

Stochastic learning in a neural network with adapting synapses

G. Lattanzi,^{1,*} G. Nardulli,^{1,†} G. Pasquariello,^{2,‡} and S. Stramaglia^{2,§}

¹*Dipartimento di Fisica dell'Università di Bari, Istituto Nazionale di Fisica Nucleare, Sezione di Bari, via Amendola 173, 70126 Bari, Italy*

²*Istituto Elaborazione Segnali ed Immagini, Consiglio Nazionale delle Ricerche, via Amendola 166/5, 70126 Bari, Italy*
(Received 19 May 1997)

We consider a neural network with adapting synapses whose dynamics can be analytically computed. The model is made of N neurons and each of them is connected to K input neurons chosen at random in the network. The synapses are n -state variables that evolve in time according to stochastic learning rules; a parallel stochastic dynamics is assumed for neurons. Since the network maintains the same dynamics whether it is engaged in computation or in learning new memories, a very low probability of synaptic transitions is assumed. In the limit $N \rightarrow \infty$ with K large and finite, the correlations of neurons and synapses can be neglected and the dynamics can be analytically calculated by flow equations for the macroscopic parameters of the system. [S1063-651X(97)08710-2]

PACS number(s): 87.10.+e, 05.20.-y

I. INTRODUCTION

In the dynamics of attractor neural networks (ANNs) two processes take place: the time evolution of the neuron states and the change in strength of the connections (synapses) between neurons. Most of the prior research has focused separately on the first dynamical process (retrieval of learned patterns) or the other (learning of patterns) [1]. In this work we deal with the problem of building up models where neurons and synapses maintain the same dynamics whether the network is engaged in computation or in learning new memories; the two stages (computation and learning) differ only for the presence of external stimuli. This problem is important for the description of the *short term* memory in the human brain (see, e.g., [2]) and has been studied mainly in the context of Hebbian learning with a decay term [3–5]: Shinomoto [3] presented a rule for the synaptic modification whose stable solutions are the Hopfield couplings [6]; Dong and Hopfield [4] considered analog neurons in a deterministic network described by a system of differential equations and applied their model to the problem of the development of synaptic connections in the visual cortex; D'Autilia and Guerra [5] studied an ANN with adapting synapses in order to model conditioned reflex and rhythm recognition.

A very interesting approach to the coupled dynamics of fast spins and slow interactions, in a fully connected network, has been proposed by Coolen, Penney, and Sherrington [7]; in their model the interactions are assumed to change adiabatically and the thermodynamic equilibrium corresponds to a system of averaged replicas, where the replica number represents the ratio of the temperatures of the spin and interaction systems. It is shown that the influence of the spins on the interaction dynamics can change the order of the phase transitions in the model. Moreover the case of a

finite and negative replica number (describing an overfrustrated system) has been studied by Dotsenko, Franz, and Mézard [8]. Caticha considered a spin model where the disorder evolves on a time scale intermediate between the annealed and quenched behaviors [9].

We consider here an ANN with asymmetric adapting synapses whose dynamics consists of a stochastic learning mechanism [10,11]; in particular we refer to Amit and Fusi [11] for a discussion on the possible sources of stochasticity in the synaptic learning. We present a dynamical theory in terms of deterministic flow equations for macroscopic order parameters, which holds if the connections are strongly diluted. For quenched synapses the dynamical approach is exact in the case of strong dilution [12]; it may be mentioned, in passing, that recently Coolen and Sherrington described dynamically the fully connected Hopfield model [13].

Our model is made of N neurons (Ising spins) $s_i(t) \in \{-1, 1\}$, $i=1, \dots, N$. For each neuron s_i , K input sites $j_1(i), \dots, j_K(i)$ are chosen at random among the N sites, and NK synaptic interactions $J_{ij}(t)$ are introduced. We assume that the synapses are n -state variables $J_{ij} \in \{1, (n-3)/(n-1), \dots, -1\}$. Let us call α the index identifying the different values of the synapses, so that $\alpha=1$ implies $J=1$ and $\alpha=n$ implies $J=-1$; in general

$$J_\alpha = \frac{n+1-2\alpha}{n-1}, \quad \alpha=1, \dots, n. \quad (1)$$

The dynamical rule for the evolution of the synapses is assumed as follows. Each synaptic variable J_{ij} connecting s_i with one of its K input neurons grows by one unit $2/(n-1)$, i.e., $J_\alpha \rightarrow J_{\alpha-1}$, with probability q if the product $s_i s_j$ of the spins connected by J_{ij} is positive; if such a product is negative the synapse decreases by one unit ($J_\alpha \rightarrow J_{\alpha+1}$) with the same probability. If a synapse is at one of the extreme limits and should be pushed off, its value remains unchanged.

A parallel stochastic dynamics with inverse temperature $\beta = T^{-1}$ is assumed for neurons, where the local field acting on neuron s_i is given by

*Electronic address: lattanzi@axpba4.ba.infn.it

†Electronic address: nardulli@axpba0.ba.infn.it

‡Electronic address: guido@iesi.ba.cnr.it

§Electronic address: sebino@iesi.ba.cnr.it

$$h_i(t) = \sum_j J_{ij}(t) s_j(t) + H_i(t); \quad (2)$$

the sum is taken over the K input neurons and external stimuli are represented by local magnetic fields H_i . Therefore the rules for the parallel dynamics in our model are

$$s_i(t+1) = \text{sgn}\left(\frac{1}{2} + \frac{1}{2} \tanh[\beta h_i(t)] - \eta_i\right), \quad (3)$$

$$J_{ij}(t+1) = J_{ij}(t) + \frac{2}{n-1} s_i(t) s_j(t) \theta(q - \eta_{ij}) \\ \times \theta\left(1 - \left|J_{ij}(t) + \frac{2}{n-1} s_i(t) s_j(t)\right|\right), \quad (4)$$

where η_i and η_{ij} are random numbers in $[0,1]$, and θ is the Heaviside function.

As to the value of q , we assume $q = O(1/K)$. This choice deserves a comment. It is known [10,11] that if q is order $\sqrt{\ln K/K}$ or greater, then it takes only one iteration to imprint the pattern corresponding to a given neuronal state; since the synapses maintain the same dynamics during the retrieval of a previously learned pattern, such an occurrence would destroy the associative capability of the network; therefore in the following we assume that q is of order $1/K$.

Let us first describe qualitatively the behavior of our model in the absence of external stimuli. At low temperature the dynamics drives the system towards one of the 2^N configurations satisfying $J_{ij} = s_i s_j$ (separable configurations). If the initial configurations of neurons and synapses are completely disordered, then at late times the network converges to one state chosen among the separable configurations; in other words the network spontaneously breaks the symmetry among the separable configurations (a similar spontaneous symmetry breaking is discussed in [4]). If, on the other hand, the initial conditions are sufficiently polarized with respect to a given separable configuration, we expect the network to converge to that configuration.

Our aim is to describe macroscopically the evolution of the system by flow equations for the macroscopic parameters describing the state of neurons and synapses. Let us take $H_i = 0$ (no external stimuli on the network) and consider an arbitrary pattern $\{\xi_i\}$ for the neurons. We introduce the variables $\tilde{s}_i = \xi_i s_i$ and $\tilde{J}_{ij} = J_{ij} \xi_i \xi_j$, which will be useful to measure the degree of polarization of neurons and synapses with respect to the pattern $\{\xi_i\}$. The statistical ensemble we deal with consists of all the histories of the network with the initial conditions $\{s_i(0)\}$ $\{J_{ij}(0)\}$ sampled independently as follows:

$$\text{Prob}[s_i(0) = \xi_i] = \frac{1}{2} (1 + m_0), \quad (5)$$

$$\text{Prob}[J_{ij}(0) = J_\alpha \xi_i \xi_j] = \rho_0(\alpha). \quad (6)$$

Since the initial conditions are uniform (i.e., independent of the site index i), the statistical properties of our model remain uniform at every later time t . We also remark that at $t=0$ both neurons and synapses are uncorrelated.

We study the model in the limit $N \rightarrow \infty$ with K large and finite. It can be shown that in such a limit the neurons never

correlate [12]. Indeed the K^t sites that belong to the tree of ancestors of a given neuron s_i are all different (this is true as long as $t \ll \ln N$). Also the correlations between synaptic variables can be neglected: the correlation between two synapses J_{ij} and J_{ik} sharing a common site increases when they are simultaneously updated while it decreases when one out of the two is updated. The probability for the simultaneous updating is of order q^2 , while the probability for the single updating is of order q . Since q is of order $1/K$, it follows that, at least in the early stage, the synapses can be treated as uncorrelated. In the following, analyzing in particular the cases $n=2$ and $n=3$, we will show that actually the correlation between synapses can *always* be neglected. We also remark that s_j is independent of J_{ij} because there are no loops in the tree of ancestors of s_i . Therefore neurons (as well as synapses) can be treated as independent and identically distributed stochastic variables.

We now introduce the order parameter $m(t) = \langle \tilde{s}_i(t) \rangle = (1/N) \sum_i \langle \tilde{s}_i(t) \rangle$ for neurons and the probability $\rho(\alpha, t)$ that $\tilde{J}_{ij}(t)$ is equal to J_α . We have, from Eqs. (5) and (6), $\rho(\alpha, 0) = \rho_0(\alpha)$ and $m(0) = m_0$. The flow equation for $\rho(\alpha, t)$ is easily found to be

$$\rho(\alpha, t+1) = \sum_{\alpha'} T_{\alpha\alpha'}[m(t)] \rho(\alpha', t), \quad (7)$$

where the $n \times n$ stochastic transition matrix T is tridiagonal with the following structure:

$$T(m(t)) = \begin{pmatrix} 1-a & b & 0 & \dots & 0 \\ a & 1-q & b & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & a & 1-q & b \\ 0 & 0 & \dots & a & 1-b \end{pmatrix}, \quad (8)$$

where $a = (q/2)[1 - m^2(t)]$ and $b = (q/2)[1 + m^2(t)]$.

In order to calculate the flow equation for the order parameter $m(t)$, we observe that the projection on $\{\xi\}$ of local field (2) can be written as

$$\tilde{h}_i(t) = \xi_i h_i(t) = \sum_j \tilde{J}_{ij}(t) \tilde{s}_j(t) = \sum_j J'_{ij}(t), \quad (9)$$

where the variable $J'_{ij} = \tilde{J}_{ij} \tilde{s}_j$ is the contribution to the synaptic input from a single input neuron. The variable \tilde{h} (the total synaptic input) is the sum of K independent and identically distributed stochastic variables J' . The probability distribution for J' reads

$$\rho'(\alpha, t) = \text{Prob}[J'_{ij}(t) = J_\alpha] \\ = \frac{1}{2} [1 + m(t)] \rho(\alpha, t) + \frac{1}{2} [1 - m(t)] \\ \times \rho(n+1 - \alpha, t), \quad (10)$$

where the first term on the right-hand side (RHS) of Eq. (10) is the probability that $\tilde{s} = 1$ and $\tilde{J} = J_\alpha$, the second term is the probability that $\tilde{s} = -1$ and $\tilde{J} = -J_\alpha$. The stochastic

variable $\tilde{h}_i(t)$ can assume $(n-1)K+1$ possible discrete values in $[-K, K]$; its distribution $\mathcal{H}_i(a) = \text{Prob}[\tilde{h}_i(t) = a]$ can be calculated by performing K times the convolution of the distribution (10). For example, in the case $n=2$, we have

$$\mathcal{H}_i(a) = \binom{K}{\frac{K-a}{2}} \rho'(1, t)^{(K+a)/2} \rho'(-1, t)^{(K-a)/2}. \quad (11)$$

It is now straightforward to evaluate, in the general case, the flow equation for $m(t)$ using Eqs. (3):

$$m(t+1) = \sum_{a=-K}^K \mathcal{H}_i(a) \tanh(\beta a) = \langle \tanh(\beta \tilde{h}) \rangle_{\mathcal{H}_i}. \quad (12)$$

Flow equations (7) and (12), with initial conditions m_0 and $\rho_0(\alpha)$, describe the coupled dynamics of $m(t)$ and $\rho(\alpha, t)$.

After this general introduction and the description of the main dynamical equations, we will devote the next section to study the fixed points of the dynamics for our model. In Sec. III we will report the results obtained by simulating numerically the flow equations in the cases $n=2$ and $n=3$. In Sec. IV it is shown that the correlations among synapses can always be neglected. In Sec. V we study the learning properties of the network. In Sec. VI the simulations of an ANN with adapting synapses are studied and compared to the theory. Section VII summarizes the conclusions.

II. STATIONARY SOLUTIONS

Let us now discuss the stationary solutions of the flow equations. First of all we observe that the invariant distribution with respect to Eq. (7) is given by

$$\rho_m(\alpha) = \frac{2m^2}{(1+m^2)^n - (1-m^2)^n} (1-m^2)^{\alpha-1} (1+m^2)^{n-\alpha}. \quad (13)$$

We remark that the stationary distribution does not depend on q . We call $\mathcal{H}^{(m)}(a)$ the probability distribution of \tilde{h} corresponding to ρ_m . For stationary solutions the following equation holds:

$$m = \langle \tanh(\beta \tilde{h}) \rangle_{\mathcal{H}^{(m)}}. \quad (14)$$

It can be shown that $m=0$ is always a stable solution for Eq. (14), and it is unique for small values of β (high temperature). As β increases, Eq. (14) displays a first-order transition, i.e., for $\beta > \beta_c(K)$ two solutions with $m > 0$ appear discontinuously, the one with larger m being locally stable for variations in m . This solution corresponds to a state that deviates only slightly from pattern $\{\xi\}$; $m=1$ is a stationary and stable solution only in the limit $\beta \rightarrow \infty$. We have thus shown that the model possesses a stable fixed point in correspondence to any pattern $\{\xi\}$. If the initial conditions m_0 and $\rho_0(\alpha)$ are sufficiently polarized with respect to $\{\xi\}$, then, by iterating the flow equations, the stable solution $m_s > 0$ is asymptotically achieved, i.e.,

$$\lim_{t \rightarrow \infty} m(t) = m_s, \quad \lim_{t \rightarrow \infty} \rho(\alpha, t) = \rho_{m_s}(\alpha). \quad (15)$$

In the case of a large number of connections (large K) a useful approximation can be used. Since the local field $\tilde{h}_i(t)$ is the sum of K independent and identically distributed stochastic variables, we use the central limit theorem to approximate it by a Gaussian random variable with mean and variance given respectively by

$$\mu(t) = Km(t) \langle \tilde{J}(t) \rangle \quad (16)$$

and

$$\sigma^2(t) = K[\langle \tilde{J}^2(t) \rangle - \langle \tilde{J}(t) \rangle^2 m^2(t)]. \quad (17)$$

The flow equation (12) can be approximated as follows:

$$m(t+1) = \int \frac{dz}{\sqrt{2\pi}} e^{-(1/2)z^2} \tanh\{\beta[\mu(t) + \sigma(t)z]\}. \quad (18)$$

In the zero-temperature limit ($\beta \rightarrow \infty$) Eq. (18) reads

$$m(t+1) = \text{erf}\left(\frac{\mu(t)}{\sigma(t)}\right), \quad (19)$$

where μ/σ may be seen as the signal-to-noise ratio (SNR) for the synaptic input of a neuron.

III. ANALYSIS OF THE CASES $n=2$ AND $n=3$

In this section we study numerically the behavior of the flow equations (7) and (18) in the cases $n=2$ and $n=3$. The probability distribution for two-state synapses is determined by its average $\mathcal{J}(t) = \langle \tilde{J}(t) \rangle$ and flow equations reduce to

$$\mathcal{J}(t+1) = (1-q)\mathcal{J}(t) + qm^2(t), \quad (20)$$

$$m(t+1) = \int \frac{dz}{\sqrt{2\pi}} e^{-(1/2)z^2} \tanh\{\beta[K\mathcal{J}(t)m(t) + \sqrt{K[1 - \mathcal{J}^2(t)m^2(t)]}z]\}. \quad (21)$$

The stationarity conditions are

$$\mathcal{J} = m^2, \quad (22)$$

$$m = \int \frac{dz}{\sqrt{2\pi}} e^{-(1/2)z^2} \tanh\{\beta[Km^3 + \sqrt{K(1-m^6)}z]\}. \quad (23)$$

Equation (23) displays a first-order transition for β greater than a critical coupling, which has the behavior $\beta_c(K) \sim 2.017/K$ for large K . Hence for $\beta > \beta_c$ Eqs. (20) and (21) have two stable fixed points, a fixed point with $m > 0$ and the trivial one. By numerical simulations we found that the recurrence equations (20) and (21) never show complex behavior and by iterating them, starting from any initial condition, it always happens that one of the two stable fixed points is approached.

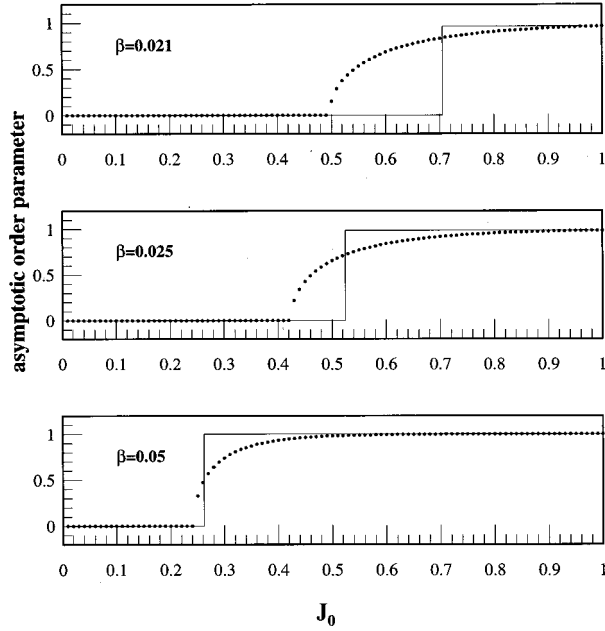


FIG. 1. Numerical evaluation of the asymptotic order parameter, vs the initial polarization \mathcal{J}_0 of the synapses, with $K=100$, $q=0.01$, $n=2$ and three different values of the temperature. The continuous line and the dotted one correspond respectively to AR and FSR (see the definitions in the text).

It is interesting to compare the retrieval capability of our neural network with adapting synapses to the case of a strongly diluted network with fixed synapses, where each neuron has K random input sites, described by Eq. (21) with $\mathcal{J}(t)=\mathcal{J}_0$ for every t . The process of retrieval in our model will be referred to as *adaptive retrieval* (AR) whereas the retrieval with fixed synapses will be called *fixed synapses retrieval* (FSR). We remark that FSR actually corresponds to the early stage of AR, due to the very low value of q . In Fig. 1 we depicted the asymptotic value $m(\infty)$ of the order parameter versus the initial condition \mathcal{J}_0 for the two cases and various temperatures. We see that in the case of AR a first-order transition occurs; in other words, for \mathcal{J}_0 greater than a threshold value \mathcal{J}_{th} , the nontrivial fixed point of Eq. (23) is asymptotically achieved; on the other hand a second-order transition occurs in the case of FSR. For \mathcal{J}_0 greater than \mathcal{J}_{th} and $\mathcal{J}_0 - \mathcal{J}_{th}$ small, AR performs better than FSR. We observe that a threshold value exists also for m_0 , i.e., for m_0 smaller than a threshold value m_{th} (numerically very small), the flow equations (20) and (21) lead to the trivial fixed point even for $\mathcal{J}_0=1$.

We remark that the change from second-order (fixed synapses) to first-order transition (adapting synapses), which we found in our model, is similar to the crossover phenomenon, for the case of the paramagnetic-ferromagnetic transition, which is observed in [7], where the interaction dynamics changes the transition to first order when the influence of the spins on the interaction dynamics is dominant with respect to the noise term.

Let us now consider the case $n=3$, when synapses take values in $\{1,0,-1\}$. Two independent parameters are needed in order to specify the distribution of the synapses; let us define $\rho^+(t)=\rho(1,t)$, $\rho^-(t)=\rho(-1,t)$, and $\rho^0(t)=\rho(0,t)$.

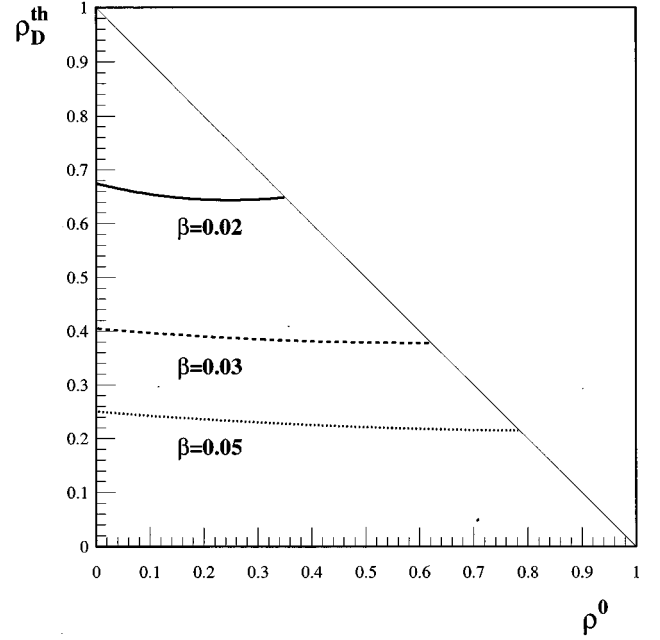


FIG. 2. The threshold value of ρ_D is plotted vs ρ^0 in the case $n=3$, $K=100$, $q=0.01$ and with three different values of the temperature.

The stationarity conditions can be computed in this case with the results

$$\rho^+ = \frac{(1+m^2)^2}{3+m^4}, \quad \rho^0 = \frac{1-m^4}{3+m^4}, \quad \rho^- = \frac{(1-m^2)^2}{3+m^4}, \quad (24)$$

$$m = \int \frac{dz}{\sqrt{2\pi}} e^{-(1/2)z^2} \tanh(\beta\{Km(\rho^+ - \rho^-) + \sqrt{K[1-\rho^0 - m^2(\rho^+ - \rho^-)^2]}z\}). \quad (25)$$

The same qualitative description of the $n=2$ case applies to the behavior of the flow equations in this case. The critical coupling has the behavior $\beta_c(K) \sim 1.8/K$ for large K . In Fig. 2 we depict, for three values of the temperature, the threshold value of $\rho_D = \rho^+ - \rho^-$ versus ρ^0 . The threshold ρ_D^{th} is such that the nontrivial fixed point is achieved if the initial conditions satisfy $\rho_D > \rho_D^{th}$.

IV. THE SYNAPTIC CORRELATIONS

In this section we show that, during the time evolution of the network, the correlations among the synapses can be neglected. This assumption is tested by taking into account the correlation between pairs of synapses sharing a common site. Consider the local field \tilde{h}_i acting on the neuron s_i given by Eq. (9). The mean and variance of \tilde{h}_i are given by

$$\mu = K\langle \tilde{J} \rangle m, \quad (26)$$

$$\sigma^2 = K(\langle \tilde{J}^2 \rangle - \langle \tilde{J} \rangle^2 m^2) + K(K-1)m^2 \mathcal{C}, \quad (27)$$

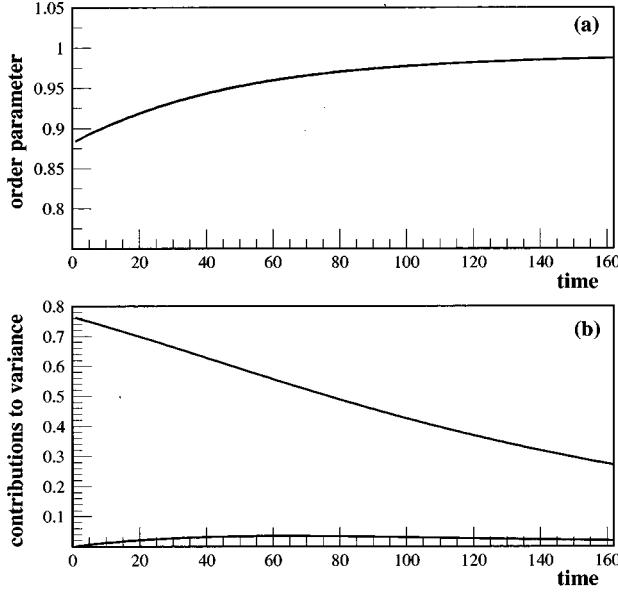


FIG. 3. (a) The time evolution of the order parameter evaluated by the flow equations that neglect the correlations and by the flow equations that take into account the pair synaptic correlation, in the case $n=2$: the two curves are indistinguishable. The initial conditions are $m_0=0.88$, $\mathcal{J}_0=0.55$; $K=100$, $q=0.01$, $\beta=0.03$. (b) We depict, with the same initial conditions, the time evolution of the two contributions to the variance of the local field (see the text): the contribution due to the synaptic correlations, i.e., $(K-1)m^2\mathcal{C}$ (lower curve), compared to the term that is computed as if the synapses were uncorrelated, i.e., $1-\mathcal{J}^2m^2$ (upper curve).

where $\mathcal{C}=\langle\tilde{\mathcal{J}}_{ij}\tilde{\mathcal{J}}_{ik}\rangle-\langle\tilde{\mathcal{J}}_{ij}\rangle\langle\tilde{\mathcal{J}}_{ik}\rangle$ is the pair synaptic correlation. The first term on the RHS of Eq. (27) is the variance of the local field, which is computed as if the synapses were uncorrelated.

In the $n=2$ case we consider the probability distribution for pairs of synapses $\{\tilde{\mathcal{J}}_{ij}, \tilde{\mathcal{J}}_{ik}\}$ sharing a common site, which evolves according to a 4×4 transition matrix. After a little algebra we obtain the flow equations for \mathcal{J} ($=\langle\tilde{\mathcal{J}}_{ij}\rangle = \langle\tilde{\mathcal{J}}_{ik}\rangle$) and \mathcal{C} :

$$\mathcal{J}(t+1)=(1-q)\mathcal{J}(t)+qm^2(t), \quad (28)$$

$$\mathcal{C}(t+1)=(1-q)^2\mathcal{C}(t)+q^2[m^2(t)-m^4(t)], \quad (29)$$

while Eq. (18) can be used for the evolution of the order parameter $m(t)$; the initial condition for the correlation is $\mathcal{C}(0)=0$. Equation (29) has a clear meaning: the pair correlation does not change if the two synapses are not updated [probability $(1-q)^2$]; it goes to zero if one out of the two is updated [probability $2q(1-q)$]; it becomes m^2-m^4 if both synapses are updated (probability q^2). We studied numerically Eqs. (28), (29), and (18) and found that the second term on the RHS of Eq. (27) is always very small compared to the first term. In Fig. 3(a) we depict the time evolution of the order parameter, for a particular set of initial conditions, in two cases: taking into account the pair correlations and not. The two curves are indistinguishable; the same behavior is observed for any choice of the initial conditions. In Fig. 3(b)

we depict, with the same set of initial conditions, the time evolution of the two contributions to the variance of the local field, given by Eq. (27). The contribution due to the synaptic correlations is always very small and reaches a maximum value during the retrieval process. The stationarity conditions for Eqs. (28) and (29) read:

$$\mathcal{J}=m^2, \quad \mathcal{C}=\frac{q}{2-q}(m^2-m^4), \quad (30)$$

and, at leading order in K , the equation for the fixed points is

$$m=\int\frac{dz}{\sqrt{2\pi}}e^{-(1/2)z^2}\times\tanh\left\{\beta\left[Km^3+\sqrt{K\left(1-m^6+\frac{m^4-m^6}{2}\right)}z\right]\right\} \quad (31)$$

to be compared with the equation that is computed as if the synapses were uncorrelated, i.e., Eq. (23). We see that the distance between the nontrivial fixed points of Eqs. (31) and (23) vanishes for large K . Hence the synaptic correlations can be neglected also with respect to the equilibrium properties of the network.

We tested the assumption also in the $n=3$ case (see the Appendix). We find that also in this case the contribution to σ^2 due to the correlations is always very small and can be neglected.

V. THE LEARNING STAGE

The learning stage is characterized by the presence of external stimuli represented by local magnetic fields in Eq. (2). We recall the results obtained in [11] about the capacity of a stochastic learning network with q of order $\sqrt{\ln K}/K$ and with n arbitrary. Let us suppose that the network receives an uninterrupted flow of uncorrelated patterns to be learned:

$$\dots,\{\xi\}_{-2},\{\xi\}_{-1},\{\xi\}_0,\{\xi\}_1,\{\xi\}_2,\dots$$

In [11] it is shown, by a signal-to-noise analysis, that the network acts as a palimpsest [14,15], i.e., patterns learned far in the past are erased by new patterns, and that the maximum number of patterns that can be stored is proportional to \sqrt{K} [16].

In our case q is of order $1/K$ and each pattern $\{\xi\}_p$ ($p=0,\pm 1,\pm 2,\dots$) has to be presented at least for $\ell=O(\sqrt{K}\ln K)$ iterations in order to be stored. Let us consider, for example, the pattern $\{\xi\}_0$, and let us set $t=0$ in correspondence of the first presentation of $\{\xi\}_0$. Initially the synapses have a completely random distribution with respect to $\{\xi\}_0$, i.e., $\rho(\alpha,0)=1/n$. During the presentation of pattern $\{\xi\}_0$ the synapses are polarized due to the action of matrix (8) with $m=1$, because $\{s\}=\{\xi\}_0$. Subsequently the network begins to learn other patterns and the synaptic distribution becomes depolarized by the action of matrix (8) with $m=0$. After the presentation of z other patterns the synaptic distribution is given by

$$\rho[t=(z+1)\ell]=T_q^{z\ell}(m=0)T_q^\ell(m=1)\rho(0), \quad (32)$$

where the dependence of T on q has been made explicit. Incidentally we remark that during the learning stage the correlation among synapses is identically zero due to the \pm symmetry. For fixed m and q of order $1/K$ we have

$$T'_q = T_{/q} + O(\ell^2 q^2) = T_{/q} + O\left(\frac{\ln K}{K}\right) \quad (33)$$

as can be easily checked by performing the Taylor expansion of the matrix T_q in the neighborhood of $q=0$. Therefore rescaling the time by the factor ℓ and changing q into $\tilde{q} = \ell q$ leads to the same problem studied in [11]. Hence we conclude that our network is a palimpsest and its capacity is proportional to \sqrt{K} .

VI. SIMULATIONS

Our theory works as long as t is small with respect to $\ln N$, so as to avoid the effects due to the neuron correlations. In order to build up finite-size networks, where the neuron correlations can be neglected for longer times, and to test our dynamical equations, we adopt the following approach: we implement a model of N neurons, where each neuron has M input neurons chosen at random ($M \ll N$). The NM synapses obey the stochastic learning rules and, for every neuron s_i and for every time t , K input neurons are chosen at random among the M corresponding to s_i and only those K inputs contribute to the local field acting on s_i at time t . This is a generalization of our model (recovered when $K=M$) which is ruled by the same dynamical equations. We implemented a $N=10\,000$ network with $M=200$, $K=21$ and two-states synapses. The time evolution of the macroscopic parameters of the system is found to be in agreement with the theoretical estimate until $t \sim 20$. In Fig. 4 we depict the time evolution of the order parameter and of the synaptic distribution and compare it with the theoretical estimate; the initial conditions were $m_0=1$ and $\mathcal{J}_0=0.3$. The agreement with the theory has been tested with many different initial conditions. We also implemented a network with $N=10\,000$ and $M=K=21$ but the time evolution in this case was in agreement with the theory only for a few time steps.

VII. CONCLUSIONS

In this paper we have considered a neural network with clipped synapses and stochastic learning rules whose learning capabilities have been studied in [11]. We extend the analysis of this model and investigate the consequences of the synaptic dynamics in the process of retrieval. We find that, in order to preserve the associative capability of the system, the synaptic transition probability q must be very small; moreover, for strong dilution, the dynamics of the network can be analytically calculated because for very small values of q the correlations among the synaptic variables can be neglected. As to the learning properties, the network acts as a palimpsest and the maximum number of storable patterns coincides with the result obtained in [11], the only difference being that a pattern has to be presented for many iterations in order to be stored.

In this framework the two stages (computation and learning) differ for the duration of the external stimulus corre-

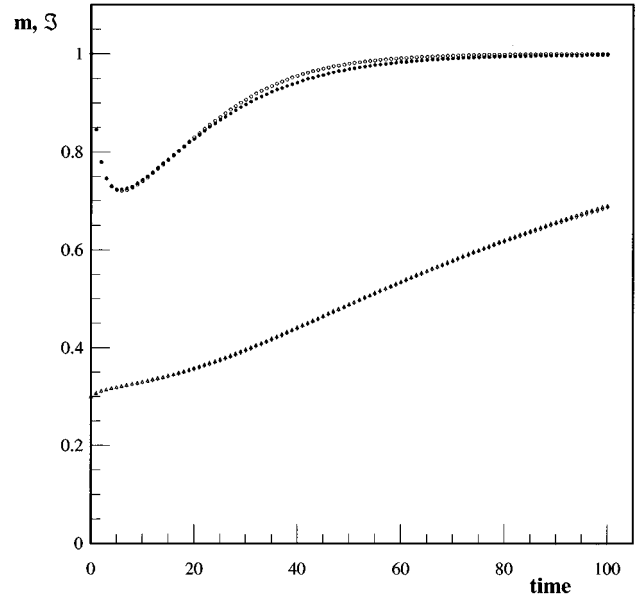


FIG. 4. Time evolution of the order parameter $\langle s \rangle$ and the mean of the synapses $\langle J \rangle$ in a ANN with two-states adapting synapses. The numerical results for $\langle s \rangle$ (\bullet) and $\langle J \rangle$ (\diamond) are averaged over 80 histories of a network with $N=10\,000$, $M=200$, $K=21$, $q=0.01$ and zero temperature; the initial conditions are $m_0=1$ and $\mathcal{J}_0=0.3$. The theoretical estimate by the flow equations is represented by \circ ($\langle s \rangle$) and by \triangle ($\langle J \rangle$).

sponding to a given pattern. If the pattern is presented for a sufficiently long time, the network stores it in the synaptic couplings. If the pattern (or a damaged version of it) is presented for few iterations, the dynamics of the network is capable of reconstructing the pattern provided it had been learned previously (unless new external stimuli impinge on the network during the retrieval).

APPENDIX

We write here the flow equations, in the $n=3$ case, which take into account the pair synaptic correlations. Let us consider the distribution of a pair of synapses $\{\tilde{\mathcal{J}}_{ij}, \tilde{\mathcal{J}}_{ik}\}$ sharing a common site, which evolves according to a 9×9 transition matrix. We define

$$\rho^+ = \text{Prob}[\tilde{\mathcal{J}}_{ij}=1] = \text{Prob}[\tilde{\mathcal{J}}_{ik}=1],$$

$$\rho^- = \text{Prob}[\tilde{\mathcal{J}}_{ij}=-1] = \text{Prob}[\tilde{\mathcal{J}}_{ik}=-1],$$

$$\rho^0 = \text{Prob}[\tilde{\mathcal{J}}_{ij}=0] = \text{Prob}[\tilde{\mathcal{J}}_{ik}=0],$$

$$\rho^{+0} = \text{Prob}[\tilde{\mathcal{J}}_{ij}=1, \tilde{\mathcal{J}}_{ik}=0] = \text{Prob}[\tilde{\mathcal{J}}_{ij}=0, \tilde{\mathcal{J}}_{ik}=1],$$

$$\rho^{+-} = \text{Prob}[\tilde{\mathcal{J}}_{ij}=1, \tilde{\mathcal{J}}_{ik}=-1] = \text{Prob}[\tilde{\mathcal{J}}_{ij}=-1, \tilde{\mathcal{J}}_{ik}=1],$$

$$\rho^{-0} = \text{Prob}[\tilde{\mathcal{J}}_{ij}=-1, \tilde{\mathcal{J}}_{ik}=0] = \text{Prob}[\tilde{\mathcal{J}}_{ij}=0, \tilde{\mathcal{J}}_{ik}=-1].$$

Next, we have

$$\langle \tilde{J}^2 \rangle = \rho_+ + \rho_-,$$

$$\langle \tilde{J} \rangle = \rho_+ - \rho_-,$$

$$\begin{aligned} C &= \langle \tilde{J}_{ij} \tilde{J}_{ik} \rangle - \langle \tilde{J}_{ij} \rangle \langle \tilde{J}_{ik} \rangle \\ &= \rho^+ + \rho^- - \rho^{+0} - \rho^{-0} - 4\rho^{+-} - (\rho^+ - \rho^-)^2. \end{aligned}$$

The flow equations for ρ^+, ρ^-, ρ^0 are given by Eq. (7), while the flow equation for $m(t)$ is Eq. (18). The other flow equations are

$$\begin{aligned} \rho^{+0}(t+1) &= (a_1 - a_3)\rho^{+0}(t) + (a_2 - a_3 + a_4 - a_5)\rho^{+-}(t) \\ &\quad + (a_4 - a_2)\rho^{-0}(t) + (a_3 + a_5)\rho^+(t) + a_2\rho^0(t), \end{aligned} \quad (\text{A1})$$

$$\begin{aligned} \rho^{+-}(t+1) &= a_3\rho^{+0}(t) + (a_1 + a_2 + a_3 + a_5)\rho^{+-}(t) \\ &\quad + a_2\rho^{-0}(t) + a_5\rho^0(t), \end{aligned} \quad (\text{A2})$$

$$\begin{aligned} \rho^{-0}(t+1) &= (a_5 - a_3)\rho^{+0}(t) + (a_3 - a_2)\rho^{+-}(t) + (a_1 \\ &\quad - a_2)\rho^{-0}(t) + a_3\rho^0(t) + (a_2 + a_5)\rho^-(t), \end{aligned} \quad (\text{A3})$$

where $a_1 = (1-q)^2$, $a_2 = \frac{1}{2}q(1-q)(1+m^2)$, $a_3 = \frac{1}{2}q(1-q)(1-m^2)$, $a_4 = \frac{1}{4}q^2(1+3m^2)$, and $a_5 = \frac{1}{4}q^2(1-m^2)$.

-
- [1] D. J. Amit, *Modeling Brain Function* (Cambridge University Press, 1989); J. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation* (Santa Fe Institute, Santa Fe, 1990).
- [2] F. E. Bloom, A. Lazerson, and L. Hofstadter, *Brain, Mind and Behavior* (W.H. Freeman, San Francisco, 1985).
- [3] S. Shinomoto, J. Phys. A **20**, L1305 (1987).
- [4] D. W. Dong and J. J. Hopfield, Network **3**, 267 (1992).
- [5] R. D'Autilia and F. Guerra, in *Representations of Musical Signals*, edited by G. De Poli, A. Piccialli, and C. Roads (MIT Press, Cambridge, 1991).
- [6] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).
- [7] R. W. Penney, A. C. C. Coolen, and D. Sherrington, J. Phys. A **26**, 3681 (1993); A. C. C. Coolen, R. W. Penney, and D. Sherrington, Phys. Rev. B **48**, 16 116 (1993); R. W. Penney and D. Sherrington, J. Phys. A **27**, 4027 (1994).
- [8] V. Dotsenko, S. Franz, and M. Mézard, J. Phys. A **27**, 2351 (1994).
- [9] N. Caticha, J. Phys. A **27**, 5501 (1994).
- [10] M. Tsodyks, Mod. Phys. Lett. B **4**, 713 (1990).
- [11] D. J. Amit and S. Fusi, Neural Comput. **6**, 957 (1994).
- [12] B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987).
- [13] A. C. C. Coolen and D. Sherrington, Phys. Rev. Lett. **71**, 3886 (1993).
- [14] J. P. Nadal, G. Toulouse, J. P. Changeus, and S. Dehaene, Europhys. Lett. **10**, 535 (1986).
- [15] G. Parisi, J. Phys. A **15**, L617 (1986).
- [16] Actually in Ref. [8] a fully connected network is considered and the capacity is proportional to \sqrt{N} , where N is the number of neurons in the network. The same analysis holds in a diluted network where each neuron is connected to K random input neurons and the capacity in this case is proportional to \sqrt{K} .