Thermodynamic properties of the Q-state Potts-glass neural network

D. Bollé,* P. Dupont, and J. Huyghebaert
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3001 Leuven, Belgium
(Received 29 August 1991)

The Q-state Potts model of neural networks, extended to include biased patterns, is studied for extensive loading α. Within the replica-symmetric approximation, mean-field equations are written down for general Q and arbitrary temperature T. The critical storage capacity is discussed for $Q=3$ and two classes of representative bias parameters. The complete T-α phase diagram is presented. A tricritical point is found in the spin-glass transition for $Q>6$, depending on α. Contrary to the Hopfield model, the critical lines do not converge to the same T as $\alpha\to 0$. A stability analysis is made.

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

Neural networks with multistate neurons have attracted recent interest in order to study the storage and retrieval properties of grey-toned patterns and also to examine how these properties change if the number of states increases. A model of this type based upon Potts spins has been introduced in Ref. [1]. Its storage capacity and retrieval of information have been discussed for unbiased patterns in the limit of zero temperature. This model has recently been extended to include a finite number of biased patterns [2]. Other multistate models have been considered in Refs. [3–11]. They are only concerned with unbiased patterns. Finally, we remark that Potts-type models are especially useful for performing multiclass classification tasks [11,12]. A systematic study of this Q-state Potts-glass neural network with extensive loading of biased patterns at finite temperatures and for an arbitrary number of states Q is quite tedious and, to our knowledge, has not been given so far.

The purpose of this Brief Report is to contribute to such a study. Here we mainly focus on the thermodynamic properties of the model since they are of independent interest and, at least to us, partly unexpected. A more detailed analysis is planned to be given elsewhere [12].

Consider the Hamiltonian

$$H=-\tfrac{1}{2}\sum_{\substack{i,j=1\\ i\neq j}}^{N}\sum_{\gamma,\rho=1}^{Q} J_{ij}^{\gamma\rho}u_{\sigma_{i,\gamma}}u_{\sigma_{j,\rho}} \tag{1}$$

with the Potts operator $u_{\sigma_{i,\gamma}}=Q\delta_{\sigma_i,\gamma}-1$ and synaptic couplings satisfying the symmetry constraint $J_{ij}^{\gamma\rho}=J_{ji}^{\rho\gamma}$. The stable states are given by those configurations $\{\sigma_i\}$ that are the local minima of H. In the presence of noise, taken into account by the introduction of a temperature $T=\beta^{-1}$, there is a finite probability of having configurations other than the local minima.

The stable configurations of the network must be correlated with the p patterns $\{k_i^a\}$, $\alpha=1,\ldots,p$ fixed by the learning process. The learning rule is given by

$$J_{ij}^{\gamma\rho}=\frac{1}{Q^2N}\sum_{a=1}^{p}(u_{k_i^a,\gamma}-B_\gamma)(u_{k_j^a,\rho}-\beta_\rho)\ , \tag{2}$$

where the $\{B_\gamma\}$ are the bias parameters of the patterns. These parameters determine the probability distribution of the independent random variables k_i^a, which can take the values $1,\ldots,Q$, as

$$P(\sigma)=\sum_{\gamma=1}^{Q}\frac{1+B_\gamma}{Q}\delta(\sigma-\gamma)\ . \tag{3}$$

Mean-field theory for the Hamiltonian (1) and (2) is derived using the replica technique [13,14]. The calculation of the free energy per neuron is quite tedious [12]. The result is, within the replica-symmetric approximation

$$\begin{aligned} f=&\tfrac{1}{2}\sum_{\nu=1}^{s}m_\nu^s-\tfrac{1}{2}\alpha\beta rq+\tfrac{1}{2}\alpha\tilde{q}-\frac{\alpha q}{2[1-\beta(\tilde{q}-q)]}\\ &+\frac{\alpha}{2\beta}\ln[1-\beta(\tilde{q}-q)]\\ &+\tfrac{1}{2}\alpha\beta\tilde{r}\tilde{q}-\frac{1}{\beta}\left\langle\!\!\left\langle\int_{-\infty}^{+\infty}D\mathbf{z}\ln\sum_{\sigma=1}^{Q}[\beta\mathcal{H}_\sigma(\mathbf{z})]\right\rangle\!\!\right\rangle\ , \end{aligned} \tag{4}$$

where $\langle\!\langle\ \rangle\!\rangle$ denotes a quenched average over the patterns $\{k_i^a\}$. The measure $D\mathbf{z}$ is given by

$$D\mathbf{z}=\prod_{l=1}^{Q}dz_l(2\pi)^{-1/2}\exp(-z_l^2/2)\ . \tag{5}$$

Furthermore, $\mathcal{H}_\sigma(\mathbf{z})$ reads

$$\begin{aligned} \mathcal{H}_\sigma(\mathbf{z})=&\tfrac{1}{2}\alpha\beta(\tilde{r}-r)(1+B_\sigma)(Q-1-B_\sigma)\\ &+\sum_{\nu=1}^{s}(u_{k^\nu,\sigma}-B_\sigma)(m_\nu+h_\nu)\\ &+\sum_{l=1}^{Q}[\alpha r(1+B_l)/Q]^{1/2}(u_{l,\sigma}-B_\sigma)z_l\ , \end{aligned} \tag{6}$$

with h_ν the couplings of the external field terms, introduced in order to describe the possible macroscopic condensation of a finite number s of patterns. The free energy (4) is a function of five order parameters: the macroscopic overlap with a condensed pattern, $m_\nu=\langle\!\langle\ \langle u_{k^\nu,\sigma}-B_\sigma\rangle\ \rangle\!\rangle$, where $\langle\ \rangle$ stands for thermal average, the extended Edwards-Anderson (EA) parameter, $q=\langle\!\langle\ \langle u_{k^\nu,\sigma}-B_\sigma\rangle^2\ \rangle\!\rangle$, the total mean-square random overlap with the noncondensed patterns,

$r=\sum_{\mu>s}\langle\!\langle\langle m_\mu\rangle^2\rangle\!\rangle/\alpha$, the autocorrelation of the overlap, $\tilde{q}=\langle\!\langle\langle(u_{k^\nu,\sigma}-B_\sigma)^2\rangle\rangle\!\rangle$, and the total autocorrelation of the noncondensed patterns, i.e., $\tilde{r}=\sum_{\mu>s}\langle\!\langle\langle m_\mu^2\rangle\rangle\!\rangle/\alpha$.

At this point, several remarks are in order. First, compared with the Hopfield model and the Potts neural network without bias, we find two extra order parameters, i.e., $\tilde{q}$ and $\tilde{r}$. This is due to the fact that in the replica matrix for q, the diagonal term is now a function of σ. Secondly, the usual EA parameter has to be extended by including explicitly the bias parameter in its definition. For $Q=2$ this definition simplifies to the usual EA parameter scaled with the factor $(1-B_1^2)$.

The fixed-point equations for the order parameters are

$$m_\nu=\left\langle\!\left\langle\int_{-\infty}^{+\infty}Dz\frac{\sum_\sigma(u_{k^\nu,\sigma}-B_\sigma)\exp[\beta\mathcal{H}_\sigma(z)]}{\sum_\sigma\exp[\beta\mathcal{H}_\sigma(z)]}\right\rangle\!\right\rangle, \tag{7}$$

$$q=\tilde{q}-\frac{1}{\beta}\sum_{l=1}^{Q}\left[\frac{1+B_l}{\alpha rQ}\right]^{1/2}\left\langle\!\left\langle\int_{-\infty}^{+\infty}Dz\frac{\sum_\sigma(u_{l,\sigma}-B_\sigma)z_l\exp[\beta\mathcal{H}_\sigma(z)]}{\sum_\sigma\exp[\beta\mathcal{H}_\sigma(z)]}\right\rangle\!\right\rangle, \tag{8}$$

$$\tilde{q}=\left\langle\!\left\langle\int_{-\infty}^{+\infty}Dz\frac{\sum_\sigma(Q-1-B_\sigma)(1+B_\sigma)\exp[\beta\mathcal{H}_\sigma(z)]}{\sum_\sigma\exp[\beta\mathcal{H}_\sigma(z)]}\right\rangle\!\right\rangle \tag{9}$$

$$r=q[1-\beta(\tilde{q}-q)]^{-2}, \tag{10}$$

$$\tilde{r}=-\beta^{-1}+(2q-\tilde{q}+\beta^{-1})[1-\beta(\tilde{q}-q)]^{-2}. \tag{11}$$

We note that in general two of the order parameters, i.e., r and $\tilde{r}$, are algebraic. For zero bias also $\tilde{q}$ becomes algebraic. For $T=0$, $\beta(\tilde{q}-q)$ equals a finite limit C [cf. Eq. (8)] and $\beta(\tilde{r}-r)=C(1-C)^{-1}$. In that case, Eqs. (7)–(11) can be reduced to two coupled nonlinear equations. For zero bias only one single nonlinear equation describes the relevant problem. A detailed discussion of these equations is contained in Ref. [12].

By solving the $T=0$ equations, e.g., for $Q=3$ and Mattis retrieval states $m_\nu=m\delta_{\nu,1}$, $m\neq0$, we get the critical storage capacities. We have selected two representative classes of bias parameters, i.e., $\mathbf{B}_1=a(2,-1,-1)$ and $\mathbf{B}_2=a(1,0,-1)$ where $a\in(0,1]$ and $\mathbf{B}=\{B_\gamma\}$, $\gamma=1,2,3$. The form $\mathbf{B}_1$ indicates that one state is privileged and the other two states have equal probability to appear. In the other case, $\mathbf{B}_2$, all three states have different probability [2].

In Fig. 1 the critical storage capacity α_c at $T=0$ is shown for the $Q=3$ $\mathbf{B}_1$ and $\mathbf{B}_2$ system as a function of the bias amplitude a. At this point we recall that in the definition of the order parameter m_ν, the overlap due to the nonzero correlations between the patterns is subtracted such that α_c is a measure of the information content of the system. In that way we can understand that $\alpha_c(a=1)=0$ for $\mathbf{B}=\mathbf{B}_1$. We also see that $\alpha_c=0.415$ for $a=0$ [1]. Since $\mathbf{B}_1$ is the most extreme choice of bias structure possible [2], all $Q=3$ systems have a storage capacity α_c at zero temperature lying between the curve for $\mathbf{B}_1$ and the line $\alpha_c=0.415$. It should be possible, like in the Hopfield model, to keep the storage capacity for biased patterns at the level of unbiased patterns by introducing soft and rigid constraints on the mean activity of the network or by adding a ferromagnetic term to the learning rule (cf. Ref. [15], respectively, Ref. [16] in the case of the Hopfield model). We finally remark that at $T=0$ both systems $\mathbf{B}_1$ and $\mathbf{B}_2$ have an overlap m that is very close to its maximum attained for $\alpha=0$, for all values of the bias amplitude a.

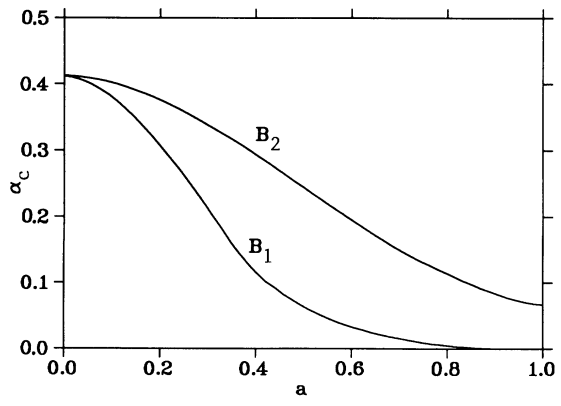

FIG. 1. The critical capacity α_c as a function of a at $T=0$.

We next turn to the complete phase diagram shown, for $Q=3$, in Fig. 2. Because it is so rich in structure we restrict ourselves in this paper to a discussion in the case of unbiased patterns. For biased patterns, the basic structure of the phase diagram is roughly unchanged. The stability properties, however, are different [12].

The transition from the disordered paramagnetic phase to the spin-glass (SG) phase can be studied analytically for general Q. Expanding the fixed-point equations (8) and (10) in powers of r and q we determine the transition temperature T_g and the nature of the transition. We arrive at $T_g=Q-1+[\alpha(Q-1)]^{1/2}$. Furthermore, for $Q\leq6$ the transition is always of second order. For $Q>6$ we find that if $\alpha<\alpha_0=16(Q-1)(Q-6)^{-2}$ the transition is of second order, while for $\alpha>\alpha_0$ the transition is of first order. For $Q\to\infty$ the tricritical point $\alpha_0\to0$ mean-

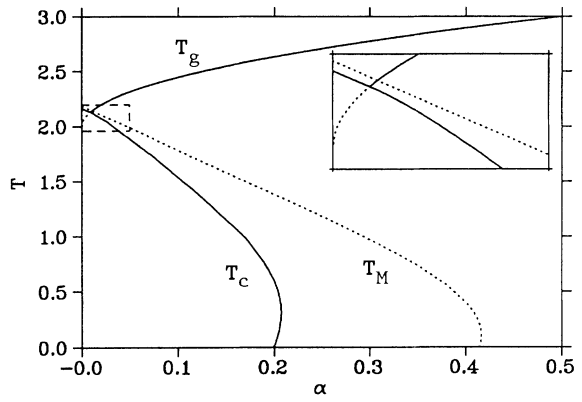

FIG. 2. T-α phase diagram for $Q=3$. The solid lines indicate the thermodynamic transitions.

ing that the transition is of first order for all values of α.

We now turn to an explicit discussion of the $Q=3$ results (Fig. 2). Below the critical line T_g, the spin-glass solutions appear. The fixed-point equations (7)–(11) are solved numerically. This yields the other critical lines depicted in Fig. 2. When crossing the line T_M from above Mattis retrieval states show up as local minima of the free energy. At this point the overlap with the built-in patterns jumps from zero to a finite macroscopic value. So the system functions as an associative memory and the critical storage capacity for a given temperature can be read off through the line T_M.

When further lowering T, the retrieval states become global minima of the free energy. This happens along the line T_c. The transition at T_c is a first-order transition. Several additional remarks are in order. First, in contrast with the Hopfield model [14] the critical lines end in different temperature points. The reason is that for finite patterns, and thus $\alpha\to 0$ in the limit $N\to\infty$, the Potts model has a discontinuous transition at T_c, as has been shown in [2]. Secondly, for a growing number of states Q this "crossover" region for small α is expected to become more extended, e.g., for $Q=4$, $T_M=3.728$ [17]. Third, if we compare the values of α_0 with a critical storage capacity α_c at $T=0$ for the Q-state model, $\alpha_c=Q(Q-1)0.138/2$, we see that $\alpha_0>\alpha_c$ for $Q\le 10$. This implies that if $Q\le 10$ then the tricritical point is certainly situated outside the crossover region. Finally, for low T we find weak reentrant spin-glass behavior, analogous to what has been seen recently [18] in the fully connected Hopfield model. The maximum value of α_c is obtained at $\alpha_c(T_M=0.099)=0.416(2)$ while $\alpha_c(T_M=0.021)=0.415(0)$.

To make this picture complete the free energies are shown in Figs. 3 and 4 for values of α in different regimes. On these figures we have also indicated the results about the stability of the different solutions to the fixed-point equations with respect to replica symmetric fluctuations. These results have been obtained by studying à la de Almeida-Thouless [19] the eigenvalues of the Hessian matrix formed by the second derivative of the free energy with respect to the order parameters m, q, and r. More details are planned to be given in Ref. [12]. For small α, e.g., $\alpha=0.005$, the situation is the following. A paramagnetic solution of the fixed-point equations exists for $T>2$. In the region $2<T<T_g=2.1$ it is unstable, for $T>T_g$ it is stable. At $T=T_g$ it bifurcates and a spin-glass solution shows up. The latter lies above the paramagnetic solution but it is locally stable down to $T=0$. At $T=0$ the SG order parameters are given by $C=3(3/4\alpha r\pi)^{1/2}$ and $r=2+3(6/\alpha\pi)^{1/2}+(27/4\pi\alpha)$. At $T=T_M=2.182$ a Mattis state starts to exist. It is locally stable for $T_M>T>T_c=2.146$ and globally stable below T_c.

A similar picture can be drawn for $\alpha=0.01$ but the temperatures T_g and T_c are interchanged, i.e., $2.141=T_g>T_c=2.127$. For $\alpha=0.10$ the results of the stability study are given in Fig. 4. From $\alpha\ge 0.199$ onwards the Mattis line of free energies lies above the spin-glass line. The results above are confirmed by studies of the bifurcation of the overlap order parameter m in function of T for different values of α [12].

For $Q>6$ and $\alpha>\alpha_0$ where the SG transition is of first order, we expect that similar to the results of Gross, Kanter, and Sompolinsky [20] there is a region around T_g where both the spin-glass and paramagnetic solution are

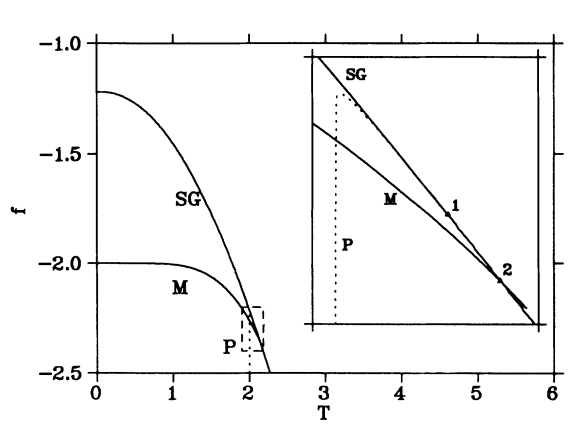

FIG. 3. The free energy f as a function of T for $\alpha=0.005$. The solid lines represent stable solutions of Eqs. (7)–(11). The points 1 and 2 indicate T_g and T_c, respectively.

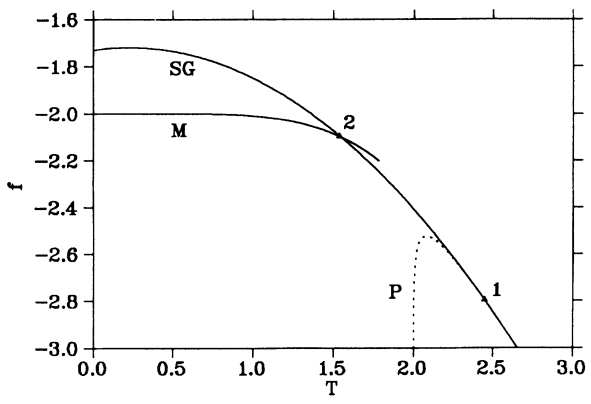

FIG. 4. Same as Fig. 3 for $\alpha=0.1$.

stable. This is currently being studied.

Finally, to examine the internal consistency of the replica symmetric theory we have calculated the entropy at $T=0$,

$$S_\alpha = -(\alpha/2)[\ln(1-C)+C(1-C)^{-1}] .$$

Compared with the Hopfield model where $S_{\alpha_c} = -0.07$ in the SG phase and with the Sherrington and Kirkpatrick model where $S_\infty = -0.16$ we find that $S_{\alpha_c} = -0.32$. For the ferromagnetic phase the Hopfield model leads to $S_{\alpha_c} = -1.4\times10^{-3}$ while we find $S_{\alpha_c} = -3.3\times10^{-3}$ in agreement with the statement about the $Q=3,4$ clock model [3]. All this suggests that replica symmetric breaking is still weak in the retrieval states for the $Q=3$ Potts model.

In conclusion, the Q-state Potts glass neural network shows a very rich structure within the replica-symmetric approximation. It is certainly interesting to study how this model behaves when breaking the replica symmetry.

We thank A. Canning, R. Kühn, and L. van Hemmen for stimulating discussions and R. Dekeyser for a critical reading of the manuscript. We are indebted to R. Cools of the Department of Computer Science for calculating part of the phase diagram in Fig. 2. D. B. and J. H. would like to thank the Belgian National Fund for Scientific Research, and the Inter-University Institute for Nuclear Sciences, for financial support.

*Electronic address: fgbda18@blekul11.bitnet.

[1] I. Kanter, Phys. Rev. A **37**, 2739 (1988).
[2] D. Bollé, P. Dupont, and J. van Mourik, J. Phys. A **24**, 1065 (1991).
[3] J. Cook, J. Phys. A **22**, 2057 (1989).
[4] A. J. Noest, Phys. Rev. A **38**, 2196 (1988).
[5] C. Meunier, D. Hansel, and A. Verga, J. Stat. Phys. **55**, 859 (1989).
[6] J. S. Yedidia, J. Phys. A **22**, 2265 (1989).
[7] D. Bollé, P. Dupont, and J. van Mourik, Europhys. Lett. **15**, 893 (1991).
[8] S. Mertens, H. M. Köhler, and S. Bös, J. Phys. A **24**, 4941 (1991).
[9] H. Reiger, J. Phys. A **23**, L1273 (1990).
[10] G. M. Shim, D. Kim, and M. Y. Choi, Phys. Rev. A **45**, 1238 (1992).
[11] J. P. Nadal and A. Rau, J. Phys. I France **1**, 1109 (1991); T. Watkin, A. Rau, D. Bollé, and J. van Mourik, *ibid.* (to be published).
[12] D. Bollé, R. Cools, P. Dupont, and J. Huyghebhert (unpublished).
[13] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975); S. Kirkpatrick and D. Sherrington, Phys. Rev. B **17**, 4384 (1978).
[14] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985); Ann. Phys. (N.Y.) **173**, 30 (1987).
[15] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **35**, 2293 (1987).
[16] M. V. Feigelman and L. B. Joffe, Int. J. Mod. Phys. B **1**, 51 (1987); S. Bös, R. Kühn, and J. L. van Hemmen, Z. Phys. B **71**, 261 (1988).
[17] D. Bollé and F. Mallezie, J. Phys. A **22**, 4409 (1989).
[18] J. P. Naef and A. Canning (unpublished).
[19] J. R. de Almeida and D. J. Thouless, J. Phys. A **11**, 983 (1978).
[20] D. J. Gross, I. Kanter, and H. Sompolinsky, Phys. Rev. Lett. **55**, 304 (1985).