Stability and Statistical Properties of Second-Order Bidirectional Associative Memory

Chi-Sing Leung, Lai-Wan Chan, *Member, IEEE,* and Edmund Man Kit Lai

***Abstract*— In this paper, a bidirectional associative memory (BAM) model with second-order connections, namely second-order bidirectional associative memory (SOBAM), is first reviewed. The stability and statistical properties of the SOBAM are then examined. We use an example to illustrate that the stability of the SOBAM is not guaranteed. For this result, we cannot use the conventional energy approach to estimate its memory capacity. Thus, we develop the statistical dynamics of the SOBAM. Given that a small number of errors appear in the initial input, the dynamics shows how the number of errors varies during recall. We use the dynamics to estimate the memory capacity, the attraction basin, and the number of errors in the retrieved items. Extension of the results to higher-order bidirectional associative memories is also discussed.**

***Index Terms*—Associative memory, BAM, neural network, stability.**

I. INTRODUCTION

ASSOCIATIVE memories [1], [2] have been intensively studied in the past decade. An important feature of associative memories is the ability to recall the stored items from partial or noisy inputs. One form of associative memories is the bivalent additive bidirectional associative memory (BAM) [3]. It is a two-layer heteroassociator that stores a prescribed set of vector pairs. We will refer to these pairs as *pattern pairs*. A BAM network is very similar to a Hopfield network but has two layers of neurons in which layer F_X has n neurons and layer F_Y has p neurons. The recall process of the BAM is an iterative one starting with a stimulus pair in F_X and F_Y. After a number of iterations, the patterns in F_X and F_Y converge to a fixed point which is desired to be one of the pattern pairs.

The BAM has three important features [3]. First, it performs both heteroassociative and autoassociative data recalls: the final state in F_X represents the autoassociative recall, while the final state in F_Y represents the heteroassociative recall. Second, the initial input can be presented in any one of the two layers. Last, the BAM is stable during recall.

To encode the pattern pairs, Kosko used the outer-product rule [3]. However, with the outer-product rule the memory capacity is very small if the pattern pairs are not orthogonal. Several modifications have been proposed to improve the memory capacity. These modifications fall into two categories: 1) modifying the encoding methods [4]–[6] and 2) introducing second-order connections to form the second-order bidirectional associative memory (SOBAM) [7]–[9]. The memory capacity of the SOBAM has been empirically studied [7] but the theoretical memory capacity has not yet been derived. The SOBAM has also been proven to be stable during recall [7].

This paper describes the stability and statistical properties of the SOBAM. Contrary to Simpson's works [7], we demonstrate that the stability of the SOBAM is not guaranteed during recall. We also point out a mistake in [7]. This mistake has led to the wrong conclusion that the stability of the SOBAM is guaranteed. Hence, we cannot use the energy approach [10] to estimate the statistical properties of the SOBAM, especially the memory capacity and the attraction basin. In this paper, we are interested in knowing whether each pattern pair can attract all the initial inputs within a certain distance from it. If so, we can obtain the attraction basin. Another important performance index is memory capacity, i.e., the maximum number of pattern pairs that can be stored in the SOBAM as attractors. Also of interest is the number of errors in the retrieved pairs. The question now is: given any $\rho_x^{(0)}n$ errors in the initial input (an arbitrary error pattern with $\rho_x^{(0)}n$ errors in the initial input), how does the number of errors vary during recall? To answer this question, we develop the statistical dynamics of the SOBAM. From this dynamics, the number of errors in the retrieved pairs, the attraction basin, and the memory capacity can be estimated.

Section II reviews the SOBAM and discusses its stability. The statistical dynamics of the SOBAM is developed in Section III, using the theory of large deviation [10]. Section IV discusses the way to estimate the memory capacity, the attraction basin, and the number of errors in the retrieved items. Numerical examples are given in Section V. Section VI shows how the results can be generalized to higher order bidirectional associative memories (HOBAM's).

II. SOBAM AND STABILITY

There are m pattern pairs $\{(X_1, Y_1), \cdots, (X_m, Y_m)\}$, where $X_h = (x_{1h}, \cdots, x_{nh})^T$ and $Y_h = (y_{1h}, \cdots, y_{ph})^T$. The components of X_h and Y_h are bipolar ($+1$ or -1). The SOBAM encodes the pattern pairs in two matrices. The first matrix, U, is a $n \times n \times p$ lattice that holds the second-order connections from F_X to F_Y. The second matrix, V, is a $p \times p \times n$ lattice that holds the second-order connections from F_Y to F_X. The matrix $U = [u_{kji}]$ is given by

$$u_{kji} = \sum_{h=1}^{m} y_{kh}\, x_{jh}\, x_{ih} \quad \forall j \& i = 1, \cdots, n, \text{ and } k = 1, \cdots, p. \tag{1}$$

Manuscript received September 19, 1994; revised May 31 1995, May 19, 1996, and October 18, 1996.

C.-S. Leung and L.-W. Chan are with Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

E. M. K. Lai is with Department of Computer and Communication Engineering, Edith Cowan University, Joondalup Campus, Perth, Western Australia.

Publisher Item Identifier S 1045-9227(97)01748-7.

The matrix $V = [v_{jkl}]$ is given by

$$v_{jkl} = \sum_{h=1}^{m} x_{jh}\, y_{kh}\, y_{lh} \quad \forall l \& k = 1, \cdots, p, \text{ and } j = 1, \cdots, n. \tag{2}$$

The F_X state at time t is denoted as $X^{(t)} = (x_1^{(t)}, \cdots, x_n^{(t)})^T$. The F_Y state at time t is denoted as $Y^{(t)} = (y_1^{(t)}, \cdots, y_p^{(t)})^T$. The recall process is

$$y_k^{(t+1)} = \text{sgn}\left(\sum_{j=1}^{n}\sum_{i=1}^{n} u_{kji}\, x_j^{(t)}\, x_i^{(t)}\right) \tag{3}$$

for $k = 1, \cdots, p$, where

$$\text{sgn}(x) = \begin{cases} +1 & x > 0 \\ -1 & x < 0 \\ \text{state unchanged} & x = 0. \end{cases}$$

Similarly

$$x_j^{(t+1)} = \text{sgn}\left(\sum_{k=1}^{p}\sum_{l=1}^{p} v_{jkl}\, y_k^{(t+1)}\, y_l^{(t+1)}\right) \tag{4}$$

for $j = 1, \cdots, n$. *Equations (3) and (4) imply that the initial input $X^{(0)}$ recalls $Y^{(1)}$; $Y^{(1)}$ recalls $X^{(1)}$ and so on.*

The SOBAM is a finite-state autonomous system whose state converges to either a stable state or a limit cycle. Unlike the original BAM, the stability of the SOBAM is not guaranteed during recall. To illustrate this, we use a SOBAM network to store the following pattern pairs:

$$\begin{aligned}
X_1 &= (-1, 1, -1, 1, 1, -1)^T \\
Y_1 &= (-1, 1, 1, -1, 1, -1)^T \\
X_2 &= (1, 1, -1, -1, 1, -1)^T \\
Y_2 &= (-1, -1, 1, -1, -1, -1)^T \\
X_3 &= (-1, -1, -1, -1, 1, -1)^T \\
Y_3 &= (-1, -1, 1, -1, -1, 1)^T \\
X_4 &= (-1, 1, -1, -1, 1, -1)^T \\
Y_4 &= (1, -1, -1, -1, 1, -1)^T \\
X_5 &= (1, 1, -1, 1, 1, 1)^T \\
Y_5 &= (1, 1, -1, 1, -1, -1)^T.
\end{aligned}$$

With the initial state $X^{(0)} = (-1, -1, 1, -1, 1, 1)^T$ and $Y^{(0)} = (1, 1, -1, 1, 1, 1)^T$, the following states can be obtained, shown in the equation at the bottom of the page.

Clearly, the network converges to a limit cycle. Thus, the stability of the SOBAM is not guaranteed.

Simpson [7] used an energy function to explain the stability of the SOBAM. The energy function is expressed as

$$\begin{aligned}
E_2 &= (E_{2Y} + E_{2X}) \\
&= -\sum_{h=1}^{m}(X_h^T X)^2 (Y_h^T Y) - \sum_{h=1}^{m}(Y_h^T Y)^2 (X_h^T X)
\end{aligned} \tag{5}$$

where (X, Y) are the current states, $E_{2Y} = -\sum_{h=1}^{m}(X_h^T X)^2 (Y_h^T Y)$ represents the F_Y energy, and $E_{2X} = -\sum_{h=1}^{m}(Y_h^T Y)^2 (X_h^T X)$ represents the F_X energy. According to the recall process, either F_X or F_Y is updated first. If F_X is updated first, the change in energy is

$$\Delta E_2 = \Delta E_{2X} = -\sum_{h=1}^{m}(Y_h^T Y)^2 (X_h^T\, \Delta X) \tag{6}$$

where $\Delta X = X^{new} - X$ and X^{new} is the new state in F_X. Conversely, if F_Y is updated first, the change in energy is

$$\Delta E_2 = \Delta E_{2Y} = -\sum_{h=1}^{m}(X_h^T X)^2 (Y_h^T\, \Delta Y) \tag{7}$$

where $\Delta Y = Y^{new} - Y$ and Y^{new} is the new state in F_Y.

Simpson showed that the values of ΔE_{2X} and ΔE_{2Y} are either negative or zero. He then claimed that the SOBAM is always stable [7]. However, from our previous counter example, it can be seen that the stability is not guaranteed. This discrepancy is due to the omission of some terms on the right-hand side of (6) and (7). Actually, if F_X is updated first, the total change in energy is

$$\begin{aligned}
\Delta E_2 = &-\sum_{h=1}^{m}(Y_h^T Y)^2 (X_h^T\, \Delta X) - \sum_{h=1}^{m}(X_h^T\, \Delta X)^2 (Y_h^T Y) \\
&- \sum_{h=1}^{m} 2(X^T \Delta X)(X_h^T\, \Delta X)(Y_h^T Y).
\end{aligned} \tag{8}$$

$$\begin{array}{lcll}
Y^{(1)} & = & \text{sgn}(-8, 0, 8, -8, -8, -8)^T & = (-1, 1, 1, -1, -1, -1)^T \\
X^{(1)} & = & \text{sgn}(-8, 32, -40, -8, 40, -40)^T & = (-1, 1, -1, -1, 1, -1)^T \\
Y^{(2)} & = & \text{sgn}(-12, -52, 12, -84, 20, -52)^T & = (-1, -1, 1, -1, 1, -1)^T \\
X^{(2)} & = & \text{sgn}(8, 48, -56, 8, 56, -24)^T & = (1, 1, -1, 1, 1, -1)^T \\
Y^{(3)} & = & \text{sgn}(-12, 12, 12, -20, -12, -52)^T & = (-1, 1, 1, -1, -1, -1)^T \\
X^{(3)} & = & \text{sgn}(-8, 32, -40, -8, 40, -40)^T & = (-1, 1, -1, -1, 1, -1)^T \\
Y^{(4)} & = & \text{sgn}(-12, -52, 12, -84, 20, -52)^T & = (-1, -1, 1, -1, 1, -1)^T \\
X^{(4)} & = & \text{sgn}(8, 48, -56, 8, 56, -24)^T & = (1, 1, -1, 1, 1, -1)^T \\
Y^{(5)} & = & \text{sgn}(-12, 12, 12, -20, -12, -52)^T & = (-1, 1, 1, -1, -1, -1)^T \\
X^{(5)} & = & \text{sgn}(-8, 32, -40, -8, 40, -40)^T & = (-1, 1, -1, -1, 1, -1)^T \\
Y^{(6)} & = & \text{sgn}(-12, -52, 12, -84, 20, -52)^T & = (-1, -1, 1, -1, 1, -1)^T \\
X^{(6)} & = & \text{sgn}(8, 48, -56, 8, 56, -24)^T & = (1, 1, -1, 1, 1, -1)^T \\
\vdots & \vdots & \vdots & \vdots \quad \ddots
\end{array}$$

In this equation, the first term on the right-hand side is the change in E_{2X} due to a change of the F_X state. The other two terms, which represent the change in E_{2Y} due to a change of the F_X state, are either negative or positive. Hence, we cannot draw any conclusion regarding the stability based on the energy function proposed by Simpson [7].

The above discussion is valid for layer-synchronous recall process in which all neurons in a layer are updated simultaneously. Since layer-synchronous recall process is a special case of asynchronous recall processes whereby the neurons in a layer are updated sequentially, the stability of the SOBAM is not guaranteed under both layer-synchronous and asynchronous recall processes.

III. Statistical Dynamics

A. Notations and Outline

This section outlines how the statistical dynamics is derived. We first define some terminologies and state the assumptions used in the rest of the paper.

- $p = rn$, where r is a positive constant.
- $m = \alpha n^2$, where α is a positive constant.
- The dimensions, n and p, are large. This assumption is often used [10]–[15].
- For analytical purposes, we assume that each component of the pattern pairs is a ± 1 equiprobable independent random variable. Though this assumption is not always being satisfied in most real-life data, it is difficult to analyze associative memories without making such an assumption. In fact, this assumption has been widely used [10]–[15].
- The Hamming distance between two bipolar vectors, X and X', is denoted as $d(X, X')$.
- *Attraction Basin*: It is required that each pattern pair is stored as a stable state (or at least there is a stable state at a small Hamming distance). Otherwise, the pattern pairs cannot be recalled. Besides, we expect a SOBAM network to have the following error correction property. If the network is started at a state $X^{(0)}$ where $d(X_h, X^{(0)}) \leq \rho_x^{(0)} n$, the F_X state will reach a stable state within a distance of $\rho_x^f n$ from the stored pattern X_h after a sequence of state transitions where $\rho_x^{(0)} > \rho_x^f$ (the F_Y state should also reach a stable state within a distance of $\rho_y^f p$ from the stored pattern Y_h where $\rho_x^{(0)} > \rho_y^f$). We are interested in knowing whether each pattern pair is able to attract all the initial inputs $X^{(0)}$ within a distance of $\rho^{(0)} n$ for some positive constants $\rho^{(0)}$. The maximum value of such $\rho^{(0)}$ denotes the attraction basin of each pattern pair. Also of interest is the number of errors in the retrieved items. This number measures the quality of the retrieved items. Since we are considering "all possible initial inputs within a certain distance," the above definition of the attraction basin is for *worst case errors*. In the rest of the paper, the term "attraction basin" refers to the attraction basin for *worst case errors*. Instead of estimating the attraction basin directly, we will estimate the number of errors after each state transition.
- Given that $p = rn$ and $m = \alpha n^2$, P_Y^{**} is the probability that for each pattern pair (X_h, Y_h) and for any error pattern with $\rho_x^{(t)} n$ errors in F_X in the present state $(d(X_h, X^{(t)}) = \rho_x^{(t)} n)$, the number of errors in F_Y in the next state is less than $\rho_y p$ $(d(Y_h, Y^{(t+1)}) < \rho_y p)$. It should be noticed that the phrase "for any error pattern with $\rho_x^{(t)} n$ errors in F_X" in the definition of P_X^{**} reflects the concept of *worst case errors*.
- Given that $p = rn$ and $m = \alpha n^2$, P_X^{**} is the probability that for each pattern pair (X_h, Y_h) and for any error pattern with $\rho_y^{(t+1)} p$ errors in F_Y in the present state $(d(Y_h, Y^{(t+1)}) = \rho_y^{(t+1)} p)$, the number of errors in F_X in the next state is less than $\rho_x n$ $(d(X_h, X^{(t+1)}) < \rho_x n)$.
-
$$\text{The error rate in } F_X = \frac{\text{The number of errors in } F_X}{n}.$$
$$\text{The error rate in } F_Y = \frac{\text{The number of errors in } F_Y}{p}.$$
- To estimate the value of P_Y^{**}, we first introduce the event $EA_{h,g}$. It is the event that
$$d(Y^{(t+1)}, Y_h) < \rho_y p$$
for a given pattern pair (X_h, Y_h) and for a given present state $X^{(t)} \in S_{h,t}$, where
$$S_{h,t} = \left\{ X \in \{+1, -1\}^n \text{ such that } d(X, X_h) = \rho_x^{(t)} n \right\}. \quad (9)$$
The index g refers to a particular error pattern. For a given $\rho_x^{(t)}$, the number of error patterns is $\binom{n}{\rho_x^{(t)} n}$. Thus, the range of g is from one to $\binom{n}{\rho_x^{(t)} n}$. Also, $\overline{EA_{h,g}}$ is the complement event of $EA_{h,g}$. It is the event that
$$d(Y^{(t+1)}, Y_h) \geq \rho_y p$$
for a given pattern pair (X_h, Y_h) and for a given present state $X^{(t)} \in S_{h,t}$.
- In the above, each event $EA_{h,g}$ only refers to an error pattern and a pattern pair. To consider each pattern pair and all the possible error patterns, we need to introduce the event EA which is the intersection of all possible $EA_{h,g}$'s
$$EA = \bigcap_{h,g} EA_{h,g}.$$
It is the event that
$$d(Y^{(t+1)}, Y_h) < \rho_y p$$
for each pattern pair (X_h, Y_h) and for any $X^{(t)} \in S_{h,t}$. Also, $\overline{EA}$ is defined as the complement event of EA
$$\overline{EA} = \bigcup_{h,g} \overline{EA_{h,g}}.$$

From the definitions of EA and P_Y^{**}

$$\begin{aligned} P_Y^{**} &\equiv \text{Prob}(EA) \\ &= 1 - \text{Prob}(\overline{EA}) \\ &= 1 - \text{Prob}(\bigcup_{h,g} \overline{EA_{h,g}}) \\ &\geq 1 - \sum_{h=1}^{m} \sum_{g=1}^{\binom{n}{\rho_x^{(t)} n}} \text{Prob}(\overline{EA_{g,h}}). \end{aligned} \tag{10}$$

In Part B, we will first estimate the values of P_X^{**} and P_Y^{**} (Lemmas 4 and 5). From the two lemmas, an upper bound on the error rate in the next state is obtained (Corollaries 2 and 4). Based on Corollary 2, given that $\rho_x^{(t)}$ is the upper bound on the error rate in F_X at time t, we can derive an upper bound $\rho_y^{(t+1)}$ on the error rate in F_Y at time $(t+1)$. Similarly, from Corollary 4, we can estimate $\rho_x^{(t+1)}$ from $\rho_y^{(t+1)}$. As a result, two sequences $\{\rho_x^{(t)}\}$ and $\{\rho_y^{(t)}\}$ are constructed to represent the statistical dynamics of the SOBAM. In Section IV, we will discuss how to use the features of these two sequences to estimate the memory capacity, the number of errors in the retrieved pairs, and the attraction basin.

B. Construction of the Dynamics

To estimate the values of P_Y^{**} and P_X^{**}, we make use of Stirling's formula and the theory of large deviation [10]. Here, we restate them as the following two lemmas.

Lemma 1—Stirling's Asymptotic Formula for Factorial: Let n be a large integer and $\delta \in (0, 0.5)$. Then

$$\binom{n}{\delta n} \sim \exp\{n\hbar(\delta)\}$$

where

$$\hbar(\delta) = -\delta \log \delta - (1-\delta)\log(1-\delta).$$

$\diamondsuit$

Lemma 2—Newman's Lemma: Suppose $\chi_{1,N}, \chi_{2,N}, \cdots$ are, for each N, independent, identically distributed, and symmetric random variables satisfying.

1)
$$\lim_{N\to\infty} \text{Var}(\chi_{1,N}) = \sigma^2 \in (0, \infty). \tag{11}$$

2) For some real $L > 2$ and $t_o > 0$,
$$\limsup_{N\to\infty} \left\{ E\left[\exp(t_o \mid \chi_{1,N} \mid^{2/L})\right]\right\} < \infty \tag{12}$$

where $E[\cdot]$ is the expectation operator.

For any $\gamma \in (0, \infty)$ and

$$\Re = \frac{\gamma^2}{2\sigma^2} \tag{13}$$

a sufficient condition for

$$\text{Prob}\left(M^{-1} \sum_{s=1}^{M} \chi_{s,N} \geq \gamma M^{-\frac{L-2}{2L-2}}\right) \leq \exp\left(-\Re M^{\frac{1}{L-1}}\right) \tag{14}$$

as $M, N \to \infty$, is

$$\gamma^{2L-2} < 2^{L-2}(\sigma^2 t_o)^L. \tag{15}$$

$\diamondsuit$

From Lemmas 1 and 2, we first estimate a bound on $\text{Prob}(\overline{EA_{h,g}})$.

Lemma 3:

$$\text{Prob}(\overline{EA_{g,h}}) < \exp\left\{ rn\hbar(\rho_y) - \frac{\rho_y rn(1-2\rho_x^{(t)})^4}{6\alpha}\right\}$$

provided that

$$\left(\frac{\sqrt{\rho_y r}(1-2\rho_x^{(t)})^2}{\sqrt{3}\alpha^{3/4}}\right)^4 < \frac{3}{2}$$

for $g = 1, \cdots, \binom{n}{\rho_x^{(t)} n}$ and $h = 1, \cdots, m$.

$\diamondsuit$

Proof of Lemma 3: Without loss of generality, we assume that all the components of the pattern pair (X_h, Y_h) are positive: $X_h = (1, \cdots, 1)^T$ and $Y_h = (1, \cdots, 1)^T$. Let J be the set of indexes at which $X^{(t)}$ and X_h differ. For a given $X^{(t)} \in S_{h,t}$, there is only one J and $|J| = \rho_x^{(t)} n$. Let K be a set of indexes of $Y^{(t)}$ where $|K| = \rho_y p$. Note that there are $\binom{p}{\rho_y p}$ such sets. Event $\overline{EA_{g,h}}$ implies that there is at least one K such that

$$\sum_{k\in K} \sum_{j=1}^{n} \sum_{i=1}^{n} u_{kji}\, x_i^{(t)}\, x_j^{(t)} < 0.$$

Hence

$$\begin{aligned} &\text{Prob}(\overline{EA_{g,h}}) \\ &\leq \text{Prob}\Bigg(\text{there is at least one } K \text{, where } |K| = \rho_y p \text{, such that } \\ &\qquad \sum_{k\in K} \sum_{j=1}^{n} \sum_{i=1}^{n} u_{kji}\, x_i^{(t)}\, x_j^{(t)} < 0\Bigg) \\ &\leq \binom{p}{\rho_y p} \text{Prob}\left(\sum_{k\in K} \sum_{j=1}^{n} \sum_{i=1}^{n} u_{kji}\, x_i^{(t)}\, x_j^{(t)} < 0 \text{ for a given } K\right). \end{aligned} \tag{16}$$

Let

$$P'' = \text{Prob}\left(\sum_{k\in K} \sum_{j=1}^{n} \sum_{i=1}^{n} u_{kji}\, x_i^{(t)}\, x_j^{(t)} < 0 \text{ for a given } K\right). \tag{17}$$

From (1), we have

$$\begin{aligned} P'' = \text{Prob}\Bigg(&\rho_y p(1-2\rho_x^{(t)})^2 n^2 \\ &+ \sum_{h'\neq h}^{m} \sum_{k\in K} y_{kh'} \left(\sum_{j\notin J} x_{jh'} - \sum_{j\in J} x_{jh'}\right)^2 < 0\Bigg). \end{aligned} \tag{18}$$

One can easily find that

$$E\left[\sum_{k\in K} y_{kh'}\left(\sum_{j\notin J} x_{jh'} - \sum_{j\in J} x_{jh'}\right)^2\right] = 0$$

and

$$E\left[\left(\left(\sum_{k\in K} y_{kh'}\left(\sum_{j\notin J} x_{jh'} - \sum_{j\in J} x_{jh'}\right)\right)^2\right)^2\right] = \rho_y p(3n^2 - 2n).$$

Hence

$$P'' = \text{Prob}\left(\frac{1}{m-1}\sum_{h'\neq h}^{m} \chi_{h'} > \gamma(m-1)^{-1/4}\right) \tag{19}$$

where

$$\gamma = \frac{\sqrt{\rho_y r}(1-2\rho_x^{(t)})^2 n^{(5/2)}}{\sqrt{3n^2-2n}(m-1)^{(3/4)}} \tag{20}$$

and

$$\chi_{h'} = \frac{\sum_{k\in K} y_{kh'}\left(\sum_{j\notin J} x_{jh'} - \sum_{j\in J} x_{jh'}\right)^2}{\sqrt{(3n^2-2n)(\rho_y p)}}. \tag{21}$$

As n and $p \to \infty$

$$\frac{m-1}{n} \to \alpha \quad \text{and } \gamma \to \frac{\sqrt{\rho_y r}(1-2\rho_x^{(t)})^2}{\sqrt{3}\alpha^{(3/4)}}.$$

By substituting Lemma 2 (put $L = 3$) into (19), we obtain

$$P'' \le \exp\left\{-\frac{\rho_y rn(1-2\rho_x^{(t)})^4}{6\alpha}\right\} \tag{22}$$

provided that

$$\left(\frac{\sqrt{\rho_y r}(1-2\rho_x^{(t)})^2}{\sqrt{3}\alpha^{3/4}}\right)^4 < 2t_o^3$$

for some real $t_o > 0$. From the Appendix, t_o must be less than $2^{-2/3}3^{1/3}$. To obtain a valid value of t_o, we have

$$\left(\frac{\sqrt{\rho_y r}(1-2\rho_x^{(t)})^2}{\sqrt{3}\alpha^{3/4}}\right)^4 < \frac{3}{2}.$$

The procedures for checking whether $\chi_{h'}$ satisfies the two conditions (11) and (12) are provided in the Appendix. By substituting (22) into (16), we obtain

$$\text{Prob}(\overline{EA_{g,h}}) \le \binom{p}{\rho_y p} \exp\left\{-\frac{\rho_y rn(1-2\rho_x^{(t)})^4}{6\alpha}\right\}. \tag{23}$$

From Lemma 1, we have

$$\text{Prob}(\overline{EA_{g,h}}) \le \exp\left\{rn\hbar(\rho_y) - \frac{\rho_y rn(1-2\rho_x^{(t)})^4}{6\alpha}\right\}. \square$$

Recall that

$$P_Y^{**} \ge 1 - \sum_{h=1}^{m} \sum_{g=1}^{\binom{n}{\rho_x^{(t)} n}} \text{Prob}(\overline{EA_{g,h}}). \tag{24}$$

From Lemma 3

$$P_Y^{**} \ge 1 - m\binom{n}{\rho_x^{(t)} n} \exp\left\{rn\hbar(\rho_y) - \frac{\rho_y rn(1-2\rho_x^{(t)})^4}{6\alpha}\right\}. \tag{25}$$

With Lemma 1, we can immediately obtain the bound on P_Y^{**} as Lemma 4 below.

Lemma 4:

$$P_Y^{**} \ge 1 - \exp\left\{n\hbar(\rho_x^{(t)}) + \log\alpha n^2 + rn\hbar(\rho_y) - \frac{\rho_y rn(1-2\rho_x^{(t)})^4}{6\alpha}\right\} \tag{26}$$

provided that

$$\left(\frac{\sqrt{\rho_y r}(1-2\rho_x^{(t)})^2}{\sqrt{3}\alpha^{3/4}}\right)^4 < \frac{3}{2}. \tag{27}$$

$\diamondsuit$

Let ρ_y^* be the minimum value of ρ_y such that the right-hand side of (26) tends to one (as $n \to \infty$). From Lemma 4, ρ_y^* is the minimum value of ρ_y such that

$$\hbar(\rho_x^{(t)}) + r\hbar(\rho_y) - \frac{\rho_y r(1-2\rho_x^{(t)})^4}{6\alpha} < 0. \tag{28}$$

Let ρ_y' be the intersection of the line

$$L_1 : y = \frac{\rho_y(1-2\rho_x^{(t)})^4}{6\alpha} - \frac{\hbar(\rho_x^{(t)})}{r} \tag{29}$$

and the curve

$$C_1 : y = \hbar(\rho_y). \tag{30}$$

Then

$$\rho_y^* = \rho_y' + \varepsilon$$

where ε is an arbitrary small positive constant. Apparently, for a given $\rho_x^{(t)} \in [0, 0.5)$, ρ_y^* can be solved numerically (see Fig. 1). Hence, for each pattern pair (X_h, Y_h) and for any error pattern with $\rho_x^{(t)} n$ errors in $X^{(t)}$ $(d(X_h, X^{(t)}) = \rho_x^{(t)} n)$, the probability that the number of errors in $Y^{(t+1)}$ is less than $\rho_y^* p$ $(d(Y_h, Y^{(t+1)}) < \rho_y^* p)$ tends to one (as $n \to \infty$). As ε can be any small positive constant, one can restate the above statement as: the probability that the number of errors in $Y^{(t+1)}$ is less than or equal to $\rho_y' p$ $(d(Y_h, Y^{(t+1)}) \le \rho_y' p)$ tends to one (as $n \to \infty$). Furthermore, according to the feature of L_1 and C_1, the following corollary can be obtained.

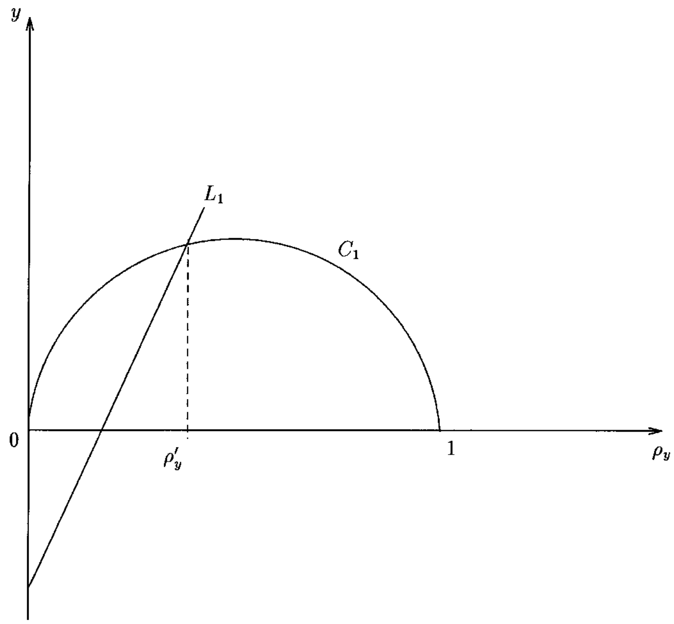

Fig. 1. Graphical implication of solving ρ_y'.

Corollary 1: If $\rho_{x1}^{(t)} < \rho_{x2}^{(t)} < 0.5$, then $\rho_{y1}' < \rho_{y2}'$.
$\diamondsuit$

Proof of Corollary 1: From (29) and (30) (see Fig. 1), if the value of $\rho_x^{(t)} \in (0, 0.5)$ is reduced, the line L_1 shifts up and its slope increases. Hence, the intersection shifts toward left and a smaller ρ_y' is obtained. $\square$

The above corollary implies that a smaller ρ_y' is obtained if the value of $\rho_x^{(t)}$ is reduced. Thus, given $\rho_x^{(t)} \in [0, 5)$, for each pattern pair (X_h, Y_h) and for any $X^{(t)}$ such that $d(X_h, X^{(t)}) \leq \rho_x^{(t)} n$ (the error rate in F_X in the present state is less than or equal to $\rho_x^{(t)}$), *the probability that the error rate in F_Y in the next state is less than or equal to ρ_y' tends to one (as $n \to \infty$)*. We capture the above statement as Corollary 2.

Corollary 2: Given that ρ_y' is the intersection of L_1 and C_1 [see (29) and (30)], for each pattern pair (X_h, Y_h) and for any $X^{(t)}$ such that $d(X_h, X^{(t)}) \leq \rho_x^{(t)} n$, the probability that $d(Y_h, Y^{(t+1)}) \leq \rho_y' p$ tends to one (as $n \to \infty$).
$\diamondsuit$

It follows from Corollary 2 that, if the error rate in F_X in the present state is less than or equal to $\rho_x^{(t)}$, the error rate in F_Y in the next state is less than or equal to ρ_y'. Thus, ρ_y' defines an upper bound on the error rate in F_Y in the next state. We denote this upper bound as $\rho_y^{(t+1)}$.

Similarly, we can easily obtain the lower bound on P_X^{**} as Lemma 5.

Lemma 5:

$$P_X^{**} \geq 1 - \exp\left\{ p\hbar(\rho_y^{(t+1)}) + \log \alpha (\frac{p}{r})^2 + \frac{p}{r}\hbar(\rho_x) - \frac{\rho_x r p (1 - 2\rho_y^{(t+1)})^4}{6\alpha} \right\} \tag{31}$$

provided that

$$\left(\frac{\sqrt{\rho_x} r (1 - 2\rho_y^{(t+1)})^2}{\sqrt{3}\alpha^{3/4}} \right)^4 < \frac{3}{2}. \tag{32}$$

$\diamondsuit$

From Lemma 5, we can estimate the error rate ρ_x' in F_X in the next state (given the error rate in F_Y in the present state) by considering the intersection of the line

$$L_2 : y = \frac{\rho_x r^2 (1 - 2\rho_y^{(t+1)})^4}{6\alpha} - r\hbar(\rho_y^{(t+1)}) \tag{33}$$

and the curve

$$C_2 : y = \hbar(\rho_x). \tag{34}$$

Hence, Corollaries 3 and 4 are obtained.

Corollary 3: If $\rho_{y1}^{(t+1)} < \rho_{y2}^{(t+1)} < 0.5$, then $\rho_{x1}' < \rho_{x2}'$.
$\diamondsuit$

Corollary 4: Given that ρ_x' is the intersection of L_2 and C_2 [see (33) and (34)], for each pattern pair (X_h, Y_h) and for any $Y^{(t+1)}$ such that $d(Y_h, Y^{(t+1)}) \leq \rho_y^{(t+1)} p$, the probability that $d(X_h, X^{(t+1)}) \leq \rho_x' n$ tends to one (as $n \to \infty$).
$\diamondsuit$

Corollary 4 shows that if the error rate in F_Y in the present state is less than or equal to $\rho_y^{(t+1)}$, the error rate in F_X in the next state is less than or equal to ρ_x'. Thus, ρ_x' defines an upper bound on the error rate in F_X in the next state. We denote this upper bound as $\rho_x^{(t+1)}$.

By solving ρ_x' (i.e., $\rho_x^{(t+1)}$) and ρ_y' (i.e., $\rho_y^{(t+1)}$) iteratively, we can construct two sequences $\{\rho_x^{(t)}\}$ and $\{\rho_y^{(t)}\}$. Fig. 2 shows an example of them. These sequences $(\{\rho_x^{(t)}\}, \{\rho_y^{(t)}\})$ form the statistical dynamics of the upper bounds on the error rates. Given arbitrary $\rho_x^{(0)} n$ errors in the initial input, if the sequences $(\{\rho_x^{(t)}\}, \{\rho_y^{(t)}\})$ converge to ρ_x^f and ρ_y^f, respectively (where $\rho_x^{(0)} > \rho_x^f$), the noisy version of the desired pattern pair can be recalled. Besides, in the retrieved item, the number of errors in F_X and the number of errors in F_Y are less than $\rho_x^f n$ and $\rho_y^f p$, respectively. If a few errors are allowed in the retrieved pairs, we can use the above dynamics to estimate the memory capacity, the attraction basin, and the number of errors in the retrieved pairs.

IV. Estimation of Statistical Properties

A. Memory Capacity

Clearly, if the sequences $(\{\rho_x^{(t)}\}, \{\rho_y^{(t)}\})$, with some $\rho_x^{(0)}$, converge to two small numbers individually, each pattern pair can attract all initial inputs within a certain distance. In other words, each pattern pair or its noisy versions can be stored as a stable state. Therefore, we can use the following method to estimate the memory capacity.

For a given r, let α_r be the maximum value of α such that the sequences $(\{\rho_x^{(t)}\}, \{\rho_y^{(t)}\})$ with some $\rho_x^{(0)}$ converge to ρ_x^f and ρ_y^f, respectively (where $\rho_x^{(0)} > \rho_x^f$). Thus, $\alpha_r n^2$ can be considered as a lower bound on the memory capacity.

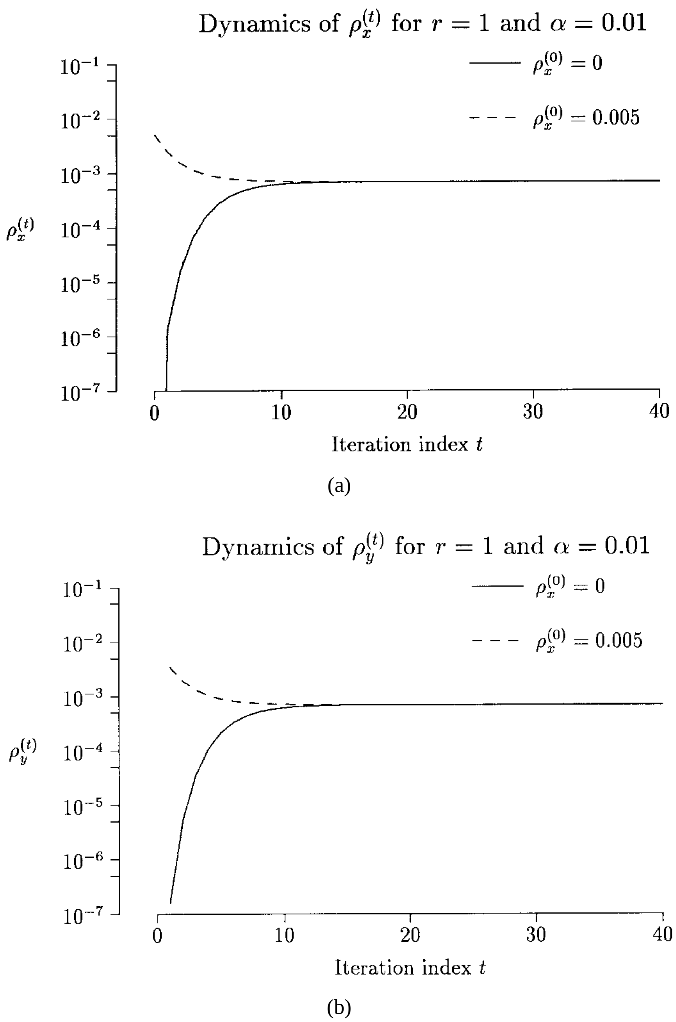

Fig. 2. The statistical dynamics of the SOBAM where $\alpha = 0.01$ and $r = 1$. The initial conditions are $\rho_x^{(0)} = 0$ and 0.005. (a) The dynamics of $\rho_x(t)$ and (b) the dynamics of $\rho_y(t)$. Since all sequences converge, the attraction basin at least equals $0.005n$.

B. Number of Errors in the Retrieved Items

As the sequences $(\{\rho_x^{(t)}\}, \{\rho_y^{(t)}\})$ reflect the upper bounds on the number of errors during recall, the final values of the sequences (ρ_x^f, ρ_y^f) reflect the upper bounds on the number of errors in the retrieved pairs.

C. Attraction Basin

Given arbitrary $\rho_x^{(0)}n$ errors in the initial input, if the sequences converge to (ρ_x^f, ρ_y^f) where $\rho_x^{(0)} > \rho_x^f$, the desired pattern pair can be recalled with no more than $\rho_x^f n$ and $\rho_y^f p$ errors in F_X and F_Y respectively. Hence, the maximum value of $\rho_x^{(0)}$, with which the sequences converge, corresponds to the lower bound on the attraction basin. We denote the maximum value as $\rho_{maxinit}$. The phrases "for any $X^{(t)}$ such that $d(X_h, X^{(t)}) \leq \rho_x^{(t)} n$" in Corollary 2 and "for any $Y^{(t+1)}$ such that $d(Y_h, Y^{(t+1)}) \leq \rho_y^{(t+1)} p$" in Corollary 4 lead directly to the fact that the estimated attraction basin refers to *worst case errors*. In this case, we need to find out if the stored item can attract "every" initial input within a certain distance.

Another error correction index is the attraction basin for *random errors*. In this case, we need to find out if a stored item can attract "an" initial input within a certain distance. The terms "an" and "every" mark the difference between *worst case errors* and *random errors*. It should be noticed that using simulations to study the attraction basin for *worst case errors* is impractical. The reason is that the number of error patterns is very large (for example $\binom{1000}{2} \approx 5 \times 10^5$). Simulations can only reflect the properties of models in the presence of *random errors*.

In the case of *random errors*, if the number of the pattern pairs is less than

$$\min\left(\frac{(1-2\rho)^4 n^2}{18\log n}, \frac{p^2}{18\log p}\right) \tag{35}$$

an initial input (in F_X) within distance ρn from X_h will be attracted to the desired pattern pair in two recall steps with high probability. The proof of the above behavior is based on the theory of large deviation (see [19, Lemma A.7]). As this paper is mainly concerned with *worst case errors*, the behavior of the models in the presence of *random errors* will not be discussed further.

D. Approximations of ρ_x^f and ρ_y^f

It is not difficult to see the following relationship between ρ_x^f and ρ_y^f.

Corollary 5: If the sequences $(\{\rho_x^{(t)}\}, \{\rho_y^{(t)}\}$) converge to small $\rho_x^f (<< 1)$ and small $\rho_y^f (<< 1)$, respectively, then

$$\rho_y^f \approx r\rho_x^f. \tag{36}$$

$\diamondsuit$

Proof of Corollary 5: From (29), (30), (33), and (34), we have

$$\hbar(\rho_y^f) = \frac{\rho_y^f (1-2\rho_x^f)^4}{6\alpha} - \frac{\hbar(\rho_x^f)}{r}$$
$$\hbar(\rho_x^f) = \frac{\rho_x^f r^2 (1-2\rho_y^f)^4}{6\alpha} - r\hbar(\rho_y^f).$$

Thus

$$\frac{\rho_y^f (1-2\rho_x^f)^4}{\rho_x^f (1-2\rho_y^f)^4} = r.$$

As the values of ρ_x^f and ρ_y^f are small

$$\rho_y^f \approx r\rho_x^f. \qquad \square$$

Similarly, we can easily obtain Corollary 6, which can be used for the direct estimation of ρ_x^f and ρ_y^f.

Corollary 6: If the values of ρ_x^f and ρ_y^f are small ($<< 1$), we have

$$\rho_x^f \approx \exp\left\{-\frac{r^2}{(1+r^2)6\alpha} + 1 - \frac{r^2}{1+r^2}\log r\right\} \quad \text{and} \tag{37}$$

$$\rho_y^f \approx \exp\left\{-\frac{r^2}{(1+r^2)6\alpha} + 1 + \frac{1}{1+r^2}\log r\right\}. \tag{38}$$

$\diamondsuit$

TABLE I
THE LOWER BOUND ON THE MEMORY CAPACITY OF THE SOBAM

r	α_r
10	0.0211
5	0.0226
2	0.0204
1	0.0128
0.5	0.00510
0.2	0.00090
0.1	0.00021

Proof of Corollary 6: From (28), we have

$$\hbar(\rho_x^f) + r\hbar(r\rho_x^f) \approx \frac{\rho_x^f r^2(1-2\rho_x^f)^4}{6\alpha}.$$

Using the approximation "$\log(1-\rho) \approx -\rho$ for small positive ρ," the corollary follows. □

Remark: Since we are concerned with the conditions under which the lower bounds on P_X^{**} and P_Y^{**} tend to one (as $n \to \infty$), we only obtain the bounds on the three statistical properties: the memory capacity, the number of errors in the retrieved pairs and the attraction basin. The actual values of these three statistical properties considered in this section are better than the estimated bounds. It should also be noticed that during the construction of the sequences $(\{\rho_x^{(t)}\}, \{\rho_y^{(t)}\})$, we should check whether both conditions (27) and (32) are satisfied.

V. NUMERICAL RESULTS

Numerical Example a: Based on the theoretical work presented in the previous section, we estimate the lower bound on the memory capacity. The estimated results are summarized in Table I. The lower bound increases with r until $r = 10$ where it starts to decrease. Also

$$\alpha_{\frac{1}{r}} \approx \frac{\alpha_r}{r^2}.$$

This symmetrical property means that inverting the ratio of the dimensions (interchange p and n) does not affect the overall estimated lower bound.

One should be aware that when the values of n and p are small, the bound may become meaningless. For example, if $n = p = 10$ and $r = 1$, the lower bound is about 1.28 (which is less than $n = 10$). However, for large n and p, the result is different. For example, if $n = 10^3$ and $r = 1$, the lower bound is about 1.28×10^4 (which is much greater than the dimension $n = 10^3$). For image processing problems [17], the dimensions are usually greater than 10^3.

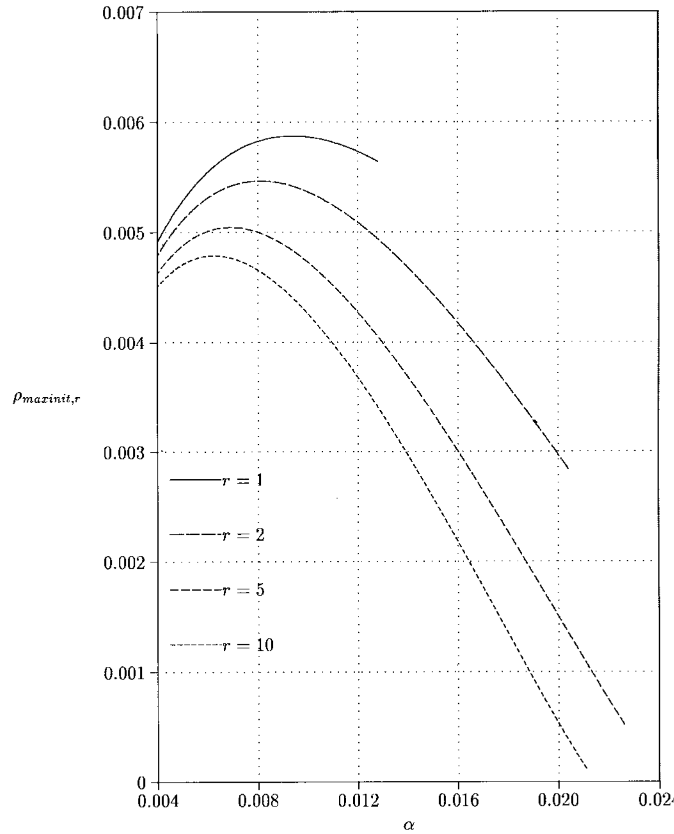

Fig. 3. The lower bound on the attraction basin where $r = 1, 2, 5, 10$.

Numerical Example b:

Fig. 3 summarizes the lower bounds on the attraction basin at $r = 1, 2, 5, 10$. When the value of α is small, the lower bound first increases as α decreases. However, when the value of α is too small, the lower bound starts to decrease as α further decreases. This unnatural trend is due to the constraints (27) and (32), which limit the searching range of $(\rho_x^{(t)}, \rho_y^{(t)})$. However, it is rational to accept the claim that for a smaller α, a larger attraction basin is obtained. We take the maximum point in the figure as the lower bound on the attraction basin for small values of α. Table II summaries the above claim. From Fig. 3 and Table II, for a meaningful attraction basin, the dimension n should be larger than 10^3. The estimated attraction basin is quite small. This is not surprising because the estimated lower bounds refer to *worst case error*.

Figs. 4 and 5 show the behavior of ρ_x^f and ρ_y^f. From the figures, the upper bound on the number of errors in the retrieved pairs (ρ_x^f or ρ_y^f) decreases exponentially as α decreases (this feature matches Corollary 6). Also, the value of ρ_y^f is approximately equal to that of $r\rho_x^f$ (as was indicated in Corollary 5). Since it is desired that the number of retrieved errors should be as small as possible, the estimated upper bounds are more attractive.

VI. HOBAM'S

Although we are mainly concerned with the properties of the SOBAM, we can apply a similar method to analyze HOBAM's. The only required change in the assumptions is

$$m = \alpha n^q \tag{39}$$

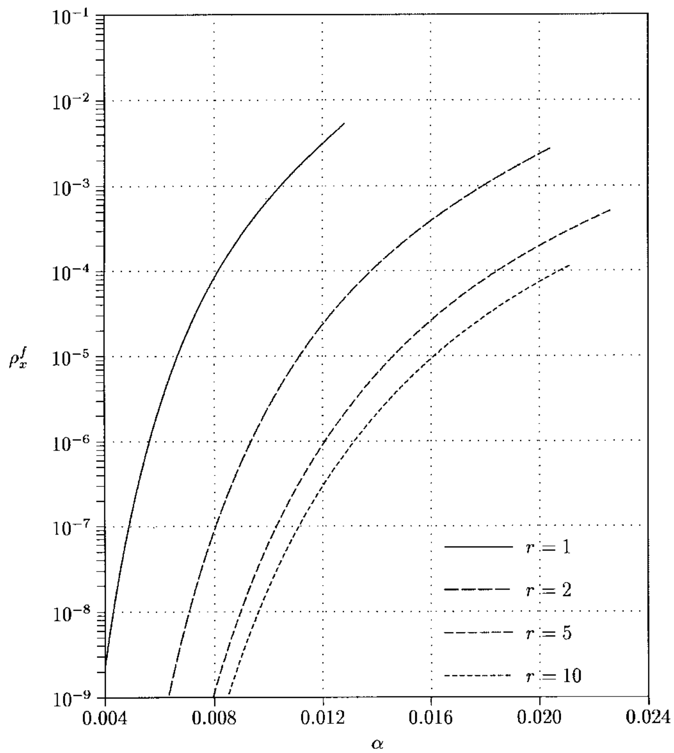

Fig. 4. The upper bound on the error rate in F_X in the retrieved pairs where $r = 1, 2, 5, 10$.

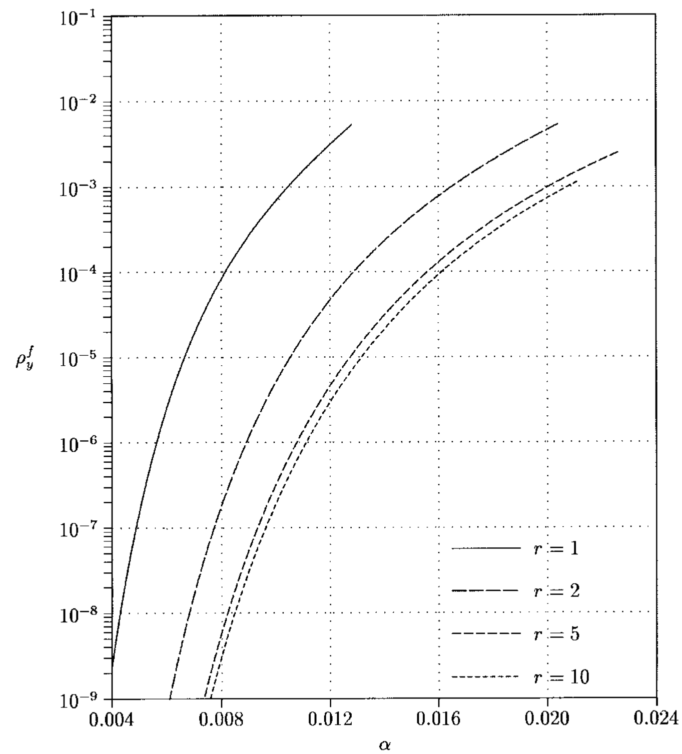

Fig. 5. The upper bound on the error rate in F_Y in the retrieved pairs where $r = 1, 2, 5, 10$.

where q is a positive integer. For the q-order BAM, the connections from F_X to F_Y are

$$u_{k,i_1,i_2,\cdots,i_q} = \sum_{h=1}^{m} y_{kh}\, x_{i_1 h}\, x_{i_2 h} \cdots x_{i_q h} \tag{40}$$

where $k = 1, \cdots, p$; $i_1 = 1, \cdots, n$; $\cdots$ and $i_q = 1, \cdots, n$. The connections from F_Y to F_X are

$$v_{j,l_1,l_2,\cdots,l_q} = \sum_{h=1}^{m} x_{jh}\, y_{l_1 h}\, y_{l_2 h} \cdots y_{l_q h} \tag{41}$$

where $j = 1, \cdots, n$; $l_1 = 1, \cdots, p$; $\cdots$ and $l_q = 1, \cdots, p$. The corresponding recall equations are

$$y_k^{(t+1)} = \operatorname{sgn}\left(\sum_{i_q=1, i_{q-1}=1, \cdots, i_1=1}^{n} u_{k,i_1,i_2,\cdots,i_q}\, x_{i_1}^{(t)}\, x_{i_2}^{(t)} \cdots x_{i_q}^{(t)} \right) \tag{42}$$

and

$$x_j^{(t+1)} = \operatorname{sgn}\left(\sum_{l_q=1, l_{q-1}=1, \cdots, l_1=1}^{p} v_{j,l_1,l_2,\cdots,l_q}\, y_{l_1}^{(t+1)}\, y_{l_2}^{(t+1)} \cdots y_{l_q}^{(t+1)} \right). \tag{43}$$

We can obtain similar results for the q-order BAM based on the following lemma.

Lemma 6: Let $\xi_1, \xi_2, \cdots, \xi_n$ be ± 1 equiprobable independent random variables and $S_n' = \frac{\sum_{i=1}^{n} \xi_i}{\sqrt{n}}$ where n is a positive integer. As $n \to \infty$

$$E\left[(S_n')^{2q} \right] \to \frac{(2q)!}{2^q q!}. \tag{44}$$

TABLE II
THE LOWER BOUND ON THE ATTRACTION BASIN OF THE SOBAM FOR SMALL VALUES OF α

r	$\rho_{maxinit,r}$
10	0.00478 ($\alpha < 0.00641$)
5	0.00504 ($\alpha < 0.00701$)
2	0.00546 ($\alpha < 0.00814$)
1	0.00587 ($\alpha < 0.00933$)

$\diamondsuit$

Proof of Lemma 6: As $n \to \infty$, the distribution of S_n' tends to be normal. Since the $2q$th moment of a normal random variable [18] is $1 \cdot 3 \cdots (2q-1)$

$$\begin{aligned} E\left[(S_n')^{2q} \right] &\to 1 \cdot 3 \cdots (2q-1) \\ &= \frac{(2q)!}{2^q q!}. \end{aligned}$$

$\square$

From Lemma 6, we can obtain the four corollaries for the q-order BAM.

Corollary 7: For every pattern pair (X_h, Y_h) and for any $X^{(t)}$ such that $d(X_h, X^{(t)}) \le \rho_x^{(t)} n$, the probability that $d(Y_h, Y^{(t+1)}) \le \rho_y' p$ tends to one (as $n \to \infty$), provided that

$$\left(\frac{\sqrt{\rho_y r}(1 - 2\rho_x^{(t)})^q}{\sqrt{\lambda_q}\alpha^{(q+1)/(2q)}} \right)^{2q} < \frac{\lambda_q}{2}$$

$$\frac{t_o^k}{k!}E\left[|\ \chi\ |^{\omega k}\right] \le C_o e^{-(((q+1)\omega)/2-1)k}\,(t_o 2^{q\omega/2}\lambda_q^{\omega/2})^k\,(\omega k+1)^{(\omega k)/2}\,(\frac{q\omega k+1}{2})^{(q\omega k)/2}k^{-k-1/2}. \tag{54}$$

where

$$\lambda_q = \frac{(2q)!}{2^q q!}$$

and ρ'_y is the intersection of L_{1q} and C_{1q}

$$L_{1q}\,:\,y = \frac{\rho_y(1-2\rho_x^{(t)})^{2q}}{2\lambda_q\alpha} - \frac{\hbar(\rho_x^{(t)})}{r}$$
$$C_{1q}\,:\,y = \hbar(\rho_y).$$

◇

Corollary 8: For every pattern pair (X_h, Y_h) and for any $Y^{(t+1)}$ such that $d(Y_h, Y^{(t+1)}) \le \rho_y^{(t+1)}p$, the probability that $d(X_h, X^{(t+1)}) \le \rho'_x n$ tends to one (as $n \to \infty$), provided that

$$\left(\frac{\sqrt{\rho_x}r^{q/2}(1-2\rho_y^{(t+1)})^q}{\sqrt{\lambda_q}\alpha^{(q+1)/(2q)}}\right)^{2q} < \frac{\lambda_q}{2}$$

where ρ'_x is the intersection of L_{2q} and C_{2q}

$$L_{2q}\,:\,y = \frac{\rho_x r^q(1-2\rho_y^{(t+1)})^{2q}}{2\lambda_q\alpha} - r\hbar(\rho_y^{(t+1)})$$
$$C_{2q}\,:\,y = \hbar(\rho_x).$$

◇

Corollary 9: If the sequences $(\{\rho_x^{(t)}\}, \{\rho_y^{(t)}\}\ )$ converge to small $\rho_x^f (<< 1)$ and small $\rho_y^f (<< 1)$, respectively, then

$$\rho_y^f \approx r^{q-1}\rho_x^f. \tag{45}$$

◇

Corollary 10: If the values of ρ_x^f and ρ_y^f are small $(<< 1)$, we have

$$\rho_x^f \approx \exp\left\{-\frac{r^q}{(1+r^q)2\lambda_q\alpha} + 1 - \frac{r^q}{1+r^q}\log r^{q-1}\right\} \tag{46}$$

and

$$\rho_y^f \approx \exp\left\{-\frac{r^q}{(1+r^q)2\lambda_q\alpha} + 1 + \frac{1}{1+r^q}\log r^{q-1}\right\}. \tag{47}$$

◇

VII. Concluding Remarks

In this paper, we have studied several properties of the SOBAM. An example has been given which shows that the SOBAM may not be stable during recall. We have also derived the statistical dynamics of the SOBAM. Based on this dynamics, we have estimated the memory capacity, the attraction basin and the number of errors in the retrieved pairs. Numerical examples have also been presented. Last, we have briefly discussed how our results can be extended to HOBAM's. One significant advantage of the methodology presented is that we can analyze some associative memories whose stability is not guaranteed.

Appendix
Verification of the Conditions of Newman's Lemma

Here, we show the conditions under which the random variable

$$\chi = \frac{(\sum_{k=1}^{\rho_y p} y_k)(\sum_{j=1}^{n} x_j)^q}{\sqrt{\rho_y p\lambda_q n^q}}$$

satisfies (11) and (12), where y_k's and x_j's are ± 1 equiprobable independent random variables. Clearly, χ is symmetric and

$$E[\chi] = 0. \tag{48}$$

Also, from Lemma 6

$$\mathrm{Var}(\chi) = E\left[\chi^2\right] = 1. \tag{49}$$

Hence, (11) is satisfied.

To check whether χ satisfies (12), we use an existing result about the sum of ± 1 equiprobable independent random variables [19].

Lemma 7: Let $\xi_1, \xi_2, \cdots, \xi_n$ be ± 1 equiprobable independent random variables. For $z > 0$ and large n

$$E\left[\frac{|\sum_{i=1}^n \xi_i|^z}{n^{z/2}}\right] \le 2^{z/2+1}\pi^{-1/2}\Gamma(\frac{z+1}{2}) \tag{50}$$

where $\Gamma(a)$ is the gamma function

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx. \tag{51}$$

◇

The above lemma is part of [19, Lemma A.6]. From Lemma 7, we have

$$E[|\ \chi\ |^z] \le 2^{z/2+1}\pi^{-1/2}\Gamma(\frac{z+1}{2})\,2^{(qz+1)/2}\pi^{-1/2}\Gamma(\frac{qz+1}{2})\lambda^{-z/2}. \tag{52}$$

Let $z = \omega k$, where k is a positive integer and $\omega > 0$. Then

$$\frac{t_o^k}{k!}E[\chi^z] \le C_o\frac{t_o^k 2^{((q+1)\omega k)/2}}{k!}\Gamma(\frac{\omega k+1}{2})\Gamma(\frac{q\omega k+1}{2})\lambda^{-q\omega k/2} \tag{53}$$

where C_o is a positive constant. For large k

$$k! \approx \sqrt{2\pi}e^{-k}k^{k+1/2}$$
$$\Gamma(\frac{\omega k+1}{2}) \approx \sqrt{2\pi}e^{-(\omega k+1)/2}(\frac{\omega k+1}{2})^{(\omega k+1)/2-1/2}$$

and

$$\Gamma(\frac{q\omega k+1}{2}) \approx \sqrt{2\pi}e^{-(q\omega k+1)/2}(\frac{q\omega k+1}{2})^{(q\omega k+1)/2-1/2}.$$

Hence we have (54), shown at the top of the page.

For large k, the kth term of the sum

$$S = \sum_{k=0}^{\infty}\frac{t_o^k}{k!}E\left[|\ \chi\ |^{\omega k}\right] \tag{55}$$

decreases exponentially provided that $\omega = \frac{2}{q+1}$ and $t_o < 2^{-(q)/(q+1)} \lambda_q^{1/(q+1)}$. As S converges to

$$E\left[\exp\left\{t_o \mid \chi \mid^{2/(q+1)}\right\}\right]$$

it follows that

$$\limsup_{n\to\infty} E\left[\exp\left\{t_o \mid \chi \mid^{2/(q+1)}\right\}\right] < \infty.$$

For the SOBAM, $q = 2$ and $t_o < 2^{-2/3}3^{1/3}$.

REFERENCES

[1] T. Kohonen, "Correlation matrix memories," *IEEE Trans. Comput.*, vol. 21, pp. 353–359, 1972.

[2] G. Palm, "On associative memory," *Biol. Cybern.*, vol. 36, pp. 19–31, 1980.

[3] B. Kosko, "Bidirectional associative memories," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, pp. 49–60, 1988.

[4] C. S. Leung, "Encoding method for bidirectional associative memory using projection on convex x sets," *IEEE Trans. Neural Networks*, vol. 4, pp. 879–881, Sept. 1993.

[5] ——, "Optimum learning for bidirectional associative memory in the sense of capacity," *IEEE Trans. Syst., Man, Cybern.*, vol 24, pp. 791–796, May 1994.

[6] Y. F. Wang, J. B. Cruz, Jr., and J. H. Mulligan, Jr., "Two coding strategies for bidirectional associative memory," *IEEE Trans. Neural Networks*, vol. 1, pp. 81–92, 1990.

[7] P. K. Simpson, "Higher-ordered and intraconnected bidirectional associative memories," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, pp. 637–653, 1990.

[8] D. Psaltis, C. H. Park, and J. Hong, "Higher order associative memories and their optical implementations," *Neural Networks*, vol. 1, pp. 149–163, 1988.

[9] J. M. Kinser, H. J. Caulfield, and J. Shamir, "Design for a massive all-optical bidirectional associative: The big BAM," *Appl. Opt.*, vol. 27, no. 16, pp. 3442–1344, 1988.

[10] C. M. Newman, "Memory capacity in neural models: Rigorous lower bounds," *Neural Networks*, vol. 1, pp. 223–238, 1988.

[11] J. Komlos and R. Paturi, "Convergence results in an associative memory model," *Neural Networks*, vol. 1, pp. 229–250, 1988.

[12] R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh, "The capacity of the hopfield associative memory," *IEEE Trans. Inform. Theory*, vol. 33, pp. 461–482, 1987.

[13] S. Amari, "Mathematical foundations of neurocomputing," *Proc. IEEE*, vol. 78, pp. 1443–1463, 1990.

[14] S. Amari and K. Maginu, "Statistical neurodynamics of associative memory," *Neural Networks*, vol. 1, pp. 63–73, 1988.

[15] C. S. Leung, "The performance of dummy augmentation encoding," in *Proc. IJCNN 93 Nagoya*, vol. 3, 1993, pp. 2674–2677.

[16] K. Haines and R. Hecht-Nielsen, "A BAM with increased information storage capacity," in *Proc. 1988 IEEE Int. Conf. Neural Networks*, 1988, pp. 181–190.

[17] G. A. Kohring, "On the Q-state neuron problem in attractor neural networks," *Neural Networks*, vol. 6, pp. 573–581, 1993.

[18] A. Papoulis, *Probability, Random Variables, and Stochastic Process*. New York: McGraw-Hill, 1985.

[19] S. S. Venkatesh and P. Bald, "Programmed interactions in higher-order neural networks: the outer-product algorithm," *J. Complexity*, vol. 7, pp. 443–479, 1991.

Chi-Sing Leung received the B.Sc. degree in electronics in 1989, the M.Phil. degree in information engineering in 1991, and the Ph.D. degree in computer science and engineering in 1995, all from the Chinese University of Hong Kong, Shatin, Hong Kong.

In 1995, he was a Research Fellow at the Chinese University of Hong Kong. He was a Lecturer with the School of Science and Technology at the Open Learning Institute of Hong Kong. Currently, he is a Visiting Fellow at the department of Computer Science, University of Wollongong, Australia. His research interests include learning and modeling of artificial neural networks, and digital communications.

Lai-Wan Chan (S'87–M'88) received the B.A. and M.A. degrees in electrical science and the Ph.D. degree in engineering from Cambridge University, Cambridge, U.K.

She joined the Computer Science and Engineering Department of the Chinese University of Hong Kong, Shatin, Hong Kong, in 1989, and is currently and Associate Professor. Her research interests include the learning and modeling of artificial neural networks, in particular recurrent neural networks and applications of neural networks in time series prediction, image recognition, and Cantonese speech recognition.

Dr. Chan was Chairperson of the IEEE Computer Chapter in the Hong Kong Section, is a Governing Board Member of the Asian-Pacific Neural Network Assembly, and was the Organizing Cochair of the International Conference on Neural Information Processing in 1996.

Edmund Man Kit Lai, photograph and biography not available at the time of publication.