ELSEVIER

Pattern Recognition Letters 19 (1998) 149–159

Pattern Recognition Letters

A Bayesian interpretation for the exponential correlation associative memory

Edwin R. Hancock [a,*], Marcello Pelillo [b,1]

[a] Department of Computer Science, University of York, York YO1 5DD, United Kingdom
[b] Dipartimento di Matematica Applicata e Informatica, Università ''Ca' Foscari'' di Venezia, Via Torino 155, 30173 Venezia Mestre, Italy

Received 12 June 1997; revised 22 October 1997

Abstract

The exponential correlation associative memory (ECAM) is a recurrent neural network model which has large storage capacity and is particularly suited for VLSI hardware implementation. Our aim in this paper is to show how the ECAM model can be entirely derived within a Bayesian framework, thereby providing more insight into the behaviour of this algorithm. The framework for our study is a novel relaxation method which involves direct probabilistic modelling of the pattern corruption mechanism. The parameter of this model is the memoryless probability of error on nodes of the network. This bit-error probability is not only important for the interpretation of the ECAM model, but allows also us to understand some more general properties of Bayesian pattern reconstruction by relaxation. In addition, we demonstrate that both the Hopfield memory and the Boolean network model developed by Aleksander can be regarded as limits of the presented relaxation approach with precise physical meaning in terms of this parameter. To study the dynamical behaviour of our relaxation model, we use the Hamming distance picture of Kanerva which allows us to understand how the bit-error probability evolves during the relaxation process. We also derive a parameter-free expression for the storage capacity of the model which, like a previous result of Chiueh and Goodman, scales exponentially with the number of nodes in the network.

Keywords: Associative memories; Exponential correlation associative memory; Bayes treatment; Hopfield network; Discrete relaxation; Storage capacity

1. Introduction

Since Hopfield's influential work (Hopfield, 1982), there has been an explosion of interest in the study of neural network models of associative memory. These are parallel computational networks in which idealised versions of corrupted patterns are recalled by locally updating individual nodes in the network to optimise a global cost function. The original model proposed by Hopfield, however, was soon recognised to have severe limitations associated with its restricted storage capacity (McEliece et al., 1987), and this has then motivated a number of

* Corresponding author. Email: erh@minster.york.ac.uk.
[1] Email: pelillo@dsi.unive.it. Work done while the author was visiting the Department of Computer Science at the University of York, United Kingdom.

0167-8655/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII S0167-8655(97)00164-5

investigators to develop alternative memory models in an attempt to overcome this fundamental drawback. One of the most successful such attempts has recently been reported by Chiueh and Goodman (1991) who developed an exponential correlation associative memory (ECAM) model which not only turns out to have large storage capacity, but can also be easily implemented in VLSI hardware (Chiueh and Goodman, 1990, 1991). The model has also been recently extended to include bidirectional (Jeng et al., 1990; Wang and Don, 1995; Wang and Lee, 1995), multi-valued (Chiueh and Tsai, 1993), and real-valued memories (Geva and Sitte, 1991).

Viewed as systems that achieve global configurational optimisation via local iterative computations, these associative memories have many features in common with relaxation algorithms widely employed in the field of computer vision and pattern recognition (Hummel and Zucker, 1983; Geman and Geman, 1984; Hancock and Kittler, 1990; Wilson and Hancock, 1997). Our aim in this paper is to demonstrate not only that there is a strong conceptual similarity between the two approaches, but that the ECAM and the Hopfield memories can be regarded as particular instances of discrete relaxation (Hancock and Kittler, 1990; Wilson and Hancock, 1997), providing them with a direct Bayesian interpretation. The analysis provides an empirical understanding of the dynamical properties of these models.

Basic to the methodology described in this paper is the direct modelling of the pattern corruption mechanism which is objective yet simple (Hancock and Kittler, 1990; Wilson and Hancock, 1997). We assume that each node in the network is subject to memoryless corruption, and adopt as our criterion for global labelling the joint probability of the binary node configuration on the network. This turns out to be a compound exponential function of the Hamming distances to the idealised stored patterns, the parameter of these exponentials being the memoryless probability of bit-errors acting on the binary patterns. Node updating takes place so as to perform gradient ascent on the global configurational probability. The ECAM, too, implements an exponential update rule, and we will show how our relaxation scheme is identical to the ECAM model when the base of its exponentials is chosen as a certain function of the bit-error probability. This corresponds also to the best possible choice in terms of storage capacity. Moreover, when the bit-error probability approaches $\frac{1}{2}$, our relaxation model becomes equivalent to the Hopfield update rule. Conversely, when the bit-error probability becomes asymptotically small, the relaxation approach implements a type of look-up operation, selecting the global pattern of minimum Hamming distance. This is similar in function to the Boolean network proposed by Aleksander (1989) and extensively analysed by Wong and Sherrington (1989).

In order to understand some of the dynamical properties of our discrete relaxation model of associative memory, we use the Hamming distance picture of Kanerva (1988). This model allows us to predict how the bit-error probability evolves as the corrupted patterns undergo iterative associative recall. In particular, the fixed points of the bit-error probability play a crucial role in limiting effective pattern recovery. If the bit-error probability exceeds a certain value termed the ‘‘critical distance’’ by Kanerva, then iterations only result in disruption of memory patterns. Moreover, if the critical distance altogether vanishes then the memory capacity of the network has been exceeded. We use this latter property to show that the storage capacity of our memory is exponential with the number of nodes. This is exactly the finding of Chiueh and Goodman (1991). However, our result for the storage capacity depends only on the number of nodes in the network, i.e., the length of memory patterns. This contrasts with the theoretical result of Chiueh and Goodman which relates the storage capacity to a threshold parameter which is effectively a trade-off between storage capacity and bit-error probability. Furthermore, it is interesting to note that our Bayesian framework naturally leads us to a choice of the exponential base that gives the ultimate upper bound for the asymptotic capacity of an associative memory.

To summarize, the novel contributions of this paper are twofold. The first contribution is to provide a Bayesian framework which furnishes a unifying viewpoint for understanding of a number of different binary associative memories. This framework allows us to identify the Aleksander, Hopfield and ECAM models as specific instances based on the prevailing value of the bit-error probability. The second contribution is to exploit the Kanerva picture to compute the limiting storage capacity of such networks. This allows us to relate the storage capacity of the ECAM to the length of the stored patterns. The main conclusion is that the storage capacity of the ECAM is exponential.

The outline of this paper is as follows. In Section 2 we briefly describe the ECAM model as introduced in (Chiueh and Goodman, 1991). Section 3 develops a model for the global configurational probability and derives the ECAM entirely within a Bayesian framework, thereby making explicit the relation between discrete relaxation and the ECAM model. In Section 4 we additionally describe the relationship between the relaxation approach, the Hopfield memory and the Aleksander Boolean network. Following Kanerva, Section 5 presents a model of the underlying pattern space which allows us to provide an interpretation for these relationships; it also allows us to understand some of the properties of Bayesian pattern reconstruction by configurational relaxation. Finally, in Section 6 we present our conclusions.

2. Exponential correlation associative memories

The ECAM is an instance of a more general associative memory model which Chiueh and Goodman Chiueh and Goodman (1991) 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 $S=\{s_j \in \Omega \mid j=1,\dots,N\}$.

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 $\Lambda^{\mu}=\{\lambda_j^{\mu} \in \Omega \mid j=1,\dots,N\}$. 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 μth training pattern.

The dynamical behaviour of an RCAM is governed by the following updating rule

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

for all $j=1,\dots,N$, where $f(\cdot)$ is an appropriate weighting function. Chiueh and Goodman (1991) showed that the dynamical system described by Eq. (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 behaviour. In addition, Chiueh and Goodman (1991) 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.,

$$f(x) = x, \tag{2}$$

then the model becomes essentially identical to the original Hopfield memory (Hopfield, 1982), with the weights initialised by the Hebb rule. Of particular interest is the case when the weighting function has an exponential form, i.e.,

$$f(x) = a^x, \tag{3}$$

where a is a predetermined constant greater than unity. This class of memories constitutes the ECAM original model. This was shown to have large storage capacity and turns out to be especially suited for VLSI implementation (Chiueh and Goodman, 1990, 1991).

3. Bayesian derivation of the ECAM

We commence our analysis of the ECAM model by considering an arbitrary input pattern S. The basic information relevant to the ‘‘consistency’’ of the configuration S is conveyed by the set of Hamming distances to the prototype patterns Λ^{μ}, $\mu=1,\dots,Z$. With the binary node variables defined in Section 2, the Hamming distance to the prototype pattern indexed μ can be written as

$$H_\mu = \frac{1}{2}\left(N - \sum_{i=1}^{N} s_i \lambda_i^\mu\right). \tag{4}$$

The configurational relaxation procedure is based on maximising the joint probability of the binary label configuration, i.e., $P(S)$. It is therefore necessary to find a way of enumerating $P(S)$ when the label configuration is highly inconsistent, i.e., when there are no prototype patterns for which the Hamming distance is zero. The approach is to adopt a Bayesian viewpoint in which it is assumed that only prototype patterns are legal and have uniform non-zero a priori probabilities of occurrence $P(\Lambda^\mu)$. Other configurations do not occur a priori but are the corrupted realisations of the prototype patterns. This idea is realised by applying the axiomatic property of joint probability to expand $P(S)$ over the space of consistent configurations

$$P(S) = \sum_{\mu=1}^{Z} P(S|\Lambda^\mu) P(\Lambda^\mu). \tag{5}$$

To develop this idea into a useful update rule requires a model of the bit corruption process, that is of the conditional probabilities of the inconsistent configurations given the Z possible prototypes $P(S|\Lambda^\mu)$. We adopt a very simple viewpoint: bit-errors are assumed to be memoryless and to occur with uniform probability $p \leqslant \frac{1}{2}$.

The first consequence of the assumed absence of memory is that the errors are independent. As a result we can factorise the conditional probabilities over the individual nodes in the network, i.e.,

$$P(S|\Lambda^\mu) = \prod_{i=1}^{N} P(s_i|\lambda_i^\mu). \tag{6}$$

Our next step is to propose a model for the bit corruption mechanism at each node in the network. Again, taking recourse to the memoryless assumption, the probability of error on individual nodes is independent of the class of label. This leads us to the following assignment of probability

$$P(s_i|\lambda_i^\mu) = \begin{cases} 1-p & \text{if } s_i = \lambda_i^\mu, \\ p & \text{otherwise}. \end{cases} \tag{7}$$

The model components given in Eqs. (6) and (7) naturally lead to the following expression for $P(S)$ in terms of the set of Hamming distances to the prototype patterns

$$P(S) = \frac{\beta}{Z} \sum_{\mu=1}^{Z} \alpha^{H_\mu}, \tag{8}$$

where $\alpha = p/(1-p)$ and $\beta = (1-p)^N$. We can re-write the above expression for the configurational probability in terms of the more familiar exponential function

$$P(S) = \frac{\beta}{Z} \sum_{\mu=1}^{Z} \exp\left(-H_\mu \ln \frac{1}{\alpha}\right). \tag{9}$$

Expressed in this form it is tempting to draw an analogy between our probability criterion and the exponential Gibbs distributions employed in the stochastic relaxation approach of Geman and Geman (1984). It can also be regarded as a symbolic analogue of the functional used in the deformable template models of Durbin et al. (1989) and of Yuille (1990). An extensive experimental comparison between the more general exponential cost function described by Hancock and Kittler (1990) and Markov random fields has recently been reported by Milun and Sher (1993).

According to our picture, the Hamming distance plays the role of configurational potential while the bit-error probability plays the role of the controlling temperature of the Gibbs distribution. In fact the Gibbsian temperature, T, is given by

$$T = \frac{1}{\ln(1/\alpha)}. \tag{10}$$

Conditions of large error probability therefore correspond to those of high temperature while conditions of small error probability correspond to low temperature. In Section 5 we will discuss some of the issues related to the control of the bit-error probability in configurational relaxation, presenting an analytical model describing its behaviour.

Our aim in making node updates is to locate an optimum value of the configurational probability $P(S)$. The optimisation procedure is based on a conventional gradient ascent approach. When rather than being binary values the node variables are real-valued quantities in the range $[-1,1]$, the gradient of $P(S)$ is

$$\frac{\partial P(S)}{\partial s_j} = \frac{\beta}{Z} \ln \alpha \sum_{\mu=1}^{Z} \alpha^{H_\mu} \frac{\partial H_\mu}{\partial s_j}, \tag{11}$$

where, from Eq. (4), we have

$$\frac{\partial H_\mu}{\partial s_j} = -\frac{1}{2} \lambda_j^\mu .$$

IIn the case of binary values, the update rule for gradient ascent becomes

$$s_j = \operatorname{sgn}\left\{ \sum_{\mu=1}^{Z} \lambda_j^\mu \alpha^{H_\mu} \right\}. \tag{12}$$

Now, observe that from Eq. (4) this update rule is equivalent to

$$s_j = \operatorname{sgn}\left\{ \sum_{\mu=1}^{Z} \lambda_j^\mu \alpha^{-\frac{1}{2}\sum_{i=1}^{N} s_i \lambda_i^\mu} \right\} \tag{13}$$

which is identical to the ECAM update rule defined by Eqs. (1) and (3) when the base of exponentiation a is taken to be $\alpha^{-1/2}$ or, in other words, when

$$a = \left(\frac{1-p}{p} \right)^{\frac{1}{2}}. \tag{14}$$

Note that under the assumption that the error probability p is less than $\frac{1}{2}$, the base constant a is greater than unity as required by the ECAM model. This provides, therefore, a direct Bayesian interpretation for the ECAM model. When the base of the exponentials a is chosen as in Eq. (14), the ECAM can be viewed as a method for maximising the configurational probability $P(S)$, thereby acting as a type of Bayesian pattern reconstruction process.

It may be of some interest to discuss how our Bayesian framework relates to the storage capacity result derived by Chiueh and Goodman (1991). At this point our analysis will be rather informal; a more rigorous discussion concerning the storage capacity of the ECAM will be presented in Section 5. Consider an input pattern S that is k bits away from the nearest memory pattern from which S is assumed to be derived through a process of memoryless bit corruption. Let $k = \kappa N$, where κ is the fraction of incorrect bits in the pattern S, and define κ' as $\kappa + \frac{1}{N}$. Chiueh and Goodman (1991) showed that when $\kappa' \geqslant (1+a^2)^{-1}$ the asymptotic storage capacity of the ECAM is the largest possible for an associative memory, as determined by Chou (1988). Observe that, under our memoryless corruption assumption, the bit-error probability p represents the expected fraction of incorrect bits in any input pattern, and is therefore intimately related to the parameter κ defined above. Specifically, κ is the actual realisation of a random variable whose expected value is p. Now, from Eq. (14), we obtain

$$p = (1+a^2)^{-1} \tag{15}$$

and thus $p' \equiv p + \frac{1}{N} > (1+a^2)^{-1}$ for all N, which means that Chiueh and Goodman's condition for meeting the largest possible capacity is expected to hold, at least in a statistical sense. We can therefore conclude that our Bayesian framework naturally leads to a choice of the exponentiation base a which is (on the average) the best possible: it allows us to achieve the ultimate upper bound for the asymptotic storage capacity of an associative memory.

The control of the bit-error probability is an important issue for configurational optimisation by gradient ascent. One strategy which is suggested by our analogy with Gibbs distributions is to exert control via a deterministic schedule similar to temperature annealing (Geman and Geman, 1984). In other words, we would reduce the exponential constant towards zero according to a predefined annealing schedule with each time step of associative recall. However, unlike the temperature of an annealing schedule, the bit-error probability has a direct physical meaning in terms of our pattern model: it is related to the Hamming distance between the network configuration and the actual pattern to have undergone corruption. If we assume that the actual pattern is that closest to the network configuration, then we can form the following estimator for bit-error probability

$$p = \frac{\min_{\mu=1,\ldots,Z} H_\mu}{N}. \tag{16}$$

In this way the relaxation scheme can adjust to the prevailing error conditions on the network. In Section 5 we will present a model of the underlying pattern space which allows us to understand how the conditions on the network change after iterative application of the configurational relaxation method.

4. Bayesian interpretation of the Hopfield network

We are now interested in comparing our update rule with its counterpart for the Hopfield network model (Hopfield, 1982). This network consists of an arrangement of nodes in which the inter-nodal connections are characterised by continuous-valued weights which store the training patterns. If w_{ij} is the interconnection weight between the nodes indexed i and j, then the network optimises the criterion $C(S)=\sum_{i,j} w_{ij} s_i s_j$, provided that the weights are symmetric, i.e., $w_{ij}=w_{ji}$. Training is frequently performed using the following Hebbian learning rule:

$$w_{ij}=\frac{1}{N}\sum_{\mu=1}^{Z}\lambda_i^{\mu}\lambda_j^{\mu}. \quad (17)$$

By contrast with our exponential probability criterion, the Hopfield network effectively minimises a quadratic function of the Hamming distances. Operating in pattern recall mode, the network performs gradient descent on $C(S)$ with stepsize proportional to H_{μ}. From Eq. (4), the rule for updating the discrete node variables can be written as

$$s_j=\mathrm{sgn}\left\{\sum_{\mu=1}^{Z}\lambda_j^{\mu}\left(1-\frac{2H_{\mu}}{N}\right)\right\}. \quad (18)$$

This clearly differs from the update rule for configurational relaxation in that it is linear in Hamming distance rather than exponential.

Under conditions in which α is close to unity, the exponentials appearing in Eq. (12) can be approximated in a linear way using the Taylor expansion with the result

$$s_j=\mathrm{sgn}\left\{\sum_{\mu=1}^{Z}\lambda_j^{\mu}\left[1-(1-\alpha)H_{\mu}+\cdots\right]\right\}. \quad (19)$$

This is identical to the Hopfield dynamical rule when $\alpha=1-\frac{2}{N}$. The Hopfield approach can therefore be regarded as a valid approximation to configurational relaxation when $p\simeq\frac{1}{2}$. Later on we will comment on the meaning of this observation in terms of the structure of the pattern space in which we are operating. At this point we will simply note that in the limit when $p=\frac{1}{2}$, the update rule becomes $s_j=\mathrm{sgn}\{\sum_{\mu=1}^{Z}\lambda_j^{\mu}\}$. This means that the update decision is completely devoid of the contextual information conveyed by the global pattern configuration; it means that each node is assigned its most frequent value from the set of training patterns. In other words, the Hopfield memory only draws extremely weakly upon the available contextual information.

Before proceeding, it is worth considering the limit of the relaxation approach when the bit-error probability p is close to zero. In this case the update rule simply selects the bit corresponding to the pattern which is closest in Hamming distance to the configuration of the network, i.e.,

$$s_j=\lambda_j^{\mu},\quad \text{such that } H_{\mu}=\min_{\nu=1,\dots,Z}H_{\nu}. \quad (20)$$

This update rule is close in spirit to the operations performed by the class of Boolean networks proposed by Aleksander (1989) which have been thoroughly analysed by Wong and Sherrington (1989). In this type of network the binary node values are used to construct an address to a memory location which contains an update value. This type of system could clearly handle the minimum Hamming distance criterion described above. We mention that Rohwer (1995) has recently performed a more detailed Bayesian analysis of a related Boolean n-tuple recognition model.

5. The Hamming distance picture

Both configurational relaxation and the Hopfield memory use functions of Hamming distance as the basis for node update decisions. To understand how the two approaches compare, 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 inter-pattern Hamming distance.

Consider a noise corrupted pattern indexed ν. 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_ν follows the binomial distribution:

$$P_s(H_\nu = t) = \frac{N!}{(N-t)!t!} p^t (1-p)^{N-t}. \tag{21}$$

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. Actual bit values at different configuration sites can be treated as independent events occurring with a uniform probability distribution. According to this model the high and low bits are assigned with probabilities q and $1-q$, respectively. The probability that the bits on the same configuration site of any two patterns disagree is $r = 2q(1-q)$. It is easy to show using the independence and uniformity assumptions that the probability distribution for the inter-pattern Hamming distance is again binomial with mean Nr and variance $Nr(1-r)$:

$$P_b(H_\mu = t) = \frac{N!}{(N-t)!t!} r^t (1-r)^{N-t}. \tag{22}$$

In what follows we will confine our attention to the case when the binary labels occur with equal probability, i.e., $r = \frac{1}{2}$.

We can regard pattern recovery as the task of locating the signal distribution $P_s(H_\nu = t)$ above a competing background process $P_b(H_\mu = t)$. Irrespective of how the node updates are made, success can only be anticipated if the overlap between the two distributions is small and the signal distribution is not swamped by background.

At this point it is interesting to note that the Hopfield and configurational relaxation approaches become computationally equivalent under the worst possible case for successful pattern recovery. This is the case when $p \simeq \frac{1}{2}$, which corresponds to the case of signal and background having identical mean Hamming distance.

The two-component model described above allows us to view the ECAM as a decision rule in the measurement domain of Hamming distance. A similar framework for studying the dynamical behaviour of associative memories has already been explored intuitively by Kanerva (1988) and has been formalized by Wong and Sherrington (1989). We are interested in using this picture to understand how the bit-error probability evolves with the iterative node update process. The first step towards meeting this aim is to use the signal and background Hamming distributions to compute the probability that any single stored pattern has a Hamming distance smaller than that for the correct pattern which is indexed ν, i.e., $P(H_\mu \leqslant H_\nu)$. This goal is achieved by performing a discrete convolution of binomial distributions given in Eqs. (21) and (22),

$$P(H_\mu \leqslant H_\nu) = \sum_{t=0}^{N} P_s(H_\nu = t) \sum_{u=0}^{t} P_b(H_\mu = u). \tag{23}$$

If we confine our attention to the case when the binary labels occur with equal probability, i.e., $r = \frac{1}{2}$, then with the binomial forms of $P_s(H_\mu = t)$ and $P_b(H_\mu = u)$ we obtain

$$P(H_\mu \leqslant H_\nu) = \frac{1}{2^N} \sum_{t=0}^{N} \frac{N! p^t (1-p)^{N-t}}{(N-t)!t!} \sum_{u=0}^{t} \frac{N!}{(N-u)!u!}. \tag{24}$$

We are interested in rearranging this series to obtain a polynomial expression in p so that we can approximate $P(H_\mu \leqslant H_\nu)$ for small values of the bit-error probability. Performing a binomial expansion of the first three terms in the series and collecting together terms of order up to p^2, we obtain the following result:

$$P(H_\mu \leqslant H_\nu) = \frac{1}{2^N}\left[1 + N^2 p + \frac{1}{4} N^2 (N-1)(N-3) p^2 + \cdots \right]. \tag{25}$$

Our aim is to compute the probability of bit-errors after relaxational updating has taken place. The condition for error is that at least one of the $Z-1$ competing background patterns produces a configuration with Hamming distance less than H_ν. The background patterns can be regarded as independent events each of which has a uniform probability $P(H_\mu \leqslant H_\nu)$ 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, i.e., $P(\min_\mu H_\mu \leqslant H_\nu)$, is equal to

$$P_{\text{error}}(p) \simeq P\left(\min_{\mu \neq \nu} H_\mu \leqslant H_\nu\right) = 1-\left(1-P\left(H_\mu \leqslant H_\nu\right)\right)^{Z-1}. \tag{26}$$

Substituting Eq. (25) into Eq. (26), and applying the Taylor expansion for small values of p we obtain the following expression for the bit-error probability which applies in the vicinity of the origin:

$$P_{\text{error}}(p) \simeq \frac{Z-1}{2^N}\left[1+N^2p+\frac{1}{4}N^2(N-1)(N-3)p^2\right]. \tag{27}$$

The effect of performing configurational relaxation on the network is equivalent to modifying the bit-error probability. This can be viewed as an iterative process of the form $P_{\text{error}}(p) \mapsto p$. It is the fixed points of this iteration scheme that determine the behaviour of the relaxation process and the feasibility of pattern recovery. These fixed points are the values of the bit-error probability for which $P_{\text{error}}(p) = p$. A detailed numerical analysis of Eq. (27) reveals the existence of three such points; the lower one is close to the origin, i.e., $p \simeq 0$, the upper one occurs at $p=1$ while the third occurs at an intermediate value of p.

An important property of the fixed points of the iterative scheme is whether they are convergent or divergent. Convergent points are those towards which the iterative scheme migrates; divergent points are those from which it migrates. The condition for a fixed point to exhibit the convergent property is

$$\left|\frac{\partial P_{\text{error}}(p)}{\partial p}\right|_{p=P_{\text{error}}(p)} < 1. \tag{28}$$

Numerical analysis reveals that the lower and upper fixed points are both convergent while the intermediate fixed point is divergent. According to Kanerva's terminology, the position of the intermediate fixed point is referred to as the ‘‘critical distance’’. As a concrete example, Fig. 1 shows the curve $P_{\text{error}}(p)$ superimposed on the line of unit gradient for a configuration of 15 nodes; moving from left to right the individual curves correspond to $Z=300$, 150, 50, 25, and 5 stored patterns.

The intersections of curve and line are the fixed points of the iterative scheme $P_{\text{error}}(p) \mapsto p$. For successful pattern recovery the relaxation scheme must be initialised below the intermediate fixed point in order for the bit-error probability to monotonically

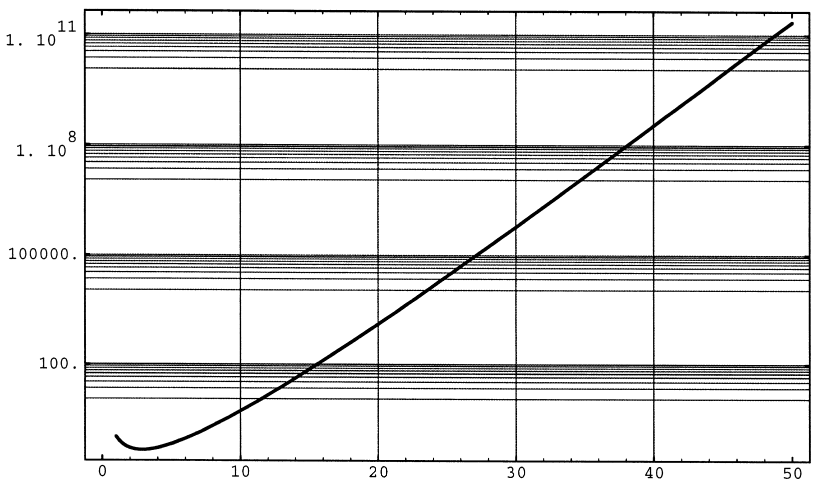

Fig. 1. Fixed points of bit-error probability.

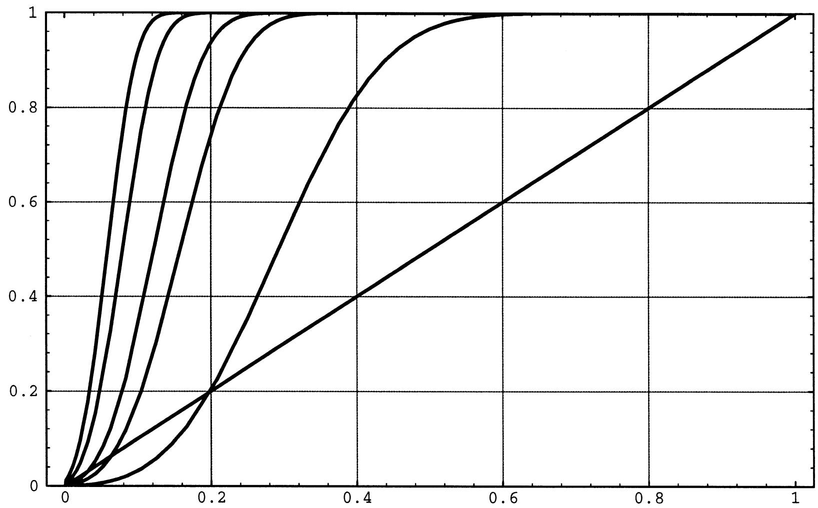

Fig. 2. Degeneracy property for a network with 15 nodes. The curve corresponds to the behaviour of $P_{error}(p)$ with $Z=73$, the theoretical maximum storage capacity for a network with 15 nodes.

decrease until coming to rest at the lower fixed point. In other words, the initial Hamming distance must be less than the critical distance. If this is not the case, then the bit-error probability will grow iteration by iteration, eventually coming to rest at the upper fixed point; this will have catastrophic results on quality of the recovered configuration.

The Kanerva picture of associative recall has been formally used by Wong and Sherrington (1989) to compute the storage capacity of Aleksander's memory (Aleksander, 1989). The basic idea is to compute the conditions under which the lower and intermediate fixed points become degenerate. Under these circumstances, there is no range of initial Hamming

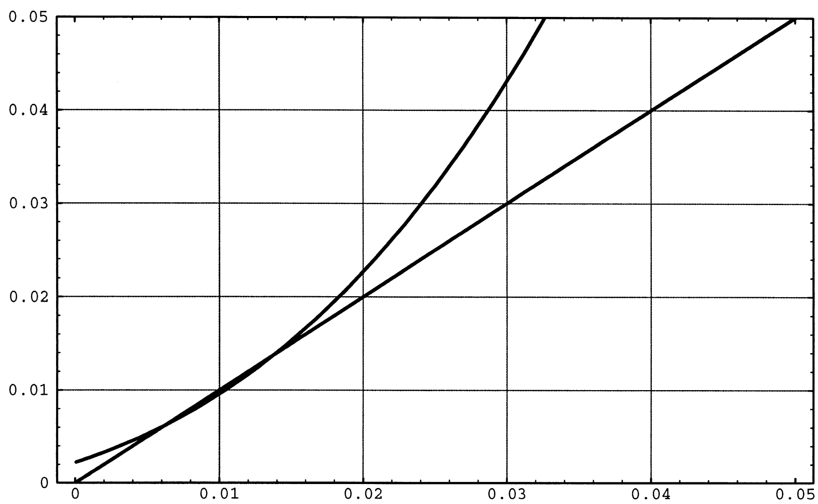

Fig. 3. Theoretical storage capacity as a function of the number of nodes in the network.

distance for which iterative improvement in the bit-error probability is possible. This degeneracy condition is illustrated in Fig. 2. Following Wong and Sherrington (1989), the conditions for degeneracy may be determined by approximating the polynomial expression for $P_{\mathrm{error}}(p)$ appearing in Eq. (27) to second-order in p. The fixed points are the two solutions of the quadratic equation

$$1+\left(N^2-\frac{2^N}{Z-1}\right)p+\frac{1}{4}N^2(N-1)(N-3)p^2=0. \tag{29}$$

The degeneracy condition for this quadratic gives the following expression for the number of stored patterns Z in terms of the numbers of nodes of the network N

$$Z=1+\frac{2^N}{N^2\pm N\sqrt{(N-1)(N-3)}}. \tag{30}$$

In the limit of large N, the storage capacity is approximately equal to

$$Z\simeq\frac{2^{N-1}}{N^2}. \tag{31}$$

Fig. 3 shows a plot of this theoretical storage value as a function of the length of memory patterns. Note how, once $N>4$, the storage capacity rises exponentially. It should be noted that our approach to computing storage capacity is somewhat different to that adopted by Chiueh and Goodman (1991). Rather than imposing a threshold on acceptable error rate, we draw on Kanerva's picture which attempts to understand associative recall in terms of a critical Hamming distance above which successful pattern recovery is impossible. Using a simple model of the pattern space in which the associative memory operates, we effectively compute the condition for the critical distance to vanish. In consequence, our storage result depends only upon the size of the network. By contrast, Chiueh and Goodman's result depends on the prespecified error threshold. However, both results share the common feature of being exponential in the number of nodes.

It is interesting to note that the relaxation process has a non-zero saturated error $P_{\mathrm{error}}(0)$. This corresponds to the probability of incorrectly recovering one of the competing stored patterns. From Eq. (27), this probability is equal to

$$P_{\mathrm{error}}(0)=\frac{Z-1}{2^N}. \tag{32}$$

This is clearly in accordance with our intuitions; there are $Z-1$ possible mistakes that can be made in the global pattern classification process which have to be drawn from a configuration space of 2^N possibilities. The existence of this saturated error means that if the relaxation scheme is initialised with an unrealistically low bit-error probability, then it will be attracted upwards in value to the lower fixed point.

6. Conclusions

We have demonstrated how Chiueh and Goodman's ECAM model can be entirely derived within a Bayesian framework, and can therefore be interpreted as performing a type of Bayesian pattern reconstruction process. The basis for this work has been a novel relaxation scheme developed in the context of pattern recognition and computer vision, which involves direct probabilistic modelling of the pattern corruption process. We have additionally shown that the Hopfield memory can be given a Bayesian interpretation, too. Viewed from the standpoint of a pattern space model, this interpretation corresponds to the case when the probability of individual bit-errors on the network is approximately $\frac{1}{2}$. In terms of pattern recovery this is almost the worst possible case; it corresponds to complete overlap between the Hamming distance distribution for corrupt and competing stored patterns. In the limit of small bit-error probability our relaxation scheme becomes equivalent to the Boolean network proposed by Aleksander.

Based on a simple model of the pattern space suggested by Kanerva (1988), we have theoretically explored the behaviour of the bit-error probability under iterative node updating. This leads us to several conclusions. Firstly, the ECAM iterative scheme has convergent fixed points for values of the bit-error probability close to zero and unity. A further interesting property is that the model has a saturated error which originates from the effect of interpattern competition. Finally, the ECAM approach is completely inoperable when its fixed points become degenerate. This degeneracy condition led us to derive an expression for the storage capacity of the model which turns out to be exponential in the number of nodes in the network. A similar result was also derived by Chiueh and Goodman (1991) but, in contrast with theirs, our result does not depend on an adjustable parameter which affects both the bit-error probability and the storage capacity of the model.

A number of alternative network models have recently been proposed which use exponential response computational units as their basic building block (Jeng et al., 1990; Wang and Don, 1995; Wang and Lee, 1995; Chiueh and Tsai, 1993; Geva and Sitte, 1991). As a matter of future investigation, it would be interesting to exploit the configurational relaxation approach described in this paper (or simple variations thereof) in an attempt to provide these models with a Bayesian interpretation. This would allow us to better understand some of their fundamental dynamical properties.

References

Aleksander, I., 1989. The logic of connectionist systems. In: Aleksander, I. (Ed.), Neural Computing Architectures: The Design of Brain-Like Machines. North Oxford Academic Publishers, London, UK, pp. 133–155.

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

Chiueh, T.D., Goodman, R.M., 1991. Recurrent correlation associative memories. IEEE Trans. Neural Networks 2, 275–284.

Chiueh, T.D., Tsai, H.K., 1993. Multi-valued associative memories based on recurrent networks. IEEE Trans. Neural Networks 4, 364–366.

Chou, P.A., 1988. The capacity of the Kanerva associative memory is exponential. In: Anderson, D.Z. (Ed.), Neural Information Processing Systems. American Institute of Physics, New York, NY, pp. 184–191.

Durbin, R., Szeliski, R., Yuille, A., 1989. An analysis of the elastic net approach to the travelling salesman problem. Neural Comput. 1, 348–358.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Machine Intell. 6, 721–741.

Geva, S., Sitte, J., 1991. An exponential response neural net. Neural Comput. 3, 623–632.

Hancock, E.R., Kittler, J., 1990. Discrete relaxation. Pattern Recognition 23, 711–733.

Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proc. Nat. Acad. Sci. USA 79, 2554–2558.

Hummel, R.A., Zucker, S.W., 1983. On the foundations of relaxation labeling processes. IEEE Trans. Pattern Anal. Machine Intell. 5, 267–287.

Jeng, Y.J., Yeh, C.C., Chiueh, T.D., 1990. Exponential bidirectional associative memories. Electron. Lett. 26, 717–718.

Kanerva, P., 1988. Sparse Distributed Memory. MIT Press, Cambridge, MA.

McEliece, R.J., Posner, E.C., Rodemich, E.R., Venkatesh, S.S., 1987. The capacity of the Hopfield associative memory. IEEE Trans. Inform. Theory 33, 461–482.

Milun, D., Sher, D., 1993. Improving sampled probability distributions for Markov random fields. Pattern Recognition Lett. 14, 781–788.

Rohwer, R.J., 1995. Two Bayesian treatments of the *n*-tuple recognition method. In: Proc. 4th IEE Internat. Conf. Artif. Neural Networks, Cambridge, England, pp. 171–176.

Wang, C.C., Don, H.S., 1995. An analysis of high-capacity discrete exponential BAM. IEEE Trans. Neural Networks 6, 492–496.

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

Wilson, R.C., Hancock, E.R., 1997. Structural matching by discrete relaxation. IEEE Trans. Pattern Anal. Machine Intell. 19, 634–648.

Wong, K.Y.M., Sherrington, D., 1989. Theory of associative memory in randomly connected Boolean neural networks. J. Phys. A 22, 2233–2264.

Yuille, A., 1990. Generalised deformable models, statistical physics and matching problems. Neural Comput. 2, 1–24.