

Dynamics of the CA3 Pyramidal Neuron Autoassociative Memory Network in the Hippocampus

M. R. Bennett, W. G. Gibson and J. Robinson

Phil. Trans. R. Soc. Lond. B 1994 **343**, 167-187
doi: 10.1098/rstb.1994.0019

References

Article cited in:

<http://rstb.royalsocietypublishing.org/content/343/1304/167#related-urls>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: <http://rstb.royalsocietypublishing.org/subscriptions>

Dynamics of the CA3 pyramidal neuron autoassociative memory network in the hippocampus

M. R. BENNETT, W. G. GIBSON AND J. ROBINSON

The Neurobiology Research Centre, Department of Physiology, and The School of Mathematics and Statistics, University of Sydney, New South Wales 2006, Australia

CONTENTS

	PAGE
1. Introduction	168
2. Nomenclature	169
3. Theory	169
(a) The network	169
(b) Time evolution of the network	171
(c) Average behaviour of the network	171
(d) Summary of progressive recall equations	172
(e) Behaviour of the network with probabilistic secretion of quanta	172
4. The CA3 pyramidal neuron recurrent network	172
(a) Neuron types and their connections in CA3	172
(b) Evaluation of connection parameters in the theory	173
(c) Choice of remaining parameters in the theory	174
5. Recall in the CA3 recurrent network	174
(a) The dynamics of progressive recall	174
(b) The dependence of recall on the setting of the membrane potential by inhibitory interneurons	176
(c) The dependence of recall on initial overlap with the target memory	177
(d) The maximum number of memories that can be stored in CA3	179
6. Effects of probabilistic secretion of quanta on recall in the CA3 recurrent network	180
7. Discussion	182
(a) The role of inhibition	182
(b) Other differences	183
8. Conclusion	183
Appendix 1. Progressive recall equations	184
Appendix 2. Formulas for average connectivities	186
References	186

SUMMARY

A theory for the dynamics of sparse associative memory has been applied to the CA3 pyramidal recurrent network in the hippocampus. The CA3 region is modelled as a network of pyramidal neurons randomly connected through their recurrent collaterals. Both the elliptical spread of the axonal systems and the exponential decrease in connectivity with distance are taken into account in estimating the connection probabilities. Pyramidal neurons also receive connections from inhibitory interneurons which occur in large numbers throughout the network; these in turn receive inputs from other inhibitory interneurons and from pyramidal neurons. These inhibitory neurons are modelled as rapidly acting linear devices which produce outputs proportional to their inputs; they perform an important regulatory function in the setting of the membrane potentials of the pyramidal neurons. The probability of a neuron firing in a stored memory, which determines the average number of neurons active when a memory is recalled, can be set at will. Memories are stored at the recurrent collateral synapses using a two-valued Hebbian. Allowance is made in the theory both for the spatial correlations between the learned strengths of the recurrent collateral synapses and temporal correlations between the state of the network and these synaptic strengths. The recall of a memory begins with the firing of a set of CA3 pyramidal neurons that overlap with the memory to be recalled as well as the firing of a set of pyramidal neurons not in the

memory to be recalled; the firing of both sets of neurons is probably induced by synapses formed on CA3 neurons by perforant pathway axons. The firing of different sets of pyramidal neurons then evolves by discrete synchronous steps.

The CA3 recurrent network is shown to retrieve memories under specific conditions of the setting of the membrane potential of the pyramidal neurons by inhibitory interneurons. The adjustable parameters in the theory have been assigned values in accord with the known physiology of the CA3 region. Certain levels of overlap between the input and the memory to be retrieved must also be satisfied for almost complete retrieval. The number of memories which can be stored and retrieved without degradation is primarily a function of the number of active neurons when a memory is recalled and the degree of connectivity in the network. The inhomogeneity in the connectivity of the pyramidal cells improves both capacity and overlap of the final state with the memory. The probabilistic secretion of quanta at the recurrent collateral synapses improves the recall mechanism when there is only partial overlap in the input with the memory to be retrieved and the input contains incorrect elements, at the expense of a slight deterioration in the fidelity of recall.

1. INTRODUCTION

The pyramidal neurons of the CA3 region of the hippocampus receive excitatory synaptic connections from a variety of sources. These include the mossy axons of granule cells and the perforant pathway axons of entorhinal neurons (Yeckel & Berger 1990). However, the greatest sources of excitatory synaptic connections on CA3 pyramidal neurons comes from the recurrent collaterals within the CA3 pyramidal system itself (Amaral *et al.* 1990). Marr (1971) was the first to suggest that such a recurrent network could act as an autoassociative memory, if the efficacy of the excitatory synapses were modifiable and if the membrane potentials of the CA3 pyramidal neurons were set by inhibitory interneurons that measure the total activity of the network (for a detailed description of Marr's ideas, see McNaughton & Morris (1987), McNaughton & Nadel (1990) and Willshaw & Buckingham (1990)). The storage of memories in this system is then possible by modification of the recurrent excitatory collateral synapses. Subsequent presentation of part of a stored memory pattern then allows for recall of the whole pattern. Marr (1971) pointed out that the recall process could occur in either one step or in a series of steps within the recurrent system. In the latter case, pyramidal neurons excited by the presentation of part of the stored memory cause other pyramidal neurons to fire or to be silent until a steady firing pattern of pyramidal neurons is reached. The one step process can be called simple recall (Gardner-Medwin 1976) whereas the multiple step process has been named the collateral effect (Marr 1971) or progressive recall (Gardner-Medwin 1976). It has been argued that the strong granule cell synaptic input to the CA3 pyramidal recurrent system provides the pattern of pyramidal neuron firing during a learning period; this occurs after the granule cells have already transformed their input from the entorhinal cortex by, for example, reducing the overlap between patterns to be presented to the CA3 pyramidal neuron system (Rolls 1989; Treves & Rolls 1990; Gibson *et al.* 1991). The direct perforant pathway input to the CA3 pyramidal neurons may then subserve the role of initiating the retrieval process (Treves & Rolls 1992).

The analysis of autoassociative memory systems has proceeded along several different paths. One of these, based on analogies with statistical mechanical systems (Little 1974; Hopfield 1982) involves a network in which all the excitatory neurons receive connections from each other, there is a high level of activity in the system, and the thresholds for firing of the neurons are set in artificial ways (for a description, see Amit 1989). A related approach is the method of statistical neurodynamics (Amari & Maginu 1988; Amari 1989). Another approach is to try to analyse more biologically realistic networks, similar to that in the CA3 region of the hippocampus (Treves & Rolls 1991), by considering a progressive recall process in a system of randomly connected pyramidal neurons in which there is a constant level of activity in the stored patterns. In this case the membrane potential of the pyramidal neurons is regulated by inhibitory interneurons (Gardner-Medwin 1976, 1989). In the present work, a specific mechanism is given for this regulation. It involves assuming that the inhibitory interneurons process signals more rapidly than the pyramidal neurons; an input applied at one instant of time to the CA3 region thus provides both an excitatory and an inhibitory input to the pyramidal neurons; it is the balancing of these that gives stability to the system and allows the orderly retrieval of memories under the progressive recall process. The biological evidence for such a different operation of the excitatory and inhibitory neurons in CA3 is given in §4a. The mathematical consequence is that the inhibitory interneurons act as linear devices, their output being proportional to their summed input. This linearity allows a theory developed for a network involving a single linear inhibitory neuron (Gibson & Robinson 1992) to be adapted in a straightforward way to the present case of many inhibitory neurons.

A shortcoming of the early theories (Marr 1971; Gardner-Medwin 1976) is that a number of correlations are neglected, thus making the region of applicability uncertain. These correlations are of two types. First, correlations occur between the learned synaptic connection strengths even though they have been formed from memories whose elements are random and independent. Second, correlations develop

between the state of the CA3 pyramidal system and the learned synaptic connection strengths because, after the first recall step, the state becomes dependent on these connections through the updating procedure; subsequent recall steps then depend on both the current state and the connection strengths. A theory that takes these correlations into account has recently been developed (Gibson & Robinson 1992) and in this work it is applied to the CA3 pyramidal neuron recurrent collateral system, so as to extend the work of Marr (1971) and Gardner-Medwin (1976), see also Willshaw & Buckingham (1990). Although for certain networks the correlations can severely alter the recall behaviour (for examples of this, see Gibson & Robinson 1992), it is found that for the parameter choices appropriate for the CA3 region their effect is generally small, the exception being some parameter regions lying on the border between recall and non-recall. All calculations in this paper have been done using the full theory, thus avoiding the need to continually check that the simpler theories are adequate. The theory has been used to find the capacity of the CA3 system and to investigate recall as a function of the properties of the inhibitory interneurons and as a function of the input initiating the recall. In the latter case, it is the overlap of the input with a stored memory that is the deciding factor.

The stochastic variability of synaptic transmission has been shown to greatly improve the ability of granule cell networks, such as those in the hippocampus, to separate overlapping patterns of activity on their inputs and to maintain a steady low level of activity for varying levels on the inputs (Gibson *et al.* 1991). As there is considerable stochastic variation in quantal transmission in the CA3 pyramidal cell system (Miles & Wong 1984, 1986), it is of interest to see if this has a role in memory recall. It is shown below that the probabilistic secretion of quanta enhances the ability of the recurrent collateral system to retrieve memories when the associative input contains many incorrect elements. This enhancement is, however, at the expense of a reduction in the total number of memories that can be stored.

2. NOMENCLATURE

- n : the number of excitatory neurons E .
 n^* : the number of inhibitory neurons I .
 N : $= n + n^*$, the total number of neurons.
 c_{ij} : the probability of an intrinsic connection from neuron j to neuron i .
 \bar{c} , \bar{c}^2 : mean and mean square connectivities for the E neurons.
 W : the connectivity matrix, whose elements W_{ij} are unity if there is an intrinsic connection from neuron j to neuron i , and zero otherwise.
 a : the activity, being the probability that a randomly chosen neuron in a memory pattern is active.
 m : the number of memories stored in the network. (Strictly, $m + 1$ memories are stored, as the numbering starts from 0.)

- Z^p : One of the memory vectors, $p = 0, 1, \dots, m$. It is of dimension n , and its entries are 0s and 1s according to $P(Z_i^p = 1) = a$.
 Z^0 : The memory singled out for retrieval. It has the first na places 1s followed by the remaining $n(1-a)$ places 0s.
 J : the matrix of learned connections.
 $X(t)$: the vector giving the state of the system at the discrete times $t = 0, 1, 2, \dots$. It is of dimension $N = n + n^*$; its first n entries are 0s and 1s, and the remaining n^* places are real numbers.
 $h(t)$: the vector of inputs to the neurons at time t .
 $h^m(t)$: the vector of postsynaptic membrane potentials of the neurons at time t .
 μ_N, σ_N : mean and standard deviation of the change in membrane potential due to the arrival of an impulse at a neuron.
 g_0, g_1 : parameters governing the strength of the inhibition from the I neurons.
 \mathcal{E} : expectation.
 $E_i(t)$: expectation of (input-threshold) for i th excitatory neuron at time t .
 $\sigma_i(t)$: standard deviation of (input-threshold) for i th excitatory neuron at time t .
 $r_0(t)$: overlap at time t of the current state of the E neurons with the memory Z^0 .
 x_t : the average valid firing rate; nax_t is the expected number of valid firings at time t .
 y_t : the average spurious firing rate; $n(1-a)y_t$ is the expected number of spurious firings at time t .
 z_t : the average activity in the inhibitory neurons at time t .
 ρ : expectation of J_{ij} .
 ρ' : expectation of J_{ij} conditional on another J_{ij} .
 γ : covariance of J_{ij} 's.
 γ' : covariances of J_{ij} 's conditional on other J_{ij} 's.

3. THEORY

(a) The network

The network consists of N neurons, of which n (numbered 1, . . . , n) are excitatory (E) and n^* (numbered $n + 1, \dots, N \equiv n + n^*$) are inhibitory (I) (figure 1).

The connections between these neurons are specified by an $N \times N$ connectivity matrix W , where $W_{ij} = 1$ if neuron j sends a collateral to neuron i and $W_{ij} = 0$ otherwise. It is assumed that $W_{ii} = 0$ (that is, neurons do not connect directly to themselves) and further that the W_{ij} are independent random variables with $P(W_{ij} = 1) = c_{ij}$. The values of the c_{ij} will depend on the types of neurons being connected and on their relative spatial location; this is treated in detail below. The connectivity matrix is thus divided into four blocks, with each block representing one of the four types of connections that can occur:

$$W = \begin{pmatrix} E \rightarrow E & I \rightarrow E \\ E \rightarrow I & I \rightarrow I \end{pmatrix}. \quad (1)$$

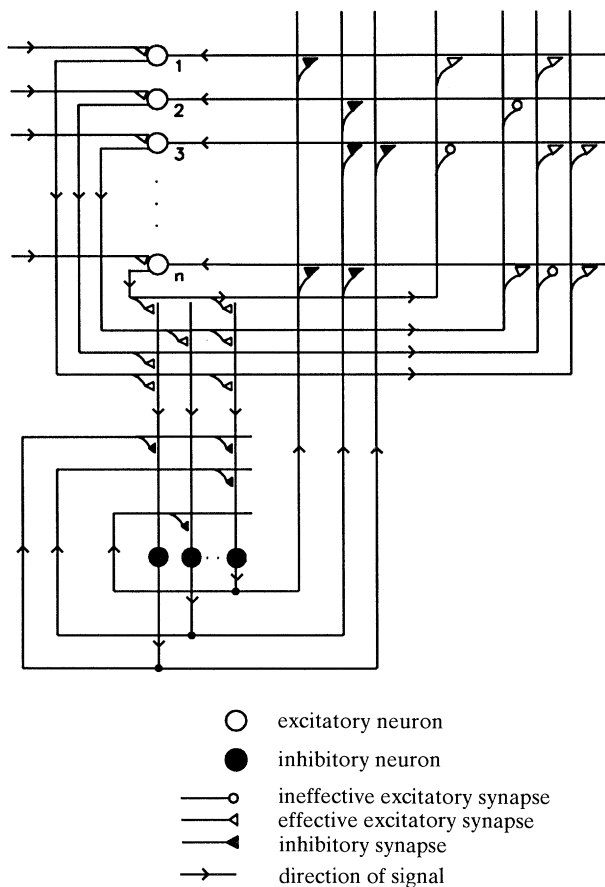


Figure 1. The basic CA3 network, consisting of n pyramidal neurons (open circles) and n^* inhibitory interneurons (filled circles). The pyramidal neurons make random connections with each other through their recurrent collaterals, the mean probability of any one connection existing being \bar{c}_{EE} . Before learning, these connections are ineffective; after learning, a subset of them becomes effective and in the final state of the network there are excitatory synaptic connections whose strengths are taken to be unity (open triangles) and others whose strengths have remained at zero (open circles). The inhibitory interneurons receive random connections from many pyramidal neurons and also from other inhibitory neurons, the mean probabilities of these connections existing being \bar{c}_{IE} and \bar{c}_{II} respectively. The inhibitory neurons in turn project to pyramidal neurons with a mean connection probability of \bar{c}_{EI} . No learning occurs at any synapses involving inhibitory interneurons, so such synaptic strengths are taken as fixed. The initial state of the system is set by a firing pattern coming onto the pyramidal neurons from either the mossy fibres or the direct perforant pathway to stratum moleculare, and this is shown by the lines entering from the left. Once the initial state has been set, the external source is removed. The CA3 recurrent network then updates its internal state cyclically and synchronously.

The synaptic strengths are contained in another $N \times N$ matrix \mathbf{J} , and again this is divided into the same four blocks. For connections involving I neurons these strengths are taken to be uniform: thus for three of the blocks J_{ij} takes the constant values J_{IE} , $-J_{EI}$ and $-J_{II}$ respectively, where $(-)J_{AB}$ denotes the connection strength from a neuron of type B to one of type A . However, the values taken in the $E \rightarrow E$ block are non-uniform, and are related to the patterns that the system has memorized.

These memory patterns occur on the E neurons only. They are denoted by $\{\mathbf{Z}^p: p = 0, 1, \dots, m\}$ where \mathbf{Z}^p is a vector of length n , whose i th element is 0 or 1 according to whether the i th neuron is inactive or active respectively, and the elements are randomly and independently chosen with $P(Z_i^p = 1) = a$, $i = 1, \dots, n$; $p = 0, \dots, m$. During a learning phase, these $m + 1$ patterns are applied from an external source, to the E neurons in the network. This learning phase gives a Hebbian modification of the strength, J_{ij} , of a synaptic connection from neuron j to neuron i , such that $J_{ij} = J_{EE}$ if for any of the $m + 1$ patterns that i th and j th neurons are simultaneously active; otherwise $J_{ij} = 0$. Thus for $i, j = 1, 2, \dots, n$,

$$J_{ij} = \begin{cases} J_{EE} & \text{if } Z_i^p Z_j^p = 1 \text{ for any } p = 0, 1, \dots, m, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

This is the so-called 'clipped' Hebbian, introduced by Willshaw *et al.* (1966) (see also Palm 1980, 1988; Faris & Maier 1988; Amit 1989). The total connection strengths between the neurons are then contained in the matrix with elements $J_{ij}W_{ij}$. Note that this matrix will not in general be symmetric, as although \mathbf{J} is symmetric, \mathbf{W} is not.

Of central concern is the recall of a stored pattern, starting from some initial state of the network. Because the patterns are random, without loss of generality this target pattern can be \mathbf{Z}^0 , and further, as the ordering of the neurons is also random, the numbering of the E neurons can be arranged so that

$$\mathbf{Z}_i^0 = \begin{cases} 1 & \text{for } i = 1, \dots, na, \\ 0 & \text{for } i = na + 1, \dots, n. \end{cases} \quad (3)$$

That is, \mathbf{Z}^0 is a vector consisting of na ones followed by $n - na$ zeros:

$$\mathbf{Z}^0 = \underbrace{(1, 1, \dots, 1)}_{na}, \underbrace{(0, 0, \dots, 0)}_{n(1-a)}. \quad (4)$$

The network starts in a particular state at time $t = 0$, and then updates synchronously at the discrete times $t = 1, 2, \dots$. The state $\mathbf{X}(t)$ of the system at time t is a vector of length N which includes the states of both the E and I neurons (in contrast to the memory vectors \mathbf{Z}^p which contain the state of the E neurons only). The first n entries are ones and zeros, according to whether the corresponding E neuron fires or not at time-step t . These E neurons fire if their summed input, which includes excitatory and inhibitory contributions, exceeds their threshold. The I neurons also sum their input at each discrete time-step, but their behaviour differs from that of the E neurons in two fundamental ways. First, their output is proportional to their summed input. This follows from the assumption that an I neuron responds to an excitatory input by producing a relatively large number of spikes during the interval before the next discrete time step, the number of such spikes being proportional to the strength of the input. To a first approximation, each spike produces the same decrease in membrane potential at the soma of a neuron to which the I neuron is connected, which implies that the effective output of

an I neuron can be approximated by a continuous linear function of its total input. The result is that the remaining n^* entries in $\mathbf{X}(t)$, rather than being 0 or 1, are real numbers proportional to the inputs to the I neurons. Second, the I neurons act more quickly than the E neurons in the sense that a signal arriving simultaneously at the inputs to an E and an I neuron is processed so rapidly by the I neuron that its output is also summed by the E neuron (see § 4a). In the framework of a synchronous network, this difference is accounted for by making the I neurons act instantaneously, whereas the E neurons respond after one time-step.

(b) Time evolution of the network

The total input to the i th neuron at time t is $nh_i(t)$, where

$$h_i(t) = \frac{1}{n} \sum_{j=1}^N W_{ij} J_{ij} X_j(t), \quad i = 1, 2, \dots, N. \quad (5)$$

Because each term in the above sum is either 0 or 1 this assumes that each impulse received by the neuron provides the same change in postsynaptic potential or, equivalently, each impulse releases the same number of quanta of transmitter. Note that as the connections from inhibitory neurons will involve negative J_{ij} s, $h_i(t)$ can take positive or negative values. From the discussion in the previous section, the E neurons update their state according to

$$X_i(t+1) = \begin{cases} 1, & h_i(t) > g_0 \\ 0, & \text{otherwise,} \end{cases} \quad i = 1, \dots, n, \quad (6)$$

where g_0 is the threshold of an E neuron, taken to be the same for all of them. On the other hand, the I neurons update according to

$$X_i(t) = \kappa h_i(t), \quad i = n+1, \dots, N, \quad (7)$$

where κ is a positive constant. The initial input pattern $\mathbf{X}(0)$ is taken to be a random distortion of the target pattern \mathbf{Z}^0 . Specifically, $\mathbf{X}(0)$ is defined by

$$P(X_i(0) = 1) = \begin{cases} x_0, & i = 1, \dots, na, \\ y_0, & i = na+1, \dots, n, \end{cases} \quad (8)$$

where x_0 and y_0 are parameters whose values are to be specified (for the particular case $x_0 = 1$, $y_0 = 0$ the input $\mathbf{X}(0)$ is exactly the target memory \mathbf{Z}^0); the remaining n^* places in $\mathbf{X}(0)$ are then determined by equation (7). Thus the total input vector has the form

$$\mathbf{X}(0) = (\underbrace{1, 0, 0, 1, 0, \dots, 0, 1, 1, 0, 1, 0, 0, \dots, 0}_{na}, \underbrace{0, \kappa h_{n+1}(0), \dots, \kappa h_N(0)}_{n^*}), \quad (9)$$

where the first na positions contain an average of $na x_0$ ones and the next $n(1-a)$ positions contain an average of $n(1-a)y_0$ ones. The time evolution of the network is now deterministic, the state at the times $t = 1, 2, \dots$ being given by repeated applications of equations (6) and (7), using the input calculated from equation (5).

(c) Average behaviour of the network

Although the time evolution of any given network is deterministic, nevertheless it will be different for each particular realization of the network, as the connectivities are determined probabilistically. For a given set of parameters n, m, \dots , it is necessary to consider an ensemble of networks with different detailed structures compatible with these parameter values. Thus the state of the network $\mathbf{X}(t)$ at a given time t is a random variable, and the average behaviour of the network is found by calculating its expectation.

This calculation is straightforward for the I neurons since here there is no learning and hence correlations between connection strengths are negligible. Define

$$\begin{aligned} x_i &= \mathcal{E} X_i(t), & i = 1, \dots, na \\ y_i &= \mathcal{E} X_i(t), & i = na+1, \dots, n \\ z_i &= \mathcal{E} X_i(t), & i = n+1, \dots, N, \end{aligned} \quad (10)$$

where \mathcal{E} denotes expectation. These are the average activity levels for the E and I neurons: the division of $\mathcal{E} X_i(t)$ for the E neurons into x_i and y_i corresponds to a division into 'valid' and 'spurious' firings for the recall of the target pattern \mathbf{Z}^0 ; for perfect recall, $x_i = 1$ and $y_i = 0$. Then (see Appendix 1)

$$z_i = g_i^* (ax_i + (1-a)y_i), \quad (11)$$

where $g_i^* = \kappa \overline{c_{IE}} J_{IE}$, $\overline{c_{IE}} = \sum_{j=1}^n c_{ij}/n$, ($i > n$)

being the average connectivity of E to I neurons. Equation (11) shows that the average firing activity of the inhibitory neurons is directly proportional to the total activity of the excitatory neurons.

Now consider the E neurons. These receive input from both E and I neurons, and from equation (6) their expected firing rate at the $t+1$ step is

$$\mathcal{E} X_i(t+1) = P(h_i(t) > g_0), \quad i = 1, \dots, n, \quad (12)$$

which can be rewritten as

$$\mathcal{E} X_i(t+1) = P\left(\frac{h_i(t) - \mathcal{E} h_i(t)}{\sqrt{\text{var } h_i(t)}} > \frac{g_0 - \mathcal{E} h_i(t)}{\sqrt{\text{var } h_i(t)}}\right), \quad (13)$$

where $\text{var } h_i(t)$ is the variance of $h_i(t)$. (This variance is in fact the expectation of the conditional variance of $h_i(t)$, given $\mathbf{X}(t)$, as is made explicit in equation (37) of Appendix 1). A naïve application of the central limit theorem then gives

$$\mathcal{E} X_i(t+1) = \Phi\left(\frac{\mathcal{E} h_i(t) - g_0}{\sqrt{\text{var } h_i(t)}}\right), \quad (14)$$

where $\Phi(\cdot)$ is the normal distribution function. The above is basically correct for the $E \rightarrow I$ and $I \rightarrow E$ contributions, but there are subtleties concerning the $E \rightarrow E$ interactions, arising because the associated J_{ij} s

are constructed from the memories (see equation (2)) and hence are not independent random variables. This has two consequences. First, the J_{ij} s themselves are correlated and this gives rise to additional terms in $\text{var}h_i(t)$ involving the covariances of these J_{ij} s. Second, the J_{ij} s become correlated with the state $\mathbf{X}(t)$ during the recall process, and this must be taken into account in evaluating both expectations and variances. The detailed treatment of these correlations is given in Appendix 1 and the final computational scheme is summarized in the next section.

(d) Summary of progressive recall equations

The principal quantities of interest are x_i and y_i ; these are the average firing levels, at time-step t , for correct and spurious cells respectively. (Recall that 'correct' and 'spurious' are defined in relation to the target memory \mathbf{Z}^0 , given by equation (3), so that x_i is the average proportion of ones in the first na places and y_i is the average number of ones in the next $n(1-a)$ places; for exact recall of \mathbf{Z}^0 , $x_i \rightarrow 1$ and $y_i \rightarrow 0$ as $t \rightarrow \infty$.) The quantities x'_i and y'_i which appear in the equations below are conditional expectations of the state $\mathbf{X}(t)$, given particular values of the connection strengths, and do not have a simple observational meaning.

The updating process is given by a set of four coupled difference equations:

$$x_{t+1} = \Phi(E_1(t)/\sigma_1(t)), \quad (15)$$

$$y_{t+1} = \Phi(E_n(t)/\sigma_n(t)), \quad (16)$$

$$x'_{t+1} = \Phi(E'_1(t)/\sigma'_1(t)), \quad (17)$$

$$y'_{t+1} = \Phi(E'_n(t)/\sigma'_n(t)). \quad (18)$$

All the quantities on the right-hand sides can be expressed in terms of x_t, y_t, x'_t, y'_t ; for example,

$$E_1(t) = \bar{c}(ax_t + (1-a)\rho y'_t) - g_1(ax_t + (1-a)y_t) - g_0, \quad (19)$$

$$E_n(t) = \bar{c}\rho(ax_t + (1-a)y'_t) - g_1(ax_t + (1-a)y_t) - g_0, \quad (20)$$

where

$$\bar{c} \equiv \overline{c_{EE}} = \frac{1}{n} \sum_{j=1}^n c_{ij}, \quad i = 1, \dots, n, \quad (21)$$

$\rho \equiv 1 - (1 - a^2)^m$ and J_{EE} has been set equal to 1. g_1 is a parameter governing the strength of the inhibition provided by the I neurons. It is given explicitly by

$$g_1 = (n^*/n)\overline{c_{EI}}J_{EI}g_1^*, \quad (22)$$

where $\overline{c_{EI}}$ is the average connectivity from I to E neurons and g_1^* is defined following equation (11). The remaining quantities required in equations (15) to (18) are given in Appendix 1.

The above equations are almost the same as those given in Gibson & Robinson (1992, § 5.1), the principal difference being the incorporation of non-homogeneous connections between the E neurons. The extension to many I neurons (as opposed to one) has not caused a significant change, since they act as

linear devices and the assumption that each I neuron receives inputs from many E neurons means that the variance in their outputs can be neglected. (This variance is included in the $O(1/n)$ term in (40).)

(e) Behaviour of the network with probabilistic secretion of quanta

From equation (5) the presynaptic input to neuron i at time t is $nh_i(t)$, where

$$h_i(t) = \frac{1}{n} \sum_{j=1}^n C_{ij} X_j(t), \quad i = 1, 2, \dots, N, \quad (23)$$

where $C_{ij} \equiv W_{ij}J_{ij}$ is the total connection strength from neuron j to neuron i . This assumes that each action potential received by neuron i has the same effect. To generalize this, let N_j be the change in membrane potential due to an impulse arriving from neuron j . This will be proportional to the random number of quanta of transmitter released by the action potential. It follows that the postsynaptic membrane potential is

$$h_i^m(t) = \frac{1}{n} \sum_{j=1}^n C_{ij} X_j(t) N_j, \quad i = 1, 2, \dots, N. \quad (24)$$

The N_j will be taken to be independent random variables with mean μ_N and variance σ_N^2 . To incorporate this synaptic noise into the formalism, it is necessary to calculate the expectation and variance of $h_i^m(t)$ and the details are given in Appendix 1. It turns out that all the required quantities have already been calculated, so the inclusion of synaptic noise involves making a choice for μ_N and σ_N .

4. THE CA3 PYRAMIDAL NEURON RECURRENT NETWORK

(a) Neuron types and their connections in CA3

The CA3 region of the hippocampus may be considered as a two-dimensional rectangular sheet, about 10 000 μm long and 2700 μm wide (Braitenberg & Schuz, 1983; Finch *et al.* 1983). In one-month- to one-year-old Sprague-Dawley rats, the CA3 region contains 330 000 pyramidal neurons (Boss *et al.* 1987). These neurons have very long dendritic trees, of cumulative length about 16 000 μm but they are compact, spreading out over only about 200 μm from a 20 μm diameter soma (Finch *et al.* 1983); as spines occur on these dendrites at intervals of a little over 1 μm along their length, the CA3 pyramidal neurons possess about 12 000 synapses, over half of which mediate synaptic connections between pyramidal neurons (Amaral *et al.* 1990). These synapses arise from a very extensive pyramidal-cell axon collateral system, which spreads for about a quarter to a half the longitudinal extent of the hippocampus along the septotemporal axis (up to 4 mm: Ishizuka *et al.* 1986; Tamamaki & Nojyo 1991) and along about half of the transverse axis (about 1 mm; Finch *et al.* 1983).

The probability that two pyramidal cells are connected decays exponentially with the longitudinal distance between them, presumably because of a

decrease in the number of synapses. The rate of this decay has been estimated at $600 \mu\text{m}^{-1}$ (Miles *et al.* 1988; Traub & Miles 1991).

There are about 33 000 inhibitory interneurons distributed throughout the CA3 region of the hippocampus, which is about 10% of the pyramidal cell number (Misgeld & Frotscher 1986). There are two principal classes of these. One set are the vertically oriented basket cell interneurons in the stratum pyramidale that synapse on the soma of the pyramidal cells as well as on other inhibitory interneurons; the others are horizontally oriented stratum oriens-alveus interneurons that also synapse on both pyramidal cells and other inhibitory interneurons (Lacaille & Schwartzkroin 1988; Schlander & Frotscher 1986). Both types of inhibitory interneurons receive an excitatory input onto their dendrites from the pyramidal cells (Lacaille & Schwartzkroin 1988).

Excitatory synaptic transmission from CA3 pyramidal neurons to inhibitory interneurons occurs much more quickly than does that between CA3 pyramidal neurons (2 to 3 ms compared with 10 to 15 ms; Miles 1990; Miles & Wong 1987). The result is that an action potential in a pyramidal neuron can initiate disynaptic inhibitory postsynaptic potentials in other pyramidal neurons with a latency of between 3 and 5 ms (Miles 1990); this may be compared with the longer latency of 10 to 15 ms for the monosynaptic connections between pyramidal neurons. This provides the physiological basis for the theory (§ 3*a*) that an action potential arriving simultaneously at an excitatory terminal on a pyramidal neuron and on an inhibitory interneuron is processed so rapidly by the interneuron that its output can be summed by the pyramidal neuron together with its excitatory input.

(b) Evaluation of connection parameters in the theory

This section is concerned with obtaining numerical values for the quantities $\bar{c} \equiv \overline{c_{EE}}$ and $\bar{c}^2 \equiv \overline{c_{EE}^2}$, as defined by equations (21) and (43). These are for connections from *E* to *E* neurons. The corresponding quantities are also defined for connections involving *I* neurons and similar estimates could be made for them, but in the current model they are combined with other factors and their individual values are not required.

By definition, c_{ij} is the probability that an axon of neuron *j* connects to a dendrite of neuron *i*. It is assumed that c_{ij} falls off exponentially with distance (§ 4*a*) and can be written

$$c_{ij} = C \exp(-\lambda|\mathbf{r}_i - \mathbf{r}_j|), \quad (25)$$

where \mathbf{r}_i and \mathbf{r}_j are the positions of the neurons and *C* is a constant satisfying $0 < C \leq 1$. This exponential dependence on distance will only be an approximation to the true situation; a more precise model would need detailed information on such factors as axonal branching and density of terminals as a function of distance from the soma. The mean connectivity \bar{c} , as given by equation (21), is a sum over these exponentials.

The *E* neurons are the pyramidal cells in the CA3

region of the hippocampus. These lie in a curved sheet that is only two or three cells thick, and thus is essentially a two-dimensional structure. It is therefore appropriate to approximate the sum in equation (21) by a two-dimensional integral (Appendix 2) over the effective area of an axonal tree, where this effective area takes into account the finite extent of the dendritic tree. This area is approximately elliptical (Ishizuka *et al.* 1986; Tamamaki & Nojyo 1991; see also the discussion in § 4*a* above) and assuming that the connection probability c_{ij} is constant on ellipses of constant eccentricity centred on \mathbf{r}_i , then (Appendix 2)

$$\bar{c} = C^2 \frac{\pi\sigma}{n} \frac{1}{(\lambda_1 R_1)^2} (1 - (1 + \lambda_1 R_1)e^{-\lambda_1 R_1}) R_1 R_2, \quad (26)$$

where σ is the density (number per unit area) of the *E* cells, R_1 , R_2 are the lengths of the semimajor and semiminor axes respectively of the effective axonal area and λ_1 is the value of λ along the major axis. (Note that, whereas for a circular area λ would be isotropic, in the elliptical case it must be a function of distance if one makes the reasonable assumption that the connection probability falls smoothly to the same value at the edge of the axonal tree. This concept is made precise in Appendix 2.) A similar calculation gives

$$\bar{c}^2 = C^2 \frac{2\pi\sigma}{n} \frac{1}{(2\lambda_1 R_1)^2} (1 - (1 + 2\lambda_1 R_1)e^{-2\lambda_1 R_1}) R_1 R_2. \quad (27)$$

The numerical values for the constants appearing in equations (26) and (27) can be obtained by a consideration of the values given in § 4*a*. The CA3 region is a two-dimensional rectangular lamina of dimension $2700 \mu\text{m}$ by $10\,000 \mu\text{m}$ and so $\sigma/n = (2700 \times 10\,000)^{-1} = 3.70 \times 10^{-8} \mu\text{m}^{-2}$. The axon collateral system of a CA3 pyramidal neuron has a longitudinal spread of about $4000 \mu\text{m}$ and a transverse spread of about $1000 \mu\text{m}$. The dendritic spread is about $200 \mu\text{m}$ so it is reasonable to take $R_1 = 2100 \mu\text{m}$ and $R_2 = 600 \mu\text{m}$. The value of λ_1 must accord with the requirement that $\exp(-\lambda_1 r)$ be reasonably small when $r = R_1$, otherwise one would go abruptly from a region of high connectivity to one of zero connectivity. The choice $\lambda_1 = 1/1200 \mu\text{m}^{-1}$ gives $\exp(-\lambda_1 R_1) = 0.17$ and also leads to $\lambda_2 = 1/343 \mu\text{m}^{-1}$ where λ_2 is the value of λ along the minor axis; this does not conflict with the value $\lambda = 1/600 \mu\text{m}^{-1}$ quoted in § 4*a* above, assuming that it is an average value. Inserting the above values into equation (26) gives $\bar{c} = C \times 0.050$, and it remains to choose a value for *C*. In one case, the value of \bar{c} estimated by Smith *et al.* (1988) was 0.05 (range 0.01 to 0.1; Traub & Miles 1991). If one makes the choice $C = 1$, then from equation (26) $\bar{c} = 0.050$ and from equation (27) $\bar{c}^2 = 0.021$. This choice of *C* is not unreasonable, as it means that two *E* neurons very close together are almost certainly connected. Thus the final choice for the connectivity parameters is $\bar{c} = 0.05$, $\bar{c}^2 = 0.021$.

It is worth remarking that the above results do not depend crucially on the assumption of an exponential fall-off in connection probability, as in equation (25).

Similar values of c and c^2 can be obtained using other expressions, for example $(1 + \lambda|\mathbf{r}_i - \mathbf{r}_j|)^{-1}$. The exponential form is chosen for convenience, and also because it is the standard one that experimentalists use to fit their results.

The above procedure involves estimates of parameters that are not precisely experimentally determined, but at least two things of significance emerge. First, using values for these parameters that certainly lie within acceptable limits, a value for \bar{c} has been obtained that is in agreement with experimental estimates. Second, a relationship has been obtained between \bar{c} and \bar{c}^2 , and this is important in the calculation of the variances in the progressive recall equations. Note that if no spatial dependence has been assumed in the connectivity, then one would have $\bar{c}^2 = \bar{c}^2$. The effect of the spatial dependence is to reduce the variances $\sigma_1(t)$ and $\sigma_n(t)$ (see equations (41) and (42)); this leads to both an increase in the storage capacity of the network and an increase in the overlap of the final state and the memory being recalled.

(c) *Choice of remaining parameters in the theory*

The parameters determined so far are the total number of CA3 pyramidal neurons $n = 330\,000$ (§ 4a) and the connection parameters for these neurons $\bar{c} = 0.05$, $\bar{c}^2 = 0.021$ (§ 4b). The remaining parameters are a , g_0 , g_1 , m , μ_N , σ_N , x_0 and y_0 .

The activity a is not precisely determined experimentally. A large a allows the network to store only a small number of memories; as a decreases the number of memories increases (although the information content of each memory decreases; Palm 1988; Amari 1989). a cannot become too small, otherwise too few neurons are active in a memory, and statistical fluctuations make the recall process unreliable. Marr (1971) chose to have an average of 200 neurons active in a total neuron population of $n = 100\,000$ when a memory is recalled, giving the value $a = 0.002$. In the present case n is larger so the choice $a = 0.001$ is made, giving an average of 330 active neurons when a memory is recalled exactly. Limited variation of a about this value gives similar results; however a much larger a (for example, 0.002) considerably reduces the storage capacity and a much smaller a (for example, 0.0005) does not give recall from a partial input, so $a = 0.001$ is a reasonable compromise.

The threshold on a pyramidal neuron, g_0 , will be taken to be a positive constant, the same for all cells. Experimental values cannot be used as the scaling in the equations is of necessity arbitrary. Because the neuron's membrane potential is set by a sum of excitatory and inhibitory inputs, the theory will work with g_0 set to zero; however, this choice does allow the network to exhibit oscillatory behaviour in certain parameter regions that are on the border between retrieval and non-retrieval of a memory. Thus a non-zero g_0 is desirable both physiologically and mathematically. An estimate of the size of g_0 can be obtained from equation (19) by considering the case where recall occurs with the spurious firings kept very low

($y_i \approx 0$). Recall implies that $E_1(t)$ must be greater than zero, leading to $0.00003x_i - g_0 > 0$, where the typical value $g_1 = 0.02$ has been used. Thus recall from a starting value of, for example, $x_0 = 0.25$ requires $g_0 < 7.5 \times 10^{-6}$. In the following calculations, $g_0 = 7 \times 10^{-6}$; this value damps out most of the undesirable oscillations without causing much degradation in recall performance.

The strength of the inhibitory input to a pyramidal neuron is determined by g_1 . It is given explicitly by equation (22), but again this involves quantities whose values are not determined experimentally so this cannot be used to evaluate g_1 . The procedure adopted below (§ 5b) is to evaluate the progressive recall equations for a wide range of values of g_1 ; it is usually found that for part of this range recall occurs, and a suitable value of g_1 can then be chosen for further calculations. The value of m , the number of stored memories, is also unknown experimentally, and a similar procedure of investigating a range of possible values can be used (§ 5d).

The stochastic nature of quantal release at synapses is governed by the random variables N_j , which have mean μ_N and standard deviation σ_N . Without loss of generality, μ_N can be taken to be unity; the noise due to the stochastic equal release is then determined entirely by σ_N , and the effect of taking various values of this parameter is investigated (§ 6).

The final parameters, x_0 and y_0 , give the number of valid and spurious firing in the input signal to the CA3 network. Again, the theory is evaluated for a range of reasonable values of both these parameters (§ 5c).

Broadly speaking, the parameters can be divided into three categories: those in the first (n, \bar{c}, \bar{c}^2) have reasonably well determined values; those in the second (a, g_0) are less well determined, but reasonable estimates can be made and the final results are qualitatively similar for a range of values; those in the third ($g_1, m, \sigma_N, x_0, y_0$) are largely unknown from the experimental point of view, but the behaviour of the model network has been explored for a comprehensive range of values for each parameter.

5. RECALL IN THE CA3 RECURRENT NETWORK

(a) *The dynamics of progressive recall*

The progressive recall process begins with a certain input onto the CA3 pyramidal neurons, this input being possibly derived from the monosynaptic connections formed by the perforant pathway on the pyramidal neurons in stratum moleculare. This input is represented by a vector $\mathbf{X}(0)$ structured according to equation (9), with the first na positions showing the firings contained in the memory \mathbf{Z}^0 (the valid firings) and the next $n(1 - a)$ positions showing the firings outside this memory (the spurious firings). The remaining n^* positions are the inputs to the inhibitory neurons and these are automatically set according to the prescription given by equation (7). At the first step of the recall process, this input is converted into a set

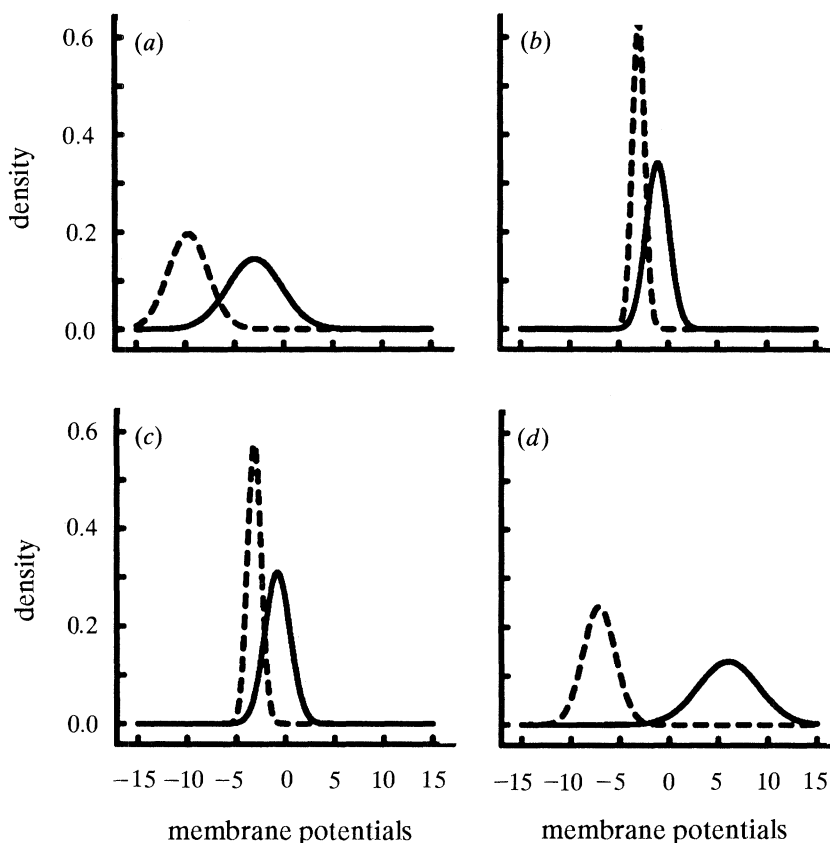


Figure 2. Progressive recall in the CA3 pyramidal neuron recurrent network. (a) Step $t=0$; (b) step $t=1$; (c) step $t=2$; (d) step $t=8$. The parameters used are: number of pyramidal neurons $n=330\,000$; connectivity parameters for the pyramidal neurons $\bar{c}=0.05$, $\bar{c}^2=0.021$; probability that a pyramidal neuron is active in a memory $a=0.001$; total number of memories stored $m=200\,000$; threshold parameter for the pyramidal neurons $g_0=7 \times 10^{-6}$; strength parameter for the output of the inhibitory interneurons $g_1=0.024$; initial valid firing rate $x_0=0.5$; initial spurious firing rate $y_0=0.001$. (Thus the initial input contains 165 valid firings and 330 spurious ones.) There is no synaptic noise ($\sigma_N=0$). The curves show the distribution of $n(h_i(t) - g_0)$ for time steps $t=0, 1, 2$ and 8 . $h_i(t)$ is the total synaptic input to the i th pyramidal neuron at time t , and is a measure of the membrane potential at that time. These inputs can be divided into two sets: $hx(t) \equiv \{h_1(t), \dots, h_{na}(t)\}$, the set of inputs for the neurons that fire under memory \mathbf{Z}^0 , represented by a solid line, and $hy(t) \equiv \{h_{na+1}(t), \dots, h_n(t)\}$, the set of inputs for the neurons that are silent under memory \mathbf{Z}^0 , represented by a broken line. Both curves are normal density functions with means and variances that are calculated using the progressive recall equations. Because g_0 has been subtracted, the threshold for firing is now zero. The membrane potentials shown at step $t=0$ (a) represent the initial input of 165 valid and 330 spurious firings, together with the output from the inhibitory interneurons. For recall in a single step one would require the two distributions to be separated by zero; clearly this has not occurred, and at this stage only a small part of the memory has been retrieved. The output from the pyramidal neurons is now fed back via the recurrent collaterals, giving the distributions shown at step $t=1$ (b). The distribution for the non-memory set $hy(t)$ (broken line) is still almost entirely below 0, but the distribution for the memory set $hx(t)$ (solid line) has moved to the right, and this will give more valid firings. This trend continues with subsequent steps, until by the $t=8$ step (d) the two distributions are almost entirely separated by the threshold at 0, giving 322 valid and two spurious firings. At this stage a steady state has been reached and no changes occur at subsequent time steps.

$\{h_1(0), h_2(0), \dots, h_n(0)\}$, of membrane potentials for the pyramidal neurons, according to equation (5). This set can be divided into two subsets $hx(0) \equiv \{h_1(0), \dots, h_{na}(0)\}$ and $hy(0) \equiv \{h_{na+1}(0), \dots, h_n(0)\}$; thus $hx(0)$ is the set of membrane potentials for those neurons that fire when the memory \mathbf{Z}^0 is recalled and $hy(0)$ is the set of membrane potentials for all other neurons. The memory \mathbf{Z}^0 will be exactly recalled in one step if all the members of $hx(0)$ are above the threshold g_0 and all the members of $hy(0)$ are below. In general, this will not occur: the two sets of membrane potentials will form two overlapping distributions that are not cleanly separated by g_0 .

An example of this is given in figure 2, where the distribution of membrane potentials at selected times in the progressive recall process is shown for a specific choice of parameters (as detailed in the figure legend).

Let $hx(t)$, $hy(t)$ be the sets of membrane potentials for the 'memory' and 'non-memory' neurons, respectively, at time t . These membrane potentials are discrete random variables which by the central limit theorem tend to become normally distributed for n large. Since in the present case $n=330\,000$ their histograms would be almost indistinguishable from smooth normal distribution curves. Thus in figure 2 the curves shown are normal density functions with

Table 1. *Progressive recall in the CA3 pyramidal neuron recurrent network*

(The network parameters are the same as those for figure 2. The table shows the number of valid and spurious firings for each of the first eight time steps, together with the overlap $r_0(t)$ (see equation (28)) of the current state with the memory \mathbf{Z}^0 . The initial input to the pyramidal neurons contains 165 valid active neurons and 330 spurious active neurons. At step $t=8$ the network contains 322 valid and two spurious active neurons, and this situation does not change with further steps. Exact recall of the memory would require 330 valid and 0 spurious firings, but this cannot be achieved under progressive recall with this number of memories.)

time step	valid firings	spurious firings	overlap with memory
0	165	330	0.408
1	47	0	0.375
2	57	0	0.415
3	86	1	0.508
4	158	4	0.684
5	261	8	0.876
6	311	4	0.965
7	321	3	0.982
8	322	2	0.984

means and variances calculated using the theory of § 3 and the Appendix, the solid lines being for the set $hx(t)$ and the broken lines for the set $hy(t)$. More precisely, the curves represent $n(h_i(t) - g_0)$ and the effect of subtracting g_0 is to make zero the threshold for firing. The initial input onto the pyramidal neurons consists of 165 valid firings and 330 spurious ones (see table 1). The membrane potentials shown at the $t = 0$ step in figure 2 give 47 valid firings and 0 spurious firings at the next step. This reduction in activity has been caused by the inhibitory interneurons; these are assumed to act much more rapidly than the pyramidal neurons (§ 4a) and so although the initial signal arrives simultaneously at the pyramidal cells and at the inhibitory interneurons, the latter process this signal so rapidly that their inhibitory contribution is added to the membrane potential of the pyramidal cells at the $t = 0$ time step. The recurrent circuit of collaterals now comes into play: the output from the pyramidal cells is fed back, with again the inhibitory interneurons acting rapidly, and the membrane potentials shown at step $t = 1$ in figure 2 are obtained. The set $hy(1)$ is almost entirely below 0, so there are almost no spurious firings; the set $hx(1)$ has a substantial portion above 0, and so the valid firings increase. This trend continues with subsequent steps, and by $t = 8$ an equilibrium state has been reached, with 322 valid firings and two spurious ones (table 1). Exact recall of the memory would be 330 valid and 0 spurious firings, but this is not achievable by the progressive recall process in a network storing this number of memories.

(b) The dependence of recall on the setting of the membrane potential by inhibitory interneurons

The setting of the membrane potential of the pyramidal neurons in CA3 by the inhibitory interneurons can vary from one recall step to another. The magnitudes of the inhibitory synaptic potentials responsible for this are linearly related to the activity in the network, the constant of proportionality being the parameter g_1 (see equations (19), (20) and (22)) and so changes in g_1 determine the membrane potential for any particular level of activity. g_1 cannot be determined from known experimental data, but some insight into its significance can be obtained by finding the final state of the network for a range of values. A convenient measure of the closeness of the current state $\mathbf{X}(t)$ to the target memory \mathbf{Z}^0 is provided by the overlap defined by

$$r_0(t) = \frac{\sum_{i=1}^n (X_i(t) - \bar{X}(t))(Z_i^0 - \bar{Z}^0)}{\sqrt{\left[\sum_{i=1}^n (X_i(t) - \bar{X}(t))^2 \sum_{j=1}^n (Z_j^0 - \bar{Z}^0)^2 \right]}}, \quad (28)$$

where

$$\bar{X}(t) = \sum_{i=1}^n X_i(t)/n \equiv S(t)/n \quad \text{and}$$

$$\bar{Z}^0 = \sum_{i=1}^n Z_i^0/n = a.$$

This expression shows that $r_0(t)$ is related to the cosine of the angle between the vector formed by the first n places of $\mathbf{X}(0)$ (that is, those entries referring to the E neurons) and the vector \mathbf{Z}^0 . It is unity if and only if there is complete retrieval of \mathbf{Z}^0 . Introducing

$$S_1(t) = \sum_{j=1}^{na} X_j(t),$$

it can be written more simply as

$$r_0(t) = \frac{S_1(t) - aS(t)}{\sqrt{[S(t)(1 - S(t)/n)]\sqrt{[na(1 - a)]}}}. \quad (29)$$

The solid curve in figure 3 shows the final overlap $r_0(t = \infty)$, in the absence of synaptic noise, for g_1 in the range 0.015 to 0.035.

There is a 'window' of retrieval, outside of which the overlap with the target memory is negligible. For small values of g_1 there is too little inhibition and this allows the number of spurious firings to build up. The result is either unstable oscillations, leading to all neurons going to the dead state, or a stable oscillatory pattern involving many spurious firings; the overlap with the memory \mathbf{Z}^0 is thus either zero or very small. As g_1 increases, there is an abrupt onset of recall (but with some oscillation) at $g_1 = 0.0166$ and an even more abrupt cessation at $g_1 = 0.0245$. (These discontinuities correspond to bifurcations that are a consequence of the nonlinear nature of the model: a brief discussion is given in Gibson & Robinson (1992); a more extensive discussion of bifurcations in a related model can be found in Amari (1971).) For values of g_1

larger than 0.0245 the inhibition is too strong and the activity in the network quickly dies. The quality of recall is not uniform in the window. The overlap $r_0(\infty)$ varies from a low of 0.557 to a high of 0.984, with $g_1 = 0.024$ giving optimum recall. The value of g_1 , determined as it is by the number and strengths of the connections between inhibitory and excitatory neurons, is presumably set during the evolution of the network which accompanies development.

(c) *The dependence of recall on initial overlap with the target memory*

The capacity of the CA3 pyramidal neuron recurrent network to recall a memory depends on the initial level of overlap with this memory. This in turn depends on the number of valid and spurious firings in the initial state; for an input characterized by parameters x_0, y_0 , these are nax_0 and $n(1-a)y_0$, respectively. In the absence of definite information about the values of x_0 and y_0 , recall has been investigated for a range of possible values.

In figure 4a, recall is investigated as a function of the inhibition strength parameter g_1 and of the number of valid firings nax_0 in the initial state. The number of spurious firings in the initial state has been fixed at 330 ($y_0 = 0.001$) and the number of memories stored is $m = 200\,000$. The height of the graph represents the overlap of the final state with the target memory Z^0 . The raised 'plateau' thus gives the region of recall, and the steepness of the walls surrounding it reflects the abrupt onset of recall, as discussed in the previous section. Again, recall is possible for a range of values of g_1 , with the choice becoming more restricted as the initial overlap becomes less. In other words, a good setting of g_1 will enable memory retrieval from very incomplete initial information.

Figure 4b shows the corresponding situation when the number of initial valid firings is fixed at 99 (corresponding to $x_0 = 0.3$) and the number of initial spurious ones is varied. A similar situation holds, and again an optimum choice of g_1 , not incompatible with the optimum choice based on figure 4a, will enable retrieval from very incomplete input information.

In figure 5 the recall region is shown as a function of both valid and spurious initial inputs. A specific choice $g_1 = 0.02$ has been made, and also the number of memories has been reduced to 100 000 to show more clearly the effect of synaptic noise (to be discussed in the following section). In the noiseless case, the solid line gives the division between the regions of memory retrieval (above) and non-retrieval (below). The quality of retrieval is uniform over the entire upper region, being an overlap of 0.996 (328 valid, 0 spurious); below the line the overlap is 0.0 (no neurons firing). This uniformity exists because, for a given set of network parameters, the system evolves to a fixed point whose properties are independent of the initial state of the system: the initial state only determines which fixed point is the final state. The upper region can be thought of as representing the basin of attraction for the memory Z^0 : provided the

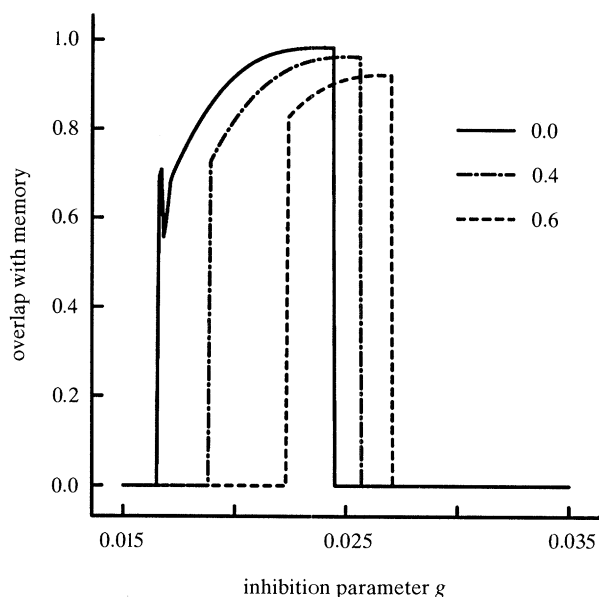


Figure 3. Recall as a function of the inhibition strength parameter g_1 for various amounts of synaptic noise. The values of $n, \bar{c}, \bar{c}^2, a, m$ and g_0 are the same as for figure 2; the initial state parameters are $x_0 = 0.5, y_0 = 0.001$ which means that the initial state contains 165 valid and 330 spurious firings; the parameter g_1 , which is a proportionality constant governing the strength of the inhibition provided by the inhibitory neurons, varies over the range 0.015 to 0.035. The progressive recall equations (15–18) are iterated until a stable final state is reached and then the overlap with the stored memory state Z^0 is calculated using equation (29). This overlap is related to the cosine of the angle between the vector representing the final state of the pyramidal neurons and the vector representing the memory Z^0 ; for exact recall of this memory it would be equal to 1.0. Results are given for the case of no synaptic noise ($\sigma_N = 0.0$) and for the cases where the noise has standard deviations $\sigma_N = 0.4$ and 0.6. In each case there is a recall 'window', outside of which the overlap is essentially zero. Inside a window the overlap can vary from a low of about 0.6 to a high of 0.9 or more; for optimal recall in a given network the value of g_1 which maximizes this overlap should be chosen. The main effect of increasing synaptic noise is to shift the window to the right, and diminish both its width and height. In the zero noise case there is some oscillatory behaviour in a small region on the edge of the recall window ($g_1 \approx 0.017$). This disappears under a moderate amount of synaptic noise ($\sigma_N = 0.2$ is enough to remove it completely).

initial state lies in this basin, then the memory is retrieved with the same final overlap.

Figures 3–5 have dealt only with the final state of the network. However, the progressive recall equations give all the intermediate steps in the recall process and an example of this is shown in figure 6, where recall is initiated from states with different overlaps with the memory Z^0 . These overlaps are $r_0(0) = 0.182, 0.262, 0.337, 0.474$ and were obtained by fixing the number of initial spurious firings at 330 and varying the number of initial valid firings. For the noiseless case (solid lines), retrieval fails for the two smaller initial overlaps, and succeeds for the two larger ones. There is a certain 'threshold' value of $r_0(0)$ (about 0.325 in the present case) at which retrieval

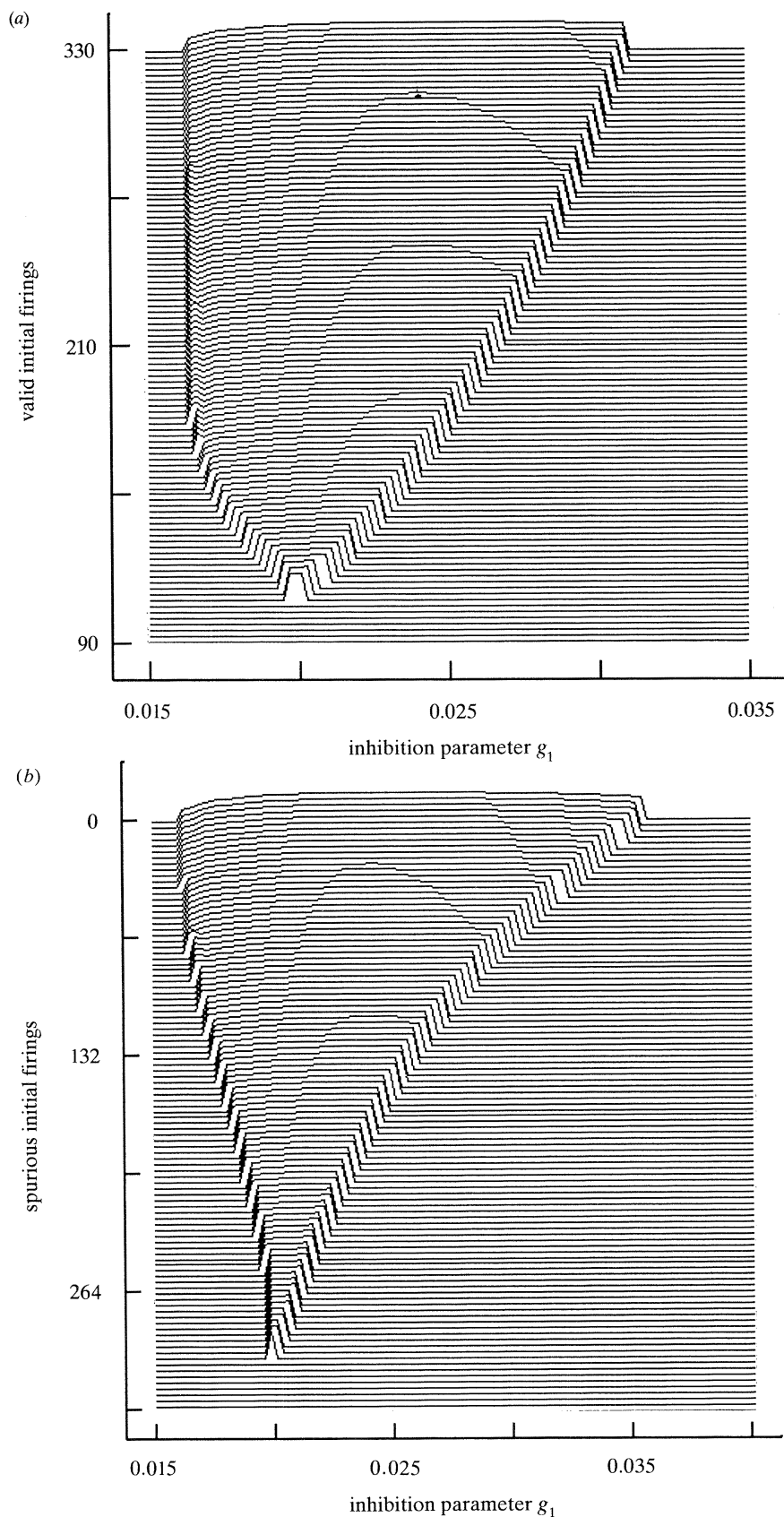


Figure 4. (a) Recall as a function of the initial state x_0 and the inhibition strength parameter g_1 . The values of n , \bar{c} , \bar{c}^2 , a , m and g_0 are the same as for figure 2; $y_0 = 0.001$, which means that the input contains 330 spurious firings; x_0 ranges from 0.273 to 1.0, which means that the number of valid firings in the input ranges from 90 to 330; g_1 varies over the range 0.015 to 0.035. The height of the graph is the overlap of the final state, obtained by iterating the progressive recall equations, with the target memory Z^0 . The raised plateau represents the region of recall; the 'walls' surrounding it reflect the abrupt nature of the onset of memory recall. (Note that the solid curve in figure 3 is simply the cross-section of (a) for the valid initial firings equal to 165.) (b) This shows a similar situation to (a): the difference is that the input signal now has $x_0 = 0.30$ and thus always contains 99 valid firings; y_0 varies from 0.001 to 0, so the input contains from 330 down to 0 spurious firings.

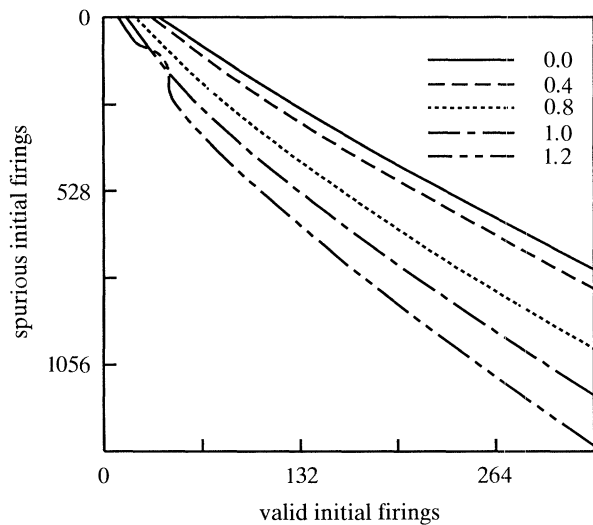


Figure 5. Recall as a function of the initial state in the presence of various amounts of synaptic noise. The values of n , \bar{c} , \bar{c}^2 , a and g_0 are the same as for figure 2; the number of stored memories has been reduced to $m=100\,000$ and g_1 has been fixed at 0.02. The axes show the number of valid and the number of spurious firings in the initial state $\mathbf{X}(0)$. The lines give the division between the regions of retrieval and non-retrieval for each value of the standard deviation of the noise ($\sigma_N=0,0.4,0.8,1.0,1.2$). In each case, the lower region represents non-retrieval (zero overlap with the target memory) and the upper region represents retrieval (non-zero overlap with the target memory: this varies from 0.996 for the noiseless case to 0.875 for the $\sigma_N=1.2$ case). It is seen that the presence of synaptic noise can significantly increase the number of initial states that lead to recall; for example, an initial state with 200 valid and 630 spurious firings will not recall a memory until the noise standard deviation exceeds 0.8. However, there is a limit to the amount of noise that is beneficial: too much noise leads to a much reduced overlap of the final state with the target memory, and also the actual retrieval region will start to diminish, as shown by the 'kink' in the $\sigma_N=1.2$ curve.

commences, with the final overlap jumping from $r_0(\infty) = 0$ to $r_0(\infty) = 0.996$. This threshold value will depend on the mixture of valid and spurious firings in the initial state; however, the value of the final overlap, $r_0(\infty)$, depends only on whether retrieval occurs and not on the specific construction of the initial state.

(d) The maximum number of memories that can be stored in CA3

Several definitions of capacity are commonly used in connection with memory storage networks (Amari & Maginu 1988; Treves & Rolls 1991; Gibson & Robinson 1992). The first is absolute capacity, and is measured by the maximum number of memories that can be stable states of the system. This requires that the memories are completely stable in the sense that if the system is started in a memory state it will stay there precisely. This is too stringent a requirement for a biological memory; a better quantity is the relative capacity which is a measure of the maximum number of memories for which the state of the system will

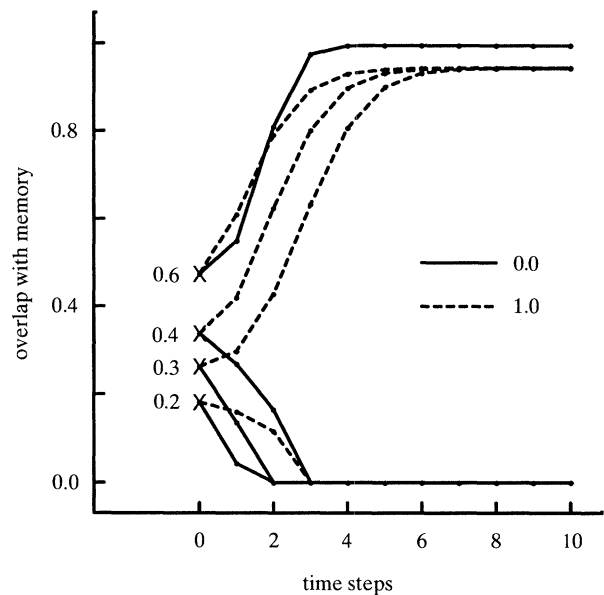


Figure 6. The time steps in the progressive recall process, starting from various initial overlaps with the memory \mathbf{Z}^0 . The values of n , \bar{c} , \bar{c}^2 , a and g_0 are the same as for figure 2, $m=100\,000$, $g_1=0.02$ and $y_0=0.001$ (thus the number of initial spurious firings is 330). The steps in the recall process are shown for four different starting overlaps (shown as X in the figure), formed by taking $x_0=0.2, 0.3, 0.4$, and 0.6 , corresponding to 66, 99, 132 and 198 valid firings, respectively. The solid lines are for the case of zero synaptic noise, and the broken lines are for $\sigma_N=1.0$. As shown by the $x_0=0.3$ and $x_0=0.4$ cases, the presence of synaptic noise can retrieve a memory that would otherwise be lost; the price to be paid is a lower final overlap: in the present case, 0.946 compared with the noiseless value of 0.996.

approach and then remain close to a memory during progressive recall. A necessary condition for this to occur is that the stored memories be stable and this can be tested by starting the network in the state $\mathbf{X}(0) = \mathbf{Z}^0$ and observing whether the overlap remains close to 1 or goes to 0. Figure 7a shows the results of performing this test for a range of values of m and g_1 , the solid curve corresponding to zero synaptic noise being the relevant one for this section. A precise choice of g_1 is necessary for optimal storage: the choice $g_1 = 0.031$ enables a noiseless network to store the order of 340 000 memories as stable fixed points.

Treves & Rolls (1991) measure capacity (α_c) as the maximum number of retrievable patterns per synapse; thus, in our notation, α_c is the value of $m/n\bar{c}$ when m is maximum. For the above values ($m = 340\,000$, $n = 330\,000$, $\bar{c} = 0.05$) this is approximately 20, which is comparable to Treves & Rolls' value of about 30 for $a = 0.001$ (Treves & Rolls 1991, figure 5a). It is to be noted that their model is different in that it uses a covariance Hebbian, rather than a clipped one.

The ability of the network to act as a memory device depends not only on the stability of the stored memories, but also on the basins of attraction surrounding these memories; if these are too small it will really only function as a recognition device and will not act as a retrieval device in the usual sense of performing pattern completion. Some insight into the size of these basins can be obtained by starting the

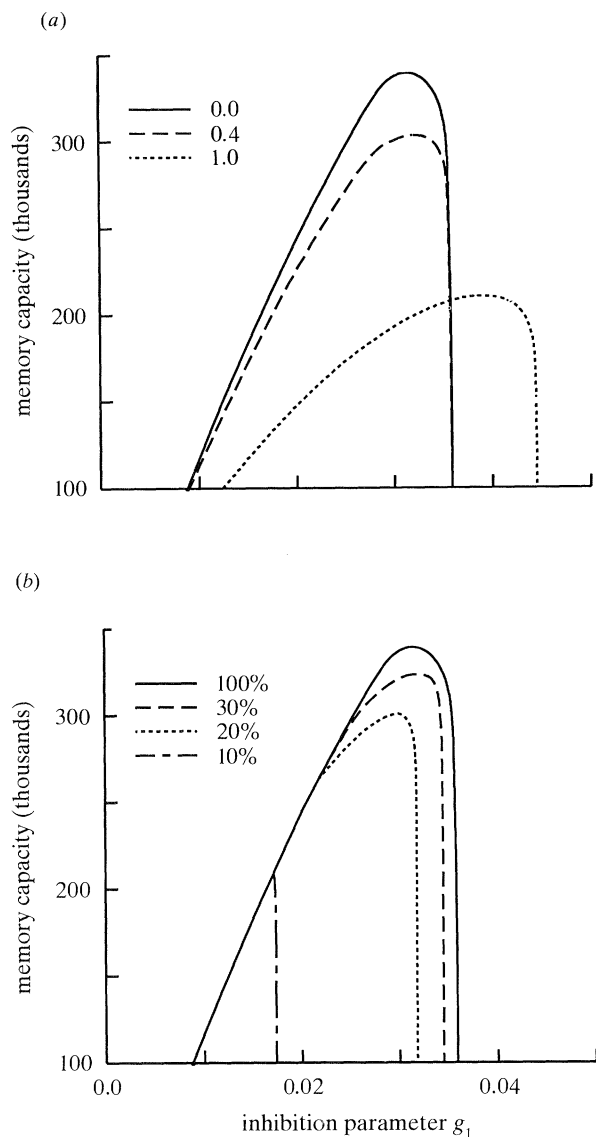


Figure 7. (a) Memory stability, as a function of the inhibition strength parameter g_1 for various levels of synaptic noise. The values of n , \bar{c} , \bar{c}^2 , a and g_0 are the same as for figure 2. The initial state is the target memory (parameters $x_0=1$, $y_0=0$); the final state, obtained by iterating the progressive recall equations until no further change occurs, is either near this memory, or else has negligible overlap with it. The calculation is repeated for a range of values of m and g_1 and the curves give the maximum number of memories that can be fixed points for each value of g_1 . (Thus the region under each curve corresponds to parameter values for which memory storage is possible.) The range of possible values of g_1 is severely reduced as the network approaches its maximum capacity. The effect of synaptic noise is to reduce the possible number of stable memories: for zero noise it is about 340 000; this reduces to about 300 000 for $\sigma_N=0.4$ and to about 210 000 for $\sigma_N=1.0$. (b) Memory capacity, as a function of the inhibition strength parameter g_1 for various initial states (zero synaptic noise). The values of n , \bar{c} , \bar{c}^2 , a and g_0 are the same as for figure 2. Results are shown for four different initial states, containing no spurious firings and 10%, 20%, 30% and 100% valid firings, respectively (that is, $y_0=0, x_0=0.1, 0.2, 0.3, 1.0$). The retrieval region is very limited in the 10% case, but improves rapidly for initial inputs containing 20% or more valid firings.

network in a state which only partially overlaps a stored memory. Figure 7b shows the retrieval regions for four cases where the initial state contains no spurious firings and 10%, 20%, 30% and 100% valid firings, respectively (thus $y_0=0$ and $x_0=0.1, 0.2, 0.3, 1.0$). In the 10% case the capacity is limited to a maximum of 200 000 memories and the choice of g_1 is critical. A 20% input considerably improves the situation and for inputs of 40% or greater there is little difference to the 100% case, indicating that by this stage most initial points lie within the basin of attraction of the relevant memory.

6. EFFECTS OF PROBABILISTIC SECRETION OF QUANTA ON RECALL IN THE CA3 RECURRENT NETWORK

Quanta are secreted from boutons stochastically (for a review, see Redman 1990). The question arises as to the effects of this phenomenon on recall in a recurrent network in the hippocampus. The progressive recall formalism was extended in § 3e and Appendix 1 to include the effects due to probabilistic secretion of quanta at synapses in the CA3 region. The final equations are (48) to (53), involving the new parameters μ_N and σ_N . As explained in § 4c, there is no loss of generality in taking $\mu_N=1$. The value of the standard deviation, σ_N , is not known experimentally, so a representative selection of values will be used below to illustrate the possible effects.

One effect of the stochastic fluctuations in quantal secretion at the recurrent collateral terminals of CA3 pyramidal neurons is to require the g_1 membrane potential setting parameters to take on higher values if successful recall of memories is to be obtained (figure 3). The width of the window is also diminished, the effect being moderate for $\sigma_N=0.4$ but becoming large for $\sigma_N=0.6$. There is also some degradation in the extent to which the memory can be recalled at any setting of g_1 , the amount varying with the value of σ_N and the choice of g_1 . Overall, it can be concluded that a reasonable amount of synaptic noise ($\sigma_N < 0.4$, say, in the present case) does not cause much degradation in the recall properties of the network.

Although probabilistic secretion of quanta at recurrent excitatory synapses leads to a decline in the amount of memory that can be recalled, this is compensated for by an increase in the ability of such recurrent networks to give good recall from a relatively poor initial overlap with the memory to be recalled. This is particularly true if the number of memories stored is well below the maximum capacity of the network. Figure 5 shows this effect in a network storing 100 000 memories, which is well below its maximum (noiseless) capacity of over 300 000 memories. Here, the addition of quite a large amount of noise (up to $\sigma_N=1.0$) causes a significant increase in the ability of the network to recall memories from partial information. This is very noticeable when the input contains a large number of spurious firings; for example, an input containing 330 valid firings and more than 825 spurious ones would not lead to recall in a noiseless system, but with $\sigma_N=1.0$ up to 1130 spurious input

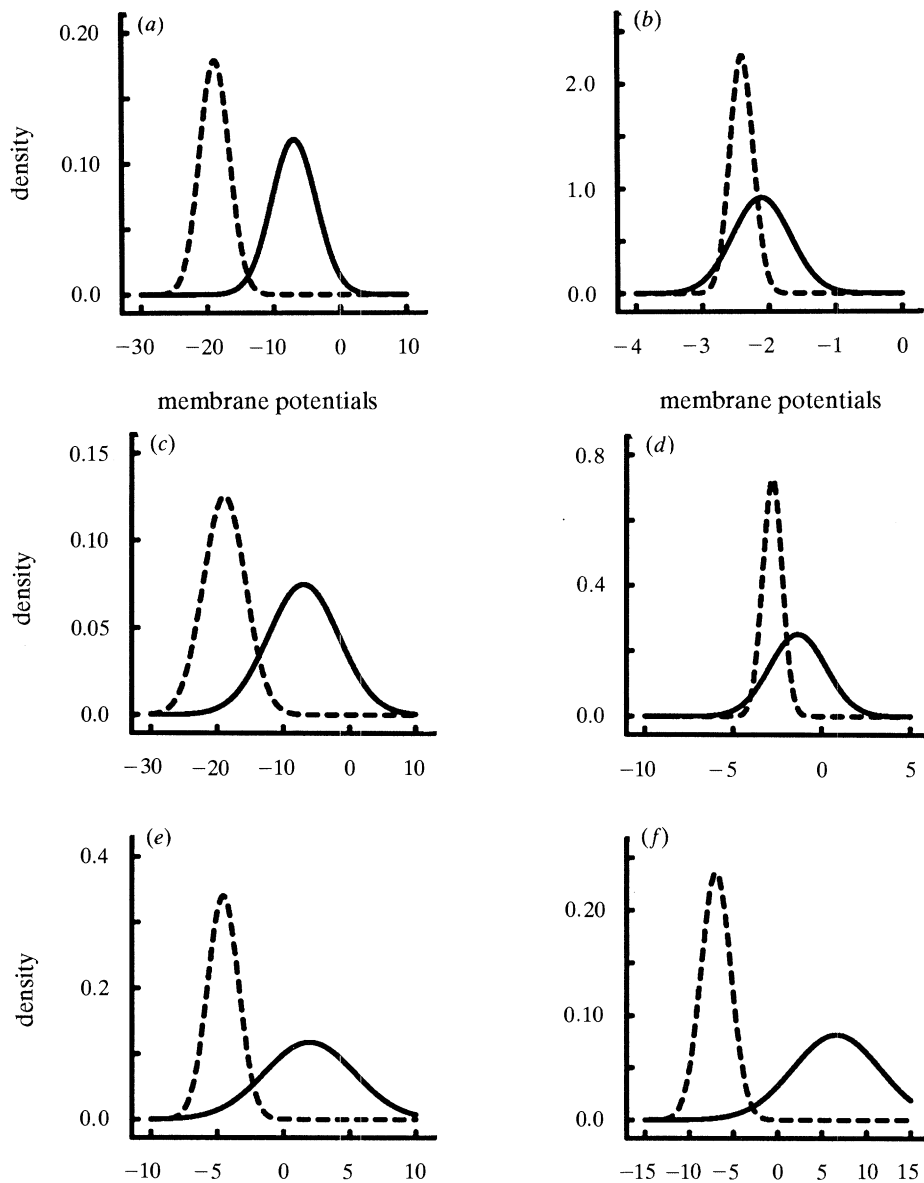


Figure 8. (*a,b*) Retrieval failure in the absence of noise. (*a*) Step $t=0$; (*b*) step $t=1$. The values of n , \bar{c} , \bar{c}^2 , a and g_0 are the same as for figure 2, $m=100\,000$, $g_1=0.02$, $\sigma_N=0$, $x_0=0.8$ and $y_0=0.0025$ (thus the number of initial valid firings is 264 and number of initial spurious firings is 824). The membrane potentials shown at step $t=0$ (*a*) are calculated from this input together with the output from the inhibitory interneurons, with g_0 subtracted so that the threshold for firing is at 0. The result is only seven valid and no spurious firings, leading to the membrane potentials shown at step $t=1$ (*b*); these are now completely below threshold, and no further firing occurs. (*c-f*) Successful retrieval in the presence of synaptic noise. (*c*) Step $t=0$; (*d*) step $t=1$; (*e*) step $t=3$; (*f*) step $t=8$. The parameters have the same values as for (*a,b*) except that now $\sigma_N=1.0$. The effect of the noise has been to increase the variances of the membrane potentials, and there are now 32 valid (and no spurious) firings at the first step. This is enough to keep the network alive, and the progressive recall mechanism gradually separates the two distributions, until by the $t=8$ step the network has settled into a stable pattern of 301 valid and six spurious firings.

firings would still give retrieval. There is, however, a limit to the amount of noise that is beneficial. In the present case, the choice $\sigma_n=1.2$ continues to give an improvement in cases where the input contains many valid firings, but leads to a deterioration in cases where there are a small number of valid firings, as shown by the 'kink' in the upper left hand region.

If the network stores more memories, the beneficial effect of synaptic noise decreases. For example, if the number of memories in the network of figure 5 is increased to 200 000 then only noise up to about $\sigma_N=0.4$ is uniformly useful; for larger value 'kinks'

start appearing in the upper left hand region of the corresponding graph. Thus there is a tradeoff between the storage capacity and the recall ability of the network, with increasing synaptic noise decreasing the former as it improves the latter.

Synaptic noise is known to be beneficial for recall in other types of memory storage networks. For example, in Hopfield networks the addition of a certain amount of noise destabilizes the 'spurious' states (which are certain mixtures of the stored memories) and thus improves recall of pure memories (Amit 1989). However, it is not clear that these effects should also hold in

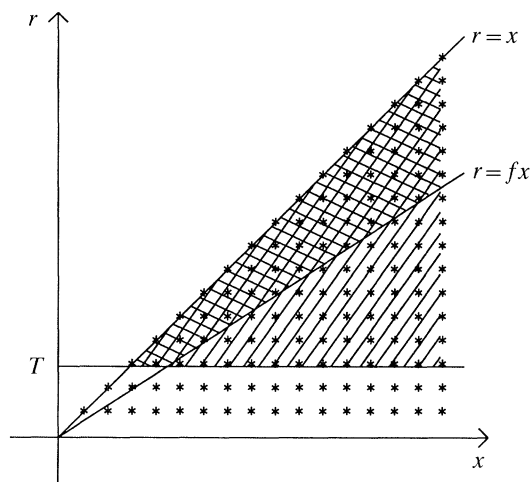


Figure 9. Different methods for setting the threshold on the excitatory neurons. x is the number of active inputs to an excitatory neuron coming from other excitatory neurons; r is the number of these inputs that connect to synapses that have been strengthened in the learning process (thus $r \leq x$). Each point * therefore represents a possible input state of a neuron. In the present theory and in that of Gardner-Medwin (1976), the threshold T is fixed for all the neurons at a particular time step. Thus the neurons with representative points above the line $r = T$ have inputs that exceed threshold and so will be active at the next time step; those with points below will be inactive. In Marr's theory (Marr 1970), only those neurons whose active input exceeds both T and fx are active at the next time step; that is, the 'active' region has been reduced to the double-hatched area. Clearly, this gives more flexibility in setting the threshold. For example, if a memory is almost recalled, then an f close to 1 will be appropriate, since almost all active inputs will be modified ones. However, as explained in the text, there seems to be no biological justification for this more elaborate mechanism.

the present type of network. One significant difference is the asymmetry of the connections. The original Hopfield network was completely symmetric; the relaxation of this condition is equivalent to the addition of noise (Hopfield 1982; Amit 1989) so in this sense the current model is already 'noisy', even without the stochastic release of transmitter at synapses.

Some insight into the role of noise can be obtained by studying a case where retrieval occurs only for $\sigma_N > 0$. Figure 8*a,b* shows a case of retrieval failure when no noise is present. The input signal contains 264 valid and 824 spurious firings; at the first step this is reduced to 7 valid and 0 spurious firings, and this is insufficient to keep the network active. Figure 8*c-f* shows the same situation, only now with $\sigma_N = 1.0$. The effect of noise is to increase the variance of the membrane potential (see equations (47) and (52)) and hence spread out the distributions. This causes them to overlap more, which is undesirable, but it also pushes more neurons over threshold. In the present case this leads to 32 valid firings at the first step, and this is enough to keep the progressive recall process going; by the 8th step the network has settled into a pattern of 301 valid and six spurious firings.

It has already been remarked that there must be some reduction in the number of memories stored if

noise is to have a beneficial effect on the recall process. Related to this is the relative capacity of the network, as discussed in § 5*d* for the noiseless case. The same testing procedure leads to the broken curves in figure 7: as expected, there is a reduction that is not too drastic for $\sigma_N = 0.4$ but is severe when $\sigma_N = 1.0$.

7. DISCUSSION

The present study has much in common with the work of Marr (1970) and Gardner-Medwin (1976). Marr's third layer, P_3 , is a recurrent network which also models memory storage in the CA3 region of the hippocampus (see also Willshaw & Buckingham 1990). The progressive recall theory of Gardner-Medwin is concerned with the synchronous updating of a feed-back network, and was the starting-point for our present theory. However, in spite of these similarities, there are a number of important differences which we wish to highlight.

(a) *The role of inhibition*

All the above theories agree on the importance of inhibition as a mechanism for regulating the activity in the network by setting appropriate thresholds on the excitatory neurons. We suggest a quite specific and biologically plausible mechanism for this threshold setting, based on a system of inhibitory neurons which samples the activity of the network and produces an inhibitory effect dependent on this activity. (The mathematical theory is given in § 3*a,b*; the biological justification is discussed in § 4*a*.)

Gardner-Medwin (1976) simply chooses a threshold T using the criterion that the average number of spurious firings should be kept below a certain level. No biological mechanism is suggested and it is difficult to conceive of one, as this would require a neuron to recognize that it should not be active for the particular pattern being recalled. Such a mechanism is also unsatisfactory for the purposes of a theoretical analysis, as there is no explicit formula for T in terms of the network parameters.

Marr (1970) has a more complicated mechanism in which the threshold R on an excitatory neuron is given by

$$R = \max(T, fx), \quad (30)$$

where x is the number of active afferent synapses on the neuron (whether modified or not), and T, f are parameters termed the subtraction threshold and division threshold, respectively. One problem with this formula is that x is not a quantity which can be measured by the neuron. It is not a measure of the average activity of the network: this is proportional to the expected value of x . The only conceivable way in which x itself could be measured would be for an inhibitory neuron to be associated with each excitatory neuron and to have exactly the same inputs, but of course this does not occur. Another problem is that no biologically plausible mechanism is given for choosing values for T and f . Marr states that the recovery of a memory 'depends on suitable juggling of T and f '

(Marr 1970, p. 44); Willshaw & Buckingham (1990) give several schemes for choosing T and f , but none has a biological basis.

In view of the influential nature of Marr's theories it is worthwhile investigating his threshold-setting mechanism a little further. In figure 9 the axes are the number of active inputs to an excitatory neuron from other excitatory neurons (x) and the number of such active inputs that connect to the neuron via a synapse that has been modified (that is, strengthened) in the learning process (r). In our notation, for the i th neuron,

$$x = \sum_{j=1}^n W_{ij} X_j(t), \quad r = \sum_{j=1}^n W_{ij} J_{ij} X_j(t). \quad (31)$$

Thus, necessarily, $r \leq x$ and in figure 9 the points * correspond to the possible input states of a neuron. If a fixed threshold T is chosen for each time step, then all neurons whose representative points lie in the single- and double-hatched regions above the line $r = T$ are active at the next step; this region can be termed the active region. Conversely, all neurons whose points lie in the inactive region below $r = T$ do not fire at the next time step. This case of fixed T corresponds to the theory of Gardner-Medwin (1976) (although an explicit formula is not given for T) and to the present theory (where $T = n[g_0 + g_1(ax_i + (1-a)y_i)]$; see (19) and (20)).

In Marr's theory, the threshold varies with each neuron depending on the number of active afferent synapses it possesses, as given by equations (3.7) and (3.8) of Marr (1970). The active region is bounded by $r = x$, $r = T$ and $r = fx$, where T and f are chosen arbitrarily, and is shown as the double-hatched region in figure 9. There is now greater flexibility in choosing the active region and this can be advantageous in the recall process. Consider recall from a partial-memory input. In the initial stages many valid neurons will only have a few of their modified inputs active, so it will be appropriate to take $f \ll 1$ and adjust T so as to increase the valid firings and decrease the spurious ones. Later, when the memory is almost recalled, a choice of f close to 1 will be appropriate, as almost all active inputs to a neuron will be to modified synapses ($r \approx x$). The subtractive threshold T is then only needed to prevent neurons with a very small number of active inputs from firing.

More precisely, using Marr's notation, the conditional expectation of r given x is $x(C_0 + C_1\Pi_3)/(C_0 + C_1)$ for valid neurons and $x\Pi_3$ for spurious ones, where C_0 and C_1 are the number of valid and spurious firings, respectively, and Π_3 is the probability that a synapse has been modified. (The relation to our notation is $C_0 = nax$, $C_1 = n(1-a)y$, $\Pi_3 = \rho$.) So if C_1 is very small, taking f close to 1 will cause most neurons which should be active to fire and, if ρ is small and x is not too small, the number of spurious firings will be very small. When x is small it is not possible to keep the number of spurious firings small enough so it is necessary to use the subtractive threshold T to avoid this.

Thus it is clear that, by choosing f and T in some

optimal way at each step, Marr's method will allow for a greater number of stored patterns to be recalled than will a method such as ours, where the threshold is fixed for all neurons at each step. However, the lack of a plausible biological mechanism to account for the dependence on x , and for the choices made for f and T , is a drawback to Marr's method.

(b) Other differences

Other neural network theories either assume complete connectivity (for example, Hopfield 1982; Amari 1989) or incorporate sparseness by assuming a uniform probability that two randomly chosen neurons are connected (Marr 1970; Gardner-Medwin 1976; Derrida *et al.* 1987; Canning & Gardner 1988). The present theory incorporates some of the structure of the CA3 region by using a limited spatial spread for the axonal and dendritic trees and also assuming a falloff in connection probability within the overlap area. The effect of this is to reduce the variance of the connectivity parameter c and hence to reduce the variances of the inputs to the valid and spurious neurons. This in turn means a higher correlation between the final state of the system and the memory being recalled and also some increase in memory capacity. For example, using the same parameters as for figure 2 but with $c^2 = (\bar{c})^2 = 0.0025$ (that is, the homogeneous case) the final overlap is reduced to 0.955 (306 valid, five spurious) compared to 0.984 (322 valid, two spurious) for the non-homogeneous case. Also, the maximum number of stable memories is now about 310 000, compared to 340 000 for the non-homogeneous case (figure 7a).

The full (Level 2) equations take account of the correlations between the elements J_{ij} of the random Hebbian and the state of the system at various time steps; the earlier theories of Marr (1970) and Gardner-Medwin (1976) do not do this. These effects can be large for certain choices of the parameters (Gibson & Robinson 1992); in the extremely sparse networks considered here they are usually small, except in certain restricted parameter regions. The main effect of neglecting these correlations is to falsely increase the range of parameter values for which the network will retrieve memories.

Finally, the present theory allows for the probabilistic secretion of quanta at the synapses to the excitatory neurons. This effect is certainly present in the biological neurons of the CA3 region, and it is of great interest to investigate the effects that it has on model networks. An expected result is that it will degrade the capacity of the network, and this indeed is the case (figure 7a). An unexpected result is that in certain cases it can actually improve the recall of a memory (figures 5, 6 and 8), although this effect is too small to account for the presence of large amounts of synaptic noise.

8. CONCLUSION

A theory for the dynamics of a sparse associative memory (Gibson & Robinson 1992) has been applied to the CA3 recurrent collateral system. The theory is

biologically realistic in that it allows for sparse coding (storage of memories in which few neurons are active) and for the correct level of connectivity between the excitatory neurons. The non-homogeneous nature of this connectivity has also been taken into account by incorporating the elliptical spread of the pyramidal cells' axonal system and the experimentally observed decrease of connection probability with distance. No symmetry is imposed on the connections, and inhibitory and excitatory neurons are distinguished. The stochastic aspect of quantal release of transmitter at synapses is incorporated into the formalism. Less satisfactory aspects are the use of synchronous updating and the implementation of learning in the form of a two-valued Hebbian rule, but these assumptions are at present necessary to obtain analytic formulas for the progressive recall process.

The pyramidal cells are taken to be threshold elements which fire an action potential whenever their membrane potential exceeds a fixed amount. The membrane potential is set by the summed inputs to the cell, and this contains contributions both from other pyramidal cells and from the inhibitory interneurons. A specific model has been proposed for these inhibitory neurons: under assumptions concerning their spiking frequency and the speed with which they process their inputs, it is reasonable to take them to be linear devices whose input is added to the input to the excitatory neurons at the same time step. It follows that they provide an inhibitory input to each pyramidal cell that is proportional to the total firing activity in all the pyramidal cells. With an appropriate choice for the proportionality constant, this input serves to stabilize the system (in the sense of avoiding the extremes of the neurons all firing or all silent, or oscillations between these cases) and allows the progressive recall process to proceed smoothly. This treatment of the inhibitory neurons can be compared with earlier ones, which either do not propose a detailed model or else give a scheme which may be more efficient for memory retrieval, but is biologically untenable.

The equations describing the progressive recall process are a modification of those given in Gibson & Robinson (1992), the modifications being the inclusion of many (as opposed to one) inhibitory neurons, the non-homogeneity of the connections between the excitatory neurons and the inclusion of synaptic noise. It is to be emphasized that these equations describe the dynamics of the recall process, and do not just give the final states (or fixed points) of the system; that is, the number of valid and spurious firings is given for each time step. The theory takes into account both spatial correlations between the connection strengths and the temporal correlations which develop between the states of the system and the connection strengths; in general, however, these are small for very sparse systems of the type treated here. The full theory consists of a set of four coupled nonlinear difference equations. These are mathematically complex and their bifurcation and stability structure is a subject of current investigation. However, their use in an actual calculation is completely straightforward and follow-

ing the system for hundreds of time steps requires only seconds of computer time; thus it is easy to investigate the behaviour of the system for wide ranges of parameter values.

The functioning of the CA3 region as an associative memory has been investigated using a specific theory of memory recall, and with a choice of parameters based as far as possible on the known physiology of this region. The conclusion is that the recurrent collateral pyramidal cell system is indeed capable of such a function, with a large storage capacity and good retrieval from partial memory input, even when this input contains a large number of spurious firings. The average connectivity of \bar{c} is about as low as one would like to go; smaller values of \bar{c} give a rapidly diminishing capacity. (Larger values of \bar{c} do increase the capacity, and in fact generally enhance the performance of the network, but these appear to be ruled out by the physiology of the CA3 region.) The limited range of connections, leading to $\bar{c}^2 = 0.021$ (compared to the homogeneous value of $\bar{c}^2 = \bar{c}^2 = 0.0025$) gives a significant improvement in both capacity and final state overlap. The assumed firing rate of $a = 0.001$ (leading to 330 neurons firing when a memory is recalled) seems a reasonable compromise: a larger value would give better recall, but would diminish the capacity; a smaller value would lead to poor recall.

Inclusion of the noise introduced by stochastic secretion of transmitter at the recurrent collateral synapses leads to improved retrieval, although one must pay a price in terms of reduced storage capacity and slightly reduced quality of recall. Previous analysis of the role of quantal secretion in the performance of biological networks has shown that the stochastic properties of this process extend the ability of granule cell networks to perform pattern separation and activity regulation (Gibson *et al.* 1991). The probabilistic secretion of quanta therefore has important implications for the functioning of neuronal computational systems (Burnod & Korn 1989).

We thank the referees for stressing the importance of comparison with earlier work and also for suggesting the use of an elliptical axonal spread for the pyramidal cells in CA3. Support under ARC Grant AC9031997 is acknowledged.

APPENDIX 1. PROGRESSIVE RECALL EQUATIONS

The total input to the i th neuron at time t is $nh_i(t)$, where $h_i(t)$ is given by equation (5). First, consider the input to the I neurons, and for the present assume that there is no inhibition of I neurons by I neurons (i.e. $J_{II} = 0$); then the input to the i th I neuron can be written

$$h_i(t) = \frac{1}{n} \sum_{j=1}^n W_{ij} J_{IE} X_j(t), \quad i = n+1, \dots, N. \quad (32)$$

Taking the expected value of each side gives

$$\mathcal{E}h_i(t) = \frac{1}{n} \sum_{j=1}^n c_{ij} J_{IE} \mathcal{E}X_j(t), \quad i = n+1, \dots, N, \quad (33)$$

and the use of equation (7) gives equation (11).

Suppose now that $J_{II} \neq 0$. Substituting equation (5) in (7) and regrouping terms leads to

$$\sum_{j=n+1}^N A_{ij} X_j(t) = \frac{\kappa}{n} \sum_{j=1}^n W_{ij} J_{ij} X_j(t), \quad i = n+1, \dots, N, \quad (34)$$

where $A_{ij} = \delta_{ij} + (\kappa J_{II}/n) W_{ij}$. The inverse of the matrix has elements that can be written $\delta_{ij} + B_{ij}/n$ where $B_{ij} = O(1)$ as $n \rightarrow \infty$. Substituting this into equation (34) and taking the expectation gives an expression of exactly the same form as (11), except that $g_1^* = \kappa \bar{c}_{IE} J_{IE}$ is replaced by $g_1^* = \kappa \bar{c}_{IE} J_{IE} (1 + n^* \beta/n)$ where $\beta = \mathcal{E} B_{ij}$. The variances remain the same as before, since the additional term involving B_{ij} gives a negligible contribution for large n .

For the E neurons, it remains to include the correlations discussed in § 3c. The systematic approach is to do the calculation of the expectation of $\mathbf{X}(t)$ in two steps, first conditioning on the previous state of the system. This leads to (14) being replaced by

$$\mathcal{E} X_i(t+1) = \Phi(E_i(t)/\sigma_i(t)), \quad i = 1, \dots, n, \quad (35)$$

where

$$E_i(t) \equiv \mathcal{E} h_i(t) - g_0 = \mathcal{E}[\mathcal{E}(h_i(t) - g_0 | \mathbf{X}(t))], \quad (36)$$

$$[\sigma_i(t)]^2 = \mathcal{E}[\text{var}(h_i(t) | \mathbf{X}(t))]. \quad (37)$$

It remains to find expressions for $E_i(t)$ and $\sigma_i(t)$ which can be used in an actual calculation. Consider first the case of valid firings, for which $i = 1, \dots, na$. The typical input term $h_1(t)$ can be written as the sum of two terms, the first coming from the E neurons and the second from the I neurons:

$$h_1(t) = \frac{1}{n} \sum_{j=1}^n W_{1j} J_{1j} X_j(t) + \frac{1}{n} \sum_{j=n+1}^N W_{1j} (-J_{EI}) X_j(t). \quad (38)$$

The calculation of the expectation of the first term parallels a calculation in Gibson & Robinson (1992; § 6). The second term is similar to (32), and taking the expectation gives a term similar to equation (11). Putting these together gives

$$\mathcal{E} h_1(t) = \bar{c}(ax_1 + (1-a)\rho y_1) - g_1(ax_1 + (1-a)y_1), \quad (39)$$

where x_i, y_i are defined by equation (10), y_i' denotes an expectation conditioned on $J_{ij} = 1$, and ρ, g_1 are defined in § 3d. A similar expression can be found for $\mathcal{E} h_n(t)$, where the n stands for a typical index in the range $na+1, \dots, n$.

The calculation of the variance of $h_1(t)$ starts from equation (38) with $X_j(t)$ in the second term replaced by equations (7) and (32). It is then straightforward to show that

$$n^2 \text{var}(h_1(t) | \mathbf{X}(t)) = \text{var} \left(\sum_{j=1}^n W_{1j} J_{1j} X_j(t) | \mathbf{X}(t) \right) + O(1/n). \quad (40)$$

For n large the $O(1/n)$ term can be neglected and the remaining term has already been evaluated in Gibson & Robinson (1992).

The progressive recall process is given by the four coupled difference equations (15) to (18), involving expectations and variances of $h_i(t)$. The expectations are given by equations (19) and (20), and these also give $E_1'(t)$ and $E_n'(t)$ if ρ is replaced by ρ' where $\rho' = (1 - 2(1 - a^2)^m + (1 - 2a^2 + a^3)^m)/\rho$. The standard deviations are given by

$$[n\sigma_1(t)]^2 = na(\bar{c} - \bar{c}^2)x_1 + n(1-a)\rho y_1' \left(\bar{c} - \bar{c}^2 \rho \frac{y_1'}{y_1} \right) + n^2(1-a)^2 \bar{c}^2 \gamma (y_1')^2, \quad (41)$$

$$[n\sigma_n(t)]^2 = na\rho x_1' \left(\bar{c} - \bar{c}^2 \rho \frac{x_1'}{x_1} \right) + n(1-a)\rho y_1' \times \left(\bar{c} - \bar{c}^2 \rho \frac{y_1'}{y_1} \right) + n^2 \gamma (ax_1' + (1-a)y_1')^2 \bar{c}^2, \quad (42)$$

where $\gamma = (1 - 2a^2 + a^3)^m - (1 - a^2)^{2m}$ and

$$\bar{c}^2 \equiv \overline{c_{EE}^2} = \frac{1}{n} \sum_{j=1}^n c_{ij}^2, \quad i = 1, \dots, n. \quad (43)$$

Equations (41) and (42) also give $\sigma_1'(t)$ and $\sigma_n'(t)$ if everywhere ρ is replaced by ρ' and γ by γ' where $\gamma' = [1 - 3(1 - a^2)^m + 3(1 - 2a^2 + a^3)^m - (1 - 3a^2 + 3a^3 - a^4)^m]/\rho - \rho'^2$. The above set of equations implements the Level 2 approximation in Gibson & Robinson (1992), in which both spatial and temporal correlations are taken into account.

It remains to show how the formalism is modified to incorporate the effects of synaptic noise. The noise arises because repeated arrivals of a signal at the same synapse can have different effects. This is a temporal phenomenon, so one starts by considering a system with all the connections C_{ij} fixed and calculating the conditional expectation and variance of $h_i^m(t)$ as given by equation (24):

$$\mathcal{E}(h_i^m(t) | \mathbf{X}(t), C) = \mu_N h_i(t), \quad (44)$$

$$\text{var}(h_i^m(t) | \mathbf{X}(t), C) = (1/n) \sigma_N^2 h_i(t), \quad (45)$$

where μ_N, σ_N^2 are the mean and variance of N_j . From these it follows that

$$\mathcal{E}(h_i^m(t) | \mathbf{X}(t)) = \mu_N \mathcal{E}(h_i(t) | \mathbf{X}(t)), \quad (46)$$

$$\text{var}(h_i^m(t) | \mathbf{X}(t)) = (1/n) \sigma_N^2 \mathcal{E}(h_i(t) | \mathbf{X}(t)) + \mu_N^2 \text{var}(h_i(t) | \mathbf{X}(t)). \quad (47)$$

All the quantities on the right hand sides have already been calculated. In summary, probabilistic secretion of quanta can be included by replacing equations (15) and (16) by

$$x_{i+1} = \Phi(E_1^m(t)/\sigma_1^m(t)), \quad (48)$$

$$y_{i-1} = \Phi(E_n^m(t)/\sigma_n^m(t)), \quad (49)$$

where

$$E_1^m(t) = \bar{c}(ax_1 + (1-a)\rho y_1') \mu_N - g_1(ax_1 + (1-a)y_1) - g_0, \quad (50)$$

$$E_n^m(t) = \bar{c}\rho(ax_1 + (1-a)y_1') \mu_N - g_1(ax_1 + (1-a)y_1) - g_0, \quad (51)$$

$$[n\sigma_1^m(t)]^2 = n\sigma_N^2 \bar{c}(ax_i + (1-a)\rho y_i') + \mu_N^2 [n\sigma_1(t)]^2, \quad (52)$$

$$[n\sigma_n^m(t)]^2 = n\sigma_N^2 \bar{c}\rho(ax_i + (1-a)y_i') + \mu_N^2 [n\sigma_n(t)]^2, \quad (53)$$

and the corresponding expression for x'_{i+1} and y'_{i+1} are obtained from these by making the usual replacement $\rho \rightarrow \rho'$, $\gamma \rightarrow \gamma'$.

APPENDIX 2. FORMULAS FOR AVERAGE CONNECTIVITIES

From equations (21) and (25),

$$\bar{c} = C \frac{1}{n} \sum_{j=1}^n \exp(-\lambda |r_i - r_j|). \quad (54)$$

If edge effects are neglected, then r_i can be set equal to zero in this formula. Approximating the sum by an integral gives

$$\bar{c} = C \frac{\sigma}{n} \int e^{-\lambda r} dA, \quad (55)$$

where σ is the density of the E neurons, taken to be constant, and dA is the element of area; the integral is over the effective axonal area.

If the axonal area is taken to be circular, then $\lambda = \text{constant}$, $dA = 2\pi r dr$ and the integral is from 0 to R , where R is the extent of the axonal tree. If the area is elliptical with eccentricity e , then as $A = \pi x^2 \sqrt{(1-e^2)}$ is the area of an ellipse with semimajor axis x , $dA = 2\pi \sqrt{(1-e^2)} x dx$ and equation (55) gives

$$\bar{c} = C \frac{\sigma}{n} 2\pi \sqrt{(1-e^2)} \int_0^{R_1} e^{-\lambda_1 x} x dx, \quad (56)$$

where R_1 is the length of the semimajor axis and λ_1 is the value of λ in the direction of the major axis. Thus the main difference with the circular case is the presence of the multiplicative factor $\sqrt{(1-e^2)} = R_2/R_1$, where R_2 is the length of the semiminor axis. The assumption made in employing this procedure is that the connection probability on the ellipse with semimajor axis x depends only on x and this means that λ must be anisotropic (in fact, $\lambda = \lambda_1 x/r$). Performing the integral in equation (56) gives equation (26). A similar calculation gives the mean squared connectivity as

$$\bar{c}^2 = C^2 \frac{\sigma}{n} 2\pi \sqrt{(1-e^2)} \int_0^{R_1} e^{-2\lambda_1 x} x dx, \quad (57)$$

and performing the integral gives equation (27).

REFERENCES

- Amaral, D.G., Ishizuka, N. & Claiborne, B. 1990 Neurones, numbers and the hippocampal network. *Prog. Brain Res.* **83**, 1–11.
- Amit, D.J. 1989 *Modeling brain function*. Sydney: Cambridge University Press.
- Amari, S. 1979 Characteristics of randomly connected threshold-element networks and network systems. *Proc. IEEE* **59**, 35–47.
- Amari, S. 1989 Characteristics of sparsely encoded associative memory. *Neural Networks* **2**, 451–457.
- Amari, S. & Maginu, K. 1988 Statistical neurodynamics of associative memory. *Neural Networks* **1**, 63–73.

- Boss, B.D., Turlejski, K., Stanfield, B.B. & Cowan, W.M. 1987 On the number of neurones in fields CA1 and CA3 of the hippocampus of Sprague-Dawley and Wistar rats. *Brain Res.* **406**, 280–287.
- Braitenberg, V. & Schuz, A. 1983 Some anatomical comments on the hippocampus. In *Neurobiology of the hippocampus* (ed. W. Seifert.), pp. 21–37.
- Burnod, Y. & Korn, H. 1989 Consequences of stochastic release of neurotransmitters for network computation in the central nervous system. *Proc. natn. Acad. Sci. U.S.A.* **86**, 352–365.
- Canning, A. & Gardner, E. 1988 Partially connected models of neural networks. *J. Phys. A (Math. Gen.)* **21**, 3275–3284.
- Derrida, E., Gardner, E. & Zippelius, A. 1987 An exactly solvable asymmetric neural network model. *Europhys. Lett.* **4**, 167–173.
- Faris, W.G. & Maier, R.S. 1988 Probabilistic analysis of a learning matrix. *Adv. appl. Prob.* **20**, 695–705.
- Finch, D.M., Nowlin, N.L. & Babb, T.L. 1983 Demonstration of axonal projections of neurones in the rat hippocampus and subiculum by intracellular injection of HRP. *Brain Res.* **271**, 201–216.
- Gamrani, H., Onteniente, P., Seguela, P., Geffard, M. & Calas, A. 1986 Gammaaminobutyric acid-immunoreactivity in the rat hippocampus. A light and electron microscopic study with anti-GABA antibodies. *Brain Res.* **364**, 30–38.
- Gardner-Medwin, A.R. 1976 The recall of events through the learning of associations between their parts. *Proc. R. Soc. Lond. B* **194**, 375–402.
- Gardner-Medwin, A.R. 1989 Doubly modifiable synapses: a model of short and long term auto-associative memory. *Proc. R. Soc. Lond. B* **238**, 137–154.
- Gibson, W.G. & Robinson, J. 1992 Statistical analysis of the dynamics of a sparse associative memory. *Neural Networks* **5**, 645–661.
- Gibson, W.G., Robinson, J. & Bennett, M.R. 1991 Probabilistic secretion of quanta in the central nervous system: granule cell synaptic control of pattern separation and activity regulation. *Phil. Trans. R. Soc. Lond. B* **332**, 199–220.
- Hopfield, J.J. 1982 Neural networks and physical systems with emergent collective computational abilities. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2554–2558.
- Ishizuka, N., Krzemhewska, K. & Amaral, D.G. 1986 Organization of pyramidal cell axonal collaterals in field CA3 of the rat hippocampus. *Soc. Neurosci. Abst.* **12**, 1254.
- Lacaille, J.C., Mueller, A.L., Kimkel, D.D. & Schwartzkroin, P.A. 1987 Local circuit interactions between oriens/alveus interneurons and CA1 pyramidal cells in hippocampal slices: electrophysiology and morphology. *J. Neurosci.* **7**, 1979–1995.
- Lacaille, J.C. & Schwartzkroin, P.A. 1988 Stratum lacunosum – moleculare interneurons of hippocampal CA1 region. Intracellular and intradendritic recordings of local circuit synaptic interactions. *J. Neurosci.* **8**, 1411–1424.
- Little, W.A. 1974 The existence of persistent states in the brain. *Math. Biosci.* **19**, 101–120.
- Little, W.A. & Shaw, G.L. 1975 A statistical theory of short and long term memory. *Behav. Biol.* **14**, 115–133.
- Marr, D. 1971 Simple memory: a theory for archicortex. *Phil. Trans. R. Soc. Lond. B* **262**, 23–81.
- Mason, A. & Larkman, A. 1990 Correlations between morphology and electrophysiology of pyramidal neurones in slices of rat visual cortex: II. *Electrophysiology. J. Neurosci.* **10**, 1415–1428.
- McNaughton, B.L. & Morris, R.G.M. 1987 Hippocampal

- synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* **10**, 408–415.
- McNaughton, B.L. & Nadel, L. 1990 Hebb-Marr networks and the neurobiological representation of action in space. In *Neuroscience and connectionist theory* (ed. M. A. Gluck & D. E. Rumelhart), pp. 1–63. New Jersey: Lawrence Erlbaum Assoc.
- Miles, R. 1990 Synaptic excitation of inhibitory cells by single CA3 hippocampal pyramidal cells of the guinea-pig in vitro. *J. Physiol.* **428**, 61–77.
- Miles, R., Traub, R.D. & Wong, R.K.S. 1988 Spread of synchronous firing in longitudinal slices from the CA3 region of the hippocampus. *J. Neurophysiol.* **60**, 1481–1496.
- Miles, R. & Wong, R.K.S. 1984 Unitary inhibitory synaptic potentials in the guinea pig hippocampus in vitro. *J. Physiol.* **356**, 97–113.
- Miles, R. & Wong, R.K.S. 1986 Excitatory synaptic interactions between CA3 neurones in the guinea pig hippocampus. *J. Physiol.* **373**, 397–418.
- Miles, R. & Wong, R.K.S. 1987 Inhibitory control of local excitatory circuits in the guinea-pig hippocampus. *J. Physiol.* **388**, 611–629.
- Misgeld, U. & Frotscher, M. 1986 Post synaptic-GABAergic inhibition of non-pyramidal neurones in the guinea-pig hippocampus. *Neurosci.* **19**, 193–206.
- Palm, G. 1980 On associative memory. *Biol. Cyber.* **36**, 19–31.
- Palm, G. 1988 Local synaptic rules with maximal information storage capacity. In *Neural and synergetic computers* (ed. H. Haken), pp. 100–110. Berlin: Springer Verlag.
- Redman, S.J. 1990 Quantal analysis of synaptic potentials in neurons of the central nervous system. *Physiol. Rev.* **70**, 165–198.
- Rolls, E.T. 1989 Functions of neural networks in the hippocampus and neocortex in memory. In *Neural models of plasticity* (ed. J. H. Byrne & W. D. Berry), pp. 240–265. San Diego: Academic Press.
- Schlandler, M. & Frotscher, M. 1986 Non-pyramidal neurones in the guinea-pig hippocampus. *Anal. Embryol.* **174**, 35–47.
- Smith, K.L., Turner, J. & Swann, J.W. 1988 Paired intracellular recordings reveal mono- and polysynaptic excitatory interactions in immature hippocampus. *Soc. Neurosci. Abstr.* **14**, 833.
- Tamamaki, N.Y. & Nojyo, Y. 1991 Cross-fiber arrays in the rat hippocampus as demonstrated by three-dimensional reconstruction. *J. comp. Neurol.* **303**, 435–442.
- Traub, R.D. & Miles, R. 1991 *Neuronal networks of the hippocampus*. New York: Cambridge University Press.
- Treves, A. & Rolls, E.T. 1990 Neuronal networks in the hippocampus involved in memory. In *Statistical mechanics of neural networks* (ed. L. Garrido) (*Lect. Notes Physics* **368**), pp. 81–95. Berlin: Springer.
- Treves, A. & Rolls, E.T. 1991 What determines the capacity of autoassociative memories in the brain? *Network* **2**, 371–397.
- Treves, A. & Rolls, E.T. 1992 Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* **2**, 189–200.
- Willshaw, D.J. & Buckingham, J.T. 1990 An assessment of Marr's theory of the hippocampus as a temporary memory store. *Phil. Trans. R. Soc. Lond. B* **329**, 205–215.
- Willshaw, D.J., Buneman, O.P. & Longuet-Higgins, H.C. 1969 Non-holographic associative memory. *Nature, Lond.* **222**, 960–962.
- Yeckel, M.F. & Berger, T.W. 1990 Feedforward excitation of the hippocampus by afferents from the entorhinal cortex: redefinition of the role of the pathway. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5832–5836.

Received 10 May 1993; accepted 28 May 1993