Attractor neural networks storing multiple space representations: A model for hippocampal place fields

F. P. Battaglia and A. Treves
Neuroscience, SISSA Interactional School for Advanced Studies, Via Beirut 2-4, 34014 Trieste, Italy
(Received 8 July 1998)

A recurrent neural network model storing multiple spatial maps, or ''charts,'' is analyzed. A network of this type has been suggested as a model for the origin of place cells in the hippocampus of rodents. The extremely diluted and fully connected limits are studied, and the storage capacity and the information capacity are found. The important parameters determining the performance of the network are the *sparsity* of the spatial representations and the degree of connectivity, as found already for the storage of individual memory patterns in the general theory of autoassociative networks. Such results suggest a quantitative parallel between theories of hippocampal function in different animal species, such as primates (episodic memory) and rodents (memory for space). [S1063-651X(98)09112-0]

PACS number(s): 87.10.+e, 05.90.+m

I. INTRODUCTION

Knowledge of space is one of the main objects of computation by the brain. It includes the organization of information of many kinds and many origins (memory, the different sensory channels, and so on) into mental constructs, i.e., *maps*, which retain a geometrical nature and correspond to our perception of the outside world.

Every animal species appears to have specialized systems for spatial knowledge in some region of the brain, more or less well developed and capable of performing sophisticated computations. For rodents there is a large amount of experimental evidence that the hippocampus, a brain region functionally situated at the end of all the sensory streams, is involved in spatial processing. Many hippocampal cells exhibit place related firing, that is, they fire when the animal is in a given restricted region of the environment (the ''place field''), so that they contain a representation of the position of the animal in space.

The hippocampus, one of the most widely studied brain structures, shares the same gross anatomical features across mammalian species; nevertheless, it is known to have different functional correlates, for example, in primates and humans (where it is believed to be involved in *episodic memory*, roughly, memory of events) and in rodents, in which it is mainly associated with spatial representation.

One relevant feature of the hippocampus which is maintained across species is a region, named CA3, characterized by massive intrinsic recurrent connections. It was appealing for many theorists to model this region as an *autoassociative memory* storing information in its intrinsic synaptic structure, information which can be retrieved from small cues by means of an attractor dynamics, and which is represented in the form of activity configurations.

Within the episodic memory framework, each attractor configuration of activity is the internal representation of some memory item. The CA3 autoassociative network can be seen as the heart of the hippocampal system, containing the very complex, intermodal representations peculiar to episodic memory. Autoassociative, or attractor neural networks have been extensively studied with the tools of statistical physics [1]. Efficiency measures, like the number of storable memory items or the quality of retrieval, can be computed for any appropriately defined formal model. An intensive effort was performed to embed in these idealized models more and more elements of biological realism, trying to capture the relevant anatomical and functional features affecting the performance of the network. It was then found that a theory of the hippocampus (or, more precisely, CA3) as an autoassociative network has as its fundamental parameters the degree of intrinsic connectivity, i.e., the average number of units which send connections to a given unit, and the *sparsity* of the representations, roughly the fraction of units which are active in one representation. These parameters have biological correlates that are measurable with anatomical and neurophysiological techniques.

Spatial processing, as it is performed by the rodent hippocampus, also involves memory of some kind; recent experimental evidence supports the idea that spatially related firing is not driven exclusively by sensory inputs but also reflects some internal representation of the explored environment. First, place fields are present and stable also in the dark, and in conditions of deprived sensory experience. Second, completely different arrangements of place fields are found in different environments or even in the same environment in different behavioral conditions.

These findings have led to the hypothesis that CA3 (like, perhaps, other brain regions with dense connectivity) stores ''charts,'' representations of environments in the form of abstract manifolds, on which each neuron corresponds to a point, that is the center of its place field. Place fields arise as a result of an attractor dynamics, whose stable states are ''activity peaks'' centered, in the chart space, at the animal's position. It is important to note that the localization of each neuron on a chart does not appear to be related to its physical location in the neural tissue.

The positions of place fields are encoded by the recurrent connections, and it is possible to store many different charts in the same synaptic structure, just as many different patterns are stored in a Hopfield net, for example, and different activity peaks can be successively evoked by appropriate in-

puts, just as it happens with autoassociative memories.

It is interesting to address the issue of whether an episodic memory network and the ''spatial multichart'' memory network share the same functional constraints, so that a biological brain module capable of performing one of the tasks is also adequate for the other. Here we present a statistical mechanics analysis of the multichart network, focusing on the parallel with autoassociative memory in the usual (episodic memory) sense. It is found that the performances of these two networks are governed by very similar laws, if the parallel between them is drawn in the appropriate way.

In Sec. II the case of a single attractor chart stored is studied, then in Sec. III the case of multiple stored charts is analyzed and the storage capacity is found, first for a simplified model and then for a more complex model which makes it possible to address the issue of sparsity of representations. In Sec. IV the storable information in a multichart network is calculated, making more precise the sense in which such a network is a store of information, and completing the parallel with autoassociative memories.

II. THE SINGLE MAP NETWORK

As a first step, we consider the case of a single attractor map encoded in the synaptic structure, as was proposed in [2]. We focus here on the shape and properties of the attractor states, as a useful comparison for the following treatment of the multiple charts case.

The neurons are modeled as threshold linear units, with firing rate:

$$V_i = g[h_i - \theta]^+ = g(h_i - \theta)\Theta(h_i - \theta), \tag{1}$$

i.e., equal to zero if the content of the square brackets is negative. h represents the synaptic input current, coming from other cells in the same module, θ is a firing threshold, which may incorporate the effect of a subtractive inhibitory input, common to all the cells, as it will be illustrated later on. The connectivity within the module is shaped by the selectivity of the units. If $\mathbf{r}_i$ is the position of the center of the place field of the ith cell in a manifold M of size $|M|$, corresponding to the environment, the connection between cells i and j may be expressed as

$$J_{ij} = \frac{|M|}{N} K(|\mathbf{r}_i - \mathbf{r}_j|), \tag{2}$$

where K is a monotone decreasing function of its argument.

The synaptic input to the ith cell is therefore given by

$$h_i = \sum_j J_{ij} V_j = \sum_j \frac{|M|}{N} K(|\mathbf{r}_i - \mathbf{r}_j|) V_j . \tag{3}$$

If the number N of cells is large, and the place fields centers (pfc) are homogeneously distributed over the environment M (be it one or two dimensional), we can replace the sum over the index j with an integration over the coordinates of the pfc's:

$$h(\mathbf{r}) = \int_M d\mathbf{r}' K(|\mathbf{r} - \mathbf{r}'|) V(r') . \tag{4}$$

Note that the normalization in Eq. (2) is chosen so as to keep the synaptic input to a given unit fixed when $|M|$ varies and the number of units is kept fixed, that is, the density of pfc's $N/|M|$ varies (the $|M|$ factor will then compensate for the fewer input units within the range of substantial K strength). A fixed-point activity configuration must have the form

$$V(\mathbf{r}) = g\left[\int_M d\mathbf{r}' K(|\mathbf{r} - \mathbf{r}'|) V(\mathbf{r})' - \theta \right]^+ . \tag{5}$$

We could write Eq. (5) as

$$V(\mathbf{r}) = \begin{cases} g\left[\int_\Omega d\mathbf{r}' K(|\mathbf{r} - \mathbf{r}'|) V(\mathbf{r}') - \theta \right] & \mathbf{r} \in \Omega \\ 0, & \mathbf{r} \notin \Omega , \end{cases} \tag{6}$$

where Ω is a domain for which there exists a solution of Eq. (2) that is zero on the boundary.

If only solutions for which Ω is a convex domain are considered, the fact that $V(r)$ is zero on $\partial\Omega$ will ensure that units with pfc's outside Ω are under threshold, therefore their activity is zero and solutions of Eq. (6) are guaranteed to be solutions of Eq. (5). The size and the shape of the domain Ω in which activity is different from zero is determined by Eq. (2). As a first remark, we notice that it is independent of the value of the threshold θ. In fact, if V_θ is a solution of Eq. (6) with threshold θ, given the linearity of Eq. (6) within Ω,

$$V_{\theta'} = \frac{\theta'}{\theta} V_\theta$$

will be a solution of the same equation with θ' instead of θ, with the same null boundary conditions on Ω. Rescaling the threshold will then have the effect of rescaling the activity configuration by the same coefficient. This means that subtractive inhibition cannot shape, e.g., shrink or enlarge, this stable configuration, and therefore it is not relevant for a good part of the subsequent analysis. Some form of inhibition is nevertheless necessary to prevent the activity from exploding. Moreover, there are fluctuation modes which cannot be controlled by overall inhibition as they leave the total average activity constant. They will be treated in Sec. III C. It is found that, at least in the one-dimensional (1D) case, these modes do not affect stability in the single chart case.

In the absence of an external input, any solution can be at most marginally stable, because a translation of the solution is again a solution of Eq. (6). An external, ''symmetry breaking'' input, taken as small when compared to the contribution of recurrent synapses, is therefore implicit in the following analysis.

A. The one-dimensional case

The case of a recurrent network whose attractors reflect the geometry of a one-dimensional manifold, besides being a conceptual first step in approaching the two-dimensional case, is relevant by itself, for example, in modeling other brain systems showing direction selectivity, e.g., in head di-

rection cells [3,4], and also for place fields on one-dimensional environments [5].

In this case Eq. (6) reads

$$V(r)=g\left(\int_{-R}^{R}K(|r-r'|)V(r')-\theta\right),$$

$$V(R)=V(-R)=0. \tag{7}$$

For several specific forms of the kernel K it is possible to solve explicitly Eq. (7), yielding interesting conclusions. For example, if

$$K(|r-r'|)=e^{-|r-r'|}, \tag{8}$$

(see, also, [2]) differentiating Eq. (7) twice yields

$$V''(r)=-\gamma^2V(r)+g\theta, \tag{9}$$

where $\gamma=\sqrt{2g-1}$.

Solutions vanishing at $-R$ and R (and not vanishing in $]-R,R[$), have the form

$$V(r)=A\cos(\gamma r)+\frac{g\theta}{2g-1} \tag{10}$$

with

$$A=-\frac{g\theta}{(2g-1)\cos(\gamma R)}. \tag{11}$$

The value of R for which Eq. (10) is a solution of Eq. (7) is determined by the integral equation itself: for example, by evaluating $V'(R)$ or $V'(-R)$ from Eq. (7) we get

$$V'(-R)=-V'(R)=g\theta. \tag{12}$$

Substituting Eqs. (10) and (11) in Eq. (12) we have

$$\tan(\gamma R)=-\gamma,$$

so that

$$R=\frac{\tan^{-1}(-\gamma)+n\pi}{\gamma}.$$

Requiring R to be positive and $V(x)$ to be positive for $-R<r<R$, leads us to choose

$$R=\frac{-\tan^{-1}(\gamma)+\pi}{\gamma}. \tag{13}$$

Note that $A>0$. R is then a monotone decreasing function of γ, and therefore of the gain g.

This is also true for other forms of the connection kernel K. As an example, consider the kernel

$$K(r-r')=\cos(r-r'). \tag{14}$$

By a similar treatment it is shown that a solution is obtained with

$$R=\frac{1}{g}. \tag{15}$$

The kernel

$$K(r-r')=\Theta(1-|r-r'|)(1-|r-r'|) \tag{16}$$

will result in a peak of activity of semiwidth

$$R=\frac{\pi}{\sqrt{2g}}. \tag{17}$$

Equations of this type (5) have more solutions in addition to the ones considered above, representing a single activity peak. For example, if we consider an infinite environment, periodic solutions will be present as well, representing a row of activity peaks separated by regions of zero activity. These solutions can be verified to be unstable if we model the inhibition as an homogeneous term acting on all cells in the same way and depending on the average activity. Intuitively, if we perturb the solution by infinitesimally displacing one of the peaks, it will tend to collapse with the neighbor that has come closer.

B. The two-dimensional case

To model the place cells network in the hippocampus, we need to extend this result to a two-dimensional environment. The equation for the neural activity will be

$$V(\mathbf{r})=g\left[\int_M d\mathbf{r}'K(|\mathbf{r}-\mathbf{r}'|)V(\mathbf{r}')-\theta\right]^+. \tag{18}$$

The generalization to 2D is straightforward if for the kernel $K(|\mathbf{r}-\mathbf{r}'|)$ we consider the one with Fourier transform

$$\hat{K}(\mathbf{p})=\frac{2}{1+\mathbf{p}^2}, \tag{19}$$

[the two-dimensional analog of the kernel of Eq. (8)] that is, a kernel resembling the propagator of a Klein-Gordon field in Euclidean space. The fact that this kernel is divergent for $(\mathbf{r}-\mathbf{r}')\to 0$ does not give rise to particular problems, since, in the continuum limit of Eq. (4), the contribution to the field h coming from the nearby points will stay finite, and in fact two units will be assigned pfc's so close to each other as to yield an overwhelmingly high connection only with a small probability. Let us look for a solution with circular symmetry such that activity $V(\mathbf{r})$ is zero outside the circle of radius R, $\mathcal{C}(R)$. If we apply the Laplacian operator on both sides of

$$V(\mathbf{r})=g\int_{\mathcal{C}(R)}d\mathbf{r}'K(\mathbf{r}-\mathbf{r}')V(\mathbf{r}')-\theta, \tag{20}$$

we obtain

$$\nabla^2V(\mathbf{r}')=-\gamma^2V(\mathbf{r}')+g\theta \tag{21}$$

(again, $\gamma^2=2g-1$), which in polar coordinates reads

$$V''(r)+\frac{1}{r}V'(r)=-\gamma^2V(r)+g\theta. \tag{22}$$

The solution is

$$V(r)=AJ_0(\gamma r)+\frac{g\theta}{2g-1}. \tag{23}$$

J_0 is the Bessel function of order 0. For the solution to vanish on the boundary of $\mathcal{C}(R)$ one must take

$$A=\frac{g\theta}{(2g-1)J_0(\gamma R)}.$$

The other condition that determines R may be found by substituting Eq. (23) in Eq. (20). Here again, $R(g)$ is a monotone decreasing function. As in the one-dimensional case, solutions with a nonconnected (or even nonconvex) support can be seen not to be stable.

III. STORING MORE THAN ONE MAP

Let us imagine now that the pfc's for each cell are drawn with uniform distribution on the environment manifold M, and connections are formed according to Eq. (2). Several ''space representations'' may be created by drawing again at random the pfc of each cell from the same distribution. The connection between each pair of cells will then be the sum of a number of terms of the form (2), one for every ''space representation,'' or ''map,'' or ''chart.'' With $p=\alpha N$ maps, and the pfc of the ith cell in the μth map indicated by $\mathbf{r}_i^{(\mu)}$:

$$J_{ij}=\sum_{\mu=1}^{p}\frac{|M|}{N}K(|\mathbf{r}_i^{(\mu)}-\mathbf{r}_j^{(\mu)}|). \tag{24}$$

The question that immediately arises is: what is the capacity of this network, that is, how many maps can we store, so that stable activity configurations, corresponding to some region in the environment described by one map, like the ones described by the solutions of Eq. (6), are present? The problem resembles the classic attractor neural network problem [1], with threshold linear units. A standard treatment has been developed [6] allowing us to calculate the capacity of a network of threshold linear units with patterns drawn from a given distribution and stored by means of a hebbian rule. The treatment is very simplified in the extreme dilution limit [7,8]. In the next sections it will be shown how this treatment can be extended to the map case, first for one particular form of the kernel K, leading to the solution of the capacity problem for a fully connected network; in the following, the solution is extended to more general kernels, first in the diluted limit, then for the fully connected network.

Another related question is: how much information is the synaptic recurrent structure encoding, and in which sense is the synaptic structure a store of information? The aim is to develop a full parallel between the multichart network and autoassociative networks, and if possible to characterize the parameters constraining the performance of this system.

A. The fully connected network: ''Dot product'' kernel

Let us consider a manifold M with periodic boundary conditions, that is, a circle in one dimension and a torus in two dimensions. The pfc of a cell $\mathbf{r}_i$ can then be described by a two-dimensional unit vector $\boldsymbol{\eta}_i$ for the one-dimensional case and by a pair of unit vectors $\boldsymbol{\eta}_i^{1,2}$ for the two-dimensional case. Suppose now that the contribution from the μth map to the connection between cell i and cell j is given by

$$K(|\mathbf{r}_i^{(\mu)}-\mathbf{r}_j^{(\mu)}|)=\sum_{l=1}^{d}(\boldsymbol{\eta}_i^{l(\mu)}\cdot\boldsymbol{\eta}_j^{l(\mu)}+1), \tag{25}$$

so that

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}\sum_{l=1}^{d}(\boldsymbol{\eta}_i^{l(\mu)}\cdot\boldsymbol{\eta}_j^{l(\mu)}+1), \tag{26}$$

where d is the dimensionality.

$p=\alpha N$ is the number of stored charts. Equation (25) describes an excitatory, very widespread form for the kernel (2) (the contribution to the connectivity is zero only if the rpc's of the two cells are at the farthest points apart, i.e., at 180°). This spread of connectivity would lead to configurations of activity that are large in the rpc space, which translated in autoassociative memory language would be very ''unsparse,'' i.e., very distributed representations. It is therefore plausible that this will severely limit the capacity of the net. In any case, the form of Eq. (25), factorizable in one term depending on $\boldsymbol{\eta}_i$ and one term depending on $\boldsymbol{\eta}_j$, after incorporating the constant part in a function $b^0(x)$, makes it possible to perform the free-energy calculation through Gaussian transformations as in [6]. A similar model has been studied in [9] with McCulloch-Pitts neurons.

A Hamiltonian useful to describe the thermodynamics of such a system is

$$H=-\frac{1}{2}\sum_{i,j(\neq i)}J_{ij}V_iV_j-NB\left(\sum_i\frac{V_i}{N}\right)$$
$$-\sum_l\sum_i\sum_\mu\mathbf{s}^{l(\mu)}\cdot\boldsymbol{\eta}_i^{l(\mu)}V_i, \tag{27}$$

where $B(x)=\int^x b(y)dy$, and $b(x)$ is a function describing a uniform inhibition term depending on the average activity in the net. $\mathbf{s}^{l(\mu)}$ is a symmetry breaking field, pointing in a direction in the μth map space. The mean-field free energy in the replica-symmetric approximation can be calculated (the partition function is calculated as the trace over a measure that implements the threshold-linear transfer function, see [6]). The presence of a phase with spatially specific activity correlated with one map will be signaled by solutions of the mean-field equations with a nonzero value for the order parameter

$$\mathbf{x}^{l(\mu)}=\frac{1}{Nd}\sum_{i=1}^{N}\boldsymbol{\eta}_i^{l(\mu)}V_i, \tag{28}$$

which plays the role of the overlap in an autoassociative memory. This parameter has the meaning of a population vector [10], that is, the animal position is indicated by an average over pfc's of the cells weighted by cells activity.

The set of resulting mean-field equations can be reduced to a set of two equations, Eqs. (A7) and (A8), in two variables, the ''nonspecific'' signal-to-noise ratio, w, and the ''specific,'' space-related signal-to-noise ratio v. The details of the calculation are reported in Appendix A.

The critical value α_c indicating the storage capacity of the network is the maximum value for which Eq. (A7) still admits solutions corresponding to space-related activity (nonzero v) and may be found numerically. At this value α_c the system undergoes a first-order phase transition towards a state in which no space-related activity is possible. Equation (A8) gives the range of gain values for which there exist solutions at a given $\alpha \leqslant \alpha_c$ [6].

In this model there is no possibility of modulating the spread of connections in the chart space. As we anticipated, the activity configurations that one obtains are very wide, with a large fraction of units active at the same time. Cells will have very large place fields, covering a large part of the environment (of the order of roughly one half for the one-dimensional case, and roughly one quarter for the two-dimensional case). As one would infer from the analysis of autoassociative memories storing patterns, for example binary, these ''unsparse'' representations of space will lead to a very small capacity of the net.

For the model defined on the one-dimensional circle the capacity found is $\alpha_c \sim 0.03$. At this value the system undergoes a first-order transition. As α increases beyond α_c, $\mathbf{x}$ jumps discontinuously from a finite value to zero.

The capacity for the diluted analog of this model (see [8], Appendix A and Sec. III B) is given by the equation

$$E_1(w,\mathbf{v}) \equiv [(1+\delta)A_2]^2 - \alpha A_3 = 0. \tag{29}$$

Remember that in this case $p = \alpha c N$ where c is the connectivity fraction parameter; see Sec. III B. In this case $\alpha_c \sim 0.25$. At α_c the transition is second order, with the ''spatial overlap'' $\mathbf{x}$ approaching continuously zero, verified at least with the precision at which it was possible to solve numerically Eq. (29). For the 2D case, storage capacities are $\alpha_c \sim 0.0008$ for the fully connected network and $\alpha \sim 0.44$ for the diluted network.

To get a larger capacity, and to provide a possible comparison with the experimental data from the hippocampus, in which the tuning of place fields is generally narrow, we must extend our treatment to more general kernels, and this will be done in the following two sections.

B. Generic kernel: Extremely diluted limit

Consider a network in which every threshold-linear unit, whose activity is denoted by V_j, senses a field

$$h_i = \frac{1}{c} \sum_{j \neq i} C_{ij} J_{ij} V_j, \tag{30}$$

where J_{ij} is given by Eq. (24). From now on the kernel K is defined as

$$K(\mathbf{r}-\mathbf{r}') = \hat{K}(\mathbf{r}-\mathbf{r}') - \overline{K},$$

$$\overline{K} = \langle\langle \hat{K}(\mathbf{r}-\mathbf{r}') \rangle\rangle \tag{31}$$

for any $\mathbf{r}$, where $\langle\langle\ \rangle\rangle$ means averaging over $\mathbf{r}$. With this notation, whatever the original kernel $\hat{K}$, K is the subtracted kernel which averages to zero. The manifold M is taken with periodic boundary condition (that is a circle in one dimension and a torus in the two-dimensional case).

C_{ij} is a ''dilution matrix''

$$C_{ij} = \begin{cases} 1 & \text{with prob} \quad c, \\ 0 & \text{with prob} \quad 1-c, \end{cases} \tag{32}$$

and $Nc/\ln N \to 0$ as $N \to \infty$. In the thermodynamic limit $N \to \infty$ the activity of any two neurons V_i and V_j will be uncorrelated [7]. A number of charts $p = \alpha c N$ is stored. Looking for solutions with one ''condensed'' map, that is, solutions in which activity is confined to units having pfc for a given chart in a certain neighborhood, it is possible to write the field h_i as the sum of two contributions, a ''signal,'' due to the condensed map and a ''noise'' term, ρz (z being a random variable with Gaussian distribution and variance one) due to all the other, uncondensed, maps. In the continuum limit, labeling units with the position $\mathbf{r}^1$ of their pfc in the condensed map,

$$h(\mathbf{r}^1) = g \int_M d\mathbf{r}^{1'} K(\mathbf{r}^1 - \mathbf{r}^{1'}) V(\mathbf{r}^{1'}) + \rho z; \tag{33}$$

the noise will have a variance

$$\rho^2 = \alpha y |M|^2 \langle\langle K^2(\mathbf{r}-\mathbf{r}') \rangle\rangle, \tag{34}$$

where

$$y = \frac{1}{N} \sum_{i=1}^{N} \langle V_i^2 \rangle. \tag{35}$$

The fixed-point equation for the average activity profile $x^1(\mathbf{r})$ is

$$x^1(\mathbf{r}) = g \int^{+} Dz (h(\mathbf{r}) - \theta), \tag{36}$$

where again Dz is the Gaussian measure, and

$$h(\mathbf{r}) = \int d\mathbf{r}' K(\mathbf{r}-\mathbf{r}') x^1(\mathbf{r}') + b(x) - \rho z \tag{37}$$

and

$$x = \int \frac{d\mathbf{r}}{|M|}\ x^1(\mathbf{r}) \tag{38}$$

is the average overall activity. The average squared activity (entering the noise term) will read

$$y = g^2 \int \frac{d\mathbf{r}}{|M|} \int^{+} Dz (h(\mathbf{r}) - \theta)^2. \tag{39}$$

The fixed-point equations may be solved introducing the rescaled variables

$$w = \frac{b(x) - \theta}{\rho}, \tag{40}$$

$$v(\mathbf{r}) = \frac{x^1(\mathbf{r})}{\rho}. \tag{41}$$

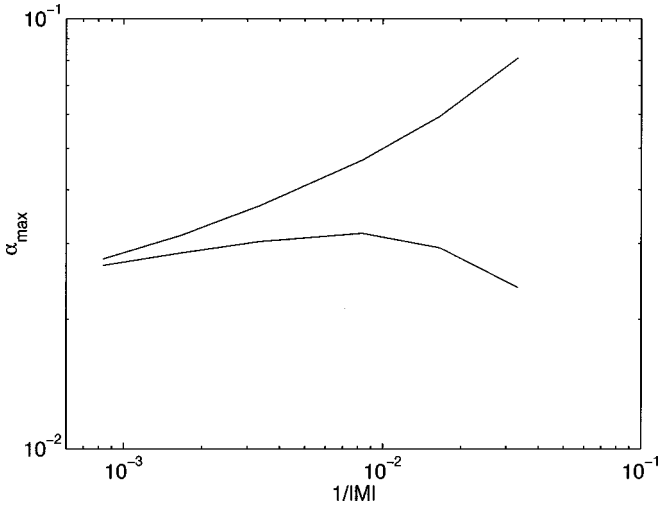

FIG. 1. The storage capacity plotted as a function of the ''map sparsity'' a_m, for the 1D model, for the extremely diluted (upper curve) and the fully connected (lower curve) limits.

The fixed-point equation for $v(\vec{r})$ is

$$v(\mathbf{r})=g\mathcal{N}\left(\int d\mathbf{r}'K(\mathbf{r}-\mathbf{r}')v(\mathbf{r}')+w\right), \tag{42}$$

where

$$\mathcal{N}(x)=x\Phi(x)+\sigma(x), \tag{43}$$

[$\Phi(x)$ and $\sigma(x)$ are defined in Eq. (A15) and Eq. (A16)] is a ''smeared threshold linear function,'' monotonically increasing, with

$$\lim_{x\to-\infty}\mathcal{N}(x)=0$$

and

$$\lim_{x\to+\infty}\mathcal{N}(x)/x=1.$$

In terms of w and $v(\mathbf{r})$, y reads

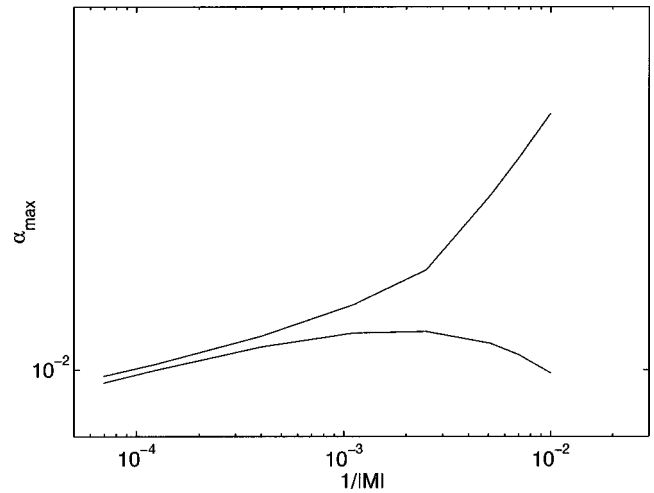

FIG. 2. Same as Fig. 1, for the 2D model. The capacity is smaller than for the 1D model for the same a_m.

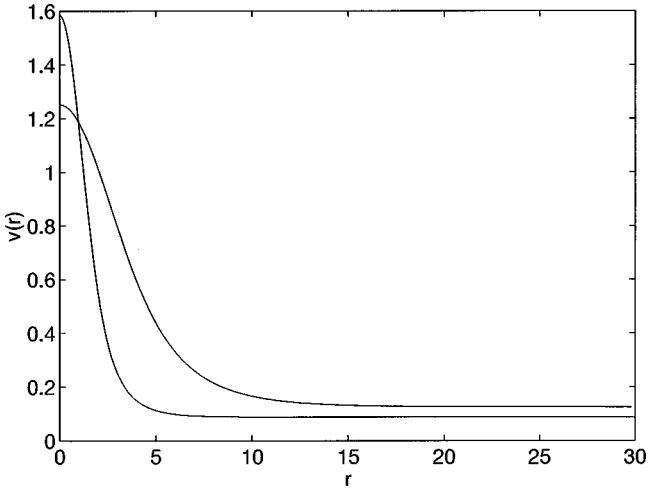

FIG. 3. The ''activity peak'' profile corresponding to the solution of Eq. (42) at the maximal storage level at $|M|=30$ and $|M|=15$. The second case is plotted expanded to match the environment size of the first one and to show the effect of more widespread connections.

$$y=\rho^2g^2\int\frac{d\mathbf{r}}{|M|}\mathcal{M}\left(\int d\mathbf{r}'K(\mathbf{r}-\mathbf{r}')v(\mathbf{r}')+w\right), \tag{44}$$

where

$$\mathcal{M}(x)=(1+x^2)\Phi(x)+x\sigma(x). \tag{45}$$

By substituting Eq. (44) in Eq. (34), we obtain

$$\frac{1}{\alpha}=g^2|M|\langle\langle K\rangle\rangle\int d\mathbf{r}\mathcal{M}\left(\int d\mathbf{r}'K(\mathbf{r}-\mathbf{r}')v(\mathbf{r}')+w\right). \tag{46}$$

If we can solve Eq. 42 and find $v(\mathbf{r})$ as a function of w and g, a solution is found corresponding to a value of α given by Eq. (46). To find the critical value of α, we have to maximize α over w and g. The mathematical solution of Eq. (42) is treated in Appendix B.

With this model, we can modulate the spread of connections by acting on $K(\mathbf{r}-\mathbf{r}')$ or alternatively, by varying the size of the environment. The results are depicted in Fig. 1 for the 1D circular environment and in Fig. 2 for the 2D toroidal environment (upper curves). Examples of the solutions of Eq. 42 are displayed in Fig. 3 for the 1D environment and in Fig. 4 for the 2D environment.

We note that, as the environment gets larger in comparison to the spread of connections (therefore, to the size of the activity peak), the capacity decreases approximately as

$$\alpha_c\sim-1/\ln(a_m), \tag{47}$$

where a_m is the *map sparsity* and it is equal to

$$a_m=\frac{k_d}{|M|}, \tag{48}$$

where k_d is a factor roughly equal to $\sim$4.5 for the 1D model and $\sim$3.6 for the 2D model.

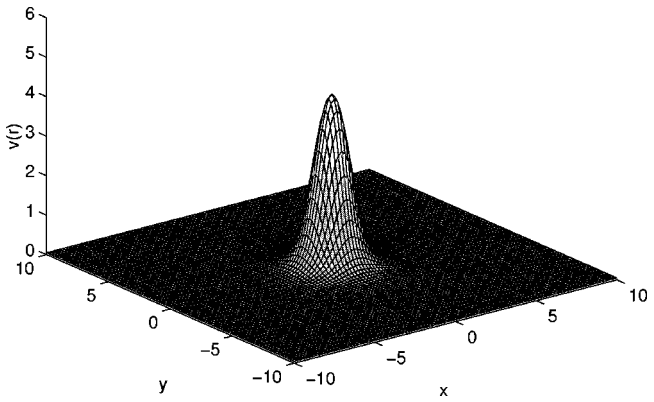

FIG. 4. The maximal storage activity peak profile in 2D at $|M|=400$.

That is, the sparser the coding, the less the capacity. This is, at first glance, in contrast with what is known from the theory of autoassociative networks, in which sparser representations usually lead to larger storage capacities.

For comparison, keeping the formalism of [6], for threshold-linear networks with hebbian learning rule, encoding memory patterns $\{r_i\}_{i=1\ldots N}$ with sparsity a defined as

$$a_p=\frac{\langle\langle r\rangle\rangle^2}{\langle\langle r^2\rangle\rangle}$$

(for binary patterns this is equal to the fraction of active units), and for small a, the capacity is given by

$$\alpha_p\sim\frac{1}{a_p\ln(1/a_p)}. \tag{49}$$

The apparent paradox (larger capacity with sparser patterns, smaller with sparser charts) is solved as one recognizes that each chart can be seen as a collection of configurations of activity relative to different points in space covering, as in a tiling, the whole environment. Each configuration is roughly equivalent to a pattern in the usual sense. Intuitively, and in a sense that will be made clearer below, a chart is equivalent, in terms of ''use of synaptic resources,'' to a number proportional to a_m^{-1} of patterns of sparsity a_m.

The proportionality coefficient or, equivalently, the distance at which different configurations are to be considered to establish a correct analogy, will be dealt with in Appendix D. These considerations and the comparison of Eqs. (47) and (49) make clear that α_c is the exact analog of the pattern autoassociators' α_p.

C. Inhibition independent stability

The dynamical stability of the solutions of Eq. (42) is in general determined by the precise functional form chosen for the inhibition, which we assumed to be a function of the average overall activity in the net. Nevertheless, there are fluctuation modes that leave the average activity unaltered. Stability against these modes is therefore unaffected by the inhibition and may be checked already for a general model. Let us consider a ''synchronous'' dynamics, that is, all the neurons are updated simultaneously at each time step. The evolution operator for the variables $V(\mathbf{r},t)$ and $\rho(t)$ is

$$V(\mathbf{r},t+1)=g\rho(t)\times\mathcal{N}\left(\int_M\frac{dr'}{|M|}K(\mathbf{r}-\mathbf{r}')\frac{V(\mathbf{r}',t)}{\rho(t)}+\frac{b(x(t))}{\rho(t)}\right), \tag{50}$$

$$\rho^2(t+1)=g^2\alpha|M|\rho^2(t)\langle\langle K^2\rangle\rangle\int_M d\mathbf{r}\times\mathcal{M}\left(\int_M\frac{dr'}{|M|}K(\mathbf{r}-\mathbf{r}')\frac{V(\mathbf{r}',t)}{\rho(t)}+\frac{b(x(t))}{\rho(t)}\right). \tag{51}$$

This evolution operator has as its fixed points $V_0(\mathbf{r})=\rho_0 v_0(\mathbf{r})$ and ρ_0 where $v_0(\mathbf{r})$ and ρ_0 are the solutions of Eqs. (42), (34), and (44), i.e., the stable states of our system.

We can linearize the evolution operator around $[V_0(\mathbf{r}),\rho_0]$ and look for fluctuation modes (eigenvectors) $[\delta V(\mathbf{r}),\delta\rho]$ with

$$\int_M dr\ \delta V(\mathbf{r})=0. \tag{52}$$

We obtain the following equations:

$$\lambda\,\delta V(\mathbf{r})=g\Phi(u_0(\mathbf{r}))\times\left[\int_M d\mathbf{r}'K(\mathbf{r}-\mathbf{r}')\,\delta V(\mathbf{r}')\right]+g\sigma(u_0(\mathbf{r}))\delta\rho, \tag{53}$$

$$\lambda\,\delta\rho=\left(1-\frac{1}{2}g\alpha|M|\langle\langle K^2\rangle\rangle\int_M d\mathbf{r}\ u_0(\mathbf{r})v_0(\mathbf{r})\right)\times\delta\rho+\frac{1}{2}g\alpha|M|\langle\langle K^2\rangle\rangle\int_M d\mathbf{r}\ u_0(\mathbf{r})\,\delta V(\mathbf{r}), \tag{54}$$

where

$$u_0(\mathbf{r})=\mathcal{N}^{-1}\left(\frac{v_0(\mathbf{r})}{g}\right).$$

Inserting Eq. (53) in Eq. (52):

$$\delta\rho=-\frac{\int_M d\mathbf{r}'\Phi(u_0(\mathbf{r}))\left[\int_M dr\ K(\mathbf{r}-\mathbf{r}')\,\delta V(\mathbf{r}')\right]}{\int_M d\mathbf{r}\ \sigma(u_0(\mathbf{r}))}. \tag{55}$$

Equation (55) can be inserted again in Eq. (53), obtaining a closed integral equation in δV. Unfortunately, this equation is very difficult to solve, but we can derive a stability condition by making an ansatz in the form of the eigenfunction $\delta V(\mathbf{r})$. More precisely, let us concentrate on the 1D case. We look for solutions with even symmetry (we know there must be an eigenfunction with odd symmetry, and an eigenvalue equal to 1, corresponding to a coherent displacement of the activity peak). This kind of solution corresponds to

spreading and shrinking of the activity peak. Let us assume that the even eigenfunction with the highest eigenvalue (the most unstable) has only two nodes [an even eigenfunction must have at least two nodes because of Eq. (52)], at r_0 and $-r_0$. Let us take the sign of the eigenfunction $\delta V(r)$ such that $\delta V(0)>0$. From Eqs. (52) and (55) we see that

$$\delta\rho<0.$$

Now, from Eq. (54)

$$\lambda=\left(1-\frac{1}{2}g\alpha|M|\langle\langle K^2\rangle\rangle\int_M dr\ u_0(r)v_0(r)\right)+\frac{1}{2}g\alpha|M|\langle\langle K^2\rangle\rangle\int_M dr\ u_0(r)\frac{\delta V(r)}{\delta\rho}, \tag{56}$$

and we recognize that

$$\int_M dr\ u_0(r)\frac{\delta V(r)}{\delta\rho}<0.$$

Thus,

$$\lambda<1-\frac{\Gamma}{2} \tag{57}$$

with

$$\Gamma=g\alpha|M|\langle\langle K^2\rangle\rangle\int_M dr\ u_0(r)v_0(r). \tag{58}$$

Thus, if the ansatz we formulated holds, we have a stability condition $\Gamma>0$, which is found to be fulfilled for all the solutions we found relative to maximal storage capacity. This implies that the storage capacity result is not affected by instability of the solutions, provided of course that an appropriate form for inhibition is chosen. This stability result is also related to the correlation in the static noise for two solutions centered at different pfc's, as we will show in Appendix D.

It can also be shown that by taking the $\alpha\to 0$ limit (i.e., the single chart case), one always has $\Gamma>0$ since it is $v_0(r)=0$ when $u_0(r)<0$.

D. The fully connected model

The treatment of the model with the fully connected network and a kernel K for connection weights satisfying the condition (31) will use the replica trick to average over the disorder (the realizations of the **r**'s) and will eventually lead to a nonlinear integral equation for the average activity profile in the space of the ''condensed map'' very similar to Eq. (42). Let the Hamiltonian of the system be

$$H=-\frac{1}{2}\sum_{i,j(\neq i)}J_{ij}^c V_i V_j - NB\left(\sum_i\frac{V_i}{N}\right)-\sum_i\sum_\mu \mathbf{s}^{l(\mu)}\cdot\mathbf{r}_i^{(\mu)}V_i, \tag{59}$$

where now the J_{ij}^c are given by Eq. (24) with a generic kernel

$$K(\mathbf{r}-\mathbf{r}')=\hat{K}(\mathbf{r}-\mathbf{r}')-\overline{K}, \tag{60}$$

where, again,

$$\overline{K}=\langle\langle\hat{K}(\mathbf{r}-\mathbf{r}')\rangle\rangle.$$

The free-energy calculation is sketched in Appendix C. Again, the stable states of the system are governed by mean-field equations. The mean-field equation (C16) is an integral equation in the functional order parameter $v(\mathbf{r})$, the average space profile of activity.

If we are able to solve Eq. (C16) and find $v^\sigma(\mathbf{r})$ as a function of w and g', by substituting Eqs. (C17) and (C18) in Eq. (C11) we have an equation that gives us the value of α corresponding to that pair (g',w). α_c is then the maximum of α over the possible values of (g',w).

To solve Eq. (C16), it is easy to verify that if $\tilde{v}(\mathbf{r})$ is a solution of

$$\tilde{v}(\mathbf{r})=g'\mathcal{N}\left(\int_M d\mathbf{r}'\hat{K}(\mathbf{r}-\mathbf{r}')\tilde{v}(\mathbf{r}')+\hat{w}\right) \tag{61}$$

with

$$\hat{w}=w-\overline{K}\int_M d\mathbf{r}\tilde{v}(\mathbf{r}),$$

that is, the same equations as Eqs. (B2) and (B3), then

$$\tilde{v}(\mathbf{r})=\int_M d\mathbf{r}'[L(\mathbf{r}-\mathbf{r}')-\overline{L}]\tilde{v}(\mathbf{r}')$$

is a solution of Eq. (C16). $\tilde{v}$ can therefore be interpreted as the average activity profile, apart from a constant. Equation (61) can be solved with the same procedure used for Eq. (42), and the maximum value of α can be found by maximizing over g' and $\hat{w}$.

The results for 1D and 2D environment are depicted in Figs. 1 and 2 (lower curves). As we may expect from pattern autoassociator theory, the capacity is much lower than for the diluted model, due to an increased interference between different charts.. As the sparsity $a\sim 1/|M|$ gets smaller, the capacities of the two models get closer, both being proportional to $1/\ln(K_d|M|)$. Reducing the sparsity parameter of space representations has therefore the effect of minimizing the difference between nets with sparse and full connectivity.

E. Sparser maps

A possible extension of this treatment is inspired from the experimental finding that, in general, not all the cells have place cells in a given environment. Reference [11], e.g., reported that $\sim$28–45% of pyramidal cells of CA1 have a place field in a certain environment. We would like to see how this fact could affect the performance of the multichart autoassociator. It is then natural to introduce a new sparsity parameter, the *chart sparsity* a_c indicating the fraction of cells which participate in a chart. We will show that, for the capacity calculation, a_c^{-1} ''sparse'' charts are equivalent to a

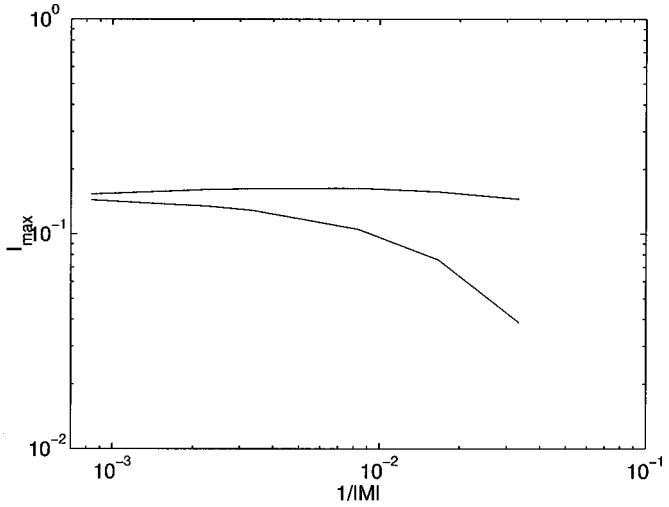

FIG. 5. The maximum stored information per synapse, as a function of $1/|M|$.

single ''full'' chart, of size a_c^{-1}. We will present the argument for the diluted case; the fully connected case is completely analogous.

Let m_i^μ be equal to 1 if cell i participates in chart μ, that is, with probability a_c. Thus, the synaptic coupling J_{ij} will read

$$J_{ij}=\sum_{\mu=1}^{p}\frac{|M|}{a_c N}K(\mathbf{r}_i^{(\mu)}-\mathbf{r}_j^{(\mu)})m_i^\mu m_j^\mu . \tag{62}$$

Let us consider a solution with one condensed map: cells participating with pfc in r in that map will have a space-related signal-to-noise ratio

$$v(\mathbf{r})=g\mathcal{N}\left(\int d\mathbf{r}'K(\mathbf{r}-\mathbf{r}')v(\mathbf{r}')+w\right) \tag{63}$$

for all the neurons *not* participating in the condensed map we will have

$$v=\mathcal{N}(w). \tag{64}$$

The noise will have a variance

$$\rho^2=\alpha a_c y\left(\frac{|M|}{a_c}\right)^2\langle\langle K^2(\mathbf{r}-\mathbf{r}^\mu\rangle\rangle, \tag{65}$$

that is a_c times the value we would get for the same number of ''full'' charts with size $|M|/a_c$, and now

$$y=\rho^2 g^2\left\{a_c\int_M\frac{d\mathbf{r}}{|M|}\mathcal{M}\left(\int d\mathbf{r}'K(\mathbf{r}-\mathbf{r}')v(\mathbf{r}')+w\right)+(1-a_c)\mathcal{M}(w)\right\}. \tag{66}$$

By comparing Eqs. (44) and (66), and remembering that for $\vec{r}$ far from the activity peak $v(\mathbf{r})\sim\mathcal{N}(w)$, we realize that this y value is approximately equivalent to the y value we would get for ''full'' charts of size $|M|/a_c$.

Inserting Eqs. (65) and (66) in Eq. (46), one finds, for the maximal capacity:

$$\alpha_{c\ (\text{sparse charts})}\sim\frac{1}{a_c\ln(K_d|M|/a_c)}. \tag{67}$$

As we anticipated, one may interpret this result as follows: the capacity is the same as if we had taken a_c^{-1} ''sparse'' charts, including $\sim N$ cells, and put them side by side to form one single ''full'' chart. If we have started with αC ''sparse'' charts we now have $a_c\alpha C$ ''full charts.'' From Eq. (46) we see that we can store at most $\alpha_{c(\text{full charts})}C$ full charts and

$$\alpha_{c(\text{full charts})}\sim\frac{1}{\ln(K_d|M|/a_c)},$$

and this explains Eq. (67). Therefore, this network is *as efficient* in terms of spatial information storage as the one operating with full charts.

IV. INFORMATION STORAGE

Like a pattern autoassociator, the chart autoassociator is an information storing network. The cognitive role of such a module could be to provide a spatial context for information of a nonspatial nature contained in other modules, which connect with the multichart module. Each chart represents a different spatial organization, possibly related to a different environmental/behavioral condition. Within each chart, a cell is bound to a particular position in space, thus being the means for attaching some piece of knowledge to a particular point in space, through intermodule connections. To give a very extreme, unrealistic, but perhaps useful, example, let us assume that each cell encodes a particular discrete item, or the memory of some events happened somewhere in the environment, in ''grandmother cell'' fashion, encoding ''the grandmother sitting in the armchair in the dining room.'' The encoding of the ''grandmother'' may be accomplished by some set of afferents from other modules. The multichart associator can then attach a spatial location to that memory of the ''grandmother.'' The spatial location encoded is ideally represented for each cell by its pfc.

In this sense, the information encoded in the network, which can be extracted by measures of the activity of the units, is the information about the spatial tuning of the units, that is their pfc's.

To restate this concept in a formal way, we look for

$$I_s=\lim_{S\to\infty}\frac{1}{CN}\sum_\mu\sum_i I(\mathbf{r}_i^\mu,\{V_i^{\mu(k)}\}_{k=1\ldots S}), \tag{68}$$

that is, the information per synapse that can be extracted from S different observations of activity of the cells with the animal in S different positions, and the system in activity

states related to chart μ. This quantity does not diverge as $S\to\infty$, since repeated observations of activity with the animal in nearby positions do not yield independent information, because of correlations between activity configurations, correlations that decrease with the distance at which the configurations are sampled.

The full calculation of this quantity involves a functional integration over the distribution of noise affecting cell activity as the animal is moving and exploring the whole environment. In Appendix D we suggest a procedure to approximate this quantity based on an ''information correlation length'' l_I such that samples corresponding to animal positions at a distance l_I yield approximately independent information.

I_s is the amount of spatial information which is stored in the module. It is the exact analog of the stored information for pattern autoassociators [6]. As for storage capacity, it is to be found numerically, by maximization over w and g.

As for the capacity, one can find the solution that maximizes I_s. The resulting $I_{\max}$ is a function of the size of the relative spread of connections $a=1/|M|$, and it amounts to a fraction of bit per synapse (see Fig. 5).

As with pattern autoassociators, the information stored increases with sparser representations. The increase is more marked for the fully connected network. For very sparse representations the performance of the fully connected model approaches the extreme dilution limit.

V. DISCUSSION

We have studied the multichart threshold linear associator as a spatial information encoding and storage module. We have given the solution for the dot-product kernel model, then introduced a formalism in which the generic kernel problem is soluble.

The second treatment has the advantage of providing a form for the average activity peak profile, which can be compared with the experimental data (see, for example, [9], Fig. 1).

We have shown that the nonlinear integral mean-field equation [Eq. (42)] can be solved at least for one class of connection kernels $K(\mathbf{r}-\mathbf{r}')$.

The storage capacity for both models has been found. We note that the capacity for the dot-product model is compatible with the wide kernel (nonsparse) limit of the generic model in one and two dimensions in the fully connected and in the diluted condition.

The generic kernel treatment makes it possible to manipulate the most relevant parameter for storage efficiency, i.e., the spread of connections. It is shown that this parameter plays a very similar role as sparsity for pattern autoassociators. In the multichart case, moreover, the effective sparsity of the *stable configurations* is determined also by the value of the gain parameter g, as shown analytically for the noiseless case. Nevertheless, the capacity of the network depends on the spread of connection parameter $a_c=k_d/|M|$ through a relation which is the exact analog of the relation between sparsity and capacity for the pattern autoassociator, at least in the very sparse limit.

We have only considered here the capacity problem for one form of the connection kernel, although the treatment we propose is applicable, at least, to the other kernels considered for the noiseless case. Our hypothesis is that a similar law for sparsity is to be found as Eq. (47), at least in the high sparsity limit, for more general forms of the kernel.

We have then shown that the capacity scales in such a way that the information stored is not changed when only a fraction of the cells participate in each chart. In this case the firing of a cell carries information not only about the position of its pfc in the chart environment, but also about *which* environment the cell has a place field in. This information adds up, so that $1/a_c$ charts can be assembled in a single larger chart of size $1/a_c$ times larger.

We have introduced a definition of stored information for the multichart memory network, which measures the number of effective different locations which can be discriminated by such a net: representations of places at a distance less than l_I are confused, because of the finite width of the activity peaks, and because of the static noise.

l_I does not vary much when $|M|$ varies. This is consistent with the fact that the storage capacity is well fitted by Eq. (47) with $k_d\sim 4.5$. l_I turns out to be ~ 3.5 for the 1D model, with the arbitrary value for f of 0.95. l_I is therefore similar to the ''radius'' of the activity peak which should correspond to the ''pattern'' in the parallel between the chart autoassociator and the pattern autoassociator.

It was not possible to carry over the calculation of r_{12} and I_2 in the 2D model as it turned out to be too computationally demanding. Therefore, we are not able to show the values of the storable information. The fact that the storage capacity follows Eq. (47) also in this case is an indirect hint of a behavior very similar to what is found in 1D.

APPENDIX A: REPLICA-SYMMETRIC FREE ENERGY FOR THE ''DOT PRODUCT'' KERNEL MODEL

The replica symmetry free energy reads

$$f=-T\left\langle\left\langle \int Dz\ \ln \mathrm{Tr}(h,h_2)\right\rangle\right\rangle-\frac{1}{2}\sum_{(\sigma),l}|\mathbf{x}^{(\sigma),l}|^2-B(x)-\sum_{(\sigma),l}(|\mathbf{s}^{(\sigma),l}|x+\mathbf{s}^{(\sigma),l}x^{(\sigma),l})$$
$$-\sum_{(\sigma),l}\mathbf{t}^{(\sigma),l}\mathbf{x}^{(\sigma),l}-tx-r_0y_0+r_1y_1+\frac{\alpha d}{2\beta}\left(\ln[1-T_0\beta(y_0-y_1)]-\frac{\beta y_1}{1-T_0\beta(y_0-y_1)}\right) \tag{A1}$$

very much like Eq. (19) in 6 and with the same meaning for symbols, except that the population vector $\mathbf{x}^{(\sigma),l}$ plays the role of the overlap x^{σ}, the vector Lagrange multiplier $\mathbf{t}^{(\sigma),l}$ appears instead of its scalar counterpart t^{σ}, and the dimensionality d appears multiplying the last term. h and h_2 are

$$h=-t-\sum_{(\sigma),l}\mathbf{t}^{(\sigma),l}\cdot\boldsymbol{\eta}^{(\sigma),l}-z(-2Tr_1)^{1/2}, \tag{A2}$$

$$h_2=r_1-r_0. \tag{A3}$$

$\langle\langle\ldots\rangle\rangle$ means averaging over the distribution of pfc's $\boldsymbol{\eta}$. T is the noise level in the thermodynamic analysis. T_0 is defined here as

$$\langle\langle(\mathbf{t}_1^l\cdot\boldsymbol{\eta}^{(\mu),l})(\mathbf{t}_2^l\cdot\boldsymbol{\eta}^{(\mu),l})\rangle\rangle=\frac{T_0}{d}\mathbf{t}_1^l\cdot\mathbf{t}_2^l, \tag{A4}$$

and it is found to be equal to 1/2 in 1D and to 1 for the 2D torus.

The saddle-point equations can be found from this equation, and t and $\mathbf{t}^{(\sigma),l}$ can be eliminated, in the same way as in [6]. Carrying on the calculation the $T=0$ equations eventually reduce to two equations in the two variables (in the case of a single ''condensed'' map):

$$w=\frac{[b(x)-\theta]}{\rho}, \tag{A5}$$

$$\mathbf{v}^l=\frac{(\mathbf{x}^l+\mathbf{s}^l)}{\rho}. \tag{A6}$$

Take, for simplicity, $|\mathbf{v}^l|=v$ (while the direction is set by $\mathbf{v}^l\propto\mathbf{s}^l$). The two equations read

$$E_1(w,\mathbf{v})\equiv(A_1+\delta A_2)^2-\alpha A_3=0, \tag{A7}$$

$$E_2(w,\mathbf{v})v\equiv(A_1+\delta A_2)\left(\frac{1}{gT_0(1+\delta)}+\alpha-A_2\right)-\alpha A_2=0, \tag{A8}$$

where $\delta=|\mathbf{s}^1|/|\mathbf{x}^1|$ is the relative importance of the external field and

$$A_1(w,v)=\frac{1}{v^2T_0}\left\langle\left\langle\mathbf{v}^l\cdot\boldsymbol{\eta}^l\int^+Dz\left(w+\sum_l\mathbf{v}^l\cdot\boldsymbol{\eta}^l-z\right)\right\rangle\right\rangle-\left\langle\left\langle\int^+Dz\right\rangle\right\rangle, \tag{A9}$$

$$A_2(w,v)=\frac{1}{v^2T_0}\left\langle\left\langle\mathbf{v}^{(1)}\cdot\boldsymbol{\eta}^{(1)}\int^+Dz\left(w+\sum_l\mathbf{v}^l\cdot\boldsymbol{\eta}^l-z\right)\right\rangle\right\rangle, \tag{A10}$$

$$A_3(w,v)=\left\langle\left\langle\int^+Dz\left(w+\sum_l\mathbf{v}^l\cdot\boldsymbol{\eta}^l-z\right)^2\right\rangle\right\rangle. \tag{A11}$$

Dz is the Gaussian measure $(2\pi)^{-1/2}e^{-z^2/2}dz$. The $+$ sign on the integral means that integration extremes are chosen such that $(w+\Sigma_l\mathbf{v}^l\cdot\boldsymbol{\eta}^l-z)>0$.

When the quenched average on the $\boldsymbol{\eta}$'s is performed, A_1, A_2, A_3 reduce to (for the d-dimensional torus $\mathcal{C}^d$):

$$A_1(w,v)=\frac{1}{(2\pi)^dvT_0}\int d\theta^l\left(\sum_l\cos\theta^l\right)\times\left[\left(w+v\sum_l\cos\theta^l-vT_0\right)\Phi\left(w+v\sum_l\cos\theta^l\right)+\left(w+v\sum_l\cos\theta^l\right)\sigma\left(w+v\sum_l\cos\theta^l\right)\right], \tag{A12}$$

$$A_2(w,v)=\frac{1}{(2\pi)^dvT_0}\int d\theta^l\left(\sum_l\cos\theta^l\right)\times\left[\left(w+v\sum_l\cos\theta^l\right)\Phi\left(w+v\sum_l\cos\theta^l\right)+\left(w+v\sum_l\cos\theta^l\right)\sigma\left(w+v\sum_l\cos\theta^l\right)\right], \tag{A13}$$

$$A_3(w,v)=\frac{1}{(2\pi)^d}\int d\theta^l\left[1+\left(w+v\sum_l\cos\theta^l\right)^2\right]\times\Phi\left(w+v\sum_l\cos\theta^l\right)+\left(w+v\sum_l\cos\theta^l\right)\times\sigma\left(w+v\sum_l\cos\theta^l\right), \tag{A14}$$

where

$$\Phi(x)=\int_{-\infty}^{x}\frac{dz}{\sqrt{2\pi}}e^{-z^2/2}, \tag{A15}$$

$$\sigma(x)=\frac{e^{-x^2/2}}{\sqrt{2\pi}}. \tag{A16}$$

APPENDIX B: GENERIC KERNEL, EXTREME DILUTION

Let us consider the one-dimensional case first, and consider the kernel

$$K(\mathbf{r}-\mathbf{r}')=\hat{K}(\mathbf{r}-\mathbf{r}')-\frac{2}{|M|}=e^{-|\mathbf{r}-\mathbf{r}'|}-\frac{2}{|M|}. \tag{B1}$$

Equation (42) can be written

$$v(\mathbf{r})=g\mathcal{N}\left(\int d\mathbf{r}'\hat{K}(\mathbf{r}-\mathbf{r}')v(\mathbf{r}')+\hat{w}\right), \tag{B2}$$

where

$$\hat{w}=w-\frac{2}{|M|}\int d\mathbf{r}'v(\mathbf{r}'). \tag{B3}$$

For the purpose of finding α_c, maximizing with respect to $\hat{w}$ is equivalent to maximizing with respect to w.

To solve Eq. (42), the transformation

$$u(\mathbf{r})=\mathcal{N}^{-1}\left(\frac{v(\mathbf{r})}{g}\right) \tag{B4}$$

is used, which results in

$$u(\mathbf{r})=g\int d\mathbf{r}'\hat{K}(\mathbf{r}-\mathbf{r}')\mathcal{N}[u(\mathbf{r})]+\hat{w}. \tag{B5}$$

By differentiating twice we get

$$u''(\mathbf{r})=-2g\mathcal{N}[u(\mathbf{r})]+u(\mathbf{r})-\hat{w}=-\frac{d}{du}\mathcal{U}[u(\mathbf{r})], \tag{B6}$$

where

$$\mathcal{U}=\int^{u}du'(2g\mathcal{N}(u')-u'+\hat{w}). \tag{B7}$$

The differential equation (B6) is *locally* equivalent to the nonlinear integral equation (B5). Equation (B6) must be solved numerically. As in the single-map case, not all the solutions of the differential equation (B6) are a solution of the integral equation (B5). Solutions of Eq. (B6) are a solution of Eq. (B5), strictly speaking, only in the case $M\equiv\mathbb{R}$. Nevertheless, we force the equivalence since, also in the case of limited environments, with periodic boundary conditions, possible pathologies are not important for solutions with activity concentrated far from the boundaries.

In order to classify the solutions of Eq. (B6) it is useful to study the ''potential function'' $\mathcal{U}$. If w is negative and large enough in absolute value, $\mathcal{U}(u)$ has a maximum and a minimum at the two roots of equation

$$\frac{d}{du}\mathcal{U}(u)=2g\mathcal{N}(u)-u+\hat{w}=0, \tag{B8}$$

or, in terms of v

$$v=g\mathcal{N}(2v+\hat{w}), \tag{B9}$$

corresponding to constant solutions of Eq. (42). We look for solutions representing a single, symmetric peak of activity centered in $r=0$. We therefore need to solve the Cauchy problem given by Eq. (B6) with the initial conditions:

$$u(0)=u_0, \tag{B10}$$

$$u'(0)=0. \tag{B11}$$

From Fig. 6 it is clear that if $u_0>u^*$ the solution will escape to $-\infty$ for r tending to infinity. This will correspond to v tending asymptotically to 0, and this solution cannot be a solution for the integral equation (42) as the asymptotic value must be a root of Eq. (B9).

The solutions of the problem with $u_0<u^*$ are periodic, corresponding to multiple peaks of activity, and they are discarded as unstable with the same arguments holding for the single map case. There is also the constant solution

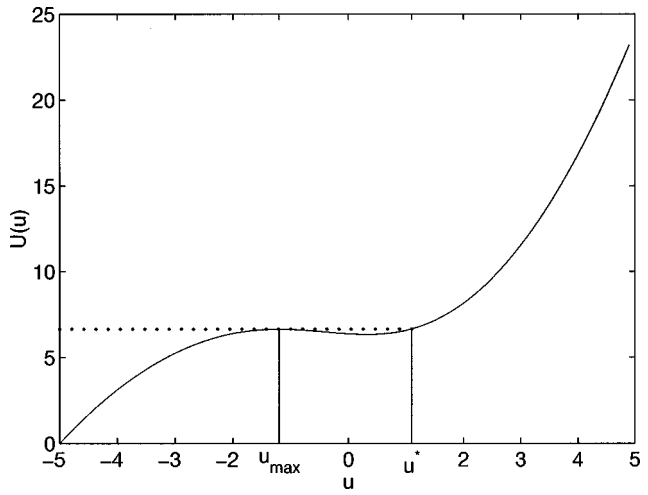

FIG. 6. The ''potential'' function $\mathcal{U}(u)$ defined by Eq. (B7) and entering the differential equation Eq. (B6). Solutions with $u'(0)=0$ and $u(0)=u_0$, with $u_{\max}<u_0<u^*$ are oscillating. The solution with $u_0=u^*$ is the one we seek, asymptotically approaching $u_{\max}$ as $r\to\infty$.

$$u(r)=u_{\min}, \tag{B12}$$

which obviously will not correspond to space-related activity. The solution corresponding to the single activity peak can only be the one with $u_0=u^*$. It tends asymptotically to $u_{\max}$. This solution can be found numerically and inserted in Eq. (46) to find the value of α associated with the pair $(g,\hat{w})$. The solution will only be present for values of $\hat{w}$ for which $\mathcal{U}(u)$ has the extremal points $u_{\max}$ and $u_{\min}$, that is

$$\hat{w}<\hat{w}^*, \tag{B13}$$

where $\hat{w}^*$ is equal to $-2g\mathcal{N}(u^*)+u^*$ and u^* is the root of the equation:

$$\Phi(u)=\frac{1}{2g} \tag{B14}$$

obtained by derivating twice $\mathcal{U}$, and this shows that Eq. (B5) cannot have solutions for $g<1/2$, as in the single-map case.

In the two-dimensional case, we can consider the kernel

$$K(\mathbf{r}-\mathbf{r}')=\hat{K}(\mathbf{r}-\mathbf{r}')-\frac{2}{|M|}, \tag{B15}$$

where $\hat{K}$ is the kernel having the Fourier transform

$$\hat{K}(\mathbf{p})=\frac{2}{1+\mathbf{p}^2}. \tag{B16}$$

The solution is worked out in the same way with the transformation (B4) and application of the Laplacian. If we con-

sider solutions with circular symmetry and pass to polar coordinates (r,ϕ), the equation for r reads

$$u''(r)+\frac{1}{r}u'(r)=-2g\mathcal{N}(u(r))+u(r)-w. \quad \text{(B17)}$$

We still have a single peak solution which tends asymptotically to $u_{\max}$, but in this case we cannot rely on the $\mathcal{U}$ function argument to find the initial condition at $r=0$, which has to be found numerically.

APPENDIX C: REPLICA FREE-ENERGY CALCULATION FOR THE GENERIC KERNEL

Again we will consider an environment M with periodic boundary conditions. We assume that there exists a kernel L such that

$$\int d\mathbf{r}''L(\mathbf{r}-\mathbf{r}'')L(\mathbf{r}''-\mathbf{r}')=|M|K(\mathbf{r}-\mathbf{r}'). \quad \text{(C1)}$$

Instead of the vector order parameter $\mathbf{x}^\mu$ that we used for the dot-product kernel case (or of the scalar overlap x^μ of [6]), we can use the functional order parameter

$$x^\mu(\vec{r})=\frac{1}{N}\sum_i [L(\mathbf{r}-\mathbf{r}_i^\mu)]V_i \quad \text{(C2)}$$

in terms of which the interaction part of the Hamiltonian (59) reads

$$\begin{aligned}&\frac{1}{2}\sum_i\sum_{j\neq i}J_{ij}V_iV_j\\&=\frac{|M|}{2N}\sum_\mu\sum_i\sum_j[K(\mathbf{r}_i^\mu-\mathbf{r}_j^\mu)-\overline{K}]V_iV_j\\&\quad-\frac{\alpha|M|}{2}[K(0)-\overline{K}]\sum_i V_i^2\\&=\frac{1}{2}N\sum_\mu\int d\mathbf{r}[x^\mu(\vec{r})]^2-\frac{\alpha|M|}{2}[K(0)-\overline{K}]\sum_i V_i^2.\end{aligned} \quad \text{(C3)}$$

Introducing the ''square root'' kernel L allows us to perform the standard Gaussian transformation manipulation and to carry out the mean-field free-energy calculation in the replica-symmetry approximation:

$$\begin{aligned}f=&-T\left\langle\left\langle\int Dz\ \ln \mathrm{Tr}(h,h_2)\right\rangle\right\rangle-\frac{1}{2}\sum_\sigma\int_M d\mathbf{r}[x^\sigma(\mathbf{r})]^2-\frac{\alpha|M|}{2}[K(0)-\overline{K}]y_0+B(x)\\&-\sum_\sigma\int_M d\mathbf{r}\ \ t^\sigma(\mathbf{r})x^\sigma(\mathbf{r})-tx-r_0y_0+r_1y_1+\frac{\alpha}{2\beta}\sum_{\mathbf{p}}\left(\ln[1-T_0(\mathbf{p})\beta(y_0-y_1)]-\frac{\beta y_1}{1-T_0(\mathbf{p})\beta(y_0-y_1)}\right),\end{aligned} \quad \text{(C4)}$$

where now $T_0(\mathbf{p})$ is the Fourier transform of the kernel $|M|\hat{K}$

$$T_0(\mathbf{p})=|M|\int_M d\mathbf{r}\ e^{-i\mathbf{p}\mathbf{r}}K(\mathbf{r}). \quad \text{(C5)}$$

We now have

$$\begin{aligned}h=&b(x)+\sum_\sigma\int_M d\mathbf{r}'x^\sigma(\mathbf{r}')[L(\mathbf{r}^\sigma-\mathbf{r}')-\overline{L}]\\&-z(-2tr_1)^{1/2},\end{aligned} \quad \text{(C6)}$$

$$h_2=-r_0+r_1. \quad \text{(C7)}$$

The $T=0$ mean-field equations are much like in [6] apart from the $x^\sigma(\mathbf{r})$ equation which reads

$$\begin{aligned}x^\sigma(\mathbf{r})=g'\Bigg\langle\Bigg\langle&[L(\mathbf{r}^\sigma-\mathbf{r}')-\overline{L}]\\&\times\int^+Dz\Bigg\{\int_M d\mathbf{r}'[L(\mathbf{r}^\sigma-\mathbf{r}')-\overline{L}]x^\sigma(\mathbf{r}')\\&+b(x)-\theta-\rho z\Bigg\}\Bigg\rangle\Bigg\rangle,\end{aligned} \quad \text{(C8)}$$

where now the + sign on the integral means that the limits of integration over z are chosen such that

$$\int d\mathbf{r}'[L(\mathbf{r}^\sigma-\mathbf{r}')-\overline{L}]x^\sigma(\mathbf{r}')+b(x)-\theta>0. \quad \text{(C9)}$$

g' is a renormalized gain, which takes into account the effect of static noise, defined by

$$(g')^{-1}=g^{-1}-\alpha\sum_{\mathbf{p}}T_0(\mathbf{p})\frac{\overline{\Psi}}{1-T_0(\mathbf{p})\overline{\Psi}}, \quad \text{(C10)}$$

where $\overline{\Psi}$ is given by Eq. (C18).

The noise variance ρ^2 is given by

$$\rho^2=-2Tr_1=\alpha\sum_{\mathbf{p}}\frac{[T_0(\mathbf{p})]^2y_0}{[1-T_0(\mathbf{p})\overline{\Psi}]^2}, \quad \text{(C11)}$$

where

$$\begin{aligned}y_0=(g')^2\Bigg\langle\Bigg\langle&\int^+Dz\Bigg\{\int_M d\mathbf{r}'[L(\mathbf{r}^\sigma-\mathbf{r}')-\overline{L}]\\&\times x^\sigma(\mathbf{r}')+b(x)-\theta\Bigg\}^2\Bigg\rangle\Bigg\rangle\end{aligned} \quad \text{(C12)}$$

and

$$\overline{\Psi} = g' \left\langle \left\langle \int^{+} Dz \right\rangle \right\rangle . \tag{C13}$$

We now pass to the rescaled variables

$$v^{\sigma}(\mathbf{r}) = \frac{x^{\sigma}(\mathbf{r})}{\rho}, \tag{C14}$$

$$w = \frac{b(x)}{\rho}, \tag{C15}$$

obtaining

$$v^{\sigma}(\mathbf{r}) = g' \int_M d\mathbf{r}^{\sigma} [L(\mathbf{r}^{\sigma} - \mathbf{r}) - \overline{L}] \times \left(w + \int_M d\mathbf{r}' [L(\mathbf{r}^{\sigma} - \mathbf{r}') - \overline{L}] v^{\sigma}(\mathbf{r}) \right), \tag{C16}$$

$$\frac{y_0}{\rho^2} = (g')^2 \int_M \frac{d\mathbf{r}^{\sigma}}{|M|} \times \mathcal{M} \left(w + \int_M d\mathbf{r}' [L(\mathbf{r}^{\sigma} - \mathbf{r}') - \overline{L}] v^{\sigma}(\mathbf{r}) \right), \tag{C17}$$

$$\overline{\Psi} = \int_M \frac{d\mathbf{r}^{\sigma}}{|M|} \Phi \left(w + \int_M d\mathbf{r}' [L(\mathbf{r}^{\sigma} - \mathbf{r}') - \overline{L}] v^{\sigma}(\mathbf{r}) \right). \tag{C18}$$

APPENDIX D: GENERIC KERNEL: STORABLE INFORMATION CALCULATION

First, the information per synapse we get from a single observation of activity, with the animal in a certain position times the number of stored charts is

$$I_1 = \alpha \int \frac{d\mathbf{r}}{|M|} \left\{ \int_{-\infty}^{u(\mathbf{r})} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \times \ln \left(\frac{e^{-z^2/2}}{\int (d\mathbf{r}'/|M|)\, e^{-[z - u(\mathbf{r}) + u(\mathbf{r}')^2]/2}} \right) + [1 - \phi(u(\mathbf{r}))] \times \ln \left(\frac{[1 - \phi(u(\mathbf{r}))]}{\int (d\mathbf{r}'/|M|)\, [1 - \phi(u(\mathbf{r}'))]} \right) \right\}. \tag{D1}$$

Next, we wish to calculate the joint information from two measures of activity, from the same cells, from all charts, while the rat is in two different locations, at a distance ϵ. These two measures are correlated random variables: let

$$V_1 = [h_1 - \rho z_1]^{+}$$

be the activity of a cell measured while the rat is in position 1, and

$$V_2 = [h_2 - \rho z_2]^{+}$$

be the activity of the same cell while the rat is in position 2.

The two noise variables are distributed according to a joint bivariate Gaussian distribution:

$$p(z_1, z_2) = \frac{1}{2\pi \sqrt{1 - r_{12}^2}} \times \exp \left(- \frac{1}{2(1 - r_{12}^2)} (z_1^2 + z_2^2 - 2 r_{12} z_1 z_2) \right). \tag{D2}$$

The correlation coefficient r_{12} is a function of the distance ϵ, implicitly defined through the equation

$$\rho^2 r_{12}(\epsilon) = \alpha |M| \langle \langle K^2 \rangle \rangle y_{12}(\epsilon), \tag{D3}$$

where y_{12} is defined as

$$y_{12}(\epsilon) = \frac{1}{N} \sum_i \langle V_{i,1} V_{i,2} \rangle \tag{D4}$$

and assuming periodic boundary conditions:

$$y_{12}(\epsilon) = \rho^2 g^2 \int \frac{d\mathbf{r}}{|M|} \int^{++} Dz_{12} \tag{D5}$$

$$\times \left(\int \frac{dr'}{|M|} K(\mathbf{r} - \mathbf{r}') v(\mathbf{r}') + w - z_1 \right) \times \left(\int \frac{dr''}{|M|} K(\mathbf{r} - \mathbf{r}'') v(\mathbf{r}'' + \epsilon) + w - z_2 \right) \tag{D6}$$

or

$$y_{12}(\epsilon) = \rho^2 g^2 \int \frac{d\mathbf{r}}{|M|} \int^{++} Dz_{12} u(\mathbf{r}) u(\mathbf{r} + \epsilon), \tag{D7}$$

where $u(r)$ is defined by Eq. (B4). The integration measure for the noise variable is defined as

$$\int^{++} Dz_{12} = \int_{u(\mathbf{r}) - z_1 > 0, u(\mathbf{r} + \epsilon) - z_2 > 0} dz_1 dz_2 p(z_1, z_2). \tag{D8}$$

Inserting Eq. (D7) in Eq. (D3) yields

$$r_{12} = \alpha |M| \langle \langle K^2 \rangle \rangle g^2 \int dr\ \mathcal{Q}[u(\mathbf{r}), u(\mathbf{r} + \epsilon), r_{12}], \tag{D9}$$

where

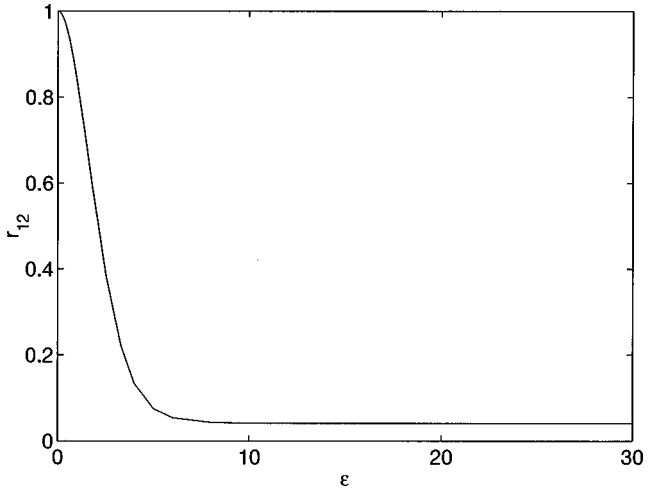

FIG. 7. The r_{12} function plotted as a function of the distance between the two pfc's, ϵ in the $|M|=30$ case.

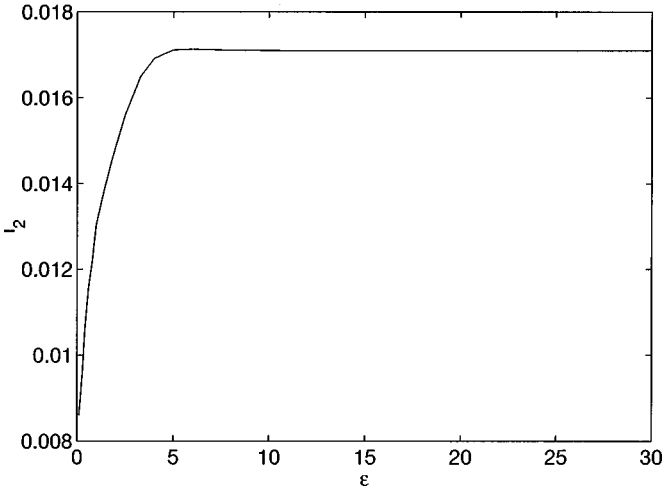

FIG. 8. The I_2 function plotted as a function of the distance between the two pfc's, ϵ in the $|M|=30$ case. Note that l_I, with $f=0.95$ [see Eq. (D17)] would be approximately 3.5. This is seen not to change much when $|M|$ varies (not shown).

$$\mathcal{Q}(x,y,r_{12})=\int^{++} Dz_{12}(x-z_1)(y-z_2).$$

Equation (D9) can be solved numerically. An example is provided in Fig. 7, but a few features can be explored analytically, in the neighborhood of $\epsilon=0$. $r_{12}=1,\epsilon=0$ is a solution, but now consider what happens when ϵ increases.

The derivatives of

$$\mathcal{D}(r_{12},\epsilon)=\alpha\langle\langle K^2\rangle\rangle g^2\int\frac{dr}{|M|}\mathcal{Q}[u(\mathbf{r}),u(\mathbf{r}+\boldsymbol{\epsilon}),r_{12}]-r_{12} \tag{D10}$$

with respect to ϵ and r_{12} must be taken into consideration. One has

$$\frac{\partial}{\partial\boldsymbol{\epsilon}}\mathcal{D}(r_{12}=1,\boldsymbol{\epsilon}=0)$$
$$=\alpha\langle\langle K^2\rangle\rangle g^2$$
$$\times\int\frac{dr}{|M|}\frac{\partial}{\partial y}\mathcal{Q}[x=u(\mathbf{r}),y=u(\mathbf{r}+\boldsymbol{\epsilon}),1]u'(\mathbf{r})=0, \tag{D11}$$

$$\frac{\partial^2}{\partial\boldsymbol{\epsilon}^2}\mathcal{D}(r_{12}=1,\boldsymbol{\epsilon}=0)<0, \tag{D12}$$

and

$$\frac{\partial}{\partial r_{12}}\mathcal{D}(r_{12}\to 1,\boldsymbol{\epsilon}=0)$$
$$=\alpha\langle\langle K^2\rangle\rangle g^2\int\frac{d\mathbf{r}}{|M|}\Phi(u(\mathbf{r}))-1$$
$$=\int\frac{dr}{|M|}\Phi(u(\mathbf{r}))\left(\int\frac{dr}{|M|}\mathcal{M}(u(\mathbf{r}))\right)^{-1}-1. \tag{D13}$$

From Eq. (D12) it turns out that when the derivative in Eq. (D13) is greater than zero, the solution $r_{12}=1$ disappears as one moves from $\epsilon=1$, but another solution is still present so that

$$\lim_{\epsilon\to 0^+} r_{12}(\epsilon)<1. \tag{D14}$$

Note that the condition

$$\frac{\partial}{\partial r_{12}}\mathcal{D}(r_{12}\to 1,\boldsymbol{\epsilon}=0)>0$$

is equivalent to

$$\Gamma=g\alpha|M|\langle\langle K^2\rangle\rangle\int_M dr\,u(r)v(r)<0, \tag{D15}$$

and the quantity Γ enters in the stability analysis considerations we sketched in Sec. III C, at least for the 1D case. Solutions with $\Gamma>0$ are stable against inhibition orthogonal fluctuations, so that it is likely that the possible pathology implied by Eq. (D14) reflects an instability of the solution. We have always found numerically that for the solution corresponding to the maximal storage capacity and information, $\Gamma>0$.

Once we know the joint probability distribution for z_1 and z_2, we can calculate the information we can extract about the pfc of a cell from two measurements of activity, while the rat is standing in two positions at a distance ϵ, from all charts:

$$
\begin{aligned}
I_2(\epsilon)=\frac{\alpha}{\ln 2}\int_M \frac{d\mathbf{r}}{|M|}\Bigg\{ & \int^{++} Dz_{12}\Bigg[-\frac{1}{2(1-r_{12}^2)}(z_1^2+z_2^2-2r_{12}z_1z_2) \\
& -\ln\Bigg(\int_M \frac{d\mathbf{r}'}{|M|}\exp\Bigg(-\frac{1}{2(1-r_{12}^2)}[(u(\mathbf{r}')-u(\mathbf{r})+z_1)^2+(u(\mathbf{r}'+\boldsymbol{\epsilon})-u(\mathbf{r}+\boldsymbol{\epsilon})+z_2)^2 \\
& -2r_{12}(u(\mathbf{r}')-u(\mathbf{r})+z_1)(u(\mathbf{r}'+\boldsymbol{\epsilon})-u(\mathbf{r}+\boldsymbol{\epsilon})+z_2)]\Bigg)\Bigg)\Bigg] \\
& +2\int^{+-} Dz_{12}\Bigg[\ln\Bigg(\int_{z_2'>u(\mathbf{r}+\boldsymbol{\epsilon})} \frac{dz_2'}{2\pi}\exp\Bigg(-\frac{1}{2(1-r_{12}^2)}(z_1^2+z_2'^2-2r_{12}z_1z_2')\Bigg)\Bigg) \\
& -\ln\Bigg(\int_M \frac{d\mathbf{r}'}{|M|}\int_{z_2'>u(\mathbf{r}'+\boldsymbol{\epsilon})} dz_2'\exp\Bigg(-\frac{1}{2(1-r_{12}^2)}[(u(\mathbf{r}')-u(\mathbf{r})+z_1)^2+(u(\mathbf{r}'+\boldsymbol{\epsilon})-u(\mathbf{r}+\boldsymbol{\epsilon})+z_2')^2 \\
& -2r_{12}(u(\mathbf{r}')-u(\mathbf{r})+z_1)(u(\mathbf{r}'+\boldsymbol{\epsilon})-u(\mathbf{r}+\boldsymbol{\epsilon})+z_2')]\Bigg)\Bigg)\Bigg] \\
& +\int^{--} Dz_{12}\Bigg[\ln\Bigg(\int_{z_1'>u(\mathbf{r}),z_2'>u(\mathbf{r}+\boldsymbol{\epsilon})} \frac{dz_1'dz_2'}{2\pi\sqrt{1-r_{12}^2}}\exp\Bigg(-\frac{1}{2(1-r_{12}^2)}(z_1'^2+z_2'^2-2r_{12}z_1'z_2')\Bigg)\Bigg) \\
& -\ln\Bigg(\int_M \frac{d\mathbf{r}'}{|M|}\int_{z_1'>u(\mathbf{r}'),z_2'>u(\mathbf{r}'+\boldsymbol{\epsilon})} \frac{dz_1'dz_2'}{2\pi\sqrt{1-r_{12}^2}}\exp\Bigg(-\frac{1}{2(1-r_{12}^2)}[(u(\mathbf{r}')-u(\mathbf{r})+z_1')^2+(u(\mathbf{r}'+\boldsymbol{\epsilon})-u(\mathbf{r}+\boldsymbol{\epsilon})+z_2')^2 \\
& -2r_{12}(u(\mathbf{r}')-u(\mathbf{r})+z_1')(u(\mathbf{r}'+\boldsymbol{\epsilon})-u(\mathbf{r}+\boldsymbol{\epsilon})+z_2')]\Bigg)\Bigg)\Bigg]\Bigg\}. \qquad \text{(D16)}
\end{aligned}
$$

The minus signs $(-)$ beside the integration signs mean that, respectively, the first, or the second condition determining the integration intervals in Eq. (D8) is reversed. The first term in the sum accounts for the contribution coming from measurement in which both activity values are positive. The second term is the contribution from measurements in which one value is zero and the other is positive. The third term comes from measurements in which both values are zero. For $\epsilon=0$, $I_2=I_1$, since the two measures are identical.

For large ϵ one has $I_2\sim 2I_1$, because the noise decorrelates and because in general the two measures will give nonzero results in distinct regions of the environment. The behavior of I_2 as a function of ϵ is exemplified in Fig. 8. We define as ''information correlation length'' the value l_I of ϵ for which

$$I_2-I_1=fI_1, \qquad \text{(D17)}$$

where f is a fixed fraction, say 0.95. We may say that measurements of activity with the rat in two positions at a distance l_I give independent information.

This allows us to define as the *stored information* I_s the quantity

$$I_s=I_1\frac{|M|}{l_I^d}, \qquad \text{(D18)}$$

that is, sampling the activity of a cell $|M|/l_I^d$ times, with the animal spanning a lattice with size l_I, we may effectively add up the information amounts we get from each single sample, as if they were independent.

[1] D. Amit, *Modeling Brain Function* (Cambridge University Press, New York, 1989).

[2] M. Tsodyks and T. Sejnowski, in Proceedings of the Third Workshop on Neural Networks: from Biology to High Energy Physics [Int. J. Neural Syst. **6** (Suppl.), 81 (1994)].

[3] R. Muller, J. Ranck, and J. Taube, Curr. Opin. Neurobiol. **6**, 196 (1996).

[4] D. Touretzky and A. Redish, Hippocampus **6**, 247 (1996).

[5] K. Gothard, W. Skaggs, and B. McNaughton, J. Neurosci. **16**, 8027 (1996).

[6] A. Treves, Phys. Rev. A **42**, 2418 (1990).

[7] B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987).

[8] A. Treves, J. Phys. A **24**, 327 (1991).

[9] A. Samsonovich and B. McNaughton, J. Neurosci. **17**, 5900 (1997).

[10] A. Georgopoulos, R. Kettner, and A. Schwartz, J. Neurosci. **8**, 2928 (1988).

[11] M. Wilson and B. McNaughton, Science **261**, 1055 (1993).