Quantitative analysis of a Schaffer collateral model

Simon Schultz
Stefano Panzeri
Edmund Rolls
University of Oxford, Department of Experimental Psychology
South Parks Rd., Oxford OX1 3UD, U.K.
Alessandro Treves
Programme in Neuroscience, SISSA,
via Beirut 2-4, 34013 Trieste, Italy

To appear in R. Baddeley, P. Hancock and P. Földiák (Eds.)
Information Theory and the Brain, Cambridge University Press,
Cambridge, U.K., 1998.

1 Introduction

Recent advances in techniques for the formal analysis of neural networks (Amit et al., 1987; Gardner, 1988; Tsodyks and Feigelman, 1988; Treves, 1990; Nadal and Parga, 1993) have introduced the possibility of detailed quantitative analyses of real brain circuitry. This approach is particularly appropriate for regions such as the hippocampus, which show distinct structure and for which the microanatomy is relatively simple and well known.

The hippocampus, as archicortex, is thought to pre-date phylogenetically the more complex neocortex, and certainly possesses a simplified version of the six-layered neocortical stratification. It is not of interest merely because of its simplicity, however: evidence from numerous experimental paradigms and species points to a prominent role in the formation of long-term memory, one of the core problems of cognitive neuroscience (Scoville and Milner, 1957; Weiskrantz, 1987; Gaffan, 1992; Cohen and Eichenbaum, 1993; McNaughton and Morris, 1987; Rolls, 1991). Much useful research in neurophysiology and neuropsychology has been directed qualitatively, and even merely categorially, at understanding hippocampal function. Awareness has dawned, however, that the analysis of *quantitative* aspects of hippocampal organisation is essential to an understanding of why evolutionary pressures have resulted in the mammalian hippocampal system being the way it is (Amaral et al., 1990; Treves et al., 1996; Stephan, 1983; Witter and Groenewegen, 1992). Such an understanding will require a theoretical framework (or formalism) that is sufficiently powerful to yield quantitative expressions for meaningful parameters, that can be considered valid for the real hippocampus, is parsimonious with known physiology, and is simple enough to avoid being swamped by details that might obscure phenomena of real interest.

The foundations of at least one such formalism were laid with the notion that the recurrent collateral connections of subregion CA3 of the hippocampus allow it to function as an autoassociative memory (Rolls, 1989, although many of the ideas go back to Marr, 1971), and with subsequent quantitative analysis (reviewed in Treves and Rolls, 1994). After the laying of foundations, it is important to begin erecting a structural framework. In this context, this refers to the modelling of further features of the hippocampal system, in a parsimonious and simplistic way. Treves (1995) introduced a model of the Schaffer collaterals, the axonal projections which reach from the CA3 pyramidal cells into subregion CA1, forming a major part of the output from CA3 and of the input to CA1. The Schaffer collaterals can be seen clearly in Figure 1, a schematic drawing of the hippocampal formation. This paper introduced an information theoretic formalism similar to that of Nadal and Parga (1993) to the analysis. As will become apparent, this approach to network analysis appears to be particularly powerful, and is certain to find diverse application in the future.

Once the rudiments of a structural framework have been erected, it is possible to begin to add to the fabric of the theory – to begin to consider the effect of additional details of biology that were not in themselves necessary to its structural basis. This is where the contribution of the work described in this chapter lies. The analysis described in (Treves, 1995) assumed, for the purposes of simplicity of analysis, that the distribution of patterns of firing of CA3 pyramidal neurons was binary (and for one case ternary), although it considered threshold-linear (and thus analogue) model neurons. Here we shall consider in more detail the effect on information transmission of the possible graded nature of neuronal signalling. Another simple assumption made was that the pattern of convergence (the number of connections each CA1 neuron receives from CA3 neurons) of the Schaffer collaterals was either uniform, or alternatively bi-layered. The real situation is slightly more complex, and a better approximation of it is considered here.

2 A model of the Schaffer collaterals

The Schaffer collateral model describes, in a simplified form, the connections from the N CA3 pyramidal cells to the M CA1 pyramidal cells. Most Schaffer collateral axons project into the stratum radiatum of CA1, although CA3 neurons proximal to CA1 tend to project into the stratum oriens (Ishizuka et al., 1990); in the model these are assumed to have the same effect on the recipient pyramidal cells. Inhibitory interneurons are considered to act only as regulators of pyramidal cell activity. The perforant path synapses to CA1 cells are at this stage ignored (although they have been considered elsewhere; see Fulvi-Mari et al., this volume), as are the few CA1 recurrent collaterals. The system is considered for the purpose of analysis to operate in two distinct modes: storage and retrieval. During storage the Schaffer collateral synaptic efficacies are modified using a Hebbian rule reflecting the conjunction of pre- and post-synaptic activity. This modification has a slower time-constant than that governing neuronal activity, and thus does not affect the current CA1 output. During retrieval the Schaffer collaterals relay a pattern of neural firing with synaptic efficacies which reflect all previous storage events. In the following, the superscript S is used to indicate the storage phase, and R to indicate the retrieval phase.

- $\{\eta_i\}$ are the firing rates of each cell i of CA3. The probability density of finding a given firing pattern is taken to be:

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

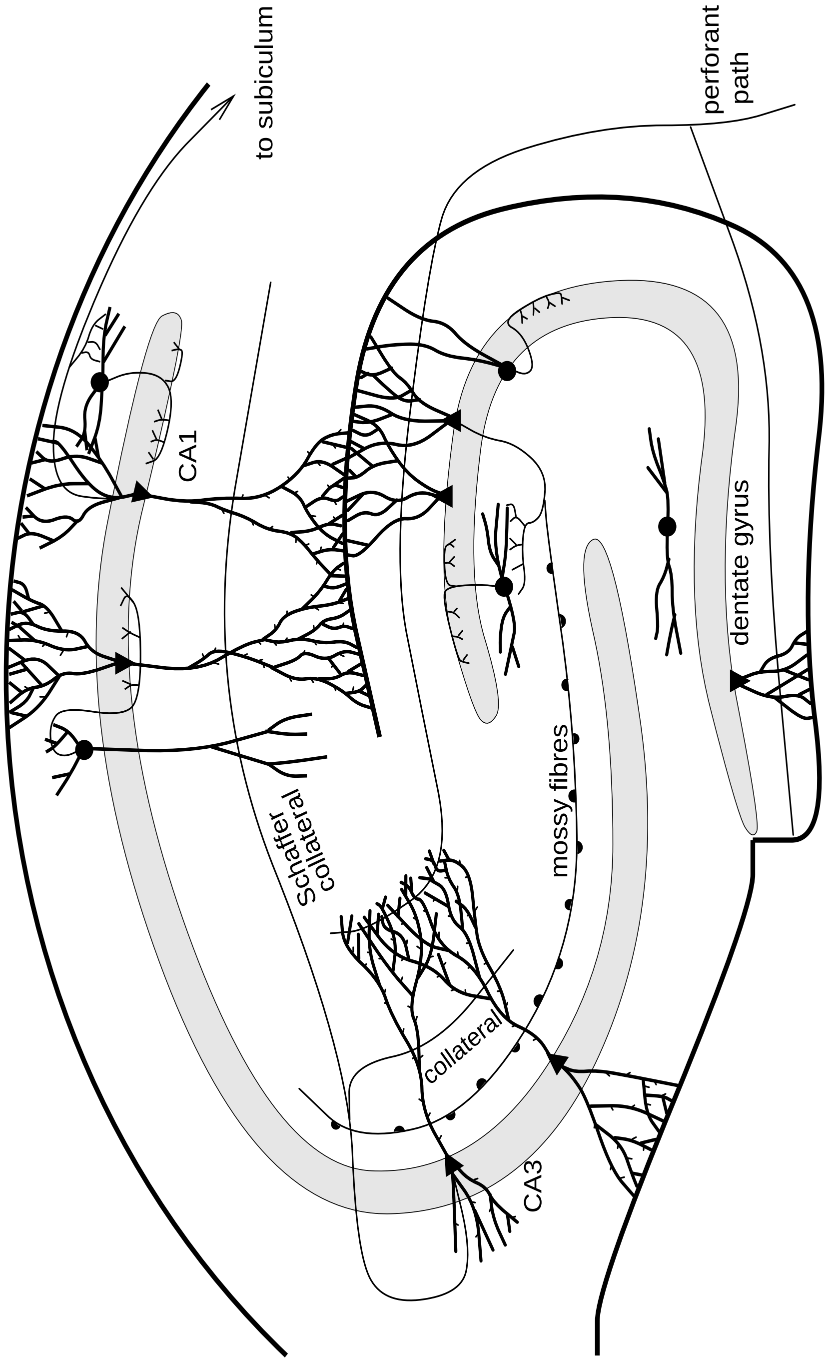

Figure 1: A schematic diagram of the hippocampal formation. Information enters the hippocampus from layer 2 entorhinal cells by the perforant path, which projects into dentate gyrus, CA3 and CA1 areas. In addition to its perforant path inputs, CA3 receives a lesser number of mossy fibre synapses from the dentate granule cells. The axons of the CA3 pyramidal cells project commissurally, recurrently within CA3, and also forward to area CA1 by the Schaffer collateral pathway. Information leaves the hippocampus via backprojections to the entorhinal cortex from CA1 and the subiculum, and also via the fornix to the mammillary bodies and anterior nucleus of the thalamus.

where η is the vector of the above firing rates. This assumption means that each cell in CA3 is taken to code for independent information, an idealised 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 rectification

$$\begin{aligned} V_i &= [\eta_i + \delta_i]^+ \\ \left\langle (\delta_i)^2 \right\rangle &= \sigma_\delta^2 \end{aligned} \tag{2}$$

(the rectifying function $[x]^+ = x$ for $x > 0$, and 0 otherwise, ensures that a firing rate is a positive quantity. This results in the probability density of V_i having a point component at zero equal to the sub-zero contribution, in addition to the smooth component). σ_δ can be related (e.g.) to interference effects 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.

- $\{\xi_j\}$ 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

$$\begin{aligned} \xi_j &= \left[\xi_0 + \sum_i c_{ij} J_{ij}^S \eta_i + \epsilon_j^S \right]^+ \\ \left\langle (\epsilon_j^S)^2 \right\rangle &= \sigma_{\epsilon^S}^2 \\ \left\langle (J_{ij}^S)^2 \right\rangle &= \sigma_J^2. \end{aligned} \tag{3}$$

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

$$\begin{aligned} c_{ij} &= \{0, 1\} \\ \langle c_{ij} \rangle N &= C_j \qquad (C \equiv \langle C_j \rangle) \end{aligned} \tag{4}$$

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

$$\begin{aligned} U_j &= \left[U_0 + \sum_i c_{ij} J_{ij}^R V_i + \epsilon_j^R \right]^+ \\ \left\langle (\epsilon_j^R)^2 \right\rangle &= \sigma_{\epsilon^R}^2 \\ \left\langle (J_{ij}^R)^2 \right\rangle &= \sigma_J^2 \end{aligned} \tag{5}$$

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

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

consists of

1. the original weight during storage, J_{ij}^S, damped by a factor $cos(\theta_\mu)$, where $0 < \theta_\mu < \pi/2$ parameterises the time elapsed between the storage and retrieval of pattern μ (μ is a shorthand for the pattern quadruplet $\{\eta_i, V_i, \xi_j, U_j\}$).

2. the modification due to the storage of μ itself, represented by a Hebbian term $H(\eta_i, \xi_j)$, normalised so that

$$\left\langle (H(\eta,\xi))^2 \right\rangle = \sigma_J^2; \tag{7}$$

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 normalised such that

$$\left\langle (J_{ij}^N)^2 \right\rangle = \sigma_J^2. \tag{8}$$

The mean value across all patterns of each synaptic weight is taken to be equal across synapses, and is therefore taken into the threshold term. The synaptic weights J_{ij}^R and J_{ij}^S are thus of zero mean, and variance σ_J^2 (all that affects the calculation is the first two moments of their distribution).

A plasticity model is used which corresponds to gradual decay of memory traces. Numbering memory patterns from $1, ..., \lambda, ..., \infty$ backwards, the model sets $\cos(\theta_\lambda) = exp(-\lambda\gamma_0/2)$ and $\gamma(\theta_\lambda) = \gamma_0 exp(-\lambda\gamma_0)$. Thus the strength of older memories fades exponentially with the number of intervening memories. The same forgetting model is assumed to apply to the CA3 network, and for this network, the maximum number of patterns can be stored when the plasticity $\gamma_0^{CA3} = 2/C$ (Treves, 1995).

For the Hebbian term the specific form

$$H(\eta_i, \xi_j) = \frac{h}{\sqrt{C}}(\xi_j - \xi_0)(\eta_i - \langle\eta_i\rangle) \tag{9}$$

is used, where h ensures the normalisation given in Eq. 7.

The thresholds ξ_0 and U_0 are assumed to be of fixed value in the following analysis. This need not be the case, however, and as far as the model represents (in a simplified fashion) the real hippocampus, they might be considered to be tuned to constrain the sparseness of activity in CA1 in the storage and retrieval phases of operation respectively, reflecting inhibitory control of neural activity.

The block diagram shown in Fig. 2 illustrates the relationships between the variables described in the preceding section.

3 Technical comments

The aim of the analysis is to calculate how much, on average, of the information present in the original pattern $\{\eta_i\}$ is still present in the effective output of the system, the pattern $\{U_j\}$, i.e. to average the mutual information

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

over the variables $c_{ij}, J_{ij}^S, J_{ij}^N$. The details of the calculation are unfortunately too extensive to present here, and the reader will have to be satisfied with an outline of the technique used. Those not familiar with replica calculations may refer to the final chapter of (Hertz et al., 1991), the appendices of (Rolls and Treves, 1997), or, less accessibly, the book by Mezard et al. (1987) for background material.

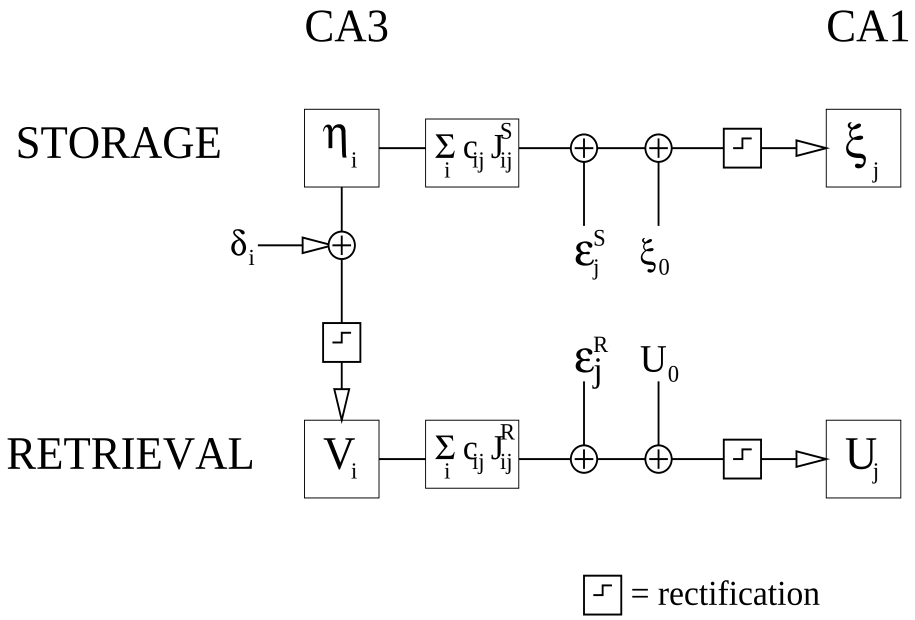

Figure 2: A block diagram illustrating the relationships between the variables present in the model. The input of the system could be considered to be the CA3 pattern during storage, η, and the output the CA1 pattern during retrieval, U. J^R_{ij} depends on J^S_{ij}, ξ_j and η_i as described in the text.

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

$$P(\eta, U) = P(U \mid \eta)P(\eta) = \int_V \int_\xi dV d\xi P(U \mid V, \xi, \eta)P(V \mid \eta)P(\xi \mid \eta)P(\eta) \quad (11)$$

where the probability densities implement the model defined above.

The average mutual information is evaluated using the replica trick, which amounts to

$$\log P = \lim_{n \to 0} \frac{(P^n - 1)}{n} \quad (12)$$

which involves a number of subtleties, for which (Mezard et al., 1987) can be consulted for a complete discussion. The important thing to note is that an assumption regarding replica symmetry is necessitated, and the stability of resulting solutions must be checked. This has been reported in (Schultz and Treves, 1997) for a simplified version of the neural network analysed here: a single layer of threshold-linear neurons with a single phase of operation (transmission) rather than the dual (storage and retrieval) modes in the model presented here. The conclusions of that study were that the replica-symmetric solution is stable for sufficiently sparse codes, but that for more distributed codes the solution was unstable below a critical noise variance. These conclusions can be assumed to carry across to the current model in at least a qualitative sense. In those regions (low noise, distributed codes) where the replica-symmetric solution is unstable, a solution with broken replica symmetry is required. It should be noted that it is not known what quantitative difference such a solution would bring: it may be very little, as is the case with the Hopfield network(Amit et al., 1987).

The expression for mutual information thus becomes

$$\langle i(\eta, U)\rangle_{c,J^S,J^N} = \lim_{n\to 0}\frac{1}{n}\left\langle \int d\eta dU P(\eta, U)\left\{\left[\frac{P(\eta, U)}{P(\eta)}\right]^n - [P(U)]^n\right\}\right\rangle_{c,J^S,J^N}. \tag{13}$$

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

The core of the calculation then is the calculation of the probability density $\left\langle P(\eta, U)^{n+1}\right\rangle$. The key to this is "self-consistent statistics" (Rolls and Treves, 1997, appendix 4): all possible values of each firing rate in the system are integrated, subject to a set of constraints that implement the model. The constraints are implemented using the integral form of the Dirac δ-function. Another set of Lagrange multipliers introduces macroscopic parameters

$$\begin{aligned}
x^{\alpha} &= \frac{1}{N}\sum_i \frac{(\eta_i^{\alpha} - \langle\eta\rangle)}{\langle\eta\rangle} V_i^{\alpha}\theta(V_i^{\alpha}) \\
w^{\alpha\beta} &= \frac{1}{N}\sum_i \eta_i^{\alpha} V_i^{\beta}\theta(V_i^{\beta}) \\
y^{\alpha\beta} &= \frac{1}{N}\sum_i V_i^{\alpha} V_i^{\beta}\theta(V_i^{\alpha})\theta(V_i^{\beta}) \\
z^{\alpha\beta} &= \frac{1}{N}\sum_i \eta_i^{\alpha}\eta_i^{\beta}
\end{aligned} \tag{14}$$

where $\theta(x)$ is the Heaviside function, and α, β are replica indices. Making the assumption of replica symmetry, and performing the integrals over all microscopic parameters, with some algebra an integral expression is obtained for the average mutual information per CA3 cell. This integral over the macroscopic parameters and their respective Lagrange multipliers is evaluated using the saddle-point approximation, which is exact in the limit of an infinite number of neurons (see, for example, Jeffreys and Jeffreys, 1972) to yield the expression given in Appendix A; the saddle-points of the expression must in general be found numerically.

4 How graded is information representation on the Schaffer collaterals?

Specification of the probability density $P(\eta)$ allows different distributions of firing rates in CA3 to be considered in the analysis. Clearly the distribution of firing rates that should be considered in the analysis is that of the firing of CA3 pyramidal cells, computed over the time-constant of storage (which we can assume to be the time-constant of LTP), during only those periods where biophysical conditions are appropriate for learning to occur. Unfortunately this last caveat makes a simple fit of the firing-rate distribution from single-unit recordings fairly meaningless unless the correct assumptions regarding exactly what these conditions are *in-vivo* can be made. It would be fair to assume that cholinergic modulatory activity is a pre-requisite, and unfortunately we cannot know directly from single-unit recordings from the hippocampus when the cells recorded from are receiving significant cholinergic modulation. Note that it might be possible to discover this indirectly. In any event, possibly the most useful thing we can do for the present is to assume that the distribution of firing rates during storage is graded, sparse, and exponentially tailed. This accords with the observations of neurophysiologists (Barnes et al., 1990; Rolls and Treves, 1997). The easiest way to introduce this to the current investigation is by means of a discrete approximation to the exponential distribution, with extra weight given to low firing rates. This allows quantitative investigation of the effects of analogue resolution on the information transmission capabilities of the Schaffer collateral model.

The required CA3 firing rate distributions were formed by the mixture of the unitary distribution and the discretised exponential, using as mixture parameters the offset ϵ between their origins, and relative weightings. The distributions were constrained to have first and second moments $\langle\eta\rangle$, $\langle\eta^2\rangle$, and thus sparseness $\langle\eta\rangle^2 / \langle\eta^2\rangle$, equal to a. In the cases considered here, a was allowed values of 0.05, 0.10 and 0.20 only. The width of the distribution examined was set to 3.0, and the number of discretised firing levels contained in this width parameterised as l. The binary distribution was completely specified by this; for distributions with a large number of levels, there was some degree of freedom, but its numerical effect on the resulting distributions was essentially negligible. Those distributions with a small number of levels ≥ 2 were non-unique, and were chosen fairly arbitrarily for the following results, as those that had entropies interpolating between the binary and large l situations. Some examples of the distributions used are shown in Fig. 3a.

The total entropy per cell of the CA3 firing pattern, given a probability distribution characterised by L levels, is

$$h(\eta) = -\sum_{l=1}^{L} P_{\eta_l}(\eta_l) \ln P_{\eta_l}(\eta_l). \tag{15}$$

The results are shown in Fig. 3b–d. The entropy present in the CA3 firing rate distributions is marked by asterisks. The mutual information conveyed by the retrieved pattern of CA1 firing rates, which must be strictly less than the CA3 entropy, is represented by circles. It is apparent that maximum information efficiency occurs in the binary limit. More remarkably, even in absolute terms the information conveyed is maximal for low resolution codes, at least for quite sparse codes. The results are qualitatively consistent over sparsenesses a ranging from 0.05 to 0.2; obviously with higher a (more distributed codes), entropies are greater. For more distributed codes (i.e. with signalling more evenly distributed over neuronal firing rates), it appears that there may be some small absolute increase in information with the use of analogue signalling levels.

For comparison, the crosses in the figures show the information stored in CA1. This was computed using a simpler version of the calculation, in which the mutual information $i(\{\eta_i\}, \{\xi_j\})$ was calculated. Obviously, in this calculation, the CA3 and CA1 retrieval noises σ_δ and σ_ϵ^R are not present; on the other hand, neither is the Schaffer collateral memory term. Since the retrieved CA1 information is in every case higher than that stored, we can conclude that for the parameters considered, the additional Schaffer memory effect outweighs the deleterious effects of the retrieval noise distributions.

It follows from the forgetting model defined by Eq. 6, that information transmission is maximal when the plasticity (mean square contribution of the modification induced by one pattern) is matched in the CA3 recurrent collaterals and the Schaffer collaterals (Treves, 1995). It can be seen in Fig. 3e that this effect is robust to the use of more distributed patterns.

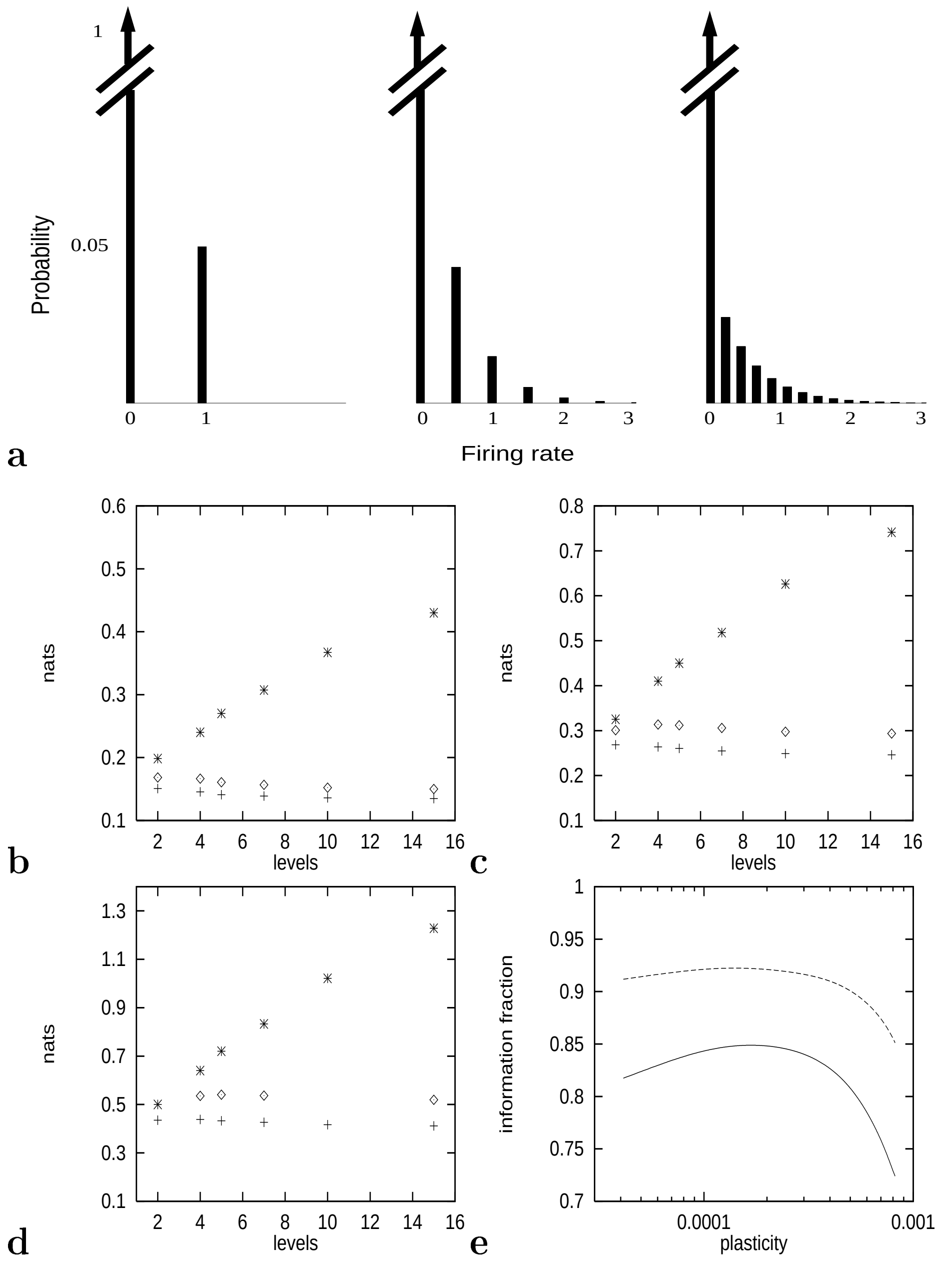

Figure 3: **a** Some of the CA3 firing rate distributions used in the analysis. These are, in general, formed by the mixture of a unitary distribution and a discretised exponential. **b** – **d** The mutual information between patterns of firing in CA1 and patterns of firing in CA3, expressed in natural units (nats). Asterisks represent the entropy of the CA3 pattern distribution, diamonds the CA1 retrieved mutual information, and crosses the CA1 information during the storage phase. The horizontal axis parameterises the number of discrete levels in the input distribution: for codes with fine analogue resolution, this is greater. **b** is for $a = 0.05$ (sparse), **c** for $a = 0.10$, and **d** for $a = 0.20$ (slightly more distributed). **e** The dependence of information transmission on the degree of plasticity in the Schaffer collaterals, for $a = 0.05$ (solid) and $a = 0.10$ (dashed). A binary pattern distribution was used in this case.

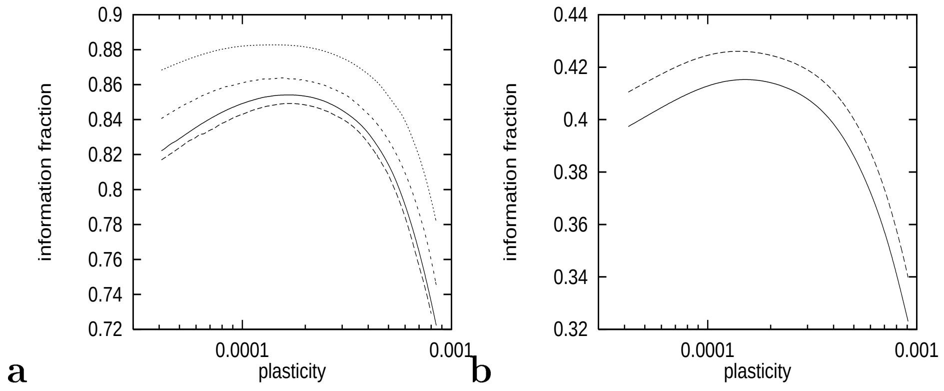

Figure 4: Information transmitted as a function of Schaffer collateral plasticity. **a** Binary CA3 firing rate distributions. The solid line indicates the result for the realistic convergence model. The dashed lines indicate, in ascending order: (i) uniform convergence, (ii) two-tier convergence model with $C_j \in \{5000, 15000\}$, (iii) two-tier convergence model with $C_j \in \{2000, 18000\}$. **b** With more realistic CA3 firing-rate distributions (the 10-level discrete exponential approximation from the previous section). The solid line indicates the result for uniform connectivity, and the dashed line the two-tier convergence model with $C_j \in \{5000, 15000\}$.

5 Non-uniform Convergence

It is assumed in (Treves, 1995) that there is uniform convergence of connections from CA3 to CA1 across the extent of the CA1 subfield. In reality, each CA1 pyramidal neuron does not receive the same number of connections from CA3: this quantity varies across the transverse extent of CA1 (although this transverse variance may be less than that within CA3; Amaral et al. 1990). Bernard and Wheal (1994) investigated this with a connectivity model constructed by simulating a *Phaseolus vulgaris* leucoagglutinin labelling experiment, matched to the available anatomical data. Their conclusion was that mid CA1 neurons receive more connections (8000) than those in proximal and distal CA1 (6500). The precise numbers are not important here; what *is* of interest is to consider the effect on information transmission of this spread in the convergence parameter C_j about its mean C.

In this analysis σ_J^2 is set to $1/C$ for all cells in the network. C is set using the assumption of parabolic dependence of C_j upon transverse extent, on the basis of Fig. 5 of (Bernard and Wheal, 1994). In order to facilitate comparison with the results reported in Treves (1995), C is held at 10,000 for all results in this section. The model used (which we will refer to as the 'realistic convergence' model) is thus simply a scaled version of that due to Bernard and Wheal, with $C_j = 7,143$ at the proximal and distal edges of CA1, and $C_j = 11,429$ at the midpoint. Note that this refers to the number of CA3 cells contacting each CA1 cell; each may do so via more than one synapse.

The saddle-point expression (16) was evaluated numerically while varying the plasticity of the Schaffer connections, to give the relationships shown in Fig. 4a between mutual information and γ_0^{CA1}. The information is expressed in the figure as a fraction of the information present when the pattern is stored in CA3 (15).

Two phenomena can be seen in the results. The first, as mentioned in the previous section (and discussed at more length in Treves, 1995), is that information transmission is maximal when the plasticity of the Schaffer collaterals is approximately matched with that of the preceding stage of information processing. The second phenomenon is the increase in information throughput with spread in the convergence about its mean. This is an effect which is not immediately intuitive: it means that the increase in mutual information provided by those CA1 neurons with a greater number of connections than the mean more than compensates for the decrease in those with less than the mean. It must be remembered that what is being computed is the information provided by *all* CA1 cells about patterns of activity in CA3. This increase in information is a network effect that has no counterpart in the information a single CA1 cell could convey. In any case, the effect is rather small: the realistic convergence model allows the transmission of only marginally more information than the uniform model. The uniform convergence approximation might be viewed as a reasonable one for future analyses, then.

Fig. 4b shows that the situation for graded pattern distributions is almost identical. The numerical fraction of information transmitted is of course lower (but total transmitted information is similar – see previous section). The uniform and two-tier convergence models provide bounds between which the realistic case must lie.

6 Discussion and summary

This chapter examined quantitatively the effect of analogue coding resolution on the total amount of information that can be transmitted in a model of the Schaffer collaterals. The tools used were analytical and numerical, and the focus was upon relatively sparse codes. What can these results tell us about the actual code used to signal information in the mammalian hippocampus? In themselves, of course, they can make no definite statement. It could be that there is a very clear maximum for information transmission in using binary codes for the Schaffer collaterals, and yet external constraints, such as CA1 efferent processing, might make it more optimal overall to use analogue signalling. So results from a single component study must be viewed with duep caution. However, these results can provide a clear picture of the operating regime of the Schaffer collaterals, and that is after all a major aim of any analytical study.

The results from this paper reiterate some previously known points, and bring out others. For instance, it is very clear from Fig. 3 that, while nearly all of the information in the CA3 distribution can be transmitted using a binary code, this information fraction drops off rapidly with analogue level. The total amount of information transmitted is similar regardless of the amount of analogue level to be signalled – but this is a well known and relatively general fact, and accords with common sense intuition. However, the total amount of information that can be transmitted is only *roughly* constant. It appears, from this analysis, that while the total transmitted information drops off slightly with analogue level for very sparse codes, the maximum moves in the direction of more analogue levels for more evenly distributed codes. This provides some impetus for making more precise measurements of sparseness of coding in the hippocampus.

Another issue which this model allows us to address is the expansion ratio of the Schaffer collaterals, i.e. the ratio between the numbers of neurons in CA1 and CA3, M/N. It can be seen in Fig. 5 that an expansion ratio of 2 (a 'typical' biological value) is sufficient for CA1 to capture most of the information of CA3, and that while the gains for increasing this are diminishing, there is a rapid drop-off in information transmission if it is reduced by any significant

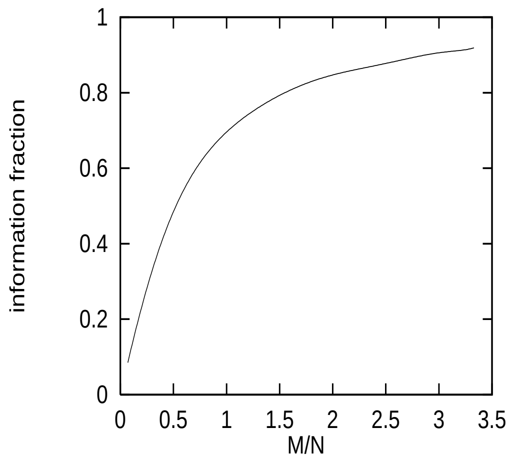

Figure 5: The dependence of information transmission on the expansion ratio $r_{CA1,CA3} = M/N$.

amount. The actual expansion ratio for different mammalian species reported in the literature is subject to some variation, with the method of nuclear cell counts giving ratios between 1.4 (Long Evans rats) and 2.0 (humans) (Seress, 1988), while stereological estimates range from 1.8 (Wistar rats) to 6.1 (humans) (West, 1990). It should be noted that in all these estimates, and particularly with larger brains, there is considerable error (L. Seress, personal communication). However, in all cases the Schaffer collateral model appears to operate in a regime in which there is at least the scope for efficient transfer of information.

Clearly it is essential to further constrain the model by fitting the parameters as sufficient neurophysiological data becomes available. As more parameters assume biologically measured values, the sensible ranges of values that as-yet unmeasured parameters can take will become clearer. It will then be possible to address further issues such as the quantitative importance of the constraint upon dendritic length (i.e. the number of synapses per neuron) upon information processing.

In summary, we have used techniques for the analysis of neural networks to quantitatively investigate the effect of a number of biological issues on information transmission by the Schaffer collaterals. We envisage that these techniques, developed further and applied in a wider context to networks in the medial temporal lobe, will yield considerable insight into the organisation of the mammalian hippocampal formation.

Appendix A. Expression from the Replica Evaluation

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

$$
\begin{aligned}
&-\frac{N}{2}(y_B\tilde{y}_B + 2w_B\tilde{w}_B + z_B\tilde{z}_B)\\
&+N\int D\tilde{s}_1 D\tilde{s}_2 \left\langle F(\tilde{s}_1,\tilde{s}_2,\eta,\tilde{y}_B,\tilde{w}_B,\tilde{z}_B)\right\rangle_\eta\\
&\times \ln\left\langle F(\tilde{s}_1,\tilde{s}_2,\eta,\tilde{y}_B,\tilde{w}_B,\tilde{z}_B)\right\rangle_\eta \Bigg\}
\end{aligned} \tag{16}
$$

where taking the extremum means evaluating each of the two terms, separately, at a saddle-point over the variables indicated. 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

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

and has to be averaged over P_η and over the Gaussian variables of zero mean and unit variance $\tilde{s}_1, \tilde{s}_2$.

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

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

$$
\tilde{s}_\pm = \sum_{k=1}^{2}(\mp 1)^{(k-1)}\sqrt{\frac{\left[\sqrt{(\tilde{y}-\tilde{z})^2+4\tilde{w}^2}\mp(-1)^k(\tilde{y}-\tilde{z})\right](\tilde{y}\tilde{z}-\tilde{w}^2)}{\left[\tilde{y}+\tilde{z}+(-1)^k\sqrt{(\tilde{y}-\tilde{z})^2+4\tilde{w}^2}\right]\sqrt{(\tilde{y}-\tilde{z})^2+4\tilde{w}^2}}}\,\tilde{s}_k \tag{19}
$$

in the last two lines of Eq. 16, but 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

$$
\begin{aligned}
\Gamma(y,w,z,C_j,\gamma) &= \int \frac{ds_1 ds_2}{2\pi\sqrt{\det \mathbf{T}'_j}}\exp -\begin{pmatrix} s_1 & s_2 \end{pmatrix}\frac{(\mathbf{T}'_j)^{-1}}{2}\begin{pmatrix} s1 \\ s2 \end{pmatrix}\\
&\times\left[\int_{-\infty}^{0} dU G(U)\ln\int_{-\infty}^{0} dU' G(U')\right.\\
&\left. + \int_{0}^{\infty} dU G(U)\ln G(U)\right],
\end{aligned} \tag{20}
$$

where

$$
\begin{aligned}
G(U) &= G(U;s_1,s_2,y,w,z,C_j,\gamma)\\
&= \phi\left[\frac{(\xi_0-s_2)(T_{yj}+2g_jT_{wj}+g_j^2T_{zj})+(U-U_0+s_1+g_js_2)(T_{wj}+g_jT_{zj})}{\sqrt{(T_{yj}T_{z_j}-T_{wj}^2)(T_{yj}+2g_jT_{wj}+g_j^2T_{zj})}}\right]\\
&\times\frac{1}{\sqrt{2\pi(T_{yj}+2g_jT_{wj}+g_j^2T_{zj})}}\exp -\frac{(U-U_0+s_1+g_js_2)^2}{2(T_{yj}+2g_jT_{wj}+g_j^2T_{zj})}
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad \phi \left[\frac{-(\xi_0 - s_2)T_{yj} - (U - U_0 + s_1 + g_j\xi_0)T_{wj}}{\sqrt{(T_{yj}T_{z_j} - T_{wj}^2)T_{yj}}} \right] \\
&\quad \times \frac{1}{\sqrt{2\pi T_{yj}}} \exp - \frac{(U - U_0 + s_1 + g_j\xi_0)^2}{2T_{yj}}, \qquad (21)
\end{aligned}
$$

and

$$
\begin{aligned}
T_{yj} &= \sigma_{\epsilon^R}^2 + \sigma_J^2 C_j (y^0 - y) \\
T_{wj} &= \sigma_J^2 C_j (w^0 - w) \cos(\theta) \\
T_{zj} &= \sigma_{\epsilon^S}^2 + \sigma_J^2 C_j (z^0 - z) \qquad (22) \\
\mathbf{T}_j' &= \sigma_J^2 C_j \begin{pmatrix} y & w\cos(\theta) \\ w\cos(\theta) & z \end{pmatrix}
\end{aligned}
$$

are effective noise terms.

$$
g_j = h \frac{C_j}{C} x^0 \left\langle \eta \right\rangle_\eta \sqrt{C\gamma(\theta)}, \qquad (23)
$$

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

$$
\begin{aligned}
x^0 &= \frac{1}{N} \sum_i \left\langle \frac{(\eta_i - \langle \eta \rangle_\eta)}{\langle \eta \rangle_\eta} V_i \right\rangle \\
&= \left\langle \frac{(\eta - \langle \eta \rangle_\eta)}{\langle \eta \rangle_\eta} \left[\eta \phi \left(\frac{\eta}{\sigma_\delta} \right) + \frac{\sigma_\delta}{\sqrt{2\pi}} \exp - \frac{1}{2} \left(\frac{\eta}{\sigma_\delta} \right)^2 \right] \right\rangle_\eta \\
y^0 &= \frac{1}{N} \sum_i \langle V_i^2 \rangle = \left\langle \left[\sigma_\delta^2 + \eta^2 \right] \phi \left(\frac{\eta}{\sigma_\delta} \right) + \frac{\eta \sigma_\delta}{\sqrt{2\pi}} \exp - \frac{1}{2} \left(\frac{\eta}{\sigma_\delta} \right)^2 \right\rangle_\eta \\
w^0 &= \frac{1}{N} \sum_i \langle \eta_i V_i \rangle = \left\langle \eta \left[\eta \phi \left(\frac{\eta}{\sigma_\delta} \right) + \frac{\sigma_\delta}{\sqrt{2\pi}} \exp - \frac{1}{2} \left(\frac{\eta}{\sigma_\delta} \right)^2 \right] \right\rangle_\eta \\
z^0 &= \frac{1}{N} \sum_i \eta_i^2 = \langle \eta^2 \rangle_\eta . \qquad (24)
\end{aligned}
$$

7 Appendix B. Parameter Values

Parameters used were, except where otherwise indicated in the text:

σ_δ	0.30
σ_ϵ^S	0.20
σ_ϵ^R	0.20
C	10000
σ_J^2	0.0001
ξ_0	-0.4
U_0	-0.4
M/N	2.0

References

Amaral, D. G., Ishizuka, N. and Claiborne, B. (1990). Neurons, numbers and the hippocampal network, *in* J. Storm-Mathisen, J. Zimmer and O. P. Ottersen (eds), *Understanding the brain through the hippocampus*, Vol. 83 of *Progress in Brain Research*, Elsevier Science, chapter 17.

Amit, D., Gutfreund, H. and Sompolinsky, H. (1987). Statistical mechanics of neural networks near saturation, *Ann. Phys. (N.Y.)* **173**: 30–67.

Barnes, C. A., McNaughton, B. L., Mizumori, S. J., Leonard, B. W. and Lin, L. H. (1990). Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing, *Prog. Brain Res.* **83**: 287–300.

Bernard, C. and Wheal, H. V. (1994). Model of local connectivity patterns in CA3 and CA1 areas of the hippocampus, *Hippocampus* **4**(5): 497–529.

Cohen, N. J. and Eichenbaum, H. (1993). *Memory, Amnesia and the Hippocampal System*, MIT Press, Cambridge, MA.

Fulvi-Mari, C., Panzeri, S., Rolls, E. T. and Treves, A. (1996). A quantitative model of information processing in CA1, *Abstracts of the Information Theory and the Brain II conference.*

Gaffan, D. (1992). The role of the hippocampus-fornix-mamillary system in episodic memory, *in* L. R. Squire and N. Butters (eds), *Neuropsychology of Memory*, Guilford, New York, pp. 336–346.

Gardner, E. (1988). The space of interactions in neural network models, *J. Phys. A: Math. Gen.* **21**: 257–270.

Hertz, J., Krogh, A. and Palmer, R. G. (1991). *Introduction to the theory of neural computation*, Addison-Wesley, Wokingham, U.K.

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

Jeffreys, H. and Jeffreys, B. S. (1972). *Methods of Mathematical Physics*, third edn, Cambridge University Press, Cambridge, U.K.

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

McNaughton, B. L. and Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system, *Trends Neurosci.* **10**: 408–415.

Mezard, M., Parisi, G. and Virasoro, M. (1987). *Spin glass theory and beyond*, World Scientific, Singapore.

Nadal, J.-P. and Parga, N. (1993). Information processing by a perceptron in an unsupervised learning task, *Network* **4**: 295–312.

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

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

Rolls, E. T. and Treves, A. (1997). *Neural networks and brain function*, Oxford University Press, Oxford, U.K.

Schultz, S. and Treves, A. (1997). Stability of the replica-symmetric solution for the information conveyed by a neural network, *submitted* .

Scoville, W. B. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions, *J. Neurol. Neurosurg. Psychiatry* **20**: 11–21.

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.* **3**: 335–340.

Stephan, H. (1983). Evolutionary trends in limbic structures, *Neuroscience and Biobehavioral Reviews* **7**: 367–374.

Treves, A. (1990). Threshold-linear formal neurons in auto-associative nets, *J. Phys. A: Math. Gen.* **23**: 2631–2650.

Treves, A. (1995). Quantitative estimate of the information relayed by the Schaffer collaterals, *J. Comput. Neurosci.* **2**: 259–272.

Treves, A., Barnes, C. A. and Rolls, E. T. (1996). Quantitative analysis of network models and of hippocampal data, *in* T. Ono, B. L. McNaughton, S. Molotchnikoff, E. T. Rolls and H. Nishijo (eds), *Perception, Memory and Emotion: Frontier in Neuroscience*, Elsevier, Amsterdam, chapter 37, pp. 567–579.

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

Treves, A. and Rolls, E. T. (1994). A computational analysis of the role of the hippocampus in learning and memory, *Hippocampus* **4**: 373–391.

Tsodyks, M. V. and Feigelman, M. V. (1988). The enhanced storage capacity in neural networks with low activity level, *Europhys. Lett.* **6**(2): 101–105.

Weiskrantz, L. (1987). Neuroanatomy of memory and amnesia: a case for multiple memory systems, *Human Neurobiology* **6**: 93–105.

West, M. J. (1990). Stereological studies of the hippocampus: a comparison of the hippocampal subdivisions of diverse species including hedgehogs, laboratory rodents, wild mice and men, *in* J. Storm-Mathisen, J. Zimmer and O. P. Ottersen (eds), *Progress in Brain Research*, Vol. 83, Elsevier Science, chapter 2, pp. 13–36.

Witter, M. P. and Groenewegen, H. J. (1992). Organizational principles of hippocampal connections, *in* M. R. Trimble and T. G. Bolwig (eds), *The Temporal Lobes and the Limbic System*, Wrightson Biomedical, Petersfield, U.K., chapter 3, pp. 37–60.