of the First International Conference on Genetic Algorithms and Their Applications*. Lawrence Erlbaum Associates, Publishers, 1985.

[6] Baker, J. E. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application*, pages 14–21, Hillsdale, NJ, USA, 1987. L. Erlbaum Associates Inc.

[7] Bar-Yam, Y. Westview Press, 1997.

[8] Clancey, W. J. *Situated cognition : on human knowledge and computer representations*. Learning in doing. Cambridge University Press, Cambridge, U.K., 1997.

[9] Deco, G., Rolls, E. T., and Romo, R. Stochastic dynamics as a principle of brain function. *Progress in Neurobiology*, 88(1):1–16, 2009.

[10] Du, S., Chen, Z., Yuan, Z., and Zhang, X. Sensitivity to noise in bidirectional associative memory (BAM). *IEEE Transactions on Neural Networks*, 16(4):887–898, July 2005.

[11] Edelman, G. M. *Neural darwinism: The theory of neuronal group selection*. Basic Books, New York, 1987.

[12] Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1989.

[13] Gomes, R. M., Braga, A. P., and Borges, H. E. A Model for Hierarchical Associative Memories via Dynamically Coupled GBSB Neural Networks. In *Proceeding of Internacional Conference in Artificial Neural Networks - ICANN 2005*, volume 3696, pages 173–178, Warsaw, Poland, August 2005. Springer-Verlag.

[14] Gomes, R. M., Braga, A. P., and Borges, H. E. Energy Analysis of Hierarchically Coupled Generalized-Brain-State-in-Box GBSB Neural Network. In *Proceeding of V Encontro Nacional de Inteligência Artificial - ENIA 2005*, pages 771–780, São Leopoldo, Brazil, Julho 2005. Sociedade Brasileira de Computação - SBC.

[15] Gomes, R. M., Braga, A. P., and Borges, H. E. Storage capacity of hierarchically coupled associative memories. In Canuto, A. M. P., de Souto, M. C. P., and da Silva, A. C. R., editors, *International Joint Conference 2006, 9th Brazilian Neural Networks Symposium, Ribeirão Preto - SP, Brazil, October 23-27, 2006, Proceedings*, Ribeirão Preto, Brazil, 2006. IEEE.

[16] Gomes, R. M., Braga, A. P., and Borges, H. E. Information storage and retrieval analysis of hierarchically coupled associative memories. *Information Sciences*, 195:175–189, July 2012.

[17] Hopfield, J. J. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Science U.S.A.*, 81:3088–3092, May 1984.

[18] Hui, S. and Zak, S. H. Dynamical analysis of the brain-state-in-a-box (BSB) neural models. *IEEE Transactions on Neural Networks*, 3(5):86–94, 1992.

[19] Janikow, C. Z. and Michalewicz, Z. An experimental comparison of binary and floating point representations in genetic algorithms. In Belew, R. and Booker, L., editors, *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 31–36, San Mateo, CA, 1991. Morgan Kaufman.

[20] Kasabov, N. Evolving Intelligence in Humans and Machines: Integrative Evolving Connectionist Systems Approach. *Computational Intelligence Magazine, IEEE*, 3(3):23–37, august 2008.

[21] Lee, D.-L. and Chuang, T. Designing asymmetric Hopfield-type associative memory with higher order hamming stability. *IEEE Transactions on Neural Networks*, 16(6):1464–1476, November 2005.

[22] Li, J., Michel, A. N., and Porod, W. Analysis and synthesis of a class of neural networks: Variable structure systems with infinite gains. *IEEE Transactions on Circuits and Systems*, 36:713–731, May 1989.

[23] Lillo, W. E., Miller, D. C., Hui, S., and Zak, S. H. Synthesis of Brain-State-in-a-Box (BSB) Based Associative Memories. *IEEE Transactions on Neural Network*, 5(5):730–737, September 1994.

[24] Michel, A. N., Farrell, J. A., and Porod, W. Qualitative analysis of neural networks. *IEEE Transactions on Circuits and Systems*, 36:229–243, 1989.

[25] Muezzinoglu, M., Guzelis, C., and Zurada, J. An energy function-based design method for discrete hopfield associative memory with attractive fixed points. *IEEE Transactions on Neural Networks*, 16(2):307–378, March 2005.

[26] Mühlenbein, H. and Schlierkamp-Voosen, D. Predictive Models for the Breeder Genetic Algorithm: I. Continuous Parameter Optimization. *Evolutionary Computation*, 1(1):25–49, 1993.

[27] O'Kane, D. and Sherrington, D. A Feature Retrieving Attractor Neural Network. *J. Phys. A: Math. Gen.*, 26(21):2333–2342, May 1993.

[28] O'Kane, D. and Treves, A. Short- and Long-Range Connections in Autoassociative Memory. *J. Phys. A: Math. Gen.*, 25(19):5055–5069, October 1992.

[29] Palm, G. On associative memory. *Biological Cybernetics*, 36:19–31, 1980. 10.1007/BF00337019.

[30] Pavloski, R. and Karimi, M. The self-trapping attractor neural network-part II: properties of a sparsely connected model storing multiple memories. *IEEE Transactions on Neural Networks*, 16(6):1427–1439, November 2005.

[31] Personnaz, L., Guyon, I., and Dreyfus, G. Collective Computational Properties of Neural Networks: New Learning Mechanisms. *Physical Review A*, 34:4217–4228, 1986.

[32] Rachkovskij, D. A. and Kussul, E. M. Binding and normalization of binary sparse distributed representations by context-dependent thinning. *Neural Computation*, 12:411–452, 2001. 10.1162/089976601300014592.

[33] Reis, A. G., Acebal, J. L., Gomes, R. M., and Borges, H. E. Space-vector structure based synthesis for hierarchically coupled associative memories. In Canuto, A. M. P., de Souto, M. C. P., and da Silva, A. C. R., editors, *International Joint Conference 2006, 9th Brazilian Neural Networks Symposium, Ribeirão Preto - SP, Brazil, October 23-27, 2006, Proceedings*, Ribeirão Preto, Brazil, 2006. IEEE.

[34] Sussner, P. and Valle, M. E. Gray-Scale Morphological Associative Memories. *IEEE Transactions on Neural Networks*, 17(3):559–570, November 2006.

[35] Sutton, J. P., Beis, J. S., and Trainor, L. E. H. A hierarchical model of neocortical synaptic organization. *Mathl. Comput. Modeling*, 11:346–350, 1988.

[36] TSODYKS, M. V. Hierarchical associative memory in neural networks with low activity level. *Modern Physics Letters B*, 04(04):259–265, 1990.

[37] Westbrook, L. Mental models: a theoretical overview and preliminary study. *J. Information Science*, 32(6):563–579, 2006.

[38] Wright, A. H. Genetic Algorithms for Real Parameter Optimization. In Rawlins, G. J. E., editor, *Proceedings of the First Workshop on Foundations of Genetic Algorithms*, pages 205–220, San Mateo, July 15–18 1991. Morgan Kaufmann.

[39] Zak, S. H., Lillo, W. E., and Hui, S. Learning and Forgetting in Generalized Brain-state-in-a-box (BSB) Neural Associative Memories. *Neural Networks*, 9(5):845–854, 1996.