Neural networks with fast time-variation of synapses

J J Torres, P L Garrido and J Marro

Instituto Carlos I de Física Teórica y Computacional, and Departamentos de Física Aplicada y Moderna, Facultad de Ciencias, Universidad de Granada, 18071-Granada, España

Abstract. We study a kinetic neural network in which the intensity of synaptic couplings varies on a timescale of order $p(1 - p)^{-1}$ compared with that for neuron variations. We describe some exact and mean-field results for $p \to 0$. This includes, for example, the Hopfield model with random fluctuations of synapse intensities such that neurons couple each other, on average, according to the Hebbian learning rule. The consequences of such fluctuations on the performance of the network are analysed in detail for some specific choices of the rate and fluctuation distribution, including the case in which couplings are asymmetric.

1. Introduction

The Hopfield model for associative memory [1] and its variations comprises a set of N binary *neurons*, $s_x = \pm 1$, whose activity state evolves with time by some either deterministic or stochastic process (corresponding to zero and finite temperature, respectively). The neurons interact with each other according to the Hebb's rule [2], for example. That is, the neuron at x is connected with the neuron at y by a *synapse* whose intensity is

$$J_{xy} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_x^\mu \xi_y^\mu \tag{1}$$

where $\{\xi_x^\mu = \pm 1\} \equiv \boldsymbol{\xi}^\mu$ represent $\mu = 1, \ldots, P$ *memorized* patterns. The values J_{xy} are independent of time. One may interpret either that the synapse intensities have been determined during a previous learning (*plasticity*) process, or else that two different, well defined timescales exist for the time evolution of synapses and neurons, respectively, the former being much larger than the latter.

One may argue, however, that neglecting time variations of the synapses further than those during the learning process is not realistic. Indeed a typical situation in biological systems is that the synapse intensities vary with time also on a timescale smaller than that for neuron activity. That is, in a time interval, Δt, of the order of the time elapsed between the generation of two consecutive action potentials, the synapses may change substantially, for example due to variations in the number of vesicles in the synaptic buttons that induce local fluctuations of the neurotransmitters concentration [3]. Another motivation for (fast) local variations of J_{xy} in the models is the fact that biological neurons are, in practice, connected by more than one synapse [4], each having a different nature, either chemical or electrical, which, consequently, transmit the action potential at different speeds. These and many other facts have been described that make synapses very noisy [5–10], and a rather general belief exists that such *noise* might be at the origin of the observed robustness and

efficacy of biological networks [5, 9, 11, 12]. As a further effort towards the analysis of such possibility, in this paper we report on a neural-network model that, in addition to (slow) learning plasticity, involves relatively fast local fluctuations of the synapse intensities. We consider the simplest situation consistent with the observation in biological systems, that is, we assume that the intensities J_{xy} vary randomly with time in such a way that its average over Δt, $\bar{J}_{xy}$, has the value corresponding to the involved learning rule. For example, assuming the Hebbian case, J_{xy} constantly *visits* at random the memorized patterns during time evolution taking instant values $\xi_x^\mu \xi_y^\mu$ with some probabilities, and $\bar{J}_{xy}$ is given by (1).

The influence on emergent properties of such fluctuations happens to be interesting. In this paper we focus on specific distributions for the fluctuations that produce exact results, namely, effective Hamiltonians for both symmetric couplings, $J_{xy} = J_{yx}$ (section 3), and asymmetric ones, namely, $J_{xy} \neq J_{yx}$ (section 4; see also section 6). In section 5 we then use the replica trick formalism to obtain explicit results from these effective Hamiltonians. The most general description here consists of a kinetic mean-field approach (section 6), which reveals a varied system behaviour. In particular, we show explicitly that allowing for fluctuations significantly affects the property of associative memory (section 3). For example, as compared with the ordinary case lacking these fluctuations, the occurrence of the spin-glass phase at finite temperature is substantially restricted in our model, and it does not appear at zero temperature (i.e. in the absence of thermal noise) above a critical value for the number of stored patterns. On the other hand, this version of the model is not critically affected by the *Almeida–Thouless* (AT) *line* or limit of stability for the replica symmetry solution. Also interesting is the fact that our model (section 2) admits more general distributions for the fluctuations, and one may devise more realistic time processes for the synapses than considered explicitly below. We plan to report soon on further realizations of the same system [13].

2. Definition of the model

Let us denote by $s = \{s_x = \pm 1\}$ and $J = \{J_{xy} \in \mathbb{R}\}$ neuron and synapse configurations, respectively. The probability of state (s, J) at time t changes stochastically according to the master equation:

$$\partial_t P_t(s, J) = \sum_{s', J'} [P_t(s', J') c(s', J' \to s, J) - P_t(s, J) c(s, J \to s', J')]. \tag{2}$$

We assume that some undetermined agents that include a heat bath at temperature T induce transitions $(s, J) \to (s', J')$ whose probability per unit time (rate) is

$$c(s, J \to s', J') = p c_1(s \to s'|J) \delta_{J, J'} + (1 - p) c_2(J \to J') \delta_{s, s'} \tag{3}$$

where $\delta_{X, X'}$ represents the Kronecker delta function,

$$c_1(s \to s'|J) = \sum_y \varpi_J(s; y) \delta_{s_y, -s'_y} \prod_{x \neq y} \delta_{s_x, s'_x} \tag{4}$$

and

$$c_2(J \to J') = \sum_{\substack{x, y \\ x \neq y}} \varpi(J_{xy} \to J'_{xy})(1 - \delta_{J_{xy}, J'_{xy}}) \prod_{\substack{u, v \\ (u, v) \neq (x, y)}} \delta_{J_{uv}, J'_{uv}}. \tag{5}$$

That is, time evolution proceeds by superposition of two kinds of elementary processes (one per unit time), namely, transitions $s_x \to -s_x$ that occur with rate $p \varpi_J(s; x)$. For given synapse configurations and transitions $J_{xy} \to J'_{xy} \neq J_{xy}$ that occur with rate

$(1 - p)\varpi(J_{xy} \to J'_{xy})$. For simplicity, we shall assume the latter to be independent of the current neuron configuration. One may write

$$\partial_t P_t(\mathbf{s}, \mathbf{J}) = p \sum_{\mathbf{x}} [\varpi_{\mathbf{J}}(\mathbf{s}^{\mathbf{x}}; \mathbf{x}) P_t(\mathbf{s}^{\mathbf{x}}, \mathbf{J}) - \varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{x}) P_t(\mathbf{s}, \mathbf{J})]$$

$$+ (1 - p) \sum_{\substack{x,y \\ x \neq y}} \sum_{J'_{xy}} [\varpi(J'_{xy} \to J_{xy}) P_t(\mathbf{s}, \mathbf{J}^{xy}) - \varpi(J_{xy} \to J'_{xy}) P_t(\mathbf{s}, \mathbf{J})] \quad (6)$$

where $\mathbf{s}^{\mathbf{x}}$ represents the configuration obtained from $\mathbf{s}$ after the change $s_x \to -s_x$, and $\mathbf{J}^{xy}$ stands for $\mathbf{J}$ after the change $J_{xy} \to J'_{xy}$. We restrict ourselves hereafter to the case

$$\varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{x}) = \Psi[2\beta s_x h_x(\mathbf{s}, \mathbf{J})] \quad (7)$$

where $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant, and

$$h_x(\mathbf{s}, \mathbf{J}) = \sum_{y \neq x} J_{xy} s_y - \theta_x \quad (8)$$

is a local field with θ_x, the threshold energy needed to activate the neuron at $\mathbf{x}$. Ψ is an arbitrary function except that $\Psi(0) = 1$, $\lim_{X \to +\infty} \Psi(X) = 0$, and $\Psi(X) = \Psi(-X) \exp(-X)$; the latter is the so-called detailed balance condition. The functions

$$\Psi(X) = \begin{cases} e^{-\frac{X}{2}} & \text{rate } V \\ 2(1 + e^X)^{-1} & \text{rate } K \\ \min\{1, e^{-X}\} & \text{rate } M \end{cases} \quad (9)$$

are explicitly considered below.

Under well defined limits, these simplifying assumptions, namely (7), (9), and independence of $\varpi(J_{xy} \to J'_{xy})$ on $\mathbf{s}$, reduce the model to familiar cases, which is interesting for reference purposes. For $p \equiv 1$, one has $c(\mathbf{s}, \mathbf{J} \to \mathbf{s}', \mathbf{J}') = c_1(\mathbf{s} \to \mathbf{s}'|\mathbf{J})\delta_{\mathbf{J},\mathbf{J}'}$, and (6) transforms into

$$\partial_t P_t(\mathbf{s}, \mathbf{J}) = \sum_{y} [\varpi_{\mathbf{J}}(\mathbf{s}^{\mathbf{y}}; \mathbf{y}) P_t(\mathbf{s}^{\mathbf{y}}, \mathbf{J}) - \varpi_{\mathbf{J}}(\mathbf{s}; \mathbf{y}) P_t(\mathbf{s}, \mathbf{J})]. \quad (10)$$

This describes synapses that remain frozen-in during the time interval that characterizes the variation of $\mathbf{s}$. The *quenched* value of $\mathbf{J}$ is determined, for example, by previous storage of P patterns, $\boldsymbol{\xi}^{\mu}$, and no further synaptic modification occurs (on the timescale of $\mathbf{s}$). Several cases associated with different symmetries and learning rules were considered before, such as for example, as follows.

(i) $J_{xy} = J_{yx}$. This symmetry is not realistic for all biological systems [4], but it reduces the original problem to a simpler one. That is, detailed balance then implies that the steady state corresponds to thermodynamic equilibrium at temperature T for energy

$$\mathcal{H}_{\mathbf{J}}(\mathbf{s}) = -\frac{1}{2} \sum_{\substack{x,y \\ x \neq y}} J_{xy} s_x s_y + \sum_{x} \theta_x s_x. \quad (11)$$

This is the Ising Hamiltonian if $J_{xy} = J$ for any pair of nearest-neighbour pairs and zero otherwise. Assuming instead a random spatial distribution for the J_{xy}'s, one obtains the Edwards–Anderson model of spin glasses [14], or else one of the familiar variations of it. Alternatively, the Hebbian learning rule (1) leads to the Hopfield model [1].

(ii) $J_{xy} \neq J_{yx}$. There is no Hamiltonian description, in general, but one may obtain some information from the master equation. For example, one has from (10) that

$$\partial_t \langle s_x \rangle = -2 \langle s_x \Psi \{ [2\beta s_x h_x(\mathbf{s}, \mathbf{J})] \} \rangle. \quad (12)$$

After introducing the mean local field $m_x(J) \equiv \sum_y J_{xy} \langle s_y \rangle$ (for simplicity, we assume $\theta_x = 0 \forall x$ hereafter), and the mean-field assumption $s_x \approx \langle s_x \rangle$, which implies $h_x(s, J) \approx m_x(J)$, one obtains for the steady state that $\langle s_x \rangle \approx \tanh\{\beta m_x(J)\}$, where the mean local field follows by self-consistency from

$$m_x(J) = \sum_{y \neq x} J_{xy} \tanh\{\beta m_y(J)\} \qquad J_{xy} \neq J_{yx}. \tag{13}$$

This is the solution obtained by Crisanti and Sompolinsky [15].

The preceding discussion indicates that equation (6) characterizes a class of Hopfield-like models that may allow us to illustrate the influence on emergent behaviour of some features (a certain type of synaptic noise) not considered before; it may perhaps lead to an evaluation of the role of fluctuating patterns of synaptic connections in biological systems. As a first step, we are particularly interested here in the limit $p \to 0$. This corresponds to a situation in which, once the learning process is complete, the synapses intensities change (fluctuate) very fast as compared with neuron changes, in such a way that one may distinguish two well defined timescales. That is, there is a microscopic timescale, τ, for the fluctuations of the synapses, in which neurons do not appreciably evolve, and a different scale, $t = p\tau$ for $p \to 0$ and $\tau \to \infty$, in which neurons evolve under a steady distribution for the synapses; we refer elsewhere [16] for a detailed study of such separation of timescales. Under such condition, which, as discussed in section 1, can be interpreted as an oversimplification of the actual situation in neurobiology, (6) transforms into

$$\partial_\tau f_\tau(J) = \sum_{\substack{x,y \\ x \neq y}} \sum_{J'_{xy}} [\varpi(J'_{xy} \to J_{xy}) f_\tau(J^{xy}) - \varpi(J_{xy} \to J'_{xy}) f_\tau(J)] \tag{14}$$

for the synapses, and

$$\partial_t P_t(s) = \sum_x [\bar{\varpi}(s^x; x) P_t(s^x) - \bar{\varpi}(s; x) P_t(s)] \tag{15}$$

for the neurons, where

$$\bar{\varpi}(s; x) = \int dJ \, f(J) \varpi_J(s; x) \tag{16}$$

with $f(J)$ the stationary solution of (14), $f(J) = \lim_{\tau \to \infty} f_\tau(J)$. In other words, the system reduces in the limit $p \to 0$ to the Markovian process (15) with effective rate (16). Note that such a result relies on the assumption that $\varpi(J_{xy} \to J'_{xy})$ in (6) is independent of the current neuron configuration; therefore, the stationary solution of (14) for the fast variables does not depend on the more slowly varying neurons [16]. The only motivation for our restriction to such a *non-generic* case here is simplicity.

This is the system that we have studied in detail for different choices of $f(J)$ and $\varpi_J(s; x)$. We only consider here elementary rates $\varpi_J(s; x)$ such that (7)–(9) hold (though other simple cases may be worked out as well). The choice for $f(J)$ needs to take into account the symmetric or asymmetric nature of J_{xy}. In order to deal with variations of the most familiar case, we assume that the fluctuations described by $f(J)$ are around mean values corresponding to the involved learning rule, i.e.

$$\bar{J}_{xy} \equiv \int dJ \, f(J) J_{xy} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_x^\mu \xi_y^\mu \tag{17}$$

for the Hebbian case. A choice consistent with this is

$$f(J) = \prod_{\substack{x,y \\ x \neq y}} \delta\left(J_{xy} - \frac{1}{N} \sum_{\mu=1}^{P} \xi_x^\mu \xi_y^\mu\right) \tag{18}$$

however, this impedes any fluctuation in practice, and the model reduces to the Hopfield one with $J_{xy} = J_{yx}$. More interesting is

$$f(\boldsymbol{J}) = \prod_{\substack{x,y \\ x \neq y}} \sum_{\mu=1}^{P} a_\mu \delta(J_{xy} - \eta_{xy}^\mu) \tag{19}$$

which implies (17) for

$$\eta_{xy}^\mu = \lambda_\mu \xi_x^\mu \xi_y^\mu \qquad \lambda_\mu = (a_\mu N)^{-1}. \tag{20}$$

This corresponds to the case in which stored patterns contribute η_{xy}^μ with probability a_μ to the intensity J_{xy}, independently of the others, so that no spatial correlation exists between different synapses. A series of results for this case are reported below. Another case of interest that implies (17) is

$$f(\boldsymbol{J}) = \sum_{\mu=1}^{P} a_\mu \prod_{\substack{x,y \\ x \neq y}} \delta(J_{xy} - \eta_{xy}^\mu) \tag{21}$$

i.e. each pattern $\boldsymbol{\xi}^\mu$ contributes with probability a_μ to the synapse configuration $\boldsymbol{J}$, so that spatial correlations exist. A detailed study of this situation by approximate methods for finite P will be reported elsewhere [13].

3. Exact results for symmetric synapses

Consider synapse fluctuations such that, *on average*, neurons couple each other according to some learning rule. In order to determine the consequences of this on the performance of the model, we first deal with one of its simplest realizations. The aim is to obtain exact results, namely, to describe the system in terms of an *effective Hamiltonian* [17, 18], which severely restricts the choices for $\Psi(X)$ and $f(\boldsymbol{J})$. In fact, we only consider rate V (cf equation (9)) and

$$f(\boldsymbol{J}) = \prod_{\substack{x,y \\ x \neq y}} f(J_{xy}). \tag{22}$$

It then follows that an effective Hamiltonian exists that, as in the Hopfield case, contains the relevant information about the stored patterns. Both symmetric and non-symmetric couplings may be considered; the latter are studied in section 4.

Let us assume (19), i.e. $f(J_{xy}) = \sum_\mu a_\mu \delta(J_{xy} - \lambda_\mu \xi_x^\mu \xi_y^\mu)$, so that fluctuations are around mean values given by the Hebb's rule. One has $\lambda_\mu = \alpha$ with $\alpha \equiv PN^{-1}$ from normalization $\sum_\mu a_\mu = 1$ for equal probability, $a_\mu = P^{-1}$; cf (20). The fluctuations around the Hebbian rule are then characterized by

$$\sigma_{xy}^2(\boldsymbol{\xi}) \equiv \overline{(J_{xy} - \bar{J}_{xy})^2} = \alpha^2 - \frac{1}{N^2} \sum_{\mu,\nu} \xi_x^\mu \xi_y^\mu \xi_x^\nu \xi_y^\nu \tag{23}$$

which varies between 0 and $2\alpha^2$. That is, in the thermodynamic limit, $N \to \infty$, with $\alpha \neq 0$ (the number P of patterns that the system stores is comparable with the lattice size), the largest value of $\sigma_{xy}^2(\boldsymbol{\xi})$ is for $\xi_x^\mu \xi_y^\mu \xi_x^\nu \xi_y^\nu = -1 \quad \forall \mu, \nu, x, y$. Otherwise, $\alpha = 0$, and the last term of (23), which measures correlations between learnt patterns, also vanishes. We sometimes assume below probabilities

$$p(\xi_x^\mu) = \tfrac{1}{2}[\delta(\xi_x^\mu - 1) + \delta(\xi_x^\mu + 1)] \tag{24}$$

it then follows $\sigma_{xy}^2(\xi) \simeq \alpha^2$. A first conclusion is that the effect of fluctuations depends importantly on the degree of correlation between stored patterns, increasing as the patterns tend to become independent from one another. Furthermore, fluctuations increase with the number α of stored patterns relative to the size of the system. This indicates a great interest for studying cases in which correlations occur between the fluctuations at different places [13].

Let us consider in detail the influence of fluctuations on the steady state. For rate V and (22), one may prove [19, 20] that the stationary solution of (15) is

$$P_{\text{st}}(s) \propto \exp(-\beta \mathcal{H}_{\text{eff}}) \tag{25}$$

with

$$\mathcal{H}_{\text{eff}}(\mathcal{J}, s) = -\tfrac{1}{2} \sum_{\substack{x,y \\ x \neq y}} \mathcal{J}_{xy} s_x s_y \tag{26}$$

where $\mathcal{J} = \{\mathcal{J}_{xy}\}$ is a set of *effective* coupling intensities given by

$$\mathcal{J}_{xy} = \frac{1}{2\beta} \ln \frac{\overline{\exp(\beta J_{xy})}}{\overline{\exp(-\beta J_{xy})}}. \tag{27}$$

For (19) with $a_\mu = P^{-1}$ so that (17) is guaranteed, one has

$$\mathcal{J}_{xy} = \frac{1}{2\beta} \ln \left\{ \frac{1 + \rho_{xy} \tanh(\beta\alpha)}{1 - \rho_{xy} \tanh(\beta\alpha)} \right\} \tag{28}$$

with $\rho_{xy} \equiv \frac{1}{P} \sum_\mu \xi_x^\mu \xi_y^\mu$. If, in addition to being equiprobable, the pattern elements are statistically independent as in (24), one expects small values for ρ_{xy} in the limit $N, P \to \infty$, and it follows

$$\mathcal{J}_{xy} \approx \frac{\mathcal{A}(\alpha)}{N} \sum_{\mu=1}^P \xi_x^\mu \xi_y^\mu \qquad \mathcal{A}(\alpha) = (\beta\alpha)^{-1} \tanh(\beta\alpha). \tag{29}$$

That is, the effective intensities are then in accordance with the Hebbian rule except for the factor $\mathcal{A}(\alpha)$ that contains the relevant information about fluctuations.

The resulting partition function is

$$Z = \sum_{\{s\}} \exp[-\beta \mathcal{H}_{\text{eff}}(s)] = \sum_{\{s\}} \exp[-\beta \mathcal{A}(\alpha) \mathcal{H}_H(s)] \tag{30}$$

where $\mathcal{H}_H(s)$ stands for the Hopfield Hamiltonian. Therefore, under the (very special) conditions stated, the system may be characterized by an *effective temperature*, namely,

$$\tilde{T} \equiv \frac{T}{\mathcal{A}(\alpha)} = \frac{\alpha}{\tanh(\alpha T^{-1})}. \tag{31}$$

In general, $\mathcal{A}(\alpha) < 1$. That is, allowing for fluctuations of the synapse intensities, in a sense amounts to adding an extra noise to the thermal one thus inducing a larger (effective) temperature. One has $\mathcal{A}(\alpha) \to 1$ as $\alpha T^{-1} \to 0$, so that the effective temperature tends to the ordinary one, $\tilde{T} \to T$, either for $\alpha \neq 0$ as $T \to \infty$ (the excess noise becomes irrelevant) or for $T \neq 0$ as $\alpha \to 0$ (the number of stored patterns remains finite in the thermodynamic limit); in both cases, the model, for which the excess noise is proportional to α^2, as shown above, reduces to the Hopfield one. Also noticeable is the fact that $\tilde{T} = \alpha$ for $T = 0$. It is to be remarked that, contrary to the effect of thermal noise, synaptic fluctuations in our model do not simply add up to the existing noise. In fact, even in its present simple version, fluctuations modify the effective strength of synaptic couplings, as demonstrated in (26) and (27).

4. Asymmetric synapses

Certain biological synapses are known to be consistent with the symmetry $J_{xy} = J_{yx}$; however, $J_{xy} \neq J_{yx}$ is thought to correspond better to many realistic situations [4]. Therefore, investigating the implications of such lack of symmetry is interesting. Motivated by experiments [3], and seeking a simple enough case that allows analytical results, let us assume that asymmetry originates during the learning process. A rather general prescription consistent with this is

$$J_{xy} = \frac{1}{N} \sum_{\mu=1}^{P} [A_1 \xi_x^\mu \xi_y^\mu + A_2 \xi_x^\mu + A_3 \xi_y^\mu + A_4] \tag{32}$$

for which no Hamiltonian description exists. In this section we discuss the possible extension of the previous results to deal with such asymmetric couplings (cf section 6 for a further study of asymmetry).

For asymmetric synapses, the steady state happens to have the canonical structure in (25) for rate V, some choices of $f(J)$ that we accomplish with the factorization property (22), and

$$\bar{\varpi}(s; x) = \left[\prod_{y \neq x} \mathcal{N}_{xy} \right] \exp[-\beta \Delta \mathcal{H}_{\text{eff}}^x(\mathcal{J}, s)]. \tag{33}$$

The first factor here is for normalization purposes, and $\Delta \mathcal{H}_{\text{eff}}^x(\mathcal{J}, s) = \mathcal{H}_{\text{eff}}(\mathcal{J}, s^x) - \mathcal{H}_{\text{eff}}(\mathcal{J}, s)$. For simplicity, we require $\mathcal{H}_{\text{eff}}(\mathcal{J}, s)$ to be of the Ising type, (26).

Asymmetry of synapses implies that J_{xy} and J_{yx} are independent variables. Under this condition, one obtains that (27) still holds for the coefficients $\mathcal{J}_{xy}$, and that

$$\mathcal{N}_{xy} = \{ \overline{\exp(\beta J_{xy})} \ \overline{\exp(-\beta J_{xy})} \}^{\frac{1}{2}}. \tag{34}$$

This follows by algebra after combining (11), (15), (25)–(27), (32) and (33); cf the more explicit derivation in section 3. A key difference with the symmetric case is that the choice for $f(J_{xy})$ in (22) needs to be restricted in order to accomplish with $J_{xy} \neq J_{yx}$. The simplest situation occurs when the latter holds whilst the effective synapses are symmetric, $\mathcal{J}_{xy} = \mathcal{J}_{yx}$, which guaranties that (26) is the effective Hamiltonian. With this in mind, we choose

$$f(J_{xy}) = \sum_{\mu=1}^{P} a_\mu \delta(J_{xy} - \eta_{xy}^\mu) \tag{35}$$

with $\eta_{xy}^\mu \neq \eta_{yx}^\mu$. Assuming here $a_\mu = P^{-1} \ \forall \mu$, and

$$\eta_{xy}^\mu = \alpha(A_1 \xi_x^\mu \xi_y^\mu + A_2 \xi_x^\mu + A_3 \xi_y^\mu + A_4) \qquad \alpha = PN^{-1} \tag{36}$$

one obtains fluctuations around mean values given by the learning rule (32).

We obtain

$$\overline{\exp(\pm \beta J_{xy})} = \exp(\pm \mathcal{A}_4) \left[\prod_{i=1}^{3} \cosh \mathcal{A}_i \right] \sum_{j=1}^{4} \lambda_\pm^j \kappa_{xy}^j \tag{37}$$

where $\mathcal{A}_i = \beta \alpha A_i$, $i = 1, 2, 3, 4$, and

$$\kappa_{xy}^1 = \sum_{\mu=1}^{P} a_\mu \xi_x^\mu \xi_y^\mu \qquad \kappa_{xy}^4 = \sum_{\mu=1}^{P} a_\mu = 1$$

$$\kappa_{xy}^2 = \sum_{\mu=1}^{P} a_\mu \xi_x^\mu \qquad \kappa_{xy}^3 = \sum_{\mu=1}^{P} a_\mu \xi_y^\mu. \tag{38}$$

Therefore, the coefficients $\mathcal{J}_{xy}$ in (26) as given by (27) are

$$\mathcal{J}_{xy} = \frac{A_4}{\beta} + \frac{1}{2\beta} \ln\left[\frac{\lambda_+^4}{\lambda_-^4}\mathcal{K}_{xy}\right] \tag{39}$$

where

$$\mathcal{K}_{xy} = \left(\sum_{i=1}^{4}\theta_+^i\kappa_{xy}^i\right)\left(\sum_{i=1}^{4}\theta_-^i\kappa_{xy}^i\right)^{-1} \tag{40}$$

$\theta_\pm^i = \lambda_\pm^i/\lambda_\pm^4$, and

$$\lambda_\pm^1 = \pm\tanh\mathcal{A}_1 + \tanh\mathcal{A}_2\tanh\mathcal{A}_3$$
$$\lambda_\pm^2 = \pm\tanh\mathcal{A}_2 + \tanh\mathcal{A}_1\tanh\mathcal{A}_3$$
$$\lambda_\pm^3 = \pm\tanh\mathcal{A}_3 + \tanh\mathcal{A}_2\tanh\mathcal{A}_1 \tag{41}$$
$$\lambda_\pm^4 = 1 \pm \prod_{i=1}^{3}\tanh\mathcal{A}_i.$$

In order to have $\mathcal{J}_{xy} = \mathcal{J}_{yx}$, it is to be required that $\mathcal{K}_{xy} = \mathcal{K}_{yx}$, which depends on the symmetry of κ_{xy}^1, namely, $\kappa_{xy}^1 = \kappa_{yx}^1$; $\kappa_{xy}^4 = \kappa_{yx}^4$; $\kappa_{xy}^2 = \kappa_{yx}^3$; $\kappa_{yx}^2 = \kappa_{xy}^3$. It turns out to be convenient to distinguish the following cases (note that the symmetry in (32) is controlled by parameters A_2 and A_3).

(i) $\theta_+^2 = \theta_+^3$ and $\theta_-^2 = \theta_-^3$ and, therefore, $\mathcal{A}_2 = \mathcal{A}_3$. Two possibilities arise: (a) $\mathcal{A}_2 = \mathcal{A}_3 = 0$; this occurs either for $\alpha, \beta \neq 0$ if $A_2 = A_3 = 0$ (symmetry), or else for $\alpha = 0$, i.e. P is finite in the thermodynamic limit. In the latter case one has $\mathcal{J}_{xy} \sim \mathcal{O}(1/N)$, while in the former case, (39) reduces to

$$\mathcal{J}_{xy} = \alpha A_4 + \frac{1}{2\beta}\ln\left[\frac{1 + \rho_{xy}\tanh(\beta\alpha A_1)}{1 - \rho_{xy}\tanh(\beta\alpha A_1)}\right] \tag{42}$$

which is similar to (28) for symmetric synapses. (b) $\mathcal{A}_2 = \mathcal{A}_3 \neq 0$, which requires $A_2 = A_3 \neq 0$, i.e. η_{xy}^μ is symmetric, namely, $\alpha^{-1}\eta_{xy}^\mu = A_1\xi_x^\mu\xi_y^\mu + A_2(\xi_x^\mu + \xi_y^\mu) + A_4$. One obtains

$$\mathcal{J}_{xy} = \alpha A_4 + \frac{1}{2\beta}\ln\left[\frac{1 + \prod_i\tanh(\beta\alpha A_i)}{1 - \prod_i\tanh(\beta\alpha A_i)}\right] + \frac{1}{2\beta}\ln\left[\frac{1 + \rho_{xy}\theta_+^1 + (\eta_x + \eta_y)\theta_+^2}{1 + \rho_{xy}\theta_-^1 + (\eta_x + \eta_y)\theta_-^2}\right] \tag{43}$$

$i = 1, 2, 3$, with $\eta_x \equiv P^{-1}\sum_\mu\xi_x^\mu$.

(ii) $\mathcal{A}_2 \neq \mathcal{A}_3$, i.e. the situation is originally asymmetric, and $\alpha \neq 0$, so that a macroscopic number of memories, $P = \alpha N$, is stored. Then one needs to introduce conditions on the stored patterns in order to have $\mathcal{K}_{xy} = \mathcal{K}_{yx}$. For example, one may assume $\eta_x = \eta_y = \eta\forall x, y$, which leads to

$$\mathcal{J}_{xy} = \alpha A_4 + \frac{1}{2\beta}\ln\left[\frac{1 + \prod_i\tanh(\beta\alpha A_i)}{1 - \prod_i\tanh(\beta\alpha A_i)}\right] + \frac{1}{2\beta}\ln\left[\frac{1 + \rho_{xy}\theta_+^1 + \eta(\theta_+^2 + \theta_+^3)}{1 + \rho_{xy}\theta_-^1 + \eta(\theta_-^2 + \theta_-^3)}\right]. \tag{44}$$

This is the simplest case with intrinsic asymmetry of synapses that may be described by an effective Hamiltonian of the Hopfield type. Again, this result indicates that a main effect of fluctuations is to induce effective values for the synaptic intensities. Interesting enough, the above shows that symmetric effective synapses may result even for $J_{xy} \neq J_{yx}$. To analyse the implications of (39) in detail is beyond our objectives in this paper; however, we use this result in section 6. Other, more general cases may be worked out under a mean-field hypothesis, for example, as shown below.

5. The replica method

The replica trick may be used to study further the implications of exact results in the previous sections. We are concerned here with the symmetric case (the case of asymmetric couplings is considered in section 6); in particular, we wish to solve the partition function (30). The latter has been computed for $\mathcal{A}(\alpha) \equiv 1$ within a mean-field approximation using the replica trick and the saddle-point method [21–23]. After replacing β by $\tilde{\beta} \equiv \beta \mathcal{A}(\alpha)$ in the known solution, one has for the free energy per spin:

$$f = \frac{\alpha}{2} + \frac{\alpha}{2\tilde{\beta}n} \operatorname{Tr} \ln[(1-\tilde{\beta})\boldsymbol{I} - \tilde{\beta}\boldsymbol{q}] + \frac{1}{2n} \sum_{\rho=1}^{n} \left[\sum_{\nu=1}^{k} (m_\rho^\nu)^2 + \alpha\tilde{\beta} \sum_{\substack{\sigma=1 \\ \sigma \neq \rho}}^{n} r_{\rho\sigma} q_{\rho\sigma} \right]$$

$$- \frac{1}{n\tilde{\beta}} \langle \ln \operatorname{Tr}_{s} e^{\tilde{\beta} H_\xi(s)} \rangle_\xi \tag{45}$$

as $n \to 0$, where ρ and σ describe the n replicas, $\boldsymbol{s} \equiv \{s^1, s^2, \ldots, s^n\}$,

$$H_\xi(\boldsymbol{s}) \equiv \sum_{\rho=1}^{n} \left[\frac{\alpha\tilde{\beta}}{2} \sum_{\substack{\sigma=1 \\ \sigma \neq \rho}}^{n} r_{\rho\sigma} s^\rho s^\sigma + \sum_{\nu=1}^{k} (m_\rho^\nu + h^\nu) \xi^\nu s^\rho \right] \tag{46}$$

and $\langle \cdots \rangle_\xi$ is an average over the distribution of random patterns, (24); $\boldsymbol{I}$ is the identity matrix of elements $\delta_{\rho,\sigma}$, and $\boldsymbol{q}$ has elements $q_{\rho\sigma}$ for $\rho \neq \sigma$ and zeros for $\rho = \sigma$. The physical meaning of the parameters m_ρ^ν, $q_{\rho\sigma}$ and $r_{\rho\sigma}$ in (45) is given by saddle-point equations. That is, one obtains

$$m_\rho^\nu = \frac{1}{N} \left\langle \sum_{y} \xi_y^\nu \langle s_y^\rho \rangle \right\rangle_\xi \tag{47}$$

as a measure of the overlap of a given state with a given stored pattern, and

$$q_{\rho\sigma} = \frac{1}{N} \left\langle \sum_{y} \langle s_y^\rho \rangle \langle s_y^\sigma \rangle \right\rangle_\xi \qquad r_{\rho\sigma} = \frac{1}{\alpha} \sum_{\mu=k+1}^{P} \langle m_\rho^\mu m_\sigma^\mu \rangle_\xi \tag{48}$$

which are the so-called Edwards–Anderson [14] and AGS [22] parameters, respectively. The index ν in these expressions describes a (finite) number, k, of *condensed patterns*, defined as those for which the overlap m^ν is nonvanishing in the thermodynamic limit. The *non-condensed* patterns, described by the index μ, for example in (48) and in the expression for $p(\xi_x^\mu)$, have overlap $\mathcal{O}(\frac{1}{\sqrt{N}})$. We next study some of the consequences of (45).

Consider replica symmetry, namely, $m_\rho^\nu = m^\nu$, and $q_{\rho\sigma} = q$ and $r_{\rho\sigma} = r$ for $\rho \neq \sigma$. It follows from (45) as $n \to 0$ that

$$m^\nu = \left[\left[\xi^\nu \tanh \tilde{\beta} \left\{ z\sqrt{\alpha r} + (\boldsymbol{m} + \boldsymbol{h}) \cdot \boldsymbol{\xi} \right\} \right]\right] \tag{49}$$

$$q = \left[\left[\tanh^2 \tilde{\beta} \left\{ z\sqrt{\alpha r} + (\boldsymbol{m} + \boldsymbol{h}) \cdot \boldsymbol{\xi} \right\} \right]\right] \qquad r = q(1 - \tilde{\beta} + \tilde{\beta}q)^{-2}. \tag{50}$$

Here $\boldsymbol{m} \equiv \{m^\nu; \nu = 1, \ldots, k\}$ and $\boldsymbol{h} \equiv \{h^\nu; \nu = 1, \ldots, k\}$ are the corresponding conjugate fields that appear as an extra term, $-\sum_{\nu=1}^{k} h^\nu \sum_{x} \xi_x^\nu s_x$, in the Hamiltonian (26), and

$$[[\cdots]] \equiv \frac{1}{\sqrt{2\pi}} \int \mathrm{d}z \, e^{-z^2/2} \langle \cdots \rangle_{\xi^\nu}. \tag{51}$$

The last average here is similar to the one in (45) but over the condensed patterns only; the integral shown explicitly corresponds to the Gaussian noise whose origin is the set of non-condensed patterns.

The following are principal implications of equations (49) and (50) for $h^\nu = 0 \forall \nu$. For $T = 0$, one has that $\tilde{T} = \alpha$. In this case, one finds a spin-glass solution, i.e. $m^\nu = 0$ with $q \neq 0$ for any $\alpha < \alpha_q = 2.618$; the resulting dependence of q on α is shown in figure 1. This essentially differs from the result $q = 1$, independent of the number of stored memories, for the Hopfield model [22]. That is, synapse fluctuations tend to impede the spin-glass phase, and the associated noise is large enough when the number of stored memories exceeds about $3N$ the number of sites so that the spin-glass, $q \neq 0$ solution is impeded. In a sense, this is an interesting quality of the present version of the model, because the spin-glass phase is undesirable for its performance, i.e. for the efficiency of processes such as storing and retrieving memories; the improvement would be real if one could simultaneously extend the Mattis region closer to α_q (cf below). As expected, the fluctuations vanish as $\alpha \to 0$, and we then recover the Hopfield result.

Mattis solutions [24],

$$m^\nu = m \delta_{\nu, \nu_0} \tag{52}$$

i.e. those that exhibit a unique non-vanishing overlap, occur (twice for each memory) at $T = 0$ for $\alpha < \alpha_M \simeq 0.132$, slightly smaller than the Hopfield value of 0.138. That is, fluctuations decrease the efficiency of associative memory somewhat; figure 2 includes a comparison. As $\alpha \to 0$, both systems recover without error, i.e. $m \to 1$. The error increases with α, being slightly larger in the presence of fluctuations. Differences between the two models are now smaller than for the spin-glass phase, because the Mattis solutions occur for smaller values of α and the two models tend to each other as $\alpha \to 0$. The phase transition is discontinuous, as suggested in figure 2. One has $q \neq 0$ for the Mattis solution, i.e. it involves spin-glass order. The stability of solutions is similar to that for the Hopfield model, namely, the ground state only consists of Mattis states for $\alpha < \alpha_C \simeq 0.052$, and it consists of Mattis states and spin-glass states for $\alpha_M > \alpha > \alpha_C$ (while only spin-glass states occur for $\alpha > \alpha_M$). The value of α_C is practically the same for the two models, because fluctuations are relatively small for such small values of α.

For $T \neq 0$, this version of the model corresponds to the Hopfield one with effective temperature (31). The spin-glass phase occurs for $T < T_{sg}(\alpha)$ with

$$T_{sg}^{-1}(\alpha) = \begin{cases} \dfrac{1}{2\alpha} \ln \left\{ \dfrac{1 - \alpha\sqrt{\alpha}}{1 - 2\alpha + \alpha\sqrt{\alpha}} \right\} & \text{for } \alpha \neq 1 \\ \tanh^{-1} \frac{1}{2} & \text{for } \alpha = 1. \end{cases} \tag{53}$$

As $\alpha \to 0$, this transforms into the Hopfield result, $T_{sg}(\alpha) = 1 + \sqrt{\alpha}$. The phase transition is of second order, i.e. q is continuous at T_{sg} for any α; cf figure 3. In addition to this transition, we find two more transition lines (cf figure 4): one is $T_M(\alpha)$ such that no Mattis states occur for $T > T_M(\alpha)$; the other is $T_C(\alpha)$ $(< T_M(\alpha) \forall \alpha)$ such that the Mattis states do not correspond to global minimum of the free energy. The phase transitions at $T_M(\alpha)$ and $T_C(\alpha)$ are discontinuous for any α; these lines are very close to the corresponding ones for the Hopfield model, given that they occur for small α, when fluctuations are small.

Therefore, the phase diagram for the present version of our model in the case of replica symmetry (cf figure 3) is qualitatively similar to the one for the Hopfield model as long as α is small. However, some important differences occur for large α when synapse fluctuations are significant. In particular, it follows that the Hopfield spin-glass states are strongly influenced by fluctuations, and do not occur for large enough α.

The solution with replica symmetry is, in general, unstable. This is reflected in the Sherrington–Kirkpatrick model by the fact that its entropy is negative for $T = 0$ [21]. The Hopfield model has a parallel behaviour, i.e. the spin-glass solution with $q_{\rho,\sigma} = q$
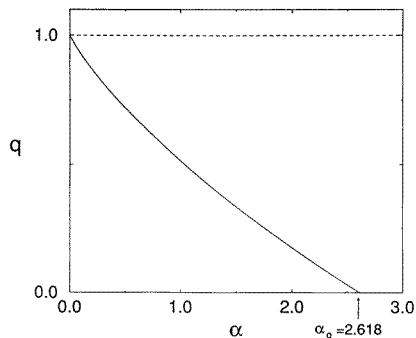
Figure 1. Comparison of the behaviour of the model presented here with the familiar Hopfield case. The full curve represents the order parameter $q = q(\alpha)$ obtained as the solution of equations (49) and (50) for $m^\nu = 0$ and $h^\nu = 0 \quad \forall \nu$, and $T = 0 \; (\tilde{T} = \alpha)$. The states below this line, which corresponds to a continuous phase transition, have a spin-glass, mixture nature. Unlike the Hopfield model (for which $q = 1$, represented by the broken line), q decreases with increasing α, and vanishes for $\alpha > \alpha_q = 2.618$ in the present model with fluctuating synapses.
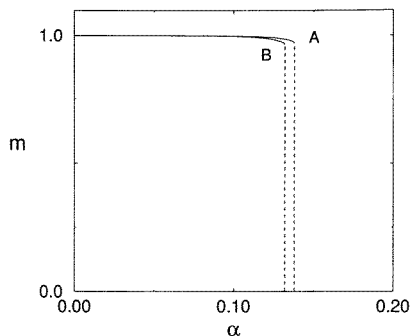

Figure 2. The overlap parameter associated to the Mattis solutions for $T = 0$ in the cases of (A) the Hopfield model, and (B) the model with fluctuating synapses.
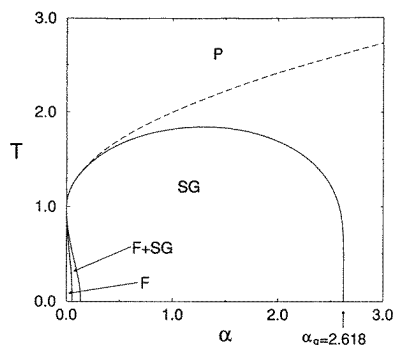

Figure 3. The phase diagram for the network defined by (30), for which an effective Hamiltonian of the Hopfield type exists, namely, (26) with (29), when replica symmetry holds. Using the magnetic language, the different regions exhibited here correspond to paramagnetic (P), spin-glass (SG), and ferromagnetic (F) phases. The upper full curve, which corresponds to equation (53), separates paramagnetic from spin-glass states; the broken curve is the corresponding Hopfield result. No spin-glass solutions occur for $\alpha > \alpha_q$ in the model with fluctuations, unlike for the Hopfield case (cf figure 1). The region near the origin is represented in more detail in figure 4.
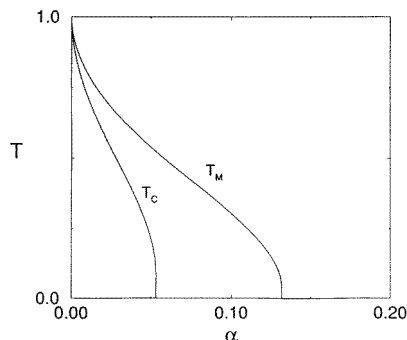

Figure 4. Detail near the origin of the phases in figure 3 to illustrate the functions $T_M(\alpha)$ and $T_C(\alpha)$, as defined in the main text, corresponding to discontinuous phase transitions.

for $\rho \neq \sigma$ is unstable for $T < T_{sg} = 1 + \sqrt{\alpha}$, and the zero-$T$ entropy associated with either spin-glass or Mattis solutions is negative [22]. A principal question is the stability of the ferromagnetic solution with replica symmetry, which is the most relevant one for

associative memory. The limit of stability of the ferromagnetic solution is known as the AT line [25]. For the Hopfield model [22], the AT line separates stable states of the solution with replica symmetry from a region in which stability breaks down. The latter is relatively small, occurs at low temperature near $\alpha_M (= 0.138)$, and its extension decreases quickly as $\alpha \to 0$, so that it does not significantly affect the associative memory properties. In the presence of synapse fluctuations, the AT line does not occur, even for $T = 0$, because of the excess noise that increases with α. The entropy for the solutions of (49) and (50) of the Mattis type is

$$S(\beta, q, m) = \beta^2(\tilde{\beta}^{-2} - \alpha^2)S_H(\tilde{\beta}, q, m) \tag{54}$$

where $S_H(\tilde{\beta}, q, m)$ is the Hopfield result with β replaced by $\tilde{\beta}$. This expression goes to zero as $\beta \to \infty$.

6. Further cases of the model

In this section we report on mean-field time evolution equations that provide a detailed description of several phase transitions. The focus is on simple versions of the model in section 2, in the limit $p \to 0$, for functions $\varpi_J(s; x)$ and $f(J)$ that, in general, do not allow for the canonical structure (25), (26). One obtains from (15) for the time variation of the mean neuron activity, $\partial_t \langle s_x \rangle = -2\langle s_x \varpi(s; x) \rangle$, that

$$\partial_t \langle s_x \rangle = -2\langle s_x \bar{a}_x(s) \rangle - 2\langle \bar{b}_x(s) \rangle \tag{55}$$

where $\bar{a}_x(s)$ and $\bar{b}_x(s)$ stand for the average with respect to $f(J)$, as in (17), of

$$a_x(s, J) \equiv \tfrac{1}{2}\{\Psi[2\beta h_x(s, J)] + \Psi[-2\beta h_x(s, J)]\} \tag{56}$$

and

$$b_x(s, J) \equiv \tfrac{1}{2}\{\Psi[2\beta h_x(s, J)] - \Psi[-2\beta h_x(s, J)]\} \tag{57}$$

respectively. (Note a dependence on x made explicit here that originates on the spatial dependence exhibited by memory patterns; cf (8) and (1).) Under the mean-field approximation $s_x \approx \langle s_x \rangle$, (55) transforms into

$$\partial_t \langle s_x \rangle \approx -2\bar{a}'_x \langle s_x \rangle - 2\bar{b}'_x \tag{58}$$

where $\bar{a}'_x$ and $\bar{b}'_x$ stand, respectively, for $\bar{a}_x(s)$ and $\bar{b}_x(s)$ with $s_x \approx \langle s_x \rangle$. The corresponding stationary solution is $m_x(J) = \sum_{y \neq x} J_{xy} \langle s_y \rangle$ with

$$\langle s_x \rangle \approx -\frac{\bar{b}'_x}{\bar{a}'_x} = \frac{\overline{\Psi[-2\beta m_x(J)] - \Psi[2\beta m_x(J)]}}{\overline{\Psi[-2\beta m_x(J)] + \Psi[2\beta m_x(J)]}} \tag{59}$$

where we have used (56), (57) and the fact that one has $h_x(s, J) \approx m_x(J)$ under the mean-field hypothesis.

Note some general features before one specifies the rate in (59). The latter is satisfied by $m_x(J) = 0$; therefore, non-zero solutions follow by developing around this trivial one in powers of $m_x(J)$. It leads to first order to $m_x(J) \approx \beta \sum_{y \neq x} J_{xy} \overline{m_y(J)}$, or

$$\overline{m_x(J)} \approx \beta \sum_{y \neq x} \overline{J_{xy} m_y(J)} \tag{60}$$

after averaging with respect to $f(J)$. Therefore, there is a phase transition, namely, $\overline{m_x(J)} \neq 0$, for

$$\det\{\delta_{x,y} - \beta \bar{J}_{xy}\} = 0. \tag{61}$$

The following are some illustrative cases.

(i) $f(\mathbf{J})$ is such that $\bar{J}_{xy} = J/N$ with J independent of $N \forall x, y$. One has that $\overline{m_x(\mathbf{J})} = Jm \forall x$, where $m \equiv N^{-1} \sum_x \langle s_x \rangle$, and (60) gives $m \approx \beta Jm$. That is, the system exhibits a phase transition from ferromagnetic-like to paramagnetic-like behaviour at $T_c \approx J$.

(ii) $f(\mathbf{J})$ is such that $\bar{J}_{xy}$ is given by (17) with $\{\xi_x^\mu; \mu = 1, \ldots, P\}$ a set of $N \times P$ statistically independent random variables with probabilities (24). After using detailed balance for $\Psi(X)$ and developing the latter around $X = 0$, one has from (59) to first order in X that

$$\langle s_x \rangle \approx \beta \alpha \sum_{y \neq x} \left(P^{-1} \sum_\mu \xi_x^\mu \xi_y^\mu \right) \langle s_y \rangle. \tag{62}$$

For $N, P \gg 1$, the limit central theorem reduces this to

$$\sum_{y \neq x} \langle s_y \rangle \delta_{xy} (1 - \alpha \beta) \approx 0. \tag{63}$$

This is satisfied by $\langle s_x \rangle = 0 \forall x$, i.e. $m = 0$; non-trivial solutions occur for $T < T_c$, where the critical temperature follows from $\det[\delta_{xy}(1 - \alpha \beta)] = 0$ as $T_c = \alpha$. This corresponds to a phase transition of second order from paramagnetic- to ferromagnetic-like behaviour. For the same $f(\mathbf{J})$, we have studied the order parameter

$$m^\mu = \frac{1}{N} \sum_y \xi_y^\mu \langle s_y \rangle \tag{64}$$

which describes the overlap with a given pattern $\boldsymbol{\xi}^\mu$. One has from (59) that $m^\mu = -N^{-1} \sum_y \xi_y^\mu (\bar{b}_x'/\bar{a}_x')$, and it follows to first order that

$$\sum_{\mu=1}^P m^\mu \left[\delta_{\mu,\nu} - \frac{\beta}{N} \sum_y \xi_y^\mu \xi_y^\nu \right] \approx 0 \tag{65}$$

after developing $\Psi(X)$. The trivial solution, for which the system cannot retrieve any of the stored patterns, is $m^\mu = 0 \forall \mu$. Non-trivial solutions which mark a phase transition to states that exhibit associative memory may only occur below a critical temperature that satisfies $\det\{\delta_{\mu,\nu} - \beta N^{-1} \sum_y \xi_y^\mu \xi_y^\nu\} = 0$. For $N \gg 1$, one has $N^{-1} \sum_y \xi_y^\mu \xi_y^\nu \approx \delta_{\mu,\nu}$ leading to $T_c = 1$.

We next consider (55) for different rates when $f(\mathbf{J})$ is given by (19) and (20). For rate K, we obtain

$$\bar{a}_x(s) = 1 \qquad \bar{b}_x(s) = -\overline{\tanh[\beta h_x(s, \mathbf{J})]} \tag{66}$$

and, consequently, $\langle s_x \rangle = \langle \overline{\tanh[\beta h_x(s, \mathbf{J})]} \rangle$. The latter reads

$$\langle s_x \rangle = \left\langle \sum_{\mu_1=1}^P \ldots \sum_{\mu_{N-1}=1}^P a_{\mu_1} \ldots a_{\mu_{N-1}} \tanh[\beta E] \right\rangle \tag{67}$$

where $E \equiv \eta_{xy_1}^{\mu_1} s_{y_1} + \cdots + \eta_{xy_{N-1}}^{\mu_{N-1}} s_{y_{N-1}}$. For rate M, one has

$$\bar{a}_x(s) = \tfrac{1}{2}[1 + \exp(-2\beta|h_x(s, \mathbf{J})|)] \tag{68}$$

and

$$\bar{b}_x(s) = \tfrac{1}{2} \operatorname{sign}(h_x(s, \mathbf{J}))(1 - e^{-2\beta|h_x(s, \mathbf{J})|})$$

and it follows that

$$\langle s_x \rangle = -B/A \tag{69}$$

where

$$A = \frac{1}{2}\left[1 + \frac{1}{P^{N-1}} \sum_{\mu_1=1}^{P} \cdots \sum_{\mu_{N-1}=1}^{P} e^{-2\beta|\tilde{E}|}\right] \tag{70}$$

and

$$B = \frac{1}{2P^{N-1}} \sum_{\mu_1=1}^{P} \cdots \sum_{\mu_{N-1}=1}^{P} (1 - e^{-2\beta|E|}) \operatorname{sign} \tilde{E} \tag{71}$$

with $\tilde{E} \equiv \eta_{xy_1}^{\mu_1} \langle s_{y_1} \rangle + \cdots + \eta_{xy_{N-1}}^{\mu_{N-1}} \langle s_{y_{N-1}} \rangle$. Cases (67) and (69) would require some drastic simplification before one obtains explicit results.

The case rate V may be worked out explicitly. Consider first the symmetric choice (19) so that (17) holds. One obtains

$$\bar{a}_x(s) = K_x \cosh\left(\beta \sum_{y \neq x} \mathcal{J}_{xy} s_y\right) \qquad \bar{b}_x(s) = -K_x \sinh\left(\beta \sum_{y \neq x} \mathcal{J}_{xy} s_y\right) \tag{72}$$

where

$$K_x = \prod_{y \neq x} [\cosh^2(\beta\alpha) - \rho_{xy}^2 \sinh^2(\beta\alpha)]^{\frac{1}{2}} \tag{73}$$

and

$$\mathcal{J}_{xy} = \frac{1}{2\beta} \ln\left[\frac{1 + \rho_{xy} \tanh(\beta\alpha)}{1 - \rho_{xy} \tanh(\beta\alpha)}\right] \tag{74}$$

with $\rho_{xy} \equiv P^{-1} \sum_{\mu=1}^{P} \xi_x^\mu \xi_y^\mu$, so that $\mathcal{J}_{xy} = \mathcal{J}_{yx}$. Under the mean-field approximation $s_x \approx \langle s_x \rangle$, the corresponding stationary solution is

$$\langle s_x \rangle = \tanh\left(\beta \sum_{y \neq x} \mathcal{J}_{xy} \langle s_y \rangle\right) \tag{75}$$

and it follows for the overlap that

$$m^\mu = \frac{1}{N} \sum_x \xi_x^\mu \tanh\left(\beta \sum_{y \neq x} \mathcal{J}_{xy} \langle s_y \rangle\right). \tag{76}$$

This is the Hopfield result for finite P with $\mathcal{J}_{xy}$ replacing the Hebb value (1). For P infinite, one may still obtain (49) and (50) heuristically from (76) assuming that $\mathcal{J}_{xy}$ is given by (29) for $N, P \to \infty$. For the asymmetric choice (36), one has after some algebra that

$$\bar{a}_x(s) = \left[\prod_{i=1}^{3} \cosh(\beta\alpha A_i)\right]\left(\prod_{y \neq x} c_{xy}\right) \cosh\left(\beta \sum_{y \neq x} \mathcal{J}_{xy} s_y\right) \tag{77}$$

and

$$\bar{b}_x(s) = -\left[\prod_{i=1}^{3} \cosh(\beta\alpha A_i)\right]\left(\prod_{y \neq x} c_{xy}\right) \sinh\left(\beta \sum_{y \neq x} \mathcal{J}_{xy} s_y\right) \tag{78}$$

with

$$c_{xy}^2 = [1 + \kappa_{xy}^1 \tanh \mathcal{A}_2 \tanh \mathcal{A}_3 + \kappa_{xy}^2 \tanh \mathcal{A}_1 \tanh \mathcal{A}_3 + \kappa_{xy}^3 \tanh \mathcal{A}_1 \tanh \mathcal{A}_2]^2$$

$$- \left[\prod_{i=1}^{3} \tanh \mathcal{A}_i + \kappa_{xy}^1 \tanh \mathcal{A}_1 + \kappa_{xy}^2 \tanh \mathcal{A}_2 + \kappa_{xy}^3 \tanh \mathcal{A}_3\right]^2$$

cf section 4. Under the mean-field approximation $s_x \approx \langle s_x \rangle$, the steady state turns out to be described by the same equations as for the symmetric case, namely, (75) and (76), with $\mathcal{J}_{xy}$ given by (39). The detailed study of the implications of these equations is beyond our objectives in this paper. It is to be remarked that these equations generalize our result at the end of section 4 to the case in which the stored patterns do not satisfy the (restrictive) condition $\eta_x = \eta_y = \eta \forall x, y$; the latter was introduced there to obtain a canonical structure which is not required here.

7. Conclusion

As an extension of previous work on disordered systems [16, 20, 26], we have presented and studied a model for associative memory. This is a generalized, kinetic version of the Hopfield neural network in the sense that the synapse intensities do not remain constant after the learning process but fluctuate with time during neuron activity. Synaptic fluctuations have been reported to occur rather generally in biological systems, and they are assumed to play a role essentially different from that of thermal noise, and to influence (in a way to be determined) on the system performance. In this paper, we have considered explicitly the case of fast fluctuations around mean values that are determined by the Hebbian learning rule. As a consequence, the stationary solution of the master equation that characterizes the system is determined by a complex transition-rate function that enters this equation. (In fact, due to fluctuations, the steady state is, in general, a non-equilibrium one, as if the system was acted on by non-Hamiltonian agents.) We have shown that this may induce interesting effects on the learning, recall and other associative-memory processes.

For certain choices of the rate function, the model reduces to familiar cases such as the Ising and Hopfield models, for which the synapse intensities do not fluctuate, corresponding to $p \equiv 1$ in equation (3). In the limit of very fast fluctuations, $p \to 0$, synapse intensities become random variables of distribution $f(\boldsymbol{J})$. Different choices for the latter allow for the consideration of symmetric, $J_{xy} = J_{yx}$, and asymmetric, $J_{xy} \neq J_{yx}$, couplings. Anyhow the steady state strongly depends on both $f(\boldsymbol{J})$ and the rate. We have studied some general properties of the model, and have presented here the analytical solution for some choices of the parameters within a mean-field type of theory. In particular, we have considered situations in which each synapse is associated to one of the elements of each memory pattern, for example (19) where the matrix η_{xy}^μ may be either symmetric or asymmetric. We present closed equations for order parameters for rate V—cf equation (9)—if $\eta_{xy}^\mu = \eta_{yx}^\mu$. The system in this case (but not more generally, e.g. rates K and M) may be described by an effective Hamiltonian of the Hopfield type, and synaptic fluctuations are shown to modify the original temperature, T, so that there is an effective temperature, $\tilde{T} > T$; more generally, fluctuations imply different values for the synaptic strengths. We have applied the replica trick to obtain the resulting explicit behaviour.

A principal effect of synapse fluctuations is that the critical value α_C for the ratio between the number of stored patterns and the size of the system, $\alpha \equiv P/N$, below which there is efficient associative memory is reduced by some 4%. It is remarkable that the solution with replica symmetry remains stable in our model near α_C, due to the fact that $\tilde{T} > T$, in contrast to the Hopfield case where the AT line excludes stability in a (small) region of the phase diagram sufficiently close to α_C and $T = 0$. On the other hand, fluctuations notably reduce the region of the phase diagram where associative memory might be hampered by the occurrence of spin-glass states. Unlike for the Hopfield model, this region is finite in our case; there is spin-glass phase only below $T_{\mathrm{sg}}(\alpha)$, which becomes zero (so that no spin-glass states may occur) for $\alpha > 2.618$. Closed equations for the steady

values of the order parameters have been found for rate V if $\eta_{xy}^{\mu} = \eta_{yx}^{\mu}$. It then follows an effective Hamiltonian under certain conditions on the stored memories; however, no such effective Hamiltonian description exists, in general, because the resulting *effective* synapses are asymmetric, which is often the actual case in nature. We show that a kinetic mean-field formalism allows for a more general approach to the case of asymmetric couplings. Further details on the performance of our system (6) will be reported elsewhere [13].

Therefore, we have shown explicitly that time variations, for example fluctuations of synaptic intensities during neuron activity, which are known to occur in biological systems, affect importantly the behaviour of a neuronal network, in general. The interest to study, by other techniques, versions of our model that are not amenable to the analytical treatments here is therefore suggested.

Acknowledgments

References

[1] Hopfield J J 1982 *Proc. Natl Acad. Sci., USA* **79** 2554
[2] Hebb D O 1949 *The Organization of Behavior: A Neurophysiological Theory* (New York: Wiley)
[3] Peretto P 1992 *An Introduction to the Modeling of Neural Networks* (Cambridge: Cambridge University Press) and references therein
[4] Noback C R and Demarest R J 1975 *The Human Nervous System: Basic Principles of Neurobiology* (New York: McGraw-Hill)
[5] Otmakhov N, Shirke A M and Malinow R 1993 *Neuron* **10** 1101
[6] Sorra K E and Harris K M 1993 *J. Neurosci.* **13** 3736
[7] Allen C and Stevens CF 1994 *Proc. Natl Acad. Sci., USA* **91** 10 380
[8] Goda Y and Stevens C F 1994 *Proc. Natl Acad. Sci., USA* **91** 12 942
[9] Stratford K J, Tarczy-Hornoch K, Martin K A C, Bannister N J and Jack J J B 1996 *Nature* **382** 258
[10] Ferster D 1996 *Science* **273** 1812
[11] Smetters D K and Zador A 1996 *Current Biology* **6** 1217
[12] Moss F 1997 *Phys. World* **10** 15
[13] Garrido P L, Marro J and Torres J J *Preprint*
[14] Edwards F and Anderson P W 1975 *J. Phys. F: Met. Phys.* **5** 965
[15] Crisanti A and Sompolinsky H 1988 *Phys. Rev.* A **37** 4865
[16] Garrido P L and Marro J 1994 *J. Stat. Phys.* **74** 663
[17] Garrido P L and Marro J 1989 *Phys. Rev. Lett.* **62** 1929
[18] López-Lacomba A I, Garrido P L and Marro J 1990 *J. Phys. A: Math. Gen.* **23** 3809
[19] Garrido P L and Muñoz M A 1993 *Phys. Rev.* E **48** R4153
[20] López-Lacomba A I and Marro J 1994 *Europhys. Lett.* **25** 169
[21] Kirkpatrick S and Sherrington D 1975 *Phys. Rev. Lett.* **35** 1972
[22] Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys., NY* **173** 30
[23] Pastur L, Shcherbina M and Tirozzi B 1994 *J. Stat. Phys.* **74** 1161
[24] Mattis D C 1976 *Phys. Lett.* **56A** 421
[25] de Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983
[26] Garrido P L and Marro J 1991 *Artificial Neural Networks (Lecture Notes in Computer Science 540)* ed A Prieto (Berlin: Springer) pp 25–33