Pattern retrieval in threshold-linear associative nets

Martin W Simmen†, Alessandro Treves‡ and Edmund T Rolls§

† Centre for Cognitive Science, University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, UK
‡ Biophysics and Cognitive Science, SISSA, via Beirut 2, 34013 Trieste, Italy
§ Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK

Abstract. Networks of threshold-linear neurons have previously been introduced and analysed as distributed associative memory systems. Here, results from simulations of pattern retrieval in a large-scale, sparsely connected network are presented. The storage capacity lies near $\alpha = 0.8$ and 1.2 for binary and ternary patterns respectively, in reasonable accordance with theoretical estimates. The system is capable of retrieving states strongly correlated with one of the stored patterns even when the initial state is a highly degraded version of one of these patterns. This pattern completion ability holds for an extensive number of memory patterns, up to $\alpha \approx \alpha_c/2$, thereby increasing the credibility of the model as an effective associative memory.

1. Introduction

The design and study of networks capable of functioning as associative memory systems has long been a central theme in the field of theoretical neural networks. Speculation that brain regions rich in feedback loops, such as CA3 in the hippocampus, the olfactory cortex, and neocortex, might implement associative memories has led to the topic also featuring prominently in computational neuroscience. In most of the early models, each neuron is simply described by a binary variable representing its output activity, with the neurons linked by a network of binary-valued synapses (Willshaw *et al* 1969, Marr 1971, Gardner-Medwin 1976). The class of alternative models ushered in by Hopfield (1982, 1984) had the great virtue of being amenable to analysis by techniques from statistical physics (see Amit (1989) for a review), though they too had many unbiological features. Recent years have seen much progress, for both classes of model, in making the models more biologically realistic; in the binary-synapse case, see for example Buckingham (1991) and Bennett *et al* (1994).

The current paper is concerned with a model developed by Treves (1990, 1991a) from the Hopfield model, in which the neuronal response function is taken to be threshold-linear in form. Although the system can be specified in completely formal terms, it is as a simple model of the pyramidal cells in CA3 that most discussion of it has occurred (Treves and Rolls 1991, 1992), this being in the context of the hypothesis that the hippocampus serves as a short-term memory store (Marr 1971, Rolls 1987, Treves and Rolls 1994). The present study uses computer simulations to investigate pattern retrieval in a large-scale network of threshold-linear cells. To mimic conditions in CA3 (Amaral *et al* 1990), the cells are interconnected in a sparse, asymmetric manner. Modelling neurons as graded response

units with no intrinsic saturation in their rates introduces qualitative differences in relaxation dynamics, such as the disappearance of spin-glass states (Treves 1991b). Sparse connectivity is expected to render the dynamics even less rigid. It is interesting therefore to examine the effects on static properties for which related analytical quantifications are available, such as storage capacity, and on dynamical properties explorable only with simulations, such as attraction basins. Previous studies have investigated the storage and retrieval of low activity patterns in symmetrically connected nets of graded-response neurons, e.g. Amit and Tsodyks (1991b), Kühn and Bös (1993), but with an emphasis on seeking retrieval states with low rates. Our emphasis is on studying the effects of diluted connectivity and asymmetry.

After a definition of the model in section 2, some relevant analytical results regarding its performance are reviewed in section 3, followed by a description of the simulation aims in section 4. Relevant aspects of the simulation method are covered in section 5. The results are presented in section 6 and the paper closes with a brief discussion of models which use more realistic dynamics.

2. The model

The model is similar to that analysed in Treves (1991a). Readers interested in a fuller justification of the biological relevance of the model than that provided below should consult Treves (1991a) and Treves and Rolls (1991). There are N cells, indexed by i, each of which crudely represents a cortical pyramidal cell. Associated with each cell is a single dynamical variable V_i, a positive scalar denoting the cell's spiking rate averaged over a short interval. The cells are sparsely interconnected by an asymmetric network of 'synapses'. The value of V_i is determined (via a response function) by the sum of the inputs to the cell, denoted h_i. This sum is over inputs from the other pyramidal cells, external sources, and inhibitory plus other regulatory cells. Specifically

$$h_i = \sum_{j \neq i} J_{ij}^c V_j + \frac{s^\mu(\eta_i^\mu - a)}{a} + b\left(\frac{1}{N}\sum_j V_j\right). \tag{1}$$

The synapses between real pyramidal cells are excitatory; in the current model this positive synaptic efficacy is regarded as the sum of a positive baseline term plus a term, J_{ij}^c, describing the effect on the synapse connecting cell j to cell i of encoding a set of firing patterns according to some learning rule. We restrict ourselves to the particular case of using a covariance rule to store random binary or ternary patterns, as these cases are amenable to analysis. The case of ternary encoding also serves as a check that the network's operation is not dependent on the use of binary representations. Each pattern η^μ, $\mu = 1, \ldots, p$, represents a set of firing rates across the N cells. Each η_i^μ is drawn from the probability distribution $P(\eta)$. For binary patterns

$$P(\eta) = a\delta(\eta - 1) + (1 - a)\delta(\eta) \tag{2}$$

and for ternary patterns we use the representative distribution (Treves 1990)

$$P(\eta) = \tfrac{1}{3}a\delta\left(\eta - \tfrac{3}{2}\right) + a\delta\left(\eta - \tfrac{1}{2}\right) + \left(1 - \tfrac{4}{3}a\right)\delta(\eta). \tag{3}$$

In both cases a is the expected value of the activity ($\langle\eta\rangle$) and the *sparsity* ($\langle\eta\rangle^2/\langle\eta^2\rangle$) in the patterns. The covariance rule prescription for the modifiable component of the synaptic efficacies is

$$J_{ij}^c = c_{ij} \frac{1}{Ca^2} \sum_{\mu}^{p} (\eta_i^\mu - a)(\eta_j^\mu - a) \tag{4}$$

where c_{ij} is a 0/1 variable indicating whether a synaptic connection exists or not. Each c_{ij} ($i \neq j$) is drawn from the probability distribution

$$P(c) = \frac{C}{N-1}\delta(c-1) + \left[1 - \frac{C}{N-1}\right]\delta(c) \tag{5}$$

where the parameter C is the expected fan-in and fan-out of each pyramidal cell. So, for example, setting $C = N - 1$ yields the fully connected network studied in Treves (1990), whereas setting $C \ll N$ yields the strongly diluted asymmetric network studied in Treves (1991a). We shall consider the regime of asymmetric moderate dilution, taking $C = 0.1N$. This is a step towards modelling the effects of the low probability (~ 0.02) of contact between any two pyramidal cells in the CA3 region of the hippocampus (Amaral *et al* 1990).

The second term in the h_i expression is the effect of an external signal evoking one of the stored patterns, $\boldsymbol{\eta}^\mu$, with strength s^μ. The inclusion of this term allows us to present cues to the network during the retrieval phase in two distinct ways: either via the external signal, or by initializing V to correlate with a pattern, setting s^μ to zero. Further details on these approaches are given in section 5.

The function $b(x)$ models the net effect of a number of terms, including the result of activity being transmitted through the baseline component of the pyramidal–pyramidal synapses, fast inhibition from local inhibitory interneurons, and the influence of various modulatory factors. As none of these terms is pattern specific, the simplifying assumption is made that the function $b(x)$ acts solely to regulate the mean activity in the network, keeping it near to a, the mean activity in the stored patterns. As in earlier work (Treves 1990), a cubic function is used:

$$b(x) = \kappa(a' - x)^3 \tag{6}$$

with a' being an adjustable parameter set near or equal to a.

A threshold-linear response function is assumed, namely

$$V = \Theta[h - T_{\text{hr}}]\, g\,(h - T_{\text{hr}}) \tag{7}$$

where T_{hr} is the threshold below which the input elicits no output, g is the gradient of response above T_{hr}, and $\Theta[\cdots]$ is the Heaviside function. Evidence for the biological validity of such a response function comes from data on how the adapted firing rate of pyramidal cells varies with the strength of current injected into the soma (see Treves and Rolls (1991) and references therein), plus attempts to derive theoretical $V(h)$ functions for cells modelled as 'integrate-and-fire' devices (Amit and Tsodyks 1991a).

3. Relevant analytical results

Using standard methods from statistical physics, Treves (1990, 1991a) has obtained results for the storage capacity as a function of the pattern sparsity parameter a in the limit $C \to \infty$ for the extremal cases of $C = N - 1$ and $C/N \to 0$, respectively. The storage capacity (α_{c}) is the maximum value of the memory loading parameter $\alpha \equiv p/C$ for which the network possesses 'retrieval states', i.e. solutions in which the overlap between the network state V and one of the stored patterns is both stationary and macroscopic. The overlap with the μth pattern is defined as: $\widehat{x^\mu} = \frac{1}{Na}\sum_i V_i(\eta_i^\mu - a)$. The quality of the retrieval can be gauged by the overlap size or by how much information the state conveys about the stored pattern. Both of these measures disappear discontinuously at $\alpha = \alpha_{\text{c}}$. Furthermore, retrieval states are only found for values of g lying in a certain interval, which shrinks and vanishes as $\alpha \to \alpha_{\text{c}}$ from below.

There are no analytical results available for the moderately diluted asymmetric network studied in this paper. However, the analyses of the two extremal architectures discussed above serve as a guide as to how well this intermediate architecture might perform as an associative memory. In this regard, it is important to note that in the limit of sparse coding ($a \rightarrow 0$) the storage capacities of the fully connected net and the highly dilute net become equivalent. The current work is concerned with the storage and retrieval of moderately sparse patterns, specifically those characterized by $a = 0.1$. By interpolation from the α_c versus a plots in Treves (1990, 1991a), the α_c values in the case of binary stored patterns with $a = 0.1$, using $s^\mu = 0$, are approximately 0.35 and 1.0 for the fully connected and highly dilute nets, respectively.

Furthermore, results have been obtained for the *symmetrically* diluted case, i.e. for networks in which $c_{ij} = c_{ji}$ (D O'Kane, personal communication). In the notation of Treves (1990, 1991a), the result is expressed in terms of two equations

$$1 = \alpha A_3 \left[\gamma / A_1^2 + (1 - \gamma) / A_2^2 \right]$$

$$1/(gT_0) = A_2 + \alpha(A_2 - A_1) \left[\gamma / A_1 + (1 - \gamma) / A_2 \right]$$

which have to be satisfied simultaneously for retrieval states to exist. Here, $\gamma = C/N$ is the degree of symmetric dilution, T_0 measures the variance of the postsynaptic term of the learning rule (see equation (16) of Treves and Rolls (1991)), and A_1, A_2, A_3 are the relevant averages over P_η which enter the capacity equations (see equations (21) of Treves and Rolls (1991)). For further clarification, for example on how to extract the information capacity, we refer the reader to Treves (1990) and Treves and Rolls (1991). Unsurprisingly, the predicted performance is qualitatively the same as that described above, with $\alpha_c = 0.79$ for $C/N = 0.1$, $a = 0.1$ binary patterns, and $s^\mu = 0$.

In performing the analysis the assumption is made that in the retrieval phase $\langle V \rangle = a$. The sparsity of the retrieved representation ($\langle V \rangle^2 / \langle V^2 \rangle$), denoted a_r, is not explicitly constrained. For ternary encodings, the analysis predicts that a_r is usually less than a. Lowering the sparsity generally allows the storage capacity to rise (Treves 1990, 1991): for the ternary distribution in equation (3), with $a = 0.1$, analysis of the symmetrically diluted system yields $\alpha_c \approx 1.2$.

4. Simulation aims

Broadly, the purpose of the numerical simulations is to assess how well a large, sparsely connected network of threshold-linear units can retrieve sparse patterns previously encoded in the synaptic network via the covariance rule. Specifically, there are five aims.

(i) to study how retrieval quality varies with the number of patterns stored.
(ii) to estimate the storage capacity of the system.
(iii) to study the types of final states produced by the network.
(iv) to study how retrieval quality depends on the size of the cue presented, i.e. to assess the network's 'pattern completion' ability.
(v) to study the effect of a persistent external cue during retrieval.

A subsidiary aim (where possible) is to compare these simulation results with analytical results for the related symmetrically dilute system. In so doing, the emphasis will be on checking for broad consistency rather than exact correspondence. This is due to the two models differing not just in their connectivity structure, but also in their dynamics (see below), plus the necessity of taking a mean-field approximation and the $C \rightarrow \infty$ limit in the analysis.

5. Simulation method

Dynamics

The simulation program implements the model defined in section 2, but with slightly different dynamics: cells are updated in parallel according to

$$V(t + \Delta t) = (1 - \Delta t)V(t) + \Theta[h(t) - T_{\text{hr}}] \; g(h(t) - T_{\text{hr}})\Delta t. \tag{8}$$

This is a forward Euler integration scheme for the differential equation

$$\frac{\text{d}V(t)}{\text{d}t} = -V(t) + \Theta[h(t) - T_{\text{hr}}] \; g(h(t) - T_{\text{hr}}). \tag{9}$$

These dynamics were chosen because the dynamics of real neurons are also governed by continuous equations (albeit more complex equations than (9)) and because they can be implemented more efficiently on parallel computers than serial dynamics†. A value of 0.2 was used for Δt (comparison runs using $\Delta t = 0.02$ required far longer run times but produced similar results, both in terms of average retrieval rates and with respect to whether individual trials led to retrieval or not).

Parameter values

As mentioned earlier, C/N is set equal to 0.1, as is a. Given these two sources of sparsity, a large network is required in order to get sensible statistical results. A study of finite-size effects (summarized in figure 2) suggested that $N = 8192$ would be an acceptable compromise between the desire for a biologically realistic large N and the constraint of limited computing resources, thus $N = 8192$ was the standard network size. The value of g was chosen (at each α value) to be that expected to maximize the information carried by the final state about the stored pattern, according to the analysis of the symmetric $C/N = 0.1$ case. In the case of binary patterns, for example, these values decrease monotonically from 0.36 at $\alpha = 0.1$ to 0.15 at $\alpha = 0.79$‡. (In trials with α values above 0.79, g was set to 0.15.) The threshold parameter T_{hr} was set to zero. A value of 10^5 was used for κ; a high value is needed to keep the activity near to a but a simple stability analysis using equations (1) and (8) indicates that values above O(10^5) lead to large oscillations of the activity. Even with $\kappa = 10^5$ some small oscillations remain but these do not seriously disturb the operation of the system. For binary stored patterns a' was set to 0.1; for ternary patterns setting $a' = 0.05$ yielded better results.

Presentation of the cue

There are two ways of supplying the network with a cue, that is, a pattern related to one (taken to be the μth) pattern encoded in the synaptic matrix. In the 'external' mode, the initial cell activities V_i are drawn from the pattern element distribution $P(\eta)$ so that they are random with respect to the activities in the target pattern η^μ, and a persistent signal evoking the μth pattern is applied by setting $s^\mu > 0$ in (1). This signal represents an idealization of how external afferents to a network might convey a cue.

---

† The simulations were performed on a Thinking Machines CM-200 SIMD parallel computer.

‡ Varying g in this manner is biologically reasonable, as the gain is considered to reflect not just the intrinsic response properties of the cells, but also their modulation, e.g., by inhibition, and as such it should be dependent on current conditions, including the memory load. At a computational level, theoretical work on associative networks of graded-response cells has also shown that lowering the gain as α increases is a sensible strategy (Waugh *et al* 1991).

In the 'internal' mode, the cue defines the initial V_i, and $s^\mu = 0$. Clearly for an associative memory to function in a wider system, any such 'internal' cue must itself have been generated by activity external to the network (Treves and Rolls 1992). Nevertheless the internal cueing mode is simple to study and we shall focus primarily on it. A cue is formed from one of the stored patterns η^μ by replacing a randomly selected fraction $(1 - f)$ of the pattern elements with values drawn from the appropriate probability distribution $P(\eta)$.

Performance measures

Several measures can be used to quantify how well the state V matches a stored pattern η^μ. One is the mean information I_c (in bits) carried by an element of V about the corresponding element of η^μ. The procedure used for computing I_c is outlined in Treves (1990). The overlap $\widehat{x^\mu}$ defined in section 3 is easier to compute, but unfortunately has an upper bound that is rather sensitive to the characteristics of the vectors V and η^μ. In practice, a more useful measure is the Pearson correlation coefficient

$$r^\mu = \frac{\langle V\eta^\mu \rangle - \langle V \rangle \langle \eta^\mu \rangle}{\sqrt{(\langle V^2 \rangle - \langle V \rangle^2)(\langle (\eta^\mu)^2 \rangle - \langle \eta^\mu \rangle^2)}} \tag{10}$$

which has maximum value 1.0. This correlation is computed after every update and is used in determining when to terminate a trial. Each trial runs for a minimum of 50 iterations and a maximum of 200, and may be terminated between these limits if r^μ exceeds 0.95, or if the average r^μ value over the previous 20 iterations differs from the current r^μ value by less than 0.02. (A limited set of comparison runs in which trials were only terminated after 200 iterations gave similar levels of retrieval performance.)

A useful measure of the system's overall information capacity is the total amount of information (about the ensemble of p patterns) stored in and retrievable from the network. This can be estimated as pNI_c; normalizing by the number of synapses yields the retrievable information per synapse αI_c, henceforth simply denoted I, as in Treves (1990, 1991a).

Network generation

With the exception of the results on finite-size effects and those involving ternary patterns, all of the results reported herein were obtained from networks possessing the same connectivity matrix $\{c_{ij}\}$; additional test runs using different realizations of this matrix yielded similar statistical results. A single synaptic strength matrix $\{J_{ij}^c\}$ was generated for each α value studied, using a completely different ensemble of patterns at each α value. Unless stated otherwise, for each (α, f) combination studied, twenty trials were run, each using a cue generated from a different member of the pattern ensemble.

6. Results

For simplicity we shall focus first on results obtained using storage of binary patterns, leaving a brief presentation of equivalent results from the ternary case until the end of the section.

Storage capacity

To assess whether an auto-associative network is at all capable of producing quality retrieval states we consider the 'best-case scenario' for retrieval, in which the network is set off in a state matching a stored pattern and we observe whether the dynamics moves the state away
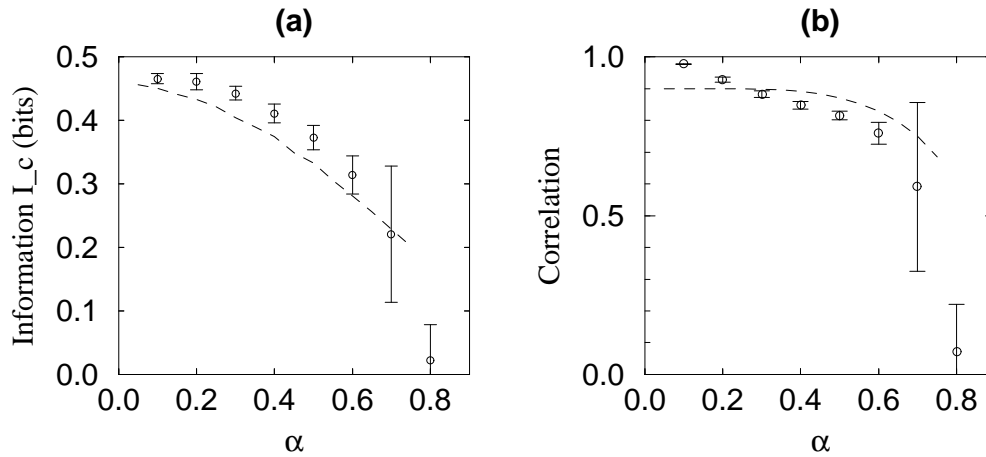
Figure 1. Retrieval quality measures as a function of loading α, given full cue. Data points are mean values over 20 trials, error bars denote standard deviations. Dashed lines denote analytical results. (*a*) Information retrieved per pattern per cell (I_c) versus α. To get a sense of scale, the mean information (in bits) per cell in a stored pattern is $-(a \ln a + (1 - a) \ln(1 - a))/\ln 2$, which for $a = 0.1$ equals 0.469. (*b*) r^μ versus α. Note that the analytical curve in (*b*) is for $\widehat{x^\mu}$, not r^μ. Although the two data sets are thus not directly comparable, the $\widehat{x^\mu}$ data are still shown for the purpose of illustration.
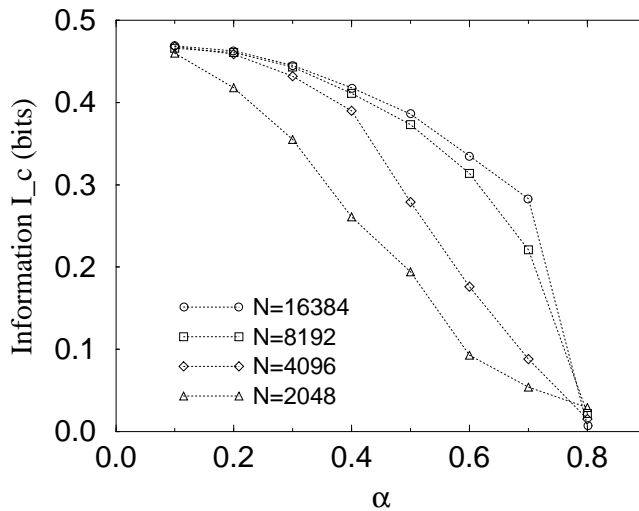
Figure 2. The effect of the network size (N) on retrieval quality. The individual plots are each of the form shown in figure 1(*a*). Error bars have been omitted for clarity; the lines connecting data points are purely for visual aid. Networks with $N = 8192$ were used in simulations reported elsewhere in this paper.

from this point. In the terminology described earlier this corresponds to taking $f = 1$ in the internal cueing mode. The results, shown in figure 1, indicate a storage capacity value (α_c) near 0.8, consistent with the theoretical value of 0.79 for the related symmetrically dilute network.

Both the information retrieval and correlation measures show an increasingly steep
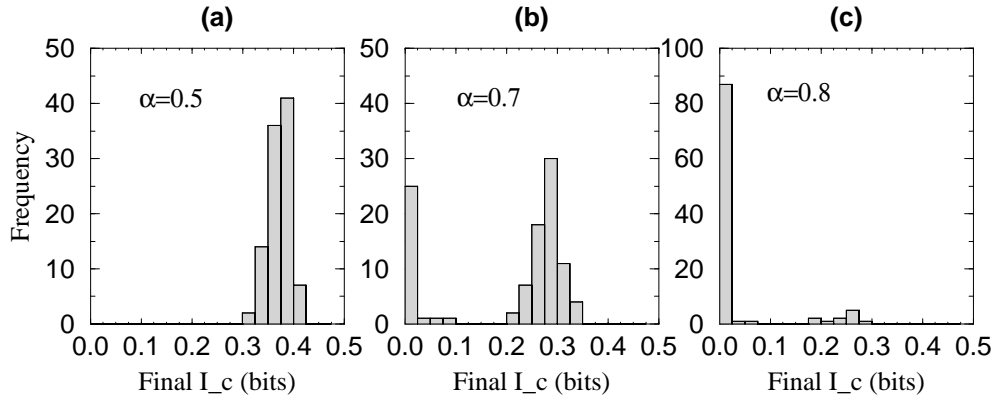
Figure 3. Frequency distribution of I_c values in the final states for various levels of the loading α. At each α level, 100 trials were performed, each using a different stored pattern as the cue. The bin width is 0.025. (*a*) $\alpha = 0.5$, (*b*) $\alpha = 0.7$, (*c*) $\alpha = 0.8$.

decline as α approaches α_c, suggestive of the first-order transition predicted analytically. Further evidence for this comes from two directions. First, a study of finite size effects (figure 2) shows that the 'transition' becomes more sharply defined as the system size increases. Second, an examination of the final states shows that the quality of individual retrieval states does not decay continuously to zero as $\alpha \rightarrow \alpha_c$. Indeed this is already suggested by the large standard deviations in figure 1 at $\alpha = 0.7$, which indicate that the final states are non-homogeneous. Figure 3 shows the distribution of I_c values found in the final states for $\alpha = 0.5$, 0.7, and 0.8, again using $f = 1$. In going from $\alpha = 0.5$ to 0.7 the frequency distribution shifts slightly to lower I_c values but more importantly a second peak opens up around $I_c \approx 0$; by $\alpha = 0.8$ this new peak dominates. Thus, as $\alpha \rightarrow \alpha_c$, the number of stored patterns having corresponding retrieval states decreases sharply towards zero, whilst the information content of those retrieval states which do exist decreases gradually, remaining at approximately half the value in the stored patterns even at $\alpha \approx \alpha_c$†. Note that qualitatively this behaviour is similar to that observed in studies of the standard Hopfield model, although quantitatively the retrieval states just below α_c have much higher pattern correlations in the Hopfield model (Amit 1989). For the trials which fail to reach a retrieval state, scrutiny of the individual cell activities V_i reveals that they follow a wandering trajectory over time, suggestive of the chaotic behaviour expected in networks with asymmetric connections (see, for example, Parisi (1986) and Tirozzi and Tsodyks (1991)).

Basins of attraction

The results just presented demonstrate that the network possesses stationary states strongly correlated with the firing patterns embedded in the synaptic matrix. In practice, for the network to be useful as an auto-associative memory, these high correlation states need not just to exist but also to be reachable from initial states with low correlation. In other words, the high correlation stationary states need to be attractors with wide basins of attraction.

To test whether this pattern completion property exists, trials were conducted using

---

† The analytical I_c values in figure 1(*a*) are those of the retrieval states, therefore they do not reflect the fact that for finite N systems the number of retrieval states diminishes as $\alpha \rightarrow \alpha_c$.
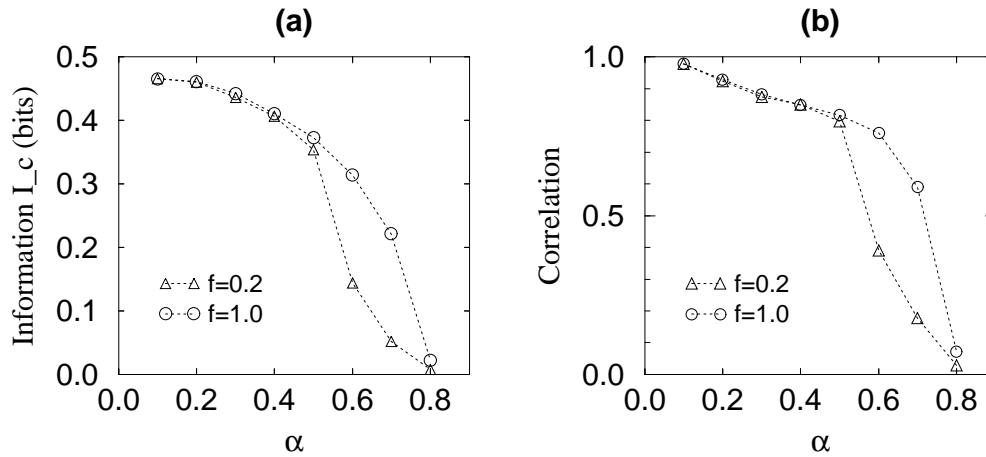
Figure 4. Retrieval quality measures as a function of loading α for two cue sizes, $f = 0.2$ and $f = 1.0$ (redrawn from figure 1). Data points are mean values over 20 trials, error bars omitted for the sake of clarity. (*a*) Information retrieved per pattern per cell (I_c) versus α. (*b*) r^{μ} versus α.
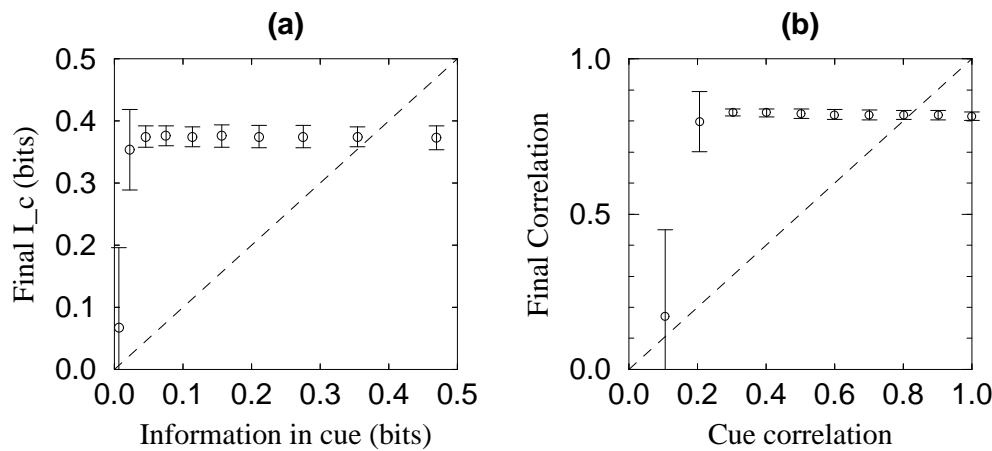
Figure 5. Retrieval quality measures as a function of the initial cue, for $\alpha = 0.5$. f is varied from 0.1 to 1.0, in increments of 0.1. Data points are mean values over 20 trials, error bars denote standard deviations. Dashed lines denote the hypothetical case in which the network performs no processing. (*a*) Information measure I_c. (*b*) Pearson correlation measure r^{μ}.

small cues. Specifically, f was set to 0.2, such that 20% of the elements in the initial V matched those in a particular pattern, the remaining 80% being assigned randomly. The results are shown in figure 4, alongside the $f = 1$ data from figure 1 for comparison. Up to $\alpha = 0.5$ retrieval quality is undiminished by the reduction in cue size; beyond $\alpha = 0.5$ the combination of little cue information and interference in the synaptic weights due to the many stored patterns renders the evolution of high correlation states unlikely.

From the data in figure 4(*a*) it is straightforward to compute values of $I(\alpha)$. For the 'best-case' scenario of $f = 1$, $I(\alpha)$ reaches a maximal value of 0.19 ± 0.02 bits per synapse at $\alpha = 0.6$. It is notable that in the much more realistic case of noisy cues ($f = 0.2$) the maximal $I(\alpha)$ value is essentially the same: 0.18 ± 0.03 bits per synapse, at $\alpha = 0.5$.

Figure 5 shows the effect of systematically varying the cue size at $\alpha = 0.5$. The plateau of good retrieval indicates that pattern completion† is possible over a wide range of cue sizes, i.e. there is a wide basin of attraction to the retrieval state. As in the earlier case, in conditions where the mean retrieval quality is poor (here, when $f = 0.1$), scrutiny of the individual states reveals a sharply bimodal distribution, with no states having an intermediate level of quality (data not shown).

Retrieval with a persistent external cue

Analysis of the fully connected network led to the conclusion that the effect of a persistent external cue was (not surprisingly) to aid pattern retrieval and raise α_c (Treves 1990). Simulations with the current model also show these effects when s^μ is assigned a positive value. We follow the approach of Treves and Rolls (1992), and use neuroanatomical data to select a biologically relevant value for s^μ. In their model, the network of CA3 pyramidal cells in the hippocampus can function as an auto-associative memory, with perforant path fibres from the entorhinal cortex serving to convey cues during retrieval. Now, data from the rat (Amaral *et al* 1990) suggest that a CA3 pyramidal cell has approximately four times as many dendritic spines innervated by CA3 recurrent collaterals as it does spines from perforant path fibres. To fix s^μ we therefore demand that if $V = \eta^\mu$, the contribution to h_i in equation (1) from the recurrents is four times that from the external cue. This leads to the prescription $s^\mu = (1 - a)/4$.
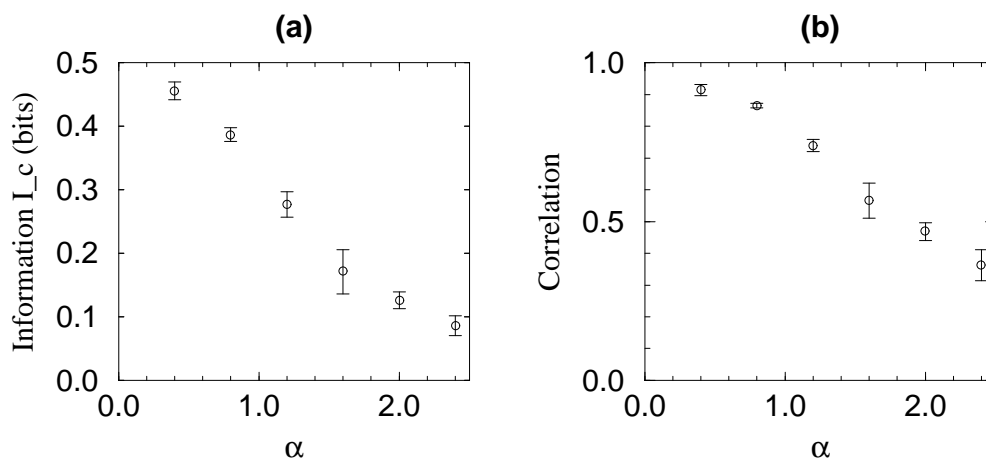
Figure 6. Retrieval quality measures as a function of loading α, given external cueing. Data points are mean values over 20 trials, error bars denote standard deviations. The g values used are the same as in figure 1, with $g = 0.15$ for $\alpha \geqslant 0.8$; i.e. they are not optimized for external cueing. (*a*) Information retrieved (I_c) versus α. (*b*) r^μ versus α.

Simulation results using this external cueing mode are shown in figure 6. There are two features to note. Firstly, that the performance is better than with internal cueing, both in terms of the range of α values over which some retrieval is possible, and with respect to the absolute retrieval quality in the range of α values in which internal cueing shows any retrieval at all. Secondly, even at high α values both retrieval measures seem to be falling only gradually, a feature also seen in the standard Hopfield model (Amit 1989).

† Recall is not strictly 'complete', in that due to the extensive number of loaded patterns even the best final states do not exactly match their corresponding pattern.

Sensitivity to gain parameter

As reported in section 3, it is expected that the gain (g) of the neuronal response function needs to lie in a certain range for retrieval to occur. It is important to check this, for if the model required exquisitely fine tuning of g in order for it to function well, then the model's relevance to biological memory would be greatly weakened. In the simulations reported so far, g was set to the value predicted to maximize I_c for the symmetric $C/N = 0.1$ network. A limited number of trials were performed to check sensitivity to g, at $\alpha = 0.5$ (using $f = 1$ internal cueing). Theoretically, for this α value, g values in the range 0.1– 0.33 should permit the existence of retrieval states in the symmetric system (D O'Kane, personal communication). The simulations displayed a sharp onset of retrieval behaviour at $g = 0.12$, with I_c peaking at $g = 0.2$, followed by a steady decrease in I_c with g, falling to effectively zero by $g = 0.6$†. Thus, as with the storage capacity, analytical results pertaining to the symmetrically dilute C/N network give a reasonable, though not exact, guide to the performance of a finite-sized, asymmetrically dilute network.

Retrieval of ternary encoded patterns

It might be argued that the above demonstration of the network's ability to function as an associative memory is irrelevant to the question of biological memories, on the grounds that it assumed a binary distribution of neuronal firing during learning, whereas a continuous distribution is much more likely. Ternary distributions represent the first step towards more structured encodings. Furthermore analytical results suggest that they already give a good estimate of the performance obtainable with continuous patterns (Treves and Rolls 1991).

Figure 7 displays the results obtained when the system is used to retrieve ternary-encoded patterns of the type described by equation (3). Qualitatively the data trends in figure 7(*a*) are similar to those found for binary data (see figure 4(*a*)); quantitatively they differ, in that the transition to no retrieval now occurs at a higher loading level ($\alpha \approx 1.2$), and the absolute amount of information retrieved per pattern per cell, I_c, is (for low α) smaller than for binary patterns, this being despite the fact that the ternary patterns encoded in the synaptic matrix contain more information than binary ones. Computing the information capacity I we find that in the $f = 1.0$ condition it has maximal value 0.139 ± 0.01 at $\alpha = 0.8$, and for $f = 0.2$, maximal value 0.100 ± 0.03 at $\alpha = 0.6$. These values are considerably lower than those for binary stored patterns, indicating, at least for the parameter set studied here, that from an information storage perspective there is nothing to be gained from storing ternary as opposed to binary patterns. Indeed analytical work suggests that this is the case across a wide range of pattern sparsity and connectivity levels in networks of threshold-linear neurons (Treves and Rolls 1991).

Finally, it was noted that the strength of the regulatory term $b(x)$ had a considerable effect on the retrieval of ternary patterns. In the simulations summarized in figure 7(*a*), a', the key parameter in $b(x)$, was set to 0.05 as this led to activity ($\langle V \rangle$) levels close to 0.1, the value in the stored patterns, and values of a_r rising from 0.023 at $\alpha = 0.2$ to 0.08 at $\alpha = 1.0$. Varying the value of a' alters the activity, sparsity, and information contained in the retrieved representation. This is demonstrated in figure 7(*b*) and 7(*c*) for networks with loading level $\alpha = 0.6$: allowing the activity of the retrieved representation to rise above that in the original patterns facilitates slightly higher levels of information recall. The generality of this effect was not investigated systematically because its magnitude does not appear

† Low Δt values were needed in the simulations with high g to prevent excessive oscillations of the sort discussed in section 5.
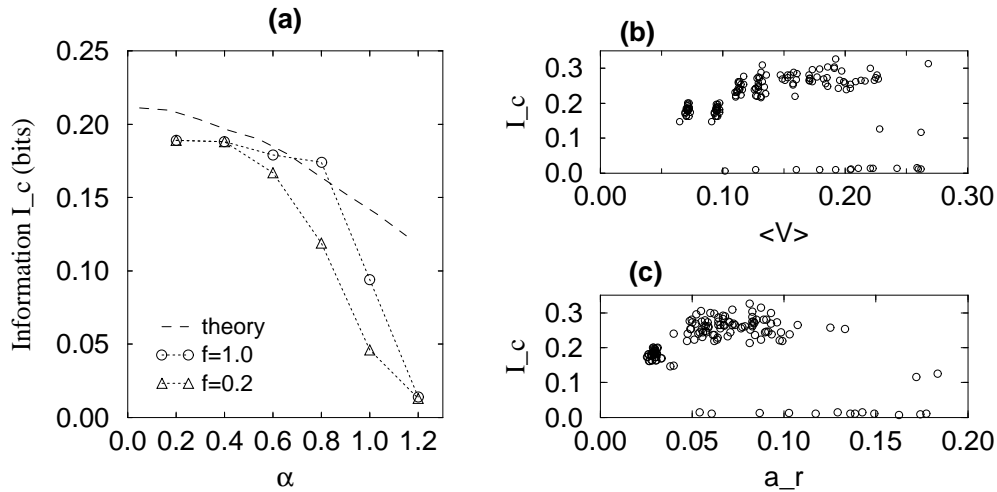
Figure 7. The quality of information retrieval for ternary stored patterns. To get a sense of scale, the mean information (in bits) per cell in a pattern prescribed by equation (3) is 0.675 for $a = 0.1$. (*a*) Information retrieved per pattern per cell (I_c) versus α for two cue sizes, $f = 0.2$ and $f = 1.0$. Data points are mean values over 20 trials, the error bars follow a pattern similar to that in figure 1(*a*) and are omitted for clarity. (*b*) and (*c*) are scatterplots of individual trial results obtained using $f = 1.0$ at $\alpha = 0.6$ with various values of the parameter a' between 0.03 and 0.16.

sufficient to alter the above conclusions about ternary versus binary coding.

7. Discussion

In summary, the pattern retrieval capabilities of a large partially connected network of formal threshold-linear 'neurons' have been investigated in a series of simulation studies. Low activity random patterns were stored by the covariance rule. The storage capacities were found to be near $\alpha = 0.8$ and 1.2 for binary and ternary patterns respectively. Although analytical estimates of α_c are not available for the asymmetric dilute network studied here, the experimental values are close to the estimates of α_c in the tractable case of symmetric dilution. The experimental transitions had first-order-like characteristics, consistent with the analysis of the symmetric network which predicts a first-order transition at α_c. For binary stored patterns the retrieval attractors were found to possess wide basins for α values up to ≈ 0.5; e.g. at $\alpha = 0.5$, initial states carrying on average only 5% of the information in a stored pattern evolved into final states retrieving on average 75% of the pattern information. Persistent external stimulation matching one of the binary stored patterns facilitates the evolution of states having non-zero pattern correlations even at $\alpha \approx 2$. However, in this high-α regime the pattern correlations are fairly low, therefore it is more appropriate to view the system as performing recognition rather than pattern recall in this case (see, for example, Amit (1989) for a discussion of recognition/recall in attractor networks).

A major motivation for studying associative memory of low activity patterns in an asymmetric low-connectivity network is that these are conditions found in both the neocortex and hippocampus—two brain regions often believed to function, at least partly, as associative memories. Despite the satisfactory performance of the model in the current simulations, many other factors need to be examined—analytically where possible, otherwise

by simulations—to assess whether such a model network can be linked efficiently to suitable input and output modules. Work in this direction, in the context of the hippocampal system, has been undertaken (Treves and Rolls 1992).

Much recent work has focussed on the associative memory capabilities of networks of formal neurons modelled with spiking dynamics, e.g. Amit *et al* (1990), Gerstner and van Hemmen (1992), or simulated by compartmental modelling, e.g. Wilson and Bower (1989), Lansner and Fransén (1992). Models such as these using more realistic dynamics allow one to study the timescales of retrieval processes. Given the speed with which some aspects of cortical processing occur (Thorpe and Imbert 1989, Oram and Perrett 1992, Tovée *et al* 1993), data on pattern retrieval timescales in models might serve as a stringent test of the biological feasibility of associative memory models. Recent progress in analysing the dynamics of a network of spiking neurons described by potential and synaptic conductance variables has also led to some novel conclusions (Treves 1993). Work is currently underway on modifying the simulation model reported here to incorporate single-cell dynamics of this more sophisticated type.

Acknowledgments

References

Amaral D G, Ishizuka N and Claiborne B 1990 Neurons, numbers and the hippocampal network *Progress in Brain Research* vol 83, ed J Storm-Mathisen *et al* (Amsterdam: Elsevier) ch 1, pp 1–11

Amit D J 1989 *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge: Cambridge University Press)

Amit D J, Evans M R and Abeles M 1990 Attractor neural networks with biological probe records *Network* **1** 381–405

Amit D J and Tsodyks M V 1991a Quantitative study of attractor neural network retrieving at low spike rates: I. Substrate—spikes, rates and neuronal gain *Network* **2** 259–73

—— 1991b Quantitative study of attractor neural network retrieving at low spike rates: II. Low-rate retrieval in symmetric networks *Network* **2** 275–94

Bennett M R, Gibson W G and Robinson J 1994 Dynamics of CA3 pyramidal neuron autoassociative memory network in the hippocampus *Phil. Trans. R. Soc.* B **343** 167–87

Buckingham J T 1991 Delicate nets, faint recollections: a study of partially connected associative network memories *PhD Thesis* University of Edinburgh

Gardner-Medwin A R 1976 The recall of events through the learning of associations between their parts *Proc. R. Soc.* B **194** 375–402

Gerstner W and van Hemmen J L 1992 Associative memory in a network of 'spiking' neurons *Network* **3** 139–64

Kühn R and Bös S 1993 Statistical mechanics for neural networks with continuous-time dynamics *J. Phys. A: Math. Gen.* **26** 831–57

Lansner A and Fransén E 1992 Modelling Hebbian cell assemblies comprised of cortical neurons *Network* **3** 105–19

Marr D 1971 Simple memory: A theory for archicortex *Phil. Trans. R. Soc.* B **262** 23–81

Oram M W and Perrett D I 1992 Time course of neural responses discriminating different views of the face and head *J. Neurophysiol.* **68** 70–84

Parisi G 1986 Asymmetric neural networks and the process of learning *J. Phys. A: Math. Gen.* **19** L675–80

Rolls E T 1987 Information representation, processing and storage in the brain: analysis at the single neuron level *The Neural and Molecular Bases of Learning* ed J-P Changeux and M Konishi (New York: Wiley) pp 503–40

Thorpe S J and Imbert M 1989 Biological constraints on connectionist models *Connectionism in Perspective* ed R Pfeifer *et al* (Amsterdam: North-Holland) pp 63–92

Tirozzi B and Tsodyks M 1991 Chaos in highly diluted neural networks *Europhys. Lett.* **14** 727–32

Tovée M J, Rolls E T, Treves A and Bellis R P 1993 Information encoding and the responses of single neurons in the primate temporal visual cortex *J. Neurophysiol.* **70** 640–54

Treves A 1990 Graded-response neurons and information encodings in associative memories *Phys. Rev.* A **42** 2418–30

—— 1991a Dilution and sparse coding in threshold linear nets *J. Phys. A: Math. Gen.* **24** 327–35

—— 1991b Are spin-glass effects relevant to understanding realistic auto-associative networks? *J. Phys. A: Math. Gen.* **24** 2645–54

—— 1993 Mean-field analysis of neuronal spike dynamics *Network* **4** 259–284

Treves A and Rolls E T 1991 What determines the capacity of autoassociative memories in the brain? *Network* **2** 371–97

—— 1992 Computational constraints suggest the need for 2 distinct input systems to the hippocampal CA3-network *Hippocampus* **2** 189–200

—— 1994 Computational analysis of the role of the hippocampus in memory *Hippocampus* **4** 374–91

Waugh F R, Marcus C M and Westervelt R M 1991 Reducing neuron gain to eliminate fixed-point attractors in an analog associative memory *Phys. Rev.* A **43** 3131–42

Willshaw D J, Buneman O P and Longuet-Higgins H C 1969 Non-holographic associative memory *Nature* **222** 960–2

Wilson M A and Bower J M 1989 The simulation of large-scale neural networks *Methods in Neuronal Modeling* ed C Koch and I Segev (Cambridge, MA: MIT Press) pp 291–333