Correlation Matrix Memories

TEUVO KOHONEN

Abstract—A new model for associative memory, based on a correlation matrix, is suggested. In this model information is accumulated on memory elements as products of component data. Denoting a *key vector* by $q^{(p)}$, and the *data* associated with it by another vector $x^{(p)}$, the pairs $(q^{(p)}, x^{(p)})$ are memorized in the form of a matrix

$$c\sum_p x^{(p)} q^{(p)T} = M_{xq}$$

where c is a constant. A randomly selected subset of the elements of M_{xq} can also be used for memorizing. The recalling of a particular datum $x^{(r)}$ is made by a transformation $x^{(r)} = M_{xq}q^{(r)}$. This model is failure tolerant and facilitates associative search of information; these are properties that are usually assigned to holographic memories. Two classes of memories are discussed: a *complete correlation matrix memory* (CCMM), and randomly organized *incomplete correlation matrix memories* (ICMM). The data recalled from the latter are stochastic variables but the fidelity of recall is shown to have a deterministic limit if the number of memory elements grows without limits. A special case of correlation matrix memories is the *auto-associative* memory in which any part of the memorized information can be used as a key. The memories are selective with respect to accumulated data. The ICMM exhibits adaptive improvement under certain circumstances. It is also suggested that correlation matrix memories could be applied for the classification of data.

Index Terms—Associative memory, associative net, associative recall, correlation matrix memory, nonholographic associative memory, pattern recognition.

I. Introduction

FOR the associative search of memorized information, optical holography has been suggested [1]–[4]. Some specific mathematical models for simulated holographic memories have recently been presented [5]–[8]. It has also been assumed that the recording of information in biological memories might be based on holography [9]–[11]. Steinbuch [12], [13], and Willshaw, Buneman, and Longuet-Higgins [14], [15] suggested a nonholographic associative memory for the same purpose, based on a switching matrix.

In this paper we replace the "Lernmatrix" of Steinbuch by a correlation matrix of component signals. It is assumed that products of signals can be formed and memorized by network elements. The possibility of the formation of products of neural signals has been analyzed by Rapoport [16], Jenik [17], [18], and Küpfmüller and Jenik [19]. In this paper we discuss some analytical properties of matrix transformations used for associative recall but not the possible role of such networks in neural systems.

The correlation matrix might become immensely large with a large number of input signals. We can show that it will then suffice to take a set of random samples of its elements for the representation of the information stored in the matrix, and the rest of the elements are put equal to zero. In addition to such a randomly sampled matrix, we discuss a randomly generated associative network in which a set of memory elements is interconnected at random to two input elements. If the number of memory elements is sufficiently large, this model still reconstructs information stored in it. Such a matrix is both failure tolerant, and completely randomly organized.

In this paper, after setting up the structure of the model, we will make a formal mathematical approach to the problem in which the statistical nature of this model is analyzed.

Manuscript received July 27, 1971; revised October 27, 1971.
The author is with the Department of Technical Physics, Helsinki University of Technology, Otaniemi, Finland.

The Model

The correlation matrix model has the structure depicted in Fig. 1. Here we have an input field which consists of two parts: the set of input elements denoted by an index set I comprises a *key field* used for the encoding of data, and the set denoted by an index set J is a *data field*. All input elements, called *receptors* in what follows, receive a set of simultaneous input signals. For simplicity, we are working in discrete-time representation in which the signals are assumed to be present at sampling instants. All signals of the key field taken together form a *key vector* denoted by $q^{(p)}$; here the superscript p is a discrete-time index, or the label of a particular *pattern*. All signals of the data field taken together form a *datum vector* denoted by $x^{(p)}$ where p labels the pattern. In *component form*, elements of the set $\{q_i^{(p)} \mid i \in I\}$ constitute the components of the key vector whereas $\{x_j^{(p)} \mid j \in J\}$ is the set of data signals.

Yet another field of *memory elements* or *associators* consists of elements labeled by a pair (i, j) corresponding to the ith element of the key field and the jth element of the data field to which the associator is connected. If there are connections for all possible pairs (i, j) and only one of a type, we speak of a *complete correlation matrix memory* (CCMM). If connections exist only for a randomly selected subset of all possible pairs (i, j), or if they are created at random without prior examination for the existence of a pair of connections, we speak of *incomplete correlation matrix memories* (ICMM's). The *memorization* of data is made by increasing the value of every associator element M_{ji} by an amount directly proportional to $q_i^{(p)}x_j^{(p)}$: for a set

$$P = \{1, 2, \cdots, p, \cdots\}$$

of patterns,

$$M_{ji} = c \sum_{p \in P} q_i^{(p)} x_j^{(p)} \tag{1}$$

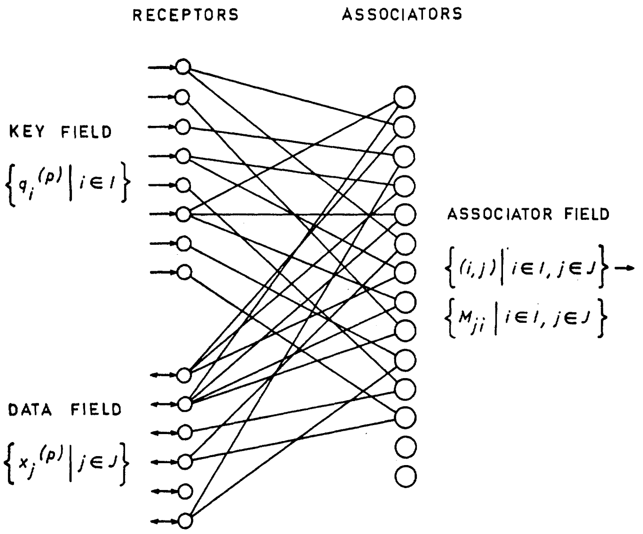

Fig. 1. Associative network.

where c is a normalizing constant. The *recall* of a particular $x_j{}^{(r)}$ associated with a key vector $q^{(r)}$ is made by transformation

$$\hat{x}_j{}^{(r)} = \sum_i M_{ji} q_i{}^{(r)}. \tag{2}$$

The memorization–recall–transformation defined by (1) and (2) has a bearing on the well-known Gauss–Markov estimator.

First we show that the recalled data indeed have a certain structural conformity to the memorized information.

II. Complete Correlation Matrix Memory (CCMM)

The complete correlation matrix is an array that has one and only one memory element for every pair (i, j) of indices, for $i \in I$, $j \in J$. The contents of the associator field are described by the matrix

$$M_{xq} = c \sum_p \mathbf{x}^{(p)} \mathbf{q}^{(p)^T} \tag{3}$$

where c is a constant and T denotes the transpose. A particular datum $\mathbf{x}^{(r)}$ will be recalled by a transformation

$$\hat{\mathbf{x}}^{(r)} = M_{xq} \mathbf{q}^{(r)}. \tag{4}$$

Let us substitute M_{xq} from (3) to (4):

$$\hat{\mathbf{x}}^{(r)} = c \sum_p \mathbf{x}^{(p)} \mathbf{q}^{(p)^T} \mathbf{q}^{(r)}$$
$$= c\mathbf{x}^{(r)} [\|\mathbf{q}^{(r)}\|^2] + c \sum_{p \neq r} \mathbf{x}^{(p)} \mathbf{q}^{(p)^T} \mathbf{q}^{(r)} \tag{5}$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector. If the inner products of any two key vectors $\mathbf{q}^{(p)}$ and $\mathbf{q}^{(r)}$ are zero, i.e., if all key vectors are *orthogonal*, we see that $\hat{\mathbf{x}}^{(r)}$ is directly proportional to $\mathbf{x}^{(r)}$. If the Euclidean norms of different key vectors are equal, c can further be selected as

$$c = \|\mathbf{q}^{(r)}\|^{-2} \qquad \text{(constant with } r\text{)} \tag{6}$$

in which case the recall is perfect for every $\mathbf{x}^{(r)}$,

$$\hat{\mathbf{x}}^{(r)} = \mathbf{x}^{(r)}.$$

The nonorthogonality of key vectors gives rise to crosstalk that may have the same polarity as the recorded data or the opposite. The relative crosstalk level for a particular datum is defined as

$$L^{(p,r)} = \frac{\mathbf{q}^{(p)^T} \mathbf{q}^{(r)}}{\|\mathbf{q}^{(r)}\|^2}. \tag{7}$$

Equation (7) gives a measure to *selectivity*.

III. Incomplete Correlation Matrix Memories (ICMM)

A. Stochastically Sampled Correlation Matrix

The number of matrix elements, mn in total, may grow impractically large if the dimensions of q and x are increased. Let us first discuss in this section a hypothetical case in which a randomly selected subset of the elements of M_{xq} is used to represent the complete correlation matrix. Let us define *sampling coefficients* s_{ij} that take the value 1 at all sampled elements of the correlation matrix and 0 elsewhere. The *sampled correlation matrix* is then defined as

$$(M_{xq})_{ji} = c \sum_p s_{ij} q_i{}^{(p)} x_j{}^{(p)} \tag{8}$$

and the recalled pattern reads, in analogy with (5), for all j,

$$\hat{x}_j{}^{(r)} = c \sum_p \sum_{i=1}^{m} s_{ij} q_i{}^{(p)} q_i{}^{(r)} x_j{}^{(p)}$$
$$= c \left[\sum_{i=1}^{m} s_{ij} (q_i{}^{(r)})^2 \right] x_j{}^{(r)} + c \sum_{p \neq r} \sum_{i=1}^{m} s_{ij} q_i{}^{(p)} q_i{}^{(r)} x_j{}^{(p)}$$
$$= K_j{}^{(r)} x_j{}^{(r)} + \sum_{p \neq r} K_j{}^{(p,r)} x_j{}^{(p)}. \tag{9}$$

The gain factors $K_j{}^{(r)}$ and $K_j{}^{(p,r)}$ are *stochastic variables*. Notice that if we were dealing with estimation problems or the classification of statistically distributed input signals, we should discuss the input signals $q_i{}^{(p)}$ and $x_j{}^{(p)}$ as stochastic variables. This, however, is not the case with the present study. In our model every signal has a unique value at every sampling instant, i.e., a value that must be memorized as such. To put it in another way, because the signals are not regarded as stochastic variables, there is no need to discuss their statistical distributions or correlations between these signals. In the following we thus assume that the sampling of matrix elements is independent of all $q_i{}^{(p)}$ in which case the $q_i{}^{(p)}$ can be regarded as *constant parameters* with values attained in a particular realization, and only the $s_{ij} \in \{0, 1\}$ are stochastic variables; the probability for s_{ij} being 1 is denoted by w $(0 \leq w \leq 1)$ where

$$w = \frac{s}{mn} \tag{10}$$

and

- s Number of sampled matrix elements.
- m Number of components in q.
- n Number of components in x.

Because it is our main objective to analyze what sort of noise (or statistical error) is introduced by the use of a randomly sampled incomplete matrix instead of a complete one, we shall derive expressions for the expectation values and variances of signals recalled from the memory, whereby these statistical operations refer only to the sampling process.

It is known from elementary statistics that if $Y_1, Y_2, \cdots, Y_m$ are independent stochastic variables with means $M_1, M_2, \cdots, M_m$ and variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_m^2$, respectively, we have for a linear combination X of the Y's,

$$X = \sum_{i=1}^{m} a_i Y_i$$

$$E(X) = \sum_{i=1}^{m} a_i M_i$$

$$\mathrm{var}\,(X) = \sum_{i=1}^{m} a_i^2 \sigma_i^2.$$

Because all s_{ij} can be assumed independent since n is a large number, the probability distribution Pr $(s_{ij}=1)$, as is well known, is *binomial*. The *mean* of s_{ij} is then

$$E(s_{ij}) = w \tag{11}$$

and the *variance* of s_{ij} is

$$\mathrm{var}\,(s_{ij}) = w(1-w). \tag{12}$$

For $K_j^{(r)}$ and $K_j^{(p,r)}$ we have now, regarding $q_i^{(p)}$ as parameters, the means and variances

$$E(K_j^{(r)}) = cw \sum_{i=1}^{m} (q_i^{(r)})^2 \tag{13}$$

$$\mathrm{var}\,(K_j^{(r)}) = c^2 w(1-w) \sum_{i=1}^{m} (q_i^{(r)})^4. \tag{14}$$

The fidelity of recall is conveniently described in terms of a *relative standard deviation* which for the noise due only to the searched pattern is

$$\frac{\sqrt{\mathrm{var}\,(K_j^{(r)})}}{E(K_j^{(r)})} = \sqrt{\frac{1-w}{w}}\;\frac{\sqrt{\sum_{i=1}^{m} (q_i^{(r)})^4}}{\sum_{i=1}^{m} (q_i^{(r)})^2}\,. \tag{15}$$

The crosstalk from another pattern (p) is expressed by the average crosstalk level,

$$E(K_j^{(p,r)}) = cw \sum_{i=1}^{m} q_i^{(p)} q_i^{(r)} \tag{16}$$

and the variance of this level is

$$\mathrm{var}\,(K_j^{(p,r)}) = c^2 w(1-w) \sum_{i=1}^{m} (q_i^{(p)} q_i^{(r)})^2. \tag{17}$$

The relative standard deviation of the recalled data due to all patterns is obtained from (9) and the previous considerations, for all j,

$$\frac{\sqrt{\mathrm{var}\,(\hat{x}_j^{(r)})}}{E(\hat{x}_j^{(r)})} = \sqrt{\frac{1-w}{w}}\;\frac{\sqrt{\sum_{i=1}^{m}\left(\sum_{p} q_i^{(p)} q_i^{(r)} x_j^{(p)}\right)^2}}{\sum_{p}\sum_{i=1}^{m} q_i^{(p)} q_i^{(r)} x_j^{(p)}}\,. \tag{18}$$

If other parameters are finite, the relative standard deviation approaches zero for $w \to 1$.

Example 1: Let us take a representative case with one memorized pattern only and $q_i^{(p)} \in \{+1, -1\}$ for which we obtain

$$\frac{\sqrt{\mathrm{var}\,(\hat{x}_j^{(r)})}}{E(\hat{x}_j^{(r)})} = \sqrt{\frac{1-w}{mw}} = \sqrt{\frac{mn-s}{ms}}\,. \tag{19}$$

Since (19) gives the relative standard deviation of a recalled pattern, we obtain an order of magnitude rule. If w is much smaller than 1, as is usually the case, if the maximum allowed relative standard deviation is N, and if the number of components in the key vector is $m \gg 1$, then the minimum number of sampled matrix elements must be

$$s = mnw \geq \frac{n}{N^2} \tag{20}$$

which thus depends on the number n of components in the data vector x only. For example, if $N=0.1$, we must have $s \geq 100n$.

It is not difficult to generalize the above results and show that the number of sampled matrix elements for a given percentage noise is obviously directly proportional to the number of components in the data vector, for any amount of them. This, of course, is advantageous with a large number of elements in q.

Example 2: If we again consider the recall of a memorized single pattern, and $q_i^{(p)} \in \{+1, 0, -1\}$, we have a distribution of $q_i^{(p)}$ which is less uniform than in Example 1. Denoting the fraction of nonzero components in the key vectors by $\alpha (0 \leq \alpha \leq 1)$ we have

$$\frac{\sqrt{\mathrm{var}\,(\hat{x}_j^{(r)})}}{E(\hat{x}_j^{(r)})} = \sqrt{\frac{1-w}{\alpha m w}} = \sqrt{\frac{mn-s}{\alpha m s}}\,. \tag{21}$$

In this case the necessary number of sampled matrix elements for relative standard deviation N is

$$s \geq \frac{n}{\alpha N^2}\,. \tag{22}$$

For example, if there are 80 percent zeros in the key and $N=0.1$, we must have $s \geq 500n$.

B. Randomly Generated Correlation Matrix Memory

We will now make a more realistic approach and assume that connections are generated stochastically *without prior examination for the existence of connections.* A matrix M_{xq} formed of the M_{ji} may thus include *multiple elements* if two associators have identical connections. The only difference with respect to the cases of Section III-A is that we must now use *occupation numbers* z_{ij} of matrix elements (instead of s_{ij}), $z_{ij} \in \{0, 1, 2, \cdots\}$.

$$(M_{xq})_{ji} = c \sum_p z_{ij} q_i^{(p)} x_j^{(p)} \tag{23}$$

where the distribution of z_{ij} is *Poissonian,*

$$\Pr(z_{ij} = \zeta) = \frac{\mu^\zeta}{\zeta!} e^{-\mu} \tag{24}$$

and the parameter of this distribution is

$$\mu = \frac{s}{mn} \tag{25}$$

where s is the number of associators.

The distribution defined by (24) is obviously the same as the distribution of hits in s stochastic throws on squares of a board with mn squares. The recalled pattern is for all j,

$$\begin{aligned}\hat{x}_j^{(r)} &= \left[c \sum_{i=1}^m z_{ij} (q_i^{(r)})^2 \right] x_j^{(r)} + c \sum_{p \neq r} \sum_{i=1}^m z_{ij} q_i^{(p)} q_i^{(r)} x_j^{(p)} \\ &= K_j^{(r)} x_j^{(r)} + \sum_{p \neq r} K_j^{(p,r)} x_j^{(p)}. \end{aligned} \tag{26}$$

Now we have

$$E(z_{ij}) = \mu = \frac{s}{mn} \tag{27}$$

$$\operatorname{var}(z_{ij}) = \mu \tag{28}$$

and then

$$E(K_j^{(r)}) = c\mu \sum_{i=1}^m (q_i^{(r)})^2 \tag{29}$$

$$\operatorname{var}(K_j^{(r)}) = c^2 \mu \sum_{i=1}^m (q_i^{(r)})^4 \tag{30}$$

and

$$E(K_j^{(p,r)}) = c\mu \sum_{i=1}^m q_i^{(p)} q_i^{(r)} \tag{31}$$

$$\operatorname{var}(K_j^{(p,r)}) = c^2 \mu \sum_{i=1}^m (q_i^{(p)} q_i^{(r)})^2. \tag{32}$$

The relative standard deviation of gain factor $K_j^{(r)}$ due to the searched pattern only is

$$\frac{\sqrt{\operatorname{var}(K_j^{(r)})}}{E(K_j^{(r)})} = \mu^{-1/2} \frac{\sqrt{\sum_{i=1}^m (q_i^{(r)})^4}}{\sum_{i=1}^m (q_i^{(r)})^2} \tag{33}$$

and the relative standard deviation of a recalled signal, due to all patterns, is

$$\frac{\sqrt{\operatorname{var}(\hat{x}_j^{(r)})}}{E(\hat{x}_j^{(r)})} = \mu^{-1/2} \frac{\sqrt{\sum_{i=1}^m \left(\sum_p q_i^{(p)} q_i^{(r)} x_j^{(p)} \right)^2}}{\sum_p \sum_{i=1}^m q_i^{(p)} q_i^{(r)} x_j^{(p)}} \cdot \tag{34}$$

If other parameters are finite, the relative standard deviation approaches zero for $s \to \infty$, i.e., $\mu \to \infty$.

Example 3: Let us repeat the problem of Example 1 for randomly generated connections. The only difference is in the expression for variance and we obtain

$$\frac{\sqrt{\operatorname{var}(K_j^{(r)})}}{E(K_j^{(r)})} = \sqrt{\frac{1}{\mu m}} \cdot \tag{35}$$

For any value of μ, we must now have

$$s \geq \frac{n}{N^2} \cdot \tag{36}$$

Example 4: Repeating the problem of Example 2 for randomly generated connections, we have

$$\frac{\sqrt{\operatorname{var}(K_j^{(r)})}}{E(K_j^{(r)})} = \sqrt{\frac{1}{\alpha \mu m}} \cdot \tag{37}$$

Thus, without any restriction on μ, we must have

$$s \geq \frac{n}{\alpha N^2} \cdot \tag{38}$$

IV. Recall by an Incomplete Key

On account of the fact that information in a correlation matrix memory is stored in redundant form, the recalling of a stored item can also be made by an incomplete key which has a high correlation with the key used during memorization. To show this fact by a specific example, we define deterministic *projection parameters* $P_i \in \{0, 1\}$ which for known elements of $\boldsymbol{q}^{(r)}$ are 1 and otherwise 0. The memorization is still defined by (1) but during recall, the key vector has the components

$$q_i^{(r)} = P_i q_i^{(p_o)}, \qquad p_o \in P. \tag{39}$$

Then we have for an ICMM

$$\begin{aligned}\hat{x}_j^{(r)} &= c \sum_p \sum_{i=1}^m z_{ij} P_i q_i^{(p)} x_j^{(p)} q_i^{(p_o)} \\ &= \left[c \sum_{i=1}^m z_{ij} P_i q_i^{(p_o)} q_i^{(p_o)} \right] x_j^{(p_o)} \\ &\quad + c \sum_{p \neq p_o} \sum_{i=1}^m z_{ij} P_i q_i^{(p)} q_i^{(p_o)} x_j^{(p)}. \end{aligned} \tag{40}$$

The expectation value of recalled data is

$$E(\hat{x}_j^{(r)}) = c\mu \sum_{i=1}^{m} P_i (q_i^{(p_o)})^2 x_j^{(p_o)} + c\mu \sum_{p \neq p_o} \sum_{i=1}^{m} P_i q_i^{(p)} q_i^{(p_o)} x_j^{(p)} \quad (41)$$

and the variance is

$$\mathrm{var}\,(\hat{x}_j^{(r)}) = c^2\mu \sum_{i=1}^{m} P_i \left(\sum_p q_i^{(p)} q_i^{(p_o)} x_j^{(p)} \right)^2. \quad (42)$$

(Note $P_i^2 = P_i$.)

V. Auto-Associative Correlation Matrix Memory

There is nothing in the foregoing which would restrict us from implementing an auto-associative memory in which the pattern itself or any part of it could be used as the key. In this case during memorization $q^{(p)} = x^{(p)}$ and thus

$$M_{xq} = c \sum_p \mathbf{x}^{(p)} \mathbf{x}^{(p)T} \quad (43)$$

is the correlation matrix. Every memory element is now connected to two different elements of an input field. (The key and data fields are thus mixed up.) During recall, $q^{(r)}$ is replaced by a part of $x^{(p_0)}$. The projection parameters P_i are now 1 for the known elements of the key pattern and otherwise 0. Then for the CCMM

$$\hat{x}_j^{(r)} = c \sum_p \sum_{i=1}^{m} P_i x_i^{(p)} x_j^{(p)} x_i^{(p_o)}. \quad (44)$$

For the randomly generated ICMM

$$\hat{x}_j^{(r)} = c \sum_p \sum_{i=1}^{m} P_i z_{ij} x_i^{(p)} x_j^{(p)} x_i^{(p_o)}. \quad (45)$$

Now we obtain for the recalled data in ICMM the expectation value

$$E(\hat{x}_j^{(r)}) = c\mu \sum_{i=1}^{m} P_i (x_i^{(p_o)})^2 x_j^{(p_o)} + c\mu \sum_{p \neq p} \sum_{i=1}^{m} P_i x_i^{(p)} x_j^{(p)} x_i^{(p_o)} \quad (46)$$

and the variance

$$\mathrm{var}\,(x_j^{(r)}) = c^2\mu \sum_{i=1}^{m} P_i \left(\sum_p x_i^{(p)} x_j^{(p)} x_i^{(p_o)} \right)^2 \quad (47)$$

where

$$\mu = \frac{s}{n(n-1)}.$$

s Number of associators.
n Number of input elements.

If all memorized patterns, or at least the key parts of them are orthogonal, we can recall a pattern by its part with true conformity, and the noise is given by (47). Unless the key parts of the patterns are orthogonal, there will be biased crosstalk from other patterns given by the last terms in (46). Notice that if there are many zeros in the pattern, the requirement of approximate orthogonality is usually fulfilled.

VI. Unsupervised Learning in the ICMM

When the key vectors contain many zero components, there will be an appreciable probability that a large number of the matrix elements M_{ji} are zero or very small. In the recall, the contribution of these elements is small and, therefore, we may guess that without causing appreciable errors, all elements smaller than a certain limit could be deleted from the set of associators. Now we can stochastically generate a corresponding amount of new connections (associators) and obviously we will have found new large matrix elements of the complete correlation matrix. Using these for memorization and recall, we can deduce, e.g., from Example 5 that the relative noise will have been reduced. By intermittently breaking old connections and generating new ones during the memorization, unsupervised learning in the stochastically generated memory takes place. Achievable results depend strongly on the statistics of key vectors.

Example 5: Let us take an illustrative simplified example, the memorization and recall of a repeating single pattern. The unnormalized correlation matrix of a single pattern is (with $c = 1$)

$$M_{xq} = \mathbf{x}\mathbf{q}^T. \quad (48)$$

Let us assume that q_i, $x_j \in \{0, 1\}$ and denote the relative fraction of 1's in the key by α and in the data by β. We shall inspect the recall of a component x_j:

$$\hat{x}_j = \sum_{i=1}^{m} z_{ij} q_i^2 x_j. \quad (49)$$

In the beginning,

$$E(\hat{x}_j) = m\alpha\mu x_j \quad (50)$$

$$\mathrm{var}\,(\hat{x}_j) = m\alpha\mu x_j. \quad (51)$$

(Note that $x_j^2 = x_j \in \{0, 1\}$.)

Now we start a process of rejecting all $M_{ji} = 0$. For further simplicity, we clear the contents of the surviving M_{ji}, generate new connections, and repeat the memorization of (q, x). At the νth stage of this process, denote the number of associators with the contents $M_{ji} = 1$ by γ^ν ($\gamma^0 = \alpha\beta s$). At the $(\nu+1)$th stage, we reject $s - \gamma^\nu$ associators and generate an equal number of new ones. Of these, $\alpha\beta(s - \gamma^\nu)$ will hit places at which the complete correlation matrix would have 1's. Therefore,

$$\gamma^{\nu+1} - \gamma^\nu = \alpha\beta(s - \gamma^\nu). \quad (52)$$

The solution of (52) reads for M processes,

$$\gamma^M = [1 - (1 - \alpha\beta)^{M+1}]s. \quad (53)$$

Here

$$\lim_{M \to \infty} \gamma^M = s. \quad (54)$$

Finally, after infinitely many renewing processes, all associators will have connections for which the complete correlation matrix would have 1's. There are now s elements distributed over $\alpha\beta mn$ possible places corresponding to components of q that are 1. Therefore, the mean and variance of a recalled pattern at j are, respectively,

$$E(\hat{x}_j) = \alpha m \cdot \frac{s}{\alpha\beta mn} x_j = \frac{s}{\beta n} x_j \quad (55)$$

$$\text{var}\,(\hat{x}_j) = \frac{s}{\beta n} x_j. \quad (56)$$

Thus, comparing (50), (51), (55), and (56), we see that the relative standard deviation has been reduced by a factor $\sqrt{\alpha\beta}$. This example shows us that the incomplete correlation matrix memory is capable of unsupervised learning, and the result is the better, the more zeros there are in M_{xq}. On the other hand, the speed of learning is slower in this case.

VII. Use of Correlation Matrix Memories for Pattern Classification

By the classification of patterns we mean that a set of patterns consists of subsets each of which is mapped on a single element. Classification with the aid of correlation matrix memories means here that we assign one element j in the x field to each occurring class and the patterns to be classified are used as the q vectors. The memory is "taught" by assigning a value $x_h = 1$ for the element h if the exposed figure belongs to the class h, but 0 otherwise (supervised learning). During recalling, a pattern q is used as the key and $\hat{x}$ is formed as before. The class h is now found by the decision

$$\hat{x}_h^{(r)} = \max_j x_j^{(r)}. \quad (57)$$

VIII. Computer Simulations

The noise level of recalled patterns, due to the random structure of the associative network, can easily be computed from (18), (34), (42), and (47) for any type of memorized patterns. It might be desirable to have a direct demonstration of the reliability and efficiency of these memories. Therefore, some computer simulations on the auto-associative model have been performed and are shown in Fig. 2. The number of input elements was 140 in this experiment and these elements formed a two-dimensional retina; the number of memory elements was 4000 and every element was connected to two elements of the input retina using a random number generator for the designation of connections. The source data consisted of binary patterns. The stored items were recalled by key patterns using the transformation defined by (45). To standardize the output signals, all $\hat{x}_j^{(r)} \geq 5.5$ were displayed as binary 1's whereas smaller output signals were displayed as binary 0's. In Fig. 2(c) and (e), the key pattern had a high correlation with a stored item so that the stored information was reconstructed as shown in Fig. 2(d) and (f), respectively. The key pattern used in Fig. 2(g) had a low correlation with stored patterns so that only minor crosstalk from memorized items is present. The missing portions of the recalled patterns in Fig. 2(d) and (f) are due to the incompleteness of the correlation matrix. This effect manifests itself in all finite networks but its contribution decreases with increasing size of the memory.

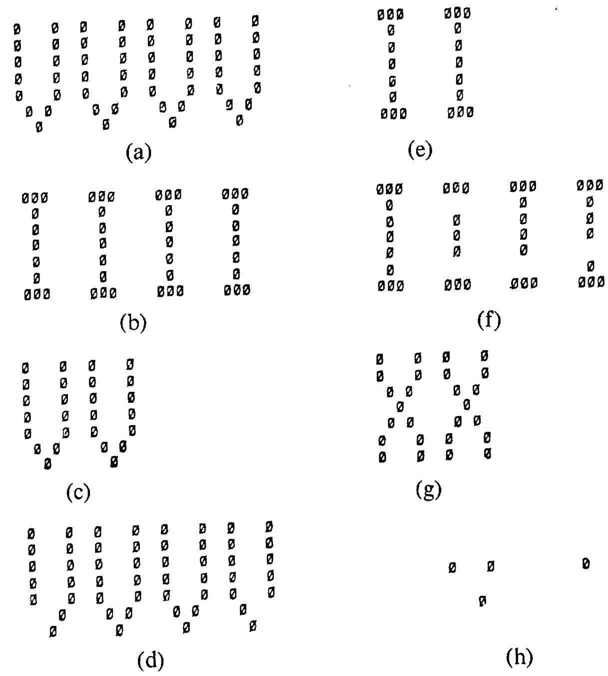

Fig. 2. (a) First memorized pattern, shown in a retina of 140 elements. Signals denoted by the symbol ø have the value 1 whereas signals denoted by blanks have the value 0. (b) Second memorized pattern, the memory traces of which have been superimposed on those of case (a). (c) First key pattern used for recall when the same symbols as in case (a) have been used. (d) Result of recall when all $\hat{x}_j^{(r)} \geq 5.5$ have been denoted by the symbol ø, and recalled signals smaller than 5.5 by a blank. (e) Second key pattern used for recall when the same symbols as in case (a) have been used. (f) Result of recall when the same symbols as in case (d) have been used. (g) Third key pattern which is uncorrelated with the memorized items. (h) Result of recall when the same symbols as in case (d) have been used.

Acknowledgment

The author wishes to thank M. Kilpi for programming the computer simulations.

References

[1] P. J. van Heerden, "A new optical method of storing and retrieving information," *Appl. Opt.*, vol. 2, pp. 387–392, 1963.
[2] ——, "Theory of optical information storage in solids," *Appl. Opt.*, vol. 2, pp. 393–400, 1963.
[3] ——, *The Foundation of Empirical Knowledge, with a Theory of Artificial Intelligence.* Wassenaar, Netherlands: Wistik, 1968.
[4] R. J. Collier, "Some current views on holography," *IEEE Spectrum*, vol. 3, pp. 67–74, July 1966.
[5] H. C. Longuet-Higgins, "Holographic model of temporal recall," *Nature*, vol. 217, p. 104, 1968.
[6] P. T. Chopping, "Holographic model of temporal recall," *Nature*, vol. 217, pp. 781–782, 1968.
[7] D. J. Willshaw and H. C. Longuet-Higgens, in *Machine Intelligence*, vol. 4, B. Meltzer and D. Michie, Eds. Edinburgh, Great Britain: Edinburgh University Press, 1969, pp. 349–357.
[8] D. Gabor, "Associative holographic memories," *IBM J. Res. Develop.*, vol. 13, pp. 156–159, 1969.

[9] I. T. Diamond and K. L. Chow, in *Psychology, A Study of a Science*, vol. 4, S. Koch, Ed. New York: McGraw-Hill, 1962, pp. 158–241.
[10] E. B. Carne, *Artificial Intelligence Techniques.* Washington, D. C.: Spartan 1965, p. 18.
[11] P. R. Westlake, "The possibilities of neural holographic processes within the brain," *Kybernetik*, vol. 7, pp. 129–153, 1970.
[12] K. Steinbuch, "Die Lernmatrix," *Kybernetik*, vol. 1, pp. 36–45, 1961.
[13] ——, *Automat und Mensch.* Berlin: Springer-Verlag, 1963, pp. 213–241.
[14] D. J. Willshaw, O. P. Buneman, and H. C. Longuet-Higgins, "Non-holographic associative memory," *Nature*, vol. 222, pp. 960–962, 1969.
[15] D. J. Willshaw and H. C. Longuet-Higgins, in *Machine Intelligence*, vol. 5, B. Meltzer and D. Michie, Eds. Edinburgh, Great Britain: Edinburgh University Press, 1970, pp. 351–359.
[16] A. Rapoport, "Addition and multiplication theorems for the input of two neurons converging on a third," *Bull. Math. Biophys.*, vol. 13, pp. 179–188, 1951.
[17] F. Jenik, in *Ergebnisse der Biologie*, H. Autrum, Ed. Berlin: Springer-Verlag, 1961, pp. 206–245.
[18] ——, in *Neural Theory and Modelling*, R. F. Reiss, Ed. Stanford, Calif.: Stanford University Press, 1964, pp. 190–212.
[19] K. Küpfmüller and F. Jenik, "Ueber die Nachrichtenverarbeitung in der Nervenzelle," *Kybernetik*, vol. 1, pp. 1–6, 1961.
[20] T. Kohonen, "A class of randomly organized associative memories," *Acta Polytech. Scandinavica*, no. El 25, 1971.

Dynamic Memories with Enhanced Data Access

HAROLD S. STONE, MEMBER, IEEE

Abstract—Dynamic memories are commonly constructed as circulating shift registers, and thus have access times that are proportional to the size of memory. When each word in a dynamic memory is connected to r words, $r \geq 2$, access time can be proportional to the base r logarithm of the size of memory. A memory that achieves minimum access time for $r = 2$ is described. The memory can also be operated in an efficient binary search mode. Slight variations of the interconnection patterns lead to a memory that is well suited for FFT and certain matrix computations.

Index Terms—Access time, binary search, dynamic memories, perfect shuffle, shift register memories.

I. Introduction

IN SOME MEMORY technologies, the storage medium inherently requires that there be a continuous circulation of data. Examples of such memories include magnetic drums and disks, MOS shift registers, and magnetic bubble memories. In this paper, we shall refer to such memories as *dynamic memories.*

For practical reasons, data movement in dynamic memories is normally cyclic. In the case of the magnetic drum, data are stored on the circumference of the drum, so that the rotation of the drum relative to a fixed head produces the cyclical movement of the data. MOS shift register memories are commonly constructed as circulating shift registers although there is no constraint that forces such memories to use the cyclic interconnection pattern.

Given the constraint that data must be moved continuously in a dynamic memory, the cyclical structure of the memory causes difficulty in achieving simultaneously both a large storage capacity and a short access time. In a cyclic memory, the access time to a randomly selected item increases linearly with the size of the memory. In this paper we investigate dynamic memories in which access time increases logarithmically with the size of memory. In particular, we embed an interconnection pattern called the *perfect shuffle* into the memory, and dispense with the more usual cyclic interconnection pattern. For the purposes of this paper we assume it takes a unit time to move data from one position to the next position in a cyclic memory, and that it also takes a unit time to permute data in memories that use noncyclic interconnections such as the perfect shuffle. This assumption is invalid for technologies in which noncyclic interconnections require data transfer times somewhat longer than a unit time due to larger physical transfer distances.

In Section II we derive the lower bound on access time that can be achieved in dynamic memories with enhanced interconnections. In Section III, we describe a dynamic memory which actually meets this bound. By modifying the control of this memory, it can also be used in a search mode with an efficiency that rivals the efficiency of random access memories. The search mode of operation is discussed in Section IV. Another type of shift register, which is de-

Manuscript received March 25, 1971; revised July 22, 1971. This work was supported by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract N00014-67-A-0112-0044, and by the National Science Foundation under Grant GJ-1180.

The author is with Digital Systems Laboratory, Stanford Electronics Laboratories, Stanford University, Stanford, Calif. 94305.