

Quantitative Estimate of the Information Relayed by the Schaffer Collaterals

ALESSANDRO TREVES

Biophysics and Cognitive Neuroscience, SISSA, via Beirut 2-4, 34013 Trieste, Italy
ale@limbo.sissa.it

Received April 3, 1994; Revised May 2, 1995; Accepted May 2, 1995

Action Editor: J. Miller

Abstract. Within the theory that describes the hippocampus as a device for the on-line storage of complex memories, the crucial autoassociative operations are ascribed mainly to the recurrent CA3 network. The CA3-to-CA1 connections may still be important, both in completing information retrieval and in re-expanding, with minimal information loss, the highly compressed representation retrieved in CA3. To quantify these effects, I have defined a suitably realistic formal model of the relevant circuitry, and evaluated its performance in the sense of information theory. Analytical estimates, calculated with mean-field, replica and saddle-point techniques, of the amount of information present in the model CA1 output, reveal how such performance depends on different parameters characterising these connections. In particular, nearly all the stored information can be preserved if the model Schaffer collaterals are endowed with an optimal degree of Hebbian plasticity, matching that of the CA3 recurrent collaterals.

Keywords: associative memory, hippocampus, CA1 field, information theory

Introduction

Several investigators have offered their own perspective on the idea that the crucial contribution of the hippocampus is in memory (Marr, 1971; Eichenbaum et al., 1992; Gaffan, 1992; Kesner and Dakis, 1992; McClelland et al., 1992; Squire, 1992)—in contrast to others who have emphasised a putative role in spatial computations (O'Keefe, 1990; Zipser, 1985; but see Treves et al., 1992). In particular, at least one version of the memory theory holds that the hippocampus acts as a sort of intermediate, *buffer* store, where certain types of memories (characterized in different species as mainly spatial, configural, relational, declarative) are initially formed, *on-line*, and from where they can be, later, retrieved, and used for recall as well as for consolidation of neocortical long-term memory.

From an abstract, information-theoretical point of view, one can point at the most stringent requirements that the biological machinery performing such a memory task would have to satisfy:

1. generate, on-line, new neuronal representations for memories that can be very complex (i.e., contain

- very many bits of specific information)

2. store these representations (again, on-line), even on the basis of a single exposure
3. hold very many different representations simultaneously in storage
4. ensure that each one can be retrieved by a partial, occasionally very small, cue
5. make the full retrieved information available back in the neocortex

Most of these requirements—nos. 2, 3 and 4—can be met by an autoassociative memory network; and Rolls (1987, 1989) has suggested that the CA3 region, with its extensive network of recurrent collaterals, would operate precisely as an autoassociator, utilizing the cooperative mechanisms illustrated by several theoretical models (Hopfield, 1982; Kohonen, 1984). This hypothesis obviously implies that the synapses on the recurrent collaterals be endowed with some kind of Hebbian plasticity.

We have been investigating the computational requirements imposed by the role assigned to CA3 in quantitative detail, using the mathematical analysis of formal network models that incorporate what we believe are necessary elements of biological realism

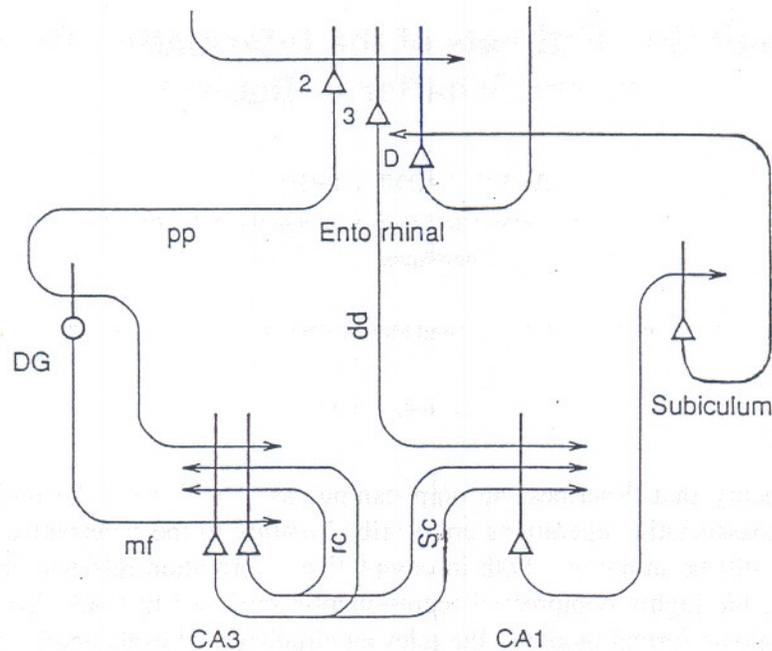


Fig. 1. Schematic overview of hippocampal circuitry, including entorhinal cortex, the dentate gyrus (DG), the CA3 and CA1 regions of the hippocampus proper, and the subiculum. CA3 receives the perforant path (pp) connections from entorhinal cortex, the mossy fibers (mf) projections from dentate granule cells, as well as recurrent collaterals (rc) from CA3 itself. CA1 also receives direct inputs from entorhinal, alongside the Schaffer collaterals from CA3. Cells projecting efferents from entorhinal cortex are from largely non-overlapping layers, as indicated (D stands for deep).

(Treves and Rolls, 1991). This has led us to suggest specific roles for additional components of hippocampal circuitry (Treves and Rolls, 1992), with arguments based on a purely computational quantitative analysis. Thus, the perforant path connections to CA3 should help satisfy requirement 4, by relaying undegraded to the CA3 cells the cue that initiates retrieval of a given representation. The mossy fibers instead, and with them the whole dentate network, might be regarded as a device for meeting requirement 1, by setting up new information-rich representation in CA3 with minimal interference from previously stored memories.

We have also put forward hypotheses concerning the operation of networks downstream in the hippocampal system, including the backprojections to neocortex (Treves and Rolls, 1994). Here, the analysis will focus on the first component to follow the putative CA3 autoassociator: the Schaffer collateral connections to CA1. One may wonder what the role of this additional stage of hippocampal processing might be, if indeed the CA3 network accomplishes most of what is required of the hippocampus as a whole. In fact, why should any signal be further transduced through this other network, where it might be expected to only degrade (lose information) or, at best, pass through unaffected? True, there is a direct perforant path projection to CA1 that

is likely to be there for some good reason, but is it only by considering this direct projection that we shall be able to understand what CA1 might actually contribute?

We have proposed (Treves and Rolls, 1994) that the Schaffer collaterals may serve two purposes, independently of the fact that CA1 cells eventually integrate their inputs with those carried by the perforant path. First, the Schaffer collaterals implement a heteroassociator, that helps to complete the retrieval of a stored representation, to the extent that this task was left unfinished by the CA3 autoassociator. Thus, they should be expected to contribute, in certain conditions, a net information gain rather than an information loss. Second, they are also the first link in the chain back to neocortex, which has to make sure that the information retrieved from the buffer memory is not lost on the way (what we denoted as requirement 5). This simple requirement favours expanding the CA3 representation onto a larger number of cells (as present in CA1), so that each cell has to code for less, and its message is therefore more robust to noise and degradation. It is clear that the expansion is useful only if it can be accomplished while preserving the overall quality of the signal relayed (i.e. its information content distributed over the network).

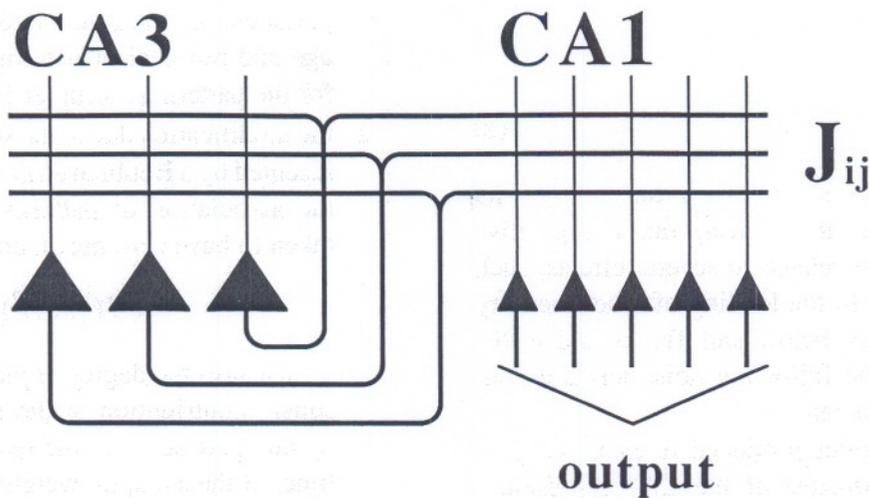


Fig. 2. The components included in the formal model: inputs are provided by the CA3 region, where the recurrent collateral network is considered to store a firing pattern $\{\eta_i\}$ and retrieve an approximate version of it, $\{V_i\}$; these patterns are multiplied by the matrix of synaptic weights J_{ij} on the Schaffer collaterals, added noise and thresholded, resulting in the CA1 firing patterns $\{\zeta_j\}$ (during storage) and $\{U_j\}$ (during retrieval). The aim is to analyse how the organization and plasticity of the Schaffer connections affect the information $\{U_j\}$ carries about $\{\eta_i\}$.

As a way of elucidating these constraints in a quantitative form, I propose to consider, as a basic index of performance, the amount of information present in the activity of CA1 cells. The calculation estimating the information in the CA1 representation is what is reported in this paper.

Methods

Model

The information content of a 'CA1' firing pattern has been evaluated analytically using a formal model. The model describes, in simplified form, only the Schaffer collateral connections from the N pyramidal cells of CA3 to the M pyramidal cells of CA1, as in the scheme of Fig. 2. It considers the effect of inhibitory interneurons only insofar as they exert a general regulation of pyramidal cells activity, and it neglects altogether the weak (Amaral and Witter, 1989) CA1 recurrent collateral system as well as other minor connections. Moreover, it does not include the direct perforant path to CA1, in order to dissociate its contribution from the effects that, as discussed in the Introduction, are to be quantified here. Two distinct modes of operation of the system are envisioned: storage and retrieval. During storage the synaptic efficacies on the Schaffer collaterals are modified in a Hebbian way reflecting the conjunction of pre- and post-synaptic activity—but the modification is not immediate and thus does not influence the current CA1 output. During retrieval, the

Schaffer collateral relay a pattern of activity retrieved from CA3, and their synaptic efficacies, while not being presently modified, reflect all previous storage events. Four different patterns of neuronal firing are then considered (Table 1)

- $\{\eta_i\}$ are the firing rates of each cell i of CA3, which together code for the information to be stored, for one particular memory, in the CA3 network. Statistically, the probability density of finding a given firing pattern is taken to be a product, for each cell, of a certain firing rate distribution

$$P(\{\eta_i\}) \prod_i d\eta_i = \prod_i P_{\eta}(\eta_i) d\eta_i \quad (1)$$

This assumption means that each cell in CA3 is taken to code for independent information, an idealized version of the idea that by this stage most of the redundancy present in earlier representations has been removed.

- $\{V_i\}$ are the firing rates in the pattern retrieved from CA3, and they are taken to reproduce the $\{\eta_i\}$ with some Gaussian distortion (noise), followed by

Table 1. The four different firing patterns appearing in the analysis. Each symbol denotes the firing rate of the cell indexed by the subscript.

| | CA3 | CA1 |
|-----------|--------------|---------------|
| Storage | $\{\eta_i\}$ | $\{\zeta_j\}$ |
| Retrieval | $\{V_i\}$ | $\{U_j\}$ |

rectification

$$V_i = [\eta_i + \delta_i]^+ \\ \langle (\delta_i)^2 \rangle = \sigma_\delta^2 \quad (2)$$

(the rectifying function $[x]^+ = x$ for $x > 0$, and 0 otherwise, ensures that a firing rate is a positive quantity). σ_δ can be related to several effects, such as interference due to the loading of other memory patterns in CA3 (see below and Treves and Rolls, 1991). This and the following noise terms are all taken to have zero means.

- $\{\zeta_j\}$ are the firing rates produced in each cell j of CA1, during the storage of the CA3 representation; they are determined by the matrix multiplication of the pattern $\{\eta_i\}$ with the synaptic weights J_{ij} —of zero mean, as explained below, and variance σ_j^2 —followed by Gaussian distortion, (inhibition-dependent) thresholding and rectification

$$\zeta_j = \left[\zeta_0 + \sum_i c_{ij} J_{ij}^S \eta_i + \epsilon_j^S \right]^+ \\ \langle (\epsilon_j^S)^2 \rangle = \sigma_{\epsilon^S}^2 \\ \langle (J_{ij}^S)^2 \rangle = \sigma_j^2 \quad (3)$$

The synaptic matrix is very sparse as each CA1 cell receives inputs from only C_j (say, 10^4) cells in CA3. The average of C_j across cells is denoted as C

$$c_{ij} \in \{0, 1\} \\ \langle c_{ij} \rangle N = C_j \quad (C \equiv \langle C_j \rangle) \quad (4)$$

- $\{U_j\}$ are the firing rates produced in CA1 by the pattern $\{V_i\}$ retrieved in CA3

$$U_j = \left[U_0 + \sum_i c_{ij} J_{ij}^R V_i + \epsilon_j^R \right]^+ \\ \langle (\epsilon_j^R)^2 \rangle = \sigma_{\epsilon^R}^2 \\ \langle (J_{ij}^R)^2 \rangle = \sigma_j^2 \quad (5)$$

Each weight of the synaptic matrix during retrieval of a specific pattern,

$$J_{ij}^R = \cos(\theta_\mu) J_{ij}^S + \gamma^{1/2}(\theta_\mu) H(\eta_i, \zeta_j) + \sin(\theta_\mu) J_{ij}^N \quad (6)$$

consists of

1. the original weight during storage, J_{ij}^S , damped by a factor $\cos(\theta_\mu)$, where $0 < \theta_\mu < \pi/2$

parametrizes the time elapsed between the storage and retrieval of pattern μ (μ is a shorthand for the pattern quadruplet $\{\eta_i, V_i, \zeta_j, U_j\}$).

2. the modification due to the storage of μ itself, represented by a Hebbian term $H(\eta_i, \zeta_j)$ —reflecting the association of patterns $\{\eta_i\}$ and $\{\zeta_j\}$ —also taken to have zero mean, and normalized so that

$$\langle (H(\eta, \zeta))^2 \rangle = \sigma_j^2 \quad (7)$$

γ measures the degree of *plasticity*, i.e. the mean square contribution of the modification induced by one pattern, over the overall variance, across time, of the synaptic weight.

3. the superimposed modifications J^N reflecting the successive storage of new intervening patterns, again normalized such that

$$\langle (J_{ij}^N)^2 \rangle = \sigma_j^2 \quad (8)$$

The mean value of each synaptic weight has been taken to be equal across synapses, and, since it would multiply the average over all input cells of the firing rates, which is fairly constant, it has been collapsed with the threshold term (Treves, 1990). Indeed, Eq. 1 implies that fluctuations in those averaged CA3 rates are small, of order $1/\sqrt{C}$, during storage, and Eq. 2 implies the same for retrieval. Formulas aside, the approximation is likely to be valid when the mean firing levels are strictly regulated by inhibition. The gain of the threshold-linear transfer function (see Treves, 1990) has been set to one by rescaling the weights. It is convenient also to set for the Hebbian term the specific form

$$H(\eta_i, \zeta_j) = \frac{h}{\sqrt{C}} (\zeta_j - \zeta_0)(\eta_i - \langle \eta_i \rangle) \quad (9)$$

where the parameter h ensures the normalization given in Eq. 7; similar forms for H would yield similar results. Note that the firing rates of the presynaptic CA3 cell i and of the postsynaptic CA1 cell j are taken to vary independently across memory patterns, up to very small correlations, because of the extensive convergence of the Schaffer collaterals; this implies that the variance in the synaptic weights is just a sum of terms from each memory pattern, and justifies the interpretation of $1/\gamma$ as the effective number of memory patterns in storage.

The aim is to measure how much, on average, of the information present in a given original pattern $\{\eta_i^v\}$ is still present in the effective output of the system at the

time ν is retrieved, i.e. in the pattern $\{U_j^\nu\}$, that is, to average¹ the mutual information

$$i(\{\eta_i^\nu\}, \{U_j^\nu\}) = \int \prod_i d\eta_i^\nu \int \prod_j dU_j^\nu P(\{\eta_i^\nu\}, \{U_j^\nu\}) \times \ln \frac{P(\{\eta_i^\nu\}, \{U_j^\nu\})}{P(\{\eta_i^\nu\})P(\{U_j^\nu\})} \quad (10)$$

over the *quenched* variables c_{ij} , J_{ij}^S , J_{ij}^N (quenched here means independent of the specific distribution realized in pattern ν , which they are, as the connectivity c_{ij} is taken to be independent of all memory patterns, whereas the synaptic strengths J_{ij}^S and J_{ij}^N depend on all other patterns but ν , which has been singled out in Eq. 6; for the origin of the term quenched see Mezard et al., 1987).

The calculation yields the expression reported in the Appendix, and which depends on a number of parameters, as explained in the following.

Parameters

A very important parameter included in the model is γ , which represents, as stated above, the degree of plasticity of the Schaffer connections, expressed as the ratio between the mean square change in synaptic strength due to the storage of one memory pattern, and the overall variance in synaptic strength. If such variance is entirely due to memory storage, one can say, inversely, that the memory holds of the order of γ^{-1} patterns at any one time. This is only a rough measure, as in fact patterns are not necessarily ever completely effaced in the model, but rather their traces may be only gradually overwritten by other intervening patterns, depending on the chosen dependence of θ on real time, as parametrized by the factor $\cos(\theta)$. The role of γ

Table 2. Meaning of the parameters affecting the amount of information available in CA1, as evaluated here.

| | |
|-----------------------|--|
| γ | degree of (Schaffer collaterals) synaptic plasticity |
| a | sparseness of the patterns stored in CA3 |
| a_ζ | sparseness of the patterns stored in CA1 |
| C | mean convergence of the Schaffer collaterals |
| N | number of CA3 pyramidal cells |
| M | number of CA1 pyramidal cells |
| σ_J | variance in the distribution of synaptic efficacies |
| σ_δ | noise level in CA3 (during retrieval) |
| σ_{ϵ^S} | noise level in CA1 (during storage) |
| σ_{ϵ^R} | noise level in CA1 (during retrieval) |
| ζ_0 | mean non-specific contributions (thresholds) to |
| U_0 | the firing of CA1 cells during storage and retrieval |

and θ is discussed explicitly in the Results section, while most other parameters are fixed once for all as follows.

The evaluation has been carried out for an arbitrary distribution P_η . For simplicity, to get numerical results this is now given a binary form, in which a cell is firing at a rate η^* with probability a , and silent otherwise

$$P_\eta(\eta) = (1 - a)\delta(\eta) + a\delta(\eta - \eta^*) \quad (11)$$

where the firing rate of all CA3 active cells has been set to η^* , and a is a *sparse coding* parameter (in general $a = \langle \eta \rangle^2 / \langle \eta^2 \rangle$, cf. Treves, 1990). An alternative (ternary) form for P_η will be briefly considered later.

σ_δ is measured on the η^* scale (i.e., as a frequency), and is chosen to account both for actual noise in retrieval from CA3 (which corresponds to the standard deviation in the rates recorded during successive trials of an already learned task, as e.g. in Rolls et al., 1989; *fast* noise in thermodynamics jargon), and for interference (the so-called *quenched* noise) caused by memory loading (which would be observable by comparing responses during and after one-shot learning, and which most theoretical models would predict to grow with the square root of the load; Treves, 1990).

The variance in the synaptic efficacies is set to $\sigma_J^2 = 1/C$, which is consistent with the gain of the threshold-linear transfer function having been set to 1, as the variance in the CA3 firing rates contributes, in the linear range, an equal variance to the rates in CA1. Note that this is purely a condition of internal consistency in the formal model, that has to do with the normalization adopted for synaptic weights and transfer functions, and has no import on the results; a quantity, instead, that would be in principle accessible to experiments, is the ratio of standard deviation to mean synaptic strength, and this quantity is not even defined in the present model because the mean baseline strength has been subtracted away.

Moreover, CA1 rates are sensitive to fast noise of s.d. σ_{ϵ^S} , σ_{ϵ^R} , again measured in $H\zeta^2$. The threshold terms ζ_0 , U_0 , also measured as frequencies (given the unit gain), are given negative values such as to produce activity distributions in CA1 with a given sparseness (e.g., roughly matching the sparseness of those in CA3, or slightly more distributed; Barnes et al., 1990).

Finally, C is set to 10^4 (similar to the estimates available for the rat; Amaral et al., 1990) and the ratio M/N of CA1 to CA3 cells, which is higher than 1 in all species studied (Seress, 1988), is here given the value $M/N = 2$. The number of CA3 cells is itself irrelevant

if one considers only the amount of information per CA3 cell.

Results

I have evaluated numerically the expression giving the amount of information in the CA1 output, for several choices of the parameters. Two representative cases are reported here, that exemplify fairly different situations. In both, the result is given as a function of the plasticity γ , and is expressed as a fraction (denoted as \mathcal{I}) of the total information present in the representation stored in CA3. This maximal information, given the choice of a binary probability distribution, is just

$$i(\eta) = -a \ln a - (1 - a) \ln(1 - a) \quad (12)$$

I have compared $\langle i \rangle$ (or, rather, the fraction of the stored information that can be extracted, during retrieval, at the CA1 stage, $\mathcal{I}_{CA1}(\gamma) = \langle i \rangle / i(\eta)$) with two relevant quantities. The first is the fraction of information $\mathcal{I}_{CA3} = i(\eta, V) / i(\eta)$ already retrieved at the CA3 stage, as determined by the noise level in CA3. This allows us to quantify any improvement ($\mathcal{I}_{CA1} - \mathcal{I}_{CA3}$) due to the refining of the retrieval operation at the CA3-to-CA1 stage. The second is the fraction of information³ $\mathcal{I}_{CA1}(\gamma = 0)$ contained in the CA1 output in the absence of plasticity in the Schaffer connections. It is interesting that this can be a substantial fraction of $i(\eta)$, indicating that even fixed connections would produce an output that reflects, to some extent, the information retrieved in CA3. At the same time, the difference $\mathcal{I}_{CA1}(\gamma) - \mathcal{I}_{CA1}(0)$ quantifies the effect of plasticity on the amount of information.

The possible effects of the peculiar one-layer arrangement of pyramidal cells in the CA regions (including CA1) have been studied by manipulating, in the model, the pattern of convergence, given by the distribution C_j of the number of CA3 cells projecting to each cell j of CA1. In one case, \mathcal{C}^1 , intended to approximate the real situation, $C_j = C$ for all j , whereas a second hypothetical case \mathcal{C}^2 has been designed to correspond to a two-layer arrangement of principal cells, with the two layers differing widely in the number of synaptic contacts each cell in CA1 receives: $C_j = C/2$ for half the cells, $C_j = 3C/2$ for the other half⁴.

The first parameter set chosen (Fig. 3) corresponds to a very accurate retrieval operation carried out by the CA3 net, which is imagined to operate well below its memory loading capacity. The sparseness of

the $\{\eta_i\}$ representation is $a = 0.1$ and the noise level is set at $\sigma_\delta^2 = 0.04 \times (\eta^*)^2$. CA1 is taken to operate under more noisy conditions, and with more distributed representations, $\sigma_{\epsilon^S}^2 = \sigma_{\epsilon^R}^2 = 0.2 \times (\eta^*)^2$, and $\zeta_0 = U_0 = -0.1\eta^*$, which results in $a_\zeta = 0.267$. Moreover, a model of plasticity is chosen in which each pattern λ is encoded with the same mean strength onto synaptic efficacies, $\gamma(\theta_\lambda) \equiv \gamma_0$, until it is cancelled abruptly, at $\theta_\lambda = \pi/2$, by some active forgetting mechanism; as for the specific pattern ν under consideration $\cos(\theta_\nu) = 0.8$, corresponding to more than $2/5$ ($\simeq 2 \arccos(0.8)/\pi$) of the memory patterns already in storage when storing ν having been forgotten, and substituted with new ones, by the time ν is retrieved.

One sees from Fig. 3 that, while retrieval in CA3 is good (\mathcal{I}_{CA3} is close to 1) given the low noise level in CA3 itself, in CA1 it is worse, i.e. $\mathcal{I}_{CA1}(\gamma)$ is only around $3/4$; further, it grows monotonically with increasing plasticity. The solid curve, for the convergence model \mathcal{C}^1 , corresponding to unimodal distribution of number of inputs per CA1 cell, shows a peculiar feature at around $\gamma = 0.0002$. This feature is a remnant of the phase transition, or abrupt failure in retrieval performance, occurring in either feedback or multilayer feedforward associative memories when the number of patterns stored exceeds a critical value (see e.g. Treves and Rolls, 1991). With these parameters the critical value would apparently be around $p_c = 1/\gamma = 5000$ patterns. In our single-layer feedforward network performance degrades only gradually (or, conversely, improves only gradually as γ grows), but still one detects vestiges of the transition in a sensitive enough indicator such as the information content. The dashed curve, for model \mathcal{C}^2 , is smoother, because the feature occurs at different γ values for the two classes of cells (p_c is proportional to the convergence C_j) and is thus spread over. Apart from such tiny details, differences between the two models of the connectivity pattern, \mathcal{C}^1 and \mathcal{C}^2 , are minimal, of the order of 1%. \mathcal{C}^1 is slightly superior for high plasticity but, contrary to my own early findings (Treves and Rolls, 1993), this trend can be reversed, e.g. at low plasticity (see also Fig. 4).

For $\gamma = 0$ the information that survives in the CA1 output is less than half, for this choice of parameters. On the other hand, $\mathcal{I}_{CA1}(\gamma)$ would continue to grow with γ up to $\gamma = 1$ — which corresponds to a single memory pattern and is thus outside the range meaningful to this analysis. One can summarize Fig. 3 by saying that the CA3-to-CA1 connections act, with these parameters, as a relay dissipating information. The dissipation is contained if an elevated plasticity

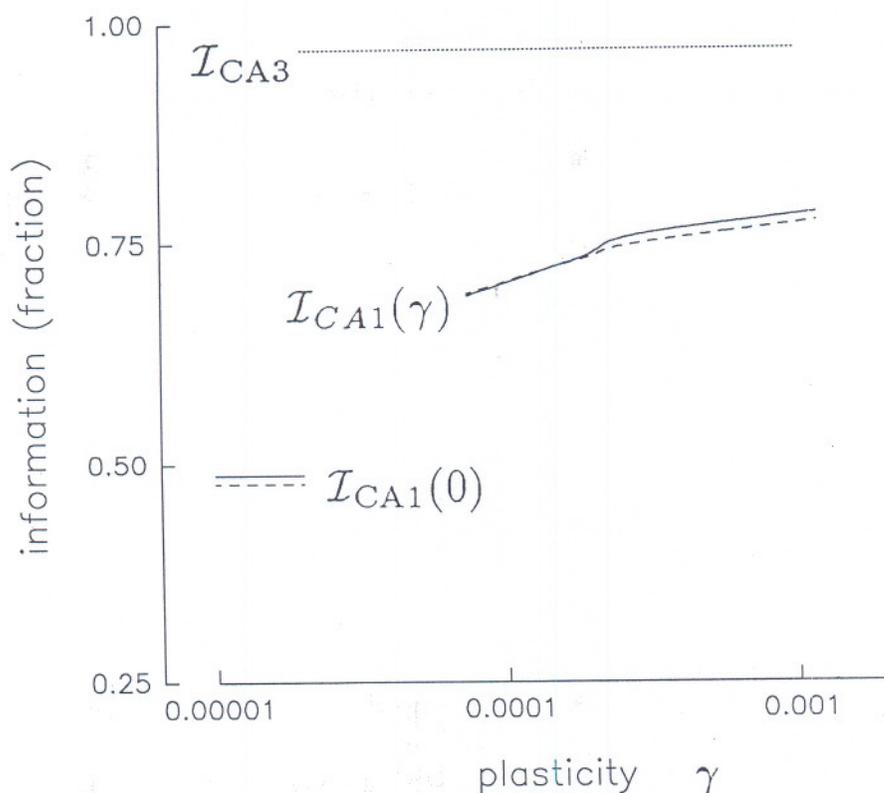


Fig. 3. The fraction of CA3 information preserved in CA1 when retrieval in CA3 is very accurate, and the plasticity model assumes abrupt forgetting and moderate memory loading. The solid curve is for the 'one layer' connectivity model (C^1), the dashed curve for the 'two layers' one (C^2). The result for nonmodifiable connections $\gamma = 0$ is indicated at the left, and the dotted line stands for the fraction of information retrieved in CA3.

level endows the connections with some memory ability, which serves to enhance, essentially, the signal-to-noise ratio; however, the high noise levels in CA1 prevent it from ever attaining the retrieval performance of CA3, where the noise level is (with the particular arbitrary choice of parameters of Fig. 3) much lower.

The second parameter set is chosen to correspond to probably more realistic conditions. $a = 0.05$ and the noise level in CA3 is set at $\sigma_{\delta}^2 = 0.09 \times (\eta^*)^2$, which is meant to include actual noise as well as interference from extensive memory storage, of roughly equal weights. The noise level in CA1, instead, is set to a level similar to only one component (the actual noise) of that CA3, $\sigma_{\epsilon_s}^2 = \sigma_{\epsilon_r}^2 = 0.04 \times (\eta^*)^2$, and the thresholds $\zeta_0 = U_0 = -0.4\eta^*$, are chosen to result in a sparseness, in CA1, $a_{\zeta} = 0.052$, i.e. also the sparseness is similar to the CA3 value.

The model of plasticity is chosen to correspond to gradual decay of memory traces. Numbering different memory patterns from the most recent backwards, $1, \dots, \lambda, \dots, \infty$, one way to model gradual decay is by setting $\cos(\theta_{\lambda}) = \exp -\lambda\gamma_0/2$ (Nadal et al., 1986; cf. Treves and Rolls, 1992) and $\gamma(\theta_{\lambda}) = \gamma_0 \exp -\lambda\gamma_0$, that is, the strength of encoding for older memories

fades exponentially with the number of intervening memories. Within this plasticity model, high plasticity automatically implies rapid forgetting. Although all memories leave their traces on indefinitely, the number of memories that at any moment in time have traces of substantial strength is effectively of order $1/\gamma_0$. Here I consider this plasticity model for the Schaffer collaterals, and obviously I also assume it to be valid within CA3, for the recurrent collaterals. In fact, for an autoassociative network of the type supposedly implemented in CA3, one could analyse the retrieval capacity (Treves and Rolls, 1991). With the same plasticity model $\gamma^{\text{CA3}}(\theta_{\lambda}) = \gamma_0^{\text{CA3}} \exp -\lambda\gamma_0^{\text{CA3}}$ and the same parameters—including the sparseness—as here, one would find (details to be published elsewhere) that to allow retrieval of only the very most recent patterns, γ_0^{CA3} has to be higher than a critical value $\gamma_c \simeq 0.7C^{-1}$. Above this value, a higher plasticity allows more of the last patterns to be retrieved, i.e. a total of $[1/\gamma_0^{\text{CA3}}] \ln[1.4C\gamma_0^{\text{CA3}}]$. This number reaches a maximum of precisely $1/\gamma_0^{\text{CA3}}$ for $1/\gamma_0^{\text{CA3}} \simeq 0.5C$. For even higher plasticity, the effect of rapid forgetting dominates and the number of retrievable patterns decreases. Let us now turn to quantifying the effect of

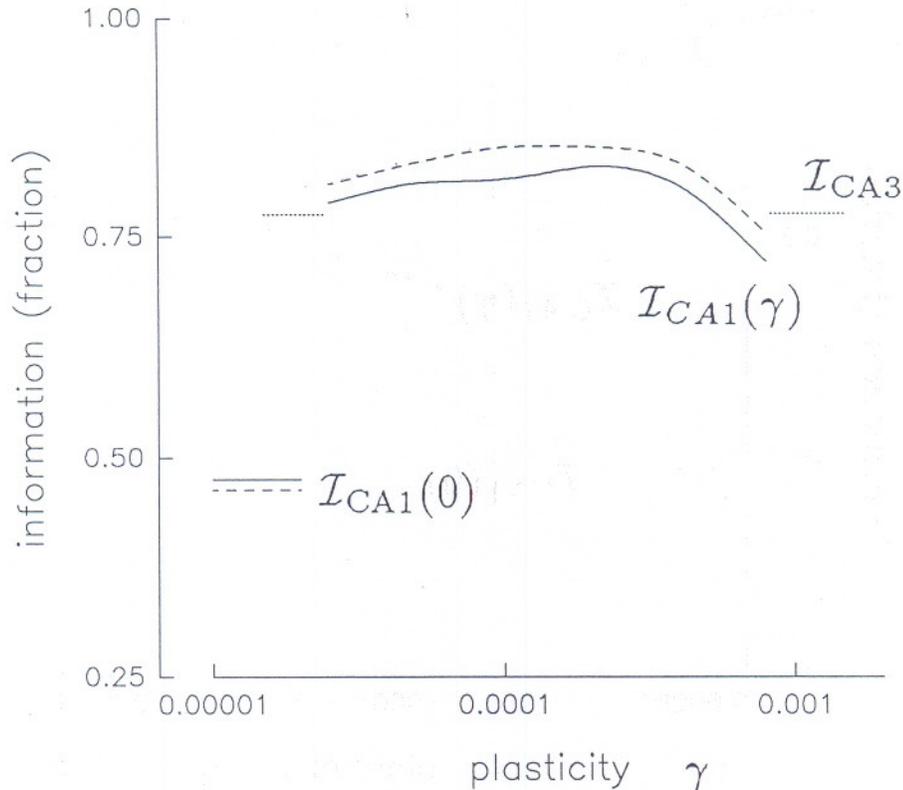


Fig. 4. The fraction of CA3 information preserved in CA1 when the noise levels in CA3 and CA1 are comparable, and the plasticity model assumes gradual forgetting and nearly maximal memory loading. Notation as in Fig. 3.

exponential forgetting on the combined CA3 and CA1 model.

A similar behaviour to that of the isolated CA3 network, with a performance maximum for an optimal degree of plasticity, appears to characterise the model CA3-to-CA1 network when analysed in terms of information content, as shown in Fig. 4. There, the plasticity in CA3 has been chosen close to optimal, $\gamma_0^{\text{CA3}} \simeq 2/C$, and the pattern ν to be retrieved has been chosen such that $\nu = 1/\gamma_0^{\text{CA3}} \simeq 0.5C$, i.e. close to the limits of retrieval capacity by the CA3 network (C denotes here both the mean convergence of recurrent collaterals in CA3 and of Schaffer collaterals in CA1, taken to be roughly equal). As γ_0 , (that is, γ_0^{CA1}) grows, one spots, in the solid curve for model \mathcal{C}^1 , a slight feature corresponding to the remnant of the phase transition. Past this feature, the fraction of information retrievable reaches a maximum (in both connectivity models \mathcal{C}^1 and \mathcal{C}^2). For even higher plasticity, the $\exp -\lambda\gamma_0$ forgetting term dominates, and performance, in terms of information, degrades rapidly. Not surprisingly, the optimal plasticity roughly matches the plasticity assumed in the CA3 recurrent connections. The two-layer model \mathcal{C}^2 performs better than the one-layer one \mathcal{C}^1 , in this case (because of the way the two classes

of cells with different number of inputs are affected by the forgetting term), but again by relatively small amounts. It is interesting to note that, given the comparable noise levels set for CA3 and CA1, $\mathcal{I}_{\text{CA1}}(\gamma)$ exceeds \mathcal{I}_{CA3} within a certain plasticity range, i.e. one can extract more information from CA1 cells than from CA3 cells, indicating that in this case the contribution of the Schaffer connections to memory retrieval more than compensates for the dissipation in information content they cause. One may generalise the results exemplified in Fig. 4 by concluding that, at least for some plasticity models that include forgetting effects,

- there is an optimal degree of plasticity that is close to the degree of plasticity in the preceding autoassociator. Moreover,
- the feedforward network can not only, if appropriately designed, limit the loss of information retrieved in the autoassociator, but also further improve the quality of retrieval.

The behaviours exemplified in Figs. 3 and 4 are quite general. At a quantitative level, the efficacy with which plasticity in the Schaffer connections enhances signal-to-noise, and thereby the information content in the firing pattern, depends substantially on the reference

Table 3. The effect of varying different parameters on \mathcal{I}_{CA3} and $\mathcal{I}_{CA1}(0)$. σ_δ , σ_{eS} , σ_{eR} , ζ_0 and U_0 are expressed in units of η^* , $i(\eta)$ in bits and the other quantities are pure numbers. Boldface values are those that differ from the ones in the line just above. The * denotes the representation stored as a ternary distribution of rates.

| M/N | σ_δ^2 | $\sigma_{eS}^2 = \sigma_{eR}^2$ | $\zeta_0 = U_0$ | a_η | $i(\eta)$ | \mathcal{I}_{CA3} | a_ζ | $\mathcal{I}_{CA1}(0)$ |
|------------|-------------------|---------------------------------|-----------------|-------------|-----------------------------|---------------------|--------------|------------------------|
| 2.0 | 0.04 | 0.20 | -0.1 | 0.1 | 0.469 N | 0.97 | 0.267 | 0.49 |
| 2.0 | 0.20 | 0.20 | -0.1 | 0.1 | 0.469 N | 0.51 | 0.267 | 0.28 |
| 2.0 | 0.20 | 0.20 | -0.4 | 0.1 | 0.469 N | 0.51 | 0.138 | 0.24 |
| 1.5 | 0.20 | 0.20 | -0.4 | 0.1 | 0.469 N | 0.51 | 0.138 | 0.20 |
| 2.0 | 0.09 | 0.04 | -0.4 | 0.1 | 0.469 N | 0.79 | 0.082 | 0.51 |
| 2.0 | 0.09 | 0.04 | -0.4 | 0.05 | 0.286 N | 0.77 | 0.052 | 0.47 |
| 2.0 | 0.04 | 0.04 | -0.4 | 0.05 | 0.286 N | 0.97 | 0.052 | 0.60 |
| 2.0 | 0.09 | 0.04 | -0.4 | 0.05* | 0.407 N | 0.48 | 0.052 | 0.30 |

levels set by \mathcal{I}_{CA3} (how much information can be retrieved already from CA3) and $\mathcal{I}_{CA1}(0)$ (how much information results in the CA1 output in the absence of plasticity). These levels depend in turn on the different parameters of the model. To get a feeling for such relationships, Table 3 presents the values of \mathcal{I}_{CA3} and $\mathcal{I}_{CA1}(0)$ resulting from different choices of the set of parameters. To simplify matters, only some of the parameters have been varied independently of others. The table refers to the one-layer connectivity model \mathcal{C}^1 , and shows the following quantitative effects. The first line corresponds to Fig. 3. The second line shows that increasing the noise affecting CA3 from the very low level assumed in Fig. 3 up to the same level as in CA1, almost halves both \mathcal{I}_{CA3} and $\mathcal{I}_{CA1}(0)$. The CA1 representation is still much more distributed than the CA3 one ($a_\zeta > a_\eta$). In the third line, the CA1 representation is made sparser by acting on the threshold, thereby moving the operating regime of CA1 cells deeper in the subthreshold region. This reduces $\mathcal{I}_{CA1}(0)$ as more cells remain silent. In the fourth line, the ratio of CA1 cells to CA3 cells is reduced to 1.5, and this further decreases $\mathcal{I}_{CA1}(0)$ (which, let us remember, is the information extracted from the whole CA1 population). In the fifth line the noise levels in both CA3 and CA1 are lowered, taking values as in Fig. 4. This obviously enhances \mathcal{I}_{CA3} , but it also enhances $\mathcal{I}_{CA1}(0)$, despite lowering the CA1 sparseness a_ζ which would tend, in itself, to degrade the information content in CA1 (cf. the third versus the second line). The sixth line corresponds to Fig. 4, as also the sparseness in CA3 has been brought to $a_\eta = 0.05$. This implies a lower absolute value for $i(\eta)$, but in relative terms the differences with the line above are minor. In the seventh line, the noise in CA3 has been decreased to model a situation of low memory loading, and correspondingly

no quenched contribution to the noise. Information ratios go up markedly. Finally, in the last line the case is considered, the only instance in this analysis, of a non-binary distribution of firing rates for the memories stored in CA3. In particular the *ternary* form

$$P_\eta(\eta) = (1 - 4a/3)\delta(\eta) + a\delta(\eta - \eta^*/2) + (a/3)\delta(\eta - 3\eta^*/2) \quad (13)$$

has been chosen (cf. the cases considered in Treves, 1990), that maintains the same first two moments and the same sparseness as the binary form, and allows isolating the effects of a non-binary (and in particular multimodal) distribution of firing rates. These are very relevant, as real distributions of firing rates in CA3, however they may be defined and measured, are certainly far from binary. Table 3 shows that, while the stored information $I(\eta)$ grows (cf. with the sixth line), the information content that can be retrieved goes down both in CA3 and CA1 (at $\gamma = 0$), in fact not only relative to how much was stored ($i(\eta)$) but even in absolute terms. This is a manifestation of the difficulty with which a network of simple threshold-linear units can discriminate among multiple firing levels in the input units. It is consistent with (i) the observation that the total information retrievable from an autoassociator does not exceed the value attained when the stored patterns are in binary form (Treves, 1990) and (ii) the fact that even non-binary patterns stored in a model CA3 net are estimated to convey not much more *mutual* information about their inputs than binary ones (Treves and Rolls, 1992), the rest being just noise. For comparison, the maximum value of the information in CA1 for the non-binary form of encoding in CA3 has also been evaluated, with the plasticity turned on and set to its optimal level, yielding an information $\mathcal{I}_{CA1}(\gamma) = 0.49$

to be compared with $\mathcal{I}_{CA1}(\gamma) = 0.83$ in the binary case of Fig. 4.

Discussion

Level of Abstraction of the Mathematical Model

Can simplified network models produce meaningful quantitative results? It is obvious that a reasonable answer to this general question depends entirely on the properties investigated. For example, network models made up of units described by simple transfer functions (which approximate some kind of neuronal steady rate response to a steady current input) are of course inappropriate to study dynamical issues (Treves, 1992, 1993). Nevertheless, they may be useful in other contexts; possibly even when the notion of 'steady response' to 'steady inputs' is intrinsically ill-defined, as it appears to be when modelling complex-spike cells, and in a region such as the hippocampus prone in some species to temporal modulation (Bland, 1986). The analysis presented here is based on the hypothesis, which can be checked e.g. through extensive computer simulation, that, in fact, basic information-theoretic properties, as the ones studied here, characterizing the network in its simplified abstract form, carry over to more realistic descriptions.

Accepting the quantitative validity of simplified formal models is further complicated, in the case of the hippocampus, by the precedent of the hippocampal theory of the late David Marr (1971). As in his other early papers, but much more so in the one on the hippocampus, Marr originated, organized or summarized many of the system-level ideas to which a majority of current views subscribe. Yet the mathematical apparatus on which he built his formal model—which could have been thought to be his prime contribution—was perforce crude and rudimentary. Both his model neurons and his model synapses are represented as binary elements, the only ones that could be handled by the semianalytic methods available at the time. His 'neurophysiological' predictions, based therefore on a scheme at gross variance with the real structure, are groundless at a quantitative level (Brown, 1990). Although for curious historical reasons neural modelling based on binary elements is still a cherished practice, the advent of modern analytical techniques (mainly derived from statistical physics; Amit, 1989) has made it possible to consider mathematical models considerably closer to their neurobiological counterparts (Treves and

Rolls, 1991). The utility of such models has to be now reevaluated.

Model Parameters and Experimental Quantities

Although there is admittedly a lot of scope for improvements, models of the type considered here are distinctive in the way their parameters can be related to experimentally observed quantities. In practice this may require resolving some ambiguities, but in principle it is straightforward. Thus, the degree of plasticity may be measurable, although a direct measure would have to be both at the single synapse level, and interleaved with behaviour *in vivo*, a combination which is beyond present techniques. Note that this parameter, which plays a crucial role in the present analysis, and which emphasizes the obviously graded nature of synaptic plasticity, does not even have a meaning in binary models *à la* Marr. Likewise, the various noise parameters have direct experimental correlates, once an appropriate way of measuring firing rates is clarified (in the sense of unmasking hidden correlates of firing variability, that could otherwise be interpreted as noise). Other parameters, such as the thresholds, which summarize a variety of effects and have no immediate correlate, can be given appropriate values inferred indirectly from such measurable parameters as the sparseness (Skaggs and McNaughton, 1992).

Quantitative Results and Experimental Quantities

Similar considerations apply to the output of the analysis: in this case, essentially, the information values. Information is a difficult quantity to extract from mammalian cell recordings (Optican et al., 1991; Tovee et al., 1993; Treves and Panzeri, 1995), and special care must be taken to define it correctly in terms of adequate *a priori* conditions. Nevertheless, experimental estimates of the information content of neuronal responses are starting to become available, even for quite large populations of hippocampal cells recorded simultaneously (Wilson and McNaughton, 1993). Estimates of the information, in bits per cell, that can be deduced from these large scale recordings (e.g. roughly 0.05 bits for firing rates measured over 500 ms) are broadly compatible, at least in order of magnitude, with the observed sparseness and noise of the firing patterns (quantitative comparisons will be published separately). Much finer measurements will likely be made in the near future, together with an elucidation of how information values depend on factors, such as

aging, that modulate sparseness (Barnes et al., 1983, 1995; Mizumori et al. 1993) and plasticity (Barnes and McNaughton, 1985). Concurrent refinements of the present formal model may enable us to account explicitly for factors not included here, such as the role of the direct perforant path projection to CA1⁵, more detailed connectivity patterns (Amaral and Witter, 1989), and so on. It is possible to envisage a situation, therefore, in which the formal analysis will become, through the simultaneous tuning of the experimental techniques and of the models, relevant in its full quantitative import.

System Level Overview

Already in the present situation, however, a quantitative analysis is important, and that is because a qualitative feature—plasticity of the Schaffer connections—produces what is only a quantitative effect—essentially, enhancement of the signal-to-noise ratio in memory retrieval. To dissect the phenomenon, a mathematically formalised model is needed⁶. The signal-to-noise enhancement depends on the degree of plasticity. When the plasticity model ('learning rule') that incorporates gradual forgetting is assumed, the model explicitly indicates that *there is an optimal degree of associative plasticity of the Schaffer collaterals*, i.e. the degree that matches that of the recurrent collaterals in CA3 (as the case would presumably be if they shared similar mechanisms). It is easy to see, however, that even with more general plasticity models it is optimal to have matching degrees of plasticity in CA3 and CA1. For example, in the simplest situation considered first in this analysis, although it is true that the information in one stored pattern increases monotonically with increasing CA1 plasticity, it is also true that if the plasticity in CA1 is higher than in CA3, not all of the patterns retrievable from CA3 will be retrievable, or at least benefit from the signal-to-noise enhancement, in CA1.

At the system level, the enhancement of the signal-to-noise ratio at the CA3-to-CA1 stage can, depending on such conditions as the noise levels, result in either a minimization of information loss (if the CA1 representation ends up containing less than the one retrieved from CA3) or even an actual improvement, or refinement, of the information retrieved by the CA3 autoassociator. These appear to be two qualitatively different outcomes, but in fact are only two different regimes in the competition between signal and noise. Independently of which of the two may happen to be the operating regime, the role of CA1, formulated in this way, seems to overlap greatly with that ascribed to CA3. After all, we have ourselves (Treves and Rolls, 1991)

described the CA3 type of recurrent collateral autoassociator as an information efficient device for content-addressed memory retrieval (Hopfield, 1982). One may wonder then, why add a further feedforward stage.

What is crucial to note, here, is the larger number of pyramidal cells in CA1. Estimates of the ratio M/N between the number of pyramidal cells in CA1 and in CA3 vary from 1.4 in rats (Amaral et al., 1990) to 5.9 in humans (West and Gundersen, 1990). Assuming the same information content in the *whole* representation retrieved from CA1 as in the one from CA3, the average information content *per cell* would be much reduced in CA1. It is important to realize that even a reduction of, say, a factor 2 is quite remarkable. If, for example, CA3 cells coded information independently, each by firing at one of 9 distinct, equiprobable rates (which clearly they do not, but it is a conveniently visualizable example), then a double number of CA1 cells could convey the same information each by firing at one of just 3 different rates, making the CA1 representation very much more robust to various forms of noise. This effect is a simple recoding by expansion, but it may be precisely what is required in order to make the hippocampal memory output ready to be backprojected into the more complex circuitry of the neocortex (Treves and Rolls, 1994) and to be interfaced, already in the subiculum and entorhinal cortex, with different information channels.

Acknowledgments

I wish to thank Edmund Rolls for extensive discussions and Stefano Panzeri for a careful check of the calculations. Partial support from the Human Frontier Science Program, the Human Capital and Mobility Program of the EEC, the MRC of the UK, and CNR, INFM and INFN of Italy.

Appendix

Replica Evaluation

$P(\{\eta_i\}, \{U_j\})$ is written (simplifying the notation) as

$$\begin{aligned} P(\eta, U) &= P(U | \eta)P(\eta) \\ &= \int_V \int_\zeta dV d\zeta P(U | V, \zeta, \eta) \\ &\quad \times P(V | \eta)P(\zeta | \eta)P(\eta) \end{aligned} \quad (14)$$

where the different probability densities implement the model defined above.

The (average) amount of information is evaluated using the replica trick (Nadal and Parga, 1993) as

$$\begin{aligned} \langle i(\eta, U) \rangle_{c, J^S, J^N} \\ = \lim_{n \rightarrow 0} \frac{1}{n} \left\langle \int d\eta dUP(\eta, U) \right. \\ \left. \times \left[\left[\frac{P(\eta, U)}{P(\eta)} \right]^n - [P(U)]^n \right] \right\rangle_{c, J^S, J^N} \quad (15) \end{aligned}$$

where one needs to introduce $n + 1$ replicas of the variables $\delta_i, \epsilon_j^S, \epsilon_j^R, V_i, \zeta_j$ and, for the second term in curly brackets only, η_i .

The calculation is a standard, if tedious, mean-field exercise, and leads to the expression

$$\begin{aligned} \langle i \rangle = \text{extr}_{y_A, \tilde{y}_A} \left\{ \sum_j \Gamma(y_A, w^0, z^0, C_j, \gamma) - \frac{N}{2} y_A \tilde{y}_A \right. \\ \left. + N \int D\tilde{s}_1 \langle F(\tilde{s}_1, 0, \eta, \tilde{y}_A, 0, 0) \right. \\ \left. \times \ln F(\tilde{s}_1, 0, \eta, \tilde{y}_A, 0, 0) \rangle_\eta \right\} \\ - \text{extr}_{y_B, \tilde{y}_B, w_B, \tilde{w}_B, z_B, \tilde{z}_B} \\ \times \left\{ \sum_j \Gamma(y_B, w_B, z_B, C_j, \gamma) \right. \\ \left. - \frac{N}{2} (y_B \tilde{y}_B + 2w_B \tilde{w}_B + z_B \tilde{z}_B) \right. \\ \left. + N \int D\tilde{s}_1 D\tilde{s}_2 \langle F(\tilde{s}_1, \tilde{s}_2, \eta, \tilde{y}_B, \tilde{w}_B, \tilde{z}_B) \rangle_\eta \right. \\ \left. \times \ln \langle F(\tilde{s}_1, \tilde{s}_2, \eta, \tilde{y}_B, \tilde{w}_B, \tilde{z}_B) \rangle_\eta \right\} \quad (16) \end{aligned}$$

where taking the extremum means evaluating each of the two terms, separately, at a saddle-point over the variables indicated (and dividing by $\ln 2$ to yield a result in bits). The notation is as follows. N is the number of CA3 cells, whereas the sum over j is over M CA1 cells. F is given by

$$\begin{aligned} F(\tilde{s}_1, \tilde{s}_2, \eta, \tilde{y}, \tilde{w}, \tilde{z}) \\ = \left\{ \phi \left[\frac{\eta + \sigma_\delta^2 (\tilde{s}_+ - \tilde{w}\eta)}{\sigma_\delta \sqrt{1 + \sigma_\delta^2 \tilde{y}}} \right] \times \frac{1}{\sqrt{1 + \sigma_\delta^2 \tilde{y}}} \right. \\ \left. \times \exp \left[\frac{[\eta + \sigma_\delta^2 (\tilde{s}_+ - \tilde{w}\eta)]^2}{2\sigma_\delta^2 (1 + \sigma_\delta^2 \tilde{y})} \right] + \phi \left[\frac{-\eta}{\sigma_\delta} \right] \exp \frac{\eta^2}{2\sigma_\delta^2} \right\} \\ \times \exp \left[\eta \tilde{s}_- - \frac{\eta^2}{2\sigma_\delta^2} (1 + \sigma_\delta^2 \tilde{z}) \right] \quad (17) \end{aligned}$$

and has to be averaged over P_η (a *quenched* average in the second line of Eq. 16, an *annealed* average in the last line), and over the Gaussian variables of zero mean and unit variance \tilde{s}_1, \tilde{s}_2 .

$$Ds \equiv (ds/\sqrt{2\pi}) \exp -s^2/2 \quad \phi(x) \equiv \int_{-\infty}^x Ds \quad (18)$$

\tilde{y}, \tilde{w} and \tilde{z} are saddle-point parameters, and \tilde{s}_+, \tilde{s}_- are linear combinations of \tilde{s}_1, \tilde{s}_2 :

$$\begin{aligned} \tilde{s}_\pm = \sum_{k=1}^2 (\mp 1)^{(k-1)} \\ \times \sqrt{\frac{[\sqrt{(\tilde{y}-\tilde{z})^2 + 4\tilde{w}^2} \mp (-1)^k (\tilde{y}-\tilde{z})] (\tilde{y}\tilde{z} - \tilde{w}^2)}{[\tilde{y} + \tilde{z} + (-1)^k \sqrt{(\tilde{y}-\tilde{z})^2 + 4\tilde{w}^2}] \sqrt{(\tilde{y}-\tilde{z})^2 + 4\tilde{w}^2}}} \tilde{s}_k \quad (19) \end{aligned}$$

(notice that in the second line of Eq. 16 one has $\tilde{s}_+ = \tilde{s}_1 \sqrt{\tilde{y}_A}, \tilde{s}_- = 0$).

Γ is effectively an entropy term for the CA1 activity distribution, given by

$$\begin{aligned} \Gamma(y, w, z, C_j, \gamma) \\ = \int \frac{ds_1 ds_2}{2\pi \sqrt{\det \mathbf{T}'_j}} \exp -(s_1 s_2) \frac{(\mathbf{T}'_j)^{-1}}{2} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \\ \times \left[\int_{-\infty}^0 dUG(U) \ln \int_{-\infty}^0 dU' G(U') \right. \\ \left. + \int_0^\infty dUG(U) \ln G(U) \right] \quad (20) \end{aligned}$$

where

$$\begin{aligned} G(U) \\ = G(U; s_1, s_2, y, w, z, C_j, \gamma) \\ = \phi \left[\frac{(\xi_0 - s_2) (T_{yj} + 2g_j T_{wj} + g_j^2 T_{zj}) + (U - U_0 + s_1 + g_j s_2) (T_{wj} + g_j T_{zj})}{\sqrt{(T_{yj} T_{zj} - T_{wj}^2) (T_{yj} + 2g_j T_{wj} + g_j^2 T_{zj})}} \right] \\ \times \frac{1}{\sqrt{2\pi (T_{yj} + 2g_j T_{wj} + g_j^2 T_{zj})}} \\ \times \exp -\frac{(U - U_0 + s_1 + g_j s_2)^2}{2(T_{yj} + 2g_j T_{wj} + g_j^2 T_{zj})} \quad (21) \end{aligned}$$

$$\begin{aligned} + \phi \left[\frac{-(\xi_0 - s_2) T_{yj} - (U - U_0 + s_1 + g_j \xi_0) T_{wj}}{\sqrt{(T_{yj} T_{zj} - T_{wj}^2) T_{yj}}} \right] \\ \times \frac{1}{\sqrt{2\pi T_{yj}}} \exp -\frac{(U - U_0 + s_1 + g_j \xi_0)^2}{2T_{yj}} \quad (22) \end{aligned}$$

and

$$\begin{aligned}
 T_{yj} &= \sigma_{\epsilon_R}^2 + \sigma_j^2 C_j (y^0 - y) \\
 T_{wj} &= \sigma_j^2 C_j (w^0 - w) \cos(\theta) \\
 T_{zj} &= \sigma_{\epsilon_S}^2 + \sigma_j^2 C_j (z^0 - z) \\
 \mathbf{T}'_j &= \sigma_j^2 C_j \begin{pmatrix} y & w \cos(\theta) \\ w \cos(\theta) & z \end{pmatrix}
 \end{aligned} \tag{23}$$

are effective noise terms.

$$g_j = h \frac{C_j}{C} x^0 \langle \eta \rangle_\eta \sqrt{C \gamma(\theta)} \tag{24}$$

y, w, z are saddle-point parameters (conjugated to \tilde{y}, \tilde{w} and \tilde{z}), and x^0, y^0, w^0, z^0 are corresponding single-replica parameters fixed as

$$\begin{aligned}
 x^0 &= \frac{1}{N} \sum_i \left\langle \frac{(\eta_i - \langle \eta \rangle_\eta) V_i}{\langle \eta \rangle_\eta} \right\rangle = \left\langle \frac{(\eta - \langle \eta \rangle_\eta)}{\langle \eta \rangle_\eta} \right. \\
 &\quad \left. \times \left[\eta \phi \left(\frac{\eta}{\sigma_\delta} \right) + \frac{\sigma_\delta}{\sqrt{2\pi}} \times \exp -\frac{1}{2} \left(\frac{\eta}{\sigma_\delta} \right)^2 \right] \right\rangle_\eta \\
 y^0 &= \frac{1}{N} \sum_i \langle V_i^2 \rangle = \left\langle [\sigma_\delta^2 + \eta^2] \phi \left(\frac{\eta}{\sigma_\delta} \right) \right. \\
 &\quad \left. + \frac{\eta \sigma_\delta}{\sqrt{2\pi}} \exp -\frac{1}{2} \left(\frac{\eta}{\sigma_\delta} \right)^2 \right\rangle_\eta \\
 w^0 &= \frac{1}{N} \sum_i \langle \eta_i V_i \rangle = \left\langle \eta \left[\eta \phi \left(\frac{\eta}{\sigma_\delta} \right) \right. \right. \\
 &\quad \left. \left. + \frac{\sigma_\delta}{\sqrt{2\pi}} \exp -\frac{1}{2} \left(\frac{\eta}{\sigma_\delta} \right)^2 \right] \right\rangle_\eta \\
 z^0 &= \frac{1}{N} \sum_i \eta_i^2 = \langle \eta^2 \rangle_\eta
 \end{aligned} \tag{25}$$

Notes

1. An average is sought, because no meaning could possibly be assigned to a result specific to certain values of each c_{ij}, J_{ij}^S and J_{ij}^N , and because extensive quantities like information are expected anyway to coincide with their average.
2. Data on trial to trial variability in firing rates in CA3 and CA1 are presently too scarce to allow a systematic analysis; but these model parameters can be set to reproduce experimental data as it becomes available.
3. This is given by a much simpler expression derivable from Eq. 16 for $\gamma = 0$.
4. This was conceived as an idealization of the arrangement in supragranular and infragranular layers of pyramidal cells in neocortical areas; in order to check whether it affected significantly the informational property considered here.

5. In very general terms, the direct perforant path input may allow CA1 to integrate the complete but compressed representation retrieved from the CA3 memory, with the partial but information-richer representation, available in entorhinal cortex, of *only* those elements of the memory that served as the cue. It has even been proposed (Levy, 1989) that the two representation may be time-shifted, so that information can be associated together across time.
6. The fact that the information estimate is obtained analytically (as allowed by the relative abstractedness of the model) permits us to determine directly the effect of varying several model parameters. The insight gained through an analytical procedure cannot be matched by that affordable with the computer simulation of more realistic model systems.

References

Amaral DG, Ishizuka N, Claiborne B (1990) Neurons, numbers and the hippocampal network. *Progress in Brain Research* 83:1-11.

Amaral DG, Witter MP (1989) The three-dimensional organization of the hippocampal formation: A review of anatomical data. *Neuroscience* 31:571-591.

Amit DG (1989) Modeling Brain Function. Cambridge University Press, New York.

Barnes CA, McNaughton BL (1985) An age comparison of the rates of acquisition and forgetting of spatial information in relation to long-term enhancement of hippocampal synapses. *Behavioral Neuroscience* 99:1040-1048.

Barnes CA, McNaughton BL, Mizumori SJY, Leonard BW, Lin LH (1990) Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Progress in Brain Research* 83:287-300.

Barnes CA, McNaughton BL, O'Keefe J (1983) Loss of place specificity in hippocampal complex spike cells of senescent rat. *Neurobiology of Aging* 4:113-119.

Barnes CA, Treves A, Rao G, Shen J (1994) Electrophysiological markers of cognitive aging: region specificity and computational consequences. *Seminars in the Neurosciences* 6:359-367.

Bland BH (1986) The physiology and pharmacology of hippocampal formation theta rhythms. *Progress in Neurobiology* 26:1-54.

Brown MW (1990) Why does the cortex have a hippocampus? In: M Gabriel, J Moore, eds. Learning and Computational Neuroscience: Foundations of Adaptive Networks. MIT Press, Boston. pp. 233-282.

Eichenbaum H, Otto T, Cohen NJ (1992) The hippocampus—what does it do? *Behavioural and Neural Biology* 57:2-36.

Gaffan D (1992) The role of the hippocampo-fornix-mammillary system in episodic memory. In: LR Squire, N Butters, eds. *Neuropsychology of Memory*. Guilford, New York. pp. 336-346.

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational capabilities. *Proceedings of the National Academy of Science (USA)* 79:2554-2558.

Kesner RP, Dakis M (1992) Memory for objects, spatial locations and motor responses: Triple dissociations among the hippocampus, caudate nucleus and medial extrastriate visual cortex. *Society for Neuroscience Abstracts* 18:1423.

Kohonen T (1984) Self-organization and associative memory. Springer Berlin.

- Levy WB (1989) A Computational approach to hippocampal function. In: RD Hawkins, GH Bower, eds. *Computational Models of Learning in Simple Neural Systems*. Academic Press, Orlando FL. pp. 243–305.
- Marr D (1971) Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London B* 262:24–81.
- McClelland JL, McNaughton BL, O'Reilly R, Nadel L (1992) Complementary roles of hippocampus and neocortex in learning and memory. *Society for Neuroscience Abstracts* 18:1216.
- Mezard M, Parisi G, Virasoro M (1987) *Spin Glass Theory and Beyond*. World Scientific Singapore.
- Mizumori SJY, Lavoie AM, Kalyani A (1993) Redistribution of place representation within the hippocampus of aged rats. *Society for Neuroscience Abstracts* 19:793.
- Nadal JP, Parga N (1993) Information processing by a perceptron in an unsupervised learning task. *Network* 4:295–312.
- Nadal JP, Toulouse G, Changeux JP, Dehaene S (1986) Networks of formal neurons and memory palimpsests. *Europhysics Letters* 1:535–542.
- O'Keefe J (1990) A computational theory of the hippocampal cognitive map. *Progress in Brain Research* 83:301–312.
- Optican LM, Gawne TJ, Richmond BJ, Joseph PJ (1991) Unbiased measures of transmitted information and channel capacity from multivariate neuronal data. *Biological Cybernetics* 65:305–310.
- Rolls ET (1987) Information representation, processing and storage in the brain, Analysis at the single neuron level. In: JP Changeux, M Konishi, eds. *The Neural and Molecular Bases of Learning*. Wiley, New York. pp. 503–540.
- Rolls ET (1989) Functions of neuronal networks in the hippocampus and neocortex in memory. In: JH Byrne, WO Berry, eds. *Neural models of plasticity: Experimental and theoretical approaches*. Academic Press, San Diego, CA. pp. 240–265.
- Rolls ET, Miyashita Y, Cahusac PMB, Kesner RP, Niki H, Feigenbaum J, Bach L (1989) Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. *Journal of Neuroscience* 9:1835–1845.
- 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. *Journal für Hirnforschung* 29:332–340.
- Skaggs WE, McNaughton BL (1992) Quantification of what it is that hippocampal cell firing encodes. *Society for Neuroscience Abstracts* 18:1216.
- Squire LR (1992) Memory and the hippocampus: A synthesis from findings with rats, monkeys and humans. *Psychological Review* 99:195–231.
- Treves A (1990) Graded-response neurons and information encodings in autoassociative memories. *Physical Review A* 42:2418–2430.
- Treves A (1992) Local neocortical processing: a time for recognition. *International Journal of Neural Systems* 3(Supp.):115–119.
- Treves A (1993) Mean-field analysis of neuronal spike dynamics. *Network* 4:259–284.
- Treves A, Miglino O, Parisi D (1992) Rats, nets, maps and the emergence of place cells. *Psychobiology* 20:1–8.
- Treves A, Panzeri S (1995) The upward bias in measures of information derived from limited data samples. *Neural Computation* 7:398–406.
- Treves A, Rolls ET (1991) What determines the capacity of autoassociative memories in the brain? *Network* 2:371–397.
- Treves A, Rolls ET (1992) Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* 2:189–199.
- Treves A, Rolls ET (1993) Computational constraints that may inform hippocampal and hippocampo-cortical connections. *Society for Neuroscience Abstracts* 19:1608.
- Treves A, Rolls ET (1994) A computational analysis of the role of the hippocampus in memory. *Hippocampus* 4:373–391.
- Tovee MJ, Rolls ET, Treves A, Bellis RP (1993) Information encoding and the responses of single neurons in the primate temporal visual cortex. *Journal of Neurophysiology* 70:640–654.
- West MJ, Gundersen HJG (1990) Unbiased stereological estimation of the number of neurons in the human hippocampus. *Journal of Comparative Neurology* 296:1–22.
- Wilson M, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. *Science* 261:1055–1058.
- Zipser D (1985) A computational model of hippocampal place fields. *Behavioral Neuroscience* 99:1006–1018.