Comparative study of spurious-state distribution in analogue neural networks and the Boltzmann machine

1992 J. Phys. A: Math. Gen. 25 2873

(http://iopscience.iop.org/0305-4470/25/10/015)

J. Phys. A: Math. Gen. 25 (1992) 2873–2887. Printed in the UK

Comparative study of spurious-state distribution in analogue neural networks and the Boltzmann machine

T Fukai† and M Shiino‡

† Department of Electronics, Tokai University, Kitakaname, Hiratsuka, Kanagawa, Japan
‡ Department of Applied Physics, Tokyo Institute of Technology, Ohokayama, Meguro-ku, Tokyo, Japan

Received 15 July 1991, in final form 23 December 1991

Abstract. We conducted a comparative study of the density distribution of metastable states in analogue neural networks and the Boltzmann machine by evaluating number densities of the attractors of the networks as functions of storage capacity, analogue gain or temperature and pattern overlap. The analysis is based on the fact that the Boltzmann machine and the analogue neural network can be described by the Thouless–Anderson–Palmer equations with and without the Onsager reaction term, respectively. We found the remarkable result that the spurious-state density around spin glass equilibrium states is much larger for the Boltzmann machine than for the analogue neural network for a reasonably wide range of analogue gain or temperature, which leads to an expectation that the analogue neural network should possess a much better potential for memory retrieval than the Boltzmann machine.

1. Introduction

The theory of content addressable memory based on binary neurons with symmetric synaptic couplings has been extensively developed within the context of statistical mechanics of spin glasses. Several problems exist, however, in the process of memory retrieval for neural networks consisting of discrete formal neurons including the existence of spurious states which is a severe obstacle to memory retrieval. Most of the methods which have been proposed to reduce the number of spurious states can be classified into two categories. The first category involves the use of analogue-valued neurons (Hopfield 1984, Hopfield and Tank 1985) in place of the discrete-valued formal ones and the second the introduction of stochastic updating to the network dynamics (Geman and Geman 1984, Ackey *et al* 1985). Using analogue neurons smooths the landscape of the energy function and decreases the density of the spurious states whereas introducing stochastic fluctuations allows the network states to escape from the local minima of the spurious states. It is, therefore, of interest to conduct a quantitative analysis to see whether these models are superior to those of discrete formal neurons with deterministic updating, which shows better performance and to what extent they share common properties. We have partially answered these questions by performing thermodynamic calculations to obtain the retrieval phase boundaries, i.e. the critical storage capacity (Shiino and Fukai 1990), as well as by estimating the number density of spurious states in the analogue neural networks (Fukai and Shiino 1990). The results we have obtained so far can be summarized as follows (Fukai and Shiino 1990, Shiino and Fukai 1990, Waugh *et al* 1990, 1991, Kühn *et al* 1991):

(i) The manner in which the storage capacity of the analogue neural network with a sigmoid-type response function decreases with decreasing analogue gain is qualitatively the same as that for stochastic Ising-spin neural networks with decreasing stochastic noise. (Amit *et al* 1985; hereafter referred to as the AGS theory.) In what follows, we call the stochastic Ising-spin neural network the Boltzmann machine (Ackey *et al* 1985).

(ii) The analogue neural network at a finite analogue gain exhibits a slightly larger critical storage capacity than that of the Boltzmann machine with the corresoponding temperature.

(iii) There exist, in general, three kinds of metastable state with exponentially large densities in the analogue neural networks: the first is a concentrated distribution around the embedded patterns; the second one occurs around the spin glass states; and the third one around unstable states, which does not manifest itself in the deterministic Ising-spin neural networks (Gardner 1986), i.e. an analogue neural network with an infinite analogue gain.

(iv) Decreasing the analogue gain dramatically suppresses the number of spurious states in comparison with the deterministic Ising-spin neural network. This implies that the use of an analogue neural network with an appropriately reduced analogue gain would considerably improve the network performance in return for a slight decrease in the storage capacity.

In the present paper, we would like to complete our comparative studies of the performances of the analogue neural network and the Boltzmann machine by comparing the densities of the metastable states of the two types of network model. Our analysis reveals that the metastable state density of the analogue network is much smaller than that of the Boltzmann machine over a reasonably wide range of parameter space.

We make full use of the concept of Thouless–Anderson–Palmer (TAP) equations by which both types of network can be related to each other. In the framework of the statistical mechanics of spin systems, the equilibrium states in the Boltzmann machine at a given temperature (stochastic noise) are characterized by a set of equations for the averaged spins (formal neurons) called the TAP equations with the so-called Onsager reaction term. The analogy between analogue networks with a sigmoid response and the Boltzmann machine is established by the fact that the equation obtained by discarding the Onsager term in the TAP equation coincides with the fixed-point condition for the analogue neural networks. In other words, analogue neural networks correspond to 'naive' mean-field models of Ising-spin neural networks. This enables us to evaluate the number of metastable states in the Boltzmann machine by counting the number of solutions to the TAP equation as in the case of the analogue neural networks where the number of solutions for the fixed-point condition was calculated analytically by use of the saddle-point method (Bray and Moore 1980, Gardner 1986, Fukai and Shiino, 1990, Waugh *et al* 1990, 1991).

The paper is organized as follows. In section 2, we present an analogue neural network with a sigmoid response function and a Boltzmann machine and then show that the two models can be discussed in a unified manner in terms of the corresponding TAP equations. In section 3, we analytically calculate the densities of the metastable states of both neural networks in the limit $N \to \infty$. The resultant expressions for the metastable state densities are numerically evaluated in section 4 in order to make a comparative study of the two neural networks. Section 5 is devoted to discussion and concluding remarks.

2. Neural network model

2.1. Nonlinear analogue neural network

Let $\{u_i\}$ $(-\infty < u_i < \infty, i = 1, \ldots, N)$ be a set of real variables which represent the membrane potentials of neurons having graded responses $z_i = f(u_i)$; the time evolution of the state u_i of analogue neuron i is described by

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = -u_i + \sum_j J_{ij} z_j \tag{1}$$

where J_{ij} is a synaptic coupling from neuron j to i. When, as will be assumed later, the synaptic coupling is symmetric and the response function f is monotonically increasing, the state of the network evolves so as to minimize the energy function (Hopfield 1984)

$$E_A = -\frac{1}{2}\sum_i \sum_j J_{ij} z_i z_j + \sum_i \int_{-1}^{z_i} \mathrm{d}z\, f^{-1}(z). \tag{2}$$

To establish a formal analogy between the analogue neural network and the Boltzmann machine of Ising spins, we choose

$$f(u) = \tanh(\beta u) \tag{3}$$

as the response function of analogue neurons with analogue gain β.

2.2. The Boltzmann machine

State $\{S_i\}$ of the neural network is usually assumed to be represented by a set of spin variables (formal neurons) taking either $+1$ or -1. The Boltzmann machine is defined by stochastic dynamics obeying the following master equation for the probability distribution $P(\{S\}, t)$ of the states of neurons:

$$\frac{\partial P(\{S\}, t)}{\partial t} = \sum_i w(-S_i \to S_i) P(\{S'\}, t) - \sum_i w(S_i \to -S_i) P(\{S\}, t) \tag{4}$$

where $\{S'\}$ is the state obtained from $\{S\}$ by reversing the state of neuron i and the transition rate at the inverse temperature β is given by

$$w(S_i \to -S_i) = \tfrac{1}{2}(1 - \tanh(\beta S_i h_i))$$

$$h_i(t) = \sum_{j \neq i} J_{ij} S_j(t). \tag{5}$$

Here h_i is the local field for neuron i and β is the measure of external noise. Since for $t \to \infty P(\{S\}, t)$ approaches the equilibrium distribution which is proportional to $e^{-\beta E_B}$ (the Boltzmann distribution) with the energy function

$$E_B = -\frac{1}{2} \sum_{i \neq j} J_{ij} S_i S_j$$

this model is called the Boltzmann machine.

In the present paper, we assume the synaptic coupling to be given by the conventional Hebb rule

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{(\mu)} \xi_j^{(\mu)} \qquad i \neq j, J_{ii} = 0 \tag{7}$$

where $p \equiv \alpha N$ random patterns $\{\xi_i^{(\mu)}\}$ $(\mu = 1, \ldots, p)$ are embedded with $\xi_i^{(\mu)}$ taking either $+1$ or -1.

2.3. TAP equations

These two seemingly different types of neural networks can be dealt with in a unified manner in terms of the TAP equations. We first note that the fixed-point condition $du_i/dt = 0$ for the analogue networks is nothing other than the TAP equation without the Onsager reaction term for the naive mean-field model of stochastic Ising-spin networks, on which the replica symmetric calculation to determine the storage capacity of the analogue networks was based (Shiino and Fukai 1990). The Onsager reaction term for the Boltzmann machine can be obtained by evaluating the free energy either by a diagramatical technique similar to that developed for the spin glass model (Thouless *et al* 1977) or by the cavity method. The resultant expression for the free energy is different from equation (2), as has been previously mentioned, by an amount arising from the Onsager reaction term. The following expression for E covers the energy (or the free energy) functions both for the analogue network with the sigmoid response function given by equation (3) ($\gamma = 0,\ E = E_A$) and for the Boltzmann machine ($\gamma = 1$) (Mezard *et al* 1987):

$$E = -\frac{1}{2} \sum_{i \neq j} J_{ij} z_i z_j + \frac{1}{2\beta} \sum_i [(1 + z_i) \ln(1 + z_i) + (1 - z_i) \ln(1 - z_i)]$$
$$+ \gamma \left(\frac{\alpha N}{2\beta} \ln \left[1 - \beta \left(1 - \sum_k z_k^2 / N \right) \right] - \frac{1}{2} \alpha \sum_i z_i^2 \right). \tag{8}$$

Here z_i stands for the averaged spin $\langle S_i \rangle$ for the Boltzmann machine. The last term in (8) yields the Onsager reaction term for the Boltzmann machine with the coupling given by equation (7). The metastable states of the networks are given by $\partial E / \partial z_i = 0$:

$$-\sum_{j \neq i} J_{ij} z_j + \frac{1}{\beta} \tanh^{-1} z_i + \gamma \left[\frac{\alpha z_i}{1 - \beta + (\beta/N) \sum_k z_k^2} - \alpha z_i \right] = 0$$
$$i = 1, \ldots, N \tag{9}$$

which is equivalent to the fixed-point condition $du_i/dt = 0$ for the analogue neural network ($\gamma = 0$) and is the TAP equation for the Boltzmann machine ($\gamma = 1$) with the last term representing the so-called Onsager reaction term. In fact, based on this TAP equation, we could obtain the same equations for determining the storage capacity of the Boltzmann machine as in the AGS theory by applying the recently developed self-consistent signal-to-noise analysis (Shiino and Fukai 1991).

3. Densities of the metastable states

We proceed to evaluate the density of the metastable states $\{z_i\}$ which is given by the stationary-state condition $\partial E/\partial z_i = 0$ together with stability condition $\det(\partial^2 E/\partial z_i \partial z_j) > 0$. Such a density n_{ms} for the metastable states having an overlap m with an embedded pattern $\{\xi_i^{(r)}\}$ is given by

$$n_{\mathrm{ms}}(m,\alpha,\beta,\gamma) \equiv \left\langle\!\!\left\langle \int_{-1}^{1} \prod_i \mathrm{d}z_i\, \delta\left(m - \frac{1}{N}\sum_i \xi_i^{(r)} z_i\right) \prod_i \delta\left(\frac{\partial E}{\partial z_i}\right) \left|\det \frac{\partial^2 E}{\partial z_i \partial z_j}\right| \right\rangle\!\!\right\rangle \tag{10}$$

where the bracket $\langle\!\langle \cdots \rangle\!\rangle$ denotes averaging over the random patterns. It is well known (Bray and Moore 1980, Gardner 1986) that n_{ms} takes the exponential form in the limit $N \to \infty$, i.e.

$$n_{\mathrm{ms}} \stackrel{N\to\infty}{\longrightarrow} \exp[NG(m,\alpha,\beta,\gamma)] \tag{11}$$

which implies that metastable states exist only for parameter values satisfying $G(m,\alpha,\beta,\gamma) > 0$. In the evaluation of equation (10), we will restrict the integration range for z_i to the region $\Omega \subset (-1,1)^N$ where $\det(\partial^2 E/\partial z_i \partial z_j) > 0$ to ensure the (necessary) stability condition for the local minima of the (free) energy function E.

The calculation of equation (10) for the Boltzmann machine (or for non-zero γ) parallels that for the analogue network ($\gamma = 0$). For the sake of clarity, we start from the expression in which averaging over the random embedded patterns is taken separately for the determinant factor and for the residual part of the right-hand side of equation (10), although we could proceed without assuming this separation as in the analogue network case and recover the same expression as derived later from equation (12) in the leading order of N.

$$n_{\mathrm{ms}}(m,\alpha,\beta,\gamma) = \left\langle\!\!\left\langle \int_{\Omega} \prod_i \mathrm{d}z_i\, \delta\left(m - \frac{1}{N}\sum_i \xi_i^{(r)} z_i\right) \prod_i \delta\left(\frac{\partial E}{\partial z_i}\right) \right\rangle\!\!\right\rangle \left\langle\!\!\left\langle \det \frac{\partial^2 E}{\partial z_i \partial z_j} \right\rangle\!\!\right\rangle. \tag{12}$$

It is noted that this factorization may not be justified in the next leading order of N since the distribution of the two factors in equation (12) might significantly overlap in this order.

Noting

$$\frac{\partial^2 E}{\partial z_i \partial z_j} = -(1-\delta_{ij})\frac{1}{N}\sum_{\mu\neq r}\xi_i^{(\mu)}\xi_j^{(\mu)} + \delta_{ij}\left[\frac{1}{\beta}\frac{1}{1-z_i^2} + \gamma\alpha\left(\frac{1}{1-\beta+\frac{\beta}{N}\sum_k z_k^2} - 1\right)\right] - \frac{2\alpha\beta\gamma}{N}\frac{z_i z_j}{(1-\beta+(\beta/N)\sum_k z_k^2)^2} \tag{13}$$

we first evaluate the determinant part in equation (12). The last term in equation (13), which is of $O(1/N)$ and hence is easily seen to yield a negligible contribution to the determinant in the limit when N goes to infinity, can be omitted in the calculation. Introducing two sets of anticommuting variables $\{\eta_i, \eta_i^*\}$ and $\{\theta_\mu, \theta_\mu^*\}$, we can rewrite the determinant factor as follows:

$$\begin{aligned}\det\left(\frac{\partial^2 E}{\partial z_i \partial z_j}\right) &= \int \prod_i \mathrm{d}\eta_i\,\mathrm{d}\eta_i^* \exp\left[\sum_{i,j}\eta_i^*\frac{\partial^2 E}{\partial z_i \partial z_j}\eta_j\right] \\ &= \int \prod_i \mathrm{d}\eta_i\,\mathrm{d}\eta_i^* \exp\left[\sum_i \kappa_i \eta_i^* \eta_i - \frac{1}{N}\sum_{\mu\neq r}\left(\sum_i \eta_i^*\xi_i^{(\mu)}\right)\left(\sum_j \eta_j \xi_j^{(\mu)}\right)\right] \\ &= \int \prod_i \mathrm{d}\eta_i\,\mathrm{d}\eta_i^* \int \prod_{\mu\neq r}\frac{\mathrm{d}\theta_\mu\,\mathrm{d}\theta_\mu^*}{N}\exp\left[N\sum_{\mu\neq r}\theta_\mu^*\theta_\mu + \sum_i \kappa_i\eta_i^*\eta_i \right. \\ &\quad \left. + \sum_{\mu\neq r, i}(\theta_\mu^*\eta_i + \eta_i^*\theta_\mu)\xi_i^{(\mu)}\right] \end{aligned} \tag{14}$$

$$\kappa_i \equiv \frac{1}{\beta}\frac{1}{1-z_i^2} + \frac{\gamma\alpha}{1-\beta+(\beta/N)\sum_k z_k^2} + (1-\gamma)\alpha. \tag{15}$$

After averaging this expression over $\{\xi_i^\mu\}$ ($\mu \neq r$) and a short manipulation, we obtain

$$\begin{aligned}&\int \prod_i \mathrm{d}\eta_i\,\mathrm{d}\eta_i^* \int \prod_{\mu\neq r}\frac{\mathrm{d}\theta_\mu\,\mathrm{d}\theta_\mu^*}{N}\exp\left[N\sum_{\mu\neq r}\theta_\mu^*\theta_\mu + \sum_i \kappa_i \eta_i^*\eta_i - \sum_{i,\mu\neq r}\eta_i^*\eta_i\theta_\mu^*\theta_\mu\right] \\ &\quad = \int \prod_i \mathrm{d}\eta_i\,\mathrm{d}\eta_i^* \exp\left[\sum_i \kappa_i\eta_i^*\eta_i\right]\left(1 - \frac{1}{N}\sum_i \eta_i^*\eta_i\right)^p. \end{aligned}$$

Introducing a parameter $n = (1/N)\sum_i \eta_i^*\eta_i$ with a constraint variable $\hat{n}$, this expression can be written as

$$\int_{-\infty}^{\infty}\mathrm{d}n\int_{-\mathrm{i}\infty}^{\mathrm{i}\infty}\frac{\mathrm{d}\hat{n}}{2\pi\mathrm{i}/N}\exp(Nn\hat{n} + p\ln(1-\hat{n}))\int\prod_i \mathrm{d}\eta_i\,\mathrm{d}\eta_i^* \exp\left[\sum_i L_i\eta_i^*\eta_i\right] \tag{17}$$

$$L_i = \frac{1}{\beta(1-z_i)^2} + \frac{\alpha\gamma}{1-\beta+(\beta/N)\sum_k z_k^2} + (1-\gamma)\alpha - \hat{n}. \tag{18}$$

In equation (17) the integrations over anticommuting variables can be easily performed to yield

$$\int \prod_i \mathrm{d}\eta_i \, \mathrm{d}\eta_i^* \exp\left[\sum_i L_i \eta_i^* \eta_i\right] = \prod_i L_i. \tag{19}$$

We carry out the n integration by the saddle-point method and obtain the final expression for the determinant:

$$\left\langle\!\!\left\langle \det \frac{\partial^2 E}{\partial z_i \partial z_j} \right\rangle\!\!\right\rangle = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}\hat{n}}{2\pi\mathrm{i}/N} \exp\left[N\left(\hat{n} - \alpha + \alpha \ln \frac{\alpha}{\hat{n}} + \frac{1}{N}\sum_i \ln L_i\right)\right]. \tag{20}$$

We note that in view of counting the stable fixed points which are ensured by the positive definiteness of matrix $\partial^2 E/\partial z_i \partial z_j$, the condition $L_i > 0$ $(\forall i)$ should be imposed to obtain the restricted integration range Ω.

The remaining part of equation (12) can be evaluated in a similar manner to that adopted in the analogue neural network case. Using the integral representation of the delta functions, we rewrite that part as

$$\begin{aligned}
&\int_\Omega \prod_i \mathrm{d}z_i \, \delta\left(m - \frac{1}{N}\sum_i \xi_i^{(r)} z_i\right) \prod_i \delta\left(\frac{\partial E}{\partial z_i}\right) \\
&\quad = \int_\Omega \prod_i \mathrm{d}z_i \int_{-\infty}^{\infty} \prod_i \frac{\mathrm{d}x_i}{2\pi} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}v}{2\pi\mathrm{i}/N} \exp\left[- Nv(m - \frac{1}{N}\sum_i \xi_i^{(r)} z_i)\right. \\
&\qquad + \mathrm{i}\sum_i x_i \left(- m\xi_i^{(r)} + \frac{1}{\beta_i} \tanh^{-1} z_i \right. \\
&\qquad \left. + \alpha\left(1 - \gamma + \frac{\gamma}{1 - \beta + \frac{\beta}{N}\sum_j z_j^2}\right) z_i \right) \\
&\qquad \left. - \frac{\mathrm{i}}{N} \sum_{\mu\neq r} \sum_i x_i \xi_i^{(\mu)} \sum_j z_j \xi_j^{(\mu)} \right].
\end{aligned} \tag{21}$$

We decouple the products involving the random memory patterns in the exponent of the integrand by means of the Gaussian integration formula with variables q_μ and λ_μ and then take the avarage over the random patterns:

$$\begin{aligned}
&\left\langle\!\!\left\langle \exp\left[-\frac{\mathrm{i}}{N} \sum_{\mu\neq r} \sum_i x_i \xi_i^{(\mu)} \sum_j z_j \xi_j^{(\mu)}\right] \right\rangle\!\!\right\rangle_{\xi^{(\mu\neq r)}} \\
&\quad = \left\langle\!\!\left\langle \int_{-\infty}^{\infty} \prod_{\nu\neq r} \frac{\mathrm{d}q_\nu \, \mathrm{d}\lambda_\nu}{2\pi} \exp\left[\mathrm{i} \sum_{\mu\neq r} \lambda_\mu q_\mu - \right.\right.\right.
\end{aligned}$$

$$
-\mathrm{i} \sum_{\mu\neq r,i} \left(q_\mu x_i + \frac{1}{N}\lambda_\mu z_i \right) \xi_i^{(\mu)} \Bigg] \Bigg\rangle\!\!\Bigg\rangle_{\xi(\mu\neq r)}
$$

$$
= \int_{-\infty}^{\infty} \prod_{\nu\neq r} \frac{\mathrm{d}q_\nu\,\mathrm{d}\lambda_\nu}{2\pi} \exp\Bigg[\mathrm{i}\sum_{\mu\neq r} \lambda_\mu q_\mu
+ \sum_i \sum_{\mu\neq r} \ln\cos\left(x_i q_\mu + \frac{1}{N} z_i \lambda_\mu \right) \Bigg]. \tag{22}
$$

Noting that $x_i q_\mu + (1/N) z_i \lambda_\mu \sim \mathrm{O}(1/\sqrt{N})$, we expand this equation up to $\mathrm{O}(1/N)$ and then perform the resultant Gaussian integration over q_μ and λ_μ to obtain

$$
\left\langle\!\!\left\langle \int_\Omega \prod_i \mathrm{d}z_i\, \delta\left(m - \frac{1}{N}\sum_i \xi_i^{(r)} z_i \right) \prod_i \delta\left(\frac{\partial E}{\partial z_i} \right) \right\rangle\!\!\right\rangle
$$

$$
\begin{aligned}
&= \int_\Omega \prod_i \mathrm{d}z_i \int_{-\infty}^{\infty} \prod_i \frac{\mathrm{d}x_i}{2\pi} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}v}{2\pi\mathrm{i}/N} \exp\Bigg[-Nv\left(m - \frac{1}{N}\sum_i z_i \xi_i^{(r)} \right) \\
&\quad + \mathrm{i}\sum_i x_i \Bigg(-m\xi_i^{(r)} + \frac{\tanh^{-1} z_i}{\beta} \\
&\quad + \alpha\left(1 - \gamma + \frac{\gamma}{1 - \beta + (\beta/N)\sum_j z_j^2} \right) z_i \Bigg) \\
&\quad - \frac{1}{2}\alpha N \ln\left(\frac{1}{N}\sum_i x_i^2 \frac{1}{N}\sum_j z_j^2 + \left(1 + \mathrm{i}\sum_i \frac{x_i z_i}{N} \right)^2 \right) \Bigg].
\end{aligned} \tag{23}
$$

Introducing parameters

$$
A = \frac{1}{N}\sum_i z_i^2 \qquad B = \frac{1}{N}\sum_i x_i^2 \qquad C = \mathrm{i}\frac{1}{N}\sum_i x_i z_i \tag{24}
$$

with the corresponding constraint variables a, b and c, respectively, and carrying out the Gaussian integration of x_i, we get the following expression for the left-hand side of equation (23).

$$
\begin{aligned}
&\int_\Omega \prod_i \mathrm{d}z_i \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \prod_i \frac{\mathrm{d}a\,\mathrm{d}b\,\mathrm{d}c\,\mathrm{d}v}{(2\pi\mathrm{i}/N)^4} \int_{-\infty}^{\infty} \mathrm{d}A\,\mathrm{d}B\,\mathrm{d}C \exp\Bigg[-Nv\left(m - \frac{1}{N}\sum_i \xi_i^{(r)} z_i \right) \\
&\quad - a\sum_i z_i^2 - \frac{1}{4b}\sum_i \Bigg[\left(c + \alpha(\gamma - 1) - \frac{\alpha\gamma}{1 - \beta + \beta A} \right) z_i \\
&\quad - \frac{1}{\beta_i}\tanh^{-1} z_i + m\xi_i^{(r)} \Bigg]^2 \\
&\quad - \frac{N}{2}\alpha\ln(AB + (1 + C)^2) + NaA + NbB + NcC - N\ln(2\sqrt{\pi b}) \Bigg].
\end{aligned} \tag{25}
$$

The integrations of variables B and C can be evaluated by the steepest descent method since the saddle-point equations for these variables become algebraic:

$$\frac{1}{N}\frac{\partial G}{\partial B} = b - \frac{\alpha}{2}\frac{A}{AB+(1+C)^2} = 0 \tag{26}$$

$$\frac{1}{N}\frac{\partial G}{\partial C} = c - \alpha\frac{1+C}{AB+(1+C)^2} = 0. \tag{27}$$

Note, however, that the saddle-point equation for the Edwards–Anderson-like parameter A is not algebraic unlike the analogue network case due to the existence of the Onsager term.

Combining this with the determinant factor (20), the result becomes

$$\begin{aligned}\frac{1}{N}\ln n_{\mathrm{ms}}(m,\alpha,\beta,\gamma) &= G(m,\alpha,\beta,\gamma)\\ &= \ln F + \hat{n} - \alpha\ln\hat{n} + \frac{\alpha}{2}\ln\frac{\alpha}{A} + \frac{\alpha-1}{2}\ln(2b) + \left(a+\frac{c^2}{4b}\right)A - c - \frac{\alpha}{2}\end{aligned} \tag{28}$$

$$\begin{aligned}F(m,\alpha,\beta,\gamma) = \int_\omega \frac{\mathrm{d}z}{\sqrt{2\pi}}\exp\Bigg[-\frac{1}{4b}\Bigg(&\left(c+(\gamma-1)\alpha-\frac{\gamma\alpha}{1-\beta(1-A)}\right)z\\ &-\frac{1}{\beta}\tanh^{-1}z+m\Bigg)^2 - az^2 + \ln\kappa(z) + vz - vm\Bigg]\end{aligned} \tag{29}$$

$$\kappa(z) = \frac{1}{\beta(1-z^2)} + \frac{\gamma\alpha}{1-\beta+\beta A} + (1-\gamma)\alpha - \hat{n}$$

where $\omega \subset (-1,1)$ represents the restricted integration range implied by the positivity of the determinant factor $\kappa(z) > 0$. Note that this representation is free of a detailed specification of the referenced pattern $\xi_i^{(r)}$ which was eliminated by scaling the integration variable $\xi_i^{(r)}z_i \to z_i$. The six integration parameters $a, b, c, v, \hat{n}$ and A should be determined by the saddle-point equations $\partial G/\partial a = \partial G/\partial b = \cdots = \partial G/\partial A = 0$.

4. Numerical analysis

The saddle-point equations for the parameters $a, b, c, v, \hat{n}$ and A were numerically solved and comparative studies of the metastable state distributions for the Boltzmann machine ($\gamma = 1$) and analogue neural network ($\gamma = 0$) were conducted. We obtained results showing remarkable differences between the two models in certain ranges of the parameters α and β. In general there exist three kinds of metastable state, each of which is characterized by the values of the pattern overlap m yielding $G > 0$ (figure 1). The first group comprises metastable states distributed around spin glass type thermal equilibrium states with vanishing pattern overlaps ($m \simeq 0$) or with no correlation with the embedded patterns. The second group (the retrieval states)

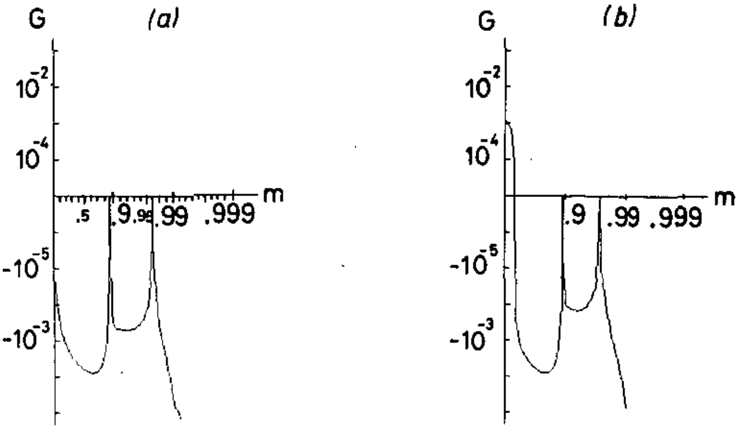

Figure 1. The G profile as a function of the pattern overlap m for the analogue neural network (a) and the Boltzmann machine (b). The parameters used are $\alpha = 0.09$ and $\beta = 3$. The heights of the sharp peaks which narrowly touch the horizontal axis in the figures are of $O(10^{-8})$. In the Boltzmann machine, we observe the existence of a large population of the metastable states around the spin glass states ($m = 0$) which are the obstacles for memory retrieval.

appears in the neighbourhood of the embedded memory patterns with m being almost unity. The third group, which corresponds to unstable states appearing in thermodynamic calculations (Shiino and Fukai 1990), is located between the other two groups of metastable states. These three peaks are separated from each other by intervals giving negative G.

Figure 1 shows a clear difference in the profiles of the metastable state distributions for the two models under consideration: the peak corresponding to the spin glass metastable states of the Boltzmann machine (figure 1(b)) is much larger than that of the analogue neural network (figure 1(a)). This implies that the Boltzmann machine possesses a much larger population of metastable states around the spin glass states (which is a severe obstacle to memory retrieval) than the analogue neural network does. This marked difference is commonly observed when the values of α and β are not very large and the models are meaningful practical applications for memory machines. It is expected from naive inspection that both models will coincide with the Ising-spin Hopfield neural network in the limit $\beta \to \infty$, the population of the spin glass metastable states of the analogue network also becomes larger and the difference between the two models becomes smaller as β increases (figure 2).

To see the changes in the population of the spurious metastable states of the two models in the α–β parameter space, we calculated the total densities of the metastable states $n_{\mathrm{t}} = \int \mathrm{d}m\, n_{\mathrm{ms}}(m, \alpha, \beta) = \exp[NG_{\mathrm{t}}(\alpha, \beta)]$ and plotted $\log_{10} G_{\mathrm{t}}(\alpha, \beta)$ in figure 3. (Region $\beta < 1$ is of no interest since we know, from the thermodynamic calculation, that no retrieval occurs in such a region.) Since n_{t} is mainly comprised from the spin glass metastable states, figure 3 shows that there exist parameter regions where the density of such a metastable state suddenly drops for the analogue neural network while, for the Boltzmann machine, the change is gradual over the whole parameter space.

It is noted, however, that there exists a large-β region where the total density for the analogue network becomes larger than that for the Boltzmann machine. The

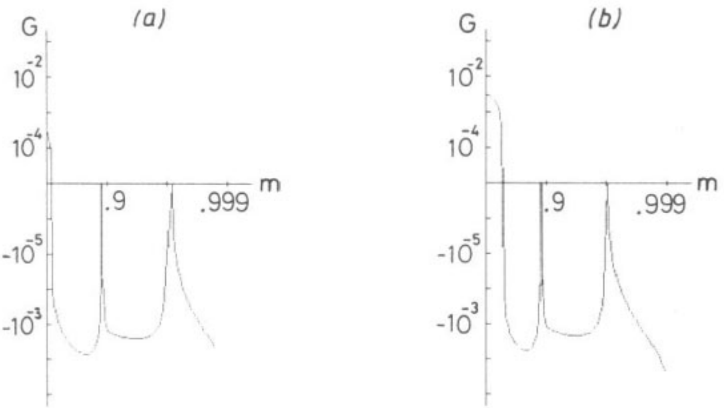

Figure 2. G profiles for the analogue neural network (a) and the Boltzmann machine (b) with the parameters $\alpha = 0.09$ and $\beta = 4.5$. The population of the spin glass metastable states in the analogue network is seen to be considerably increased for such a value of β.

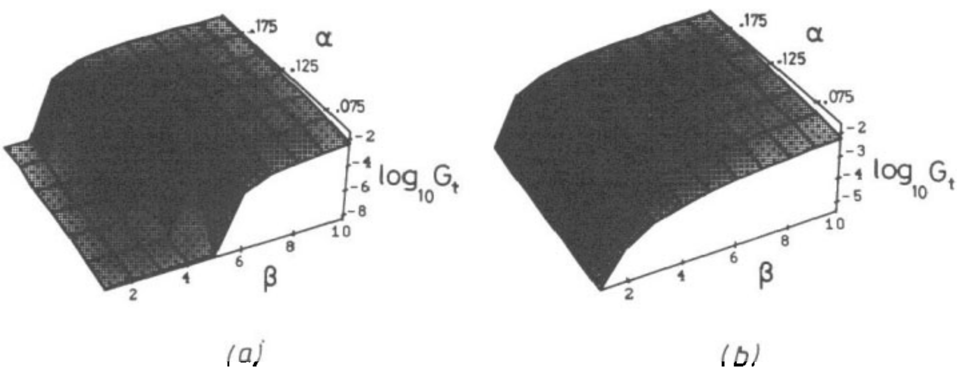

Figure 3. The total density of the metastable states $G_t(m, \alpha, \beta) = \ln(\int \mathrm{d}m\, n_{ms}(m, \alpha, \beta))/N$ for the analogue neural network (a) and the Boltzmann machine (b). We only show the results for $\beta > 1$ since otherwise there are no retrieval states. In the analogue neural network, the population of the metastable states is drastically reduced for small β.

boundary separating the region in the α–β^{-1} plane is displayed by the chain curve in figure 4. This can be clearly seen for the densities of the metastable states in the spin glass limit $\alpha \to \infty$ of both models. In fact we found that the $\alpha \to \infty$ limit of equation (28) yields the $\sigma = 0$ solution of Takayama and Nemoto (1990) for the spin glass models which, for $\gamma = 1$, is just the Bray and Moore's solution (see the appendix). It was shown by Takayama and Nemoto that the presence of the Onsager reaction term in the TAP equation reduces the number of metastable states (Waugh *et al* 1990).

We found that the G profile around $m \simeq 0$ is unchanged as γ is varied so long as γ is not very close to unity. Figure 5 shows the G profile for an imaginary neural network with $\gamma = 0.95$ with the same values for α and β as those used in figure 1(a). As mentioned earlier, the peak of the spin glass type metastable states around $m \simeq 0$ is identical to the corresponding one in figure 1(a) for the analogue neural network

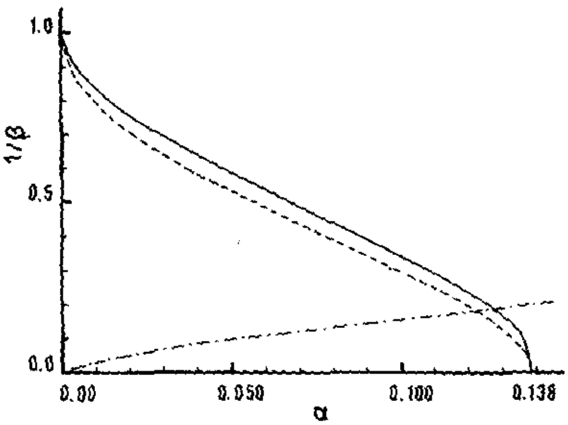

Figure 4. Storage capacities of the analogue network and the Boltzmann machine and $n_{t(\text{Boltzmann})} = n_{t(\text{analogue})}$ line. The full curve stands for the critical storage capacity of the analogue network (Shiino and Fukai 1990) while the broken curve for the Boltzmann machine (Amit *et al* 1985). Both curves are obtained by replica calculation. In the region above the chain curve, the density of the metastable states is smaller for the analogue network than for the Boltzmann machine. Thus the analogue network will exhibit better performances in most of the region relevant to practical uses of these neural network models.

although the other two peaks are, as is naively expected from the value of γ, almost identical to the corresponding ones in figure 1(b) for the Boltzmann machine.

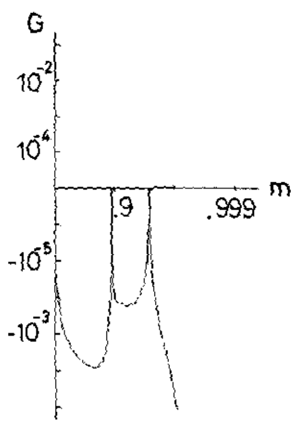

Figure 5. The G profile for an imaginary neural network defined by $\gamma = 0.95$. The values of α and β are the same as those used in figure 1 ($\alpha = 0.09$ and $\beta = 3$). The peak of the spin glass metastable states remains almost unchanged from that plotted for $\gamma = 0$ while the other two peaks are almost identical to those plotted for $\gamma = 1$.

In figure 6 we show G profiles of the analogue network for large values of β. The profiles for the Boltzmann machine with large β were found to be almost the same as those of the analogue network within numerical calculations. The populations of the three types of metastable state increase as the analogue gain is increased and the intermediate metastable state is absorbed by the broad peak of the spin glass

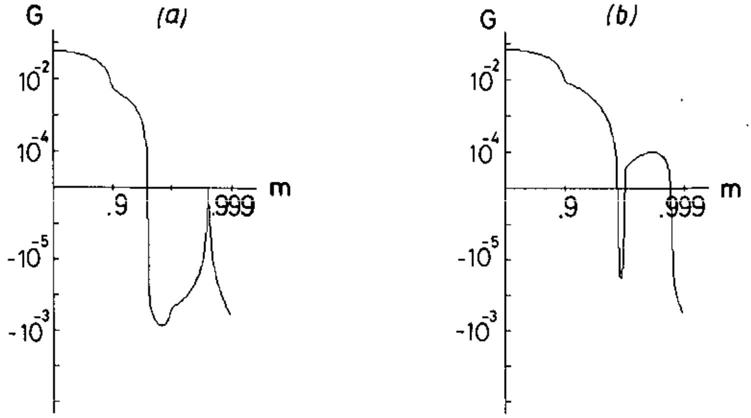

Figure 6. The G profiles for the analogue neural networks with (*a*) $\beta = 100$ and (*b*) $\beta = 10\,000$ when $\alpha = 0.113$ which was found by Gardner (1986) as the critical storage capacity of the deterministic Ising-spin neural network. We see that the intermediate metastable states are absorbed by the broad peak of the spin glass metastable states at sufficiently large values of β. The profiles for the Boltzmann machine with large β were found to be almost the same as those of the analogue network within numerical calculations.

metastable states at sufficiently large values of β. After absorption of the intermediate peak, the gap separating the remaining two peaks corresponding to the retrieval and the spin glass metastable states becomes narrower as the analogue gain is increased further until it finally disappears at $\beta \to \infty$.

The disappearance of the separation between the two kinds of metastable state with increasing α was adopted by Gardner (1986) as a criterion for estimating the critical storage capacity ($\alpha_c = 0.113$) for the deterministic Ising-spin neural network when calculating the number of metastable states. When, on the other hand, the value of β for the analogue neural network or the Boltzmann machine is not so large, the behaviour of the change in the G profile with increasing α differs from that for sufficiently large β cases: we observe that, as α increases, the two peaks of the retrieval and intermediate metastable states approach one another until they finally merge and become negative-valued. Accordingly the critical storage capacities of the networks should be given by the values of α at which the disappearance of the retrieval states takes place. In fact, for both networks the critical storage capacities so obtained in the metastable state distribution analysis were found to coincide with those determined in the thermodynamic analysis based on the replica symmetric theory, which are shown in figure 4 (Amit *et al* (1985) for the Boltzmann machine and Shiino and Fukai (1990) for the analogue network).

5. Discussion and concluding remarks

We have conducted comparative studies of the analogue neural network and the Boltzmann machine by estimating the densities of the metastable states. Our analysis revealed the remarkable fact that the population of the spurious states in the analogue neural network is significantly less than that in the Boltzmann machine for parameter ranges of practical importance with respect to storage level and analogue gain or temperature. The key role in discriminating the two types of neural network is played by the Onsager reaction term of the spin glass theory: for β not so large, the absence of the Onsager reaction term in the TAP equation makes the metastable state density around the spin glass states significantly small ($n_{t(\text{Boltzmann})} > n_{t(\text{analogue})}$) while for large β, the situation is reversed ($n_{t(\text{analogue})} > n_{t(\text{Boltzmann})}$). We have also made it clear that the spin glass limit $\alpha \to \infty$ of the metastable state densities for the Boltzmann machine and the analogue neural network yields solutions corresponding to Bray and Moore's (1980) and Takayama and Nemoto's (1990), respectively.

The difference found theoretically in the present paper for the spurious-state distributions of the two types of neural network should be seen in practical applications when the number of neurons exceeds several hundreds or thousands, depending on the values used for α and β. It is necessary, however, to clarify the effect of finite size on the spurious-state distributions as well as basins of attraction of the two types of neural network in order to make a close correspondence between theory and practical network performances.

Finally we comment on the $\beta \to \infty$ limit of the G profiles of the analogue network and Boltzmann machine. As shown in a previous paper (Fukai and Shiino 1990), it is easy to take the $\beta \to \infty$ limit of equations (28) and (29) for the analogue network to recover Gardner's result. By contrast, the Onsager reaction term of the Boltzmann machine makes rigorous analysis of equations (28) and (29) too complicated to see whether or not it is still effective in the $\beta \to \infty$ limit. However, we have found numerically that the Onsager reaction term hardly affects the spurious-state distribution of neural networks with sufficiently large β.

Acknowledgments

This work is partly supported by Grant-in-Aid for Encouragement of Young Scientists (No 02740195) and Grant-in-Aid for General Scientific Research (No 02460289). One of the authors (TF) is obliged to Kazuo Akashi and Asahi Corporation for their technical support on the numerical computation.

Appendix. Spin glass limit of the density distribution

Setting $\bar{\beta} = \sqrt{\alpha}\beta$ in equation (29), we derive the spin glass limit $\alpha \to \infty$ of equation (29) with $\bar{\beta}$ representing the inverse temperature of the corresponding spin glass system. We first note

$$\frac{\gamma\alpha}{1-\beta(1-A)} - \gamma\alpha \to \sqrt{\alpha}\bar{\beta}\gamma(1-A). \tag{A1}$$

Changing variables

$$\begin{aligned}
B &= \frac{\bar{\beta}}{\sqrt{\alpha}}(\alpha - \hat{n} + \sqrt{\alpha}\bar{\beta}\gamma(1-A)) \\
\Delta &= \frac{\bar{\beta}}{\sqrt{\alpha}}(c - \alpha - \sqrt{\alpha}\bar{\beta}\gamma(1-A)) \\
2b &= \alpha q \\
a &= -\lambda
\end{aligned} \tag{A2}$$

and setting

$$A = q + x/\sqrt{\alpha} \tag{A3}$$

we rewrite equation (29) to obtain in the limit $\alpha \to \infty$

$$G = \frac{1}{2}\left(\bar{\beta}\gamma(1-q) - \frac{B}{\bar{\beta}}\right)^2 + \frac{x^2}{4q^2} - q\lambda + \frac{1}{2}\left(\bar{\beta}\gamma(1-q) + \frac{\Delta}{\bar{\beta}}\right)^2 + \frac{x}{q}\left(\bar{\beta}\gamma(1-q) + \frac{\Delta}{\bar{\beta}}\right) + \ln I \tag{A4}$$

with

$$I = \int_{(1-z^2)^{-1}+B>0} \frac{\mathrm{d}z}{\bar{\beta}\sqrt{2\pi q}}\left(\frac{1}{1-z^2} + B\right)\exp\left[-\frac{(\tanh^{-1} z - \Delta z)^2}{2\bar{\beta}^2 q} + \lambda z^2\right]. \tag{A5}$$

Since the saddle-point equation $\partial G/\partial x = 0$ yields

$$x = -2q\left(\bar{\beta}\gamma(1-q) + \frac{\Delta}{\bar{\beta}}\right) \tag{A6}$$

the function G in the spin glass limit with the inverse temperature $\bar{\beta}$ is determined by

$$G = \frac{1}{2\bar{\beta}^2}(B^2 - \Delta^2) - q\lambda - \gamma(B + \Delta)(1 - q) + \ln I. \tag{A7}$$

We now see that G with $\gamma = 1$ is just what Bray and Moore (1980) obtained for the TAP mean-field equation from the Sherrington and Kirkpatrick (1975) spin glass model.

References

Ackey D H, Hinton G E and Sejnowski T J 1985 *Cog. Sci.* 9 147
Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* 55 1530
Bray A J and Moore M A 1980 *J. Phys. C: Solid State Phys.* 13 L469
Fukai T and Shiino M 1990 *Phys. Rev.* A 42 7459
Gardner E 1986 *J. Phys. A: Math. Gen.* 19 L1047
Geman S and Geman D 1984 *IEEE Trans.* PAMI-6 p 721
Hopfield J J 1984 *Proc. Natl Acad. Sci. USA* 81 3088
Hopfield J J and Tank D W 1985 *Biol. Cybern.* 52 141
Kühn R, Bös S and Hemmen J L 1991 *Phys. Rev.* A 43 2084
Mezard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* 32 1792
Shiino M and Fukai T 1990 *J. Phys. A: Math. Gen.* 23 L1009
—— 1992 *J. Phys. A: Math. Gen.* 25 L375–81
Takayama H and Nemoto K 1990 *J. Phys. C: Solid State Phys.* 2 1997
Thouless D J, Anderson P W and Palmer R G 1977 *Phil. Mag.* 35 593
Waugh F R, Marcus C M and Westervelt R M 1990 *Phys. Rev. Lett.* 64 1986
—— 1991 *Phys. Rev.* A 43 3131