

Computational Constraints Suggest the Need for Two Distinct Input Systems to the Hippocampal CA3 Network

Alessandro Treves and Edmund T. Rolls

University of Oxford, Department of Experimental Psychology,
South Parks Road, Oxford OX1 3UD, England

ABSTRACT

The CA3 network in the hippocampus may operate as an autoassociator, in which declarative memories, known to be dependent on hippocampal processing, could be stored, and subsequently retrieved, using modifiable synaptic efficacies in the CA3 recurrent collateral system. On the basis of this hypothesis, the authors explore the computational relevance of the extrinsic afferents to the CA3 network. A quantitative statistical analysis of the information that may be relayed by such afferent connections reveals the need for two distinct systems of input synapses. The synapses of the first system need to be strong (but not associatively modifiable) in order to force, during learning, the CA3 cells into a pattern of activity relatively independent of any inputs being received from the recurrent collaterals, and which thus reflects sizable amounts of new information. It is proposed that the mossy fiber system performs this function. A second system, with a large number of associatively modifiable synapses on each receiving cell, is needed in order to relay a signal specific enough to initiate the retrieval process. This may be identified, we propose, with the perforant path input to CA3.

Key words: hippocampus, autoassociative memory, attractor neural networks, associative synapses, information storage

Evidence from a number of different disciplines has increasingly shed light on the role that the mammalian hippocampal formation plays in memory (for reviews, see Morris et al., 1988; Squire et al., 1990; Storm-Mathisen et al., 1990). The hippocampus appears to be particularly implicated in the formation of long-term episodic and semantic memories (Squire et al., 1989; Rolls, 1989b; 1990), although it does not seem to constitute the unique final storage site for such memories (Zola-Morgan and Squire, 1990). A number of attempts have been made to integrate the available evidence into a theoretical understanding and, more specifically, to relate the hypothesized function of the hippocampal formation in memory to its anatomical and physiological characteristics. One of us has described a theory of the hippocampus that relates its function to the operation of the neuronal networks that comprise it (Rolls, 1989a; 1989b; 1990). The theory assigns a crucial role to the CA3 network in forming memories of discrete events. It is postulated that any new event to be memorized is given a representation as the firing pattern of CA3 pyramidal cells, and that it is the extensive intrinsic connectivity formed by the axon collaterals from the CA3 cells to other CA3 cells that allows for the retrieval of a whole representation to be initiated by the activation of some small

part of the same representation. The CA3 network is thus considered to operate, in part at least, as an autoassociative memory, in which neuronal representations of different events, or episodes, experienced by the organism may be stored (on the modifiable synapses of the collateral connections), and from which they may be retrieved following a partial cue.*

In addition to the extensive and numerous collateral connections, another anatomically observed salient feature of CA3 circuitry is the presence of two distinct major input systems: the perforant path (PP) and the mossy fibers (MF). The perforant path originates in the entorhinal cortex, the structure that serves as the main input station to the hippocampal formation, while the mossy fibers are the axonal projections of the granule cells of the dentate gyrus. As the granule cells receive most of their extrinsic input from the entorhinal cortex, the mossy fibers appear to carry, by and large, the same information, although relayed by the dentate gyrus, as that

* Note that while it is convenient to distinguish conceptually between storage and retrieval, retrieval might be useful even during the storage phase of episodic memories, because it would allow the whole of the episodic memory to be held stable if some of its individual components were to change. This stability could help the storage of long-term memory elsewhere, for example in the cerebral neocortex (see Rolls, 1989b). Note also that the theory considered here does not require that memories be stored in the long-term in the CA3 net. The theory is *neutral* on this issue.

Correspondence and reprint requests to: A. Treves, University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, England.

conveyed directly by the perforant path. It is intriguing, therefore, to try to understand the function of this "double" projection. This paper attempts to show that within the theory that regards CA3 as an autoassociative memory, a requirement naturally arises for two extrinsic input systems that would have the distinct qualitative features apparently characterizing the perforant path and mossy fiber connections. This further demonstrates, in a manner that is predictive and in principle testable, how specific elements of a given brain structure could be associated with specific computational subtasks that are part of the function ascribed to the structure as a whole.

Organization of CA3 circuitry

The most distinctive feature of hippocampal circuitry is the sequence of three unidirectional, nonreciprocated systems of (excitatory) projections from the entorhinal cortex to the dentate gyrus (DG), from the DG to CA3 (the mossy fibers), and from CA3 to CA1 (the Schaeffer collaterals). The three stages: DG, CA3, and CA1, therefore receive a sort of natural ordering in terms of information flow, notwithstanding the fact that other systems of projections also exist, for example, direct entorhinal projections to both CA3 and CA1 (Fig. 1). It is noteworthy that, along this sequence, the number of cells involved first decreases from DG to CA3 and then increases from CA3 to CA1, in a way that is relatively consistent across different species, as shown in Table 1.

Among the three systems of projections mentioned above, the mossy fiber projection is by far that with the smallest degree of divergence and convergence: it has been estimated that in the rat (Amaral et al., 1990) each granule cell could reach of the order of 14 CA3 cells, and conversely (based on the data in Table 1) that each CA3 cell would not receive from more than 46 granule cells, on average. Within CA3, a far more extensive input system is that provided by the perforant path connections, which, originating in entorhinal cor-

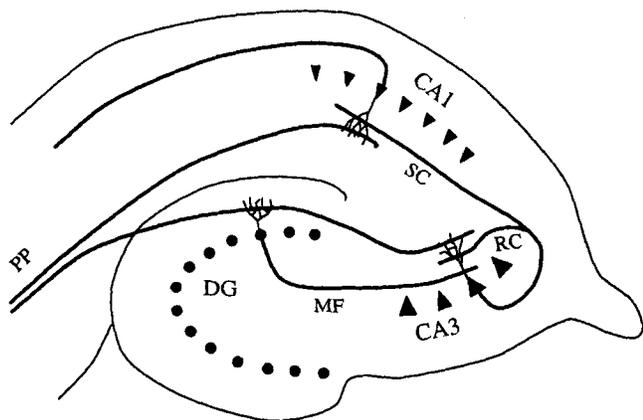


Fig. 1. Main systems of excitatory connections within the hippocampus. Perforant path (PP) inputs from the entorhinal cortex reach the granule cells in the dentate gyrus (DG), and the pyramidal cells of CA3 and CA1. The granule cells project to CA3 via the mossy fibers (MF), while CA3 cells project both to CA1 via the Schaeffer collaterals (SC) and to other CA3 cells via the recurrent collaterals (RC).

Table 1. Number of Cells in the Hippocampus of Rats and Primates

| | DG | CA3 | CA1 | Reference |
|-----------------|-------------------|--------------------|--------------------|-------------------------|
| Rat (SD strain) | 1.0×10^6 | 0.33×10^6 | 0.42×10^6 | Boss et al., 1985; 1987 |
| Rhesus monkey | 4.8×10^6 | 0.86×10^6 | 1.4×10^6 | Seress, 1988 |
| Human | 8.8×10^6 | 2.3×10^6 | 4.7×10^6 | Seress, 1988 |

The numbers are for one side of the brain, and those for CA3 include the CA2 subfield. Interestingly, it was recently reported (West and Gundersen, 1990) that, when correcting appropriately for post-mortem shrinkage of tissues, the numbers for humans increase to $15 \times 10^6 \pm 28\%$ (DG), $2.7 \times 10^6 \pm 22\%$ (CA3), and $16 \times 10^6 \pm 32\%$ (CA1).

tex, make synapses on the apical dendrites of CA3 pyramidal cells. Based on the average extent of dendritic arborization in the layer innervated by the perforant path (stratum lacunosum/moleculare), it has been calculated that each CA3 pyramidal cell may receive up to 3750 PP synapses in the rat (Amaral et al., 1990), possibly from as many different cells in entorhinal cortex. The largest number of synapses (about 12,000) on the dendrites of CA3 pyramidal cells is, however, provided by the axon collaterals of CA3 cells themselves. (The numbers of synapses cited have been estimated for the rat by Amaral et al. (1990), and comparable data are not yet available for primates.) The CA3 system therefore is, uniquely within the hippocampus, a system in which intrinsic, recurrent excitatory connections are, at least numerically, dominant with respect to excitatory afferents. In addition to the above, there are also intrinsic connections with a variety of numerically limited and mainly inhibitory populations of interneurons, and also extrinsic connections with sublimbic structures. These latter systems are all postulated, in the present work, to subservise generic regulation of neural activity in CA3, as opposed to providing signals specific to the information being processed in the system, and their operation is not considered in detail in the following.

The focus, then, is on the population of CA3 pyramidal cells and on three systems of synapses they receive: those from the recurrent collaterals (RC), of which there are relatively many per cell; those from the entorhinal cortex via the PP, of which there are also relatively many; and MF synapses from the dentate gyrus, of which there are relatively few (Fig. 2). The ratio between the average number of PP synapses per cell and that of RC synapses is important in determining their relative influence on the firing of CA3 cells, and will be denoted as K . On the basis of the information available for the rat, it will be assumed that realistic values of K (possibly also in the case of primates) are around, say, $K = \frac{1}{4}$.

Two important features characterizing these different sets of synapses are their relative strength and their plasticity. MF synapses are presumed to be individually strong and to display long-term activity-dependent synaptic enhancement (LTP). This form of enhancement appears not to be associative (Zalutsky and Nicoll, 1990). RC and PP synapses, by contrast, are thought to be individually weak, and of the NMDA type, implicated elsewhere in an associative type of plasticity (Miles, 1988; Brown et al., 1990). What is meant

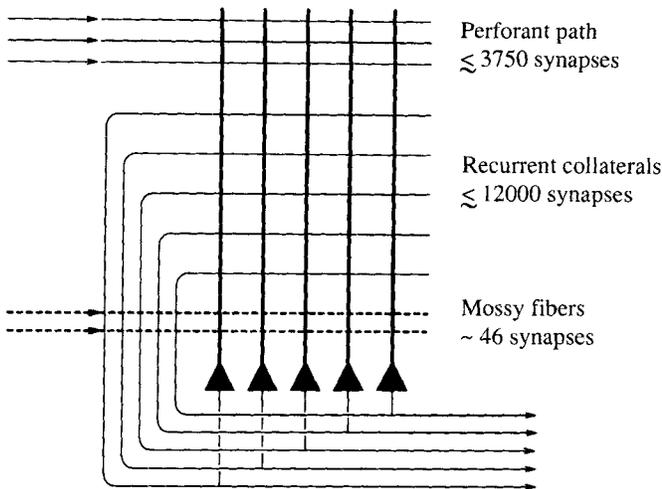


Fig. 2. The CA3 network. This schematic diagram depicts the main systems of excitatory synaptic connections to pyramidal cells, with the estimates, for the Sprague-Dawley rat, of the number of synapses per cell for each system (Amaral et al., 1990). Inputs come from the left and outputs leave to the right of the diagram.

here by associative and nonassociative plasticity is that in the first case, and not in the second, the change in the efficacy of the synapse due to a learned event depends (in a conjunctive, nonadditive manner) on both the activity in the presynaptic terminal during the event, and on the conditions prevailing at the postsynaptic site, as determined by other factors (including the activity at different synapses). A more quantitative definition of associative plasticity is given below.

Another potentially important feature of synaptic organization is the connectivity pattern determining the probability of contact between any two given cells in the pre- and postsynaptic population. Indeed, definite gradients in the topographical distribution of the contact probability have been described, for example, in the recurrent collateral system of the rat (Ishizuka et al., 1990). Thus, the projections from CA3 cells located close to the DG may be restricted in their destination to the same CA3 portion proximal to the DG and to the same septotemporal level of origin, whereas cells located closer to CA1 tend to have more widespread projections, both along the DG-CA1 direction and along the septotemporal axis. The mossy fiber system, by contrast, appears to have a much more restricted divergence, in that each granule cell tends to contact only cells in the same septotemporal "slice" of the CA3 field. It should be noted that some of these trends vary considerably between different species. Thus, while in the rat it has been observed that among the recurrent collaterals in CA3 the ipsilateral connections and the commissural fibers from the contralateral hippocampus carry roughly the same weight, in the monkey only a quantitatively and topographically limited portion of the collateral connections comes through from the contralateral side (Amaral et al., 1984). These detailed connectivity patterns in CA3, however, will not substantially affect the arguments that follow, inasmuch as the recurrent collateral connections are divergent enough to guarantee that excitation can potentially be spread from any given pyramidal cell to any other pyramidal cell in

CA3 within a few synaptic steps, or within a few tens of milliseconds. This property seems indeed to hold, with the possible proviso that in some species that include primates, due to the limited commissural connections mentioned above, the CA3 fields in the two hippocampi should be more appropriately regarded as distinct systems, with possibly some degree of lateralization of function (Rolls, 1990).

Collateral connections and memory retrieval

The hippocampal theory, on which our considerations are based, hypothesizes that the role of CA3 in the formation of episodic memories is as follows. When the animal experiences a particular episode, a multisensory constellation of stimuli runs through and is processed along various sensory pathways to form an ensemble of neural representations of different elements of the episode. These highly elaborated signals, rich in information content, reach the hippocampus through the overlying areas of cortex and (possibly upon concurrent excitation or disinhibition by nonspecific activating systems) elicit a definite pattern of firing responses in the CA3 pyramidal cells. This firing pattern, which constitutes a unitary representation of the episode, results specifically from the direct entorhinal input (Yeckel and Berger, 1990), from the DG-relayed mossy fiber input and, after a brief delay, from the detailed feedback effect of recurrent connections. Labelling CA3 pyramidal cells $1, \dots, i, \dots, N$, and assigning the label μ to the particular episode being represented, the set of firing rates, when averaged over some prescribed period after the transient, is denoted as $\{\eta_i^\mu\}$. The theory then postulates that, if representation $\{\eta_i^\mu\}$ is to be stored, modifications occur in the efficacies of different systems of synaptic connections, including the CA3 recurrent connections, with a change in the efficacy of each particular synapse dependent on the activity of the pre- and postsynaptic cell during learning. Such synaptic modifications constitute a long-term memory trace and persist, superimposed with subsequent modifications, while the system proceeds to learn different episodes (in the way specified below). Later, a partial cue, in the form of an afferent signal having some correlation with the signal that reached CA3 during episode μ , is postulated to be able, by virtue of the trace embedded in the recurrent synapses and once activity is circulated through those synapses, to elicit a firing pattern $\{V_i\}$ strongly correlated with $\{\eta_i^\mu\}$. This cued *memory retrieval* may subservise, among other functions, the consolidation of longer-term forms of memory storage outside the hippocampus, as hypothesized in the early model of Marr (1971).†

The crucial property in the proposed mechanism is the ability of the CA3 system to perform as an autoassociative memory in selectively retrieving a specific firing pattern among several stored on the same set of synaptic efficacies (Gardner-Medwin, 1976). This is a highly nontrivial property, if the system is assumed to operate in the presence of both intrinsic

† Curiously, while Marr's view of the hippocampus has inspired much of the later research, his own specific formal model, which is only distantly related to actual hippocampal circuitry, does not rely crucially on the operation of recurrent collaterals, as elucidated in a recent critique (Willshaw and Buckingham, 1990).

sources of noise and interference effects due to the storage of many patterns. It has been the merit of formal models of autoassociative memory (Hopfield, 1982) to show that this ability indeed pertains, at least, to simplified systems that mimic networks with recurrent connections. A quantitative analysis of such formal systems (Amit et al., 1987) shows that the ability to retrieve persists up to a critical storage load, the storage capacity. Extending previous analyses, which had dealt mainly with systems of binary units, to systems whose units represent more realistically neuronal (graded) firing behavior (Treves, 1990), we have elaborated quantitatively some of the factors that may contribute to determining the storage capacity of actual autoassociative networks in the brain, in particular in the CA3 network (Treves and Rolls, 1991).

The formal models mentioned above, by considering steady-state input output relations, describe only the conditions that enable recurrent processing through the collateral connections to make the current firing pattern $\{V_i\}$ strongly correlated with a stored pattern $\{\eta_i^\mu\}$. The initiation of the retrieval process, however, and prior to that, the storage of the patterns themselves, depend critically on the input systems to the autoassociative network, and it is on these two problems that the present analysis is focused. The basic assumption underlying our arguments is that a biologically plausible solution to these two relatively simpler problems has to be at least good enough to fully exploit the ability of the network to cope with the harder problem of completing retrieval. Thus, for example, given that it is possible to estimate the total amount of information that can be retrieved from a given autoassociative network, it will be considered here to be reasonable to require that the network be able to store at least that amount of information in the first place. In order to specify these requirements quantitatively, we summarize next the results concerning retrieval in an autoassociative network such as may be implemented by CA3 (Treves and Rolls, 1991).

Assume that p firing patterns $\{\eta_i^\mu\}$ are stored with equal weight, and that each pattern has the same statistical distribution of firing rates, in particular that the firing rate of each cell in each pattern follows, independently, a common probability distribution P_η . This is not necessarily a realistic situation, but serves to define a convenient framework for capacity measures. Note that the assumption that each cell carries statistically independent signals is in line with the hypothesis that the total information expressed by the firing pattern is a quantity to be maximized with respect to the fixed number of cells in CA3. This may involve removing redundancy implicit in early sensory representations (Barlow, 1960; Rolls, 1987).

Within these assumptions, one can obtain, analytically, the maximum value of p , above which interference effects make retrieval impossible. p_{max} turns out to be proportional to C^{RC} , the number of associatively modifiable synapses per cell (on the recurrent collaterals), and, roughly, inversely proportional to the sparseness a of the neuronal representation. The sparseness is defined as

$$a = \frac{\langle \eta \rangle^2}{\langle \eta^2 \rangle} \quad (1)$$

where $\langle \cdot \rangle$ denotes an average over the distribution P_η characterizing the patterns $\{\eta_i^\mu\}$. Sparser distributions, $a \ll 1$, imply less interference among different patterns, and therefore allow for higher values of p . To a first approximation one finds

$$p_{max} \approx \frac{C^{RC}}{a \ln(1/a)} \times \text{slowly varying factor} \quad (2)$$

where the factor depends on the detailed structure of the rate distribution, on the connectivity pattern, etc., but is broadly speaking in the order of 0.2–0.3 (see Treves and Rolls, 1991, for a more extensive discussion \ddagger).

Another, possibly more meaningful, measure of capacity is I_{max} , the maximal amount of information (in bits per synapse) that can be retrieved from the network. It is defined as follows. Each pattern carries, with respect to the *a priori* measure defined by P_η , a certain information content (Shannon and Weaver, 1949) per cell, i_p . During retrieval, an amount i_i of information per cell will be lost, as the retrieved pattern $\{V_i\}$ will have high but not full correlation with $\{\eta_i^\mu\}$, due to noise effects including those derived from extensive storage of other patterns. Multiplying the subtracted quantity by p/C^{RC} , one obtains a measure of the total retrievable information per synapse as

$$I \equiv \frac{p}{C^{RC}} (i_p - i_i) \quad (3)$$

Its maximal value I_{max} was calculated (Treves and Rolls, 1991) and found to be in several interesting cases \S around 0.2–0.3 bits per synapse, with only a mild dependency on parameters such as the sparseness of coding a . Note that as more patterns are stored with a fixed P_η , obviously I grows with p , but also goes down as i_i increases due to more and more severe interference effects. In fact the maximum I_{max} is obtained for p much below p_{max} .

Here we can derive a reasonable estimate of how much information has to be stored in each pattern for the network to exploit its capacity I_{max} efficiently. Considering the network to operate with p well below p_{max} , in which case i_i may not constitute a large fraction of i_p , leads to the requirement on i_p :

$$i_p > a \ln(1/a) \quad (4)$$

that has to be satisfied, in some meaningful range of the sparseness a , for efficient information retrieval. i_p depends on P_η , which itself depends on the conditions prevailing during information storage in CA3. Moreover, these conditions determine how much of the information contained in the afferent inputs to CA3 actually goes into i_p , and how much of i_p , instead, is "fake" information resulting from the storage process, as will be clear below. An analysis of the storage process may therefore yield constraints on the inputs to CA3, as shown in the next section.

We consider, for the arguments that follow, that geological evolution, in order to optimize the organization of the CA3

\ddagger As a concrete example, if $a = 0.02$ and $C^{RC} = 12,000$, one has $p_{max} \approx 36,000$.

\S See also the discussion by Nadal (1991).

network as an autoassociative memory, had at its disposal two basic types of synapses to use on afferent inputs. One, identified *a posteriori* with the PP synapse, is taken to be an associatively modifiable synapse with characteristics, in particular average strength similar to the RC synapse. The other is a potentially much stronger synapse, but lacking the type of plasticity suitable for associative learning. The conclusion we shall arrive at is that, indeed, an autoassociator needs *both* types of inputs each with a distinct role, and that these are in roughly the same qualitative proportions as PP and MF synapses on the dendritic tree of CA3 cells.

RESULTS

Strong specific inputs are necessary for information storage

The first argument concerns what type of input is necessary during storage. We will show that an input of the PP type alone is unable to direct efficient information storage. Such an input is too weak, it will turn out, to drive the firing of the cells into a pattern strongly correlated with itself, as the "dynamics" of the network is dominated by the randomizing effect of the recurrent collaterals.

Suppose then that the mossy fiber system is not present, or inactive, during the storage of a particular episode, and that the only information-rich input to every CA3 cell comes through thousands of individually weak synapses located on the apical dendrites. Even though the efficacy of each of these synapses may have been associatively modified while learning previous episodes, it can be regarded as a random variable with normal (Gaussian) distribution, provided the present distribution of activity on the axonal fibers is unrelated to those activity distributions occurring in past learning events. More important, due to the large number of synapses, the *integrated synaptic current* coming down from the apical dendrites, h_i^{PP} , can also be regarded, whatever the activity distribution on the axons, as a random Gaussian variable whose variance is proportional to the number of synapses C^{PP} (Table 2).

Once the response of CA3 cells has been elicited, the total synaptic current reaching the soma will also include the contribution mediated by recurrent collaterals h_i^{RC} . There will also be contributions from nonspecific afferents and inhibitory interneurons but, as these are assumed to be poor in information content, the variance in their distributions over different CA3 cells may be neglected. As noted above, the

mossy fibers are supposed, just for the sake of this argument, not to contribute to h_i during storage. Just as with h_i^{PP} , the distribution of h_i^{RC} may be taken as Gaussian, with a variance proportional to the number of recurrent collateral synapses C^{RC} . This time, however, the variance may be shown to be enhanced also by a factor originating from the feeding back of correlations between the current firing pattern and previously learned patterns. Without further elaboration, for which we refer to our previous work (Treves and Rolls, 1991), we denote this enhancement factor as $(1 - \psi)^{-2}$, where $0 \leq \psi \leq 1$ and ψ is a measure of the degree of feedback, which depends, e.g., on the details of the recurrent connectivity.

The total synaptic current is therefore expressed as

$$h_i = \bar{h} + \Delta h_i^{PP} + \Delta h_i^{RC} \quad (5)$$

where \bar{h} is the input averaged over the CA3 cell population. The important point is that the distribution of Δh_i^{PP} and Δh_i^{RC} are independent, because the distribution of efficacies on the RC synapses reflects only previously learned activity patterns, unrelated to the current one. If, moreover, the PP and RC synapse have similar average strengths, and if the afferent PP axons also carry similar average activity as the RC axons, then the ratio between the variances will be determined essentially by the ratio between C^{PP} and C^{RC} ,

$$\frac{\rho_{PP}^2}{\rho_{RC}^2} \approx K(1 - \psi)^2, \quad (6)$$

where

$$\rho_{PP}^2 = \overline{(\Delta h^{PP})^2} \quad \rho_{RC}^2 = \overline{(\Delta h^{RC})^2}. \quad (7)$$

Consider now learning to occur, in that the efficacies of both PP and RC synapses are modified. If it is a reasonable approximation to consider the cells as point-like (and indeed the large CA3 pyramidal cells are electrotonically quite compact), then the modifications at all synapses may be taken to reflect, in their postsynaptic dependence, the time-averaged membrane potential. This, in turn, is reflected directly in the time-averaged firing rate η_i of each cell (at least above threshold, but on the other hand relatively large depolarizations appear to be necessary to induce synaptic modification; Collingridge and Singer, 1990); so that the *information* that is being encoded on the synaptic efficacies may be taken to be i_p , the information contained in the firing pattern $\{\eta_i\}$. Now $\{\eta_i\}$ is determined by both h_i^{PP} and h_i^{RC} , but the only information that matters is that contained in the input, and therefore in h_i^{PP} , while h_i^{RC} only adds a random distortion determined by those memory traces that preexisted at the time of learning.^{||} This is an important point: if the only purpose of future memory retrievals were to identify the firing pattern stored in CA3 most correlated with the cue, randomly generated representations would suffice; but if, on the other

Table 2. Main Symbols Used in the Formal Analysis

| | |
|--------------|---|
| η_i^μ | Firing rate of cell i in the μ^{th} stored pattern |
| V_i | Firing rate of cell i in retrieving a pattern |
| p | Number of patterns concurrently stored |
| C | Number of synapses per cell |
| a | Sparseness of a set of firing patterns (see equation 1) |
| I | Information stored in the network, per synapse |
| i | Information content of one pattern, per cell |
| h_i | Integrated synaptic current into cell i |
| J | Synaptic efficacy |
| q | Cue: correlation in the afferent firing pattern |
| S | Signal: correlation in the synaptic currents |

^{||} Note that, in the presence of spin-glass effects, in particular the spin-glass-like multiplicity of stationary uncorrelated patterns characterizing, for example, the standard Hopfield network (Amit, 1989), $\{\eta_i\}$ would not be determined uniquely by external inputs and existing synaptic efficacies. However, this type of multiplicity has been shown not to be present when realistic features such as graded response and inhibitory control of rates are taken into account (Treves, 1991).

hand, the information contained in the retrieved pattern is to be used, or in other words if the response profiles of CA3 cells are to have some meaning, then this information has to be supplied by the input lines.

The amount of information present in the input $\{h_i^{PP}\}$ that survives in the pattern $\{\eta_i\}$, and is therefore *storable*, is easy to quantify. If the fraction of cells that receive input between h^{PP} and $h^{PP} + dh^{PP}$ is denoted as $c^-(h^{PP})dh^{PP}$, the fraction of cells that fire at rates between η and $\eta + d\eta$ is $c_-(\eta)d\eta$, and the fraction of cells that satisfy *both* conditions is $c(h^{PP}, \eta)dh^{PP}d\eta$, then the storable information per cell is (cf. Treves, 1990)

$$i_s = \iint dh^{PP}d\eta c(h^{PP}, \eta) \ln_2 \left[\frac{c(h^{PP}, \eta)}{c^-(h^{PP})c_-(\eta)} \right]. \quad (8)$$

Note that $c^-(h^{PP}) = \int d\eta c(h^{PP}, \eta)$ and $c_-(\eta) = \int dh^{PP} c(h^{PP}, \eta)$. If the pattern $\{\eta_i\}$ is *independent* of the input distribution $\{h_i^{PP}\}$, then $c(h^{PP}, \eta) = c^-(h^{PP})c_-(\eta)$ and no information survives, $i_s = 0$. If on the other there is one-to-one correspondence between input and firing activity, $c^-(h^{PP}) = c(h^{PP}, \eta) = c_-(\eta)$, then all the information contained in the input, $i_0 = - \int dh^{PP} c^-(h^{PP}) \ln_2 c^-(h^{PP})$, is passed on into $\{\eta_i\}$.

Using expressions 5 and 6, it is easy to compute i_s in the case in which a purely *linear* relation holds between h_i and η_i :

$$i_s = \frac{1}{2} \ln_2 [1 + K(1 - \psi)^2] \approx \frac{K(1 - \psi)^2}{2 \ln 2}.$$

Note that a linear relation is equivalent to having essentially the whole (Gaussian) h_i distribution above the firing threshold: a situation that produces a distribution $P_\eta(\eta_i)$ with limited relative variance, i.e., tending to the limit of no sparseness, $a \rightarrow 1$ (recall equation 1). If the relation between h_i and η_i is more realistically considered to include a threshold nonlinearity within the normal range, a decreases, but also i_s can only decrease. As an example, an expression is calculated in Appendix A for i_s when a simple *threshold-linear* relation is assumed, with threshold T_{hr} . The result depends on the ratio r between the mean suprathreshold current $\bar{h} - T_{hr}$ and its standard deviation $(\rho_{PP}^2 + \rho_{RC}^2)^{1/2}$. Different values of r also produce rate distributions $P_\eta(\eta_i)$ with different degrees of sparseness a , so that what is displayed (bottom of Figure 3) is a plot of i_s versus a . No matter how sparse the coding is in the autoassociator (and the logarithmic scale of the plot emphasizes possibly relevant ranges), i_s appears to be below the range required for i_p , as expressed in equation 4. Therefore one may conclude that only an insufficient amount of information can be stored in these circumstances. The reason for that is clear: *the information contained in $\{h_i^{PP}\}$ is washed away by the intervening feedback $\{h_i^{RC}\}$* . The argument is even more compelling when considering that, the PP synapse being actually located on the apical dendrites further away from the soma than the RC synapses, their effective strength might be reduced, bringing down further the ratio of equation 6, and thus further decreasing i_s .

We now reintroduce the mossy fibers. Can an input of the MF type force efficient information storage? The indications are that it can, although it is difficult to derive a simple formal

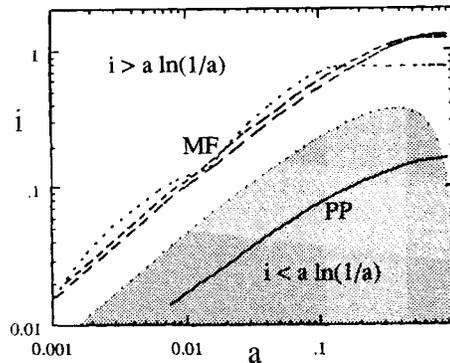


Fig. 3. Information encoded in the firing rate of a CA3 cell. As a function of the resulting sparseness a of the CA3 representation, the solid line is the amount of information per cell that can be stored of that carried by perforant path inputs, as given by the expression calculated in Appendix A, with $K = 0.25$ and $\psi = 0$; the dashed lines are the amounts that can be stored of the information carried by a binary distribution of activity on the mossy fibers, for $\lambda = 5.0$, $C^{MF} = 50$, and $a_{GC} = 0.1$ (long dash), $a_{GC} = 0.02$ (medium dash), and $a_{GC} = 0.004$ (short dash). The shaded area is where not enough information is encoded in the rate distribution, according to the requirement of equation 4. The requirement is clearly met by the distributions produced by the model mossy fiber inputs and not by those produced by the model perforant path inputs. Note the logarithmic scales on both axes of the graph.

argument, not only because the small number of such synapses *invalidates* the Gaussian approximation, but also because one needs to put forward more assumptions concerning, for example, the DG activity carried by the mossy fibers. A number of factors, however, may work in the right direction. The first, obvious, one is the presumed strength of MF synapses, which could result both from their size and from their proximity to the soma. The second factor is that the relatively small number of synapses allows a suitably *sparse* firing activity in the DG to be converted into a signal, h_i^{MF} , whose distribution over different cells may be highly structured (far from Gaussian). This has been referred to as an "orthogonalizing" effect, and has been proposed (Rolls, 1989a) as one function of DG processing.⁴ Finally, we propose that the observed plasticity of mossy fiber synapses, both in the form called post-tetanic potentiation and in that described as nonassociative long-term enhancement (Brown et al., 1989), might be useful because it would result in greatly increased efficacy whenever several spikes are carried by the *same* mossy fiber. In other words, a consistently firing mossy fiber would, in contrast to an occasionally firing one, produce nonlinearly amplified currents in the postsynaptic CA3 cell.

Merely as an illustration of the effects of the first and second factor mentioned above, and not as a statement about the nature of the mossy fiber input, consider the following example. Let the distribution of activity among dentate gran-

⁴ While this effect produces a sparser signal, thereby allowing more patterns to be stored but also reducing the information content of h_i^{MF} itself, it also enhances its variance with respect to ρ_{RC}^2 , pushing up i_s .

ule cells, during a learning event, be binary, with a fraction a_{GC} of the cells firing at rate η_{GC} , and the rest silent; and let all C^{MF} mossy fiber synapses to a given CA3 cell i have uniform efficacy J . Then the distribution of the input $\{h_i^{MF}\}$ is binomial, with average $C^{MF} a_{GC} J \eta_{GC}$ and variance $\rho_{MF}^2 = C^{MF} a_{GC} (1 - a_{GC}) J^2 \eta_{GC}^2$. Adding the Gaussian spread given by $\{h_i^{RC}\}$, assuming some value for $\lambda = \rho_{MF}^2 / \rho_{RC}^2$, and neglecting this time $\{h_i^{PP}\}$, one can again compute i_s and plot it versus the resulting a of the distribution. One sees from the top of Figure 3 that for $\lambda = 5$, $C^{MF} = 50$ and a broad range of values for a_{GC} , i_s is within the range implied by requirement 4, indicating that the MF inputs may produce in the CA3 cells a pattern of activity that contains sufficient information for learning. This is in contrast to the more extensive PP inputs, as discussed above, and shown in Figure 3.

While the latter is only an indication, the tentative conclusion of the argument as a whole is that an input system with the characteristics of the mossy fibers is essential during learning, in that it may act as an (unsupervised) teacher, establishing a neuronal representation of the episode, in the CA3 network, rich in information about the input itself.

Inputs mediated by associative synapses are necessary to initiate retrieval

Once a pattern μ has been stored, our hippocampal theory postulates that the CA3 network is able, stimulated with a partial cue, to retrieve the pattern. Very many other memories might have been stored in the meantime, inducing superimposed synaptic modifications that might have completely changed the distribution of synaptic efficacies; but retrieval can still occur, as formal models demonstrate, solely on the basis of the traces encoded while learning the pattern μ . The partial cue takes the form of an input, which could be carried in principle by PP or MF axonal activity, or both, weakly correlated with the input present during storage. The minimal correlation q in the distribution of firing activity on the afferent axons, between storage and retrieval, has to be assumed weak, $q \ll 1$: otherwise, there would be no point in having information stored in CA3, as a memory of the episode would clearly already exist elsewhere, earlier on. (Marr (1971) suggested $q \approx 0.3$ as a reasonable value.) The correlation in the input can be interpreted for example as corresponding to a memory for a *subset of elements* of the episode, serving as a key to access the whole stored event. The primary task of the synapses on the afferent fibers during retrieval is to take the axonal input and convert it into an integrated synaptic current, denoted here as h_i^{AFF} . In its distribution over different cells, h_i^{AFF} has to carry a *signal* specific enough to enable retrieval of pattern μ by the recurrent collaterals.

In formal network models that describe retrieval (Amit, 1989), h_i^{AFF} is often taken to determine the initial distribution of firing activities, from which the network then evolves driven by recurrent connections; or, if h_i^{AFF} is considered to persist during the time it takes to complete retrieval, it is also included as a fixed external contribution to the input current of each cell. In any case, biologically plausible analytical models describe, so far, only the "steady-state" distribution of activities emerging at retrieval, rather than the full dynamic evolution of activities leading to that retrieval state (Treves

and Rolls, 1991). The latter will require, in the future, more sophisticated mathematical analyses and numerical simulations, and a more complete appraisal of the processes involved. Therefore, it is difficult to estimate how specific the signal in h_i^{AFF} has to be in order to initiate retrieval consistently. It is, however, obvious that what is needed is for the signal pointing toward pattern μ to stand out clearly above any "spurious" correlations between h_i^{AFF} and other stored patterns. It is possible to put forward the following simple argument to show that this requirement strongly favors an input system in which very many axonal fibers contact each CA3 cell through associatively modifiable synapses.

The aim is to estimate the correlation, or signal, S between the distribution of the current h_i^{AFF} among CA3 cells, and that of the total synaptic current h_i^μ during learning of pattern μ . The expression obtained is then compared for the cases of associative and nonassociative synapses, to evaluate the role that the synapses may play in initiating retrieval. Denote with A_i the set of factors that contribute to determine h_i^{AFF} , and with B_i^μ those that determine h_i^μ . Even though some of these factors, as they will be considered later, are common to both sets A_i and B_i^μ , they can to a good approximation be taken as independent from cell to cell, because h_i^{AFF} does not reflect the activity of cells different from i , and because at the time h_i^μ was produced, during learning, different cells communicated only through effectively random synaptic efficacies. Therefore, one may consider a single CA3 cell, drop index i , and estimate the desired correlation as

$$S = \frac{\langle h^{AFF} h^\mu \rangle_{A,B} - \langle h^{AFF} \rangle_A \langle h^\mu \rangle_B}{\sqrt{(\langle h^{AFF} \rangle_A^2 - \langle h^{AFF} \rangle_A^2) \sqrt{(\langle h^\mu \rangle_B^2 - \langle h^\mu \rangle_B^2)}} \quad (10)$$

where $\langle \cdot \rangle_A$ denotes an average over the distribution of values taken by the factors included in A .

Labelling $1, \dots, j, \dots, C^{AFF}$ the axons afferent to the cell, their contribution to its synaptic current is written

$$h^{AFF} = \sum_j J_j V_j \quad (11)$$

with V_j the firing rate in axon j at retrieval and J_j the synaptic efficacy. If η_j^μ is the firing rate while learning pattern μ , assuming that each CA3 cell effectively samples the afferent pattern of activity means that even for one cell

$$\frac{\langle V \eta^\mu \rangle - \langle V \rangle \langle \eta^\mu \rangle}{\sqrt{\langle V^2 \rangle - \langle V \rangle^2} \sqrt{\langle (\eta^\mu)^2 \rangle - \langle \eta^\mu \rangle^2}} \approx q. \quad (12)$$

The efficacy J_j is written $J + J_j^c$ to include a baseline efficacy J , plus superimposed modifications induced by learning. If new patterns are being learned continuously, it is natural to assume that the modifications they individually induce decay with time. Considering a simple exponential decay (this is *not* an essential assumption), and labelling patterns $1, \dots, \lambda, \dots$ from the *last* one backward, one may attach to each modification a factor $\exp(-\lambda p)$, with p a measure of effectively how many traces are stored at the same time.

If each cell is viewed as a point-like element, the factors that determine modification at one synapse while learning pattern λ may be taken to include, for present purposes, the presynaptic firing rate η_j^λ and some time average of the postsynaptic depolarization. The latter is in a one-to-one corre-

spondence with the average synaptic current h^λ . An *associative* modification is then defined as one that depends multiplicatively on these two factors, i.e., of the form $F(h^\lambda)G(\eta_i^\lambda)$, with F and G some unspecified functions. A *nonassociative* modification, instead, may depend on one or even both factors, but in a purely additive form, $F(h^\lambda) + G(\eta_i^\lambda)$.

For simplicity, the associative synapses considered here will be those whose efficacies are expressed by learning rules of the type

$$J_j^c = \epsilon \sum_\lambda e^{-\lambda/p} F(h^\lambda)(\eta_j^\lambda - \bar{\eta}) \quad (13)$$

where the form of the presynaptic term ensures that the contribution of each learned pattern on average vanishes, enhancing the storage capacity (Rolls and Treves, 1990; Willshaw and Dayan, 1990). By contrast, nonassociative synapses will be considered, in line with some current hypotheses on the mossy-fiber type of plasticity (Brown et al., 1989; Johnston et al., 1989), to undergo efficacy modifications dependent only on postsynaptic conditions,

$$J_j^c = \epsilon \sum_\lambda e^{-\lambda/p} F(h^\lambda). \quad (14)$$

In both cases, $F(h)$ is assumed of order 1, and the "learning rate" prefactor ϵ , which fixes the correct dimensions, will be taken such as to enhance the specificity of the signal, within the constraint that the total synaptic efficacy stays positive (what this requirement implies in each case is clarified in Appendix B). Note that one may consider also nonassociative synapses in which efficacy changes depend on just the presynaptic rate (Zalutsky and Nicoll, 1990),

$$J_j^c = \epsilon \sum_\lambda e^{-\lambda/p} G(\eta_j^\lambda), \quad (15)$$

but those, it is easy to realize, produce no specific signal at all, in the limit in which C^{AFF} is large enough to allow, as implied by equation 12, effective sampling of the afferent pattern by each cell.

The synaptic current h^μ is determined only in part by the afferent system under consideration, the remaining component being due to other classes of synapses; it is written

$$h^\mu = h^{\mu,0} + \sum_j J_j^{(\mu)} \eta_j^\mu \quad (16)$$

where $J_j^{(\mu)}$, the efficacy at the time of learning pattern μ , includes only the modifications produced by previously learned patterns, $\lambda \geq \mu + 1$.

A simple statistical analysis, sketched in Appendix B, allows estimation of the size of the specific signal that systems of synapses of the two types are able to produce:

$$S_{ASS} \sim \frac{\sqrt{C^{AFF}} e^{-\mu/p}}{\sqrt{p}} q \quad S_{NONASS} \sim \frac{q}{\sqrt{C^{AFF}}} \quad (17)$$

Synapses of type 13 can produce a signal of essentially the same order of the cue, provided that C^{AFF} is itself of the same order of p , and storage did not occur too long ago (the trace decay factor $e^{-\mu/p}$). With synapses of type 14, instead, the cue is transformed into a much smaller signal, inversely proportional to the square root of the number of synapses in-

involved—accurate sampling of the input is actually counterproductive in this case, just as with synapses of type 15.

Associative synapses appear thus to be essential in order to relay a sizable signal, and therefore initiate retrieval; moreover, the signal grows proportionally to the square root of the number of such synapses.

If the total number of associatively modifiable synapses on a given cell is constrained, for example by the necessity to keep the cell electrotonically short, one may envisage a sort of competition between allocating synapses to the input system initiating retrieval and to the recurrent collaterals affecting it. Increasing the number of synapses on the former allows the same signal to be produced by smaller and smaller cues, while the capacity for retrieval is proportional to the number of recurrent collateral synapses. It may be possible to try to understand the factor K in these terms.

DISCUSSION

The arguments we have presented indicate that, on computational grounds, it is advantageous for a recurrent collateral autoassociative memory to have two input systems available. One, mediated by very strong but not necessarily plastic synapses, has to drive information storage, while the second, mediated by an extensive number of associatively modifiable synapses, conveys the signals initiating memory retrieval. In CA3, it is tempting to identify the first system with the mossy fibers and the second with the perforant path. When considering the validity of both the arguments and the identifications, it is important to be clear about their implications.

One implication is that during retrieval, or at least while initiating retrieval, the mossy fiber input should be absent or strongly reduced with respect to its amplitude during learning, in order not to blur the signal relayed by the perforant path. The "opposite" is not true: the perforant path input does not have to be inactive during storage. In fact, the above hypothesis *requires* it to be active, to enable associative modifications to take place at its synapses.

Another implication is that CA3 cells should be compact enough to allow synaptic modification on the distal dendrites to reflect depolarizations produced close to the soma by the mossy synapses. Conversely, active dendritic processes are useful, in this picture, only insofar as they balance the attenuation associated with passive cable conduction. No specific function is ascribed here to highly local processing at selected dendritic sites.

A third implication is that the time scale for information storage, i.e., for the depolarizations produced by mossy fiber inputs, be consistent with the time scale for associative plasticity at recurrent collateral and perforant path synapses. These two time scales are, in principle, independent from a third time scale, that for retrieval, which itself has to be such as to allow sufficient circulation of activity through the collaterals—a requirement that has to be explored by studying the dynamics of retrieval, but which is likely to be strongly dependent on the size of the available cue.

An appealing feature of the proposed scheme is its falsifiability. Clearly, the above arguments would immediately be shown to be wrong or irrelevant if it were found that the synapses on the perforant path to CA3 are not modifiable in an associative manner. The whole autoassociative memory

theory would be invalidated if the synapses on the collaterals were found not to be associative.

Moreover, to the extent that a degree of specialization, in the sense indicated above, does exist between the two input systems, our arguments predict selective behavioral impairments to result from temporary inactivation of the two pathways. Thus, temporary inactivation of the mossy fibers could later result, barring generic dysfunctions, in amnesia specific to events stored during the time window of inactivation. This suggests an experimental testing procedure that might be feasible in the immediate future (R. G. M. Morris, personal communication, 1991). Similar effects should be produced by temporary selective blocking of associative plasticity at perforant path synapses, or, in fact, at recurrent collateral synapses. Alongside this later impairment, inactivation of the terminal portion (following DG) of the perforant path, if it were to be technically feasible, might be also expressed *immediately* as an impairment in the performance of certain computationally demanding memory tasks of the type for which the hippocampus is required (Rolls, 1990; 1991), due to a failure at retrieving previously stored information.

The roles we propose for the input systems to CA3 are consistent with the general hypothesis (Rolls, 1989a) that one of the functions of the processing occurring in the dentate gyrus is to transform certain neuronal representations into convenient inputs for the autoassociative memory assumed to be implemented in the CA3 network. They are also consistent with experimental data concerning the effects of destroying dentate granule cells (McNaughton et al., 1989).

The scheme presented here is neutral, it should be noted, with respect to a number of hypotheses that have been put forward as part of various hippocampal theories. These concern, for example, the nature of the information processed in the hippocampus as a whole (Kesner, 1990; Rolls, 1990), and the question of whether it is predominantly spatial or contextual (O'Keefe and Nadel, 1978), related to working (Olton et al., 1979) or personal (Gaffan, 1987) memory, to object-place associations (Parkinson et al., 1988), relational (Eichenbaum and Cohen, 1988), more generally associatively supramodal (McNaughton and Nadel, 1990), or even representing an "index" of neocortical locations** (Teyler and DiScenna, 1986). The scheme is also essentially neutral with respect to the detailed temporal features of hippocampal activity during different behaviors, and the hypotheses they have suggested, based in particular on the study of those species that display Θ -rhythmicity (Buzsáki, 1989). It is neutral, moreover, with respect to ideas concerning the role of CA1 in the memory processes (Levy, 1989) and the important question of the return projections to neocortex (Rolls, 1989a; 1989b).

On the other hand, what has been presented here is a fur-

ther development of the hypothesis that one of the functions of CA3 is that of an autoassociative memory. An autoassociative memory implemented in CA3 would be useful within the context of a number of different theories of hippocampal function. Whatever the overall function of the hippocampus, we believe that such detailed quantitative analyses of how parts of it might operate are important complements to empirical investigations.

ACKNOWLEDGMENTS

This work was supported in part by grants from the Medical Research Council (PG 8513790) and from the EEC BRAIN Initiative (88300446/JU1). The authors acknowledge many discussions with D. Amaral, of the Salk Institute, on his quantitative neuroanatomical studies. The assistance of R. D. W. Payne and useful discussions with D. J. Amit, R. G. M. Morris, H. Sompolinsky, M. Tsodyks, and D. Willshaw are also gratefully acknowledged.

References

- Amaral, D. G., R. Insausti, and W. M. Cowan (1984) The commissural connection of the monkey hippocampal formation. *J. Comp. Neurol.* 224:307-336.
- Amaral, D. G., N. Ishizuka, and B. Claiborne (1990) Neurons, numbers and the hippocampal network. *Prog. Brain Res.* 83:1-11.
- Amit, D. J. (1989) *Modelling Brain Function*, Cambridge University Press, New York.
- Amit, D. J., H. Gutfreund, and H. Sompolinsky (1987) Statistical mechanics of neural networks near saturation. *Ann. Phys.* 173:30-67.
- Barlow, H. B. (1960) The coding of sensory messages. In *Current Problems in Animal Behaviour*, W. H. Thorpe and O. L. Zangwill, eds. pp. 331-360, Cambridge University Press, Cambridge.
- Boss, B. D., G. M. Peterson, and W. M. Cowan (1985) On the number of neurons in the dentate gyrus of the rat. *Brain Res.* 338:144-150.
- Boss, B. D., K. Turlejski, B. B. Sanfield, and W. M. Cowan (1987) On the number of neurons in fields CA1 and CA3 of the hippocampus of Sprague-Dawley and Wistar rat. *Brain Res.* 406:280-287.
- Brown, T. H., A. H. Ganong, E. W. Kairiss, C. L. Keenan, and S. R. Kelso (1989) Long-term potentiation in two synaptic systems of the hippocampal brain slice. In *Neural Models of Plasticity*, J. H. Byrne and W. O. Berry, eds., pp. 266-306, Academic Press, San Diego, CA.
- Brown, T. H., E. W. Kairiss, and C. L. Keenan (1990) Hebbian synapses: biophysical mechanisms and algorithms. *Ann. Rev. Neurosci.* 13:475-511.
- Buzsáki, G. (1989) Two-stage model of memory trace formation: A role for "noisy" brain states. *Neuroscience* 31:551-570.
- Collingridge, G. L., and W. Singer (1990) Excitatory amino acid receptors and synaptic plasticity. *Trends. Pharm. Sci.* 11:290-296.
- Eichenbaum, H., and N. J. Cohen (1988) Representation in the hippocampus: What do hippocampal neurons code? *Trends Neurosci.* 11:244-248.
- Gaffan, D. (1987) Amnesia, personal memory and the hippocampus: Experimental neuropsychological studies in monkeys. In *Cognitive Neurochemistry*, S. M. Stahl, S. D. Iversen, and E. C. Goodman, eds., pp. 46-56, Oxford University Press, Oxford.
- Gardner-Medwin, A. R. (1976) The recall of events through the learning of associations between their parts. *Proc. Roy. Soc. Lond.* B194:375-402.
- Hopfield, J. J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* 79:2554-2558.

** It should be noted, though, that a very reductive interpretation of the "index" hypothesis views the retrieval operation considered in this paper solely as a way to identify one pattern among those stored in CA3, and to use it as a pointer to a memory stored elsewhere. In that case, the CA3 patterns, which would not be true memories, would not need to contain more than the small amount of information $i_s = (\ln_2 p)/N$ necessary for identification, and our arguments based on information would not be valid.

Ishizuka, N., J. Weber, and D. G. Amaral (1990) Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *J. Comp. Neurol.* 295:580–623.

Johnston, D., W. F. Hopkins, and R. Gray (1989) The role of norepinephrine in long-term potentiation at mossy-fiber synapses in the hippocampus. In *Neural Models of Plasticity*, J. H. Byrne and W. O. Berry, eds., pp. 307–328, Academic Press, San Diego, CA.

Kesner, R. P. (1990) Learning and memory in rats with an emphasis on the role of the hippocampal formation. In *Neurobiology of Comparative Cognition*, D. S. Olton and R. P. Kesner, eds., pp. 179–204, Erlbaum, Hillsdale, NJ.

Levy, W. B. (1989) A computational approach to hippocampal function. In *Computational Models of Learning in Simple Neural Systems*, R. D. Hawkins and G. H. Bower, eds., pp. 243–305, Academic Press, Orlando, FL.

Marr, D. (1971) Simple memory: A theory for archicortex. *Phil. Trans. Roy. Soc. Lond.* B262:24–81.

McNaughton, B. L., and L. Nadel (1990) Hebb-Marr networks and the neurobiological representation of action in space. In *Neuroscience and Connectionist Theory*, M. A. Gluck and D. E. Rumelhart, eds., pp. 1–63, Erlbaum, Hillsdale, NJ.

McNaughton, B. L., C. A. Barnes, J. Meltzer, and R. J. Sutherland (1989) Hippocampal granule cells are necessary for normal spatial learning but not for spatially-selective pyramidal cell discharge. *Exp. Brain Res.* 76:485–496.

Miles, R. (1988) Plasticity of recurrent excitatory synapses between CA3 hippocampal pyramidal cells. *Soc. Neurosci. Abstr.* 14:19.

Morris, R. G. M., E. R. Kandel, and L. R. Squire (1988) Learning and Memory (special issue editors). *Trends Neurosci.* 11(4).

Nadal, J. P. (1991) Associative memory: On the (puzzling) sparse coding limit. *J. Phys.* A24:1093–1101.

O'Keefe, J., and L. Nadel (1978) *The Hippocampus as a Cognitive Map*, Clarendon, Oxford.

Olton, D. S., J. T. Becker, and G. H. Handelman (1979) Hippocampus, space and memory. *Behav. Brain Sci.* 2:313–365.

Parkinson, J. K., E. A. Murray, and M. Mishkin (1988) A selective memory role for the hippocampus in monkeys: Memory for the location of objects. *J. Neurosci.* 8:4159–4167.

Rolls, E. T. (1987) Information representation, processing and storage in the brain: Analysis at the single neuron level. In *The Neural and Molecular Bases of Learning*, J. P. Changeaux and M. Konishi, eds., pp. 503–540, Wiley, New York.

Rolls, E. T. (1989a) Functions of neuronal networks in the hippocampus and neocortex in memory. In *Neural Models of Plasticity*, J. H. Byrne and W. O. Berry, eds., pp. 240–265, Academic Press, San Diego, CA.

Rolls, E. T. (1989b) The representation and storage of information in neuronal networks in the primate cerebral cortex and hippocampus. In *The Computing Neuron*, R. Durbin, C. Miall, and G. Mitchison, eds., pp. 125–159, Addison-Wesley, Wokingham.

Rolls, E. T. (1990) Functions of the primate hippocampus in spatial processing and memory. In *Neurobiology of Comparative Cognition*, D. S. Olton and R. P. Kesner, eds., pp. 339–362, Erlbaum, Hillsdale, NJ.

Rolls, E. T. (1991) Functions of the primate hippocampus in spatial and nonspatial memory. *Hippocampus* 1:258–261.

Rolls, E. T., and A. Treves (1990) The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network* 1:407–421.

Seress, L. (1988) Interspecies comparison of the hippocampal formation shows increased emphasis on the regio superior in the Ammon's horn of the human brain. *J. Hirnforsch.* 29:335–340.

Shannon, C. E., and W. Weaver (1949) *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL.

Squire, L. R., M. Mishkin, and A. P. Shimamura (1990) (eds.) *Learn-*

ing and memory. In *Discussions in Neuroscience*, Elsevier, Amsterdam.

Squire, L. R., A. P. Shimamura, and D. Amaral (1989) Memory and the hippocampus. In *Neural Models of Plasticity*, J. H. Byrne and W. O. Berry, eds., pp. 208–239, Academic Press, San Diego, CA.

Storm-Mathisen, J., J. Zimmer, and O. P. Ottersen (eds.) (1990) *Understanding the brain through the hippocampus*. *Prog. Brain Res.* 83.

Teyler, T. J., and P. DiScenna (1986) The hippocampal memory indexing theory. *Behav. Neurosci.* 100:147–154.

Treves, A. (1990) Graded-response neurons and information encodings in autoassociative memories. *Phys. Rev.* A42:2418–2430.

Treves, A. (1991) Are spin-glass effects relevant to understanding realistic autoassociative networks? *J. Phys.* A24:2645–2654.

Treves, A., and E. T. Rolls (1991) What determines the capacity of autoassociative memories in the brain? *Network*, 2:371–397.

West, M. J., and H. J. G. Gundersen (1990) Unbiased stereological estimation of the number of neurons in the human hippocampus. *J. Comp. Neurol.* 296:1–22.

Willshaw, D. J., and J. T. Buckingham (1990) An assessment of Marr's theory of the hippocampus as a temporary memory store. *Phil. Trans. R. Soc. Lond.* 329:205–215.

Willshaw, D. J., and P. Dayan (1990) Optimal plasticity from matrix memories: What goes up must come down. *Neural. Comp.* 2:85–93.

Yeckel, M. F., and T. W. Berger (1990) Feedforward excitation of the hippocampus by afferents from the entorhinal cortex: Redefinition of the role of the trisynaptic pathway. *Proc. Natl. Aca. Sci. USA* 87:5832–5836.

Zalutsky, R. A., and R. A. Nicoll (1990) Comparison of two forms of long-term potentiation in single hippocampal neurons. *Science* 248:1619–1624.

Zola-Morgan, S. M., and L. R. Squire (1990) The primate hippocampal formation: Evidence for a time-limited role in memory storage. *Science* 250:288–290.

APPENDIX A

In a threshold-linear unit the output firing rate is determined as

$$\eta = \begin{cases} g(h - T_{hr}) & \text{for } h > T_{hr} \\ 0 & \text{for } h \leq T_{hr}, \end{cases} \quad (18)$$

with g a gain parameter, possibly dependent on the overall network activity.

From their definitions, one has the probability densities

$$c^-(h^{PP}) = \frac{\sigma(\Delta h^{PP}/\rho_{PP})}{\rho_{PP}},$$

$$c(h^{PP}, \eta) = c^-(h^{PP}) \left[\frac{\sigma(z(h^{PP}, \eta))}{g\rho_{RC}} + \phi(z(h^{PP}, \eta))\delta(\eta) \right],$$

$$c_-(\eta) = \int dh^{PP} c(h^{PP}, \eta) \quad (19)$$

where

$$z(h^{PP}, \eta) = \frac{\eta}{g\rho_{RC}} - \frac{\bar{h} + \Delta h^{PP} - T_{hr}}{\rho_{RC}} \quad (20)$$

and

$$\sigma(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad \phi(x) = \int_{-\infty}^x \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy, \quad (21)$$

(and $\delta(x)$ is Dirac's δ -function).

Then, defining $r = (\bar{h} - T_{hr})/(\rho_{RC}^2 + \rho_{PP}^2)^{1/2}$, and carrying out the integrals in Eq. 8, one finds

$$i_s = \frac{1}{2}\phi(r)\ln_2[1 + K(1 - \psi)^2] - \frac{1}{2\ln 2}r\sigma(r)\frac{K(1 - \psi)^2}{1 + K(1 - \psi)^2} + \int_{-\infty}^{+\infty} ds\sigma(s)\phi(-\tilde{r}(s))\ln_2\phi(-\tilde{r}(s)) - \phi(-r)\ln_2\phi(-r), \quad (22)$$

where $\tilde{r}(s) \equiv r\sqrt{1 + K(1 - \psi)^2} + s\sqrt{K(1 - \psi)^2}$.

To compute a , one notes that

$$\langle \eta \rangle = g(\rho_{RC}^2 + \rho_{PP}^2)^{1/2}[r\phi(r) + \sigma(r)] \quad (23)$$

$$\langle \eta^2 \rangle = g^2(\rho_{RC}^2 + \rho_{PP}^2)[(1 + r^2)\phi(r) + r\sigma(r)].$$

A similar treatment can be applied to the example involving h_i^{MF} .

APPENDIX B

Consider the signal pointing to pattern μ , provided by a system of synapses of the type of eq. 13. Among the factors included in set A, that determine h^{AFF} , those whose distribution is correlated with that of h^μ include V_j and h^μ itself, which enters in the corresponding modification term. Moreover, h^μ depends on $J_j^{(\mu)}$, which includes terms surviving in the efficacy value J_j determining, later, h^{AFF} . Using the linearity of eqs. 13, 16 one estimates the numerator of S as

$$\langle h^{AFF}h^\mu \rangle_{A,B} - \langle h^{AFF} \rangle_A \langle h^\mu \rangle_B$$

$$\approx \langle \sum_k \hat{J}_k V_k \sum_j J_j^{(\mu)} \eta_j^\mu \rangle_{A,B} - \langle \sum_k \hat{J}_k V_k \rangle_A \langle \sum_j J_j^{(\mu)} \eta_j^\mu \rangle_B$$

$$+ \epsilon e^{-\mu p} \left\{ \langle F(h^\mu) \left[\sum_k (\eta_k^\mu - \bar{\eta}) V_k \right] h^\mu \rangle_{A,B} \right.$$

$$\left. - \langle F(h^\mu) \left[\sum_k (\eta_k^\mu - \bar{\eta}) V_k \right] \rangle_A \langle h^\mu \rangle_B \right\} \quad (24)$$

where \hat{J} does not include the contribution of pattern μ , and minor sources of correlation have been neglected. As different axons carry independent information, the first line in the r.h.s. contributes

$$\sum_j \left\{ \langle \hat{J}_j J_j^{(\mu)} \rangle \langle V_j \eta_j^\mu \rangle - \langle \hat{J}_j \rangle \langle J_j^{(\mu)} \rangle \langle V_j \rangle \langle \eta_j^\mu \rangle \right\} = C^{AFF}$$

$$\times \left[J^2 (\langle V \eta^\mu \rangle - \bar{V} \bar{\eta}) + \epsilon^2 e^{-\mu p} \frac{p}{2} \langle F(h)^2 \rangle (\bar{\eta}^2 - \bar{\eta}^2) \langle V \eta^\mu \rangle \right], \quad (25)$$

while the second line yields

$$C^{AFF} \epsilon e^{-\mu p} [\langle hF(h) \rangle - \bar{h} \langle F(h) \rangle] (\langle V \eta^\mu \rangle - \bar{V} \bar{\eta}). \quad (26)$$

An exam of expressions (25–26) shows that the general requirement on ϵ implies in this case that the size of the modifications be such as to fully exploit the available synaptic range J . This results in an estimate for ϵ :

$$\epsilon^2 \frac{p}{2} \langle F(h)^2 \rangle (\bar{\eta}^2 - \bar{\eta}^2) \sim J^2; \quad (27)$$

while \bar{h} , which includes other currents, has to be at least of order

$$\bar{h} \sim C^{AFF} J \bar{\eta}. \quad (28)$$

Using these two estimates, one can see that the term (26) dominates by a factor of order C^{AFF}/\sqrt{p} the other ones (25), which are therefore discarded.

Similarly, one can evaluate the quantity appearing in the denominator

$$\langle (h^{AFF})^2 \rangle_A - \langle h^{AFF} \rangle_A^2$$

$$= \sum_{j,k} \left\{ \left\langle \left[\hat{J}_j V_j + \epsilon e^{-\mu p} F(h^\mu) (\eta_j^\mu - \bar{\eta}) V_j \right] \left[\hat{J}_k V_k + \epsilon e^{-\mu p} F(h^\mu) (\eta_k^\mu - \bar{\eta}) V_k \right] \right\rangle_A - \langle \hat{J}_j V_j + \epsilon e^{-\mu p} F(h^\mu) (\eta_j^\mu - \bar{\eta}) V_j \rangle_A \langle \hat{J}_k V_k + \epsilon e^{-\mu p} F(h^\mu) (\eta_k^\mu - \bar{\eta}) V_k \rangle_A \right\} \quad (29)$$

and find that the dominant terms when $j = k$ are, now, those involving the products of \hat{J}^{c^2} 's, but there are also nonnegligible terms with $j \neq k$, to yield

$$C^{AFF} \left[J^2 (\bar{V}^2 - \bar{V}^2) + \epsilon^2 \frac{p}{2} \langle F(h)^2 \rangle (\bar{\eta}^2 - \bar{\eta}^2) \bar{V}^2 \right]$$

$$+ C^{AFF} (C^{AFF} - 1) \epsilon^2 e^{-2\mu p} \quad (30)$$

$$\times [\langle F(h)^2 \rangle - \langle F(h) \rangle^2] (\langle V \eta^\mu \rangle - \bar{V} \bar{\eta})^2.$$

Putting pieces together, and considering q to be small enough that the first line of (30) dominates over the second, one finds that the signal mediated by associative synapses is, discarding factors of $O(1)$,

$$S_{ASS} \sim \frac{\sqrt{C^{AFF}} e^{-\mu p}}{\sqrt{p}} q. \quad (31)$$

When this procedure is repeated for nonassociative efficacies of the type of eq. 14, the term corresponding to the second line of eq. 24 is now independent of q , i.e. it is the same if pattern μ is substituted with some other pattern uncorrelated with $\{V_j\}$ (this is precisely the result of nonassociativity, of course). The first line of eq. 24 does, however, produce terms proportional to q , as well as others proportional only to $\bar{V} \bar{\eta}$. Ensuring that the former terms, carrying the specificity to μ , dominate in size (a) the others from the same line and (b) that from the second, can be translated into requirements for ϵ :

$$C^{AFF} (J + \epsilon p \langle F(h) \rangle)^2 \quad (32)$$

$$\geq \begin{cases} (C^{AFF} \epsilon)^2 p (\langle F(h)^2 \rangle - \langle F(h) \rangle^2) & (a) \\ C^{AFF} \epsilon (\langle hF(h) \rangle - \bar{h} \langle F(h) \rangle) / \bar{\eta} & (b) \end{cases}$$

which are both satisfied if ϵ is sufficiently small (for fixed J). Using then the estimate, equivalent to (28),

$$\bar{h} \sim C^{AFF} (J + \epsilon p \langle F(h) \rangle) \bar{\eta} \quad (33)$$

one may evaluate the maximum signal that can be obtained in this case as

$$S_{NONASS} \quad (34)$$

$$\approx \frac{C^{AFF} (J + \epsilon p \langle F(h) \rangle)^2 (\langle V \eta^\mu \rangle - \bar{V} \bar{\eta})}{(\bar{h}^2 - \bar{h}^2)^{1/2} [C^{AFF} (J + \epsilon p \langle F(h) \rangle)^2 (\bar{V}^2 - \bar{V}^2)]^{1/2}} \sim \frac{q}{\sqrt{C^{AFF}}}$$

discarding again factors of $O(1)$.