Biol Cybern (2009) 100:395–407
DOI 10.1007/s00422-009-0310-0

ORIGINAL PAPER

Biological Cybernetics

Reduced order multiport parallel and multidirectional neural associative memories

Abdul Aziz Bhatti

Received: 25 August 2008 / Accepted: 31 March 2009 / Published online: 21 April 2009

Abstract This paper proposes multiport parallel and multidirectional intraconnected associative memories of outer product type with reduced interconnections. Some new reduced order memory architectures such as k-directional and k-port parallel memories are suggested. These architectures are, also, very suitable for implementation of spatio-temporal sequences and multiassociative memories. It is shown that in the proposed memory architectures, a substational reduction in interconnections is achieved if the actual length of original N-bit long vectors is subdivided into k sublengths. Using these sublengths, submemory matrices, T_s or W_s, are computed, which are then intraconnected to form k-port parallel or k-directional memories. The subdivisions of N-bit long vectors into k sublengths save $\frac{(k-1)\times 100}{k}\%$ of interconnections. It is shown, by means of an example, that more than 80% reduction in interconnections is achieved. Minimum limit in bits on k as well as maximum limit on subdivisions in k is determined. The topologies of reduced interconnectivity developed in this paper are symmetric in structure and can be used to scale up to larger systems. The underlying principal of construction, storage and retrieval processes of such associative memories has been analyzed. The effect of complexity of different levels of reduced interconnectivity on the quality of retrieval, signal to noise ratio, and storage capacity has been investigated. The model possesses analogies to biological neural structures and digital parallel port memories commonly used in parallel and multiprocessing systems.

A. A. Bhatti (✉)
School of Science and Technology,
University of Management and Technology,
Lahore, Pakistan
e-mail: drabhatti@gmail.com

Keywords Neural network memories · Parallel memory · Multidirectional memory · Multi associative memory · Reduced interconnections · Retrieval issues

1 Introduction

The neural network based memories, inspired by the human brain, have been the topic of research for many years. The human brain is the underlying biological structure that is responsible for human intelligence. It has complex networks of neurons which collaborate in a highly nonlinear manner to create a massive information processing systems (Teddy et al. 2008). Therefore, there has always been a great interest to mimic brain-like information processing characteristics that are critical to develop such intelligent information processing systems. However, the brain has highly complex, nonlinear, dynamic and evolutionary structure that makes it extremely difficult to model these properties. The cerebellum is one brain region located at the bottom rear of the head in which neuronal connectivity is sufficiently regular to facilitate the learning and associative memory functions and employs error correction signals to drive the network learning and memory tasks (Teddy et al. 2008). However, there are several major architectural limitations are associated with cerebellar models. As human memories are both associative and content-addressable as well as they exhibit massive parallelism, so these characteristics have provided a great impetus to explore the connectionist and neural network models, which, hopefully, will help develop the new methods and models for achieving the massive parallelism, and new ideas for their implementation in software, VLSI or optical hardware, techniques to reduce the level of complexity as well as nonlinearity. In practice to mimic the brain-like characteristics of neural structure that performs the associative

memory task, the neural network based memories of outer product type have been the topic of research for many years (Bhatti 1987; Bhatti 1990, 1992; Daniele et al. 2006; Bhatti and Ouyang 1990; Hirai 1982; Hopfield 1984; Kosko 1988; Simpson 1990; Wang et al. 1994; Wang 2000). The most of research has concentrated on to explore and develop methods for constructing autoassociative and bidirectional associative memories because they possess the properties of massive parallelism (Bhatti 1992; Daniele et al. 2006; Cottrel 1988; Kosko 1988; Simpson 1990; Wang 2000). Hopfield model and a number of its variants are examples of the model of outer product type which have been used for many practical applications including associative memories and optimization of NP-complete problems (Tank and Hopfield 1986).

2 Need for reduced order memories

Over the years, there has been a lot of interest in reducing the dimensionality of input data, while preserving the information content contained therein (Guan et al. 2007; Hopfield 1984; Suykens 2008). The introduction of kernels has proved useful in reducing the dimensionality of data and pruning the support vector machines (Liang et al. 2008). In memory algorithms using scalar product operation, the kernel functions compute the scalar product implicitly in feature space without explicitly using or even knowing the mapping (Muller et al. 2001).

This paper investigates into the outer product based neural associative memories that are completely connected models, and requires $O(N^2)$ interconnections. However, as the long information vectors are split into short ones, and the outer product based memories are constructed separately from each short length, a substantial reduction in interconnections as well as in the size of memory is achieved. The size of the memory matrix (T or W), reduces as the inverse square function of the length, in bits, of information vectors.

The multiport parallel and multidirectional intraconnected memory architectures, proposed in this paper, are very suitable to implement the spatio-temporal sequences (Simpson 1990; Wang 2000), and multiassociative memories in which a number of interassociated sets of different associations are stored and retrieved as information vectors simultaneously in parallel.

The number of neurons in the system is equal to N, the length in bits, of stored vector. For implementation in VLSI or optical hardware of these models, the number of neurons that can be integrated on a chip is limited by the area required for interconnections. However, neural network systems are expected to process information in a manner similar to a biological brain in which a neuron is connected on average to only a small fraction of other neurons. Also, the interest in the economical implementation and efficient use of resources such as communication links further suggests the importance of reduced interconnections or partially connected models, because only these topologies can be expected to scale up to large systems. Hence, there is a significant motivation to investigate the computational abilities, quality of retrieval, convergence time and the storage capacity of the reduced connected models.

3 Development of reduced interconnection memories

In practice, many practical applications use large scale data bases or information and data processing systems that require very large amount of data or information with very long N-bit binary words. Consider such a set of M N-bit long bipolar binary vectors (memory words). For faster retrieval, all N bits must be retrievable simultaneously in parallel. Such a memory is named as an autoassociative or one-port parallel memory, and its construction requires $O(N^2)$ interconnections.

It is shown that a substational reduction in interconnections is achieved if the actual N-bits length of original vectors to be stored, is split into k sublengths, and then using each of these sublengths separately, the submemory matrices, T_s or W_s, are computed. These submemories are arranged as k-port parallel or k-directional memories.

The reduced order two-port, and k-port parallel as well as both parallel and sequential versions of bidirectional and k-directional memory structures are developed, analyzed, and their performance, connectivity, complexity, and information storage capacities are compared with the one-port parallel memory.

All of these memory structures are considered next.

3.1 Autoassociative or one-port parallel memory

One-port parallel memory is a unidirectional, and commonly known as autoassociative memory. In the construction of such a memory, an outer product operation is performed on a set of M N-bit long bipolar binary vectors with itself. The resulting memory matrix with nonzero diagonal elements is shown in Fig. 1.

Such memories are massively parallel in nature, and perform parallel search to retrieve the complete desired vector.

The storage and retrieval algorithms, analysis of performance, storage capacities, complexity, and interconnection requirements have been developed and analyzed in this paper.

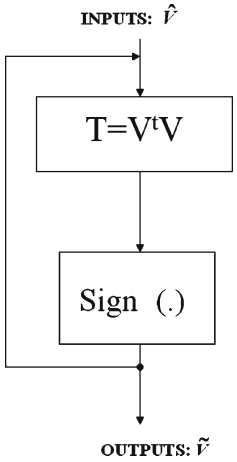

Fig. 1 Autoassociative memory

3.2 Storage algorithm

In the construction of an outer product type neural memory, the bipolar vectors are used. A Unipolar vector X is changed into a bipolar binary vector as

$$V_i = 2X_i - 1 \tag{1}$$

where V_i and X_i are the ith bit of vectors V and X, respectively, and $V_i = 1$ if $X_i = 1$, and $V_i = -1$ if $X_i = 0$.

The memory matrix T for a set of M N-bit long bipolar vectors is constructed as

$$\begin{aligned} T &= \sum_{m=1}^{M} V^{(m)} V^{(m)t} \\ T_{ij} &= \sum_{m=1}^{M} V_i^m V_j^{(m)t} \quad 1 \le (i, j) \le N \end{aligned} \tag{2}$$

This memory matrix T, given in Eq. (2) has non-zero diagonal elements and is superior in performance to that of Hopfield memory matrix T which has zero diagonal elements and is given as

$$T = \left\{ \sum_{m=1}^{M} V^{(m)} V^{(m)t} \right\} - MI \tag{3}$$

However, the memory model given in Eq. (2) is used for demonstration of the underlying principle.

3.3 Retrieval analysis

Let $\hat{V}^{m0}$ is the input probe vector which is closest, in terms of the Hamming distance, to V^{m0} that is one of the stored vectors. The output estimate of the ith bit, $\tilde{V}_i^{m0}$, of the stored vector $m0$-th is given as

$$\hat{V}_i^{m0} = \text{sgn}\left[\sum_{j=1}^{N} T_{ij} \tilde{V}_j^{m0} \right]$$

or

$$\tilde{V}_i^{m0} = \text{sgn}\left[\sum_{j=1}^{N} \left(\sum_{m=1}^{M} V_i^m V_j^m \right) \hat{V}_j^{m0} \right] \tag{4}$$

As the desired output is V_i^{m0}, the Eq. (4) can be written in the form as

$$\begin{aligned} \tilde{V}_i^{m0} = \text{sgn}\Bigg[& \sum_{j=1}^{N} V_j^{m0} V_j^{m0} \hat{V}_i^{m0} + \sum_{\substack{m \neq m0 \\ i=j}}^{M} V_i^m V_i^m \hat{V}_j^{m0} \\ & + \sum_{j \neq i}^{N} \sum_{m \neq m0}^{M} V_i^m V_j^m \hat{V}_j^{m0} \Bigg] \end{aligned} \tag{5}$$

The first term in Eq. (5) is a signal term. The second term modifies the signal by adding or subtracting a value of $(M-1)$ depending on if $\hat{V}_j^{m0}$ is a correct or an incorrect bit (Wang et al. 1990). Now, let "d" is the Hamming distance between the probe vector $\hat{V}^{m0}$ and the corresponding stored vector. V^{m0}. Therefore, Eq. (5) is rewritten as

$$\begin{aligned} \tilde{V}_i^{m0} = \text{sgn}\Bigg[& (N - 2d) V_i^{m0} + (M - 1) \hat{V}_i^{m0} \\ & + \sum_{i \neq j}^{N} \sum_{m \neq m0}^{M} V_i^m V_j^m \hat{V}_j^{m0} \Bigg] \end{aligned} \tag{6}$$

The third term in Eq. (6) is noise.

3.4 Signals to noise ratio and capacity analysis

Noting that the components of stored vectors are statistically independent. According to central limit theorem, and for large N and M, the third term consists of a sum of $(N-1)$ $(M-1)$ independently identically distributed (i,i,d) random variables each of which is $+1$ or -1 with equal probability of $1/2$ (Papoulis et al. 2002). This can be approximated by a Gaussian distribution with mean zero and variance, σ^2, which is given as $\sigma^2 = (N-1)(M-1)$

Then signal to noise ratio, SNR, is given as

$$\text{SNR} = \frac{S}{\sigma} = \frac{(N - 2d) \pm (M - 1)}{\sqrt{(N-1)(M-1)}}$$

And for $N, M >> 1$,

$$\text{SNR} = \frac{(N - 2d) \pm M}{\sqrt{NM}} \tag{7}$$

Now if $\hat{V}_i^{m0} = V_i^{m0}$, then ith bit in the ith bit position of a probe vector corresponds to ith diagonal elements T_{ii} in the memory matrix T, and it will increase the signal component. If $\hat{V}_i^{m0} \neq V_i^{m0}$, then it will decrease the signal component (Wang et al. 1990).

However, for a correct retrieval of a desired vector, the number of correctly known bits in a probe vector must be greater than the incorrect number of bits. Hence, a considerable improvement is possible for a non-zero trace of the T matrix.

With the Hamming distance $d = 0$, the SNR from Eq. (7) may be approximated as

$$\text{SNR} = \frac{(N - 2d) + M}{\sqrt{NM}} \simeq \sqrt{\frac{N}{M}} \tag{8}$$

If the original N-bit length is split into k equal sublengths, the SNR_k due to one sublength is given by Eq. (8) as

$$\text{SNR}_k = \sqrt{\frac{N/k}{M}} \tag{9}$$

Equation (9) indicates that as the original length of N-bit is reduced by a factor of k, the signal to noise ratio, SNR_k, for kth sublength decreases by a factor of $\sqrt{k}$. Therefore, the long information vectors can only be split into several short ones, if the original SNR is high and after the split it must be at least well above one for each sublength.

Next, the analysis of storage capacity of M binary vectors as a function of their length, N bits, is carried out using the same approach as given in McEliece (1987).

Assume that the correct bit positions in a store memory vector are represented as $+1$ and the incorrect ones as -1. As a result, an error free memory vector consists of all $+1$s i.e. (1 1 ... 1 1 ... 1 1 1). The noise Z_m is assumed to be the sum of $(M - 1)$ i.i.d. random variables. The normalized noise is given as

$$Z_m = \frac{(M-1)N}{\sqrt{(M-1)N}} = \sqrt{(M-1)\,N}$$

If N is fixed and M is large, the noise, Z_m, approaches a standard normal random variable. With $d = 0$, the probability that the ith bit is in error (or is negative) is approximately given as

$$\phi\left(\frac{N}{\sqrt{(M-1)N}}\right) \simeq \phi\left(-\sqrt{\frac{N}{M}}\right) = Q\left(\sqrt{\frac{N}{M}}\right)$$

where $Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty \mathrm{e}^{-t^2/2} \mathrm{d}t$

Following the approach used in McEliece (1987), the capacity of M binary vectors as a function of their N-bit length is given as

$$M = \frac{N}{2 \log N} \tag{10}$$

All logs are natural logs.

Now let the original N-bit length of M memory vectors is split into k smaller sublengths $N_1, N_2, \ldots, N_k$ of N/k bits each.

As a result, there are k sets each consisting of M memory subvectors of N/k bits length.

Now substituting sublength N/k bits in Eq. (10), yields

$$M_i = \frac{(N/k)}{2\,(\log\,(N/k))} = \frac{(N/k)}{2\left[\log N - \log k\right]} \tag{11}$$

Next, the storage and retrieval of information is demonstrated by means of two numerical examples, by constructing completely connected and reduced interconnection autoassociative memories.

4 Comparison and analysis of interconnections and retrieval process

Here, the analysis of retrieval process, and the interconnection requirements are presented and compared with the reduced order memories.

4.1 Analysis of completely connected memory

In order to demonstrate the storage and retrieval of information, as well as interconnection requirements in a completely connected autoassociative memory, consider a set of $M = 4$ unipolar vectors $[A\ B\ C\ D]$, each 20 bit long ($N = 20$) given as (Bhatti and Ouyang 1990; Bhatti 1990, 1992; Daniele et al. 2006; Farhat 1986; Hopfield 1984)

$$\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \tag{12}$$

Using Eq. (1), the vector set $[A\ B\ C\ D]$ are converted into bipolar from, and using the storage prescription given by Eq. (2), the corresponding memory matrix T is computed as

$$T = \begin{bmatrix}
4 & -2 & 0 & -2 & -2 & 2 & 0 & 2 & 0 & 4 & 0 & 4 & 0 & 0 & -2 & 2 & 2 & 0 & -4 & 0 \\
-2 & 4 & -2 & 0 & 0 & 0 & -2 & 0 & 2 & -2 & -2 & -2 & 2 & -2 & 4 & 0 & 0 & 2 & 2 & -2 \\
0 & -2 & 4 & 2 & -2 & 2 & 4 & -2 & 0 & 0 & 4 & 0 & -4 & 0 & -2 & -2 & 2 & 0 & 0 & 0 \\
-2 & 0 & 2 & 4 & 0 & 0 & 2 & 0 & -2 & -2 & 2 & -2 & -2 & 2 & 0 & -4 & 0 & -2 & 2 & 2 \\
-2 & 0 & -2 & 0 & 4 & -4 & -2 & 0 & -2 & -2 & -2 & -2 & 2 & 2 & 0 & 0 & -4 & -2 & 2 & 2 \\
2 & 0 & 2 & 0 & -4 & 4 & 2 & 0 & 2 & 2 & 2 & 2 & -2 & -2 & 0 & 0 & 4 & 2 & -2 & -2 \\
0 & -2 & 4 & 2 & -2 & 2 & 4 & -2 & 0 & 0 & 4 & 0 & -4 & 0 & -2 & -2 & 2 & 0 & 0 & 0 \\
2 & 0 & -2 & 0 & 0 & 0 & -2 & 4 & -2 & -2 & 2 & 2 & 2 & 2 & 0 & 0 & 0 & -2 & -2 & 2 \\
0 & 2 & 0 & -2 & -2 & 2 & 0 & -2 & 4 & 0 & 0 & 0 & 0 & -4 & 2 & 2 & 2 & 4 & 0 & -4 \\
4 & -2 & 0 & -2 & -2 & 2 & 0 & 2 & 0 & 4 & 0 & 4 & 0 & 0 & -2 & 2 & 2 & 0 & -4 & 0 \\
0 & -2 & 4 & 2 & -2 & 2 & 4 & -2 & 0 & 0 & 4 & 0 & -4 & 0 & -2 & -2 & 2 & 0 & 0 & 0 \\
4 & -2 & 0 & -2 & -2 & 2 & 0 & 2 & 0 & 4 & 0 & 4 & 0 & 0 & -2 & 2 & 2 & 0 & -4 & 0 \\
0 & 2 & -4 & -2 & 2 & 2 & -4 & 2 & 0 & 0 & -4 & 0 & 4 & 0 & 2 & 2 & -2 & 0 & 0 & 0 \\
0 & -2 & 0 & 2 & 2 & -2 & 0 & 2 & -4 & 0 & 0 & 0 & 0 & 4 & -2 & -2 & -2 & -4 & 0 & 4 \\
-2 & 4 & -2 & 0 & 0 & 0 & -2 & 0 & 2 & -2 & -2 & -2 & 2 & -2 & 4 & 0 & 0 & 2 & 2 & -2 \\
2 & 0 & -2 & -4 & 0 & 0 & -2 & 0 & 2 & 2 & -2 & 2 & 2 & -2 & 0 & 4 & 0 & 2 & -2 & -2 \\
2 & 0 & 2 & 0 & -4 & 4 & 2 & 0 & 2 & 2 & 2 & 2 & -2 & -2 & 0 & 0 & 4 & 2 & -2 & -2 \\
0 & 2 & 0 & -2 & -2 & 2 & 0 & -2 & 4 & 0 & 0 & 0 & 0 & -4 & 2 & 2 & 2 & 4 & 0 & -4 \\
-4 & 2 & 0 & 2 & 2 & -2 & 0 & -2 & 0 & -4 & 0 & -4 & 0 & 0 & 2 & -2 & -2 & 0 & 4 & 0 \\
0 & -2 & 0 & 2 & 2 & -2 & 0 & 2 & -4 & 0 & 0 & 0 & 0 & 4 & -2 & -2 & -2 & -4 & 0 & 4
\end{bmatrix} \tag{13}$$

Consider a probe vector $\hat{V}^{m0}$ given as

$$\hat{V}^{m0} = [0\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$$

This $\hat{V}^{m0}$ probe vector is closest, in terms of Hamming distance to V^{m0} which is one of the stored vectors. The initiation of retrieval process starts when a probe vector $\hat{V}^{m0}$ is applied as input to the memory matrix T.

Estimate of the total input activation, $\tilde{V}_i^{m0}$, to the ith neuron is given as

$$\tilde{V}_i^{m0} = \sum_{j=1}^{N} T_{ij} \hat{V}_j^{m0} \tag{14}$$

The next state of $\tilde{V}_i^{m0}$ is given as

$$\tilde{V}_i^{m0}(t + \Delta t) = \left\{ \begin{matrix} 0 & \text{if } \sum_{j=1}^{N} T_{ij} \hat{V}_j^{m0} < 0 \\ 1, & \text{otherwise} \end{matrix} \right\} \tag{15}$$

The interactive process terminates when the estimate of the current state is equal to the previous state, and is given as

$$\tilde{V}_i^{m0}(t + \Delta t) = \tilde{V}_i^{m0}(t) \tag{16}$$

The iterative retrieval process using the threshold condition from Eq. (15), gives the result after first iteration as

1st iteration

$$[-2\ -4\ 6\ 4\ -8\ 4\ 6\ -4\ 2\ -2\ 10\ -2\ -10\ -2\ 0\ -8\ 8\ 2\ 2\ -2]$$

After thresholding

$$[0\ 0\ 1\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ 0]$$

2nd iteration

$$[-4\ 4\ 12\ 4\ -16\ 12\ 12\ -12\ 8\ -4\ 12\ -4\ -16\ -12\ 0\ -8\ 12\ 8\ 0\ -12]$$

After thresholding

$$[0\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ 0]$$

Stable state is reached as

$$[0\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ 0]$$

This is one of the stored vectors.

4.1.1 Interconnection requirements

Note that the memory matrix T is a fully connected matrix. It is 20 by 20, has 400 number of interconnections, and the strength of each interconnection ranges between $-M$ and M, where M is number of stored vectors.

4.2 Analysis of reduced interconnection autoassociative memories

Consider the same set $[A\ B\ C\ D]$ of $M = 4$ unipolar vectors, each 20 bit long given in Eq. (12). The bipolar version of those vectors is given as

$$\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \left[\begin{array}{cccccccccc|cccccccccc} 1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 \end{array}\right] \tag{17}$$

4.2.1 Construction of two-port parallel memory

The original length of $M = 4$, 20-bit long vectors given by Eq. (17), is split into two parts. The resulting two sets consist of $M = 4$ vectors and the length of vectors in each set is now 10 bits long. These are given as

$$\begin{bmatrix} A_1 \\ B_1 \\ C_1 \\ D_1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 \end{bmatrix} \tag{18}$$

and

$$\begin{bmatrix} A_2 \\ B_2 \\ C_2 \\ D_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 \end{bmatrix} \tag{19}$$

The two-port parallel memory is constructed by performing an outer product operation on a set of subvectors with itself.

The reduced interconnection autoassociative memories computed from each set are given as

$$T_1 = [A_1 B_1 C_1 D_1] \begin{bmatrix} A_1 \\ B_1 \\ C_1 \\ D_1 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & -2 & 0 & -2 & -2 & 2 & 0 & 2 & 0 & 4 \\ -2 & 4 & -2 & 0 & 0 & 0 & -2 & 0 & 2 & -2 \\ 0 & -2 & 4 & 2 & -2 & 2 & 4 & -2 & 0 & 0 \\ -2 & 0 & 2 & 4 & 0 & 0 & 2 & 0 & -2 & -2 \\ -2 & 0 & -2 & 0 & 4 & -4 & -2 & 0 & -2 & -2 \\ 2 & 0 & 2 & 0 & -4 & 4 & 2 & 0 & 2 & 2 \\ 0 & -2 & 4 & 2 & -2 & 2 & 4 & -2 & 0 & 0 \\ 2 & 0 & -2 & 0 & 0 & 0 & -2 & 4 & -2 & -2 \\ 0 & 2 & 0 & -2 & -2 & 2 & 0 & -2 & 4 & 0 \\ 4 & -2 & 0 & -2 & -2 & 2 & 0 & 2 & 0 & 4 \end{bmatrix} \tag{20}$$

and

$$T_2 = [A_2 B_2 C_2 D_2] \begin{bmatrix} A_2 \\ B_2 \\ C_2 \\ D_2 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 0 & -4 & 0 & -2 & -2 & 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & -2 & 2 & 2 & 0 & -4 & 0 \\ -4 & 0 & 4 & 0 & 2 & 2 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & -2 & -2 & -2 & -4 & 0 & 4 \\ -2 & -2 & 2 & -2 & 4 & 0 & 0 & 2 & 2 & -2 \\ -2 & 2 & 2 & -2 & 0 & 4 & 0 & 2 & -2 & -2 \\ 2 & 2 & -2 & -2 & 0 & 0 & 4 & 2 & -2 & -2 \\ 0 & 0 & 0 & -4 & 2 & 2 & 2 & 4 & 0 & -4 \\ 0 & -4 & 0 & 0 & 2 & -2 & -2 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 & -2 & -2 & -2 & -4 & 0 & 4 \end{bmatrix} \tag{21}$$

4.2.2 Parallel retrieval of a desired output vector

When in a two-port parallel memory, two probe vectors are applied simultaneously in parallel to each of the two one-port parallel memories of a two-port parallel memory, the desired vectors one from each submemory, are retrieved together in parallel, and then these two retrieved vectors are concatenated together to form the desired 20 bits long output vector.

Using threshold condition given in Eq. (15), and the following probe vector, $\hat{V}^{D1}$ with memory matrix T_1, the retrieval process is given as

$$\hat{V}^{D1} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

The result after first iteration is

$$\begin{bmatrix} -2 & 2 & 12 & 6 & -10 & 8 & 10 & -6 & 6 & -2 \end{bmatrix}$$

After thresholding

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

This is the stable state, and is one of the stored vectors D_1.

Next, using the following probe vector $\hat{V}^{D2}$, with the memory matrix T_2, the retrieval process is given as

$$\hat{V}^{D2} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

After first iteration, the result is given as

$$\begin{bmatrix} 4 & -4 & -4 & -8 & 6 & -2 & 6 & 8 & 4 & -8 \end{bmatrix}$$

After thresholding

$$\tilde{V}^{D2} = [1 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 0]$$

The is the final stable state and is given as

$$[1 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 0]$$

The two retrieved vector $\tilde{V}^{D1}$ and $\tilde{V}^{D2}$ are concatenated together to form the original 20-bit long desired vector which is one of the stored vectors, that is vector number 4 in the set.

The concatenation of these vectors is

$$\tilde{V}^{D1} + \tilde{V}^{D2} = \tilde{V}^{D} \tag{22}$$

$\tilde{V}^{D} = V^{D}$, and is given as

$$V^{D} = [0 \; 1 \; 1 \; 1 \; 0 \; 1 \; 1 \; 0 \; 1 \; 0 \; 1 \; 0 \; 0 \; 0 \; 1 \; 0 \; 1 \; 1 \; 1 \; 0]$$

4.2.3 Analysis of interconnection requirements

This is a 20-bit long one of the original stored vectors in the memory matrix T given in Eq. (13).

Note that the number interconnections of T_1 and T_2 are = $10 \times 10 + 10 \times 10 = 200$, as compared with 400 interconnections of the completely connected matrix. Clearly, the actual length of 20 bits of the original stored vector is divided into two parts, and therefore, 50% reduction in interconnections is achieved.

5 Construction of k-port parallel memories

Consider a set of M N-bit long binary vectors (memory words). For faster retrieval, these binary vectors should be stored in an autoassociative or one-port parallel memory that requires $O(N^2)$ interconnections. To use less memory or reduce the number of interconnections, the original N-bit length of M vectors is subdivided into k smaller sublengths $N_1, N_2 \ldots N_k$ bits long. As a result, there are k sets each consisting of M vectors, and the sublengths of vectors in each set are $N_1, N_2, \ldots N_k$ bits. Using these smaller sublengths, the memory matrices are computed by performing an outer product operation on any of the kth set of subvectors with itself. All of these k submemory matrices are arranged in the same order in which the k groups of subvectors will be concatenated together to form the original set of stored vectors.

These multiport parallel memories can be used as multiassociative memory in which interassociated sets of multiassociations can be retrieved simultaneously in parallel.

The k probe vectors are applied simultaneously in parallel to each of the k one-port parallel memories of a k-port parallel memory, the desired vectors one from each submemory, are retrieved together in parallel, and are concatenated together to form the desired N-bit long output vector.

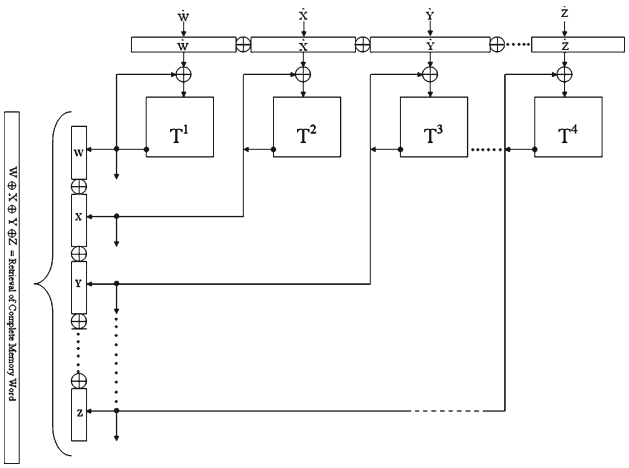

Fig. 2 Reduced order autoassociative memories as a k-port parallel memory

However, all of these k memory matrices can be concatenated together in an orderly manner, and collectively function like a k-port parallel memory matrix which uses far less interconnections i.e. $(\frac{k-1}{k}) \times 100\%$ collectively, and is shown in Fig. 2.

Such memory models are useful in large scale databases, warehousing, data mining systems, parallel and multiprocessing systems, fault tolerant and real time computing systems, and many other practical applications.

6 Development of parallel k-directional memories

The bidirectional and intraconnected memories (Daniele et al. 2006;Cottrel 1988; Farhat 1986; Kosko 1988; Simpson 1990; Wang et al. 1993; Wang 2000) commonly reported in the literature are simple variants of reduced order k-directional memories which require far less interconnections, and are considered in this section. Its functionality is demonstrated by means of an example.

6.1 Construction of reduced order parallel bidirectional memories

Let the N bits length of a set of M stored vectors is divided into two parts, and the resulting two sets, say X and Y, each consisting of M vectors of lengths say N_1, and N_2 bits long are obtained. A reduced order bidirectional memory is constructed by taking the outer product of these two subsets X and Y. It is shown in Fig. 3.

This proposed parallel BAM operates in parallel and when two probe vectors are applied simultaneously in parallel to each of the two parts of the parallel bidirectional memories, the desired vectors one from each submemory, are retrieved together in parallel, and then these two retrieved vectors are

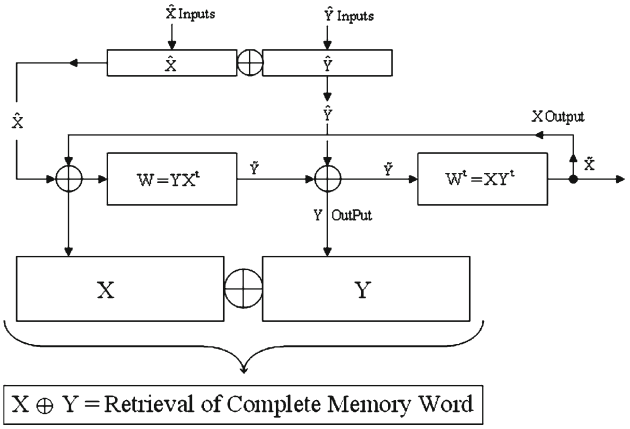

Fig. 3 Reduced order parallel bidirectional memory

concatenated together to form the desired N bits long output vector.

It has improved storage capacity and superior performance to that of sequential BAM commonly reported in the literature (Bhatti 1992; Daniele et al. 2006; Kosko 1988; Simpson 1990; Wang et al. 1993; Wang 2000).

6.2 Analysis of interconnection requirements

The number of interconnections or the memory matrix size is $N_1 N_2$. If $N_1 = N_2 = N/2$, the memory size is $N^2/4$. The total number of interconnections required for bidirectional retrieval is equal to $(N^2/4) + (N^2/4) = N^2/2$. As a result, the number of interconnections reduces to one half or 50%.

The storage and retrieval of information in a reduced order bidirectional memory is demonstrated by means of a numerical example.

6.3 BAM storage algorithm

The bidirectional associative memory (BAM) is an outer product type neural memory, and computed by using bipolar binary vectors.

Consider two sets X and Y, each consisting of M bipolar vectors, and the length of vectors in set X is N_1 bits long, and in set Y is N_2 bits long. The BAM memory matrix, W, is written as

$$W = \sum_{m=1}^{M} Y^{(m)} X^{(m)t} \tag{23}$$

$$W^t = \sum_{m=1}^{M} X^{(m)} Y^{(m)t} \tag{24}$$

$$W_{ij} = \sum_{m=1}^{M} Y_i^m X_j^{(m)t}, \quad i = 1, \ldots, N_2, \quad j = 1, \ldots, N_1 \tag{25}$$

where

$$X^{(m)} = \left(X_1^{(m)}, \ldots, X_j^{(m)}, \ldots, X_{N_1}^{(m)}\right)$$

and

$$Y^{(m)} = \left(Y_1^{(m)}, \ldots, Y_i^{(m)}, \ldots, Y_{N_2}^{(m)}\right)$$

6.4 Retrieval analysis

Let $\hat{X}^{m0}$ is an imperfect input probe vector which is closest, in terms of Hamming distance, to X^{m0} that forms an association pair with Y^{m0}. The output estimate of the ith bit, $\tilde{Y}_i^{m0}$, is given as

$$\tilde{Y}_i^{m0} = \text{Sgn}\left[\sum_{j=1}^{N_1} W_{ij} \hat{X}_j^{m0}\right]$$

or

$$\tilde{Y}_i^{m0} = \text{Sgn}\left[\sum_{j=1}^{N_1}\left[\sum_{m=1}^{M} Y_i^m X_j^m\right] \hat{X}_j^{m0}\right] \tag{26}$$

As the desired output bit is, Y_i^{m0}, Eq. (26), can be written as

$$\tilde{Y}_i^{m0} = \text{Sgn}\left[y_i^{m0} \sum_{j=1}^{N_1} X_j^{m0} \hat{X}_j^{m0} + \sum_{j=1}^{N_1} \sum_{m \neq m0}^{M} Y_i^m X_j^m \hat{X}_j^{m0}\right] \tag{27}$$

Similarly, the estimate of the ith bit $\tilde{X}_i^{m0}$, when $\hat{Y}^{m0}$, an imperfect probe vector, which is closest, in terms of Hamming distance, to Y^{m0} which is one of the stored vectors, is given as

$$\begin{aligned}\tilde{X}_i &= \text{Sgn}\left[\sum_{j=1}^{N_2} W_{ij} \hat{Y}_j^{m0}\right] \\ &= \text{Sgn}\left[\sum_{j=1}^{N_2}\left[\sum_{m=1}^{M} X_i^m Y_j^m\right] \hat{Y}_j^{m0}\right]\end{aligned} \tag{28}$$

As the desired output bit is X_i^{m0}, the ith bit of the desired stored vector X^{mo}, Eq. (28), can be written as

$$\tilde{X}_i^{m0} = \text{Sgn}\left[X_i^{m0} \sum_{j=1}^{N_2} Y_j^{m0} \hat{Y}_j^{m0} + \sum_{j=1}^{N_2} \sum_{m \neq m0}^{M} X_i^m Y_j^m \hat{Y}_j^{m0}\right] \tag{29}$$

Now, consider Eq. (27), the first term is the signal term, and the second term is a noise.

Let d_x is the Hamming distance between the probe vectors $\hat{X}^{m0}$ and the corresponding stored vector X^{m0}. Therefore,

Eq. (27) can be written in the form as

$$\tilde{Y}_i^{m0} = \mathrm{Sgn}\left[(N_1 - 2d_x)Y_i^{m0} + \sum_{j=1}^{N_1}\sum_{m\neq m0}^{M} Y_i^m x_j^m \hat{x}_j^{m0}\right] \quad (30)$$

Likewise, the estimate of ith bit $\tilde{X}_i^{mo}$ of the stored vector x^{m0}, can be written as

$$\tilde{X}_i^{m0} = \mathrm{Sng}\left[(N_2 - 2d_y)x_i^{m0} + \sum_{j=1}^{N_2}\sum_{m\neq m0}^{M} x_i^m Y_j^m \hat{Y}_j^{m0}\right] \quad (31)$$

where d_y the Hamming distance between the probe vectors $\hat{Y}^{m0}$ and the corresponding stored vector Y^{m0}.

The initiation of the retrieval process starts when a probe vector $\hat{X}^{m0}$ is applied as input to the memory matrix W. the total input estimate, $\hat{Y}_i^{m0}$, the ith bit of Y^{m0} is given as

$$\tilde{Y}_i^{m0} = \sum_{j=1}^{N} W_{ij}\hat{X}_j^{m0} \quad (32)$$

The next state of $\tilde{Y}_i^{m0}$ is given as

$$\tilde{X}_i^{m0} = \sum_{j=1}^{P} W_{ij}\tilde{Y}_j^{m0} \quad (33)$$

$$\tilde{Y}_i^{m0}(t+\Delta t) = \left\{\begin{array}{ll} 0 & \text{if } \sum_{j=1}^{N_1} w_{ij}\hat{X}_j^{m0} < 0 \\ 1, & \text{otherwise}\end{array}\right\} \quad (34)$$

Similarly,

$$\tilde{X}_i^{m0}(t+\Delta t) = \left\{\begin{array}{ll} 0 & \text{if } \sum_{j=1}^{N_2} w_{ij}\hat{Y}_j^{mo} < 0 \\ 1, & \text{otherwise}\end{array}\right\} \quad (35)$$

The interactive process terminates when the estimate of the current state is equal to the previous state

$$\tilde{X}_i^{m0}(t+\Delta t) = \tilde{X}_i^{m0}(t) \quad (36)$$

6.5 Signals to noise ratio and capacity analysis of a parallel BAM

Now, considering Eq. (30), the first term is a signal and the 2nd term is a noise.

Next, note that the components of stored vectors are statistically independent. According to central limit theorem (Papoulis et al. 2002), and for large N and M, the second term consists of a sum of $N_1(M-1)$ independently and identically distributed (iid) random variables each of which is +1 or −1 with equal probability of 1/2, and can be approximated by a Gaussian distribution with mean zero and variance, σ^2, which is given as

$$\sigma^2 = N_1(M-1) \quad (37)$$

The signal to noise ratio, SNR, is given as

$$\mathrm{SNR} = \frac{S}{\sigma} = \frac{(N_1 - 2d_x)}{\sqrt{N_1(M-1)}}$$

And for $N_1, M >> 1$,

$$\mathrm{SNR} \simeq \frac{(N_1 - 2d_x)}{\sqrt{N_1 M}} \simeq \frac{N_1}{\sqrt{N_1 M}} = \sqrt{\frac{N_1}{M}} \quad (38)$$

Similarly, from Eq. (31), one can get

$$\mathrm{SNR} = \frac{S}{\sigma} = \frac{(N_2 - 2d_y)}{\sqrt{N_2(M-1)}}$$

And for large $N_2, M >> 1$, then

$$\mathrm{SNR} = \frac{(N_2 - 2d_y)}{\sqrt{N_2 M}} \simeq \frac{N_2}{\sqrt{N_2 M}} \simeq \sqrt{\frac{N_2}{M}} \quad (39)$$

In sequential BAMs, X input produces Y as output and Y input produces X as output, but not both together. Therefore, the signal to noise ratio SNR, in such memories (Daniele et al. 2006; Cottrel 1988; Kosko 1988; Simpson 1990; Wang et al. 1994, 1993) is given by Eqs. (38) and (39).

Since the proposed BAM operates in parallel, and requires either $\hat{X}_i$ or $\hat{Y}_i$ or both $(\hat{X}_i, \hat{Y}_i)$ probe vectors to initiate the retrieval process. These two probe vectors $\hat{X}_i$ and $\hat{Y}_i$ used in a sequential BAM, forming a pair $(\hat{X}_i, \hat{Y}_i)$ are concatenated together to form a compound probe vector as $\hat{V}_i = \hat{X}_i \oplus \hat{Y}_i$, and are given as input. Both X_i and Y_i memory vectors are produced simultaneously in parallel as output.

Note that even though both memory vectors X_i and Y_i are retrieved in parallel, yet the internal structure of the parallel BAM is obtained by intraconnecting the two sequential BAMs W and W^t. Therefore, signal to noise ratio, SNR, of the intraconnected parallel BAM is equivalent to the sequential BAM, commonly proposed in the literature (Daniele et al. 2006; Cottrel 1988; Shen and Cruz 2005; Kosko 1988; Simpson 1990; Wang et al. 1994; Wang 2000), and is given by Eqs. (38) and (39).

Let $N_1 = N_2 = N_k$. After substitution of N_k for N_1 and N_2, the SNR_k is given as

$$\mathrm{SNR}_k = \frac{\sqrt{N_k}}{\sqrt{M}} \quad (40)$$

Note that by splitting the N-bit long vectors into $k = 2$, the SNR is reduced by a factor of $\sqrt{2}$, and therefore, the performance of the sequential BAMs, commonly proposed (Daniele et al. 2006; Cottrel 1988; Shen and Cruz 2005; Kosko 1988; Simpson 1990; Wang et al. 1994; Wang 2000), fall short of Hopfield's condition. Consequently, for improved performance, the continuity restriction is commonly proposed. This continuity condition can also be applied to improve the performance of reduced ordered memories considered in this paper. The continuity restriction implies that the Hemming distance should be uniformly distributed along

the length of the memory vectors being stored. However, there are some encoding strategies which can maximize the noise tolerance (Shen and Cruz 2005), and therefore, improve the storage capacity.

6.6 Determination of limits on k

6.6.1 Minimum limit in bits on k

In neural associative memories a neuron has two states, on or off, (1 or -1), at least two bits are essential to distinguish between two states. Also, it can be postulated that $\frac{N}{2}+1$ are the minimum number of correct bits required in a probe vector to retrieve an information vector from the memory. Therefore, the minimum number of bits, N_m, required to retrieve an entity can be computed as

$$N_m = \left\lceil \frac{N}{2} \right\rceil + 1 \tag{41}$$

where $\lceil \bullet \rceil$ is a ceiling function.

For $N = 1$ or 2, the minimum number of bits $N_m = 2$.

For reliable operation, the signal to noise ratio, $\gamma = \text{SNR}$, must be much grater than 1.

From Eq. (8), the numbers of memory vectors M that can be stored and retrieved are given as

$$\sqrt{M} < \left\lfloor \frac{\sqrt{N}}{\gamma} \right\rfloor \tag{42}$$

where $\lfloor \bullet \rfloor$ is a floor function.

For $N_m = 2$ and $\gamma > 1$, Eq. (42) gives $M = 1$.

Therefore, with $M = 1$, the minimum length in bits of a subdivision for any k is two or greater.

However, error correction capability for any associative memory is vital. For correction of one error, the minimum Hamming distance among vectors must be at least 3 or greater, but $N_m = 2$ does not have any error correction ability.

6.6.2 Limit on maximum subdivisions in k

As before, for reliable operation, the signal to noise ratio, $\text{SNR} = \gamma$, must be much greater than 1. Using Eq. (9), for N bits length of M vectors, the maximum subdivisions in k are given as

$$k \leq \left\lfloor \frac{N}{\gamma^2 M} \right\rfloor \tag{43}$$

6.6.3 A note on SNR

Hopfield (1984) suggested the value of SNR about 2.5 so that the vectors to be stored should be approximately orthogonal. Lee and Chuang (2005) and Hwang and Hsiao (2003) used SNR about 1.5 for $M = 4$, and Kosko (1988) used as low as 1.3 for $M = 3$. However, our simulation results suggest the value of SNR about 1.4 to 1.7 should be used.

7 Example of reduced interconnection bidirectional memory

Consider the same set of $M = 4$ binary vectors, each $N = 20$ bits long that is given in Eq. (12). The bipolar version of these vectors is given in Eq. (17). These vectors are subdivided into two parts, and are given as

$$\begin{bmatrix} A_1 \\ B_1 \\ C_1 \\ D_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 \end{bmatrix} \tag{44}$$

and

$$\begin{bmatrix} A_2 \\ B_2 \\ C_2 \\ D_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 \end{bmatrix} \tag{45}$$

The bidirectional memory W is constructed as

$$[W_1] = [A_1 B_1 C_1 D_1] \begin{bmatrix} A_2 \\ B_2 \\ C_2 \\ D_2 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 4 & 0 & 0 & -2 & 2 & 2 & 0 & -4 & 0 \\ -2 & -2 & 2 & -2 & 4 & 0 & 0 & 2 & 2 & -2 \\ 4 & 0 & -4 & 0 & -2 & -2 & 2 & 0 & 0 & 0 \\ 2 & -2 & -2 & 2 & 0 & -4 & 0 & -2 & 2 & 2 \\ -2 & -2 & 2 & 2 & 0 & 0 & -4 & -2 & 2 & 2 \\ 2 & 2 & -2 & -2 & 0 & 0 & 4 & 2 & -2 & -2 \\ 4 & 0 & -4 & 0 & -2 & -2 & 2 & 0 & 0 & 0 \\ -2 & 2 & 2 & 2 & 0 & 0 & 0 & -2 & -2 & 2 \\ 0 & 0 & 0 & -4 & 2 & 2 & 2 & 4 & 0 & -4 \\ 0 & 4 & 0 & 0 & -2 & 2 & 2 & 0 & -4 & 0 \end{bmatrix} \tag{46}$$

$$[W_2] = [W_1]^t$$

The retrieval process is a cyclic process and is given as

$$W_1 \hat{X} \rightarrow \tilde{Y}$$
$$W_1^t \tilde{Y} \rightarrow \tilde{X}$$
$$W_1 \tilde{X} \rightarrow Y$$
$$W_1^t Y \rightarrow X$$

The process continues until the stable states are reached.

This method may be named as intraconnected bidirectional memories.

In this example, A_1 and A_2 were used as probe vectors and the correct results were obtained. However, when the original 20 bit long vectors were divided into 2 sets of 12 bits and 8 bits long vectors, there was one bit in error. Therefore, further research work is needed to insure the proper bounds.

7.1 Analysis of interconnection requirements

For k number of subparts of the original length of N bits, the number of interconnections for consecutive pairs can be computed as

$$\text{INTS} = \sum_{\substack{i\neq j \\ i=1}}^{k} N_i N_j, \quad 1 \leq (i, j) \leq k \tag{47}$$

The k subparts constitute, in consecutive order, k number of intraconnected or concatenated bidirectional memories termed as k-directional associative memory.

Saving in interconnections can be computed as

$$\text{Saving in INTS} = N^2 - \left(\sum_{\substack{i\neq j \\ i=1}}^{k} N_i N_j \right) \tag{48}$$

$$= \left(1 - \frac{1}{N^2} \left(\sum_{\substack{i=1 \\ i\neq j}}^{k} N_i N_j \right) \right) \times 100\% \tag{49}$$

Following example demonstrate the amount of saving in the number of interconnections.

7.1.1 Example: saving in interconnections

Consider a k-directional memory with $k = 5$. Let a set of $M = 12$ bipolar vectors and the length, N in bits, of each of these vectors is 100 bits. Now let this $N = 100$ bits is divided into five, $k = 5$, unequal subparts as $N_1 = 25$ bits, $N_2 = 15$ bits, $N_3 = 20$ bits, $N_4 = 30$ bits, and $N_5 = 10$ bits long, respectively. Interconnection requirements for $k = 5$ directional memories:

$$\text{INTS} = (N_1N_2 + N_2N_3 + N_3N_4 + N_4N_5 + N_5N_1)$$

Saving in interconnections = N^2 − INTS = 8175 interconnections = 81.75% reduction in interconnections.

If $k = 2$, the original length of N bits of M vectors in a set is divided into two parts and these two parts are used to construct a bidirectional memory. However, this can also be used as two-directional parallel associative memory that can retrieve the complete vector, and at the same time results in 50% savings in interconnections. Therefore, the bidirectional or intraconnected memories (Bhatti 1992; Hwang and Hsiao 2003; Kosko 1988; Simpson 1990) are merely a class of reduced order memories of outer product type.

8 Construction of k-directional memories

Like k-port parallel memories, the construction of k-directional memories requires to split the original N-bit length of M vectors into k smaller sublengths $N_1, N_2, \ldots, N_k$ bits long. As a result, there are k sets each consisting of M vectors, and the sublengths of vectors in each set are $N_1, N_2, \ldots N_k$ bits.

In the construction of a k-directional memory, the outer product operation is performed, in an ordered sequence, on pairwise consecutive sublengths N_i and N_j bits long, where $i \neq j$, and $j = i + 1$, if $i = k$ then $j = 1$, and if $i = k + 1$ then $i = 1$, for all $1 \leq (i, j) \leq k$.

When in a k-directional memory, k probe vectors are applied simultaneously in parallel to each of the k concatenated submemories of a k-directional memory, the desired vectors one from each submemory, are retrieved simultaneously in parallel, and then these k retrieved vectors are concatenated together to form the desired N bits long output vector.

Clearly, like k-port parallel memories, the k-directional memory can be used to implement the spatio-temporal sequences Wang (2000), and multiassociations that can be retrieved simultaneously in parallel, and this is one of the important aspects the way human memory seems to work. Next, the reduced connection k-directional memory is constructed for $K = 4$.

The original N bits length of M vectors is divided into $K = 4$ smaller sublengths N_1, N_2, N_3 and N_4 bits long, then there are $k = 4$ sets, say $[WXYZ]$ sets, each consisting of M vectors, and the sublengths of vectors in each of these four sets $[WXYZ]$ are N_1, N_2, N_3 and N_4, respectively. The outer product operation is performed in an ordered sequence on pairwise consecutive sets $W\ X\ Y\ Z\ W$ in that order. The resulting memory matrix is a k-directional, here $k = 4$ and it consists of four intraconnected bidirectional memories which are given as

$$\overset{W}{\rightarrow} WX \overset{X}{\rightarrow} XY \overset{Y}{\rightarrow} YZ \overset{Z}{\rightarrow} ZW \overset{W}{\rightarrow} WX \rightarrow \cdots$$

It is shown in Fig. 4.

8.1 Analysis of interconnection requirements

Next, the original length of N bits is divided into k equal parts, the length of each of k sets of vectors is N/k, and the interconnections for consecutive pairs are N^2/k^2. This

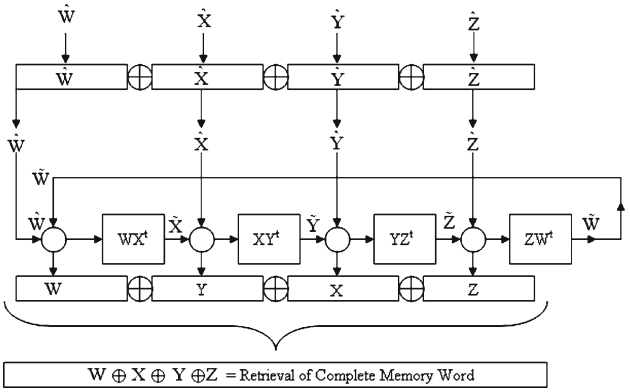

Fig. 4 k-Directional associative memory for $k = 4$

constitutes a k-directional memory and is shown in Fig. 4. The total number of interconnections needed for k subparts is given as INTS $= k(N^2/k^2) = N^2/k$.

Clearly, when the original length of N bits is divided into k equal parts, the required number of interconnections is inversely proportional to the number of subparts and is equal to $1/k$, and the reduction in interconnections is given as

$$\text{Reduction in interconnect} = \left(\frac{k-1}{k}\right) N^2$$

9 Retrieval constraints on k-port parallel and k-directional memories

When M number of memory vectors are stored, the retrieval of a desired memory vector will be disturbed by the noise introduced by other $M-1$ memory vectors. Therefore, the memory vectors too close, in terms of Hamming distance, to each other will introduce greater amount of noise that makes the retrieval difficult. Consequently, for better quality of retrieval, the memory vectors to be stored should be approximately orthogonal, and the Hamming distance should be uniformly distributed along the length of memory vectors.

9.1 Hopfield's constraint

Hopfield, through extensive simulation studies, found that for the binary vectors to be pseudo orthogonal, and the length of binary vectors should be about six to seven times the number of vectors, M, to be stored (Hopfield 1984). However, when the N-bit length of M vectors is divided into k smaller sublengths of $N_1, N_2, \ldots, N_k$ bits long, and using these smaller sublengths, the memory matrices are computed, the reduced length of vectors should preferably satisfy the Hopfield's condition.

In this respect most of the bidirectional memories, reported in the literature (Lee and Chuang 2005; Hwang and Hsiao 2003; Kosko 1988), do not satisfy Hopfield's condition, and the additional restrictions, such as continuity condition, are imposed which restrict their applications.

9.2 Orthogonality and continuity constraints

In constructing the reduced order memories the length of vectors is subdivided into smaller segments, but the number of stored vectors remains unchanged. As a result, the signal to noise ratio, SNR, and the level of orthogonality or resolution among them reduces. In k-port parallel, bidirectional and k-directional memories, the length of stored vectors, generally, falls short of Hopfield's condition.

Therefore, for reduced order memories, the fundamental constraint for improved performance is that the Hamming distance must be uniformly distributed along the lengths of vectors in all of k-sets.

In k-port parallel memories there is no intrafield connectivity present, and all k submemories operate independent of each other. Therefore, for kth order collective retrieval, the Hamming distance must be uniformly distributed along the lengths of all of k sets of M subvectors for $i = 1, \ldots, M$, for every k. This constraint is termed as continuity requirement and is given as

$$\frac{1}{N_p}\left(H\left(X_i^p, X_j^p\right)\right) = \frac{1}{N_q}\left(H\left(X_i^q, X_j^q\right)\right) \tag{50}$$

for $i, j = 1, 2, \ldots, M$, and $M < N$; and $1 \le (p, q) \le k$.

For equal sublengths $N_p - N_q$, and for orthogonal vectors the continuity condition reduces to

$$H\left(X_i^p, X_j^p\right) = H\left(X_i^q, X_j^q\right) = 0 \tag{51}$$

For retrieval of a vector in each of these k memory matrices, the retrieval key should have at least $(N_i + 2)/2$ correct elements in the ith probe vector for $1 \le i \le k$. Each of the k memory matrices is independent of each other, and for retrieval requires its own probe vector.

For the case of a bidirectional memory, consider the two sets X and Y. The X_i and Y_i for $i = 1, 2, \ldots, M$, forming an associated pair of vectors to be stored in a BAM. This requirement is termed as continuity constraint and for a sequential BAM is given as (Bhatti 1992; Hopfield 1984; Hwang and Hsiao 2003; Kosko 1988; Simpson 1990; Wang et al. 1994, 1993),

$$\frac{1}{N}H(X_i, X_j) \cong \frac{1}{P}H(Y_i, Y_j), \quad \text{for } i, j = 1, 2, \ldots M, \tag{52}$$

Where N and P are the dimensions of X_i and Y_i, respectively, and $H(,)$ denotes the Hamming distance.

10 Conclusions

In this paper, the importance of reducing the number of interconnections has been investigated. It is shown that subdividing a set of N-bit long binary vectors into k sublengths saves $\frac{(k-1)}{k} \times 100\%$ of interconnections. The construction of reduced interconnection autoassociative and k-port parallel, both sequential and parallel versions of bidirectional, and k-directional memories have been demonstrated by means of two numerical examples. Some new memory architectures are suggested for reduced interconnection memory structures. Minimum limit in bits on k as well as maximum limit on subdivisions in k is determined. The underlying principal of construction, storage, retrieval processes and operation of an autoassociative memory, k-port parallel, bidirectional, and k-directional memories have been described. The effect of complexity of different levels of reduced interconnectivity on the quality of retrieval, signal to noise ratio, and storage capacity has been investigated.

The multiport parallel and multidirectional memory architectures, proposed in this paper, are very suitable to implement the spatio-temporal sequences (Simpson 1990; Wang 2000), and multiassociative memories in which a number of interassociated sets of different associations are stored and retrieved as information vectors simultaneously in parallel.

Acknowledgements The skillful efforts of Mr. Raza Nasir for typing the entire manuscript and for constructing figures in this paper are appreciated. The author would like to thank the reviewers for their valuable contributions that have improved the quality of this paper.

References

Bhatti AA (1987) State of the art research, development and applications of neural networks. D905-106065-1 High Technology Center, the Boeing Electronics Company, Bellevue, Washington

Bhatti AA (1990) Analysis of performance issues in neural network based associative memory models. IJCNN, San Diego, California, June

Bhatti AA (1992) Analysis of the effectiveness of using unipolar and bipolar binaries, and Hamming distance in neural network computing. J Opt Eng 31(9):1972–1975

Bhatti AA, Ouyang YC (1990) Thresholding, hamming distance, unipolar/bipolar binaries, and retrieval in neural network based memories. IEEE 22nd SSST Conference Proceedings, Cookeville, Tennessee, March 11–13

Cottrel M (1988) Stability and attractivity in associative memory networks. Biol Cybern 58:129–139

Daniele C, Costantini G, Perfetti R, Ricci E (2006) Associative memory design using support vector machines. IEEE Trans Neural Netw 17(5):1165–1174

Farhat N (1986) Optoelectronics builds viable neural-net memory. Electronics, pp 41–44

Guan S-U, Chunyn B, TseNgee N (2007) Reduced pattern training based on task decomposition using pattern distributor. IEEE Trans Neural Netw 18(6):1738–1749

Hirai Y (1982) A learning network resolving multiple match in associative memory. In: Proceedings of 6th international conference on pattern recognition

Hopfield JJ (1984) Neurons with graded response, have computational properties like those of two state neurons. Proc Natl Acad Sci 81:3080

Hwang J-D, Hsiao F-H (2003) Stability analysis of neural-network interconnected systems. IEEE Trans Neural Netw 14(1):201–208

Kosko B (1988) Bi-directional associative memories. IEEE Trans SMC 18:49–60

Lee D-L, Chuang TC (2005) Designing asymmetric Hopfield-type associative memory with higher order Hamming stability. IEEE Trans Neural Netw 16(6):1464–1476

Liang X, Chen R-C, Guo X (2008) Pruning support vector machines without altering performances. IEEE Trans Neural Netw 19(10): 1793–1803

McEliece RJ, Posner EC, Rodemich ER, Venkatesh SS (1987) The capacity of the hopfield associative memory. IEEE Trans Inform Theory 33(4):461–482

Muller K-R, Mike S, Ratsch G, Tsuda K, Scholkopf B (2001) An introduction to kernel-based learning algorithms. IEEE Trans Neural Netw 12(2):181–201

Papoulis A, Pillai SU (2002) Probability, random variables and stochastic processes, 4th edn. McGraw Hill, New York

Simpson PK (1990) Higher-ordered and interconnected bidirectional associate memories. IEEE Trans SMS 20(3):637–653

Suykens JAK (2008) Data visualization and dimensionality reduction using kernel maps with reference point. IEEE Trans Neural Netw 19(9):1501–1517

Shen D, Cruz JB Jr (2005) Encoding strategy for maximum noise tolerance bidirectional associative memory. IEEE Trans Neural Netw 16(2):293–300

Tank DW, Hopfield J (1986) Simple neural optimization network an A/D converter, signal decision circuit, and a linear programming circuit. IEEE Trans Circuits Syst 33(5):533–541

Teddy SD, Quek C, Lai EM-K (2008) PSECMAC: a novel self-organizing multiresolution associative memory architecture. IEEE Trans Neural Netw 19(4):689–712

Wang L (2000) Heteroassociatve of spatio-temporal sequences with the bidirectional associative memory. IEEE Trans Neural Netw 11(6):1503–1505

Wang J-H, Krile TF, Walkup JF (1990) Determination of the Hopfield associative memory characteristics using a single parameter. Neural Netw 3(3):319–331

Wang T, Xinhua Z, Xiaoliang X (1994) Designing bidirectional associative memories with optimal stability. IEEE Trans Syst Man Cybern 24(5):778–790

Wang T, Xinhua Z, Xiaoliang X, Xipeng X (1993) A neuron-weighted learning algorithm and its hardware implementation in associative memories. IEEE Trans Comput 42(5):636–640