

# Computational Constraints Between Retrieving the Past and Predicting the Future, and the CA3-CA1 Differentiation

Alessandro Treves\*

**ABSTRACT:** The differentiation between the CA3 and CA1 fields of the mammalian hippocampus is one of the salient traits that set it apart from the organization of the homologue medial wall in reptiles and birds. CA3 is widely thought to function as an autoassociator, but what do we need CA1 for? Based on evidence for a specific role of CA1 in temporal processing, I have explored the hypothesis that the differentiation between CA3 and CA1 may help solve a computational conflict. The conflict is between pattern completion, or integrating current sensory information on the basis of memory, and prediction, or moving from one pattern to the next in a stored sequence. CA3 would take care of the former, while CA1 would concentrate on the latter. I have found the hypothesis to be only weakly supported by neural network simulations. The conflict indeed exists, but two mechanisms that would relate more directly to a functional CA3-CA1 differentiation were found unable to produce genuine prediction. Instead, a simple mechanism based on firing frequency adaptation in pyramidal cells was found to be sufficient for prediction, with the degree of adaptation as the crucial parameter balancing retrieval with prediction. The differentiation between the architectures of CA3 and CA1 has a minor but significant, and positive, effect on this balance. In particular, for a fixed anticipatory interval in the model, it increases significantly the information content of hippocampal outputs. There may therefore be just a simple quantitative advantage in differentiating the connectivity of the two fields. Moreover, different degrees of adaptation in CA3 and CA1 cells were not found to lead to better performance, further undermining the notion of a functional dissociation. © 2004 Wiley-Liss, Inc.

**KEY WORDS:** Schaffer collateral; associative memory; spatial localization; recurrent network; mossy fibers; information measures

## INTRODUCTION

The hippocampus develops out of the medial pallium, the medial portion of the superior wall of each cerebral hemisphere. Reptilians and birds, presumed in this respect to be closer than modern mammals to their common ancestors, present a simple cortical organization of the hippocampus. The hippocampus in reptilians includes a dorsomedial band with larger cells, and a more medial band with smaller cells, particularly rich in zinc (Ulinski, 1990). In mammals, the medial band detaches from the rest of the pallium (Gloor, 1997), forming the dentate gyrus (DG), whose granule cells no longer project to the dorsolateral cortex, but only to the cornu ammonis, the structure derived from the folding of the dorsomedial band (Braitenberg and

Schüz, 1991). The cornu ammonis differentiates into two main fields, CA3 and CA1, which are distinct in a number of features, and most prominently in their connectivity. CA3 pyramidal cells receive sparse but powerful synapses from the mossy fibers (MF), the axons of the DG granule cells. They also receive extensive direct perforant path (PP) projections from entorhinal cortex (EC), the cortical input station to the hippocampus. These PP fibers, arising mainly from EC layer II, are also the main source of inputs to DG granule cells. Therefore, PP inputs to CA3 cells, on the order of 3,600 per cell in the rat (Amaral et al., 1990), carry, qualitatively, cortical information of the same nature as MF inputs; however, they have distinct properties (Ji and Staubli, 2002) and are much fewer (on the order of 50 per cell in the rat). Most synapses onto CA3 pyramidal cells, though, are from the recurrent collaterals (RC) of the axons of the CA3 cells themselves, in what comprises one of the clearest recurrent systems in the brain, i.e., roughly 12,000 per cell in the rat (Amaral et al., 1990). In contrast, CA1 cells do not receive MF inputs, and their PP inputs mainly arise from EC layer III (Amaral and Witter, 1995). In addition, rather than extensive recurrent collaterals, they receive extensive Schaffer collaterals, which are branches of the same CA3 axons that give rise to the recurrent system in CA3. CA1 pyramidal cells are also typically smaller and less prone to bursting (Traub et al., 1991), and may differ in a number of other details, e.g., the inhibitory network that regulates their activity (Freund and Buzsáki, 1996; Morishita and Alger, 2000).

## WHAT WAS THE DIFFERENTIATION BETWEEN CA3 AND CA1 MEANT TO ACHIEVE?

As a first, and possibly oversimplified, approximation, we could schematize the transition from the undifferentiated (or “uniform”) to differentiated cornu ammonis, as including two main, perhaps simultaneous, steps: (1) MF “backprojections” from DG, earlier extending throughout CA (or even to dorsolateral cortex, as bona fide backprojections; Ulinski, 1990), are now restricted to CA3; and (2) the local CA collaterals, which earlier constituted an extensive recurrent network, now originate solely from CA3 cells, whereas the axons of CA1 cells leave CA after minimal collateral branching. This is the scheme I

SISSA, Cognitive Neuroscience Sector, Trieste, Italy; and NTNU, Centre for the Biology of Memory, Trondheim, Norway

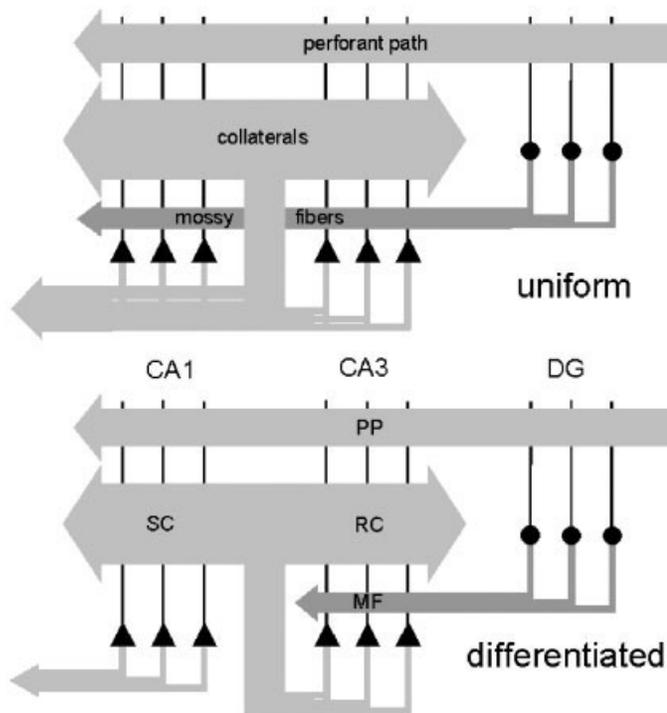
Grant sponsor: Human Frontier; Grant number: RGB001-98; Grant number: ST00080/2000-B.

\*Correspondence to: Alessandro Treves, SISSA, Neuroscience, via Beirut 4, I-34014 Trieste, Italy. E-mail: ale@sisssa.it

Accepted for publication 4 July 2003

DOI 10.1002/hipo.10187

Published online 13 February 2004 in Wiley InterScience (www.interscience.wiley.com).



**FIGURE 1.** Scheme of connectivity of the differentiated CA3-CA1 network, and of the “uniform” network used for comparison.

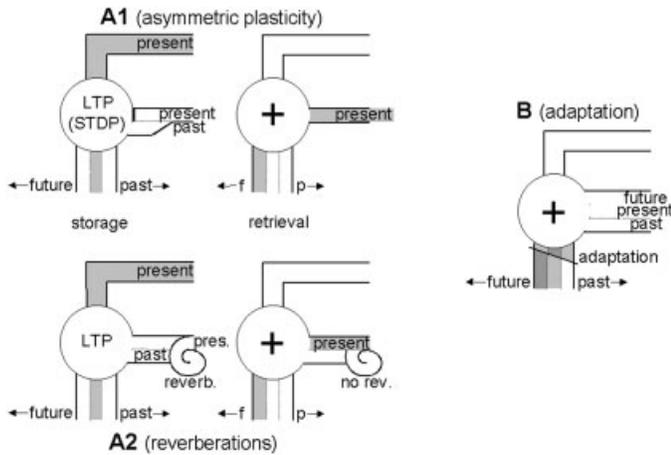
shall adopt in the neural network simulations presented here (Fig. 1).

What is the functional significance of such a differentiation? In earlier work with Edmund Rolls (Treves and Rolls, 1992), I had argued, within the framework of the ideas formulated by Marr (1971) and others (McNaughton and Morris, 1987; Rolls, 1989), that the contribution of the “duplicated” cortical input to CA3, relayed via the MF system, was in establishing new information-rich representations. The resulting prediction, that MF inactivation should disrupt new hippocampal learning, but not retention, is consistent with recent behavioral evidence (Lassalle et al., 2000). However, this “DG” functional hypothesis does not explain the role of the differentiation among CA fields. In subsequent work based purely on the mathematical analysis of network models (Treves, 1995; Schultz et al., 2000), we had noted how CA1 outputs, merely by being cascaded through one more associative synaptic matrix than CA3 outputs, could be richer in information, especially if SC inputs to CA1 are integrated with direct PP inputs (Fulvi Mari et al., 2000). This simple quantitative advantage in adding one more processing step does not appear to be a compelling explanation for the differentiation, however. Moreover, it was derived from very oversimplified models, of the type that could be treated analytically. Further, the analysis essentially compared a CA3 standalone model with a model with CA1 added to CA3, rather than comparing a CA3-CA1 uniform model with a differentiated one. Since the evolutionary “cost” of adding more elements (neurons and synapses) cannot be assessed in any reasonable way, it seems better to stick to comparisons between networks, uniform and differentiated, with the same number of components.

Indications of a more interesting qualitative distinction of roles between CA3 and CA1 arise from lesion studies in rats, reviewed by Kesner et al. (2002b). As summarized in the table at the end of their review, field CA3 is said to be critical for spatial pattern association and completion, while field CA1 for temporal pattern separation and completion. Nevertheless, an analysis of the Kesner review indicates that the table at the end is a well-meaning simplification. Their Figure 31.2 suggests that CA3 may be involved in temporal pattern separation just as much as CA1. Moreover, the role of either DG or CA3 in temporal pattern association has never really been assessed. Further, available studies on the role of CA1 fail to make a clear distinction between tasks in which massive hippocampal outputs to the cortex are crucial, and tasks in which a more limited hippocampal influence on the cortex may be sufficient. In the first case, lesioning CA1 should have an effect independent of the specific contribution of CA1 to information processing, simply because one is severing the main hippocampocortical output pathway. In the second case, CA3 outputs through the fimbria/fornix could enable hippocampal-mediated influences to be felt, deprived, though, of the specific CA1 contribution. Finally, these behavioral/selective-lesion findings have largely been confined to the rat (but see, e.g., Gaffan et al., 2001), and one wonders how they could be replicated with primates and other species.

Temporal pattern association can lead to the capability to predict future events. The neuronal mechanisms that might subservise this capability have been discussed by Levy and coworkers in an extensive series of modeling studies (Levy, 1989, 1990; Minai et al., 1994; Levy et al., 1995). The central result of these studies is that a recurrent network like CA3 can be used to predict the next firing pattern in a previously stored sequence, can hold multiple sequences concurrently in storage, and can disambiguate among overlapping sequences. These models therefore suggest that CA3 alone may be sufficient for rather complex tasks of temporal pattern separation and completion. What remains unclear is whether this CA3 capability would be augmented, unaffected, or perhaps even undermined, by a subsequent feedforward CA1 network.

The possibility that the advantage of differentiating CA1 from CA3 has to do with the temporal dimension may fit with a specific hypothesis about the mechanisms for temporally predictive encoding, as manifested in the theta-phase precession of place fields (Mehta, 2001). According to a quite general hypothesis, one can think of the “original” place field of a CA1 cell as the one resulting from the integration of cortical inputs. This place field is expanded by associative synaptic modification, particularly in the RC and SC systems, to produce additional firing by the cell, also outside the original field. Because the firing outside the field is typically driven by weaker inputs, it tends to occur later in the theta period, once the inhibitory wave has reached its minimum. At this point in the reasoning, the generic hypothesis bifurcates into several specific hypotheses. Some of these (Fig. 2A; see Levy, 1989, for an early discussion) maintain that Schaffer collateral plasticity essentially associates postsynaptic activity related to present experience (driven mainly by PP inputs) with presynaptic activity related to the recent past, e.g., a few tens of ms ago, coming from CA3. This may come about with a spike-timing-dependent plasticity (STDP) rule (Bi and Poo, 1998)—hypothesis A1—or alternatively by as-



**FIGURE 2.** Scheme of some of the mechanisms that may underlie a predictive ability in the hippocampus. Mechanisms A both rely on a time shift between the representations associated together by CA1 synaptic plasticity: in variant A1 the shift results from the asymmetry in the plasticity rule, in variant A2 from reverberations in CA3. Mechanism B is based solely on generically associative plasticity and on firing frequency adaptation.

sociating with CA1 activity CA3 activity which has been reverberated and delayed along recurrent loops—hypothesis A2 in Figure 2. With the reverberations of the latter hypothesis, there is no strong constraint on the time span of the temporal association, which might last for seconds, as in the experiments by Huerta et al. (2000). Once presynaptic activity is “replayed” without delay, it can evoke postsynaptic activity that codes for future events (Fig. 2). The asymmetry between past and future implicit in either the STDP rule or the reverberation results in the asymmetric expansion of the field (Mehta et al., 1997, 2000), which extends mainly to locations visited before the original field, and in a gradual precession of location specific firing with theta phase (O’Keefe and Recce, 1993; Skaggs et al., 1996). Alternative and simpler specific hypotheses (Fig. 2B) maintain that the place field expansion is basically symmetric between past and future but, since the additional firing after the original field is suppressed by neuronal and/or synaptic fatigue, what is more evident is the additional firing before the field, resulting in an apparent asymmetric expansion of the field and, on average, in the gradual precession with theta phase. The overall result, in both cases A and B, is predictive encoding: from neural activity late in the theta phase one can decode where the rat will be, a few tens of ms after its current location, while the current location is better coded, instead, by activity early in the theta phase. Interestingly, neither the specific mechanism A nor B has much to do with the theta rhythm, although they may be manifested in the theta precession phenomenon. In species such as primates, which do not seem to show evident theta rhythm (but see Jensen and Tesche, 2002), hippocampal encoding may still become predictive of future experience, through associative learning, with either mechanism A or B (Levy et al., 1990).

If mechanism A2 applies, the one relying on delays due to collateral reverberation, this may account for the advantage of differentiating CA3, where the reverberation occurs, from CA1,

where encoding should be predictive. Such a line of reasoning would be consistent with the indications from the lesion data, and would predict that neural activity in CA1, at least in some conditions, is more predictive than in CA3. A difference in theta precession along these lines has not been reported, but this may be due to the activity having been recorded in conditions inappropriate to reveal a difference, e.g., along unidirectional tracks rather than genuine 2D environments (there are limited 2D precession data for CA3). If mechanisms A1 or B apply, the potential advantage of the differentiation is less apparent, but it might still be present. The parameters that optimize memory retrieval will in general be different from those that optimize prediction in time, and it may well be that connectivity parameters be among those that differ. CA3 connectivity may be closer to optimal for memory retrieval, while CA1 may be closer to an optimal predictive network.

To explore this issue and help discriminate between the possible mechanisms A1, A2, and B, all consistent with the available evidence, it appears that the most direct approach is to use neural network simulations. In fact, the analytical approaches that I have developed in the past (Treves, 1995) seem impractical due to the complexity of the system, while experimental approaches seem premature, before the exact hypothesis to be tested and the range of possible confounds has been clarified by simulation studies. In neural network simulations, at the price of some necessary simplification, one can compare the performance of the differentiated CA3-CA1 circuit with a uniform CA circuit of equal number and type of components (as in Treves, 2003).

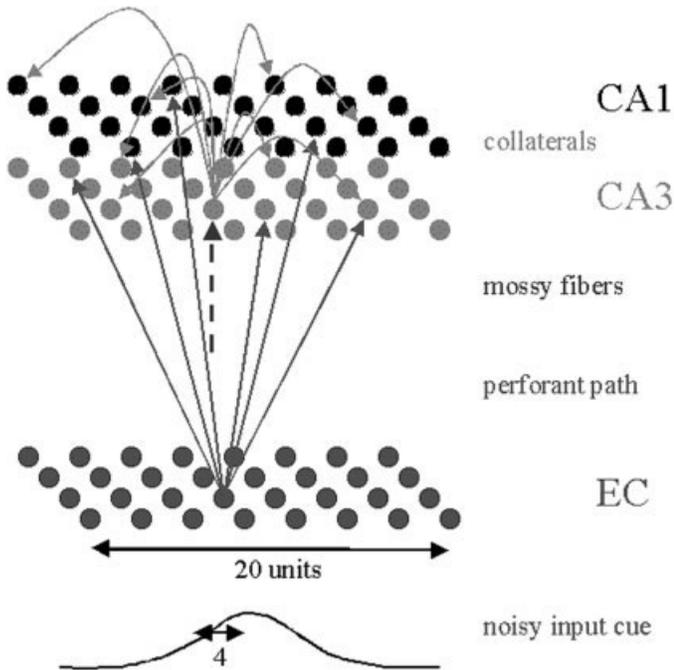
Neurophysiological recordings, even when combined with lesion studies, can only compare the normal circuit with others with missing components, and it is thus difficult for them to say the last word on the meaning of a differentiation. Nevertheless, experimental predictions arising from the modeling work might be testable with neurophysiological experiments, which have already begun to indicate differences in memory-related properties between CA3 and CA1 (Robertson et al., 1998).

## METHODS

The model, shown schematically in Figure 3, simulates a rat that runs around a two-dimensional (2D) environment, and whose position  $p$  is coded in the inputs to the CA fields. These inputs generate activity, which is then decoded by a given procedure, to extract a “read-out” position  $q$ . The latter is compared to the actual position  $p$ , and their correspondence, averaged over a running session, is quantified by percentage correct and information measures. The same model could be applied to spatial view maps in the monkey, with some modification (Rolls et al., 1998). Note that the environment is taken to be a torus, to limit finite size effects, which are more serious with definite boundaries.

### Circuitry

The inputs are modeled as activity on a grid of  $N \times N$  units, representing entorhinal cortex, EC. In all simulations reported here,  $N = 20$ . The location of the units in the grid is not intended



**FIGURE 3.** Scheme of the network used in the simulations. DG units (not represented) project each to a single CA3 unit in register with it (and, in the uniform version, also to the corresponding CA1 unit).

to correspond to their position in the cortex, but rather to map the particular 2D environment the model rat is running in. CA3 and CA1 fields are also modeled as 2 grids of  $N \times N$  units, in which however, initially, the position of each unit is merely a convention of the computer code. Later on, with changes in connection weights, the spatial selectivity of CA units can, if the system “learns” successfully, come to be in register with that of EC units, thus comprising two more maps of the same environments. This occurs in some of the simulations, but not in all, and in the differentiated model only in CA3, as discussed below. Periodic boundary conditions are used, as for the environment, so each array is in fact a torus.

Each CA unit receives  $C_{pp}$  feedforward connections from the EC array, and  $C_c$  collateral connections from other CA units. The collateral connections originate from both CA arrays in the uniform model, and only from the CA3 array in the differentiated model, in which case one may call recurrent collaterals those synapsing onto CA3 units, and Schaffer collaterals those onto CA1 units (Fig. 1). Both sets of connections are assigned to each receiving unit at random. In the uniform model, the array of origin of each collateral connection, whether “CA3” or “CA1,” is also random.

Weights are originally set at a uniform constant value (normalized so that the total strength of PP afferents equals  $W_{pp}$ , and that of collaterals equals the parameter  $W_c$ ), to which is added a positive random component of similar mean square amplitude, to generate an approximately exponential distribution of initial weights onto each unit (exponential above the minimum, e.g., above  $W_{pp}/C_{pp}$ ).

Both models also include a very simplified representation of the effect of the dentate gyrus (DG), which follows the ideas (McNaughton and Morris, 1987) at the basis of our earlier analyses (Treves and Rolls, 1992). DG inputs to each CA unit are represented as arriving from a single unit in the EC array, the one in the corresponding grid position, and are relayed by a synapse of fixed strength, which transmits only during training. When testing the system, there is no DG input. In the uniform model, such DG inputs arrive to both CA fields, while in the differentiated model only to CA3 (Fig. 1). The strengths of DG synapses,  $W_{mf}$ , is one of the parameters studied, and in particular setting it to 0 obviously enables studying systems without this particular feature. A plausible model of the real hippocampus, however, requires these synapses to be quite strong, with a potential capability to discharge, or “detonate,” CA3 cells (McNaughton and Morris, 1987; Henze et al., 2002).

### Trajectories and Reverberations

EC inputs are taken to represent a continuous trajectory in a particular 2D environment, isomorphic to the arrangement of the  $N \times N$  units in the toroidal grid. At each time step, the new position of the virtual rat is selected by repeating the direction of the last spatial step (the difference between the last and the penultimate position) with a little bit of noise added, so as to make the itinerary smoothly curved, with a large radius of curvature. In the simulations reported here, the spatial step was 0.2 grid units, and taken to correspond to 1 cm. Noise was added as a random vector of mean amplitude 0.08 cm. Time steps were taken to correspond to 12.5 ms, resulting in a virtual rat speed of 80 cm/s (trajectories are curved at random by the spatial noise, but given the large radius of curvature, most of the time the virtual rat approximately loops around the torus). The EC activation in a given location was calculated as a Gaussian function of the distance of each unit from the rat position, with a spread  $\sigma_{EC} = N\sqrt{(a_{EC}/4\pi)}$  and maximum activation equal to 2. With this, the sparseness (Treves and Rolls, 1991) of EC activity is set to  $\sim a_{EC}$ , and for typical values of the sparseness ( $a_{EC} = 0.5$ ) activity spreads over about 4 grid units. Since positions are calculated in advance of when the virtual rat steps there, one can compare the position  $q$  decoded from neural activity with a whole list of positions  $p(k)$  including the current one ( $k = 0$ ), those taken in the past ( $k < 0$ ), and future ones ( $k > 0$ ). In practice,  $k$  is considered from +5 to -4 (which would correspond to 10 equispaced times between +62.5 ms and -50 ms relative to the present). DG activity is co-localized with EC activity on the grid, but sparser (for typical values of sparseness  $a_{DG} = 0.05$  and, using the same formula, activity in DG spreads over 1.3 grid units). The above applies during training. During testing, partial cues are used (cf. Nakazawa et al., 2002); i.e., random trajectories are generated in the same fashion, but DG activity is irrelevant, and EC activity is generated with the above formula with probability (independent for each unit) set equal to “cue size”  $Q$ . With probability  $1 - Q$ , EC activity is randomly assigned to each unit, from an approximately exponential distribution of sparseness  $a_{EC}$ . In most simulations reported here, it was set  $Q = 0.2$ , implying that only 20% of EC units have activity correlated with the spatial

position of the virtual rat, at any time step during testing. Because of reverberation along collateral connections, however, the proportion of CA units correlated with space is higher, even before any learning takes place, in part because of transduction through the synaptic matrix, and in part because CA activity reflects EC activity over a number of different time steps, in each of which the fraction  $Q$  of spatially informative units is randomly selected anew.

The activity of CA units is, in fact, updated at each time step and, since part of their inputs are from collateral connections, reverberations effectively influence the output (as in standard attractor dynamics; Amit, 1995), even while the PP input reflects a constantly changing external input. Each updating of unit  $i$  amounts to summing all excitatory inputs into the variable  $h_i$ , followed by a competitive algorithm defined as having the effect of normalizing the mean activity of each CA field, and setting its sparseness to  $a_{CA}$  (Treves and Rolls, 1991). The algorithm represents a combination of subtractive and divisive feedback inhibition, and operates simply by iteratively adjusting the gain  $g$  and threshold  $\vartheta$  of the threshold-linear transfer function characterizing all units in a field,

$$r_i = g(h_i - \vartheta) \quad (\text{if } h_i > \vartheta, \text{ and } r_i = 0 \text{ otherwise})$$

until their mean activity and sparseness take the desired values. This algorithm converges very quickly, and I have used it routinely in other simulations (see, e.g., Treves, 2003) in which it does not seem necessary to model inhibition in detail. In most of the current simulations,  $a_{CA} = 0.2$ .

## Training and Testing

The network operates in successive training and testing phases. At each time step, the EC and DG pattern of activation is fed forward to the CA units, whose activity is reverberated along collateral connections, and (during testing only) it is read out to decode the position of the rat. In a training phase, the EC and CA activation values are used to modify connections weights according to a model associative rule. The exact “learning rule” used to modify connection weights was found not to affect results substantially. Those reported here were obtained with the general rule

$$\Delta w_{ij} \propto r_j^{post} \cdot (r_i^{pre} - \bar{r}_i^{pre})$$

where  $r$  denotes the firing rates of the pre- and postsynaptic units, and  $\bar{r}$  an average over the corresponding array. The rule is applied, at each time step in each training phase, to weight  $w_{ij}$ . Different variants, used in particular to model STDP, were obtained by varying the activity  $r_i^{pre}$ . The default option was to use the presynaptic activity at the previous time step (i.e., hypothesizing a 12.5-ms optimal lag between pre- and postsynaptic activity for potentiation); alternative options were explored by using for  $r_i^{pre}$  a trace over previous time steps, integrated with a difference-of-exponentials kernel of time constants (in inverse time steps)  $\beta_1$  and  $\beta_2$

$$r_i^{pre}(\text{trace}) = \left[ \frac{\beta_1 \beta_2}{(\beta_1 - \beta_2)} \right] \sum_{k=-\infty}^{-1} \times [\exp(\beta_2 k) - \exp(\beta_1 k)] r_i^{pre}(k)$$

In a testing phase, the activity of CA3 and CA1 units is fed into a decoding algorithm, external to the cortical network. The algorithm attempts to predict the location of the virtual rat, by comparing the current pattern of activation (in each CA field) with stored templates of the corresponding activation when the rat is stationary at each location on the  $N \times N$  grid. The comparison results simply in the choice of the best matching location (the one whose activation template has the largest normalized dot product with the current activation). Templates are recalculated after each training phase, to reflect the updated weights. The result of each testing time step is a decoded location  $\mathbf{q}$  for each CA field. In addition, the location decoded by considering both fields together was calculated separately, but it is not discussed in the present work, since the corresponding localization accuracy was trivially just a bit better than the average for CA3 and CA1. Further, since percentage correct and information measures are extracted from a number of units already well in the saturation regime (Treves, 2001), even decoding only, say, one-half of the units in a single field does not alter results significantly.

A training or testing phase, or running session, included 50,000 time steps in the simulations reported here, and an entire run for a network with given connections was comprised of 4 testing phases interleaved with three training phases. To generate statistically reliable results, 10 runs were averaged for each set of parameters in the figures. A difference between training and testing is that while weights are updated at each time step during training, testing only occurs every tenth time step, to avoid wasting CPU time on assessing localization accuracy for nearby, and hence strongly correlated positions.

Localization accuracy is averaged over a testing phase (5,000 decoding steps along a random trajectory) to generate a table of joint probabilities  $P(\mathbf{p}(k), \mathbf{q})$  computed from actual positions at each time step from  $k = -4$  to  $k = 5$  around the decoding time step. From  $P(\mathbf{p}(k), \mathbf{q})$ , for each  $k = -4, \dots, 5$ , one can extract a measure of percentage correct localization

$$f(k) = \sum_{\mathbf{p}} P(\mathbf{p}(k), \mathbf{p})$$

and one could extract a localization information

$$I'(k) = \sum_{\mathbf{p}(k), \mathbf{q}} P(\mathbf{p}(k), \mathbf{q}) \log_2 \frac{P(\mathbf{p}(k), \mathbf{q})}{P(\mathbf{p}(k))P(\mathbf{q})}$$

However, to avoid limited sampling problems, I simply computed the difference  $\delta(k) = \mathbf{p}(k) - \mathbf{q}$  each time, and calculated the reduced information

$$I(k) = \sum_{\delta(k)} P[\delta(k)] \log_2 P[\delta(k)] \times N \times N$$

which assumes translation invariance of the original probability matrix, and does not require impractically long sampling times. The source code is available by e-mail for further details about the simulations. Figure 4 demonstrates place representations with examples from a simulation that uses default values of the parameters, as in Table 1.

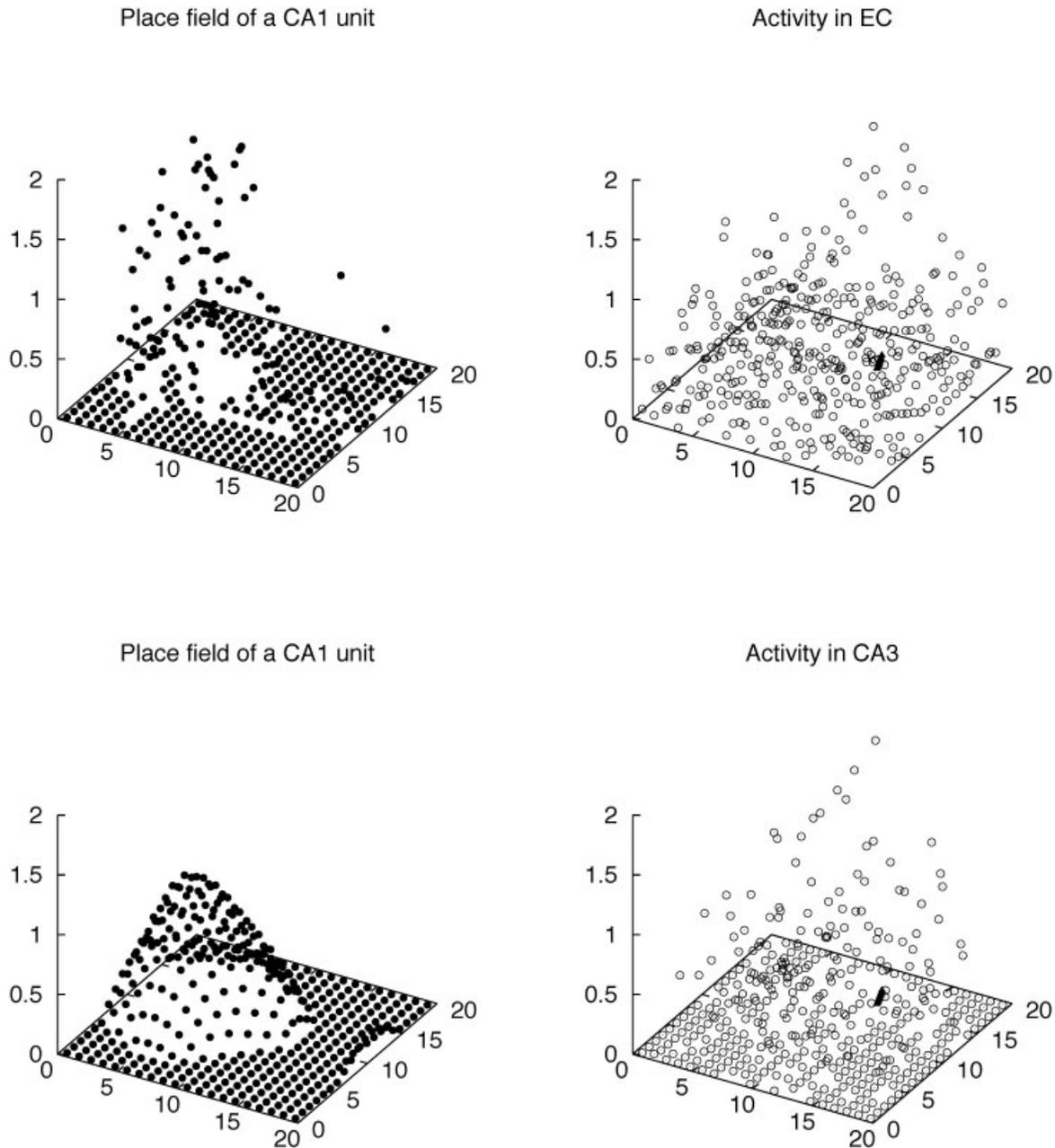


FIGURE 4. Example of the “place field” of a CA1 unit, before (top left) and after (bottom left) three training sessions. The place field is obtained by testing the system with the virtual rat at each possible location on the grid, and  $Q = 1$ . Note that the center of the field is not in register with the “physical” location of the unit in the computer code, at position (10,10), as in this differentiated model CA1 fields are initially randomly located and noisy (but with some

structure given by the nonuniform connectivity), and training refines them without displacing them. The right shows the activity in EC (top) and CA3 (bottom), at the end of training, when the virtual rat follows the trajectory indicated by the black triangles (from  $k = -4$  to  $k = 5$ ), and a partial cue ( $Q = 0.2$ ) is given as input. Note that CA3 activity is in register with EC activity (but sparser) and with the position of the rat.

## RESULTS

To test the implementability of mechanisms A1, A2, and B, I carried out a series of simulations in which certain critical parameters were varied, while the position decoded from neural activity was compared with actual positions at different time steps

relative to present. In order to explore mechanisms that may be functioning effectively only with separate CA3 and CA1 fields, I first report, in Figures 5–7, simulations that were carried out in the differentiated model. The graphs illustrate the average percentage correct localization, after three training sessions, extracted from decoding CA1 activity, and comparing the decoded position to actual position up to 4 time steps in the past, and up to 5 in

TABLE I.

Default Values Used in Most Simulations, Except Where Noted Otherwise

Size	Weights			Sparseness			Others
$N \times N$	$20 \times 20$	$W_{pp}$	1	$a_{EC}$	0.5	$Q$	0.2
$C_{pp}$	40	$W_c$	1 (but when testing $W_c = 3$ )	$a_{DG}$	0.05	$d$	0.1
$C_c$	120	$W_{mf}$	2 (but when testing $W_{mf} = 0$ )	$a_{CA}$	0.2	$\tau_{opt}$	7 time steps (i.e., 87 ms)

the future. Decoding CA3 activity results in all cases in lower values.

### Only Mechanism B Produces Genuine Prediction

Figure 5 illustrates the attempt to implement mechanism A1, the one based on STDP on collateral connections. As a model STDP rule, I used the modified form of the generic rule where, as mentioned under Methods, the presynaptic term is a trace of presynaptic activity over previous time steps. The trace is a difference of exponentials with time constants  $\beta_1$  and  $\beta_2$ , which makes the convolution kernel peaked at time step

$$\tau_{opt} = (\beta_1 - \beta_2)^{-1} \ln(\beta_1/\beta_2)$$

In Figure 5, curves are shown for three values of  $\beta_1 \equiv 2\beta_2$ , 0.4, 0.2, and 0.1 (time steps) $^{-1}$ , resulting in kernel peaks at  $\tau_{opt} \cong 3.5, 7,$  and 14 time steps, or, converting to “physiological” peak times, at  $\sim 43, 87$  and 173 ms. In addition, the curve for the standard rule, in which the presynaptic term is simply the activity at the previous

time step, is shown for comparison. As shown in Figure 5, in this latter case percentage correct localization is peaked three time steps in the past, and declines rapidly when moving into the future. The network is therefore unable of genuine prediction: whatever percentage correct localization above chance, in the future, is just the result of the correlation between past and future positions.

With trace rules extending over increasing time delays, there is a shift of the peak percentage localization toward the present, but never beyond it, even for a delay extending over  $\sim 14$  time steps (173 ms). My conclusion of these and many similar simulations, relying solely on a modified version of the synaptic learning rule to model STDP effects, is that despite my best efforts I was unable to produce prediction based on mechanism A1 alone. The same conclusion holds for uniform networks, not differentiated between CA3 and CA1 (not shown).

Figure 6 illustrates the attempt to implement mechanism A2, the one based on differential usage of recurrent collateral connections in CA3. The mechanism would be based on the hypothesis, illustrated in Figure 2, that somehow collateral connections produce an effective delay in transmitting a representation along the Schaffer collaterals, a delay which is longer during storage (in the model, training) than during retrieval (i.e., testing). The modulation of the effective delay is obtained indirectly, in the model, by varying the overall strength of collateral connections relative to perforant path connections, both in CA3 and CA1, during storage and during retrieval. Stronger collateral during storage would result in more reverberation of activity from previous time steps, in CA3, and thus in a longer effective delay, as required by mechanism A2. This hypothesis, it should be noted, is thus in contrast with the notion, supported by neurophysiological measures in slices (Hasselmo and Schnell, 1994), that the effect of collateral connections is suppressed, via acetylcholine, during storage of new memories. Further, in the absence of definite experimental indications to the contrary, I have applied the same strength modulation in CA3 and CA1; the effect of strengthening Schaffer collaterals is not to lengthen an effective delay, and therefore it may partly contrast the effect of the modulation of CA3 recurrent collaterals. In the figures, the default values are for collaterals to have the same overall strength as perforant path connections during training ( $\times 1$ ), and to be three times stronger during testing ( $\times 3$ ), i.e., the default values for all other simulations. In addition, Figure 6 (left) includes curves where the strength of collaterals is modulated only during training. It can be seen that strengthening collaterals during training lowers the localization performance of the network, and also shifts the peak localization further in the past. This is in agree-

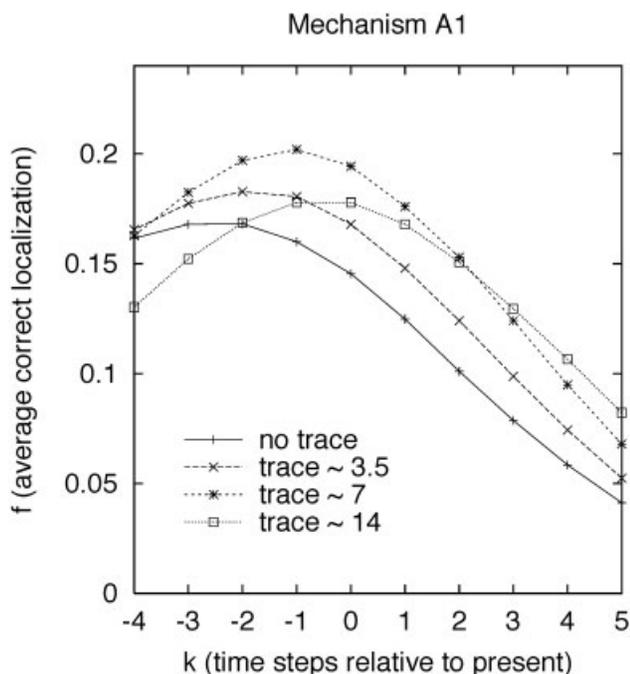


FIGURE 5. Percentage correct localization for different time steps relative to the present, with learning rules varying in their presynaptic trace kernel length, corresponding to peak times of  $\sim 43, 87,$  and 173 ms.

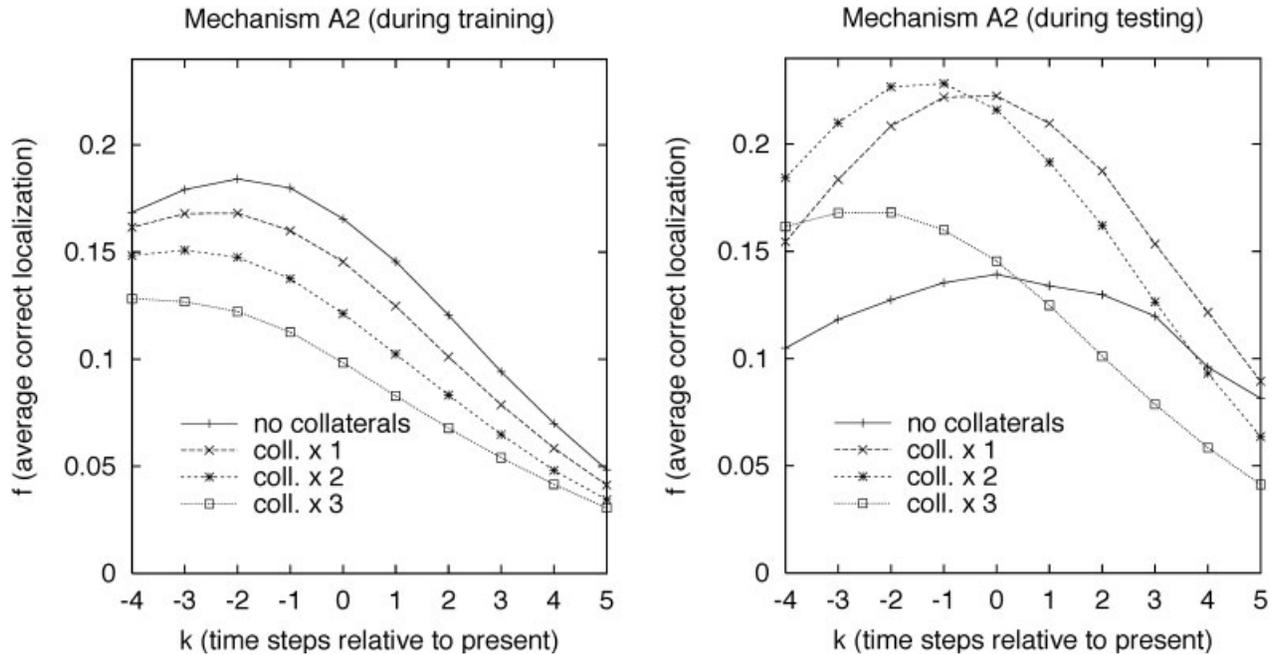


FIGURE 6. Localization when the strength of collateral connections is modulated during training (left) and during testing (right).

ment with the physiologically based notion that collaterals should be suppressed, not enhanced, during storage, and runs counter to the hypothesis expressed by mechanism A2. Figure 6 (right) includes curves where the strength of collaterals is modulated only during testing. It can be seen that the absolute performance of the network is not monotonic in this parameter, having a shallow maximum when collaterals are overall between one and two times as strong as perforant path connections. The peak of the curves, however, monotonically shifts toward the past with increasing collateral strength. Thus, these simulations indicate that having less collateral reverberation during retrieval than that during storage might shift the peak localization ability toward the future, as hypothesized with mechanism A2, but not to the point of turning into a genuine prediction, since even with zero collaterals at testing the peak localization occurs for the present time step. Further, this occurs at the price of a gross reduction in the accuracy of overall localization. The overall conclusion from simulations like those reported in Figure 6 is that mechanism A2, which I stress runs counter to the current understanding of the role of acetylcholine modulation, would not be effective anyway at producing predictions into the future. It is important to note, also, that while the simulations support the opposite notion, which reflects our current understanding, that localization accuracy benefits from suppressing collaterals during storage, still even silencing collaterals at training does not result in genuine prediction (Fig. 6, left, solid curve). Note that, again, the same conclusion holds for uniform networks, not differentiated between CA3 and CA1 (not shown).

Figure 7 illustrates the effects of a simple model that implements mechanism B.

The model amounts to subtracting from the sum of inputs to each unit a fraction  $d$  of its own recent output activity. The trace of its "recent" activity is calculated with a convolution kernel, nor-

malized to 1, expressed as a difference of two exponentials, exactly as for the presynaptic term of the STPD rule, with inverse time constants  $\beta_1 = 2\beta_2 = 0.2$  (time steps) $^{-1}$ . As shown in Figure 7, introducing modest amounts of adaptation rapidly shifts the peak of localization accuracy toward the future, thus resulting in genuine predictions. This results (for values beyond  $d = 0.02$ ) in a decrease of peak percentage correct localization, but the decrease becomes substantial only for large  $d$  values. The conclusion is that adaptation is an effective mechanism for generating a predictive ability, and although this comes at the expense of overall localization performance, the conflict between the two functions is not dramatic. The shift toward future locations brought about by adaptation comes, essentially, from weakening the signal pointing toward past and present locations, which are all close by, not by globally altering signals toward other more remote locations in the environment. Therefore, one expects a measure like percentage correct, sensitive to the exact localization among nearby locations, to decrease more with this weakening of the signal than a more global measure such as the amount of localization information. This is apparent in Figure 7 (right), which reports the decoded information in bits.

The conclusion from this set of simulations is that an extremely simple mechanism, which can be based on a model of firing rate adaptation but also perhaps of synaptic fatigue, produces robust prediction at a limited cost in terms of accuracy.

### Differentiation Does Not Help Very Much With Prediction

Having established that adaptation is the one mechanism, among the three considered, which effectively generates prediction in the CA3-CA1 differentiated network, one can ask to what extent

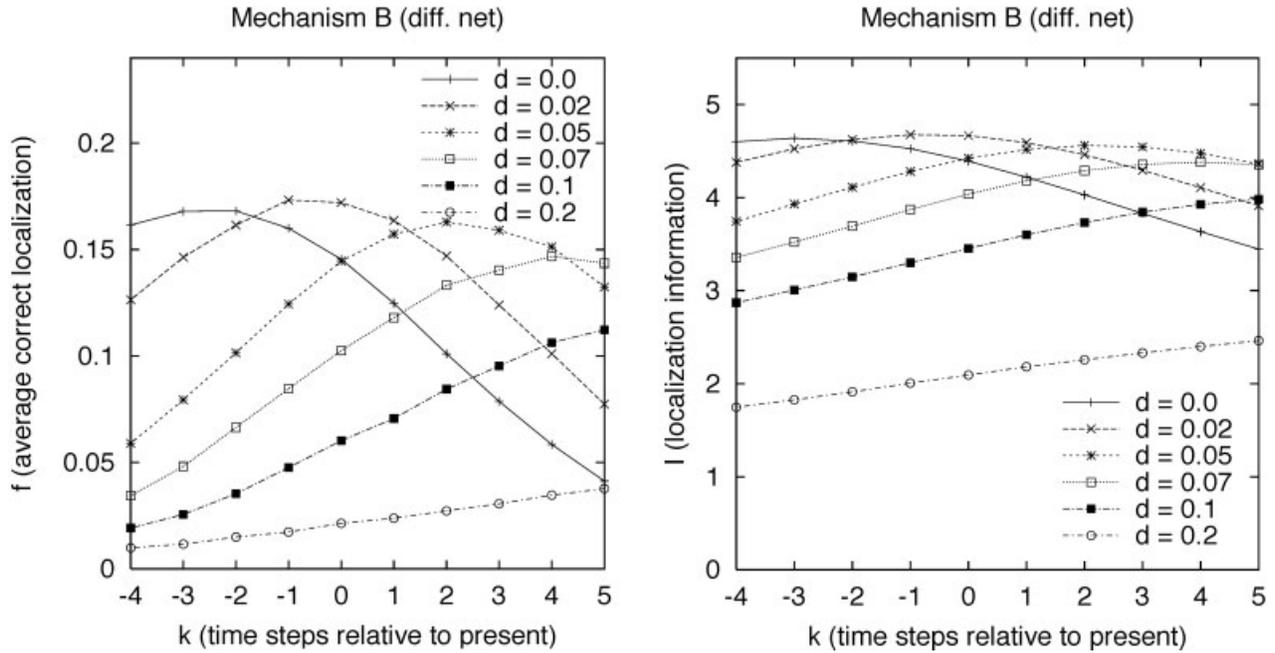


FIGURE 7. Percentage correct localization (left) and localization information (right) as a function of time step, for different values of the adaptation parameter  $d$ , for the differentiated model.

is the mechanism, which seemingly does not depend on differentiation, effective in the uniform network. Figure 8 shows the results of simulations identical to those of Figure 7, except that they were performed with the uniform network. It is apparent that differentiation does not interact much with the predictive ability: the two pairs of graphs look very similar. Close inspection of the graphs,

however, reveals a degree of interaction. Whereas in the absence of adaptation the uniform network is slightly superior in performance to the differentiated one, the opposite is true for values of the adaptation parameter above  $d = 0.05$ . As a result, the decrease in performance with increasing adaptation is significantly less in the differentiated network: the peak percentage correct localization

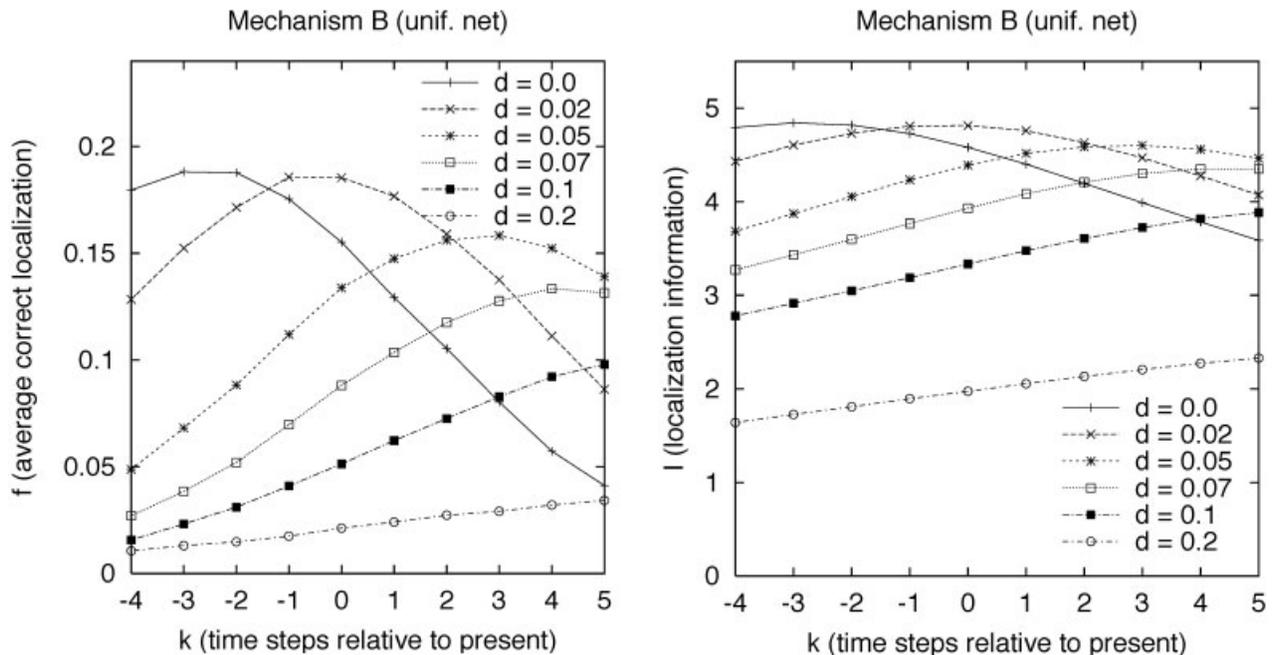
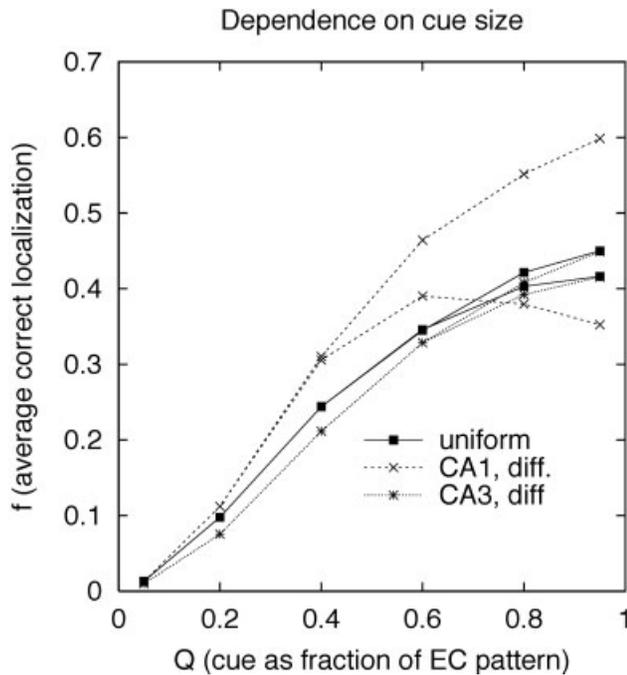


FIGURE 8. Percentage correct localization (left) and localization information (right) as in Fig. 6, but for the uniform model of Fig. 1.



**FIGURE 9.** Localization as a function of cue size,  $Q$ , for the differentiated and uniform model. In the differentiated model, separate measures are extracted by decoding the activity of the CA3 and of the CA1 array (these measures are statistically equivalent in the uniform model). The lower branch of each curve reports the value at  $k = 5$ , when this differs from the peak value for  $-4 \leq k \leq 5$  (in the upper branch).

decreases by 13% instead of 29% going from  $d = 0.0$  to  $d = 0.07$ . These values were obtained by decoding activity in CA1. Decoding activity in CA3 (temporal detail not shown, but see the following figures for the values at  $k = 5$ ) produced instead consistently lower values in the differentiated network. Thus, one can understand the effect of the differentiation of Figure 1, at least in this paradigmatic case, as quantitatively separating the performance of the two networks along the direction of information flow: lower in CA3, higher in CA1. With the uniform network, performance is approximately equal in the two sub-fields, and in most cases roughly an average of the values obtained with differentiation. There is, however, a limited but positive interaction between differentiation and firing rate adaptation: adaptation brings about the same predictive ability (the location of the peak) in the differentiated network, at a lesser price in terms of localization performance (the height of the peak) than in the uniform network.

The above results correspond to rather poor localization but were obtained with relatively small cue size,  $Q = 0.2$ . It is important to know whether the main findings apply to other cue sizes. A confounding factor is that with larger cue sizes the shift toward locations different from the present, both in the direction of the past and of the future, is much reduced, simply because the more precise signal coming through the perforant path “binds” the network to the current location. Figure 9 considers only the case  $d = 0.1$ , which is taken from now on as the standard value for adaptation, and shows the performance of the uniform and differentiated networks as a function of  $Q$ . For the differentiated network, results

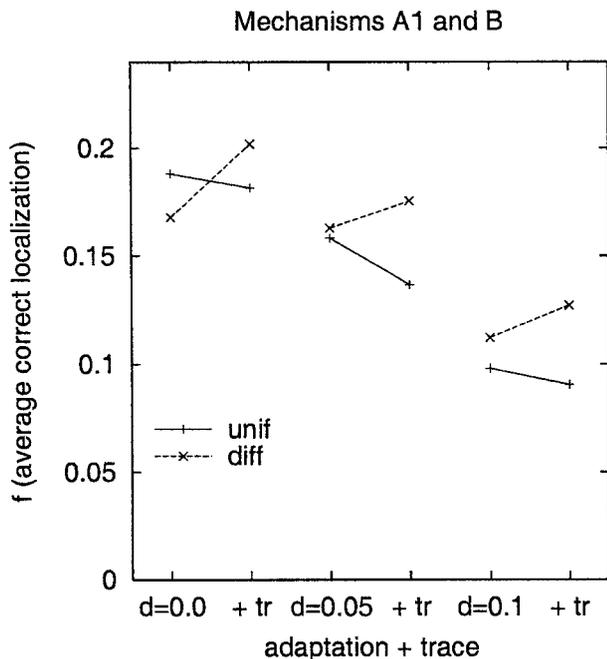
are plotted separately for CA3 and CA1. Moreover, either at  $Q = 0.4$  or  $Q = 0.6$ , each curve bifurcates. The lower branch reports the values obtained at  $k = 5$ , i.e., at the (future) end of the time interval considered, where performance is best for small cues; while the upper branch reports the peak performance, which for larger cues occurs at lower  $k$  values, i.e., less projected into the future. It is apparent from Figure 9 that with respect to peak performance, the uniform network is in between CA3 and CA1 of the differentiated network, but closer to the former. The very same effect (one more stage of processing) that enhances peak performance in CA1, brings down performance at  $k = 5$  for larger cues: where the signal is already highly informative, further processing sharpens it toward the current location, and with respect to the location which will be visited in 5 time steps apparently it does more damage than good. In the following, the cue size is set at  $Q = 0.2$ , and in most cases results are plotted for  $k = 5$ , whenever this is the maximum performance within the interval considered (the true maximum occurs for larger  $k$  values).

One may wonder whether different degrees of adaptation of CA3 and CA1 units would lead to better performance. This would indicate an advantage in differentiating not just the connectivity, but also the biophysics of cells in the two fields, possibly compatible with the evidence (see, e.g., Traub et al., 1991). Extensive simulations with different degrees of adaptation show that this is not the case. Making CA1 units either less or more adapting than CA3 units leads to somewhat worse performance. The comparison can be made by choosing parameters that keep constant the location of the peak performance, and contrasting the height of the peak. For example, if  $d = 0.1$  in CA3 and  $d = 0.05$  in CA1, the peak is at  $k \approx 4.5$ , with percentage correct localization  $f = 0.131$  and information  $I = 4.18$  (these values are similar for  $k = 4$  and  $k = 5$ ). With  $d = 0.08$  in both CA3 and CA1, again the peak is at  $k \approx 4.5$ , with percentage correct localization  $f = 0.133$  and information  $I = 4.23$ . A similar decrease is found when making CA1 units more adapting than CA3 units. Optimal performance seems to occur for identical  $d$  values in the two arrays of units.

### Interaction With Mechanisms A1 and A2

Mechanism A1 on its own was unable, in the model tested, to produce genuine prediction. When paired with firing rate adaptation, however, an STDP (or “trace”) plasticity rule can enhance performance and also slightly improve the time span of prediction. Interestingly, the increase in percentage correct localization occurs only in the differentiated model. Figure 10 shows the peak correct localization for different values of the adaptation parameter, with and without the trace rule. Adding the trace rule decreases performance in the uniform model and increases it in the differentiated model.

The results presented for mechanism A2, although they failed to produce the desirable prediction effect, illustrated the importance of tuning the strength of collateral connections. It is therefore of interest to understand how performance depends on such strength, when it is modulated during training or during testing. Figure 11 shows the values obtained (for  $Q = 0.2$  and  $k = 5$ , as a function of the strength of collaterals—again, in both CA3 and CA1) in the



**FIGURE 10.** Localization performance as a function of adaptation and of the use of a trace rule. Data points correspond to  $k = 5$  for the adaptation parameter  $d = 0.1$ , and to peak correct localization for  $d = 0.0$  (the peak occurs at  $k = -3$  without trace rule, and shifts to  $k = -1$  with the trace rule, in both models), and for  $d = 0.05$  (here the peak is at  $k = 3$  in the uniform model and  $k = 2$  in the differentiated one, and does not shift).

training phase, if they are set at three times the strength of perforant path projections during testing. It is apparent that the best performance is when collaterals are suppressed during training, in line with the Hasselmo and Schnell (1994) argument. It is also seen that there is, again, a positive interaction with differentiation: the increase in performance is much less pronounced in the uniform model and, conversely, when the strength of collaterals during training is substantial (e.g., three times that of the perforant path, as during testing) the uniform model outperforms the differentiated one. Finally, there is a minor but interesting discrepancy between the percentage correct and information measure; standard error values are reported for both, for an appraisal of its significance.

If the strength of collaterals is modulated, instead, during testing, a non-monotonic behavior appears. In other words, there is an optimal value, which in Figure 12 is around 2–2.5 times the strength of both the perforant path, and the collaterals during training. The existence of such a maximum is indirectly indicative of the validity of the model: if maximum performance had been when the collaterals are fully suppressed both during training and testing, obviously they should have been discarded from the model. Also, if the optimal value of their overall strength during testing had been very large, that would have suggested something ill-designed about the model. The fact that the optimal ratio is not too dissimilar from the ratio of collateral-to-perforant path synapses in the real hippocampal circuit (Amaral et al., 1990) suggests

that within this highly simplified model the main parameters may have been at least chosen consistently with each other.

### Learning, and the Role of Mossy Fibers

So far, the results presented were always those obtained after 3 training sessions. How this stage of training is reached depends crucially, in the model, on the mossy fiber system. In the model, mossy fibers are represented quite succinctly by one-to-one connections from units in the “DG” array to units in the CA arrays (or to units in the CA3 array only in the case of the differentiated model). This single connection is only active during training, and its strength is given by the parameter  $W_{mf}$  which, when  $W_{mf} = 1$ , implies that the single connection is as powerful as all (“direct”) perforant path connections from EC summed together (i.e., their baseline value before training). Values of the parameter  $W_{mf}$  around unity are compatible with experimental evidence (see, e.g., Henze et al., 2002) and with our earlier modeling work on the function of the mossy fiber input to CA3 (Treves and Rolls, 1992).

The DG array represents a sparser map fully in register with the EC map of the environment, and the teaching role of the mossy fibers derives precisely from this sparseness. Before training starts, CA fields do not represent any organized map, in either the differentiated or the uniform model, and whatever amount of spatial information in their activity is simply the result of the random connectivity between EC and CA. This connectivity is uniform statistically, but not in unit-to-unit detail; therefore, responses to inputs that are very close to each other are also close, and this results in some amount of information and in percentage correct slightly above chance (Fig. 4). With training, the inputs to CA3 units as the virtual rat visits nearby locations, along a trajectory, include the reliable DG signal from the corresponding units, which tends to make responses in nearby locations similar, and thus to instruct learning by imposing a map in register with the DG map. This teaching signal is available also to CA1 units only in the uniform model. The EC signal acts in a partially similar fashion, but activity in EC is much more distributed, and more importantly the PP connectivity is not topographically ordered. Nevertheless, even in the absence of mossy fibers, responses in nearby locations tend to be similar, and this provides a basis for synaptic plasticity to refine spatial selectivity in the CA fields. Therefore simulations are necessary to assess quantitatively the benefits of training carried out with and without the assistance of the mossy fibers.

Figure 13 reports localization performance at different stages of training with varying mossy fiber strength, primarily for the differentiated model. In the absence of mossy fibers, there is an improvement in performance after the first training session, which, however, almost vanishes with further training. The reversal is less marked in the uniform model, whose performance is plotted for comparison in Figure 13 (left). With mossy fiber strength set to 1, there is a much larger improvement in localization after the first session, which, however, stays constant thereafter. With stronger mossy fibers, the initial improvement is less dramatic, but more sustained over successive sessions, and performance reaches consistently above that of the uniform model. The behavior is clarified in Figure 13 (right) (referring to the differentiated model only),

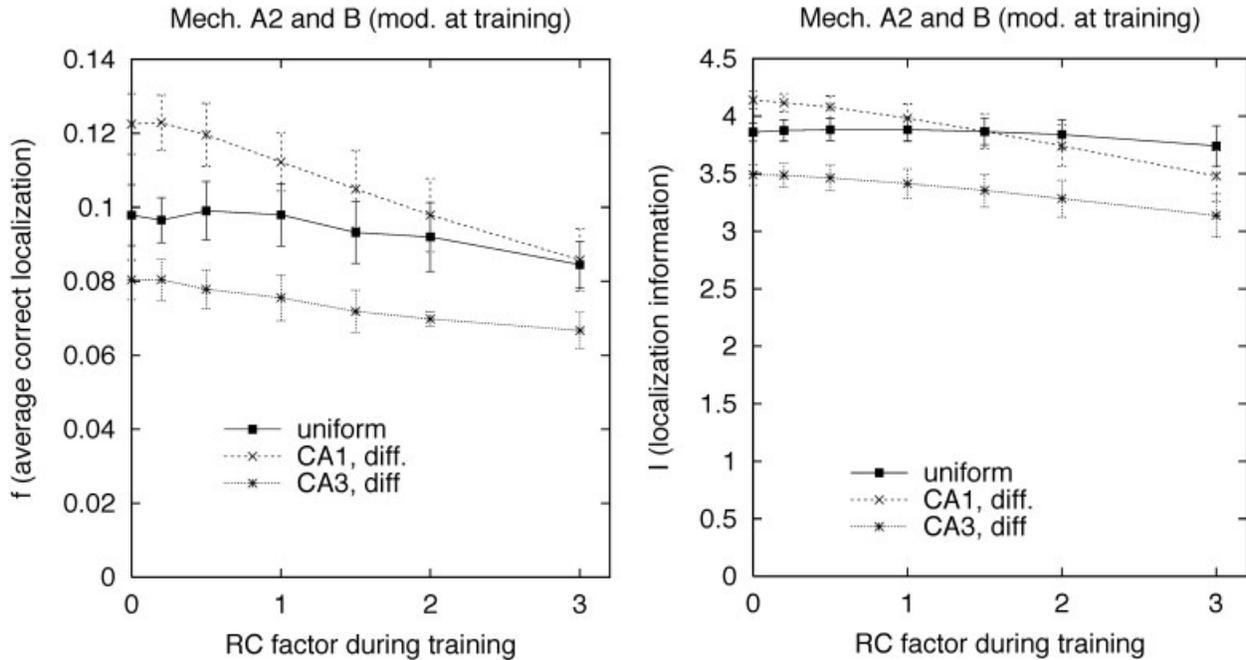


FIGURE 11. Localization as a function of the strength of collaterals during training, when they are set at three times that of the perforant path during testing. Percentage correct values in the left, information in the right. The recurrent collateral (RC) strength factor is defined in Methods.

which shows that performance is optimal, after one learning session, with mossy fibers of strength 1, in what is a rather peaked dependence. After successive training sessions the peak smoothens and shifts to stronger mossy fibers. It seems that, asymptotically for

extensive training, performance is monotonically increasing with mossy fiber strength.

The effect of the mossy fibers, in the model, can be understood as imparting structure into the emerging spatial representation, organizing it along the metric structure of the environment. If this is the case, one expects the localization information to increase with learning even more than percentage correct measures, because information measures depend also on how tightly clustered wrongly decoded locations are, around the correctly decoded one. The relationship between information and percentage correct measures, discussed in (Treves, 1997), can be quantified by the metric content index. The index can take values from 0 to around 1 (the precise maximum depends on several assumptions), and for certain model cases it is found to be roughly constant for a given representation, in the sense that sampling more and more units and applying the same decoding algorithm, both percentage correct and information increase, tending to co-vary along a curve of constant metric content (see Treves, 2001). An increase in metric content, instead, signifies the emergence of more structured representations, embodying more of the metric structure (the distance relationships) of the spatial environment. Higher metric content implies that even if localization is not accurate (not contributing to percentage correct) it is more often approximately accurate and less often random than when using a representation of lower metric content.

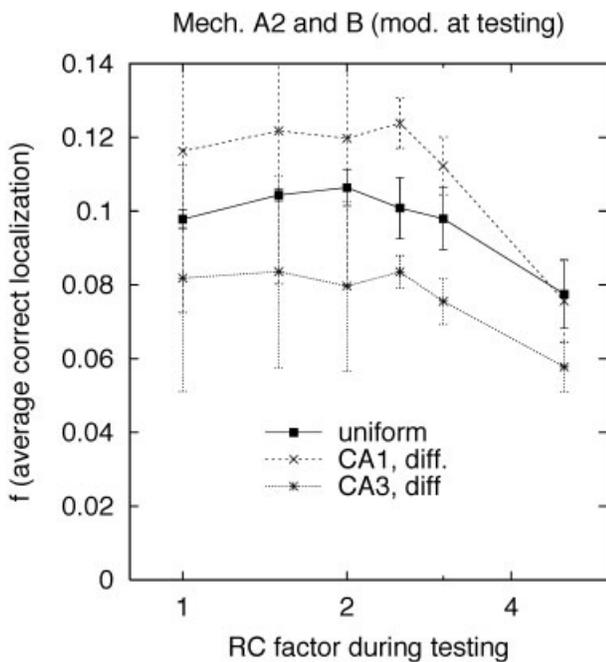


FIGURE 12. Localization as a function of the strength of collaterals during testing, when they are set to be as strong as the perforant path during training.

Figure 14 shows the relationship between information and percentage correct, in CA3 and CA1, for the differentiated model. Each data point along a broken line corresponds to a training stage. The CA1 data are the same as in Figure 13. In CA3, results happen

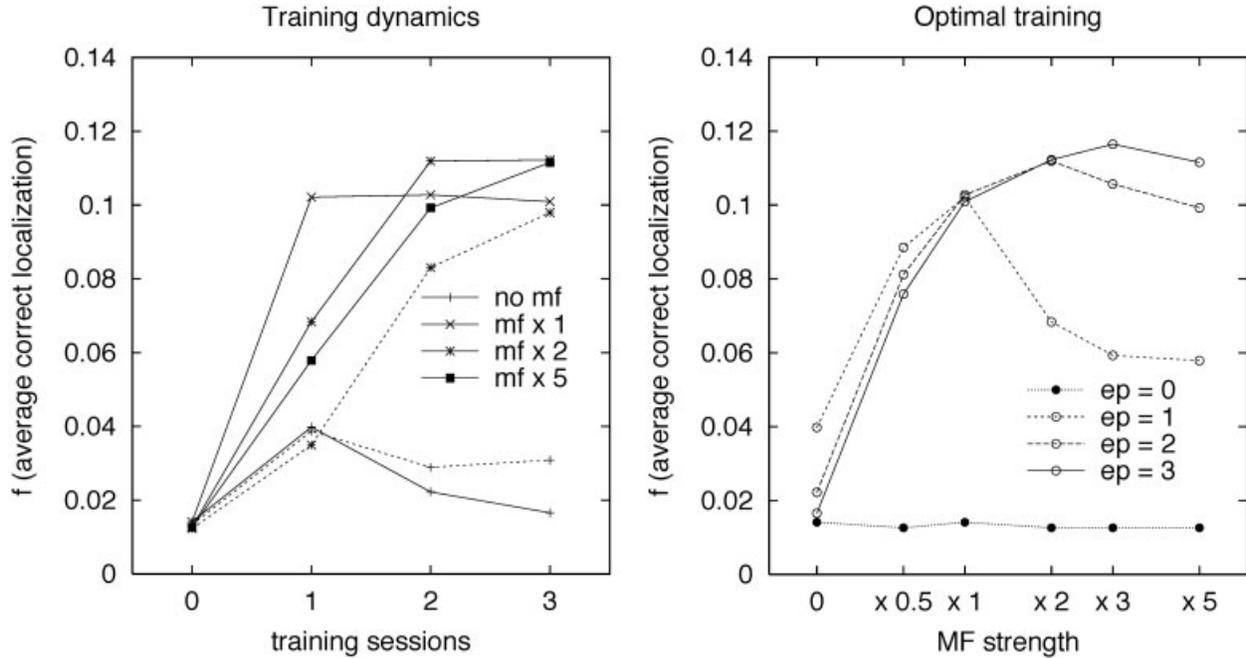


FIGURE 13. Localization as a function of training session (left) and as a function of mossy fiber strength (right), extracted from CA1 activity in the differentiated model. The left also reports two curves for the uniform model (dotted lines), for no mossy fibers and when their strength is set to 2.

to be identical after the first training session, but over successive session they diverge in opposite directions: better localization and increasing metric content with mossy fibers, decreasing metric

content and worse localization (percentage correct returning almost to baseline) without mossy fibers. In CA1 the trend is the same, but the enhanced performance produced by the mossy fibers (which only connect to CA3 units in this differentiated model) is salient already in the first training session. Thus, mossy fibers serve a crucial role in the model, and their effect is felt more strongly where it is indirect, in CA1, thanks to the additional stage of associative “polishing” of the spatial structure retrieved in CA3.

Before training, one expects the CA fields to enable much worse localization than is possible from decoding the inputs, at the EC stage; one may ask whether the effect of training is sufficient to produce an effective gain in localization, which would thus represent a genuine net contribution of the information retrieved from synaptic weights. The answer depends on many factors, for example, already at the input stage, on the sparseness  $a$  of the representation, on the cue size  $Q$ , and on the exact procedure used to model partial cues. In general, though, after training the CA fields reach, in this model, the same ballpark of localization performance as available at the inputs; sometimes improving on it, sometimes not. For example, for  $Q = 0.2$ , with  $a_{EC} = 0.5$  and the procedure used here, one has 19.6% correct localization ( $I = 4.77$  bits) in EC, and with  $a_{EC} = 0.2$  one would have 17.6% correct ( $I = 4.49$  bits). In CA1, with no adaptation, no trace rule and collateral strength set equal to perforant path both at training and testing, one finds (with  $a_{CA} = 0.2$ , and at  $k = 0$ ) 22.2% correct ( $I = 5.05$  bits), whereas with adaptation one obtains prediction but forfeits localization.

The comparison between decoding the inputs and the outputs should be taken with great caution, however, as it depends heavily on the unrealistic assumption, useful in the present model, that both the EC and DG input arrays represent ab initio an orderly

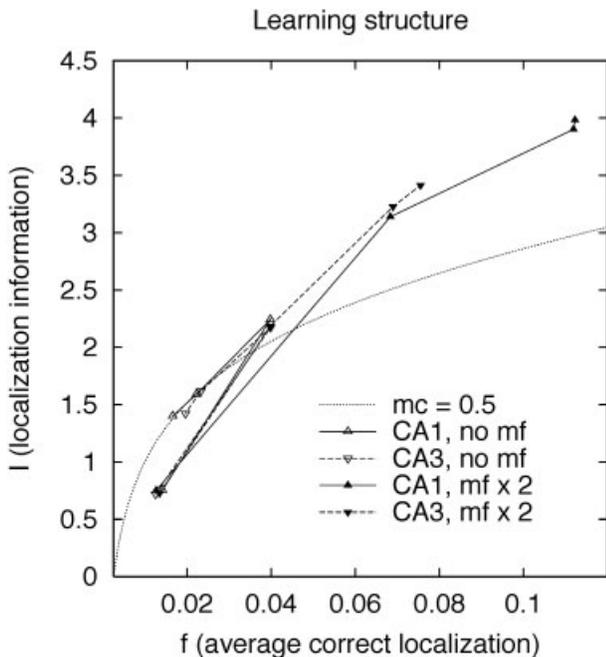


FIGURE 14. Information vs percentage correct. The dotted curve indicates where they would co-vary, for metric content 0.5. Actual metric content values increase, with mossy fibers, from about 0.3 before training to about 0.7 after 3 training session, in both CA3 and CA1. Without mossy fibers, metric content reverts to  $<0.5$ .

map of the environment. Thus, if one were to leave the MF inputs active also during testing, one would obtain even higher performance (36.3% correct localization from CA1, and  $I = 6.01$  bits, with  $Q = 0.2$ ), which is, however, a spurious effect of modeling DG as a topographic map. The rationale of these simulations is that such an unrealistic assumption has its restricted range of validity when simulating the training of CA fields, but not as a simulation of activity in EC or DG themselves, or as a way to extract better localization from the CA fields, when they reflect DG activity also during testing.

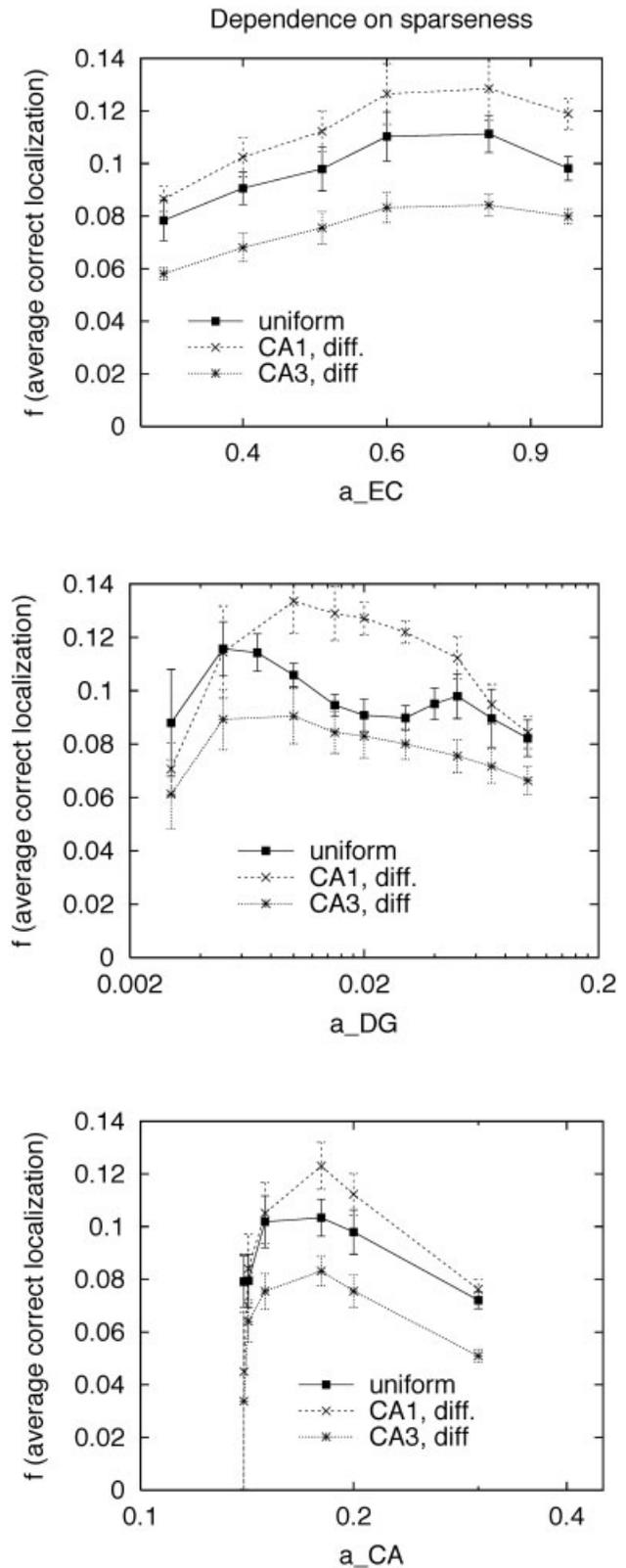
**Each Representation Has an Optimal Sparseness**

The performance of the network model depends on several parameters, and crucial among these is the sparseness of the representations in the different fields. Sparseness is set a priori in EC and DG; it results from the algorithm representing competitive inhibition in the CA fields. Over most of the simulations, it was set to  $a_{DG} = 0.05$ ,  $a_{CA} = 0.2$ ,  $a_{EC} = 0.5$ , in qualitative agreement with neurophysiological measures in the rat hippocampus (Jung and McNaughton, 1993; Frank et al., 2001). Figure 15 illustrates the effect of varying each of these parameters, separately from the others, while using the default values for all other parameters, and implementing mechanism B. The main observation is that for each field there is an optimal sparseness, whose value is, of course, related to that of all the other parameters, above and below which performance drops. Curiously, in the uniform model there seem to be 2 local maxima in performance as a function of  $a_{DG}$ , almost as if the system tried to indicate one particular value of dentate sparseness optimal for mossy fibers inputs to CA3, and a different one optimal for mossy fiber inputs to CA1, even though CA3 and CA1 are statistically equivalent in the uniform model. The differentiated model has a clear single peak, instead. In general, the exact values of the optimal values are likely not very significant, but their general relation  $a_{DG} \ll a_{CA} \ll a_{EC}$  is consistent with observation in the real system, and the very existence of maxima is suggestive of a well-designed system.

I have not explored different sparseness values for CA3 and CA1, which might lead to better performance than forcing a single value for both fields. One should also consider the possibility that with different values for sparseness, it might become advantageous to differentiate the degrees of adaptation, which was found not to be advantageous (see above) when the sparseness is the same. This possibility significantly enlarges the space of parameters to be studied, and is left for further work.

**DISCUSSION**

Within a traditional neuropsychological approach, i.e., based on lesions and on behavioral measures, one may ask what is the function of the CA1 field and whether it has a disproportionate role, as some evidence may suggest, in hippocampal processing of information “across time” (Rawlins, 1985). A neurocomputational approach attempts to refine this question by first of all removing trivial confounds. Thus, the trivial confound of studying a dam-



**FIGURE 15.** Localization for varying sparseness in the EC array (top), in the DG array (middle) and in the CA array (bottom), when otherwise  $a_{DG} = 0.05$ ,  $a_{CA} = 0.2$  and  $a_{EC} = 0.5$ , for the uniform and differentiated models.

aged system is avoided by using as reference a system without the region of interest (CA1), but otherwise fully functional; and the trivial confound of comparing systems of different sizes is removed by using the same number of components in each term of the comparison, and differentiating circuitry or other properties, instead. Second, such an approach is not limited to issues where strong qualitative discriminations are possible, but can address also issues where distinctions are subtler and finely quantitative. This seems particularly useful in the case of the hippocampus, where the full differentiation of its sub-fields, in mammals, may well have followed long after the evolution of a distinct role in memory for the whole medial and dorsomedial cortex. If indeed the refinement of the structure has lagged behind the function, a quantitative impact is all that can be expected of such refinement. The main price associated with neural network methods is, of course, that one studies a synthetic and highly simplified system whose relation with the part of the brain being modeled remains unclear, and results must be interpreted with care (see also Treves and Samengo, 2002), and corroborated by direct neurophysiological measures, most useful if combined with selective lesions (see, e.g., Brun et al., 2002). One important indication that emerges from the network simulations, then, is the encouragement not to be deterred by the lack of gross differences in the activity recorded in the CA3 and CA1 fields: relatively small differences may be all that has to be expected.

The notion that hippocampal processing of sequences (Fortin et al., 2002; Kesner et al., 2002a), and in particular phase precession (Jensen and Lisman, 1996), may be a manifestation of an ability to predict is not a novel idea here, nor are the three mechanisms that are posited to underlie such ability (see Levy, 1989, for an early and insightful discussion of mechanisms A1 and A2). The “trace” rule of mechanism A1, for example, has been proposed long ago as a mechanism to set up invariant representations (Foldiák, 1991), and, while its asymmetry was experimentally characterized more precisely as STDP, it was proposed to underlie the learning of temporal sequences (Minai and Levy, 1993; Abbott and Blum, 1996; Roberts, 1999) and in particular of spatial trajectories (Blum and Abbott, 1996; see also Foster et al., 2000), by reinforcing temporal order among neuronal pairs. Thus, conceptually, it seems an ingredient crucial for a system able to predict. The differential modulation of collateral connections in encoding and retrieval has been discussed extensively by Hasselmo et al. (1996) as a possible function of acetylcholine; though our mechanism A2 would require collaterals in CA3 to be suppressed at retrieval, contrary to experimental evidence (Kametani and Kawamura, 1990; Hasselmo et al., 1995; Marrosu et al., 1995)—and, indeed, its implausibility is matched by its failure to generate a predictive ability.

What is novel here is the quantitative comparison of different mechanisms within a well-defined task. It is this comparison that allows us to assess mechanisms A1 and A2, which conceptually make sense, as ineffective, and mechanism B, which does not interact much with the specifics of hippocampal circuitry, as effective. At the same time, the superiority of mechanism B is not claimed here to generalize to other models, e.g., to networks that mimic neocortical circuitry in a particular task. The relevant boundary conditions may be quite different, and a separate study

would be needed in each case. Even when the same virtual rat is restricted to 1D trajectories, for example, I find results that are not a simple special limit of those, reported here, holding in 2D. This is intuitive: an important factor to consider, in 1D, is that future and past locations around a given present location are equally well defined, given a constant velocity of the virtual rat, while in 2D they are progressively more spread out with increasing time difference, and the corresponding signal is thus weaker in 2D. Therefore the comparison among the three mechanism is claimed here to be valid only for the specific task simulated, with the underlying hope that the simulations offer an appropriate model to help discuss the possible rationale for differentiation in the real system, given the presumed importance of memory-based navigation in 2D environments by early mammals.

Having said that, the observed effectiveness of neuronal (or synaptic) adaptation in generating prediction, which is a main finding of this study, does correlate well with its similar effectiveness in other contexts. For example, adaptation leads to robust prediction of the next symbol in a sequence, in an artificial grammar learning task (S. Womble and A. Treves, in preparation). As it often happens, the robustness of the mechanism appears to be reinforced by its simplicity and generality.

Does the success of mechanism B really offer a rationale for the differentiation between CA3 and CA1 in the mammalian hippocampus? At first sight, it does not. The effect is quite small. Yet, there is a limited but nonzero interaction between differentiation and the balance between localization and prediction, as seen by comparing Figures 7 and 8, and a further interaction with mechanism A1, as shown in Figure 10. These results are in line with those of my earlier analytical work (Treves, 1995), but extend them further, in two main directions. There, CA1 was shown to be able to add some more information to that retrieved from memory by CA3, but the comparison was between systems of unequal size. Second, the analysis was purely in terms of information content, with discrete attractor states, without any time dimension and the possibility to model prediction along continuous attractors. The computational advantage associated here with the differentiation is small, but it comes at zero cost or, to be more precise, at the presumably vanishing cost of the additional genetic specification of a differentiated circuitry. Its quantitative nature should be stressed: the figures show how misleading it would be to think of CA1 as simply adding a new qualitative facility, be it prediction or temporal pattern association or completion. Given the quantitative nature of the approach, it would in fact be important to be able to extend the analytical treatment of Battaglia and Treves (1998) to corroborate simulation results, but this seems impractical at present, in view of the analytical complexity of even the simple model considered here.

An important possibility to be considered (emphasized to me by Yasser Roudi, personal communication) is that the modest advantage reported here may have provided an initial bias for a process of differentiation that may have been stabilized later by other intervening changes in the organization of mammalian brains. For example, comparative anatomy (Seress, 1988) suggests that the CA3 and CA1 fields may have included similar numbers of neurons initially, with CA1 then expanding in cell counts more than CA3,

particularly in advanced species like primates. This differentiation in numbers could bring benefits that would have effectively piggy-backed on the earlier differentiation in connectivity. Likewise, neocortical expansion may have reinforced the usefulness of a postprocessing stage, CA1, gradually reexpanding the compressed CA3 representation being fed back to neocortex (Treves, 1995).

The model reported here does not include theta rhythm, which was considered in preliminary versions of the simulations. The inclusion of theta rhythm requires fixing a number of additional free parameters, thereby complicating the conceptual framework (Harris et al., 2002; Mehta et al., 2002; Lengyel et al., 2003). The expectation, however, is that the results obtained here would carry over to a model with theta, which has long been shown to easily display phase precession (Tsodyks et al., 1996). This could be the object of further study.

Another issue to be considered in future work is the difference in origin between PP inputs to CA3 and CA1, which arise from layer II and, respectively, layer III of entorhinal cortex. Introducing this further aspect of differentiation would obviously be important given hypotheses about key differences in the activity in the two layers. Even in the absence of any difference in activity, though, considerable effects might result from differences in connectivity. Medial and lateral perforant path projections from layer II reach the same cells in CA3 (and in DG), at different portions of their dendritic trees; whereas medial and lateral layer III projections tend to reach different cells in CA1 (Amaral and Witter, 1995). A partial cue activating only a limited portion of entorhinal cortex, therefore, might be relayed directly to only a fraction of CA1 cells, while the others would be entirely dependent on Schaffer collateral activation. This mechanism (similar to the model considered in Fulvi Mari et al., 2000) appears interesting, but requires a considerable extension of the simulations used here, which would have to include a representation of the position of cells along the CA3-CA1 axis of the hippocampus. A further extension would include representing the septotemporal axis, where differences in inputs (Amaral and Witter, 1995) might also produce interesting differences in response properties (Jung et al., 1994), including a probably apparent dissociation between storage and retrieval (Small et al., 2001). Representing cell position in the tissue involves a major departure from the simplicity of the present simulations, which rely on mapping units directly to place fields, and not to their position within the hippocampus. Simulations with both representations would have to be significantly larger in scale.

Are there predictions from these simulations that could be falsified experimentally? The effect of differentiation seems rather inaccessible to experiment at present, or at least until genetic manipulations are found that abolish the differentiation between CA3 and CA1 while leaving the overall system intact and functioning. Even then, the small effect predicted here would be difficult to assess experimentally. More directly accessible are the effects that apply independent of differentiation. The simulations predict a direct and positive relation between the extent of prediction and the degree of firing rate adaptation, and the latter might be manipulated, e.g., by acting on potassium channels or on the cholinergic system (Barkai and Hasselmo, 1994). This would be particularly interesting if contrasted with a negative relation of adaptation with

localization accuracy, indicated by the finding (e.g., Barkai et al., 1994) that suppression of adaptation enhances coding in simple static associative memories. Thus, the prediction ability would provide a rationale for the usefulness of adaptation, which would then be regulated by modulation to enhance coding precision when necessary. This balancing act between coding and prediction might apply in general, also outside the hippocampus, and lead to experimental tests of the opposing effects of adaptation in a number of systems. Within the hippocampus, the natural experimental model would, however, be the rat with its theta rhythm, which implies that the inclusion of theta in the model would have to be considered before extrapolating an interpretation of experimental results.

The existence of optimal sparseness values in the different fields might be investigated experimentally, if modulation of inhibition is found to be effective in varying sparseness, delicately, within a given field. Decoding and information measures could then be used to check for a localization capability non-monotonic with inhibition, in each field. All these are, in a sense, side aspects, which do not address the original contribution of these simulations, the comparison between the uniform and differentiated model. Still, their experimental validation would corroborate the idea that the general framework of the simulations—a computational approach that at the moment seems the only one able to deal with differentiation itself—is appropriate to indirectly study the operations carried out by the real hippocampus, at least in the rat.

## Acknowledgments

The helpful advice of Francesco Battaglia, Ines Samengo, and Mate Lengyel is gratefully acknowledged, as well as discussions with Ray Kesner, Edmund Rolls, and Bruce McNaughton. I am grateful also for the hospitality of the Institute for Theoretical Physics in Santa Barbara and of the Neural Systems, Memory and Aging Division of the Arizona Research Laboratories in Tucson.

## REFERENCES

- Abbott LF, Blum KI. 1996. Functional significance of long-term potentiation for sequence learning and prediction. *Cereb Cortex* 6:406–416.
- Amaral DG, Ishizuka N, Claiborne B. 1990. Neurons, numbers and the hippocampal network. *Prog Brain Res* 83:1–11.
- Amaral DG, Witter MP. 1995. Hippocampal formation. In: Paxinos C, editor. *The rat nervous system*. San Diego, CA: Academic Press. p 443–493.
- Amit DJ. 1995. The Hebbian paradigm reintegrated: local reverberations as internal representations. *Behav Brain Sci* 18:617–657.
- Barkai E, Hasselmo ME. 1994. Modulation of the input/output function of rat piriform cortex pyramidal cells. *J Neurophysiol* 72:644–658.
- Barkai E, Bergman RE, Horwitz G, Hasselmo ME. 1994. Modulation of associative memory function in a biophysical simulation of rat piriform cortex pyramidal cells. *J Neurophysiol* 72:658–677.
- Battaglia FP, Treves A. 1998. Attractor neural networks storing multiple space representations: a model for hippocampal place fields. *Phys Rev E* 58:7738–7753.

- Bi G-Q, Poo M-M. 1998. Activity-induced synaptic modifications in hippocampal culture, dependence on spike-timing, synaptic strength and cell type. *J Neurosci* 18:10464–10472.
- Blum KI, Abbott LF. 1996. A model of spatial map formation in the hippocampus of the rat. *Neural Comput* 8:85–93.
- Braitenberg V, Schüz A. 1991. *Anatomy of the cortex*. Berlin: Springer-Verlag.
- Brun VH, Otnæss MK, Molden S, Steffenach H-A, Witter MP, Moser M-B, Moser EI. 2002. Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. *Science* 296:2243–2246.
- Földiák P. 1991. Learning invariance from transformation sequences. *Neural Comput* 3:193–199.
- Fortin NJ, Agster KL, Eichenbaum HB. 2002. Critical role of the hippocampus in memory for sequences of events. *Nat Neurosci* 5:458–462.
- Foster DJ, Morris RGM, Dayan P. 2000. A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* 10:1–16.
- Frank LM, Brown EN, Wilson MA. 2001. A comparison of the firing properties of putative excitatory and inhibitory neurons from CA1 band the entorhinal cortex. *J Neurophysiol* 86:2029–2040.
- Freund TF, Buzsáki G. 1996. Interneurons of the hippocampus. *Hippocampus* 6:347–470.
- Fulvi Mari C, Panzeri S, Rolls ET, Treves A. 2000. A quantitative model of information processing in CA1. In: Baddeley R, Földiák P, Hancock P, editors. *Information theory and the brain*. Cambridge, UK: Cambridge University Press. p 273–289.
- Gaffan D, Parker A, Easton A. 2001. Dense amnesia in the monkey after transection of fornix, amygdala and anterior temporal stem. *Neuropsychologia* 39:51–70.
- Gloor P. 1997. *The temporal lobe and limbic system*. New York: Oxford University Press.
- Harris KD, Henze DE, Hirase H, Leinekugel X, Dragol G, Czurkó A, Buzsáki G. 2002. Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature* 417:738–741.
- Hasselmo ME, Schnell E. 1994. Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1: computational modeling and brain slice physiology. *J Neurosci* 14:3898–3914.
- Hasselmo ME, Schnell E, Barkai E. 1995. Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *J Neurosci* 15:5249–5262.
- Hasselmo ME, Wyble B, Wallenstein G. 1996. Encoding and retrieval of episodic memories: role of cholinergic and GABAergic modulation in hippocampus. *Hippocampus* 6:693–708.
- Henze DA, Wittner L, Buzsáki G. 2002. Single granule cells reliably discharge targets in the hippocampal CA3 network in vivo. *Nat Neurosci* 5:790–795.
- Huerta PT, Sun LD, Wilson MA, Tonegawa S. 2000. Formation of temporal memory requires NMDA receptors within CA1 pyramidal neurons. *Neuron* 25:473–480.
- Jensen O, Lisman JE. 1996. Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells. *Learn Mem* 3:279–287.
- Jensen O, Tesche CD. 2002. Frontal theta activity in humans increases with memory load in a working memory task. *Eur J Neurosci* 15:1395–1399.
- Ji Z, Staubli U. 2002. Presynaptic kainite receptors play different physiological roles in mossy fiber and associational-commissural synapses in CA3 of hippocampus of adult rats. *Neurosci Lett* 331:71–74.
- Jung MW, McNaughton BL. 1993. Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus* 3:165–182.
- Jung MW, Wiener SI, McNaughton BL. 1994. Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *J Neurosci* 14:7347–7356.
- Kametani H, Kawamura H. 1990. Alterations in acetylcholine release in the rat hippocampus during sleep-wakefulness detected by intracerebral dialysis. *Life Sci* 47:421–426.
- Kesner RP, Gilbert PE, Barua LA. 2002a. The role of the hippocampus in memory for the temporal order of a sequence of odours. *Behav Neurosci* 116:286–290.
- Kesner RP, Gilbert PE, Lee I. 2002b. Subregional analysis of hippocampal function in the rat. In: Squire LR, Schacter DL, editors. *Neuropsychology of memory*. 3rd ed. New York: Guilford Press. p 395–411.
- Lassalle J-M, Bataille T, Halley H. 2000. Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the Morris navigation task. *Neurobiol Learn Mem* 73:243–257.
- Lengyel M, Szatmáry Z, Érdi P. 2003. Dynamically detuned oscillations account for the coupled rate and temporal code of place cell firing. *Hippocampus* 13:700–714.
- Levy WB. 1989. A computational approach to hippocampal function. In: Hawkins RD, Bower GH, editors. *Computational models of learning in simple neural systems*. (The psychology of learning and motivation. vol 23.) San Diego, CA: Academic Press. p 243–305.
- Levy WB. 1990. Maximum entropy prediction in neural networks. In: Washington, DC: International joint conference on neural networks. p 17–110.
- Levy WB, Colbert CM, Desmond NL. 1990. Elemental adaptive processes of neurons and synapses: a statistical/computational perspective. In: Gluck M, Rumelhart D, editors. *Neuroscience and connectionist theory*. Hillsdale, NJ: Lawrence Erlbaum Associates. p 187–235.
- Levy WB, Wu XB, Baxter RA. 1995. Unification of hippocampal function via computational encoding considerations. *Int J Neural Syst* 6(suppl):71–80.
- Marr D. 1971. Simple memory: a theory for archicortex. *Philos Trans Roy Soc Lond B* 262:24–81.
- Marrosu F, Portas C, Mascia MS, Casu MA, Fa M, Giagheddu M, Imperato A, Gessa GL. 1995. Microdialysis measurement of cortical and hippocampal acetylcholine release during sleep-wake cycle in freely moving cats. *Brain Res* 671:329–332.
- McNaughton BL, Morris RGM. 1987. Hippocampal synaptic enhancement and information storage. *Trends Neurosci* 10:408–415.
- Mehta MR. 2001. Neuronal dynamics of predictive encoding. *Neuroscientist* 7:490–495.
- Mehta MR, Barnes CA, McNaughton BL. 1997. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc Natl Acad Sci USA* 94:8918–8921.
- Mehta MR, Quirk MC, Wilson MA. 2000. Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* 25:504–506.
- Mehta MR, Lee AK, Wilson MA. 2002. Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* 417:741–746.
- Minai AA, Levy WB. 1993. Sequence learning in a single trial. *INNs World Congr Neural Netw* 2:505–508.
- Minai AA, Barrows GL, Levy WB. 1994. Disambiguation of pattern sequences with recurrent networks. *INNs World Congr Neural Netw* 4:178–181.
- Morishita W, Alger BE. 2000. Differential effects of the group II mGluR agonist, DCG-IV, on depolarization-induced suppression of inhibition in hippocampal CA1 and CA3 neurons. *Hippocampus* 10:261–268.
- Nakazawa K, Quirk MC, Chitwood RA, Watanabe M, Yeckel MF, Sun LD, Kato A, Carr CA, Johnston D, Wilson MA, Tonegawa S. 2002. Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science* 297:211–218.
- O'Keefe J, Recce ML. 1993. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3:317–330.
- Rawlins JNP. 1985. Associations across time: the hippocampus as a temporary memory store. *Behav Brain Sci* 8:479–496.

- Roberts PD. 1999. Computational consequences of temporally asymmetric learning rules. I. Differential Hebbian learning. *J Comput Neurosci* 7:235–246.
- Robertson RG, Rolls ET, Georges-Francois P. 1998. Spatial view cells in the primate hippocampus: effects of removal of view details. *J Neurophysiol* 79:1145–1156.
- Rolls ET. 1989. Functions of neuronal networks in the hippocampus and cerebral cortex in memory. In: Cotterill R, editor. *Models of brain function*. New York: Cambridge University Press. p 15–33.
- Rolls ET, Treves A, Robertson RG, Georges-Francois P, Panzeri S. 1998. Information about spatial view in an ensemble of primate hippocampal cells. *J Neurophysiol* 79:1797–1813.
- Schultz SR, Panzeri S, Rolls ET, Treves A. 2000. Quantitative analysis of a Schaffer collateral model. In: Baddeley R, Földiák P, Hancock P, editors. *Information theory and the brain*. Cambridge, UK: Cambridge University Press. p 257–272.
- Seress L. 1988. Interspecies comparison of the hippocampal formation shows increased emphasis on the regio superior in the Ammon's horn of the human brain. *J Hirnforsch* 29:335–340.
- Skaggs WE, McNaughton BL, Wilson MA, Barnes CA. 1996. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6:149–172.
- Small SA, Nava AS, Perera GM, DeLaPaz R, Mayeux R, Stern Y. 2001. Circuit mechanisms underlying memory encoding and retrieval in the long axis of the hippocampal formation. *Nat Neurosci* 4:442–448.
- Traub RD, Wong RK, Miles R, Michelson H. 1991. A model of a CA3 hippocampal pyramidal neuron incorporating voltage-clamp data on intrinsic conductances. *J Neurophysiol* 66:635–650.
- Treves A. 1995. Quantitative estimate of the information relayed by the Schaffer collaterals. *J Comput Neurosci* 2:259–272.
- Treves A. 1997. On the perceptual structure of face space. *Biosystems* 40:189–196.
- Treves A. 2001. Information coding in higher sensory and memory areas. In: Moss F, Gielen S, editors. *Neuro-informatics and neural modeling*. (Handbook of biological physics. vol 4.) Amsterdam: Elsevier. p 825–852.
- Treves A. 2003. Computational constraints that may have favoured the lamination of sensory cortex. *J Comput Neurosci* 14:271–282.
- 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, Samengo I. 2002. Standing on the gateway to memory: shouldn't we step in? *Cogn Neuropsychol* 19:557–575.
- Tsodyks MV, Skaggs WE, Sejnowski TJ, McNaughton BL. 1996. Population dynamics and theta rhythm phase precession of hippocampal place cell firing: a spiking neuron model. *Hippocampus* 6:271–270.
- Ułinski PS. 1990. The cerebral cortex of reptiles. In: Jones EG, Peters A, editors. *Cerebral cortex*. (Comparative structure and evolution of cerebral cortex. vol 8A.) New York: Plenum Press. p 139–215.