An extremely diluted asymmetric network with graded response neurons

1991 J. Phys. A: Math. Gen. 24 337

(http://iopscience.iop.org/0305-4470/24/1/039)

J. Phys. A: Math. Gen. 24 (1991) 337–351. Printed in the UK

An extremely diluted asymmetric network with graded response neurons

S Mertens

Institut für Theoretische Physik der Universität Göttingen, Bunsenstraße 9, 3400 Göttingen, Federal Republic of Germany

Received 18 July 1990, in final form 2 October 1990

Abstract. The properties of an extremely diluted asymmetric network of neurons with a sigmoidal response function are investigated. It is shown that in the absence of thermal noise the storage capacity increases with decreasing gain of the response function. The information capacity of the network in the case of multistate patterns is discussed. It turns out that the network can process only a finite amount of information in spite of the unbounded amount of information that can be encoded in multistate patterns.

1. Introduction

The close relationship between neural networks with binary formal neurons and well known Ising spin glass systems [1] has in the past lead to many insights into the nature of attractor neural networks [2]. From a physiological point of view however, the firing rate S of a neuron should be a *continuous* function of the post synaptic potential (PSP) h:

$$S = \mathrm{dyn}(h) \tag{1}$$

where the input/output relation dyn is of sigmoidal shape. The use of binary variables is equivalent to the approximation of dyn by the sign-function.

A network of graded response neurons was proposed by Hopfield in 1984 [3] but an analytical investigation of this model has not been carried out until recently: Marcus *et al* [4] considered the Hopfield–Little model with deterministic parallel dynamics and $\mathrm{dyn}(h) = \tanh(h/\Gamma)$. They found that the inverse gain Γ of the sigmoidal response function and the temperature in the usual binary Hopfield–Little model with stochastic dynamics [5] play a very similiar role. In particular, the storage capacity of the network decreases with decreasing gain. Though their calculations are only approximative, their findings are supported by numerical simulations†.

In a series of papers, Treves [7–9] discusses a network with a threshold-linear I/O relation

$$\mathrm{dyn}(h) = \begin{cases} \Gamma^{-1}(h - h_{\mathrm{th}}) & \text{for } h > h_{\mathrm{th}} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

† After this work was completed, I became aware of the work of Kühn *et al* [6] who presented a replica calculation of this model which qualitatively confirmes the results of [4].

0305-4470/91/010337+15$03.50

and additional terms in the PSP to include inhibition. Among other things he calculated the information capacity of this network. This is a quantity of considerable interest in a network with graded response neurons since such a network can process more complex patterns than just binary ones.

With the exception of [9] all work cited so far relies on the symmetry of the synaptic coupling matrix. This assumption clearly contradicts physiology. In biological networks a large fraction of the synaptic connections are even unidirectional. In the limit of low mean connectivity, an analytical treatment of such asymmetric networks becomes feasible, as has been demonstrated for the Hopfield–Little model [10, 11].

The technique developed in [11] can easily be generalized to apply to a rather broad range of network models—including models with an arbitrary I/O relation dyn(h) This fact is used in the present contribution where the properties of an extremely diluted asymmetric network with sigmoidal I/O relation are discussed. The paper is organized as follows. Section 2 contains the definition of the model and its solution in terms of evolution equations for the order parameters that describe the network on a macroscopic level. Section 3 comprises a discussion of the retrieval properties of the network. In particular it will be shown that the storage capacity increases with decreasing gain, which is exactly the opposite effect one can observe in fully connected, symmetric networks. In section 4 we store multistate patterns in the network and discuss the achievable information capacity. It will turn out that, under certain conditions, binary patterns yield the best transfer of information even in a network of analogue neurons that could represent any multistate pattern. After the summary in the final section, a short appendix is added where the evolution equations for a network with unspecified I/O relation, a generalized Hebbian learning rule and an arbitrary distribution of pattern-activities are depicted. This appendix elucidates the broad range of network models to which the techniques of [11] are applicable.

In considering a network of continous variables in the high dilution limit, this work has some overlap with that of Treves [9]. There are, however, some differences concerning the model as well as the questions being addressed. The linear I/O relation (2) used by Treves does not take neural saturation into account and has no simple 'binary limit' to compare the effects of continous firing rates directly to discrete models. These features are provided by the present model. Rather than concentrating on the role of the graded response, Treves concentrates on the effects of the learning rule and the distribution of pattern activities, whereas we will use only the simple Hebbian learning rule and equally distributed pattern activities.

At this point it should be mentioned that, independently of this work, Rieger [12] has developed another formalism to treat extremely diluted asymmetric networks with arbitrary I/O functions based on a master equation with appropriate transition probabilities. In the absence of thermal noise and for random sequential dynamics, both formalisms should yield the same results—e.g. the phase diagram (figure 1) in section 3 has also been found by Rieger.

2. The model

The model consists of N continuous variables $S_i(t) \in [-1, 1]$ ('firing rates') which obey the discrete time dynamics

$$S_i(t_{n+1}) = \tanh\left(\frac{h_i(t_n) + \Phi_i(t_n)}{\Gamma}\right) \qquad (3)$$

$$h_i(t_n) = \sum_{j=1}^{N}{}' J_{ij} S_j(t_n) \tag{4}$$

where J_{ij} is the efficacy of the synaptic junction from neuron j to neuron i. The primed sum indicates summation over $j \neq i$. To capture the influence of thermal noise, the term $\Phi_i(t)$ is added to the PSP. The $\Phi_i(t)$ are stochastic variables, drawn independently from a distribution $w(\Phi)$ with $\langle \Phi \rangle_w = 0$. The updating according to (3) proceeds either simultaneously for all S_i (parallel dynamics) or for one randomly selected S_i at each time step (random sequential dynamics).

The parameter Γ is the inverse gain of the sigmoidal I/O function tanh. The binary Hopfield–Little model is recovered for $\Gamma \to 0$ ('high gain limit') and the special choice [11, 13]

$$w(\Phi) = \frac{\beta}{2} \frac{1}{\cosh^2 \beta \Phi}$$

since in this case (3) can be rewritten as

$$S_i(t_{n+1}) = \pm 1 \qquad \text{with probability} \quad \tfrac{1}{2}[1 \pm \tanh \beta h_i(t_n)]$$

which is nothing but the well known Glauber dynamics of an Ising spin in the local field h_i at temperature β^{-1}.

It should be noted that the choice of a sigmoidal I/O relation like the hyperbolic tangent fixes the values $S = \pm 1$ as the maximum (minimum) inducable firing rates of a neuron.

To act as an associative network, the system has to 'learn' a set of P prescribed patterns ξ^μ, $\mu = 1, \ldots, P$ with $\xi^\mu = (\xi_1^\mu, \ldots, \xi_N^\mu)$. As in the Hopfield–Little model the ξ_i^μ are assumed to be independent stochastic quantities, distributed according to

$$\rho(\xi) = \tfrac{1}{2}\delta(\xi - 1) + \tfrac{1}{2}\delta(\xi + 1) \tag{5}$$

and the learning process is realized through a modification of the synaptic efficacies for each new pattern ξ^μ according to Hebb's rule

$$\Delta J_{ij} \propto \xi_i^\mu \xi_j^\mu. \tag{6}$$

If the couplings J_{ij} were symmetric (i.e. $J_{ij} = J_{ji}$) before any patterns have been learned, then the learning rule (6) clearly preserves this symmetry. In biological networks the couplings are however supposed to be unidirectional, i.e. if $J_{ij} \neq 0$ then $J_{ji} = 0$. To account for this feature in the present model, the couplings are written as

$$J_{ij} = \frac{c_{ij}}{K} \sum_{\nu=1}^{P} \xi_i^\nu \xi_j^\nu. \tag{7}$$

The sum arises from (6) and the assumption that the neurons were completely disconnected before the learning process ('tabula rasa'). The c_{ij} are independent random $(0, 1)$ variables with

$$c_{ij} = \begin{cases} 1 & \text{with probability } K/N \\ 0 & \text{with probability } (1 - K/N) \end{cases} \tag{8}$$

i.e. the mean number of synapses per neuron is K. Note that the independence of the c_{ij} can imply unidirectionality, i.e. $J_{ij}J_{ji} \neq 0$ with probability $(K/N)^2$. It is now assumed that the connectivity K is small compared to the number of neurons in the system:

$$\lim_{N\to\infty} \frac{\log K}{\log N} = 0. \tag{9}$$

Then each coupling in the network is unidirectional with probability 1. There is however yet another motivation for this somehow artificial constraint: analytical solvability. As was first pointed out by Hilhorst and Nijmeijer [14], equation (9) guarantees that no neuron appears more than once in the tree of dynamical ancestors of any neuron. This fact is used to reduce the dynamics of the network to the dynamics of a single neuron in an effective PSP

$$h_i(t_n) = \sum_{\nu=1}^{s} q_\nu(t_n)\xi_i^\nu + \sqrt{\alpha}y(t_n) \tag{10}$$

where $\alpha = P/K$ is the storage ratio and $y(t)$ is a Gaussian variable with zero mean and

$$\langle y(t_n)y(t_m)\rangle = C_{nm} \tag{11}$$

where q_ν and C_{nm} have to be determined self-consistently from

$$\begin{aligned} q_\nu(t_n) &= \frac{1}{N}\sum_i \langle \xi_i^\nu S_i(t_n)\rangle_{\Phi,K,\xi} \\ &= \langle \xi^\nu S(t_n)\rangle_{\Phi,y,\xi} \end{aligned} \tag{12}$$

and

$$\begin{aligned} C_{nm} &= \frac{1}{N}\sum_i \langle S_i(t_n)S_i(t_m)\rangle_{\Phi,K,\xi} \\ &= \langle S(t_n)S(t_m)\rangle_{\Phi,y,\xi}. \end{aligned} \tag{13}$$

Note that the limit $K \to \infty$ has to be taken to achieve this result. q_ν is called the overlap of the network with pattern ξ^ν and $C_n := C_{nn} = \langle S^2(t_n)\rangle$ the mean activity of the network. Note that for uncorrelated patterns the system can have a non-vanishing overlap with only a finite number s of patterns. Starting from the single neuron dynamics in the effective local field (10) it is straightforward to obtain evolution equations for the order parameter q_μ and C. For parallel dynamics they are

$$q_\mu(t_{n+1}) = \int_{-\infty}^{\infty} \frac{dy}{\sqrt{2\pi}} e^{-y^2/2} \left\langle \xi^\mu \tanh\left(\frac{\xi^\nu \sum_{\nu=1}^{s} q_\nu(t_n) + \sqrt{\alpha C_n}y + \Phi}{\Gamma} \right) \right\rangle_{\xi,\Phi} \tag{14}$$

$$C_{n+1} = \int_{-\infty}^{\infty} \frac{dy}{\sqrt{2\pi}} e^{-y^2/2} \left\langle \tanh^2\left(\frac{\xi^\nu \sum_{\nu=1}^{s} q_\nu(t_n) + \sqrt{\alpha C_n}y + \Phi}{\Gamma} \right) \right\rangle_{\xi,\Phi}. \tag{15}$$

Note that due to the Gaussian character of y and the locality in time of the effective PSP (10), the C_{nm} with $n \neq m$ do not appear in these evolution equations.

3. Retrieval properties

The long-time behaviour of the system is governed by the fixed points of the evolution equations for q and C. For both parallel and random sequential updating the fixed points are given by

$$q = \int_{-\infty}^{\infty} \frac{dy}{\sqrt{2\pi}} e^{-y^2/2} \tanh\left(\frac{q + \sqrt{\alpha C} y}{\Gamma}\right) \tag{16}$$

$$C = \int_{-\infty}^{\infty} \frac{dy}{\sqrt{2\pi}} e^{-y^2/2} \tanh^2\left(\frac{q + \sqrt{\alpha C} y}{\Gamma}\right) \tag{17}$$

where it has been assumed that the network has a finite overlap with only one of the patterns: $q_\nu = q\delta_{\nu\nu_0}$ (Mattis state) and for the time being the thermal noise is set to zero.

It is worth noting that (16) and (17) are formally identical with the self-consistent equations found in the replica symmetric solution of the SK model with ferromagnetic interaction [15].

The stable solutions of (16) and (17) depend on α and Γ and are summarized in the (α, Γ) phase diagram in figure 1. A trivial fixed point is $(q = 0, C = 0)$. It is stable provided $\Gamma > 1$ and $\sqrt{\alpha} < \Gamma$. These inequalities determine phase II ('zero activity phase') of figure 1. Beside this trivial fixed point there are two possible kinds of stable fixed points, both with $C > 0$. One is the 'non-retrieval fixed point' $q = 0$, where the network does not remember any pattern (phase III), the other one is the retrieval fixed point $\pm q > 0$ (phase I), where the system has finite correlation with a pattern ξ or its complement $-\xi$. The basins of attraction of the q and the $-q$ fixed point are separated by the unstable fixed point $q = 0$. The line that separates the retrieval phase from the non-retrieval phase is denoted $\alpha_c(\Gamma)$ and is given implicitly by

$$\int_{-\infty}^{\infty} \frac{dy}{\sqrt{2\pi}} e^{-y^2/2} \cosh^{-2}\left(\frac{\sqrt{\alpha_c(1-\Gamma)} y}{\Gamma}\right) = \Gamma. \tag{18}$$

An expansion of this equation in powers of Γ yields

$$\alpha_c(\Gamma) = \frac{2}{\pi}\left[1 + \Gamma + \left(1 - \frac{\pi^3}{24}\right)(\Gamma^2 + \Gamma^3 + \Gamma^4) + \frac{\pi^6}{1920}\Gamma^4\right] + O(\Gamma^5). \tag{19}$$

$\alpha_c(\Gamma = 0) = 2/\pi$ as it should be since $\Gamma = 0$ corresponds to the well known case of binary neurons [10]. $\alpha_c(\Gamma)$ is monotonically increasing with Γ up to its maximum value $\alpha_c(\Gamma = 1) = 1$. It should be noted that the phase boundaries between the phases II and III and between the phases I and II as well as the endpoints of the line $\alpha_c(\Gamma)$ are independent of the special choice of the hyperbolic tangent as the I/O function as long as this function is sigmoidal, odd and has its maximum slope Γ^{-1} at argument zero.

The retrieval overlap q goes to zero continously if one approaches the phase boundaries α_c or $\Gamma = 1$ (figure 2(*a*)). Note that the second-order phase transition of $q(\alpha)$ at $\alpha = \alpha_C$ reflects the sign symmetry of the underlying dynamics. For I/O relations

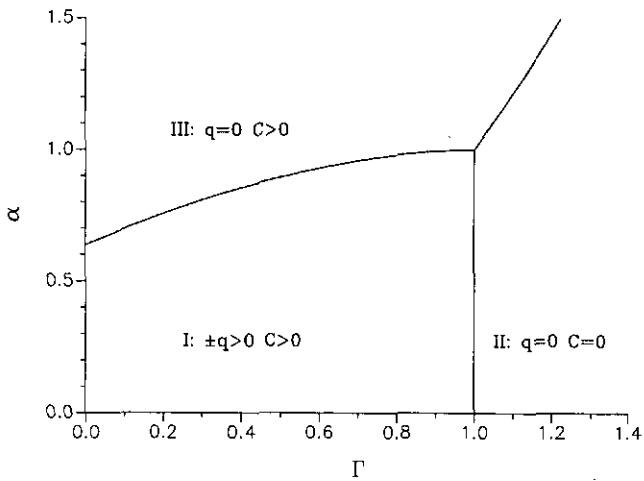

Figure 1. Zero-temperature phase diagram. Γ: inverse gain of the response-function; α: number of learned pattern per synapse. $\Gamma = 0$ corresponds to the binary Hopfield–Little model in its extremely diluted version.

without this symmetry (like the threshold linear function of Treves [9]) this phase transition is found to be of first order.

For $\Gamma < 1$ the activity C decreases with increasing α, takes on its minimum value $C_{\min} = 1 - \Gamma$ for $\alpha = \alpha_c$ and increases with α for $\alpha > \alpha_c$ up to $C = 1$ for $\alpha \to \infty$ (figure 2(*b*)). This behaviour can easily be understood considering the local field $h = \xi q + \sqrt{\alpha C}y$. h is Gaussian distributed with mean ξq and variance αC. For $\alpha > \alpha_c$, $q = 0$ and any increase of α increases the fluctuations of the PSP around zero, leading to an enhancement of $C = \langle S^2 \rangle$. For $\alpha < \alpha_c$ this effect is more than compensated by the decrease of q with increasing α which pushes the mean value of h closer to zero and therefore tends to decrease C.

$C < 1$ means that the neurons are operating away from their saturated levels ± 1. By the choice of the hyperbolic tangent as I/O relation for a single neuron one is forced to interpret the firing rates ± 1 as the maximum (minimum) firing rate that a postsynaptic potential can induce. Therefore $C < 1$ means that the neurons operate below their maximum and above their minimum firing rate. The former is desirable from a physiological point of view since real neurons appear to operate far below their maximum firing rate [16, 17]. The fact that the lower firing rates are increased at the same time is due to the symmetric treatment of inhibitory (negative) and excitatory (positive) contributions to the postsynaptic potential in our network.

It is interesting to note that the correlation with the pattern vanishes discontinuously at $\Gamma = 1$. This can be seen in figure 2(*c*) where the cosine Q of the angle between $\boldsymbol{\xi}$ and $\boldsymbol{S} = (S_1, \ldots, S_N)$

$$Q := \frac{q}{\sqrt{C}} \tag{20}$$

is plotted as a function of α and Γ. An expansion of (16) and (17) around $q = 0 = C$ yields

$$Q(\Gamma = 1) = \sqrt{1 - \alpha}. \tag{21}$$

This discontinouity of Q indicates that q goes down to zero continously for $\Gamma \to 1$ only because C does.

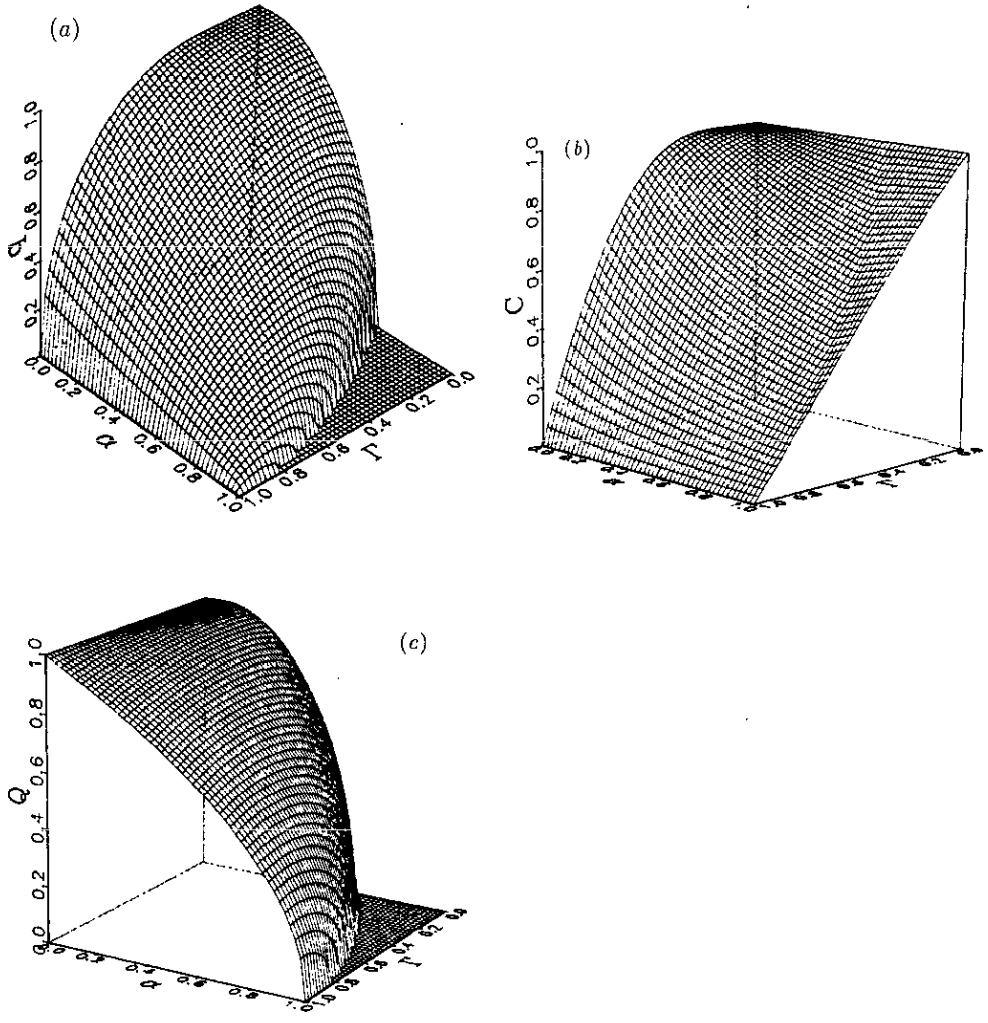

Figure 2. Fixed-point values of the order parameters q (a), C (b) and the normalized overlap $Q = q/\sqrt{C}$ (c) as functions of the storage ratio α and the inverse gain Γ for zero temperature.

α_c is a measure for the maximum number of patterns that can be stored and retrieved. The phase diagram indicates that this storage capacity can be increased by allowing the neurons to react more smoothly on the postsynaptic potential. This is exactly the opposite effect that one can observe in the fully connected version of this model, where $\alpha_c(\Gamma)$ is a monotonically decreasing function [4,6]. The root of this distinct behaviour lies in the way the noise enters the PSP in both models. Here we have a Gaussian noise whose variance scales with C and the net effect of decreasing q (bad!) and decreasing C (good!) with increasing Γ allows for an enhancement of α_c. In the fully connected model the noise depends in a more complicated manner on the order parameters. It has been shown [4,6] that the inverse gain Γ of the analogue neuron model with deterministic dynamics and the temperature in the usual binary Hopfield–Little model with thermal noise play a very similiar role. With this analoguey in mind it should not come as a surprise that $\alpha_c(\Gamma)$ is a decreasing function.

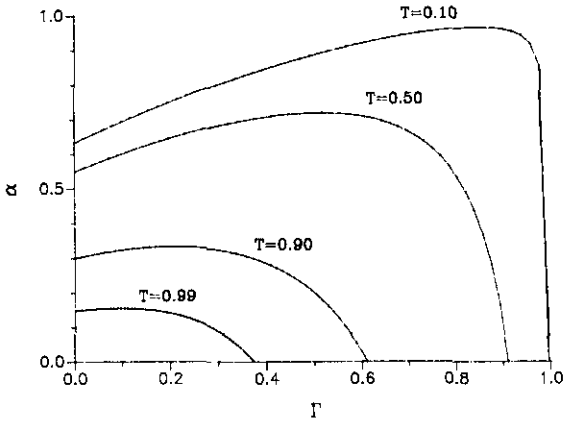

Figure 3. Phase diagram for four different temperatures. The area underneath each curve is the retrieval phase with $q \neq 0$. Note that the storage capacity $\alpha_c(\Gamma)$ increases for small values of Γ and decreases for larger values of Γ.

In the presence of additional noise, the monotonicity of $\alpha_c(\Gamma)$ is no longer guaranteed. This can be seen most easily by adding thermal noise with finite strength to the PSP. Now there is of course no fixed point with $C = 0$ and we distinguish only between two phases: the retrieval phase with $q \neq 0$ and the non-retrieval phase with $q = 0$. Figure 3 shows the effect of thermal noise distributed according to

$$w(\Phi) = \begin{cases} 1/2T & \text{if } -T \leq \Phi \leq T \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

on the phase diagram. The 'temperature' T is a measure for the fluctuations in the synaptic noise: $\langle \Phi^2 \rangle_w = T^2/3$. The retrieval phase gets smaller with increasing temperature and above a critical temperature $T_c = 1$ there is no retrieval phase at all. The endpoints of the phase boundary $\alpha_c(\Gamma)$ resp. $\Gamma_c(\alpha)$ are given by

$$\sqrt{2\alpha_c(\Gamma = 0)} = \frac{T}{\mathrm{erf}^{-1}(T)} \tag{23}$$

and

$$\Gamma_c(\alpha = 0) = \frac{T}{\tan^{-1} T} \tag{24}$$

where $\mathrm{erf}^{-1}(x)$ is the inverse of the error function

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{d}t\, \mathrm{e}^{-t^2}. \tag{25}$$

For finite temperature the curve $\alpha_c(\Gamma)$ is still monotonically increasing up to a certain value of Γ, but for larger values of Γ it decreases! The explanation of this behaviour is clear: in the presence of thermal fluctuations C cannot become arbitrary small as Γ increases, while $q(\Gamma)$ still tends to zero. The balance between $q(\Gamma)$ and $C(\Gamma)$ is disturbed and α_c decreases. For small Γ, the pattern-induced noise is much more important than the thermal fluctuations and the behaviour of $\alpha_c(\Gamma)$ is the same as for $T = 0$.

4. Information capacity

A network with analogue neurons can represent in principle more complex patterns than just 'on/off' vectors. The equations in the appendix show that the retrieval properties of the network are to a great extent independent of the distribution of pattern activities—at least for Mattis states and as long as each ξ_i^ν is drawn independently from the same distribution $\rho(\xi)$. We denote the second moment of this distribution with $\langle\xi^2\rangle$ and define the renormalized inverse gain Γ^* and temperature T^* to be

$$\Gamma^* := \frac{\Gamma}{\langle\xi^2\rangle} \qquad T^* := \frac{T}{\langle\xi^2\rangle}. \tag{26}$$

Then the phase diagrams figures 1 and 3 are valid for any distribution $\rho(\xi)$, provided Γ and T are replaced by their renormalized values. For fixed gain and zero temperature this implies that the storage capacity α_c increases with decreasing $\langle\xi^2\rangle$.

For a network that processes more complex patterns than binary ones, the information capacity of the network is an interesting measure of performance. It is defined as the total amount of information that can be stored and retrieved in the network. The concept of information transfer is borrowed from the theory of communication via a noisy channel [18] and has been applied in the context of neural networks by several authors—see for example [19].

Suppose we have p-state patterns, i.e. $\rho(\xi)$ consists of p δ-functions

$$\rho(\xi) = \sum_{k=1}^{p} w_k \delta(\xi - \eta_k) \tag{27}$$

with $\sum_k w_k = 1$. Then the information content in bits of one pattern is given by

$$-N \sum_{k=1}^{p} w_k \log_2 w_k. \tag{28}$$

In the absence of thermal noise and for $\alpha = 0$ and $0 < \Gamma^* \leq 1$ we have a one-to-one correspondence between the fixed point response level S_i and the corresponding activity ξ_i of the condensed pattern: the total amount of information contained in the pattern is preserved in the retrieval state. This information content can be very high for multistate patterns: it grows roughly like $\log p$. Note that this preservation of information for $\alpha = 0$ holds for any degree of synaptic dilution—even for the fully connected model.

Thermal noise or a finite α destroys this strict one-to-one correspondence between the response levels and the pattern activities η_k: information is lost in the process of storage and retrieval. The amount of lost information is defined as the information needed to reconstruct the correct ξ_i from the response levels S_i. To calculate this we divide the interval $[-1, +1]$ of response levels into n sub-intervals and denote with w_k^l the probability that the response level of a neuron falls in the l^{th} interval *and* the corresponding activity of the condensed pattern was η_k. Then the information needed to reconstruct the pattern from this joined probability distribution is given by

$$-N \sum_{l=1}^{n} \sum_{k=1}^{p} w_k^l \log_2 \frac{w_k^l}{w^l} \tag{29}$$

where $w^l = \sum_k w_k^l$ is the probability of finding a response level in the l^{th} interval non-regarding the pattern activity. This amount of information has to be subtracted from (28) to yield the amount of information that can be stored and retrieved for a single pattern. Since we can store and retrieve αK patterns in the network, we arrive at

$$I = \alpha \sum_{l=1}^{n} \sum_{k=1}^{p} w_k^l \log_2 \frac{w_k^l}{w_k w^l} \tag{30}$$

for the total information capacity of the network measured in bits/synapse.

Of course I depends on the way the continuous range of response levels is divided into n intervals. Without any *a priori* knowledge about how the pattern is encoded in the distribution of response levels it seems reasonable to choose an equidistant subdivision into n intervals $[U_{l-1}, U_l]$ with

$$U_l := -1 + \frac{2l}{n}. \tag{31}$$

Now I still depends on the resolution, i.e. on the number of intervals. We will study two cases. The high-resolution limit $n \to \infty$ and the limit of minimal resolution $n = p$. It follows from the convexity of $-x \log x$ that I increases with increasing n. In that sense, the information capacity in the high-resolution limit (denoted I_∞) is the optimal value that can be achieved. $n = p$ is the minimal resolution that allows the network *in principle* to recover the total information contained in a p-state pattern. This choice of n can be motivated by the assumption that learning and retrieving may proceed with the same resolution.

In the absence of thermal noise I can be calculated from

$$w_k^l = \frac{w_k}{2} \operatorname{erf}\left(\frac{\Gamma \tan^{-1} U_l - \eta_k q}{\langle \xi^2 \rangle \sqrt{2\alpha C}}\right) - \frac{w_k}{2} \operatorname{erf}\left(\frac{\Gamma \tan^{-1} U_{l-1} - \eta_k q}{\langle \xi^2 \rangle \sqrt{2\alpha C}}\right) \tag{32}$$

for finite n. In the high-resolution limit we have

$$I_\infty = -\frac{\alpha}{\ln 2}\left(\frac{1}{2} + \int_{-\infty}^{\infty} \frac{dy}{\sqrt{2\pi}} e^{-y^2/2} H(y)\right) \tag{33}$$

with

$$H(y) = \sum_{k=1}^{p} w_k \ln \sum_{k'=1}^{p} w_{k'} \exp\left[-\frac{1}{2}\left(y + \frac{(\eta_k - \eta_{k'})q}{\langle \xi^2 \rangle \sqrt{\alpha C}}\right)^2\right] \tag{34}$$

where q and C are the usual order parameters. One could easily study the evolution of I in time during the process of retrieval but we confine ourselves to the long time behaviour, i.e. we insert the stationary values of the order parameters into the expressions for I.

For simplicity we assume that the pattern activities can take on the values

$$\eta_k = -1 + 2\frac{k-1}{p-1} \tag{35}$$

with equal probability, i.e. $w_k = 1/p$. Figure 4 shows the information capacity for 3-state patterns over the whole retrieval phase for both maximum and minimum resolution. In both cases, I varies strongly with α and reaches its maximum value somewhere halfway in-between 0 and α_c. The dependence on Γ^* is much weaker in the high-resolution limit than it is for minimum resolution. This can easily be understood by taking into account that Γ^* enters I_∞ only indirectly via the value of $q/\sqrt{C}$, which in turn varies very slowly with Γ (figure 2(c)), whereas for minimum resolution there is a direct interference between Γ^* and the fixed read out intervals $[U_{l-1}, U_l]$. Consider for example the case Γ^* close to 1: Though the ratio $q/\sqrt{C}$ is finite, the PSP is close to zero for all possible pattern activities and due to the finite resolution most of the response levels S_i fall into the same interval—the one that contains the zero activity response level. This uniform assignment implies of course a complete loss of information about the condensed pattern which is contained in the fine structure of the response level distribution. The same mechanism works in the regime of small Γ^*, i.e. very steep I/O function. Here most of the response levels fall into the intervals $[-1, -1 + 2/p]$ and $[1 - 2/p, 1]$ and it is only the information about the sign of the pattern activities which is retrieved. This explains the stronger dependence of I on Γ^* for minimal resolution. However, even I_∞ has a flat maximum as a function of Γ^*.

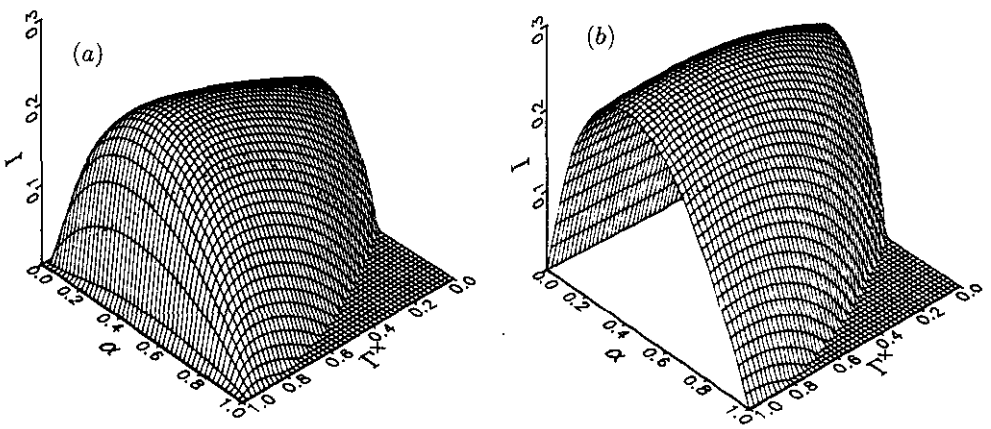

Figure 4. Information capacity for 3-state patterns in the high resolution limit (a) and for minimum resolution (b) in bits/synapse.

Table 1. Maximum information capacity in bits/synapse for the storage and retrieval of p-state patterns in the limits of high (I_∞) and minimal (I) resolution.

p	2	3	4	5	10	15
I_∞	0.283	0.271	0.269	0.269	0.268	0.268
I	0.232	0.227	0.241	0.249	0.262	0.266

In table 1 the value of the maximum (as a function of α and Γ^*) information content is monitored for different values of p for both resolution regimes. Let us consider first the high-resolution limit. Despite the fact that both the information content of a pattern and the storage capacity α_c increase with increasing p, the information capacity I_∞ decreases. For large p the 'signals' $\xi_i q$ in the PSP are very close for

adjacent values of the discrete pattern activities. This enhances the impact of the noise on the process of filtering out the correct η_k. According to the table, this deterioration of the retrieval quality more than compensates for the higher *a priori* information content of the pattern and the increase of α_c. In terms of information capacity, the optimal representation of information seems to be the binary one—at least in the high-resolution limit. This result agrees well with the findings of Treves [9] for the threshold linear model and biased pattern distributions.

For minimal resolution the situation is different. Here the enhancement of the resolution in the readout process with growing p is the dominant effect, leading to an increase of I with increasing p. For large values of p both readout procedures converge to the same value of maximum information content:

$$p \gg 2 \qquad I \simeq I_\infty \simeq 0.268. \tag{36}$$

In both cases this value is achieved for $\alpha \simeq 0.4$ and $\Gamma^* \simeq 0.75$. Note that even in the case of minimal resolution and $p = 2$, i.e. when the information is processed completely in terms of the sign of the S_i, the maximum information capacity of the graded response network ($I \simeq 0.232$ for $\Gamma \simeq 0.4$) exceeds that of the usual binary version ($I \simeq 0.216$ for $\Gamma = 0$).

5. Summary

Networks of analogue neurons have recently attracted much interest. A common result of all the research in that field seems to be that well known analytical tools, originally developed for discrete variables, can very easily be extended to cope with continuous quantities. The present work is no exception in that it demonstrates how the technique of tackling extremely diluted asymmetric networks as was presented in [11], can be used to investigate such networks with a sigmoidal I/O relation.

We have seen that the storage capacity α_c increases with decreasing gain of the response function as long as the thermal fluctuations can be neglected with respect to the pattern-induced noise. In the regime where the thermal noise becomes more important, $\alpha_c(\Gamma)$ decreases.

We further addressed the question of the amount of information that can be transferred by the network from the learned patterns to the retrieval state. This information capacity depends on the amount of information encoded in the patterns, the retrieval properties of the network and the procedure to read out the information. We have seen that the maximum information capacity for unbiased p-state patterns converges for large p rather quickly to a value of about 0.268 bits/synapse, though the information content of a single pattern grows like $\log_2 p$. For a readout procedure with high resolution, the best information transfer is achieved for binary patterns.

Acknowledgments

It is a pleasure for me to thank Siegfried Bös, Reimer Kühn, Reiner Kree, Heiko Rieger, Alessandro Treves and Annette Zippelius for stimulating discussions and valuable hints. I am further indebted to the Studienstiftung des deutschen Volkes for financial support.

Appendix.

The formalism of extremely diluted, asymmetric networks as has been presented recently by Kree and Zippelius [11] permits the analytical treatment of a very broad class of networks, avoiding at the same time the cumbersome combinatorial arguments of the original formalism [10]. Consider parallel or random sequential dynamics with an arbitrary I/O relation dyn(h):

$$S_i(t_{n+1}) = \mathrm{dyn}\left[h_i(t_n) + \Phi_i(t_n)\right]. \tag{A1}$$

It should be noted that no constraints like for instance monotonicity, continuity or boundedness have to be imposed on dyn in the course analytical solution. The PSP $h_i(t)$ is composed of two parts:

$$h_i(t) = \frac{1}{K}\sum_{j=1}^{N}{}' c_{ij} J_{ij} S_j(t) + \frac{1}{N}\sum_{j=1}^{N} f[S_j(t)]. \tag{A2}$$

The first term is the usual weigthed sum of the incoming signals (the c_{ij} are the dilution coefficients) and the second term is added to capture the influence of the mean network activity on the local field. As an example for the usefulness of such a term the reader may consider $f(S) = -\gamma S - U$/with $\gamma, U > 0$. This adds a constant threshold and an inhibitory term proportional to the overall activity $1/N\sum_j S_j$ to the local field. In a network that operates with (0,1) variables (choose $\mathrm{dyn}(h) = \Theta(h)$) such terms are indispensable to stabilize the 0 state.

We want the system to learn αK uncorrelated random patterns $\boldsymbol{\xi}^\nu$, i.e. the ξ_i^ν are drawn independently from a distribution $\rho(\xi)$. The learning rule is written as

$$J_{ij} = \sum_{\nu=1}^{\alpha K} R(\xi_i^\nu, \xi_j^\nu) \tag{A3}$$

which is the most general learning rule with the properties of *locality* (i.e. J_{ij} depends exclusively on the pattern activities at sites i and j), *homogeneity* (i.e. the learning mechanism is the same for all patterns and all sites) and *additivity* (i.e. each new pattern adds a modification to the already existing couplings). In the Hopfield–Little model and in the main text we have simply $R(x, y) = xy$.

A simple signal-to-noise analysis suggests that R should obey

$$\langle R(x,\xi)\rangle_\xi = 0 \qquad \text{for all } x$$

in order to allow the storage ratio α to be O(1). $\langle\cdot\rangle_\xi$ denotes the average over $\rho(\xi)$. A further constraint on R required by the analytical solution is a factorization of R into a pre- and a postsynaptic part. Both conditions are matched by assuming R to be of the form

$$R(x,y) = \mathrm{post}(x)\left[\mathrm{pre}(y) - \langle \mathrm{pre}(\xi)\rangle_\xi\right] \tag{A4}$$

where we have split the synaptic plasticity into a post- and a presynaptic factor. Consult [9, 20] for neural networks with other choices of post(x) and pre(x) than just

post(x) = pre(x) = x. In particular, [9] discusses the effects of non-linear functions post(x) on the storage and retrieval of multistate patterns.

Equations (A1)–(A4) define a class of networks that can be solved analytically by simply following step-by-step the calculation outlined in [11]. Here we only quote the results. Introducing the order parameters

$$q_\nu(t_n) = \left\langle \frac{1}{N}\sum_{j=1}^{N}\left[\,\mathrm{pre}(\xi_j^\nu) - \langle \mathrm{pre}(\xi)\rangle_\xi\,\right] S \right\rangle_{c,\xi,\Phi} \tag{A5}$$

$$m(t_n) = \left\langle \frac{1}{N}\sum_{j=1}^{N} f[S_j(t_n)] \right\rangle_{c,\xi,\Phi} \tag{A6}$$

$$C_{mn} = \left\langle \frac{1}{N}\sum_{j=1}^{N} S_j(t_n) S_j(t_m) \right\rangle_{c,\xi,\Phi} \tag{A7}$$

and the abbreviation

$$\langle R^2\rangle := \left\langle R^2(\xi,\tilde{\xi})\right\rangle_{\xi,\tilde{\xi}}$$

one obtains as the effective PSP for the single-neuron dynamics ($K \to \infty$)

$$h(t_n) = \sum_{\nu=1}^{s} \mathrm{post}(\xi^\nu) q_\nu(t_n) + m(t_n) + \sqrt{\alpha\langle R^2\rangle}\, y(t_n) \tag{A8}$$

where $y(t_n)$ is Gaussian with zero mean and variance

$$\langle y(t_n) y(t_m)\rangle = C_{nm}. \tag{A9}$$

From this single-neuron dynamics one can easily construct evolution equations for the order parameters $m(t_n)$, $q_\nu(t_n)$ and $C(t_n) := C_{nn}$. For parallel dynamics these equations are

$$m(t_{n+1}) = \int_{-\infty}^{\infty} \frac{\mathrm{d}y}{\sqrt{2\pi}} \mathrm{e}^{-y^2/2} \left\langle f[\mathrm{dyn}(h(t_n) + \Phi)] \right\rangle_{\xi,\Phi} \tag{A10}$$

$$q_\nu(t_{n+1}) = \int_{-\infty}^{\infty} \frac{\mathrm{d}y}{\sqrt{2\pi}} \mathrm{e}^{-y^2/2} \left\langle \left(\mathrm{pre}(\xi^\nu) - \langle \mathrm{pre}(\xi)\rangle_\xi \right) \mathrm{dyn}(h(t_n) + \Phi) \right\rangle_{\xi,\Phi} \tag{A11}$$

$$C(t_{n+1}) = \int_{-\infty}^{\infty} \frac{\mathrm{d}y}{\sqrt{2\pi}} \mathrm{e}^{-y^2/2} \left\langle \mathrm{dyn}^2(h(t_n) + \Phi) \right\rangle_{\xi,\Phi} \tag{A12}$$

with

$$h(t_n) = \sum_{\nu=1}^{s} q_\nu(t_n)\,\mathrm{post}(\xi^\nu) + m(t_n) - \sqrt{\alpha\langle R^2\rangle C(t_n)}\, y.$$

Setting $f \equiv 0$, $\mathrm{post}(x) \equiv \mathrm{pre}(x) := x$ and $\mathrm{dyn}(h) \equiv \tanh(h/\Gamma)$ one can easily verify that the resulting equations lead to the phase diagram shown in figure 1 with the scalings of (26)—independently of the pattern distribution.

References

[1] Mezard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)

[2] Amit D J 1989 *Modelling Brain Function* (Cambridge: Cambridge University Press)

[3] Hopfield J J 1984 *Proc. Natl Acad. Sci., USA* **81** 3088

[4] Marcus C M, Waugh F R and Westervelt R M 1990 *Phys. Rev* **41** *3355*

[5] Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev.* A **32** 1007; 1985 *Phys. Rev. Lett.* **55** 1530; 1987 *Ann. Phys., NY* **173** 30

[6] Kühn R, Bös S and van Hemmen L 1990 *Neural Networks, Proc. XI Sitges Conf.* ed L Garrido (Berlin: Springer)

[7] Treves A 1990 *J. Phys. A: Math. Gen.* **23** 2631

[8] Treves A 1990 *Phys. Rev.* A **42** to appear

[9] Treves A 1991 Dilution and sparse coding in threshold-linear nets *J. Phys. A: Math. Gen.* **24** 327

[10] Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* **4** 167

[11] Kree R and Zippelius A 1990 *Physics of Neural Networks* ed E Domany, J L van Hemmen and K Schulten (Berlin: Springer)

[12] Rieger H 1990 *Neural Networks, Proc. XI Sitges Conf.* ed L Garrido (Berlin: Springer)

[13] Peretto P 1984 *Biol. Cyb.* **50** 51

[14] Hilhorst H and Nijmeijer M 1987 *J. Physique* **48** 185

[15] Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1792

[16] Rubin N and Sompolinsky H 1989 *Europhys. Lett.* **10** 465

[17] Treves A and Amit D J 1989 *J. Phys. A: Math. Gen.* **22** 2205

[18] Shannon C E and Weaver W 1949 *The Mathematical Theory Of Communication* (Urbana, IL: University of Illinois Press)

[19] Amit D J, Gutfreund H and Sompolinsky H 1987 *Phys. Rev.* A **35** 2293
Buhmann J, Divko R and Schulten K 1989 *Phys. Rev.* A **39** 2689
Gardner E 1987 *Europhys. Lett.* **4** 481; 1988 *J. Phys. A: Math. Gen.* **21** 257

[20] Gardner E, Mertens S, and Zippelius A 1989 *J. Phys. A: Math. Gen.* **22** 2009