Statistical mechanics of a Hopfield neural-network model in a transverse field

Yu-qiang Ma
*Chinese Center of Advanced Science and Technology (World Laboratory), P.O. Box 8730, Beijing 100080, People's Republic of China and Department of Physics, Nanjing University, Nanjing 210008, People's Republic of China**

Yue-ming Zhang
Department of Physics, Nanjing Medical College, Nanjing 210029, People's Republic of China

Yu-gang Ma
Institute of Nuclear Research, Academia Sinica, P.O. Box 800-204, Shanghai 201800, People's Republic of China

Chang-de Gong
Chinese Center of Advanced Science and Technology (World Laboratory), P.O. Box 8730, Beijing 100080, People's Republic of China and Department of Physics, Nanjing University, Nanjing 210008, People's Republic of China

(Received 11 January 1993)

We construct a soluble model to describe the retrieval properties of the quantum transverse Hopfield model in neural networks. A set of equations for the order parameters describing retrieval and spin-glass phases of the Hopfield model in a transverse field are obtained. Based on these equations, phase diagrams are examined and the memory storage capacity of the networks is analyzed as a function of the temperature and transverse field.

PACS number(s): 64.60.Cn, 87.10.+e, 75.10.Hk

Since Hopfield's work [1] on the modeling of neural networks for associative memory, physical models of neural networks have been extensively investigated [2–16]. In many of the models, the network consists of N neurons whose internal connections (synapses) are updated to facilitate the storage and the retrieval of information. Usually the synapses are designed so that a given set of states of the system becomes fixed attractors of its dynamic evolution.

In the Hopfield model of a neural network, each neuron is represented by two-state (active-passive) Ising spins $S_i=\pm 1$. A set of $p=\alpha N$ uncorrelated patterns $\{\xi_i^\mu\}$ $\{i=1,\ldots,N,\ \mu=1,\ldots,p\}$ in which ξ_i^μ is either $+1$ or -1 with equal probability is encoded in the interaction matrix by the Hebb rule,

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

Here, the p patterns $\{\xi_i^\mu\}$ constitute the embedded memories. The retrieval of a particular memory is achieved when the system starting from some initial configuration (imposed by an external stimulus) evolves under its own dynamics to a stationary configuration $\{S_i\}$, which is strongly correlated with that memory. Some of the investigations have concentrated on the ability of these networks to function as associative memories which retrieve a complete stored information if an external longitudinal field is exerted [17–20]. However, the Hopfield model for N interacting neurons in transverse field Γ, with Hamiltonian

$$\mathcal{H}=-\sum_{i,j}J_{ij}S_i^zS_j^z-\Gamma\sum_i S_i^x \ , \tag{2}$$

is a theoretical construct which introduces quantum effects to a classical problem in a natural way [21]. Here S_i^z and S_i^x are the Pauli matrices at the ith spin and Γ is a transverse field which represents the tunneling frequency of the neurons. Since a transverse field Γ applied in the spin system brings about a quantum effect by causing spin flips, the requisite noncommutativity of operators in the Hamiltonian creates a potentially difficult technical problem. On the other hand, the study of the models described above is interesting not only in the context of neural-network models but also in the context of the statistical mechanics of quantum disordered magnetic systems. The Hamiltonian defined in (1) and (2) with $p=1$ is an infinite-range Mattis model [22] in a transverse field [23]. The transition can be driven by both temperature as well as the transverse field. The critical behavior (exponents) for a d-dimensional system is the same as that of the d-dimensional Ising system for finite temperature transitions, and at $T=0$ K is identical to that of a pure Ising system in $(d+1)$ dimensions. The model (1) and (2) with $p>1$ represents an intermediate case. In the classical case ($\Gamma=0$), van Hemmen [24] introduced and solved a related model with $p=2$ and his mean-field equation has been extended to arbitrary p by Provost and Vallee [25]. However, as far as we know, up to now no studies have made contact with the quantum version of above models. In this paper, we will use a simple method to study the effects of the transverse field on the retrieval states of quantum Hopfield models.

Let us introduce the effective Hamiltonian for the ith spin,

$$\mathcal{H}_i=-h_iS_i^z-\Gamma S_i^x \ , \tag{3}$$

1063-651X/93/47(6)/3985(3)/$06.00

where the local field $h_i=\sum_j J_{ij}m_j$ and m_j is the average magnetization at site j. The corresponding partition function becomes

$$Z=\mathrm{Tr}e^{-\beta\mathcal{H}_i}=2\cosh[\beta(h_i^2+\Gamma^2)^{1/2}] , \quad (4)$$

where β is the inverse temperature. Then the average magnetization m_i at site i is given by the set of equations

$$m_i=h_i(h_i^2+\Gamma^2)^{-(1/2)}\tanh[\beta(h_i^2+\Gamma^2)^{1/2}] . \quad (5)$$

When storing p patterns in the network, one monitors the degree of associative recall of a single pattern μ by the overlap,

$$q^\mu=N^{-1}\sum_i \xi_i^\mu m_i(h_i,\Gamma) , \quad \mu=1,\ldots,\alpha N . \quad (6)$$

In the large-N limit, Eq. (6) can be written as an integral over the distribution of local fields $P(H^\mu)$,

$$q^\mu=\int dH^\mu P(H^\mu)m_i(H^\mu,\Gamma) , \quad (7)$$

where $H_i^\mu=\xi_i^\mu h_i$. We will limit ourselves to the thermodynamic limit, and study the configuration $\{S_i\}$ having a macroscopic overlap with only one of the stored patterns (the first, say) and microscopic overlaps with all the other $p-1$ patterns. This Mattis-state solution is described by the order parameter of the form $q^\mu\equiv q\delta_{\mu 1}$ or $\mathbf{q}=(q,0,\ldots,0)$ in the vector notation. Then, the local field for pattern 1,

$$H_i^1=\xi_i^1\sum_{j(\neq i)} J_{ij}m_j , \quad (8)$$

can be split into two parts,

$$H_i^1=q+\xi_i^1\sum_{k>1}\xi_i^k q^k . \quad (9)$$

Equation (9) acts as a Gaussian noise with mean q and variance σ^2 given by

$$\sigma^2=\sum_{k>1}(q^k)^2 . \quad (10)$$

To evaluate the sum of squares in Eq. (10), we first write the overlaps with the uncondensed patterns using Eq. (6),

$$q^k=N^{-1}\sum_i \xi_i^k m_i\left[\sum_{\rho(\neq k)}\xi_i^\rho q^\rho+\xi_i^k q^k,\Gamma\right] . \quad (11)$$

Here the single term ($\rho=k$) is small compared to the sum over all the rest ($\rho\neq k$). We expand $m_i(h_i,\Gamma)$ to first order in q^k giving

$$q^k=N^{-1}\sum_i \xi_i^k\left[m_i\left[\sum_{\rho(\neq k)}\xi_i^\rho q^\rho,\Gamma\right]+\xi_i^k q^k m_i'\left[\sum_{\rho(\neq k)}\xi_i^\rho q^\rho,\Gamma\right]\right] , \quad (12)$$

where $m_i'(h_i,\Gamma)=dm_i(h_i,\Gamma)/dh_i$. The missing $\rho=k$ term in the argument of $m_i'(\Gamma,\sum_{\rho(\neq k)}\xi_i^\rho q^\rho)$ only affects the value $m_i'(\Gamma,\sum_\rho \xi_i^\rho q^\rho)$ to order $O(1/N)$ which we neglect by taking the argument to be the whole h_i. We now define the quantity

$$c\equiv N^{-1}\sum_i m_i'(h_i,\Gamma) , \quad (13)$$

and the Edwards-Anderson order parameter Q,

$$Q\equiv N^{-1}\sum_i m_i^2(h_i,\Gamma) . \quad (14)$$

From Eqs. (12)–(14), we have

$$(q^k)^2=Q/N(1-c)^2 . \quad (15)$$

From Eqs. (10) and (15), the variance σ^2 of the local-field distribution is given by

$$\sigma^2=\alpha Q/(1-c)^2 . \quad (16)$$

Because $m^2(h_i,\Gamma)$ and $m'(h_i,\Gamma)$ are both even functions, we can write the averages in Eqs. (13) and (14) in terms of H_i^1 rather than h_i, and finally as integrals over the distribution of local fields $P(H^1)$. We have

$$q=\int(2\pi)^{-(1/2)}e^{-y^2/2}dy(q+\sqrt{\alpha r}y)[\Gamma^2+(q+\sqrt{\alpha r}y)^2]^{-(1/2)}\tanh\{\beta[\Gamma^2+(q+\sqrt{\alpha r}y)^2]^{1/2}\} , \quad (17)$$

$$c=\int(2\pi)^{-(1/2)}e^{-y^2/2}dy(\Gamma^2[\Gamma^2+(q+\sqrt{\alpha r}y)^2]^{-(3/2)}\tanh\{\beta[\Gamma^2+(q+\sqrt{\alpha r}y)^2]^{1/2}\}+\beta(q+\sqrt{\alpha r}y)^2[\Gamma^2+(q+\sqrt{\alpha r}y)^2]^{-1}\mathrm{sech}^2\{\beta[\Gamma^2+(q+\sqrt{\alpha r}y)^2]^{1/2}\}) , \quad (18)$$

$$Q=\int(2\pi)^{-(1/2)}e^{-y^2/2}dy(q+\sqrt{\alpha r}y)^2[\Gamma^2+(q+\sqrt{\alpha r}y)^2]^{-1}\tanh^2\{\beta[\Gamma^2+(q+\sqrt{\alpha r}y)^2]^{1/2}\} , \quad (19)$$

$$r=Q/(1-c)^2 . \quad (20)$$

Phase diagrams can be calculated from Eqs. (17)–(20), and the results reduce to those for the Hopfield neural-network model discussed in Ref. [3] when $\Gamma=0$.

We first investigate the phase transition of second order from the disordered, paramagnetic phase ($q=Q=0$) to the spin-glass phase ($q=0$, $Q\neq 0$). To derive an equation for the spin-glass transition temperature, we expand Eqs. (19) and (20) with $q=0$, in powers of Q and r. The transition temperature T_g is determined by

$$(1+\sqrt{\alpha})\Gamma^{-1}\tanh(\beta_g\Gamma)=1 . \quad (21)$$

In the classical case ($\Gamma=0$), Eq. (21) yields the known result $T_g=1+\sqrt{\alpha}$ [3].

The transition from the spin-glass phase to the retrieval phase can be studied by numerically solving (17)–(20), and is shown in Fig. 1, where the spin glass to paramagnetic phase transition is also shown. As in the second-order transition theory, critical storage capacity α is determined from the condition of a sudden disappearance of retrieval phase ($q \neq 0$) existing with the spin-glass phase ($Q \neq 0$, $q=0$) as α is changed with Γ and β kept fixed. We now turn to an explicit discussion of phase diagram (Fig. 1). Below the critical line T_g, the spin-glass solutions appear. When crossing the line T_M from above Mattis retrieval states show up as local minima of the free energy. At this point the overlap with the embedded patterns jumps from zero to a finite macroscopic value. So the system functions as an associative memory and the critical storage capacity for a given temperature can be read off through the line T_M. On the other hand, we see that the regions of the retrieval states decrease with the increase of the strength of transverse field Γ, and when $\Gamma=1$ the retrieval states disappear. In the present case, the main effect of quantum fluctuations is reduction of the critical storage capacity α beyond which no solution with $q \neq 0$ exists.

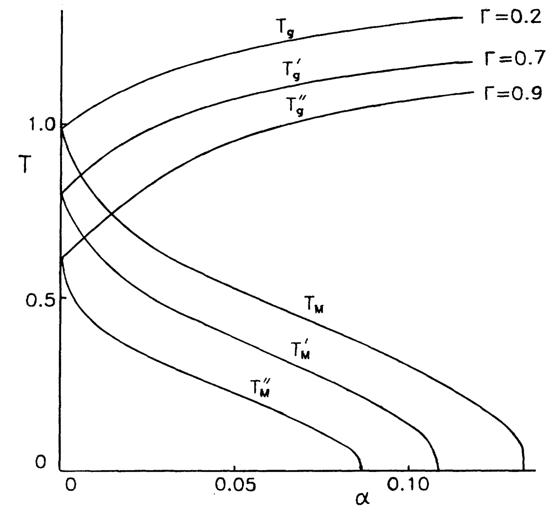

FIG. 1. T-α phase diagram for several values of transverse field Γ. The spin-glass states appear below the line T_g and the memory states below the line T_M.

So far, a Hopfield neural-network model with transverse fields that induces tunneling among the neurons is presented. The full phase diagram is obtained, and the possibilities of the existence of the retrieval, paramagnetic, and spin-glass phases are examined. Quantum fluctuations make retrieval states unstable, and even the retrieval states, which depend strongly on Γ, can be destroyed. It is obvious that Eqs. (17)–(20), which give the expression for the dependence of critical storage on transverse field for quantum Hopfield model, are brief and applicable.

*Address for correspondence.

[1] J. J. Hopfield, Proc. Nat. Acad. Sci. USA **79**, 2554 (1982); **81**, 3088 (1984).
[2] W. Kinzel, Z. Phys. B **60**, 205 (1985).
[3] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985); Phys. Rev. Lett. **55**, 1530 (1985); Ann. Phys. (N.Y.) **173**, 30 (1987).
[4] H. Sompolinsky and I. Kanter, Phys. Rev. Lett. **57**, 2861 (1986).
[5] H. Sompolinsky and A. Crisanti, Phys. Rev. Lett. **61**, 259 (1988).
[6] M. Y. Choi, Phys. Rev. Lett. **61**, 2809 (1988).
[7] For a review, see D. J. Amit, *Modelling Brain Function* (Cambridge University Press, New York, 1989).
[8] E. Barkai, D. Hansel, and I. Kanter, Phys. Rev. Lett. **65**, 2312 (1990).
[9] C. M. Marcus, F. R. Waugh, and R. M. Westervelt, Phys. Rev. A **41**, 3355 (1990).
[10] H. Eissfeller and M. Opper, Phys. Rev. Lett. **68**, 2094 (1992).
[11] B. K. Chakrabarti and P. K. Dasgupta, Physica A **186**, 32 (1992).
[12] D. Sherrington, M. Wong, and A. Rau, Philos. Mag. B **65**, 1303 (1992).
[13] A. Engel, H. M. Kohler, F. Tschepke, H. Vollmayr, and A. Zippelius, Phys. Rev. A **45**, 7590 (1992).
[14] D. Bolle, P. Dupont, and J. Huyghebaert, Phys. Rev. A **45**, 4194 (1992).
[15] E. Barkai, D. Hansel, and H. Sompolinsky, Phys. Rev. A **45**, 4146 (1992).
[16] Jean-Pierre Naef and A. Canning, J. Phys. I (France) **2**, 247 (1992).
[17] A. Engel, H. Englisch, and A. Schutte, Europhys. Lett. **8**, 393 (1989).
[18] A. Rau and D. Sherrington, Europhys. Lett. **11**, 499 (1990).
[19] A. Engel, M. Bouten, A. Komoda, and R. Serneels, Phys. Rev. A **42**, 4998 (1990).
[20] A. Rau, D. Sherrington, and K. Y. M. Wong, J. Phys. A **24**, 313 (1991).
[21] For a review, see P. Sen and B. K. Chakrabarti, Int. J. Mod. Phys. B **6**, 2439 (1992).
[22] D. C. Mattis, Phys. Lett. **56A**, 421 (1976).
[23] B. K. Chakrabarti, Phys. Rev. B **24**, 4062 (1981).
[24] J. L. van Hemmen, Phys. Rev. Lett. **49**, 409 (1982).
[25] J. P. Provost and G. Vallee, Phys. Rev. Lett. **50**, 598 (1983).