Analytic solution of neural network with disordered lateral inhibition

Kosuke Hamaguchi* and J. P. L. Hatchett
Mathematical Neuroscience Laboratory, RIKEN BSI, 2-1 Hirosawa Wako, Saitama, Japan

Masato Okada
Department of Complexity Science and Engineering, University of Tokyo, Kashiwanoha 5-1-5, Kashiwa, Chiba, 277-8561, Japan and PRESTO, JST, Japan

(Received 23 November 2005; published 4 May 2006)

The replica method has played a key role in analyzing systems with disorder, e.g., the Sherrington-Kirkpatrick (SK) model, and associative neural networks. Here we study the influence of disorder in the lateral inhibition type interactions on the cooperative and uncooperative behavior of recurrent neural networks by using the replica method. Although the interaction between neurons has a dependency on distance, our model can be solved analytically. Bifurcation analysis identifies the boundaries between paramagnetic, ferromagnetic, spin-glass, and localized phases. In the localized phase, the network shows a bump like activity, which is often used as a model of spatial working memory or columnar activity in the visual cortex. Simulation results show that disordered interactions can stabilize the drift the of bump position, which is commonly observed in conventional lateral inhibition type neural networks.

DOI: 10.1103/PhysRevE.73.051104 PACS number(s): 05.20.−y, 75.10.Nr, 87.19.La

I. INTRODUCTION

Human and animal brains can capture, store, and retrieve complex patterns in the real world. The retrieval process was first successfully described by Hopfield [1]. Advances in mean-field spin-glass theory have revealed the stability of the associative memory embedded with extensive numbers of patterns, and the memory capacity of the network [2,3]. In a capturing process such as visual perception, Hubel and Wiesel have found columnar activity in the cat visual cortex responding to a bar stimuli with a specific orientation [4]. Neural networks that can model local excitation are often described by lateral inhibition type networks. The lateral inhibition denotes recurrent excitation with nearby neurons and inhibition between distant neurons [5]; it is also referred to as the Mexican-hat-type interaction. In these models, a bump activity, which is the locally activated network state, is stable depending on the input and configuration of the network interaction. The stability of one bump has been analyzed extensively [6–10] in several neuron models ranging from analog neurons to spiking neurons.

Recent progress in the analysis of lateral inhibition type networks without connection randomness shows that the mean field theory can be applicable [8] even to neural networks that have distance-dependent interactions. The trick is to represent the connection function with a combination of linearly independent functions, e.g., a Fourier series. lateral inhibition is usually described by the difference of Gaussians, or the second derivative of a Gaussian, but Fourier series expansion of the interaction function with lower order and periodic boundary conditions allow us to describe the system state using order parameters. Even though most of the lateral inhibition type models do not include random interactions, it was clear from the start that deterministic interactions is just a simplification of the real system. To analyze a system with random interactions, one finds the replica method is a powerful tool. In this paper, we study the stable states of an Ising spin neural network with disordered lateral-inhibitory interactions.

II. MODEL DEFINITIONS

We study an Ising spin neural network, which is modeled by an N-neuron state vector $\boldsymbol{S}=(S_{\theta_1},S_{\theta_2},\ldots,S_{\theta_N})\in\{-1,1\}^N$. Here $S_{\theta_i}=1$ if neuron i fires, and $S_{\theta_i}=-1$ if it is at rest. Neuron i is located at angle $\theta_i=\frac{2\pi i}{N}-\pi$ on a one-dimensional ring indexed by $\theta_i\in[-\pi,\pi)$ as shown in Fig. 1. The Hamiltonian of the system we are going to study is

$$H(\boldsymbol{S})=-\frac{1}{2}\sum_{\theta_i\neq\theta_j}J_{\theta_i\theta_j}S_{\theta_i}S_{\theta_j}-h\sum_i S_{\theta_i}, \tag{1}$$

where h is a common external input to neurons. The interaction $J_{\theta_i\theta_j}$ is defined to be disordered lateral inhibition, it is a function only of $(\theta_i-\theta_j)$ plus noise:

$$J_{\theta_i\theta_j}=\frac{J_0}{N}+\frac{J_1}{N}\cos(\theta_i-\theta_j)+\xi_{\theta_i\theta_j}, \tag{2}$$

$$\xi_{\theta_i\theta_j}\sim\mathcal{N}\left(0,\frac{J^2}{N}\right), \tag{3}$$

where J_0 is a uniform ferromagnetic interaction, J_1 is a lateral inhibition type interaction, and $\xi_{\theta_i\theta_j}=\xi_{\theta_j\theta_i}$ are quenched disordered interactions term independently drawn from an identical Gaussian distribution with mean 0, and variance J^2/N. Thus, there is no correlation between $\xi_{\theta_i\theta_j}$ and $\xi_{\theta_i\theta_k}$. Here we assume that J_0 and J_1 are non-negative real numbers, and J can be either a positive or negative real number.

*Electronic address: hammer@brain.riken.jp; URL: http://www.brain.riken.jp/labs/mns/hammer/

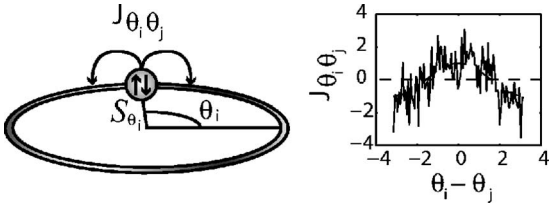

FIG. 1. Network model: N spins indexed with $\theta_i \in [-\pi, \pi)$ on a ring are interacting with recurrent disordered lateral inhibition.

III. REPLICA CALCULATION OF THE FREE ENERGY AND ORDER PARAMETERS

In this section, we calculate the free energy per neuron and relevant order parameters by using the replica method. To calculate the partition function Z, let us circumvent averaging $\ln Z$ by using

$$\ln Z = \lim_{n\to 0} \frac{Z^n - 1}{n}, \tag{4}$$

which allows us to compute $\ln Z$ from Z^n. The latter is a partition function of n copies, or *replicas*, of the original system. We write

$$Z^n = \mathrm{Tr}_S \exp\left(-\beta \sum_{\alpha=1}^{n} H(S^\alpha) \right), \tag{5}$$

where α is the replica index running from 1 to n. The free energy $F=-\beta^{-1}\ln Z$ is a function of quenched disorder $\{\epsilon_{\theta_i\theta_j}\}$. Instead of studying free energy based on each realization of disorder $\{\epsilon_{\theta_i\theta_j}\}$, we study averaged F over the probability distribution of $\epsilon_{\theta_i\theta_j}$. We refer to this average as the configurational average and write it as $[\cdots]$.

The configurational average of the free energy is given via

$$[Z^n] = \mathrm{Tr}_S \int \prod_{\alpha<\beta} dq^{\alpha\beta} \prod_\alpha dm_0^\alpha \prod_\alpha dm_c^\alpha \prod_\alpha dm_s^\alpha \times \exp\left\{ -\frac{NJ^2\beta^2}{2} \sum_{\alpha<\beta} (q^{\alpha\beta})^2 - \frac{N\beta J_0}{2} \sum_\alpha (m_0^\alpha)^2 - \frac{N\beta J_1}{2} \sum_\alpha \{(m_c^\alpha)^2 + (m_s^\alpha)^2\} + L + \frac{\beta^2 J^2 Nn}{4} \right\}, \tag{6}$$

where

$$L = \sum_{\theta_i} \left\{ J^2\beta^2 \sum_{\alpha<\beta} S_{\theta_i}^\alpha S_{\theta_i}^\beta q^{\alpha\beta} + \beta J_0 \sum_\alpha S_{\theta_i}^\alpha m_0^\alpha + \beta J_1 \sum_\alpha [m_c^\alpha \cos(\theta_i) + m_s^\alpha \sin(\theta_i)] S_{\theta_i}^\alpha + \beta h \sum_\alpha S_{\theta_i}^\alpha \right\}. \tag{7}$$

Here we defined the following order parameters:

$$m_0^\alpha = N^{-1} \sum_i S_{\theta_i}^\alpha, \quad q^{\alpha\beta} = N^{-1} \sum_i S_{\theta_i}^\alpha S_{\theta_i}^\beta,$$

$$m_c^\alpha = N^{-1} \sum_i \cos(\theta_i) S_{\theta_i}^\alpha, \quad m_s^\alpha = N^{-1} \sum_i \sin(\theta_i) S_{\theta_i}^\alpha. \tag{8}$$

where m_0 is the magnetization, q is the spin-glass order parameter, and m_c and m_s are order parameters which show how spins are aligned in the same direction locally, but not globally, around $\theta=0, \pi/2$. The network state is said to be in the bump state, or locally activated, if m_c or m_s are nonzero. We note that m_c and m_s depend on the position of the bump, but our interests are rather the size and the position of the bump. So we introduce the following transformation to separate the size and position of the bump:

$$(m_1^\alpha)^2 = (m_c^\alpha)^2 + (m_s^\alpha)^2, \tag{9}$$

$$\phi^\alpha = \tan^{-1}(m_s^\alpha / m_c^\alpha). \tag{10}$$

The alternative order parameters, m_1 and ϕ, are global measure of the activity profiles that indicate the degree of activity localization and its angle, respectively. It is easy to find that the third term in Eq. (7) can be rewritten as $\beta J_1 \Sigma_\alpha m_1^\alpha [\cos(\theta_i - \phi^\alpha)] S_{\theta_i}^\alpha$. The difference between our model and the SK model is the localized magnetization, i.e., the (m_1, ϕ) terms.

A. Replica symmetry ansatz

First, we derive the replica symmetric solution, and then consider the replica symmetry breaking conditions later. Here we assume replica symmetry, that is $q^{\alpha\beta}=q, m_0^\alpha=m_0$, $m_1^\alpha=m_1$, $\phi^\alpha=\phi$. The introduction of replica symmetry assumption and Gaussian identity $\exp\left(\frac{b}{2}S_{\theta_i}^2\right) = \int Dz_{\theta_i} \exp(\sqrt{b} S z_{\theta_i})$, where $Dz=(2\pi)^{-1/2}\exp(-z^2/2)$ simplify $\mathrm{Tr}_S \exp(L)$ in Eq. (6) to

$$\begin{aligned} \mathrm{Tr}_S \exp(L) &= \mathrm{Tr}_S \prod_{\theta_i} \int Dz_{\theta_i} \exp\left\{ [\beta J \sqrt{q} z_{\theta_i} + \beta J_0 m_0 + \beta J_1 m_1 \cos(\theta_i - \phi) + \beta h] \left(\sum_\alpha S_{\theta_i}^\alpha \right) \right\} \\ &\quad \times \exp\left(-\frac{Nn\beta^2 J^2}{2} q + \frac{Nn\beta^2 J^2}{4} \right) \\ &= \prod_{\theta_i} \int Dz_{\theta_i} [2\cosh \beta \tilde{H}(z_{\theta_i}, \theta_i - \phi)]^n \\ &\quad \times \exp\left(-\frac{Nn\beta^2 J^2}{2} q + \frac{Nn\beta^2 J^2}{4} \right), \end{aligned} \tag{11}$$

where $\tilde{H}(z_{\theta_i}, \theta_i) = J\sqrt{q} z_{\theta_i} + J_0 m_0 + J_1 m_1 \cos(\theta_i) + h$. From this, we get

$$[Z^n] = \int dq dm_0 dm_c dm_s \exp\left\{ \frac{NnJ^2\beta^2}{4} [(1-n)q^2 - 2q + 1] - \frac{Nn\beta J_0}{2} m_0^2 - \frac{Nn\beta J_1}{2} m_1^2 + \sum_{\theta_i} \ln\left(\int Dz_{\theta_i} [2\cosh \beta \tilde{H}(z_{\theta_i}, \theta_i - \phi)]^n \right) \right\}. \tag{12}$$

The free energy per neuron $\beta[f]=-\frac{1}{N}[\ln Z]=-\lim_{n\to 0}\frac{[Z^n]-1}{Nn}$ is evaluated in the thermodynamic limit. The summation of θ_i is now replaced by an integral over θ. Free energy per neuron is

$$-\beta[f]=-\frac{\beta J_0}{2}m_0^2-\frac{\beta J_1}{2}m_1^2+\frac{\beta^2 J^2}{4}(1-q)^2+\frac{1}{2\pi}\int_{-\pi}^{\pi}d\theta\int_{-\infty}^{\infty}Dz\,\ln[2\cosh\beta\tilde{H}(z,\theta-\phi)]. \tag{13}$$

The order parameters are determined through the saddle point equations as below:

$$m_0=\int\frac{d\theta}{2\pi}\int Dz\tanh[\beta\tilde{H}(z,\theta-\phi)], \tag{14}$$

$$m_1=\int\frac{d\theta}{2\pi}\int Dz\cos(\theta-\phi)\tanh[\beta\tilde{H}(z,\theta-\phi)], \tag{15}$$

$$q=\int\frac{d\theta}{2\pi}\int Dz\tanh^2[\beta\tilde{H}(z,\theta-\phi)]. \tag{16}$$

The bump is neutrally stable in the direction of ϕ because the partial derivative of the free energy with respect to ϕ is zero,

$$\frac{\partial f}{\partial\phi}=-\int_{-\pi}^{\pi}\frac{d\theta}{2\pi}\int_{-\infty}^{\infty}Dz\tanh\beta\tilde{H}(z,\theta-\phi)\sin(\theta-\phi)=0. \tag{17}$$

This is natural because the physical meaning of ϕ is the position of the bump, and the bump is stable anywhere on the ring layer as long as the external input is spatially uniform, i.e., $h=$const. In this case, the Hamiltonian is rotationally invariant, and the network is said to have the line attractor. Henceforth, we consider only the m_0, m_1 and q order parameters when discussing the stability of the system. Equations (14)–(16) have four types of solutions, which are locally stable in the direction of m_0, m_1, and q. The stable states of the network are (1) paramagnetic (P): $m_0=m_1=q=0$, (2) ferromagnetic (F): $m_0\neq 0$, $m_1=0$, $q\neq 0$, (3) localized (L): $m_0=0$, $m_1\neq 0$, $q\neq 0$, and (4) spin-glass (SG): $m_0=m_1=0$, $q\neq 0$.

IV. REPLICA-SYMMETRY BREAKING CONDITION

In this section, we study the condition for the replica-symmetry breaking (RSB) condition via a de Almeida–Thouless-type argument [11]. We come back to the partition function in Eq. (6). Since terms in the exponential in Eq. (6) are proportional to N, we can evaluate the integral with saddle-point method:

$$[Z^n]\approx\exp\left\{-\frac{NJ^2\beta^2}{2}\sum_{\alpha<\beta}(q^{\alpha\beta})^2-\frac{N\beta J_0}{2}\sum_{\alpha}(m_0^{\alpha})^2-\frac{N\beta J_1}{2}\sum_{\alpha}(m_1^{\alpha})^2\right\}\exp(\ln\mathrm{Tr}_S e^L)\exp\left(\frac{\beta^2J^2Nn}{4}\right)\approx 1+Nn\left\{-\frac{J^2\beta^2}{2n}\sum_{\alpha<\beta}(q^{\alpha\beta})^2-\frac{\beta J_0}{2n}\sum_{\alpha}(m_0^{\alpha})^2-\frac{\beta J_1}{2n}\sum_{\alpha}(m_1^{\alpha})^2+\frac{1}{Nn}\ln\mathrm{Tr}_S e^L+\frac{\beta^2J^2}{4}\right\}. \tag{18}$$

By using the transformation $y^{\alpha\beta}\equiv\beta Jq^{\alpha\beta}$, $x_0^{\alpha}\equiv(\beta J_0)^{1/2}m_0^{\alpha}$, $x_1^{\alpha}\equiv(\beta J_1)^{1/2}m_1^{\alpha}$, the free energy per neuron can be written as

$$-\beta[f]=\lim_{n\to 0}\frac{[Z^n]-1}{nN}=-\frac{1}{2n}\sum_{\alpha<\beta}(y^{\alpha\beta})^2-\frac{1}{2n}\sum_{\alpha}(x_0^{\alpha})^2-\frac{1}{2n}\sum_{\alpha}(x_1^{\alpha})^2+\frac{1}{Nn}\ln\mathrm{Tr}_S e^L+\frac{\beta^2J^2}{4}. \tag{19}$$

From the saddle-point condition, we have $\frac{\partial[f]}{\partial y^{\alpha\beta}}=0$, $\frac{\partial[f]}{\partial x_0^{\alpha}}=0$, and $\frac{\partial[f]}{\partial x_1^{\alpha}}=0$. To check the stability of the replica-symmetry condition, we first show second partial derivatives of $N^{-1}\ln\mathrm{Tr}_S e^L$ around the replica-symmetric point in the directions of $y^{\alpha\beta}$ and $y^{\gamma\delta}$,

$$\begin{aligned}&\frac{\partial^2}{\partial y^{\alpha\beta}\partial y^{\gamma\delta}}N^{-1}\ln\mathrm{Tr}\,e^L\\&=(\beta J)^2\left\{\frac{\mathrm{Tr}_S[N^{-1}\sum_\theta S_\theta^\alpha S_\theta^\beta S_\theta^\gamma S_\theta^\delta\exp(L_0)]}{\mathrm{Tr}_S\exp(L_0)}-\frac{\{\mathrm{Tr}_S[N^{-1}\sum_\theta S_\theta^\alpha S_\theta^\beta\exp(L_0)]\}\{\mathrm{Tr}_S[N^{-1}\sum_\theta S_\theta^\gamma S_\theta^\delta\exp(L_0)]\}}{[\mathrm{Tr}_S\exp(L_0)]^2}\right\}\\&=(\beta J)^2\{\langle S_\theta^\alpha S_\theta^\beta S_\theta^\gamma S_\theta^\delta\rangle_{L_0}-\langle S_\theta^\alpha S_\theta^\beta\rangle_{L_0}\langle S_\theta^\gamma S_\theta^\delta\rangle_{L_0}\}.\end{aligned} \tag{20}$$

Here we defined $\langle\cdots\rangle_{L_0}=\int\frac{d\theta}{2\pi}\int Dz\exp(L_0)$, and L_0 corresponds to the replica-symmetry (RS) L case. The details of the calculation of taking traces in Eq. (20) is shown in the Appendix. From Eq. (20), expansion of $[f]$ around the RS to the second order, using $y^{\alpha\beta}=y+\eta^{\alpha\beta}$ gives

$$[f] = [f]_{L_0} + \frac{1}{2}\sum_{\alpha<\beta}\sum_{\gamma<\delta}\frac{\partial^2[f]}{\partial y^{\alpha\beta}\partial y^{\gamma\delta}}\eta^{\alpha\beta}\eta^{\gamma\delta}$$

$$= [f]_{L_0} + \sum_{\alpha<\beta}\sum_{\gamma<\delta}\frac{1}{2}[G_{(\alpha\beta)(\gamma\delta)}\eta^{\alpha\beta}\eta^{\gamma\delta}], \tag{21}$$

where G is the Hessian matrix, and

$$G_{(\alpha\beta)(\alpha\beta)} = 1 - (\beta J)^2(1 - \langle S_\theta^\alpha S_\theta^\beta\rangle_{L_0}) \equiv P,$$

$$G_{(\alpha\beta)(\alpha\gamma)} = -(\beta J)^2(\langle S_\theta^\beta S_\theta^\gamma\rangle_{L_0} - \langle S_\theta^\alpha S_\theta^\beta\rangle_{L_0}^2) \equiv Q,$$

$$G_{(\alpha\beta)(\gamma\delta)} = -(\beta J)^2(\langle S_\theta^\alpha S_\theta^\beta S_\theta^\gamma S_\theta^\delta\rangle_{L_0} - \langle S_\theta^\alpha S_\theta^\beta\rangle_{L_0}\langle S_\theta^\gamma S_\theta^\delta\rangle_{L_0}) \equiv R. \tag{22}$$

We look for an eigenvector that are symmetric under interchange of all but two of the indices. That is the eigenvector μ which has $\eta^{\zeta\xi}=c$ for specific replicas ζ, ξ, and $\eta^{\zeta\alpha} = \eta^{\xi\alpha}=d$, for any α, and $\eta^{\alpha\beta}=e$ for the other replicas. From the orthogonality conditions from the other two eigenvectors that are symmetric under interchange of all indices, and all but one indices, the eigenvalue λ is given as

$$\lambda = P - 2Q + R. \tag{23}$$

The condition that the eigenvalue given by Eq. (23) is positive can be written in the form

$$\frac{1}{\beta^2 J^2} > \int_{-\pi}^{\pi}\frac{d\theta}{2\pi}\int Dz\ \mathrm{sech}^4[\beta H(z,\theta)]. \tag{24}$$

This gives the replica symmetry breaking condition.

V. RESULTS

A. Phase diagrams and order parameters

It is of our main interest to see how networks with lateral inhibition type interactions differ from those with ferromagnetic interactions. For this purpose, hereafter, we set $h=0$ for simplicity. From Eqs. (14)–(16), we analyze the stability of P, F, L, and SG phases. First, we apply a bifurcation analysis to compute the second-order transition lines away from the P phase. We assume that $m_0=m_1=q=0$, and that the deviation from P state in the direction of F, L, SG is very small. By expanding Eqs. (14)–(16) up to first order, we can obtain the stability condition of P through F, L, and SG direction as follows:

$$P \to F: \beta J_0 = 1, \tag{25}$$

$$P \to L: \frac{\beta J_1}{2} = 1, \tag{26}$$

$$P \to \mathrm{SG}: \beta J = 1. \tag{27}$$

Similar bifurcation analyses gives the transitions from SG $\to\{F,L\}$ directions as follows:

$$\mathrm{SG} \to F: \beta J_0(1-q) = 1, \tag{28}$$

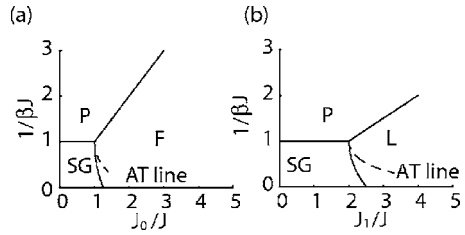

FIG. 2. Phase diagrams showing the limit of stability of the AT solution in the absence of an external input. (a) No lateral inhibition ($J_1=0$) which is equivalent to the SK model. (b) No ferromagnetic interaction ($J_0=0$) in which the localized phase exists instead of the ferromagnetic phase in (a) At low temperature and small disorder.

$$\mathrm{SG} \to L: \beta J_1(1-q)/2 = 1. \tag{29}$$

The local stability of F and L phases are also obtained as

$$L \to F: \beta J_0(1-q) = 1, \tag{30}$$

$$F \to L: \beta J_1(1-q)/2 = 1. \tag{31}$$

Phase diagrams in Fig. 2 show the stability of the F and L phases in $(1/\beta J, J_0/J)$ and $(1/\beta J, J_1/J)$ space. Transitions from P to L occur at lower temperatures, indicating that L phase is less stable than the F phase. The AT line also shifts in the direction of J_1/J, indicating again that lateral-inhibitory interactions are less effective at maintaining the L state than ferromagnetic interactions are in the F state.

Next we study the interaction between the ferromagnetic (J_0) and lateral-inhibitory (J_1) interaction. The phase diagram is analyzed in $(\beta J_0, \beta J_1)$ plane with fixed disorder $J=\{0, 0.5, 2\}$ as shown in Figs. 3(a)–3(c). Depending on the relative strength of βJ_0 and βJ_1, F or L phases are stable once they exceed certain thresholds. Between these two,

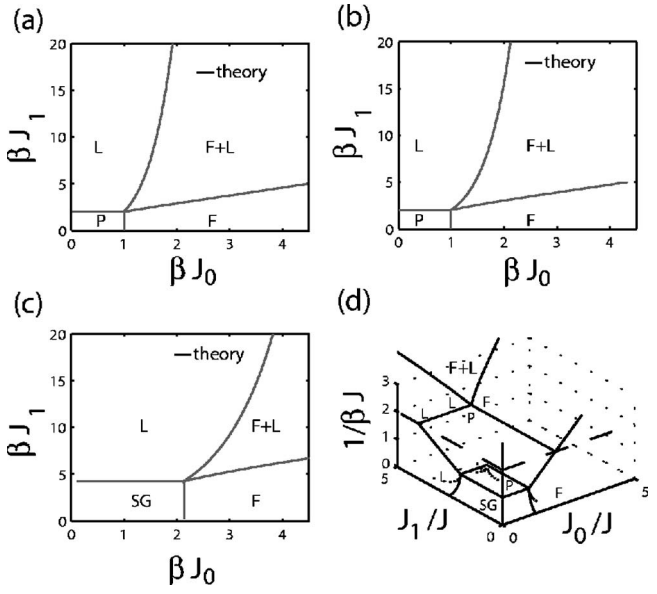

FIG. 3. Phase diagrams with fixed J in the intersection of $(\beta J_0, \beta J_1)$ plane. From (a) to (c), J is set to $J=\{0, 0.5, 2\}$, respectively. (d) The three-dimensional phase diagram in $(J_0/J, J_1/J, 1/\beta J)$.

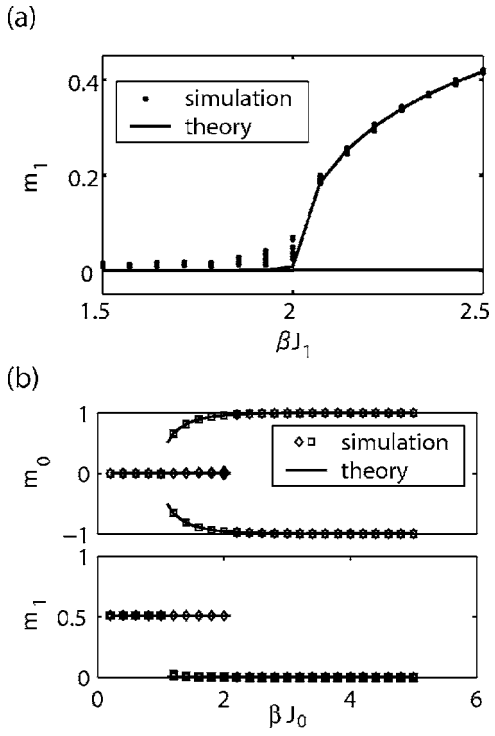

FIG. 4. (a) Phase transition from P to L, with fixed parameters $\beta=1$, $J=0$, $J_0=0.5$ with varying J_1 parameter from 1.5 to 2.5. Solid line is obtained by solving the self-consistent Eqs. (14)–(16). In simulations, $N=40\,000$ spins are evolved with Glauber dynamics with initial state $m_0=0$, $m_1=1/\pi$ until it reaches stable states, and m_1 values obtained from 10 trials are plotted. (b) Phase transition from L to F, with fixed parameters $\beta=1$, $J=0$, $J_1=3$ with varying J_0 parameter from 0 to 5. Theoretical results are obtained as in the same way in (a). Diamonds and squares represent order parameters calculated in simulations with $N=8000$, and initial states are set to different values; one is ferromagnetic state, and the other is localized state. Squares initial states are sets to either $m_0=\pm 1$, $m_1=0$, and initial states of diamonds are $m_0=0$, $m_1=1/\pi$.

there are bistable regions where both F and L phases are locally stable. The bistable region gradually shrinks as J increases. The P phase region does not change its size in $(\beta J_0, \beta J_1)$ space for $\beta J<1$. Once the spin-glass interaction exceeds the threshold value $\beta J=1$, the P phase is replaced by SG phase. As βJ increases, as one would expect, the SG region expands. Since we have three free order parameters in our model, we give the full three-dimensional phase diagram. $(J_0/J, J_1/J, 1/\beta J)$ space is shown in Fig. 3(d).

In Fig. 4, we compare the results of simulations and theory in the region of the phase boundary. In simulations spins evolved at maximum 2000 time steps with Glauber dynamics until they reached a meta-stable equilibrium point. The time is in units of updates per spin. Here, parameters are set to $\beta=1$, $J=0$. Figure 4(a) shows the order parameters m_1 along $J_0=0.5$ line where a phase transition occurs from P to L. The simulations confirm our theoretical prediction that the phase transition point is at $\beta J_1=2$ from P to L. In Fig. 4(b), we also show the phase transitions from L to F phase. We set initial states to different values in the simulations, so that we can observe bistable $F+L$ states. Parameter of this network is the same as above except that $J_1=3$ and J_0 is varied. In relatively small βJ_0 region $(0<\beta J_0<1)$, the L phase is stable. In relatively high βJ_0 region $(2>\beta J_0)$, the F phase is stable. In between, we observe bistable states $F+L$.

B. Motion of the bump position

In this subsection, we study the relationship between the disorder and the stability of the bump state through rotational direction. Throughout this subsection, we consider the case where the network is in a L phase where the bump state is stable, otherwise we cannot define the position of the bump ϕ. The lateral inhibition network without disorder has neutrally stable states in the direction of ϕ, unless there is spatially modulated external input to the system. The bump state is stable anywhere in the ring network, and a shift in the ϕ direction does not change the free energy of the state. Therefore, even small noise from the external heat bath or finite system size causes the bump position fluctuate and move around the ring network as a Brownian particle diffuses on a frictionless surface.

Here, our interest is on the bump state in the brain. The model of working memory requires the network to keep active without an external input, and these lateral inhibition type networks are often adopted as a model of spatial working memory [8]. It is known, however, that the bump position in lateral inhibition network is unstable and easily fluctuates, which corresponds to memory loss. In leaky integrate-and-fire neuron models of spatial working memory, the bump state shows systematic drift for a small synaptic time constant [9]. Therefore it is of wide interest to prevent the bump state drift in spatial working memory models. In the model studied here, the noise from the system itself vanishes in the thermodynamic limit, and the bump position is stable. However, in a biologically realistic situation, neural network consists of a large but finite number of neurons. Taking this point into account, we study the drift of bump position in a finite sized system by using Monte Carlo simulations.

What happens when we introduce disorder? It is known that disorder leads to a spinglass phase, and energy landscape takes on many-valley structure. Under the region where replica symmetry is broken, the system is nonergodic because some valleys are separated by infinitely high energy barriers. Therefore, we expect that bump position can be stabilized by the introduction of disorder, because it embeds many-valley structure in the flat-energy landscape of neutrally-stable line attractor, and the bump state is trapped in local minima along the ring network.

To test this hypothesis, we studied how disorder affected the variance of $\phi(t)$ in a finite size network of $N=4000$ neurons. Figure 5(a) shows the dynamics of ϕ, with fixed parameters, $\beta=1$, $J_0=0$, $J_1=8$, and various J which give $J_1/J=[1.8, 2.5, 3, 4, 8, 80, 800, \infty]$. Here the disorder $\{\epsilon_{\theta_i\theta_j}\}$ is quenched to one realization throughout the simulations. At $t=0$, we set the network to a bump state with $m_0=0$, $m_1=1/\pi$ at 6 different positions of the ring, which are $\phi(0)=[-\pi, -2/3\pi, -1/3\pi, 0, 1/3\pi, 2/3\pi]$. For each initial position $\phi(0)$, we calculate 5 trials of $\phi(t)$ dynamics for 2000

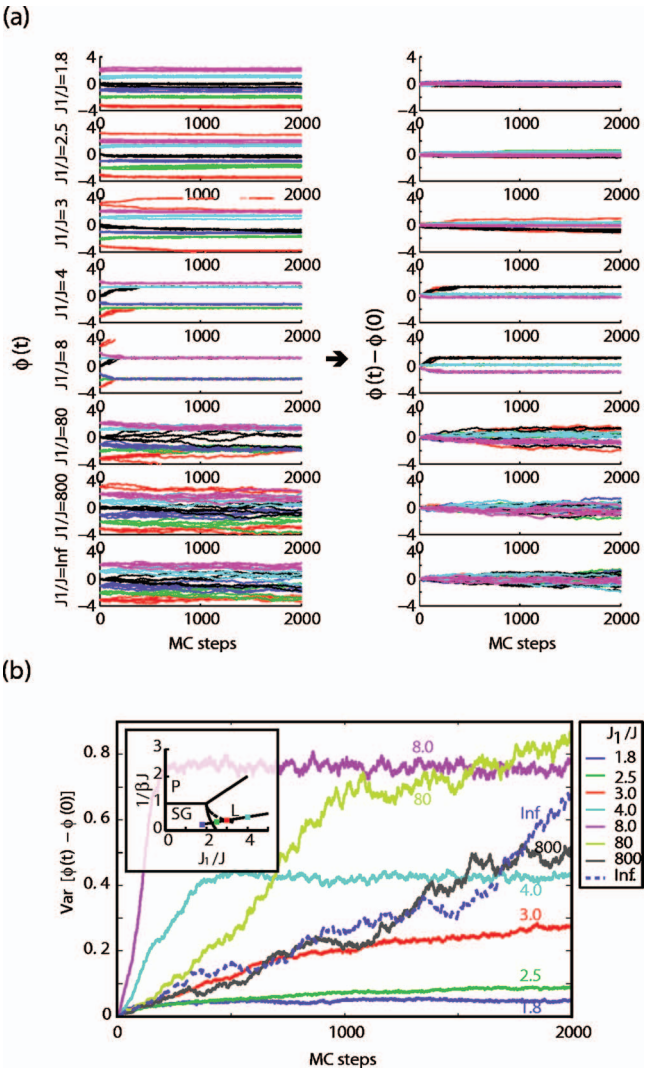

FIG. 5. (Color) (a) Dynamics of $\phi(t)$ on left column and $\phi(t)-\phi(0)$ on the right column. Some parameters are fixed, $\beta=1$, $J_0=0$, $J_1=8$, and the order of disorder J is varied to give $J_1/J=[1.8,2.5,3,4,8,80,\infty]$. For every 6 initial positions $\phi(0)$, 5 trials of spin dynamics are performed, therefore 30 trials are performed for one parameter. The different line colors represent different initial position $\phi(0)$. (b) Dynamics of variance of $\phi(t)-\phi(0)$ calculated from the 30 trials. In the insets, a line of $(\beta J)^{-1}=8J^{-1}$ is shown. The squares on the line corresponds to the parameter $J_1/J=[1.8,2.5,3,4]$ with fixed $J_1=8$, which colors correspond to the line colors.

steps. As a total, 30 trials of simulations are calculated for one fixed J.

In Fig. 5(a), we plotted $\phi(t)$ on the left column, and $\phi(t)-\phi(0)$ on the right. Different line colors represent different initial positions $\phi(0)$. The bottom panels in Fig. 5(a) is the case where $J_1/J=\infty$ $(J=0)$ and the network has line attractor, thus the dynamics of ϕ is Brownian motion. Introduction of disorder to some small extent, such as $J_1/J=800$ $(J=0.01)$ case does not change the network dynamics qualitatively because the fluctuation of the dynamics is stronger than the energy barrier. Further increasing the disorder reduces the fluctuation of ϕ as shown in Fig. 5, $J_1/J=4$ or 8. However, $\phi(t)$ converges to specific positions. Such limited, small number of stable states in the network dynamics is undesired property in the light of configuring spatial working memory model because spatial working memory requires the model to have continuous, or at least very large number of, stable points in ϕ. When disorder J exceeds a certain threshold, RSB occurs and the bump can stay in nearby region from their initial positions. For this parameter set, RSB occurs at $J_1/J\simeq 3$. The dynamics of $\phi(t)$ after the RSB are shown in top three panels of Fig. 5(a) and their parameters in the $(1/\beta J, J_1/J)$ space are shown in the inset of Fig. 5(b).

Figure 5(b) shows the dynamics of variance of $\phi(t)-\phi(0)$, var$[\phi(t)-\phi(0)]$, which is calculated from 30 trials of simulations with different initial positions $\phi(0)$'s. They illustrate how the bumps deviate from their initial positions. In the neutrally stable states with $J_1/J=\infty$ $(J=0)$, variance increases linearly, which indicates that the bump undergoes Brownian motion. Introduction of a small degree of disorder such as $J_1/J=80$ $(J=0.1)$ leads to higher variance in the early phase of the dynamics, because a small number of attractors accelerates the drift of bumps. When the disorder becomes stronger $(J_1/J=3,4,8)$, the transition of bumps to those attracting points becomes fast and the bumps quickly settle down to those points. Such behavior is clearly shown in the plateau-like dynamics of the variance. After the disorder has reached a certain value, the bump states can settle to nearby stable equilibrium points and do not deviate largely from their initial position. $J_1/J=2.5$ is the region where replica symmetry is breaking and localized activity is stable.

So far, we considered the dynamics of the bump in the L phase. We actually show a case of SG phase, $J_1/J=1.8$. In the analogy of remnant magnetization in SG phase, the full transitions from L to SG state is extremely slow [12]. Since the initial state of the network is set to bump states $(m_1=1/\pi)$, the network states still shows nonvanishing m_1 values during this simulation duration. This parameter gives the least bump fluctuation in the parameters used. We confirmed that all the states shown in above have a nonzero value of m_1. These results indicate that disorder can change the energy landscape, and the bump positions become stabilized, which might be a useful property for maintaining spatial working memory in the brain.

VI. CONCLUSION

In this paper, the stability of four phases and the stability of the replica-symmetry ansatz in an Ising spin neural network with disordered lateral inhibition are studied. Similarly to the SK model, the network has a stable spin-glass phase even in a regime where lateral inhibitory interactions are dominant compared to ferromagnetic interactions. We used a Fourier expansion method to analyze the stability of a spin-system which interaction depends on distance of two spins. This method is applicable to any system with distance-dependent interaction.

The disorder changes the energy landscape, and in a phase where replica-symmetry breaking coexists with an L phase, we found that the drift of bump state, which represents a

memory state of the brain, can be stabilized. We also note that analysis of associative memory models have their basis in the SK model [2]. Recently, associative memory with spatially localized structure with threshold-linear neurons [13] was studied with self-consistent signal-to-noise analysis (SCSNA) [14]. Since replica-symmetry breaking cannot be studied with the method of SCSNA, it would be intriguing task to study the spatially localized associative memory by using our method.

ACKNOWLEDGMENTS

This work is partially supported by Grant-in-Aid Nos. 14084212 and 16500093 from the Ministry of Education, Culture, Sports, Science, and Technology, the Japanese Government.

APPENDIX: CALCULATION OF TRACE IN EQ. (20)

$$\begin{aligned}
\mathrm{Tr}_S\left(N^{-1}\sum_\theta S_\theta^\alpha S_\theta^\beta\right)\exp(L_0) &= \mathrm{Tr}_S\left(N^{-1}\sum_\theta S_\theta^\alpha S_\theta^\beta\right)\int Dz\exp\left\{\sum_\theta\left[\beta J\sqrt{q}z+\beta J_1 m_1\cos(\theta-\phi)\left(\sum_\alpha S_\theta^\alpha\right)\right]\right\}\\
&= N^{-1}\mathrm{Tr}_S\left(S_{\theta_1}^\alpha S_{\theta_1}^\beta\int Dz\exp[\beta H(z,\theta_1)(S_{\theta_1}^\alpha+S_{\theta_1}^\beta)]\exp\left\{\sum_{\theta\neq\theta_1}\left[\beta H(z,\theta)\left(\sum_\alpha S_\theta^\alpha\right)\right]\right\}\right.\\
&\quad\times\exp\left[\beta H(z,\theta_1)\left(\sum_{\gamma\neq\alpha,\beta}S_{\theta_1}^\gamma\right)\right]+S_{\theta_2}^\alpha S_{\theta_2}^\beta\int Dz\exp[\beta H(z,\theta_2)(S_{\theta_2}^\alpha+S_{\theta_2}^\beta)]\\
&\quad\times\exp\left\{\sum_{\theta\neq\theta_2}\left[\beta H(z,\theta)\left(\sum_\alpha S_\theta^\alpha\right)\right]\right\}\exp\left[\beta H(z,\theta_2)\left(\sum_{\gamma\neq\alpha,\beta}S_{\theta_2}^\gamma\right)\right]+\cdots+S_{\theta_N}^\alpha S_{\theta_N}^\beta\int Dz\\
&\quad\times\exp[\beta H(z,\theta_N)(S_{\theta_N}^\alpha+S_{\theta_N}^\beta)]\\
&\quad\left.\times\exp\left\{\sum_{\theta\neq\theta_N}\left[\beta H(z,\theta)\left(\sum_\alpha S_\theta^\alpha\right)\right]\right\}\exp\left[\beta H(z,\theta_N)\left(\sum_{\gamma\neq\alpha,\beta}S_{\theta_N}^\gamma\right)\right]\right). \qquad \text{(A1)}
\end{aligned}$$

After taking the trace,

$$=N^{-1}\sum_{\theta_i}\int Dz\{2\sinh[\beta H(z,\theta_i)]\}^2\left(\prod_{\theta\neq\theta_i}2\cosh[\beta H(z,\theta)]\right)^n(2\cosh[\beta H(z,\theta_i)])^{n-2}.$$

In the limit of $n\to 0$ and $N\to\infty$, we use $\mathrm{Tr}_S e^{L_0}\to 1$ to get

$$\langle S_\theta^\alpha S_\theta^\beta\rangle=\int\frac{d\theta}{2\pi}\int Dz\tanh^2[\beta H(z,\theta)]. \qquad \text{(A2)}$$

Four-pair correlation is also calculated as

$$\langle S_\theta^\alpha S_\theta^\beta S_\theta^\gamma S_\theta^\delta\rangle=N^{-1}\mathrm{Tr}_S\left(\sum_\theta S_\theta^\alpha S_\theta^\beta S_\theta^\gamma S_\theta^\delta\right)\exp(L_0)\to\int\frac{d\theta}{2\pi}\int Dz\tanh^4[\beta H(z,\theta)]. \qquad \text{(A3)}$$

[1] J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).
[2] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985).
[3] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985).
[4] D. H. Hubel and T. N. Wiesel, J. Physiol. (London) **160**, 106 (1962).
[5] C. Blakemore, R. Carpente, and M. Georgeso, Nature (London) **228**, 37 (1970).
[6] H. R. Wilson and J. D. Cowan, Biophys. J. **12**, 1 (1972).
[7] S. Amari, Biol. Cybern. **27**, 77 (1977).
[8] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, Proc. Natl. Acad. Sci. U.S.A. **92**, 3844 (1995).
[9] C. R. Laing and C. C. Chow, Neural Comput. **13**, 1473 (2001).
[10] O. Shriki, D. Hansel, and H. Sompolinsky, Neural Comput. **15**, 1809 (2003).
[11] J. de Almeida and D. Thouless, J. Phys. A **11**, 983 (1978).
[12] W. Kinzel, Phys. Rev. B **33**, 5086 (1986).
[13] Y. Roudi and A. Treves, J. Stat. Mech.: Theory Exp., 2004 P07010.
[14] M. Shiino and T. Fukai, Phys. Rev. E **48**, 867 (1993).