Capacity of networks with correlated attractors

J. Phys. A: Math. Gen. 27 (1994) 741–756. Printed in the UK

Capacity of networks with correlated attractors

L F Cugliandolo† and M V Tsodyks‡

† Dipartimento di Fisica, Università di Roma, La Sapienza, INFN Sezione di Roma I, Roma, Italy
‡ Racah Institute of Physics and Center for Neural Computation, Hebrew University, Jerusalem, Israel

Received 27 April 1993, in final form 7 October 1993

Abstract. We analyse the extensive loading of the neural network model proposed to describe neurophysiological experiments in which correlated attractors associated to uncorrelated patterns are found. The phase diagram is obtained and discussed. Some generalizations of the original model are also considered. In all the cases we demonstrate the existence of a region in the phase diagram with correlated attractors. Results from numerical simulations which confirm the mean-field theory results are also presented.

1. Introduction

Griniasty *et al* [1] have recently proposed an attractor neural network model which describes the findings of the experiments done by Miyashita and Chang [2], conversion of temporal correlations between stimuli to spatial correlations between attractors. In these experiments a monkey was trained to recognize and match a set of visual pictures. On the one hand, a selective increase in neural activity which lasted as long as 16 seconds after the removal of the picture was found. This fact was interpreted in [1] as a manifestation of attractor dynamics. On the other hand, these persistent activities present a striking feature which cannot be described in the framework of the standard Hopfield model of associative memory [3]: attractors corresponding to different stimuli are spatially correlated. More specifically, correlations between attractors associated to temporally close stimuli in the training session were observed. These correlations do not reflect the geometrical properties of the stimuli but are a consequence of learning.

In [1] a simple modification of the Hopfield model was proposed which can capture these basic experimental features. For the two-state neuron network described by the variables $s_i(t) = \pm 1$, the following synaptic matrix was proposed:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} [\, \xi_i^\mu \xi_j^\mu + a\,(\xi_i^{\mu+1} \xi_j^\mu + \xi_i^{\mu-1} \xi_j^\mu)\,] \tag{1.1}$$

which is supplemented by the usual schematized spike emission dynamics (see e.g. [4])

$$s_i(t+\delta t) = \mathrm{sgn}[h_i(t)] \qquad h_i(t) = \sum_{i \neq j=1}^{N} J_{ij} s_j(t)\,. \tag{1.2}$$

Here, the p uncorrelated patterns are N-vectors with components given by $\xi_i^\mu = \pm 1$ with probabilities $P(\xi_i^\mu = \pm 1) = \frac{1}{2}$. The index μ labels the stored patterns and signals the order in the sequence that corresponds to the temporal order of presentation in the training phase. The parameter a reflects the strength of association between consecutive patterns.

As was shown in [1, 5], in the low loading level ($p/N \to 0$ as $N \to \infty$) and for $\frac{1}{2} < a < 1$, the matrix (1.1) produces a set of correlated attractors. The mutual correlations are a decreasing function of the distance among the corresponding stimuli in the sequence of presentation.

The analysis of the papers [1, 5] has two obvious limitations: only the network of binary neurons was considered, as well as the limit of finite loading. The behaviour of a biologically more realistic network consisting of analog elements, representing neuronal spiking rates, is the topic of a parallel study [6]. In this paper we extend the analysis of the model to the case of extensive loading, where the number of stored patterns is proportional to the number of neurons. Our aim is to demonstrate that the correlated attractors found in the finite loading regime are not destroyed by the extensive loading, which is not obvious *a priori*, and furthermore that the overlaps and correlations are not drastically modified. The phase diagram of the network in the variables a and $\alpha \equiv p/N$ and, in particular, the critical storage capacity for the specially interesting regime with correlated attractors are found. Some generalizations of the matrix (1.1) for different structures of association among the patterns are also discussed.

The paper is organized as follows. In section 2 the analysis of the finite loading is recapitulated. In section 3 the mean-field theory analysis and the phase diagram for the extensive loading of the various proposed models are discussed. In section 4 numerical simulation results are presented. Finally, a section with conclusions is included.

2. Finite loading

2.1. Mean-field equations

We consider here a generalization of the dynamics given by (1.2) to incorporate the stochastic nature of neural activity. This is done by inserting the network in a thermal bath of temperature $T = 1/\beta$. The new dynamics is given by (see e.g. [4])

$$s_i(t+\delta t) = \begin{cases} +1 & \text{with probability } [1+\exp(-2\beta h_i(t))]^{-1} \\ -1 & \text{with probability } [1+\exp(2\beta h_i(t))]^{-1}\,. \end{cases} \tag{2.1}$$

The natural variables describing the similarity between the state of the network and the stored patterns are the overlaps $m_\mu(t)$ defined as

$$m_\mu(t) = \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu \langle s_i(t)\rangle \tag{2.2}$$

where $\langle\ldots\rangle$ denotes the thermal average. The mean-field equations in the finite loading regime in the limit of large N (i.e. $p/N \to 0$) at finite temperature are

$$m^\mu = \left\langle\!\!\left\langle \xi^\mu \tanh\beta \sum_{\nu=1}^{p} m^\nu(\xi^\nu + a(\xi^{\nu+1} + \xi^{\nu-1}))\right\rangle\!\!\right\rangle_\xi . \tag{2.3}$$

$\langle\langle\cdots\rangle\rangle_\xi$ represents the mean over the probability distribution of ξ^μ variables. The dynamics of the network, when stimulated by a pure pattern, is described by the iteration of these equations starting from a state identical to the stimulating pattern. This initial state corresponds to an overlap vector $m_\mu^\alpha(t=0) = (m_\mathrm{I})_\mu^\alpha = \delta_\mu^\alpha$. The RHS of equation α, evaluated in this configuration reads

$$\mathrm{RHS}^\alpha = \frac{1}{2}\left[\frac{1-\tanh^2 2\beta a}{1-\tanh^2\beta\tanh^2 2\beta a}+1\right]\tanh\beta \tag{2.4}$$

while the RHS of equations $\alpha+1$ and $\alpha-1$ read

$$\mathrm{RHS}^{\alpha+1} = \mathrm{RHS}^{\alpha-1} = \frac{1}{2}\left[\frac{1-\tanh^2\beta}{1-\tanh^2\beta\tanh^2 2\beta a}\right]\tanh 2\beta a\,. \tag{2.5}$$

In order to have a pure retrieval state, i.e. a configuration such that $m^\mu = m\,\delta_\mu^\alpha$, as a fixed point, $\mathrm{RHS}^{\alpha+1}$ and $\mathrm{RHS}^{\alpha-1}$ should be equal to zero while the iteration of the α equation should reach a fixed point. As can be seen from (2.5), the pure retrieval state does not exist if both a and T are non-zero. When $a=0$ the Hopfield network is recovered.

2.2. *Zero temperature*

The zero temperature limit of system (2.3) has been studied in [1,5]. Auto-associative retrieval states exist as fixed points if the parameter a belongs to the interval $[0,\frac{1}{2})$, i.e. the uncorrelated patterns are exact attractors. If $a\in(\frac{1}{2},1)$, the system evolves to an attractor which has an overlap different from zero with exactly nine patterns, independently of the number of stored patterns. The overlap with the pattern used as stimulus is the 'highest' and the overlaps with the neighbouring patterns in the stored sequence decay symmetrically until vanishing at a distance of 5. The whole set of attractors can be obtained by cyclic rotations of

$$m^\mu = \frac{1}{2^7}(0,\ldots,0,1,3,13,51,77,51,13,3,1,0,\ldots,0)\,. \tag{2.6}$$

These attractors are mutually correlated. Correlation between attractors is defined as

$$C(\alpha,\beta) = \frac{\sum_{i=1}^N(\sigma_i^\alpha-\overline{\sigma}^\alpha)(\sigma_i^\beta-\overline{\sigma}^\beta)}{\sqrt{\sum_{i=1}^N(\sigma_i^\alpha-\overline{\sigma}^\alpha)^2\sum_{i=1}^N(\sigma_i^\beta-\overline{\sigma}^\beta)^2}} \tag{2.7}$$

where σ_i^α is the fixed point of the spike emission (1.2) $\sigma_i^\alpha = s_i$, i.e. the activity of neuron i when the network is in attractor α. $\overline{\sigma}^\alpha$ is the average activity in attractor α which is zero in this limit. Thus, replacing (1.2) in (2.7) and using the definition of the overlaps m_μ (2.2), at zero temperature the correlations read

$$C(\alpha,\beta) = \left\langle\!\!\left\langle \mathrm{sign}\left[\sum_\mu^p m_\mu^\alpha(\xi^\mu + a(\xi^{\mu+1}+\xi^{\mu-1}))\right]\mathrm{sign}\left[\sum_\nu^p m_\nu^\beta(\xi^\nu + a(\xi^{\nu+1}+\xi^{\nu-1}))\right]\right\rangle\!\!\right\rangle.$$

Due to the structure of the attractors, (2.6), their correlation $C(\alpha,\beta)$ only depends on the separation of the corresponding stimulating patterns in the memorized sequence, $d=|\alpha-\beta|$. For $\alpha>\beta$, $C(\alpha,\beta)=C(\alpha-\beta)=C_d$ and, furthermore, $C_d=C_{p-d}$, because of the cyclic property. The computation of the correlations is straightforward and up to a distance $d=5$, they are

$$C_0 = 1 \quad C_1 = 0.66 \quad C_2 = 0.33 \quad C_3 = 0.12 \quad C_4 = 0.04 \quad C_5 = 0.01 \tag{2.8}$$

while more distant attractors are not significantly correlated. If $p\geqslant 22$, attractors separated at distances $10<d<p-10$ are not correlated at all, $C_d=0$.

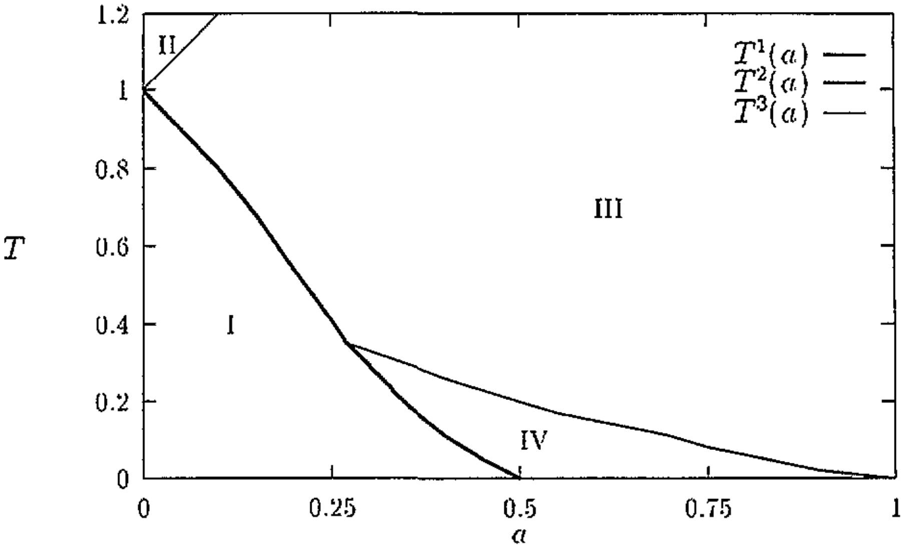

Figure 1. Phase diagram for a network loaded with $p = 13$ patterns in a thermal bath. I Mattis-like states; II Paramagnetic phase; III Symmetric states; IV Correlated attractors.

2.3. *Finite temperature*

If the temperature T is different from zero, the fixed points of the mean-field equations can be found by iterating the system (2.3). It is then easily found that the qualitative behaviour of the network when in a thermal bath does not change. The phase diagram (a, T) is plotted in figure 1.

If $a < 0.5$ and the temperature is small (sector I), the network behaves as a Hopfield network in a thermal bath. Although, if a is not strictly zero, attractors with just one non-zero overlap do not exist, for small temperatures the fixed points have only one overlap close to one and all others negligible. These solutions are modified by temperature. When it is increased, the main overlap decreases while others increase. The critical curve $T^1(a)$ determines the transition temperature at which these attractors disappear.

If $a = 0$ and the temperature is $T > T^1(0) = 1$ the system evolves to a state completely uncorrelated with all attractors, $m^\mu = 0, \forall\mu$. These are the fixed points in sector II, i.e. the paramagnetic phase.

For small non-zero a and $T > T^1(a)$ (sector III) the system evolves to an approximately symmetric state. These states become more symmetric increasing the temperature until T reaches $T^2(a)$ and the transition to the paramagnetic phase takes place.

For $a > a_{\mathrm{cr}}$ correlated attractors appear. The critical a at non-zero temperature is smaller than 0.5 and it depends on p ($a_{\mathrm{cr}} = 0.27$ for $p = 13$). Correlated attractors exist in sector IV. The typical values of the overlaps are similar to those given by (2.6). Although for general T they are all non-zero, they rapidly decay with the distance from the initial pattern (in particular, for $a \in (\frac{1}{2}, 1)$ the transition temperature is just 0 and at this temperature the overlaps (2.6) are exact).

At a further increment of the temperature the system evolves to a symmetric state (sector III).

3. Extensive loading

In this section we extend the analysis of the model to the case of extensively many stored patterns. To this end, we obtain the mean-field equations for a general set of 'quadratic' models which include, as special cases, the original model and some generalizations. We then present the models and analyse their phase diagrams.

3.1. Mean-field equations

Following [7], we start by computing the averaged free-energy per spin

$$f = -\frac{1}{\beta}\lim_{N\to\infty}\frac{1}{N}\left\langle\!\left\langle \log \mathrm{Tr}_s \exp\left[-\beta\sum_{i\neq j}^{N} J_{ij}s_is_j\right]\right\rangle\!\right\rangle_\xi \tag{3.1}$$

where $\langle\langle\cdots\rangle\rangle_\xi$ is the quenched average over ξ variables.

The whole set of interesting models can be described with a general 'quadratic' synaptic matrix

$$J_{ij} = \frac{1}{2N}\sum_{\mu\eta}^{p}\xi_i^\mu X_{\mu\eta}\xi_j^\eta . \tag{3.2}$$

The indices $\mu, \eta = 1, \ldots, p$, label patterns and X is a $p \times p$ matrix. To simplify notation, it is useful to write the matrix X and its inverse $B \equiv X^{-1}$ as

$$X = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix} \qquad B = \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix} \tag{3.3}$$

where $x_1(b_1), x_2(b_2), x_3(b_3)$ and $x_4(b_4)$ are $s\times s$, $s\times(p-s)$, $(p-s)\times s$ and $(p-s)\times(p-s)$ matrices, respectively. Each synaptic matrix of the form (3.2) constitutes a particular way of associating patterns; thus, different matrices X give rise to different models.

The following derivation of the mean-field equations relies on the assumption that patterns can be separated into condensed ('low'), i.e. with the overlaps remaining finite in the $N \to \infty$ limit, and non-condensed ('high'), i.e. with the overlaps of magnitude $O(1/\sqrt{N})$. The correctness of this assumption depends on the form of the matrix X. It is justified if the matrix does not couple all the patterns; for instance, if it is a block-matrix. If it is a more general matrix the derivation should be justified *a posteriori*.

The averaged free-energy (3.1) can be calculated using the replica method [7]. Some details of this computation are presented in appendix A. Assuming replica symmetry, the final expression for the averaged free-energy per spin, still in terms of a general matrix X, is

$$\begin{aligned} f = {} & \lim_{N\to\infty}\frac{1}{2N}\mathrm{Tr}X + \frac{1}{2}\sum_{\nu\lambda}^{s}\overline{m}^\nu O[q]_{\nu\lambda}\overline{m}^\lambda + \frac{1}{2}\alpha\overline{r}\beta(1-q) \\ & - \frac{\alpha}{2\beta}\int_0^1 \mathrm{d}u\left[\frac{-\beta q}{\Lambda^{-1}(u)-\beta(1-q)} + \log(\Lambda^{-1}(u)-\beta(1-q))\right] \\ & - \frac{1}{\beta}\left\langle\!\left\langle \log 2\cosh\beta\Big(\sqrt{\alpha\overline{r}}z + \sum_{\lambda}^{s}(\overline{m}^\lambda + h^\lambda)\xi^\lambda\Big)\right\rangle\!\right\rangle_{z,\xi} . \end{aligned} \tag{3.4}$$

The mean $\langle\langle\cdots\rangle\rangle_{z,\xi}$ represents a combined average over discrete ξ^ν and Gaussian z variables. α denotes the memory loading fraction, $\alpha \equiv p/N$. The operator $O[q]$ acts on low-pattern space $\mathcal{L}$ and its matrix representation is

$$O[q]_{\nu\lambda} \equiv b_{1\nu\lambda} - \sum_{\gamma\delta}^{p} b_{2\nu\gamma}(b_4 - \beta(1-q)I)^{-1}{}_{\gamma\delta} b_{3\delta\lambda}$$

$(\nu, \lambda = 1, \ldots, s)$. A sum over the eigenvalues Λ_γ of the high-pattern block of the matrix X has been translated into an integral assuming that they depend on γ/p, $\Lambda_\gamma = \Lambda(\gamma/p)$. The variables $\overline{m}_\nu$ are related to the overlaps $m_\nu = -\delta f/\delta h^\nu$ through

$$m^\nu = \sum_{\lambda}^{s} O[q]_{\nu\lambda}\, \overline{m}^\lambda\,.$$

The general mean-field equations at finite temperature T in terms of the actual overlaps m^λ are

$$m^\nu = \left\langle\!\!\left\langle \xi^\nu \tanh\beta(\sqrt{\alpha\overline{r}}z + \sum_{\lambda\overline{\lambda}}^{s} m^\lambda O[q]^{-1}{}_{\lambda\overline{\lambda}}\xi^{\overline{\lambda}}) \right\rangle\!\!\right\rangle_{z,\xi} \tag{3.5}$$

$$q = \left\langle\!\!\left\langle \tanh^2\beta(\sqrt{\alpha\overline{r}}z + \sum_{\lambda\overline{\lambda}}^{s} m^\lambda O[q]^{-1}{}_{\lambda\overline{\lambda}}\xi^{\overline{\lambda}}) \right\rangle\!\!\right\rangle_{z,\xi} \tag{3.6}$$

$$\overline{r} = -\frac{1}{\alpha}\sum_{\lambda\overline{\lambda}}^{s} m^\lambda \frac{\delta O[q]^{-1}{}_{\lambda\overline{\lambda}}}{\delta q} m^{\overline{\lambda}} + \frac{1}{2\alpha\beta}\frac{\delta\mathcal{I}}{\delta q} \tag{3.7}$$

with

$$\mathcal{I} \equiv \alpha\int_0^1 \mathrm{d}u \left[\frac{-\beta q}{\Lambda^{-1}(u) - \beta(1-q)} + \log(\Lambda^{-1}(u) - \beta(1-q))\right]. \tag{3.8}$$

These equations reduce to the known mean-field equations of .[1, 5, 7]. In the Hopfield model, $X = I$, $\Lambda_\gamma = 1\ \ \forall\gamma$, and $O[q] = O[q]^{-1} = I$; thus, the mean-field equations of [7] are immediately recovered. The finite p situation of section 2 can also be obtained. The matrix X has in this case a finite sector different from zero with components given by $Y^{(s)}{}_{\nu\lambda} = \delta_{\nu\lambda} + a(\delta_{\nu\,\lambda+1} + \delta_{\nu\,\lambda-1} + \delta_{\nu 1}\delta_{\lambda s} + \delta_{\nu s}\delta_{\lambda 1})$. $O[q]^{-1}{}_{\nu\lambda}$ reduces to $Y^{(s)}{}_{\nu\lambda}$. In the $\alpha \to 0$ limit the first mean-field equation decouples and reduces to (2.3).

3.2. The models

The models to be discussed are defined by particular matrices X, (cf (3.3)). Calling $Y^{(t)}$ a $t \times t$ matrix with components given by

$$Y^{(t)}{}_{\mu\eta} = \delta_{\mu\eta} + a(\delta_{\mu\,\eta+1} + \delta_{\mu\,\eta-1} + \delta_{\mu\,t}\delta_{\eta\,1} + \delta_{\mu\,1}\delta_{\eta\,t})$$

the cases under study are represented by

(I) $\quad x_1 = Y^{(s)} \quad x_2 = 0 \quad x_3 = 0 \quad x_4 = I$

(II) $\quad X = Y^{(p)}$.

The separation of patterns in condensed and non-condensed is justified in the first case since 'low' and 'high' learnt patterns are disconnected inside the synaptic matrix, X being a block-matrix. In the second case this assumption is not justified *a priori* although it can be justified *a posteriori*. In fact, although stored patterns form an infinite cycle of dimension $p = \alpha N$, results from the finite p model suggest that this approximation may be valid to describe attractors in a certain range of values of the parameter a. In the finite loading regime, if $0.5 < a < 1$, the network stimulated by each of the stored patterns evolves to an attractor correlated with a *small* number of patterns concentrated around the stimulating pattern (see section 2). More precisely, at zero temperature the overlaps vanish *exactly* at a distance of 5 (cf (2.6)). If, even in the extensive loading regime, the structure of attractors is similar and the decay of the overlaps is fast enough, i.e. $m^s \sim 1/\sqrt{N}$, the assumption will be justified *a posteriori*. Indeed, the results from the mean-field calculation do not contradict this assumption and they are also in good accord with numerical simulations (see section 4).

Network I. The first model, which will be called network I, corresponds to learning a finite cycle of patterns in the presence of an infinite number of patterns already learnt in Hopfield's style, which act as a noise. The eigenvalues of the high block are $\Lambda_\gamma = 1 \forall \gamma$, and $O[q]^{-1} = Y^{(s)}$. Thus, the mean-field equations (3.5)–(3.7) reduce to

$$m^\nu = \left\langle\!\!\left\langle \xi^\nu \tanh \beta\Big(\sqrt{\alpha r}z + \sum_{\lambda\bar{\lambda}}^{s} m^\lambda Y^{(s)}{}_{\lambda\bar{\lambda}} \xi^{\bar{\lambda}}\Big) \right\rangle\!\!\right\rangle_{z,\xi} \tag{3.9}$$

$$q = \left\langle\!\!\left\langle \tanh^2 \beta\Big(\sqrt{\alpha r}z + \sum_{\lambda\bar{\lambda}}^{s} m^\lambda Y^{(s)}{}_{\lambda\bar{\lambda}} \xi^{\bar{\lambda}}\Big) \right\rangle\!\!\right\rangle_{z,\xi} \tag{3.10}$$

$$r = \frac{1}{(1-c)^2} \tag{3.11}$$

where $c \equiv \beta(1-q)$.

Network II. In the second model, called network II, the eigenvalues Λ_γ are those of the matrix $Y'^{(p-s)}{}_{\gamma\delta} = \delta_{\gamma\delta} + a(\delta_{\gamma\,\delta+1} + \delta_{\gamma\,\delta-1})$, which read

$$\Lambda(l/t) = 1 - 2a \cos \frac{t+1-l}{t+1}\pi \qquad l = 1, \ldots, t$$

with $t = p - s$. The integral $\mathcal{I}$ (3.8) can be explicitly computed:

$$\mathcal{I} \equiv \alpha \left[\frac{c-\beta}{c(1-c)} \frac{1}{\sqrt{1-y^2}} (1 - (1-c)\sqrt{1-y^2}) + \log(1-c) \right.$$
$$\left. + \frac{1}{2} \log \left(\frac{(1+\sqrt{1+y^2})(1+\sqrt{1-(2a)^2})}{(1+\sqrt{1-y^2})(1+\sqrt{1+(2a)^2})} \right) \right] \tag{3.12}$$

with

$$y \equiv \frac{2ac}{1-c}.$$

The inverse of matrix $O[q]$ has the following form (see appendix B)

$$O[q]^{-1}{}_{\nu\lambda} = Y'^{(s)}{}_{\nu\lambda} + f_\nu\, \delta_{\lambda 1} + \overline{f}_\nu\, \delta_{\lambda s}\,.$$

The contribution of the last two terms to the mean-field equations (3.5)–(3.7) is neglected since the interesting solutions have rapidly decreasing overlaps, such that m^1 and m^s behave as $1/\sqrt{N}$ when $N \to \infty$. The system to be solved reduces to

$$m^\nu = \left\langle\!\!\left\langle \xi^\nu \tanh \beta\Big(\sqrt{\alpha\overline{r}}z + \sum_{\lambda\overline{\lambda}}^{s} m^\lambda Y'^{(s)}{}_{\lambda\overline{\lambda}}\, \xi^{\overline{\lambda}}\Big)\right\rangle\!\!\right\rangle_{z,\xi} \qquad (3.13)$$

$$q = \left\langle\!\!\left\langle \tanh^2\beta\Big(\sqrt{\alpha\overline{r}}z + \sum_{\lambda\overline{\lambda}}^{s} m^\lambda Y'^{(s)}{}_{\lambda\overline{\lambda}}\, \xi^{\overline{\lambda}}\Big)\right\rangle\!\!\right\rangle_{z,\xi} \qquad (3.14)$$

$$\overline{r} = \frac{-1}{2\alpha}\frac{\delta\mathcal{I}}{\delta c} \qquad (3.15)$$

where s should be taken big enough to ensure $m^s \sim 1/\sqrt{N}$.

Other examples could be considered, such as two disconnected cycles, one of dimension s and the other one of dimension $p-s$, an s-cycle of patterns learnt in the presence of a finite number of infinite cycles, a s-cycle of patterns learnt in the presence of an infinite number of finite cycles, etc. These models do not present a different qualitative behaviour from the ones already described. Indeed, their mean-field equations for the overlaps and q coincide with the ones presented for networks I and II. The mean-field equation related to the noise created by the background patterns, which depends on the high-pattern block eigenvalues Λ_γ, differs from model to model. We will not analyse these examples here but just comment that they have a similar phase diagram.

3.3. *Zero temperature*

Network I. In the zero temperature limit, the mean-field equations of network I read

$$m^\nu = \left\langle\!\!\left\langle \xi^\nu \mathrm{erf}\left(\frac{\sum_{\lambda\overline{\lambda}}^{s} m^\lambda Y^{(s)}{}_{\lambda\overline{\lambda}}\xi^{\overline{\lambda}}}{\sqrt{2\alpha\overline{r}}}\right)\right\rangle\!\!\right\rangle_{\xi^\nu} \qquad (3.16)$$

$$c = \sqrt{\frac{2}{\pi\alpha\overline{r}}}\left\langle\!\!\left\langle \exp\left[-\left(\frac{\sum_{\lambda\overline{\lambda}}^{s} m^\lambda Y^{(s)}{}_{\lambda\overline{\lambda}}\xi^{\overline{\lambda}}}{\sqrt{2\alpha\overline{r}}}\right)^2\right]\right\rangle\!\!\right\rangle_{\xi^\nu} \qquad (3.17)$$

$$\overline{r} = \frac{1}{(1-c)^2}\,. \qquad (3.18)$$

The dynamics of the network stimulated by a pure pattern is described by the iteration of these equations starting from the initial configuration $m_I^\nu = \delta_\nu^\alpha$. Proceeding in this way, the phase diagram (a, α) can be studied. Figure 2 represents it for a network with $s = 13$ (the

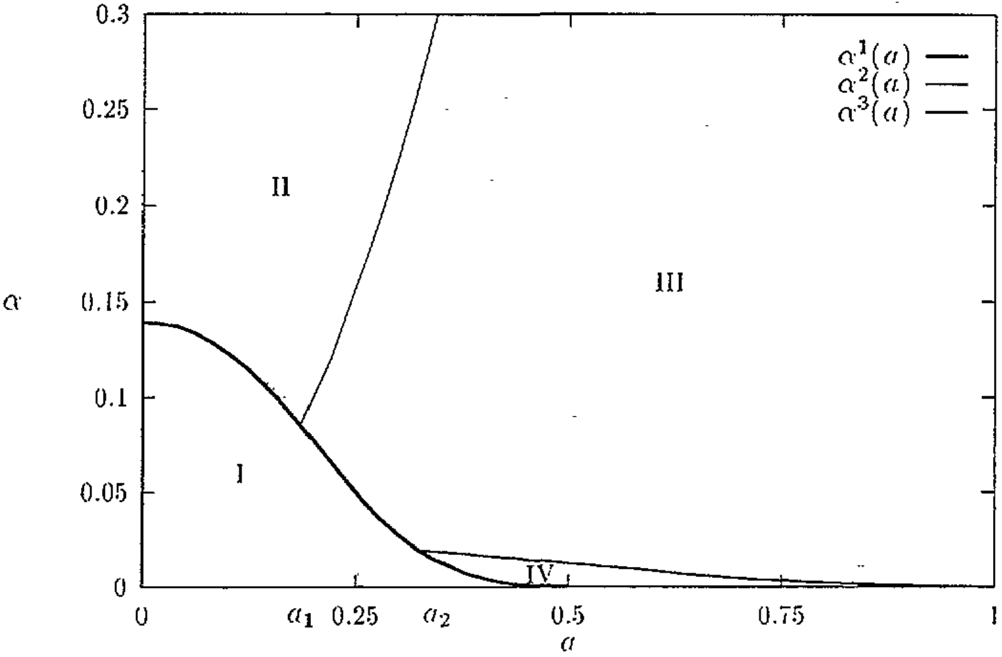

Figure 2. Phase diagram for network I; $T = 0$, $s = 13$.

dependence on the dimension s of the low block is discussed below). Note the similarity with the phase diagram (a, T) for the network with finite loading.

In zone I the system behaves as a Hopfield network. Starting the network from a pure pattern state it evolves to a state having overlap close to one with the selected initial pattern and small, though different from zero, overlaps with the other low patterns. Such attractors exist only for $a < \frac{1}{2}$ and below the critical line $\alpha = \alpha^1(a)$. This line does not depend on the number of condensed patterns s.

For every fixed a inside the interval $[0, \frac{1}{2})$ the attractor correlation with the stimulating pattern goes to one when α approaches zero while it slightly decreases when it gets closer to the critical line $\alpha^1(a)$. Conversely, the overlaps with the neighbouring patterns go to zero when α goes to zero and increase when α approaches the critical capacity line. Nevertheless, the value of the main overlap is close to one while the other components are close to zero in the whole of sector I. The attractors belonging to different stimuli are essentially uncorrelated.

For α above the critical line $\alpha^1(a)$ and $a < a_1$, i.e. zone II, the network stimulated by a pure pattern evolves to a spin glass state corresponding to a null overlap vector. Evidently, when $a = 0$ the critical capacity is that found in [7] for a Hopfield network, $\alpha^1(0) = 0.138$. The spin glass solutions satisfy

$$c = \left(\sqrt{\frac{\pi\alpha}{2}} + 1\right)^{-1} \qquad \bar{r} = \left(1 + \sqrt{\frac{2}{\pi\alpha}}\right)^2.$$

Since these equations are independent of the parameter a the spin glass phase behaves as in the Hopfield model.

If the parameter $a_1 < a < a_2$ and $\alpha > \alpha^1(a)$, the system evolves to a symmetric state. Indeed, a symmetric ansatz $m^\nu = m$, $\nu = 1, \ldots, s$ simplifies the system (3.16)–(3.18) and allows it to be written as a single equation

$$\left(\frac{s}{1+2a}\right)u = \frac{\langle\langle \eta_s \mathrm{erf}[\,u\eta_s]\rangle\rangle_{\eta_s}}{\sqrt{2\alpha} + 2/\sqrt{\pi}\langle\langle \exp[-(u\eta_s)^2]\rangle\rangle_{\eta_s}} \tag{3.19}$$

where $\eta_s \equiv \sum_\nu^s \xi^\nu$ and $u \equiv m(1+2a)/\sqrt{2\alpha\bar{r}}$. The factor $1 + 2a$ implies a dependence on the parameter a. For each s (3.19) defines a critical curve $\alpha^2(a)$ above which solutions with $u \neq 0$ do not exist. It marks a transition between the symmetric (sector III) and the spin glass (sector II) regions. If $a = 0$ (3.19) reduces to the equation for symmetric solutions of the Hopfield model presented in [7]. Hence, the $\alpha^2(a)$ curve actually starts at $a = 0$ and $\alpha^2(0)$ is different from zero (it depends on s, for instance if $s = 3$, $\alpha^2(0) \simeq 0.03$) and increases with a. This means that symmetric solutions exist even for $a < a_1$ though in that region they are not reached by the dynamics of the system, starting from the pure pattern. As for the dependence on the size of the low block, both the $\alpha^2(a)$ curve and a_1 depend on s. The curve $\alpha^2(a)$ is a decreasing function of s and in the big s limit it goes to zero everywhere. Thus, a_1 increases with s.

When the parameter a is bigger than a_2, a new sector appears in the phase diagram. After reaching the curve $\alpha^1(a)$ the network evolves to a state having non-zero and significant overlap with various patterns. The values of the overlaps decay with the distance in the stored sequence from the one used as stimulus. The typical values for the overlap vector corresponding to attractors in zone IV are

$$m^\nu \simeq (0, \ldots, 0, 0.01, 0.1, 0.4, 0.6, 0.4, 0.1, 0.01, 0, \ldots, 0)\,. \tag{3.20}$$

These attractors are similar to those found for the finite loading regime and the finite temperature case (see figure 1). Just as in the finite p limit these attractors are correlated.

For $a > a_2$ a further increase in α implies a new transition, now between zones IV and III. This defines a new critical line called $\alpha^3(a)$ (see Figure 2).

Network II. As for network II at zero temperature, it is described by the following set of mean-field equations:

$$m^\nu = \left\langle\!\!\left\langle \xi^\nu \mathrm{erf}\left(\frac{\sum_{\lambda\bar\lambda}^s m^\lambda {Y'^{(s)}}_{\lambda\bar\lambda}\xi^{\bar\lambda}}{\sqrt{2\alpha\bar{r}}}\right)\right\rangle\!\!\right\rangle_{\xi^\nu} \tag{3.21}$$

$$c = \sqrt{\frac{2}{\pi\alpha\bar{r}}}\left\langle\!\!\left\langle \exp\left[-\left(\frac{\sum_{\lambda\bar\lambda}^s m^\lambda {Y'^{(s)}}_{\lambda\bar\lambda}\xi^{\bar\lambda}}{\sqrt{2\alpha\bar{r}}}\right)^2\right]\right\rangle\!\!\right\rangle_{\xi^\nu} \tag{3.22}$$

$$\bar{r} = \frac{1}{(1-c)^2\sqrt{1-y^2}}\left[1 + \frac{4a^2}{1-y^2}(1-c)^2\right] \tag{3.23}$$

where $y \equiv 2ac/(1-c)$.

According to the discussion in the beginning of the section, this system should be understood as an infinite set of equations for condensed patterns. The only legitimate solutions are those with the overlaps m^ν peaked around one pattern ν and whose values fall off rapidly enough.

This network has a phase diagram similar to that of network I. As we showed earlier, the critical line $\alpha^2(a)$, separating symmetric and spin glass states, moves down to zero as $s \gg 1$, thus there are no symmetric attractors for this network. Thus the curve $\alpha^3(a)$ here marks a transition between correlated and spin glass phases. It can be estimated by iterating (3.21)–(3.22). For example, for network II $\alpha^3(a = 0.6) = 0.0016$, while for network I, when $s = 13$, $\alpha^3(a = 0.6) = 0.0083$.

4. Numerical simulations

The results obtained from the mean-field theory calculation can be checked by performing numerical simulations. On the one hand, the finite loading regime ($p/N \to 0$ when $N \to \infty$) is represented by loading the network with a small number of random patterns, such that $p/N \ll 1$. On the other hand, the extensive loading regime is represented by loading the network with $p = \alpha N$ random patterns. The initial configuration of the network is taken to be one of the memorized patterns. Its state is then updated using asynchronous dynamics. In this scheme units are checked in a prescribed order and whenever one of them flips all the local fields are recomputed and changed accordingly. Finally, when the network stabilizes the overlap between the attractor state and each of the learnt patterns is computed from (2.2). Repeating this procedure with all the learnt patterns used as stimuli the correlations between attractors are calculated at the end using (2.7).

4.1. Finite loading

The finite p situation at zero temperature has been simulated by loading networks of $N = 5000$, 10 000, 20 000 and 30 000 neurons with $p = 15$ patterns. Two values of the parameter a have been selected: $a = 0.45$ and $a = 0.6$, below and above the critical value $a_{\text{cr}} = 0.5$, respectively. For each network the evolution of $N_i = 300$ stimulating patterns has been studied.

When the number of neurons is increased and the number of patterns is kept fixed in such a way that the relation p/N decreases approaching the limit $p/N \ll 1$, the results from simulations are expected to reproduce the mean-field theory results of section 2 more accurately. In the case $a = 0.45$, $a < a_{\text{cr}} = 0.5$, the network should behave as a Hopfield network; the stimulating pattern should evolve to an attractor highly correlated with it, namely a Mattis-type attractor. Results from simulations show that the network with $N = 5000$ neurons behaves in a quite different way, deviations from the theoretical predictions due to the relatively small number of neurons considered ($p/N = 0.003$). The networks with $N = 10\,000$ and $N = 20\,000$ units present an intermediate behaviour. Finally, in the network with $N = 30\,000$ neurons, all initial configurations evolve to Mattis-type attractors, and the network behaves as a Hopfield one, in accord with the mean-field theory prediction.

In the case $a = 0.6$, $a > a_{\text{cr}} = 0.5$, the initial configurations are expected to evolve to attractors mainly correlated with the stimulating pattern but also with its neighbours in the learnt sequence, namely correlated attractors. As for $a = 0.45$, the network presents a quite different behaviour when it has $N = 5000$ neurons, it improves its behaviour when the number of units is $N = 10\,000$ and for $N = 20\,000$ it already behaves in agreement with the theoretical predictions. Furthermore, the values of the overlaps and correlations between attractors for the network with $N = 20\,000$ are also in good accord with those presented in (2.6) and (2.8). The averaged overlaps and correlations are

$$m^{\mu} \simeq (\ldots, 0.02, 0.10, 0.40, 0.60, 0.40, 0.10, 0.02, \ldots) \tag{4.1}$$

$$C_0 \simeq 1 \quad C_1 \simeq 0.67 \quad C_2 \simeq 0.33 \quad C_3 \simeq 0.13 \quad C_4 \simeq 0.05 \quad C_5 \simeq 0.02\,. \tag{4.2}$$

Figure 3 shows the number of stimulating patterns that evolve to a Mattis-type (Mattis), correlated (Corr) or a different (Diff) configuration depending on the number of neurons of the network, both for $a = 0.45$ and $a = 0.6$. Note that the size of the network which must be chosen to reproduce accurately the results of mean-field analysis corresponds roughly to the critical capacity of the model, as should be expected (see also the next section).

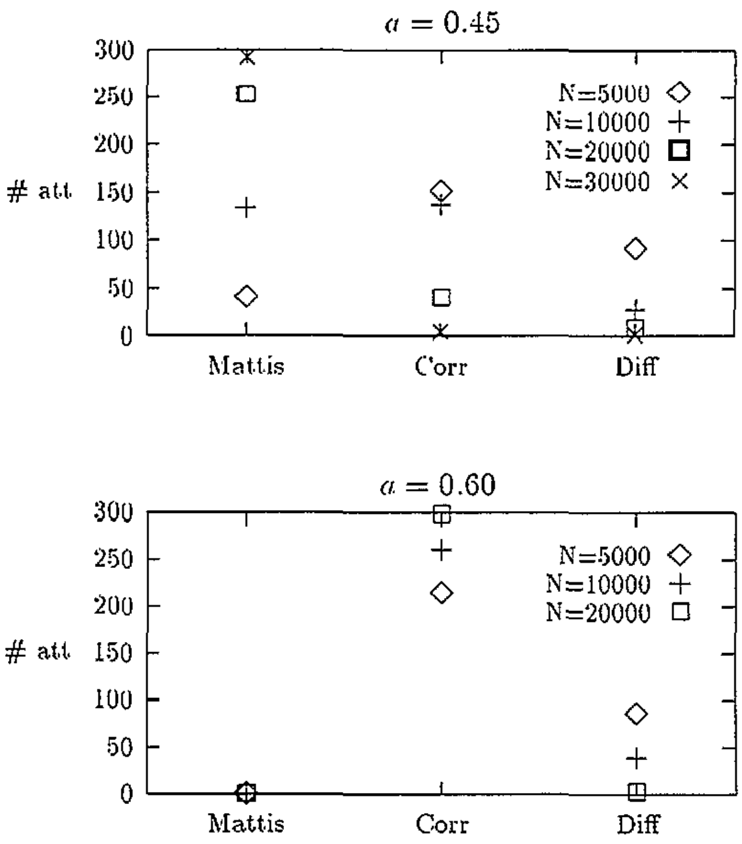

Figure 3. The number of various types of attractors, obtained in the numerical simulations of the networks with fixed number of patterns ($p = 15$) and varying number of neurons; $a = 0.45$ and $a = 0.6$, respectively.

4.2. *Extensive loading*

In this section we shall concentrate on the outcomes from simulations of network II. We chose to simulate this model since an *a priori* unjustified assumption has been made in section 3 to derive its mean-field equations which needs special confirmation from numerical results.

Simulations have been carried out on networks with 20 000 and 40 000 neurons. The networks have been loaded with $p = \alpha N$, α fixed, random patterns. Again, two different values of the parameter a have been selected, $a = 0.45$ and $a = 0.6$. For $a = 0.6$, the mean-field theory analysis predicts that no Mattis-type attractors exist and that the critical storage capacity at which correlated attractors disappear is $\alpha^3 = 0.0016$. In figure 4 we plot the fraction of runs leading to attractors different from correlated ones, for capacities $\alpha = 0.0014$, 0.0016 and 0.0018 around the critical α^3.

The points in each graph correspond to a different choice of p and N such that α is fixed. It can be seen that as the number of neurons and patterns is increased proportionally, the probability of obtaining the correlated type of attractors increases (decreases) if $\alpha < \alpha^3$ ($\alpha > \alpha^3$). We believe that the remaining deviations from the theory are due to finite size effects.

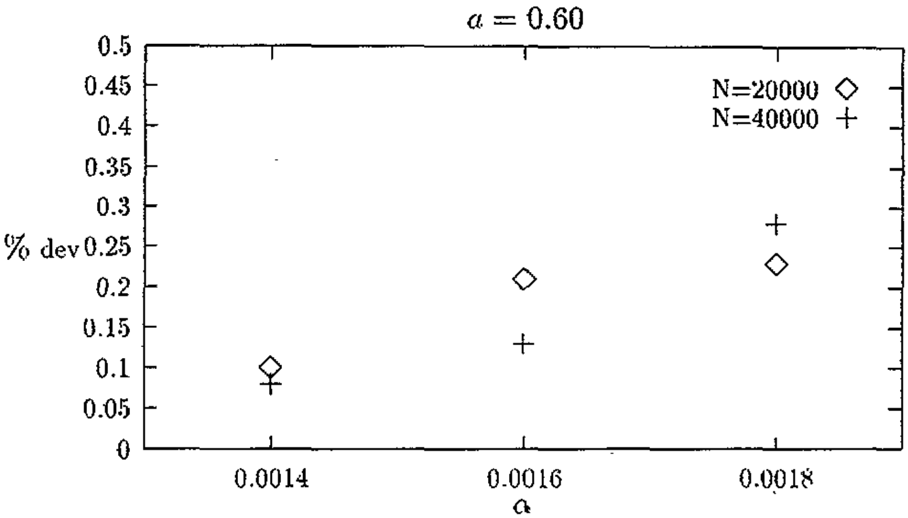

Figure 4. Simulations of network II with finite loading. Probability of obtaining an attractor of the type different from (4.2), for different values of α.

The averaged overlaps and correlations obtained from simulations in this regime have similar values to those given by (4.1), (4.2).

5. Conclusions

In conclusion, it is important to stress the qualitative difference between the present model, introduced in [1], and the previous work on the Hopfield model with correlated attractors (see e.g. [8–11]). In all these previous works the patterns presented to the network have fixed correlations, which are inherited by the attractors due to the learning algorithm. In the present approach, the patterns are uncorrelated, and the correlations between attractors are due to the associations of different patterns at the learning stage. It opens the possibility of *investigating the influence of various learning protocols on the correlation structures of the* learned attractors.

In this work we have studied the extensive loading of the neural network with correlated attractors associated to uncorrelated patterns introduced in [1]. We have shown that the peculiar correlated attractors of the finite loading [1, 5] are not destroyed by the loading of an extensive number of patterns. The phase diagram of this network, as well as the ones corresponding to related ways of associating patterns in the learning session, have a region with correlated attractors. The values of the overlaps and correlations are not dramatically modified by the extensive loading. Finally, numerical simulations described in section 4 confirm the mean-field theory results both for the overlaps and the correlations and for the critical capacity lines of the phase diagrams.

Acknowledgment

We wish to thank Professor D J Amit for helpful discussions and careful reading of the manuscript.

Appendix A

In order to compute the free energy per spin f, (3.1), with the replica method [7] one starts by calculating the mean of the replicated partition function $Z'^{n} = \mathrm{Tr}_{s^\rho}\exp[-\beta\sum_{ij}^{N}\sum_{\rho}^{n} J_{ij} s_i^\rho s_j^\rho]$ (Greek indices $\rho, \sigma = 1, \ldots, n$ label replicas). The ξ-variable dependence is then linearized through the Hubbard–Stratonovich (HS) identity and the variables $\overline{m}^{\mu\rho}$, $\mu = 1, \ldots, p$ are introduced.

Assuming that X is such that s patterns $\nu = 1, \ldots, s$ 'condense', i.e. have a finite overlap when N goes to infinity, the pattern space $\mathcal{P}$ can be written as a direct sum $\mathcal{P} = \mathcal{L} \oplus \mathcal{H}$ with $\mathcal{L}$ the condensed or low- pattern space and $\mathcal{H}$ the non-condensed or high-pattern space ($\gamma, \delta = s + 11, \ldots, p$ label high patterns). X represents an operator acting on $\mathcal{P}$.

The average over non-condensed patterns can then be explicitly computed. Moreover, changing variables $\overline{m}^{\gamma\rho} \to \overline{m}^{\gamma\rho}/\sqrt{\beta N}$ and expanding around big N, the integrals over $\overline{m}^{\gamma\rho}$ are quadratic and can be performed. Thus

$$\langle\!\langle Z'^{n}\rangle\!\rangle_{\xi^\mu} = (\det X)^{-n/2}(\sqrt{\beta N})^{ns}(\alpha\beta^2)^{n(n-1)/2}\int \mathcal{D}_{\nu\rho}\overline{m}\int \mathcal{D}_{[\rho\sigma]}\overline{r}\int \mathcal{D}_{[\rho\sigma]}q$$
$$\times \exp\Bigg[-\tfrac{1}{2}\beta N\overline{m}^{\nu\rho}\overline{L}[Q]_{\nu\lambda,\rho\sigma}\overline{m}^{\lambda\sigma} - \tfrac{1}{2}\mathrm{Tr}\log\overline{K}[Q] - \tfrac{1}{2}\alpha\beta^2 N\sum_{\rho\neq\sigma}^{n}\overline{r}_{\rho\sigma}q_{\rho\sigma}$$
$$+\left\langle\!\!\left\langle \log\mathrm{Tr}_{s_\rho}\exp\left(\tfrac{1}{2}\alpha\beta^2\sum_{\rho\neq\sigma}^{n}\overline{r}_{\rho\sigma}s^\rho s^\sigma + \beta\sum_{\rho}^{n}\sum_{\nu}^{s}(\overline{m}^{\nu\rho} + h^\nu)\xi^\nu s^\rho\right)\right\rangle\!\!\right\rangle_{\xi^\nu}\Bigg]. \tag{A.1}$$

The measure corresponds to $\mathcal{D}_{\nu\rho}\overline{m} \equiv \prod_{\nu}^{s}\prod_{\rho}^{n}(\mathrm{d}\overline{m}^{\nu\rho}/\sqrt{2\pi})$. $\overline{r}_{\rho\sigma}$ ($\rho \neq \sigma$) is a symmetric Lagrange multiplier which enforces the relation $q_{\rho\sigma} = \sum_i^N s_i^\rho s_i^\sigma$. Q is then a symmetric operator acting on replica space ($q_{\rho\sigma}, \rho \neq \sigma$, with 0 elements in the diagonal). $\mathcal{D}_{[\rho\sigma]}\overline{r} \equiv \prod_{\rho<\sigma}^{n} d\overline{r}_{\rho\sigma}$ and $\mathcal{D}_{[\rho\sigma]}q \equiv \prod_{\rho<\sigma}^{n} dq_{\rho\sigma}$.

$\overline{L}[Q]$ and $\overline{K}[Q]$ are operators acting on the spaces $\mathcal{S} = \mathcal{L}\otimes\mathcal{R}$ and $\mathcal{S}' = \mathcal{H}\otimes\mathcal{R}$, defined as direct products between low-pattern and replica space ($\mathcal{R}$) and between high-pattern and replica space, respectively. The operators $\overline{L}[Q]_{\nu\lambda,\rho\sigma}$ and $\overline{K}[Q]_{\gamma\delta,\rho\sigma}$ have the following matrix representations:

$$\begin{aligned}\overline{L}[Q]_{\nu\lambda,\rho\sigma} &\equiv b_{1\nu\lambda}\delta_{\rho\sigma} - \sum_{\gamma\delta}^{p} b_{2\nu\gamma}\overline{K}[Q]^{-1}{}_{\gamma\delta,\rho\sigma}b_{3\delta\lambda}\\ \overline{K}[Q]_{\gamma\delta,\rho\sigma} &\equiv (b_4 - \beta I)_{\gamma\delta}\delta_{\rho\sigma} - \beta\delta_{\gamma\delta}q_{\rho\sigma}\,.\end{aligned} \tag{A.2}$$

The external fields h^ν, $\nu = 1, \ldots, s$ have been introduced to signal the s patterns expected to condense.

In the $n \to 0$ limit the multiplying constant factor in (A.1) reduces to one. Thus, the averaged free-energy per spin (3.1) reads

$$f = \lim_{N\to\infty}\frac{1}{2N}\mathrm{Tr}X + \alpha\beta\lim_{n\to 0}\frac{1}{2n}\sum_{\rho\neq\sigma}^{n}\overline{r}_{\rho\sigma}q_{\rho\sigma} + \lim_{n\to 0}\frac{1}{2n}\sum_{\nu\lambda}^{s}\sum_{\rho\sigma}^{n}\overline{m}^{\nu\rho}\overline{L}[Q]_{\nu\lambda,\rho\sigma}\overline{m}^{\lambda\sigma}$$
$$-\frac{1}{\beta}\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{2nN}\mathrm{Tr}\log\overline{K}[Q]$$
$$-\lim_{n\to 0}\frac{1}{\beta n}\left\langle\!\!\left\langle \log\mathrm{Tr}_{s^\rho}\exp\left[\frac{\alpha\beta^2}{2}\sum_{\gamma=1\rho\neq\sigma}^{n}\overline{r}_{\rho\sigma}s^\rho s^\sigma\frac{1}{p} + \beta\sum_{\nu}^{s}\int_0\sum_{\rho}^{n}\mathrm{d}(\overline{m}^{\nu\rho}(\Lambda + h^\nu))\xi^\nu\right.\right.$$

Assuming replica symmetry, this expression can be simplified and the limits $n \to 0$ taken. In order to do so, it will be useful to adopt a compact notation to describe the operators acting on product spaces, more precisely, as direct products denoted by $\otimes$ between operators acting on pattern space (first factors) and operators acting on replica space (second factors). $\overline{K}[Q]$ can then be written as

$$\overline{K}[Q] \equiv (b_4 - \beta I) \otimes I - \beta I \otimes Q$$

and assuming replica symmetry

$$\overline{K}[Q] = (b_4 - \beta I) \otimes I - \beta q I \otimes (1 - I)$$

where operator 1 has a matrix representation given by $1_{\rho\sigma} = 1$, $\rho, \sigma = 1, \ldots, n$. The inverse $\overline{K}[Q]^{-1}$ can be written in the form

$$\overline{K}[Q]^{-1} = C \otimes I + D \otimes (1 - I)$$

with C and D acting on $\mathcal{H}$ and satisfying $C - D = (b_4 - \beta(1-q)I)^{-1}$, independently of n. The operator $\overline{L}[Q]$ is described by

$$\overline{L}[Q] = b_1 \otimes I - b_2 \overline{K}[Q]^{-1} b_3 .$$

Finally, $\overline{m}^{\nu\rho}$ *is a vector in* S *and can be represented by* $\overline{m} \otimes 1$.

In the replica symmetry approach the third term in (A.3) reads

$$\begin{aligned}
\text{3rd Term} &\equiv \lim_{n \to 0} \frac{1}{n} (\overline{m} \otimes 1) \overline{L}[Q] (\overline{m} \otimes 1) \\
&= \overline{m} b_1 \overline{m} - \lim_{n \to 0} \frac{1}{n} (\overline{m} \otimes 1) b_2 [C \otimes I + D \otimes (1 - I)] b_3 (\overline{m} \otimes 1) \\
&= \overline{m}[b_1 - b_2 (b_4 - \beta(1-q)I)^{-1} b_3] \overline{m}
\end{aligned}$$

As regards to the fourth term in (A.3)

$$\text{4th Term} \equiv \lim_{N \to \infty} \lim_{n \to 0} \frac{1}{2nN\beta} \operatorname{Tr} \log \overline{K}[q]$$

it can be computed as follows

$$\det \overline{K}[q] = \det[(b_4 - \beta I) \otimes I - \beta q I \otimes (1 - I)] = \det[(\Lambda^{-1} I - \beta I) \otimes I - \beta q I \otimes (1 - I)]$$

where Λ^{-1} is a vector in $\mathcal{H}$ with components given by X^{-1}-eigenvalues. It is easy to see that the operator in the RHS has $p - s$ eigenvalues $\Lambda_\gamma^{-1} - \beta - \beta q(n-1)$ and $(p-s)(n-1)$ eigenvalues $\Lambda_\gamma^{-1} - \beta + \beta q$. Thus,

$$\lim_{N \to \infty} \lim_{n \to 0} \frac{1}{2nN\beta} \operatorname{Tr} \log \overline{K}[q] = \frac{1}{2\beta} \lim_{N \to \infty} \frac{1}{N} \sum_{\gamma}^{p} \frac{-\beta q}{\Lambda_\gamma^{-1} - \beta + \beta q} + \log[\Lambda_\gamma^{-1} - \beta + \beta q] .$$

If Λ_γ depends on γ/p, this sum can be transformed into the integral $\mathcal{I}$ (cf (3.8)) since

$$\lim_{N \to \infty} \frac{1}{N} \sum_{\gamma}^{p} f(\Lambda(\frac{\gamma}{p})) = \lim_{N \to \infty} \frac{p}{N} \sum_{\overline{\gamma} = s + 1/p \, \delta\overline{\gamma} = 1/p}^{1} f(\Lambda(\overline{\gamma})) \frac{1}{p} = \alpha \int_0^1 \mathrm{d}u \, f(\Lambda(u)) .$$

Appendix B

Using the relations between $b_1, \ldots, b_4$ and $x_1, \ldots, x_4$ coming from $BX = I$, the operator $O[c]^{-1}$ can be written as

$$O[c]^{-1} = x_1[(I - x_1^{-1}x_2x_4^{-1}x_3)^{-1} + x_1^{-1}x_2b_4(I - cb_4^{-1})^{-1}x_3]^{-1}$$

with $b_4 = (x_4 - x_3x_1^{-1}x_2)^{-1}$. The form of $O[c]^{-1}$ can be understood knowing the form of the matrix A^{-1},

$$A \equiv (I - x_1^{-1}x_2x_4^{-1}x_3)^{-1} + x_1^{-1}x_2b_4(I - cb_4^{-1})^{-1}x_3 = A_1^{-1} + A_2 .$$

Since x_2 and x_3 are matrices with only two components different from zero , the product x_2Zx_3, for any $p - s \times p - s$ matrix Z, is an $s \times s$ matrix with four components different from zero, namely components 11,1s,s1 and ss. Thus, A_2 has only two columns different *from ze , columns 1 and s and* A_1 *is a diagonal matrix with columns 1 and s modified by* the second term. Finally, since ${A_1}^{-1}$ has the same form as A_1, the same happens to A and it is easy to see that $O[c]^{-1}$ is

$$O[c]^{-1}{}_{\nu\lambda} = Y'^{(s)}{}_{\nu\lambda} + f_\nu\,\delta_{\lambda 1} + \overline{f}_\nu\,\delta_{\lambda s} .$$

References

[1] Griniasty M, Tsodyks M and Amit D J 1993 Convertion of temporal correlations between stimuli to spatial correlations between attractors *Neural Comput.* 5 1

[2] Miyashita Y 1988 Neuronal correlate of visual associative long-term memory in the primate temporal cortex *Nature* 335 817

Miyashita Y and Chang H S 1988 *Neural correlate of pictorial short-term memory in the primate temporal cortex Nature* 331 68

[3] Hopfield J J 1982 Neural networks and physical systems with the emergent selective computational abilities *Proc. Natl Acad. Sci. USA* 79 (1982) 2554

[4] Amit D J *Modeling Brain Function* (Cambridge: Cambridge University Press)

[5] Cugliandolo L F 1994 Correlated attractors from uncorrelated stimuli *Neural Comput.* 6 220

[6] Amit D J, Brunel N and Tsodyks M 1993 Correlations of cortical Hebbian reverberations: experiment and theory *J. Neurophysiol.* submitted

[7] Amit D J, Gutfreund H and Sompolinsky H 1987 Statistical mechanics of neural networks near saturation *Ann. Phys., NY* 173 30

[8] Feigelman M V and Ioffe L B 1987 The augmented models of associative memory: asymmetric interaction and hierarchy of patterns *Int. J. Mod. Phys.* 1 51

[9] Gutfreund H 1988 Neural networks with hierarchically correlated patterns *Phys. Rev.* B 37 570

[10] Tsodyks M V 1990 Hierarchical associative memory in neural networks with low activity level *Mod. Phys. Lett.* B 4 259

[11] Fontanari J F and Theumann W K 1990 On the storage of correlated patterns in Hopfield's model *J. Physique* 51 375