Localized activity profiles and storage capacity of rate-based autoassociative networks

Yasser Roudi[1,*] and Alessandro Treves[1,2]
[1]*Scuola Internazionale Superiore di Studi Avanzati (SISSA), Settore di Neuroscienze Cognitive, Trieste, Italy*
[2]*NTNU, Centre for the Biology of Memory, Trondheim, Norway*

(Received 13 May 2005; revised manuscript received 21 February 2006; published 5 June 2006)

We study analytically the effect of metrically structured connectivity on the behavior of autoassociative networks. The steady state equations are derived for a generic input-output function, and then we focus on their solutions in the case of networks composed of three alternative simple rate-based model neurons: threshold-linear, binary or smoothly saturating units. For a connectivity which is short range enough the threshold-linear network shows localized retrieval states. The saturating and binary models also exhibit spatially modulated retrieval states if the highest activity level that they can achieve is above the maximum activity of the units in the stored patterns. We show that this saturation level together with the linear gain of the transfer function are important parameters that determine the possibility of localized retrieval. If the ratio of the number of stored patterns to the number of connections per unit goes to zero, while the latter goes to infinity, it is possible to derive an analytical formula for the critical value of the connectivity width, below which one observes spatially nonuniform retrieval states. The formula is also shown to offer a good first approximation for higher storage loads. We show that even in the case of localized retrieval the storage capacity remains proportional to the number of connections per neurons, with the proportionality constant lower by a factor of 3–4 compared to uniform retrieval. The approach that we present here is generic in the sense that there are no specific assumptions on the single unit input-output function nor on the exact connectivity structure.

DOI: 10.1103/PhysRevE.73.061904 PACS number(s): 87.19.La, 87.18.Sn, 87.18.Hf

I. INTRODUCTION

Recurrent neuronal networks are able to store patterns of activity and retrieve them later when provided with partial cues—a property called autoassociative retrieval. This is believed to play an important role in memory, as a function of the real brain. During the past 20 years, several neuronal network models of autoassociative retrieval have been studied along the lines of the seminal work by Amit *et al.* [1]. An important feature of this approach is that interesting quantities like storage capacity can actually be calculated, using methods of statistical physics. There have been various extension of the model studied by Amit *et al.*, pointing in the direction of more biologically plausible autoassociative networks. For instance, associative memory retrieval has been extensively analyzed in networks with analog input-output transfer functions [2,3] and spiking neuron models [4–7].

Nevertheless, most models of associative retrieval rely on very simplified assumptions about the pattern of connectivity in the network. It is usually assumed, in fact, that the probability of existence of a connection between two units does not depend on their distance. This is a reasonable assumption for modelling memory retrieval in parts of the brain—like the CA3 field of the hippocampus—where the connectivity between neurons is, to a first approximation, uniform. The cerebral isocortex, on the other hand, while also thought to retrieve memories autoassociatively [8], shows substantially metrical organization in its connectivity [9]. One study in rat visual cortex, for instance, suggests that the probability of connection decreases from 50%-80% for directly adjacent neurons to 0%–15% for neurons 500 μm apart [10].

The analytical treatment of even very simplified models with spatially organized connectivity is however difficult [11]. First, the distance dependence in the connectivity forces one to introduce "field" order parameters in the model [12]. Moreover, asymmetric connectivity makes inapplicable those methods of equilibrium statistical mechanics which were originally used to solve classical models of associative retrieval [13]. Therefore, associative networks with metrically structured connectivity have been recently studied only through simulations, considering special cases, e.g., of binary units, or with some approximation [12,14–16]. Even though the effect of a metric connectivity in autoassociative networks is just starting to be seriously considered, there is a large literature on the effect of neuronal connectivity patterns in networks *without* quenched memory patterns. The localization of activity, for instance, has been extensively analyzed in ring models for orientation selectivity [17], head direction cells [18,19] and spatial working memory [20]. The effect of geometrically structured connectivity and transmission delays on the spatiotemporal properties of the activity of large neuronal networks has been recently studied, too [21]. However, as we said, such models do not include stored memory patterns. In the ring models, for instance, all the relevant information carried by neuronal activity is that related to its location on the ring. Here, instead, we consider a network with metrically organized connectivity *and* stored memory patterns, and we study the interplay between these two features.

In this context, we have previously derived the steady state equations of an autoassociative network comprised of threshold-linear model neurons [12]. We have shown that in such a network the spatial organization of the connectivity can modulate the attractor states that correspond to stored memory patterns. The interesting scenario arises when the

[*] Present address: Gatsby Computational Neuroscience Unit, University College London, London, UK.

connectivity is such that a sufficiently large number of connections to each unit come from nearby units. Then, even though memory patterns are stored all over the network and none of them has any preferential spatial location, retrieval activity can be localized in a certain region, while still being pattern selective, that is it can convey both spatial information and information about the memory pattern that has been evoked by the external cue. This analysis has been based on approximating the solutions of the steady state equations in the case of threshold-linear neurons, together with computer simulations.

The purpose of this paper is to extend the previous analysis to generic rate-based neuronal networks, describe how they can generally be solved and analyze their solutions in the case of different neuronal models in order to figure out the effect of single neuron parameters on the system. We also confirm our previous results through a more accurate method presented here. This will provide us with a more accurate picture of the system.

We derive the steady state equations for a rate-based network with spatially organized connectivity. We show that the equations can be analyzed analytically if the ratio of the number of stored patterns to the number of connections per unit goes to zero, as the latter goes to infinity. When this ratio is finite, we provide a numerical method to solve the fixed-point equations to arbitrary accuracy. We analyze the equations for three alternative model neurons: binary, threshold-linear, and smoothly saturating. Including various neuronal types is an important step forward for two reasons. The first reason is that it has been argued that it is not possible to get the same kind of localized retrieval states in a network of binary neurons [15,22] and integrate-and-fire neurons [14]. The result that we derive here is that it is important to set the linear gain and level of saturation of the neurons in the right regime to get localized memory states. This explains the failure in getting localized retrieval states in other studies and makes the picture drawn from different studies more coherent. Second, since real neurons do show firing rate saturation, it is important so see whether localized retrieval states can be observed in network comprised of such saturating model neurons and to understand how it affects the localization of the activity.

In the case of threshold-linear networks, numerical solutions of the equations show that, as a result of short range connectivity, localized retrieval states can be observed. This confirms our previous analysis in Ref. [12]. These localized solutions are absent in a network with 0–1 binary units when the maximum activity level in memory patterns is chosen to be 1. This is consistent with the results of [15,22]. However, for binary and smoothly saturating units, but with a maximum activity level above their maximum rates in the stored patterns, the existence of spatially modulated retrieval states depends on this saturation level itself and, for smoothly saturating units, on how it is approached, that is, on the gain of their transfer function.

Numerical solutions show that the level of quenched noise has a minor effect on the properties of the spatially modulated solutions; and suggest, therefore, that our analytical solution in the case of low storage load is a good first order approximation.

II. MODEL

Consider a network of N neurons, in which the firing rate of unit i is represented by a variable $v_i \geqslant 0$. The activity of each neuron is determined through a transfer function $v_i = F(h_i - \mathrm{Th})$ where h_i is the input to the neuron and Th is a threshold such that $F(h_i - \mathrm{Th}) = 0$ for $h_i < \mathrm{Th}$, i.e., subthreshold inputs elicit no output.

We further assume that the input (local field) to the unit i takes the following form:

$$h_i = \sum_{j \neq i} J_{ij} v_j + b(x), \tag{1}$$

where the first term enables the memories encoded in the weights to determine the dynamics. In the second term, $x = \frac{1}{N}\Sigma_i v_i$ and the function denoted by $b(x)$, with an appropriate functional form, can regulate the activity of the network, so that at any moment in time $\frac{1}{N}\Sigma_i v_i = \mathrm{const}$. While in simulations $b(x)$ will be assigned an explicit functional form, such exact form is not important for our study of the fixed-point equations. The reason is that the contribution from $b(x)$ is constant across units, when the fixed-point equations are satisfied and $\frac{1}{N}\Sigma_i v_i = \mathrm{const}$, and it can therefore be absorbed in the threshold. This effective threshold is used by the network to regulate its activity.

We assume that each unit receives C inputs from the other units in the network. We take the limits $C \to \infty$ and $N \to \infty$ so that finite size effects are not important in our analysis.

The Hebbian learning rule prescribes that the synaptic weight between units i and j be given as [23,24]:

$$J_{ij} = \frac{1}{Ca^2} \sum_{\mu=1}^{p} \varpi_{ij} (\eta_i^\mu - a)(\eta_j^\mu - a), \tag{2}$$

where p is the number of stored patterns, η_i^μ represents the activity of the unit located at i in memory pattern μ and $\varpi_{ij} = 1$, with probability $\wp_{ij}$, if there is a connection between units i and j, and $\varpi_{ij} = 0$ otherwise.

Each η_i^μ is taken to be a *quenched variable* drawn independently from a distribution $p(\eta)$, with the constraints $\eta \geqslant 0$, $\langle \eta \rangle = \langle \eta^2 \rangle = a$, where $\langle\,\rangle$ stands for the average over the distribution $p(\eta)$ [25]. The parameter a measures the *sparsity* of the memory patterns: they are sparsely coded if $a \ll 1$. Here we concentrate on the binary coding scheme $p(\eta) = a\delta(\eta - 1) + (1 - a)\delta(\eta)$, but the calculation can be easily applied to any probability distribution.

Throughout this paper we assume that the b term in Eq. (1), or equivalently Th, is chosen in such a way that $\frac{1}{N}\Sigma_i v_i = a$ at all times, i.e., the activity in the network is regulated at the same level as in the memory patterns (note, however, that for the memory patterns also the average of the square activity is set to a, which turns a into a sparsity parameter). Fixing the activity guarantees, among other things, that it will not blow up during the retrieval operation. The mean activity level can still be set to a different constant value, but we find it convenient to stick to the simple ansatz that the network maintains the same mean activity at retrieval as it had at storage, that is, the same mean activity level as in the stored memory patterns. Our conclusions require a straightforward rephrasing to be applied to a network operating at a different activity level during retrieval. Further, we assume, as mentioned above, that inhibition—that we do not model explicitly—may effectively act to keep the *overall* mean activity constant. Even though in most studies of autoassociative networks inhibition has been assumed to act globally (in parallel to the assumption of a nonmetric excitatory connectivity), it will be quite interesting to study how short range inhibition would change our view of associative retrieval in the cortex. This issue remains to be analyzed: for the purpose of the current study we do not worry about it, and just assume that the various classes of inhibitory neurons in the cortex can, one way or another, ensure constant overall activity at all times.

III. FIXED-POINT EQUATIONS

Following Ref. [12], we start our analysis by defining as order parameter the *local overlap* m_i^μ defined as

$$m_i^\mu = \frac{1}{C}\sum_j \varpi_{ij}(\eta_j^\mu/a - 1)v_j. \qquad (3)$$

This parameter measures the degree of retrieval of pattern μ, i.e., if pattern μ is retrieved, then $\frac{1}{N}\Sigma_j m_j^\mu = \frac{1}{Na}\Sigma_j(\eta_j^\mu - a)v_j = O(1)$. However if it is not retrieved, this sum will be $O(1/\sqrt{N})$ [13].

We hope to derive equations that relate these parameters to each other and then see whether it is possible to have a solution in which the sum $\frac{1}{N}\Sigma_j m_j^\mu$ is large for one pattern (without loss of generality we take it to be $\mu=1$) and not for the others. This could be done through the signal-to-noise analysis which is a classical tool in analyzing autoassociative networks and has been used extensively in the literature [26–31]. In signal-to-noise analysis one assumes that the effect of nonretrieved patterns on the retrieval is a Gaussian noise to each neuron i. The variance of this noise is denoted by ρ_i^2. One then writes the input to each unit as a sum of signal plus a random Gaussian variable with variance ρ_i^2 and then uses the result to relate ρ_i^2 and m_i^1. Details of such calculation are highlighted in Appendix A.

By going to a continuous limit with a straightforward redefinition of the above parameters and using a self-consistent signal-to-noise analysis (SCSNA [12,26]), the fixed-point equations for our network can be derived in all generality (see Appendix A and also Ref. [12]). It is convenient to adjust our notation to the continuum limit. $\mathbf{r}$ henceforth represents the position of each of N units on a continuum manifold of dimension d, while the index i previously used to indicate discrete unit positions will be recycled in later sections to denote, instead, distinct Fourier components along each spatial dimension; and C continues to denote the number of units connected to a given unit. Later on, we shall assume for simplicity our manifold to be a d-dimensional hypertorus of linear size $2L$, with unitary spacing among the units, i.e., $(2L)^d = N$. The SCSNA yields in the continuum limit the equations

$$m(\mathbf{r}_2) = \frac{1}{C}\int d\mathbf{r}_1 \wp(\mathbf{r}_2;\mathbf{r}_1) I_2(\mathbf{r}_1),$$

$$\rho^2(\mathbf{r}_2) = \frac{\alpha T_0^2}{C}\int d\mathbf{r}_1 A(\mathbf{r}_2;\mathbf{r}_1) I_3(\mathbf{r}_1),$$

$$\psi(\mathbf{r}_2;\mathbf{r}_1) = \int d\mathbf{r} K(\mathbf{r}_2;\mathbf{r})\wp(\mathbf{r};\mathbf{r}_1) + \int d\mathbf{r}d\mathbf{r}' K(\mathbf{r}_2;\mathbf{r})K(\mathbf{r};\mathbf{r}')\wp(\mathbf{r}';\mathbf{r}_1) + \cdots ,$$

$$K(\mathbf{r}_2;\mathbf{r}) = \frac{T_0}{C}\wp(\mathbf{r}_2;\mathbf{r})\left\langle \int DzG'(\mathbf{r}) \right\rangle \equiv \wp(\mathbf{r}_2;\mathbf{r})\Phi(\mathbf{r}),$$

$$\Gamma(\mathbf{r}) = \alpha T_0 \psi(\mathbf{r};\mathbf{r}), \quad x = \frac{1}{N}\int d\mathbf{r} DzG(\mathbf{r};\Gamma), \qquad (4)$$

where $\alpha = p/C$ is the storage load, $T_0 = \frac{1-a}{a}$ and

$$I_2(\mathbf{r}) = \left\langle [\eta(\mathbf{r})/a - 1]\int DzG(\mathbf{r};\Gamma) \right\rangle,$$

$$I_3(\mathbf{r}) = \left\langle \int DzG(\mathbf{r};\Gamma)^2 \right\rangle,$$

$$A(\mathbf{r}_2;\mathbf{r}_1) = \wp(\mathbf{r}_2;\mathbf{r}_1) + 2\wp(\mathbf{r}_2;\mathbf{r}_1)\psi(\mathbf{r}_2;\mathbf{r}_1) + \psi(\mathbf{r}_2;\mathbf{r}_1)^2,$$

and $Dz \equiv dz(e^{-z^2/2}/\sqrt{2\pi})$; while $v(\mathbf{r}) = G(\mathbf{r};\Gamma) \equiv \hat{G}[\hat{h}(\mathbf{r});\Gamma]$ is the self-consistent solution of $v(\mathbf{r}) = F[\hat{h}(\mathbf{r}) + \Gamma(\mathbf{r})v(\mathbf{r})]$, and, finally, $\hat{h}(\mathbf{r}) \equiv h(\mathbf{r}) - \Gamma(\mathbf{r})v(\mathbf{r}) - \mathrm{Th}$ is the part of the local field at $\mathbf{r}$ which does not directly depend on $v(\mathbf{r})$, minus the threshold Th.

Among the above order parameters, ψ has a nice physical interpretation. It measures how the reverberation of the noise through loops in the network affects retrieval. In the simple case of nonmetric connectivity, ψ vanishes if the network is extremely diluted. Each term in the sum in the equation for ψ represents one level of a "loop expansion," which in some cases, e.g., for a network without structure, can be closed [12].

IV. NONUNIFORM SOLUTIONS: THE LIMIT $\alpha \to 0$

When $\alpha = 0$, analyzing the formation of nonuniform solutions becomes simple even for a d-dimensional network. It is convenient, and instructive for the latter treatment of the $\alpha \neq 0$ case, to carry out the analysis in Fourier space. In the $\alpha = 0$ case, the fixed-point equations in the continuum limit read

$$\tilde{m}_{i_1,\ldots,i_d} = \frac{\tilde{\wp}_{i_1,\ldots,i_d}}{C}\int d\mathbf{r}\prod_{n=1}^{d}\cos\left(\frac{\pi i_n r_n}{L}\right) \times \left\langle \left(\frac{\eta^\mu(\mathbf{r})}{a} - 1\right) F[h(\mathbf{r},\eta^\mu)] \right\rangle,$$

$$x = \frac{1}{N}\int d\mathbf{r}\langle F[h(\mathbf{r},\eta^\mu)]\rangle,$$

$$h(\mathbf{r},\eta^{\mu})=\left(\frac{\eta^{\mu}(\mathbf{r})}{a}-1\right)m(\mathbf{r})-\mathrm{Th}, \tag{5}$$

where $\widetilde{m}_{i_1,i_2,\ldots,i_d}$ and $\widetilde{\wp}_{i_1,\ldots,i_d}$ are the Fourier modes of $m(\mathbf{r})$ and $\wp(\mathbf{r},\mathbf{r}')$, respectively. Note that in this Fourier decomposition, it is assumed that $\wp(\mathbf{r},\mathbf{r}')$ depends just on $|r_i-r_i'|$, $\forall i$.

We can first try to see whether these equations admit any solution which is uniform in space, i.e., $m(\mathbf{r})=\text{const}$. In Fourier space, a constant solution has $\widetilde{m}_{0,\ldots,0}\neq 0$ and all the others modes equal to zero.

It is easy to check that the above equations admit the *spatially uniform* solution

$$\widetilde{m}_{i_1,i_2,\ldots,i_d}=(1-a)\,\delta_{i_1 0}\delta_{i_2 0}\cdots\delta_{i_d 0},$$
$$\mathrm{Th}=(1-a)(1/a-1)-F^{-1}(1),$$

provided $F^{-1}(1)<(1-a)/a$ [45]. Even though this solution exists, its stability is not guaranteed. In fact we can show that the stability of this solution depends on the structure of the connectivity. For instance, considering a Gaussian connectivity distribution,

$$\wp(\mathbf{r},\mathbf{r}')=\frac{C}{(2\pi\sigma^2)^{d/2}}\prod_{i=1}^{d}\exp\left(\frac{-(r_i-r_i')^2}{2\pi\sigma^2}\right), \tag{6}$$

a linear stability analysis shows that the uniform solution is stable only for $\sigma>\sigma_c$, where

$$\sigma_c=\frac{L}{\pi}\sqrt{2\ln\{a(1/a-1)^2F'[F^{-1}(1)]\}}, \tag{7}$$

and L, we remind, denotes the half-length of each dimension.

For $\sigma<\sigma_c$ the uniform solution becomes unstable, and the instability can be in the direction of any of the first Fourier modes $\widetilde{m}_{10\ldots0},\widetilde{m}_{01\ldots0},\ldots,\widetilde{m}_{00\ldots1}$ [46].

Here without loss of generality and for simplicity we take $\widetilde{m}_{10\ldots0}$, among all d possible directions of instability, as the one that becomes unstable first. In a network of threshold-linear units, for which $F(x)=gx\Theta(x)$, where g is the linear gain, the equation for $\widetilde{m}_{10\ldots0}$—provided $\widetilde{m}_{10\ldots0}<a/[g(1-a)]$—reads

$$\widetilde{m}_{10\cdots0}=g(\widetilde{\wp}_{10\cdots0}/C)a(1/a-1)^2\widetilde{m}_{10\cdots0}. \tag{8}$$

This equation, exactly at σ_c, is satisfied for any $\widetilde{m}_{10\ldots0}$. This means that the system is marginally stable at this point in the direction of $\widetilde{m}_{10\ldots0}$, resulting in a jump to $\widetilde{m}_{10\ldots0}=a/[g(1-a)]$ for $\sigma<\sigma_c$. This is shown in Fig. 1, which presents numerical solutions of the equation for $\widetilde{m}_1$, for different values of σ/L, when the network is on a ring. It is worth noting that such trivial equation for $\widetilde{m}_{10\ldots0}$ comes directly from the linear nature of the threshold-linear function above threshold. Adding quenched noise to a network of threshold-linear units or using any other function (see the next sections), would change this trivial equation to a nonlinear form, resulting in the disappearance of the jump [47]. This is important for the analysis of the networks with $\alpha\neq 0$, as one can expect that the transition to nonuniform solution will not be abrupt in the presence of quenched noise. This is verified also by numerically solving the fixed-point equations in the next section.

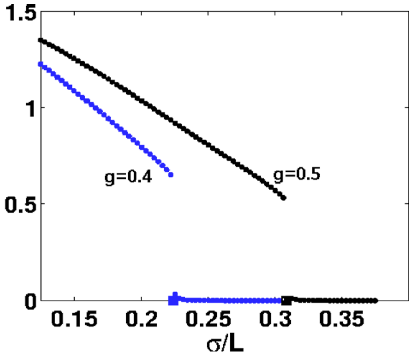

FIG. 1. (Color online) The dependence of the first Fourier mode $\widetilde{m}_1$ on σ/L, from numerically solving the equation for $\widetilde{m}_1$ for threshold-linear units on a ring with a Gaussian connectivity pattern, $a=0.2$ and $\alpha=0$. Black, $g=0.5$ and blue, $g=0.4$. The squares on the σ/L axis indicate the transition points to nonuniform solution as predicted by Eq. (7).

An important point about the limit $\alpha\to 0$ is that, as we shall see later by numerically solving the fixed-point equations, Eq. (7) is a good approximation for the critical σ even when $\alpha\neq 0$. We shall show that the effect of adding quenched noise is mainly to change the order of the transition (from sharp to smooth), with only a minor change in the critical width σ_c, and on the degree of "bumpiness" of the solution for $\sigma<\sigma_c$.

V. FIXED-POINT EQUATIONS WITH $\alpha\neq 0$

In the preceding section we assumed that the ratio $\alpha=p/C$ goes to zero when $C\to\infty$. Without this assumption, in order to find out where nonuniform solutions appear, one approach is to use a perturbation around the uniform solution, and see when that perturbation destabilizes the uniform solution. We previously used this approach to analyze the behavior of an associative network with Gaussian connectivity and threshold-linear model neurons [12]. The problem with this perturbative analysis is that even though it does tell us where the uniform solution becomes unstable, it does not tell us how the solution behaves below σ_c, unless when we are very close to σ_c, and even in that case the analysis will be very involved.

Here, alternatively, we choose to numerically solve the steady state equations, Eqs. (4), for many values of σ, and thus try to find the transition point σ_c, and what happens below it. One advantage of numerically solving these equations is the possibility to estimate the storage capacity. In order to obtain this estimation, one needs to find out for which values of α the self-consistent equations (4) admit a solution with $\int d\mathbf{r}\,m(\mathbf{r})\neq 0$ or, alternatively in Fourier space, $\widetilde{m}_{0\cdots0}\neq 0$.

Solving Eqs. (4) is in general very time consuming, given the complexity of the equations and the fact that they are functional equations. One can, however, assume that the spatially modulated overlaps can be approximated by their first Fourier modes. This assumption will be shown to be reasonable, at least when the connectivity probability distribution is a Gaussian function of the distance between neurons [32].

With this assumption, one should of course write Eqs. (4) in Fourier space. For $\alpha \neq 0$ we focus, for simplicity, on a network which lies on a 1D ring with half-length L; the analysis could be extended, though, to arbitrary dimension (at a heavier computational cost to calculate the integrals below).

Assuming that $\wp(r,r')$ depends on $|r-r'|$, we write the connectivity matrix, m and ρ^2 in their Fourier modes and, after a little bit of algebra, find the following fixed-point equations [from now on $c(r) \equiv \cos(\pi r/L)$, $s(r) \equiv \sin(\pi r/L)$ and Δ is a dummy function label that can be c, representing a cosine, or s, representing a sine]:

$$\wp(r_2,r_1) = \sum_k \tilde{\wp}_k c[k(r_2 - r_1)],$$

$$\tilde{m}_k = \frac{\tilde{\wp}_k}{C} \int drc(kr) I_2(r),$$

$$\rho^2{}_k = \frac{\alpha T_0^2}{(1+\delta_{k0})LC}(Y_1^k + Y_2^k + Y_3^k),$$

$$\psi(r_2;r_1) = \sum_{kj\Delta} \psi_{kj}^{\Delta} \Delta(kr_2)\Delta(jr_1), \qquad (9)$$

where ψ_{kl}^{Δ} and Y_i^k are dummy variables defined in Appendix B and for calculating them one should calculate a 1D integral. After transforming to Fourier space, the above equations for $\tilde{m}_k, \tilde{\rho}_k, \tilde{\psi}_r^c, \ldots$ can be solved iteratively. The number of terms that one includes in the sum in Eq. (B1) (defining ψ_{rs}^{Δ}), together with the number of modes that one considers to approximate the connectivity structure, determine the accuracy of the calculation. Note that the above equations do not have any new physics in them, but to calculate the integrals involved in these equations is much easier than those involved in the original real space equations, Eqs. (4). Having to deal with a finite number of Fourier modes, instead of the functions that satisfy the original equations, makes the numerics much easier.

Now we can use the above equations (9) to see how the results of the preceding section can be generalized to the case of $\alpha \neq 0$. We again concentrate on the threshold-linear neuron model for which we have

$$G(r;\Gamma)_{TL} = \frac{g}{1-g\Gamma}[\hat{h}(r)]\Theta[\hat{h}(r)],$$

where again $\hat{h}(r) = h(r) - \Gamma(r)v(r) - \mathrm{Th}$ is the part of the local field which does not directly depend on $v(r)$, minus the threshold Th. We also again assume a Gaussian connectivity on a ring

$$\wp(r_2 - r_1) = (C/\sqrt{2\pi\sigma^2})\exp[-(r_2-r_1)^2/2\pi\sigma^2]. \qquad (10)$$

In Fig. 2 we plot the amplitude of the first Fourier mode $\tilde{m}_1$—an indication of the deviation from the uniform solution—as a function of σ for threshold-linear units (dotted, $\alpha=0$; dashed-dotted, $\alpha=0.1$; and full line, $\alpha=0.15$; in this figure and others, always $a=0.2$ and $C/N=0.05$). With threshold-linear units, for small enough σ/L, the solution is essentially a localized bump. Therefore, we conclude that even when $\alpha \neq 0$ nonuniform retrieval is possible. Moreover, we can make two important observations at this point. One is that the value of σ_c at which the uniform solution is destabilized and the nonuniform solution appear, seems to be just mildly affected by α. As a result, the analytical equation (7) that we derived previously for the case of $\alpha \to 0$ appears to be a good first approximation for $\alpha \neq 0$. The second point is that even for small values of σ, the three curves in Fig. 2 remain close to each other. In fact the only major effect that decreasing α to zero induces is that it changes the transition to nonuniform retrieval from a smooth to an abrupt one, as has been mentioned in Sec. IV. These points suggest that, in order to analyze the formation of nonuniform activity profiles, we can concentrate on the $\alpha \to 0$ limit, in which case the fixed-point equations simplify considerably.

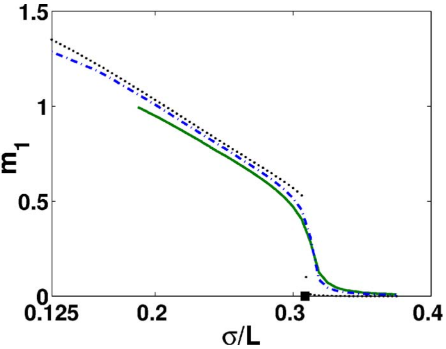

FIG. 2. (Color online) Dependence of the first Fourier mode $\tilde{m}_1$ on σ/L, for a network of threshold-linear units on a ring. In this graph, we have used $g=0.5$, $a=0.2$, and $\alpha=0$ (black, dotted), $\alpha=0.1$ (blue, dashed-dotted) and $\alpha=0.15$ (green, full line). The black square on the σ/L axis indicates the transition point to nonuniform solution as predicted by Eq. (7).

To understand the behavior of other modes, we have plotted the amplitude of the second and the third Fourier modes in addition to the first one, as a function of σ/L, in Fig. 3. As one can appreciate from this graph, the second and third Fourier modes become nonzero for smaller values of σ/L, compared to the first mode. Note that at $\sigma/L=0.125$, and for

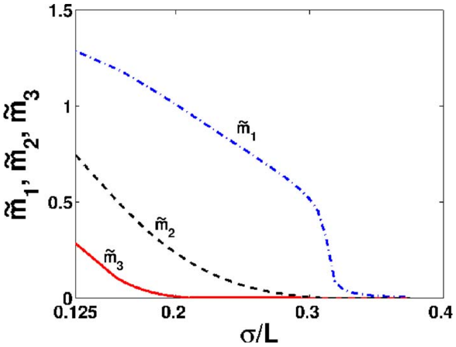

FIG. 3. (Color online) Dependence of the first Fourier mode $\tilde{m}_1$ (blue, dashed-dotted line), second Fourier mode $\tilde{m}_2$ (black, dashed line), and third Fourier mode $\tilde{m}_3$ (red, full line) on σ/L, for a network of threshold-linear units on a ring. In this graph we have used $g=0.5$, $a=0.2$, and $\alpha=0.1$. The first Fourier mode becomes nonzero first and is the significant mode for a range of σ/L. For smaller values of σ/L, however, the second and then the third modes also become nonzero, and gradually important in describing the shape of the solution.

the $C=0.05N=0.1L$ that we have used in this paper, the peak of the Gaussian probability distribution Eq. (10) is ~ 0.32. This is six times larger than the probability of connectivity if neurons were connected randomly with uniform probability $C/N=0.05$. From Fig. 3, one can see that for this localized connectivity with $\sigma/L=0.125$, the first mode is approximately twice the second mode and therefore the solution cannot be approximated solely by the first mode. Thus the graph shows that even though for less localized connectivity patterns the only significant mode is the first one, for more localized connectivity structures other modes also become important in defining the shape of the solution.

VI. LOCALIZED SOLUTIONS: ALTERNATIVE NEURON MODELS

An intriguing question is to understand whether it is possible to get localized retrieval in networks comprised of model neurons other than threshold-linear. This question has been investigated by other authors previously. Koroutchev and Korutcheva [15,22] analyzed a network of 0–1 binary neurons with symmetric, but metrically organized connections using the replica trick. These authors found out that such networks are not capable of showing localized retrieval if the mean activity of the network is kept fixed during retrieval to the same value as the mean activity of the stored patterns. Simulation studies of a network comprised of integrate-and-fire neurons also show that it seems unlikely to have both localized activity and retrieval of a pattern [14]. In this part of the paper, we study how the single neuron model affects the possibility of having a nonuniform retrieval state.

First, we concentrate on the binary transfer functions, for which we have

$$G(r;\Gamma)_B = \xi\Theta[\hat{h}(r)], \tag{11}$$

where ξ represents the value of the high state of the unit, e.g., $\xi=1$ for a classical 0–1 binary unit. As in the preceding section, a Gaussian connectivity on a ring is again assumed, $\wp(r_2-r_1)=(C/\sqrt{2\pi\sigma^2})\exp[-(r_2-r_1)^2/2\pi\sigma^2]$.

The results of solving the fixed-point equations is shown in Fig. 4. As opposed to threshold-linear units, a network of 0–1 binary neurons fails to exhibit nonuniform retrieval [14,15,22]. This difference with threshold-linear network can be understood in the following way. There are two conditions to be satisfied for the retrieval state to exist: $\tilde{m}_0=\frac{1}{N}\int dr[\eta^1(r)/a-1]v(r)=\mathrm{O}(1)$ and $x=\frac{1}{N}\int dr v(r)=a$. The second condition means that, for spatially modulated retrieval states, in some parts of the network units with activity 1 in the corresponding stored pattern should have activity below 1, and in other parts above 1. The latter requirement poses no problem to the threshold-linear network, whose units can reach high levels of activity. For a network with binary units, or with units that saturate, the crucial issue is whether the *up* state, or the saturation level, is sufficiently above 1 (the arbitrarily set activity level of active units in the stored patterns; obviously the argument can be generalized to nonbinary stored patterns). Thus binary units with activity levels, say, 0 and 1.5 (relative to the *up* state in the stored pattern) should be able to show spatially modulated activity profiles, although, rather than localized *bumps*, they appear as square-shaped spatially restricted activity. This results in the cyan and green curves for $\tilde{m}_1$ in Fig. 4.

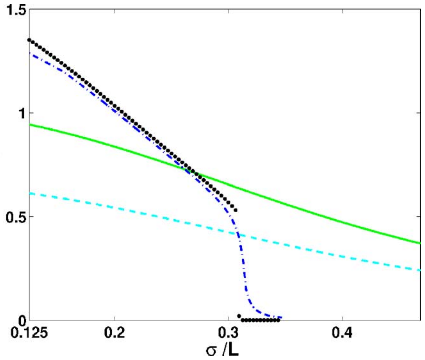

FIG. 4. (Color online) $\tilde{m}_1$ versus σ/L for 0–1.5 binary units (cyan curve, dashed line) and for 0–2 binary units (green curve, full line), both at $\alpha=0$. For comparison, we also have replotted the case of threshold-linear units with $g=0.5$ and $\alpha=0$ (black, dotted line) or $\alpha=0.1$ (blue, dashed line). The 0–1 binary unit gives $\tilde{m}_1=0$ for all values of σ.

Based on this intuitive argument, another way of getting spatially modulated retrieval in 0–1 binary network would be to relax the constraint $x=\frac{1}{N}\int dr v(r)=a$, as discussed in Ref. [15].

To further assess the effect of the saturation level on the formation of localized retrieval states, we consider the following input-output function:

$$F(h-\mathrm{Th}) = \varepsilon \tanh[g(h-\mathrm{Th})/\varepsilon]\Theta(h-\mathrm{Th}),$$

where g is the slope at threshold and ε is the saturation level. One should notice that for a sufficiently high ε this transfer function is effectively just a threshold-linear function.

For simplicity we focus on the $\alpha\to 0$ limit, as we do not expect the quenched noise to make any qualitative change in the behavior of the system, except for the smoothness of the transition. Figure 5 shows how $\tilde{m}_1$ changes with σ for fixed $g=0.5$ and for different values of the saturation level, as measured by ε. When the saturation is set at $\varepsilon=1$, for the intuitive reason sketched above the first Fourier mode does

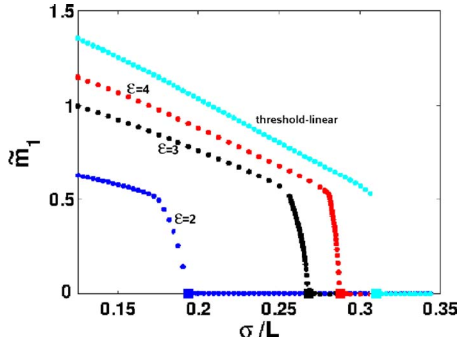

FIG. 5. (Color online) $\tilde{m}_1$ versus σ/L for $g=0.5$ for different values of the saturation level: $\varepsilon=2$ (blue), $\varepsilon=3$ (black), and $\varepsilon=4$ (red). Cyan, threshold-linear units ($\varepsilon\to\infty$).

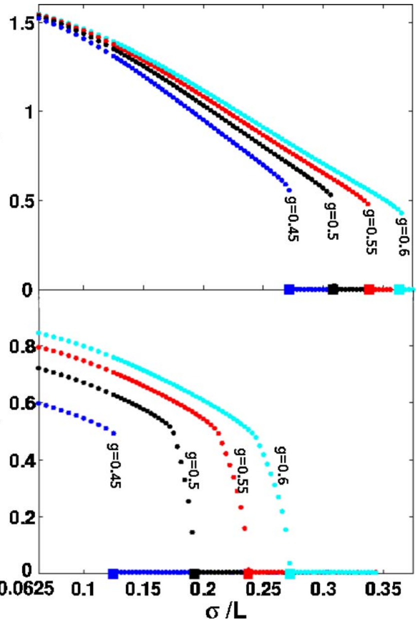

FIG. 6. (Color online) $\tilde{m}_1$ versus σ/L for different values of the gain: $g=0.45$ blue, $g=0.5$ black, $g=0.55$ red, and $g=0.6$ cyan: (upper panel) threshold-linear and (lower panel) saturating units with $\varepsilon=2$. With saturating units decreasing g (thus linearizing the input-output function close to threshold) sharpens the transition. The filled squares indicate σ_c, as predicted by Eq. (7).

not differ from zero. By increasing ε, however, one approaches the threshold-linear regime. In Fig. 6 we plot $\tilde{m}_1$ versus σ for different values of g and for both threshold-linear and saturating input-output functions. Notice the quasilinear behavior for values of σ below the transition.

VII. STORAGE CAPACITY

The next issue we address is the effect of nonuniform retrieval on storage capacity. Previously, we have found that localized connectivity decreases the storage capacity through two mechanisms [12]. First, a localized connectivity increases the number of loops, which in turn amplify the effect of quenched noise. Second, localized retrieval decreases, effectively, the number of connections available to active units to receive the retrieval signal, particularly at the flanks of the activity profile.

As a first approximation, one might assume that these two effects are independent, and estimate them in the following way. The first effect could be estimated by taking a guess function and plugging it into the fixed-point equations, and then fixing the parameters of the guess function by, for instance, optimizing the resulting storage capacity for these parameters. This gives us the storage capacity if only the first effect had contributed. The effect of loops can be estimated by putting a uniform solution in the fixed-point equations and calculating how much loops decrease the storage capacity of network assuming such a uniform solution, compared to the network with the same solution in the absence of loops. Considering these two effects as additive leads to a good approximation for the storage capacity, at least for the threshold-linear units [12].

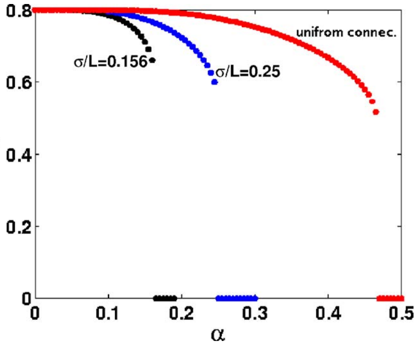

FIG. 7. (Color online) $\tilde{m}_0$ versus α for different values of σ/L in a threshold-linear network with $g=0.5$. Black for $\sigma/L=0.156$, blue for $\sigma/L=0.25$, and red for a structureless network. Note the corresponding values, $\tilde{m}_1=1.22$ for $\sigma/L=0.156$, $\tilde{m}_1=0.8$ for $\sigma/L=0.25$, and $\tilde{m}_1=0$ for the structureless network, all when $\alpha=0.05$.

Here, instead, we aim to calculate the storage capacity to arbitrary accuracy by numerically solving the fixed-point equations. This is an alternative and more accurate approach to calculate the storage capacity, when compared to the approximation scheme that we had presented before. One should bear in mind that both ways of calculating the storage capacity may be useful, but in different scenarios. Finding the storage capacity through numerically solving the fixed-point equations is more accurate, however it can be quite slow, particularly in 2D or 3D, to calculate the required integrals and obtain the convergence to a solution. The approximate method, although fast and easy, can be rather inaccurate if the function that one uses to estimate the effect of localization is not close to the real solution and/or the two effects are highly correlated. It is therefore useful to consider both methods.

In Fig. 7 we plot $\tilde{m}_0$ as a function of α, for $\sigma/L=0.156$, $\sigma/L=0.25$, and for a uniform connectivity pattern. Although the storage capacity $\alpha_c=\mathrm{Inf}_\alpha[\alpha|\tilde{m}_0(\alpha)=0]$ decreases, the decrease is not too severe, even for very localized solutions. When $\sigma/L=0.156$, for instance, the solution is quite localized, and the storage capacity is decreased by a factor ~ 3 compared to uniform retrieval. The fact that even for very localized connectivity $\alpha_c=p_c/C$ does not go to zero, but remains finite, means that the maximum number of retrievable patterns remains proportional to C.

VIII. DISCUSSION

In this paper we have studied the effect of a spatially organized connectivity pattern on pattern retrieval with rate-based model neurons. Although the steady state equations that we have derived are applicable to any transfer function, to study the behavior of their solutions we have focused on three alternative input-output transfer functions.

The results we have presented in this paper show that, in general, a network with a fairly realistic single unit input-output transfer function becomes capable of localized retrieval, if the connectivity is sufficiently concentrated at short distances, simply by manipulating the single unit saturation level and its gain. Increasing the gain, and/or the saturation level, makes retrieval states more localized. These parameters can be effectively controlled via inhibitory mechanisms. The effect of the quenched noise is minor on the qualitative behavior of the system, making the analytic formula, Eq. (7), a reasonable approximation for a wide range of parameters. The fact that, for a given value of σ, changing the saturation level or the slope at threshold (g) can force the network out of the spatially modulated retrieval regime may explain the result of Ref. [14].

It is known that autoassociative networks require sparse coding in order to benefit from a large storage capacity [33–36]. Mathematically speaking, the scaling relation between the maximum retrievable number of patterns p_c and a usually takes the form of $p_c \propto 1/[a \ln(1/a)]$. Even though the formal mathematical analyses leading to this result have been mainly carried out for networks comprised of simplified model neurons, it is expected from a simple signal-to-noise analysis that sparse codes should be beneficial even in networks comprised of more complex model neurons. Contrary to the sparse coding regime expected from such theoretical considerations, however, experimental measures of sparsity, e.g., in visual area IT, which is believed to be the storehouse for long term object memory [37], tend to yield relatively large values of a, such as $a \simeq 0.7$ [38,39]. Our results here suggest that the reported values of a may have been measured, effectively, conditional to the recorded units being active in a localized retrieval state, thus overestimated by neglecting the large quasisilent part of the network.

It is also important to notice that localized retrieval, while quantitatively decreasing local storage capacity, may considerably increase the computational power of a network with structured connectivity. This can be appreciated by noting that, in a large network, more than one memory pattern of activity may be retrieved at the same time, each in a different location, without much interference. A combination of locally retrieved memories can be thought of as a global, composite memory pattern. The number of such composite patterns would be combinatorially large, thus hugely increasing the overall storage capacity of the network. Note that this is unlikely to happen in a network without metrically organized connectivity, as a result of the instability of mixed, so-called *spurious*, states [40]. Each neuron in the isocortex receives of the order of 10^4 connections, and this number implies a storage capacity for at most a similar number of locally retrievable patterns. The fact that the number of memories stored in the isocortex seems much higher may stem from the combinatorial character of global memory patterns, allowed by the localization discussed here. This issue may be explored further by studying modular networks [41–43].

ACKNOWLEDGMENTS

The support of Human Frontier Science Program RG P0047/2004-C is gratefully acknowledged.

APPENDIX A: THE SELF-CONSISTENT SIGNAL-TO-NOISE ANALYSIS

Here we outline the main steps which leads to the fixed-point equations (4) for readers who may be unfamiliar with the self-consistent signal-to-noise analysis.

Using the definition of m_i^μ in Eq. (3), we can write the input to unit i as

$$h_i = \sum_\mu (\eta_i^\mu/a - 1)m_i^\mu - \alpha\varpi_{ii}(1/a - 1)v_i + b(x), \quad \text{(A1)}$$

where $\alpha = p/C$ is the storage load.

As a result the activity of the network can be written as

$$v_i = F\left((\eta_i^1/a - 1)m_i^1 + (\eta_i^\nu/a - 1)m_i^\nu + \sum_{\mu\neq 1,\nu}(\eta_i^\mu/a - 1)m_i^\mu \right.$$
$$\left. - \alpha\varpi_{ii}(1/a - 1)v_i - \mathrm{Th}\right). \quad \text{(A2)}$$

where we have singled out one of the nonretrieved patterns ν for the reason that will become clear soon, and have assumed that the first patterns is retrieved. We have also absorbed the effect of $b(x)$ in the threshold, Th, as discussed in the text. The sum in Eq. (A2) is taken to be comprised of a Gaussian random noise, with variance $[\rho_i^\mu]^2$, plus a term which is proportional to v_i. The variance of the noise can be taken to be independent from the pattern μ and we can simply denote it by $[\rho_i]^2$. With these assumptions—which are the main assumptions of the signal-to-noise analysis—Eq. (A2) can be solved for v_i leading to

$$v_i = G[(\eta_i^1/a - 1)m_i^1 + (\eta_i^\nu/a - 1)m_i^\nu + \rho_i z - \mathrm{Th}], \quad \text{(A3)}$$

where z is a Gaussian random variable with variance unity. The function G is defined through this equation.

We can now expand Eq. (A2) up to the first order in m_i^ν. The result can be plugged into the definition of m_i^ν in Eq. (3) to derive a self-consistent equation for m_i^ν. Solving this self-consistent equation gives us the relation between $m_i^{\nu\neq 1}$ and the variables ρ_i and $m_i = m_i^1$, and therefore we can write the right-hand side of Eq. (A3) in a form independent of $m_i^{\nu\neq 1}$. The result can be used in the definitions of ρ_i and m_i to derive the self-consistent equations (4). The other variables in these equations, i.e., ψ and Γ, are parameters which naturally appear in the course of calculating $m_i^{\mu\neq 1}$.

For a more detailed account of the calculation, and a discussion of the validity of the approximations involved, we refer the reader to Refs. [12,26,27,44].

APPENDIX B

The dummy variable ψ_{kl}^Δ and Y_i^k in Eq. (9) are defined as

$$\psi_{kl}^\Delta = \tilde{\wp}_k \tilde{\wp}_l \sum_n \Phi_{kl}^{\Delta,n},$$

$$\Phi_{kl}^{\Delta,n+1} = \sum_i \tilde{\wp}_i \Phi_{ki}^{\Delta,n} \Phi_{il}^{\Delta,1},$$

$$\Phi_{ij}^{\Delta,1} = \int dr \Phi(r)\Delta(ir)\Delta(jr) \quad \text{(B1)}$$

and

$$Y_1^k = (1 + \delta_{k0})L\tilde{\wp}_k \int drc(kr)I_3(r),$$

$$Y_2^k = 2\sum_{\Delta} \int dr Q_k^{\Delta}(r) I_3(r),$$

$$Y_3^k = \sum_{\Delta} \int dr W_k^{\Delta}(r) I_3(r),$$

$$Q_k^{\Delta}(r) = \sum_{ijl} \tilde{\wp}_i \psi_{jl}^{\Delta} \Delta(ir)\Delta(lr)\Pi_{ijk}^{\Delta},$$

$$W_k^{\Delta}(r) = \sum_{ii'jl} \psi_{ii'}^{\Delta} \psi_{jl}^{\Delta} \Delta(i'r)\Delta(lr)\Pi_{ijk}^{\Delta},$$

$$\Pi_{ijk}^{\Delta} = \int dr \Delta(ir)\Delta(jr)c(kr). \tag{B2}$$

[1] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985).

[2] D. J. Amit and M. V. Tsodyks, Network Comput. Neural Syst. **2**, 275 (1991).

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

[4] D. J. Amit and N. Brunel, Cereb. Cortex **7**, 237 (1997).

[5] P. E. Latham and S. Nirenberg, Neural Comput. **16**, 1385 (2004).

[6] A. Treves, Network **4**, 259 (1993).

[7] C. van Vreeswijk and H. Sompolinsky, in *Methods and Models in Neurophysics*, edited by C. Chow, B. Gutkin, D. Hansel, C. Meunier, and J. Dalibard (Elsevier, Amsterdam, 2005).

[8] J. M. Fuster, *Memory in the Cerebral Cortex* (The MIT Press, Cambridge, 1994).

[9] V. Braitenberg and A. Schuz, *Anatomy of the Cortex* (Springer, Berlin, 1991).

[10] B. Hellwig, Biol. Cybern. **82**, 111 (2000).

[11] A. Bovier and V. Gayrard, Markov Processes Relat. Fields **3**, 392 (1997).

[12] Y. Roudi and A. Treves, J. Stat. Mech.: Theory Exp. 1, P070102 (2004).

[13] D. J. Amit, *Modeling Brain Function* (Cambridge University Press, Cambridge, 1989).

[14] A. Anishchenko, E. Bienenstock, and A. Treves, q-bio.NC/0502003 (unpublished).

[15] K. Koroutchev and E. Korutcheva, Cent. Eur. J. Phys. **3**, 409 (2005).

[16] L. G. Morelli, G. Abramson, and M. N. Kuperman, Eur. Phys. J. B **38**, 495 (2004).

[17] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, Proc. Natl. Acad. Sci. U.S.A. **92**, 3844 (1995).

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

[19] K. Zhang, J. Neurosci. **16**, 2112 (1996).

[20] A. Compte, N. Brunel, P. S. Goldman-Rakic, and X. Wang, Cereb. Cortex **10**, 910 (2000).

[21] A. Roxin, N. Brunel, and D. Hansel, Phys. Rev. Lett. **94**, 238103 (2005).

[22] K. Koroutchev and E. Korutcheva, Phys. Rev. E **73**, 026107 (2006).

[23] J. Buhmann, R. Divko, and K. Schulten, Phys. Rev. A **39**, 2689 (1989).

[24] M. V. Tsodyks and M. V. Feigelman, Europhys. Lett. **6**, 101 (1988).

[25] E. T. Rolls and A. Treves, *Neural Networks and Brain Function* (Oxford University Press, Oxford, 1998).

[26] M. Shiino and T. Fukai, J. Phys. A **25**, L375 (1992).

[27] M. Shiino and T. Fukai, Phys. Rev. E **48**, 867 (1993).

[28] D. Bolle, J. B. Blanco, and T. Verbeiren, J. Phys. A **37**, 1951 (2004).

[29] A. C. C. Coolen, in *Handbook of Biological Physics*, edited by F. Moss and S. Gielen (Elsevier, New York, 2001), pp. 597–662.

[30] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, Oxford, 2001).

[31] P. E. Latham, in *Advances in Neural Information Processing Systems*, edited by T. G. Dietterich, S. Becker, and Z. Ghahramani (MIT Press, Cambridge MA, 2002), p. 14.

[32] C. Mehring, U. Hehl, M. Kubo, M. Diesmann, and A. Aertsen, Biol. Cybern. **88**, 395 (2003).

[33] N. Brunel, in *Methods and Models in Neurophysics*, edited by C. Chow, B. Gutkin, D. Hansel, C. Meunier, and J. Dalibard (Elsevier, Amsterdam, 2005).

[34] D. Field, Neural Comput. **6**, 559 (1994).

[35] B. A. Olshausen and D. J. Field, Curr. Opin. Neurobiol. **14**, 481 (1994).

[36] A. Treves and E. T. Rolls, Network Comput. Neural Syst. **2**, 371 (1991).

[37] Y. Miyashita, Annu. Rev. Neurosci. **16**, 245 (1993).

[38] E. T. Rolls and M. J. Tovee, J. Neurosci. **73**, 713 (1995).

[39] A. Treves, S. Panzeri, E. T. Rolls, M. Booth, and E. A. Wakeman, Neural Comput. **11**, 601 (1999).

[40] Y. Roudi and A. Treves, Phys. Rev. E **67**, 041906 (2003).

[41] C. F. Mari and A. Treves, BioSystems **48**, 47 (1998).

[42] E. Kropff and A. Treves, J. Stat. Mech.: Theory Exp. 2, P08010 (2005).

[43] D. O'Kane and A. Treves, Network Comput. Neural Syst. **3**, 379 (1992).

[44] M. Shiino and M. Yamana, Phys. Rev. E **69**, 011904 (2004).

[45] This condition would impose some constraints on the parameters that characterize the single unit input-output function. In the threshold-linear case, for example, it means $g > a/(1-a)$, where g is the linear gain of the input-output function above threshold. In a cortical network this condition may be satisfied via inhibitory mechanisms.

[46] Here we have assumed that the network is on a (hyper-) torus. If we assume hyperspherical boundary conditions instead (if the network is on a 2D sphere, for instance) the solution becomes unstable in the direction of the first radial Fourier mode. We thank Professor Haim Sompolinsky for pointing this out.

[47] This is evident in Figs. 2, 5, and 6.