Nonmonotonic behavior of the capacity in phasor neural networks

D. Bollé*,† and G. M. Shim†,‡
Instituut voor Theoretische Fysica, Katholieke Universiteit Leuven, B-3001 Leuven, Belgium
(Received 31 May 1994)

The stochastic dynamics of Q-phasor neural networks is discussed using a probabilistic approach. For layered feedforward architectures and Hebbian learning, exact evolution equations are given for arbitrary Q at both zero and finite temperatures. The capacity-temperature diagram is presented. At zero temperature a nonmonotonic behavior of the capacity is found as a function of the number of phases Q, contrary to other multistate neural network models.

PACS number(s): 87.10.+e, 64.60.Cn, 75.10.Hk

I. INTRODUCTION

In the literature there exist by now rather detailed discussions on the capacity and retrieval properties of Q-state Potts and Q-state Ising neural networks. In particular, extremely diluted, layered feedforward and fully connected architectures have been considered, the main difference between those caused by the amount of correlations among the neurons (see, e.g., [1–3]). A missing piece in these discussions on multistate models is a study of the parallel dynamics for the clock or phasor model, especially on a layered architecture. These types of models have some interest in the processing of signals with a circular symmetry such as the orientations of edges or the directions of optic flow in images or phase patterns over detector arrays. The dynamics of the extremely diluted version of this model has been discussed in [4,5]. Concerning the fully connected clock model, a replica-symmetric mean-field theory approach has been derived in [6]. The first step dynamics for zero temperature and $Q \leq 4$ has been treated in [7]. Recently this approach has been generalized to finite temperatures [8]. Finally the Gardner optimal capacity has been obtained in [9].

In this work we focus on the stochastic dynamics of the Q-phasor layered feedforward network with Hebbian couplings between adjacent layers and independently chosen representations of patterns on different layers. This allows an exact treatment of the parallel dynamics, in contrast with fully connected networks. The main underlying reason is that in the layered model as well as in extremely diluted models there are no feedback loops. A difference between the layered and the extremely diluted architecture is that in the latter, 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 can be handled exactly. Moreover, the layered architecture is especially interesting from a practical point of view, e.g., in engineering problems.

Following a probabilistic approach [10] and making a careful signal-to-noise ratio analysis we derive exact layer-to-layer evolution equations at zero and at finite temperatures for arbitrary Q. By using complex noises, all Q including the case $Q \to \infty$ can be treated explicitly in the same way without any difficulties. This is *a fortiori* true for extremely diluted architectures (compare [8]).

We present capacity-temperature diagrams and write down the nature of the retrieval transition. For zero temperature we give, as far as we are aware, a new formula to calculate the critical capacity for all architectures mentioned above. In the case of the layered feedforward and fully connected models, the critical capacity shows a nonmonotonic behavior as a function of Q, contrary to the other multistate models. For comparison we provide an exhaustive list of possible behaviors in the Q-state Potts and Q-state Ising networks.

II. MODEL

The model we consider is composed of multistate clock or phasor neurons arranged in layers, each layer containing N neurons. A neuron can take (complex) values in the set

$$\mathcal{S}=\left\{s_k=\exp\left[i\frac{2\pi}{Q}k\right]\,\middle|\,k=0,1,\ldots,Q-1\right\}. \tag{1}$$

Each neuron on layer t is unidirectionally connected to all neurons on layer $t+1$. Given a configuration $\sigma(t)$ the local field in neuron i on layer $t+1$ is given by

$$h_i(\sigma(t))=\sum_{j=1}^{N} J_{ij}(t+1)\sigma_j(t)\,, \tag{2}$$

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 $\sigma(t+1)$ of layer $t+1$ is determined by the state $\sigma(t)$ of the previous layer t according to the transition probabilities

*Electronic address: Desire.Bolle@fys.kuleuven.ac.be
†Also at Interdisciplinair Centrum voor Neurale Netwerken, Katholieke Universiteit Leuven, Belgium.
‡Electronic address: gyoung@tfdec1.fys.kuleuven.ac.be

$$\Pr[\sigma_i(t+1)=s_k\in\mathcal{S}|\boldsymbol{\sigma}(t)] = \frac{\exp\{\beta\,\mathrm{Re}[h_i(\boldsymbol{\sigma}(t))s_k^*]\}}{\sum_{l=0}^{Q-1}\exp\{\beta\,\mathrm{Re}[h_i(\boldsymbol{\sigma}(t))s_l^*]\}}\ , \tag{3}$$

where the temperature $T=\beta^{-1}$ measures the noise level. We consider parallel updating. The configuration of the first layer $\boldsymbol{\sigma}(t=1)$ is chosen as input. At the next time step, the second layer is updated according to the rule (3), and so on. At zero temperature the dynamics is then given in terms of the gain function

$$\sigma_i(t+1)=g(h_i(\boldsymbol{\sigma}(t))) =\exp\left\{i\frac{2\pi}{Q}\mathrm{Int}\left[\frac{Q}{2\pi}\mathrm{Arg}[h_i(\boldsymbol{\sigma}(t))]+\frac{\pi}{Q}\right]\right\}, \tag{4}$$

where Int[ ] denotes the integer part of the expression between large square brackets.

Patterns are stored in this network with the Hebb rule. The representation of the patterns on layer t is a collection of independent and identically distributed random variables (IIDRVs) $\{\xi_i^\mu(t)\in\mathcal{S}\}$, $\mu\in\{1,2,\ldots,p=\alpha N\}$ with zero mean and variance 1. The synaptic couplings between adjacent layers are then chosen according to

$$J_{ij}(t+1)=\frac{1}{N}\sum_{\mu=1}^{p=\alpha N}\xi_i^\mu(t+1)[\xi_j^\mu(t)]^*\ , \tag{5}$$

where the ξ_i^μ and the J_{ij} are complex.

This model generalizes the phasor model introduced in [5] in two directions. First, the stochastic dynamics is not defined via Gaussian noise, but through a finite temperature spin dynamics (3). It is easy to check that the latter obeys the detailed balance condition if $J_{ij}=J_{ji}^*$ (and $J_{ii}=0$) and hence is governed by an Hamiltonian, leading in the fully connected case to the Q-state clock Hamiltonian studied in [6]. Second, the architecture is taken to be layered feedforward such that correlations between the neurons exist because of the common ancestor problem (see, e.g., [10]), contrary to the extremely diluted architecture where any finite number of neurons have disjoint clusters of ancestors so that they are completely uncorrelated.

III. RETRIEVAL DYNAMICS

The retrieval quality of this layered Q-phasor network can be measured by the main overlap between a stored pattern and the microscopic state of the network

$$m^\mu(t)=\frac{1}{N}\sum_i[\xi_i^\mu(t)]^*\sigma_i(t)\ . \tag{6}$$

We now write down exact evolution equations for this model both at $T=0$ and $T\neq 0$ using a careful signal-to-noise ratio analysis based on the law of large numbers and the central limit theorem. We can pursue the calculation step by step in analogy with the one detailed in [10], taking into account that we are working here with complex noises. The results are the following. Supposing that the initial data $\boldsymbol{\sigma}(1)$ are a collection of IIDRVs with mean zero, variance 1, and correlated with only one stored pattern, say, the first one ($\mu=1$), i.e.,

$$E\{\sigma_i(1)[\xi_i^\mu(1)]^*\}=\delta_{\mu,1}m_0,\quad m_0>0\ . \tag{7}$$

We arrive at the $T=0$ recursion relations

$$m(t+1)=\int d^2z\frac{\exp(-|z|^2)}{\pi}g(m(t)+[\alpha D(t)]^{1/2}z)\ , \tag{8}$$

$$D(t+1)=1+\frac{1}{2\alpha}\left|\int d^2z\frac{\exp(-|z|^2)}{\pi}z^* \times g(m(t)+[\alpha D(t)]^{1/2}z)\right|^2, \tag{9}$$

where $D(t)$ is defined as the variance of the residual overlap for the noncondensed patterns

$$D(t)=\mathrm{Var}[r^\mu(t)]=\mathrm{Var}\left[\frac{1}{\sqrt{N}}\sum_i[\xi_i^\mu(t)]^*\sigma_i(t)\right],\quad \mu>1\ . \tag{10}$$

Setting $m(t+1)=m(t)=m$ and $D(t+1)=D(t)=D$ to obtain the stationary state and introducing the variable $x=m/(\alpha D)^{1/2}$ it is easy to derive that the critical capacity α_c can be found from

$$\alpha_c=\max_x\left[\left[\frac{F(x)}{x}\right]^2-G(x)^2\right], \tag{11}$$

where

$$F(x)=\int d^2z\frac{\exp(-|z|^2)}{\pi}g(x+z)\ , \tag{12}$$

$$G(x)=\int d^2z\frac{\exp(-|z|^2)}{\pi}z^*g(x+z)\ . \tag{13}$$

It is interesting to note that an analogous formula can be

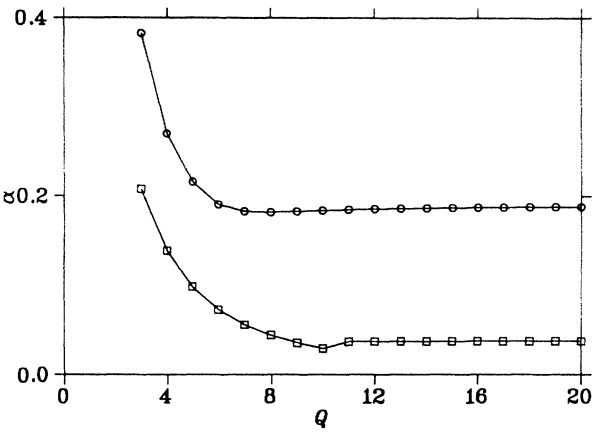

FIG. 1. The capacity α as a function of the number of phases Q for the layered feedforward (circles) and the fully connected (squares) phasor models.

TABLE I. The behavior of α_c for different multistate models as a function of Q and the nature of the retrieval transition.

	Fully connected	Layered feedforward	Extremely diluted
Potts	$\uparrow$ ($Q^{1.87}$) discontinuous	$\uparrow$ ($Q^{1.85}$) discontinuous	$\uparrow$ ($Q^{1.77}$) discontinuous ($Q\neq 2$)
Phasor	$\searrow\nearrow$ discontinuous	$\searrow\nearrow$ discontinuous	$\uparrow$ ($\pi/4$) continuous ($Q\neq 3$)
Q Ising $b=\frac{1}{2}$	$\downarrow$ (Q^{-2}) discontinuous	$\downarrow$ ($Q^{-1.91}$) discontinuous	$\uparrow$ (1) continuous ($Q\neq 3,5$)

derived for the fully connected version of this model, viz.,

$$\alpha_c^{(\mathrm{FC})}=\max_x\left[\frac{F(x)}{x}-G(x)\right]^2 . \tag{14}$$

For the extremely diluted architecture $\alpha_c^{(\mathrm{ED})}=\max_x[F(x)/x]^2$. Comparing this with (11) and (14) we find a useful inequality, relating the capacities of the different architectures: $\alpha_c^{(\mathrm{ED})}>\alpha_c>\alpha_c^{(\mathrm{FC})}$. We further remark that a formula analogous to (11) and (14) can be obtained for the Hopfield and the Q-state Potts models.

For the continuous phasor model ($Q\to\infty$), which has some particular relevance from a practical point of view, the results (12) and (13) further simplify to

$$F(x)=\int_0^\infty dr\,\exp[-(r^2+x^2)]rI_1(rx) , \tag{15}$$

$$G(x)=\int_0^\infty dr\,\exp[-(r^2+x^2)][r^2I_0(rx)-rxI_1(rx)] , \tag{16}$$

where I_0 and I_1 are modified Bessel functions. This clearly illustrates the advantage of using complex noises (compare [8]).

IV. RESULTS AND CONCLUDING REMARKS

In Fig. 1 we show the capacity as a function of the number of phases Q for both the layered feedforward and the fully connected architectures. We realize that the behavior is nonmonotonic. In particular, for the layered network we find a minimum in the capacity, $\alpha=0.182$, at $Q=8$. For $Q=7$ we have $\alpha=0.183$, for $Q=9$ we find $\alpha=0.183$, and for $Q=\infty$ we get $\alpha=0.190$. In comparison, for the fully connected network the minimum $\alpha=0.029$ occurs at $Q=10$ and the values for $Q=9$, 11, and ∞ are $\alpha=0.035$, 0.037, and 0.038, respectively. The significance of this behavior is that the capacity no longer decreases for growing values of Q greater than $Q=8$ for the layered model and $Q=10$ for the fully connected model. We remark that the minimum is relatively deeper in the fully connected case.

This behavior is in contrast with that for the multistate Ising and Potts models. For comparison with the latter we have extracted the relevant information from the literature [4–7,10–14] such that we get in Table I an overview of this behavior in Q and of the nature of the retrieval transition line.

Here several remarks have to be made. The exponents in the Q dependence for the Potts models are obtained using a log-log plot and are only given as a rough indication. Indeed they are extracted from results up to $Q=150$ only and they are rather sensitive to the range of Q values used (e.g., for the fully connected Potts model it was correctly found in [11] that the exponent is 2 when employing Q values up to 9). For the Q-state Ising model the value for the gain parameter b is taken to be optimal, i.e., such that the Hamming distance is smallest [10,14]. For the extremely diluted Q-state Ising models the capacity is always increasing when considering even and odd Q values separately (and in fact increasing for all values of Q from $Q=5$ onwards). For $Q=2$ all models become equivalent to the Hopfield model. For $Q=3$ the Potts and phasor models are equivalent and $\alpha_c^{\mathrm{Potts}}=2\alpha_c^{\mathrm{phasor}}$ and for the phasor model itself $\alpha_c(4)=\alpha_c(2)$ (if the temperature is scaled appropriately).

For $T\neq 0$ exact evolution equations can be derived by exploiting the nontrivial idea to express the stochastic dynamic within the gain function formulation of the deterministic dynamics using auxiliary thermal fields. Again an analogous detailed analysis such as that in [10] can be performed with the final results

$$m(t+1)=\int d^2z\frac{\exp(-|z|^2)}{\pi}\langle S(m(t)+[\alpha D(t)]^{1/2}z)\rangle , \tag{17}$$

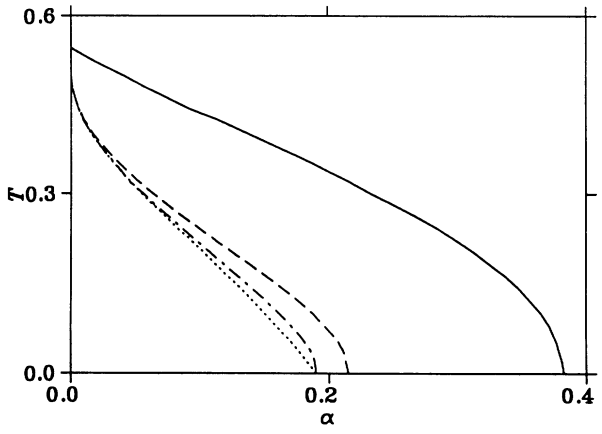

FIG. 2. Capacity-temperature phase diagram for the $Q=3$, 5, 6, and ∞ layered feedforward phasor model (solid, dashed, dash-dotted, and dotted lines, respectively).

$$D(t+1)=1+\frac{1}{2\alpha}\left|\int d^2z\frac{\exp(-|z|^2)}{\pi}z^* \times\langle S(m(t)+[\alpha D(t)]^{1/2}z)\rangle\right|^2, \tag{18}$$

where we have introduced the thermal average

$$\langle S(z')\rangle=\frac{\mathrm{tr}_{s_k}s_k\exp[\beta\,\mathrm{Re}(z's_k^*)]}{\mathrm{tr}_{s_k}\exp[\beta\,\mathrm{Re}(z's_k^*)]}. \tag{19}$$

For $Q\to\infty$ these formulas can again be simplified in terms of modified Bessel functions. Capacity-temperature diagrams for different values of Q are shown in Fig. 2. It turns out that the diagrams from $Q=8$ onward up to $Q=\infty$ coincide except for very small values of T.

In conclusion, we have solved for arbitrary temperatures exact evolution equations for the Q-phasor model on a layered architecture using a probabilistic approach. At zero temperature, we have discussed the behavior of the critical capacity in function of the Q phases in immediate comparison with other multistate neural networks.

ACKNOWLEDGMENTS

This work has been supported in part by the Research Fund of the K. U. Leuven (Grant No. OT/91/13). The authors are indebted to J. Huyghebaert, B. Vinck, and V. Zagrebnov for stimulating discussions. One of us (D.B.) would like to thank the Belgian National Fund for Scientific Research for financial support.

[1] E. Barkai, I. Kanter, and H. Sompolinsky, Phys. Rev. A **41**, 590 (1990).
[2] R. Kree and A. Zippelius, in *Models of Neural Networks,* edited by E. Domany, J. L. van Hemmen, and K. Schulten (Springer-Verlag, Berlin, 1991), p. 193.
[3] E. Domany and R. Meir, in *Models of Neural Networks* (Ref. [2]), p. 307.
[4] A. J. Noest, Europhys. Lett. **6**, 469 (1988).
[5] A. J. Noest, Phys. Rev. A **38**, 2196 (1988).
[6] J. Cook, J. Phys. A **22**, 2057 (1989).
[7] A. E. Patrick, P. Picco, J. Ruiz, and V. A. Zagrebnov, J. Phys. A **24**, L637 (1991).
[8] D. Gandolfo, J. Ruiz, and V. A. Zagrebnov, Université de Marseille Report No. CPT-92/P.2728 (unpublished).
[9] F. Gerl, K. Bauer, and U. Krey, Z. Phys. B **88**, 339 (1992).
[10] D. Bollé, G. M. Shim, and B. Vinck, J. Stat. Phys. **74**, 583 (1994).
[11] I. Kanter, Phys. Rev. A **37**, 2739 (1988).
[12] H. Rieger, J. Phys. A **23**, L1279 (1990).
[13] G. M. Shim, D. Kim, and Y. M. Choi, Phys. Rev. A **45**, 1238 (1992).
[14] D. Bollé, G. M. Shim, B. Vinck, and V. A. Zagrebnov, J. Stat. Phys. **74**, 565 (1994).