Storage Capacity of the Exponential Correlation Associative Memory

Richard C. Wilson and Edwin R. Hancock*
Department of Computer Science,
University of York
York, Y01 5DD, UK

email:erh@cs.york.ac.uk
Tel:44 1904 43 3374
Keywords: Exponential correlation associative memory, storage capacity

Abstract

In this paper we analyze the pattern storage capacity of the exponential correlation associative memory(ECAM). This architecture was first studied by Chiueh and Goodman [3] who concluded that, under certain conditions on the input patterns, the memory has a storage capacity that was exponential in the length of the bit-patterns. A recent analysis by Pelillo and Hancock [9], using the Kanerva picture of recall, concluded that the storage capacity was limited by $2^{N-1}/N^2$. Both of these analyses can be criticised on the basis that they overlook the role of initial bit-errors in the recall process and deal only with the capacity for perfect pattern recall. In other words, they fail to model the effect of presenting corrupted patterns to the memory. This can be expected to lead to a more pessimistic limit. Here we model the performance of the ECAM when presented with corrupted input patterns. Our model leads to an expression for the storage capacity of the ECAM both in terms of the length of the bit-patterns and the probability of bit-corruption in the original input patterns. These storage capacities agree closely with simulation. In addition, our results show that slightly superior performance can be obtained by selecting an optimal value of the exponential constant.

*corresponding author erh@cs.york.ac.uk

1 Introduction

Correlation memories have found widespread use in the recognition of binary pattern vectors [3, 4, 5, 6, 7]. An important class of correlation memory is the recursive correlation associative memory [3]. Such networks are typified by a node excitation function and a pattern distance measure. As recently demonstrated by Chieuh and Goodman [3], the general class of pattern recogniser subsumes a diversity of different memory architectures. If the excitation function is the identity function, then the Hopfield memory [1] is obtained. If, on the other hand, the excitation function is exponential then an interesting structure called the exponential correlation associative memory(ECAM) results. Through an analysis of the Lyapanov function, Chieuh and Goodman [3] have shown that the ECAM not only possess monotonic convergence properties, but that there is a suggestion that it has a storage capacity that is exponential in the length of the bit patterns.

In a recent series of papers, we have developed a Bayesian theory of structural pattern recognition that has many features in common with the ECAM [8, 10, 9]. The most striking of these is that label consistency is gauged by an exponential function of Hamming distance. In fact, we recently showed that the ECAM can be viewed as performing Bayesian pattern restoration via gradient ascent on a configurational probability measure [9].

Using the Bayesian framework provided by our previous work [10, 9] we focus in this paper on the storage capacity of the ECAM when the input patterns have non-vanishing error probabilities. This is in direct contrast to the approach of both Chieuh and Goodman [3], and Hancock and Pelillo [9], and leads to an expression for the storage capacity in terms of both the length of bit-patterns and the bit-corruption probability.

Finally, we demonstrate experimentally that our analysis accurately predicts the behaviour of real ECAM's. In addition we show that there is an optimal choice of the exponential constant of the ECAM for which the ECAM out-performs our storage limit. This value is broadly in line with the value observed by Milun and Sher [12], who have explored the noise sensitivity of our Bayesian pattern reconstruction method [8].

2 Recursive Correlation Associative Memories

The ECAM is an instance of a more general associative memory model which Chiueh and Goodman [3] called the recursive correlation associative memory (RCAM). The network is composed of N computational nodes, and at any particular stage of updating each node will be in one of the binary states denoted by $\Omega \equiv \{-1, 1\}$. The particular realisation of the labelling of the node indexed j is denoted by s_j. With this notation, the global state of the network is represented by the configuration of binary values $S = \{s_j \in \Omega, \forall j = 1 \dots N\}$.

Now, suppose that we have access to a set of training patterns. Typically, these would be configurations of binary labels which we want to recover from an initial inconsistent state of the network. Assume that there are Z such global patterns denoted by $\Lambda^\mu = \{\lambda_j^\mu \in \Omega, \forall j = 1 \dots N\}$. According to this notation, μ is the pattern index and λ_j^μ is the binary value assigned to the site indexed j by the μ^{th} training pattern.

With these ingredients, the dynamical behaviour of an RCAM is governed by the following updating rule

$$s_j = \operatorname{sgn}\left\{\sum_{\mu=1}^{Z} \lambda_j^\mu \ f\left(\sum_{i=1}^{N} s_i \lambda_i^\mu\right)\right\} \tag{1}$$

for all $j = 1 \dots N$, where $f(\cdot)$ is an appropriate weighting function. In [3] it was shown that the dynamical system described by equation (1) is asymptotically stable both in the asynchronous and synchronous update modes, provided that the weighting function f is continuous and monotone nondecreasing over the interval $[-N, N]$. This was proved by showing that the system has an associated Liapunov (or "energy") function that governs its dynamical behavior. In addition, Chiueh and Goodman [3] showed how some known associative memory models can be viewed as a special case of the RCAM by just choosing a suitable weighting function. For example, if f is chosen to be the identity function, i.e., $f(x) = x$. then the model becomes essentially identical to the original Hopfield memory [1], with the weights initialised by the Hebb rule. Of particular interest is the case when the weighting function has an exponential form, i.e., $f(x) = e^{kx}$, where k is a positive real-valued constant. This class of memories constitutes the original ECAM. This was shown to have large storage capacity and turns out to be especially suited for VLSI implementation [4, 3].

In a recent study Hancock and Pelillo [9] have shown that the ECAM update rule can be derived using a simple Bayesian model of iterative pattern recall. The update rule for maximum-likelihood pattern reconstruction via gradient ascent is given by

$$s_j = \operatorname{sgn}\left\{\sum_{\mu=1}^{Z} \lambda_j^\mu e^{-kH_\mu}\right\} \ . \tag{2}$$

where $H_\mu = \frac{1}{2}[N - \sum_{i=1}^{N} s_i \lambda_i^\mu]$ is the Hamming distance between the input pattern S and the stored pattern Λ^μ. Ignorring multiplicative factors which are constant for fixed-length patterns, the update rule may be simplified to

$$s_j = \operatorname{sgn}\left\{\sum_{\mu=1}^{Z} \lambda_j^\mu \exp\left[k \sum_{i=1}^{N} s_i \lambda_i^\mu\right]\right\} \ . \tag{3}$$

According to the Bayes treatment, the constant of exponentiation k is equal to $\frac{1}{2}\ln\left(\frac{1-p}{p}\right)$, where p is the bit-error probability in the pattern. Provided $p < \frac{1}{2}$, then $k > 0$. Furthermore it is interesting to note that as p approaches $\frac{1}{2}$ from below, then $\lim_{p \to \frac{1}{2}^-} k = 0$, and the exponentials appearing in the update rule can be approximated in a linear manner producing the Hopfield dynamical [1] rule.

3 Pattern Storage and Error Rates

The storage capacity of a memory is generally defined as the number of patterns for which the memory can recover the input pattern with no error. However there are some problems with applying this definition to associative memories. In the first instance, the nature of the stored patterns is random. By chance a difficult distribution of patterns may render the memory ineffective. Moreover, a badly corrupted input pattern may also make recovery impossible. In other words, there is always some probability of error when recovering a pattern from the memory. In this case, the error rate is a more appropriate measure of the performance of a memory [3].

In accordance with this observation, and in keeping with the philosophy of pattern recovery, we set the target error rate to some small value and measure the number of patterns which can be stored before the error rate of the ECAM reaches this value. In the simulations which we present later-on in Section 5 we have taken the error rate of $\frac{1}{2}$ bit per pattern to indicate the onset of recall errors.

It is informative at this stage to examine the storage capacity of some of the limits of the ECAM. In the case of a small exponential constant k, the memory performs the action of a Hopfield network. The storage capacity of this network is known to be $Z = 0.14N$ [2], but this limit only applies to uncorrupted input patterns. Indeed, the Hopfield networks ability to recover from significant input pattern error is rather limited.

The second limit of interest is the case of hard exponentials (large k) and small errors in the input pattern. For example, Chiueh and Goodman [3] assume that the input pattern is closest in terms of Hamming distance to the prototype pattern. Study of the dynamics of the ECAM reveal that, by making k arbitrarily large, convergence to the closest pattern can be guaranteed. The storage capacity in this case is either infinite or governed by multiple pattern degeneracy. The bit-error rate is zero and the probability of recovering the wrong pattern is

$$P_w = 1 - \frac{1}{1 + \frac{Z}{2^N}} \tag{4}$$

4 The Hamming Distance Picture

To understand the update dynamics of the ECAM, it is instructive to study how the Hamming distance is distributed for realistic pattern classification problems. The distribution has two components. The first reflects the effect of noise corruption on perfect patterns producing departures from zero Hamming distance. The second is due to the distribution of Hamming distance between the different stored patterns—we refer to this as the background Hamming distance. In the following we refer to the Hamming distance of the input pattern to the prototype pattern as H_s and the Hamming distance to some background pattern as H_b.

Consider a noise corrupted pattern indexed s. If the bit corruption is memoryless, then the Hamming distance to the true uncorrupted prototype of this pattern is equal to the number of bit errors. In this case H_s follows the binomial distribution

$$P_s(H_s = t) = \frac{N!}{(N-t)!t!} p^t (1-p)^{N-t} \ . \tag{5}$$

To describe the Hamming distribution between the competing stored patterns requires a model of the structure of the pattern space. The basic assumption is that the patterns are random bit patterns. If we confine our attention to the case when the bits occur with equal probability, we obtain the following distribution for unrelated patterns

$$P_b(H_b = t) = \frac{N!}{2^N (N-t)!t!} \ . \tag{6}$$

The input pattern to the ECAM is also subject to a probabilistic error process. In other words each pattern-bit has some probability p of being corrupted in a uniform memoryless corruption process. The Hamming distance from the prototype pattern is governed by equation (5). We now examine the performance of the ECAM in the limit where k is large. In this case the contribution to the energy from the nearest pattern can be made arbitrarily larger than all other patterns, and therefore the memory will converge to the pattern of minimum Hamming distance.

4.1 Pattern Error

The first step towards understanding the storage capacity of the ECAM is to calculate the probability of erroneous pattern recall P_{err}. We commence by using the signal and background Hamming distributions to compute the probability that any single stored pattern has a Hamming distance smaller or equal to that for the correct pattern which is indexed s, i.e., $P_w = P(H_s \leq H_b)$. This goal is achieved by performing a discrete convolution of binomial distributions given in (5) and (6)

$$P(H_s \leq H_b) = \frac{1}{2^N} \sum_{u=0}^{N} \frac{N!}{(N-u)!u!} \sum_{t=0}^{u} \frac{N! p^t (1-p)^{N-t}}{(N-t)!t!} \ . \tag{7}$$

At large values of N, the second summation is dominated by the term in which the signal and background patterns have identical Hamming distance. In other words, $P(H_b = H_s) >> P(H_b < H_s)$. We can then rewrite this sum as

$$P(H_s \leq H_b) = \frac{1}{2^N} \sum_{u=0}^{N} \frac{N!}{(N-u)!u!} \frac{N! p^u (1-p)^{N-u}}{(N-u)!u!} \ . \tag{8}$$

The condition for error is that at least one of the $Z-1$ competing background patterns produces a configuration with Hamming distance equal to H_s, and in this case, the probability of recovering the wrong pattern is 1/2. The background patterns can be regarded as independent events each of which has a uniform probability $P(H_b \leq H_s)$ of causing an error. These conditions specify a binomial distribution for the number of potential error-producing patterns. As a result, the probability of misclassifying the true pattern as any of the remaining $Z-1$ background patterns is equal to $P_{err} = 1 - (1 - P_w)^{Z-1}$. When P_w is small, then $P_{err} \simeq (Z-1)P_w$.

4.2 Gaussian Pattern Corruption

Unfortunately, the binomial coefficient expansions entering the expression for the recall error are not tractable in closed-form. We therefore replace the binomial expressions for the signal and background distributions by Gaussian approximations. These approximations can be expected to be valid at moderate levels of p and fail as $p \to 0$, when a Poisson distribution is more appropriate. In the case of the signal distribution which has mean Np and variance $Np(1-p)$, the approximating Gaussian is

$$P_s(H_s = t) = \frac{1}{\sqrt{2\pi Np(1-p)}} \exp\left[-\frac{1}{2}\frac{(t-Np)^2}{Np(1-p)}\right] \tag{9}$$

The background, on the other hand, has mean $N/2$ and variance $N/4$. As a result the approximating Gaussian is

$$P_b(H_b = t) = \sqrt{\frac{2}{\pi N}} \exp\left[-\frac{2(t-\frac{1}{2}N)^2}{N}\right] \tag{10}$$

After some analysis, it follows from equation (6) that the probability of a background pattern having equal Hamming distance to the prototype pattern is given by

$$P_w = \frac{1}{\sqrt{\frac{1}{2}\pi N[1+4p(1-p)]}} \exp\left[-\frac{N(2p-1)^2}{2+8p(1-p)}\right] \tag{11}$$

4.3 Storage capacity

According to our background distribution, then if an input pattern is misclassified, then there are on average $N/2$ bit errors. In our experimental work we deem recall to have succeeded if there is on average less than half a bit-error per pattern. For such patterns $P_{err}\frac{N}{2} = \frac{1}{2}$ and in consequence $P_{err} = \frac{1}{N}$. As a result, and using the small P_w approximation discussed above, the limit on the storage capacity is given by $Z = 1 + \frac{1}{NP_w}$. Substituting for P_w from equation (9), the storage capacity is given by

$$Z = 1 + \sqrt{\frac{2\pi[1 + 4p(1-p)]}{N}} \exp\left[\frac{N(2p-1)^2}{2 + 8p(1-p)}\right] \tag{12}$$

This result makes clear that there is an interplay between the length of the memory patterns and the fraction of bit-errors p on the input patterns in determining the storage capacity of the memory. This result is applicable when the exponential constant is large and the corruption probability is moderate.

As mentioned earlier, when the exponential constant is small, the memory performs the action of a Hopfield network and has no significant ability to recover from corrupt inputs. At moderate levels of k however, the memory performs well. However, the analysis of storage capacity in this case is not tractable because our assumption of convergence to the nearest pattern is not valid. Intuitively we might expect the dynamics of the ECAM to be influenced by local densities in the pattern space, and since the density around the correct pattern will on average be higher, convergence to the correct pattern will be improved. This observation is confirmed in the next section. Our results show that there is an optimal value for the constant which gives some 5-10% improvement in the storage capacity.

Before proceeding to an experimental validation of our theoretical findings, it is worth considering the large-pattern limit when the bit-error probability becomes vanishingly small. Under these conditions

$$Z_{max} \simeq \sqrt{\frac{\exp N}{N}} \tag{13}$$

This result is considerably smaller than the 2^N patterns suggested by Chieuh and Goodman [3] or the more pessimistic bound of $\frac{2^{N-1}}{N^2}$ obtained by Hancock and Pelillo [9] using the Kanerva picture.

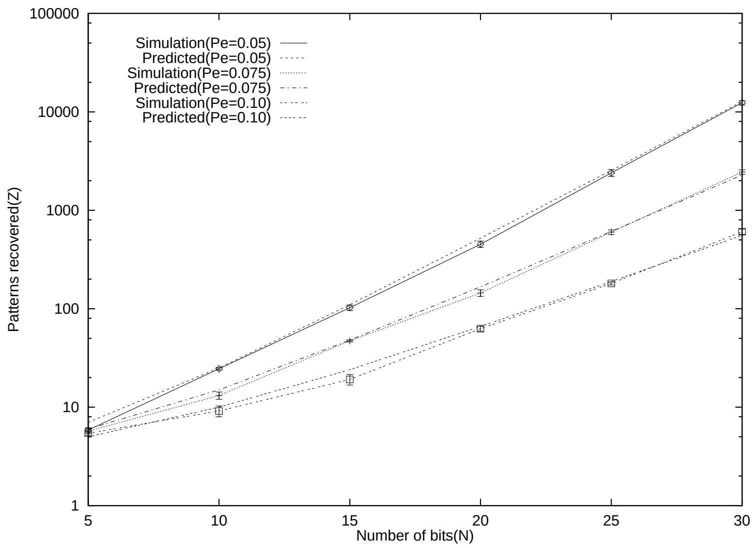

Figure 1: Comparison between theoretical and actual storage capacities of the ECAM in the large-k mode (lograithmic scale)

5 Experiments

Our experiments are based on random bit patterns containing equal proportions of of high and low bits. We load the memory with a controlled number of patterns. We select memory patterns at random and then randomly flip each bit with a probability p. The bits of the corrupted input patterns are iteratively updated according to equation 2. The final pattern of bits is compared with its uncorrupted version originally loaded into the memory. The bit error rate is the Hamming distance between the output of the ECAM and the prototype pattern. This error rate is averaged over a large number of runs to determine mean and variance. We increase the number of patterns loaded into the memory until the average error rate rises above half a bit per pattern.

P_e	N=10		N=15		N= 20		N=25		N=30	
0.05	25	27	103	117	452	554	2400	2711	12373	13568
0.075	13	16	47	52	145	182	600	667	2448	2501
0.10	9	11	19	27	63	73	181	211	607	622
0.125	8	8	12	16	25	35	71	81	206	195
0.150	6	6	7	10	14	19	26	37	58	74
0.175	5	5	6	7	9	11	12	19	22	33
0.200	3	4	4	5	5	7	7	11	8	17

Table 1: Experimental and theoretical storage capacities

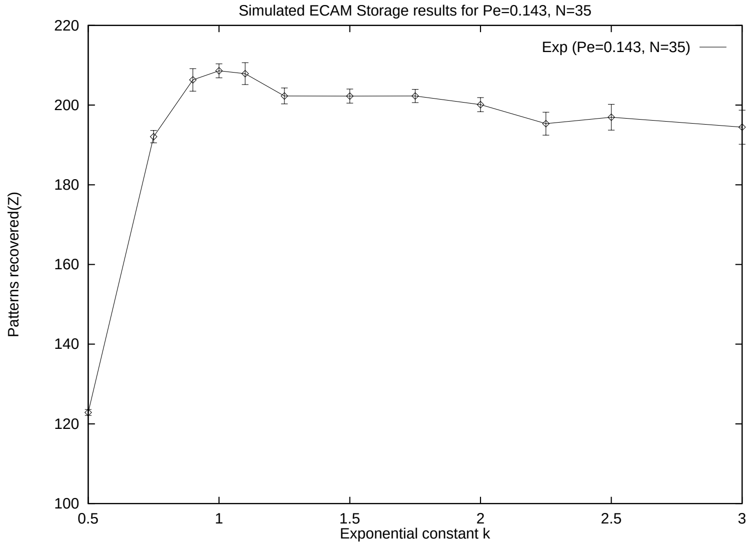

Figure 2: The recall performance of the ECAM as a function of the exponential constant k.

The results of our experiments are summarised in Table 1 and Figure 1. The table lists the experimental and theoretical values of the maximum number patterns that can be loaded into the memory before the onset of recall errors. The results are presented for various pattern vector lengths and for various levels of input pattern corruption. In Table 1 the first listed number in each column is the experimental value while the second listed value is the theoretical value. The results are also plotted in Figure 1. The main features to note are that the theoretical and experimental values are in good agreement, although the theoretical values appear to consistently overestimate the experimental values by a small amount. The second feature to note is that the storage capacity falls with increasing bit errors on the input patterns.

In Figure 2 we show a plot of the number of successful pattern recalls as a function of the exponential constant k. This experiment is based on patterns which have an initial bit-error probability of 0.143. There are several features of this plot that deserve further comment. In the first instance, as the exponential constant approaches zero, i.e. the memory approaches the soft-limit where it approximates the Hopfield network, then so the number of recalled patterns decreases rapidly. Secondly, as the exponential constant becomes increasingly large, i.e. the exponentials approach their hard-limit, then so the number of recalled patterns becomes largely insensitive to the exponential constant. Finally, there is an optimal value of k for which the storage capacity is 10% greater than the hard limit. This corresponds to a value of $k = 1.0$. This value is close to the value of 0.90 suggested by Hancock and Pelillo [9] and 0.87 suggested by Milun and Sher [12].

6 Conclusions

The analysis presented in this paper has revealed that the effective storage capacity of the ECAM is much smaller than either exponential limits suggested by both the analysis of Chieuh and Goodman [3] and the Kanerva picture [11, 13], when presented with corrupt input patterns. Moreover, our theoretical result is supported by experiment.

Our future plans are to focus more closely on the form of the excitation function of the RCAM. The simple Bayesian analysis of Hancock and Pelillo [9] has suggested that this should be an exponential if the bit errors are uniform and memoryless. The framework presented in his paper allows us to picture the process of recall in terms of the overlap between the input noise distribution and the distribution of memory patterns. By choosing the excitation function that minimises this overlap we can maximise the storage capacity of the ECAM. Early results that it is the exponential that satisfies this condition.

References

[1] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci. USA*, vol. 79, pp. 2554–2558, 1982.

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

[3] T. D. Chiueh and R. M. Goodman, "Recurrent correlation associative memories," *IEEE Trans. Neural Networks*, vol. 2, pp. 275–284, 1991.

[4] T. D. Chiueh and R. M. Goodman, "VLSI implementation of a high-capacity neural network associative memory," in D. S. Touretzky, Ed., *Advances in Neural Information Processing Systems 2*. San Mateo, CA: Morgan-Kaufmann, 1990, pp. 793–800.

[5] C. C. Wang and H. S. Don, "An analysis of high-capacity discrete exponential BAM," *IEEE Trans. Neural Networks*, vol. 6, pp. 492–496, 1995.

[6] C. C. Wang and J. P. Lee, "The decision-making properties of discrete multiple exponential bidirectional associative memories," *IEEE Trans. Neural Networks*, vol. 6, pp. 993–999, 1995.

[7] T. D. Chiueh and H. K. Tsai, "Multi-valued associative memories based on recurrent networks," *IEEE Trans. Neural Networks*, vol. 4, pp. 364–366, 1993.

[8] E.R. Hancock and J.Kittler, "Discrete Relaxation",*Pattern Recognition*, **23**, pp 711–733, 1990.

[9] E.R. Hancock and M. Pelillo, "A Bayesian Interpretation for the Exponential Correlation Associative Memory", *Pattern Recognition Letters*, **19**, pp. 149–159, 1998.

[10] R. C. Wilson and E. R. Hancock, “Structural Matching by Discrete Relaxation”, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 634–648, 1997.

[11] P. Kanerva, *Sparse Distributed Memory.* MIT Press, Cambridge, MA, 1988.

[12] D. Milun and D. Sher, “Improving sampled probability distributions for Markov random fields,” *Pattern Recognition Letters.* v14 pp. 781–788 1993.

[13] P. A. Chou, “The capacity of the Kanerva associative memory is exponential,” in D. Z. Anderson, Ed., *Neural Information Processing Systems.* New York, NY: American Institute of Physics, 1988, pp. 184–191.