Properties of associative memory analog neural networks with asymmetric synaptic couplings

Masahiko Yoshioka* and Masatoshi Shiino†
Department of Applied Physics, Tokyo Institute of Technology, Ohokayama, Meguro-ku, Tokyo 152, Japan
(Received 1 July 1996; revised manuscript received 23 December 1996)

Using the self-consistent signal-to-noise analysis, we study fully connected analog neural networks having asymmetric synaptic couplings and positive-valued nonmonotonic transfer functions. Asymmetric synaptic couplings we assume are based on random patterns with bias a and given by $J_{ij}=(1/N)\Sigma_\mu\{(\xi_i^\mu-a)(\xi_j^\mu-a)+(b/\sqrt{N})(\xi_i^\mu-a)+c(\xi_j^\mu-a)+(d/N)\}$, where N is the number of neurons. We find that the synaptic interaction term of memorizing the postsynaptic activity (b term) and that of the presynaptic activity (c term) respectively, give rise to renormalized noise and a renormalized pattern-dependent but neuron-independent component in the local field of neurons, with the latter making the network behavior sample dependent. An enhancement of the storage capacity and the super retrieval phase due to the use of nonmonotonic transfer functions are shown to occur as a result of renormalization of noise even in the presence of the asymmetric synaptic couplings. [S1063-651X(97)05306-3]

PACS number(s): 87.10.+e, 05.90.+m, 05.70.Fh, 07.05.Mh

I. INTRODUCTION

Since the important contribution in Hopfield's paper [1], great progress has been made in the field of neural network theory [2–7]. To date most studies of physical models of associative memory neural networks have been conducted by taking advantage of the existence of a Lyapunov function or energy function [8], which enables one to employ a powerful and systematic method of statistical mechanics [9–11]. Model neural networks with energy functions are constructed on the basis of the assumption of introducing symmetry in the scheme of synaptic couplings between neurons. Synaptic couplings with a certain learning rule are a major ingredient of attractor networks of associative memory for recalling stored patterns. Conveniently assumed in extensive research is the symmetric Hebb learning rule [12]. Use of it has been proved to be able to capture essential features of attractor networks of associative memory models at the cost of biological realism.

One may, however, hope to know what will happen to the behavior of networks with asymmetric couplings, which are ubiquitous in biological nervous systems. Previously Peretto [13] proposed the asymmetric couplings

$$J_{ij}=\frac{1}{N}\sum_{\mu}\{A\xi_i^\mu\xi_j^\mu+B\xi_i^\mu+C\xi_j^\mu+D\},\quad J_{ii}=0, \tag{1}$$

where N and $\{\xi_i^\mu\}$ respectively represent the number of total neurons and a set of P random patterns taking either -1 or $+1$, to investigate the storage capacity of an associative memory model that has biological relevance regarding synaptic couplings. Assuming the updating dynamics of the network to be given by the stochastic dynamics, he aimed to get insights into the effects on the behavior of memory recall of the synaptic interaction terms responsible for memorizing postsynaptic (B term) and presynaptic (C term) activities. Because of the inadequacy of the replica method in dealing with the stochastic neural networks without energy functions, Peretto resorted to a different approach without replica calculations to approximately solve the problem. He noticed the C term to be the most influential for limiting the size of the storage capacity: with increasing C, the storage capacity of the network with input-output function of tanh$[\beta u]$ decreases appreciably.

We note here that the B term and the C term may be viewed as a kind of asymmetric synaptic noise, each of which seems to contribute a noise component in the local fields of neurons, and specifically that the latter will give rise to a pattern-dependent but neuron-independent component, which Peretto did not address.

We would like to elaborate on the above issues by formulating the problem anew. Taking advantage of our systematic method of the self-consistent signal-to-noise analysis (SCSNA) [14,15], which is capable of dealing with analog networks [7,8,10,11,14–23] with a transfer function of an arbitrary shape including the case where no energy functions exist [14,15,19–22], we consider analog networks with the synaptic couplings given by Eq. (1). In view of biological relevance we choose a transfer function representing neuron's mean firing rate to be positive valued. When a transfer function is not an odd function of the membrane potential of a neuron, the scaling with N for the B term of Eq. (1) has to be taken differently from the case studied by Peretto, because the site average of neuron output does not vanish for lack of symmetry with respect to reversing the sign of the output. Taking proper scalings needed for the terms of Eq. (1), we investigate their contributions to the local fields of neurons.

Our aim in the present paper is, first, to evaluate, using SCSNA, the universal effects of the B term and the C term on the behavior of the storage capacity, and second, to examine the problem of whether the super retrieval state [15,20] is allowed to remain in existence in the presence of the above types of asymmetric couplings, when nonmonotonic transfer functions are chosen appropriately.

Analog networks with nonmonotonic transfer functions, which are obtained from sigmoidal functions by cutting off

*Electronic address: myosioka@apneuro.ap.titech.ac.jp
†Electronic address: mshiino@ap.titech.ac.jp

the output activity of a neuron, have recently attracted much attention in the light of improving network performances as an associative memory [15,19–24], although in such networks the usual type of Lyapunov function does not make sense to ensure the stability of memory retrieval. They exhibit, in general, an enhancement of the storage capacity compared to the Amit, Gutfreund, and Sompolinsky (AGS) value of 0.14 [9] for the case of sigmoidal transfer functions, an extreme reduction of the number of spurious states, and the occurrence of the super retrieval states, when the symmetric Hebb rule is incorporated into the synaptic couplings.

The super retrieval state is a retrieval state in which a pure noise component in the cross-talk noise generated by the interference among an extensive number of stored patterns vanishes [15,20]. Its occurrence is a kind of phase transition phenomenon, which directly reflects the effect of noise renormalization for the local fields of analog neurons with nonmonotonic transfer functions having a steep negative slope. The outcome of the appearance of the super retrieval state is that memory retrieval without errors is possible even in the case of an extensive memory loading under the local learning rule of the Hebb type. We are interested particularly in how a noise component arising from the B term or the C term is renormalized to vanish, leading to the occurrence of the super retrieval state for certain nonmonotonic transfer functions.

The paper is organized as follows. In Sec. II, we formulate the model of analog networks that have the asymmetric synaptic connections of Eq. (1) with proper scalings with N and positive-valued transfer functions including the nonmonotonic case. In Sec. III, we develop the SCSNA to obtain a set of order parameter equations for retrieval states. The SCSNA presented here takes a different scheme from the original one in that the noise renormalization procedure proceeds without introducing overlaps with nonretrieved patterns each of which is of order $1/\sqrt{N}$, yielding the same result as that of the original version. We show that the coexistence of the B and C terms in the synaptic couplings makes the network behavior illegitimate because of the appearance of a diverging noise term in the large N limit. In Sec. IV we present the results of our analyses on the behaviors of the networks with asymmetric synaptic couplings based on either the B term or C term. Part of the present work concerning the effect of the C term was already reported elsewhere [25]. In addition to the effects of noise from the B term or C term, we also study the possibility of removing the reversed states as a kind of spurious one by employing the effect of C term whose existence makes the synaptic couplings asymmetric with respect to reversing the sign of the stored patterns. In Sec. V we give a brief summary and remarks.

II. MODEL OF NEURAL NETWORKS WITH ASYMMETRIC COUPLINGS

The neural networks we study are analog networks [10,11,14,15,17,18] which consist of N neurons. Denoting the membrane potential of neuron i by u_i, we assume that output signals of neuron i are described by its mean firing rate $F[u_i]$, which is called a transfer function. Synaptic couplings J_{ij} connect the output of neuron j to the input of neuron i. Then, with these definitions, the dynamics of u_i is described by the differential equations

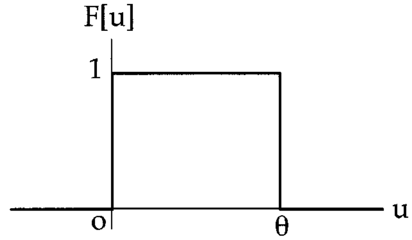

FIG. 1. The shape of the transfer function investigated in the present study.

$$\frac{d}{dt} u_i = -u_i + \sum_{j\neq i}^{N} J_{ij} F[u_j], \quad i=1,\ldots,N. \tag{2}$$

In biologically plausible neural networks, $F[u]$ should take a positive value because of what is meant by $F[u]$. In view of this, we assume the transfer function to be

$$F[u] = 1[u] - 1[u-\theta], \tag{3}$$

where $1[x]$ denotes the Heaviside function and θ a positive parameter representing cutoff of activity of a neuron. While setting $\theta=\infty$ gives a positive-valued sigmoid transfer function, the $F[u]$ with $\theta<\infty$ defines a nonmonotonic transfer function as shown in Fig. 1. The reason for considering nonmonotonic transfer functions is that, in the case of small θ, one can observe good network performance such as an enhancement of the storage capacity and occurrence of the super retrieval states [15,19–23].

The networks are assumed to memorize P random patterns, which are represented by ξ_i^μ $(i=1,\ldots,N,\ \mu=1,\ldots,P)$ taking either -1 or $+1$ according to the probability distribution

$$\Pr(\xi_i^\mu) = \frac{1-a}{2}\,\delta(\xi_i^\mu+1) + \frac{1+a}{2}\,\delta(\xi_i^\mu-1), \tag{4}$$

where bias a is the average of ξ_i^μ.

The synaptic couplings J_{ij} we assume are asymmetric ones that are obtained by modifying what were previously discussed by Peretto [13]. The asymmetric learning rule (1) proposed by Peretto has the property that, by adjusting A, B, C, and D, it can be equivalent to any learning rule of the form

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{P} \Delta J(\xi_i^\mu, \xi_j^\mu), \tag{5}$$

where $\Delta J(\xi_i^\mu, \xi_j^\mu)$ is a function of ξ_i^μ and ξ_j^μ, because ξ_i^μ and ξ_j^μ take only -1 or $+1$.

Setting $A=1$ in Eq. (1), we further scale B, C, and D appropriately with respect to N so that the network with the transfer function (3) behaves properly in the limit of $N\to\infty$. Noting that $(1/\sqrt{N})\Sigma_\mu \xi_i^\mu$ and $(1/N)\Sigma_i F[u_i]$ tend respectively to a random and constant quantity of order unity in the limit $N\to\infty$, we have

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{P}\left\{\xi_i^\mu\xi_j^\mu+\frac{b}{\sqrt{N}}\xi_i^\mu+c\,\xi_j^\mu+\frac{d}{N}\right\},\quad J_{ii}=0. \tag{6}$$

In the case of $a\neq 0$, introducing $\eta_i^\mu=\xi_i^\mu-a$, we have, instead of Eq. (6),

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{P}\left\{\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\eta_i^\mu+c\,\eta_j^\mu+\frac{d}{N}\right\},\quad J_{ii}=0, \tag{7}$$

which is the learning rule to be investigated in what follows.

We define several quantities to be used later. First we define the loading rate of memorized patterns as

$$\alpha=\frac{P}{N}. \tag{8}$$

The overlap with pattern μ is defined as

$$m^\mu=\frac{1}{N}\sum_{i=1}^{N}(\xi_i^\mu-a)F[u_i], \tag{9}$$

which takes value of order unity $[O(1)]$ when pattern μ is retrieved. This overlap is, however, inappropriate to measure the degree of quality of retrieval because of the shape of the transfer function as considered in the present study. Therefore we define the tolerance overlap [15]

$$g^\mu=\frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu\,\mathrm{sgn}[u_i], \tag{10}$$

which takes unity if pattern μ is completely retrieved.

Setting

$$x_i=F[u_i], \tag{11}$$

we define the local field of neuron i as

$$h_i=\sum_{j\neq i}J_{ij}x_j. \tag{12}$$

III. SELF-CONSISTENT SIGNAL-TO-NOISE ANALYSIS

We analyze the above-described networks in the limit $N\to\infty$. Assuming that the networks approach equilibrium for a long time to retrieve a desired pattern, we begin with solving the equations of equilibrium state

$$x_i=F[h_i],\quad i=1,\ldots,N \tag{13}$$

under the condition that

$$m^1=O(1),\quad m^\mu=O(1/\sqrt{N}),\quad \mu=2,\ldots,P, \tag{14}$$

where pattern 1 is selected as a retrieved one.

Substituting Eq. (7) into Eq. (12), we have

$$h_i=(\eta_i^1+c)m^1+\alpha\,dX+\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\eta_i^\mu+c\,\eta_j^\mu\right)x_j, \tag{15}$$

where

$$X=\frac{1}{N}\sum_i x_i. \tag{16}$$

The third term on the right-hand side (RHS) of this equation represents the cross-talk noise that varies from site to site. This noise does not obey a Gaussian distribution because of the correlations between $\eta_i^\mu\eta_j^\mu+(b/\sqrt{N})\eta_i^\mu+c\,\eta_j^\mu$, and x_j. The structure or characteristic of the noise term can be elucidated by the SCSNA [14,15], which conducts renormalization of noise.

The SCSNA has its basis on the assumption that the local fields (15) take the form [14,15]

$$h_i=(\eta_i^1+c)m^1+\alpha\,dX+\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\eta_i^\mu+c\,\eta_j^\mu\right)\tau x_j^\mu+\Gamma x_i. \tag{17}$$

In this equation, a super script μ of x_j^μ denotes that x_j^μ has almost negligible correlations with pattern μ, which can indeed be neglected in the limit $N\to\infty$, and differs only by $O(1/\sqrt{N})$ from x_j. Such a reduction of correlation is carried out by removing the contribution of pattern μ to the local field of neuron j as is done below. The crux of our analysis, SCSNA, is to determine τ and Γ self-consistently.

Using Eqs. (13) and (17), x_i is rewritten as

$$x_i=F\left[(\eta_i^1+c)m^1+\alpha\,dX+\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\eta_i^\mu+c\,\eta_j^\mu\right)\tau x_j^\mu+\Gamma x_i\right]. \tag{18}$$

We solve this equation for x_i to obtain the form

$$x_i=\tilde{F}\left[(\eta_i^1+c)m^1+\alpha\,dX+\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\eta_i^\mu+c\,\eta_j^\mu\right)\tau x_j^\mu\right]. \tag{19}$$

Substituting this into local field (15), we have

$$h_i=(\eta_i^1+c)m^1+\alpha\,dX+\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\eta_i^\mu+c\,\eta_j^\mu\right)\tilde{F}[\tilde{h}_j], \tag{20}$$

where

$$\tilde{h}_j=(\eta_j^1+c)m^1+\alpha\,dX+\frac{1}{N}\sum_{\nu\neq 1}\sum_{k\neq j}\left(\eta_j^\nu\eta_k^\nu+\frac{b}{\sqrt{N}}\eta_j^\nu+c\,\eta_k^\nu\right)\tau x_k^\nu. \tag{21}$$

Now we carry out the reduction of correlations. Performing a Taylor expansion of $\tilde{F}[\tilde{h}_j]$ of Eq. (20), we have

$$h_i=(\eta_i^1+c)m^1+\alpha\, dX+\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\,\eta_i^\mu+c\,\eta_j^\mu\right)x_j^\mu+\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\,\eta_i^\mu+c\,\eta_j^\mu\right)\times\tilde{F}'[\tilde{h}_j^\mu]\frac{1}{N}\sum_{k\neq j}\left(\eta_j^\mu\eta_k^\mu+\frac{b}{\sqrt{N}}\,\eta_j^\mu+c\,\eta_k^\mu\right)\tau x_k^\mu, \tag{22}$$

where

$$\tilde{h}_j^\mu=(\eta_j^1+c)m^1+\alpha\, dX+\frac{1}{N}\sum_{\nu\neq 1,\mu}\sum_{k\neq j}\left(\eta_j^\nu\eta_k^\nu+\frac{b}{\sqrt{N}}\,\eta_j^\nu+c\,\eta_k^\nu\right)\tau x_k^\nu \tag{23}$$

and

$$x_j^\mu=\tilde{F}[\tilde{h}_j^\mu]. \tag{24}$$

Let us take notice of the fourth term on the RHS of Eq. (22):

$$\frac{1}{N^2}\sum_{\mu\neq 1}\sum_{j\neq i}\sum_{k\neq j}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\,\eta_i^\mu+c\,\eta_j^\mu\right)\left(\eta_j^\mu\eta_k^\mu+\frac{b}{\sqrt{N}}\,\eta_j^\mu+c\,\eta_k^\mu\right)\tilde{F}'[\tilde{h}_j^\mu]\tau x_k^\mu$$
$$=\frac{1}{N^2}\sum_{\mu\neq 1}\sum_{j\neq i}\sum_{k\neq i,j}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\,\eta_i^\mu+c\,\eta_j^\mu\right)\left(\eta_j^\mu\eta_k^\mu+\frac{b}{\sqrt{N}}\,\eta_j^\mu+c\,\eta_k^\mu\right)\tilde{F}'[\tilde{h}_j^\mu]\tau x_k^\mu$$
$$+\frac{1}{N^2}\sum_{\mu\neq 1}\sum_{j\neq i}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\,\eta_i^\mu+c\,\eta_j^\mu\right)\left(\eta_j^\mu\eta_i^\mu+\frac{b}{\sqrt{N}}\,\eta_j^\mu+c\,\eta_i^\mu\right)\tilde{F}'[\tilde{h}_j^\mu]\tau x_i^\mu. \tag{25}$$

Since ξ_i^μ obeys the probability distribution (4), in the limit of $N\to\infty$, the RHS of the above equation is rewritten in the form

$$(1-a^2)U\frac{1}{N}\sum_{\mu\neq 1}\sum_{k\neq i}\left(\eta_i^\mu\eta_k^\mu+\frac{b}{\sqrt{N}}\,\eta_i^\mu+c\,\eta_k^\mu+\frac{bc}{\sqrt{N}}\right)\tau x_k^\mu+\alpha(1-a^2)^2U\tau x_i, \tag{26}$$

where

$$U=\frac{1}{N}\sum_j\tilde{F}'[\tilde{h}_j^\mu]. \tag{27}$$

Note that U does not depend on μ in the limit $N\to\infty$.

Now we focus our attention to the term

$$(1-a^2)U\frac{1}{N}\sum_{\mu\neq 1}\sum_{k\neq i}\frac{bc}{\sqrt{N}}\tau x_k^\mu \tag{28}$$

appearing in Eq. (26). This term is seen to diverge in the limit of $N\to\infty$ when one notes $x_k^\mu\approx x_k$. It means that our networks cannot retrieve any pattern when N is infinite, if $bc\neq 0$. Thus we must keep $bc=0$: $b=0$ or $c=0$ to make the network properly function, and we do so from now on.

Then, from Eqs. (17), (22), and (26), we obtain

$$\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\,\eta_i^\mu+c\,\eta_j^\mu\right)\tau x_j^\mu+\Gamma x_i$$
$$=\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\,\eta_i^\mu+c\,\eta_j^\mu\right)\times\{1+(1-a^2)U\tau\}x_j^\mu+\alpha(1-a^2)^2U\tau x_i. \tag{29}$$

Based on this equation, one can self-consistently determine τ and Γ. It follows that

$$\tau=\frac{1}{1-(1-a^2)U} \tag{30}$$

and

$$\Gamma=\alpha\frac{(1-a^2)^2U}{1-(1-a^2)U}. \tag{31}$$

Then, using Eqs. (17) and (30) we rewrite the local field as

$$h_i=(\eta_i^1+c)m+\alpha\, dX+\frac{1}{1-(1-a^2)U}\frac{1}{N}\times\sum_{\mu\neq 1}\sum_{j\neq i}\left(\eta_i^\mu\eta_j^\mu+\frac{b}{\sqrt{N}}\,\eta_i^\mu+c\,\eta_j^\mu\right)x_j^\mu+\Gamma x_i$$
$$=(\eta_i^1+c)m+\alpha\, dX+\overline{z_i}+cS+\Gamma x_i, \tag{32}$$

where

$$\overline{z_i}=\frac{1}{1-(1-a^2)U}\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\eta_i^\mu\left(\eta_j^\mu+\frac{b}{\sqrt{N}}\right)x_j^\mu \quad (33)$$

and

$$S=\frac{1}{1-(1-a^2)U}\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\eta_j^\mu x_j^\mu . \quad (34)$$

We note that when $b=0$, S equals the sum of $m^\mu(\mu \geqslant 2)$: $S=\Sigma_{\mu\neq 1}m^\mu$, since one has in general

$$m^\mu=\frac{1}{1-(1-a^2)U}\left\{\frac{1}{N}\sum_{j\neq i}\eta_j^\mu x_j^\mu + (1-a^2)U\frac{1}{N}\sum_{j\neq i}\frac{b}{\sqrt{N}}x_j^\mu\right\}. \quad (35)$$

S is not a site-dependent quantity, but a sample-dependent one. The appearance of such S in the local field (32) means that the network with $c\neq 0$ becomes sample dependent even in the limit $N\rightarrow\infty$. Then a problem arises of obtaining the probability distribution of S. As will be discussed later, calculating the distribution of S is too complicated to perform. However, once S is known, we can evaluate the properties of the network as a function of S. Therefore, for the time being, we treat S as a given parameter.

Pure noise $\overline{z_i}$ obeys a Gaussian distribution with mean 0 and variance σ^2, which we now evaluate. From Eq. (33) we have

$$\overline{z_i^2}=\frac{1}{\{1-(1-a^2)U\}^2}\frac{1}{N^2}\sum_{\mu 1\neq 1}\sum_{\mu 2\neq 1}\times\sum_{j1\neq i}\sum_{j2\neq i}\eta_i^{\mu 1}\eta_i^{\mu 2}\left(\eta_{j1}^{\mu 1}+\frac{b}{\sqrt{N}}\right)\left(\eta_{j2}^{\mu 2}+\frac{b}{\sqrt{N}}\right)x_{j1}^{\mu 1}x_{j2}^{\mu 2}. \quad (36)$$

Since we can replace the average over site by the average over patterns, we have

$$\sigma^2=\langle\langle\overline{z_i^2}\rangle\rangle=\frac{\alpha(1-a^2)}{\{1-(1-a^2)U\}^2}\{(1-a^2)q+b^2X^2\}, \quad (37)$$

where $\langle\langle ... \rangle\rangle$ denotes the average over the random patterns and

$$q=\frac{1}{N}\sum_i x_i^2 . \quad (38)$$

Finally we carry out the change of variables

$$\sigma^2=\alpha r \quad (39)$$

and

$$\frac{\overline{z}}{\sigma}=z. \quad (40)$$

Then, from Eqs. (32) and (37), we obtain the resultant SCSNA equation:

$$Y(\xi,z)=F[(\xi-a+c)m+\alpha\, dX+cS+\sqrt{\alpha r}z+\Gamma Y], \quad (41)$$

$$m=\left\langle\int_{-\infty}^{+\infty}Dz(\xi-a)Y\right\rangle, \quad (42)$$

$$U\sqrt{\alpha r}=\left\langle\int_{-\infty}^{+\infty}Dz\; zY\right\rangle, \quad (43)$$

$$X=\left\langle\int_{-\infty}^{+\infty}Dz\; Y\right\rangle, \quad (44)$$

$$q=\left\langle\int_{-\infty}^{+\infty}Dz\; Y^2\right\rangle, \quad (45)$$

$$\Gamma=\alpha\frac{(1-a^2)^2U}{1-(1-a^2)U} \quad (46)$$

and

$$r=\frac{(1-a^2)}{\{1-(1-a^2)U\}^2}\{(1-a^2)q+b^2X^2\}, \quad (47)$$

where

$$S=\sum_{\mu\neq 1}m^\mu \quad (b\neq 0). \quad (48)$$

In these equations

$$\int_{-\infty}^{+\infty}Dz\cdots=\int_{-\infty}^{+\infty}\cdots\frac{1}{\sqrt{2\pi}}\exp(-z^2/2)dz \quad (49)$$

and $\langle\cdots\rangle$ denotes the average over pattern 1, namely ξ, which obeys the probability distribution (4).

Equation (41), in general, admits multisolutions for $Y(\xi,z)$. In that case one has to select an appropriate solution of $Y(\xi,z)$ among them, using the Maxwell rule, to perform the Gaussian integrations over z. An example of the application of the Maxwell rule is demonstrated in Appendix A.

It should be noted that the solutions (m,r,q,X) obtained from the SCSNA equations (41)–(47) are not always of relevance, because the SCSNA is based on the fixed point equation (13) alone and does not take the stability of its solution into account. In the case of networks with asymmetric couplings or nonmonotonic transfer functions for which the existence of a Lyapunov function is not expected or known, the stability of SCSNA solutions has to be determined by numerical simulations of the dynamics (2).

IV. THE BEHAVIORS OF THE NETWORKS WITH ASYMMETRIC COUPLINGS

Based on the results of the SCSNA and numerical simulations, we explore properties of the networks with the asymmetric synaptic couplings (7). In the previous paper [25], we discussed the dependence of the network properties on bias

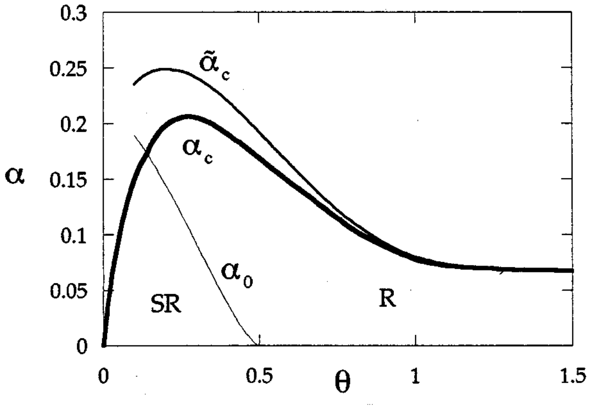

FIG. 2. Phase diagram showing the storage capacity as a function of θ in the case of $a=b=c=d=0$. $\tilde{\alpha}_c$ is the storage capacity obtained with the SCSNA, and α_c is that obtained with numerical simulations. The region below α_c is the retrieval phase (R), and that below α_0 is the super retrieval phase (SR).

a [26] in the presence of the c term. Setting $a=0$ in what follows, we are concerned mainly with the effects of the b term and the c term.

A. An enhancement of the storage capacity and the occurrence of the super retrieval in the case with small θ

We first present general properties of the analog networks with the nonmonotonic transfer function (3) by setting $b=c=d=0$. The behaviors of the associative networks exhibited by the nonmonotonic transfer function are summarized into the phase diagram on the $\theta-\alpha$ plane in Fig. 2. $\tilde{\alpha}_c$ is the critical value of the loading rate obtained by the SCSNA, and α_c is that obtained by numerical simulations with $N=1000$. While $\tilde{\alpha}_c=\alpha_c$ holds in the case of large θ, one has $\alpha_c<\tilde{\alpha}_c$ in the case of small θ. The difference between $\tilde{\alpha}_c$ and α_c arises from the fact that the networks having a nonmonotonic transfer function with $\theta<\infty$ are not expected, in general, to have Lyapunov functions ensuring stability of the memory retrieval and also that solutions of the SCSNA equations may correspond to unstable solutions of the original dynamics (2). Indeed one can observe the occurrence of oscillatory motions in the time evolution of the retrieval dynamics of analog networks having a certain type of nonmonotonic transfer functions [15]. We observe an enhancement of the storage capacity to occur: compared with the AGS value, the storage capacity α_c increases as θ decreases from $\theta=\infty$.

The enhancement of the storage capacity is caused by a reduction of the magnitude of the cross-talk noise that occurs as a result of $U<0$ in the case of nonmonotonic transfer functions [6] [see Eq. (37)]. Note that U measures the site average of the derivative of the output $Y(h)$, i.e., $\langle dY(h)/dh \rangle$.

An extreme reduction of the cross-talk noise gives rise to the occurrence of the super retrieval state as shown by the region below α_0 in Fig. 2. In the super retrieval phase one has variance $\sigma^2 \to +0$ as a result of $U\to-\infty$. In other words, noise in the local fields vanishes and memory retrieval without errors ensues for the local learning rule of the Hebb type with an extensive memory loading. One can verify this phenomenon by means of numerical simulations. Figures 3(a) and 3(b) show the distributions of the local fields obtained by numerical simulations, respectively, in the super retrieval and in the normal retrieval phase for the same value $\alpha=0.05$. One indeed observes a distribution with sharply peaked components in Fig. 3(a). Because of Γx_i with $\Gamma<0$ in the local field (17), the distribution for $h>0$ comprises two δ-function components:

$$P(h)=\tfrac{1}{2}\delta\left[h+\theta+\frac{\alpha}{2}\right]+\left(\theta+\frac{\alpha}{2}\right)\delta\left[h-\theta+\frac{\alpha}{2}\right]+\left(\tfrac{1}{2}-\theta-\frac{\alpha}{2}\right)\delta\left[h-\theta-\frac{\alpha}{2}\right], \quad (50)$$

which are a direct consequence of the application of the Maxwell rule.

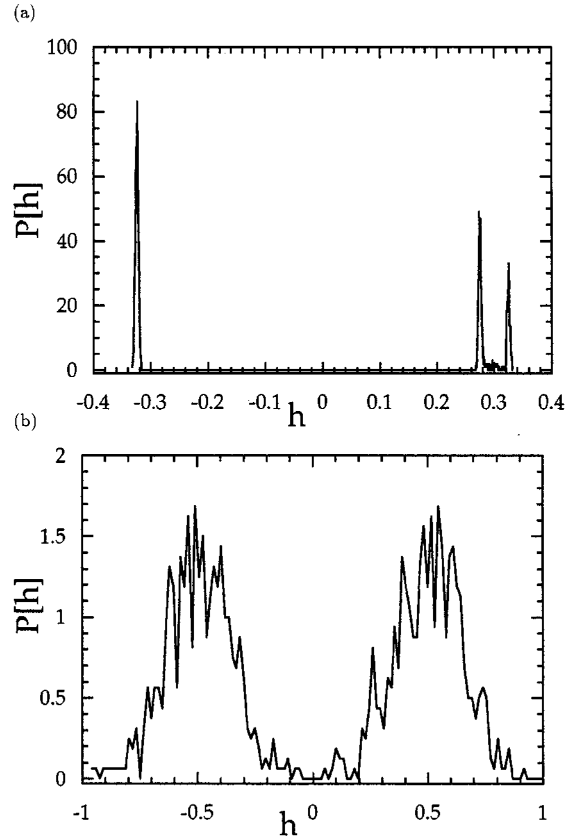

FIG. 3. Local-field distribution obtained by numerical simulations with $N=1000$ for $a=b=c=d=0$ under the condition $\alpha=0.05$. (a) $\theta=0.3$ (super retrieval) (b) $\theta=0.8$.

B. Effect of the b term

1. Increase in the variance of noise due to b

We now study the effect of the b term by setting $c=d=0$. Note that one cannot set $c\neq0$ because bc must be 0 as was mentioned in Sec. III.

In the SCSNA equation, b appears only in the equation for r Eq. (47), that is, in the variance $\sigma^2(=\alpha r)$ of noise in the local fields, which has a contribution proportional to b^2

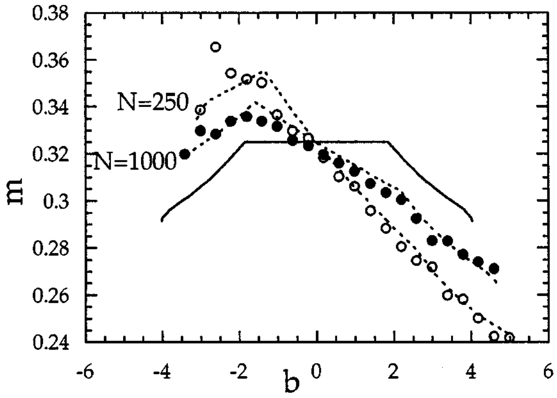

FIG. 4. Overlap m plotted against b under the condition $\alpha = 0.05$ and $\theta = 0.3$ for $a = c = d = 0$. The solid line represents the result of the SCSNA in the limit of $N \rightarrow \infty$. When N is finite, we obtain the approximations that are plotted with the broken lines. Open circles and closed circles represent the average of ten results of numerical simulations with $N = 250$ and $N = 1000$, respectively.

from the b dependent asymmetric term of J_{ij}. The b term turns out to play the role of noise in the process of memory retrieval. In view of the fact that $(b/\sqrt{N})\Sigma_{\mu}\eta_i^{\mu}$ obeys a Gaussian distribution with site i varied, the b term can be viewed as synaptic noise.

We show in Fig. 4 the overlap m obtained by the SCSNA (solid line) as a function of b under the condition $\theta = 0.3$ and $\alpha = 0.05$. The level part of the line corresponds to super retrieval states, which will be studied later. Noting that the solutions of the SCSNA equation are symmetric with respect to b, we observe that the results of numerical simulations with $N = 250$ and $N = 1000$, which are plotted with open circles and closed circles respectively, are not in good agreement with the results of SCSNA. It is because the number of neurons N we used is too small to approximate the result obtained in the thermodynamic limit $N \rightarrow \infty$. This can be seen more clearly, when we observe the finite-size effect by restoring $\xi(b/\sqrt{N})X$, which has been neglected in the local fields because of the limit $N \rightarrow \infty$: we replace Eq. (41) by

$$Y(\xi,z) = F\left[\xi m + \xi \frac{b}{\sqrt{N}} X + \alpha\, dX + \sqrt{\alpha r} z + \Gamma Y\right]. \quad (51)$$

Then we obtain the N-dependent results, which are displayed by the broken lines in Fig. 4. We see that the theoretical result for each N becomes in good agreement with the corresponding result of numerical simulations and also that the larger N becomes, the smaller is the difference between the values of m for $N = \infty$ and for finite N. This supports the theory that the SCSNA result represented by the solid line is approached in the limit $N \rightarrow \infty$.

Figures 5(a) and 5(b) are the phase diagrams on the $b - \alpha$ plane with $\theta = 0.3$ and $\theta = \infty$ respectively, where α_c is seen to slightly warp because of finite N. Taking the finite size effect into account, we can conclude $\alpha_c = \tilde{\alpha}_c$ in the case of $\theta = \infty$ when the limit of $N \rightarrow \infty$ is taken. However, in the case of $\theta = 0.3$, α_c is smaller than $\tilde{\alpha}_c$ because retrieval solutions of the SCSNA may get unstable in the case of small θ.

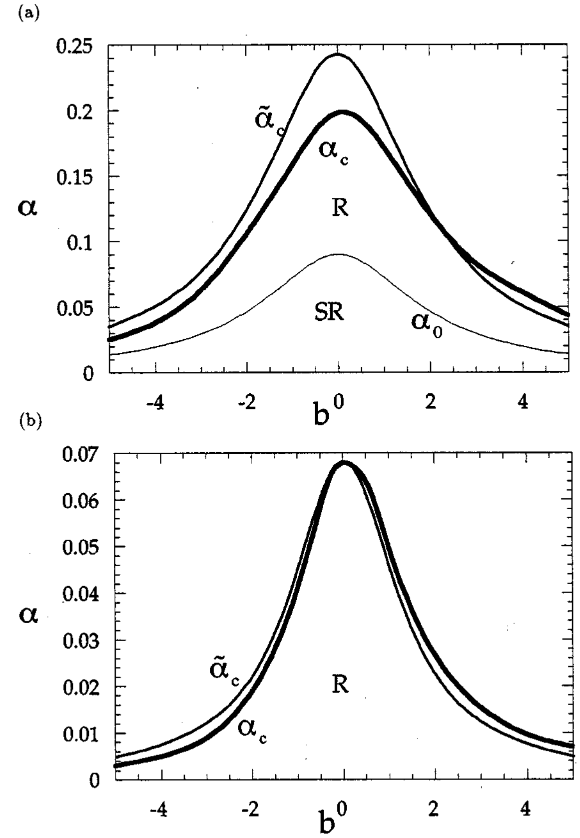

FIG. 5. Phase diagrams showing the storage capacity plotted against b with (a) $\theta = 0.3$ (b) $\theta = \infty$ for $a = c = d = 0$. α_c are obtained by numerical simulations with $N = 1000$.

Figures 5(a) and 5(b) show that α_c decreases with increasing $|b|$, as was expected from the nature of the b term. When $\alpha = 0$, this synaptic noise has no effect on the behavior of the network, as can be seen from Eq. (37).

2. *Super retrieval occurs even when* $b \neq 0$

Even in the presence of the b term yielding the synaptic noise of the form $(b/\sqrt{N})\Sigma_{\mu}\eta_i^{\mu}$, one can observe the occurrence of the super retrieval phase as has been shown in Fig. 4: in the super retrieval phase, where the variance σ^2 of noise vanishes in the local fields, overlap m does not depend on b, as is represented by the level line in Fig. 4. The occurrence of the super retrieval phase can be more directly confirmed by the local-field distribution. We show the local-field distributions obtained with numerical simulations in the case of a super retrieval and normal retrieval state in Figs. 6(a) and 6(b), respectively. One can observe quite sharply peaked components of the local-field distribution exhibited by the super retrieval state. Note that the local-field distribution is also given by Eq. (50). Since at first glance the b term simply generates the additional component $[(1/N)\Sigma_j x_j][(b/\sqrt{N})\Sigma_{\mu}\eta_i^{\mu}]$ proportional to the synaptic noise in the local fields, the occurrence of vanishing noise $\sigma^2 \rightarrow +0$ is somewhat surprising. Equation (37) implies that the pure noise arising from the b term is renormalized so that its variance is given by $[1/\{1-(1-a^2)U\}^2]\alpha b^2 X^2(1-a^2)$ instead of $\alpha X^2 b^2(1-a^2)$. This renormalization of noise is

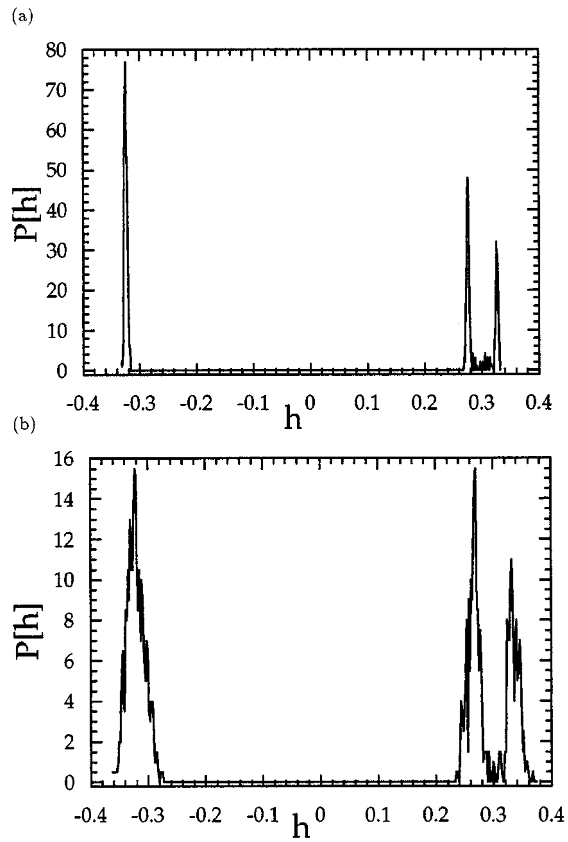

FIG. 6. Local-field distributions obtained by numerical simulations with $N=1000$ for $a=c=d=0$ under the condition $\theta=0.3$ and $\alpha=0.05$. (a) $b=1$ (super retrieval) (b) $b=3$.

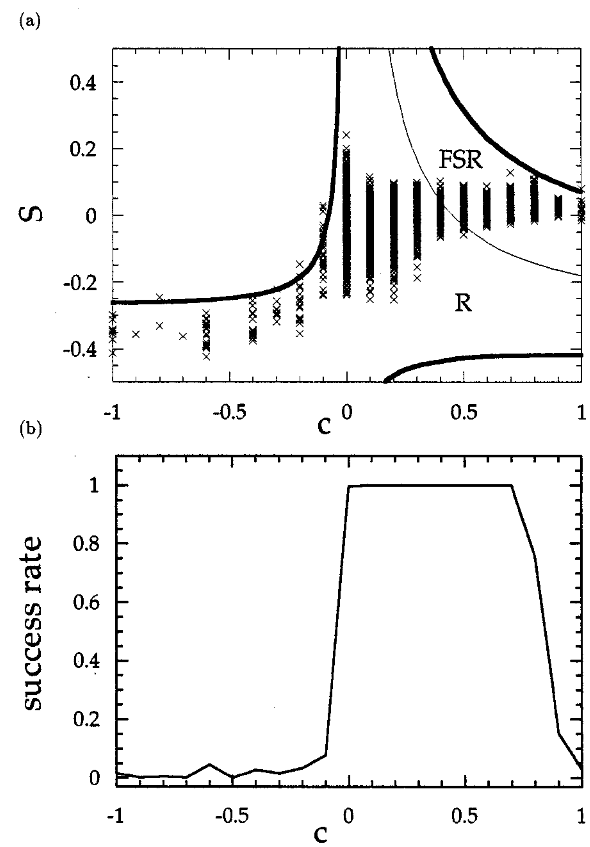

FIG. 7. (a) Phase diagram on the $c-S$ plane showing the region in which solutions of the SCSNA equation are allowed to exist under the condition $\alpha=0.1$, $\theta=0.3$, $a=b=0$, and $d=-4$; R represents the retrieval phase, and FSR the fictious super retrieval phase (see text). $S=0$ in the fictious super retrieval phase gives the super retrieval state. We have carried out 250 simulations with $N=1000$ for each c. Observing S with the aid of Eq. (48), we plot them with crosses when the network succeeded in retrieval. (Crosses sometimes look like a bar because of too many crosses.) (b) Plot of the success rate calculated from the results of numerical simulations in Fig. 7(a).

essential for the super retrieval phase to occur, since $U\to-\infty$ can make the so produced noise vanish.

It is worthwhile referring to the case where one simply adds random noise having no correlations with the stored patterns to synaptic couplings instead of the synaptic noise due to the b term. In such a case, the super retrieval phase cannot be allowed to occur, since uncorrelated random noise does not undergo the renormalization of noise [7].

The phase boundary of the super retrieval phase can easily be obtained by taking the limit $\sigma^2\to+0$ in Eqs. (41)–(47) with $c=0$ (Appendix A). A set of equations for α_0 and $z_3(1)$ reads

$$\frac{1+a}{2}\,I_0[-\infty,z_3(1)]=\frac{\theta+\dfrac{\alpha_0}{2}\,(1-a^2)}{(1-a)^2+\alpha_0 d}, \tag{52}$$

$$I_0[-\infty,z_3(1)]=\frac{1+a}{2\alpha_0}\,\{I_1[-\infty,z_3(1)]\}^2-\frac{b^2}{2(1-a)}\times\{I_0[-\infty,z_3(1)]\}^2 \tag{53}$$

together with

$$-(1-a)(1+a)+\alpha_0 d<0. \tag{54}$$

C. Sample-dependent behavior due to c

We turn to study the effect of c. We are forced to set $b=0$ as has been noted in Sec. III. In the case of $c\neq0$, we should focus our attention to the cS term in the local fields, which is a sample-dependent quantity and accordingly not definitely determined [25]. The role played by cS can be examined on the basis of the phase diagram that is obtained by assuming the cS to be a parameter in the SCSNA equations (41)–(47).

The region enclosed by the thick lines in the phase diagram of Fig. 7(a) is the expected retrieval phase on the c-S plane for which solutions with $m\neq0$ of the SCSNA equations exist in the case of $\alpha=0.1$, $\theta=0.3$, $a=0$, and $d=-4$. This expected retrieval phase includes the fictitious super retrieval phase that is expected to occur for small θ without considering the relation (34) between S and U. We carried out a number of numerical simulations for each value of c and observed S with the aid of Eq. (48). When the networks resulted in a successful retrieval, we plotted the resultant value of S against c in the c-S phase diagram of Fig.

7(a). We see that, in general, S distributes in the interval predicted by the SCSNA.

Then our interest is in knowing how S distributes, namely, the probability density of S. Although one may expect that S obeys a Gaussian distribution with mean 0, it is not the case. As has been shown in Fig. 7(a), S has some bounds and should obey a certain distribution that is not Gaussian. Obtaining the distribution of S exactly is quite a formidable task, since S appears in the local fields that determine the SCSNA equation and it also has to satisfy the relation

$$S\{1-(1-a^2)U[S]\}=\frac{1}{N}\sum_{\mu\neq 1}\sum_{j\neq i}\eta_j^\mu x_j^\mu. \tag{55}$$

Then we are forced to try to approximate the probability density. Although U and q are functions of S, we neglect their dependence on S by assuming that they only exhibit a weak dependence on S. Then, from Eq. (34), we obtain an approximation for the probability density:

$$\mathrm{Prob}(S)=\frac{1}{\sqrt{2\pi}\zeta}\exp\left(-\frac{S^2}{2\zeta^2}\right), \tag{56}$$

where

$$\zeta^2=\frac{\alpha(1-a^2)}{\{1-(1-a^2)U\}^2}\,q. \tag{57}$$

The distribution (56) can give a rough estimate for that obtained by numerical simulations in the case of large θ (although not shown here).

It is noted that one can specify an exact probability density of S in the super retrieval phase. Owing to the U dependence of S [Eq. (34)], the network with $c\neq 0$ can indeed have the genuine super retrieval states with $S=0$ in the fictitious super retrieval phase in Fig. 7(a): the required conditions $U\to-\infty$ for the occurrence of the super retrieval phase now yields $S\to 0$. Then, properties of the network do not exhibit sample dependence in the super retrieval phase. This explains the narrow distribution of S around $S=0$ obtained by numerical simulations in the super retrieval phase [Fig. 7(a)].

Figure 7(b) shows the success rate, which is the relative frequency of obtaining successful retrieval. We can expect that if an interval of S for which solutions of the SCSNA equations exist includes $S=0$, the success rate takes high value in view of the probability density (56). In the super retrieval phase, the success rate is one because the value of S is concentrated on $S=0$ with probability one. However, the success rate with $0.8\lesssim c$ is low because of instability of the solutions of the SCSNA equation. Note that, in the limit $N\to\infty$, the success rate in the super retrieval phase takes either 0 or 1 depending on the stability issue. On the other hand, in the case of interval $c\lesssim 0.5$ (which corresponds to the normal retrieval phase), as $|c|$ decreases, the success rate is expected to decrease continuously because of the sample dependence, even in the limit of $N\to\infty$.

D. Removal of reversed states

In this section we study the problem of removing a kind of spurious state, which appears as states corresponding to reversed patterns also in the usual Hopfield model, by employing the effect of the c term. First, we define a reversed state as an equilibrium one with $m<0$, which corresponds to the retrieval of reversed patterns. Whenever a network has an odd transfer function, it exhibits the property that there can coexist a retrieval state, say, with overlap m and a reversed one with overlap $-m$ at one time. The reason is that, if $x_i(t)$ is a solution of the dynamics (2) with an odd transfer function, $-x_i(t)$ is also its solution. Here we call the above property symmetry of reversed retrieval.

Furthermore, irrespective of the shape of a transfer function, networks can also have the symmetry of reversed retrieval, i.e., coexistence of a retrieval and reversed state with the same size of $|m|$. For example, a network with the well-known Hebb rule, for which

$$J_{ij}=\frac{1}{N}\sum_\mu \xi_i^\mu\xi_j^\mu=\frac{1}{N}\sum_\mu(-\xi_i^\mu)(-\xi_j^\mu) \tag{58}$$

holds with $\langle\xi\rangle=0$, exhibits the symmetry of reversed retrieval. This can easily be confirmed by an argument given below or direct inspection of Eqs. (41)–(47).

In the case of the synaptic couplings (7), we note the relation

$$J_{ij}(\{\xi\},a,c,d)=J_{ij}(\{-\xi\},-a,-c,d)\quad\text{for all}\quad i\neq j \tag{59}$$

to hold, if $b=0$. In this case, using the relation (59), we know that letting m' be an overlap for a retrieval state $\{x_i\}_{i=1,\ldots,N}$ of the network with $J_{ij}(\{\vec{\xi}'\},-a,-c,d)$, the $\{x_i\}_{i=1,\ldots,N}$ becomes a retrieval state of another network with $J_{ij}(\{\vec{\xi}\},a,c,d)$ $(\xi=-\xi')$ yielding the overlap $m\equiv 1/N\Sigma_i(\xi_i-a)x_i=-m'$. Then it turns out that the network with $J_{ij}(\{\vec{\xi}\},a,c,d)$ has a reversed state having negative-valued overlap $(-m')$. Note that one cannot expect any more that networks with $a\neq 0$ or $c\neq 0$ exhibit the property of symmetry of reversed retrieval under the use of the transfer function (3). Based on the above discussion, we consider the problem of whether one can substantially remove reversed states of the network with $c\neq 0$, which exhibits asymmetrical dependence of the strage capacity on c.

Now assuming $a=0$, we give in Fig. 8(a) the phase diagram together with the results of numerical simulations for reversed states of the network of Fig. 7. Figure 8(b) shows the success rate. Note that the phase diagram is given by the symmetric transformation of Fig. 7(a) with respect to the origin.

Let us give attention to the interval $0.1\lesssim c\lesssim 0.7$ in Figs. 7 and 8. In the interval, a high success rate for retrieval states is attained as seen in Fig. 7(b), while Fig. 8(b) shows that the success rate for the reversed state is fairly low. We see that reversed states are almost removed in this interval.

In Figs. 7 and 8, we incorporated the ferromagnetic interaction, i.e., the d term of the synaptic couplings with an appropriate magnitude d to enhance the degree of removing the reversed states. As can be seen from Eq. (32), assuming $m\approx X$ for $m>0$ we observe that d works in the local fields in

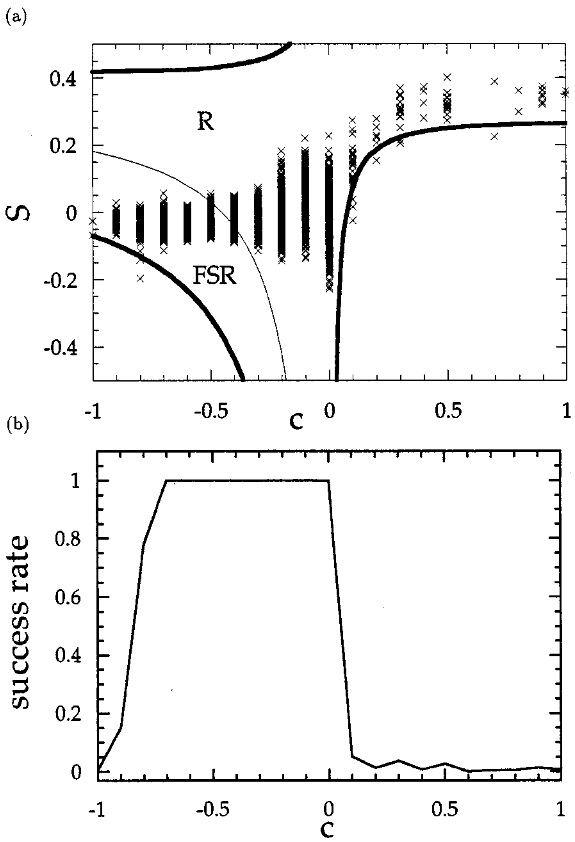

FIG. 8. (a) Same as Fig. 7(a) for reversed states. (b) Same as Fig. 7(b) for reversed states.

conjunction with c in the form $c+\alpha d$. Roughly speaking, d plays the role of shifting the phase diagram on the c-S plane for retrieval states along the c axis by an amount $-\alpha d$.

V. SUMMARY AND CONCLUDING REMARKS

We have investigated the behaviors of associative memory networks with the nonmonotonic transfer functions (3) and the asymmetric synaptic couplings (7). In the case of small θ, we have confirmed, even for asymmetric couplings with $b\neq 0$ or $c\neq 0$, the occurrence of enhancement of the storage capacity as well as of the phase transition phenomenon in which the super retrieval phase manifests itself as a result of the vanishing of noise $\sigma^2\rightarrow +0$ in the local fields.

We have elucidated the roles played by the asymmetric components with parameters b and c in the synaptic couplings. The asymmetric term with coefficient b (b term), which can be viewed as a kind of asymmetric synaptic noise, has been shown to give rise to part of the noise in the local field varying from site to site. We have observed that as $|b|$ is increased, noise variance σ^2 in general increases, letting the storage capacity decrease. In other words, introduction of the b term in the synaptic interactions, in most cases, degrades the network performance as an associative memory. We, however, would like to emphasize the effect of the synaptic noise having correlations with the stored patterns on the appearance of the super retrieval phase. Unlike the case of adding random noise having no correlation with the stored patterns to synaptic couplings, we have obtained the result that the presence of the b term still allows the super retrieval state to remain in existence. This is a consequence of the fact that the b term made of the stored patterns undergoes interference with the symmetric part of J_{ij} given by the Hebb rule and the resultant noise due to the b term gets renormalized with the reduction factor proportional to $1/\{1-(1-a^2)U\}^2$, which is crucial for the occurrence of the super retrieval [see Eq. (37)]. As long as the system is in the super retrieval state, the static properties of the network are independent of values of b [i.e., $m(b)=\text{const}$].

The asymmetric coupling term with c has been shown to have a characteristic feature of yielding a site-independent but pattern-dependent component cS in the local fields of neurons, which makes the behavior of the networks sample dependent. In the presence of such c term, the storage capacity cannot be determined within the context of the SCSNA. Instead, we have demonstrated that the SCSNA can give the bounds for the size of the cS to qualitatively explain the statistical nature of sample dependence of the network with $c\neq 0$. It is worth noting that even with this type of asymmetric coupling included, one can observe the occurrence of the super retrieval state, in which the pattern-dependent component cS vanishes.

We note that the noise-renormalization procedure involved in the SCSNA is essential to correctly evaluate the noise components in the local fields also in the presence of asymmetric couplings. In particular, it is the noise renormalization procedure that elucidates the fact that setting $bc\neq 0$ does not make any sense, because one cannot draw such a conclusion by simply looking at the expression for J_{ij}.

Finally, we discuss the stability problem of the network dynamics associated with the introduction of the synaptic couplings with b or c term in the case where a transfer function is of sigmoidal type including the case of $\theta=\infty$ in Eq. (3). As far as numerical simulations on the dynamics (2) are concerned, the network with $b\neq 0$ and $c=0$ is stable enough to exhibit relaxational motions to approach one of fixed-point-type attractors. Such mild behavior may be explained by the existence of a function that plays the role of the so-called Lyapunov function under the restricted condition that $\Sigma_{\mu\neq 1}m^{\mu}(t)=O(1)$ hold and the limit $N\rightarrow\infty$ be taken, as shown in Appendix B.

On the other hand, although numerical simulations on the network with $c\neq 0$ and the positive-valued transfer function ($\theta=\infty$) yield its convergence to fixed-point-type attractors irrespective of whether memory retrieval is successful or not, the network with $c\neq 0$ indeed happens to exhibit periodic or aperiodic oscillations [25,27] (not shown here) depending on initial conditions in the case of more general sigmoidal-type transfer functions, like $F(u)=\frac{1}{2}\{(1-\rho)\tanh(\beta u)+1+\rho\}$ with $\rho<0$. The occurrence of such temporal oscillations in the case where memory retrieval fails is, however, beyond the scope of this paper. Putting together the appearance of the sample dependence as well as the possibility of the oscillatory behavior, the network with asymmetric couplings due to the learning based on the presynaptic activity seems to be able to show potentially rich variety of behavior of memory recall dynamics.

ACKNOWLEDGMENT

This work is supported by Grant-in-Aid for General Scientific Research (07832008) from the Ministry of Education, Science and Culture.

APPENDIX A: DERIVATION OF EQUATIONS DETERMINING THE BOUNDARY OF THE SUPER RETRIEVAL PHASE

Assuming the condition $-\theta<\Gamma<0$, which applies in the region of interest, we derive the equations determining the boundary of the super retrieval phase. First we solve Eq. (41) to obtain $Y(\xi,z)$ for the transfer function (3). Defining

$$M(\xi)=(\xi-a+c)m+\alpha dX+cS, \tag{A1}$$

we have

$$Y(\xi,z)=\begin{cases}0, & z<z_1(\xi)\\ -\dfrac{M(\xi)+\sqrt{\alpha r}z}{\Gamma}, & z_1(\xi)<z<z_2(\xi)\\ 1, & z_2(\xi)<z<z_3(\xi)\\ 0, & z_3(\xi)<z\end{cases}, \tag{A2}$$

where

$$z_1(\xi)=\frac{-M(\xi)}{\sqrt{\alpha r}}, \tag{A3}$$

$$z_2(\xi)=\frac{-\Gamma-M(\xi)}{\sqrt{\alpha r}} \tag{A4}$$

and

$$z_3(\xi)=\frac{\theta-\dfrac{\Gamma}{2}-M(\xi)}{\sqrt{\alpha r}}. \tag{A5}$$

Using this, Eqs. (42)–(45) are written in the form

$$m=\left\langle(\xi-a)\left[-\frac{M(\xi)}{\Gamma}I_0[z_1(\xi),z_2(\xi)]-\frac{\sqrt{\alpha r}}{\Gamma}I_1[z_1(\xi),z_2(\xi)]+I_0[z_2(\xi),z_3(\xi)]\right]\right\rangle, \tag{A6}$$

$$U\sqrt{\alpha r}=\left\langle-\frac{M(\xi)}{\Gamma}I_1[z_1(\xi),z_2(\xi)]-\frac{\sqrt{\alpha r}}{\Gamma}I_2[z_1(\xi),z_2(\xi)]+I_1[z_2(\xi),z_3(\xi)]\right\rangle, \tag{A7}$$

$$X=\left\langle-\frac{M(\xi)}{\Gamma}I_0[z_1(\xi),z_2(\xi)]-\frac{\sqrt{\alpha r}}{\Gamma}I_1[z_1(\xi),z_2(\xi)]+I_0[z_2(\xi),z_3(\xi)]\right\rangle \tag{A8}$$

and

$$q=\left\langle\frac{M(\xi)^2}{\Gamma^2}I_0[z_1(\xi),z_2(\xi)]+\frac{2M(\xi)\sqrt{\alpha r}}{\Gamma^2}I_1[z_1(\xi),z_2(\xi)]+\frac{\alpha r}{\Gamma^2}I_2[z_1(\xi),z_2(\xi)]+I_0[z_2(\xi),z_3(\xi)]\right\rangle, \tag{A9}$$

where

$$I_0[z_1,z_2]=\int_{z_1}^{z_2}Dz\ 1, \tag{A10}$$

$$I_1[z_1,z_2]=\int_{z_1}^{z_2}Dz\ z \tag{A11}$$

and

$$I_2[z_1,z_2]=\int_{z_1}^{z_2}Dz\ z^2. \tag{A12}$$

When in the normal retrieval phase one approaches the boundary of the super retrieval phase, we have

$$r\to+0, \tag{A13}$$

$$U\to-\infty, \tag{A14}$$

$$z_1(-1),z_2(-1),z_3(-1)\to+\infty, \tag{A15}$$

$$z_1(1),z_2(1)\to-\infty \tag{A16}$$

and

$$z_3(1)\to\text{finite} \tag{A17}$$

Then, from Eqs. (A5) and (A17), one has

$$M(1)\to\theta-\frac{\Gamma}{2}. \tag{A18}$$

Substituting Eqs. (A13)–(A17) into Eqs. (46) and (A6)–(A9), we obtain the equations that hold on the boundary of the super retrieval phase:

$$m=\frac{1+a}{2}(1-a)I_0[-\infty,z_3(1)], \tag{A19}$$

$$U\sqrt{\alpha r}=\frac{1+a}{2}I_1[-\infty,z_3(1)], \tag{A20}$$

$$X=\frac{1+a}{2}I_0[-\infty,z_3(1)], \tag{A21}$$

$$q=\frac{1+a}{2}I_0[-\infty,z_3(1)] \tag{A22}$$

and

$$\Gamma=-\alpha(1-a^2). \tag{A23}$$

From Eqs. (A19) and (A21) we have

$$m=(1-a)X. \tag{A24}$$

Because of Eqs. (47), (A13), (A14), (A20), and (A21), we have

$$q=\frac{(1+a)^2}{4\alpha}\{I_1[-\infty,z_3(1)]\}^2 - b^2\frac{1+a}{4(1-a)}\{I_0[-\infty,z_3(1)]\}^2. \quad \text{(A25)}$$

From Eqs. (A1), (A18), (A23), and (A24) we have

$$m=\frac{(1-a)\left\{\theta-Sc+\frac{\alpha}{2}(1-a^2)\right\}}{(1-a)(1-a+c)+\alpha d}. \quad \text{(A26)}$$

Then, from Eqs. (A19), (A22), (A25), and (A26) we obtain

$$\frac{1+a}{2}I_0[-\infty,z_3(1)]=\frac{\theta-Sc+\frac{\alpha_0}{2}(1-a^2)}{(1-a)(1-a+c)+\alpha_0 d} \quad \text{(A27)}$$

and

$$I_0[-\infty,z_3(1)]=\frac{1+a}{2\alpha_0}\{I_1[-\infty,z_3(1)]\}^2 - \frac{b^2}{2(1-a)}\{I_0[-\infty,z_3(1)]\}^2, \quad \text{(A28)}$$

where α_0 denotes α on the boundary of the super retrieval phase.

On the other hand, because of Eqs. (A1), (A3), and $z_1(-1)\to+\infty$ [Eq. (A16)], we have

$$(-1-a+c)m+\alpha\, dX+cS<0. \quad \text{(A29)}$$

Because of Eq. (A18), this is written in the form

$$\theta-\frac{\Gamma}{2}-2m<0. \quad \text{(A30)}$$

Substituting Eqs. (A23) and (A26) into this, we have

$$\{-(1-a)(1+a-c)+\alpha_0 d\}\left\{\theta+\frac{\alpha_0}{2}(1-a^2)\right\} <-2(1-a)Sc. \quad \text{(A31)}$$

Then solving Eqs. (A27) and (A28) for α_0 and $z_3(1)$ under the condition Eq. (A31), we get the boundary of the super retrieval phase. Setting $c=0$ in Eq. (A27), (A28), and (A31) gives Eqs. (52)–(54).

APPENDIX B: EXISTENCE OF A FUNCTION PLAYING THE ROLE OF A LYAPUNOV FUNCTION FOR THE NETWORK WITH $b\neq 0$ AND $c=0$ UNDER A RESTRICTED CONDITION

Assuming $F[x]$ to be sigmoidal, we introduce the function

$$E(x_1,\ldots,x_N,X)=-\tfrac{1}{2}\sum_{i\neq j}J_{ij}^S x_i x_j+\sum_i\int^{x_i}F^{-1}[x_i]dx_i + \frac{b}{N\sqrt{N}}\sum_i\sum_\mu \xi_i^\mu\frac{x_i^2}{2} - \sum_i\frac{b}{\sqrt{N}}\left(\sum_\mu\xi_i^\mu\right)x_i X \quad \text{(B1)}$$

with $x_i=F[u_i]$, $X=(1/N)\Sigma_i x_i$, and J_{ij}^S denoting the symmetric part of J_{ij}. The dynamical equation (2) can be written as $\dot{u}_i=-\partial E/\partial x_i|_X$. We then obtain

$$\frac{dE}{dt}=\sum_i F'[u_i]\left\{-\left(\left.\frac{\partial E}{\partial x_i}\right|_X\right)^2 + \frac{1}{N}\left.\frac{\partial E}{\partial x_i}\right|_X\sum_j\frac{b}{\sqrt{N}}\left(\sum_\mu\xi_j^\mu\right)x_j\right\}. \quad \text{(B2)}$$

When the condition $\Sigma_{\mu\neq 1}m^\mu(t)=O(1)$ is satisfied as in the case of the usual retrieval condition (14), we have

$$\frac{dE}{dt}=-\sum_i F'(u_i)\left(\left.\frac{\partial E}{\partial x_i}\right|_X\right)^2\leqslant 0 \quad \text{(B3)}$$

in the limit $N\to\infty$.

It is noted, however, that setting an initial condition as $u_i=F^{-1}[(b/\sqrt{N})(\Sigma_\mu\xi_i^\mu)]$ breaks the above-mentioned condition.

[1] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).
[2] D. J. Amit, *Modelling Brain Function* (Cambridge University Press, Cambridge, England, 1989).
[3] *Models of Neural Networks*, edited by E. Domany, J. L. van Hemmen, and K. Schulten (Springer, New York, 1991).
[4] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, CA, 1991).
[5] J. Phys. A **22** (1989).
[6] M. Shiino, *Computational Physics as a New Frontier in Condensed Matter Research* (The Physical Society of Japan, Tokyo, 1995), p. 310.
[7] M. Okada, Neural Netw. **9**, 1429 (1996).
[8] J. J. Hopfield, Proc. Acad. Sci. USA **81**, 3088 (1984).

[9] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985); Ann. Phys. **173**, 30 (1987).
[10] M. Shiino and T. Fukai, J. Phys. A **23**, L1009 (1990).
[11] R. Kühn, S. Bös, and J. L. van Hemmen, Phys. Rev. A **43**, 2084 (1991); R. Kühn and S. Bös, J. Phys. A **26**, 831 (1993).
[12] D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).
[13] P. Peretto, J. Phys. (France) **49**, 711 (1988).
[14] M. Shiino and T. Fukai, J. Phys. A **25**, L375 (1992).
[15] M. Shiino and T. Fukai, Phys. Rev. E **48**, 867 (1993).
[16] M. A. Cohen and S. Grossberg, IEEE Trans. Syst., Man, Cybern. **13**, 815 (1983).
[17] T. Fukai and M. Shiino, Phys. Rev. A **42**, 7459 (1990).
[18] F. R. Waugh, C. M. Marcus, and R. M. Westervelt, Phys. Rev. Lett. **64**, 1986 (1990); Phys. Rev. A **43**, 3131 (1991).
[19] T. Fukai, J. Kim, and M. Shiino, Neural Netw. **8**, 391 (1995).
[20] M. Shiino and T. Fukai, J. Phys. A **26**, L831 (1993).
[21] M. Morita, Neural Netw. **6**, 115 (1993).
[22] S. Yoshizawa, M. Morita, and S. Amari, Neural Netw. **6**, 167 (1993).
[23] H. Yanai and S. Amari, IEEE Neural Netw. **7**, 803 (1996).
[24] G. Boffetta, R. Monasson, and R. Zecchina, J. Phys. A **26**, L507 (1993).
[25] M. Yoshioka and M. Shiino, J. Phys. Soc. Jpn. **66**, 5 (1997).
[26] D. J. Amit and M. V. Tsodyks, Network **2**, 275 (1991).
[27] H. Sompolinsky and A. Crisanti, Phys. Rev. Lett. **61**, 259 (1988).