Journal of Statistical Physics, Vol. 74, Nos. 3/4, 1994

Retrieval and Chaos in Layered Q-Ising Neural Networks

D. Bollé,[1,2] G. M. Shim,[1] and B. Vinck[1]

Received June 16, 1993

Using a probabilistic approach, we study the parallel dynamics of the Q-Ising layered networks for arbitrary Q. By introducing auxiliary thermal fields, we express the stochastic dynamics within the gain function formulation of the deterministic dynamics. Evolution equations are derived for arbitrary Q at both zero and finite temperatures. An explicit analysis of the fixed-point equations is carried out for both $Q=3$ and $Q \to \infty$. The retrieval properties are discussed in terms of the gain parameter, the storage capacity, and the temperature. Using the time evolution of the distance between two network configurations, we investigate the possibility of microscopic chaos. Chaotic behavior is always present for arbitrary finite Q. However, in the limit $Q \to \infty$ the existence of chaos depends on the parameters of the system.

KEY WORDS: Layered networks; Q-Ising neurons; parallel dynamics; chaotic behavior; probabilistic approach.

1. INTRODUCTION

Besides extremely diluted asymmetric neural networks, there is another class of networks that allows an exact treatment of its parallel dynamics: layered feedforward networks. (For a recent review and the relevant literature on this class of networks with binary neurons we refer to ref. 1 and the references cited therein.) The main underlying reason is that in both types of networks there are no feedback loops. An additional property of extremely diluted models is that any finite number of neurons have disjoint clusters of ancestors so that they are completely uncorrelated. In layered models, however, correlations among the neurons do exist precisely because of this common ancestor problem. Nevertheless these correlations

[1] Instituut voor Theoretische Fysica and Interdisciplinair Centrum voor Neurale Netwerken, K.U. Leuven, B-3001 Leuven, Belgium.
[2] e-mail: FGBDA18@BLEKUL11.BITNET.

can be handled exactly, giving rise to layer-to-layer recursions for the relevant order parameters determining the system: the main overlap for the condensed pattern and the so-called width parameter connected with the noncondensed patterns.[1] The latter has been identified explicitly in ref. 2 as the variance of the residual overlap responsible for the intrinsic noise in the dynamics of the main overlap.

In the present work we consider the parallel dynamics of Q-Ising ($Q>2$) and analog ($Q \to \infty$) layered neural networks with Hebbian couplings between adjacent layers and independently chosen representations of patterns on different layers. We provide an exact solution following a probabilistic approach.[3] In particular, employing a signal-to-noise ratio analysis based on the law of large numbers (LLN) and the central limit theorem (CLT), we derive layer-to-layer evolution equations at zero and at finite temperatures T for arbitrary Q. For finite T this requires a nontrivial generalisation of the auxiliary thermal fields method used, e.g., in ref. 2 to express the stochastic dynamics within the gain function formulation of the deterministic dynamics. (For $Q=2$ we recover the results of refs. 4–6.) A detailed study is presented of the macroscopic structure of the retrieval dynamics for $Q=3$ and $Q \to \infty$. Furthermore, we also discuss some aspects of the deterministic dynamics on the microscopic level. Especially we investigate if the distance between two arbitrarily close network configurations correlated with only one embedded pattern diverges, implying chaotic behavior. This extends the results obtained in ref. 7 to arbitrary finite Q and $Q \to \infty$ (analog) layered models.

Similar problems as outlined above have been treated for extremely diluted asymmetric Q-Ising and Q-state Potts models in refs. 8 and 9. Where appropriate we compare the results of these studies with the results obtained here. The parallel dynamics of Potts networks on layered feed-forward architectures has been solved exactly[10] using the generating function method introduced in ref. 11.

The rest of this paper is organized as follows. In Section 2 we introduce the model, its dynamics, and the Hamming distance as a macroscopic measure for the retrieval quality of the network. In Section 3 we solve the deterministic parallel dynamics for arbitrary Q using the probabilistic approach. This leads to recursion relations for the main overlap, the activity, and the variance of the residual overlap. The corresponding recursion relations solving the stochastic parallel dynamics for arbitrary Q are derived in Section 4 by introducing the appropriate auxiliary thermal fields. Section 5 is devoted to the detailed analysis of the retrieval properties of the $Q=3$ model and the analog model by discussing capacity–gain diagrams together with the macroscopic structure (attractors, repellors, saddle points) of the dynamics. In Section 6 an explicit analysis in terms of the Hamming distance tells us that chaotic behavior in the sense of diverging trajectories mentioned above occurs. For arbitrary finite Q this happens in the whole capacity–gain plane, for $Q \to \infty$ there exists a dynamical transition line toward chaos. Some concluding remarks are contained in Section 7.

2. THE MODEL

Consider a neural network composed of multistate neurons arranged in layers, each layer containing N neurons. A neuron can take values in a discrete set $\mathscr{S} \equiv \{-1 = s_1 < s_2 < \cdots < s_{Q-1} < s_Q = +1\}$. Each neuron in layer t is unidirectionally connected to all neurons on layer $t+1$. Given a configuration $\boldsymbol{\sigma}_\Lambda(t) \equiv \{\sigma_j(t)\}$, $j \in \Lambda = \{1, 2,..., N\}$, the local field $h_i(\boldsymbol{\sigma}_\Lambda(t))$ in neuron i on layer $t+1$ is

$$h_i(\boldsymbol{\sigma}_\Lambda(t)) = \sum_{j \in \Lambda} J_{ij}(t+1)\, \sigma_j(t) \tag{1}$$

where $J_{ij}(t+1)$ is the strength of the coupling from neuron j on layer t to neuron i on layer $t+1$. The state $\boldsymbol{\sigma}_\Lambda(t+1)$ of layer $t+1$ is determined by the state $\boldsymbol{\sigma}_\Lambda(t)$ of the previous layer t according to the transition probabilities

$$\Pr[\sigma_i(t+1) = s_k \in \mathscr{S} \mid \boldsymbol{\sigma}_\Lambda(t)] = \frac{\exp\{-\beta\varepsilon_i[s_k \mid \boldsymbol{\sigma}_\Lambda(t)]\}}{\sum_{s \in \mathscr{S}} \exp\{-\beta\varepsilon_i[s \mid \boldsymbol{\sigma}_\Lambda(t)]\}} \tag{2}$$

Here the temperature $T = \beta^{-1}$ measures the noise level, and the energy potential $\varepsilon_i[s \mid \boldsymbol{\sigma}_\Lambda]$ is defined by[12]

$$\varepsilon_i[s \mid \boldsymbol{\sigma}_\Lambda] = -\tfrac{1}{2}[h_i(\boldsymbol{\sigma}_\Lambda)\, s - bs^2] \tag{3}$$

where $b > 0$ is the gain parameter of the system. We take parallel updating. The configuration of the first layer, $\boldsymbol{\sigma}_\Lambda(t=1)$, is chosen as input. At the next time step, the second layer is updated according to the rule (2), and so on. At zero temperature, $\sigma_i(t+1)$ takes the value s_k according to

$$\sigma_i(t) \to \sigma_i(t+1) = s_k : \min_{s \in \mathscr{S}} \varepsilon_i(s \mid \boldsymbol{\sigma}_\Lambda(t)) = \varepsilon_i(s_k \mid \boldsymbol{\sigma}_\Lambda(t)) \tag{4}$$

which is equivalent to using a gain function $g(\cdot)$,

$$\begin{aligned} &\sigma_i(t+1) = g(h_i(\boldsymbol{\sigma}_\Lambda(t))) \\ &\qquad g(x) \equiv \sum_{k=1}^{Q} s_k[\theta(b(s_{k+1} + s_k) - x) - \theta(b(s_k + s_{k-1}) - x)] \end{aligned} \tag{5}$$

with $s_0 \equiv -\infty$ and $s_{Q+1} \equiv \infty$. For finite Q, $g(\cdot)$ is a step function. For $Q \to \infty$ the nature of $g(\cdot)$ depends on the distribution of the states chosen at finite Q. For example, if we consider the distributional limit of equidistant states $\mathcal{S} = \mathcal{S}_Q = \{s_k = -1 + 2(k-1)/(Q-1),\ k = 1,..., Q\}$, the density of states is uniform and the gain function (5) becomes the piecewise linear function

$$g(x) = \begin{cases} \text{sign}(x) & \text{if } |x| > 2b \\ x/2b & \text{otherwise} \end{cases} \tag{6}$$

The gain parameter b controls the average slope of $g(\cdot)$.

In this network, we want to store sets of patterns. The representation of the patterns on layer t is a collection of independent and identically distributed random variables (i.i.d.r.v.) $\{\xi_i^\mu(t) \in \mathcal{S}\}$, $\mu \in \mathcal{P} = \{1, 2,..., p = \alpha N\}$ with zero mean and variance $A = \text{Var}[\xi_i^\mu(t)]$. The synaptic couplings between adjacent layers are chosen according to the Hebb rule

$$J_{ij}(t+1) = \frac{1}{NA} \sum_{\mu \in \mathcal{P}} \xi_i^\mu(t+1)\, \xi_j^\mu(t) \tag{7}$$

The possibility of an analytic treatment of the dynamics mainly stems from the independent choice of the representations of the patterns on different layers.

The retrieval quality of the network can be measured by the Hamming distance between a stored pattern and the microscopic state of the network

$$d_H(\boldsymbol{\xi}^\mu(t), \boldsymbol{\sigma}_A(t)) \equiv \frac{1}{N} \sum_{i \in A} [\xi_i^\mu(t) - \sigma_i(t)]^2 \tag{8}$$

which naturally introduces the main overlap

$$m_A^\mu(t) = \frac{1}{AN} \sum_{i \in A} \xi_i^\mu(t)\, \sigma_i(t) \tag{9}$$

and the activity of the neurons

$$a_A(t) = \frac{1}{N} \sum_{i \in A} [\sigma_i(t)]^2 \tag{10}$$

In contrast to binary networks, it is necessary to know both the main overlap and the activity.

3. EXACT SOLUTION OF THE MODEL AT $T=0$

Suppose that the initial data $\boldsymbol{\sigma}_\Lambda(1)$ are a collection of i.i.d.r.v. with mean zero, variance $\mathrm{Var}[\sigma_i(1)]=a_0$, and correlated with only one stored pattern, say the first one ($\mu=1$), i.e.,

$$\frac{1}{A}\,\mathrm{E}[\xi_i^\mu(1)\,\sigma_i(1)]=\delta_{\mu,1}m_0, \qquad m_0>0 \tag{11}$$

By the LLN the main overlaps and the activity at $t=1$ get the form

$$m^\mu(1)\equiv\lim_{N\to\infty} m_\Lambda^\mu(1)\overset{\mathrm{Pr}}{=}\frac{1}{A}\,\mathrm{E}[\xi_i^\mu(1)\,\sigma_i(1)]=\delta_{\mu,1}m_0 \tag{12}$$

$$a(1)\equiv\lim_{N\to\infty} a_\Lambda(1)\overset{\mathrm{Pr}}{=}\mathrm{E}[\sigma_i^2(1)]=a_0 \tag{13}$$

where the convergence is in probability.[13] According to the evolution rule (5), the configuration of the second layer is completely determined by the local fields $\{h_i(\boldsymbol{\sigma}_\Lambda(1)), i\in\Lambda\}$. Inserting Eq. (7) into Eq. (1), we split the local field on site i into a signal term and a noise term

$$h_i(\boldsymbol{\sigma}_\Lambda(1))=\xi_i^1(2)\,m_\Lambda^1(1)+\frac{1}{\sqrt{N}}\sum_{\mu\in\mathscr{P}\backslash 1}\xi_i^\mu(2)\,r_\Lambda^\mu(1) \tag{14}$$

where we have introduced the residual overlaps[2]

$$r_\Lambda^\mu(t)=\frac{1}{A\sqrt{N}}\sum_{i\in\Lambda}\xi_i^\mu(t)\,\sigma_i(t) \qquad \text{for} \quad \mu>1 \tag{15}$$

The first term on the r.h.s. of Eq. (14) represents the contribution from the condensed pattern ($\mu=1$) and the second term describes the noise contribution from the noncondensed patterns ($\mu>1$). Since the $\{\boldsymbol{\sigma}(1)\}$ are uncorrelated with the $\{\boldsymbol{\xi}^\mu(1)\}$ for $\mu>1$, $\{\xi_i^\mu(1)\,\sigma_i(1)\}$ is a collection of i.i.d.r.v. Therefore, using CLT, the limiting residual overlaps become

$$r^\mu(1)\equiv\lim_{N\to\infty} r_\Lambda^\mu(1)\overset{\mathscr{D}}{=}\mathscr{N}(0, D(1)) \tag{16}$$

where $D(1)\equiv a_0/A$ and the convergence is in distribution (see, e.g., ref. 13). The quantity $\mathscr{N}(0, d)$ represents a Gaussian random variable with expectation 0 and variance d. Although $\xi_i^\mu(1)\,\sigma_i(1)$ and $\xi_i^\nu(1)\,\sigma_i(1)$ are correlated, the distributional limits of the residual overlaps $\{r^\mu(1)\}$ are i.i.d.r.v. because the $\{\xi_i^\mu(1)\}$ are independent.[2,14] Since the $\{\xi_i^\mu(2)\}$ are i.i.d.r.v. and independent of the $\{r_A^\mu(1)\}$, the last term in Eq. (14) becomes $\mathcal{N}_i(0, \alpha AD(1))$ in the limit $N \to \infty$ and the $\{\mathcal{N}_i\}$ are again i.i.d.r.v.

To get the main overlaps and the activity on the second layer in the limit $N \to \infty$, we apply the LLN. The main overlaps are given by

$$m^\mu(2) \equiv \lim_{N \to \infty} \frac{1}{AN} \sum_{i \in A} \xi_i^\mu(2)\, g\left(\xi_i^1(2)\, m_A^1(1) + \frac{1}{\sqrt{N}} \sum_{\mu \in \mathscr{P}\setminus 1} \xi_i^\mu(2)\, r_A^\mu(1)\right)$$

$$\overset{\mathrm{Pr}}{=} \delta_{\mu,1} \frac{1}{A} \left\langle\!\!\left\langle \xi^1(2) \int Dz\; g(\xi^1(2)\, m^1(1) + [\alpha AD(1)]^{1/2} z) \right\rangle\!\!\right\rangle \tag{17}$$

where $\langle\langle \cdots \rangle\rangle$ stands for the average taken with respect to the distribution of the first pattern, and Dz denotes a Gaussian measure, $Dz = \exp(-\frac{1}{2}z^2)\, dz/(2\pi)^{1/2}$. The Gaussian integration comes from the noise term associated with the noncondensed patterns. Since $g(\cdot)$ is just a sum of step functions, we can perform this integration explicitly, leading to

$$m^\mu(2) = \delta_{\mu,1} \frac{1}{2A} \left\langle\!\!\left\langle \xi^1(2) \sum_{k=1}^{Q} s_k \{\mathrm{Erf}[X_k(1)] - \mathrm{Erf}[X_{k-1}(1)]\} \right\rangle\!\!\right\rangle \tag{18}$$

where we have introduced

$$\mathrm{Erf}[x] \equiv \frac{2}{\sqrt{\pi}} \int_0^x dt\, e^{-t^2} \tag{19}$$

and

$$X_k(t) \equiv \frac{-\xi^1(t+1)\, m^1(t) + b(s_{k+1} + s_k)}{[2\alpha AD(t)]^{1/2}} \tag{20}$$

The activity of the neurons gets the form

$$a(2) \equiv \lim_{N \to \infty} a_A(2) \overset{\mathrm{Pr}}{=} \tfrac{1}{2} \left\langle\!\!\left\langle \sum_{k=1}^{Q} s_k^2 \{\mathrm{Erf}[X_k(1)] - \mathrm{Erf}[X_{k-1}(1)]\} \right\rangle\!\!\right\rangle \tag{21}$$

The formulas (18) and (21) are similar to those for the first-step dynamics of the recurrent Q-Ising model.[8] Since the $\{\sigma_i(1)\}$ are supposed to be a collection of i.i.d.r.v., there is no correlation effect in the main overlap and the activity of the second layer.

For the configuration of the third layer we need to derive the statistical distribution of the residual overlaps $r_A^\mu(2)$:

$$r_A^\mu(2) = \frac{1}{A\sqrt{N}} \sum_{i \in A} \xi_i^\mu(2)\, g\left(\frac{1}{\sqrt{N}} \xi_i^\mu(2)\, r_A^\mu(1) + \xi_i^1(2)\, m_A^1(1) \right.$$
$$\left. + \frac{1}{\sqrt{N}} \sum_{\nu \in \mathscr{P}\setminus\{1,\mu\}} \xi_i^\nu(2)\, r_A^\nu(1)\right) \tag{22}$$

The gain function $g(\cdot)$ [see Eq. (5)] is a step function that changes its value by $s_{k+1}-s_k$ at $b(s_k+s_{k+1})$ $(k=1,..., Q-1)$. Therefore, the term $\xi_i^\mu(2)\, r_A^\mu(1)/\sqrt{N}$ in (22) can play an important role if for some k the rest of the argument of $g(\cdot)$ satisfies

$$\left|\xi_i^1(2)\, m_A^1(1)+\frac{1}{\sqrt{N}}\sum_{\nu\in\mathscr{P}\backslash\{1,\mu\}}\xi_i^\nu(2)\, r_A^\nu(1)-b(s_k+s_{k+1})\right| < \frac{1}{\sqrt{N}}\,|\xi_i^\mu(2)\, r_A^\mu(1)| \tag{23}$$

Let us denote by I_k $(k=1,..., Q-1)$ the set of sites satisfying condition (23). In the limit $N\to\infty$, the cardinal numbers of the I_k are given by

$$\lim_{N\to\infty}\frac{|I_k|}{\sqrt{N}}=2\,|\xi_i^\mu(2)\, r^\mu(1)|\; P_2^\mu[b(s_k+s_{k+1})-\xi_i^1(2)\, m^1(1)] \tag{24}$$

where $P_2^\mu(\cdot)$ is the probability distribution of

$$\omega_i^\mu(2)\equiv\lim_{N\to\infty}\omega_{i,A}^\mu(2) \tag{25}$$

$$\omega_{i,A}^\mu(2)\equiv\frac{1}{\sqrt{N}}\sum_{\nu\in\mathscr{P}\backslash\{1,\mu\}}\xi_i^\nu(2)\, r_A^\nu(1) \tag{26}$$

We then split the residual overlap $r_A^\mu(2)$ in two sums

$$\begin{aligned} r_A^\mu(2)=&\frac{1}{A\sqrt{N}}\sum_{i\notin\bigcup_k I_k}\xi_i^\mu(2)\, g\left(\xi_i^1(2)\, m_A^1(1)+\frac{1}{\sqrt{N}}\sum_{\nu\in\mathscr{P}\backslash\{1,\mu\}}\xi_i^\nu(2)\, r_A^\nu(1)\right)\\ &+\sum_{k=1}^{Q-1}\frac{1}{A\sqrt{N}}\sum_{i\in I_k}\xi_i^\mu(2)\left(\frac{s_k+s_{k+1}}{2}+\frac{s_{k+1}-s_k}{2}\operatorname{sign}[\xi_i^\mu(2)\, r_A^\mu(1)]\right) \end{aligned} \tag{27}$$

The term $\xi_i^\mu(2)\, r_A^\mu(1)/\sqrt{N}$ in the argument of $g(\cdot)$ in the first expression of Eq. (27) is left out since it cannot change the value of $g(\cdot)$, by definition of the sets I_k. Therefore we can apply CLT on the first term and LLN on the second term with the random variable $r_A^\mu(1)$ fixed in Eq. (27). This yields

$$\begin{aligned}\lim_{N\to\infty} r_A^\mu(2) \overset{\mathscr{D}}{=} \mathscr{N}_\mu\Bigg(0, \frac{1}{A}\left\langle\!\!\left\langle\int d\omega\, P_2^\mu(\omega)\, g^2[\xi^1(2)\, m^1(1)+\omega]\right\rangle\!\!\right\rangle\Bigg)\\ +r^\mu(1)\sum_{k=1}^{Q-1}(s_{k+1}-s_k)\langle\!\langle P_2^\mu[b(s_k+s_{k+1})-\xi^1(2)\, m^1(1)]\rangle\!\rangle\end{aligned} \tag{28}$$

Since the distribution of $\omega_i^\mu(2)$ is Gaussian with mean zero and variance $\alpha AD(1)$, Eq. (28) gets the form

$$r^\mu(2) \equiv \lim_{N\to\infty} r_A^\mu(2) \stackrel{\mathscr{D}}{=} \mathscr{N}_\mu\left(0, \frac{a(2)}{A}\right) + r^\mu(1) \sum_{k=1}^{Q-1} (s_{k+1} - s_k) \left\langle\!\!\left\langle \frac{\exp[-X_k^2(1)]}{[2\pi\alpha AD(1)]^{1/2}} \right\rangle\!\!\right\rangle \tag{29}$$

The limiting residual overlaps found in (29) are sums of two independent Gaussian random variables and hence are again Gaussian. A deviation from the standard CLT for i.i.d.r.v. is caused by the dependence of the $\{\xi_i^\mu(2)\}$ and the $\{\sigma_i(2)\}$ and the mutual dependence among the $\{\sigma_i(2)\}$ in (15). However, the independence of the $\{r^\mu(2)\}$ still persists because the $\{r^\mu(1)\}$ are independent. This independence of the limiting residual overlaps $\{r^\mu(1)\}$ is indeed the main result obtained from Eqs. (18) and (21). Since this property is preserved layer by layer, the recursion relations for $t > 2$ can be read off from Eqs. (18) and (21) together with an additional recursion relation following from (29) for the variance of the limiting residual overlaps $D(t) \equiv \mathrm{Var}[r^\mu(t)]$:

$$m^\mu(t+1) = \delta_{\mu,1} \frac{1}{2A} \left\langle\!\!\left\langle \xi^1(t+1) \sum_{k=1}^{Q} s_k \{\mathrm{Erf}[X_k(t)] - \mathrm{Erf}[X_{k-1}(t)]\} \right\rangle\!\!\right\rangle \tag{30a}$$

$$a(t+1) = \frac{1}{2} \left\langle\!\!\left\langle \sum_{k=1}^{Q} s_k^2 \{\mathrm{Erf}[X_k(t)] - \mathrm{Erf}[X_{k-1}(t)]\} \right\rangle\!\!\right\rangle \tag{30b}$$

$$D(t+1) = \frac{a(t+1)}{A} + \frac{1}{2\pi\alpha A} \times \left[\left\langle\!\!\left\langle \sum_{k=1}^{Q} s_k \{\exp[-X_k^2(t)] - \exp[X_{k-1}^2(t)]\} \right\rangle\!\!\right\rangle \right]^2 \tag{30c}$$

where we recall (20). We note that the number of independent variables in the recursion relations is two, say the variables $m(t)$ and $D(t)$. The initial conditions for the recursion relations (30) are $m(1) = m_0$ and $D(1) = a_0/A$ [and $a(1) = a_0$]. We remark that the recursion relations (30) generalize the results of refs. 4–6 for $Q = 2$.

The recursion relations (30) suggest the following formulas for a general gain function $g(\cdot)$:

$$m^\mu(t+1) = \delta_{\mu,1} \frac{1}{A} \left\langle\!\!\left\langle \xi^1(t+1) \int Dz\, g(\xi^1(t+1)\, m^1(t) + [\alpha AD(t)]^{1/2} z) \right\rangle\!\!\right\rangle \tag{31a}$$

$$a(t+1)=\left\langle\!\!\left\langle \int Dz\, g^2(\xi^1(t+1)\, m^1(t)+[\alpha A D(t)]^{1/2} z)\right\rangle\!\!\right\rangle \tag{31b}$$

$$D(t+1)=\frac{a(t+1)}{A}+\frac{1}{\alpha A}\left[\left\langle\!\!\left\langle \int Dz\, zg(\xi^1(t+1)\, m^1(t) + [\alpha A D(t)]^{1/2} z)\right\rangle\!\!\right\rangle\right]^2 \tag{31c}$$

where now $\langle\!\langle \cdots \rangle\!\rangle$ denotes the average over the distribution of the patterns with the density $\rho(\xi)$,

$$\langle\!\langle f(\xi)\rangle\!\rangle \equiv \int_{-1}^{1} d\xi\, \rho(\xi)\, f(\xi) \tag{32}$$

For example, if we consider uniformly distributed patterns, the density $\rho(\cdot)$ is simply 1/2. In the case of the piecewise linear input–output function (6) the relations (31) can be derived through computation of the characteristic function for the residual overlaps. They can also be obtained as a limiting case of the solution of the model at nonzero temperatures. This is the subject of the next section.

4. EXACT SOLUTION OF THE MODEL AT $T \neq 0$

At finite temperatures, auxiliary thermal fields[2] are introduced to express the stochastic dynamics within the gain function formulation of the deterministic dynamics. For each i and t, let $\{\phi_i^l(t)\}$, $l=1,\dots, Q-1$, be a collection of i.i.d.r.v. with joint distribution

$$F_\beta(x_1,\dots, x_{Q-1}) \equiv \Pr\left[\bigcap_{l=1}^{Q-1} \{\phi_i^l(t) < x_l\}\right] = \left[1+\sum_{l=1}^{Q-1} \exp(-\beta x_l)\right]^{-1} \tag{33}$$

The transition probability (2) can then be written in the form

$$\Pr[\sigma_i(t+1)=s_k \mid \boldsymbol{\sigma}_A(t)] = \Pr\left[\bigcap_{l\neq k} \{\phi_i^l(t+1) < \varepsilon_i[s_l \mid h_i(\boldsymbol{\sigma}_A(t))] - \varepsilon_i[s_k \mid h_i(\boldsymbol{\sigma}_A(t))]\}\right] \tag{34}$$

The joint probability density for $\boldsymbol{\phi}_i(t)$ follows from partial derivatives with respect to $\phi_i^l(t)$, which yields

$$f_\beta[\boldsymbol{\phi}_i(t)] = \beta^{Q-1}(Q-1)!\, \frac{\exp[-\beta \sum_{l=1}^{Q-1} \phi_i^l(t)]}{\{1+\sum_{l=1}^{Q-1} \exp[-\beta \phi_i^l(t)]\}^Q} \tag{35}$$

Unfortunately, this straightforward representation is not convenient to deal with the Q-Ising dynamics if $Q > 2$, since the auxiliary thermal fields cannot be treated in the same way as the energy potentials in the deterministic dynamics. A convenient method to recover this approach is to introduce another set of auxiliary fields. To this end, consider for each i and t auxiliary thermal fields $\boldsymbol{\psi}_i(t) \equiv \{\psi_i^1(t),..., \psi_i^Q(t)\}$ with joint probability density

$$\tilde{f}_\beta[\boldsymbol{\psi}_i(t)] \equiv \beta^{Q-1} Q! \frac{\exp[-\beta \sum_{k=1}^{Q} \psi_i^k(t)]}{\{\sum_{k=1}^{Q} \exp[-\beta \psi_i^k(t)]\}^Q} \delta\left(\sum_{k=1}^{Q} \psi^k\right) \tag{36}$$

We observe that this definition leads to the normalization

$$\int \left[\prod_{l=1}^{Q-1} d\phi^l\right] f_\beta[\boldsymbol{\phi}]\, \Omega[\boldsymbol{\phi}] = \int \left[\prod_{l=1}^{Q} d\psi^l\right] \tilde{f}_\beta[\boldsymbol{\psi}]\, \Omega[\{\psi^l - \psi^k\}_{l \neq k}] \tag{37}$$

for any fixed $k \in \{1,..., Q\}$ and an arbitrary function $\Omega(\cdot)$ of $Q-1$ variables. Given an index k, there is a one-to-one correspondence between $f_\beta(\cdot)$ and $\tilde{f}_\beta(\cdot)$. The introduction of (36) yields a third way to describe the transition probabilities (2), i.e.,

$$\begin{aligned}
&\Pr[\sigma_i(t+1) = s_k \,|\, \boldsymbol{\sigma}_\Lambda(t)] \\
&= \mathrm{E}_{\boldsymbol{\psi}}\left[\prod_{l \neq k} \theta(\varepsilon_i[s_l \,|\, h_i(\boldsymbol{\sigma}_\Lambda(t))] - \psi_i^l(t+1) - \varepsilon_i[s_k \,|\, h_i(\boldsymbol{\sigma}_\Lambda(t))] + \psi_i^k(t+1)])\right]
\end{aligned} \tag{38}$$

To make the link of the stochastic dynamics with the deterministic dynamics in a convenient way we finally observe that for each realization of the auxiliary thermal fields $\boldsymbol{\psi}_i(t+1)$ with density $\tilde{f}_\beta(\cdot)$, the network state evolves according to the deterministic rule

$$\begin{aligned}
\sigma_i(t+1) = \sum_{k=1}^{Q} s_k \prod_{l \neq k} \theta(&\varepsilon_i[s_l \,|\, h_i(\boldsymbol{\sigma}_\Lambda(t))] - \psi_i^l(t+1) \\
&- \varepsilon_i[s_k \,|\, h_i(\boldsymbol{\sigma}_\Lambda(t))] + \psi_i^k(t+1)]) \\
\equiv g_{\boldsymbol{\psi}}[h_i(&\boldsymbol{\sigma}_\Lambda(t))]
\end{aligned} \tag{39}$$

where the local field $h_i(\boldsymbol{\sigma}_\Lambda(t))$ is given by Eq. (1). In this way the problem of deriving recursion relations for $T \neq 0$ becomes tractable.

Consider the same initial conditions as in the zero-temperature case. The main overlaps and the activity of the neurons as well as the residual overlaps on the first layer are equal to those at zero temperature because the effect of the stochastic dynamics does not yet appear on the first layer.

The main overlaps and the activity on the second layer at finite temperatures get the form

$$m_A^\mu(2) = \frac{1}{AN} \sum_{i \in A} \xi_i^\mu(2)\, g_\Psi\left(\xi_i^1(2)\, m_A^1(1) + \frac{1}{\sqrt{N}} \sum_{\mu \in \mathscr{P}\backslash 1} \xi_i^\mu(2)\, r_A^\mu(1)\right) \quad (40)$$

$$a_A(2) = \frac{1}{N} \sum_{i \in A} g_\Psi^2\left(\xi_i^1(2)\, m_A^1(1) + \frac{1}{\sqrt{N}} \sum_{\mu \in \mathscr{P}\backslash 1} \xi_i^\mu(2)\, r_A^\mu(1)\right) \quad (41)$$

Since both the limiting residual overlaps $\{r^\mu(1)\}$ and the auxiliary thermal fields $\{\psi_i^l(2)\}$ are i.i.d.r.v., the main overlaps and the activity in the limit $N \to \infty$ can be computed by applying the LLN. This requires three types of averages. First, there is the average over the noise term associated with the noncondensed patterns, which, in the limit $N \to \infty$, is Gaussianly distributed with mean zero and variance $\alpha AD(1)$. Second, the average over the auxiliary thermal fields $\psi(2)$ with the probability density given by Eq. (36) can be done explicitly using the relation (38). Third, averaging over the condensed pattern, we obtain the main overlaps and the activity in the limit $N \to \infty$:

$$m^\mu(2) = \delta_{\mu,1} \frac{1}{A} \left\langle\!\!\left\langle \xi^1(2) \sum_{k=1}^{Q} s_k \int Dz\, \Xi_k(1) \right\rangle\!\!\right\rangle \quad (42a)$$

$$a(2) = \left\langle\!\!\left\langle \sum_{k=1}^{Q} s_k^2 \int Dz\, \Xi_k(1) \right\rangle\!\!\right\rangle \quad (42b)$$

where we have introduced the notation

$$\Xi_k(t) \equiv \left[1 + \sum_{l \neq k} \exp(-\beta\{\varepsilon[s_l \,|\, h(t+1)] - \varepsilon[s_k \,|\, h(t+1)]\})\right]^{-1} \quad (43)$$

$$h(t+1) \equiv \xi^1(t+1)\, m^1(t) + [\alpha AD(t)]^{1/2}\, z \quad (44)$$

The recursion relations (42) are again similar to those of the first-step dynamics of the recurrent Q-Ising model.[8]

For the configuration of the third layer, we need to characterize the statistical distribution of the residual overlaps $r_A^\mu(2) \equiv (A\sqrt{N})^{-1} \sum_{i \in A} \xi_i^\mu(2)\, \sigma_i(2)$ with

$$\sigma_i(2) \equiv g_\Psi\left(\xi^1(2)\, m_A^1(1) + \frac{1}{\sqrt{N}} \xi_i^\mu(2)\, r_A^\mu(1) + \frac{1}{\sqrt{N}} \sum_{\nu \in \mathscr{P}\backslash\{1,\mu\}} \xi_i^\nu(2)\, r_A^\nu(1)\right) \quad (45)$$

To this end, consider the characteristic function for the distributional limit $r^\mu(2)$ with $r^\mu(1)$ fixed,

$$\begin{aligned} &\mathrm{E}[\exp\{\mathrm{i}xr^\mu(2)\}] \\ &= \lim_{N\to\infty}\left(1+\frac{\mathrm{i}x}{A\sqrt{N}}\mathrm{E}[\xi_i^\mu(2)\,\sigma_i(2)]-\frac{x^2}{2A^2N}\mathrm{E}[\{\xi_i^\mu(2)\,\sigma_i(2)\}^2]+\cdots\right)^N \end{aligned} \tag{46}$$

where the average is taken over the random patterns $\xi_i^\mu(2)$, the noise term associated with the noncondensed patterns (except pattern μ), and the auxiliary thermal fields $\psi_i(2)$. After performing this last average, we expand the expectation values in Eq. (46) with respect to the small term $\xi_i^\mu(2)\,r_A^\mu(1)/\sqrt{N}$. We arrive at

$$\begin{aligned} &\mathrm{E}[\exp\{\mathrm{i}xr^\mu(2)\}] \\ &= \exp\left[-\frac{x^2}{2A}a(2)+\mathrm{i}xr^\mu(1)\left\langle\!\!\left\langle\sum_{k=1}^{Q}s_k\int Dz\,\frac{z\Xi_k(1)}{[\alpha AD(1)]^{1/2}}\right\rangle\!\!\right\rangle\right] \end{aligned} \tag{47}$$

such that the distribution of the random variable $r^\mu(2)$ reads

$$r^\mu(2)=\mathcal{N}_\mu\left(0,\frac{a(2)}{A}\right)+r^\mu(1)\left\langle\!\!\left\langle\sum_{k=1}^{Q}s_k\int Dz\,\frac{z\Xi_k(1)}{[\alpha AD(1)]^{1/2}}\right\rangle\!\!\right\rangle \tag{48}$$

Including the fact that $r^\mu(1)$ is also Gaussian with mean zero and variance $D(1)$, we find that the variance of $r^\mu(2)$ is finally given by

$$D(2)=\frac{a(2)}{A}+\frac{1}{\alpha A}\left[\left\langle\!\!\left\langle\sum_{k=1}^{Q}s_k\int Dz\,z\Xi_k(1)\right\rangle\!\!\right\rangle\right]^2 \tag{49}$$

The mutual independence of the $\{r^\mu(2)\}$ again persists because $\{r^\mu(1)\}$ is a collection of i.i.d.r.v. Hence, analogously to the $T=0$ case, we conclude that the recursion relations at finite temperatures are

$$m^\mu(t+1)=\delta_{\mu,1}\frac{1}{A}\left\langle\!\!\left\langle\xi^1(t+1)\sum_{k=1}^{Q}s_k\int Dz\,\Xi_k(t)\right\rangle\!\!\right\rangle \tag{50a}$$

$$a(t+1)=\left\langle\!\!\left\langle\sum_{k=1}^{Q}s_k^2\int Dz\,\Xi_k(t)\right\rangle\!\!\right\rangle \tag{50b}$$

$$D(t+1)=\frac{a(t+1)}{A}+\frac{1}{\alpha A}\left[\left\langle\!\!\left\langle\sum_{k=1}^{Q}s_k\int Dz\,z\Xi_k(t)\right\rangle\!\!\right\rangle\right]^2 \tag{50c}$$

where we recall that $\Xi_k(t)$ is given by Eq. (43). One can also derive Eqs. (50) through the use of the generating function method,[11] which resembles the treatment of the Potts model on a layered architecture.[10]

The corresponding recursion relations in the limit $Q \to \infty$ are given by Eqs. (50) after replacing the discrete sum over the possible neuron states by

$$\sum_{k=1}^{Q} s_k^n \Xi_k(t) \to \frac{\int_{-1}^{1} ds\, \chi(s)\, s^n \exp\{-\beta\varepsilon[s \mid h(t+1)]\}}{\int_{-1}^{1} ds\, \chi(s) \exp\{-\beta \in [s \mid h(t+1)]\}} \tag{51}$$

$\chi(\cdot)$ is the density of the neuron states. For a uniform distribution of states, which is the distributional limit of equidistant states, the density $\chi(\cdot)$ is again simply 1/2. We remark that Eqs. (50) generalize the results of ref. 5 for arbitrary finite Q and for the limit $Q \to \infty$.

5. ANALYSIS OF THE FIXED-POINT EQUATIONS

We first consider the case $Q = 3$, i.e., a neuron can take the values ± 1 and 0. Each component $\xi_i^\mu(t)$ of the stored patterns takes the values ± 1 and 0 with probabilities $A/2$, respectively $1 - A$. At zero temperature the recursion relations (30) reduce to

$$m^\mu(t+1) = \delta_{\mu,1} \frac{1}{2} (\mathrm{Erf}[X_+(t)] - \mathrm{Erf}[X_-(t)]) \tag{52a}$$

$$D(t+1) = \frac{a(t+1)}{A} + \frac{1}{2\pi\alpha A} [A(e^{-X_+^2(t)} + e^{-X_-^2(t)}) + 2(1-A)\, e^{-X_0^2(t)}]^2 \tag{52b}$$

with

$$a(t+1) \equiv 1 - \frac{A}{2} \{\mathrm{Erf}[X_+(t)] + \mathrm{Erf}[X_-(t)]\} - (1-A)\, \mathrm{Erf}[X_0(t)] \tag{53}$$

$$X_\eta(t) \equiv \frac{b + \eta m^1(t)}{[2\alpha A D(t)]^{1/2}} \qquad \text{for } \eta \in \{+1, 0, -1\} \tag{54}$$

The fixed-point equations can be read off from Eqs. (52) by setting $m^1(t+1) = m^1(t) = m$, $D(t+1) = D(t) = D$.

As is obvious from (4), the gain function reduces to the input–output relation of binary networks if $b = 0$. Introducing the variable $x \equiv m/(2\alpha A D)^{1/2}$ we find that fixed-point equations at $b = 0$ take the reduced from

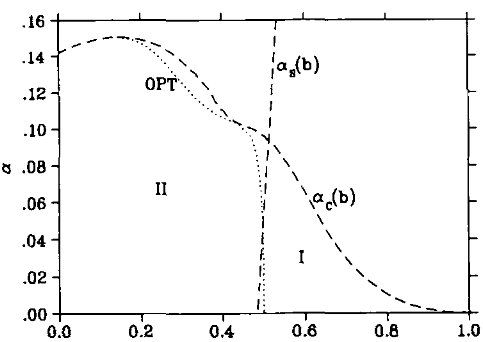
$$\alpha = \frac{1}{2} \left[\frac{\mathrm{Erf}(x)}{x}\right]^2 - \frac{2}{\pi} (A e^{-x^2} + 1 - A)^2 \tag{55}$$

Equation (55) possesses no solution for $\alpha > 0$ unless the activity of the patterns A is greater than 1/3. The maximum value of α at $b=0$, $\alpha_c(0) \approx 0.269$, is obtained for $A=1$ and it decreases as A gets smaller. For uniform patterns ($A=2/3$) we get $\alpha_c(0) \approx 0.142$. For $A=1$, $\alpha_c(0)$ is equal to that of the layered binary network.[5] We note, however, that the recursion relations (52) for $b=0$ reduce to those of the corresponding binary network only if $A=1$ (that is, if the stored patterns take the values ± 1 only). This behavior can be contrasted with the extremely diluted ($Q=3$)-state network, whose corresponding recursion relations become identical to those for the binary case at $b=0$ regardless of A.[8, 15]

At $b \neq 0$, we have solved the fixed-point equations numerically, and we have studied the stability properties of the solutions. The resulting (α, b) diagram is shown in Fig. 1. At any (α, b), $Z \equiv (m=0, D=0)$ is a stable fixed point. At the boundary $\alpha_C(b)$ a retrieval state $R \equiv (m \neq 0, D \neq 0)$ (dis)-appears discontinuously in m. It is an attractor of the dynamics, accompanied by a nonretrieval state $NR \equiv (m \neq 0, D \neq 0)$, which is a saddle point solution to the fixed-point equations. The retrieval quality is measured by the Hamming distance (8) between R and the embedded pattern. If the network recalls the pattern without error, the Hamming distance is zero. At fixed α, in the retrieval region, the retrieval quality is a nonmonotonous function of b. Therefore an optimal b can be determined. The corresponding line in the (α, b) plane is also displayed in Fig. 1.

For a sustained activity solution $S \equiv (m=0, D>0)$, Eqs. (52) can be reduced to

$$D = \frac{1}{A}\left[1 - \operatorname{Erf}\left(\frac{b}{(2\alpha AD)^{1/2}}\right)\right] + \frac{2}{\pi A \alpha} \exp\left(-\frac{b^2}{\alpha AD}\right) \tag{56}$$



	I	II
R	a	a
NR	s	s
S_1	-	a
S_2	-	r
Z	a	a

Fig. 1. The (α, b) diagram for the $Q=3$ network at zero temperature with uniform patterns ($A=2/3$). The line $\alpha_C(b)$ denotes the boundary of the retrieval region. The line $\alpha_S(b)$ is the lower bound for the existence of the sustained activity states. The line OPT is the line of the best retrieval quality. The structure of the retrieval dynamics is explained: a denotes an attractor, s a saddle point, r a repellor.

For all $\alpha > 0$ and $b < b_0 = (2/\pi e)^{1/2}$, Eq. (56) possesses a pair of nonzero solutions independent of A. The solution with the higher D (denoted S_1) is always stable, the one with the smaller D (denoted S_2) is a saddle point. The transition from S_1 and S_2 to Z is again discontinuous. For $b > b_0$, there is a value of α, indicated by $\alpha_S(b)$ in Fig. 1, below which no nonzero solutions to (56) exist. We remark that in the corresponding extremely diluted network such a pair of sustained activity solutions exist only if $\alpha \geqslant (b/B)^2$ ($B \approx 0.576$) such that in this case there is, for any b, an α interval which does not allow a sustained activity state.[8,15]

At finite temperatures, the fixed-point equations are read off from Eqs. (50). They are again solved numerically. Figure 2 shows the evolution of the dynamical (α, b) diagram with increasing temperature, in particular the retrieval boundary $\alpha_C(b)$ (Fig. 2a) as well as the line of optimal

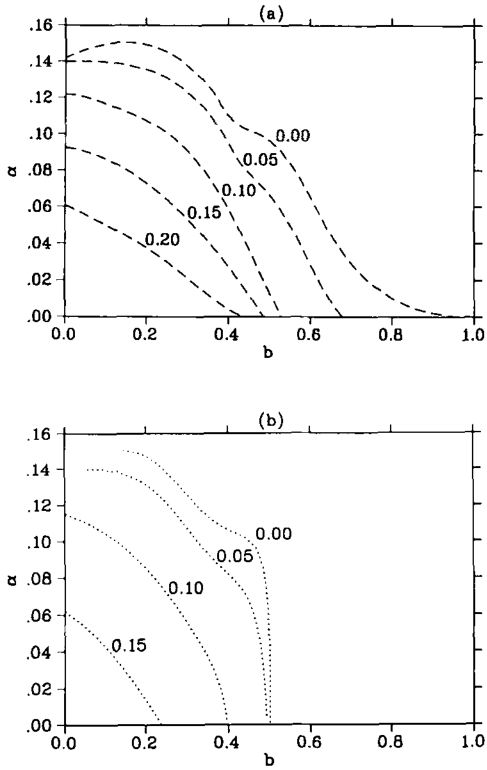


Fig. 2. The (α, b) diagrams for the $Q = 3$ network at finite temperatures with uniform patterns ($A = 2/3$). (a) The boundary of the retrieval region; (b) the line of the best retrieval quality.

retrieval (Fig. 2b). The retrieval state R (dis)appears discontinuously at the boundary $\alpha_C(b)$, where it coalesces with a saddle point NR. The zero solution, which is always stable at $T=0$, does not exist at finite T. Since the thermal noise tends to randomize the neuron states, it is turned into a sustained activity solution. At sufficiently low temperatures, the structure of the retrieval dynamics is the same as for zero temperature. However, at moderate temperatures, there is only one sustained activity solution left, which is stable. In general, the dynamical structure of the $(Q=3)$-model is hardly temperature dependent: there is always an attractor on the line $m=0$ such that the retrieval state never attracts the whole (m, D) plane. In contrast, for the extremely diluted network the retrieval state turns out to be the only attractor in the system for appropriately chosen (α, b).

Next we turn to the case of analog neurons $(Q \to \infty)$. We restrict our discussion to a uniform distribution of states and stored patterns $(A = 1/3)$ between -1 and $+1$. In this case the gain function is given by Eq. (6) and the density functions $\rho(\cdot)$ and $\chi(\cdot)$ are 1/2. The fixed-point equations at zero temperature follow immediately from Eqs. (31). At $b=0$, they can be condensed into one equation

$$\alpha = \frac{9}{8x^2}\left[\left(1-\frac{1}{2x^2}\right)\mathrm{Erf}(x) + \frac{e^{-x^2}}{\sqrt{\pi}\,x}\right]^2 - \frac{\mathrm{Erf}^2(x)}{2x^2} \tag{57}$$

with $x \equiv m/(2\alpha A D)^{1/2}$. The maximum value of α at $b=0$ is $\alpha_c(0) \approx 0.106$.

For general b, the fixed-point equations are solved numerically. The resulting (α, b) diagram is shown in Fig. 3. The zero solution Z is always a fixed point. In contrast to the case $Q=3$, Z is stable only if it is the only solution. Otherwise, Z is unstable. At the boundary $\alpha_C(b)$ a retrieval state

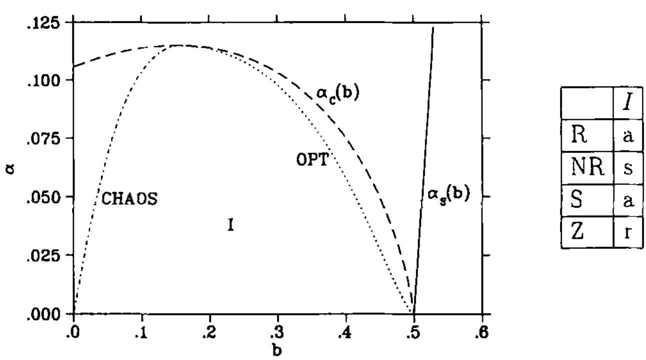


Fig. 3. As in Fig. 1, for the piecewise linear gain function at $T=0$ for uniform patterns $(A=1/3)$. Above the dynamical transition line *CHAOS* chaotic behavior occurs.

R (dis)appears discontinuously in m. It is an attractor of the dynamics, accompanied by a nontrieval state NR, which is a saddle point solution to the fixed-point equations. Again a line of optimized b at given α is displayed in Fig. 3. We observe that there is a qualitative difference from the corresponding extremely diluted network. For small b, both systems show an increasing storage capacity $\alpha_C(b)$. However, $\alpha_C(b)$ of the layered network eventually decreases to zero as b approaches 1/2, whereas in the extremely diluted case $\alpha_C(b)$ is increasing up to $b = 1/2$. Furthermore, the former shows a first-order transition, while the latter exhibits a second-order transition at the retrieval boundary.

The fixed-point equation for a sustained activity solution S is

$$D = \frac{a}{A} + \frac{D}{4b^2} \operatorname{Erf}^2\left(\frac{2b}{(2\alpha AD)^{1/2}}\right) \tag{58}$$

with

$$a = 1 - \left(1 - \frac{\alpha AD}{4b^2}\right) \operatorname{Erf}\left(\frac{2b}{(2\alpha AD)^{1/2}}\right) - \frac{(2\alpha AD)^{1/2}}{2b\sqrt{\pi}} \exp\left(-\frac{2b^2}{\alpha AD}\right) \tag{59}$$

Equation (58) possesses one nonzero solution for all α unless $b > 1/2$. For $b > 1/2$, the value of D decreases as α gets smaller and eventually vanishes at $\alpha_S(b) = 4b^2 - 1$. This means that the transition from S to Z is now second order. [The corresponding extremely diluted network also shows a continuous transition from the sustained activity state to the zero state at $\alpha = (2b)^2$.] The solution S is always stable. Consequently, we again conclude that R is never the only attractor of the dynamics such that the basin of attraction of R is always limited by NR to separate it from the attractor S on the axis $m = 0$.

The transition at $\alpha_S(b)$ is very similar to the spin-glass transition in the binary network.[16] In fact, the gain function (6) is an example of a more general sigmoid input–output relation in a network of analog neurons. Roughly speaking, the piecewise linear gain function mimics the transition probability $[1 + \tanh(h)]/2$ in networks of binary neurons at finite temperatures. The role of the temperature is then taken over by the gain parameter b.

6. CHAOTIC BEHAVIOR OF THE DYNAMICS

Until now we have described the dynamical properties of a layered multistate neural network in terms of macroscopic quantities, viz. the overlap, the activity, and the variance of the residual overlaps. More specifically, we have shown that for appropriately chosen parameters b and α the network functions as an associative memory. It is nevertheless interesting to investigate the nature of the deterministic dynamics in the retrieval regime on the microscopic level. In this context it is known[7] that a layered network of binary neurons exhibits chaotic behavior. As in refs. 7 and 8 we use the term "chaotic" in a well-defined sense. The dynamics is called chaotic if two network configurations which are initially close in Hamming distance and correlated with only one embedded pattern repel each other. It is clear that this criterion of diverging trajectories is a microscopic property of the network.

Let us consider two different initial configurations $\boldsymbol{\sigma}_\Lambda(1)$ and $\tilde{\boldsymbol{\sigma}}_\Lambda(1)$, which are collections of i.i.d.r.v. with mean zero and variance a_0 and $\tilde{a}_0$, respectively. [In the sequel, variables with a tilde are to be understood as corresponding to the initial condition $\tilde{\boldsymbol{\sigma}}_\Lambda(1)$.] They have a finite projection on one pattern, say the first one, they have zero projection on the other patterns, and are also mutually correlated. Explicitly, in the limit $N \to \infty$

$$\begin{aligned} \mathrm{E}[\sigma_i(1)\,\tilde{\sigma}_j(1)] &= C_0 \delta_{i,j} \\ \mathrm{E}[\xi_i^\mu(1)\,\sigma_i(1)] &= \delta_{\mu,1} m_0 A \\ \mathrm{E}[\xi_i^\mu(1)\,\tilde{\sigma}_i(1)] &= \delta_{\mu,1} \tilde{m}_0 A \end{aligned} \tag{60}$$

By the same reasoning as in the case of a single configuration, the main overlap, the activity, and the variance of the residual overlaps for both configurations evolve according to Eq. (30) or more generally Eq. (31). We thus obtain $m(t)$, $a(t)$, and $D(t)$ as well as the corresponding variables with tilde.

The Hamming distance between $\boldsymbol{\sigma}_\Lambda(t)$ and $\tilde{\boldsymbol{\sigma}}_\Lambda(t)$ is given by

$$d_{\mathrm{H}}[\boldsymbol{\sigma}_\Lambda(t), \tilde{\boldsymbol{\sigma}}_\Lambda(t)] = a_\Lambda(t) + \tilde{a}_\Lambda(t) - \frac{2}{N} \sum_{i \in \Lambda} \sigma_i(t)\,\tilde{\sigma}_i(t) \tag{61}$$

Consequently, to describe the time evolution of this Hamming distance it is necessary to know the evolution of the correlation between the configurations $\boldsymbol{\sigma}_\Lambda(t)$ and $\tilde{\boldsymbol{\sigma}}_\Lambda(t)$. We denote this correlation by

$$C_\Lambda(t) \equiv \frac{1}{N} \sum_{i \in \Lambda} \sigma_i(t)\,\tilde{\sigma}_i(t) \tag{62}$$

where, according to the evolution rule (5) and Eq. (14),

$$\sigma_i(t+1) \equiv g\left(\xi_i^1(t+1)\, m_\Lambda^1(t) + \frac{1}{\sqrt{N}} \sum_{\mu \in \mathscr{P}\setminus 1} \xi_i^\mu(t+1)\, r_\Lambda^\mu(t)\right) \tag{63}$$

By the LLN, the correlation on the first layer $t=1$ then gets the form

$$C(1) \equiv \lim_{N \to \infty} C_A(1) \overset{\mathrm{Pr}}{=} \mathrm{E}[\sigma_i(1)\,\tilde{\sigma}_i(1)] = C_0 \tag{64}$$

On layer $t+1$, the characteristic function of the noise terms associated with the residual overlaps in the expressions for $C_A(t+1)$ is, in the limit $N \to \infty$, given by

$$\lim_{N \to \infty} \mathrm{E}\left[\exp\left(\frac{\mathrm{i}x}{\sqrt{N}} \sum_{\mu>1} \xi_i^\mu(t+1)\, r_A^\mu(t) + \frac{\mathrm{i}y}{\sqrt{N}} \sum_{\mu>1} \xi_i^\mu(t+1)\, \tilde{r}_A^\mu(t)\right)\right] \tag{65}$$

$$= \exp\left[-\frac{1}{2}\alpha A D(t)\, x^2 - \frac{1}{2}\alpha A \tilde{D}(t)\, y^2 - \alpha A S(t)\, xy\right] \tag{66}$$

where we have introduced the residual correlation

$$S(t) \equiv \mathrm{E}[r^\mu(t)\, \tilde{r}^\mu(t)] \tag{67}$$

We see that the noise terms are Gaussian random variables with mean zero and variance $\alpha A D(t)$ and $\alpha A \tilde{D}(t)$, respectively, and that they are mutually correlated with correlation $\alpha A S(t)$. Applying LLN, the correlation between two configurations becomes

$$C(t+1) \equiv \lim_{N \to \infty} C_A(t+1) = \left\langle\!\!\left\langle \int d\omega\, d\tilde{\omega}\, \mathscr{R}_t(\omega, \tilde{\omega}) \right.\right.$$
$$\left.\left. \times\, g[\xi^1(t+1)\, m^1(t) + \omega]\, g[\xi^1(t+1)\, \tilde{m}^1(t) + \tilde{\omega}] \right\rangle\!\!\right\rangle \tag{68}$$

where

$$\mathscr{R}_t(\omega, \tilde{\omega}) \equiv \frac{\exp\{-[\tilde{D}(t)\,\omega^2 + D(t)\,\tilde{\omega}^2 - 2S(t)\,\omega\tilde{\omega}]/2\alpha A[D(t)\,\tilde{D}(t) - S^2(t)]\}}{2\pi\alpha A[D(t)\,\tilde{D}(t) - S^2(t)]^{1/2}} \tag{69}$$

We now derive the recursion relation for the quantity S defined in (67). First note that $S(t=1)$ can easily be calculated because $\sigma_i(1)$ and $\tilde{\sigma}_j(1)$ are independent if $i \neq j$:

$$S(1) = \lim_{N \to \infty} \frac{1}{A^2 N} \sum_{i \in A} \mathrm{E}[\{\xi_i^\mu(1)\}^2]\, \mathrm{E}[\sigma_i(1)\, \tilde{\sigma}_i(1)] = C_0/A \tag{70}$$

We recall that [see Eqs. (26)] in the limit $N \to \infty$, $\omega^{\mu}_{i,A}(t+1)$ converges to a Gaussian random variable with mean zero and variance $\alpha A D(t)$. The distributional limits $\omega^{\mu}_{i}(t+1)$ and $\tilde{\omega}^{\mu}_{j}(t+1)$ are independent for different sites i and j because $\{\xi^{\nu}_{i}(t+1)\}$, $\nu \in \mathscr{P}\backslash\{1, \mu\}$, and $\{\xi^{\nu}_{j}(t+1)\}$, $\nu \in \mathscr{P}\backslash\{1, \mu\}$, are independent if $i \neq j$. For $i = j$, the correlation between $\omega^{\mu}_{i}(t+1)$ and $\tilde{\omega}^{\mu}_{i}(t+1)$ is given by $\alpha A S(t)$ such that the joint density for $\omega^{\mu}_{i}(t+1)$ and $\tilde{\omega}^{\mu}_{i}(t+1)$ is exactly $\mathscr{R}_t(\cdot, \cdot)$ [see Eq. (69)].

To obtain the residual correlation $S(t+1)$, it is convenient to separate the product $r^{\mu}_{A}(t+1)\, \tilde{r}^{\mu}_{A}(t+1)$ into two contributions,

$$r^{\mu}_{A}(t+1)\, \tilde{r}^{\mu}_{A}(t+1) = \frac{1}{A^2 N} \sum_{i \in A} [\xi^{\mu}_{i}(t+1)]^2\, \sigma_i(t+1)\, \tilde{\sigma}_i(t+1)$$
$$+ \frac{1}{A^2 N} \sum_{i \neq j} \xi^{\mu}_{i}(t+1)\, \sigma_i(t+1)\, \xi^{\mu}_{j}(t+1)\, \tilde{\sigma}_j(t+1) \tag{71}$$

It is sufficient to calculate the expectation value of the first term in Eq. (71) up to the zeroth order in $O(1/\sqrt{N})$. Doing so, the noise terms related with the μth pattern eventually drop out after averaging over the joint distribution $\mathscr{R}_t(\cdot, \cdot)$. This yields $C(t+1)/A$. However, the second term in Eq. (71) contributes a term of $O(1)$. Taking these remarks into account, we find in the limit $N \to \infty$

$$S(t+1) = \frac{C(t+1)}{A} + \frac{S(t)}{\alpha A [D(t)\, \tilde{D}(t)]^{1/2}}$$
$$\times \left\langle\!\!\left\langle \int Dz\, z g(\xi^1(t+1)\, m^1(t) + [\alpha A D(t)]^{1/2} z) \right\rangle\!\!\right\rangle$$
$$\times \left\langle\!\!\left\langle \int D\tilde{z}\, \tilde{z} g(\xi^1(t+1)\, \tilde{m}^1(t) + [\alpha A \tilde{D}(t)]^{1/2} \tilde{z}) \right\rangle\!\!\right\rangle \tag{72}$$

This expression, together with Eq. (68), is our starting point for the discussion of the chaotic properties of the microscopic dynamics.

To this end, we now restrict the initial conditions to the case of two configurations having the same projection $[m^1(t) = \tilde{m}^1(t)]$ on the first pattern and the same activity $[a(t) = \tilde{a}(t)$, and hence $D(t) = \tilde{D}(t)]$. Since for identical configurations $\boldsymbol{\sigma}(t) = \tilde{\boldsymbol{\sigma}}(t)$ it is clear that $C(t) = a(t)$ and $D(t) = S(t)$, we study the stability of the fixed point $D = S$ to determine whether or not the dynamics is chaotic. The recursion relation for $D - S$ is easily obtained from Eq. (30c) and Eq. (72),

$$
\begin{aligned}
D(t+1)-S(t+1) &\\
&= \frac{1}{A}\left[a(t+1)-C(t+1)\right] \\
&\quad + \frac{D(t)-S(t)}{\alpha A D(t)} \left\langle\!\!\left\langle \int Dz\, z g(\xi^1(t+1)\, m^1(t) + [\alpha A D(t)]^{1/2} z) \right\rangle\!\!\right\rangle^2
\end{aligned}
\tag{73}
$$

Because $C(t+1)$ also depends on $D(t)-S(t)$, we need to expand $C(t+1)$ with respect to this small quantity. We first consider the case of finite Q. Integration over the variable $\tilde{\omega}$ in Eq. (68) yields

$$
C(t+1) = \frac{1}{2\sqrt{\pi}} \sum_{k,l=1}^{Q} s_k s_l \left\langle\!\!\left\langle \Gamma_{k,l}(t) + \Gamma_{k-1,l-1}(t) - \Gamma_{k-1,l}(t) - \Gamma_{k,l-1}(t) \right\rangle\!\!\right\rangle
\tag{74}
$$

where we have introduced the short-hand notation

$$
\Gamma_{k,l}(t) \equiv \int_{-\infty}^{X_k(t)} dz\, e^{-z^2}\, \mathrm{Erf}\left(\frac{D(t)\, X_l(t) - S(t)\, z}{[D^2(t)-S^2(t)]^{1/2}}\right)
\tag{75}
$$

with $X_k(t)$ given by Eq. (20). In the limit $D(t)-S(t) \to 0$ the integral $\Gamma_{k,l}(t)$ can be evaluated

$$
\begin{aligned}
\Gamma_{k,l}(t) \approx \frac{\sqrt{\pi}}{2}\{1+\mathrm{Erf}[X_k(t)]\} - \delta_{k,l}\frac{[D^2(t)-S^2(t)]^{1/2}}{D(t)\sqrt{\pi}} \exp[-X_k^2(t)] &\\
- \theta(k-l)\sqrt{\pi}\,\{\mathrm{Erf}[X_k(t)] - \mathrm{Erf}[X_l(t)]\} + O(D-S)&
\end{aligned}
\tag{76}
$$

Now from Eqs. (75) and (76), it is straightforward to show (up to leading order in $D-S$) that

$$
C(t+1) \approx a(t+1) - \frac{[D^2(t)-S^2(t)]^{1/2}}{2\pi D(t)} \sum_{k=1}^{Q-1} (s_{k+1}-s_k)^2 \left\langle\!\!\left\langle \exp[-X_k^2(t)] \right\rangle\!\!\right\rangle
\tag{77}
$$

and hence Eq. (73) becomes

$$
D(t+1)-S(t+1) \approx \frac{[D^2(t)-S^2(t)]^{1/2}}{2\pi A D(t)} \sum_{k=1}^{Q-1} (s_{k+1}-s_k)^2 \left\langle\!\!\left\langle \exp[-X_k^2(t)] \right\rangle\!\!\right\rangle
\tag{78}
$$

The last term in Eq. (73) drops out in Eq. (78) because it is linear in $D-S$. Equation (78) shows that the fixed point $D=S$ is always unstable for finite Q since an arbitrary small deviation from the fixed point will result in a rapid, square-root divergence. This type of divergence has also been found in neural networks of binary neurons[7] and in extremely diluted networks of Q-Ising neurons.[9] We note that the gain function $g(\cdot)$ for general Q is indeed a generalization of the input–output relation sign $(\cdot)$ for two-state networks. Both show a finite number of jumps in their output as the input varies, such that the state of a neuron can change at these jumping points as a consequence of a small perturbation of the input. Square-root terms are essentially contributions from these jumping points.

Therefore, it is also interesting to note that the coefficient of the r.h.s. of Eq. (78) indeed decreases to zero as Q^{-1}. More explicitly, if the possible states $\{s_k\}$ are equally distributed, then for large Q, Eq. (78) reads

$$D(t+1)-S(t+1)\approx\frac{2}{Q}\frac{[D^2(t)-S^2(t)]^{1/2}}{\pi AD(t)} \times\left\langle\!\!\left\langle\int_{-1}^{1}d\eta\exp\left(-\frac{[2b\eta-\xi^1(t+1)\,m^1(t)]^2}{2\alpha AD(t)}\right)\right\rangle\!\!\right\rangle \tag{79}$$

This implies that the behavior in the limit $Q\to\infty$ will qualitatively differ from that for finite Q. For $Q\to\infty$ and using the piecewise linear gain function (6) for uniformly distributed states, we again expand the recursion relation for D—S up to leading order starting from Eq. (73). After a lot of tedious calculations, we get

$$D(t+1)-S(t+1)=\frac{D(t)-S(t)}{4b^2}\left\{\left[\left\langle\!\!\left\langle\operatorname{Erf}\left[\frac{2b+\xi^1(t+1)\,m^1(t)}{[2\alpha AD(t)]^{1/2}}\right]\right\rangle\!\!\right\rangle\right]^2 +\alpha\left\langle\!\!\left\langle\operatorname{Erf}\left[\frac{2b+\xi^1(t+1)\,m^1(t)}{[2\alpha AD(t)]^{1/2}}\right]\right\rangle\!\!\right\rangle\right\}+O[(D-S)^2] \tag{80}$$

The leading order in $D-S$ is now linear, which implies that the fixed point $D=S$ can be either stable or unstable depending on the absolute value of the coefficient of the r.h.s. of Eq. (80) compared to unity. Figure 3 shows the dynamical transition line to the chaotic behavior. The chaotic region is relatively smaller than for the extremely diluted networks of analog neurons. More specifically, the retrieval dynamics is not chaotic for sufficiently high b. As expected, the sustained activity regime is always chaotic.

We note that in contrast to the case of finite Q, there is no square root contribution to the recursion for $D-S$. This is due to the absence of jumping points in the gain function $g(\cdot)$.

For $\alpha \to 0$, we can expand the fixed-point equations (30) for the order parameters. For finite Q, the Hamming distances read in this limit

$$d_H(\xi^1, \sigma) = A + \hat{a} - 2A\hat{m} + \frac{1}{\sqrt{\pi}} \left\langle\!\!\left\langle \sum_{k=1}^{Q-1} (s_{k+1} - s_k) \times \left(\frac{s_{k+1} + s_k}{2} - \xi^1 \right) \frac{\exp(-\hat{X}_k^2)}{\hat{X}_k} \right\rangle\!\!\right\rangle \tag{81}$$

$$d_H(\sigma, \tilde{\sigma}) = \frac{1}{\pi^2 \hat{a}} \left[\sum_{k=1}^{Q-1} (s_{k+1} - s_k)^2 \langle\!\langle \exp(-\hat{X}_k^2) \rangle\!\rangle \right]^2 \tag{82}$$

$$\hat{X}_k \equiv [-\xi^1 \hat{m} + b(s_{k+1} + s_k)]/(2\alpha\hat{a})^{1/2} \tag{83}$$

where $\hat{m}$ and $\hat{a}$ now denote the main overlap and the activity at $\alpha = 0$. Since $\hat{X}_k$ is proportional to $1/\sqrt{\alpha}$ and $A + \hat{a} - 2A\hat{m}$ is nonnegative, we conclude that $d_H(\xi^1, \sigma) \gg d_H(\sigma, \tilde{\sigma})$. This relation also holds in the limit $Q \to \infty$, although in this case there appear terms linear in α [recall Eq. (80)] instead of the exponential terms. These relations are a generalization of those found in networks of binary neurons.[7]

7. CONCLUDING REMARKS

Using a probabilistic approach, we have analyzed the retrieval regime of layered networks of multistate neurons starting from the exact solution of the dynamics. First, evolution equations have been derived for arbitrary Q at both zero and finite temperatures. By introducing the auxiliary thermal fields in such a way that they can be treated in the same manner as the energy potentials, the stochastic dynamics can be formulated within the gain function framework of the deterministic dynamics. Next, we have investigated in detail the fixed-point equations for $Q = 3$ and in the limit $Q \to \infty$. There are three different types of fixed points, depending on the parameters of the model: a retrieval fixed point, a sustained activity fixed point, and the trivial fixed point. In contrast to the extremely diluted case, the retrieval state is always accompanied by an attractor which has zero overlap with the embedded pattern. In all cases under consideration the retrieval state disappears discontinuously as the storage capacity α increases. At zero temperature, the transition from the sustained activity state to the trivial state for $Q = 3$ is first order, while for $Q \to \infty$ it is second order. Finally, we have tackled the problem of the behavior of the system with deterministic dynamics in the configuration space. A type of chaoticity in the network trajectories is always present for arbitrary finite Q. However, in the case of a piecewise linear gain function there exists a dynamical transition toward chaos in the (α, b) plane. The (α, b) region where chaos does occur is relatively smaller than in the corresponding extremely diluted networks.

ACKNOWLEDGMENTS

This work has been supported in part by the Research Fund of the K. U. Leuven (grant OT/91/13). The authors are indebted to V. A. Zagrebnov for stimulating and clarifying discussions. One of us (D.B.) thanks the Belgian National Fund for Scientific Research for support as Research Director.

REFERENCES

1. E. Domany and R. Meir, in *Models of Neural Networks*, E. Domany, J. L. van Hemmen, and K. Schulten, eds. (Springer, Berlin, 1991), p. 307.
2. A. E. Patrick and V. A. Zagrebnov, *J. Phys.* (Paris) **51**:1129 (1990).
3. A. E. Patrick and V. A. Zagrebnov, *J. Stat. Phys.* **63**:59 (1991).
4. E. Domany, R. Meir, and W. Kinzel, *Europhys. Lett.* **2**:175 (1986).
5. R. Meir and E. Domany, *Phys. Rev. Lett.* **59**:359 (1987); *Europhys. Lett.* **4**:645 (1987); *Phys. Rev. A* **37**:608 (1988).
6. E. Domany, W. Kinzel, and R. Meir, *J. Phys. A* **22**:2081 (1989).
7. B. Derrida and R. Meir, *Phys. Rev. A* **38**:3116 (1988).
8. D. Bollé, B. Vinck, and V. A. Zagrebnov, *J. Stat. Phys.* **70**:1099 (1993).
9. D. Bollé, G. M. Shim, B. Vinck, and V. A. Zagrebnov, Retrieval and chaos in extremely diluted Q-Ising neural networks, *J. Stat. Phys.*, this issue.
10. G. M. Shim, D. Kim, and M. Y. Choi, *Phys. Rev. A* **45**:1238 (1992).
11. E. Gardner, B. Derrida, and P. Mottishaw, *J. Phys.* (Paris) **48**:741 (1987).
12. H. Rieger, *J. Phys. A* **23**:L1273 (1990).
13. A. N. Shiryayev, *Probability* (Springer, New York, 1984).
14. C. Maes and V. A. Zagrebnov, private communication.
15. J. S. Yedidia, *J. Phys. A* **22**:2265 (1989).
16. D. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. Lett.* **55**:1530 (1985); *Ann. Phys.* (N.Y.) **173**:30 (1987).