Replica-symmetric theory of nonlinear analogue neural networks

1990 J. Phys. A: Math. Gen. 23 L1009

(http://iopscience.iop.org/0305-4470/23/18/012)

J. Phys. A: Math. Gen. 23 (1990) L1009–L1017. Printed in the UK

LETTER TO THE EDITOR

Replica-symmetric theory of the nonlinear analogue neural networks

M Shiino† and T Fukai†‡

† Department of Applied Physics, Tokyo Institute of Technology, Ohokayama, Meguro-ku, Tokyo, Japan
‡ Department of Management and Information Science, Gumma Women's College, Naka-Oorui-cho 501, Takasaki, Japan

Received 1 June 1990

Abstract. On the basis of the theory of the naive mean field model of spin glasses, analogue neural networks of the Hopfield type are investigated in the saturation limit. The saddle-point equations for the order parameters describing retrieval and spin glass phases of the networks are obtained by means of a statistical mechanical analysis within the framework of the replica-symmetric theory and are shown to undergo a modification to that of AGS theory, due to the absence of the Onsager reaction term in the TAP equation. Based on the equations, the memory storage capacity of the networks is analysed as a function of the analogue gain β. A small increase in the critical storage capacity is found for finite values of β, compared with that of the Ising model networks with corresponding inverse temperature, although the qualitative nature of the phase diagram is unchanged.

Since Hopfield's work (Hopfield 1982, 1984, Hopfield and Tank 1985) on the modelling of neural networks for associative memory based on nonlinear analogue neurons, physical models of neural networks have been extensively investigated (Amit 1989, Geszti 1990). In particular, the performance of neural networks of discrete-valued formal neurons with symmetric synaptic couplings has been extensively explored on the basis of statistical mechanical theory of spin glasses (Hemmen and Morgenstern 1987, Mezard *et al* 1987, Amit *et al* 1985a, b, 1987). Although Hopfield suggested that the performance of analogue networks is better than that of the Ising model networks, there has been little theoretical exploration on the properties of analogue networks to determine the extent to which the use of analogue networks has advantage over the Ising versions or how their memory capacity and the number of spurious states compare with their counterparts.

We have been concerned with studying such characteristic properties of nonlinear analogue neural networks from the viewpoint of the comparison of the network performance with discrete-valued neural networks such as the Ising model networks. In our previous paper (Fukai and Shiino 1990) which dealt mainly with the problem of counting the number of equilibrium (metastable) states of analogue neural networks of the Hopfield type, we reported that the number of the spurious states with no macroscopic correlation with the embedded patterns is remarkably suppressed, ensuring an excellent and promising performance as associative memory of the analogue neural networks. We also estimated the critical storage capacity and compared it with the result obtained from a statistical mechanical approach, which was only briefly noted in the previous paper. The aim of the present letter is then to elaborate the

0305-4470/90/181009+09$03.50

statistical mechanical analysis based on the replica-symmetric theory of obtaining the storage capacity of the analogue neural networks and to show that it is a little larger than the critical capacity of the stochastic Ising model neural networks of Amit, Gutfreund and Sompolinsky (hereafter referred to as AGS). The analysis is made with full use of the concept of the native mean-field model in the spin-glass theory.

The nonlinear analogue neural network model we consider is the set of equations describing the time change of membrane potentials u_i (Hopfield 1984)

$$C\frac{\mathrm{d}u_i}{\mathrm{d}t}=\sum_j J_{ij}V(u_j)-\frac{u_i}{R}+I_i \qquad i=1,\ldots,N \tag{1}$$

which represent the conservation law of the currents flowing through the membranes with I_i and $V(u)$ being a current from the external to the network and input–output relation of each neuron respectively. The capacitance C and resistance R can each be set to unity for simplicity without loss of generality. The synaptic couplings J_{ij} $(i\neq j)$ with $p(=\alpha N)$ sets of random embedded patterns are assumed to be

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}\xi_i^{(\mu)}\xi_j^{(\mu)} \tag{2}$$

with $\xi_i^{(\mu)}$ taking on ±1.

It is well known (Hopfield 1984) that the dynamics of (1) with symmetric J_{ij} is characterized by downhill motion on the surface of a Lyapunov function and ends up with one of the fixed-point attractors determined by

$$u_i=\sum_j J_{ij}(V(u_j)+I_i. \tag{3}$$

Equations of this kind with appropriate V will be associated with the TAP equation corresponding to the naive mean-field model in the spin-glass theory. The naive mean-field model of a spin system with interactions J_{ij} reads (Bray *et al* 1986)

$$m_i=\tanh\left[\beta\left(\sum_j J_{ij}m_j+h_i^{(\mathrm{ex})}\right)\right] \tag{4}$$

which is the so-called TAP equation (Thouless *et al* 1977) without the Onsager reaction term (Mexard *et al* 1987). Defining the effective local field $h_i^{(\mathrm{loc})}$ acting on site i as

$$h_i^{(\mathrm{loc})}=\sum_j J_{ij}m_j+h_i^{(\mathrm{ex})} \tag{5}$$

(4) is rewritten as

$$h_i^{(\mathrm{loc})}=\sum_j J_{ij}\tanh(\beta h_j^{(\mathrm{loc})})+h_i^{(\mathrm{ex})} \tag{6}$$

which will be equivalent to (3) if one sets $u_i=h_i^{(\mathrm{loc})}$, $I_i=h_i^{(\mathrm{ex})}$, $V(x)=\tanh(\beta x)$.

In writing (1) or (3), one may, however, require the $V(u)$ to represent a positive-valued monotonically increasing function of u so that the input–output relation $V(u)$ describes the dependence of the average rate of firing on the membrane potential u. In such a case, when one chooses for example $V(u)$ to be $\frac{1}{2}(1+\tanh(\beta u))$, (3) becomes

$$u_i=\sum_j\frac{1}{2}J_{ij}\tanh(\beta u_j)+\frac{1}{2}\left(\sum_j J_{ij}\right)+I_i. \tag{7}$$

We then see that the above equation virtually reduces to (6) under the condition that an external input current be $I_i=-\frac{1}{2}(\sum_j J_{ij})$. From the above discussion, it turns out that one will be allowed to deal with (4) instead of (3) for the purpose of exploring the network behaviour of the analogue neurons with symmetric connections. In what follows, assuming $V(x) = \tanh(\beta x)$ we will be concerned with (4) or (6). We then refer to the β as the analogue gain.

It was shown by Bray *et al* that the TAP equation for the naive mean-field model (4) has its basis on a specific spin system in which each site i consists of a set of M Ising spins S_{ia} ($a = 1, \ldots, M$), each of which interacts with each of the M spins at other sites. To be specific, writing the Hamiltonian (Bray *et al* 1986) for such a system as

$$\mathcal{H} = -\frac{1}{2M}\sum_{ij} J_{ij} \sum_{a,b=1}^{M} S_{ia}S_{jb} - \sum_i h_i^{(\mathrm{ex})} \sum_{a=1}^{M} S_{ia} \tag{8}$$

evaluation of the partition function $\mathrm{Tr}_{\{S_{ia}\}i=1,\ldots,N}\, \mathrm{e}^{-\beta H}$ with the use of new variables $m_i = (1/M)\sum_{a=1}^{M} S_{ia}$ was shown to yield, in the limit $M \to \infty$, the naive mean-field equation (4) with inverse temperature β. Consequently, a thermodynamical analysis of (4) should be equivalent to the statistical mechanical analysis of the Hamiltonian (8) for the specific spin system. Following the standard method developed by AGS for an analysis of the stochastic Ising model networks, we calculate the averaged free energy of the Hamiltonian (8) with J_{ij} given by (2) to obtain the ergodic components of the thermodynamical system, from which memory storage capacity of the analogue networks will be derived.

With the external fields $h_i^{(\mathrm{ex})}$ in (8) being $h_i^{(\mathrm{ex})} = \sum_{\nu=1}^{s} h^{(\nu)}$, the averaged free energy per spin of the system (8) can be calculated using the replica method (Amit *et al* 1987):

$$f = \lim_{M\to\infty} \lim_{n\to 0} \lim_{N\to\infty} \frac{-1}{\beta MnN} (\langle\langle Z^n \rangle\rangle - 1) \tag{9}$$

where the quenched average $\langle\langle Z^n\rangle\rangle$ over the ξ_s' is given by

$$\begin{aligned}\langle\langle Z^n\rangle\rangle &= \left\langle\!\!\left\langle \mathrm{Tr}_{S_\rho} \exp\left[\frac{\beta}{2MN}\sum_{\rho=1}^{n}\sum_{\mu=1}^{p}\left\{\sum_a\sum_i \xi_i^{(\mu)}S_{ia}^{(\rho)}\right\}^2 - \frac{\beta p}{2MN}\sum_{\rho=1}^{n}\sum_i\left(\sum_{a=1}^{M} S_{ia}^{(\rho)}\right)^2 \right.\right.\right. \\ &\quad \left.\left.\left. + \beta\sum_{\rho=1}^{n}\sum_{\nu=1}^{s} h^{(\nu)}\sum_i \xi_i^{(\nu)}\sum_{a=1}^{M} S_{ia}^{(\rho)}\right]\right\rangle\!\!\right\rangle \\ &= \left\langle\!\!\left\langle \mathrm{Tr}_{S_\rho}\int \prod_{\mu,\rho}\sqrt{\frac{MN\beta}{2\pi}}\,\mathrm{d}g_\rho^\mu \exp\left[-\frac{MN\beta}{2}\sum_{\rho,\mu}(g_\rho^\mu)^2 + \beta\sum_{\rho,\mu} g_\rho^\mu \sum_a\sum_i \xi_i^{(\mu)}S_{ia}^{(\rho)} \right.\right.\right. \\ &\quad \left.\left.\left. - \frac{\beta p}{2MN}\sum_\rho\sum_i\left(\sum_a S_{ia}^{(\rho)}\right)^2 + \beta\sum_\rho\sum_\nu h^{(\nu)}\sum_i \xi_i^{(\nu)}\sum_a S_{ia}^{(\rho)}\right]\right\rangle\!\!\right\rangle \end{aligned} \tag{10}$$

in which the $\{S_{ia}^{(\rho)}\}(\rho = 1, \ldots, n)$ are replica spin variables. We will perform the average over ξ's in two steps; first over the $p - s$ non-condensed patterns $\{\xi_i^{(\mu)}\}$ with $\mu > s$ and later over the s condensed patterns $\{\xi_i^{(\mu)}\}$ with $\mu < s$ (s will be kept fixed as p, $N \to \infty$).

The part associated with the averaging over the $p - s$ patterns in (10) results in

$$\begin{aligned}&\left\langle\!\!\left\langle \int \prod_{\rho,\mu>s}\sqrt{\frac{MN\beta}{2\pi}}\,\mathrm{d}g_\rho^\mu \exp\left[-\frac{Mn\beta}{2}\sum_{\rho,\mu>s}(g_\rho^\mu)^2 + \beta\sum_{\rho,\mu>s} g_\rho^\mu\sum_a\sum_i \xi_i^{(\mu)}S_{ia}^{(\rho)}\right]\right\rangle\!\!\right\rangle \\ &\quad = \int \prod_{\rho,\mu>s}\sqrt{\frac{MN\beta}{2\pi}}\,\mathrm{d}g_\rho^\mu \exp\left[-\frac{MN\beta}{2}\sum_{\rho,\mu>s}(g_\rho^\mu)^2 \right. \\ &\qquad \left. + \sum_{i,\mu>s}\ln\cosh\left(\beta\sum_{\rho,a} g_\rho^\mu S_{ia}^{(\rho)}\right)\right]. \end{aligned} \tag{11}$$

The leading terms in (11) as $N\to\infty$ are given by the Gaussian integral, which is obtained from the rescaling $G^\mu_\rho=\sqrt{MN\beta/2}g^\mu_\rho$:

$$\int\prod_{\rho,\mu>s}\sqrt{\frac{1}{\pi}}\,\mathrm{d}G^\mu_\rho\exp\left[-\sum_{\rho,\mu>s}(G^\mu_\rho)^2+\sum_{i,\mu>s}\frac{\beta}{MN}\sum_{\rho,\sigma,a,b}G^\mu_\rho G^\mu_\sigma S^{(\rho)}_{ia}S^{(\sigma)}_{ib}\right]$$
$$=[\det(-\beta M\bar{\bar{Q}}+\bar{\bar{I}}-\beta M\bar{\bar{D}})]^{-(p-s)/2}$$

where

$$(\bar{\bar{Q}})_{\rho\sigma}=q_{\rho\sigma}=\sum_{i,a,b}\frac{S^{(\rho)}_{ia}S^{(\sigma)}_{ib}}{M^2N}\qquad(\rho\neq\sigma)$$
$$(\bar{\bar{D}})_{\rho\sigma}=q_\sigma\delta_{\rho\sigma}=\delta_{\rho\sigma}\sum_{i,a,b}\frac{S^{(\rho)}_{ia}S^{(\rho)}_{ib}}{M^2N}.\tag{12}$$

An integral representation of the RHS of the above equation (12) will be

$$\int\prod_{(\rho,\sigma)}\mathrm{d}q_{\rho\sigma}\prod_\omega\mathrm{d}q_\omega\exp\left[-\frac{\rho-s}{2}\mathrm{Tr}\ln\{-\beta\bar{\bar{Q}}(q_{\rho\sigma})+\bar{\bar{I}}-\beta M\bar{\bar{D}}(q_\omega)\}\right]\int\prod_{(\rho\sigma)}\mathrm{d}r_{\rho\sigma}\prod_\omega\mathrm{d}r_\omega$$
$$\times\exp M^2N\left[-\tfrac{1}{2}\alpha\beta^2\sum_{\rho\neq\sigma}r_{\rho\sigma}q_{\rho\sigma}+\tfrac{1}{2}\alpha\beta^2\frac{1}{M^2N}\sum_{\rho\neq\sigma}\sum_{i,a,b}r_{\rho\sigma}S^{(\rho)}_{ia}S^{(\sigma)}_{ib}\right.$$
$$\left.-\tfrac{1}{2}\alpha\beta^2\sum_\omega r_\omega q_\omega+\tfrac{1}{2}\alpha\beta^2\frac{1}{M^2N}\sum_{\omega,i,a,b}r_\omega S^{(\omega)}_{ia}S^{(\omega)}_{ib}\right]\tag{13}$$

where irrelevant factors are omitted.

We now have

$$\langle\langle Z^n\rangle\rangle=\int\prod_{\rho,\mu\leqslant s}\sqrt{\frac{MN\beta}{2\pi}}\,\mathrm{d}g^\mu_\rho\int\prod_{(\rho\sigma)}\mathrm{d}q_{\rho\sigma}\prod_\omega\mathrm{d}q_\omega\prod_{(\rho\sigma)}\mathrm{d}r_{\rho\sigma}\prod_\omega\mathrm{d}r_\omega$$
$$\times\exp N\left[-\frac{M\beta}{2}\sum_{\rho,\mu\leqslant s}(g^\mu_\rho)^2-\frac{\alpha}{2}\mathrm{Tr}\ln\{-\beta M\bar{\bar{Q}}(q_{\rho\sigma})+\bar{\bar{I}}-\beta M\bar{\bar{D}}(q_\omega)\}\right.$$
$$-\frac{M^2\alpha\beta^2}{2}\sum_{\rho\neq\sigma}r_{\rho\sigma}q_{\rho\sigma}-\frac{M^2\alpha\beta^2}{2}\sum_\omega r_\omega q_\omega$$
$$+\ln\left\langle\!\!\left\langle\mathrm{Tr}\exp\left\{\frac{-\alpha\beta}{2M}\sum_\rho\left(\sum_a S^{(\rho)}_a\right)^2+\beta\sum_{\rho,\mu\leqslant s,a}(g^\mu_\rho+h^\mu)\xi^{(\mu)}S^{(\rho)}_a\right.\right.\right.$$
$$\left.\left.\left.+\frac{\alpha\beta^2}{2}\sum_{a,b,\rho\neq\sigma}r_{\rho\sigma}S^{(\rho)}_aS^{(\sigma)}_b+\frac{\alpha\beta^2}{2}\sum_{a,b,\omega}r_\omega S^{(\omega)}_aS^{(\omega)}_b\right\}\right\rangle\!\!\right\rangle\right]\tag{14}$$

where $\langle\langle\ldots\rangle\rangle$ in the RHS refers to the average over the condensed patterns. In the thermodynamic limit $N\to\infty$, the integrals over the auxiliary variables g^μ_ρ, $q_{\rho\sigma}$, $r_{\rho\sigma}$, q_ω and r_ω can be evaluated by the saddle-point method, from which the physical meaning of those variables is determined. The g^μ_ρ and $q_{\rho\sigma}$, for example, represent the overlaps with the μth ($\mu\leqslant s$) embedded pattern

$$g^\mu_\rho=\frac{1}{MN}\left\langle\!\!\left\langle\sum_i\xi^{(\mu)}_i\left\langle\sum_a S^{(\rho)}_{ia}\right\rangle\right\rangle\!\!\right\rangle$$

and the Edwards–Anderson order parameter (Edwards and Anderson 1975)

$$q_{\rho\sigma}=\frac{1}{M^2N}\left\langle\!\!\left\langle\sum_i\left\langle\left(\sum_a S^{(\rho)}_{ia}\right)\left(\sum_b S^{(\sigma)}_{ib}\right)\right\rangle\right\rangle\!\!\right\rangle$$

respectively, and so on. Then, in the context of the replica symmetric theory, setting $g_\rho^\mu = g^\mu$, $q_{\rho\sigma} = q$, $r_{\rho\sigma} = r$, $q_\omega = \bar{q}$, and $r_\omega = \bar{r}$ yields

$$\begin{aligned}\langle\langle Z^n\rangle\rangle = \exp N\bigg[& -\frac{Mn\beta}{2}\sum_{\mu=1}^{s}(g^\mu)^2 - \frac{\alpha}{2}\operatorname{Tr}\ln\{-\beta M\bar{\bar{Q}}(q) + \bar{\bar{I}} - \beta M\bar{\bar{D}}(\bar{q})\} \\ & - \frac{M^2\alpha\beta^2}{2} rqn(n-1) - \frac{M^2\alpha\beta^2}{2}\bar{r}\bar{q}n \\ & + \ln\bigg\langle\bigg\langle \operatorname{Tr}\exp\bigg\{\frac{-\alpha\beta}{2M}\sum_\rho\bigg(\sum_a S_a^{(\rho)}\bigg)^2 + \beta\sum_{\rho,\mu,a}(g^\mu + h^\mu)\xi^{(\mu)}S_a^{(\rho)} \\ & + \frac{\alpha\beta^2 r}{2}\sum_{a,b,\rho\neq\sigma} S_a^{(\rho)}S_b^{(\sigma)} + \frac{\alpha\beta^2\bar{r}}{2}\sum_{a,b,\omega} S_a^{(\omega)}S_b^{(\omega)}\bigg\}\bigg\rangle\bigg\rangle\bigg]. \end{aligned} \tag{15}$$

Introducing another set of auxiliary variables z, $y_\rho(\rho = 1, \ldots, n)$, we can calculate the $\operatorname{Tr}\exp\{\ldots\}$ part in the exponent of the $\langle\langle Z^n\rangle\rangle$,

$$\begin{aligned}\operatorname{Tr}\exp\{\ldots\} & \\ = \int & \sqrt{\frac{1}{2\pi}}\,\mathrm{d}z\prod_\rho\sqrt{\frac{M}{2\pi}}\,\mathrm{d}y_\rho\exp\bigg(-\frac{z^2}{2} - \frac{M}{2}\sum_\rho y_\rho^2\bigg)\operatorname{Tr} \\ & \times\exp\bigg[\sum_{a,\rho}\bigg\{\beta\sqrt{\alpha r}\,z + \sqrt{-\alpha\beta + M\alpha\beta^2(\bar{r} - r)}\,y_\rho \\ & + \beta\sum_\mu(g^\mu + h^\mu)\xi^{(\mu)}\bigg\}S_a^{(\rho)}\bigg] \\ = \int & \sqrt{\frac{1}{2\pi}}\,\mathrm{d}z\,\mathrm{e}^{-z^2/2}\prod_\rho\sqrt{\frac{M}{2\pi}}\,\mathrm{d}y_\rho \\ & \times\exp M\bigg\{-\frac{1}{2}\sum_\rho y_\rho^2 + \sum_\rho\ln\bigg[2\cosh\bigg\{\beta\sqrt{\alpha r}\,z \\ & + \sqrt{-\alpha\beta + M\alpha\beta^2(\bar{r} - r)}\,y_\rho + \beta\sum_\mu(g^\mu + h^\mu)\xi^{(\mu)}\bigg\}\bigg]\bigg\}. \end{aligned} \tag{16}$$

Making the ansatz (Bray *et al* 1986)

$$\gamma - r = \frac{v}{M} \qquad \bar{q} - q = \frac{u}{M} \tag{17}$$

and noting

$$\begin{aligned}\operatorname{Tr}\ln\{ & -\beta M\bar{\bar{Q}}(q) + \bar{\bar{I}} - \beta M\bar{\bar{D}}(\bar{q})\} \\ & = \ln\{1 - \beta M\bar{q} - (n-1)\beta Mq\} + (n-1)\ln\{1 - \beta M\bar{q} + \beta Mq\} \end{aligned} \tag{18}$$

we can rewrite the $\langle\langle Z^n\rangle\rangle$ in (15) as, $M\to\infty$ and $n\to 0$,

$$\begin{aligned}\langle\langle Z^n\rangle\rangle = \exp N\bigg[& -\frac{Mn\beta}{2}\sum_{\mu=1}^{s}(g^\mu)^2 + \frac{\alpha}{2}Mn\bigg(\frac{\beta q}{1 - \beta u}\bigg) - \frac{Mn}{2}\alpha\beta^2(ru + qv) \\ & + \ln\bigg\langle\bigg\langle\int\sqrt{\frac{1}{2\bar{u}}}\,\mathrm{d}z\,\mathrm{e}^{-z^2/2}\exp Mn\bigg\{-\frac{1}{2}y^2 + \ln\bigg[2\cosh\bigg\{\beta\sqrt{\alpha r}\,z \\ & + \sqrt{-\alpha\beta + \alpha\beta^2 v}\,y + \beta\sum_\mu(g^\mu + h^\mu)\xi^{(\mu)}\bigg\}\bigg]\bigg\}\bigg\rangle\bigg\rangle\bigg] \end{aligned} \tag{19}$$

where y is chosen to satisfy

$$y=\sqrt{-\alpha\beta+\alpha\beta^2 v}\cosh\left\{\beta\sqrt{\alpha r}z+\sqrt{-\alpha\beta+\alpha\beta^2 v}\,y+\beta\sum_{\mu}(g^\mu+h^\mu)\xi^{(\mu)}\right\} \tag{20}$$

which results from the saddle-point evaluation of the integrals over y_ρ in (16) in the limit $M\to\infty$ within the replica symmetric theory.

It is now straightforward to obtain the free energy (9) from (19):

$$-\beta f=\frac{-\beta}{2}\sum_{\mu=1}^{s}(g^\mu)^2+\frac{\alpha}{2}\frac{\beta q}{1-\beta u}-\frac{\alpha\beta^2}{2}(ru+qv)+\int\sqrt{\frac{1}{2\pi}}\mathrm{d}z\,\mathrm{e}^{-z^2/2}\left\langle\!\!\left\langle-\frac{1}{2}y^2\right.\right.$$
$$\left.\left.+\ln\left[2\cosh\left\{\beta\sqrt{\alpha r}\,z+\sqrt{-\alpha\beta+\alpha\beta^2 v}\,y+\beta\sum_{\mu=1}^{s}(g^\mu+h^\mu)\xi^{(\mu)}\right\}\right]\right\rangle\!\!\right\rangle. \tag{21}$$

Defining $Y=y/\sqrt{-\alpha\beta+\alpha\beta^2 v}$, (20) is rewritten as

$$Y=\tanh\left\{\beta\sqrt{\alpha r}\,z+(\alpha\beta^2 v-\alpha\beta)Y+\beta\sum_{\mu}(g^\mu+h^\mu)\xi^{(\mu)}\right\}. \tag{22}$$

Local minima of f in (21) with respect to variations in g^μ, r, v, u, and q represent the ergodic components of the thermodynamic system with naive mean-field Hamiltonian (8), each of which is separated by barriers of $O(N)$. It follows that

$$\begin{aligned}
\frac{\partial f}{\partial g^\mu}=0 \quad &: \quad g^\mu=\int\sqrt{\frac{1}{2\pi}}\mathrm{d}z\,\mathrm{e}^{-z^2/2}\langle\!\langle\xi^{(\mu)}Y\rangle\!\rangle\\
\frac{\partial f}{\partial v}=0 \quad &: \quad q=\int\sqrt{\frac{1}{2\pi}}\mathrm{d}z\,\mathrm{e}^{-z^2/2}\langle\!\langle Y^2\rangle\!\rangle\\
\frac{\partial f}{\partial r}=0 \quad &: \quad \sqrt{\alpha r}\,\beta u=\int\sqrt{\frac{1}{2\pi}}\mathrm{d}z\,\mathrm{e}^{-z^2/2}\langle\!\langle zY\rangle\!\rangle\\
\frac{\partial f}{\partial u}=0 \quad &: \quad r=\frac{q}{(1-\beta u)^2}\\
\frac{\partial f}{\partial q}=0 \quad &: \quad \beta v=\frac{1}{1-\beta u}.
\end{aligned} \tag{23}$$

As in the AGS theory, retrieval states are described by $g^\mu=g\delta_{\mu\nu}$. In that case, rescaling the variables as $\xi^{(1)}Y\to Y$ and $\xi^{(1)}z\to z$ and setting $h=0$ in (23) yields

$$Y=\tanh\left\{\beta\sqrt{\alpha r}\,z+\frac{\alpha\beta^2 u}{1-\beta u}Y+\beta g\right\} \tag{24a}$$

$$g=\int\sqrt{\frac{1}{2\pi}}\mathrm{d}z\,\mathrm{e}^{-z^2/2}Y \tag{24b}$$

$$q=\int\sqrt{\frac{1}{2\pi}}\mathrm{d}z\,\mathrm{e}^{-z^2/2}Y^2 \tag{24c}$$

$$\sqrt{\alpha r}\,\beta u=\int\sqrt{\frac{1}{2\pi}}\mathrm{d}z\,\mathrm{e}^{-z^2/2}zY \tag{24d}$$

with $r=q/(1-\beta u)^2$.

Equations ($24a$–d) constitute a basic set of equations for the description of the analogue neural network which should be contrasted with the result of AGS. Unlike their case, the Y to be averaged over the Gaussian noise is given only implicitly by ($24a$). The difference has to be attributed to the very nature of the naive mean-field model in which the Onsager reaction term is absent in the TAP equation (4). We note that equation for $Y(z)$ ($24a$) admits, in general, multisolutions. In that case, the solution should be chosen which maximizes the exponent of the integrant in (16): a kind of Maxwell's rule tells us that the available solution $Y(Z)$ as a function of z undergoes a discontinuous jump at $z=-g/\sqrt{\alpha r}$ from $-Y_*$ to Y_* with Y_* determined by $Y_*=\tanh(\alpha\beta^2 u/(1-\beta u))\,Y_*]$.

Taking the limit $\beta\to\infty$ considerably simplifies the analysis of (24). One can then easily check that the same result as obtained by AGS follows, which is consistent with the observation that the problem of the presence or absence of the Onsager reaction term in the TAP equation does not matter at zero temperature in the spin-glass theory (Bray *et al* 1986).

For finite values of β, on the other hand, a certain discrepancy from the AGS theory will be found. To see this, we first investigate the phase transition of second order from the disordered, paramagnetic phase ($g=q=0$) to the spin-glass phase ($g=0$, $q\neq 0$).

Assuming $g=0$ and q to be infinitesimally small, we expand ($24a$) to obtain in the leading order of q

$$Y=\frac{\beta\sqrt{\alpha r}}{1-\alpha\beta^2 u/(1-\beta u)}\,z \tag{25}$$

Substituting into ($24c$) and using $r=q/(1-\beta u)^2$, we have

$$q=\frac{\alpha\beta^2}{(1-\alpha\beta^2 u/(1-\beta u))^2(1-\beta u)^2}\,q+\mathrm{O}(q^2) \tag{26}$$

from which the condition determining the phase transition point is given by

$$\frac{\alpha\beta^2}{(1-\alpha\beta^2 u/(1-\beta u))^2(1-\beta u)^2}=1. \tag{27}$$

Another complementary condition will be obtained from (25) and ($24d$) as

$$u=\frac{1}{1-\alpha\beta^2 u/(1-\beta u)}. \tag{28}$$

Both (27) and (28) yield the phase transition point β_G

$$\beta_G=\frac{1}{1+2\sqrt{\alpha}} \tag{29}$$

which differs from that of AGS theory, as is expected.

Retrieval and spin-glass phases are studied by numerically solving (24), with the result that the phase diagram of the analogue network is qualitatively the same as that of the stochastic Ising model network of AGS. As in the AGS theory, critical storage capacity α_c is determined from the condition of a sudden disappearance of retrieval phase ($g\neq 0$) coexisting with the spin-glass phase ($g=0$, $q\neq 0$) as α is changed with β kept fixed. The α_c against β^{-1} curve thus obtained numerically is shown in figure

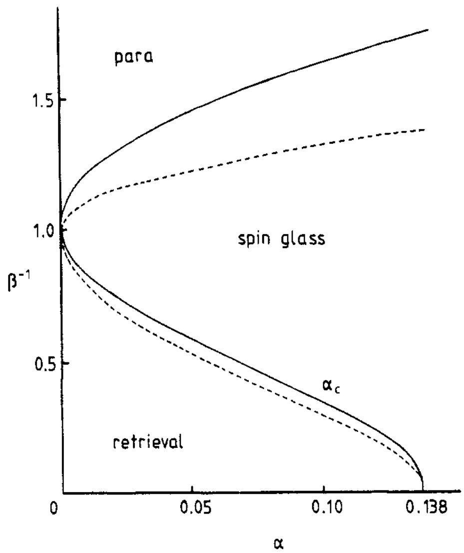

Figure 1. Phase diagram for the analogue neural network with $V(x) = \tanh(\beta x)$. Plots of critical storage capacity α_c against the inverse of the analogue gain β are displayed together with the phase transition points β_G of second order between paramagnetic and spin-glass phases. Both of the phase boundaries are based on the saddle-point equations (24). The α_c against β^{-1} curve as well as α against β_G^{-1} for the stochastic Ising model network (AGS) are also shown for comparison (broken line). The difference in the α_c as well as in the β_G is attributed to the fact that the analogue network (naive mean-field model) differs from the stochastic Ising network of AGS by the absence of the Onsager reaction term in the TAP equation.

1, where the corresponding curve in AGS theory is also presented for comparison. Apart from the two limiting case $\beta = \infty$ ($\alpha_c = 0.138$) and $\beta = 1$ ($\alpha_c = 0$), for which no discrepancy in α_c is found at all between the analogue and (stochastic) Ising model networks, the critical storage capacity of the present analogue network is observed to be a little larger than that of the stochastic Ising model network with corresponding inverse temperature, although any qualitative difference cannot be found.

As was discussed in our previous paper (Fukai and Shiino 1990), the present result is consistent with the result of another kind of analysis concerning the counting of the number of equilibrium states in the present analogue networks: The $\alpha_c - \beta^{-1}$ curve obtained from a bifurcation analysis based on the thermodynamic calculation of the naive mean-field free energy is in perfect agreement with that inferred from the evaluation of the critical number density of the equilibrium states of the dynamical equation of the analogue network.

Finally we will comment on the result reported quite recently on a similar analogue neural network by Marcus *et al* (1990). They obtained the same set of equations as in the AGS theory for the calculation of the critical storage capacity, with the result giving forth a contradiction to the present result (24) of our analysis. To be more specific, their set of equations will be inconsistent with their result for the linear stability analysis of the phase transition between the paramagnetic and spin-glass phases, which is in agreement with the result of our analysis. Their analysis was based on the signal to noise analysis (Geszti 1990), which may be inappropriate in a rigorous treatment of the noise part originating from the non-condensed patterns.

In summary, conducting the statistical mechanical analysis of the naive mean field model equivalent to the nonlinear analogue network, we have obtained, within the framework of the replica symmetric theory, a set of equations describing local minima of the averaged free energy, which has been found to differ from that for the stochastic Ising network due to the absence of the Onsager reaction term in the TAP equation. Based on the equations, critical storage capacity has been calculated numerically as a function of the analogue gain β. Although overall behaviour of the $\alpha_c - \beta^{-1}$ curve is qualitatively the same as the corresponding stochastic Ising networks, a small increase in α_c has been found compared with the counterpart of the stochastic Ising network.

References

Amit D J 1989 *Modelling Brain Function* (Cambridge: Cambridge University Press)
Amit D J, Gutfreund H and Sompolinsky H 1985a *Phys. Rev. Lett.* **55** 1530
—— 1985b *Phys. Rev.* A **32** 1007
—— 1987 *Ann. Phys.* **173** 30
Bray A J, Sompolinsky H and Yu C 1986 *J. Phys. C: Solid State Phys.* **19** 6389
Edwards S F and Anderson P W 1975 *J. Phys. F: Met. Phys.* **5** 965
Fukai T and Shiino M 1990 *Preprint* Gumma Women's College/Tokyo Institute of Technology
Geszti T 1990 *Physical Models of Neural Networks* (Singapore: World Scientific)
Hopfield J J 1982 *Proc. Natl Acad. Sci., USA* **79** 2554
—— 1984 *Proc. Natl Acad. Sci., USA* **81** 3088
Hopfield J J and Tank D W 1985 *Biol. Cybern.* **52** 141
Hemmen J L and Morgenstern I 1987 *Lecture Notes in Physics* vol 275 (Berlin: Springer)
Mezard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
Marcus C M and Waugh F R and Westervelt R M 1990 *Phys. Rev.* A **41** 3355
Thouless D J, Anderson P W and Palmer R G 1977 *Phil. Mag.* **35** 593