Dynamical phase transitions in the Little-Hopfield model

1994 J. Phys. A: Math. Gen. 27 4115

(http://iopscience.iop.org/0305-4470/27/12/018)

J. Phys. A: Math. Gen. 27 (1994) 4115–4124. Printed in the UK

Dynamical phase transitions in the Little–Hopfield model

P de Felice†‡, C Marangi†‡, G Nardulli†‡ and G Pasquariello§
† Dipartimento di Fisica, Universitá di Bari, Italy
‡ Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy
§ Istituto Elaborazione Segnali Immagini, CNR, Bari, Italy

Received 5 January 1994

Abstract. The time evolution of the distance between two random initial configurations subjected to the same thermal noise is used to study dynamical phase transitions in attractor neural networks trained by the Hebb rule. Numerical results are given for fully connected architectures, whereas, in the dilute case, both analytical and numerical outcomes are provided and a good agreement is shown to exist between the two sets of results.

It has been shown [1] that dynamical phase transitions can be observed in a large class of spin models, by measuring the time evolution of the normalized Hamming distance between two input patterns subjected to the same random noise. The same approach has been used [2] for a modification of a class of the threshold automata of the McCulloch and Pitts [3] type.

In spin models [1] two phases are observed in the long-time limit: at high temperature, independently of the initial conditions, the distance between the two states decreases or vanishes, according to the particular model (Sherrington–Kirkpatrick (SK) model, spin-glass models in arbitrary dimensions, Ising models); at low temperature the two configurations are confined, on average, at a non-zero distance which depends on the initial mutual positions. Between these two extrema there is an intermediate region, characterized by the fact that the two configurations have a finite non-zero distance in the long-time limit which, on average, is independent of the starting conditions.

In this paper we wish to extend these results to a particular class of spin models, i.e. neural networks; more precisely we apply the distance method to the Little–Hopfield model [4], where neurons are described by an Ising variable $S_i = \pm 1$, and the synaptic coupling between the neuron j and k is given by the Hebb rule [5], i.e. by the correlation between the P stored memories $\{\xi_i^\mu\}$:

$$J_{jk} = \frac{1}{N}\sum_{\mu=1}^{P} \xi_j^\mu \xi_k^\mu \tag{1}$$

for $j \neq k$ and $J_{ii} = 0$, where N is the number of neurons and $P/N = \alpha$ is the storage capacity. Moreover, we assume parallel dynamics for the network. Serial dynamics will be briefly discussed below.

At temperature T the state of the network at time $t+1$ (which we denote $S_i^a(t+\Delta t)$ with $\Delta t = 1$) is given in terms of network configuration at time t as follows:

$$S_i^a(t+\Delta t) = \operatorname{sign}\left[\frac{1}{2} + \frac{1}{2}\tanh\left(\sum_j \frac{J_{[ij}S_j^a(t)}{T}\right) - \eta_i(t)\right] \tag{2}$$

0305-4470/94/124115+10$19.50

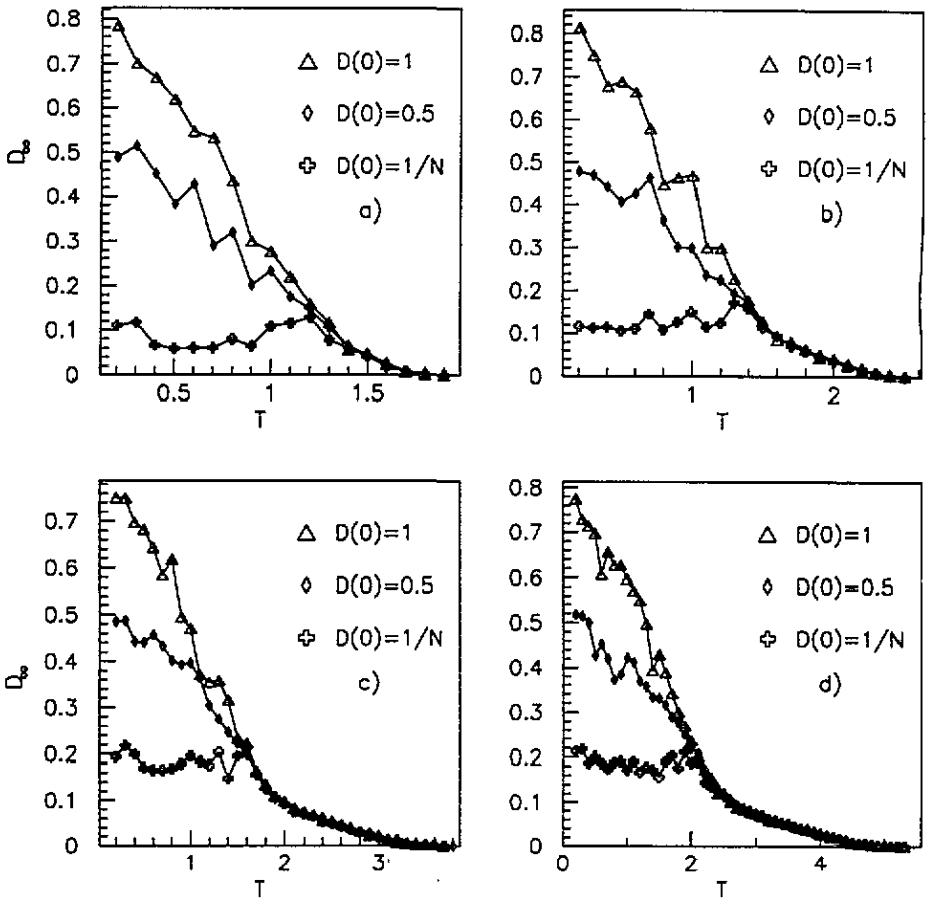

Figure 1. Hamming distance D_∞ versus temperature T in the fully connected architecture for four different values of α: $\alpha = 0.125$ (a); $\alpha = 0.25$ (b); $\alpha = 0.5$ (c); $\alpha = 1.0$ (d). The three sets of points correspond to initial distance $D(0) = 1$ (triangles, uppermost curve), $D(0) = 0.5$ (diamonds, middle curve), $D(0) = 1/N$ (crosses, lowest curve) ($N = 256$).

where η_i is a random number in the range [0, 1].

We also consider the time evolution of a different configuration S^b:

$$S_i^b(t + \Delta t) = \text{sign}\left[\frac{1}{2} + \frac{1}{2}\tanh\left(\sum_j \frac{J_{ij}S_j^b(t)}{T}\right) - \eta_i(t)\right] \tag{3}$$

subjected to the same noise η_i and we introduce the normalized Hamming distance

$$D(t) = \frac{1}{2N}\sum_i \left|S_i^a(t) - S_i^b(t)\right| . \tag{4}$$

We shall call D_∞ the long-time ($t \to \infty$) limit of $D(t)$. We wish to study the dependence of the order parameter D_∞ on T and α.

To begin with, we present numerical results of some simulations for fully connected networks. Even though, as stressed already, in all simulations we have used the parallel dynamics, where all the sites are simultaneously updated at each iteration, we have verified that comparable results are obtained adopting a serial heat-bath dynamics, where, at each

iteration, only one site i, randomly selected among the N, is updated according to (2) and (3), with $\Delta t = 1/N$. The simulations have been performed for three system sizes ($N = 256, 512$ and 1000) with no significant difference in the behaviour of D_∞ and for a wide range of the storage capacity α. In figure 1 we show the dependence of D_∞ on the temperature T for $N = 256$ and for four values of α ($\alpha = 0.125,\ 0.25,\ 0.5$ and 1). Each experimental point corresponds to the average on 100 different starting configurations for the initial states $\{S_i^a\}$ and $\{S_i^b\}$; for each case we take $t = 500$ Monte Carlo steps for the parallel dynamics and $t = 500 \times N$ for the serial case. For each value of α, three plots are shown, corresponding to different initial Hamming distances between the two configurations. On the upper curve we report the final distances when the initial configurations are opposite, (for each site i, $S_i^a = -S_i^b$), corresponding to a starting distance $D(t = 0) = 1$. The middle curve shows D_∞ when one starts with two uncorrelated states ($D(0) = 0.5$); finally the lowest curve is obtained using two initial configurations differing by a single neuron ($D(0) = 1/N$). Each initial configuration is uncorrelated with the memories $\{\xi_i^\mu\}$.

From a qualitative point of view, the results of figure 1 are similar to the results reported in [1] for the SK model, having $\langle J_{ij} \rangle = 0$ and $\langle J_{ij}^2 \rangle = 1$, where a dynamical phase transition also occurs at finite temperature. In fact in all the simulations it is possible to find a temperature T_l above which D_∞ has a finite non-zero value that is independent of the starting distance $D(0)$. The use of neural networks allows us, however, to study the dependence on the storage capacity α and provides insights into the nature of the phase transitions.

Our results are that T_l increases with α and, more generally, for each fixed $T \geqslant T_l$, D_∞ also increases with α. We cannot give a comprehensive explanation of this behaviour for the fully connected network, because, as is well known, no analytical result for such a network is available owing to the presence of correlation among the sites after the first time step (only for the dilute case can analytical results be given, as we shall discuss below). In order to analyse the dependence of T_l on α we report in figure 2 computer data for T_l obtained with several values of α and we compare them to the curve $T_g = 1 + \sqrt{\alpha}$ which gives the transition temperature between the spin glass and the paramagnetic phase. The computer data are obtained as follows; putting D_∞ equal to D_1, D_2, D_3 for $D(t = 0) = 1, \frac{1}{2}, 1/N$ respectively, we define the parameter

$$\eta = \frac{1}{3} \sum_{i,j=1}^{3} \left| \frac{D_i - D_j}{D_i + D_j} \right| . \tag{5}$$

T_l is operationally defined as the smallest temperature for which $\eta \leqslant 0.1$. The agreement between the data and the curve $T_g = 1 + \sqrt{\alpha}$ in figure 2 is rather good and suggests the identification between T_l and T_g; it can be noted that also in the Derrida's analysis of the SK model the dynamical phase-transition temperature T_l (which is ≈ 1) seems numerically equal to the spin-glass paramagnetic static phase transition $T_c = 1$. Let us finally note that our conclusion is valid regardless of the details in the definition of T_l, i.e. we may identify T_l and T_g for a rather wide class of definitions of T_l.

Although no analytical proof is available to support this identification, we can understand it as follows: passing from the spin glass to the paramagnetic phase the network loses any capacity to work as a retrieval system since both the global (magnetization) and the local (Edwards–Anderson) parameters vanish; in these conditions one expects that the dependence on the initial conditions disappears, as happens for $T > T_l$.

It is known [1] that some models show the existence of a second phase transition characterized by a critical temperature T_u such that, for $T > T_u$, $D_\infty = 0$. We have investigated this possibility for the Little–Hopfield model and we have seen evidence for the second phase transition only for small values of N, whereas the effect tends to disappear with

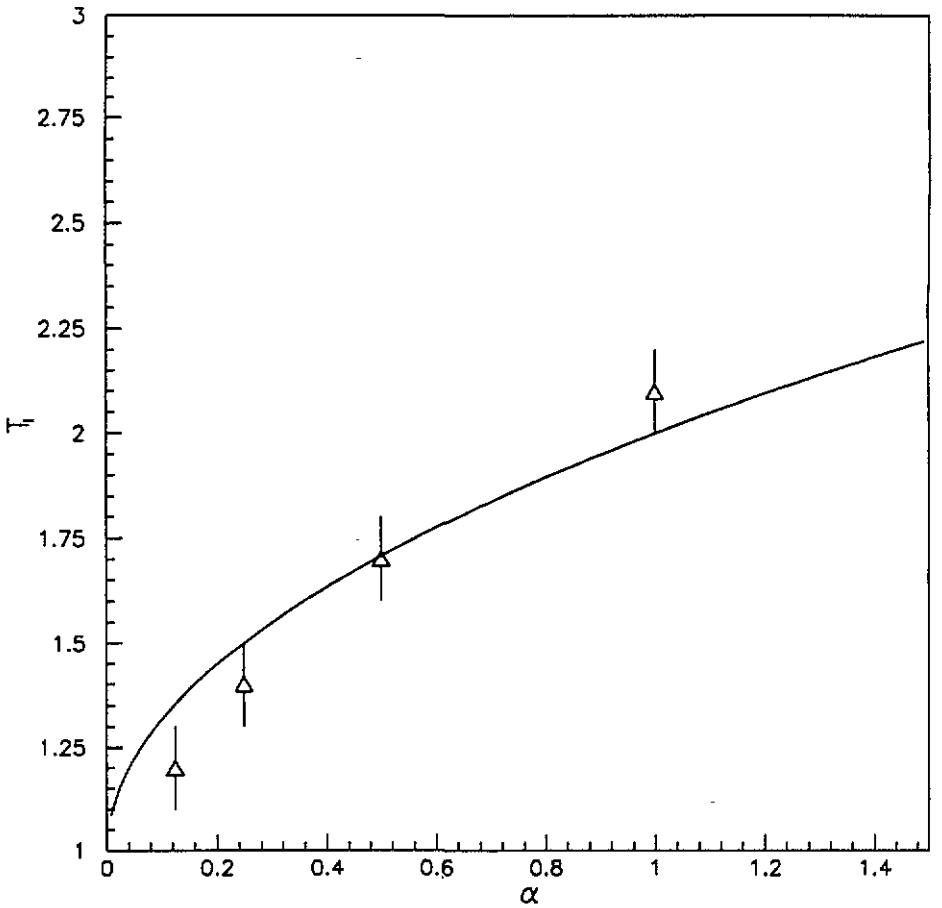

Figure 2. Dynamical transition temperature T_l versus α for the fully connected case. Simulation data for $N = 256$ are compared with the equilibrium transition temperature (full curve) $T_g = 1 + \sqrt{\alpha}$.

increasing N. In figure 3 we report D_∞ as a function of T with $\alpha = 1$ for two values of N ($N = 256$ in a and $N = 1000$ in b). We observe that, while for small N the data show a second phase transition T_u, this effect disappears at $N = 1000$. From this numerical evidence we are led to conclude that D_∞ does not vanish for any finite temperature in the thermodynamic limit; this conclusion agrees with the SK analysis [1] and is also supported by the analytical calculation performed in the dilute case (see below).

After the analysis of the fully connected network, let us now consider the dilute case, characterized by synaptic couplings given by

$$J_{jk} = \frac{1}{C} \sum_{\mu=1}^{P} \xi_j^\mu \xi_k^\mu \tag{6}$$

with probability C/N and

$$J_{jk} = 0 \tag{7}$$

with probability $1 - C/N$, with the constraint $C < \ln N$ (again we assume $J_{ii} = 0$). The dilution hypothesis is needed at zero temperature to ensure that the neurons remain uncorrelated

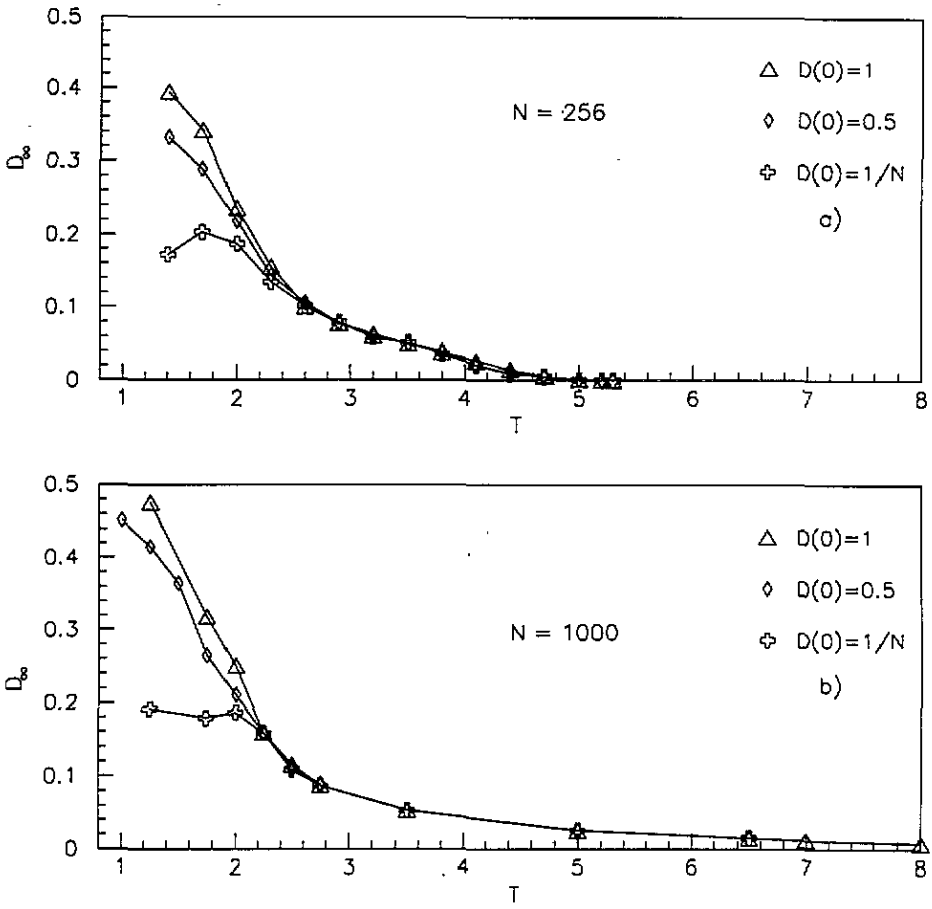

Figure 3. Hamming distance D_∞ versus temperature T in the fully connected case, for two different network sizes: (*a*) $N = 256$ and (*b*) $N = 1000$.

at any time t [6]. This will allow us to obtain a recursive relation for $D(t)$. Since we are mainly interested in the high-temperature region, where the dynamical phase transitions take place, we expect a less stringent requirement; in particular we find a good agreement between analytical and numerical results by imposing the weaker condition $C \ll N$.

In order to obtain a recursive relation for $D(t)$, we first compute the conditional probability for the Hamming distance: $D(t+1)$ having the value D', given the network configurations at time t : $S_i^a(t)$ and $S_i^b(t)$:

$$P(D(t+1) = D' | \{S^a(t)\}, \{S^b(t)\}) = \mathrm{Tr}_{S^a S^b} \prod_{i=1}^{N} \int_0^1 \mathrm{d}\eta_i \, \delta\big(S_i^{ta} - \mathrm{sign}[p_i^a - \eta_i]\big)$$
$$\times \delta\big(S_i^{tb} - \mathrm{sign}[p_i^b - \eta_i]\big) \, \delta\left(D' - \frac{1}{2N} \sum_i \left| S_i^{ta} - S_i^{tb} \right| \right) \tag{8}$$

where $p_i^a = \frac{1}{2} + \frac{1}{2} \tanh\left(\sum_j J_{ij} S_j^a / T\right)$. The result is

$$P\left(D(t+1) = D' | \{S^a(t)\}, \{S^b(t)\}\right) = \delta\left(D' - \frac{1}{N} \sum_i \left| p_i^a - p_i^b \right| \right). \tag{9}$$

Next we compute the conditional probability for $D(t+1)$ having the value D', given the value D at time t:

$$P\left(D(t+1)=D'|D(t)=D\right)=\frac{1}{\mathcal{N}}\,\mathrm{Tr}_{S^aS^b}\,\delta\left(D-\frac{1}{2N}\sum_i\left|S_i^a-S_i^b\right|\right)$$
$$\times\delta\left(D'-\frac{1}{N}\sum_i\left|p_i^a-p_i^b\right|\right) \tag{10}$$

where

$$\mathcal{N}=\mathrm{Tr}_{S^aS^b}\,\delta\left(D-\frac{1}{2N}\sum_i\left|S_i^a-S_i^b\right|\right). \tag{11}$$

In order to compute (10) we employ the method used in [7] (see also [8]). More details on the calculation are reported in the appendix. We get, in the $N\to\infty$ limit,

$$P(D'|D)=\int\frac{\mathrm{d}w}{2\pi}\mathrm{e}^{\mathrm{i}w[D'-f(D,\lambda)]}=\delta(D'-f(D,\lambda)) \tag{12}$$

where

$$\lambda=\frac{\alpha}{T^2}. \tag{13}$$

In other words we get

$$D(t+1)=f(D(t),\lambda) \tag{14}$$

$$f(D(t),\lambda)=\frac{1}{2\lambda\pi\sqrt{D(D-1)}}\int_0^{+\infty}\mathrm{d}z\,\mathrm{e}^{-A(1-D)z^2}\int_{-\infty}^{+\infty}\mathrm{d}u\,\mathrm{e}^{-ADu^2}\tanh z\frac{1-\tanh^2u}{1-\tanh^2z\tanh^2u} \tag{15}$$

where A is a function of λ and D:

$$A=\frac{1}{2\lambda D(1-D)}. \tag{16}$$

The non-trivial fixed point relative to (14) is obtained by solving the equation

$$D_\infty=f(D_\infty,\lambda) \tag{17}$$

with $D_\infty=D(t=\infty)\neq 0$. For small λ the non-trivial fixed point is given by

$$D_\infty\simeq\frac{2\lambda}{\pi}. \tag{18}$$

Let us stress that the solution of (17) depends on α and T only *via* the parameter $\lambda=\alpha/T^2$. We report the numerical solutions of (14) in figure 4 (full curve) and we compare it with computer data. We see that the agreement with theoretical expectation is, in general, satisfactory; in particular, the data also depend on α and T *via* the parameter λ. This figure also displays the independence of the results from the number of active connections C, provided that $C\ll N$; we report both $C=100$ and $C=120$ cases, with $N=1000$. For small λ the dependence of D_∞ is approximately linear: $D_\infty\simeq 2\lambda/\pi=2\alpha/(\pi T^2)$, in agreement with (18). Notice that this behaviour implies that D_∞ does not vanish for any $T\neq\infty$. The small discrepancies, for very small values of λ ($\lambda\leqslant 0.03$), between analytical results and computer data should be attributed to finite-size effects, since we have checked that in general the agreement improves with increasing N.

We cannot compare the solution of (17) with data for large values of λ, even though (17) has non-trivial fixed points also for $\lambda>1$; the reason is that only for

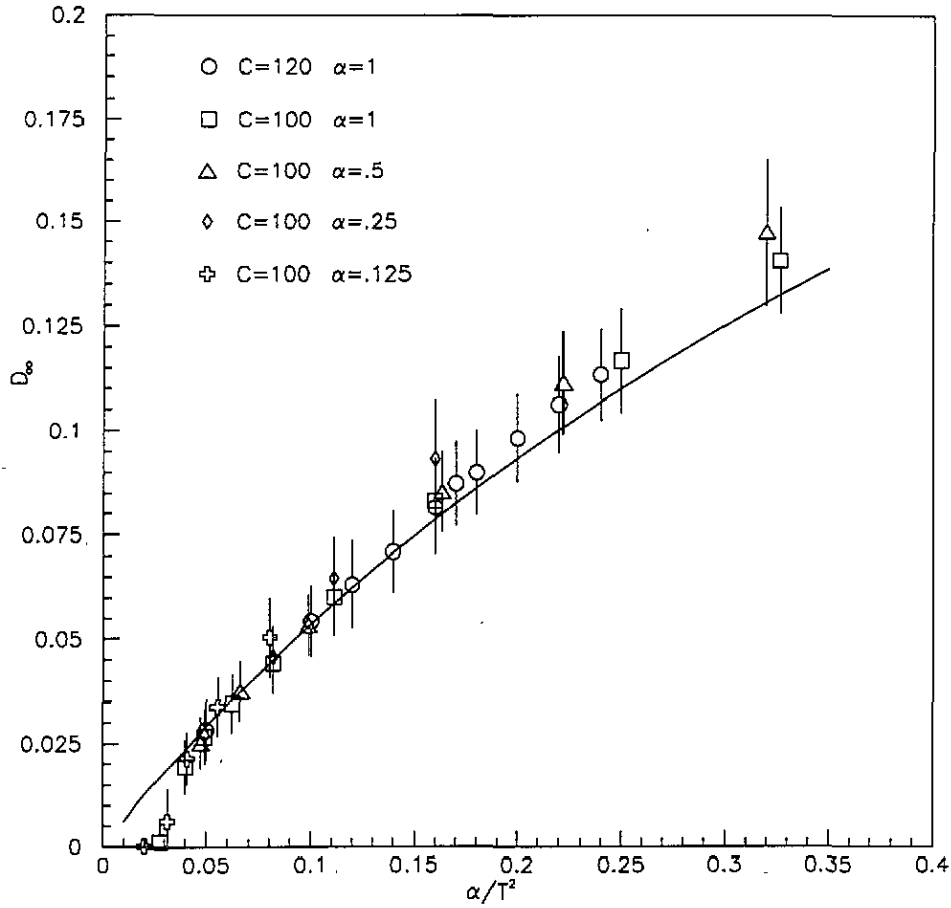

Figure 4. Hamming distance D_∞ versus the parameter $\lambda = \alpha/T^2$ for the dilute case. Analytical results (full curve) follow from (17); computer data are for different values of α in the range (0.125, 1.0) and two values of the dilution factor (corresponding to $C = 100, 120$ and $N = 1000$).

sufficiently large temperatures (i.e. not large values of λ) can we relax the strong-dilution condition: $C < \ln N$ and simply assume $C \ll N$; as a matter of fact, as pointed out already, the thermal noise has the effect of diluting the correlation among the sites induced, at any time step, by the dynamical equation and therefore to allow larger values of C; for small temperatures this does not happen and therefore we should satisfy a more stringent condition $1 \ll C < \ln N$, which is beyond our computer's present capabilities.

In figure 5 we report D_∞ versus temperature, for four values of α (here $\alpha = P/C$): $\alpha = 0.125, 0.25, 0.5$ and 1, for a dilution factor $C/N = \frac{100}{1000}$. This figure should be compared to the analogous figure 1 obtained for the fully connected case: the qualitative behaviour is similar but there are some remarkable differences. For example, we have checked that the simple dependence of D_∞ on T and α *via* the parameter $\lambda = \alpha/T^2$, which holds in this case, is not valid for the fully connected architecture; a second difference which can be noted is that the temperatures T_l (when the curve coalesces) in figure 5 are significantly

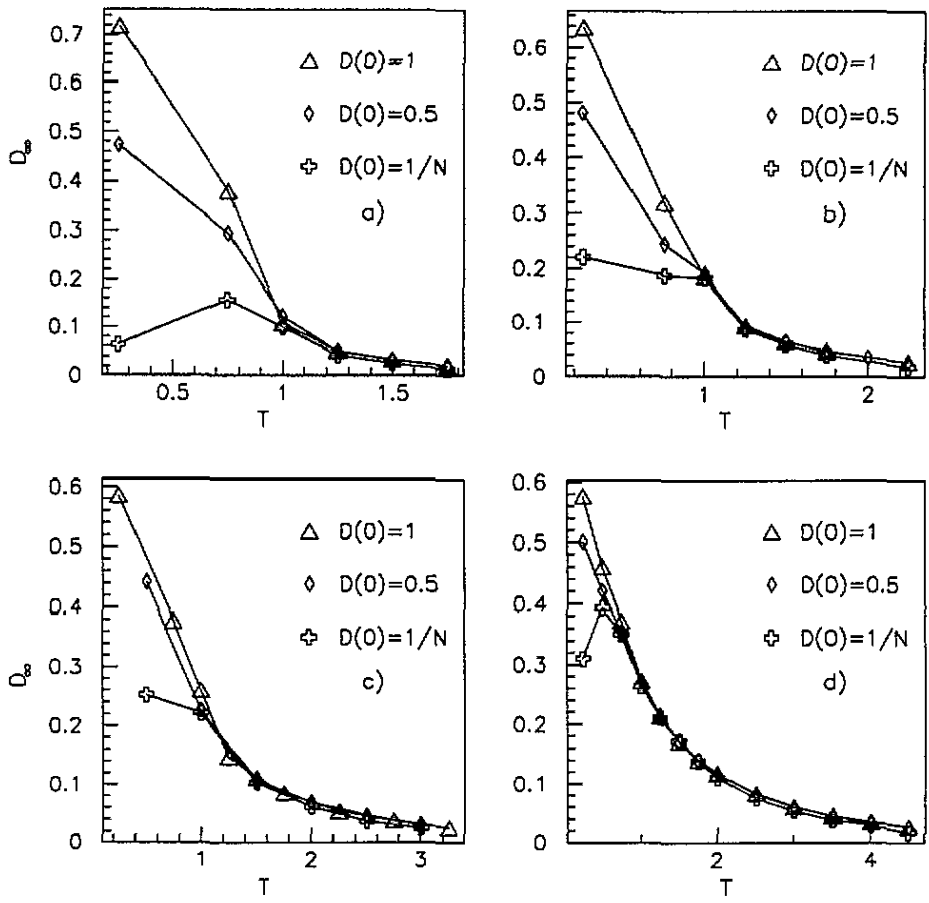

Figure 5. Hamming distance D_∞ versus temperature T in the dilute case, for four different values of α: (*a*) $\alpha = 0.125$; (*b*) $\alpha = 0.25$; (*c*) $\alpha = 0.5$; (*d*) $\alpha = 1.0$. The three sets of points correspond to initial distance $D(0) = 1$ (triangles, uppermost curve), $D(0) = 0.5$ (diamonds, middle curve), $D(0) = 1/N$ (crosses, lowest curve) ($N = 1000$ and $C = 100$).

smaller in the present case than the ones obtained from the data of fully connected networks.

In conclusion we have studied by the distance method the dynamical phase transitions of the Little–Hopfield model in the high-temperature regime. For the fully connected case our results on the T-dependence are qualitatively similar to previous analyses of spin-glass models; in the present analysis, however, the presence of the storage-capacity parameter has also allowed us to study the α-dependence. In the dilute case we have shown, both analytically and numerically, that the long-time limit of the distance depends only on the ratio $\lambda = \alpha/T^2$, with a functional dependence that for large temperatures and small α is given by $D_\infty \simeq 2\alpha/(\pi T^2)$.

Acknowledgment

We thank A Maritan for most useful discussions.

Appendix

In order to solve equation (10) auxiliary integration variables h_l^a, h_l^b $(l = 1 \ldots N)$ have been introduced

$$P(D'|D) = \prod_{l=1}^{N} \int \mathrm{d}h_l^a \, \mathrm{d}h_l^b \, \frac{N_l}{W} \, \delta\left(D' - \frac{1}{2N} \sum_j \left|\tanh \beta h_j^a - \tanh \beta h_j^b\right|\right) \tag{A1}$$

where N_l and W are given by the following expressions:

$$\begin{aligned} N_l &= \mathrm{Tr}_{S^a S^b} \, \delta\left(h_l^a - \sum_k J_{lk} S_k^a\right) \delta\left(h_l^b - \sum_k J_{lk} S_k^b\right) \delta\left(D - \frac{1}{2} + \frac{1}{2N} \sum_j S_j^a S_j^b\right) \\ &= \frac{N}{(2\pi)^3} \int \mathrm{d}x \, \mathrm{d}y_l^a \, \mathrm{d}y_l^b \exp\left[\mathrm{i}\left(Nx(2D-1)/2 + y_l^a h_l^a + y_l^b h_l^b\right)\right] G_l \end{aligned} \tag{A2}$$

$$\begin{aligned} W &= \mathrm{Tr}_{S^a S^b} \, \delta\left(D - \frac{1}{2} + \frac{1}{2N} \sum_j S_j^a S_j^b\right) \\ &= N \int \frac{\mathrm{d}x}{2\pi} \, \mathrm{e}^{\mathrm{i}Nx(2D-1)/2} \, \mathrm{Tr}_{S^a S^b} \, \mathrm{e}^{\mathrm{i}x/2 \sum_j S_j^a S_j^b} \\ &= 4^N N \int \frac{\mathrm{d}x}{2\pi} \, \mathrm{e}^{\mathrm{i}Nx(2D-1)/2 + N \ln \cos(x/2)} \end{aligned} \tag{A3}$$

and the last integral can easily be computed in the $N \to \infty$ limit by the steepest-descent method. The factor G_l in (A2) is given by

$$\begin{aligned} G_l &= \mathrm{Tr}_{S^a S^b} \, \exp\left[\mathrm{i} \sum_j \left(\tfrac{1}{2} x S_j^a S_j^b - y_l^a J_{lj} S_j^a - y_l^b J_{lj} S_j^b\right)\right] \\ &= 4^N \exp\left[\sum_j \ln\left(\cos(x/2) \cos y_l^a J_{lj} \cos y_l^b J_{lj} - \mathrm{i} \sin(x/2) \sin y_l^a J_{lj} \sin y_l^b J_{lj}\right)\right]. \end{aligned} \tag{A4}$$

Since $\langle J_{lj} \rangle = 0$ and $\langle J_{lj}^2 \rangle = \alpha/N$ one easily gets

$$G_l = 4^N \exp\left\{N \ln \cos(x/2) - \tfrac{1}{2}\alpha[y_l^{a2} + y_l^{b2} + 2\mathrm{i} y_l^a y_l^b \tan(x/2)]\right\}. \tag{A5}$$

Putting (A5) into (A2), and performing the x integration by the steepest-descent method, we get

$$\begin{aligned} \frac{N_l}{W} &= \frac{1}{(2\pi)^2} \int \mathrm{d}y_l^a \, \mathrm{d}y_l^b \exp\left\{-\frac{\alpha}{2}[y_l^{a2} + y_l^{b2}] + (2D-1)\alpha y_l^a y_l^b + \mathrm{i}(y_l^a h_l^a + y_l^b h_l^b)\right\} \\ &= \frac{1}{4\pi\alpha} \frac{1}{\sqrt{D(1-D)}} \mathrm{e}^{-h_l^{a2} + h_l^{b2} + 2h_l^a h_l^b (2D-1)/8\alpha D(1-D)}. \end{aligned} \tag{A6}$$

From this result and (A1), by changing variables as follows:

$$h_l^a = T(z_l + u_l) \qquad h_l^b = T(u_l - z_l) \tag{A7}$$

we get

$$\begin{aligned} P(D'|D) = \left(\frac{1}{4\pi\alpha\sqrt{D(1-D)}}\right)^N \int_{-\infty}^{+\infty} \frac{\mathrm{d}w}{2\pi} \mathrm{e}^{\mathrm{i}wD'} \prod_l (2T^2) \int_{-\infty}^{+\infty} \mathrm{d}z_l \int_{-\infty}^{+\infty} \mathrm{d}u_l \\ \times \exp\left\{-A[(1-D)z_l^2 + D u_l^2] - \mathrm{i}\frac{w}{2N}\left|\tanh(z_l + u_l) + \tanh(z_l - u_l)\right|\right\} \end{aligned} \tag{A8}$$

with $A = T^2/2\alpha D(1-D)$; in the $N \to \infty$ limit one gets the results written in (12)–(16).

References

[1] Derrida B 1989 *Phys. Rep.* **184** 207
[2] Kurten K E 1989 *J. Physique* **50** 2313
[3] McCulloch W S and Pitts W 1943 *Bull. Math. Biophys.* **5** 115
[4] Little W A *Math. Biosci.* **19** 74 101
Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
[5] Hebb D O 1949 *The Organization of Behaviour* (New York: Wiley)
[6] Amit D J 1989 *Modelling Brain Function* (Cambridge: Cambridge University Press)
[7] Kepler T B and Abbott L F 1988 *J. Physique* **49** 1657
[8] Nardulli G and Pasquariello G 1991 *J. Phys. A: Math. Gen.* **24** 1103