Gatsby Computational Neuroscience Unit
University College London
http://www.gatsby.ucl.ac.uk

17 Queen Square, London
WC1N 3AR, United Kingdom
+44 20 7391 1176

Funded in part by the Gatsby Charitable Foundation.

October 19, 1999

GCNU TR 1999–005

The involvement of recurrent connections in area CA3 in establishing the properties of place fields: A model.

Szabolcs Káli **Peter Dayan**

Gatsby Unit

Abstract

Strong constraints on the neural mechanisms underlying the formation of place fields in the rodent hippocampus come from the systematic changes in spatial activity patterns that are consequent on systematic environmental manipulations. We describe an attractor network model of area CA3 in which local, recurrent, excitatory and inhibitory interactions generate appropriate place cell representations from location- and direction-specific activity in the entorhinal cortex. The model has two modes of operation, learning and recall, which are switched under neuromodulatory control. During learning, mossy fiber inputs impose activity patterns on CA3. Then, through Hebbian plasticity in the recurrent excitatory connections, attractors in CA3 are sculpted appropriately, and through Hebbian plasticity in the perforant path inputs, entorhinal activity is associated with these attractors. During recall, the spatial characteristics of the place fields are controlled by the way that the perforant path input selects amongst the attractors. Depending on the training experience provided, the model generates place fields that are either directional or non-directional, and which change in accordance with experimental data when the environment undergoes simple geometric transformations. Representations of multiple environments can be stored and recalled with little interference, and these have the appropriate degrees of similarity in visually similar environments.

The involvement of recurrent connections in area CA3 in establishing the properties of place fields: A model.

Szabolcs Káli **Peter Dayan**

Gatsby Unit

1 Introduction

The hippocampus is known to be involved in spatial learning and memory in rodents. Some of the most convincing evidence for this is the presence of place cells in areas CA3 and CA1 of the hippocampus (O'Keefe and Dostrovsky, 1971; O'Keefe, 1976), and of many other types of spatially selective cells in neighboring areas (Jung and McNaughton, 1993; Quirk et al., 1992). Principal neurons in CA3 and CA1 are active only when the animal is located in a well-defined local region of the environment (a place field)(Muller et al., 1987), and collectively provide a population code for spatial position (Wilson and McNaughton, 1993).

Pyramidal cells in area CA3 receive the majority of their inputs from other CA3 pyramidal cells (Amaral and Witter, 1989; Rolls, 1996). This recurrent network underlies many attractor network models of the way that the hippocampus acts as a memory (Marr, 1971; McNaughton and Morris, 1987; Rolls, 1996; Levy, 1996; Hasselmo et al., 1996). It is reasonable to assume that the mechanisms which allow the hippocampus to function as a memory device also affect the firing patterns observed during spatial behavior. Even so, most of the models which are intended to account for various properties of place cells ignore the CA3 recurrent connections (Zipser, 1985; Sharp, 1991; Touretzky and Redish, 1996; Burgess et al., 1997), or use them in a rather abstract way (Muller et al., 1996). Therefore, it is an important issue whether a model which exploits the proposed attractor dynamics of the CA3 network can account for experimental data on place cells.

The model of Samsonovich and McNaughton (1997) was the first to explore the consequences of the CA3 attractor network for the place cell representation. Their model assumes the existence of a collection of independent continuous sets of attractors realized by the CA3 recurrent network, and successfully accounts for some of the basic experimental observations about place cells. However, in a model with fixed, independent sets of attractors, it is hard to explain the recent experimental findings by Skaggs and McNaughton (1998), who found partially overlapping place cell representations in two distinct but similar-looking parts of an apparatus. Such models generally predict either identical or completely different firing patterns in this situation. In addition, Samsonovich and McNaughton's (1997) model does not address the question how the strengths of the CA3 recurrent connections, which are critical for the emergence of appropriate attractors, become established. There is substantial evidence for synaptic plasticity in most major hippocampal pathways, and at least indirect evidence for the plasticity of the CA3 recurrent collateral connections. These activity-dependent synaptic changes may provide the means for setting up the appropriate connection strengths.

Brunel and Trullier (1998) and ourselves (Káli and Dayan, 1998) independently implemented models which rely on modifiable recurrent connections in CA3 to explain the differences in the directionality of place cells in different kinds of environment. However, the strongest challenge for models, and particularly models based on attractor networks, comes from data on the behavior of place cells in multiple environments that are similar, or are re-

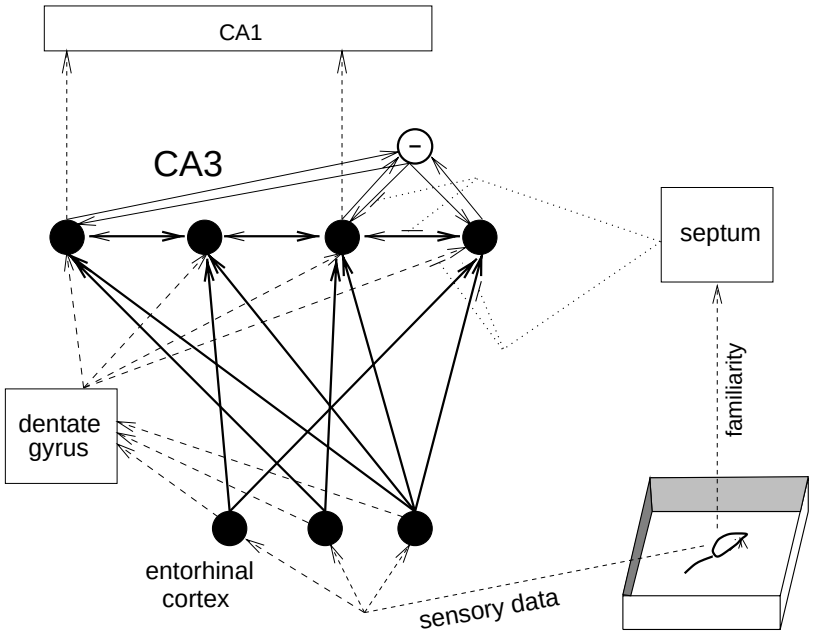

Figure 1: Model architecture. The inputs to the network are the activities of neurons in entorhinal cortex, which are determined by sensory features in the environment. This representation is then transformed by feedforward pathways (the direct perforant path connections to CA3 and the pathway through the dentate gyrus) and recurrent processing in area CA3, which involves lateral connections between CA3 pyramidal cells (filled circles) as well as their connections with an inhibitory neuron (open circle). The solid lines indicate neuronal connections that are modelled explicitly, and the thick ones (the CA3 recurrent connections and the perforant path inputs to CA3) are modifiable. Each type of connection is all-to-all in the model. All inputs to CA3 pyramidal cells are gated by neuromodulatory signals (dotted lines) from septal nuclei, whose activity depends on familiarity with the current environment.

lated by simple geometric manipulations. In this paper, we present an attractor model with appropriate behavior in these cases.

Place field formation in simple environments

We start from the inputs to the model. Instead of building a detailed model of rodent sensory processing, we consider as inputs the firing rates of pyramidal neurons in superficial layers of entorhinal cortex (EC). This structure provides most cortical inputs to the hippocampal formation. We utilize experimental data as well as computational considerations to propose some general constraints on how the EC spatial representation may depend on sensory features of the environment, and also suggest a plausible functional form for this dependence in the simple case when all the directly available information about location originates from the walls of the experimental apparatus.

The next step is to describe how the input representation is transformed into the place cell representation. This transformation involves two processes in our model: first, passage through the two feedforward pathways from EC to CA3; and second, dynamic processing by the CA3 network. The result of this dynamic processing, the measurable place cell activity, depends on the feedforward inputs to CA3, as well as the neuromodulatorily controlled global state of the network. Both the neural components involved (as depicted in Figure 1) and their dynamics, including neuronal activities, synaptic plasticity, and neuromodulation, are described in detail later in this section. The values of the parameters used in the simulations

are summarized in Table 1 at the end of the description of the model (on page 8). We then demonstrate how the model generates place fields in the simple case of a single environment surrounded by walls; more complex cases are analyzed in subsequent sections.

Input representation

Unfortunately, there is relatively little direct experimental evidence about the nature of spatial representations in EC, and especially about how these depend on details of the environment. Although entorhinal neurons are found to be spatially selective (Barnes et al., 1990; Quirk et al., 1992), they appear to be much noisier and more broadly tuned than place cells in the hippocampus. Quirk et al. (1992) also found them to be more "sensory bound" than hippocampal cells in that their firing fields transform in a smooth manner following substantial changes in the shape of the environment. This is very unlike the complete remapping seen in place cells under similar circumstances (Muller and Kubie, 1987). The anatomy of the inputs to EC is rather better understood (Burwell and Amaral, 1998). Many of the inputs to EC come from higher-order association areas, which contain complex representations of the sensory information available to the animal. In particular, cells may convey information about both the identity of a perceived object and its location with respect to the animal, or, to put it differently, about the location of the rat with respect to particular objects in the world. Such information about multiple objects may be combined in EC in order to form a more reliable view-based representation of the animal's location in space. Spatial information derived from path integration may also be available.

In the model, each EC cell is assumed to respond to a subset of the available cues. Based on the suggestion that EC is involved in conjunctive coding (Myers et al., 1995), each EC cell in our model combines the sources of spatial information it is sensitive to in a conjunctive manner. Since the animal's sensory experience depends on both its position and the direction it faces, we assume that the activity of entorhinal neurons is head direction as well as location dependent. A model EC cell fires maximally when all the cues it is sensitive to are in the position corresponding to the cell's preferred location and orientation, and activity diminishes as some or all of the sources of information signal a different location or orientation. We achieve this by multiplying together Gaussian tuning curves, each of which is tied to the location of a different cue and peaks at the preferred location of the cell. We assume that these individual tuning curves can have different variances, which may reflect the uncertainty of the animal about its location based solely on that cue.

In cases where the environment has walls, these are probably important sources of spatial information, and we assume that the tuning curve components tied to the walls of a rectangular apparatus are ridge-like functions with Gaussian dependence on the distance from the wall. The variances of these tuning functions may also depend on the location and heading of the animal – in particular, we assume that the variance is lower if the animal is closer to the wall, or facing *away* from the wall. The latter dependence is based on the assumption that the animal is coming from somewhere nearer the wall and has been able to maintain its location accurately using path integration. For EC neurons in our model, components of the tuning function tied to particular walls have the following functional form:

$$z_k^a = e^{-\frac{(d_a - d_k^{EC,a})^2}{2\sigma_{EC,a}^2}}, \tag{1}$$

where k indexes the neuron, d_a is the actual distance from wall a (a can be either N, W, S or E), $d_k^{EC,a}$ is the distance from wall a of the neuron's preferred location, and $\sigma_{EC,a}$ is the width of this component, which depends on the current position and heading of the animal like

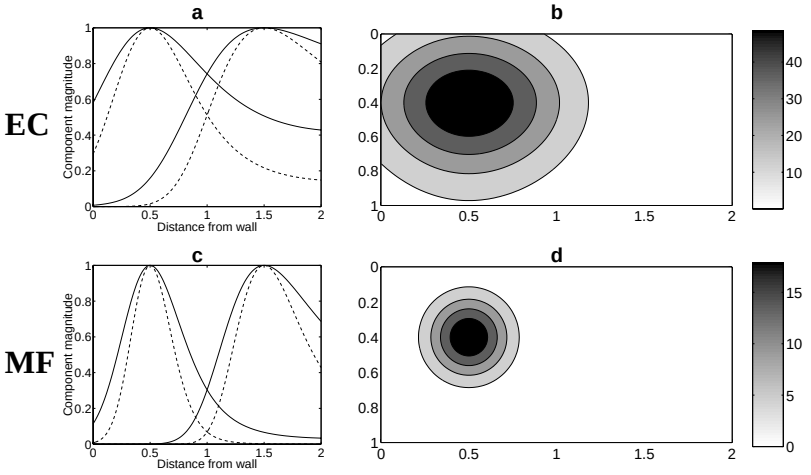

Figure 2: Input components and net spatial tuning. **a**: Two examples of how a spatial component of the tuning function of cells in entorhinal cortex (EC) depends on the distance of the rat from the wall to which that component is tied. The preferred distances from the wall are 0.5 and 1.5 (where 1 and 2 are the lengths of the walls of the rectangular box that the model rat explores), respectively; for each preferred distance, the solid curve is for the case when the rat is facing the wall, and the dashed curve is for the opposite head direction. Note how the width of the curve changes with preferred distance and actual head direction. **b**: The net two-dimensional tuning of a sample EC neuron in the rectangular box; the current heading of the rat at each location is the same as the preferred head direction of the cell. **c** and **d**: Plots similar to **a** and **b**, for the mossy fiber (MF) inputs to CA3; note that the tuning is much sharper here, due to the orthogonalization property of the dentate gyrus. For all contour plots in this article, darker shading indicates higher activity, and the contour lines are at 20, 40, 60, and 80% of the maximum activity of the given cell or set of cells.

$$\sigma_{EC,a} = \sigma_{EC}(1 + 0.35 d_a^2)(1 + 0.2\cos(\phi - \phi_a)), \tag{2}$$

where ϕ is the current head direction of the animal, ϕ_a is the direction of wall a (0, $\pi/2$, π, and $-\pi/2$ for N, E, S, and W, respectively), and σ_{EC} is constant. Equation 1 and the positional dependence in Equation 2 are similar to the expressions describing the spatial tuning of "sensory" cells in the model of Burgess et al. (1997). The numerical values of the parameters in the above equations have been chosen for environments of around the size employed in most relevant experiments.

The components tied to different walls are then multiplied together, and a net dependence on direction (which is assumed to be independent from the spatial components) is also included. This yields the following total activation of a model EC neuron as a function of the rat's location and heading:

$$z_k = b\, z_k^N z_k^W z_k^S z_k^E e^{\rho_{EC}\cos(\phi - \phi_k^{EC})}, \tag{3}$$

where b is a constant to set the scale, and the last term describes the dependence on head direction as a circular Gaussian function (with sharpness parameter ρ_{EC}) of the difference between the current head direction ϕ and the cell's preferred heading ϕ_k^{EC}. Figure 2a shows two examples of the spatial and directional dependence of input components in EC, while Figure 2b displays the resulting net spatial tuning for a sample EC neuron.

Feedforward connections

There are two separate neural pathways from EC to area CA3 (see Fig. 1) which have quite different characteristics and likely serve different computational purposes (McNaughton and Morris, 1987; Treves and Rolls, 1992). The perforant pathway (PP) provides a direct connection between these areas, and has a large degree of divergence and convergence. Thus, CA3 cells can sample the EC representation very effectively. In the model, we implement this property using all-to-all connections between EC and CA3 neurons, although this is obviously a simplification. The strengths of these connections are initially set to zero, but they are assumed to be modifiable by associative Hebbian learning.

The perforant pathway also projects to the dentate gyrus (DG), which in turn provides a second set of feedforward inputs to CA3 through the mossy fibers. Dentate granule cells are spatially selective, and they have also been found to be sensitive to direction; however, data only exist for linearly restricted environments (Jung and McNaughton, 1993). Unlike EC neurons, dentate granule cells have sharper spatial tuning than CA3 place cells, and we assume that they are also sharply tuned for head direction. Episodic memory theories of hippocampal function suggest that an important function of the DG is that of orthogonalization, i.e. reducing the similarity between input patterns in order to facilitate their discrimination (Treves and Rolls, 1994; O'Reilly and McClelland, 1994). One way the DG is thought to decrease pattern overlap is to implement a sparser representation (perhaps through direct competitive interactions), and indeed, the proportion of active cells in the DG at any given time is reported to be only about 0.5% (B.L. McNaughton, cited by O'Reilly and McClelland, 1994; see also Jung and McNaughton, 1993).

A typical CA3 pyramidal cell receives on the order of 50 mossy fiber (MF) inputs, which are thought to be relatively powerful (McNaughton and Morris, 1987). Combined with the sparseness of the DG representation, this means that a CA3 neuron is very unlikely to have more than one active mossy fiber input at any given time. In circumstances under which CA3 cells are driven primarily by these inputs, place cells essentially inherit the tuning characteristics of their afferent granule cells. We assume, for simplicity, that each CA3 cell has at most one active MF input in any given environment. This defines the preferred location and direction for that neuron, although these can, of course, be altered by the recurrent connections in CA3. Multiple active MF inputs may explain why some place cells have multiple place fields even in simple environments (Muller et al., 1987); however, we can safely ignore this complexity for the purpose of this paper. In addition, in order to make better use of the limited number of cells we can implement in our numerical simulations, all our model CA3 pyramidal cells are activated by MF inputs somewhere in any given environment.

In its current form, the model considers both the mossy fiber connections and the perforant path connections from EC to DG as being fixed. Since our goal is to model activity in CA3, and that is completely determined by its inputs and internal dynamics, we can therefore skip modeling the dentate gyrus explicitly, and proceed by characterizing how the MF input to CA3 (which results from processing in DG) depends on the characteristics of the environment. We assume that, for any single environment, the MF input to CA3 place cells has a similar functional form to the tuning function of EC cells described in the previous section, but both the spatial and the directional tuning is assumed to be sharper as a result of sparsification in DG (see Fig. 2c,d). This can be achieved by replacing the spatial spread parameter σ_{EC} with a smaller value, σ_{MF}, and by replacing ρ_{EC}, characterizing the sharpness of directional tuning, with a larger ρ_{MF} in Equations 2-3. The proposed orthogonalization property of the dentate gyrus becomes more pronounced when we look at multiple environments. We assume that, except when two environments are quite similar, the MF inputs to CA3 in two different environments are completely unrelated. We will return to the case of exceptionally similar environments in a later section.

Two-mode hypothesis

Our model is grounded in two assumptions. The first is that observed place cell activity patterns reflect the stable states of the CA3 recurrent network. The second is that the network establishes new attractors to represent novel situations. This requires activity patterns other than previously established attractors to be realized in CA3, which can only happen if the recurrent connections are suppressed, and the feedforward inputs are allowed to dominate. Several different solutions to this problem have been proposed (Treves and Rolls, 1992; Hasselmo et al., 1995, 1996; Hasselmo, 1999), and we adopt the suggestion of Hasselmo et al. (1996) which is based on experimental data on the effects of septal (cholinergic and $GABA_B$-receptor-mediated) modulation in the hippocampus.

In the model, the hippocampal network has two modes of operation. When the rat is in a familiar environment, no learning takes place in any of the connections, the MF inputs are relatively less effective than the PP connections and CA3 recurrent synapses, and the intrinsic dynamics of the recurrent network dominates activity in CA3, leading to previously established attractors. This is called "recall mode". On the other hand, when the rat first encounters a new environment, learning in both the PP inputs to CA3 and the recurrent connections is enabled, synaptic transmission through the recurrent connections is suppressed, inhibition in CA3 is reduced, and inputs through the mossy fiber connections dominate. This state of the network is called "learning mode". Later we show how the patterns of weights set up during learning produce the patterns of place cells activity seen subsequently.

The selection between the two modes is based on a general notion of familiarity with the current environment, and is presently performed by hand in the model. Recall mode is entered after a fixed amount of exploration per unit area of the environment, or immediately upon entry into the environment if it is similar enough to an environment already explored; otherwise, learning is initiated. Similarity to a familiar environment is assumed to be reflected in the activity of a brain area outside the scope of our current model; this activity would in turn affect neuromodulatory cells in the septal nuclei. In fact, Hasselmo et al. (1996) have implemented a model of associative memory in the hippocampus using feedback regulation of septal modulation by activity in area CA1, which, in turn, reflects familiarity with current input patterns. At present, we skip learning in a new environment only if it shares most sensory features with an environment which is completely familiar to the animal, i.e., one that has been thoroughly explored.

CA3 neural architecture and dynamics

The main aspect of hippocampal circuitry we actually implement is the CA3 recurrent network (Figure 1). The model CA3 contains a collection of N pyramidal cells, each connected to all the others through modifiable weights. This high degree of connectivity mimics the extensive recurrent collateral connections of CA3 pyramidal neurons (Ishizuka et al., 1990; Li et al., 1994). Due to the relatively small number of neurons in the model, the number of connections per cell is still much lower than in reality, even though the degree of connectivity is higher. However, in the case of such extensive connectivity, the actual number of connections only enters the calculations as a constant scaling factor for the individual weights as long as the cells a particular neuron connects to can be considered from a functional point of view as a random sample, and as long as the number of connections per neuron is high enough (and any one connection is weak enough) that neural responses are determined by averaged population effects.

Local feedback and feedforward inhibition are thought to play an important and complex role in neural dynamics in CA3. Inhibitory interneurons are spatially much less selective than pyramidal neurons, but their activity during locomotion changes periodically at the theta frequency. We ignore this temporal variation as well as the diversity of interneurons and patterns

of connectivity, and include in the model a single global inhibitory neuron which fosters competition between stored patterns and keeps global activity levels approximately constant. This cell receives input from all the excitatory neurons, and provides inhibitory feedback to each which is proportional to the product of the firing rate of the inhibitory neuron and the depolarization of its postsynaptic target.

We adapt the equations introduced by Wilson and Cowan (1972) to model the dynamics of the CA3 neural population. The following set of equations describes how the membrane potential of CA3 cells in our model changes over time:

$$\begin{aligned} \tau \dot{u_i} &= -u_i + A_R \sum_j J_{ij} g_u(u_j) - A_R h g_v(v) u_i + A_R I_i^{PP} + A_L I_i^{MF} \\ \tau' \dot{v} &= -v + w \sum_j g_u(u_j), \end{aligned} \tag{4}$$

where u_i is the membrane potential of the i'th pyramidal cell, v is the membrane potential of the global inhibitory cell, τ and τ' are the membrane time constants for pyramidal neurons and the inhibitory cell, respectively, J_{ij} is the strength of the connection from neuron j to neuron i, h is the efficacy of inhibition, w represents the strength of the excitatory connection from any one pyramidal cell onto the inhibitory cell, and I_i^{PP} and I_i^{MF} are the inputs to cell i through the perforant path and the mossy fibers, respectively. $g_u(u) = \beta[u - \mu]_+$ is the threshold linear activation function for the pyramidal cells, where $[...]_+$ makes all negative arguments zero while leaving positive numbers unaffected, μ stands for the threshold and β is the slope of the activation function above the threshold. Similarly, $g_v(v) = \gamma[v - \nu]_+$ for the inhibitory neuron. A_R and A_L indicate the mode of operation; in learning mode, $A_L = 1$ and $A_R = 0$, while during recall, $A_L = 0$ and $A_R = 1$.

The value of τ' plays an important role in defining the exact temporal dynamics of the network; however, it has no effect on the identity of the fixed points of the dynamics. Therefore, we set $\tau' = 0$ in the simulations, so that v is always equal to $w \sum_j g_u(u_j)$, which simplifies the theoretical treatment of the model, and makes the simulations numerically more stable. We conducted simulations to verify that, within a wide range of the parameters, this manipulation does not affect the qualitative dynamical behavior of the model and actually leads to the same stable patterns of activity (but see Discussion for further comments on unmodeled dynamical behaviors).

Synaptic learning

Perforant path and recurrent weights are acquired during the learning phase. We assume that all plastic weights change according to a Hebbian rule, that is, the weight change is proportional to both the presynaptic activity and the degree of postsynaptic depolarization. An exponential weight decay term is also included to prevent weights from growing indefinitely. We assume that weight changes only occur in the presence of postsynaptic activity. This leads to the following update equations for perforant path weights W_{ik} and recurrent weights J_{ij}:

$$\alpha \dot{W}_{ik} = \Theta(g_u(u_i))(\kappa u_i z_k - W_{ik}), \tag{5}$$

where α is the time constant of weight decay, κ sets the learning rate, Θ is the unit step function, and z_k is the activity of neuron k in entorhinal cortex; similarly,

$$\alpha \dot{J}_{ij} = \Theta(g_u(u_i))(\kappa u_i g_u(u_j) - J_{ij}). \tag{6}$$

We significantly simplify exploration during learning by imagining that the rat receives even exposure to all combinations of location and head direction allowed by the apparatus (and the movement pattern followed). If we assume that weight changes occur more slowly than the

ρ_{MF}	1.5	τ	100	β	1	κ	16
ρ_{EC}	0.5	τ'	0	μ	80	b	100
σ_{MF}	0.2	h	3	γ	1	N	1200
σ_{EC}	0.4	w	0.005	ν	12		

Table 1: Model parameters. The table displays the values of model parameters used in the simulations. σ_{MF} and σ_{EC} are in units such that the shorter side of the rectangular environment used in most simulations is of unit length. τ and τ' are given in time steps, and all other quantities are in their natural units. Note that since the parameters only appear in certain combinations in the equations, some groups of parameters can be changed together appropriately without affecting the behavior of the model.

time required to sample the environment efficiently, Equations 5 and 6 show that the weights become proportional to the average of the product of the presynaptic firing rate and the postsynaptic depolarization, taken over all locations and directions where the postsynaptic cell fires; i.e., $J_{ij} = \kappa\langle u_i g_u(u_j)\rangle$ and $W_{ik} = \kappa\langle u_i z_k\rangle$, where the angle brackets represent averaging over all positions and headings of the animal where $g_u(u_i) > 0$ (or, equivalently, where $u_i > \mu$).

The neural dynamics described by Equations 4 is simplified substantially in this phase by taking $A_R = 0$. Assuming that the MF inputs change more slowly than the membrane time constant, the membrane potential of CA3 place cells during the learning phase is given by $u_i = I_i^{MF}$. Substituting this expression into the above equations yields the following expressions for the weights at the end of the learning phase:

$$J_{ij} = \kappa \frac{\iiint\limits_{I_i^{MF}>\mu} I_i^{MF} g_u(I_j^{MF})\, dx\, dy\, d\phi}{\iiint\limits_{I_i^{MF}>\mu} dx\, dy\, d\phi} \quad \text{and} \quad W_{ik} = \kappa \frac{\iiint\limits_{I_i^{MF}>\mu} I_i^{MF} z_k\, dx\, dy\, d\phi}{\iiint\limits_{I_i^{MF}>\mu} dx\, dy\, d\phi} \tag{7}$$

The result is that the CA3 recurrent weights acquired during learning depend on the spatial correlation between MF inputs to the two neurons, while the perforant path weights depend on the correlation between MF inputs and activities in entorhinal cortex. Note that, unlike in Hasselmo et al.'s (1996) model, synapses are modified even when their efficacy is reduced to zero by neuromodulation, i.e., when the post-synaptic effect of perforant path and recurrent connections is negligible in the learning phase.

Recall dynamics

During the recall phase, the neural dynamics is governed by the full Equations 4 with $A_L = 0$ and $I_i^{PP} = \sum_j W_{ik} z_k$, and W_{ik} and J_{ij} are set to the values acquired in the learning phase. We solve these equations by numerical integration using Euler's method, doing a fixed number of iterations. We found that, within a broad range of model parameters, the network always settles into a stable state by the end of the iterations. Furthermore, for most initial CA3 activity patterns, the same final state is reached for given feedforward inputs. This shows that these states are actually attractors of the neural dynamics, and that they have suitably large basins of attraction. The final state of the network was determined for different input patterns in EC, representing different positions and head directions of the animal over a grid that covered the whole environment. The firing rate map for a given cell is defined as the final activity of that cell in recall mode as a function of the actual location and head direction of the animal.

In order to explore the activity patterns generated in our model, we first simulated the network dynamics using weights resulting from the random exploration of a rectangular box (with one side twice as long as the other). The results of these simulations are summarized in Figure 3, which displays activities in EC, net perforant path inputs and final activities in CA3

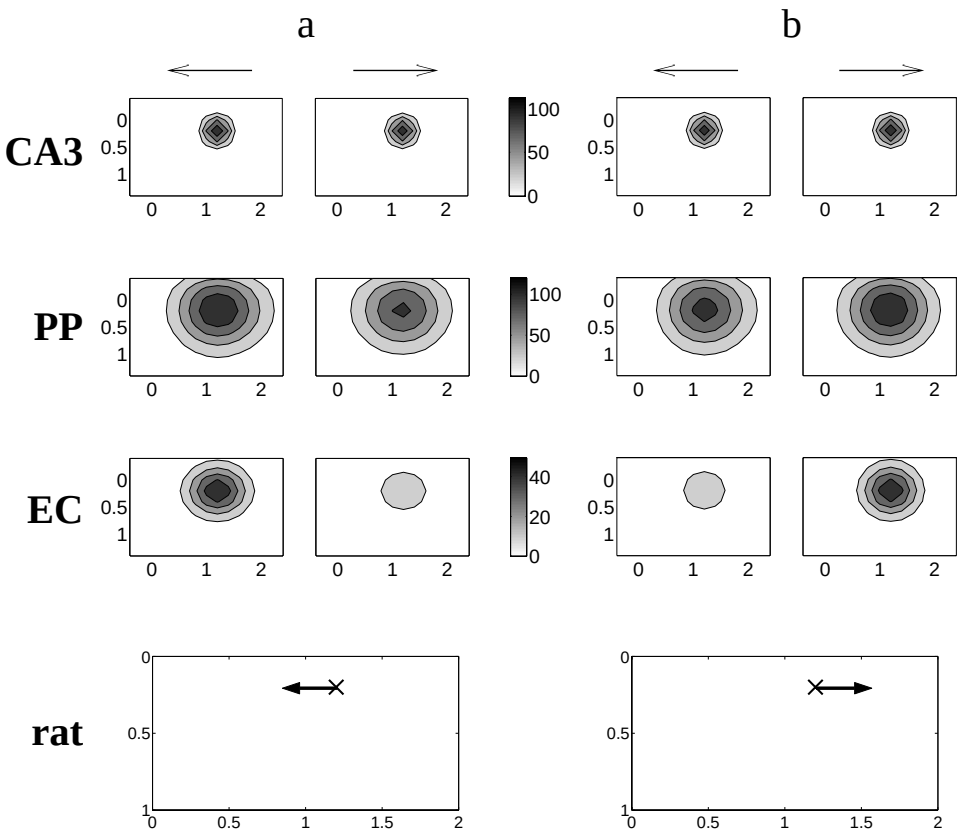

Figure 3: The formation of non-directional place fields. **a,b**: The bottom plot in each case shows the actual position (indicated by the cross) and head direction (indicated by the arrow) of the rat in the environment. The other plots show the activities of cells in entorhinal cortex (EC), the net perforant path (PP) inputs to CA3 neurons (I_i^{PP}), and the final activities of the same place cells (marked CA3), as a function of the preferred location of the neuron; the two columns in both **a** and **b** are for cells with preferred head direction indicated by the arrow above each column. The color bars are for both parts of the figure.

for all cells with two particular (opposite) preferred head directions, when the model rat is at a given location, facing in a particular direction. Figure 3 shows that the final states of the model CA3 network resemble thresholded two dimensional Gaussian bumps of activity in the space where place cells with a given preferred heading (defined as the preferred heading of their active mossy fiber input) are arranged on an imaginary plane according to their preferred locations. This type of solution can emerge spontaneously from the network dynamics even in the absence of external inputs, in which case the location of the bump is random – i.e., determined by the initial neural activities, as well as various other factors including the distribution of preferred locations and directions of the neurons. In our model, the uniform distribution of preferred locations and head directions over a grid leads to possible final activity profiles which are centered on the same grid locations. Making the grid of preferred locations finer by increasing the number of place cells would result in a more continuous set of possible final states. However, in the presence of inputs, the discrete nature of the attractor set becomes irrelevant. In the presence of even a small perforant path input to CA3, the location of the bump is determined by this input so that the activity profile provides the best possible fit to the input. The position of the peak can now vary continuously, and the shape of the activity profile is basically unaffected. This holds in our model if the net feedforward input to the most active CA3 neurons is between roughly 1% and 30% of the summed input they receive from other CA3 cells; in most simulations we set the relative efficacies of perforant path and

recurrent synapses so that this ratio is about 5%.

Figure 3a illustrates how inputs are used by the network to effectively select one of the possible final states. First of all, the EC activity pattern (which is determined by sensory features in the environment as already described) gives rise to a pattern of perforant path inputs to CA3 which is centered on neurons with preferred locations close to the actual position of the rat, although the profile is even broader than the activity profile in EC. This is the consequence of plasticity of the perforant path in the learning phase, which establishes an association between EC cells and CA3 neurons with similar preferred locations and head directions. The PP projection also reduces directionality substantially, so that inputs to CA3 already depend less on the preferred head direction of the cell than neuronal activities in EC. The shape of the final activity profile across place cells is, however, essentially determined by the CA3 internal dynamics, resulting in a spatial activity profile which is much more sharply peaked than the feedforward inputs. Further, the final activities of the cells are essentially independent of their preferred head direction. The resulting model place fields, showing the activity of a single unit when the model rat is at different locations and head direction in the environment, possess many of the characteristics of real place cell firing patterns recorded in open environments. As in the example in Figure 4a, they are unimodal, approximately Gaussian with circular symmetry, and essentially non-directional.

Figure 3 also reveals how non-directional place fields can result from a directional input representation in EC. The two parts of the figure compare the activities of EC neurons, the PP inputs to CA3 place cells, and the final activities of place cells as the model rat faces in two opposite directions at the same location. Due to the properties of the PP projection discussed above, place cells receive relatively similar inputs in the two cases. More importantly, however, this leads to the emergence of the same non-directional attractor in CA3, making the place fields independent of head direction.

In agreement with a recent modelling study by Brunel and Trullier (1998), we found that the ability of the recurrent network to suppress the directionality of the inputs depends critically on the set of locations and head directions experienced by the rat during learning. Place cells become direction independent only in situations in which the animal is exposed to a wide range of directions at a particular location. On the other hand, when the behavioral task or the environment itself constrains the set of directions experienced at a given location, as in a radial maze or when the rat is required to follow a specific route in an open field, place cells retain their intrinsic directionality. Even in these cases, the width of directional tuning can, however, be modified by the recurrent network. These results are in good agreement with experimental findings (Markus et al., 1995). The dependence of directionality on movement patterns is illustrated in Figure 4b, which shows the place field of the same model CA3 cell that appears in Figure 4a, for a rat which has performed a different behavioral task in the same environment. In this task, which can be thought of as a simplified version of the directed search task described by Markus et al. (1995), the rat is required to run back and forth between the two shorter walls of the environment to obtain reward. We model this by assuming during exploration that the rat is now exposed only to the two directions parallel to the long walls instead of all directions at each location. Everything else in the simulations is left the same. This change affects the correlations between place cells in the learning phase, resulting in altered weight structure, which, in turn, changes the attractors. In agreement with experimental data, the new attractors do not eliminate the directionality of the inputs to the place cells.

Very different environments

Experiments in which the firing rate maps of place cells are recorded in multiple environments which are similar to a controlled degree can provide valuable information about how input representations depend on details of the environment, how they are transformed into the place

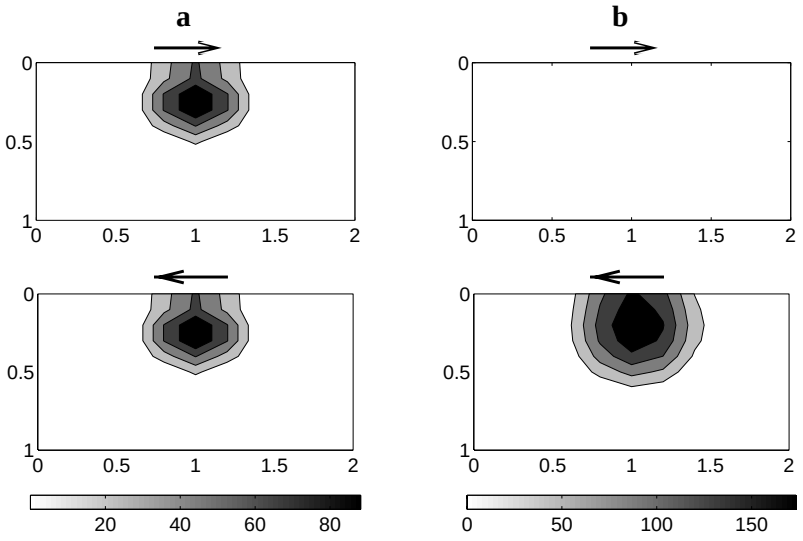

Figure 4: The task dependence of directionality. The place field of a CA3 cell which prefers the left direction, when the rat faces in the direction indicated by the arrows; **a**: in a model rat which explored the environment randomly during the learning phase; **b**: in a model animal which always ran in one of the directions parallel to the long walls of the box during learning. The top plot is empty in **b** because the cell does not fire at all in that direction in this case. Note that the plots in this figure are different in nature from the somewhat similar-looking ones in Figure 3. Here the activity of a given cell is plotted as a function of the actual position (and heading) of the animal, whereas the contour plots in Figure 3 display the quantities described there for a collection of cells with different preferred locations and headings, at a fixed location (and heading) in the environment.

cell representation, and also about possible interference between representations of different environments realized by the same network of place cells. The general pattern of results is that radically different environments give rise to very different, and apparently unrelated place cell representations (O'Keefe and Conway, 1978; Muller and Kubie, 1987; Bostock et al., 1991). On the other hand, when a previously familiar environment is subjected to subtle alterations, the place cell representation often stays basically the same (O'Keefe and Conway, 1978; Bostock et al., 1991), or changes according to the transformation of the environment (Muller and Kubie, 1987; O'Keefe and Burgess, 1996).

In order to test our model in the first type of situation, we added another model environment to the one described in the previous section, and tested whether these two environments can be learned and recalled simultaneously without interference. The two environments are very different in terms of visual appearance; the new environment has a circular shape, and is assumed to carry visual features that are dissimilar to the ones in the rectangular box. Therefore, we assume that the spatial characteristics of both EC neuronal activities and mossy fiber inputs to CA3 as well as their relations are completely independent in the two environments; i.e., for instance, knowing the relative locations of maximum activity for two EC neurons in one environment carries no information about the relation of their preferred locations in the other environment. However, as a worst case scenario, we use exactly the same neuronal populations to represent the two environments; if these populations are distinct to any extent, this can only improve the separability of the two environments.

Initial learning and recall in the rectangular environment are performed using the procedures described in the previous section. Then the weights are modified by running a learning phase in the circular environment, and spatial firing distributions during recall are determined in both environments to assess interference caused by exposure to the other environment.

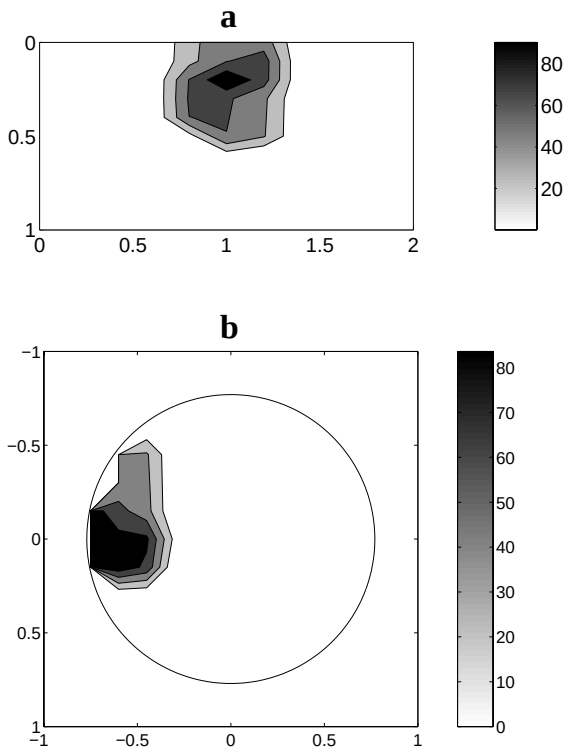

Figure 5: Very different environments. This figure shows the place fields of the same place cell in (**a**) a rectangular and (**b**) a circular apparatus which have very different sensory features, after the rat has become familiar with both environments. There is no obvious relation between place fields of the same cell in the two environments. The effect of encoding a second environment on the place cell representation in the first environment can be assessed by comparing part **a** of this figure with the top plot of Figure 4a, which shows the same place field before exploring the circular environment.

Figure 5 shows the firing rate distributions of a model CA3 cell after learning in both environments. In general, there is no systematic relation between the location of place fields in the two environments, which indicates that several different sets of attractors can be stored and recalled independently in the model. The degree of interference between the representations of different environments can be assessed by comparing place fields in one environment before and after experience in the other environment. Comparing Figures 4a and 5a reveals no significant difference between the firing rate distributions in the rectangular box of the same place cell before and after training in the circular environment. In particular, the location of maximal firing, the size, shape, and directionality (not shown) of the place field are all virtually unchanged. The only noticeable difference is that the place fields appear less regular and slightly noisier after exposure to another environment. In fact, learning to represent a new, "orthogonal" environment can be thought of as introducing noise into both the feedforward and the recurrent weights as far as the representation of the original environment is concerned. The visible decrease in the regularity of the field is, to a large extent, attributable to the complete lack of randomness in the simulation that produced Figure 4a, and would not be expected to be observed in experiments. Furthermore, since the number of neurons and connections is much larger in the real hippocampus than in the model, and not all neurons are active in any particular environment, interference between representations of different environments is likely to be less severe, and the number (and perhaps the spatial extent) of environments that can be stored is probably larger. Finally, our model would also produce orthogonal place cell representations for environments that differ only in shape (Muller and Kubie, 1987), even

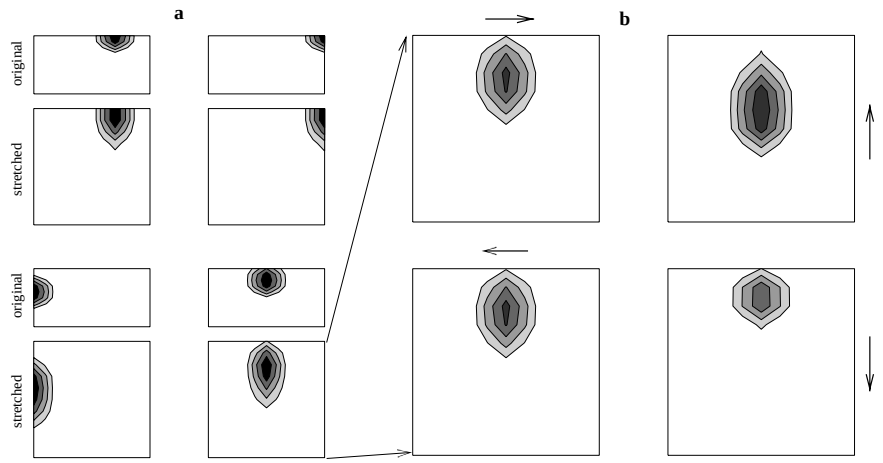

Figure 6: Place fields in transformed environments. **a**: The place fields of four selected cells in the original and the stretched environment in our simulation of the experiment by O'Keefe and Burgess (1996); the firing rates shown are averages over all head directions. **b**: Directionality of the place field shown in the bottom right corner of part a; the place field depends on the heading of the rat (indicated by the arrows). This dependence on head direction is induced by the transformation of the environment; place fields in the original environment are essentially non-directional (like the one shown in Figure 4a). Color bars were omitted from this figure to avoid clutter; firing rate ranges for all cells in both environments are similar to those shown in other figures.

from non-orthogonal input representations (Quirk et al., 1992), provided that the DG can separate the input patterns effectively, and the two environments are perceived as different so that learning is initiated in both environments.

Geometric Manipulations

We also investigated what happens to place fields in our model if the environment undergoes some simple geometric transformation. We chose to model the experiment of O'Keefe and Burgess (1996) because its relatively complex pattern of results should be useful to distinguish between different models. In this experiment a rat, which has been thoroughly familiarized with a rectangular box, is transferred into a new box that differs from the original one only in the length of one or both sides. We will concentrate on the case when the second environment is a larger square box which can be obtained by stretching the original box by a factor of two. In this case, stretching the environment had one of the following general effects (O'Keefe and Burgess, 1996): some fields remained fixed with respect to one of the walls of the apparatus; some changed their location and/or shape in correspondence with the transformation of the box; others developed a second peak in the direction of stretching. Many of the cells with two-peaked or stretched fields also developed directional dependence; i.e., the location of maximum activity depended on the heading of the rat, usually in the way that the subfield closer to a wall was more active when the rat was facing away from that wall.

We assume that learning is triggered by exposure to the novel situation of the initial, rectangular box, and that the transformed environment in this case is similar enough to the original one so that no significant learning occurs subsequently. Therefore, the attractors established in the first environment are the final states of the network dynamics in the new environment as well, and place fields are determined by the way that the inputs (as a function of location and direction) in the new environment *select* attractors established in the old environment.

Figure 6a shows the place fields of four model CA3 neurons in the rectangular box which

was used during initial learning, and in the larger square box. The place fields follow the transformation of the box; that is, their centers remain at the same relative distance from opposite walls, and their shapes become elongated along the direction of stretching. As revealed by Figure 6b, the fields consist of directional subcomponents with the observed relation between subfield position and preferred direction.

We can understand some of the characteristics of transformed place fields by analyzing our model. Attractors have a regular, compact shape if place cells are characterized by their preferred locations in the original environment; on the other hand, we have no a priori knowledge about what they look like as a function of preferred locations in the new environment. Thus, it is much easier to understand the transformations occurring in the system if we look at activities in the new environment (the square box) as a function of the neurons' preferred coordinates in the old environment (the rectangle). This is illustrated in Figure 7, which shows the activities of EC neurons, PP inputs to CA3, and final activities (after recurrent processing) in CA3 at three different locations in the square box, all as functions of preferred locations in the rectangular box. The activities of EC cells are determined by multiplying together (Gaussian-tuned) components whose activities depend on the animal's heading, and its position with respect to the walls. Since the walls have moved relative to each other, the different components lead to different estimates of position in the old coordinate system. Combining such inputs conjunctively leads to an EC activity profile which peaks somewhere between the positions indicated by individual walls. For instance, when the rat is halfway between the two walls that have been moved apart, listening to one of these walls would indicate that the animal is located at the opposite wall, and the resulting EC activity profile is centered on neurons which like the middle of the rectangular box (see the bottom left contour plot in Figure 7). Since the PP connections were established in the rectangular box, the PP input pattern to CA3 cells is centered around the same location as the EC activity pattern if both are viewed as a function of preferred coordinates in the rectangle (compare the first and the second columns of Figure 7). The recurrent connections then sharpen the activity profile considerably, but leave the location of the bump (in the old coordinate system) essentially unchanged. The final activities of CA3 cells as a function of location in the square box define the place fields in the new environment. We can see that as the rat moves around in the new environment, the activity packet also moves smoothly on the plane defined by the preferred locations of place cells in the rectangular box. This results in a smooth transformation of place fields between the two environments. In addition, the activity packet moves slower in the stretched direction in the old coordinate system than the actual speed of the rat in the new environment, or, in other words, the rat needs to travel about twice as much in the square box than in the rectangular box for the activity profile to shift by the same amount; consequently, place fields become elongated in the direction of stretching.

The emergence of directional subcomponents can be understood by looking at how the activities of EC cells and the resulting activities of CA3 neurons depend on the head direction of the rat. This is depicted in Figure 8, which shows that due to the dependence on head direction of the rat's confidence in the inputs from different walls (as described earlier), conflicting sources of information are weighed differently depending on which way the rat faces. The EC activity profile and, consequently, the CA3 activity profile, shift as the rat turns around in the square box, and the result is that a given place cell fires maximally at different locations depending on head direction.

Very similar environments

Skaggs and McNaughton (1998) conducted an experiment designed specifically to probe the relation between spatial representations in environments with a high degree of similarity. In this experiment, animals explored an apparatus which consisted of two visually identical

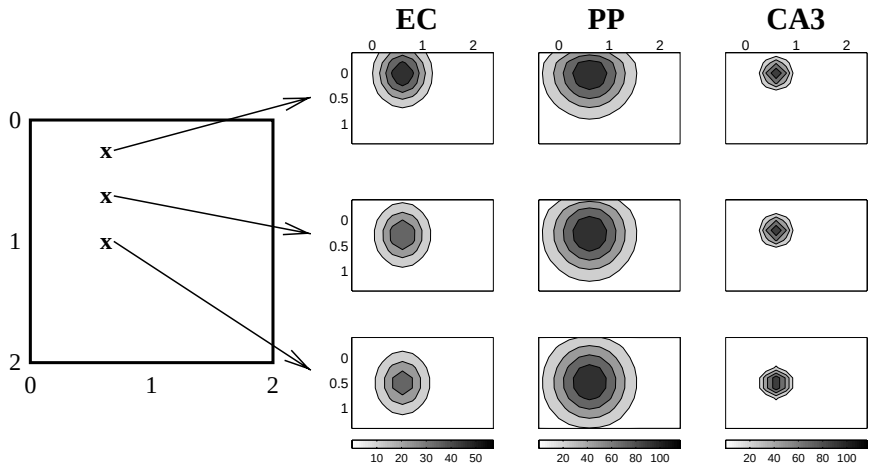

Figure 7: Place field stretching. Neuronal activities in entorhinal cortex (EC), perforant path (PP) inputs to place cells, and CA3 final activities as a function of the preferred location of the neuron in the original, rectangular box, for three different positions of the rat in the square box, indicated by the crosses in the plot on the left. The plots only show cells with preferred direction 'north', and the model rat faces 'west' in all cases.

boxes connected by a corridor. Many place cells were found to have similar place fields in the two regions, whereas others had uncorrelated place fields. This finding challenges the idea that there is a predefined set of uncorrelated attractors wired into the recurrent connections in CA3 (Samsonovich and McNaughton, 1997), because such a model would predict either identical or orthogonal firing patterns in different environments or different parts of the same environment. This particular problem may be solved by postulating a hierarchy of fixed attractors with various degrees of overlap (Samsonovich et al., 1998); however, it still remains to be explained why similar representations are selected in very similar environments. On the other hand, the attractors established in our model are input dependent, which in principle allows attractors with an arbitrary degree of similarity, and directly defines the association between attractors and environments. Therefore, we simulated the experiment by Skaggs and McNaughton (1998) in our model to study the spatial representations in very similar environments.

We still do not model the different sources of spatial information explicitly. We assume that there are some inputs (e.g., signals derived from path integration) which allow the two boxes to be distinguished, while other inputs to the system (e.g., local visual cues) are identical at corresponding locations in the two boxes. Since cells in EC are assumed to respond to different inputs to a randomly varying extent and to encode these inputs conjunctively, we applied the following scheme to determine activities in EC at locations inside the two boxes. EC cells are now characterized by a preferred location (and also a preferred head direction) based on visual inputs (this is now actually a set of two locations, one in each box), as well as a polarization index (P), which is defined as the maximum firing rate for the cell in the north box minus the maximum firing rate in the south box, divided by the maximum rate in any of the boxes. P is always between -1 and 1, its magnitude indicates how much that particular cell is influenced by cues that distinguish the two boxes, and its sign shows which box the neuron prefers. We assign P values to EC cells randomly from a uniform distribution. The firing rate of an EC neuron is then given by $z_k = (1+P_k)z_k{}'$ in the north box and $z_k = (1-P_k)z_k{}'$ in the south box, where $z_k{}'$ is a function of coordinates within the current box, and it depends on spatial position and head direction the same way as z_k in Equation 3. We assume that the MF inputs to CA3 can be characterized similarly; however, due to the orthogonalizing properties of the dentate gyrus, P values do not vary continuously, but only take the values -1, 0, and 1, each with

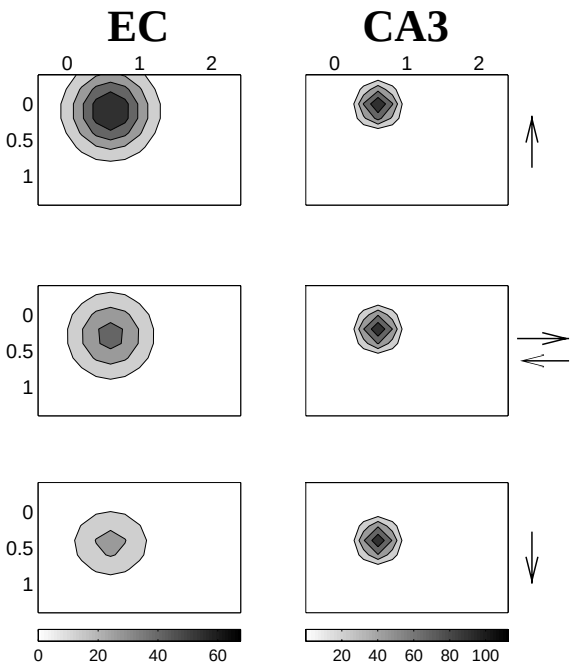

Figure 8: Directionality of stretched place fields. EC neuronal activities, and CA3 final activities (of the same cells as in Figure 7), as a function of the neurons' preferred locations in the rectangular box, for different headings of the rat (indicated by the arrows) at a single location in the square box (marked by the middle cross in Figure 7). The position of the input peak changes as the rat faces in different directions (due to the dependence on head direction of the rat's confidence in different cues), and the position of the final activity profile in CA3 changes accordingly.

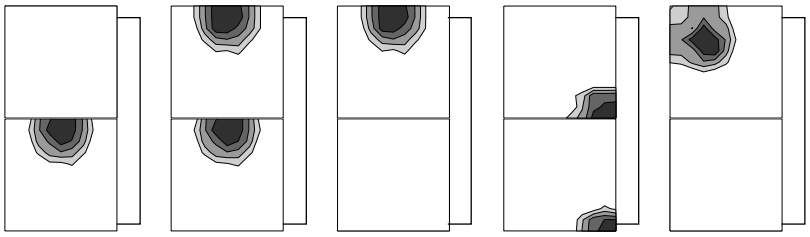

Figure 9: Place fields in our simulation of Skaggs and McNaughton (1998). The figure shows the place fields of five CA3 place cells in the two identical boxes; activity in the corridor connecting the boxes was not simulated.

probability 1/3. This means that there is a population of cells in CA3 which receives the same input at corresponding locations in the two boxes during learning, while another population receives different inputs. Since the first time the rat is introduced into the apparatus it is allowed to explore it entirely, we do not treat the two halves of the environment differently during the learning phase.

Some examples of the place fields that develop in this model are shown in Figure 9. There are cells which have similar firing rate patterns in the two boxes, while others are active in only one of the boxes, in accordance with experimental observations. In other words, our model has no difficulty storing and recalling partially overlapping spatial representations. In the model, the degree of overlap is determined by the extent of orthogonalization occurring in DG, i.e., what proportion of granule cells distinguishes between the two boxes – CA3 cells simply inherit the selectivity of their MF inputs as attractors are established during the learning phase.

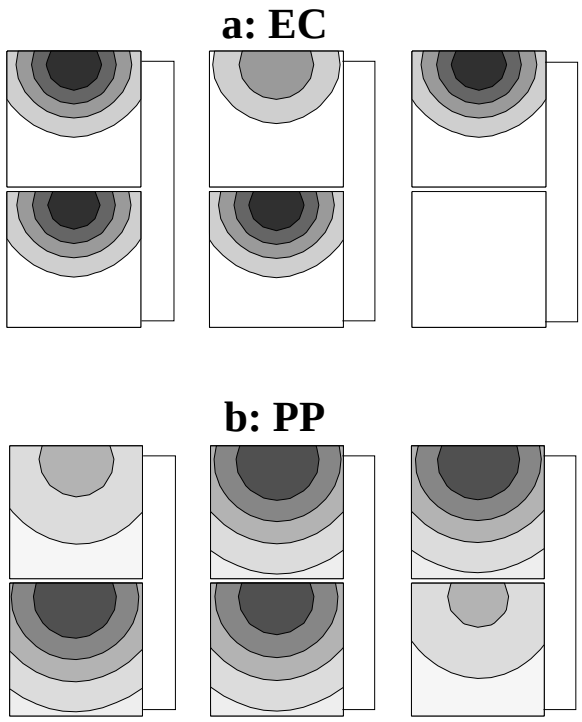

Figure 10: Input representation and inputs to CA3 in our simulation of Skaggs and McNaughton (1998). **a**: The activities, as a function of location in the apparatus, of 3 entorhinal neurons which have the same preferred (visual) location within the boxes, but different degrees of polarization (as defined in the text; the polarization indices are -0.01, 0.25, and -0.70, respectively). **b**: This part of the figure, which displays the perforant path inputs to the first three CA3 place cells of Figure 9, shows that, as a result of learning in the perforant pathway, some place cells receive similar inputs at corresponding locations in the two boxes, while others receive inputs of different magnitudes, setting the stage for the CA3 recurrent network which makes these differences much more pronounced (as seen in Figure 9).

Most EC neurons are active in both boxes, although to a different extent (see Figure 10a). Consequently, all CA3 cells that are active in this environment get a substantial PP input in both boxes (Figure 10b); however, the activity patterns encoded during learning are restored by the recurrent connections and feedback inhibition, and the PP input only determines which of these patterns emerges. The figures also show that although EC neurons have relatively broad tuning curves, and this results in CA3 cells receiving feedforward input that is even more broadly tuned, the final tuning of CA3 neurons is considerably sharper due to recurrent activity. The attractor network also renders place cells directionally nonselective, just as in the rectangular environment considered before.

Discussion

Principal findings

We have presented a plastic attractor network model of CA3 place cells which describes how a conjunctive representation of location- and direction-specific sensory information in entorhinal cortex can be transformed by feedforward pathways and recurrent processing in the hippocampus, into a place cell representation whose properties match a large number of experimental observations. In particular, our model (1) accounts for the head direction independence of place cells in open environments as well as their directionality in linearly restricted envi-

ronments, (2) demonstrates how several different environments can be stored and recalled independently by the CA3 recurrent network, (3) produces place cell activity patterns with an appropriate degree of overlap in visually similar environments, and (4) correctly captures the transformations of place fields after simple geometric manipulations of the environment.

Although the representations formed may be useful for spatial tasks such as navigation (Burgess et al., 1997; Foster et al., 1998, 1999), the major goal for our model was to show how ideas about how non-spatial information is processed by the hippocampus are in accordance with data on place fields.

Components of the model

The idea of using attractor networks for computations has been applied in various settings (Somers et al., 1995; Zhang, 1996; Pouget et al., 1998); such networks have been shown to be capable of amplifying certain facets of their inputs (Ben-Yishai et al, 1995) as well as creating invariance (Chance et al., 1999). Our model (and Brunel and Trullier's, 1998) displays both behaviors simultaneously; the recurrent network enhances the spatial tuning of place cells while suppressing their directional tuning in open field environments. Under attractor dynamics, it is unwise to invent rules describing how individual place cells respond in various situations; rather, the system is better described collectively, by identifying the attractors and specifying which attractor gets selected for any particular input. The attractor concept also helps explain the persistence of spatial firing patterns in the face of environmental manipulations such as cue removal or cue rotation (O'Keefe and Conway, 1978; Muller and Kubie, 1987) as well as the abrupt changes that ensue for changes of other kinds (e.g., changing the shape of the environment from circular to square; Muller and Kubie, 1987) or of a larger magnitude. However, feedforward models (*eg* Sharp, 1991; Burgess et al., 1996), albeit ignoring the recurrent connections, can also be made to exhibit many of the properties we have demonstrated. We have not yet modeled the pathway from CA3 to CA1, assuming that the spatial properties of the latter faithfully reflect those of the former, assuming normal plasticity. Of course, CA1 is the source of the bulk of the experimental data on place fields.

The hippocampus exhibits a wide array of dynamical behaviors, including a characteristic set of oscillations in all fields, rendering moot our description of CA3 as an attractor network. Nevertheless, the average activities of cells in oscillatory networks can exhibit attractor-like dynamics (Li and Dayan, 1999), so a correct description of place fields (which reflect average firing rates) may nevertheless be achieved.

The learning rule was chosen as a crude model of experimental long-term synaptic plasticity (LTP), and we have ignored most empirical complexities. We have not taken into account the fact that the sign and magnitude of long-term synaptic modification depends on the relative timing of pre- and postsynaptic activity (Levy and Steward, 1983; Markram et al., 1997), which has been suggested as a mechanism underlying a navigational role of place cells (Blum and Abbott, 1996). Indeed, the recurrent weights in our model ultimately learn a weight structure similar to the "cognitive graph" described by Muller et al. (1991, 1996).

Similar proposals to ours have been put forward in associative memory models of the hippocampus (Treves and Rolls, 1992) as to the separate roles for the indirect pathway to CA3 via the dentate gyrus (which defines attractors during the learning mode) and the direct perforant path (which selects attractors during recall mode). However, the activity patterns representing location and direction are intrinsically continuous, and thus strongly overlapping, so the patterns that are retrieved can differ in systematic ways from all the patterns encountered during learning (*eg* being insensitive to head direction in open field environments). Switching between the two modes need not be complete; moderate changes in relative synaptic efficacies are sufficient to boost performance in a model of episodic memory by Hasselmo et al. (1996). Incomplete switching might underlie the slow changes in the place cell representation

observed in some conditions (Bostock et al, 1991).

Exactly how entorhinal and dentate neurons encode features in the environment, and how they respond to manipulations of the environment, is not experimentally clear. Our choice was necessarily somewhat arbitrary – the aim has been to show that there exists at least one reasonable choice that results in place fields consistent with experimental data in a wide range of experimental situations. Our entorhinal representation is similar to that of Burgess et al. (1997), except that their units are directionally non-selective, and each is tied to exactly two orthogonal walls of the environment. Place cell firing patterns are then determined through the feedforward weights connecting EC to CA3; these weights are set up using a competitive learning scheme similar to the one used by Sharp (1991) to model the formation of place fields. Competitive learning supports the separation of different input patterns in these models; in our model, the same task is thought to be accomplished through processing by the dentate gyrus. It is possible, of course, that both of these processes contribute to pattern separation in the hippocampus.

Comparison with other models

The models of Samsonovich and McNaughton (1997) and Burgess et al. (1997) are closest to ours (and, at least for the former, Brunel and Trullier's 1998). The most important distinction from Samsonovich and McNaughton (1997) is that it relates the position of the activity profile (or 'packet') in CA3 (as an attractor network) to external coordinates in a different way, assuming a hard-wired system which is capable of updating the position of the CA3 activity packet based on self-motion information, and a learned association with sensory representations which can be used to correct for accumulated errors in path integration. Learning works differently in our model, and the metric of the place cell representation reflects the way in which the EC representation depends on external coordinates, including sensory features of the environment and, to account for the formation and maintenance of place fields in darkness, self-motion information. It is unlikely that the hippocampus itself is responsible for path integration (Alyan et al., 1997).

Burgess et al.'s (1997) model also accounts for some of O'Keefe and Burgess's (1996) data. Their results are complementary to the ones we presented here, in that their model captures the behavior of those place cells which remain fixed with respect to one wall or develop a second place field after stretching the environment, while our model correctly describes those place fields that follow the transformation of the environment, and also explains the acquired directionality of stretched place fields in the transformed environment. A modified version of our model, which incorporates random variations in the extent to which input cells respond to different spatial cues, reproduces all the observed classes of place field transformation. Due to its randomness, it offers less insight into the underlying mechanisms than the model described here. Our model also accounts for other properties of place cells, such as directionality and non-directionality.

Critical experiments

Various experiments could, in principle, test the key assumptions and predictions of our model. First, pharmacological or molecular biological blockade of plasticity in the CA3 recurrent connections should prevent the formation of a new representation in a novel situation. According to our model, the system would either remain trapped in learning mode, which would be indicated by, among other things, retained directionality of place fields in an open field, or recall attractors from one or more environments explored before the blockade, resulting in irregular or fragmented place fields. Direct manipulations of the neuromodulatory control mechanisms governing the choice of learning versus recall mode should have a similar

effect. Unfortunately, there exist many different forms of experimental plasticity, and it is not clear which in particular are most relevant for learning in vivo.

Our model predicts that the CA3 place cell representation should be different during the first few minutes of exploration in a new environment from the time after the animal has become familiar with its surroundings. In particular, place cells are expected to be directional in any novel environment immediately after entry, and become non-directional later in open environments.

Analysis of our model also indicates that the amount of training in a given environment might have a significant effect on the place cell representation in a similar environment encountered subsequently, since only well-established attractors are assumed to be capable of being recalled. For instance, we would expect to see a less obvious relation between place fields in different environments in the experiment of O'Keefe and Burgess (1996) if, instead of training the rat in one size of box before allowing it to explore the others, they had made it explore all four environments in quick succession, especially if the rat is prevented from using extramaze cues.

References

[1] Alyan, S.H.,Paul, B.M.,Ellsworth, E.,White, R.D., and McNaughton, B. L. (1997). Is the hippocampus required for path integration? Soc. Neurosci. Abstr., 23, 504.

[2] Amaral, D.G. and Witter, M.P. (1989). The three-dimensional organization of the hippocampal formation: a review of anatomical data. Neuroscience, 31, 571-591.

[3] Barnes, C.A., McNaughton, B.L., Mizumori S.J.Y., 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.

[4] Ben-Yishai, R., Bar-Or, R.L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. Proc. Natl. Acad. Sci. U.S.A., 92, 3844-3848.

[5] Blum, K.I. and Abbott, L.F. (1996). A model of spatial map formation in the hippocampus of the rat. Neural Comput., 8, 85-93.

[6] Bostock E., Muller, R.U., and Kubie, J.L. (1991). Experience-dependent modifications of hippocampal place cell firing. Hippocampus, 1, 193-205.

[7] Brunel, N. and Trullier, O. (1998). Plasticity of directional place fields in a model of rodent CA3. Hippocampus, 8, 651-65.

[8] Burgess, N., Donnett, J.G., and O'Keefe, J. (1997). Robotic and neuronal simulation of hippocampal navigation. Phil. Trans. Roy. Soc. B, 352, 1535-43.

[9] Burwell, R.D. and Amaral, D.G. (1998). Cortical afferents of the perirhinal, postrhinal, and entorhinal cortices of the rat. J. Comp. Neurol., 398, 179-205.

[10] Chance, F.S., Nelson, S.B., and Abbott, L.F. (1999). Complex cells as cortically amplified simple cells. Nat. Neurosci., 2, 277-82.

[11] Foster, D.J., Morris, R.G.M., and Dayan, P. (1998). Hippocampal model of rat spatial abilities using temporal difference learning. In Jordan, M.I., Kearns, M.J., Solla, S.A., eds, Advances in Neural Information Processing Systems, 10, 145-151.

[12] Foster, D.J., Morris, R.G.M., and Dayan, P. (1999). A model of hippocampally dependent navigation, using the temporal difference learning rule. Hippocampus, in press.

[13] Hasselmo, M.E. (1999). Neuromodulation: acetylcholine and memory consolidation. Trends Cognit. Sci., 3, 351-359.

[14] Hasselmo, M.E., Schnell, E., and Barkai, E. (1995). Learning and recall at excitatory recurrent synapses and cholinergic modulation in hippocampal region CA3. J. Neurosci., 15, 5249-62.

[15] Hasselmo, M.E., Wyble, B.P., and Wallenstein, G.V. (1996). Encoding and retrieval of episodic memories: Role of cholinergic and GABAergic modulation in the hippocampus. Hippocampus, 6, 693-708.

[16] 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.

[17] Jung, M.W. and McNaughton, B.L. (1993). Spatial selectivity of unit activity in the hippocampal granular layer. Hippocampus, 3, 165-182.

[18] Káli, S. and Dayan, P. (1998). The formation of direction independent place fields in area CA3 of the rodent hippocampus using Hebbian plasticity in a recurrent network. Soc. Neurosci. Abstr., 24, 931.

[19] Levy, W.B. (1996). A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. Hippocampus, 6, 579-590.

[20] Levy, W.B. and Steward O. (1983). Temporal contiguity requirements for long-term associative potentiation / depression in the hippocampus. Neuroscience, 8, 791-7.

[21] Li, Z. and Dayan, P. (1999). Computational differences between asymmetrical and symmetrical networks. Network, 10, 59-77.

[22] Li, X.-G., Somogyi, P., Ylinen, A., and Buzsaki, G. (1994). The hippocampal CA3 network: An in vivo intracellular labeling study. J. Comp. Neurol., 339, 181-208.

[23] Markram, H., Lubke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science, 275, 213-5.

[24] Markus, E.J., Qin, Y.-L., Leonard, B., Skaggs, W.E, McNaughton, B.L., and Barnes, C.A. (1995). Interactions between location and task affect the spatial and directional firing of hippocampal neurons. J. Neurosci., 15, 7079-7094.

[25] Marr, D. (1971). Simple memory: a theory for archicortex. Philos. Trans. R. Soc. Lond. B, 262, 24-81.

[26] 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.

[27] Muller, R.U. and Kubie, J.L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. J. Neurosci., 77, 1951-1968.

[28] Muller, R.U., Kubie, J.L., and Ranck, J.B. Jr. (1987). Spatial firing patterns of hippocampal complex-spike cells in fixed environment. J. Neurosci., 7, 1935-1950.

[29] Muller, R.U., Kubie, J.L., and Saypoff, R. (1991). The hippocampus as a cognitive graph. Hippocampus, 1, 243-246.

[30] Muller, R.U., Stead, M., and Pach, J. (1996). The hippocampus as a cognitive graph. J. Gen. Physiol., 107, 663-694.

[31] Myers, C.E., Gluck, M.A., and Granger, R. (1995). Dissociation of hippocampal and entorhinal function in associative learning: A computational approach. Psychobiology, 23, 116-138.

[32] O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. Exp. Neurol., 51, 78-109.

[33] O'Keefe, J. and Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. Nature, 381, 425-428.

[34] O'Keefe, J. and Conway, D.H. (1978). Hippocampal place units in the freely moving rat: why they fire where they fire. Exp. Brain Res., 31, 573-90.

[35] O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res., 34, 171-175.

[36] O'Reilly, R.C. and McClelland, J.L. (1994). Hippocampal conjunctive encoding, storage, and recall: avoiding a tradeoff. Hippocampus, 4, 661-682.

[37] Pouget, A., Zhang, K., Deneve, S., and Latham P.E. (1998). Statistically efficient estimation using population coding. Neural comput., 10, 373-401.

[38] Quirk, G.J., Muller, R.U., Kubie, J.L., and Ranck J.B. Jr. (1992). The positional firing properties of medial entorhinal neurons: description and comparison with hippocampal place cells. J. Neurosci., 12, 1945-1963.

[39] Rolls, E.T. (1996). A Theory of Hippocampal Function in Memory. Hippocampus, 6, 601-620.

[40] Samsonovich, A. and McNaughton, B.L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. J. Neurosci., 17, 5900-5920.

[41] Samsonovich, A., McNaughton, B.L., and Nadel, L. (1998). Hierarchical multichart model of the hippocampal cognitive map. Soc. Neurosci. Abstr., 24, 931.

[42] Sharp, E.P. (1991). Computer simulation of hippocampal place cells. Psychobiology, 19, 103-115.

[43] Skaggs, W.E. and McNaughton, B.L. (1998). Spatial firing properties of hippocampal CA1 populations in an environment containing two visually identical regions. J. Neurosci., 18, 8455-8466.

[44] Somers, D.C., Nelson, S.B., and Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. J. Neurosci., 15, 5448-65.

[45] Touretzky, D.S. and Redish, A.D. (1996). A theory of rodent navigation based on interacting representations of space. Hippocampus, 6, 247-270.

[46] Treves, A. and Rolls, E.T. (1992). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. Hippocampus, 2, 189-199.

[47] Treves, A. and Rolls, E.T. (1994). A computational analysis of the role of the hippocampus in memory. Hippocampus, 4, 374-391.

[48] Wilson, H.R. and Cowan, J.D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J., 12, 1-24.

[49] Wilson, M.A. and McNaughton, B.L. (1993). Dynamics of the hippocampal ensemble code for space. Science, 261, 1055-1058.

[50] Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. J. Neurosci., 16, 2112-2126.

[51] Zipser, D. (1985). A computational model of hippocampal place fields. Behav. Neurosci., 99, 1006-1018.