Fluctuating Paths and Fields, Eds. W. Janke, A. Pelster, H.-J. Schmidt, and M. Bachmann (World Scientific, Singapore, 2001).

GAUGE SYMMETRY AND NEURAL NETWORKS

T. MATSUI

Department of Physics, Kinki University, Higashi-Osaka, 8502 Japan
E-mail: matsui@phys.kindai.ac.jp

We propose a new model of neural network. It consists of spin variables to describe the state of neurons as in the Hopfield model and new gauge variables to describe the state of synapses. The model possesses local gauge symmetry and resembles lattice gauge theory of high-energy physics. Time dependence of synapses describes the process of learning. The mean-field theory predicts a new phase corresponding to the confinement phase, in which the brain loses the ability of learning and memory.

1 Introduction

The Hopfield model of neural network [1] succeeds to explain some basic functions of human brain such as the associated memory. However, to be a more realistic model, at least the following points shall be taken into account:

- effects of external stimulations through eyes and ears on neurons,
- effects of time variations of synapses on neurons.

The second point is essential to describe the function of learning, since the possible patterns to memorize are completely determined according to the strengths of synapse connections among neurons as long as they are time-independent. Their time-dependence induces the process of learning itself.

In Section 2, we review the Hopfield model briefly. Then, in Section 3, we propose a new model of neural network, in which the strengths of synapse connections are regarded as gauge connections varying in time according to the gauge principle. By using the mean-field theory, we see that the model predicts a new state of the brain in which both learning and memory are impossible. An outlook is presented in Section 4.

Fluctuating Paths and Fields, Eds. W. Janke, A. Pelster, H.-J. Schmidt, and M. Bachmann (World Scientific, Singapore, 2001).

2 Hopfield Model

Let us briefly review the framework of the Hopfield model, where the energy $E_H(\{S_i\})$ is given by

$$E_H(\{S_i\}) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} J_{ij} S_i S_j. \tag{1}$$

Here, $S_i = \pm 1$ is the Ising spin variable to describe the state of the i-th neural cell ($i = 1, 2, \ldots, N$). When excited, it is $S_i = 1$, when unexcited we have $S_i = -1$. J_{ij} is a given real constant that expresses the strength of synapse connection for the signal propagating from the j-th cell to the i-th cell.

The time evolution of $S_i(t)$ for every discrete time interval ϵ (often set unity) is governed by the following equation:

$$S_i(t+\epsilon) = \text{sgn}\left[-\frac{\partial E_H}{\partial S_i}(t)\right] = \text{sgn}\left[\sum_j J_{ij} S_j(t)\right]. \tag{2}$$

Thus, $J_{ij} > 0$ tries to proliferate (un)excited cells, while $J_{ij} < 0$ prefers mixtures of excited and unexcited ones. If the system converges into a certain configuration of $\{S_i\}$ after a sufficiently long time, it corresponds to recalling certain pattern. Such a configuration should be a stationary point of E_H, i.e. $\partial E_H/\partial S_i = 0$ for every i. All these configurations are determined once the coefficients J_{ij} are given.

Practically speaking, the rule (2) may not necessarily hold over the whole time interval due to unavoidable errors in signal propagations. Such a situation may be simulated by adding random noises $\eta_i(t)$ into the square bracket in the right-hand side of (2), whose strength can be identified as a fictitious "temperature" T. If T is large, the error in signal propagations occurs frequently. Thus it is interesting to study statistical mechanics of the system E_H by using the Boltzmann distribution. The partition function Z_H is given by

$$Z_H = \prod_i \sum_{S_i=\pm 1} \exp(-\beta E_H), \quad \beta \equiv 1/T. \tag{3}$$

In the case that all J_{ij} are positive, the system has two phases:

- a ferromagnetic phase below a certain critical temperature T_c, $T < T_c$, in which there is a long-range order, and the average $\langle S_i \rangle \neq 0$,

Fluctuating Paths and Fields, Eds. W. Janke, A. Pelster, H.-J. Schmidt, and M. Bachmann (World Scientific, Singapore, 2001).

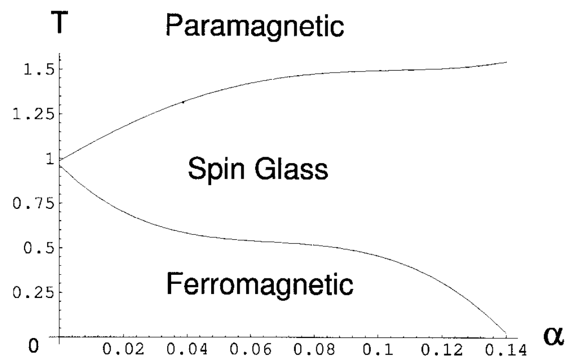

Figure 1. Phase structure of the Hopfield model in the $\alpha(\equiv M/N) - T$ plane.

Table 1. Phases of the Hopfield model.

Phase	$\sum_i \langle S_i \rangle$	$\sum_i \langle S_i \rangle^2$	Property
Ferromagnetic	$\neq 0$	$\neq 0$	memory
Spin glass	0	$\neq 0$	false memory
Paramagnetic	0	0	no memory

- a paramagnetic phase above T_c, $T > T_c$, in which S_i are random and $\langle S_i \rangle = 0$.

The ferromagnetic phase corresponds to the state of clear memory, while in the paramagnetic phase no definite patterns can persist. If J_{ij} is complicated, there arises a spin-glass phase as we shall see.

Explicitly, let us fix J_{ij} according to Hebb's rule as

$$J_{ij} = \frac{1}{N} \sum_{\alpha=1}^{M} \xi_i^\alpha \xi_j^\alpha, \tag{4}$$

where we prepare M patterns $S_i = \xi_i^\alpha$ $(\alpha = 1, \ldots, M)$ to recall. The replica method gives rise to the phase diagram shown in Fig. 1 [2]. Each phase is explained in Table 1.

Fluctuating Paths and Fields, Eds. W. Janke, A. Pelster, H.-J. Schmidt, and M. Bachmann (World Scientific, Singapore, 2001).

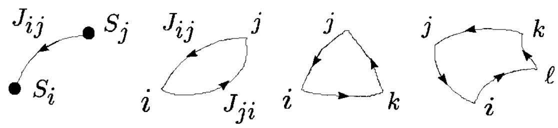

Figure 2. Graphical representation of each term in E of Eq. (6).

3 New Model with Local Gauge Symmetry

As pointed out in Section 1, one needs the time variation of J_{ij} to incorporate the function of learning. There are various approaches for this point. We regard both S_i and J_{ij} as dynamical variables and treat them on an equal footing. Let us assume that their time dependence is controlled such that they reach a local minimum of the new energy $E(\{S_i, J_{ij}\})$. In order to determine E, we impose the condition that E is locally gauge invariant under the following gauge transformation:

$$S_i \to S'_i \equiv V_i S_i,\ J_{ij} \to J'_{ij} \equiv V_i J_{ij} V_j,\ E(\{S'_i, J'_{ij}\}) = E(\{S_i, J_{ij}\}), \quad (5)$$

where $V_i = \pm 1$ is the $Z(2)$ variable associated with the i-th cell. Since J_{ij} describes the state of the synapse connecting the i-th and the j-th cells, it is natural to regard it as the connection of gauge theory. The neural network may possess certain conserved quantities in association with the long-term memory. The local gauge symmetry we address may respect such a conservation law. This point will be reported in detail in a separate publication [3]. It is often stressed that the connections J_{ij} and J_{ji} are independent (asymmetric). Then a general form of $E(\{S_i, J_{ij}\})$ and the partition function Z may be given by

$$\begin{aligned} E &= -\frac{1}{2}\sum_{i,j} S_i J_{ij} S_j + \frac{g_2}{2}\sum_{i,j} J_{ij} J_{ij} \\ &\quad + \frac{g_3}{3!}\sum_{i,j,k} J_{ij} J_{jk} J_{ki} + \frac{g_4}{4!}\sum_{i,j,k,\ell} J_{ij} J_{jk} J_{k\ell} J_{\ell i} + \dots, \\ Z &= \prod_i \sum_{S_i=\pm 1} \prod_{i\neq j} \int dJ_{ij} \exp(-\beta E), \quad \beta \equiv 1/T. \end{aligned} \quad (6)$$

Since $V_i^2 = 1$, each term of E is gauge invariant (see Fig. 2). E takes a form very similar to the lattice gauge theory in particle physics, where S_i

Fluctuating Paths and Fields, Eds. W. Janke, A. Pelster, H.-J. Schmidt, and M. Bachmann (World Scientific, Singapore, 2001).

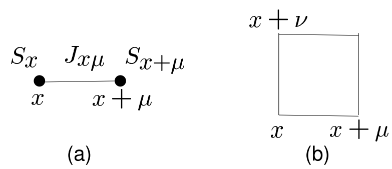

Figure 3. Graphical representation of E of Eq. (7): (a) λ term; (b) g^{-2} term.

corresponds to a matter field and J_{ij} to an exponentiated gauge field. If the parameters $g_2, g_3, g_4, ...$ are set to zero, E reduces to E_H of (1).

3.1 Model I

To be explicit, we need to specify the model further. Let us first consider the $Z(2)$ Higgs gauge model on a 3D cubic lattice,

$$E_{\rm I} = -\lambda \sum_x \sum_{\mu=1}^{3} S_{x+\mu} J_{x\mu} S_x - \frac{1}{g^2} \sum_x \sum_{\mu<\nu} J_{x\mu} J_{x+\mu,\nu} J_{x+\nu,\mu} J_{x\nu},$$
$$Z_{\rm I} = \prod_x \sum_{S_x=\pm 1} \prod_{x,\mu} \sum_{J_{x\mu}=\pm 1} \exp(-\beta E_{\rm I}) \equiv \exp(-\beta F_{\rm I}), \tag{7}$$

where x denotes the lattice site on which S_x lives, and μ $(= 1, 2, 3)$ both the direction and the unit vector. We consider only the connections between the nearest-neighbor sites $(x, x+\mu)$ and treat them as a symmetric $Z(2)$ variable on a link $(x, x+\mu)$; $J_{x,x+\mu} = J_{x+\mu,x} \equiv J_{x\mu} = \pm 1$. The λ term and the $1/g^2$-term are depicted in Fig. 3.

The time evolution of $J_{x\mu}$ may be given by the similar rule as (2),

$$S_x(t+\epsilon) = \mathrm{sgn}\left[-\frac{\partial E_{\rm I}}{\partial S_x}(t) + \eta_x(t)\right],$$
$$J_{x\mu}(t+\alpha\epsilon) = \mathrm{sgn}\left[-\frac{\partial E_{\rm I}}{\partial J_{x\mu}}(t) + \zeta_{x\mu}(t)\right], \tag{8}$$

where α sets the ratio of the two time scales for S_x and $J_{x\mu}$. We report our study of (8) elsewhere [4].

Now let us study the phase diagram of $E_{\rm I}$ by using the mean-field theory, which is formulated as a variational principle as follows. Let us introduce a variational energy E_0. Then the Jensen-Peierls inequality gives rise to

$$F_{\rm I} \le F_0 + \langle E_{\rm I} - E_0 \rangle_0,$$

Fluctuating Paths and Fields, Eds. W. Janke, A. Pelster, H.-J. Schmidt, and M. Bachmann (World Scientific, Singapore, 2001).

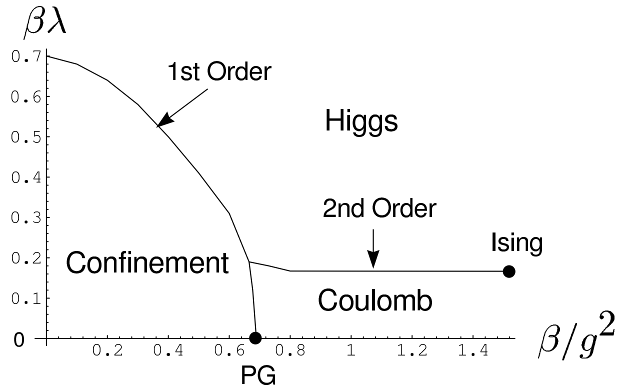

Figure 4. Phase diagram of Model I of Eq. (7). The point marked as Ising locates the second-order transition point of the Ising model. The point PG locates the first-order transition point of the $Z(2)$ pure gauge theory.

$$\begin{aligned} Z_0 &= \mathrm{Tr}\ \exp(-\beta E_0) \equiv \exp(-\beta F_0), \\ \langle O\rangle_0 &= Z_0^{-1}\ \mathrm{Tr}\ O\ \exp(-\beta E_0), \end{aligned} \tag{9}$$

where Tr implies $\prod_x \sum_{S_x=\pm 1} \prod_{x,\mu} \sum_{J_{x\mu}=\pm 1}$. We choose the variational parameters in E_0 so that the right-hand-side of inequality reaches the minimum. For E_0 we assume the translational invariance of mean fields and employ the single-site and single-link energy,

$$E_0 = -\sum_x \sum_\mu W_{x\mu} J_{x\mu} - \sum_x h_x S_x, \tag{10}$$

with the two variational parameters $W_{x\mu} = W$ and $h_x = h$. The result is given in Table 2 and Fig. 4. As shown in Table 2, one may take $\langle S_x\rangle$ as an order parameter to judge whether the system succeeds to recall definite patterns, and $\langle J_{x\mu}\rangle$ is taken to judge whether the system is able to learn some new patterns. In the confinement phase, neither memory nor learning is possible. This phase is missing in the Hopfield model. We note that Monte Carlo simulations of the 3D $Z(2)$ Higgs gauge model exhibit these three phases, but the phase boundary of Higgs and confinement phases does not continue to $\beta/g^2 = 0$ but terminates at some finite value. These two phases can reach each other smoothly by contouring the end point. This "complementarity" reflects that $|J_{x\mu}| = 1$. This is proved by a rigorous treatment [5], but not predicted correctly in our variational treatment [6].

Fluctuating Paths and Fields, Eds. W. Janke, A. Pelster, H.-J. Schmidt, and M. Bachmann (World Scientific, Singapore, 2001).

Table 2. Phases of Model I of Eq. (7).

Phase	$\langle S_x \rangle$	$\langle J_{x\mu} \rangle$	Memory	Learning	Hopfield Model
Higgs	$\neq 0$	$\neq 0$	yes	yes	Ferromagnetic
Coulomb	0	$\neq 0$	no	yes	Paramagnetic
Confinement	0	0	no	no	not available

In order to answer to what extent these results are trustworthy, we introduce and study two other models, Model II and Model III.

3.2 Model II

The framework and the energy $E_{\rm II}$ of Model II is the same as in Model I, but here we allow the additional state $J_{x\mu} = 0$ which describes the possibility that the connection between x and $x + \mu$ is missing:

$$E_{\rm II} = E_{\rm I}, \quad Z_{\rm II} = \prod_x \sum_{S_x=\pm 1} \prod_{x,\mu} \sum_{J_{x\mu}=0,\pm 1} \exp(-\beta E_{\rm II}). \tag{11}$$

The phase diagram calculated by the similar mean-field theory is shown in Fig. 5. The global structure remains the same as in Fig. 4, although the region of the confinement phase is enlarged as expected since the added states clearly favor this phase.

3.3 Model III

In Model III, we introduce two independent $Z(2)$ variables $J_{x\mu}$ and $\bar{J}_{x\mu}$ for the synapse between x and $x + \mu$ to take into account their independence as

$$J_{x\mu} \equiv J_{x,x+\mu}, \quad \bar{J}_{x\mu} \equiv J_{x+\mu,x}. \tag{12}$$

We also define $J_{x,-\mu} \equiv J_{x-\mu,x}$. The energy $E_{\rm III}$ is then given by

$$\begin{aligned} E_{\rm III} = &-\lambda \sum_x \left(\sum_{\pm\mu} \bar{J}_{x,\pm\mu} S_{x\pm\mu} \right) \left(\sum_{\pm\nu} J_{x,\pm\nu} S_{x\pm\nu} \right) \\ &- \frac{1}{g^2} \sum_x \sum_{\mu<\nu} \left[\bar{J}_{x\mu} \bar{J}_{x+\mu,\nu} J_{x+\nu,\mu} J_{x\nu} + (\mu \leftrightarrow \nu) \right]. \end{aligned} \tag{13}$$

We note that the expression $J_{ij} S_i S_j$ in $E_{\rm H,\ I,\ II}$ washes out the asymmetry $J_{ij} \neq J_{ji}$, while the first term of (13) reflects it. Each term in E is depicted in

Fluctuating Paths and Fields, Eds. W. Janke, A. Pelster, H.-J. Schmidt, and M. Bachmann (World Scientific, Singapore, 2001).

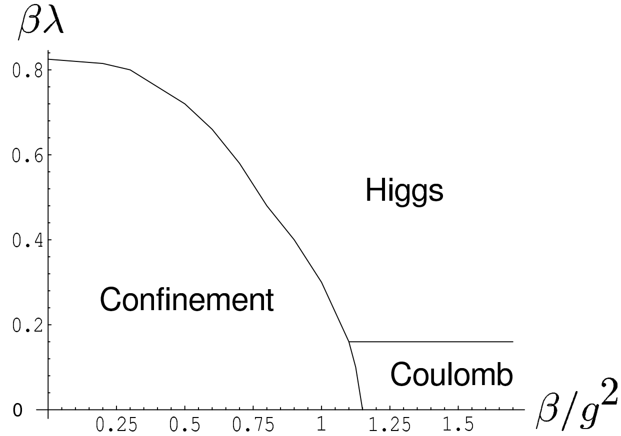

Figure 5. Phase diagrams of Model II of Eq. (11).

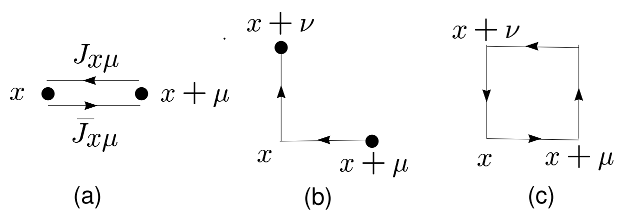

Figure 6. Graphical representation of Model III of Eq. (13): (a) $J_{x\mu}$ and $\bar{J}_{x\mu}$; (b) λ term; (c) g^{-2} term.

Fig. 6. The phase diagram in mean-field theory is shown in Fig. 7. The global structure still remains unchanged, although the region of the confinement phase is diminished considerably. This may be understood since the first term in $E_{\rm III}$ is bilinear in J_{ij}, in contrast to $E_{\rm H,\ I,\ II}$, and it favors nonvanishing J_{ij}.

4 Summary and Outlook

Our results may be summarized with some outlook as follows:

- Due to the dynamical variables J_{ij}, a new confinement phase appears at high temperatures, which describes the new state of no ability of learning

Fluctuating Paths and Fields, Eds. W. Janke, A. Pelster, H.-J. Schmidt, and M. Bachmann (World Scientific, Singapore, 2001).

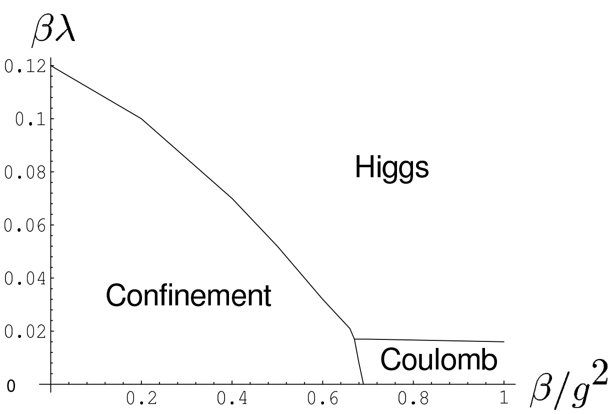

Figure 7. Phase diagrams of Model III of Eq. (13).

and memory.

- To describe the spin-glass phase, further study of long-range correlations and/or frustrations is necessary.
- Relaxing of $J_{ij} = \pm 1(,0)$ to $-\infty < J_{ij} < \infty$ may be interesting, but requires a detailed form of the energy.
- Study of the time evolution of J_{ij} and S_i may describe the mechanism of learning such as the process to forget the patterns.
- Study of the effect of local gauge symmetry on brain function on a "quantum" level is interesting. Introduction of gauged versions of quantum brain dynamics [7] and cellular automata with Penrose's idea [8] may be the first step.

Acknowledgments

I thank Prof. Hagen Kleinert for fruitful discussions on various fields of physics during my pleasant stay at the Freie Universität Berlin in 1983-1990. I also appreciate discussions with Dr. Kazuhiko Sakakibara and Mr. Motohiro Kemuriyama.

References

[1] J.J. Hopfield, *Proc. Nat. Acad. Sci. USA* **79**, 2554 (1982).
[2] D.J. Amit, H. Gutfreund, and H. Sompolinsky, *Ann. Phys. (N.Y.)* **173**,

Fluctuating Paths and Fields, Eds. W. Janke, A. Pelster, H.-J. Schmidt, and M. Bachmann (World Scientific, Singapore, 2001).

30 (1987).

[3] T. Matsui and K. Sakakibara, in preparation.

[4] M. Kemuriyama and T. Matsui, in preparation.

[5] E. Fradkin and S. Shenker, *Phys. Rev. D* **19**, 3682 (1979).

[6] E. Brézin and J.M. Drouffe, *Nucl. Phys. B* **200**, 93 (1982).

[7] C. Stuart, Y. Takahashi, and H. Umezawa, *J. Theor. Biol.* **71**, 605 (1978).

[8] S. Hameroff and R. Penrose, *J. Consciousness Study* **3**, 36 (1996).