Bump formation in a binary attractor neural network

Kostadin Koroutchev*

Departamento de Ingeniería Informática, Universidad Autónoma de Madrid, 28049 Madrid, Spain

Elka Korutcheva[†]

Departamento de Física Fundamental, Universidad Nacional de Educación a Distancia, c/Senda del Rey 9,28080 Madrid, Spain

The conditions for the formation of local bumps in the activity of binary attractor neural networks with spatially dependent connectivity are investigated. We show that these formations are observed when asymmetry between the activity during the retrieval and learning is imposed. An analytical approximation for the order parameters is derived. The corresponding phase diagram shows a relatively large and stable region where this effect is observed, although critical storage and information capacities drastically decrease inside that region. We demonstrate that the stability of the network, when starting from the bump formation, is larger than the stability when starting even from the whole pattern. Finally, we show a very good agreement between the analytical results and the simulations performed for different topologies of the network.

I. INTRODUCTION

Bump formation in recurrent neural networks has been recently reported in several investigations concerning linear-threshold units [1,2], binary units [3,4], small-world networks of integrate-and-fire neurons [5], and in a variety of spatially distributed neural network models with excitatory and inhibitory couplings between the cells [6,7].

As has been shown, the localized retrieval is due to the short-range connectivity of the networks and could explain the behavior in structures of biological relevance such as the neocortex, where the probability of connections decreases with distance [8].

In the case of the linear-threshold neural network model, signal-to-noise analysis has been recently adapted [1,2] to spatially organized networks and has shown that the retrieval states of the connected network have nonuniform activity profiles when the connections are short range enough, even without any spatial structure of the stored patterns. The increase of the gain or saturation level of neurons enhances the level of localization of the retrieval states and does not lead to a drastic decrease of the storage capacity in these networks even for very localized solutions [2].

An interesting investigation of spontaneous activity bumps in small-world (SW) networks [9,10] of integrate-and-fire neurons [5] has recently shown that the network retrieves when its connectivity is close to the random and displays localized bumps of activity when its connectivity is close to the ordered. The two regimes are mutually exclusive in the range of parameters governing the proportion of the long-range connections on the SW topology of integrate-and-

fire network, while the two behaviors coexist in the case of linear-threshold and smoothly saturated units. Moreover, it has been stated that the transition between localization and retrieval regimes, in the case of SW networks of integrate-and-fire neurons, can occur at a degree of randomness beyond the SW regime and it is not related, in general, to the SW properties of the network.

Bump formation has been also investigated in spatially distributed neural network models with excitatory and inhibitory synaptic couplings. Recently it has been shown [6] that a neural network only with excitatory couplings can exhibit localized activity from an initial transient synchrony of a localized group of cells, followed by desynchronized activity within the group. This activity may grow or may be reduced when depression with a given size of the frequency of the inputs is introduced. It is also very sensitive to the initial conditions and range of the parameters of the network.

The results of bump formation have been recently reported by us [3,4] in the case of the binary Hebb model for associative networks. We stated that these spatially asymmetric retrieval states (SAS's) can be observed if an asymmetry between the learning and the retrieval states is imposed. This means that the network is constrained to have a different activity compared to that induced by the patterns.

In the present investigation we regard a symmetric and distance-dependent connectivity for all neurons within an attractor neural network (NN) of Hebbian type formed by N binary neurons $\{S_i\}$, $S_i \in \{-1, 1\}$, $i = 1, \ldots, N$, storing p binary patterns η_i^μ, $\mu \in \{1, \ldots, p\}$. The connectivity between the neurons is symmetric, $c_{ij} = c_{ji} \in \{0, 1\}$, $c_{ii} = 0$, where $c_{ij} = 1$ means that neurons i and j are connected.

We are interested only in connectivities in which the fluctuations between the individual connectivities are small—e.g., $\forall_i \Sigma_j c_{ij} \approx cN$, where c is the mean connectivity.

The learned patterns are drawn from the following distribution

*Also at Inst. for Communications and Computer Systems, Bulgarian Academy of Sciences, 1113, Sofia, Bulgaria. Electronic address: k.koroutchev@uam.es

[†]Also at G. Nadjakov Inst. Solid State Physics, Bulgarian Academy of Sciences, 1784 Sofia, Bulgaria.

$$P(\eta_i^\mu) = \frac{1+a}{2}\delta(\eta_i^\mu - 1) + \frac{1-a}{2}\delta(\eta_i^\mu + 1),$$

where the parameter a is the sparsity of the code [11,12].

We study the Hopfield model [13]

$$H = \frac{1}{N}\sum_{ij} J_{ij}S_iS_j, \qquad (1)$$

with Hebbian learning rule [14]

$$J_{ij} = \frac{1}{c}\sum_{\mu=1}^{p} c_{ij}(\eta_i^\mu - a)(\eta_j^\mu - a). \qquad (2)$$

Further in this article we will work in terms of the variables $\xi_i^\mu \equiv \eta_i^\mu - a$.

In the case of binary networks and symmetrically distributed patterns, the only asymmetry between the retrieval and learning states that can be imposed, independent of the position of the neurons, is the total number of the neurons in a given state. Having in mind that there are only two possible states, this condition leads to a condition on the mean activity of the network, which is introduced by adding an extra term H_a to the Hamiltonian

$$H_a = NR\left(\sum_i S_i/N - a\right).$$

This term favors states with lower total activity $\Sigma_i S_i$, which is equivalent to decreasing the number of active neurons, creating an asymmetry between the learning and retrieval states.

For $R=0$, this model has been intensively studied [11,12,15] and shows no evidence of spatial asymmetric states.

For $R>0$, the energy of the system increases with the number of active neurons ($S_i=1$) and the term H_a tends to limit the number of active neurons below aN.

In the present work we show explicitly that the constraint on the network level of activity is a sufficient condition for the observation of spatially asymmetric states in binary Hebb neural networks. Similar observations have been reported in the case of the linear-threshold network [1], where in order to observe bump formation, one has to constrain the activity of the network. The same is true in the case of smoothly saturating and binary networks [2], when the highest activity level they can achieve is above the maximum activity of the units in the stored pattern.

In the present paper we discuss in detail the binary neural network model, giving a complete analytical derivation of its properties. We explain the minimal conditions for the formation of spatially asymmetric states in a recurrent network with metrically organized connectivity.

We compare the results for three different topologies of the network and show interesting multibump formation in the case of the SW-like topology. Further extension of the analysis to the information carried by the network reveals that the retrieved states can be well recovered by a limited amount of information. We show that the stability of the network's retrieval when starting from the bump is larger than the stability when starting even from the whole pattern.
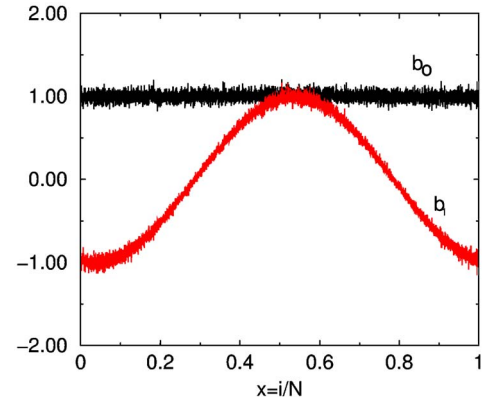

FIG. 1. (Color online) The components of the first and second eigenvectors of the connectivity matrix for $N=6400$, $c=320/N$, $\lambda_1=319.8$, and $\lambda_2=285.4$. The first eigernvector has constant components, the second one sinelike components.

This could be useful in practical retrieval tasks with minimal information transmitted.

The paper is organized as follows: In Sec. II and using replica theory, we present our detailed mean-field analysis in the case of distance-dependent connectivities. We derive the equations for the corresponding order parameters (OP's) for finite and zero temperatures. In Sec. III we present the results of the simulations, done for the case of different topologies of the network, and we compare them with the analytical results. In Sec. IV we focus on the stability of the phase diagram of the network and especially on the dependence of the SAS region on the parameters of the model. We discuss the behavior of the critical storage capacity when bump formation is present, as well as its effect on the information capacity of the network. The conclusion, drawn in the last part, Sec. V, shows that only the asymmetry between the learning and retrieval states is sufficient to observe spatially asymmetric states.

II. ANALYTICAL ANALYSIS

For the analytical analysis of the SAS states, we consider the decomposition of the connectivity matrix c_{ij} by its eigenvectors $a_i^{(k)}$:

$$c_{ij} = \sum_k \lambda_k a_i^{(k)} a_j^{(k)}, \quad \sum_i a_i^{(k)} a_i^{(l)} = \delta_{kl},$$

where λ_k are the corresponding (positive) eigenvalues. For convenience we denote $b_i^k \equiv a_i^{(k)}\sqrt{\lambda_k}$, having

$$c_{ij} = \sum_k b_i^k b_j^k.$$

We will assume that $a_i^{(k)}$ are ordered by their eigenvalues in decreasing order—e.g., for $k>l \Rightarrow \lambda_k \leq \lambda_l$. To get some intuition of what $a_j^{(k)}$ look like, we plot in Fig. 1 the first two eigenvectors.

For a wide variety of connectivities, the first three eigenvectors approximate

$$a_k^{(0)} = \sqrt{1/N}, \qquad (3)$$

$$a_k^{(1)} = \sqrt{2/N} \cos(2\pi|k - k_0|/N), \qquad (4)$$

and

$$a_k^{(2)} = \sqrt{2/N} \sin(2\pi|k - k_0|/N). \qquad (5)$$

Following the classical analysis of Amit *et al.* [15], we use Bogolyubov's method of quasiaverages [16] to take into account a finite number of overlaps that condense macroscopically. To this aim we introduce an external field, conjugate to a finite number of patterns $\{\xi_i^\nu\}$, $\nu = 1, 2, \ldots, s$, adding to the Hamiltonian the term $H_h = \sum_{\nu=1}^s h^\nu \sum_i \xi_i^\nu S_i$.

In order to impose some asymmetry in the neural network's states, we also add the term $H_a = NR(\sum_i S_i/N - a)$.

The whole Hamiltonian we are studying is

$$H = \frac{1}{cN} \sum_{ij\mu} S_i \xi_i^\mu c_{ij} \xi_j^\mu S_j + \sum_{\nu=1}^s h^\nu \sum_i \xi_i^\nu S_i + NR\left(\sum_i S_i/N - a\right). \qquad (6)$$

By using the "replica method" [17], for the averaged free energy per spin we get

$$f = -\lim_{n\to 0} \lim_{N\to\infty} \frac{1}{\beta nN}(\langle\langle Z^n \rangle\rangle - 1), \qquad (7)$$

where $\langle\langle \cdots \rangle\rangle$ stands for the average over the pattern distribution $P(\xi_i^\mu)$, n is the number of replicas, which are later taken to zero, and β is the inverse temperature.

The replicated partition function is

$$\langle\langle Z^n \rangle\rangle = \left\langle\left\langle \text{Tr}_{S^\rho} \exp\left[\frac{\beta}{2cN} \sum_{ij\mu\rho} (\xi_i^\mu S_i^\rho) c_{ij}(\xi_j^\mu S_j^\rho) - \frac{1}{2c}\beta pn \right.\right.\right.$$
$$\left.\left.\left. + \beta \sum_\nu h^\nu \sum_{i,\rho} \xi_i^\nu S_i^\rho - \sum_\rho \beta NR\left(\sum_i S_i^\rho/N - a\right)\right]\right\rangle\right\rangle. \qquad (8)$$

The decomposition of the sites, by using an expansion of the connectivity matrix c_{ij} over its eigenvalues λ^l, $l = 1, \ldots, M$, and its eigenvectors a_i^l, gives

$$\langle\langle Z^n \rangle\rangle = e^{-\beta pn/2c}\left\langle\left\langle \text{Tr}_{S^\rho} \exp\left[\frac{\beta}{2cN} \sum_{\mu\rho l} \sum_{ij} (\xi_i^\mu S_i^\rho b_i^l)(\xi_j^\mu S_j^\rho b_j^l)\right.\right.\right.$$
$$\left.\left.\left. + \beta \sum_\nu h^\nu \sum_{i\rho} \xi_i^\nu S_i^\rho - \sum_\rho \beta RN\left(\sum_i S_i^\rho/N - a\right)\right]\right\rangle\right\rangle. \qquad (9)$$

Introducing the variables $m_{\rho l}^\mu$ at each replica ρ, each configuration, and each eigenvalue, we get

$$\langle\langle Z^n \rangle\rangle = e^{-\beta pn/2c+\beta RaN}\left\langle\left\langle \text{Tr}_{S^\rho} \int \prod_{\mu l\rho} \frac{dm_l^\mu}{\sqrt{2\pi}}\right.\right.$$

$$\times \exp \beta cN\left(-\frac{1}{2}\sum_{\mu\rho l}(m_{\rho l}^\mu)^2 + \sum_{\mu\rho l} m_{\rho l}^\mu \frac{1}{cN}\sum_i \xi_i^\mu S_i^\rho b_i^l\right)$$

$$\times \exp \beta cN\left(-\frac{1}{2}\sum_{\nu\rho l}(m_{\rho l}^\nu)^2 + \sum_{\nu\rho l} m_{\rho l}^\nu \frac{1}{cN}\sum_i \xi_i^\nu S_i^\rho b_i^l\right.$$

$$\left.\left.\left. + h^\nu \frac{1}{N}\sum_i (\xi_i^\nu S_i^\rho + RS_i^\rho)\right)\right\rangle\right\rangle, \qquad (10)$$

where we have split the sums over the first s patterns and the remaining (infinite) p-s ones.

The "condensed" order parameters $m_{\rho l}^\nu$ have a clear physical meaning as an overlap between the pattern ν and the state of the neuron in the replica ρ of the system, modulated by the lth eigenvalue of the connectivity matrix:

$$m_{\rho l}^\nu = \frac{1}{cN}\sum_i \xi_i^\nu S_i^\rho b_i^l. \qquad (11)$$

In the present investigation, the main physical hypothesis is that only a finite number k of order parameters $m_{\rho,l}^\nu$, $l = 1, \ldots, k$, are macroscopically different from zero. As we will see later in this paper the simulations also confirm this hypothesis. This avoids the introduction of an infinite number of parameters, dependent on the state of the system, which would make impossible the analytical solution of the problem.

After taking the averages over the pattern, for the first term we get

$$I = \exp \beta\left(-\frac{1}{2}\sum_{\mu\rho l}(m_{\rho l}^\mu)^2 + \frac{\beta(1-a^2)}{2cN}\sum_{\rho\sigma lki\mu} m_{\rho l}^\mu m_{\sigma k}^\mu S_i^\rho S_i^\sigma b_i^l b_i^k\right). \qquad (12)$$

The integration of the last expression over the OP $m_{\rho l}^\mu$ gives

$$\int \prod_{\mu\rho l} \frac{dm_{\rho l}^\mu}{\sqrt{2\pi}} I = \int \prod_{\rho\sigma lk} dq_{\rho\sigma}^{lk} \exp\left(-\frac{p}{2}\text{Tr}\ln[A_{\rho\sigma}^{lk}]\right)\prod_{\rho\sigma lk} \delta\left(q_{\rho\sigma}^{lk}\right.$$
$$\left. - \frac{1}{cN}\sum_i S_i^\rho S_i^\sigma b_i^l b_i^k\right). \qquad (13)$$

The expression within the δ function can be regarded as a definition of the OP $q_{\rho\sigma}^{lk}$.

Now let us suppose that the order parameter $q_{\rho\sigma}^{lk}$ can split as a product of two terms: one that only depends on the replica's and the eigenvalue's indexes and another one that introduces the spatial dependence of the distribution of the eigenvalues and eigenvectors: $q_{\rho\sigma}^{lk} \equiv q_{\rho\sigma}^l \sum_i b_i^l b_i^k/cN = q_{\rho\sigma}^l \delta_{lk}(\lambda_l/cN)$. In this way we can keep only one index for the OP $q_{\rho\sigma}^l$ when taking into account the spatial distribution of the matrix of interactions.

Introducing the parameter $r_{\rho\sigma}^l$, conjugate to $q_{\rho\sigma}^l$, for the last expression we obtain

$$\int \prod_{\mu\rho l} \frac{dm^{\mu}_{\rho l}}{\sqrt{2\pi}} I = \int \prod_{\rho\sigma l} dq^l_{\rho\sigma} \prod_{\rho\sigma l} dr^l_{\rho\sigma}$$

$$\times \exp\left(-\frac{p}{2} \text{Tr} \ln[A^l_{\rho\sigma}]\right)$$

$$\times \exp cN\left(-\frac{1}{2}\alpha\beta^2(1-a^2)\sum_{\rho\sigma l} r^l_{\rho\sigma} q^l_{\rho\sigma}\right.$$

$$+ \frac{1}{2cN}\alpha\beta^2(1-a^2)\sum_{i\rho\sigma l} r^l_{\rho\sigma} S^{\rho}_i S^{\sigma}_i b^l_i b^l_i\right), \quad (14)$$

where the parameter $\alpha \equiv p/N$ is the storage capacity of the network.

The matrix $A^l_{\rho\sigma}$ is

$$A^l_{\rho\sigma} = \delta_{\rho\sigma}(1 - \beta(1-a^2)\mu_l(1-q_l)) - \beta(1-a^2)q_l\mu_l \quad (15)$$

and $\mu_l = \lambda_l/cN$.

For the replicated partition function $\langle\langle Z^n \rangle\rangle$, after taking the limit $h^{\nu} \to 0$, we obtain

$$\langle\langle Z^n \rangle\rangle = e^{-\beta pn(1-a^2)/2c + \beta RanN} \int \prod_{\nu} dm^{\nu}_{\rho l} \int \prod_{\rho\sigma l} dq^{lk}_{\rho\sigma} dr^l_{\rho\sigma} \exp cN\left(-\frac{\beta}{2}\sum_{\nu\rho l}(m^{\nu}_{\rho l})^2 - \frac{1}{2}\text{Tr}\ln[A^l_{\rho\sigma}] - \frac{1}{2}\alpha\beta^2(1-a^2)\right.$$

$$\times \sum_{\rho\neq\sigma,l} r^l_{\rho\sigma} q^l_{\rho\sigma}\right)\left\langle\left\langle \text{Tr}_{S^{\rho}} \exp cN\left[\frac{1}{2cN}\alpha\beta^2\sum_{i\rho\sigma lk} r^l_{\rho\sigma} S^{\rho}_i S^{\sigma}_i b^l_i b^l_i + \beta\sum_{\nu\rho l} m^{\nu}_{\rho l}\frac{1}{cN}\sum_i \xi^{\nu}_i S^{\rho}_i b^l_i + \beta R\sum_i S^{\rho}_i\right]\right\rangle\right\rangle. \quad (16)$$

Supposing a replica symmetry (RS) ansatz—i.e., $m^{\nu}_{\rho} = m^{\nu}$ for any replica index ρ and $q^l_{\rho\sigma} = q^l$, $r^l_{\rho\sigma} = r_l$ for $\rho \neq \sigma$—for the free energy, Eq. (7), we obtain

$$f = \frac{\alpha(1-a^2)}{2c} + Ra + \frac{\alpha}{2\beta n}\text{Tr}\ln[A^{lk}] + \frac{1}{2}\sum_{\nu l}(m^{\nu}_l)^2 - \frac{\alpha\beta(1-a^2)}{2}\sum_{lk} r_l q^l$$

$$- \frac{1}{n\beta}\left\langle\left\langle \ln \text{Tr}_{S^{\rho}} \exp\left[\frac{1}{2cN}\alpha\beta^2(1-a^2)\sum_{i\rho\sigma l} r^l_{\rho\sigma}\overline{S^{\rho}_i S^{\sigma}_i b^l_i b^l_i} + \beta\sum_{\nu\rho l} m^{\nu}_l \frac{1}{cN}\sum_i \xi^{\nu}_i S^{\rho}_i b^l_i + \beta R\sum_i S^{\rho}_i\right]\right\rangle\right\rangle. \quad (17)$$

The overbar line in the last expression has its clear physical meaning as an average over the spatial distribution $\overline{(\cdot)}_i = (1/cN)\Sigma_i(\cdot)$.

As a next step, let us suppose that the average over a finite number of patterns, ξ^{ν}, can be self-averaged [15]. In our case this is expressed by the decomposition $\overline{S^{\rho}_i S^{\sigma}_i b^l_i b^l_i} \approx S^{\rho}S^{\sigma}\overline{(b^l_i)^2}$, which is reasonable to assume as a first approximation by taking into account the effect of the spatial distribution. This approximation permits us to complete the analytical analysis.

After taking the trace $\text{Tr}_{S^{\rho}}$ and the limit $n \to 0$ in Eq. (17) and using the saddle-point method, we end up with the following expression for the free energy:

$$f = \frac{1}{2c}\alpha(1-a^2) + \frac{1}{2}\sum_k(m_k)^2 - \frac{\alpha\beta(1-a^2)}{2}\sum_k r_k q_k + \frac{\alpha\beta(1-a^2)}{2}\sum_k \mu_k r_k + \frac{\alpha}{2\beta}\sum_k \{\ln[1 - \beta(1-a^2)\mu_k + \beta(1-a^2)q_k]$$

$$- \beta(1-a^2)q_k[1 - \beta(1-a^2)\mu_k + \beta(1-a^2)q_k]^{-1}\} - \frac{1}{\beta}\int \frac{dze^{-z^2/2}}{\sqrt{2\pi}} \overline{\ln 2 \cosh \beta\left(z\sqrt{\alpha(1-a^2)\sum_l r_l b^l_i b^l_i} + \sum_l m_l\xi_i b^l_i + Rb^0_i\right)}. \quad (18)$$

The equations for the OP r_k, m_k, and q_k are, respectively,

$$r_k = \frac{q_k(1-a^2)}{[1 - \beta(1-a^2)(\mu_k - q_k)]^2}, \quad (19)$$

$$m_k = \int \frac{dze^{-z^2/2}}{\sqrt{2\pi}} \overline{\xi_i b^k_i \tanh \beta\left(z\sqrt{\alpha(1-a^2)\sum_l r_l b^l_i b^l_i} + \sum_l m_l\xi_i b^l_i + Rb^0_i\right)}, \quad (20)$$

and

$$q_k = \int \frac{dz e^{-z^2/2}}{\sqrt{2\pi}} \overline{(b_i^k)^2 \tanh^2 \beta\left(z\sqrt{\alpha(1-a^2)\sum_l r_l b_i^l b_i^l + \sum_l m_l \xi_i b_i^l + R b_i^0}\right)}. \tag{21}$$

The following analysis refers to the case $T=0$. Keeping $C_k \equiv \beta(\mu_k - q_k)$ finite and limiting the above system only to the first two m_k's, Eqs. (19)–(21) read

$$m_0 = \frac{1-a^2}{4\pi} \int_{-\pi}^{\pi} g(\phi) d\phi, \tag{22}$$

$$m_1 = \sqrt{2\mu_1} \frac{1-a^2}{4\pi} \int_{-\pi}^{\pi} g(\phi) \sin \phi d\phi, \tag{23}$$

$$C_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} g_c(\phi) d\phi, \tag{24}$$

$$C_1 = \frac{\mu_1}{\pi} \int_{-\pi}^{\pi} g_c(\phi) \sin^2 \phi d\phi, \tag{25}$$

$$r_k = \frac{\mu_k(1-a^2)}{[1-(1-a^2)C_k]^2}. \tag{26}$$

Here,

$$g(\phi) = \mathrm{erf}(x_1) + \mathrm{erf}(x_2), \tag{27}$$

$$g_c(\phi) = [(1+a)e^{-(x_1)^2} + (1-a)e^{-(x_2)^2}]/[\sqrt{\pi}y], \tag{28}$$

$$x_1 = [(1-a)(m_0 + m_1\sqrt{2\mu_1} \sin \phi) + R]/y, \tag{29}$$

$$x_2 = [(1+a)(m_0 + m_1\sqrt{2\mu_1} \sin \phi) - R]/y, \tag{30}$$

$$y = \sqrt{2\alpha(1-a^2)[r_0 + 2\mu_1(r_1 - 1 + a^2)\sin^2 \phi]}. \tag{31}$$

For values of the order parameters $m_1 = m_2 = \cdots = 0$ and $R = 0$, we obtain a result similar to that of Canning and Gardner [18]. The classical result of Amit *et al.* [15] is obtained when $\mu_1 = 0$.

The result of the numerical solution of Eqs. (22)–(26) is shown in Figs. 2(c) and 2(d), where we have rescaled the order parameters m_0 and m_1 to belong to the interval [0, 1], instead of the interval [0, 1−a], for any sparsity a. The sharp
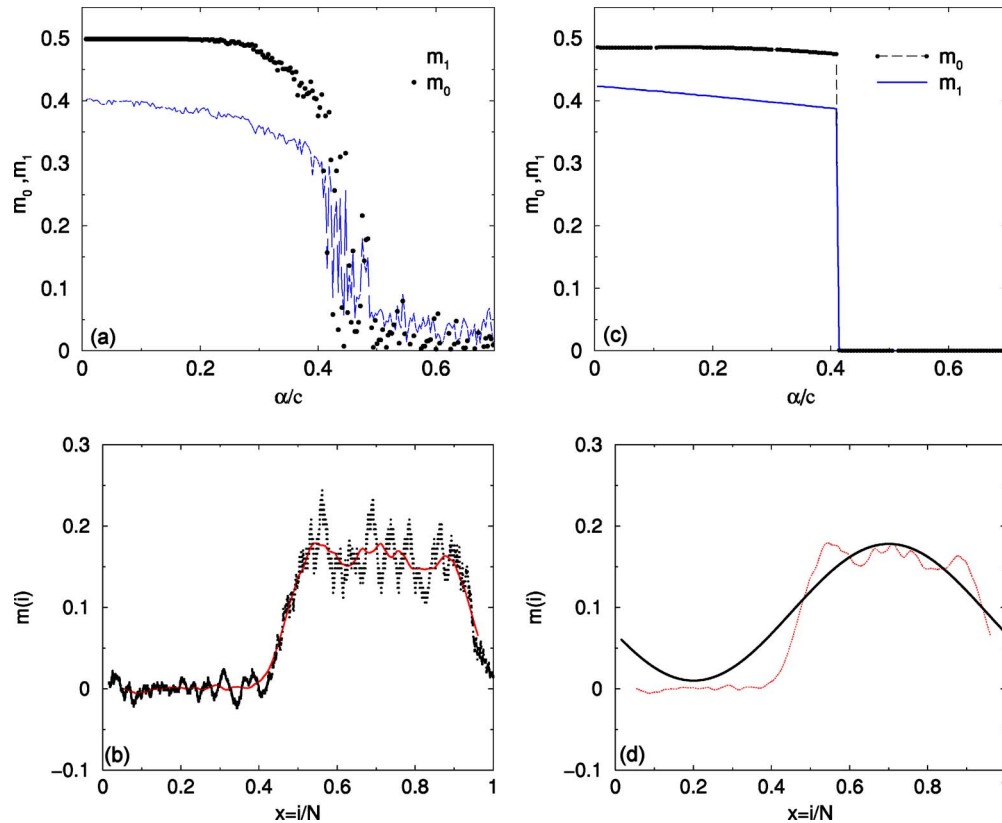

FIG. 2. (Color online) Simulation (a) and computation of m_0, m_1 according to Eqs. (22)–(26) (c). The sparsity of the code is $a = 0.8$ and $R = 0.57$. The form of the bump, represented by its local overlap, Eq. (33), according to the simulation (b) and the theory, (d) bold line. Parameters of the simulation: $p/(cN) = 0.1$, $N = 6400$, $c = 0.05$, $\sigma_x = 500$, and $m(i)$ smoothed with lengths 100 and 500; $\mu_1 = 0.95$.
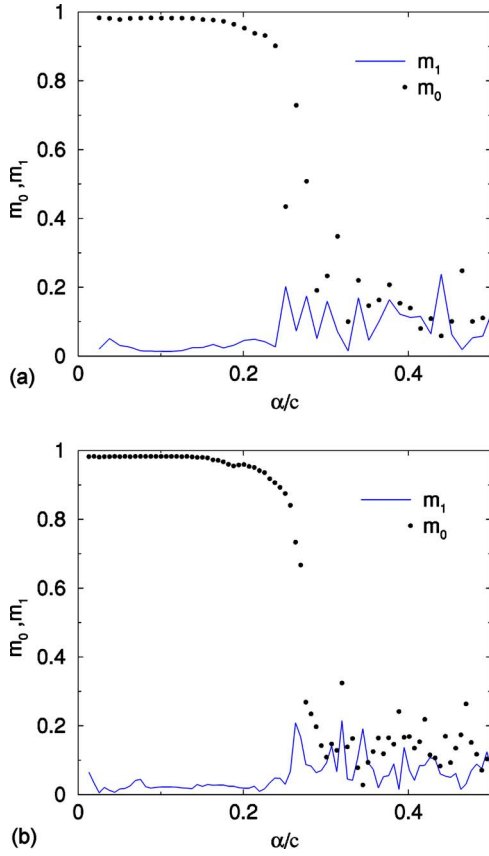
FIG. 3. (Color online) Binary attractor networks with the same distribution of the pattern and the retrieval activity ($R=0$). The overlap m and the power of its first Fourier transform are chosen as a measure of the existence of SAS. The panel (a) is for $N=6400$, $c=80/N$, $\sigma_x=100$, and $a=0.8$, the panel (b) is for $N=6400$, $c=160/N$, $\sigma_x=200$, and $a=0.8$. None of the networks present SAS. The same is true for the sparse code, different dilution, and other topologies such as, for example, those defined by Eqs. (32), (34), and (35).

bound of the phase transition is a result of taking into account just two terms of m_k, $k=0,1$, and the lack of finite-size effects in the thermodynamic limit. The corresponding behavior of m_0 and m_1, obtained by simulation, is presented in the top panels of Figs. 2(a) and 2(b). The good correspondence between the numerical result and the simulations confirms our hypothesis of the finite number of macroscopic OP $m_l \neq 0$. The physical reason for this suggestion lies in the fact that the bump formations are compact structures with a large size and therefore the low-frequency components—e.g., those corresponding to large eigenvalues of the matrix of connectivities—are responsible for the observed behavior.

III. SIMULATIONS

We performed simulations for a network in a form of a circular ring, with a distance between the neurons i and j,

$$|i-j| \equiv \min(i-j+N \bmod N, j-i+N \bmod N),$$

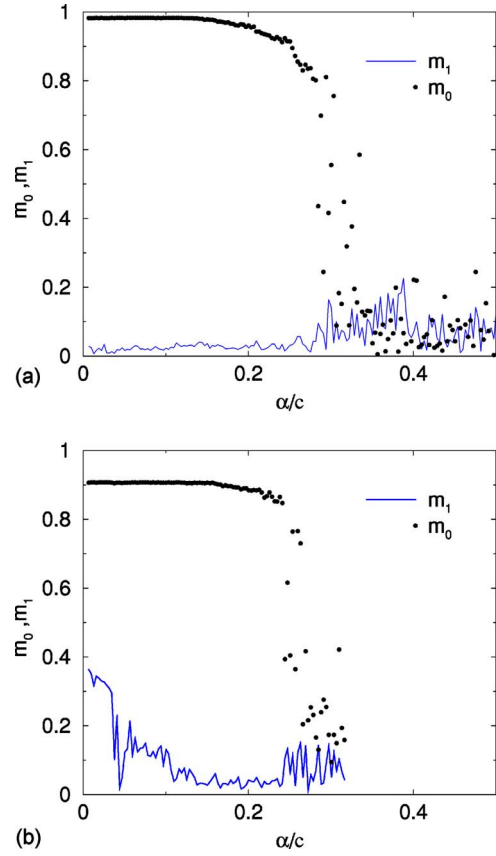using three different topologies, explained in this section.

FIG. 4. (Color online) Spatial asymmetry observed by its first Fourier component power m_1 with symmetric code ($a=0$). The panel (a) shows the network with $N=6400$, $c=0.05$, $\sigma_x=500$, and $R=0$. The panel (b) shows the network with $N=6400$, $c=0.05$, $\sigma_x=500$, and $R=0.5$. SAS is observed only when $R\neq 0$.

To compare with the results of Ref. [1], we work with a typical connectivity distance $\sigma_x N$. This Gaussian-like distribution defines the first topology we are studying:

$$P(c_{ij}=1) = c\left[\frac{1}{\sqrt{2\pi}\sigma_x N}e^{-(|i-j|/N)^2/2\sigma_x^2} + p_0\right], \quad (32)$$

where p_0 is chosen to normalize the expression in the brackets. When σ_x is small enough, then spatial asymmetry is expected.

The dynamics of the network at time $t+1$ and temperature $T=0$ for our case is

$$S_i(t+1) = \mathrm{sgn}\left(\frac{1}{N}\sum_j \sum_\mu \xi_i^\mu \xi_j^\mu c_{ij} S_j(t) - T_h\right),$$

where T_h is the threshold of the system, which is in general nonzero, due to the extra energy term H_a.

If the retrieval state corresponding to the pattern ξ_i^0 is S_i, the mean overlap is $m_0=\Sigma_i\xi_i^0 S_i/N$ and the local overlap at site i is

$$m(i) \equiv \xi_i^0 S_i. \quad (33)$$

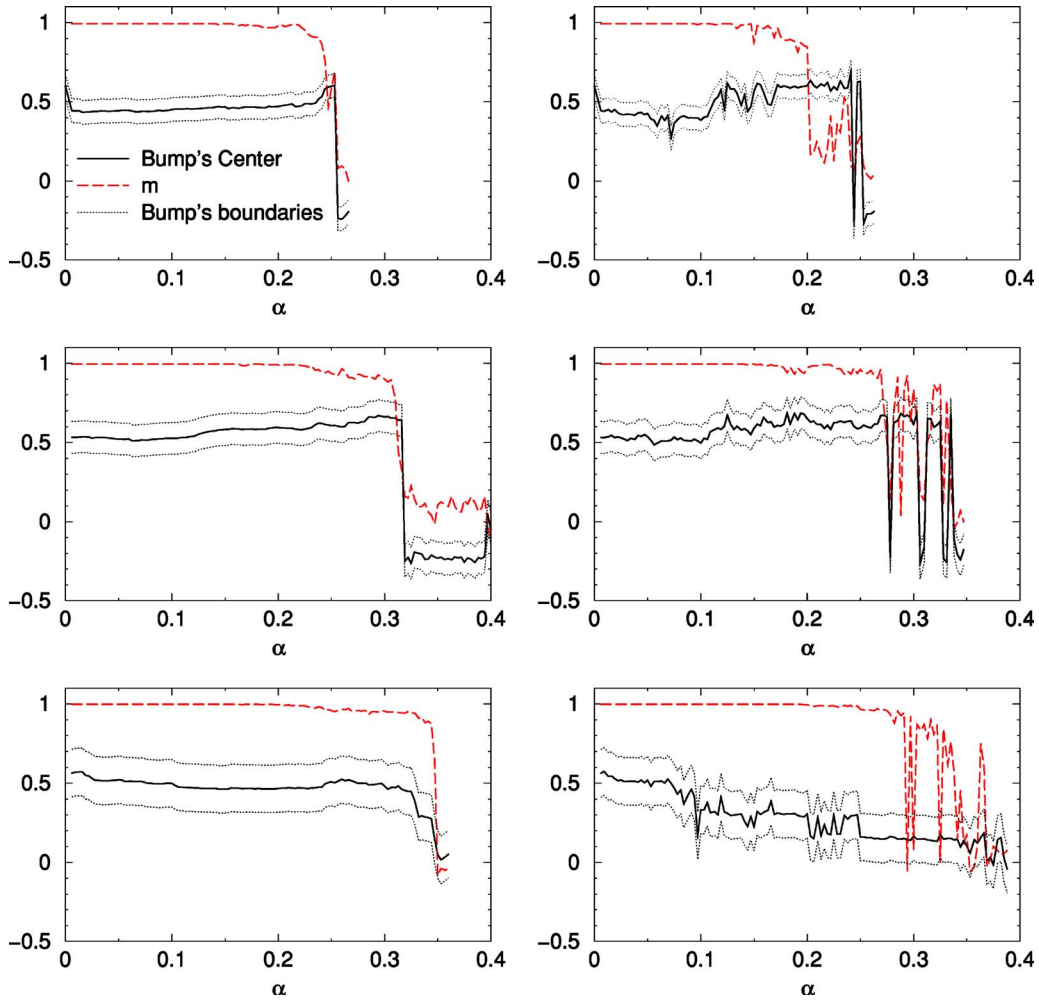In the case when S_i follows a single sine wave, the ideal measure of spatial asymmetry would be

FIG. 5. (Color online) The dynamics of the bump formations. On the vertical axis the center and boundaries of the bump, as well as the asymptotic value of m, are given as a function of the load α for three different initial conditions $m_0=0.15$, $m_0=1$ [(a) and (d)], $m_0=0.2$, $m_0=1$ [(b) and (e)], and finally $m_0=0.3$, $m_0=1$ [(c) and (f)]. More stable behavior is observed when starting from the bump [(a), (b), (c)] than starting from uniform initial condition (the pattern itself) [(d), (e), (f)]. The parameters of the simulations are $N=6400$, $c=0.05$, and $\sigma_x=500$.

$$m_1 = \frac{1}{N}\left|\sum_k \xi_k^0 S_k e^{2\pi i k/N}\right|,$$

which corresponds to the theoretically derived OP m_1 from Eqs. (11) and (20).

Simulations with more sharp localized connectivity

$$P(c_{ij}=1) \propto \frac{1-b\cos\varphi}{1-2b\cos\varphi+b^2}, \qquad (34)$$

with $\varphi \equiv 2\pi|i-j|/N$ and b being some parameter, show similar results. This connectivity defines the second type of topology we are studying. It has the advantage that the eigenvectors of the connectivity matrix are cosine waves and the eigenvalues are known. The results of the simulations for different σ and a are shown in Figs. 2–4. These two topologies, Eqs. (32) and (34), give very similar results.

If $R=0$, no asymmetry can be observed for any σ_x, up to the level of the network fragmentation (Fig. 3). No difference between asymmetric and symmetric connectivity is ob-

servable for any connectivity $c < 0.05$ and any of the topologies tested.

The sparse code increases the SAS effects, Fig. 4(b) vs Fig. 2(a), but SAS cannot be observed for any sparsity a if the proportion of the firing neurons is kept equal to a—i.e., when the retrieval and memorized states have the same level of activity. However, even without sparsity ($a=0$) the bump states can be observed, Fig. 4(b).

One could expect that the bump formations actually will deteriorate the performance of the neural networks, because the bump is localized in a limited part of the network and this effectively shrinks the active part of the network. However, it results in not being so. Namely, if we start the retrieval process from the bump $m_0 < 1$, fixing R in such a way that a bump will be produced, this retrieval is better than the retrieval achieved, starting from a totally overlapping pattern $m_0=1$.

In the case of one-dimensional networks given with a Gaussian-like distribution of the connectivity on a circular ring, we observe the behavior given in Fig. 5.
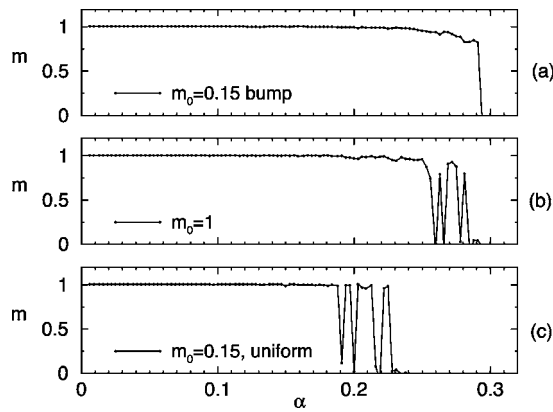
FIG. 6. The behavior of the overlap for three values of the initial conditions: starting from the bump formation with $m_0=0.15$ (a), from the pattern itself (uniform distribution $m_0=1$) (b), and from a uniform distribution with $m_0=0.15$ (c). A clear advantage of the first initial condition is observed with respect to the the size of the region of retrieval. The parameters of the simulations are $N=6400$, $c=0.05$, and $\sigma_x=500$.

In Fig. 5 we observe a pronounced stability of the position of the center and the boundaries of the bump when using the bump initial condition, while in the case of using the whole pattern as an initial condition, this is no longer true. In the last case, one observes a shrinking of the region where the SAS state is observed, given by the rapid decrease of the order parameter (red dashed line), which occurs for smaller
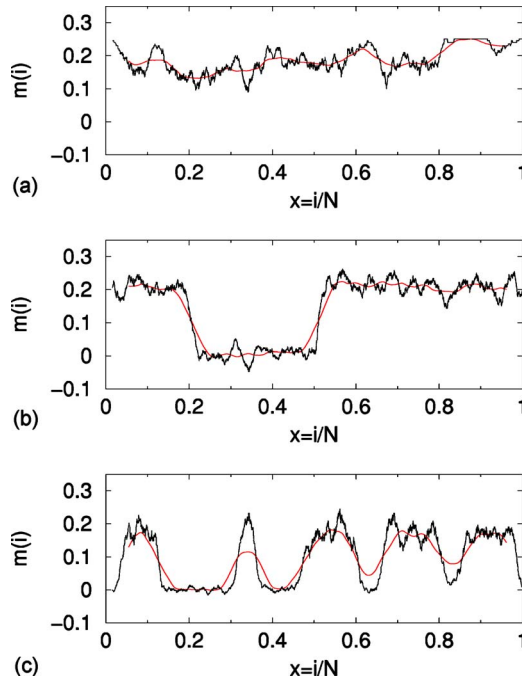

FIG. 7. (Color online) Smoothed local activity for small-world connectivity. In panel (a) no bumpiness is present, but significant fluctuations are ($m_1=0.1m_0$). In panel (b) a single bump is presented. And the last panel (c) shows multiple bump activity. $N=6400$, $c=0.05$, and $a=0.6$(a), $a=0.7$(b), and $a=0.8$(c), rewire rate $\omega=0.06$, $R=0.5$. The same effect can be achieved by changing (c,ω).
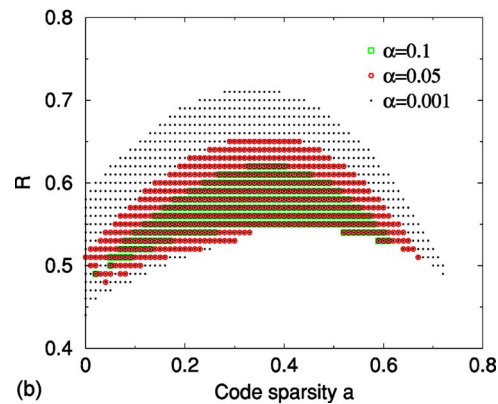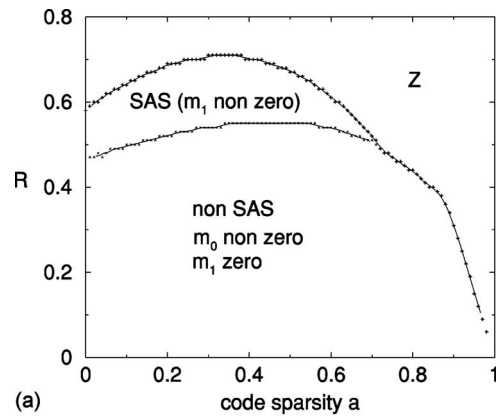

FIG. 8. (Color online) Phase diagram R versus a for $\alpha=0.001$ ($\alpha/c=0.02$). The SAS region, with $m_0 \neq 0$, $m_1 \neq 0$ is relatively large and decreases by increasing the load (bottom). Note that not all of the values of α are feasible with arbitrary mean connectivity of the network c.

values of the load compared to the case of initial bump attractor.

This behavior is repeated for three different values of the bump initial conditions $m_0=0.15$, $m_0=0.2$, and $m_0=0.3$, given in Figs. 5(a)–5(c). The corresponding behaviors when starting from the whole pattern are given in Figs. 5(d)–5(f).

Figure 5 also shows that the difference between the two behaviors is smaller when the size of the initial bump is smaller. For example, when an initial bump condition is $m_0=0.3$, then the region of stability of the solutions extends up to a value of the load $\alpha=0.35$, while it decreases up to $\alpha=0.29$ for the uniform initial condition. For $m_0=0.15$, these values are $\alpha=0.25$ and $\alpha=0.2$, respectively.

The last result clearly points out the role of bump formation on the stability during retrieval.

The comparison of the overlap in the case of initial conditions corresponding to a bump of size $m_0=0.15$, uniform distribution $m_0=1$, and uniform initial condition with $m_0=0.15$ is given in Fig. 6.

From Fig. 6 it can be observed that the uniform initial condition with $m_0=0.15$ gives the worst result for the size of the retrieval solution. The overlap shrinks to zero for a load less than $\alpha=0.2$, while this value is $\alpha=0.29$ when starting from the bump. We face again with the conclusion that the bumps are stable formations and enhance the performance of the network.
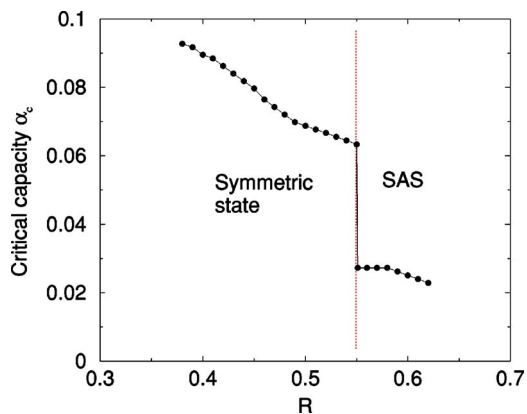
FIG. 9. (Color online) The critical storage capacity α_c for a =0.4. A drop of the capacity is observed on the transition between the normal retrieval state and the state with localized activity.

This result could serve as a basis of the biological relevance of bump formation for the whole stability of the network to learn and to retrieve information efficiently.

Finally, we also performed simulations in the case of the SW topology [9,10] rewiring the rate $\omega \in [0,1]$:

$$P(c_{ij}=1) = (1-\omega)\theta(c-|i-j|/N) + \omega c, \qquad (35)$$

where θ is the theta function. For $R=0$, no SAS behavior is observed.

However, when $R \neq 0$, the simulations show more complex behavior with the appearance of several bump forma-

tions, as well as phases with no bumps but significant m_1, Fig. 7. It seems probable that the multiple bumps are due to the roughness of the SW eigenvalue distribution [19]. This different behavior of the SW topology deserves future attention.

IV. DISCUSSION

The numerical analysis of Eqs. (22)–(26) gives a stable region for the solutions corresponding to bump formation for different values of the load α, the sparsity a, and the retrieval asymmetry parameter R, shown in Fig. 8. As can be seen, the sparsity of the code a enhances the SAS effect, although it is also observed for $a=0$.

As we expected, the asymmetry factor R between the stored and retrieved patterns is very important in order to have spatial asymmetry. The diagram in Fig. 8 shows a clear phase transition with R. For small values of R, the effect (SAS) is not present. If $R>1$, then the only stable state is the trivial Z state, as all nonzero solutions are suppressed. Only for intermediate values of R does the bump solution exist.

When increasing the load, the area corresponding to SAS states shrinks as is shown in Fig. 8(b). This is due primarily to the fact that the critical capacity α_c of the network drops when the system enters into the SAS state. More precisely, the behavior of the critical storage capacity as a function of the asymmetry parameter R, shown in Fig. 9, presents a drastic drop of its value at the transition from homogeneous retrieval (symmetric) state to spatially localized (asymmetric state). Effectively, only the fraction of the network in the
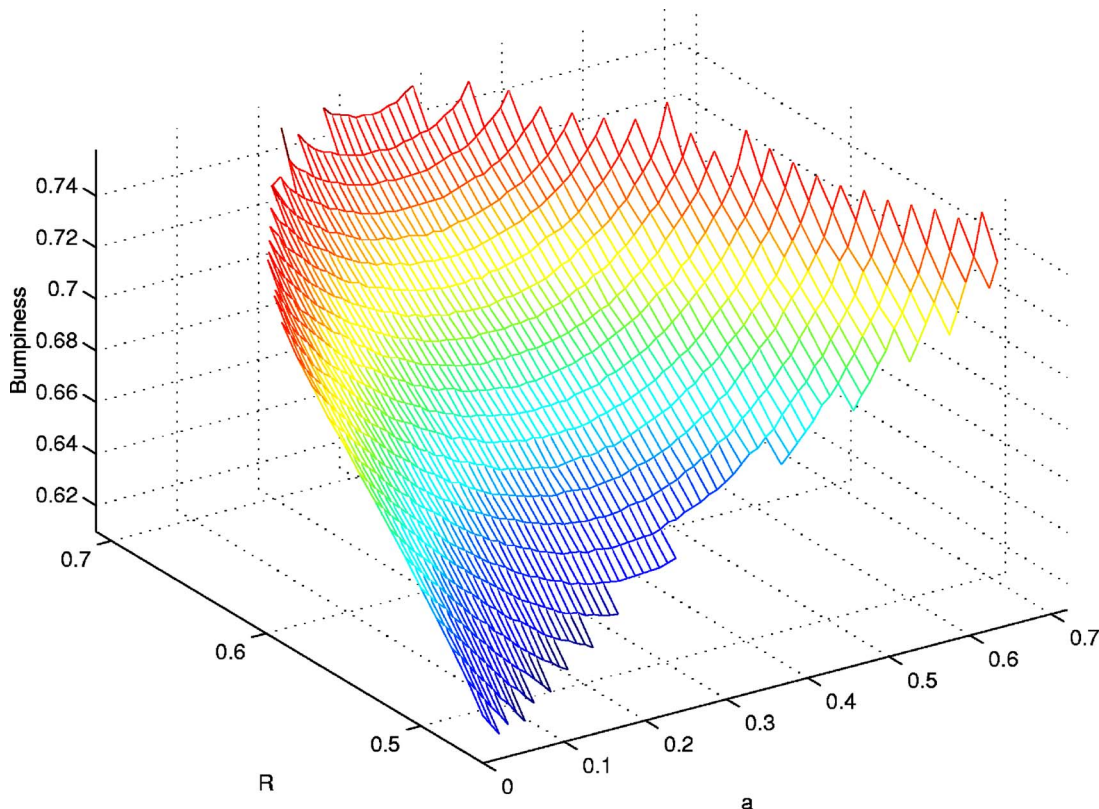

FIG. 10. (Color online) The bumpiness B of the network for different values of R and a.

bump can be excited and the storage capacity drops proportionally to the size of the bump. For the values of the sparsity of the code $a=0.4$, where the effect of the asymmetry parameter is maximal, the decrease of the critical capacity is approximately twofold.

We have also investigated the effect of the parameters R and a on bump formation. For this, we introduce a parameter that can give a criterion for the bumpiness as the relative ratio between the two order parameters m_0 and m_1, given by Eqs. (22)–(26):

$$B = \sqrt{\frac{m_1^2}{m_0^2 + m_1^2}}.$$

The behavior of B is presented in Fig. 10. It is seen that the simultaneous increase of both R and a inside the area where the SAS behavior is observed favors bumpiness in the binary network. This behavior is due to the stronger localization of the activity as the extra term H_a favors states with lower total activity. From the other side, increasing a makes the code sparser, which favors a metrical organization of both patterns and neurons.

Finally we have investigated the information capacity of the network. For this aim we calculated the corresponding expression for the mutual information [20] in the case of order parameters m_0 and m_1:

$$I_M(m_0, m_1, \alpha) = \frac{\alpha(m_0 + 1)}{2} \ln(m_0 + 1)$$

$$+ \frac{\alpha |m_1|}{\pi} F\left[\left\{\frac{1}{2}, 1, 1\right\}, \left\{\frac{3}{2}, \frac{3}{2}\right\}, \frac{m_1^2}{(1 + m_0)^2}\right]$$

$$+ \frac{\alpha(m_0 + 1)}{2} \ln\left[\frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{m_1^2}{(1 + m_0)^2}}\right].$$

$$(36)$$

Here $F[\cdot]$ is the hyper geometric function [21]. The first term is the normal expression for the mutual information (the sparsity of the code eliminates the symmetry between m and $-m$). The next two lines of the equation can be regarded as a correction to the usual expression due to OP $m_1 \neq 0$.

With the aim of Eqs. (22)–(26) and the above expression, for the information capacity I_M/cN we obtained the behavior presented in Fig. 11, which has been also reproduced by simulations. This behavior is nontrivial and shows a well-pronounced maximum for intermediate values of the sparsity of the code a. The corresponding behavior of the order parameters m_0 and m_1, presented in Fig. 11(b), shows that inside the SAS region, their values are significantly different from zero, but the information capacity is very small although nonzero. If, on the one hand, the information capacity I_M/cN measures the number of bits needed to represent the state and if, on the other hand, the overlap parameters (m_0, m_1) measure the degree to which one learned state can be recovered, then having small information capacity and large OP m_0, m_1 means that the original state can be recovered well with a very small amount of information. This result could be useful for effective retrieval of patterns by using a very small amount of information, especially in com-
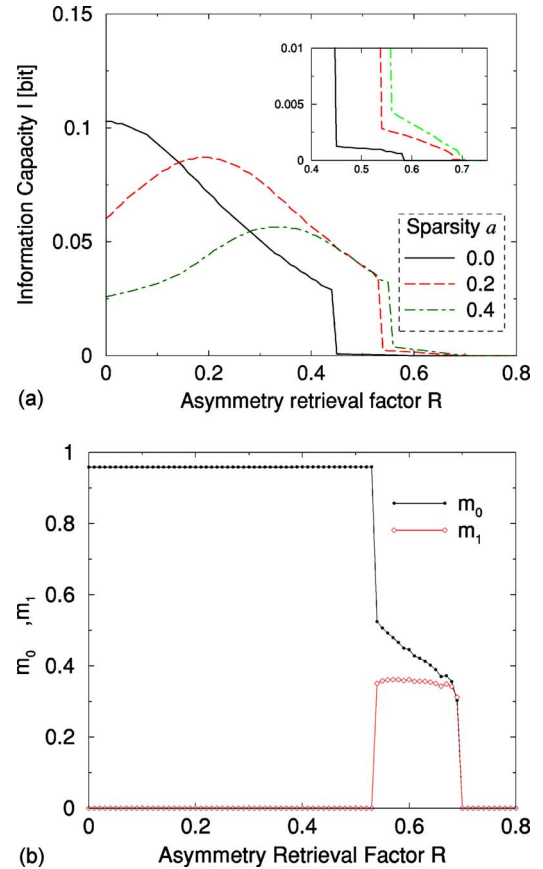

(a)

(b)

FIG. 11. (Color online) (a) The information capacity versus the asymmetry retrieval factor R for different values of the sparsity a. (b) m_0, m_1 for $a=0.2$. The inset in (a) shows the behavior of the information capacity near the point where it sharply decreases to zero.

bination with the observed improvement of the stability of the network, provided that the retrieval starts from the bump.

Similar results have been reported in [2] with respect to the increase of the computational power of the structured network. The last is expressed by the ability of the network to retrieve several patterns, each in a different location, whose combination could form the global combinatorially large pattern.

Apart from the advantage of localized states in terms of the minimal cost in transferring information, such structures could have important biological relevance for the cortical modules of the mammalian brain, which are characterized by a geometrical distribution of synaptic connections. Recent experiments of the activity of a macaque temporal cortex [22] have demonstrated that the receptive fields of the visually evoked activity patterns are restricted up to threefold when several patterns are present together. The last result seems to be consistent with the picture of localized activity in a neural network with metric organization of the synaptic activities.

V. CONCLUSION

In this paper we have studied the spatially dependent activity in a binary neural network model when retrieval and

learning states have a certain degree of asymmetry. We have shown that neither sparsity of the code nor asymmetry of the connections is necessary to achieve bump solutions for the activity of the network, but rather different symmetry between the retrieval and learning states.

The extension of the classical method of Amit *et al.* [15] in the case of spatially dependent connectivity has permitted, within some approximations, the derivation of the above results analytically. The good correspondence with simulations, tested for different topologies of the network, is also discussed in the paper.

We have shown that the region of spatially asymmetric states is relatively large in the parameter space (R,a), and we have measured quantitatively how the bumpiness of the network depends on these two parameters.

A drop of the critical storage capacity is observed in the transition from symmetric to SAS states, which is due to the fact that effectively only the fraction of the network in the bump can be excited and the storage capacity drops proportionally to the size of the bump.

When the recovering of the pattern starts from a state corresponding to the bump formation, the network behaves in a better way in terms of stability and capacity, compared to the case of uniform initial conditions. This happens even if the overlap with the uniform initial conditions is significantly larger. The latter result could argue for the biological relevance of bump formation for effective retrieval of information.

Finally we have discussed the behavior of the information capacity versus the asymmetry retrieval factor R and we have discussed its sharp decrease in the parametric region, where bump formation is present.

[1] Y. Roudi and A. Treves, JSTAT **1**, P07010 (2004).

[2] Y. Roudi and A. Treves, e-print cond-mat/0505349.

[3] K. Koroutchev and E. Korutcheva (unpublished).

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

[5] A. Anishchenko, E. Bienenstock, and A. Treves, e-print q-bio.NC/0502003.

[6] J. Rubin and A. Bose, Network Comput. Neural Syst. **15**, 133 (2004).

[7] N. Brunel, Cereb. Cortex **13**, 1151 (2003).

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

[9] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[10] D. J. Watts, *Small Worlds: The Dynamics of Networks Between Order and Randomness, Princeton Review in Complexity* (Princeton University Press, Princeton, 1999).

[11] M. Tsodyks and M. Feigel'man, Europhys. Lett. **6**, 101 (1988).

[12] D. Bollé, G. Shim, B. Vinck, and V. A. Zagrebnov, J. Stat. Phys. **14**, 565 (1994).

[13] J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).

[14] D. Hebb, *The Organization of Behavior: A Neurophysiological Theory* (Wiley, New York, 1949).

[15] D. Amit, H. Gutfreund, and H. Sompolinsky, Ann. Phys. (N.Y.) **173**, 30 (1987).

[16] N. N. Bogolyubov, Physica (Amsterdam) **26**, Suppl., S1 (1960).

[17] M. Mézard, G. Parisi, and M.-A. Virasoro, *Spin-Glass Theory and Beyond* (World Scientific, Singapore, 1987).

[18] A. Canning and E. Gardner, J. Phys. A **21**, 3275 (1988).

[19] I. Farkas, I. Derenyi, A. L. Barabasi, and T. Vicsek, Phys. Rev. E **64**, 026704 (2001).

[20] R. Blahut, *Principle and Practice of Information Theory* (Addison-Wesley, Cambridge, MA, 1988).

[21] *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun, Natl. Bur. Stand. Appl. Math. Ser. 55 (U.S. GPO, Washington, D.C., 1972).

[22] E. Rolls, N. Aggelopoulos, and F. Zheng, J. Neurosci. **23**, 339 (2003).