Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: Phase diagram

R. Monasson and S. Rosay

Laboratoire de Physique Théorique de l'ENS, CNRS & UPMC, 24 rue Lhomond, 75005 Paris, France

We study the stable phases of an attractor neural network model, with binary units, for hippocampal place cells encoding one-dimensional (1D) or 2D spatial maps or environments. Different maps correspond to random allocations (permutations) of the place fields. Based on replica calculations we show that, below critical levels for the noise in the neural response and for the number of environments, the network activity is spatially localized in one environment. For high noise and loads the network activity extends over space, either uniformly or with spatial heterogeneities due to the crosstalk between the maps, and memory of environments is lost. Remarkably the spatially localized regime is very robust against the neural noise until it reaches its critical level. Numerical simulations are in excellent quantitative agreement with our theoretical predictions.

I. INTRODUCTION

Understanding the representation of space by the brain is a long-lasting question, which has been addressed using many varied methods. This includes memory of places, localization of one's position, mental exploration, and planning of forecoming trajectories. During the last decades, the use of microelectrodes allowing single cell recordings has revolutionized our knowledge of neural networks. In 1971, O'Keefe and Dostrovsky [1] recorded neural activity in the hippocampus of rats and discovered the existence of place cells, which fire only when the animal is located in a certain position in space (called place field). This discovery suggested that the hippocampus could be the support for space representation or a "cognitive map." Since then, many experimental and theoretical studies have been carried out on the hippocampus, making it one of the most, if not the most, studied parts of the brain [2].

The properties of place cells, their conditions of formation, and the sensory and behavioral correlates of place fields have been investigated experimentally [3–5]. Place fields have the striking property to appear as randomly distributed, independently of the neurons' locations in the neural tissue: two neighboring neurons can have very distant place fields. Furthermore, several "environments" or "maps" can be learned, and a given neuron can have place fields in several environments, which are apparently randomly assigned, a property called remapping [6]. Place fields are controlled primarily by visual cues but the activity of place cells persists in the dark [7] and is also driven by self-motion signals, that is, "path integration" [8]. More recently, the discovery of grid cells [9,10] in the entorhinal cortex (that feeds input into the hippocampus) opened a new way in the comprehension of a complex system of interacting brain regions [11]. Many theoretical models have been proposed to account for these experimental results. Beyond the comprehension of the hippocampus itself, the motivation is to reach more insights about the functional principles of the brain [2].

Experiments show that the hippocampus is able to learn, memorize, and retrieve spatial maps. The massive intrinsic connectivity in hippocampal CA3 led to the hypothesis of an attractor neural network [12–14] where memorized activity patterns are the attractors of the dynamics, such as in the celebrated Hopfield model [15]. In the Hopfield model it is assumed that the patterns are additively stored in the synapses, through a Hebbian learning mechanism. A deep and quantitative understanding of the Hopfield model was made possible by the use of the statistical physics theory of mean-field spin glasses [16,17]. In the case of the rodent hippocampus, the memorized items are space manifolds called environments [6]. Neural network models for place cells have been proposed, in particular by Battaglia and Treves, who carried out a mean-field calculation of the storage performance of a network with linear threshold units [18]. Recently Hopfield proposed a similar model for mental exploration in a network with adaptation [19]. However, the crosstalk between the different environments encoded in the network, and the transitions that can occur between them as observed experimentally [20] remain poorly understood.

Here, we propose a model of interacting binary units and study the different regimes of activity in the presence of neural noise. The model is defined in Sec. II. We study the case where multiple environments are memorized in Sec. III, and derive the different regimes of activity of the network under given conditions of neural noise and memory load in Sec. IV. The phase diagram of the system is computed in Sec. V and compared to numerical simulations. We show that an activity of the network that is locally spatialized in one of the stored maps, as observed experimentally, is the stable state of the network provided that both the neural noise and the memory load are small enough. For high noise and/or loads the the activity is delocalized in all environments, either uniformly over space or with spatial heterogeneities controlled by the crosstalk between environments (glassy phase). We finally discuss the value of the parameters (Sec. V C) and the hypothesis of the model (Sec. VI) compared to previous works. The study of the free energy landscape and of the dynamics of the model will be addressed in a companion publication [21].

II. MODEL

A. Definition

The N place cells are modeled by interacting binary units σ_i equal to 0 or 1 corresponding to, respectively, silent and active states. We suppose that, after learning of the environment and

random allocation of place fields, each place cell preferentially fires when the animal is located in an environment-specific location in the D-dimensional space, defining its place field. For simplicity, space is assumed to be a segment of length N for $D = 1$, and a square of edge length $\sqrt{N}$ in $D = 2$, with periodic boundary conditions. The N centers of the place fields are assumed to be perfectly located on a D-dimensional regular grid: two contiguous centers are at unit distance from each other. This simplification allows us to concentrate on the interference between the stored spatial maps as the only source of structural noise.

Let d_c be the extension of a place field, that is, the maximal distance between locations in space recognized by the same place cell. Place cells whose place fields overlap, and, therefore, spike simultaneously as the animal wanders in the environment, are assumed to strengthen their synaptic connections. Calling d_{ij} the distance between the place field centers of cells i,j in the environment we assume that the reinforcement process results in the production of excitatory synaptic couplings given by

$$J_{ij}^0 = \begin{cases} \frac{1}{N} & \text{if} \quad d_{ij} \leqslant d_c, \\ 0 & \text{if} \quad d_{ij} > d_c \end{cases}. \tag{1}$$

The fact that all environments are equivalent (there is no privileged permutation) is basic to our theory. We choose the place extension d_c such that each cell i is connected to the same number of other cells j, independently of the space dimension D. Let $w\,N$ be this number: $w(\ll 1)$ is the fraction of the neural population any neuron is coupled to. Hence, $d_c = \frac{w}{2}N$ in dimension $D = 1$, and $d_c = \sqrt{\frac{wN}{\pi}}$ in dimension $D = 2$. The $\frac{1}{N}$ scale factor is such that the total contribution to the local field received by a place cell is finite when the number of cells N is sent to infinity. Note that we assume here that the environment is perfectly explored: couplings depend on the distance d_{ij} only, and not on the particular sequence of positions occupied by the animal during the time spent in the environment. The case of partial, nonhomogeneous explorations was studied in [14]. Couplings defined by prescription (1) are symmetric, and only reflect the local structure of the environment.

Each time the rodent explores a new environment a remapping of the place fields takes place. Let L be the number of explored environments, in addition to the environment above (hereafter called reference environment). We assume that the remapping is represented by a random permutation of the N place-cell indices associated to the place fields in the reference environment, denoted by $\ell = 0$. Let π^ℓ be the permutation corresponding to remapping number ℓ, where $\ell = 1, \dots, L$ is the index of the environment. In environment ℓ cells i,j interact if the distance $d_{\pi^\ell(i)\pi^\ell(j)}$ is smaller than d_c, and do not interact at larger distances. An obvious modification of (1) defines the coupling matrix J^ℓ corresponding to environment ℓ. We finally assume that all environments contribute equally and additively to the total synaptic matrix,

$$J_{ij} = \sum_{\ell=0}^{L} J_{ij}^\ell = J_{ij}^0 + \sum_{\ell=1}^{L} J_{\pi^\ell(i)\pi^\ell(j)}^0. \tag{2}$$

For the sake of a better understanding, we consider an example of a matrix J in the very simple case $N = 6$, $w = \frac{2}{6}$,
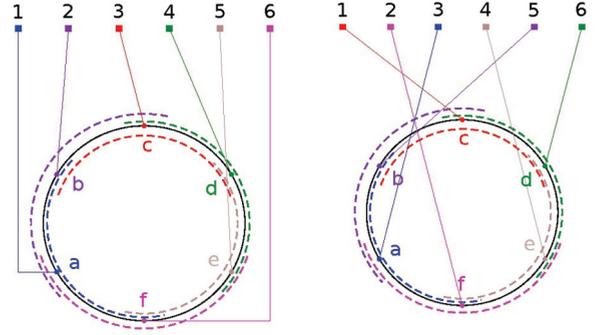

FIG. 1. (Color online) Example of remapping of the place field centers of $N = 6$ neurons (denoted by indices 1,...,6) in two different one-dimensional (1D) environments with periodic boundary conditions and $w = \frac{2}{6}$. Place fields in each environment are represented by colored dashed lines, place field centers are denoted by letters a,...,f.

$L + 1 = 2$, and $D = 1$, illustrated in Fig. 1. For the reference environment the coupling matrix is

$$J^0 = \frac{1}{6} \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{3}$$

For another environment obtained through the random permutation $\pi = (3,6,1,5,2,4)$ we obtain the coupling matrix

$$J^1 = \frac{1}{6} \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \tag{4}$$

The total coupling matrix is therefore

$$J = J^0 + J^1 = \frac{1}{6} \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 2 \\ 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 2 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}. \tag{5}$$

In addition to pyramidal cells, the network contains long-range, inhibitory interneurons whose activity is modeled by a global inhibition on place cells. We assume that the main effect of inhibition is to fix the total neural activity. We introduce the parameter f to denote the fraction of active cells:

$$\sum_{i=1}^{N} \sigma_i = f\,N. \tag{6}$$

Once the coupling matrix J_{ij} (2) and the constraint over the global activity (6) are defined the probability of a neural

activity configuration $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ is assumed to be

$$P_J(\boldsymbol{\sigma}) = \frac{1}{Z_J(T)} \exp(-E_J[\boldsymbol{\sigma}]/T), \qquad (7)$$

where the "energy" of the configuration reads

$$E_J[\boldsymbol{\sigma}] = -\sum_{i<j} J_{ij}\, \sigma_i\, \sigma_j, \qquad (8)$$

and the partition function is

$$Z_J(T) = \sum_{\boldsymbol{\sigma}\,\text{with constraint (6)}} \exp(-E_J[\boldsymbol{\sigma}]/T). \qquad (9)$$

Parameter T, which plays the role of temperature in statistical mechanics, fixes the amount of noise in the model. Large values of T corresponds to essentially flat distributions over the neural configuration space. Low T concentrate the probability distribution P_J around the configurations with lowest energies E_J. In all numerical computations hereafter we will take the parameters values $w = 0.05$ and $f = 0.1$, except in Sec. V C where the effect of those values on the results will be discussed.

B. Case of a single environment

Our model is an extension of the Hopfield model [15] to the case of space-dependent interactions. Despite this additional complexity in the model it remains exactly solvable in the infinite N limit, due to the long-range nature of the interactions [22].

We start by considering the case of a single environment, for which the coupling matrix is given by (1). To lighten notations we consider the $D = 1$ case. In the large N limit, a continuous approach can be introduced by defining the locally coarse-grained activity

$$\rho(x) \equiv \lim_{\epsilon \to 0} \lim_{N \to \infty} \frac{1}{\epsilon N} \sum_{(x-\epsilon/2)N \leqslant i < (x+\epsilon/2)N} \langle \sigma_i \rangle_J, \qquad (10)$$

where $\langle . \rangle_J$ denotes the average over distribution P_J (7). Note that the order of limits is important for the local average to be correctly defined. Due to the presence of periodic boundary conditions we choose $x \in [-\frac{1}{2}; \frac{1}{2}]$. The density of activity $\rho(x)$ is found upon minimization of the free energy functional

$$\begin{aligned} \mathcal{F}(\{\rho(x)\}) = &-\frac{1}{2} \int dx\, dy\; \rho(x) J_w(x-y) \rho(y) \\ &+ T \int dx \{\rho(x) \ln \rho(x) \\ &+ [1 - \rho(x)] \ln[1 - \rho(x)]\}, \end{aligned} \qquad (11)$$

where $J_w(u) = 1$ if $|u| < \frac{w}{2}$, and 0 otherwise. The minimum is taken over the activity densities fulfilling

$$\int dx\, \rho(x) = f. \qquad (12)$$

All integrals run over the $[-\frac{1}{2}; \frac{1}{2}]$ interval.

The minimization equation for $\rho(x)$ can be written as

$$\rho(x) = \frac{1}{1 + e^{-\mu(x)/T}}, \qquad (13)$$

$$\mu(x) = \int dy\, J_w(x-y) \rho(y) + \lambda, \qquad (14)$$

where $\mu(x)$ plays the role of a chemical potential, and the constant λ is chosen to satisfy (12). We discuss the solutions of these equations in the following sections. Note that the free energy per site,

$$F(T) = \lim_{N \to \infty} -\frac{T}{N} \ln Z_J(T), \qquad (15)$$

is simply given by the value of the free-energy functional $\mathcal{F}$ in its minimum $\rho(x)$, solution of (13) and (14).

C. Relationship with rate models

Neurons are often described by their firing rate, i.e., the short-term average of the number of spikes they emit. A straightforward relationship can be drawn with binary models [23]. The current incoming onto neuron i evolves according to

$$\tau \frac{dI_i}{dt} = -I_i + \sum_j J_{ij} g(I_j). \qquad (16)$$

Here, $g(x)$ is the characteristic function expressing the firing rate of the neuron as a function of the current. It is a sigmoidal function, running between 0 and 1 (saturation of the postsynaptic neuron at high currents), and J_{ij} includes both the positive coupling J^0 (1) between neighboring cells, and a constant, global inhibition contribution J^I, whose value is chosen to enforce (6). The dynamical equation admits a stationary state, implicitly defined through

$$I_i = \sum_j J_{ij} g(I_j). \qquad (17)$$

Identifying

$$I_i \to \mu_i, \quad g(I_i) \to \rho_i, \qquad (18)$$

and choosing

$$g(I) = \frac{1}{1 + \exp(-I/T)}, \qquad (19)$$

we observe that Eq. (17) for the stationary currents is identical to Eq. (14) for the chemical potential in the single-environment case. The constant term λ in (14) is related to the constant inhibitory contribution to J through $\lambda = J^I f$. The parameter T fixes the slope of g at the origin.

III. STATISTICAL MECHANICS OF THE MULTIPLE ENVIRONMENT CASE

A. Average over random remappings

In the presence of multiple environments the partition function Z_J becomes a stochastic variable, which depends on the L remappings, or, equivalently, on the L random permutations π^ℓ, with $\ell = 1 \ldots L$. We assume that, in the large N limit, the free energy of the system is self-averaging, i.e., concentrated around the average. To compute the average free energy we need to average the logarithm of $Z_J(T)$ over the random permutations. To do so we use the replica method: we first compute the nth moment of $Z_J(T)$, and then send $n \to 0$. The neural configuration is now a set $\vec{\boldsymbol{\sigma}} = (\boldsymbol{\sigma}^1, \ldots, \boldsymbol{\sigma}^n)$ of $n \times N$ spins σ_i^a, where $i = 1 \ldots N$ is the spin index and $a = 1 \ldots n$ is the replica index. The nth moment of the partition

function reads

$$
\overline{Z_J(T)^n} = \sum_{\vec{\sigma}} \exp\left[\beta \sum_{a=1}^n \sum_{i<j} \left(J_{ij}^0 + \sum_{\ell=1}^L J_{ij}^\ell\right) \sigma_i^a \sigma_j^a\right]
$$

$$
= \sum_{\vec{\sigma}} \exp\left[\beta \sum_{a=1}^n \sum_{i<j} J_{ij}^0 \sigma_i^a \sigma_j^a\right] \Xi(\vec{\sigma})^L, \quad (20)
$$

where $\beta = 1/T$ and the overbar denotes the average over the random remappings. The sum over $\vec{\sigma}$ is restricted to configurations with average activity equal to f (within each replica), and

$$
\Xi(\vec{\sigma}) = \frac{1}{N!} \sum_{\pi^\ell} \exp\left[\beta \sum_{i<j} J_{ij}^0 \sum_{a=1}^n \sigma_{\pi^\ell(i)}^a \sigma_{\pi^\ell(j)}^a\right]. \quad (21)
$$

Here the equivalence between permutations is explicitly exploited. The calculation of the average over the random permutation π^ℓ is not immediate, but can be done exactly in the large N limit. Details are reported in Appendix B. The result is

$$
\ln \Xi(\vec{\sigma}) = N \frac{\beta}{2} n w f^2 - \sum_{\lambda \neq 0} \text{Tr} \ln[\mathbf{Id}_n - \beta\lambda(\mathbf{q} - f^2 \mathbf{1}_n)], \quad (22)
$$

where $\mathbf{Id}_n$ denotes the n-dimensional identity matrix, $\mathbf{q}$ is the overlap matrix with entries

$$
q^{ab} \equiv \frac{1}{N} \sum_j \sigma_j^a \sigma_j^b, \quad (23)
$$

and $\mathbf{1}_n$ is the $n \times n$ matrix whose all entries are equal to 1. The sum in (22) runs over all the nonzero eigenvalues of the matrix J^0. Explicit expressions for those eigenvalues will be given in the next section for the $D = 1$ case, while the two-dimensional case is treated in Appendix A.

A key feature of (22) is that Ξ depends on the spin configuration $\vec{\sigma}$ through the overlaps q^{ab} only. Those overlaps thus play the role of order parameters for the activity in the environment $\ell \geqslant 1$, as does $\rho(x)$ for the environment 0. Calculation of the nth moment of the partition function therefore amounts to estimating the entropy of neural activity configuration $\vec{\sigma}$ at fixed $\{q^{ab}, \rho(x)\}$, which can be done exactly in the $N \to \infty$ limit.

B. Replica-symmetric theory

To perform the $n \to 0$ limit we make use of the replica symmetric ansatz, which assumes that the overlaps q^{ab} take a single value q for replica indices $a \neq b$. The validity of the ansatz will be discussed in Sec. IV. The Edwards-Anderson order parameter q, defined through

$$
q \equiv \frac{1}{N} \sum_{i=1}^N \overline{\langle \sigma_i \rangle_J^2}, \quad (24)
$$

measures the fluctuations of the local spin magnetizations from site to site. Values for q range from f^2 to f. We expect q to be equal to f^2 when the local activity $\langle \sigma_i \rangle_J$ (averaged over

the configurations with distribution P_J) is uniform over space, and to be larger otherwise.

As in the single environment case we define the order parameter $\rho(x)$ as the density of activity around point x in space, see (10),

$$
\rho(x) \equiv \lim_{\epsilon \to 0} \lim_{N \to \infty} \frac{1}{\epsilon N} \sum_{(x-\epsilon/2)N \leqslant i < (x+\epsilon/2)N} \overline{\langle \sigma_i \rangle_J}. \quad (25)
$$

The difference is that, in the multiple environment case, the density $\rho(x)$ appearing in the replica theory is averaged over the environments. Local fluctuations of the density from environment to environment can be calculated [21], but will not be considered here; only global fluctuations, averaged over space, are considered through the order parameter q.

As in the single environment case a chemical potential $\mu(x)$, conjugated to $\rho(x)$, is introduced. In addition, a new order parameter r plays the role of the conjugated force to q, and controls the fluctuations of the spin magnetizations. All order parameters are determined through the optimization of the free energy functional $\mathcal{F}(q, r, \{\rho(x)\}, \{\mu(x)\})$, see Appendix C, whose expression for the $D = 1$ case is given by

$$
\mathcal{F} = \frac{\alpha\beta}{2} r(f - q) - \frac{\alpha}{\beta} \psi(q, \beta) + \int dx \, \mu(x) \rho(x)
$$

$$
- \frac{1}{2} \int dx \int dy \, \rho(x) J_w(x - y) \rho(y)
$$

$$
- \frac{1}{\beta} \int dx \int Dz \ln(1 + e^{\beta z \sqrt{\alpha r} + \beta\mu(x)}), \quad (26)
$$

where $Dz = \exp(-z^2/2)/\sqrt{2\pi}$ is the Gaussian measure, and

$$
\psi(q, \beta) \equiv \sum_{k \geqslant 1} \left[\frac{\beta(q - f^2)\sin(k\pi w)}{k\pi - \beta(f - q)\sin(k\pi w)}\right.
$$

$$
\left. - \ln\left(1 - \frac{\beta(f - q)\sin(k\pi w)}{k\pi}\right)\right]. \quad (27)
$$

Parameter $\alpha \equiv L/N$, hereafter called load, denotes the ratio of the numbers of environments and of cells.

Extremization of the free energy functional leads to the saddle-point equations

$$
r = 2(q - f^2) \sum_{k \geqslant 1} \left[\frac{k\pi}{\sin(k\pi w)} - \beta(f - q)\right]^{-2},
$$

$$
q = \int dx \int Dz \, [1 + e^{-\beta z \sqrt{\alpha r} - \beta\mu(x)}]^{-2},
$$

$$
\rho(x) = \int Dz \, [1 + e^{-\beta z \sqrt{\alpha r} - \beta\mu(x)}]^{-1}, \quad (28)
$$

$$
\mu(x) = \int dy \, J_w(x - y) \rho(y) + \lambda,
$$

where λ is determined to enforce (12). The expression of $\mathcal{F}$ and of the saddle-point equations for the $D = 2$ case can be found in Appendix A.

IV. THE PHASES AND THEIR STABILITY

In both $D = 1$ and 2 dimensions three qualitatively different solutions are found for the extremization equations of $\mathcal{F}$, corresponding to three distinct phases of activity: a

paramagnetic phase in which the activity is uniform over space, a "clumplike" phase in which the activity is localized in one of the stored spatial maps, and a glassy phase where the activity is neither uniform nor coherent with any map. We now discuss the domains of existence and stability of each phase. We are chiefly interested in the clump phase domain, which corresponds to the experimentally observed regime where memorized maps can be retrieved. As usual all expressions given below correspond to the $D = 1$ case, while the case $D = 2$ is treated in Appendix A; all numerical results were obtained for $f = 0.1$, $w = 0.05$.

A. High noise: Paramagnetic phase

At high temperature we expect the activity to be dominated by the noise in the neural dynamics, and to show no spatial localization. The corresponding order parameters are

$$\rho(x) = f, \quad q = f^2 \quad \text{(paramagnetic phase: PM)}.$$

The activity profile is shown in Fig. 2(a). The paramagnetic phase (PM) exists for all values of the control parameters, with corresponding potentials:

$$\mu(x) = T \ln \left(\frac{f}{1 - f} \right), \quad r = 0 \quad \text{(PM)}.$$

We now discuss its stability.

1. Case of a single environment ($\alpha = 0$)

In the single environment case the stability of the paramagnetic solution is determined by computing the Hessian of the free-energy functional $\mathcal{F}$ (11). We find that

$$\frac{\delta^2 \mathcal{F}}{\delta\rho(x)\delta\rho(y)} = \frac{T}{f(1 - f)}\delta(x - y) - J_w(x - y). \quad (29)$$

The solution is stable as long as the Hessian is definite positive.

In the one-dimensional case the most unstable mode corresponds to a spin wave $\delta\rho(x) \propto \sin(2\pi k x)$, with wave number $k = 1$; note that the $k = 0$ mode is forbidden according to condition (12). The instability develops under the spinodal temperature

$$T_{\text{PM}} = f(1 - f) \frac{\sin \pi w}{\pi} \approx 0.0045. \quad (30)$$

T_{PM} and, more generally, all thermodynamic quantities are invariant under the changes $f \to 1 - f$ or/and $w \to 1 - w$,

which simply amount to reverse $\sigma_i \to 1 - \sigma_i$, i.e., to change active spins into holes and vice versa.

In dimension $D = 2$ a similar calculation shows that the first unstable mode is a one-dimensional spin wave along one arbitrary direction in the plane. The corresponding spinodal temperature is

$$T_{\text{PM}}^{2D} = f(1 - f) \sqrt{w} \frac{\sin(\pi \sqrt{w})}{\pi}. \quad (31)$$

2. Case of multiple environments ($\alpha > 0$)

The study of the stability of the PM phase in the multiple environments case is reported in Appendix E 1. As in the single environment case the PM solution is unstable at all temperatures $T < T_{\text{PM}}$ against perturbation of the activity of the type $\delta\rho(x) \propto \sin(2\pi k x)$. In addition coupled fluctuations of λ, r, q may lead to instabilities if T is smaller than $T_{\text{PM}}(\alpha)$, implicitly defined through

$$\sum_{k \geqslant 1} \left[\frac{T_{\text{PM}}(\alpha) k\pi}{f(1 - f) \sin(k\pi w)} - 1 \right]^{-2} = \frac{1}{2\alpha}. \quad (32)$$

The instabilities correspond to the transition to the glassy phase; see Sec. IV C. Note that T_{PM} defined in (30) corresponds with $T_{\text{PM}}(\alpha = 0)$. As a conclusion, in the (α, T) plane, the PM phase is stable in the region $T > T_{\text{PM}}(\alpha)$. This region is sketched in Fig. 3.

B. Moderate noise and load: The clump phase

In experiments place cells exhibit patterns of localized activity where neurons with neighboring place fields are active together. Our modeling reproduces such localized-in-space activity patterns (called "bumps" or "clumps" of activity) at sufficiently low (α, T). The corresponding phase, hereafter referred to as the "clump phase" (CL), is characterized by the order parameters

$$\rho(x) \neq f, \quad q > f^2 \quad \text{(clump phase: CL)}.$$

Correspondingly, the chemical potential $\mu(x)$ will vary over space, and the conjugated force r is strictly positive.

FIG. 3. The paramagnetic (PM) phase is stable in the upper part of the (α, T) plane, defined by $T > T_{\text{PM}}(\alpha)$. The spin glass (SG) phase exists below this line; replica symmetry is broken everywhere in the $T < T_{\text{PM}}(\alpha)$ region.
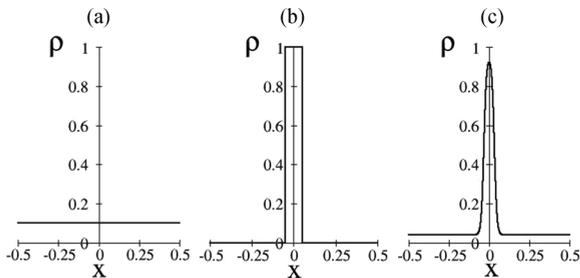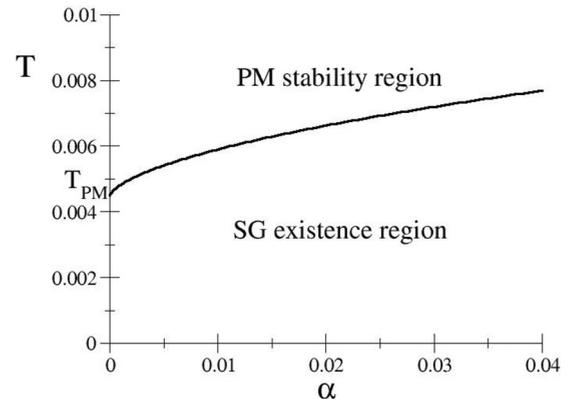

FIG. 2. Average activity $\rho(x)$ in dimension $D = 1$ in the paramagnetic phase (a) and in the clump phase ((b): temperature $T = 0$; (c): temperature $T = 0.0073$) for $\alpha = 0$, computed with $M = 2000$ bins of discretization.

1. Single environment ($\alpha = 0$)

When the temperature T is sent to 0^+, assuming that $f > w$, we find a solution to (13), (14) that is localized in space:

$$\mu(x) = \begin{cases} \frac{w}{2} & \text{if} & |x| < \frac{1}{2}(f-w) \\ \frac{f}{2} - |x| & \text{if} & \frac{1}{2}(f-w) \leqslant |x| < \frac{1}{2}(f+w) , \\ -\frac{w}{2} & \text{if} & |x| \geqslant \frac{1}{2}(f+w) \end{cases}$$

(33)

and

$$\rho(x) \rightarrow \begin{cases} 1 & \text{if} & |x| \leqslant f/2, \\ 0 & \text{if} & |x| > f/2. \end{cases}$$

(34)

Any translation $x \rightarrow x + x_0$ defines another ground state with the same free energy. The activity profile is shown in Fig. 2(b).

At small but finite temperature we have solved Eqs. (13) and (14) numerically by discretizing space with a large number M of bins of width $1/M$, such that Mw and Mf are both much larger than unity. The activity profile is now rounded off by the thermal noise; see Fig. 2(c) for a representative example. Cells far away from the center of the clump are active with some positive probability $< f$. This clump is reminiscent of a liquid phase, surrounded by its vapor. The clump persists up to some critical temperature T_{CL}, e.g., $T_{CL} \simeq 0.008$ for $f = 0.1, w = 0.05$, at which it disappears. The dependence of T_{CL} on f and w will be studied in Sec. VI. Notice that T_{CL} also slightly depends on the number of bins of discretization M as shown in Fig. 4.

The clump phase is also present for $D = 2$. An example of a two-dimensional clump is shown in Fig. 5.

2. Crosstalk between different environments ($\alpha > 0$)

We now look for a solution with localized activity in the first environment, and nonlocalized activity in the other environments. Keeping the temperature T fixed and increasing the load α has the effect of squeezing and lowering the clump (Fig. 6). Once the disorder (random remappings) is averaged out, the clump solution is translationally invariant as in the single environment case. Here we assume that no external input (which would be important for retrieval properties of the network, and would break translation invariance) is present.
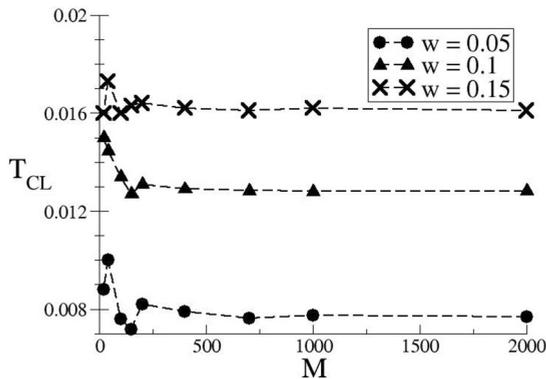

FIG. 5. (Color online) Two-dimensional clump of activity $\rho(x,y)$ for a single environment ($\alpha = 0$) at temperature $T = 0.0055$ computed with $M = 400$.

We have studied the stability of the clump solution against longitudinal and replicon modes. The longitudinal stability domain is found by determining the boundary in the (α, T) plane along which the clump abruptly collapses. This boundary, shown in Fig. 7, can be described as follows:

(i) at small α the clump phase is longitudinally stable for $T < T_{CL}(\alpha)$, a slowly decreasing function of α, which coincides with the temperature T_{CL} found for a single environment when $\alpha \rightarrow 0$;

(ii) at small temperature, the clump phase is longitudinally stable if $\alpha < \alpha_{CL}(T)$, an increasing function of T. We denote α_{CL} its value when $T \rightarrow 0$;

(iii) at intermediate temperatures a weak reentrance is present. The curves $T_{CL}(\alpha)$ and $\alpha_{CL}(T)$ merge at a point where the tangent is vertical and the reentrance begins.

Along the boundary of the clump phase the value of the Edwards-Anderson parameter increases from $q = f^2$ in ($\alpha = 0, T = T_{CL}$) to $q = f$ in ($\alpha = \alpha_{CL}, T = 0$).

Calculation of the stability against replicon modes is detailed in Appendix E 3. We find that the replica-symmetric solution is stable, except in a small region confined to small T and α close to α_{CL}. This result is shown by the dashed line in Fig. 7. It is reminiscent of the results for the "retrieval phase" in the Hopfield model [16].

FIG. 4. Highest temperature at which the clump exists, T_{CL}, as a function of the number M of discretization bins for three values of w. The average activity is $f = 0.1$ and the load vanishes, $\alpha = 0$.
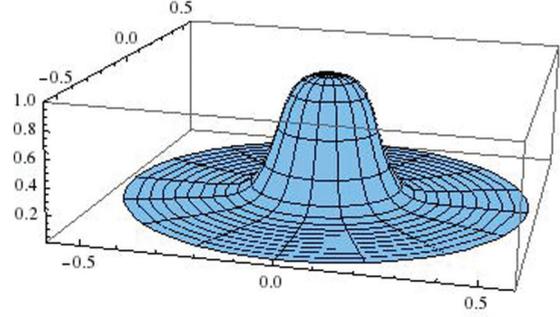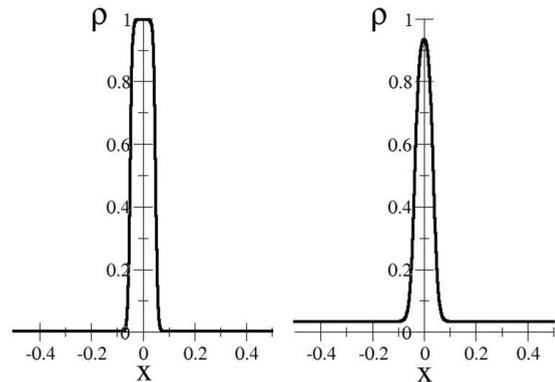

FIG. 6. Effect of the load α on the clump: average activity $\rho(x)$ in dimension $D = 1$ in the clump phase at temperature $T = 0.004$ for $\alpha = 0$ (left) and $\alpha = 0.02$ (right).
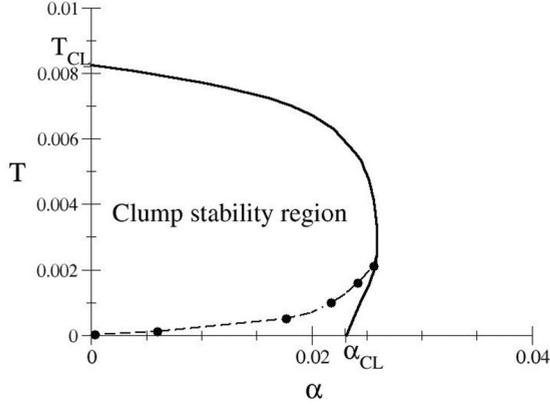
FIG. 7. Domain of stability the clump phase, computed with $M = 200$ bins. Longitudinal and replicon instability lines correspond to, respectively, the full and dashed lines. Due to the computational effort required for the calculation of the replicon eigenvalues, only a few points (black dots) were computed.

C. High load: The glassy phase

At large α the disorder in the interactions is strong enough to magnetize the spins locally, without any coherence with any spatial map. Again, the average of the activity $\langle \sigma_i \rangle_J$ will depend on the realization of the environments, while the average over the environment, $\overline{\langle \sigma_i \rangle_J}$, will be uniform in space and equal to f. In this glassy (SG) phase the order parameters will take values

$$\rho(x) = f, \quad q > f^2 \quad \text{(glassy phase: SG)}.$$

Correspondingly the chemical potential $\mu(x)$ does not depend on x, and $r > 0$.

As reported in Appendix E 2 a glassy solution is found when $T < T_{PM}(\alpha)$, corresponding to the paramagnetic stability line calculated above. Within this region, the SG phase is always stable against clumpiness (localization of the activity). The spin glass phase is unstable against the replicon mode, indicating that replica symmetry is broken, similarly to the spin glass phase in the Hopfield model [16]. Results are summarized in Fig. 3.

V. PHASE DIAGRAM

A. Transitions between phases

Transition lines between the phases described above are determined in the (α, T) plane from the comparison of their free energies:

(i) The clump-paramagnetic transition at high temperature is located slightly below the clump instability line. We denote $T_c(\alpha)$ the corresponding temperature for a given α.

(ii) The clump-glass transition occurs at a load denoted $\alpha_g(T)$ for a given temperature T. Here again, we find a slight reentrance at moderate temperature: $\alpha_g(T)$ is maximal for $T \approx 0.004$. As replica symmetry is broken in the SG phase the true free energy is expected to be higher than the RS value, and the true transition line should be slightly shifted to higher values of α.

(iii) At high α, T there is a second-order phase transition between the PM and the SG phases.

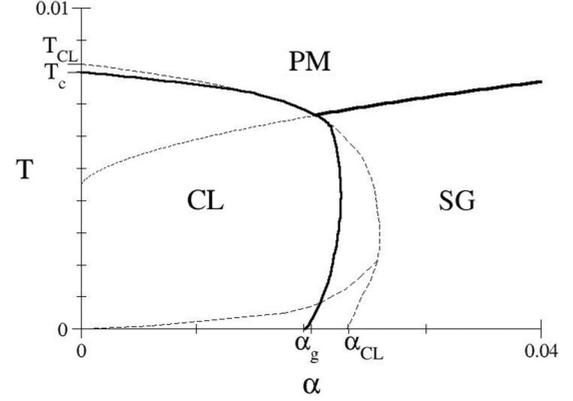FIG. 8. Phase diagram in the (α, T) plane in $D = 1$. Thick lines: transitions between phases. Thin dashed lines reproduce stability regions described above. Critical lines were computed with $M = 200$.

The phase diagram in dimension $D = 1$ is summarized in Fig. 8.

It is interesting to emphasize the differences between this phase diagram and the one of the Hopfield model computed in [16]. In the Hopfield model, the "retrieval" or "ferromagnetic" (FM) phase (which corresponds to our clump phase) has a triangular shape in the (α, T) plane. The temperature at which the FM phase becomes unstable at a given α is smaller than $T_{PM}(\alpha)$. There is no coexistence between the PM and FM phases, and both are separated by the glassy phase. Moreover, for the Hopfield model, $T_{FM}(\alpha)$ is monotonously decreasing so the capacity is maximal at zero temperature [24]. Consequently, it seems that our model of attractor neural network is much more robust to noise than the standard Hopfield model. This can be understood considering the structure of the coupling matrix. In the Hopfield model one pattern defines a single direction in the configuration space; interference with other patterns and dynamical noise may push the activity configuration in the high-dimensional orthogonal subspace, and the memory of the pattern is easily lost. In the present case, on the contrary, one map defines a whole collection of configurations (bumps) centered on different locations, thus the synaptic matrix will make the network converge to one of the attractors, even in the presence of a high level of noise.

When a first-order transition line is crossed the order parameter q is discontinuous. We have computed numerically the value of the Edwards-Anderson parameter at different points and plotted its evolution at the clump-paramagnetic transition at fixed α (Fig. 9) and at the clump-glass transition at fixed T (Fig. 10).

B. Numerical simulations

We have performed Monte Carlo simulations to check our theoretical predictions. The system is initialized with two types of conditions (respectively, uniform and clump configurations). At each time step, two neuron indices i, j are chosen such that $\sigma_i = 1 - \sigma_j$. We then calculate the change in the energy when the two spins are flipped, and accept the flip or not according to Metropolis' rule. As a consequence the activity is kept constant (and equal to fN over the neural
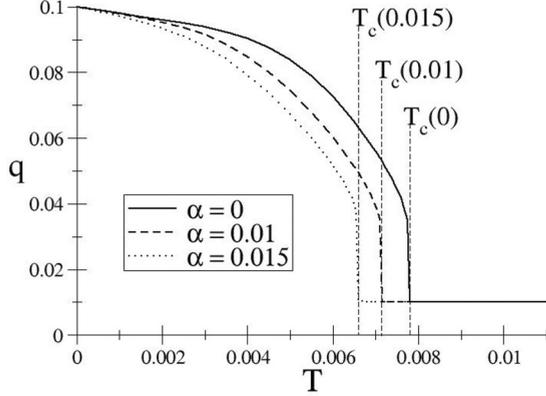
FIG. 9. q as a function of T for fixed α: $\alpha = 0$ (solid line), $\alpha = 0.01$ (dashed line), and $\alpha = 0.015$ (dots), computed with $M = 1000$. A discontinuity is observed at the clump-paramagnetic transition.

population), and the system is guaranteed to reach equilibrium for sufficiently long simulation times.

1. Single environment case

Figure 11 shows the average energy $E(T)$ vs the temperature T, for various sizes N. At high temperature, $E(T) = -\frac{1}{2}f^2w$ as expected in the paramagnetic phase. At low temperature, the shape of the activity clump varies with T, and so does $E(T)$. We find a clear signature of the first order transition as N grows. The critical temperature is in excellent agreement with the theoretical value for T_c.

We plot in Fig. 12 the spin-spin correlation $\langle \sigma_i \sigma_j \rangle$ as a function of the normalized distance, $d = \frac{|i-j|}{N}$:

$$C(d) = \langle \sigma_i \sigma_{i+dN} \rangle. \tag{35}$$

At low temperature, finite size effects are negligible and $C(d)$ is a nontrivial decreasing function of d in the large N limit. At small d, $C(d)$ is of the order of f, and then decreases to a much smaller value over a distance of the order of f. As the location of the clump is arbitrary, we expect its center x_0 to be uniformly distributed over the $[-\frac{1}{2}; \frac{1}{2}]$ interval. The
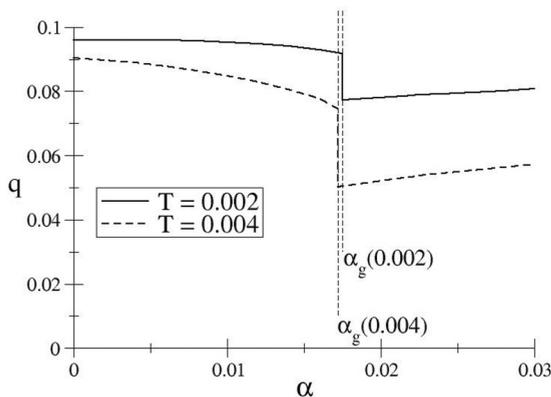

FIG. 10. q as a function of α for fixed temperature: $T = 0.002$ (solid line) and $T = 0.004$ (dashed line), computed with $M = 1000$. A discontinuity is observed at the clump-glass transition.
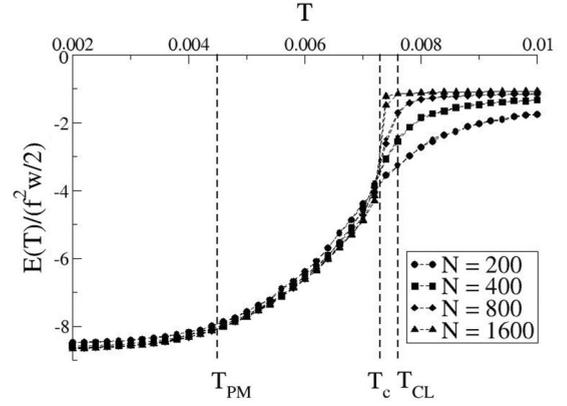

FIG. 11. Average energy for the unidimensional model with a single environment and for increasing sizes N. For each size, we plot the average energy obtained after thermalization for $10\,N$ Monte Carlo steps starting from the uniform and from the clump configurations. Each point is averaged over 1000 simulations.

correlation is therefore given, in the thermodynamic limit, by

$$C(d) = \int dx_0\, \rho(x_0)\, \rho(x_0 + d). \tag{36}$$

At zero temperature, this formula gives $C(d) = f - d$ for $d < f$, $C(d) = 0$ for $d \geqslant f$. At finite temperature, we compute ρ from the extremization equation (13), and plug the value into the right-hand side of (36). The agreement with the correlation $C(d)$ obtained from MC simulations is perfect (Fig. 12).

At high temperature and for finite N, $C(d)$ decreases over a distance $d \simeq \frac{w}{2}$ to the paramagnetic value f^2. When $N \to \infty$, $C(d)$ is uniformly equal to f^2 at all distances $d > 0$. As an additional check of the value of T_c we find that the spin-spin correlation decays quickly with increasing N for $T = 0.0074$, and saturates to a d-dependent value larger than f^2 for $T = 0.0072$ (not shown).

2. Multiple environments

We now report the outcome of Monte Carlo simulations with $L + 1$ environments ($L > 0$), obtained through random permutations of the sites. We have verified numerically the theoretical predictions for $\mu(x)$ (Fig. 13) and r (Fig. 14). This latter quantity can be accessed by measuring the local fields at different positions, $\mu(x) + \lambda + z\sqrt{\alpha r}$. The quenched
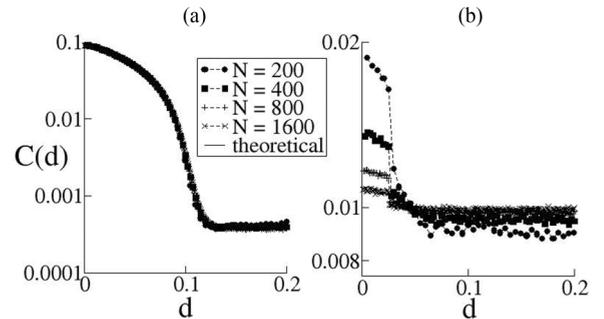

FIG. 12. Correlation $C(d)$ between spins at distance d (35) at low (left) and high (right) temperatures, and for various sizes N. (a) $T = 0.004$, (b) $T = 0.01$. Note the difference of logarithmic scale on the y axis between the two panels.
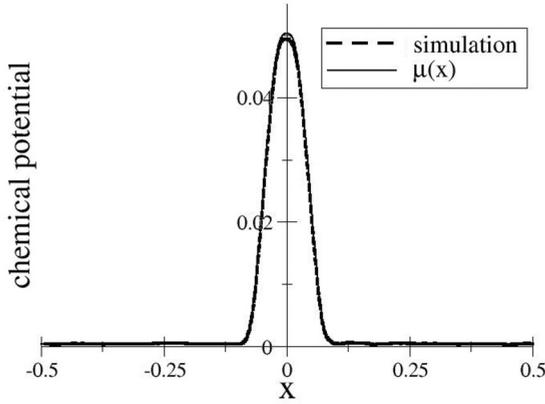
FIG. 13. Chemical potential $\frac{1}{N}\sum_j J_{ij}^0 \sigma_j$ as a function of x for $\alpha = 0.01$ and $T = 0.004$: theoretical prediction $\mu(x)$ (solid line) and result of simulation for one set of environments (dashed line) with $N = 10\,000$, averaged over 100 rounds of $10N$ steps each.

noise on the field comes from the contribution of environments $\ell \geqslant 1$: $z\sqrt{\alpha r}$ is a Gaussian random variable of mean 0 and standard deviation $\sqrt{\alpha r}$ independent of x. In our simulations we have measured the contributions $h_i \equiv \frac{1}{N}\sum_{\ell=1}^{L}\sum_j J_{ij}^{\ell}\sigma_j$ of the environments $\ell \geqslant 1$ to the local fields at different locations. The distribution of the h_i's perfectly agrees with the theoretically predicted Gaussian (inset in Fig. 14).

We have also investigated the behavior of the system for varying levels of noise and load, and compared it to the phase diagram found theoretically. In simulations we have considered the environment ℓ of lowest energy (in which the activity acquires a clumplike shape) and measured its contribution to the energy density, $E^{\ell}[\{\sigma_i\}] = -\frac{1}{N}\sum_{i<j} J_{ij}^{\ell} \sigma_i \sigma_j$. This quantity is compared with the theoretical value $-\frac{1}{2}\int dx\, dy\, \rho(x) J_w(x-y)\rho(y)$.

We have run simulations for different temperatures and numbers of environments, with $N = 2000$ and $N = 5000$ units. After thermalization, the energy of the coherent environment is recorded after 100 rounds of $10N$ Monte Carlo steps each. Results are shown in Figs. 15 and 16.
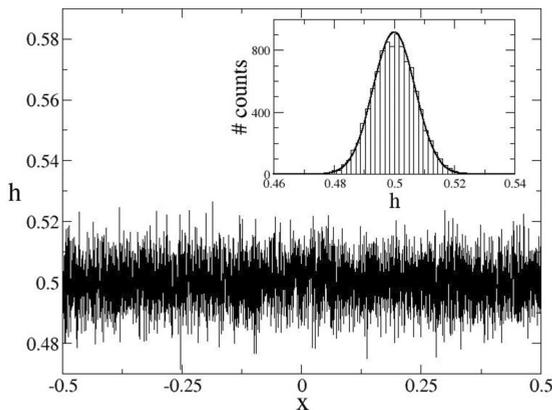

FIG. 14. Contribution h_i of environments $\ell \geqslant 1$ to the local fields as a function of $x = i/N - 0.5$ for the same model as in Fig. 13. Inset: histogram of h_i (rectangles) compared to the Gaussian distribution of mean fLw and standard deviation $\sqrt{\alpha r}$ (solid line). The value $\sqrt{\alpha r} \simeq 6.98 \times 10^{-3}$ was obtained from the resolution of (28).
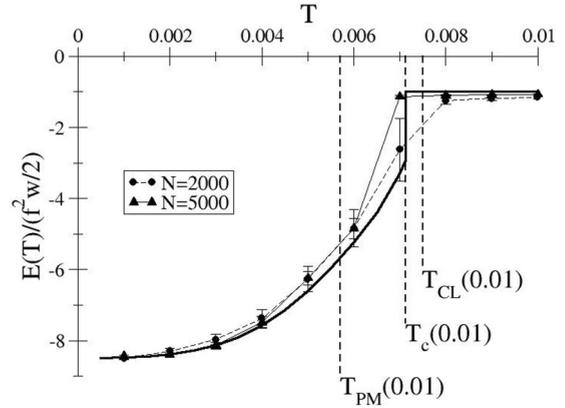

FIG. 15. Density of energy in the environment coherent with the clump for constant $\alpha = 0.01$ (same realization of the disorder): results of Monte Carlo simulations for $N = 2000$ (circles) and $N = 5000$ (triangles) with error bars, compared to theoretical result computed with $M = 1000$ (line).

The agreement with theoretical predictions is very good in the case of the clump-paramagnetic transition (Fig. 15). Concerning the clump-glass transition (Fig. 16), as we mentioned above we expect the transition to occur at larger load, $\alpha_g(T) < \alpha_g^{\mathrm{observed}} < \alpha_{\mathrm{CL}}(T)$, due to the replica-symmetry broken nature of the glass phase. This expectation is corroborated by Fig. 17, which represents the fraction of simulations ending in the glassy phase as a function of α for $T = 0.004$. We have checked that this fraction does not depend on the initial conditions of the simulation. The transition occurs around $\alpha \simeq 0.018 \pm 0.001$ (uncertainty due to long thermalization times in the simulations), while $\alpha_g \simeq 0.0173$ for $T = 0.004$ used in the simulation.

C. Dependence on parameter values

All the numerical computations above were performed with parameter values $w = 0.05$ and $f = 0.1$. To gain insight on the influence of the parameter values on the behavior of the clump phase, we focus on two quantities representing its stability domain, namely α_{CL} and T_{CL}, respectively the
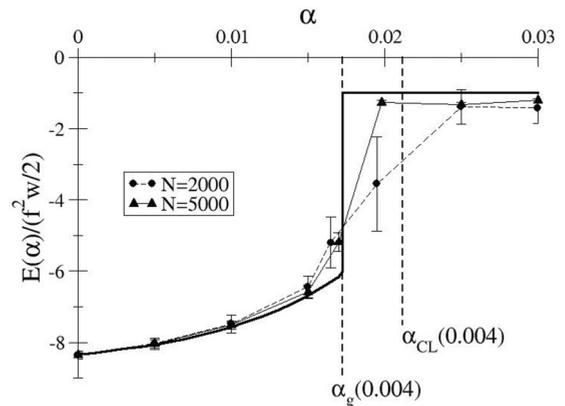

FIG. 16. Density of energy in the environment coherent with the clump for constant $T = 0.004$: results of Monte Carlo simulations for $N = 2000$ (circles) and $N = 5000$ (triangles) with error bars, compared to theoretical result computed with $M = 1000$ (line).
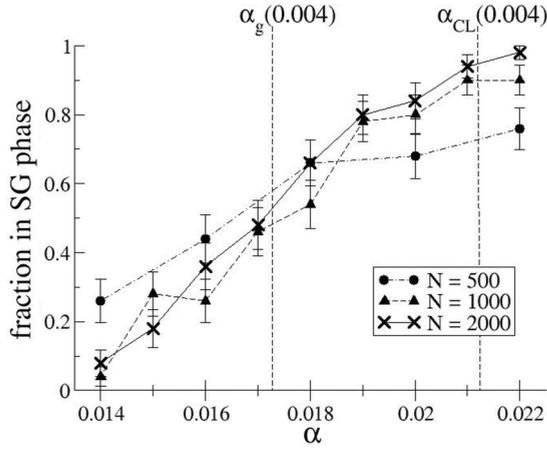
FIG. 17. Monte Carlo simulations around the clump-glass transition for $T = 0.004$: fraction of simulations found in the glassy phase after 100 rounds of $10N$ steps, as a function of α and for different N, with error bars. For each point the fraction was calculated from 50 simulations, half of which were started in a clump configuration and the other half in a uniform configuration.

load at which the clump phase becomes unstable at $T = 0$ and the temperature at which the clump phase is unstable when $\alpha = 0$. We also study the influence of w and f on first-order transitions, through α_g and T_c, respectively the load of transition to the glassy phase at $T = 0$ and the temperature of transition to the PM phase at $\alpha = 0$.

1. Reduced-distance parameter w

Parameter w controls the maximal distance d_c between the place field centers of interacting place cells; see (1) and Sec. II A. It fixes the width of the clump in the phase of localized activity. Experiments on rats have shown that the size of place fields depends on the size and complexity of the environment and on the behavioral context. A value $w = 0.05$, i.e., place fields occupying a few percent of the total space, is reasonable [25]. We have varied w for different values of f, and have found that T_{CL} is a monotonously increasing function of w (Fig. 18). This result agrees with the intuition that increasing w makes the clump phase more favorable energetically. It also appears that $\alpha_{CL}(w)$ has a maximum around $w \sim f$. In terms of storage capacity, this result suggests that there exists an optimal choice for the parameters: for a given level of inhibition hence a given number fN of active neurons, choosing $w \sim f$ maximizes the proportion of these active neurons that are located in the place field. Given that the quenched noise coming from other environments is statistically uniform over space (Fig. 14), $w \sim f$ represents a tradeoff between limiting the crosstalk and using the active neurons in the area covered by the place field.

As far as thermodynamic transitions to the PM and glassy phases are concerned we find that T_c and α_g behave similarly to, respectively, T_{CL} and α_{CL} when w varies, as shown in Fig. 19. Consequently, the qualitative aspect of the phase diagram remains the same when w varies.

2. Total activity f

Parameter f is the activity level of the network fixed by global inhibition. As expected, T_{CL} is a monotonously
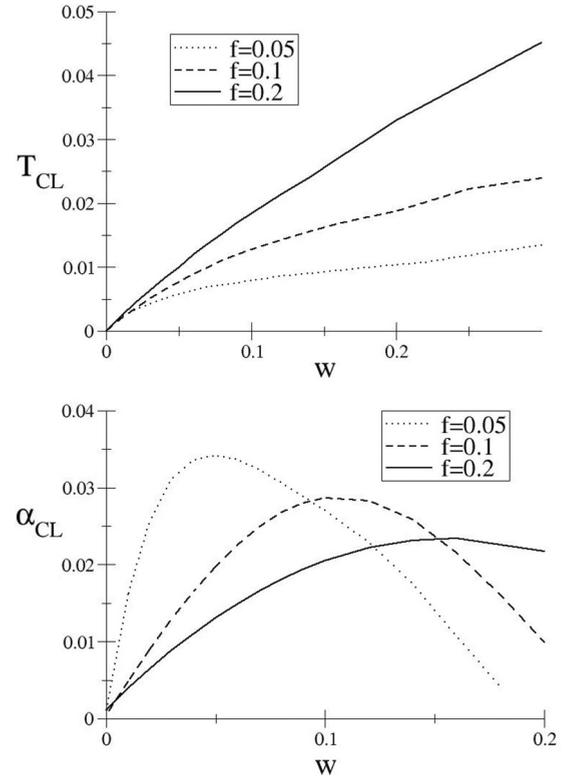

FIG. 18. Influence of w on the clump phase: T_{CL} (top) and α_{CL} (bottom) as a function of w, for different fixed values of f. Note the maximum around $w \sim f$ in the latter graph. Computations were done with $M = 1000$. The numerical error is $\delta\alpha_{CL} \sim 0.005$.

increasing function of f (Fig. 20). We find again a maximum of α_{CL} when f is of the order of w, consistently with the previous results. We also find that the boundary of the transition lines in phase diagram, α_g and T_c, behave similarly to α_{CL} and T_{CL} (Fig. 21).

VI. EXTENSIONS AND DISCUSSION

A. Taking silent cells into account

Thompson and Best [3] report that not all pyramidal cells have place fields in a given environment, a significative fraction of them (63% in their recording in CA1) being silent, i.e., with no place field, in this particular environment. To take this effect into account, our model can be further refined to incorporate partial activity of the cell ensemble. We assume that a fraction $c < 1$ of cells are active in any environment:

(i) In the reference environment (environment 0), cN given spins σ_i among the N are assigned regularly spaced place field centers $p(i)$.

(ii) For each one of the other environments, say, $\ell \geqslant 1$, each spin σ_i (among all N spins) is selected with probability c, and the place field centers are reshuffled by a random permutation π^ℓ. The set of chosen spins is encoded in the dilution variables

$$\tau_i^\ell = \begin{cases} 1 & \text{with probability } c, \\ 0 & \text{with probability } 1 - c. \end{cases} \tag{37}$$
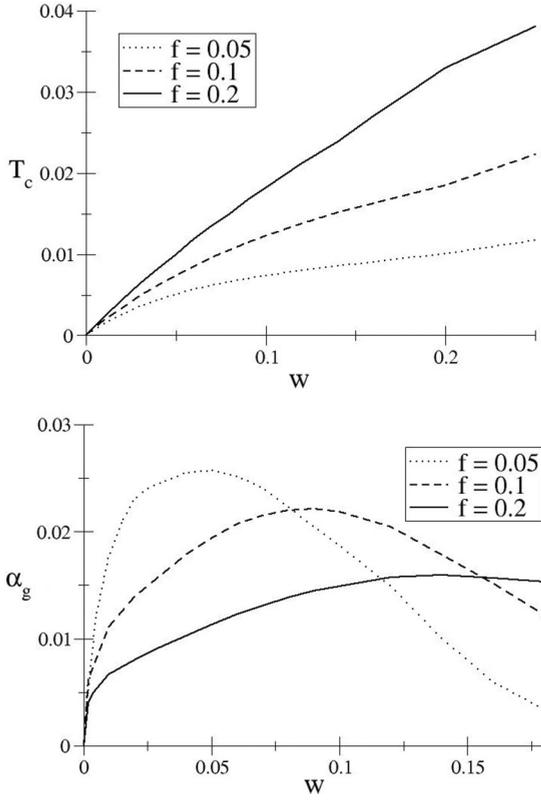
FIG. 19. Influence of w on the first-order transitions: T_c (top) and α_g (bottom) as a function of w, for different fixed values of f. Computations were done with $M = 1000$.

The resulting expression for the coupling matrix is

$$J_{ij} = J^0_{p(i)p(j)} + \sum_{\ell=1}^{L} J^0_{\pi^\ell(i)\pi^\ell(j)} \tau_i^\ell \tau_j^\ell. \qquad (38)$$

We incorporate this new hypothesis in the calculation of the average over disorder of the replicated partition function. The average is now over two types of disorder: the permutations π^ℓ and the selection of involved cells τ_i^ℓ. Neural configurations σ still satisfy (6). Moreover, we expect that the global inhibition is homogeneously distributed over the different subpopulations of neurons, and, for each realization of the τ_i^ℓ, we restrict the sum to configurations such that

$$\frac{1}{cN} \sum_{i=1}^{N} \tau_i^\ell \sigma_i = f. \qquad (39)$$

A detailed calculation, reported in Appendix D, shows that (6) implies (39) up to corrections of the order of $\frac{1}{\sqrt{N}}$. In addition we give in Appendix D the expression for the free energy in dimension $D = 1$. The corresponding extremization equations are

$$r = 2c^2(q - f^2) \sum_{k \geqslant 1} \left[\frac{k\pi}{\sin(k\pi w)} - \beta c(f - q) \right]^{-2},$$

$$q = c \int dx \int Dz [1 + e^{-\beta z \sqrt{\alpha r} - \beta \mu(x)}]^{-2}$$
$$+ (1 - c) \int Dz [1 + e^{-\beta z \sqrt{\alpha r} - \beta \mu_2}]^{-2},$$

$$\rho(x) = \int Dz [1 + e^{-\beta z \sqrt{\alpha r} - \beta \mu(x)}]^{-1},$$

$$\mu(x) = c \int dy\, J_w(x - y)\rho(y) + \lambda,$$

$$f = \int dx\, \rho(x),$$

$$f = \int Dz [1 + e^{-\beta z \sqrt{\alpha r} - \beta \mu_2}]^{-1}. \qquad (40)$$

In the partial activity model, the active spins (with activity $\rho(x)$) obey equations that are very similar to the previous case, with a dilution factor coming from the silent spins which are in a paramagnetic phase. From a qualitative point of view the behavior of the system does not differ significantly from the system with all spins active ($c = 1$). We have computed the effect of varying c on the value of T_c and α_g: T_c is found to be a linear function of c, while α_g is a monotonously increasing function of c. Results are shown in Fig. 22.

B. Relationship with linear threshold models and previous studies

Several attractor neural network models for the hippocampus have been proposed in previous works. Tsodyks and Sejnowski [14] proposed a rate model with semilinear threshold neurons, uniform inhibition, and excitatory synapses between neurons with neighboring place fields, with a strength decaying exponentially with distance. Their study was limited
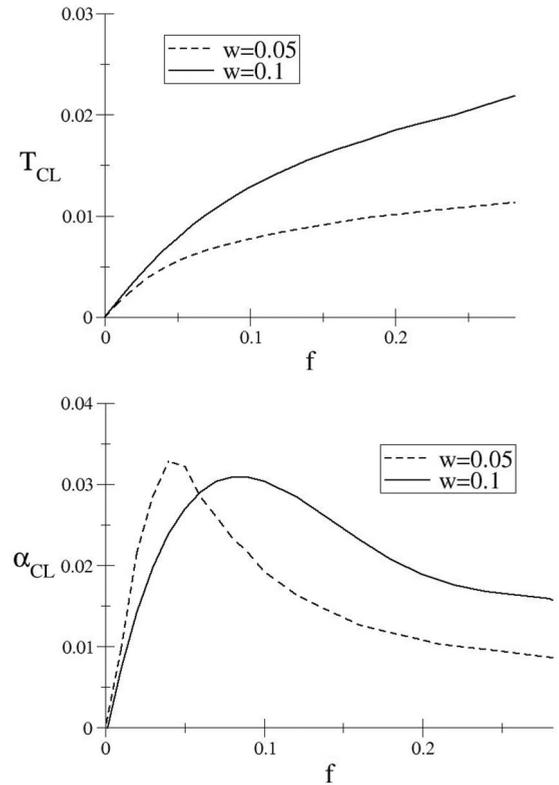

FIG. 20. Influence of f on the clump phase: T_{CL} (top) and α_{CL} (bottom) as a function of f, for different fixed values of w. Note the maximum around $f \sim w$ in the latter graph. Computations were done with $M = 1000$. The numerical error is $\delta\alpha_{CL} \sim 0.005$.
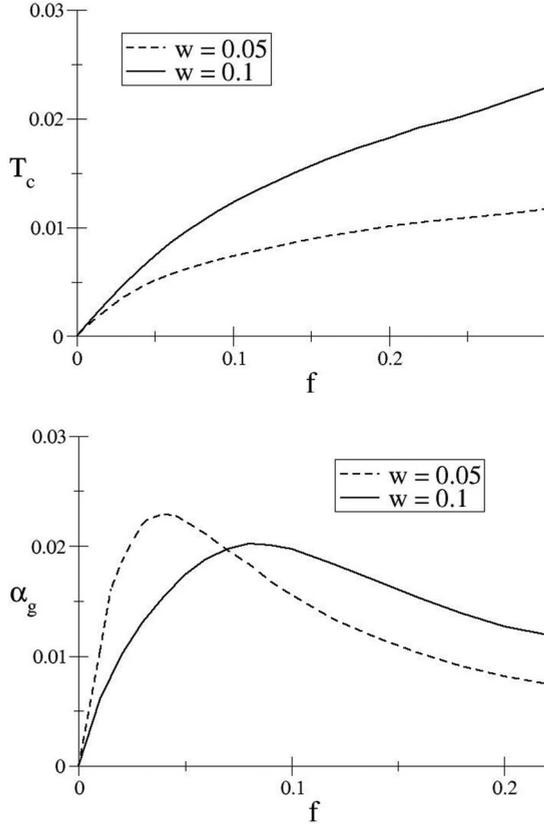
FIG. 21. Influence of f on the first-order transitions: T_c (top) and α_g (bottom) as a function of f, for different fixed values of w. Computations were done with $M = 1000$.
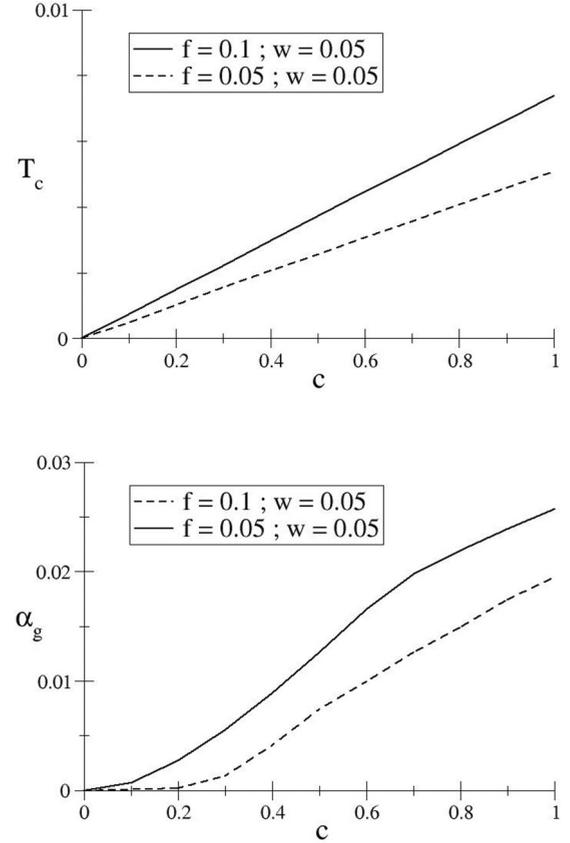

FIG. 22. Effect of partial activity: Influence of the fraction c of active cells on the clump domain: T_c (top) and α_g (bottom) as a function of c, for different fixed values of f and w. Computations were done with $M = 200$.

to the single environment, one-dimensional case. They showed the formation of localized activity. Moreover, they studied the effect of inhomogeneities in the synaptic matrix due to irregularities in the learning process, an interesting effect that we do not address here.

Battaglia and Treves [18] introduced the multiple environment storage in additive synapses. They studied the case of linear threshold neurons with generic form of kernel of connection weights. The free energy is calculated implementing the threshold linear transfer function and averaging over disorder in the replica-symmetric approximation, along the lines developed in [26]. The clump phase is studied at zero temperature, and the storage capacity is found as the maximal value of α for which localized solutions exist. Different forms of couplings and varying sparsity of the representation are considered, and an enlightening parallel with episodic memory is proposed. The issue of information storage is addressed.

Our method is in the same spirit as [18], but the model differs as we consider binary units instead of threshold linear units (i.e., without saturation) for a simple coupling matrix and an explicit form of inhibition. Nevertheless, a parallel can be drawn between the range of interaction w in our model and the "map sparsity" $\frac{1}{|M|}$ in [18]. In spite of the differences between the models, the order of magnitude of the maximal storage capacity is the same in both models: $\sim 3.10^{-2}$ in one dimension, $\sim 8.10^{-3}$ in two dimensions (see Figs. 1 and 2 in [18]). The "chart sparsity" α_c in [18] corresponds to our parameter c.

The main difference between both models lies in the way noise is taken into account. In [18], the level of noise is embedded in the rate model, in the gain g of the units, and is not taken into account in the thermodynamics since the study is carried out at zero temperature. Our model considers binary units with a level of noise T corresponding to the thermodynamic temperature. On average, binary neurons behave as rate neurons with sigmoidal transfer function of gain $\frac{1}{T}$ (see Sec. II C). From this point of view our model is more microscopic than the one in [18], as we have a description of noise at the neuron level. Furthermore, we have looked at the stability of the clump phase against replicon modes. Our study also includes the other regimes of activity of the model (i.e., the PM and SG phases) and their thermodynamic stability compared to the clump phase, summarized in the phase diagram.

C. Conclusion

In this paper we have introduced an attractor neural network model for the storage of multiple spatial maps in the hippocampus. Although very simplified, the model accounts for experimentally observed properties of place cells, such as the remapping of place fields from one environment to the other. We showed that multiple maps can be simultaneously learned in the same network, i.e., with the same synaptic coupling coefficients, even in the presence of noise in the neural response. Remarkably, moderate levels of noise can

even slightly increase the capacity storage with respect to the noiseless case. Notice that the qualitative behavior of the model is robust to changes in the value of the parameters; for instance we do not expect that changing the couplings from a square-box function into an exponentially decreasing function over the distance wN in $D = 1$ or $\sqrt{wN}$ in $D = 2$ would have much effect on the phase diagram.

The storage of a map manifests itself through the fact that the neural activity is localized, and acquires a clumplike shape in the corresponding environment. When the load (number of environments) or the noise are too high the neural activity cannot be localized any longer in any one of the environments. For high noise, the activity, averaged over time, simply becomes uniform over the space. For high loads the activity is not uniform, but is delocalized with spatial heterogeneities controlled by the crosstalks between the (too many) maps. The prevalence of the glassy phase at high load and of the uniform (paramagnetic in the physics language) phase at high noise moderately limits the extension of the clump phase. Moreover, we have found that in the glassy phase the replica symmetric assumption is not correct, and we may expect from general consideration about replica symmetry breaking that the first-order transition from the clump phase to the glassy phase occurs at higher loads α. Remarkably the clump phase is therefore the thermodynamically dominant phase in nearly all of its stability domain.

The present work is a direct offspring of spin-glass models of attractor neural networks [16], with the difference that here one pattern corresponds to one map, i.e., a whole set of coherent neural configurations, instead of a single configuration of activity. This explains the robustness of the patterns against neural noise in our case compared to the Hopfield model case, as discussed in Sec. V. This generalization of the notion of "stored pattern" is interesting and would deserve further consideration. It appears that the concept of attractor neural network can embed memory items with much richer structure than the ones originally considered; it is quite encouraging that the theoretical framework built for the original Hopfield model can be extended to deal with those structured items. In the case of the hippocampus, it is widely believed that CA3 is the support of episodic memory, that is, the memory of autobiographical events and contextual experiences. According to this view, the hippocampus could learn not only spatial relations between locations but also associate them to events, times, and emotions. In our model the coupling matrix associates nearby places together. We could imagine a generalization of it to a network which makes associations between units coding for other, more abstract nonspatial features, although characterizing the "metric" properties of general feature space is much harder than for the usual Euclidean space.

Our work would deserve to be extended along other directions. First the assumption that synaptic couplings additively sum up the contributions coming from all the environments could be lifted. We could replace the synapses J_{ij} with a nonlinear function $G(J_{ij})$. The additive case corresponds to $G(x) = x$, while a strongly nonadditive synapse is obtained with the choice $G(x) = \min(x, \frac{1}{N})$: synapses can be written only once, and contributions from different environments do not add up but saturate the synaptic coupling. It would be worth extending the study of nonlinear synapses done for the Hopfield model [27,28] to the present model.

Second, we have considered that the only source of quenched noise was the interference between the multiple environments. In other words, in the single-environment case, our synaptic matrix is translationally invariant and the center of the activity clump can be moved at no energy cost in space. This idealizing assumption was done to study the effect of multiple-environment crosstalk only. However, even in the single environment case, place fields do not define a perfectly regular covering of space. We expect that such heterogeneities in the couplings will further destabilize the clump phase, and decrease the storage capacity [27]. Quantifying those effects would be interesting.

However, the most important extension seems to us to be the study of the dynamics. The richness of the phase diagram we have unveiled here and the multiplicity of phases for the system raise the question of if and how the network activity makes transitions between those phases. Multiple environments stored in the same network not only influence the shape of the clump and lead to transitions to a glassy phase, but they can as well provoke transitions between attractors. The study of these transitions and of the corresponding reaction paths will be reported in a companion paper [21]. It could prove useful to interpret recent experiments, where changes of the hippocampal activity resulting from the "teleportation" of the rat have been recorded [20]. In addition it would be interesting to understand in a more quantitative way the activated diffusion process of the clump in an environment. In the presence of other maps, the invariance by translation is lost and the clump does not freely diffuse. Quantifying the barriers opposing motion, as well as understanding the qualitative difference between motions in 1D and 2D spaces, would be very useful.

APPENDIX A: FORMULAS FOR TWO-DIMENSIONAL MAPS

The only difference in the replica computation lies in the eigenvalues of the coupling matrix. Thus, in dimension 2, the free energy functional writes

$$
\begin{aligned}
\mathcal{F}^{2D} = {} & \frac{\alpha\beta}{2} r(f - q) - \frac{\alpha}{\beta} \psi^{2D}(q, \beta) \\
& - \frac{1}{2} \int d\vec{x}\, d\vec{y}\, \rho(\vec{x}) J_w(\vec{x} - \vec{y}) \rho(\vec{y}) + \int d\vec{x}\, \mu(\vec{x}) \rho(\vec{x}) \\
& - \frac{1}{\beta} \int d\vec{x} \int Dz \ln(1 + e^{\beta z \sqrt{\alpha r} + \beta \mu(\vec{x})}),
\end{aligned} \tag{A1}
$$

where

$$\psi^{2D}(q,\beta) \equiv 2 \sum_{\substack{(k_1,k_2) \\ \neq(0,0)}} \left(\frac{\beta(q-f^2)}{\phi(k_1,k_2) - \beta(f-q)} - \ln\left(1 - \frac{\beta(f-q)}{\phi(k_1,k_2)}\right)\right), \tag{A2}$$

with

$$\phi(k_1,k_2) \equiv \frac{k_1 k_2 \pi^2}{\sin(k_1\pi\sqrt{w})\sin(k_2\pi\sqrt{w})}. \tag{A3}$$

Hence the saddle point equations write

$$r = 4(q-f^2) \sum_{(k_1,k_2)\neq(0,0)} [\phi(k_1,k_2) - \beta(f-q)]^{-2},$$

$$q = \int d\vec{x} \int Dz [1 + e^{-\beta z\sqrt{\alpha r} - \beta\mu(\vec{x})}]^{-2},$$

$$\rho(\vec{x}) = \int Dz [1 + e^{-\beta z\sqrt{\alpha r} - \beta\mu(\vec{x})}]^{-1}, \tag{A4}$$

$$\mu(\vec{x}) = \int d\vec{y} \, J_w(\vec{x} - \vec{y})\rho(\vec{y}) + \lambda,$$

where λ is determined to enforce constraint (12).

In the $D=2$ case Eqs. (A4) can be simplified by exploiting the invariance by rotation: in polar coordinates

$$\mu(r) = 2 \int_{\substack{r+r'\geqslant\sqrt{w/\pi} \\ |r-r'|\leqslant\sqrt{w/\pi}}} dr' \rho(r)r' \arccos\left(\frac{r^2+r'^2 - \frac{w}{\pi}}{2rr'}\right) + 2\pi \int_{r+r'<\sqrt{w/\pi}} dr' \rho(r)r' + \lambda. \tag{A5}$$

We thus computed $\rho(r)$ in the clump phase and found the region in the (α,T) plane where this solution is stable against longitudinal modes. We find that this region is reduced compared to the $D=1$ case, but its shape is qualitatively similar. The result is shown in Fig. 23.
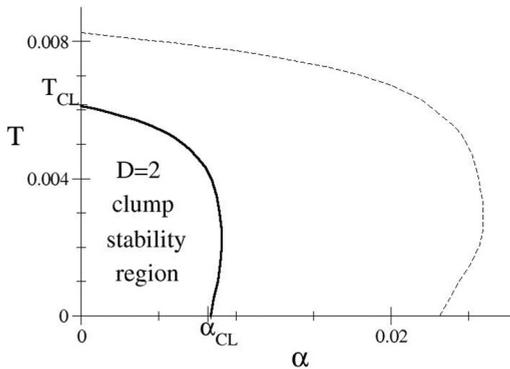

FIG. 23. Solid line: longitudinal stability region of the clump phase for $D=2$. The $D=1$ case is shown in thin dashed line for comparison.

APPENDIX B: AVERAGE OF THE BOLTZMANN FACTOR OVER A RANDOM ENVIRONMENT

The purpose of this appendix is to calculate

$$\Xi(\vec{\sigma}) = \frac{1}{N!} \sum_\pi \exp\left[\beta \sum_{i<j} J_{ij}^0 \sum_{a=1}^n \sigma_{\pi(i)}^a \sigma_{\pi(j)}^a\right]$$

$$= C\xi(\vec{\sigma}) \tag{B1}$$

with

$$C \equiv \exp\left(N\frac{\beta}{2}nwf^2\right)$$

$$\xi(\vec{\sigma}) \equiv \frac{1}{N!} \sum_\pi \exp\left[\frac{\beta}{2} \sum_{i,j} J_{ij}^0 \sum_{a=1}^n \left(\sigma_{\pi(i)}^a - f\right)\left(\sigma_{\pi(j)}^a - f\right)\right], \tag{B2}$$

where the sum is carried out over all permutations of N elements.

The eigenvectors of the matrix J^0 are plane waves. Let $v_{q,j}$ denote the jth (real-valued) component of the qth normalized eigenvector, and λ_q the associated eigenvalue. Then,

$$\sum_{i,j} J_{ij}^0 \left(\sigma_{\pi(i)}^a - f\right)\left(\sigma_{\pi(j)}^a - f\right)$$

$$= \sum_{q=1}^{N-1} \lambda_q \left(\sum_j v_{q,j}\left(\sigma_{\pi(j)}^a - f\right)\right)^2. \tag{B3}$$

Due to condition (6) we have discarded the homogeneous mode $q=0$ from the sum in (B3). Introducing a set of $n(N-1)$ independent Gaussian variables with zero mean and variance unity, denoted by Φ_q^a, we can write (all odd powers of $\sqrt{\beta}$ vanish after integration over the Gaussian measure)

$$\xi(\vec{\sigma}) = \left\langle \exp\left[\sqrt{\beta} \sum_{q,a,j} \sqrt{\lambda_q}\, v_{q,j}\, \Phi_q^a\, (\sigma_{\pi(j)}^a - f)\right]\right\rangle_{\pi,\Phi}$$

$$= 1 + \sum_{k\geqslant 1} \frac{\beta^k}{(2k)!} \sum_{\substack{q_i,a_i,j_i \\ i=1\cdots 2k}} \left(v_{q_1,j_1}\, v_{q_2,j_2}\ldots v_{q_{2k},j_{2k}}\right.$$

$$\times \sqrt{\lambda_{q_1}\lambda_{q_2}\ldots\lambda_{q_{2k}}}\, T_{j_1,j_2\ldots j_{2k}}^{a_1,a_2\ldots a_{2k}} \left\langle\Phi_{q_1}^{a_1}\Phi_{q_2}^{a_2}\ldots\Phi_{q_{2k}}^{a_{2k}}\right\rangle_\Phi\Big), \tag{B4}$$

where

$$T_{j_1,j_2,\ldots,j_{2k}}^{a_1,a_2,\ldots,a_{2k}} \equiv \left\langle\left(\sigma_{\pi(j_1)}^{a_1} - f\right)\left(\sigma_{\pi(j_2)}^{a_2} - f\right)\ldots\left(\sigma_{\pi(j_{2k})}^{a_{2k}} - f\right)\right\rangle_\pi. \tag{B5}$$

Using Wick's theorem the $2k$-point correlation function of the Φ variables is easy to calculate. The outcome is a multiplicative factor $(2k-1)!!$, and the replacement of the $2k$ sums over the indices q_m,a_m by only k independent sums over q_m,a_m. The value of T (B5) depends only on the number of distinct indices, i_m, and of their associated multiplicities. Power counting shows that $T_{i_1,i_2,\ldots,i_{2k}}^{a_1,a_2,\ldots,a_{2k}}$ vanishes in the infinite N limit unless the set $\{i_1,i_2,\ldots,i_{2k}\}$ includes exactly k distinct indices, each one with multiplicity 2. When this condition holds we write (a_m,a'_m), the replica indices attached to the mth distinct index

i, with $m = 1, 2, \ldots, k$. Then, in the large N limit,

$$T^{a_1, a_1', a_2, a_2', \ldots, a_k, a_k'}_{i_1, i_1, i_2, i_2 \ldots, i_k, i_k} = \prod_{m=1}^{k} (q^{a_m a_m'} - f^2). \quad (B6)$$

We assume that the quantities $\sqrt{\beta \lambda_q} \sum_j v_{q,j} (\sigma^a_{\pi(j)} - f)$ do not diverge when the limit $N \to \infty$ is taken for each one of the terms in the series over k in (B4). This hypothesis breaks down if the permutation π is "close" to the identity permutation, or, equivalently, if the configuration $\vec{\sigma}_\pi = \{\sigma_{\pi(i)}\}$ is coherent with the environment 0. As π is randomly chosen the probability that this is the case vanishes for large N.

We are left with the summation over the j_m indices. Using the identities

$$\sum_j v_{q,j}\, v_{q',j} = \delta_{q,q'}, \quad (B7)$$

we obtain from (B4) the following expression:

$$\xi(\vec{\sigma}) = 1 + \sum_{k \geqslant 1} \frac{(\beta/2)^k}{k!} \sum_{\mathcal{P}} w(\mathcal{P}), \quad (B8)$$

where the last sum runs over all weighted pairings among $2k$ points, described as follows:

(i) We define $2k$ points. The first k points carry the pair indices (q_m, a_m), with m running from 1 to k. The second k points carry the same pair indices. Hence, each pair index (q_m, a_m) is shared by two points.

(ii) A pairing $\mathcal{P}$ is a set of k bonds $b_\ell \equiv \{(q_{m_\ell}, a_{m_\ell}), (q_{m'_\ell}, a_{m'_\ell})\}$, $\ell = 1, 2, \ldots, k$, each joining one pair of points (dimer coverage).

(iii) The weight of the pairing is

$$w(\mathcal{P}) \equiv \sum_{\substack{a_1, \ldots, a_k \\ q_1, \ldots, q_k}} \prod_{m=1}^{k} \lambda_{q_m} \prod_{\ell=1}^{k} \delta_{q_{m_\ell}, q_{m'_\ell}} \left(q^{a_{m_\ell} a_{m'_\ell}} - f^2\right). \quad (B9)$$

We denote $\mathbf{q}$ the overlap matrix with entries q^{ab} and $\mathbf{1}_n$ the $n \times n$ matrix whose all entries are equal to 1. Let us introduce a notation for the moments of the eigenvalues:

$$\Lambda_h \equiv \sum_{q \geqslant 1} \lambda_q^h = 2 \sum_{q \geqslant 1} \left(\frac{\sin(q \pi w)}{q \pi}\right)^h. \quad (B10)$$

Two examples of pairings are shown in Fig. 24. The weight associated to the pairing $\mathcal{P}_A$ is simply

$$w(\mathcal{P}_A) = \prod_{m=1}^{k} \left[\sum_{q_m} \lambda_{q_m} \sum_{a_m} \left(q^{a_m a_m} - f^2\right)\right]$$
$$= (\Lambda_1 \, \text{Tr}[\mathbf{q} - f^2])^k = [\Lambda_1 \, n \, f (1 - f)]^k, \quad (B11)$$

as all Kronecker δ in (B9) are equal to 1 by construction. The weight associated to the second pairing in Fig. 24 is

$$w(\mathcal{P}_B) = \{\Lambda_3 \, \text{Tr}[(\mathbf{q} - f^2)^3]\} (\Lambda_1 \, \text{Tr}[\mathbf{q} - f^2])^{k-3}. \quad (B12)$$

For a given pairing,

(i) the horizontal bonds represent independent replicas: point number m leads to a factor $\sum_{q_m} \lambda_{q_m} \sum_{a_m} (q^{a_m a_m} - f^2)$ in the weight of the pairing;

(ii) the vertical and diagonal bonds couple replicas together.
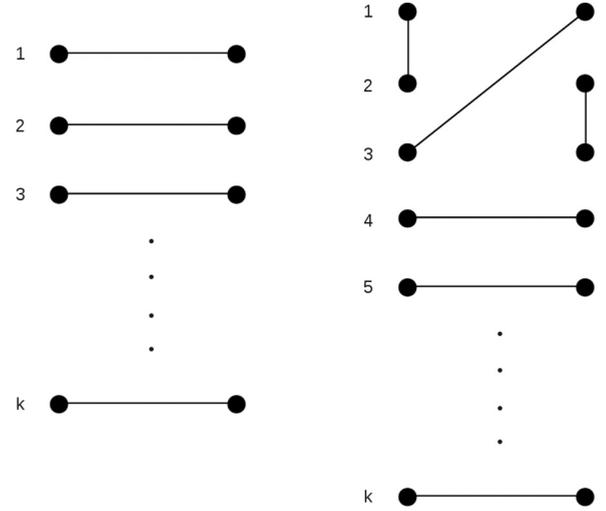

FIG. 24. Two examples of pairings between $2k$ points: $\mathcal{P}_A$ (left) and $\mathcal{P}_B$ (right).

We then have to calculate the combinatorial multiplicity of the weights, i.e., how many pairings have the same weight in the sum (B8). For a given k, a pairing associates points by groups of j coupled replicas indices (i.e., $2j$ points). Let m_j be the number of such groups in a given pairing. We have $\sum_j j m_j = k$. Pairings $\mathcal{P}$ with the same (j, m_j) have equal weights

$$w(\mathcal{P}) = w[\{(j, m_j)\}] = \prod_j (\Lambda_j T_j)^{m_j}, \quad (B13)$$

where we set $T_j \equiv \text{Tr}[(\mathbf{q} - f^2 \mathbf{1})^j]$.

Combinatorial study shows that the number of such pairings is

$$\mathcal{N}[\{(j, m_j)\}] = k! \prod_j \frac{1}{m_j!} \left(\frac{2^{j-1}}{j}\right)^{m_j}. \quad (B14)$$

Finally, using $\sum_j j m_j = k$ and (B10), we can rewrite

$$\xi(\vec{\sigma}) = 1 + \sum_{k \geqslant 1} \left(\frac{\beta}{2}\right)^k \prod_j \sum_{m_j \geqslant 0} \frac{1}{m_j!} \left(\frac{2^{j-1}}{j}\right)^{m_j} (\Lambda_j T_j)^{m_j}$$
$$= \exp\left[\sum_j \frac{1}{2} \frac{\beta^j}{j} \Lambda_j T_j\right]$$
$$= \exp\left[-\sum_{\lambda \neq 0} \text{Tr} \ln[\mathbf{Id}_n - \beta \lambda (\mathbf{q} - f^2 \mathbf{1}_n)]\right]. \quad (B15)$$

APPENDIX C: REPLICA SYMMETRIC CALCULATION OF THE FREE ENERGY

We introduce parameters r^{ab} conjugated to the overlaps q^{ab}. With this notation, we have (up to a multiplicative irrelevant constant):

$$\overline{Z_J^n} = \sum_{\vec{\sigma}} \int \prod_{a<b} dq^{ab} dr^{ab} \exp[G(\{q^{ab}, r^{ab}\}, \vec{\sigma})], \quad (C1)$$

where

$$
G(\{q^{ab}, r^{ab}\}, \vec{\sigma}) = N\alpha\beta^2 \sum_{a<b} r^{ab} \left(\frac{1}{N} \sum_i \sigma_i^a \sigma_i^b - q^{ab} \right)
$$
$$
+ \frac{\beta}{2N} \sum_a \sum_{|i-j|<wN/2} \sigma_i^a \sigma_j^a
$$
$$
- \alpha N \sum_{\lambda \neq 0} \mathrm{Tr}\, \ln[\mathbf{Id}_n - \beta\lambda(\mathbf{q} - f^2 \mathbf{1}_n)]. \tag{C2}
$$

We rewrite the sum over spin configurations as a path integral over continuous-space average densities and over the conjugated potentials,

$$
\sum_{\vec{\sigma}} \exp\left[\alpha\beta^2 \sum_{a<b} r^{ab} \sum_i \sigma_i^a \sigma_i^b + \frac{\beta}{2N} \sum_a \sum_{|i-j|<wN/2} \sigma_i^a \sigma_j^a \right]
$$
$$
= \int \prod_a \mathscr{D}\rho^a(x)\mathscr{D}\mu^a(x)d\lambda^a \exp\left[N \int dx \ln Z(\{\mu^a(x), r^{ab}\}) \right.
$$
$$
+ N \sum_a \left[\beta\lambda^a \left(\int dx\, \rho^a(x) - f \right) - \beta \int dx\, \rho^a(x)\mu^a(x) \right.
$$
$$
\left. \left. + \frac{\beta}{2} \int dx\, dy\, \rho^a(x)J_w(x-y)\rho^a(y) \right] \right], \tag{C3}
$$

where we have defined

$$
Z(\{\mu^a(x), r^{ab}\})
$$
$$
\equiv \sum_{\{\sigma^a\}} \exp\left[\alpha\beta^2 \sum_{a<b} \sigma^a \sigma^b r^{ab} + \beta \sum_a \mu^a(x)\sigma^a \right]. \tag{C4}
$$

In the replica symmetric (RS) ansatz, we assume

$$
\forall\, a \neq b, \quad \forall\, x, \quad
\begin{cases}
r^{ab} = r, \\
q^{ab} = q, \\
\rho^a(x) = \rho(x), \\
\mu^a(x) = \mu(x), \\
\lambda^a = \lambda.
\end{cases} \tag{C5}
$$

We obtain

$$
T_j = (n-1)(f-q)^j + [f - f^2 + (n-1)(q-f^2)]^j, \tag{C6}
$$

and

$$
Z(\mu(x), r) = 1 + n \int Dz \ln(1 + e^{\beta z\sqrt{\alpha r} + \beta\mu(x) - \alpha\beta^2 r/2}) \tag{C7}
$$

up to $O(n^2)$ corrections. We now make the change of variable $\mu(x) \to \mu(x) - \frac{\alpha\beta r}{2}$. The averaged partition function is, for small n,

$$
\overline{Z^n} \sim \int dq\, dr\, d\lambda\, \mathscr{D}\mu(x)\mathscr{D}\rho(x) e^{-nN\beta\mathscr{F}[\mu(x), \rho(x), q, r, \lambda]}, \tag{C8}
$$

where

$$
\mathscr{F} = \frac{\alpha\beta}{2}r(f-q) - \frac{\alpha}{\beta}\psi(q,\beta) - \lambda\left(\int dx\rho(x) - f \right)
$$
$$
+ \int dx\rho(x)\mu(x) - \frac{1}{2}\int dxdy\, \rho(x)J_w(x-y)\rho(y)
$$
$$
- \frac{1}{\beta}\int dx\, Dz\ln(1 + e^{\beta\sqrt{\alpha r}z + \beta\mu(x)}) \tag{C9}
$$

with

$$
\psi(q,\beta) \equiv \sum_j \frac{1}{2}\frac{\beta^j}{j}\Lambda_j[j(q-f)^2(f-q)^{j-1} + (f-q)^j]. \tag{C10}
$$

When $N \to \infty$ the integral is calculated through the saddle-point method, and we look for the extremum of $\mathscr{F}$ over its arguments. We now give the expression of ψ and of the order parameter $r = -\frac{2}{\beta^2}\frac{\partial\psi}{\partial q}$ in dimensions $D = 1$ and $D = 2$.

Case $D = 1$. We define $A_k \equiv \frac{\pi k}{\sin(\pi kw)}$ and write

$$
\Lambda_j^{1D} = 2\sum_{k\geqslant 1}(A_k)^{-j}. \tag{C11}
$$

We immediately obtain

$$
\psi^{1D}(q,\beta) = \sum_{k\geqslant 1} \frac{\beta(q-f^2)}{A_k - \beta(f-q)} - \ln\left(1 - \frac{\beta(f-q)}{A_k} \right) \tag{C12}
$$

and

$$
r^{1D} = 2(q-f^2)\sum_{k\geqslant 1}[A_k - \beta(f-q)]^{-2}. \tag{C13}
$$

Case $D = 2$: see Appendix A.

APPENDIX D: SILENT CELLS CASE: CALCULATION OF THE FREE ENERGY

We now consider the hypothesis that a fraction $c < 1$ of the cells are involved in each environment's representation. Two types of disorder are present: the random permutations of the place field centers as before, and the choices of the subsets of cells taking part to each map ℓ, i.e., the dilution variables τ_i^ℓ. The nth moment of the partition function reads

$$
\overline{Z_J^n} = \sum_{\vec{\sigma}} \exp\left[\beta \sum_a \sum_{i<j} J_{ij}^0 \tau_i^0 \tau_j^0 \sigma_i^a \sigma_j^a \right]
$$
$$
\times \left\langle \exp\left[\beta \sum_{\ell=1}^L \sum_a \sum_{i<j} J_{ij}^0 \tau_{\pi^\ell(i)}^\ell \tau_{\pi^\ell(j)}^\ell \sigma_{\pi^\ell(i)}^a \sigma_{\pi^\ell(j)}^a \right] \right\rangle_{\pi,\tau}, \tag{D1}
$$

where τ denotes one realization of the $L \times N$ random variables τ_i^ℓ, and the τ_i^0 are 1 if i is a multiple of $1/c$, 0 otherwise. The sum is now taken over spin configurations satisfying (6) and (39) for each replica index a.

Using the function $\mathbb{1}(x) = 1$ if $x = 0$ and 0 elsewhere, we write

$$\overline{Z_J^n} = C \sum_{\text{all } \vec{\sigma}} \mathbb{1}\left(\frac{1}{N} \sum_i \sigma_i^a - f\right) e^{\beta \sum_a \sum_{i<j} J_{ij}^0 \tau_i^0 \tau_j^0 \sigma_i^a \sigma_j^a} \chi(\vec{\sigma})^L,$$

(D2)

where C is a constant and

$$\chi(\vec{\sigma}) \equiv \left\langle \mathbb{1}\left(\frac{1}{cN} \sum_i \tau_i^\ell \sigma_i^a - f\right) \right.$$

$$\left. \times \left\langle e^{\beta \sum_a \sum_{i<j} J_{ij}^0 \tau_{\pi(i)}^\ell \tau_{\pi(j)}^\ell (\sigma_{\pi(i)}^a - f)(\sigma_{\pi(j)}^a - f)} \right\rangle_\pi \right\rangle_\tau.$$

(D3)

The average over π follows exactly the steps described in Appendix B. Defining

$$\Theta(\tilde{\mathbf{q}}) \equiv -\sum_{\lambda \neq 0} \text{Tr} \ln[\mathbf{Id}_n - \beta\lambda(\tilde{\mathbf{q}} - cf^2 \mathbf{1}_n)],$$

(D4)

where $\tilde{\mathbf{q}}$ is now the $n \times n$ matrix of elements

$$\tilde{q}_{ab} \equiv \frac{1}{N} \sum_i \tau_i \sigma_i^a \sigma_i^b,$$

(D5)

we end up with $\chi(\vec{\sigma}) = \int d\tilde{q}_{ab} \exp[\Theta(\tilde{\mathbf{q}})] \, \tilde{\chi}(\tilde{\mathbf{q}}, \vec{\sigma})$ with

$$\tilde{\chi}(\tilde{\mathbf{q}}, \vec{\sigma}) = C' \int_{i\mathbb{R}} d\lambda_a dR_{ab} e^{\sqrt{N}\left(cf \sum_a \lambda_a + \sum_{a \leqslant b} R_{ab} \tilde{q}_{ab}\right)}$$

$$\times \prod_i \left\langle \exp\left[-\tau_i \left(\sum_a \frac{\lambda_a}{\sqrt{N}} \sigma_i^a + \frac{R_{ab}}{\sqrt{N}} \sigma_i^a \sigma_i^b\right)\right]\right\rangle_{\tau_i},$$

(D6)

where we have introduced parameters R_{ab} conjugated to $\tilde{q}_{ab}$ and Lagrange multipliers λ_a to enforce the constraint on $\vec{\sigma}$.

We now perform the average over the decoupled variables τ_i. Introducing the order parameters

$$T_{abc} \equiv \frac{1}{N} \sum_i \sigma_i^a \sigma_i^b \sigma_i^c, \quad S_{abcd} \equiv \frac{1}{N} \sum_i \sigma_i^a \sigma_i^b \sigma_i^c \sigma_i^d, \quad (D7)$$

and carrying out the Gaussian integration over the leading terms (when $N \gg 1$) in λ_a and R_{ab}, we have (up to a multiplicative constant)

$$\tilde{\chi}(\tilde{\mathbf{q}}, \vec{\sigma}) = C'' e^{-(N/2) \sum_{a,b,c,d} [A^{-1}]_{abcd}(\tilde{q}_{ab} - cq_{ab})(\tilde{q}_{cd} - cq_{cd})},$$

(D8)

where

$$A_{abcd} \equiv c(1-c)\left(S_{abcd} + 2 \sum_{e,f} [Q^{-1}]_{ef} T_{abe} T_{cdf}\right). \quad (D9)$$

Hence, in the large N limit, the integral over $\tilde{q}_{ab}$ is dominated by $\tilde{q}_{ab} = c q_{ab}$.

The replica symmetric calculation of the free energy proceeds along the steps of Appendix C. The only difference is that, here, $\mu(x)$ and $\rho(x)$ describe the activity of the cN cells involved in the reference environment, while the $(1-c)N$ remaining cells have uniform activities $= f$. We obtain the expression of the energy functional

$$\mathcal{F}_c = \frac{\alpha\beta}{2} r(f-q) - \frac{\alpha}{\beta} \psi_c(q,\beta)$$

$$+ c \int dx \, \mu(x)\rho(x) + (1-c)\mu_2 f$$

$$- \frac{c^2}{2} \int dx \, dy \, \rho(x) J_w(x-y)\rho(y)$$

$$- \lambda c \left(\int dx \, \rho(x) - f\right)$$

$$- \frac{c}{\beta} \int dx \int Dz \ln\left(1 + e^{\beta z \sqrt{\alpha r} + \beta \mu(x)}\right)$$

$$- \frac{(1-c)}{\beta} \int Dz \ln\left(1 + e^{\beta z \sqrt{\alpha r} + \beta \mu_2}\right), \quad (D10)$$

where q is defined as before and

$$\psi_c(q,\beta) = \sum_{k \geqslant 1} \left[\frac{\beta c(q-f^2)\sin(k\pi w)}{k\pi - \beta c(f-q)\sin(k\pi w)}\right.$$

$$\left. - \ln\left(1 - \frac{\beta c(f-q)\sin(k\pi w)}{k\pi}\right)\right]. \quad (D11)$$

APPENDIX E: STABILITY OF THE REPLICA SYMMETRIC SOLUTION

The extremization of the free energy functional under the fixed-activity constraint and under the replica symmetric assumption leads to three solutions corresponding to three different phases. We want to study the stability of those solutions in the (α, T) space. We will limit ourselves to the one-dimensional case. A small perturbation of the solution

$$\begin{aligned}
\rho^a(x) &\to \rho^a(x) + \delta\rho^a(x), \\
\mu^a(x) &\to \mu^a(x) + \delta\mu^a(x), \\
r^{ab} &\to r^{ab} + \delta r^{ab}, \\
q^{ab} &\to q^{ab} + \delta q^{ab}
\end{aligned}$$

(E1)

results in $\mathscr{F} \to \mathscr{F} + \frac{1}{2}\delta^2\mathscr{F}$, where

$$\delta^2\mathscr{F} = \int dx dy \begin{bmatrix} \delta\rho^a(x) \\ \delta\mu^a(x) \\ \delta r^{ab} \\ \delta q^{ab} \end{bmatrix}^\dagger M(x,y) \begin{bmatrix} \delta\rho^c(y) \\ \delta\mu^c(y) \\ \delta r^{cd} \\ \delta q^{cd} \end{bmatrix}. \quad (E2)$$

The Hessian matrix $M(x,y)$ reads, in the $\{\delta\rho^a(x),\delta\mu^a(x),\delta r^{ab},\delta q^{ab}\}$ basis,

$$
M = \begin{bmatrix}
\frac{\partial^2 \mathscr{F}}{\partial\rho^a(x)\partial\rho^c(y)} & \frac{\partial^2 \mathscr{F}}{\partial\rho^a(x)\partial\mu^c(y)} & 0 & 0 \\[2mm]
\frac{\partial^2 \mathscr{F}}{\partial\mu^a(x)\partial\rho^c(y)} & \frac{\partial^2 \mathscr{F}}{\partial\mu^a(x)\partial\mu^c(y)} & \frac{\partial^2 \mathscr{F}}{\partial\mu^a(x)\partial r^{cd}} & 0 \\[2mm]
0 & \frac{\partial^2 \mathscr{F}}{\partial r^{ab}\partial\mu^c(y)} & \frac{\partial^2 \mathscr{F}}{\partial r^{ab}\partial r^{cd}} & \frac{\partial^2 \mathscr{F}}{\partial r^{ab}\partial q^{cd}} \\[2mm]
0 & 0 & \frac{\partial^2 \mathscr{F}}{\partial q^{ab}\partial r^{cd}} & \frac{\partial^2 \mathscr{F}}{\partial q^{ab}\partial q^{cd}}
\end{bmatrix}. \quad \text{(E3)}
$$

Using the notations

$$
t(x) \equiv \overline{\langle\sigma\rangle_{(x)}^3} = \int Dz[1 + e^{-\beta z\sqrt{\alpha r}-\beta\mu(x)}]^{-3},
$$

$$
s(x) \equiv \overline{\langle\sigma\rangle_{(x)}^4} = \int Dz[1 + e^{-\beta z\sqrt{\alpha r}-\beta\mu(x)}]^{-4},
$$

$$
t \equiv \int dx\, t(x); \quad s \equiv \int dx\, s(x); \quad \text{(E4)}
$$

$$
q_2 \equiv \int dx\, q^2(x),
$$

we have

$$
\frac{\partial^2 \mathscr{F}}{\partial\rho^a(x)\partial\rho^c(y)} = -J_w(x-y)\delta^{ac}, \quad \text{(E5)}
$$

$$
\frac{\partial^2 \mathscr{F}}{\partial\rho^a(x)\partial\mu^c(y)} = \delta(x-y)\delta^{ac}, \quad \text{(E6)}
$$

$$
\frac{\partial^2 \mathscr{F}}{\partial\mu^a(x)\partial\mu^c(y)} = \begin{cases} \delta(x-y)\beta[\rho^2(x)-\rho(x)] & \text{if } a=c, \\ \delta(x-y)\beta[\rho^2(x)-q(x)] & \text{otherwise,} \end{cases}
$$
(E7)

$$
\frac{\partial^2 \mathscr{F}}{\partial\mu^a(x)\partial r^{cd}} = \begin{cases} \alpha\beta^2[q(x)\rho(x)-t(x)] & \text{if } a \neq c \neq d, \\ \alpha\beta^2[q(x)\rho(x)-q(x)] & \text{otherwise,} \end{cases}
$$
(E8)

$$
\frac{\partial^2 \mathscr{F}}{\partial r^{ab}\partial r^{cd}} = \begin{cases} \alpha^2\beta^3\big(\int q^2 - q\big) & \text{if } a=c \text{ and } b=d, \\ \alpha^2\beta^3\big(\int q^2 - s\big) & \text{if } a \neq b \neq c \neq d, \\ \alpha^2\beta^3\big(\int q^2 - t\big) & \text{otherwise,} \end{cases}
$$
(E9)

and, letting

$$
B_k \equiv \frac{k\pi}{\sin(k\pi w)} - \beta(f-q), \quad C_1 \equiv \sum_{k\geqslant 1}\frac{\beta}{B_k^2},
$$

$$
C_2 \equiv \sum_{k\geqslant 1}\frac{\beta^2(q-f^2)}{B_k^3}, \quad C_3 \equiv \sum_{k\geqslant 1}\frac{\beta^3(q-f^2)^2}{B_k^4},
$$

$$
\frac{\partial^2 \mathscr{F}}{\partial q^{ab}\partial q^{cb}}
$$

$$
= \begin{cases} -2\alpha(C_1 + 2C_2 + 2C_3) & \text{if } a=c \text{ and } b=d, \\ -4\alpha C_3 & \text{if } a \neq b \neq c \neq d, \\ -2\alpha(C_2 + 2C_3) & \text{otherwise.} \end{cases}
$$
(E10)

The eigenvector equation writes

$$
M \cdot \vec{v} = \lambda \cdot \vec{v}, \quad \text{(E11)}
$$

where $\vec{v}$ is the vector of fluctuations around the saddle point:

$$
\vec{v}(x) = \begin{bmatrix} \delta\rho^a(x) \\ \vdots \\ \delta\mu^a(x) \\ \vdots \\ \delta r^{ab} \\ \vdots \\ \delta q^{ab} \\ \vdots \end{bmatrix}. \quad \text{(E12)}
$$

According to [29] the symmetry of the matrix elements under permutation of the indices imposes to look for an orthogonal set of eigenvectors with the particular forms:

$$
\vec{v}_1(x) = \begin{cases} \delta\rho^a(x) = \delta\rho(x) & \forall a, \\ \delta\mu^a(x) = \delta\mu(x) & \forall a, \\ \delta r^{ab} = \delta r & \forall a,b, \\ \delta q^{ab} = \delta q & \forall a,b, \end{cases} \quad \text{(E13)}
$$

$$
\vec{v}_2(x) = \begin{cases} \delta\rho^a(x) = \delta\hat{\rho}(x) & \text{if } a=\theta, \\ \qquad = \delta\check{\rho}(x) & \text{otherwise,} \\ \delta\mu^a(x) = \delta\hat{\mu}(x) & \text{if } a=\theta, \\ \qquad = \delta\check{\mu}(x) & \text{otherwise,} \\ \delta r^{ab} = \delta\hat{r} & \text{if } a \text{ or } b=\theta, \\ \qquad = \delta\check{r} & \text{if } a \text{ and } b \neq \theta, \\ \delta q^{ab} = \delta\hat{q} & \text{if } a \text{ or } b=\theta, \\ \qquad = \delta\check{q} & \text{if } a \text{ and } b \neq \theta, \end{cases} \quad \text{(E14)}
$$

$$
\vec{v}_3(x) = \begin{cases} \delta\rho^a(x) = \delta\tilde{\rho}(x) & \text{if } a=\theta \text{ or } \theta', \\ \qquad = \delta\rho^*(x) & \text{otherwise,} \\ \delta\mu^a(x) = \delta\tilde{\mu}(x) & \text{if } a=\theta \text{ or } \theta', \\ \qquad = \delta\mu^*(x) & \text{otherwise,} \\ \delta r^{ab} = \delta\tilde{r} & \text{if } a=\theta \text{ and } b=\theta', \\ \qquad = \delta\tilde{\tilde{r}} & \text{if } a \text{ or } b=\theta \text{ or } \theta', \\ \qquad = \delta r^* & \text{if } a \text{ and } b \neq \theta,\theta', \\ \delta q^{ab} = \delta\tilde{q} & \text{if } a=\theta \text{ and } b=\theta', \\ \qquad = \delta\tilde{\tilde{q}} & \text{if } a \text{ or } b=\theta \text{ or } \theta', \\ \qquad = \delta q^* & \text{if } a \text{ and } b \neq \theta,\theta', \end{cases} \quad \text{(E15)}
$$

where θ and θ' are two fixed replica indices. $\vec{v}_1(x)$ and $\vec{v}_2(x)$ are called longitudinal modes; $\vec{v}_3(x)$ are called transverse or "replicon" modes. Imposing the orthogonality conditions

$$
\vec{v}_1(x)\cdot\vec{v}_2(x) = \vec{v}_1(x)\cdot\vec{v}_3(x) = \vec{v}_2(x)\cdot\vec{v}_3(x) = 0, \quad \text{(E16)}
$$

and taking the $n \to 0$ limit in Eq. (E11), we end up with the eigensystem in the longitudinal sector,

$$
-\int dy\, J_w(x-y)\delta\rho(y) + \delta\mu(x) = \lambda\,\delta\rho(x),
$$

$$
\delta\rho(x) + \beta(q-\rho)(x)\delta\mu(x) + \alpha\beta^2(q-t)(x)\delta r = \lambda\,\delta\mu(x),
$$

$$
2\alpha\beta^2\int (t-q)\delta\mu + \alpha^2\beta^3(-q+4t-3s)\delta r + \alpha\beta\delta q
$$
$$
= \lambda\,\delta r,
$$

$$
\alpha\beta\delta r - 2\alpha(C_1 - 2C_2)\delta q = \lambda\,\delta q, \qquad \text{(E17)}
$$

and in the replicon sector,

$$
\alpha^2\beta^3[-q+2t-s]\delta r^* + \alpha\beta\delta q^* = \lambda\,\delta r^*,
$$
$$
\alpha\beta\delta r^* - 2\alpha C_1\delta q^* = \lambda\,\delta q^*. \qquad \text{(E18)}
$$

For each one of the three phases (PM, CL, SG) the stability region in the (α, T) plane is delimited by lines where one of the eigenvalues vanishes.

Note that the matrix of system (E17) is not symmetric while the hessian matrix $M(x,y)$ is: a $-\frac{1}{2}$ factor appears when taking the $n \to 0$ limit since there are $\frac{n(n-1)}{2}$ two-replica-index components. This multiplicative factor does not change the points where a given eigenvalue changes signs. Nevertheless, it has the effect of giving nonreal eigenvalues.

1. Paramagnetic phase stability region

Taking $\rho(x) = f$, $q(x) = f^2$, $t(x) = f^3$, and $s(x) = f^4$ for all x in (E17) leads to a very simple system, invariant under translation in the x space. The eigenmodes in the $(\delta\rho(x), \delta\mu(x))$ sector are plane waves, $e^{2i\pi kx}$, with integer wave vectors k. The eigensystem (E17) decomposed on each Fourier mode gives the following:

(i) $k > 0$ *components of the longitudinal matrix.* The corresponding determinant is

$$
\begin{vmatrix} -\frac{\sin(\pi k w)}{\pi k} & 1 \\ 1 & \beta(f^2 - f) \end{vmatrix}. \qquad \text{(E19)}
$$

It vanishes for $\beta(k) = \frac{\pi k}{\sin(\pi k w)(f - f^2)}$ which is minimal for $k = 1$. For $f = 0.1$ and $w = 0.05$, $T_1 \approx 0.0045$.

(ii) $k = 0$ *component of the longitudinal matrix.* We get a system with determinant

$$
\begin{vmatrix} f^2 - f & \beta(f^2 - f^3) & 0 \\ 2\alpha\beta(f^3 - f^2) & \alpha\beta^2(-f^2 + 4f^3 - 3f^4) & 1 \\ 0 & 1 & -2\frac{C_1}{\beta} \end{vmatrix}. \qquad \text{(E20)}
$$

These modes appear for $(\alpha, T_0(\alpha))$ at which this determinant vanishes, i.e.,

$$
\sum_{k\geqslant 1}\left[\frac{T_0(\alpha)\,k\pi}{f(1-f)\sin(k\pi w)} - 1\right]^{-2} = \frac{1}{2\alpha}. \qquad \text{(E21)}
$$

(iii) *Replicon modes.* These modes solve a system with determinant

$$
\begin{vmatrix} \alpha\beta^2(-f^2 + 2f^3 - f^4) & 1 \\ 1 & -2\frac{C_1}{\beta} \end{vmatrix}. \qquad \text{(E22)}
$$

This defines the same stability line (E21) as found above.

To sum up, the paramagnetic phase is stable at high temperatures; when T decreases at fixed α, it becomes unstable at $T_{\mathrm{PM}}(\alpha) = \max\{T_0(\alpha), T_1\}$ as depicted in Fig. 3.

2. Glassy phase stability region

We find a uniform solution to the saddle point equations (28) with $q > f^2$ only for $T < T_{\mathrm{PM}}(\alpha)$: the region of existence of the glassy phase hence corresponds to the region where the PM solution is unstable. In this region we find that the RS solution is always stable against longitudinal modes (E17) and always unstable against transverse modes (E18). The replica symmetric ansatz is therefore not correct in the case of the glassy phase.

3. Clump phase stability region

(i) *Longitudinal modes.* Due to the x dependence in this phase, we must use a numerical approach in a discretized space to study the eigenvalues of the longitudinal matrix. Since computation time for the matrix diagonalization limits dramatically the number of points in the discretization, we chose to study the longitudinal stability with a different method. Scanning the (α, T) plane, $\rho(x)$ is computed by solving iteratively the saddle-point equations (28) starting from an initial clump; the line of stability corresponds to the points where the clump collapses, i.e., the iteration converges to a uniform activity $\rho(x) = f\ \forall x$. The result is shown in Fig. 7 in the main text.

(ii) *Replicon modes.* For all α, T, we compute numerically q, t, s by solving iteratively the saddle-point equations as before, allowing us to calculate the determinant of system (E18). We looked for the line where this determinant vanishes. We found that replica symmetry breaking is limited to a small region in the low T–high α edge of the region of longitudinal stability; see Fig. 7.

[1] J. O'Keefe and J. Dostrovsky, Brain Res. **34**, 171 (1971).

[2] P. Andersen, R. Morris, D. Amaral, T. Bliss, and J. O'Keefe, *The Hippocampus Book* (Oxford University Press, New York, 2006).

[3] L. Thompson and P. Best, J. Neurosci. **9**, 2382 (1989).

[4] E. Bostock, R. U. Muller, and J. L. Kubie, Hippocampus **1**, 193 (1991).

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

[6] J. L. Kubie and R. U. Muller, Hippocampus **1**, 240 (1991).

[7] G. J. Quirk, R. U. Muller, and J. L. Kubie, J. Neurosci. **10**, 2008 (1990).

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

[9] M. Fyhn, S. Molden, M. P. Witter, E. I. Moser, and M.-B. Moser, Science **305**, 1258 (2004).

[10] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, Nature (London) **436**, 801 (2005).

[11] E. I. Moser, E. Kropff, and M.-B. Moser, Annu. Rev. Neurosci. **31**, 69 (2008).

[12] E. T. Rolls, *Neural Models of Plasticity: Experimental and Theoretical Approaches* (Academic, San Diego, 1989), pp. 240–265.

[13] M. Tsodyks, Hippocampus **9**, 481 (1999).

[14] M. Tsodyks and T. Sejnowski, Int. J. Neural Syst. **6**, 81 (1995).

[15] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).

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

[17] D. J. Amit, *Modeling Brain Function. The World of Attractor Neural Networks* (Cambridge University Press, Cambridge, England, 1989).

[18] F. P. Battaglia and A. Treves, Phys. Rev. E **58**, 7738 (1998).

[19] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **107**, 1648 (2010).

[20] K. Jezek, E. J. Henriksen, A. Treves, E. I. Moser, and M.-B. Moser, Nature (London) **478**, 246 (2011).

[21] R. Monasson and S. Rosay (unpublished).

[22] J. L. Lebowitz and O. Penrose, J. Math. Phys. **7**, 98 (1966).

[23] I. Ginzburg and H. Sompolinsky, Phys. Rev. E **50**, 3171 (1994).

[24] A very weak reentrance was found in the RS solution in subsequent works [30,31].

[25] D. G. Amaral, N. Ishizuka, and B. Claiborne, Prog. Brain Res. **83**, 1 (1990).

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

[27] H. Sompolinsky, Phys. Rev. A **34**, 2571 (1986).

[28] J. L. Van Hemmen, *Heidelberg Colloquium on Glassy Dynamics: in Proceedings of a Colloquium on Spin Glasses, Optimization and Neural Networks Held at the University of Heidelberg, June 9–13, 1986* (Springer Verlag, Berlin, 1987), Vol. 275.

[29] J. R. L. De Almeida and D. J. Thouless, J. Phys. A: Math. Gen. **11**, 983 (1978).

[30] J.-P. Naef and A. Canning, J. Phys. I **2**, 247 (1992).

[31] H. Steffan and R. Kühn, Z. Phys. B **95**, 249 (1994).