Storing an extensive number of grey-toned patterns in a neural network using multistate neurons

## LETTER TO THE EDITOR

# Storing an extensive number of grey-toned patterns in a neural network using multistate neurons

H Rieger

Institut für Theoretische Physik, Universität zu Köln, Zülpicher Strasse 77, D-5000 Köln 41, Federal Republic of Germany

**Abstract.** Grey-toned patterns are pictures composed of pixels of several shades of grey. The ability of neural networks using multistate neurons to store such patterns is systematically investigated. If conventional generalizations of Hopfield networks using analogue or soft neurons are considered, it is impossible to stabilize these grey tones. Nevertheless it is shown that it can be done with networks that use neurons which have only a discrete set of possible activities. This is demonstrated for the pseudo-inverse rule for the synaptic couplings, where only the stability of the patterns shrinks with increasing number $Q$ of grey tones one wants to store. If the patterns are uncorrelated one can use the Hebb rule and in this case the mean field theory is presented. Applying this rule the storage capacity decreases as $Q^{-2}$ with the number of grey tones.

The original Little–Hopfield model of neural networks [1, 2] consists of neurons that can take on two different states corresponding to 'firing' and not firing'. With appropriate synaptic couplings the network is able to memorize patterns, i.e. some prescribed configurations of the neuronal activities act as attractors within the dynamical evolution of the network [2, 3]. If one transposes these configurations into a two-dimensional picture, the two-valued neuronal activities correspond to black and white pixels. In this sense the Hopfield net is able to retrieve black/white pictures. The impression of more light or more darkness within certain regions of the picture can only be achieved by manipulating the density of black pixels on a white background.

Now, in a very natural way, the question arises as to whether one is able to store pictures composed of pixels of various shades of grey in a network that consists of neurons with graded response (the interest in this problem is of course not limited to vision). Neural networks using such neurons—first considered in [4]—were investigated very recently by several authors [5–11]. The main problem arising in this context is that for soft neurons, i.e. neurons with a continuous IO relation, the intermediate activities (corresponding to grey tones) are hard to stabilize, as we shall see below. One can circumvent this problem by a more abstract model, which deals only with continuously varying order parameters and suitable potentials [12], but then the neurons and their dynamics are missing, which is not advantageous, e.g. for hardware implementations.

The aim of this letter is to investigate systematically the ability of neural networks using multistate neurons to store patterns, where the individual neuronal activities take on several values between −1 and +1 (corresponding to black and white). Here only results are given, the derivation of formulae and a more detailed analysis in connection with computer simulations can be found in a forthcoming publication [13].

Multistate neurons are characterized by an IO relation $g_i(h_i)$ that prescribes for each value of the postsynaptic potential (or local field) $h_i$ the corresponding output value (or activity) $\sigma_i = g_i(h_i)$, which can be interpreted in terms of firing rates (after the transformation to $S_i = (1+\sigma_i)/2$). The local field is as usual defined to be

$$h_i = \sum_{j(\neq i)}^{N} J_{ij}\sigma_j. \tag{1}$$

The neurons are updated in a random-sequential way. When the $i$th neuron is updated the actual local field $h_i$ causes the new activity $\sigma_i$ to be

$$\sigma_i^{new} = g_i(h_i). \tag{2}$$

In the presence of noise there is a non-vanishing probability to take on other activities than prescribed by (2). In this case one has to consider transition probabilities instead of the update rule (2). Let us first concentrate on the noiseless case.

Certain configurations of the neuronal activities $\boldsymbol{\xi}^\nu = (\xi_1^\nu, \ldots, \xi_N^\nu)$, $\nu = 1, \ldots, p$—called patterns—should be fixed points of the dynamics (2) with as large as possible a basin of attraction. Therefore it is required that

$$\xi_i^\nu = g_i\left(\sum_{j(\neq i)}^{N} J_{ij}\xi_j^\nu\right) \qquad i = 1, \ldots, N. \tag{3}$$

This can be achieved by an appropriate choice of the synaptic couplings $J_{ij}$. The IO function should be subjective (otherwise cancel all activities $\sigma_i$ that do not correspond to any local field), hence it is possible to choose $\eta_i^\nu$ with $g_i(\eta_i^\nu) = \xi_i^\nu$. For linear independent patterns the pseudo inverse can be applied [14]

$$J_{ij} = \frac{1}{N}\sum_{\mu,\nu=1}^{p} \eta_i^\mu (\mathbf{C}^{-1})_{\mu\nu}\xi_j^\nu \qquad J_{ii} = 0 \tag{4}$$

where $\mathbf{C}_{\mu\nu} = (1/N)\sum_{i=1}^{N}\xi_i^\mu\xi_i^\nu$ is the correlation matrix of the patterns.

Already at this stage one observes that it seems to be impossible to provide the patterns in a network using a broad class of smooth IO functions (e.g. $\tanh(gh)$, $\mathrm{erf}(gh)$, piecewise linear, etc) with sufficient stability. If $g_i'(\eta_i^\nu)$ is of order one, a small change in the neuronal activities will not be damped out by the update rule (2) and in the presence of noise the patterns are expected to be unstable. This becomes more vivid if we concentrate now on a special class of IO functions, which arises in a natural way doing statistical mechanics with networks of multistate neurons and symmetric couplings [5].

Consider the Hamiltonian

$$H = -\tfrac{1}{2}\sum_{i\neq j} J_{ij}\sigma_i\sigma_j + \tfrac{1}{2}\sum_i \sigma_i^2 \tag{5}$$

introduced in [15] in the context of $S$-Ising spin glasses. The activities $\sigma_i$ vary either continuously within the interval $[\sigma_{min}, \sigma_{max}]$ or take on discrete values $\sigma^1, \ldots, \sigma^Q$. Transition probabilities leading to the canonical probability distribution $P(\boldsymbol{\sigma}) \propto e^{-\beta H}$ are given by

$$w(\sigma_i^{new} = \tilde{\sigma}) = \frac{e^{\beta(h_i\tilde{\sigma} - \tilde{\sigma}^2/2)}}{\mathrm{Tr}_{\tilde{\sigma}}\, e^{\beta(h_i\tilde{\sigma} - \tilde{\sigma}^2/2)}} \qquad \beta = 1/T \text{ (inverse temperature)}. \tag{6}$$

In the noiseless case one gets an IO function, which is piecewise linear for continuously

varying $\sigma_i$:

$$g(h) = \begin{cases} \sigma_{\min} & \text{for } h < \sigma_{\min} \\ h & \text{for } h \in [\sigma_{\min}, \sigma_{\max}] \\ \sigma_{\max} & \text{for } h > \sigma_{\max} \end{cases} \tag{7}$$

and composed by step functions in the case of discrete $\sigma$:

$$g(h_i) = \sigma_i^k \qquad \text{for } h_i \in [U_{k-1}, U_k[, \quad k = 1, \dots, Q \tag{8}$$

with $U_0 = -\infty$, $U_Q = +\infty$ and $U_k = (\sigma^k + \sigma^{k+1})/2$ for $k = 1, \dots, Q-1$.

Following [14] one can define the Euclidean distance between the actual network configuration $\sigma$ and the subspace spanned by the patterns $\Delta = \text{dist}(\sigma, \text{spann}\{\xi^1, \dots, \xi^p\})$ and a simple calculation leads in the case of the couplings (4) (with $\eta_i^\nu = \xi_i^\nu$) to $H = \Delta^2$. Thus for continuously varying activities all linear combinations of the patterns within the cubus $[\sigma_{\min}, \sigma_{\max}]^N$ are degenerate in energy and the network is not usable for associative memory or pattern recognition. The situation is even worse in the presence of noise.

The only way out is discrete output values (or IO functions with plateaux at certain field values). This is due to the fact that in this case the probability that spann$\{\xi^1, \dots, \xi^p\}$ contains a vector with components equal to one of the possible activities $\sigma^1, \dots, \sigma^Q$ vanishes for $N \to \infty$. This was proven in [14] for $(\pm 1)$ neurons and holds also for each finite $Q$. Nevertheless other spurious states occur with increasing $Q$, different from linear combinations of the patterns and thus higher in energy, but also metastable. Therefore the performance of the network should be checked more quantitatively, which will be done below.

The mean-field theory of the neural network using $Q$-state neurons, described by the Hamiltonian (5) and the couplings (4) is analogous to that of $(\pm 1)$ neurons [14] and yields—as one would expect—$\alpha_c = 1$ for the critical storage capacity $\alpha = p/N$, but of course, due to the new spurious states mentioned above, another temperature dependence. The advantage of the learning rule (4) lays in its ability to store also correlated patterns, its disadvantage is the fact that it is non-local.

In the case of uncorrelated random patterns we have approximately $C_{\mu\nu} = C\delta_{\mu\nu}$, with $C = \langle\!\langle \xi^2 \rangle\!\rangle$. Within this approximation the equation (4) becomes the well known Hebb rule

$$J_{ij} = \frac{1}{CN} \sum_{\nu=1}^{p} \xi_i^\nu \xi_j^\nu. \tag{9}$$

The calculation of the free energy per neuron

$$f = -\frac{1}{\beta} \lim_{N \to \infty} \frac{1}{2N} \langle\!\langle \ln \text{Tr}_\sigma \, e^{-\beta H} \rangle\!\rangle_\xi \tag{10}$$

is now standard [3] and yields (for an even distribution of $\xi$) within the replica symmetric approximation

$$f = \frac{\alpha}{2}\mathcal{H} + \frac{1}{2}\sum_\nu (m^\nu)^2 + \frac{\alpha}{2\beta}\left\{\ln[1 - \beta(\mathcal{H} - q)] - \frac{\beta q}{1 - \beta(\mathcal{H} - q)}\right\}$$
$$- \frac{\alpha\beta rq}{2} + \frac{\alpha\beta\mathcal{R}\mathcal{H}}{2} - \frac{1}{\beta}\left\langle\!\!\left\langle \int_{-\infty}^{+\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \ln \text{Tr}_\sigma \, e^{\beta(h_z\sigma - \tilde{U}\sigma^2)} \right\rangle\!\!\right\rangle_{\xi^\nu} \tag{11}$$

with $h_z = \sqrt{\alpha r}\, z + (1/C) \sum_\nu m \xi^\nu$, $2\tilde{U} = 1 - \alpha \chi/(1-\chi)$, where $\chi = \beta(\mathcal{X} - q)$. The meaning of the order parameters is the following: $\mathcal{X} = \langle\!\langle\langle\sigma^2\rangle\rangle\!\rangle$ is the mean square of the neuronal activities, $q = \langle\!\langle\langle\sigma\rangle^2\rangle\!\rangle$ is the EA-order parameter, $r = \langle\!\langle\langle m^2\rangle\rangle\!\rangle$ is the mean square of the random overlaps of the non-condensed patterns $\xi^\mu$ with the configuration $\sigma$ of the network, $\mathcal{R} = \langle\!\langle\langle m^2\rangle\rangle\!\rangle$ is the mean correlation between the local activities $\sigma_i^2$ and the pattern activity $(\xi_i^\mu)^2$ of the non-condensed patterns and $m^\nu = \langle\!\langle\langle\xi^\nu\sigma\rangle\rangle\!\rangle$ is the overlap of the network configuration with a condensed pattern $\xi^\nu$. The replica symmetry is broken below the generalized AT line, where the replicon eigenvalue

$$\lambda = [1 - \beta(\mathcal{X} - q)]^2 - \alpha\beta^2 \int_{-\infty}^{+\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \left\langle\!\!\left\langle\!\!\left\langle \left\{ \frac{\mathrm{Tr}_\sigma \sigma^2\, e^{\beta(h_z\sigma - \tilde{U}\sigma^2)}}{\mathrm{Tr}_\sigma\, e^{\beta(h_z\sigma - \tilde{U}\sigma^2)}} \right.\right.\right.\right.$$
$$\left.\left.\left.- \left[\frac{\mathrm{Tr}_\sigma \sigma\, e^{\beta(h_z\sigma - \tilde{U}\sigma^2)}}{\mathrm{Tr}_\sigma\, e^{\beta(h_z\sigma - \tilde{U}\sigma^2)}}\right]^2 \right\}^2 \right\rangle\!\!\right\rangle\!\!\right\rangle_{\xi^n} \tag{12}$$

becomes negative.

Although the overlap $m^\nu$ is the quantity that decides whether there are correlations between the state of the network and the pattern $\xi^\nu$, one is also interested in the amount of correct and incorrect activities: $a^\nu = (1/N) \sum_{i=1}^N \delta_{i,\xi_i^\nu}$. Since $\sigma$ and $\xi$ can take on only $Q$ different values it is possible to represent the Kronecker symbol $\delta_{\sigma,\delta}$ as a polynomial (here $\sigma^r$ means the $r$th power of the variable $\sigma$, the same for $\xi^s$):

$$\delta_{\sigma,\xi} = \sum_{r,s=0}^{Q-1} \lambda_{rs} \sigma^r \xi^s \tag{13}$$

where the $Q^2$ coefficients $\lambda_{rs}$ have to be determined by solving the linear system of equations that is obtained by inserting all possible values for $\sigma$ and $\xi$ into equation (13). For example in the case of $Q = 3$ and $\sigma, \xi \in \{-1, 0, +1\}$ one gets $\delta_{\sigma,\xi} = 1 - \sigma^2 - \xi^2 + \frac{1}{2}\sigma\xi + \frac{1}{2}\sigma^2\xi^2$. Thus, once the order parameters are known, the quantity $a^\nu$ can be calculated via

$$a^\nu = \sum_{r,s=0}^{Q-1} \lambda_{rs} \int_{-\infty}^{+\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \left\langle\!\!\left\langle\!\!\left\langle (\xi^\nu)^s \frac{\mathrm{Tr}_\sigma \sigma^r\, e^{\beta(h_z\sigma - \tilde{U}\sigma^2)}}{\mathrm{Tr}_\sigma\, e^{\beta(h_z\sigma - \tilde{U}\sigma^2)}} \right\rangle\!\!\right\rangle\!\!\right\rangle_{\xi^\nu}. \tag{14}$$

Of special interest are the critical storage capacities $\alpha$ for vanishing temperature up to which retrieval states $m^\nu = m\delta_{1,\nu}$ with $m \neq 0$ exist. For $T = 0$ the self-consistency equations for the order parameters read:

$$m = \frac{\sigma^Q - \sigma^1}{2} \langle\!\langle\xi\rangle\!\rangle + \frac{1}{2} \sum_{k=1}^{Q-1} \sum_{l=1}^{Q} p_l \sigma^l [\sigma^{k+1} - \sigma^k]\, \mathrm{erf}\left(\frac{C^{-1}m\sigma^l - \tilde{U}_k}{\sqrt{2\alpha r}}\right)$$

$$q = \frac{(\sigma^Q)^2 + (\sigma^1)^2}{2} + \frac{1}{2} \sum_{k=1}^{Q-1} \sum_{l=1}^{Q} p_l [(\sigma^{k+1})^2 - (\sigma^k)^2]\, \mathrm{erf}\left(\frac{C^{-1}m\sigma^l - \tilde{U}_k}{\sqrt{2\alpha r}}\right) \tag{15}$$

$$\chi = \sum_{k=1}^{Q-1} \sum_{l=1}^{Q} p_l \sigma^l [\sigma^{k+1} - \sigma^k] \frac{1}{\sqrt{2\pi\alpha r}} \exp\left(-\frac{[C^{-1}m\sigma^l - \tilde{U}_k]^2}{2\alpha r}\right)$$

where $p_l = \mathrm{prob}\{\xi_i^\nu = \sigma^l\}$ and $\tilde{U}_k = \tilde{U}(\sigma^k + \sigma^{k+1})$. Note that $q = \mathcal{X}$ for $T = 0$ and $r = q(1-\chi)^{-2}$.

Suppose that the neurons have equidistant activities $\sigma^k = -1 + 2(k-1)/(Q-1)$. If no information about the patterns is available it seems to be most adequate to assume $p_l = 1/Q$, that means each grey tone is equally probable. If the patterns have a greater number of grey tones than $Q$ and one is only interested in an information content of

In $Q$ per pixel, one has to map the superficial grey tones onto the possible activities. Otherwise another network has to be used (with $\tilde{Q}$-state neurons, where $\tilde{Q}$ is the number of grey tones). Hence, in what follows the distribution of patterns is given by

$$P(\xi) = \frac{1}{Q} \sum_{k=1}^{Q} \delta(\xi - \sigma^k). \tag{16}$$

Solving the equations (15) one gets the critical storage capacities as depicted in figure 1. This fits rather nicely with

$$\alpha_c(Q) \approx 0.3 Q^{-2} \qquad \text{for } Q \gg 1. \tag{17}$$

These are the results for the replica-symmetric theory, but although at zero temperature the replica-symmetry is broken (see (12)), the effects are extremely weak, for $Q = 3$ already one order of magnitude smaller than in the two-state case and decreasing with increasing $Q$. The phase transition at $\alpha_c$ is of first order, where $m$ jumps from a value that deviates from $C = \langle\!\langle \xi^2 \rangle\!\rangle$ less than 1% (for $Q > 2$) discontinuously to zero. As mentioned above one can calculate the relative number of neurons that are wrong and gets less than 0.3%.
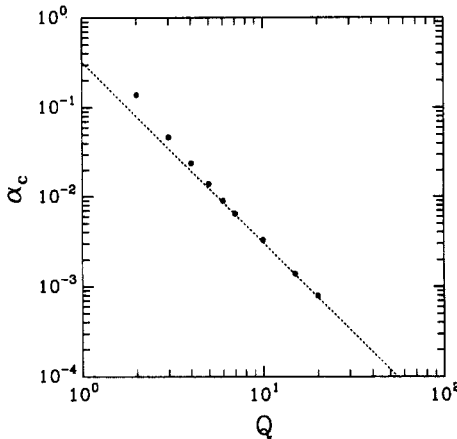


**Figure 1.** The critical storage capacity $\alpha_c$ for different values of $Q$ (at zero temperature) within a log–log plot. The dots are results obtained by solving the equations (15), the broken line is the fit mentioned in the text.

Let us have a closer look at the full phase diagram of the three-state net (i.e. $\sigma^k = -1, 0, +1$, compare with [6, 7]), which is depicted in figure 2. The most remarkable difference to the two-state net (see [3]) is (i) the much lower storage capacity $\alpha_c = 0.047$ and (ii) the paramagnetic phase ($m = 0$, $q = 0$) between the retrieval phase ($m > 0$) and the spin glass phase ($m = 0$, $q > 0$). This is due to the presence of the zero state ($\sigma_i = 0$) in connection with the second term $\Sigma_i \sigma_i^2$ in the Hamiltonian (5), which punishes high activities. Only if the variance $\alpha$ of the random couplings $J_{ij}$ is strong enough does the first term describing the interaction between the neurons become dominant and provides spin-glass behaviour. The fact that this paramagnetic phase, where the neurons have a great probability to be in the zero state, is neighbouring the retrieval phase has many advantages for practical purposes, as was detected in [6]. This feature is of course present in all networks using $Q$-state neurons, where $Q$ is an odd number. The
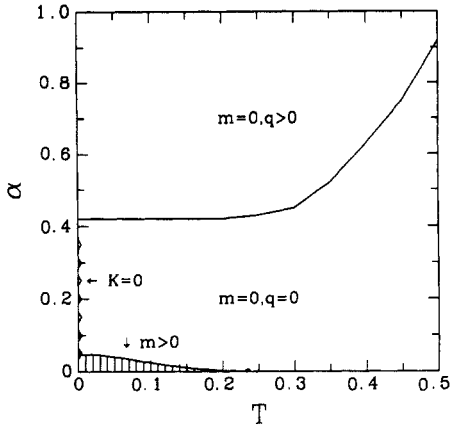
**Figure 2.** The phase diagram for the three-state network. From top to bottom there is the spin glass phase ($m = 0$, $q > 0$), the paramagnetic phase ($m = 0$, $q = 0$) and the retrieval phase ($m > 0$). On the left side at $T = 0$ the paramagnetic phase is bordered by the zero-state $\mathcal{K} = 0$. The AT line cannot be seen within the retrieval zone on this scale.

line $\alpha_c(T)$ below which retrieval states exist, has the same shape as in the two-state net [3], the AT line, below which replica-symmetry breaking occurs, cannot be seen within the scale of figure 2.

For higher values of $Q$ the phase diagram looks qualitatively the same, but the retrieval zone shrinks for increasing $Q$. Within this zone different retrieval phases occur, with different degrees of correlation between the network configurations and the patterns, which are all metastable (see [13]). It depends on the initial amount of correct neurons in which phase the network is in equilibrium.

In this letter it has been shown that pattern recognition of grey-toned patterns is possible in a neutral network that uses neurons with a discrete set of activities. Using the pseudo inverse one can store $N$ patterns in a network with $N$ neurons, but the basin of attractions for retrieval without error shrink with the number of grey tones one wants to store. Even with the conventional Hebb rule it is possible to store uncorrelated grey-toned patterns, but for this rule the storage capacity decreases as $Q^{-2}$ with $Q$. It would be interesting to calculate the maximal storage capacity of these networks (for optimal couplings) as was done for the two-state case in [16] and to ascertain whether the result $\alpha_{max} = 2$ for $Q = 2$ holds also for higher $Q$.

**References**

[1] Little W A 1974 *Math. Biosci.* **19** 101
[2] Hopfield J J 1982 *Proc. Natl Acad. Sci., USA* **79** 2554
[3] Amit D J, Gutfreund H and Sompolinsky H *Ann. Phys., NY* **173** 30
[4] Hopfield J J 1984 *Proc. Natl Acad. Sci., USA* **81** 3088
[5] Rieger H 1989 *PhD Thesis* University of Cologne; 1990 *Neural Networks (Proc. XI Sitges Conf.)* ed L Garrido (Berlin: Springer)

[6] Meunier C, Hansel D and Verga A 1989 *J. Stat. Phys.* **55** 859
[7] Yedidia J S 1989 *J. Phys. A: Math. Gen.* **22** 2265
[8] Marcus C M, Waugh F R and Westervelt R M 1990 *Phys. Rev.* A **41** 3355
[9] Treves A 1990 *Phys. Rev.* A **41** in press; *J. Phys. A: Math. Gen.* in the press
    Treves A and Rolls E T 1990 *Neural Networks (Proc. XI Sitges Conf.)* ed L Garrido (Berlin: Springer)
[10] Kühn R 1990 *Neural Networks (Proc. XI Sitges Conf.)* ed L Garrido (Berlin: Springer)
    Kühn R, Bös S and van Hemmen J L 1990 *Phys. Rev. Lett.* submitted
[11] Mertens H 1990 *Neural Networks (Proc. XI Sitges Conf.)* ed L Garrido (Berlin: Springer)
[12] Haken H 1988 *Z. Phys.* B **71** 521
[13] Rieger H to be published
[14] Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
[15] Ghatak S K and Sherrington D 1977 *J. Phys. C: Solid State Phys.* **10** 3149
[16] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257