In *ICANN 1999.*

Spatial Representations in Related Environments in a Recurrent Model of Area CA3 of the Rat

Szabolcs Káli[1,2] **and Peter Dayan**[1]
[1]Gatsby Computational Neuroscience Unit, UCL,
17 Queen Square, London WC1N 3AR, U.K.
[2]Department of Brain and Cognitive Sciences, MIT, Cambridge, U.S.A.
szabolcs@gatsby.ucl.ac.uk

Abstract

Recurrent network models of area CA3 in the hippocampus capture faithfully many of the properties of place cells. However, they seem ill suited to explaining the substantial experimental data on place cells in environments with particular visual or geometrical similarities. We show that a model in which the activities of CA3 place cells are determined mainly by modifiable recurrent connections (together with global inhibitory feedback) is capable of reproducing the major classes of behavior that are observed. In visually similar environments, the patterns of place cell activities have the appropriate degree of similarity; after geometric transformations to the environment, the model place fields undergo geometric transformations, and also remapping, induced (or uncovered) directionality and disappearance.

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, and of many other spatially selective cells in neighboring areas.[5,10] Principal neurons in hippocampal areas CA3 and CA1 are active only when the animal is located in a well-defined local region of the environment (a place field),[9] which collectively provide a population code for spatial position. The spatial relationship between place fields is typically unpredictable in grossly different environments.

To help understand information processing in the hippocampal system, we have investigated how the response characteristics of place cells may be established, and in particular how the animal's sensory experience determines the the various properties of place fields including location, shape, and directionality. For instance, place cells have an approximately circularly symmetric structure in an open environment, and a strong dependence on head direction in a linear environment.

A host of models has been suggested to account for these data, including one in a previous study of ours,[6] in which a central role is played by recurrent connections in CA3, acting as an plastic *attractor network*, and various models of other groups,[2,11,12,17] in which attractor networks of other forms are important. Experimental data from more sophisticated paradigms, such as those in which environments change, is therefore necessary to distinguish between the models.

In the next section, we summarize the experimental observations on which the model is tested. We then describe the model in some detail, and present the results obtained from simulations. Finally, we discuss some more general implications of our results, analyze some of the limitations of the current model, and compare our approach to other related work.

[0]The authors would like to thank Neil Burgess and Kathryn J. Jeffery for useful discussions.
Supported by the Gatsby Charitable Foundation

Background

Place fields are formed relatively quickly (on the order of ten minutes) upon entry into a new environment.[16] The process seems to require active exploration, and probably depends on long-term potentiation (LTP) within the hippocampus. Once they are fully developed, place fields remain stable for several days, even if the animal is absent from that particular environment for most of the intervening time.[14] Place cells have at most one place field in regular environments. The bulk of the place cell data comes from area CA1 rather than CA3 – we assume that the results will be true for CA3 too.

We model two classes of experiments that hint at some of the complexities of place cells. One studies the consequences of creating an environment in which two different regions look very similar;[13] and the other studies the effects on pre-established place fields of various environmental manipulations.

One experiment from the first class used two separate regions that were visually almost identical. In this case, many place cells turn out to have similar place fields; whereas others have uncorrelated place fields in the two regions. This finding challenges the idea that there is a predefined set of uncorrelated attractors wired into the recurrent connections in CA3, because such a model would predict either identical or orthogonal firing patterns in different environments or different parts of the same environment.

The general pattern of results from experiments designed to test how manipulating the environment affects place cells is that radical environmental change causes a completely new, and apparently unrelated, spatial representation to develop. However, subtler manipulations result in a substantial proportion of place fields undergoing simple geometric transformations which reflect the transformation of the environment. For instance, rotation of all distal cues with respect to a circular apparatus results in a corresponding rotation of the place fields.[7] Scaling the experimental apparatus leads to similar scaling of the location and size of the place fields.

In an experiment of O'Keefe and Burgess,[8] a rat, which has initially experienced a rectangular box, is transferred into a new box that differs from the original one only in the length of one side. In this case, stretching the environment had one of the following general effects: 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, sometimes in conjunction with a novel, induced, dependence on head direction in the new environment.

The Model

Our model started from the observation that most of the natural inputs that might control place cells would inevitably depend on head direction, even in open field environments. We showed that the recurrent connections in area CA3 could eliminate this directionality in open field environments, provided that they could undergo associative Hebbian plasticity.[1,6] However, in our model, CA3 forms an attractor network, a structure that had been believed to have difficulty accounting for experimental observations on the behavior of place cells in multiple, closely related, environments. In this paper we show how our model captures these results too.

Figure 1a shows the model. The main part of the hippocampal circuitry we actually implement is the CA3 recurrent network – the properties of other areas are taken into account through the way they determine the input to CA3. We model CA3 as a collection of 1200 pyramidal cells, each connected to all the others through modifiable weights. We also include a global inhibitory neuron in the model, which provides feedback inhibition for the pyramidal neurons, and keeps global activity levels approximately constant.

Based on experimental data on the effects of septal (cholinergic and GABAergic) modulation in the hippocampus and theoretical results for autoassociative memories,[15] we employ the suggestion[3] that the hippocampal network has two modes of

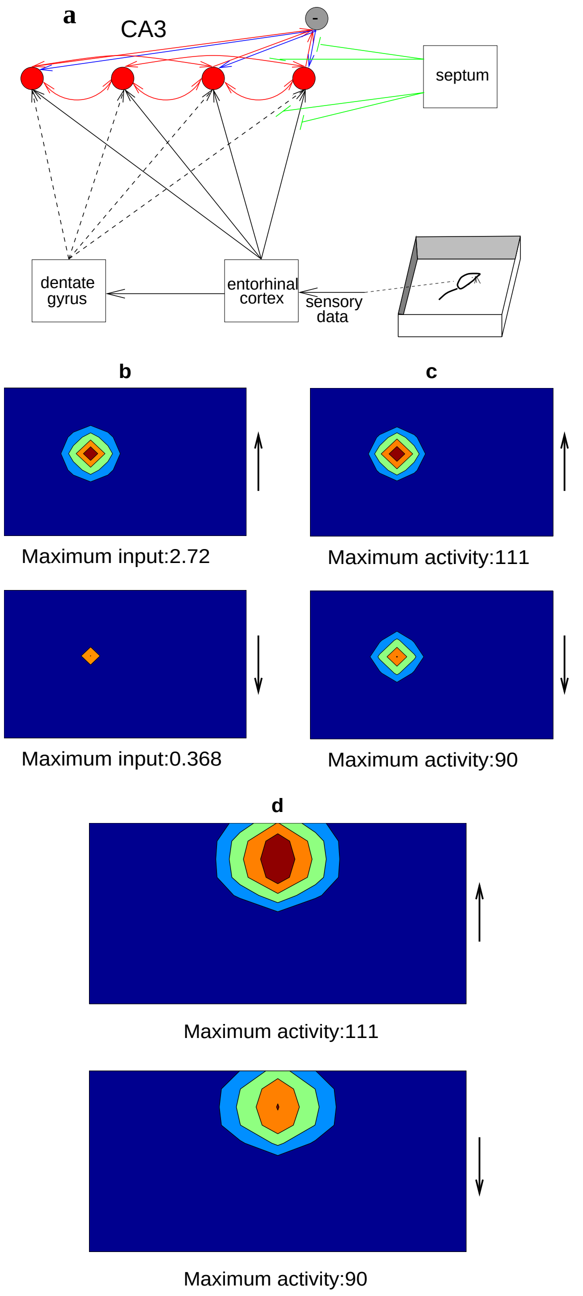

Figure 1: **a**: Model architecture; **b**: Inputs to place cells and **c**: their final activities in a rectangular box, as a function of their preferred locations, for two populations of CA3 neurons with preferred head directions indicated by the arrows; **d**: The place field of a selected cell, with the heading of the rat indicated by the arrows. Note that the directionality in the input gets suppressed by the attractor network.

operation. When the rat is in a familiar environment, no learning takes place in any of the connections, the relative efficacy of the mossy fiber inputs with respect to the perforant path connections and CA3 recurrent synapses decreases, and the intrinsic dynamics of the recurrent network dominates activity in CA3. This is called 'recall mode'. On the other hand, when the rat first encounters a new environment, learning in both the perforant path inputs to CA3 and the recurrent connections is enabled, the recurrent connections are suppressed, inhibition in CA3 is reduced, and inputs through the mossy fiber connections dominate. This state of the network is called 'learning mode'. We show how the pattern of weights set up during learning can produce the patterns of activity of place cells seen subsequently.

The dynamics of unit activities is described by the following equations:

$$\begin{aligned} \dot{u_i} &= -\alpha u_i + \sum_j J_{ij} g_u(u_j) - h g_v(v) u_i + I_i \\ \dot{v} &= -\alpha v + w \sum_j g_u(u_j) \end{aligned} \tag{1}$$

where u_i is the membrane potential of the ith pyramidal cell, v is the membrane potential of the global inhibitory cell, $1/\alpha$ is the membrane time constant, 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, I_i represents all external inputs to this cell from outside CA3, and $g_u(u) = \beta(u-\mu)\Theta(u-\mu)$ is the activation function for the pyramidal cells, where Θ is the unit step function (zero for negative arguments and one for positive ones). μ stands for the threshold and β is the slope of the activation function above the threshold. Similarly, $g_v(v) = \gamma(v-\nu)\Theta(v-\nu)$. The recurrent weights change in learning mode according to:

$$\dot{J}_{ij} = -\kappa J_{ij} + u_i g_u(u_j) \tag{2}$$

Each CA3 cell receives a large number of perforant path inputs, which initially have small random weights, and these connections undergo Hebbian synaptic modification similar to that described for the recurrent weights. We assume that there is at most one active mossy fiber input to a CA3 neuron at any given time, and we characterize CA3 pyramidal cells by the preferred location of its most active input in the environment where the rat is trained initially.

We model the two sets of inputs to CA3 as follows. In common with many models,[15] we assume that neurons in entorhinal cortex respond in a conjunctive manner to available sources of spatial information. In general, these may include local as well as distant sensory cues of all modalities, and also ideothetic information (from path integration). We do not model any of the processes by which this information becomes available; instead, we treat them equally as constraints on the set of locations where a given cell fires. Each cell responds to a subset of the available cues. It fires maximally when all the cues it is sensitive to are in the position corresponding to the cell's preferred location, and activity diminishes as some or all of the sources of information signal a different location. 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. These variances 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. Dentate granule cells are thought to have similar response properties, though their tuning is assumed to be sharper.

Perforant path and recurrent weights are acquired during the learning phase. We model exploration during learning by imagining that the rat receives even exposure across the whole apparatus. During the recall phase, the final activities of the cells are calculated as a function of location, by integrating equations 1 for a fixed time using Euler's method. The network always settles into a stable state by the end of the iterations.

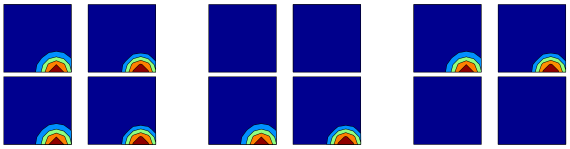

Figure 2: The place fields of three selected cells and the feedforward input they receive in the simulation of the experiment by Skaggs and McNaughton; in each pair of figures, the left one shows the input and the right one the place field.

The attractors that are learned in this model in an open environment such as a rectangular box resemble thresholded two dimensional Gaussian bumps of activity in the space where neurons are arranged on an imaginary plane according to their preferred locations. During recall, the location of the bump is determined by the feedforward input so that the activity profile provides the best possible fit to the input. This is demonstrated for the simple case of a single rectangular box by Figure 1b-d, which also shows that although inputs to CA3 strongly depend on head direction, the attractor dynamics can suppress this dependence so that the final activities vary only slightly with preferred head direction. In turn, this results in place fields with little directional sensitivity. Recall is dominated by the recurrent connections (on average, only about 5 % of the input to each cell comes from feedforward activation). In any environment that causes this particular set of attractors to appear (*ie* all similar environments), the shape of the place fields is determined by the way that the position of the bump of activity in its own coordinate frame changes as the animal moves around in the environment.

Two Similar Environments

In the experiment by Skaggs and McNaughton,[13] there are two visually identical boxes, which are likely distinguished by idiothetic information. They give rise to place cells, some of which have place fields in the same location in both environments. This is hard to account for using attractor networks with pre-wired attractors,[4] which find it difficult to have different but similar firing patterns emerge in the two parts of the apparatus.

A natural model of the input to the place system is that some of the dentate granule cells are driven primarily by visual inputs (*i.e.*, they are insensitive to the signals that distinguish the two regions), while others are driven mostly by path integration information or a combination of the two. Thus the former population of cells receives the same input at corresponding locations in the two boxes, while the latter 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 2, along with the feedforward inputs they receive. Those cells that only receive visual input have similar place fields in the two boxes, and cells whose inputs are different in the two regions only have a place field in one of the boxes. In other words, our model behaves very much like a feedforward model in this respect, although, as in figure 1, it would behave unlike a feedforward model in rendering the place fields independent of head direction. Most importantly, our model does not suffer from the problems of attractor networks with prewired attractors, and accommodates comfortably the graded similarity of the firing patterns in each environment.

Geometric Manipulations

We also investigated what happens to the place fields if the environment undergoes some simple geometric transformations. We assume that learning is triggered by exposure to a novel situation, 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) in the new environment *select* attractors established in the old environment.

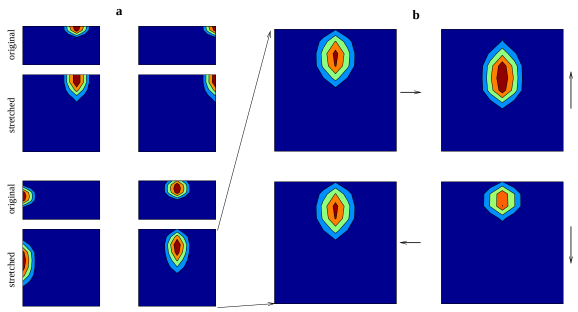

Figure 3: **a**: The place fields of four selected cells in the original and the stretched environment in our simulation of the experiment by O'Keefe and Burgess; **b**: Directionality of the place field shown in the bottom right corner of part a; the place field depends on the heading of the rat (indicated by the arrows).

Figure 3a shows the place fields of four model CA3 neurons in a rectangular box which was used during initial learning, and a larger square box which was obtained by stretching the rectangular box by a factor of two. 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 3b, and as is evident for some of the cells recorded in the experiments,[8] the fields consist of directional subcomponents. That the geometrical manipulation can induce directionality argues powerfully against models for which circular symmetry in open fields is a trivial consequence of the $\sim 300°$ field of view of rats or for which plasticity is unimportant. It arises here because the inputs have directional dependencies.

In the model just described, inputs to each cell were fully determined by its preferred location and head direction, and the actual position and head direction of the rat. As a result, all place fields behaved in essentially the same way. By contrast, when we allow random variations across cells in the degree to which they are sensitive to the locations of different walls, different cells respond differently to the transformation of the environment (Fig. 4). The figure shows examples of fields that remain at a fixed position with respect to one of the walls, or develop a second peak in the di-

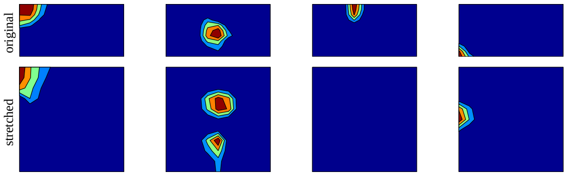

Figure 4: Four examples of place fields in the two environments in a model which allows random variations in the sensitivity of place cells to the location of different cues.

rection of stretching, or remap to a different location, or disappear completely. Therefore, the model displays the broad classes of response observed in real place cells.

Discussion

We have presented an attractor model of the recurrent network in CA3 which generates non-directionality in open field environments (and, though we did not demonstrate it here, directionality in linear track-like environments[6]), graded similarity between place fields in two visually similar environments, and appropriate place field remapping following geometric transformations.

The model avoids the pitfalls of attractor networks with fixed attractors (*eg* failing to account for the graded similarity) and, by capturing induced directionality following geometric transformations, captures experimental findings more proficiently than the (albeit more straightforward) model of O'Keefe and Burgess,[8] without any need to adjust threshold or other parameters for individual cells.

We have left several important issues unexplored. Preliminary results confirm that the model also captures the observed effects in cue rotation and removal experiments. We need to study in more detail the source of individual differences between cells with similar preferred locations. In the model, these differences come from variations in the set of features to which each cell responds, but further quantitative analysis needs to be done to answer this question. We also plan to assess the relevance of learning in the new environment, particularly in relation to possible interference. Further, our model only addresses the behavior of cells in area CA3, and we plan to model area CA1, which is a further plastic synapse away, which is actually the source of most of the experimental data.

References

[1] Brunel, N. and Trullier, O. (1998). Hippocampus, 8, 651-65.

[2] Fuhs, M.C., Goodridge, J.P. and Touretzky, D.S. (1998) Soc. Neurosci. Abstr., 24, 931.

[3] Hasselmo, M.E., Wyble, B.P., and Wallenstein, G.V. (1996). Hippocampus, 6, 693-708.

[4] McNaughton, B.L., Barnes, C.A., Gothard, K.M., Jung, M.W., Knierim, J.J., Kudrimoti, H.K., Qin, Y-L., Skaggs, W.E., Gerrard, J.L., Suster, M., and Weaver, K.L. (1996). J. Exp. Biol., 199, 173-185.

[5] Jung, M.W. and McNaughton, B.L. (1993). Hippocampus, 3, 165-182.

[6] Káli, S. and Dayan, P. (1998) Soc. Neurosci. Abstr., 24, 931.

[7] Muller, R.U. and Kubie, J.L. (1987). J. Neurosci., 77, 1951-1968.

[8] O'Keefe, J. & Burgess, N. (1996). Nature, 381, 425-428.

[9] O'Keefe, J. and Dostrovsky, J. (1971). Brain Res., 34, 171-175.

[10] Quirk, G.J., Muller, R.U., Kubie, J.L., and Ranck J.B. Jr. (1992). J. Neurosci., 12, 1945-1963.

[11] Samsonovich, A. & McNaughton, B.L. (1997). J. Neurosci., 17, 5900-5920.

[12] Samsonovich, A., McNaughton, B.L., and Nadel, L. (1998) Soc. Neurosci. Abstr., 24, 931.

[13] Skaggs, W.E. and McNaughton, B.L. (1998). J. Neurosci., 18, 8455-8466.

[14] Thompson, L.T. and Best, P.J. (1990). Brain Res., 509, 299-308.

[15] Treves, A. and Rolls, E.T. (1994). Hippocampus, 4, 374-391.

[16] Wilson, M.A. and McNaughton, B.L. (1993). Science, 261, 1055-1058.

[17] Zhang, K. (1996). J. Neurosci., 16, 2112-2126.