

This is a repository copy of *A study of pattern recovery in recurrent correlation associative memories*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/1999/>

Article:

Hancock, E.R. orcid.org/0000-0003-4496-2028 and Wilson, R.C. (2003) A study of pattern recovery in recurrent correlation associative memories. *IEEE Transactions on Neural Networks*. pp. 506-519. ISSN 1045-9227

<https://doi.org/10.1109/TNN.2003.811559>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

A Study of Pattern Recovery in Recurrent Correlation Associative Memories

Richard C. Wilson and Edwin R. Hancock

Abstract—In this paper, we analyze the recurrent correlation associative memory (RCAM) model of Chiueh and Goodman. This is an associative memory in which stored binary memory patterns are recalled via an iterative update rule. The update of the individual pattern-bits is controlled by an excitation function, which takes as its argument the inner product between the stored memory patterns and the input patterns. Our contribution is to analyze the dynamics of pattern recall when the input patterns are corrupted by noise of a relatively unrestricted class. We make three contributions. First, we show how to identify the excitation function which maximizes the separation (the Fisher discriminant) between the uncorrupted realization of the noisy input pattern and the remaining patterns residing in the memory. Moreover, we show that the excitation function which gives maximum separation is exponential when the input bit-errors follow a binomial distribution. Our second contribution is to develop an expression for the expectation value of bit-error probability on the input pattern after one iteration. We show how to identify the excitation function which minimizes the bit-error probability. However, there is no closed-form solution and the excitation function must be recovered numerically. The relationship between the excitation functions which result from the two different approaches is examined for a binomial distribution of bit-errors. The final contribution is to develop a semiempirical approach to the modeling of the dynamics of the RCAM. This provides us with a numerical means of predicting the recall error rate of the memory. It also allows us to develop an expression for the storage capacity for a given recall error rate.

Index Terms—Associative memory, error rates, recurrent correlation associative memory (RCAM), storage capacity.

I. INTRODUCTION

CORRELATION memories have proved to be powerful tools for binary pattern recognition and have been studied in the literature for over 40 years. When presented with an input pattern, the recall or output state of the memory is adjusted on the basis of similarity with a set of stored memory patterns. This idea can be traced back to the “lernmatrix” of Steinbuch [1]. Moreover, Willshaw’s early work on memory capacity has had a seminal impact on the field. There have since been a host of refinements and developments of the idea [5]–[8], [12], [3]. The memory may be operated in both autoassociative and heteroassociative modes. Of particular relevance to this paper is the recurrent correlation associative memory (RCAM) [6]. This is an autoassociative memory in which updating of the output state is specified by two ingredients. The first of

these is a measure of the similarity between the input pattern and the individual memory patterns. This measure is usually taken to be the inner product of the pattern-vectors, although tensor-products have also been used. The second ingredient is an excitation function which controls the step-size associated with different levels of pattern similarity. In fact, when the excitation function is the identity function, then the RCAM is equivalent to the Hopfield memory [3]. When the excitation function is the exponential then an interesting structure called the exponential correlation associative memory (ECAM) results [6]. This memory may be operated with both binary and multivalued pattern-vectors [16]. Recently, Hancock and Pelillo [20] have shown how both the Hopfield memory and the ECAM can be viewed as performing maximum likelihood pattern reconstruction via discrete relaxation operations [21]. A closely related structure is the bidirectional exponential associative memory [17], [18]. Finally, it is important to stress that the RCAM is of practical interest because of its suitability for very large-scale integration (VLSI) implementation [5].

Although the RCAM is functionally identical to the Hopfield memory when the excitation function is the identity, it is important to note that the philosophy underpinning the two memories is very different. In a Hopfield memory which stores M patterns each of length N bits, the memory patterns reside in an $N \times N$ matrix of weights. As a result N^2 weights are required to specify the memory. In contrast, the RCAM stores all of the M memory patterns as individual vectors and must explicitly compute the full set of inner products when a pattern is presented for recall. The model, therefore, requires MN bits of storage. Clearly, if the maximum number of patterns that can be stored in a Hopfield memory is $M < N$, then the RCAM model is more compact. There is also the added consideration of the overheads associated with amount of space required to store the memory patterns (memory data).

The amount of information that can be stored in a given number of bits of memory is clearly one of critical questions with regard to associative memories. One of the most popular approaches is to analyze the storage capacity, that is to say the number of patterns that can be stored in the memory before the recall quality is compromised. However, this approach does not address the issue of the density of information contained in the memory. A more satisfactory measure is the information density [24], or the number of pattern bits that can be stored per bit of memory data. For the Hopfield memory, the storage capacity is rather poor, being typically much less than the length of the pattern-vector [4]. For instance, McElice *et al.* [23] have shown that all of the patterns stored in the Hopfield memory are recoverable provided that $M < (N/4 \log N)$ and that that the

Manuscript received August 13, 2001; revised February 19, 2003. The work of R. C. Wilson was supported by an EPSRC Advanced Fellowship.

The authors are with the Department of Computer Science, University of York, York YO10 5DD, U.K.

Digital Object Identifier 10.1109/TNN.2003.811559

target memory pattern is within a Hamming distance $H = N/2$ of the input pattern. However, other studies [24], [28], [26] have suggested a limit of $M \simeq 0.14N$ based on vanishingly small levels of noise. The density of memory bits in this case is, therefore, 0.14 bits per synapse [24]. Since the synaptic weights of the Hopfield memory require approximately $\log N$ bits of storage, the density of memory patterns is $0.14/\log N$ stored bits per memory bit. For comparison, the density of the RCAM is 1, and is superior in these terms.

Having established that the storage density of the RCAM is larger than that of the Hopfield memory, the important question which arises is that of how many of the M memory patterns can be uniquely recalled? There is a suggestion that the exponential version of the memory (ECAM) can store a number of patterns which is exponential in the length of the pattern vector [6]. However, the analysis of Chiueh and Goodman requires that the noise-corrupted input pattern must have a Hamming distance that is closer to the target memory pattern than any of the remaining memory patterns. It is our aim here to examine the performance of the RCAM under more general models of noise. However, to approach this problem, we must abandon the idea of perfect pattern recall. Since the input pattern may be arbitrarily far from the target memory pattern, recall errors will always occur at some rate. As a result, it is the recall error rate which is of importance in the analysis of the memory.

Our overall aim in this paper is to provide an analysis of the problem of pattern recall in the RCAM. We adopt a probabilistic framework where we model the distribution of inner-products for noise corrupted pattern vectors. We show that the recall performance of the memory is governed by two probability distributions. The first of these is a model which describes distribution of patterns which lead to the correct recall of the target memory pattern. We refer to this component as the foreground distribution. The second model component describes the distribution of bit errors associated with the recall of patterns other than the target from the memory. We refer to this second process as background.

Because of the complexity of the iterative process which governs the recall of patterns from the RCAM, we can not model the foreground and background distributions at arbitrary epochs. However, we can draw some conclusions about the performance of a memory from the initial update step. Using the two component model we consider various ways in which the degree of pattern discrimination may be used to assess the recall performance of the RCAM. We investigate two alternatives. The first of these is the Fisher discriminant, the second is the probability of bit-errors on the recall pattern. With these two heuristic performance measures to hand, we can attempt to identify the excitation function which results in the best performance with respect to these measures. This is achieved by searching for the excitation function which optimizes each of our two performance measures. When recall performance is gauged using the bit-error probability, then the optimal excitation function can only be recovered in numerical form. The disadvantage is that the excitation function must be recalculated for each different configuration of the memory. On the other hand, when the Fisher discriminant is used, then the optimal excitation function may be estimated in closed-form.

Moreover, when the input bit-errors follow a binomial distribution, then the excitation function is the exponential.

Our final contribution is to present a semiempirical approximation which can be employed to find an expression for the storage capacity corresponding to a specified recall error rate. This also furnishes us with a numerical method for predicting the error rate for given memory.

The outline of this paper is as follows. Section II describes the process of pattern recall in the RCAM. Section III describes our two-component pattern model and approximations which simplify the pattern distributions. In Section IV, we consider how the probability of bit errors after one iteration can be used as a measure of pattern recall performance. We also show how this performance measure can be optimized with respect to the excitation function. Unfortunately, the problem is not tractable in closed form and the optimal excitation must be recovered numerically. For this reason, in Section V we develop a simpler performance measure which is reminiscent of the Fisher discriminant measure. We demonstrate that the simpler performance measure is optimized by an exponential excitation function. Section VI provides experiments with the two excitation functions and develops an empirical model for estimating the storage capacity under conditions of fixed output error rate. Finally, Section VII offers conclusions and suggests directions for future investigation.

II. PATTERN RECALL IN THE RCAM

In this paper, we are interested in the the RCAM. This is an associative memory that can be used for binary pattern recognition. From a computational standpoint, the memory can be used to iteratively recall previously stored binary patterns when presented with their perturbed or noisy variants as input. Suppose that $\mathbf{x}^t = (x_1^t, x_2^t, \dots, x_N^t)^T$ is the pattern vector residing on the input of the memory at iteration t . The pattern-vector is of length N and each component x_i^t is drawn from the binary set $\Omega = \{-1, +1\}$. At iteration $t+1$, the updated realization of the recalled pattern-vector is $\mathbf{x}^{t+1} = (x_1^{t+1}, x_2^{t+1}, \dots, x_N^{t+1})^T$. The RCAM contains Z stored memory patterns which are available for recall. Each memory pattern is also a binary vector of length N . If d is the pattern index, then we denote the contents of the memory by the set of binary pattern-vectors $D = \{\mathbf{s}^{(d)}; d = 1, \dots, Z\}$.

The updating of the recalled pattern vectors is controlled by two factors. The first of these is the similarity between the current recall-state of the memory and each of the stored memory patterns. The similarity between the current output pattern and the individual memory patterns, is measured using the inner-product, or dot product, $\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle$ between the pattern-vectors. The dot product is related to the number of bit-differences or Hamming distance

$$H(\mathbf{x}^t, \mathbf{s}^{(d)}) = \frac{1}{2} [N - \langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle].$$

The Hamming distance is important, since it will be used to simplify the description of the probability distributions in a later section of the paper.

The second component of the pattern update procedure is an excitation function $f(\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle)$. The excitation function controls the contributions that the individual memory patterns make to the overall update direction for the recall state of the memory. In other words, the excitation function provides a distance dependant weight for each of the memory patterns.

With these ingredients the recall pattern is iteratively modified according to the following update rule [5]:

$$\mathbf{x}^{t+1} = \text{sgn} \left\{ \sum_{d \in D} [f(\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle) - \bar{f}] \mathbf{s}^{(d)} \right\} \quad (1)$$

where

$$\bar{f} = \frac{1}{Z} \sum_{d \in D} f(\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle)$$

is the mean value of the excitation function over the set of memory patterns. This is in addition to the update rule in [5], and is intended to compensate for constant shifts in f , i.e., $f(x) \rightarrow f(x) + c$ has no effect on the memory. When the number of positive and negative bits is the same, the operation is identical to [5].

The memory is *autoassociative* in the sense that it recalls an output pattern which is close to, but not necessarily identical to, one of the memory patterns. The excitation function f is the key to the recall performance of the RCAM. Its role is to determine the weight given to different values of the inner-product between the output state of the memory and the stored memory patterns. For example, the choice $f(t) = t$ results in the Hopfield network [3]. In contrast, the choice $f(t) = \exp(kt)$ (where k is a positive real-valued constant) results in the ECAM model.

It has been shown [5] that given a suitable choice for f , then the stored memory patterns are the fixed-points of the RCAM update process. Therefore, provided that the input patterns are sufficiently close to the memory patterns, then the update process will converge to one of the memory patterns. The main requirement is that $f(\cdot)_{[-N, N]}$ is monotonically increasing. More importantly to the goals of this paper, the choice of the excitation function f is critical in determining the recall performance of the RCAM.

To develop a statistical model of the memory, the set of stored pattern $D = D_i^+ \cup D_i^-$ is subpartitioned into two subsets D_i^+ and D_i^- . The partition D_i^+ contains the items in which the i^{th} component of the memory pattern takes on the value $s_i = +1$. Similarly D_i^- contains those patterns whose i^{th} component takes on the complementary value $s_i = -1$. With this notation, the update process is captured by the two decision variables

$$\begin{aligned} W_i^+(t) &= \sum_{d \in D_i^+} f(\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle) - |D_i^+| \bar{f} \\ W_i^-(t) &= \sum_{d \in D_i^-} f(\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle) - |D_i^-| \bar{f}. \end{aligned} \quad (2)$$

The components of the recall output-vector at the site indexed i are updated as follows:

$$x_i^{t+1} = \begin{cases} +1, & \text{if } W_i^+(t) > W_i^-(t) \\ -1, & \text{otherwise} \end{cases}. \quad (3)$$

Suppose that the observed input pattern \mathbf{x}^t is generated by the target memory pattern $\mathbf{s}^{(o)}$. Since \mathbf{x}^t is a noise corrupted realization of $\mathbf{s}^{(o)}$, we must admit the possibility that \mathbf{x}^t may in-fact appear to be more similar to one of the remaining, alternative, patterns stored in the memory. In this situation, iterative recovery of the original pattern may be impossible. It is for this reason that the recall from the memory will never be completely error-free. Here we aim to relate the storage capacity of the memory to the error-rate for the recovery of patterns.

Without any loss of generality, we will study the update of the bit i when $s_i^{(o)} = +1$. In other words, the correct bit-assignment at site i is $+1$. Under these circumstances, we can write

$$\begin{aligned} W_i^+(t) &= \left\{ \sum_{d \in D_i^{*+}} f(\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle) \right\} - (|D_i^+| - 1) \bar{f} \\ &\quad + f(\langle \mathbf{x}^t, \mathbf{s}^{(o)} \rangle) - \bar{f} \\ W_i^-(t) &= \left\{ \sum_{d \in D_i^-} f(\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle) \right\} - |D_i^-| \bar{f} \end{aligned} \quad (4)$$

where $D_i^{*+} = D_i^+ - \{\mathbf{s}^{(o)}\}$ is the set of memory patterns that assign the bit $+1$ to the site i when the target pattern $\mathbf{s}^{(o)}$ is excluded. In the next section we will examine the probability distributions of these two decision variables.

III. PATTERN MODEL

As we discussed earlier, we wish to consider situations where the observed pattern-vector \mathbf{x}^t is derived from a stored memory pattern through the action of a bit-error process. The bit-corrupted pattern is presented to the memory for recall. In order to model the action of the RCAM, we must first provide a model of the bit-error process, which is responsible for generating the observed pattern-vector from the relevant memory pattern. If this pattern corruption model is known, then we can attempt to design an excitation function f that endows the RCAM with good recall properties. This endeavour is the focus of attention in the latter part of this paper.

Since we intend to pursue a probabilistic analysis of the memory, we must define some statistical properties of the stored memory patterns. Obviously, the population statistics of the stored pattern sets can vary and this in turn will significantly affect recall. With this caveat in mind, we confine our attention to sets of stored memory patterns that satisfy the following restrictions.

- The number of stored patterns is large and the probability distributions governing the bit-configurations are well behaved.
- The memory pattern-vectors are independently identically distributed (i.i.d.) random variables. In other words, the memory pattern vectors are sequences of random variables governed by an identical probability distribution.

When this is the case, we can apply the central limit theorem to the population statistics of the stored patterns, and pursue a set of approximations which will allow us to determine probability distributions for the pattern classes. It is important to be aware

that the applicability of the central limit theorem depends on the distribution of $f(\langle \mathbf{x}^t, \mathbf{s} \rangle)$. In particular, this distribution must satisfy the Lindeberg condition [29].

Our model of the pattern corruption process focusses on the distribution of the quantities $W_i^+(t)$ and $W_i^-(t)$. We can invoke the central limit theorem to determine the distributions of $W_i^+(t)$ and $W_i^-(t)$ since they are the sums of a large number of identically distributed random-variables. As a result, we can approximate the distribution of $W_i^-(t)$ by a Gaussian with mean zero and variance $\sigma_-^2 = |D^-|\sigma_f^2$. Similarly, $W_i^+(t)$ has a Gaussian distribution with mean $\Delta f = f(\langle \mathbf{x}^t, \mathbf{s}^{(o)} \rangle) - \bar{f}$ and variance $\sigma_+^2 = (|D^+| - 1)\sigma_f^2$. Again, provided that the memory patterns are large in number and randomly distributed, then we can make the following approximation:

$$\sigma_-^2 \approx \sigma_+^2 \approx \sigma^2 = \frac{|D|\sigma_f^2}{2}. \quad (5)$$

We can, therefore, write the probability distributions for the two decision variables using the Gaussian approximations

$$\begin{aligned} P(W_i^+(t) = w) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(w + \Delta f)^2}{2\sigma^2}\right] \\ P(W_i^-(t) = w) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{w^2}{2\sigma^2}\right]. \end{aligned} \quad (6)$$

These two distributions have identical variance. The distribution mean for the correct choice of bit assignment update is offset with respect to distribution for the incorrect choice, by an amount $f(\langle \mathbf{x}^t, \mathbf{s}^{(o)} \rangle) - \bar{f}$. It is this separation which makes it possible to determine which distribution gave rise to the current bit assignment. The offset and variance of the distributions determine the error probability of the bit update.

A. Error Probability After One Update

Once we have the approximate probability distributions for the bit assignment updates, we can calculate the probability of making an erroneous update at the site indexed i . The incorrect choice is made when $W_i^-(t) > W_i^+(t)$ and errors occur with probability $\nu = P(W_i^-(t) > W_i^+(t))$. This probability is equal to the cumulative overlap of the two distribution, i.e.,

$$P(W_i^-(t) > W_i^+(t)) = \int_{-\infty}^{\infty} dw \int_{-\infty}^w dv P(W_i^+(t) = w) P(W_i^-(t) = v). \quad (7)$$

Substituting for the two distributions from (7), we find

$$\begin{aligned} \nu(\sigma, \Delta f) &= \int_{-\infty}^{+\infty} dw \int_{-\infty}^w dv \frac{1}{2\pi\sigma^2} \exp\left[-\frac{v^2}{2\sigma^2}\right] \\ &\quad \cdot \exp\left[-\frac{(w - \Delta f)^2}{2\sigma^2}\right]. \end{aligned} \quad (8)$$

In the above expression, we have made explicit the fact that the error-probability ν is a function of the separation Δf between the means of the two distributions and their common variance σ^2 . We refer to the function ν as the *overlap function*. Its role is to quantify the amount of overlap between the background

and foreground distributions. By making the appropriate substitutions, we can re-write this error probability in terms of the single variable $\alpha = \Delta f/\sigma$ in the following manner:

$$\nu(\alpha) = \int_{-\infty}^{+\infty} dr \int_{-\infty}^r dq \frac{1}{2\pi} \exp\left[-\frac{q^2}{2}\right] \exp\left[-\frac{(r - \alpha)^2}{2}\right]. \quad (9)$$

It is clear that the quantity α represents the separation between the foreground and background distributions. Moreover, it is related to the Fisher class separation criterion for the probability distributions for the two decision variables. Since

$$\alpha = \frac{f(\langle \mathbf{x}^t, \mathbf{s}^{(o)} \rangle) - \bar{f}}{\sigma} \quad (10)$$

depends only on the excitation function f and the number of stored patterns $|D|$, it is these factors which determine the recall ability of the memory (given our background pattern model and subsequent assumptions). The overlap function also has the important property of being monotonically decreasing with increasing α . This means that increasing the value of α will decrease the probability of making an error when updating the bit assignments. In other words, the quantity α can be used as a heuristic measure of the recall ‘‘quality’’ of the memory. Finally, a good analytical approximation to $\nu(\alpha)$ s is

$$\nu(\alpha) = \frac{1}{2} \left[1 - \operatorname{erf} \left[\frac{\alpha}{2\sqrt{2}} \right] \right]. \quad (11)$$

Fig. 1 shows a typical examples of this function plotted as a function of the Hamming distance H_c between the original pattern and its current realization. The overlap function ν which governs the update process is a monotonically increasing function of α . This leads us to the conclusion that reducing α will lead to an overall improvement in the bit-error rate.

IV. MEASURING RECALL PERFORMANCE OF THE EXCITATION FUNCTION

The recall performance of the RCAM (or alternatively the storage capacity) has been studied previously in [5]. For the ECAM, Chieh and Goodman [5] find the storage capacity in the case where the input patterns are partially corrupted, but still closer in terms of dot-product to their original source than to any other memory pattern. There are, therefore, two limitations to these studies. First, each concentrates on a specific realization of the RCAM. In this section, we are interested in measuring the performance of RCAM with arbitrary excitation functions. Second, the restrictions on the properties of the input patterns quite limiting. In fact, in these cases, choosing the closest pattern would result in perfect recall.

In general, we would expect the memory to be exposed to a diversity of input patterns with varying levels of corruption. Viewed as a population, we can capture the collective properties of the input patterns using a probability distribution over the inner-product $\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle$. To further develop this idea, we introduce a two-component model of the distribution of inner products between the pattern vectors. The first component models the distribution of inner products between the target memory pattern and the perturbed or noise corrupted input pattern. Suppose that $q(\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle = u)$ is the noise distribution of inner-products for the source pattern indexed d . We assume that the

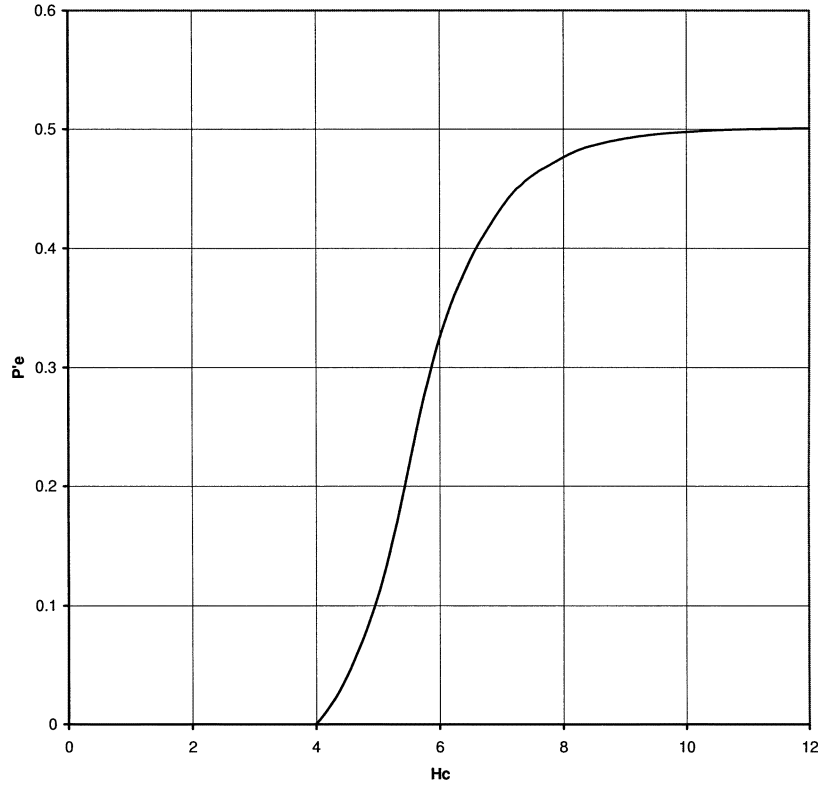


Fig. 1. Overlap function plotted against the Hamming distance H_c between the source pattern and its current realization.

source patterns are randomly distributed, i.e., there is nothing salient concerning the choice of index pattern d . The second model component involves additionally considering the distribution of the inner-product between the different patterns stored within the memory. This background distribution function is represented by $r(\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle = u)$.

Our problem is intrinsically discrete in nature since the inner product takes on integer values in the interval between $-N$ and N . Rather, than attempting to recover a continuously defined excitation function, we instead recover a discrete representation of the excitation function. Accordingly, we let $f_u = f(u)$ denote the value of the excitation function when the inner product takes on the integer value u drawn from the set $\{-N, -N + 1, \dots, N - 1, N\}$. Further, we use the shorthand q_u and r_u to denote the noise (or foreground) and background distribution functions $q(\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle = u)$ and $r(\langle \mathbf{x}^t, \mathbf{s}^{(d)} \rangle = u)$. With this notation

$$\bar{f} = \sum_{u=-N}^N r_u f_u.$$

While we cannot compute the final error rate for the output of the RCAM, one good heuristic is the expectation value of bit-error probability after one application of the RCAM. With the discrete problem representation adopted above, this quantity is given by

$$Q_f = \sum_{u=-N}^N q_u \nu(\alpha_u) \quad (12)$$

where

$$\alpha_u = \frac{(f(u) - \bar{f})}{\sigma}.$$

We can locate the optimal excitation function with respect to this measure by minimizing the value of Q_f with respect to the discrete representation of the excitation function f . One way of meeting this goal is to solve the set of stationary-point equations for the different values of f_u . This involves differentiating Q_f with respect to f_u for each value of $u \in \{-N, \dots, N\}$ in turn. The family of stationary-point equations is obtained by setting each of the resulting derivatives to zero. In other words

$$\frac{\partial Q_f}{\partial f_k} = 0 \quad (13)$$

for all values of $k \in \{-N, \dots, N\}$. To compute the required derivatives we exploit the chain rule and write

$$\frac{\partial Q_f}{\partial f_k} = \sum_{u=-N}^N q_u \frac{\partial \nu(\alpha_u)}{\partial \alpha_u} \frac{\partial \alpha_u}{\partial f_k}. \quad (14)$$

Substituting for $\nu(\alpha_u)$ and α_u , the set of stationary-point equations becomes

$$\frac{2}{\sqrt{\pi}} \sum_{u=-N}^N q_u \exp[-\alpha_u^2] \Delta f_u = 0. \quad (15)$$

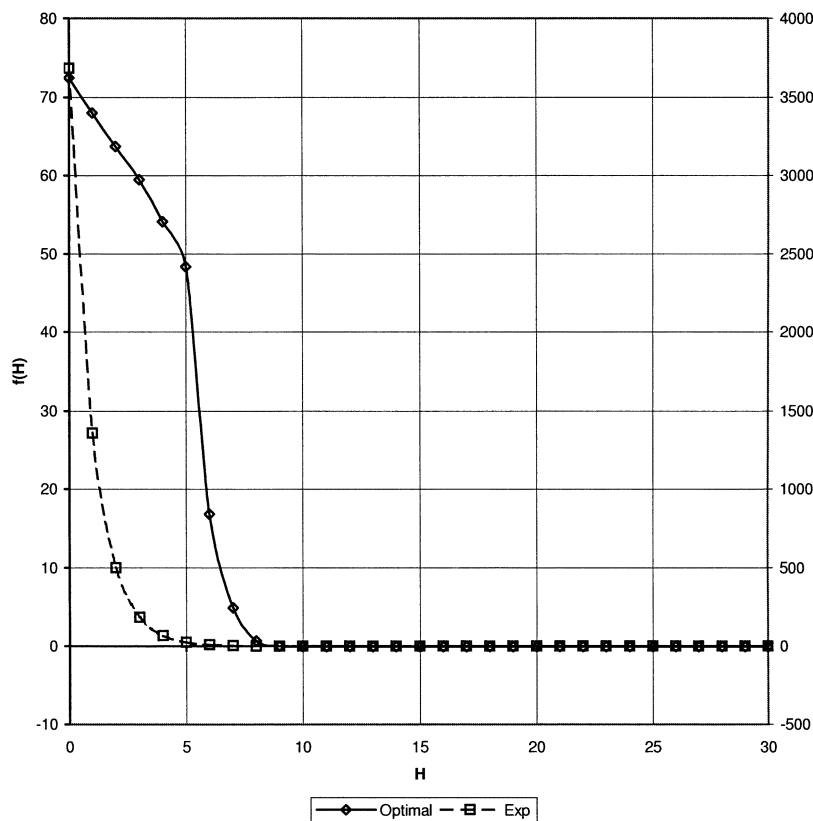


Fig. 2. Optimal functions for the RCAM. The exponential function (scale on right) maximizes Q_s , whereas the optimal function (left scale) minimizes the expectation of bit-error probability after one iteration.

We find that this condition is satisfied when

$$f_k = \frac{q_k}{r_k} \exp \left[-\frac{f_k^2}{\sigma^2} \right]. \quad (16)$$

Each value of f_k must be found by solving a transcendental equation. In fact the solution is not possible in closed-form and we must resort to numerical methods to determine the excitation function. The shape of this function is dependent on both the number of stored patterns and the number of bits in the pattern vector and, therefore, must be recalculated when these quantities change. An example of this function is shown in Fig. 2.

V. PATTERN SPACE MODELS

The main drawback to the use of the expectation-value of the bit-error probability as a measure of recall quality is that it depends on a series of error-functions and these in turn lead to a transcendental equation for the excitation function which involves exponentials. Although these transcendental equations can be solved numerically, we would like to explore a more tractable route to the excitation function. Our approach is as follows. We commence by simplifying the quality measure Q_f . This involves making a Taylor approximation to the error-functions. Next, we construct a pattern-space model for the distribution for the foreground and background inner-products.

We commence by considering the dependence of the bit-error probability $\nu(\alpha)$ on the quantity α . For small and large values of α the error probability, respectively, approaches unity and zero. The critical factor in determining the pattern recovery behavior

of the memory is the slope of the shoulder of the bit-error probability function. The shoulder of the function occurs when α is close to zero. Here, the error function can be approximated in the following linear way:

$$\operatorname{erf} z \simeq \frac{2}{\sqrt{\pi}} z. \quad (17)$$

Under this approximation, we can write

$$Q_f \simeq \frac{1}{2} \sum_{u=-N}^N q_u \left[1 - \frac{f_u - \bar{f}}{\sqrt{2\pi}\sigma} \right]. \quad (18)$$

Hence, we can turn our attention to locating the excitation function which maximizes the quantity

$$\hat{Q}_f = \frac{1}{\sigma} \sum_{u=-N}^N q_u (f_u - \bar{f}). \quad (19)$$

In other words, the measure of recall quality depends on the ratio of the separation of the foreground and background distributions patterns to their spread or width. The separation measure is, therefore, reminiscent of the Fisher discriminant.

Based on this observation, we turn our attention to finding the excitation function which maximizes the quadratic Fisher discriminant measure of pattern separation

$$Q_s = \sum_{u=-N}^N q_u \alpha^2 = \frac{1}{\sigma^2} \left[\sum_{u=-N}^N q_u (f_u - \bar{f}) \right]^2. \quad (20)$$

This performance measure is equal to the average squared offset between the foreground and background patterns, divided by the variance or squared width of the distribution W . To make these relationships explicit, we can rewrite the measure as

$$Q_s = \frac{S^2}{T} \quad (21)$$

where S is the average offset from the mean

$$S = \sum_{u=-N}^N q_u(f_u - \bar{f}) \quad (22)$$

and T is the variance of the background distribution

$$T = \sum_{u=-N}^N r_u(f_u - \bar{f})^2. \quad (23)$$

The quantity S^2 plays the role of a within-class variance for the input pattern vectors. The quantity T , on the other hand, can be viewed as the between-class variance. The quantity Q_s is, hence, closely related to the Fisher class separation measure from statistical pattern recognition. It should be noted that the quantity Q_s does not represent the performance of the memory in terms of storage capacity.

Table I lists the values of Q_s for some familiar functions. Here, we have used patterns of length 30 bits and have assumed that r_u follows a binomial distribution with a mean value of three bit-errors per pattern. These results suggest that the exponential makes a better choice than the linear function, in keeping with known storage capacity results.

As we mentioned earlier in the introduction, and central to the motivation of this paper, is the observation that the choice of excitation function is critical in determining the performance of the RCAM. For example, the Hopfield network is equivalent to an RCAM when the excitation function is linear in the inner product, i.e., $f(\langle \mathbf{x}, \mathbf{s}^{(d)} \rangle) = -\langle \mathbf{x}, \mathbf{s}^{(d)} \rangle$. It has been shown [4] that this memory has a rather poor storage limit of $0.14N$ patterns where N is the length of the bit-patterns. Other authors [11] have suggested that the exponential function results in an exponential storage capacity ($\propto e^{kN}$). By choosing the correct excitation function, it appears possible to greatly increase the storage capacity.

A. Optimal Excitation Function for Q_s

Given the separation criterion Q_s , we can attempt to find the excitation function which maximizes the separation between the foreground and background pattern distributions. We commence from the result proven in Appendix A, namely that a family of functions, not a single function, all give identical values of Q_s and, hence, identical performance for pattern recovery, subject to our model of pattern distributions. Moreover, in Appendix A we show that simplified recall performance measure Q_s (and, in fact, the overall performance of any RCAM) remains invariant under similarity transformation

$$f' = af + b \quad (24)$$

of the excitation function f . Here a and b are arbitrary constants. In order to define such a similarity transformation, we

TABLE I
VALUES OF Q_s

| Discriminant Function | Quality Factor Q_s |
|-----------------------|----------------------|
| t | 19 |
| e^{t^2} | 105273 |
| e^t | 120622 |

must specify two points which fall on the function. For the purpose of recovering the optimal excitation function, the choice $\bar{f} = 1$ and $f_0 = q_0/r_0$ is convenient.

To proceed, we now solve the saddle-point equations for Q_s with respect to the discrete function values f_k . The details are provided in Appendix B. For the simplified recall performance measure Q_s , the optimal excitation function is given by

$$f_u = \frac{q_u}{r_u}. \quad (25)$$

The general family of optimal functions is, therefore

$$f_u = a \frac{q_u}{r_u} + b. \quad (26)$$

The corresponding value of Q_s (the maximum possible value of Q_s) is given by

$$Q_{opt} = \sum_{u=-N}^N \frac{q_u^2}{r_u} - 1. \quad (27)$$

B. Pattern Space Models

To recover the optimal excitation function using the solution given in (25), we must provide concrete models of r_u and q_u . In other words, we must exploit a model of the pattern-space in which the RCAM operates. We consider the specific case in which the bit-errors on the input pattern are memoryless. Suppose that each bit of the stored pattern $\mathbf{s}^{(o)}$ is flipped with a uniform probability P_e . Under these circumstances, the distribution of Hamming distance is binomial, i.e.,

$$P_s(H(\mathbf{x}, \mathbf{s}^{(d)}) = h) = \frac{N!}{(N-h)!h!} P_e^h (1-P_e)^{N-h}. \quad (28)$$

Substituting for the definition of Hamming distance the distribution for the inner product between the input and the target memory pattern is

$$q(\langle \mathbf{x}, \mathbf{s}^{(d)} \rangle = u) = \frac{N!}{\frac{N+u}{2}! \frac{N-u}{2}!} [P_e(1-P_e)]^{N/2} \left[\frac{1-P_e}{P_e} \right]^{u/2}. \quad (29)$$

To describe the distribution of inner-product for the memory patterns we require model of the structure of the pattern space. The basic assumption is that the patterns are random configurations of bits. If we confine our attention to the case when the different bits occur with equal probability, we obtain the following distribution for Hamming distance:

$$P_B(H(\mathbf{x}, \mathbf{s}^{(d)}) = h) = \frac{N!}{2^N(N-h)!h!}. \quad (30)$$

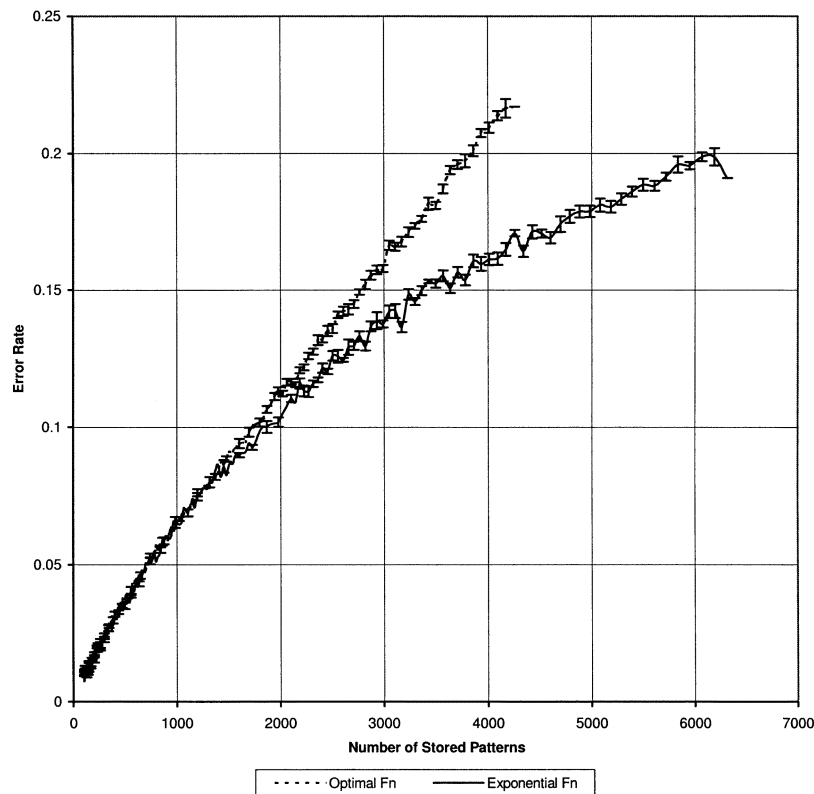


Fig. 3. Comparative recall error rates for the exponential and numerical excitation function.

Again substituting for the Hamming distance, the required distribution for the inner product is

$$r(\langle \mathbf{x}, \mathbf{s}^{(d)} \rangle = u) = \frac{N!}{2^N \frac{N-u!}{2} \frac{N+u!}{2}}. \quad (31)$$

With these two distribution models, the optimal excitation function is

$$f_u = [P_e(1 - P_e)]^{N/2} \left[\frac{1 - P_e}{P_e} \right]^{u/2} \frac{1}{2^N} \\ = \left[\frac{\sqrt{P_e(1 - P_e)}}{2} \right]^N \exp \left[\frac{u}{2} \ln \left[\frac{P_e}{1 - P_e} \right] \right]. \quad (32)$$

This is precisely the exponential function which is known to provide very good pattern recovery performance. In fact, we can state that the exponential excitation function works well because it maximizes the expectation value of the Fisher separation between the foreground and background pattern distributions. This excitation function applies only when the distribution of errors on the input pattern is binomial. It is important to stress, that after one iteration of the RCAM, the distribution of errors does not remain binomial, and the error probability P_e will change.

VI. COMPARISON OF THE EXCITATION FUNCTIONS

In this section, we provide some experimental evaluation of the proposed excitation functions. We commence by comparing the excitation function obtained by numerical optimization with the exponential function. Next, we study the overlap properties of the resulting memory. Finally, we present some results which show the error-rate achievable with the two excitation functions.

A. Optimal Functions

The more precise quality measure Q_f , which is related to the expectation value of the bit error probability, must be optimized by numerical means. Moreover, the solution to the associated equation varies with the number of patterns stored in the memory. For this reason, the resulting excitation function is less flexible than the exponential excitation function which results from the distribution analysis. In Fig. 2, we compare the two excitation functions. The two curves show the Hamming distance dependence of the excitation functions. The numerical approximation to the optimal function is computed for $P_e = 0.1$, $N = 30$ and with $Z = 400$ stored patterns. The numerical excitation function clearly differs from the exponential at lower values of the Hamming distance. It is interesting to note that the excitation is approximately linear and of negative slope for small Hamming distances. This is reminiscent of the Hopfield memory excitation function. However, in this region patterns are almost certain to converge to the correct stored pattern. The small Hamming distance behavior, therefore, has little impact on the error rate of the memory. At larger Hamming distances, the excitation functions decays almost exponentially.

To underline this point Fig. 3 compares the bit error rates for the numerical excitation function and the exponential excitation function. This quantity of interest is the bit-error rate for the recalled patterns, or in other words, the probability that any bit in a recalled pattern is in error. The bit error rate is plotted as a function of the number of stored patterns. The results have been obtained for patterns of length 30 bits. The input pattern noise is generated by randomly flipping bits with probability 0.1. For the numerical excitation function, the error rate increases lin-

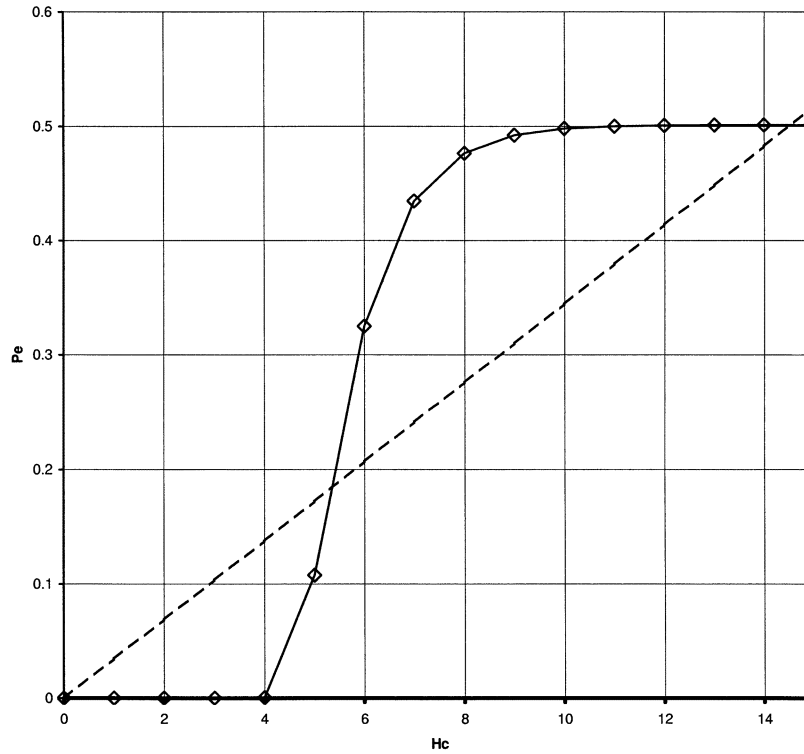


Fig. 4. Overlap function and the bit error rate before the first iteration of the RCAM.

early with the number of stored patterns. This function is tuned to $N = 30$ bits and $Z = 1000$ patterns. By contrast, for the exponential excitation function the bit error probability increases approximately with the square root of Z . When we operate the RCAM with the numerical excitation function, then it performs well provided that the appropriate value of Z is used. In fact, under these conditions it gives performance that is comparable to that obtained with the exponential excitation function. The exponential excitation function performs well for a much larger range of values of Z .

B. Predicting the Error Rate

As we discussed in Section III-A, the expectation value of the bit error probability after the first iteration of the RCAM is given by the overlap function $\nu(\alpha_u)$. We can write $\nu(\alpha_u)$ in the following analytical form:

$$\nu(\alpha_u) = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{\alpha_u}{2\sqrt{2}} \right) \right]. \quad (33)$$

Fig. 4 shows ν as function of Hamming distance. This function represents the bit-error transfer function from the original pattern to the updated pattern after one iteration of the RCAM.

We can determine the approximate error-rate of the memory using the following argument: Suppose that $\nu(\alpha_u)$ is probability of choosing the incorrect bit-assignment for an input pattern which has inner-product u with the desired target pattern. The expected number of bit-errors, i.e., the Hamming distance after application of the RCAM is given by $H' = N\nu(\alpha_u)$. As a result the expectation value of inner-product u after one iterative application of the memory is given by $u' = N[1 - 2\nu(\alpha_u)]$. Provided

that $u' < u$ then the action of the memory is to further corrupt the pattern, i.e., to increase the number of bit-errors or Hamming distance. In consequence, the memory will diverge from the target pattern and correct pattern recall will not occur. Further suppose that u_a is the break-even value of the inner-product which satisfies the equation

$$u_a = N[1 - 2\nu(\alpha_{u_a})]. \quad (34)$$

The error-probability is the cumulative area under the distribution function q_u up to this break-even point. More formally, the recall error-rate is given by

$$R = \sum_{u=-N}^{u_a} q_u. \quad (35)$$

Of course, q_u is a discrete function, and R only exists at the integer values of u_a . In order to evaluate the error-rate at noninteger values of u_a we generate a smoothly interpolated function R' which passes through the values of R . The predicted error rate is $R'(u_a)$. There is an obvious limitation of this analysis. The update process is governed by the distribution of errors and not by the expectation value of the bit-error probability. As a result, patterns that become increasingly corrupt during early iterations may be restored at later epochs. This effect becomes less apparent when the error function has a sharp shoulder, i.e., the transition is abrupt. This process is illustrated in Fig. 4. Patterns above the shoulder will (on average) diverge from the incorrect solution, whereas those below it converge to the correct solution after sufficient iterations of the RCAM update rule. Fig. 5 shows

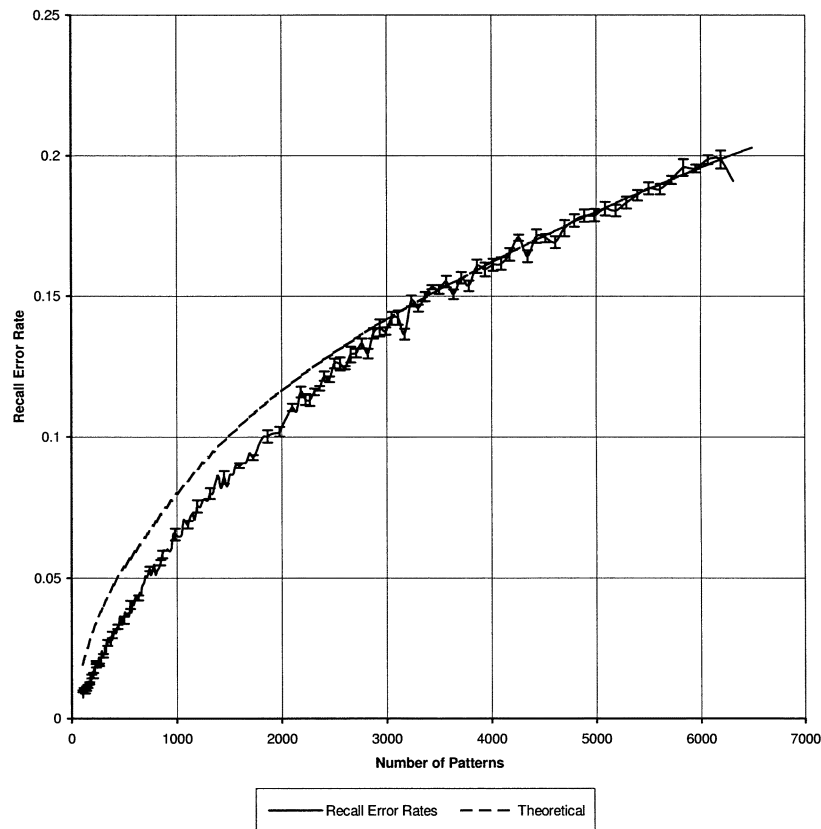


Fig. 5. Predicted and measured error rates for the exponential RCAM.

recall error rate plotted as a function of the number of stored patterns. The two curves are the observed recall error rate and the predicted recall error rate. In both cases the memory has an exponential excitation function and the RCAM has parameters $P_e = 0.1$ and $N = 30$. The predicted curve deviates from the actual error rate for low numbers of stored patterns because the Gaussian approximation is increasingly inaccurate in the region. At higher loading levels, the experimental and predicted values agree well.

C. Storage Capacity for a Fixed Error Rate

The purpose of the next experiment is to determine the relationship between the length of the pattern-vectors (N) and the number of stored patterns (Z) when the output error rate is fixed at some predetermined level. In other words, we assume that some low output error rate is acceptable and determine how many stored patterns we can accommodate for a given length of pattern-vector. This may be seen as the analog of the storage capacity studies of the Hopfield memory and the ECAM. Fig. 6 gives the results of such a study on the ECAM. The storage capacity is evaluated for an output bit-error rate of $r = 0.01$, and input patterns with random bit-flipping errors which occur with probabilities 0.05, 0.1, 0.125, and 0.15. Note the log-scale on the y -axis of the plot. These results show a very clear exponential relationship between Z and N .

We can develop a simple empirical model which provides further insight into the pattern recall process. For a predetermined error-rate $\nu(\alpha_{u_a})$, we can use (35) to determine the critical value

of the inner-product u_a . This value is plotted in Fig. 7 as a function of the pattern length N for $r = 0.01$.

From the plot, it is clear that there is an approximately linear relationship between U_a and the length of the pattern-vector N for the range of bit-pattern lengths studied. The equation of the best-fit line is $u_a = 0.1702N + 1.663$. Substituting this linear relationship into (34), we find that

$$1 - 2\nu(\alpha_{u_a}) = 0.6596. \quad (36)$$

Rewriting this equation, we find

$$Z = \left(\frac{f(0.6596N) - \bar{f}}{2\sigma_f \operatorname{erf}^{-1}(0.1702)} \right)^2. \quad (37)$$

For the ECAM, we can approximate the storage capacity by

$$Z = a \exp \left(bN \ln \left[\frac{1 - P_e}{P_e} \right] \right).$$

Using the data presented in Fig. 6, we find that $a = 0.043 \pm 0.009$ and $b = 0.117 \pm 0.003$. In other words, for an output error rate of $r = 0.01$, the storage capacity of the ECAM is

$$Z = 0.043 \exp \left(0.117N \ln \left[\frac{1 - P_e}{P_e} \right] \right).$$

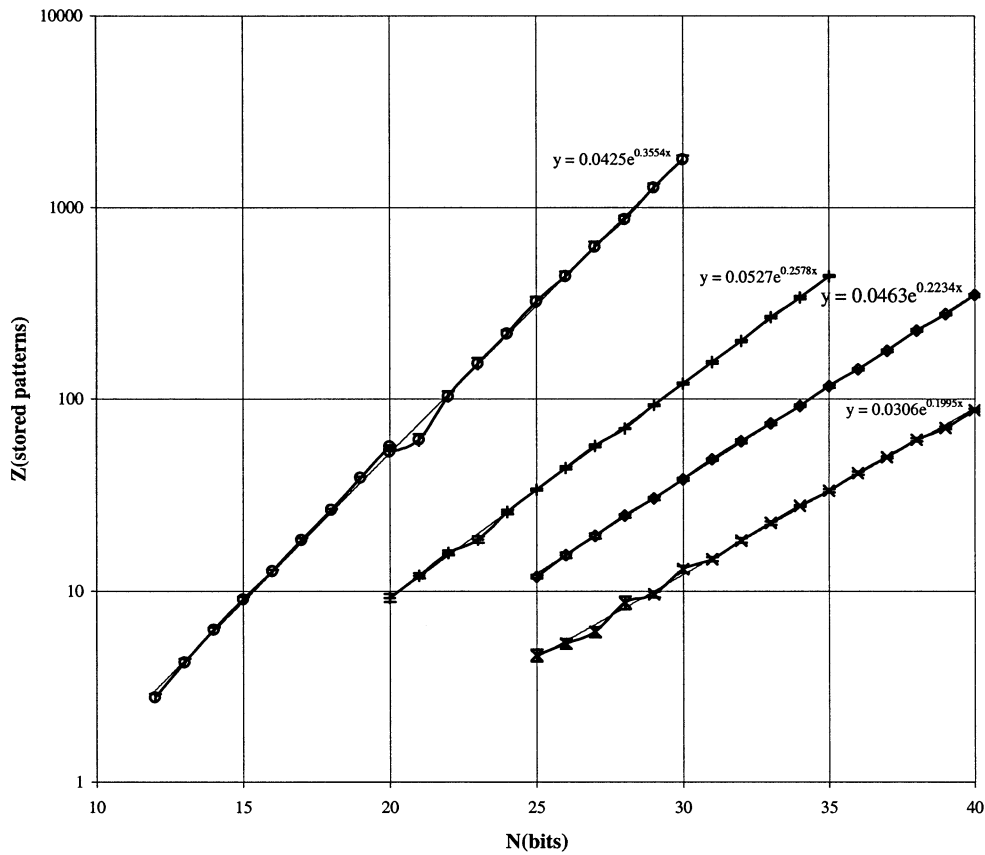


Fig. 6. Storage capacity for fixed output error rate for four levels of input corruption.

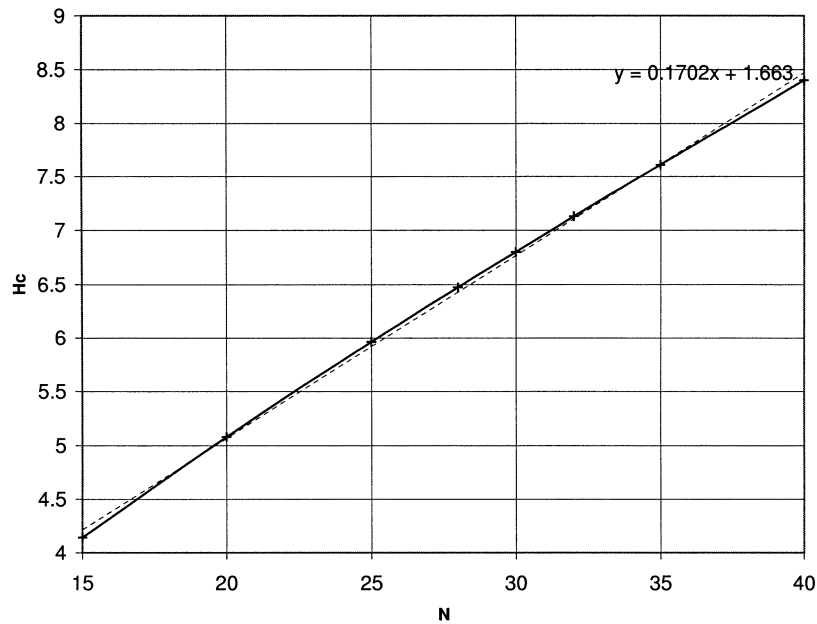


Fig. 7. Critical Hamming distance (corresponding to critical inner-product u_a) against the pattern-length. Over this range, they are close to linearly related.

VII. CONCLUSION

We have shown that the RCAM model of associative recall is governed by two distributions; a foreground distribution representing the correct label update and a background distribution which leads to an erroneous update. When a large number of patterns are stored in the memory, these distributions can be rep-

resented by Gaussians with the same variance, but separated by a factor dependent on the current number of bit errors present in the input pattern. It is this separation, which allows recovery of the correct pattern.

Study of the first iteration reveals two “quality” measures which we can use to determine the expected performance of the RCAM. The first of these involves the expectation value

of the separation between background and foreground distributions and leads us to conclude that the exponential RCAM is the best choice for large storage capacity. The second quality measure in the expectation of bit-wise error rate after the first iteration of the RCAM. Consideration of this measure leads to a numerical solution for an optimal function which performs similarly to the exponential. However, the exponential is more robust when presented with widely different storage requirements.

Finally, study of the bit error rate after one iteration of the memory revealed a method of computing the pattern recall error rate for an RCAM. This method was shown to accurately predict the error rates of the exponential RCAM. By studying the storage capacity of a memory at a fixed output error-rate, we derived an expression for the storage capacity of the ECAM at an output error rate of 0.01.

APPENDIX A FAMILY OF FUNCTIONS GIVES CONSTANT Q

An important step toward the optimal excitation function is to consider the family of functions that leave Q invariant. In this Appendix, we prove the following lemma.

Lemma: $Q \rightarrow Q$ under the transformation $f \rightarrow af + b$.

Proof: Let $f' = af + b$

$$\begin{aligned} \bar{f}' &= \sum_{t=-N}^N p(t)f'(t) \\ &= \sum_{t=-N}^N p(t)(af(t) + b) = a\bar{f} + b \\ Q' &= \frac{\left[\sum_{t=-N}^N q(t) \{f'(t) - \bar{f}'\} \right]^2}{\sum_{t=-N}^N p(t) \{f'(t) - \bar{f}'\}} \\ &= \frac{\left[\sum_{t=-N}^N q(t)a \{f(t) - \bar{f}\} \right]^2}{\sum_{t=-N}^N p(t)a^2 \{f(t) - \bar{f}\}} \\ &= Q. \end{aligned}$$

As a result of this lemma, Q remains invariant under a similarity transform (i.e., translation and scaling) of the excitation function. Any excitation function that is related to an optimal excitation function under the similarity transform is, hence, also optimal.

APPENDIX B MAXIMAL VALUE OF Q_s

In this Appendix, we consider the effect of first-order perturbations of the excitation function f . This leads to an optimal solution to the saddle-point equations for f_k .

We commence by changing just one of the discrete function values f_t by an amount Δ . The perturbed function is specified as follows:

$$f'_t = \begin{cases} f_t + \Delta, & \text{if } t = k \\ f_t, & \text{otherwise} \end{cases}. \quad (38)$$

Under this perturbation, the mean value of the excitation function changes by an amount

$$\bar{f}' = \bar{f} + r_k \Delta \quad (39)$$

while, to first-order in Δ , the numerator and denominator of the quality measure Q_s are perturbed in the following way:

$$\begin{aligned} S' &= S + (q_k - r_k)\Delta \\ T' &= T + 2r_k f_k \Delta + 2r_k \bar{f} \Delta + \dots \end{aligned} \quad (40)$$

As a result, to first-order the change in Q_s is given by

$$\begin{aligned} Q'_s &= \frac{S'^2}{T'} \left[1 + 2 \frac{(q_k - r_k)\Delta}{S} \right] \left[1 + 2 \frac{r_k(f_k + \bar{f})\Delta}{T} \right]^{-1} \\ &= Q_s \left[1 + 2 \frac{(q_k - r_k)\Delta}{S} - 2 \frac{r_k(f_k + \bar{f})\Delta}{T} \right]. \end{aligned} \quad (41)$$

Since the derivative of Q_s is zero at the local maxima of Q_s , the first-order change must also vanish. As a result

$$\frac{q_k - r_k}{S} = \frac{r_k(f_k - \bar{f})}{T}. \quad (42)$$

Earlier we proved that (42) is solved by a family of functions related by a similarity transform. The family of equations may be further simplified by constraining two of the remaining degrees of freedom. Specifically, we impose the constraints

$$\bar{f} = 1 \quad f_0 = \frac{q_0}{p_0}. \quad (43)$$

To proceed with our development, we substitute for S and T into the condition for the Q_s to be maximized. As a result

$$(q_k - r_k) \sum_{t=-N}^N p_t f_t^2 - 1 = r_k(f_k - 1) \sum_{t=-N}^N q_t f_t - 1. \quad (44)$$

We would like to simplify matters by eliminating the summation over the individual values of the inner product. Accordingly, we note that the above equation holds for arbitrary values of k . In consequence, we may write

$$\begin{aligned} (q_{k+1} - p_{k+1}) \left(\sum_{t=-N}^N p_t f_t^2 - 1 \right) &= \\ p_{k+1}(f_{k+1} - 1) \left(\sum_{t=-N}^N q_t f_t - 1 \right). \end{aligned} \quad (45)$$

It is now a simple matter to eliminate the summations by substitution between (44) and (45). After straightforward algebra, we find that

$$p_{k+1} f_{k+1} = \frac{1}{q_k - r_k} [(q_{k+1} - p_{k+1})r_k f_k - r_k q_{k+1} + p_{k+1} q_k]. \quad (46)$$

We have now obtained a recurrence relation for r_k and q_k , which holds for arbitrary k . We can, therefore, use the recurrence relation to compute the value of $p_{k+2}f_{k+2}$. As a result

$$p_{k+2}f_{k+2} = \frac{1}{q_k - r_k} [(q_{k+2} - p_{k+2})r_k f_k - r_k q_{k+2} + p_{k+2}q_k]. \quad (47)$$

More generally, we may write

$$p_{k+n}f_{k+n} = \frac{1}{q_k - r_k} [(q_{k+n} - p_{k+n})r_k f_k - r_k q_{k+n} + p_{k+n}q_k]. \quad (48)$$

When we take the case $k = 0$, then

$$p_n f_n = \frac{1}{q_0 - p_0} [(q_n - p_n)p_0 f_0 - p_0 q_n + p_n q_0]. \quad (49)$$

This formula allows us to compute the product $p_n f_n$. However, substituting for $f_0 = q_0/p_0$ in (49), we find

$$f_n = \frac{q_n}{p_n}. \quad (50)$$

The general family of optimal functions is, therefore, of the form

$$f_n = a \frac{q_n}{p_n} + b. \quad (51)$$

The corresponding maximum value of Q_s is given by

$$Q_{opt} = \sum_{n=-N}^N \frac{q_n^2}{p_n} - 1. \quad (52)$$

REFERENCES

- [1] K. Steinbuch, "Das lern matrix," *Kybernetik*, vol. 1, p. 36, 1961.
- [2] D. J. Willshaw, *Nature*, vol. 222, p. 960, 1969.
- [3] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," in *Proc. Nat. Academy Science USA*, vol. 79, 1982, pp. 2554–2558.
- [4] E. Gardner, "Structure of metastable states in the Hopfield model," *J. Phys. A*, vol. 19, no. 16, pp. 1047–1052, 1986.
- [5] T. D. Chiueh and R. M. Goodman, "Recurrent correlation associative memories," *IEEE Trans. Neural Networks*, vol. 2, pp. 275–284, Mar. 1991.
- [6] —, "VLSI implementation of a high-capacity neural network associative memory," in *Advances in Neural Information Processing Systems 2*, D. S. Touretzky, Ed. San Mateo, CA: Morgan Kaufmann, 1990, pp. 793–800.
- [7] C. C. Wang and H. S. Don, "An analysis of high-capacity discrete exponential BAM," *IEEE Trans. Neural Networks*, vol. 6, pp. 492–496, Mar. 1995.
- [8] T. D. Chiueh and H. K. Tsai, "Multi-valued associative memories based on recurrent networks," *IEEE Trans. Neural Networks*, vol. 4, pp. 364–366, Mar. 1993.
- [9] P. Kanerva, *Sparse Distributed Memory*. Cambridge, MA: MIT Press, 1988.
- [10] D. Milun and D. Sher, "Improving sampled probability distributions for Markov random fields," *Pattern Recognition Lett.*, vol. 14, pp. 781–788, 1993.

- [11] P. A. Chou, "The capacity of the Kanerva associative memory is exponential," in *Neural Information Processing Systems*, D. Z. Anderson, Ed. New York: AIP, 1988, pp. 184–191.
- [12] P. Whittle, "Artificial memories: Capacity, basis rate and inference," *Neural Networks*, vol. 10, pp. 1619–1626, 1997.
- [13] C. C. Wang, S. M. Hwang, and J. P. Lee, "Capacity analysis of the asymptotically stable multi-valued exponential directional associative memory," *IEEE Trans. Syst., Man, Cybern.*, vol. 26, pp. 733–743, June 1996.
- [14] Z. B. Xu, G. Q. Huy, and C. P. Kwong, "Asymmetric Hopfield-type networks; Theory and applications," *Neural Networks*, vol. 9, pp. 483–501, 1996.
- [15] Z. Y. Chen, C. P. Kwong, and Z. B. Xu, "Multiple valued feedback and recurrent correlation neural networks," *Neural Comput. Applicat.*, vol. 3, pp. 242–250, 1995.
- [16] C. C. Wang and J. P. Lee, "The decision making properties of discrete multiple exponential bi-directional associative memories," *IEEE Trans. Neural Networks*, vol. 6, pp. 993–999, July 1995.
- [17] H. C. Shi, Y. X. Zhao, and X. H. Zhuang, "A general model for bidirectional associative memories," *IEEE Trans. Syst., Man, Cybern. B*, vol. 128, pp. 511–519, June 1998.
- [18] C. C. Wang, C. F. Tsai, and J. P. Lee, "Analysis of radix searching of exponential bidirectional associative memory," *Proc. Inst. Elect. Eng. Comput. Digital Techniques*, vol. 145, pp. 279–285, 1998.
- [19] C. C. Wang, "The majority theorem of centralized multiple BAM's networks," *Inform. Sci.*, vol. 110, pp. 179–193, 1998.
- [20] E. R. Hancock and M. Pelillo, "The Bayesian interpretation of the exponential correlation associative memory," *Pattern Recognition Lett.*, vol. 19, pp. 149–159, 1998.
- [21] E. R. Hancock and J. Kittler, "Discrete relaxation," *Pattern Recognition*, vol. 23, pp. 711–733, 1990.
- [22] C. Mazza, "On the storage capacity of nonlinear neural networks," *Neural Networks*, vol. 10, pp. 593–597, 1997.
- [23] R. J. Elice, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh, "The capacity of the Hopfield associative memory," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 461–482, Apr. 1987.
- [24] F. T. Sommer and P. Dayan, "Bayesian retrieval in associative memories with storage errors," *IEEE Trans. Neural Networks*, vol. 9, pp. 705–713, July 1998.
- [25] A. Jagota, G. Narasimhan, and K. W. Regan, "Information capacity of binary weights associative memories," *Neurocomputing*, vol. 19, pp. 35–58, 1998.
- [26] D. Burshtein, "Typical error pattern recovery of the Hopfield memory under error tolerant conditions," *IEEE Trans. Inform. Theory*, vol. 44, pp. 861–865, Apr. 1998.
- [27] D. Casasent and B. Telfer, "High-capacity pattern-recognition associative-processors," *Neural Networks*, vol. 5, no. 4, pp. 687–698, 1992.
- [28] M. Lowe, "On the storage capacity of the Hopfield model with biased patterns," *IEEE Trans. Inform. Theory*, vol. 45, pp. 314–318, Feb. 1999.
- [29] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. New York: Wiley, 1971, vol. 2, pp. 257–258.



Richard C. Wilson received the B.A. degree in physics from the University of Oxford, Oxford, U.K., in 1992 and the D.Phil. degree from the University of York, York, U.K., for his thesis "inexact graph matching using symbolic constraints" in 1996.

From 1996 to 1998, he was a Research Associate at the University of York. After a period of postdoctoral research, he was awarded an Advanced Research Fellowship in 1998, a position which he currently holds in the Department of Computer Science at the University of York. He has published approximately 70 papers in journals, edited books, and refereed conferences. He is currently an Associate Editor of the journal *Pattern Recognition*. His research interests are in statistical and structural pattern recognition, graph methods for computer vision, high-level vision, and scene understanding.

Dr. Wilson received an outstanding paper award in the 1997 Pattern Recognition Society awards and has won the best paper prize in ACCV 2002. He is a member of the IEEE computer society.



Edwin R. Hancock received the Bachelor's degree (with honors) in physics and the Ph.D. degree in high energy physics from the University of Durham, Durham, U.K., in 1977 and 1981, respectively.

He was a Researcher for ten years in the fields of high-energy nuclear physics and pattern recognition at the Rutherford-Appleton Laboratory (now the Central Research Laboratory of the Research Councils). During this period, he also held adjunct teaching posts at the University of Surrey, U.K., and the Open University. In 1991, he moved to the

University of York as a Lecturer in the Department of Computer Science. He was promoted to Senior Lecturer in 1997 and to Reader in 1998. In 1998, he was appointed to a Chair in Computer Vision. He now leads a group of some 15 faculty, research staff and Ph.D. students working in the areas of computer vision and pattern recognition. His main research interests are in the use of optimization and probabilistic methods for high and intermediate level vision. He is also interested in the methodology of structural and statistical pattern recognition. He is currently working on graph-matching, shape-from-X, image data-bases and statistical learning theory. His work has found applications in areas such as radar terrain analysis, seismic section analysis, remote sensing and medical imaging. He has published approximately 60 journal papers and 200 refereed conference publications.

Dr. Hancock has been a member of the editorial board of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE RECOGNITION. He has also been a Guest Editor for special editions of the journals *Image and Vision Computing and Pattern Recognition*, and he is currently a Guest Editor of a special issue of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE devoted to energy minimization methods in computer vision. He has been on the program committees for numerous national and international meetings. In 1997, he established a new series of international meetings on energy minimization methods in computer vision and pattern recognition. He was awarded the Pattern Recognition Society medal in 1991 for the best paper to be published in the journal *Pattern Recognition* and an outstanding paper award in 1997.