

Statistical mechanics for neural networks with continuous-time dynamics

R Kühn† and S Böst‡

† Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 19, D-6900 Heidelberg, Federal Republic of Germany

‡ Institut für Theoretische Physik, Universität Giessen, Heinrich-Buff-Ring 16, D-6300 Giessen, Federal Republic of Germany

Received 21 July 1992, in final form 15 September 1992

Abstract. The present paper is intended to summarize our current knowledge about the long-time behaviour of networks of graded response neurons with continuous-time dynamics. We demonstrate the workings of our previously developed statistical-mechanical approach to continuous-time dynamics by applying it to networks with various forms of synaptic organization (learning rules), and neural composition (neuron-types as encoded in gain functions), as well as to networks varying with respect to the ensemble of stored data (unbiased and low-activity patterns). We present phase diagrams and compute distributions of local fields for a variety of examples. Local field distributions are found to deviate from the Gaussian form obtained for stochastic neurons in the context of the replica approach. A solution to the low firing rates problem within the framework of nets of analogue neurons is also briefly discussed. Finally, the statistical-mechanical approach to the analysis of continuous-time dynamics is extended to include effects of fast stochastic noise. Detailed-balance solutions are shown to be unique and of canonical form, governed by Hamiltonians which exhibit a reciprocity relation between potential-dynamics and firing-rate dynamics: for the potential-dynamics, the Hamiltonian is given by the Lyapounov function of the system—expressed in terms of the firing rates—and it generates a Gibbs distribution over firing-rate space. For the noisy firing-rate dynamics, the same Lyapounov function—now expressed in terms of neural potentials—generates a Gibbs distribution over the space of these potentials. As a consequence, the firing-rate dynamics will freeze in configurations saturating the neural input-output relations, whenever such saturation levels exist. Both types of stationary distribution are shown to exist only under unrealistic assumptions about the noise in the system.

1. Introduction

In previous papers (Kühn 1990, Kühn *et al* 1991), we have provided a general statistical mechanical framework for analysing the long-time behaviour of networks of graded-response neurons with a deterministic continuous-time dynamics of the form

$$C_i \frac{dU_i}{dt} = \sum_{j=1}^N J_{ij} V_j - \frac{U_i}{R_i} + I_i \quad (1)$$

$$V_j = g_j(\gamma_j U_j). \quad (2)$$

In (1), C_i denotes the input capacitance of the i th neuron, R_i its trans-membrane resistance, U_i its postsynaptic potential, and V_i its instantaneous output. The input-output characteristics of a neuron is encoded in its transfer (gain) function g_j as in (2),

γ_j denoting a gain parameter. The I_i represent the input from afferent (current) sources and the synaptic weights are as usual denoted by J_{ij} . For our statistical-mechanical approach to be applicable, the dynamics (1), (2) must be governed by a Lyapounov function, a condition that is satisfied if synapses are symmetric, and if neural gain functions are monotonically increasing functions of their argument.

Networks of graded response neurons with dynamics governed by a set of differential *RC*-charging equations as in (1), (2) were proposed by Hopfield (1984)—at that time mainly as supplying further independent evidence for the degree of robustness of collective, network-based computation. The (qualitative) statement was that ‘networks of graded response neurons have collective properties like those of two-state neurons’ (Hopfield 1984). An explicit demonstration of the range of validity of the ‘universality principle’ alluded to in that statement may well be the strongest motivation for further quantitative studies of networks described by (1), (2), as long as a proper description of the dynamics of natural nerve nets is still missing or—at best—under debate (see e.g. recent papers by Amit and Tsodyks (1991a, b), or Gerstner and van Hemmen (1992)).

There are, of course, also a number of *specific* points to be advanced in favour of (1), (2). One is that the continuous-time dynamics (1), (2) carries some potential for the inclusion of neurophysiological detail into formal neural network models, which is not available in the standard models using two-state neurons with (stochastic) synchronous or asynchronous dynamics. For instance, capacitive input delays and trans-membrane leakages are explicitly taken into account in (1)—input delays, however, certainly not as detailed as the variability of synaptic–dendritic information transport would require. Moreover, within a firing-rate description at least, gain functions as in (2) can be shown to encode neural behaviour during relative refractory periods. Finally, continuous-time dynamics avoids the most obvious shortcomings of parallel and asynchronous dynamics with respect to neural modelling: (i) there is no need for assuming agents that would enforce global synchrony in the case of parallel dynamics; (ii) neurons determine their state strictly in response to their postsynaptic potential *rather than only when the order of updating is on them*, as in asynchronous dynamics.

Alternatively, (1), (2) provide a quantitative description of the dynamics of networks of resistively coupled nonlinear amplifiers (Hopfield 1984, Mead 1989), in which case the V_j denote output voltages (rather than firing rates) of amplifiers with gain functions g_j . Such networks have been suggested (Hopfield and Tank 1985, Tank and Hopfield 1985, Koch *et al* 1986) as real-time solvers for hard optimization tasks, and a quantitative theoretical understanding of their performance would thus be of use as a guide for improving the efficiency of devices of this type. In particular, it is now known that finite gain (like finite temperature in stochastic Ising networks) can be used quite efficiently to control the number of spurious stable states (Fukai and Shiino 1990, Waugh *et al* 1990). Tuning gain parameters in analogue systems may therefore serve, e.g., as a fast deterministic substitute of simulated annealing—much like mean-field annealing (Soukoulis *et al* 1983, Peterson and Anderson 1987). Knowledge of the phase structure of a given setup, i.e. its collective properties, would certainly be of help in using this method efficiently, in that it points out ways of avoiding spurious attractors or in that it can guarantee, for a given problem, the stability of phases of interest in appropriate parameter ranges.

Dynamical properties of networks of graded-response neurons governed by (1), (2) have been investigated by Sompolinsky *et al* (1988). These authors studied networks

with random asymmetric synapses, for which (1), (2) are *not* governed by a Lyapounov function, and found a transition to chaotic behaviour at sufficiently high gains for an input–output relation of the form $V_i = \tanh(\gamma U_i)$. More recently, collective properties of networks of analogue neurons endowed with Hebb–Hopfield-type couplings have been discussed by various authors. Treves (1990a, b) investigated networks of threshold-linear neurons governed by asynchronous stochastic dynamics. Stability properties of analogue neuron systems with discrete-time iterated map dynamics have been investigated by Marcus and Westervelt (1989), by Marcus *et al* (1990), and by Herz (1991). A phase diagram describing the fixed-point structure of analogue neurons with hyperbolic-tangent response was obtained by Shiino and Fukai (1990) on the basis of a mapping of this particular problem onto a so-called naive mean-field theory for Ising spin systems (Bray *et al* 1986)—an approach, however, which is restricted to neurons with $g(x) = \tanh(x)$ response. More general transfer functions as well as networks with asymmetric couplings can be handled by a cavity-type approach recently proposed by the same authors (Shiino and Fukai 1992); for an earlier attempt, see also Marcus *et al* (1990).

We will not mention nor discuss here recent related work on multi-state neuron systems with Ising-, Potts-, Clock- or XY-symmetries and refer the reader to recent papers by Mertens *et al* (1991), Bollé *et al* (1991) and Gerl *et al* (1992) and to references therein.

The purpose of the present paper is to summarize our current knowledge about the long-time behaviour of networks of graded-response neurons with continuous-time dynamics. In order to set the scene, we shall first (section 2) provide a brief description of the statistical-mechanical approach proposed earlier (Kühn 1990, Kühn *et al* 1991), starting with a soft-neuron version of the Hopfield model (section 3.1), thereafter extending it in various ways. In particular, in section 3.4 we will be concerned with the effect of network inhomogeneities (mixtures of several neuron types and distributions of gain parameters), and self-couplings. In section 4 we investigate synaptic organizations (learning rules) different from the Hebb–Hopfield form (Hopfield 1982); that is, in section 4.1 we shall have a few words to say on low-activity low-firing-rates networks. Such networks were recently and independently also studied by Amit and Tsodyks (1991b), so here we will restrict ourselves to results which go beyond theirs. Section 4.2 is devoted to a study of networks of analogue neurons coupled via pseudo-inverse synapses. In section 5, we study the effects of fast stochastic noise on the continuous-time dynamics of networks of graded response neurons. Guided by previous approaches, we concentrate on long-time stationary states, *assuming* that they satisfy detailed balance conditions. The nature of the corresponding stationary distributions is elucidated, and it is shown that they exist only under unrealistic assumptions about the noise in the system.

2. Statistical mechanics for analogue neurons

The statistical-mechanical approach to graded-response neurons to be presented below is restricted to situations where the dynamics (1), (2) is governed by a Lyapounov function. Such is the case, if the synaptic matrix in (1) is symmetric and if the neurons have monotone increasing input–output relations. These conditions were identified by Cohen and Grossberg (1983) and by Hopfield (1984), and the

Lyapounov function was shown to be of the form

$$\mathcal{H}_N = -\frac{1}{2} \sum_{i,j=1}^N J_{ij} V_i V_j + \sum_{i=1}^N \frac{1}{\gamma_i R_i} G_i(V_i) - \sum_{i=1}^N I_i V_i \quad (3)$$

where G_i denotes the integrated inverse input-output relation

$$G_i(V) = \int^V g_i^{-1}(V') dV' \quad (4)$$

The value of the lower integration limit in (4) is arbitrary. It can be used to define the zero of the energy scale in (3). In figure 1, the interaction energy of a pair of ferromagnetically coupled neurons with $V_i = \tanh(\gamma U_i)$ in the subspace $V_1 = V_2 = V$ is depicted for various values of the gain parameter γ . In terms of (3), the dynamics (1), (2) reads

$$C_i \frac{dU_i}{dt} = -\frac{\partial \mathcal{H}_N}{\partial V_i} \quad (5)$$

entailing

$$\frac{d\mathcal{H}_N}{dt} = \sum_{i=1}^N \frac{\partial \mathcal{H}_N}{\partial V_i} \frac{dV_i}{dt} = -\sum_{i=1}^N C_i \gamma_i g'_i(\gamma_i U_i) \left(\frac{dU_i}{dt} \right)^2 \leq 0 \quad (6)$$

with equality in (6) *only* at stationary points of (1), (2). The *existence* of stationary points follows from one additional assumption on the g_i , namely that they increase, for large $|U_i|$, not faster than linearly with U_i (Marcus and Westervelt 1989). This assumption guarantees that \mathcal{H}_N is bounded from below, so that the dynamical flow generated by (1), (2) will *always* converge to fixed points, which are the global or local minima of \mathcal{H}_N .

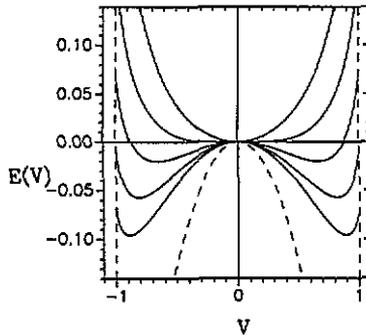


Figure 1. Interaction energy of a pair of neurons with $J_{ij} = 1$, $I_i = 0$, and $V_i = \tanh(\gamma U_i)$ in the subspace $V_1 = V_2 = V$ for various values of the gain parameter γ . From top to bottom, we have $\gamma = 0.8, 1.0, 1.2, 1.4, 1.6$, and (broken curve) $\gamma = \infty$.

One way of locating these minima is to compute the zero-temperature ($\beta \rightarrow \infty$) limit of the free energy

$$\begin{aligned}
 f_N(\beta) &= -(\beta N)^{-1} \log \text{Tr}_{\mathbf{V}} \exp[-\beta \mathcal{H}_N(\mathbf{V})] \\
 &= -(\beta N)^{-1} \log \int \prod_i d\rho(V_i) \exp[-\beta \mathcal{H}_N(\mathbf{V})]
 \end{aligned}
 \tag{7}$$

and to investigate the nature of its stable and metastable phases. This allows us to find all the attractors of the dynamics (1), (2) which are surrounded by extensive energy barriers and, as we shall see, to characterize them macroscopically (Kühn 1990, Kühn *et al* 1991).

In (7), $d\rho(V_i)$ denotes an *a priori* measure on the space of neural output states which—guided by the principle of insufficient reason—we take to be uniform (though not normalized) on its support, namely the range of g_i , and thereby avoid encoding hidden assumptions about the system’s behaviour already at the level of this output measure. Later on we shall see that (smoothness taken for granted) the support of $d\rho(V_i)$ is all that matters anyway, as long as we are interested only in zero-temperature properties of the Gibbs measure generated by \mathcal{H}_N . Moreover, the analysis of stochastic generalizations of the dynamics (1), (2) below will be seen to lend additional support to this ‘natural’ choice.

If \mathcal{H}_N is *not* bounded from below, the statistical-mechanical approach just described can nevertheless be used to exhibit and characterize the *local* minima of \mathcal{H}_N , by constraining each V_i integration in (7) to some suitably large compact subset of \mathbb{R} , i.e. by giving the $d\rho(V_i)$ in (7) a finite support which may be a proper subset of the range of g_i . By this device, we ensure that the Lyapounov function is bounded on the support of $\prod_i d\rho(V_i)$, hence that (7) exists. Such a strategy will work as long as \mathcal{H}_N is, for example, a continuous function of its arguments. More generally, for this purpose \mathcal{H}_N is only required to be of bounded variation on the integration domain.

Let us note in passing that the input capacitances C_i of the neurons do not enter \mathcal{H}_N and thus do not affect the free energy. That is, the characteristic capacitive input-delays $\tau_i = R_i C_i$ of individual neurons do *not* affect the nature or the number of fixed points of (1), (2). They must of course be expected to determine the size and shape of the basins of attraction, convergence times, the way in which fixed points are approached, and other intrinsically dynamic features of the network. Note also that self-couplings may appear in the dynamical rule (1), (2) without invalidating the statistical-mechanical approach, in contrast to the situation for binary neurons with asynchronous stochastic dynamics.

In what follows, we shall exclusively be concerned with the long-time static properties of networks described by (1), (2), i.e. we will have nothing to say about the dynamics proper.

The statistical-mechanical approach just described may at first sight appear like an unnecessary detour. One might, after all, think of searching for minima of \mathcal{H}_N directly among the solutions of the equations

$$\frac{\partial \mathcal{H}_N}{\partial V_i} = 0 \quad 1 \leq i \leq N.
 \tag{8}$$

This alternative, however, has proven to be impractical for at least two reasons. Firstly, the tools for solving the $N \times N$ -dimensional stability problem associated with

(8) were not available in all cases of interest, and secondly (8) generally do not allow a satisfactory macroscopic characterization of the relevant minima in terms of order parameters. For both types of problems, on the other hand, the arsenal of techniques provided by statistical mechanics is able to furnish satisfactory solutions, that is to say, in all cases we have so far explored. Technically, these solutions are provided by mean-field techniques of a form introduced by Amit *et al* (1987), albeit modified where necessary, in order to account for the continuous nature of our fundamental dynamical variables.

3. Hebb–Hopfield couplings

We now proceed to substantiate the above general considerations by studying a number of specific examples. We begin by briefly recalling the mean-field theory for networks of analogue neurons with Hebb–Hopfield couplings

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu} \quad i \neq j \quad (9)$$

designed to store a set of p unbiased binary random patterns $\xi_i^{\mu} \in \{\pm 1\}$, $1 \leq \mu \leq p$. Biased patterns will be discussed later on in section 4.

3.1. Homogeneous networks

For the time being, we shall take the networks to be homogeneous. That is, all neurons are assumed to have the same input–output relation $g_i = g$, with i -independent gain parameters $\gamma_i = \gamma$. Moreover we shall take $R_i = C_i = 1$ in suitable units, this latter assumption implying no loss of generality, since the C_i do not enter our theory and the R_i can always be absorbed in the gain parameters γ_i . These homogeneity assumptions are mainly adopted here for convenience. Inhomogeneous networks will be dealt with later on in section 3.4. The input–output relation $g_i = g$ need not be specified until it comes to the evaluation of the fixed-point equations for the order parameters.

For networks with synaptic couplings given by (9), the free energy (7) may be written

$$f_N(\beta) = -(\beta N)^{-1} \log \int \prod_i d\tilde{\rho}(V_i) \exp\left(\frac{N\beta}{2} \sum_{\mu} m_{\mu}^2\right) \quad (10)$$

where we have introduced the overlaps

$$m_{\mu} = \frac{1}{N} \sum_{i=1}^N \xi_i^{\mu} V_i \quad (11)$$

and where the integrated inverse input–output relation G as well as a term correcting for the absence of self-interactions have been absorbed in the single-site measure

$$d\tilde{\rho}(V) = d\rho(V) \exp[-\alpha\beta V^2/2 - \beta\gamma^{-1}G(V)] \quad (12)$$

with $\alpha = p/N$. In the limit of extensively stored patterns ($\alpha > 0$) the free energy is evaluated by the replica method (Amit *et al* 1987). For states which have macroscopic

correlations with at most finitely many (s) of the $p = \alpha N$ stored patterns, one obtains in the replica-symmetric approximation,

$$f(\beta) = \frac{1}{2} \sum_{\nu=1}^s m_\nu^2 + \frac{\alpha}{2} \left\{ \beta^{-1} \log[1 - \beta(q_0 - q_1)] + (q_0 - q_1)\bar{r} + \beta(q_0 - q_1)r \right\} \\ - \beta^{-1} \left\langle \left\langle \log \int d\tilde{\rho}(V) \exp \left\{ \beta \left[\left(\sum_{\nu=1}^s m_\nu \xi^\nu + \sqrt{\alpha r} z \right) V \right. \right. \right. \right. \\ \left. \left. \left. + \frac{1}{2} \alpha \bar{r} V^2 \right] \right\} \right\rangle \right\rangle \quad (13)$$

the double angular brackets denoting a combined average over the finitely many ξ^ν with which the system is macroscopically correlated, and a Gaussian random variable z with zero mean and unit variance. In (13),

$$r = \frac{q_1}{[1 - \beta(q_0 - q_1)]^2} \quad \text{and} \quad \bar{r} = \frac{1}{1 - \beta(q_0 - q_1)} \quad (14)$$

and the m_ν , q_0 and q_1 must be chosen to satisfy the fixed-point equations

$$m_\nu = \langle \langle \xi^\nu [V]_{\xi,z} \rangle \rangle \quad (15a)$$

$$q_0 = \langle \langle [V^2]_{\xi,z} \rangle \rangle \quad (15b)$$

$$q_1 = \langle \langle [V]_{\xi,z}^2 \rangle \rangle \quad (15c)$$

q_0 and q_1 denoting diagonal and off-diagonal elements of the matrix of Edwards–Anderson order parameters

$$q_{\sigma\sigma'} = \frac{1}{N} \sum_{i=1}^N V_i^\sigma V_i^{\sigma'} \quad 1 \leq \sigma \leq \sigma' \leq n$$

respectively. In (15), $[\dots]$ denotes the ‘thermal’ average

$$[\dots]_{\xi,z} = \frac{\int d\tilde{\rho}(V) (\dots) \exp [\beta(\sum_\nu m_\nu \xi^\nu + \sqrt{\alpha r} z) V + \frac{1}{2} \alpha \beta \bar{r} V^2]}{\int d\tilde{\rho}(V) \exp [\beta(\sum_\nu m_\nu \xi^\nu + \sqrt{\alpha r} z) V + \frac{1}{2} \alpha \beta \bar{r} V^2]} \\ = \frac{\int d\rho(V) (\dots) \exp[-\beta H(V)]}{\int d\rho(V) \exp[-\beta H(V)]} \quad (16)$$

where $\xi = (\xi^\nu)_{\nu=1}^s$, and where, using (12), we have introduced

$$H(V) = \gamma^{-1} G(V) - \left(\sum_\nu m_\nu \xi^\nu + \sqrt{\alpha r} z \right) V - \frac{\alpha}{2} (\bar{r} - 1) V^2. \quad (17)$$

In order to yield information about the nature of the local and global minima of \mathcal{H}_N , the averages (16) are to be evaluated in the deterministic $\beta \rightarrow \infty$ limit. In this limit we get (Kühn 1990, Kühn *et al* 1991)

$$[F(V)]_{\xi,z} = F(\hat{V}(\xi, z)) \quad (18)$$

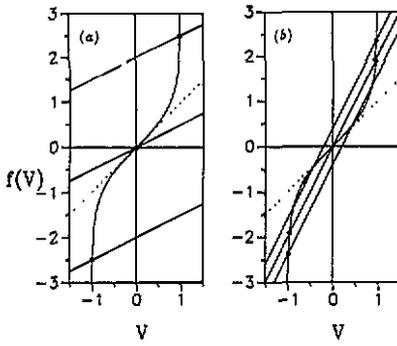


Figure 2. Solution of (19). Both, (a) and (b) show the intersection of the curve $g^{-1}(V)$ with lines of the form $\gamma(\tau V + b)$, each for three values of $b = b(z) = \sum_{\nu} m_{\nu} \xi^{\nu} + \sqrt{\alpha r} z$. Stable solutions corresponding to minima of $H(V)$ are marked by a point. (a) For $\gamma\tau < 1$, the relevant solution \hat{V} is continuous as b passes through zero. (b) For $\gamma\tau > 1$, the relevant solution jumps from $\hat{V} < 0$ to $\hat{V} > 0$ as b increases through zero, entailing a corresponding jump in the local field $\hat{U} = (1/\gamma)g^{-1}(\hat{V})$ as discussed below.

for any continuous function F , where $\hat{V}(\xi, z)$ is the point minimizing $H(V)$. It must be determined among the solutions of the transcendental fixed-point equation

$$\hat{V} = \hat{V}(\xi, z) = g\left(\gamma\left[\sum_{\nu} m_{\nu} \xi^{\nu} + \sqrt{\alpha r} z + \tau \hat{V}\right]\right) \quad (19)$$

with

$$\tau = \alpha(\bar{r} - 1) \quad (20)$$

on the support of $d\rho$; for an illustration, see figure 2.

For the purpose of a numerical solution of the fixed-point equations (15), it is advantageous to rewrite them in terms of the variables m_{ν} , q_1 and $C \equiv \beta(q_0 - q_1)$, so as to get (as $\beta \rightarrow \infty$)

$$m_{\nu} = \left\langle\left\langle \xi^{\nu} \hat{V}(\xi, z) \right\rangle\right\rangle \quad C = \frac{1}{\sqrt{\alpha r}} \left\langle\left\langle z \hat{V}(\xi, z) \right\rangle\right\rangle \quad q_1 = \left\langle\left\langle \hat{V}^2(\xi, z) \right\rangle\right\rangle \quad (21)$$

$\hat{V} = \hat{V}(\xi, z)$ being determined by (19) as before, and $r = q_1/(1 - C)^2$, while $\bar{r} = 1/(1 - C)$.

Note that up to this point, the theory could be developed in *complete generality* with respect to input-output relations. On a formal level this generality is possible, because input-output relations affect *only* single-site measures in (10)–(15), and not terms related to synaptic interactions.

The limit of finitely many stored patterns is recovered by taking the limit $\alpha \rightarrow 0$ in (13)–(19). In this limit, the m_{ν} alone are sufficient to describe the state of the system, and must be chosen to satisfy (15a). The solution \hat{V} of (19) can be determined explicitly in this case to yield $\hat{V} = g(\gamma \sum_{\nu} \xi^{\nu} m_{\nu})$, so that the $\beta \rightarrow \infty$ limit of (15a) takes the form $m_{\nu} = \langle \xi^{\nu} g(\gamma \sum_{\nu} \xi^{\nu} m_{\nu}) \rangle$, the single angular brackets denoting an average over the ξ^{ν} according to their distribution. These equations bear a strong formal similarity to those describing the stochastic Hopfield model, with the hyperbolic tangent replaced by a general input-output relation, and inverse temperature β by the gain parameter γ . Needless to say, we get a true formal equivalence for the choice $g(x) = \tanh(x)$. More generally, for input-output relations of sigmoidal shape, with $\lim_{x \rightarrow \pm\infty} g(x) = \pm 1$ and $g'(x) \leq g'(0) = 1$, the 'paramagnetic' null solution will lose local stability as γ is increased above $\gamma_c = 1$. The 'phase transition' at γ_c may be first or second order, tri- or higher order critical, depending on the shape of g in the vicinity of $x = 0$ (Kühn 1990, Kühn *et al* 1991).

3.2. Local field distributions

Before turning to a discussion of the phase structure in the limit of extensively many stored patterns ($\alpha > 0$), let us draw the readers' attention to the fact that the mean-field theory presented in the previous subsection provides more than just a tool for identifying stable stationary states of the network dynamics and for characterizing them macroscopically in terms of order parameters. To see this, one notes that the order parameters constitute, in fact, a *parametrization* of the local field distributions pertaining to the various types of attractor. In the present subsection we would like to demonstrate, in the case of the soft-neuron version of the Hopfield model, how such parametrizations can be *unfolded*, so as to obtain explicit analytic expressions for these local field distributions.

There are at least two reasons why this should be interesting. First and foremost, a local field \hat{U} translates, via the input-output relation $\hat{V} = g(\gamma\hat{U})$, into a neural output level (firing rate) \hat{V} , so that local field distributions can immediately be translated into firing-rate distributions. In a biological context it should be noted that these distributions are directly accessible to experimental techniques of the neurophysiologist—in contrast to the order parameters m_ν , C , and q_1 or, for that matter, storage capacities. Thus, local field distributions and the corresponding firing rate distributions appear to be far better candidates for providing feedback between experimental evidence and theoretical modelling than order parameters. Second, as we shall see, the computation of local field distributions can be used to obtain significant speed-ups in the numerical solution of the fixed-point equations (19)–(21).

Our computation of local field distributions starts out from (19). Local field \hat{U} and firing rate \hat{V} being related through $\hat{V} = g(\gamma\hat{U})$, one can rewrite (19) in terms of \hat{U} to read

$$\hat{U} = \sum_{\nu} m_{\nu} \xi^{\nu} + \sqrt{\alpha r} z + \tau g(\gamma \hat{U}) \tag{19'}$$

where, as before, in case of several solutions one has to choose the one minimizing $H(V) = H(g(\gamma U))$.

From this observation one obtains the \hat{U} -distribution for given $\xi = (\xi^{\nu})_{\nu=1}^s$ as follows. Instead of determining $\hat{U}(z)$ for given z through (19'), one proceeds the other way round and defines $z = z(\hat{U})$ through

$$z = z(\hat{U}) = \frac{1}{\sqrt{\alpha r}} \left(\hat{U} - \sum_{\nu} \xi^{\nu} m_{\nu} - \tau g(\gamma \hat{U}) \right). \tag{22}$$

In cases where (19') has multiple solutions, several different \hat{U} will give rise to the same z in which case one has to choose that representation which minimizes $H(g(\gamma\hat{U}))$. The \hat{U} chosen to represent $z(\hat{U})$ in (22) may occasionally jump, namely when two minima of $H(g(\gamma\hat{U}))$, i.e. two solutions of (19') exchange their relative depth—much as order parameters do at first-order phase transitions. Locally, however, $\hat{U}(z)$ is smooth and invertible, implying that $z = z(\hat{U})$ is continuous and differentiable. Hence, knowledge of

$$\mathcal{P}(z)dz = \frac{dz}{\sqrt{2\pi}} \exp(-z^2/2)$$

allows to obtain

$$\mathcal{P}_\xi(\hat{U}) \equiv \mathcal{P}(z(\hat{U})) \frac{dz}{d\hat{U}} = \frac{1}{\sqrt{2\pi\alpha r}} \exp \left\{ -\frac{1}{2\alpha r} \left(\hat{U} - \sum_\nu \xi^\nu m_\nu - \tau g(\gamma \hat{U}) \right)^2 \right\} \\ \times (1 - \gamma \tau g'(\gamma \hat{U})) \chi(\hat{U}) \quad (23)$$

where $\chi(\hat{U}) = 1$ for those \hat{U} , for which $\hat{U}(z)$ is locally smooth and invertible, whereas $\chi(\hat{U}) = 0$ on those intervals across which the solution $\hat{U}(z)$ to (19') jumps due to the minimality criterion for $H(g(\gamma \hat{U}))$. Note that $\gamma \tau g'(\gamma \hat{U}) < 1$ for a solution to (19') that represents a (local) minimum of $H(g(\gamma \hat{U}))$, so that $\mathcal{P}_\xi(\hat{U})$ is indeed non-negative. The situation is particularly transparent for input-output relations g which are of the typical sigmoid form, i.e. which are odd, and convex for positive arguments, thus having maximum slope $g'(x)$ at $x = 0$. For such g , one has to switch between a ($\hat{U} \leq 0$) representation of $z(\hat{U})$ to a ($\hat{U} \geq 0$) representation at $z = -\sum_\nu \xi^\nu m_\nu / \sqrt{\alpha r}$, where \hat{U} solves

$$\hat{U} = \tau g(\gamma \hat{U}) \quad (24)$$

(see figure 2). Thus there will be a jump in the solution and a corresponding gap in $\mathcal{P}_\xi(\hat{U})$ if $\gamma \tau > 1$, i.e. at sufficiently high gain or loading level. That is, χ is the characteristic function of the set $\mathbb{R} \setminus [-\hat{U}_0, \hat{U}_0]$, formally $\chi(\hat{U}) = \chi_{\mathbb{R} \setminus [-\hat{U}_0, \hat{U}_0]}(\hat{U})$, with \hat{U}_0 being the positive solution of (24).

In the case where the input-output relation has the multi-step structure of a Q -state neuron (Rieger 1990, Mertens *et al* 1991, Bollé *et al* 1991), the field distribution may have one or several gaps, depending on the 'average slope' of g near $x = 0$. This may give rise to an intricate structure of co-existing metastable retrieval phases (Bös 1992), a phenomenon which we will study in greater detail in a forthcoming paper (Bös and Kühn 1992).

Gaps or no gaps, (23) clearly shows that local field distributions deviate from the Gaussian for networks of stochastic two-state neurons in the context of the replica approach, and that their precise form will also depend on details of the neural gain function.

Note that knowledge of $\mathcal{P}_\xi(\hat{U})$ speeds up the numerical solution of the fixed-point equations considerably, since it allows us to rewrite them as

$$m_\nu = \langle\langle \xi^\nu g(\gamma \hat{U}) \rangle\rangle \quad C = \frac{1}{\sqrt{\alpha r}} \langle\langle z(\hat{U}) g(\gamma \hat{U}) \rangle\rangle \quad q_1 = \langle\langle g(\gamma \hat{U})^2 \rangle\rangle \quad (21')$$

where the double angular brackets now denote a combined average over the ξ and $\mathcal{P}_\xi(\hat{U})$, the latter being explicitly known once the gaps—if any—have been determined from (24). By computing local field distributions, we have thus been able to circumvent the problem of solving fixed-point equations within fixed-point equations as in the formulation (19)–(21). We now turn to results.

3.3. Results for homogeneous networks

For input-output relations g with $\lim_{x \rightarrow \pm\infty} g(x) = \pm 1$, and $g'(x) \leq g'(0) = 1$, the topology of the phase diagram is similar to that of the stochastic model, inverse gain playing a role similar to that of temperature (see figure 3). This is not surprising, since

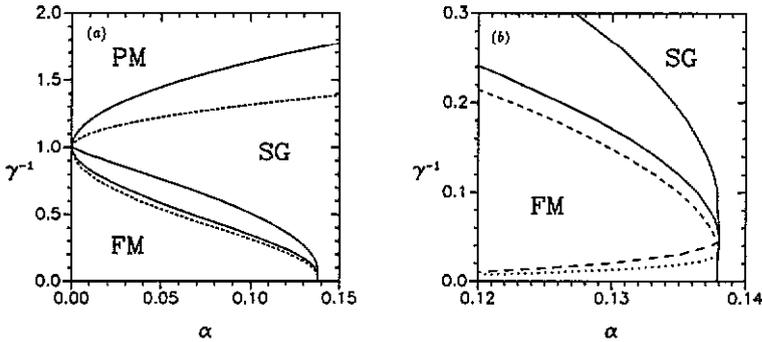


Figure 3. Phase diagram for soft-neuron versions of the Hopfield model. PM, SG, and FM denote the paramagnetic (null-), spin-glass, and retrieval phases, respectively. Results for the stochastic Ising network are shown as broken curves (upon identification of inverse gain and temperature scales) for comparison. Retrieval phase boundaries are shown for neurons with hyperbolic tangent ($g(x) = \tanh(x)$; lower full curve) and piecewise linear ($g(x) = \text{sgn}(x) \min(|x|, 1)$; upper full curve) response. (b) Enlarged portion of the phase diagram of (a). It also shows the ‘gap-line’ (dotted) below which the local field distribution for the retrieval phase has a gap at small \hat{U} , and the AT-line (long-dashed), both for neurons with hyperbolic tangent response.

finite gain, like non-zero temperature, reduces the average response of each neuron to its local field. Unlike finite gain, however, non-zero temperature also induces fluctuations around the average response which are not present in deterministic finite-gain systems. This explains why in the \tanh -soft-neuron case, phases of spontaneously broken symmetry extend to higher values of inverse gain than in the case of stochastic Ising-neurons when inverse-gain and temperature scales are identified. In the limit of finitely many stored patterns ($\alpha = 0$), on the other hand, fluctuations in the individual neurons’ response play no role and naive mean-field theory provides an appropriate description of the system. As a consequence, as noted above, at $\alpha = 0$ the system with deterministic \tanh -neurons is formally equivalent to the stochastic Ising model network if gain γ and inverse temperature β are identified.

Note that the instability of the ‘paramagnetic’ null solution with respect to spin-glass ordering occurring at $\gamma_g^{-1} = 1 + 2\sqrt{\alpha}$ and the maximum storage capacity $\alpha_c \simeq 0.138$ at $\gamma^{-1} = 0$ are *universal* properties of networks having normalized sigmoid input-output relation as defined above. As for the case of finitely many patterns, the *order* of the spin-glass transition at $\gamma_g(\alpha)$ will depend on properties of $g(x)$ in the vicinity of $x = 0$, and there will quite generally be hysteresis effects, if the transition is discontinuous.

In an enlarged portion of the phase diagram (figure 3(b)), we also exhibit the line, satisfying $\gamma\tau = 1$, that signifies the opening of a gap in the local field distribution for the retrieval phase. The other additional line signals an instability of the replica-symmetric retrieval solution in the direction of the replicon mode (de Almeida and Thouless 1978). This instability occurs when the replicon eigenvalue

$$\lambda_{\text{AT}} = \frac{\alpha\beta^2}{(1-C)^2} \left\langle\left\langle ([V^2] - [V]^2)^2 \right\rangle\right\rangle - 1 \tag{25}$$

becomes positive. Using the fact that

$$\beta ([V^2]_{\xi,z} - [V]_{\xi,z}^2) = \frac{1}{\sqrt{\alpha r}} \frac{d}{dz} [V]_{\xi,z}$$

we can conclude from (25) that replica symmetry is definitely broken in regions of parameter space where the local field distribution has a gap. To see this one notes that gaps in the local field distribution are in one-to-one correspondence with jumps of $\lim_{\beta \rightarrow \infty} [V]_{\xi,z} = \hat{V}(\xi, z)$, hence with δ -function singularities of its z -derivative.

While local field distributions as computed in the previous section do deviate from the Gaussian form typically obtained for networks of stochastic neurons in the context of the replica method, we find this effect to be small in regions of parameter space where replica symmetry is unbroken, and barely visible in graphical representations. The effect appears to be stronger for the soft-neuron version of the SK model (Bös 1992).

3.4. Network inhomogeneities and self-couplings

In the considerations of the previous sections, we have restricted our investigations to systems which are *homogeneous* in the sense that all neurons behave identically. The statistical mechanics approach, however, also applies to inhomogeneous systems in which neuronal gain functions $V_i = g(\gamma U_i)$ vary from neuron to neuron in that either gain *parameters* γ_i turn out to be i -dependent while the functional form of the neuronal gain functions remains the same for all neurons, or in that gain functions themselves belong to different classes. Moreover, self-couplings J_{ii} which may, but need not, vary with i are also covered by the general statistical mechanics approach described in section 2. Inhomogeneities of this type only create an additional element of on-site disorder, i.e. disorder of the same type as that embodied in the random patterns stored in the net, creating no additional problems of principle.

The only change required is to replace the single-site measure $d\bar{\rho}(V)$ introduced in (12) by the i -dependent measure

$$d\bar{\rho}_i(V) = d\rho(V) \exp[\beta(J_{ii} - \alpha)V^2/2 - \beta\gamma_i^{-1}G_i(V)] \quad (26)$$

in which G_i is the integrated inverse input-output relation of a neuron with gain function g_i , and γ_i and J_{ii} denote (possibly i -dependent) gain parameters and self-couplings. Assuming that the J_{ii} , the γ_i and the g_i are randomly and independently selected according to some given distribution, one can show that—formally—(13)–(21) which describe the collective behaviour of homogeneous networks remain valid in the inhomogeneous case, provided that double angular brackets are now understood as implying an *additional average* over the randomness contained in the measure $d\bar{\rho}_i(V)$. Finiteness of the family of possible input-output relations is sufficient for this self-averaging result to hold.

Let us now proceed to see how network inhomogeneities and self-couplings (random or not) will affect the overall network performance. For the sake of definiteness, here we will restrict our attention mostly to results which are universal across the class of standard sigmoid input-output relations g introduced earlier.

The simplest case to consider is a non-random self-interaction $J_{ii} = J_0$. Qualitatively it is clear that a positive J_0 will favour large values of $|V|$, whereas negative J_0 will tend to favour attractors having small values of $|V|$. In the case of finitely many stored patterns ($\alpha = 0$), a small- m_ν expansion of (15a) reveals that the local instability of the paramagnetic solution against the formation of phases with

non-zero m_ν , which for $J_0 = 0$ occurred at $\gamma_c = 1$, is now shifted to higher or lower values of γ , depending on the sign of J_0 . One finds

$$\gamma_c(J_0) = \frac{1}{1 + J_0} \quad J_0 > -1 \tag{27}$$

and a complete suppression of the retrieval phase for $J_0 \leq -1$. Further detailed investigation of the small- m_ν expansion shows that variations in J_0 do not alter the order of the phase transition at γ_c .

By a small q_1 expansion of (19) at extensive levels of loading one finds that the local instability of the paramagnetic solution with respect to the formation of spin-glass-type ordering now occurs at

$$\gamma_g = (1 + J_0 + 2\sqrt{\alpha})^{-1} \quad J_0 > -1 - 2\sqrt{\alpha} \tag{28}$$

with the paramagnetic solution extending down to $\gamma^{-1} = 0$ for $\alpha \leq (\frac{1}{2}|1 + J_0|)^2$ when $1 + J_0 < 0$.

For non-negative J_0 , the critical storage level at $\gamma^{-1} = 0$ remains at $\alpha_c \simeq 0.138$, irrespective of the value of $J_0 > 0$. Details of the retrieval phase boundary beyond the points ($\alpha = 0, \gamma = (1 + J_0)^{-1}$) and, for $J_0 > 0, (\alpha \simeq 0.138, \gamma^{-1} = 0)$ depend on details of the input-output relation.

Figure 4 shows how phase boundaries are affected by non-random self-couplings for a network with piecewise linear response. Apart from fine details, non-zero self-interactions appear to result in an overall parallel displacement of phase boundaries.

Next, let us consider networks of neurons of all which have the same input-output relation, but randomly varying gain parameters γ_i . A small- m_ν expansion of (15a) at $\alpha = 0$ shows that the instability of the paramagnetic null phase with respect to the formation of a retrieval phase is now controlled by the average gain parameter, the critical condition being $\langle \gamma \rangle_c = 1$. In case of a second-order transition (i.e. for $g'''(0) < 0$) one has

$$m_\nu \sim \left\{ \frac{3!(\langle \gamma \rangle - 1)}{\langle \gamma^3 \rangle (-g'''(0))} \right\}^{1/2} \quad \langle \gamma \rangle \searrow \langle \gamma \rangle_c \tag{29}$$

i.e. the third moment $\langle \gamma^3 \rangle$ affects the amplitude of the retrieval overlap near $\langle \gamma \rangle_c$.

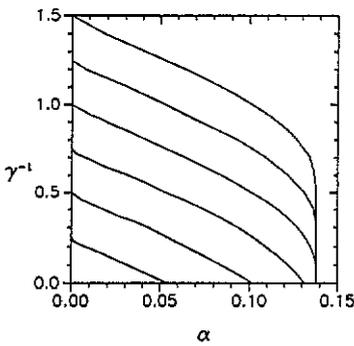


Figure 4. Phase diagram for a soft-neuron Hopfield model with non-zero self-couplings J_0 and piecewise linear response. Shown are retrieval phase boundaries for $J_0 = 0.5, 0.25, 0.0, -0.25, -0.5,$ and -0.75 (top to bottom).

In a similar vein, the condition for the local instability of the null-phase against formation of spin-glass-type ordering is given by the conditions

$$1 = \left\langle \frac{\gamma^2}{(1 - \gamma\tau)^2} \right\rangle_c \frac{\alpha}{(1 - C)^2} \quad C = \left\langle \frac{\gamma}{1 - \gamma\tau} \right\rangle_c \quad (30)$$

which can in general no longer be evaluated analytically. In (29)–(30), angular brackets denote averages over the distribution of gain parameters; the subscript *c* refers to critical conditions.

4. Other forms of synaptic organization

Having discussed the storage and retrieval of unbiased binary random patterns in networks with Hebb–Hopfield couplings (9) in some detail, we now turn to the investigation of other storage tasks and other forms of synaptic organization. First, we briefly address the storage of low-activity patterns in networks of analogue neurons. In particular, we demonstrate that low-activity networks—operating at low local firing rates if gain functions are chosen such as to reflect neural refractoriness (Amit and Tsodyks 1991a, b, Kühn 1991)—saturate the Gardner bound for the storage of low-activity patterns (Gardner 1988) by order of magnitude. The second topic of the present section is devoted to the statistical mechanics of analogue neurons coupled via pseudo-inverse synapses (Kohonen 1984, Personnaz *et al* 1985, Kanter and Sompolinsky 1987).

A third possible topic, namely the problem of storing multi-state (grey-toned) patterns in networks of multi-state neurons, is in principle covered by the structure of the general theory presented in the previous sections. However, discussing the intricate features emerging in such systems would be beyond the scope of the present article, so we leave it to a separate publication (Bös 1992, Bös and Kühn 1992).

4.1. Covariance rules and low-rates systems

In this first subsection, we turn to the problem of storing *biased* binary random patterns in networks of analogue neurons. The formal structure of the theory outlined in the previous section can be taken over nearly word for word, if we adopt the convention $\eta_i^\mu \in \{0, 1\}$ for the pattern bits to be stored, with

$$\text{Prob}\{\eta_i^\mu = 1\} = a \quad (31)$$

and if we take the storage prescription to be of the form

$$J_{ij} = \frac{1}{N\Delta_0} \sum_{\mu=1}^p (\eta_i^\mu - a)(\eta_j^\mu - a) \quad i \neq j \quad (32)$$

with a normalization constant $\Delta_0 = a(1 - a)$ chosen such as to fix the J_{ij} -scale in an a -independent manner.

This convention makes (32) a so-called covariance learning rule (Tsodyks and Feigel'man 1988, Buhmann *et al* 1990) which is well known to be an efficient learning rule saturating the Gardner bound $\alpha_c \sim (2a \ln(1/a))^{-1}$ for the storage of low-activity patterns with $a \ll 1$, if combined with a 0-1 representation of neural activities and appropriate thresholds; see also Horner (1989) and Perez-Vicente and Amit (1989).

The problem of storing and retrieving low-activity patterns in networks of analogue neurons operating at low firing rates was recently and independently addressed by Kühn (1991) and by Amit and Tsodyks (1991a, b), with basically identical conclusions: (i) retrieval of low-activity patterns at low firing rates *is* possible even at extensive levels of loading; (ii) low firing rates can be understood as a natural consequence of neural refractoriness.

Here we do not wish to repeat in detail the arguments leading to these results. Rather, we wish to address the question of how low firing rates affect network efficiency as measured, e.g., by the storage capacity. Our main result, which appears not to have been reported before, is that networks of analogue neurons retrieving at reduced firing rates continue to saturate the Gardner bound for the storage of low-activity patterns by order of magnitude, albeit with a systematic depression proportional to the *square* of the reduced average firing rate (on a scale on which $V_{\max} = 1$).

Before deriving our main result, let us briefly present an argument showing that gain functions giving rise to a mode of operation at low firing rates do, in fact, emerge naturally from models of neural refractoriness. A simple version is the following: a neuron emitting a spike, say, at time $t = 0$, cannot generate the next spike before the end of the absolute refractory period at time t_0 . At later times, spike emission is possible again but requires at first elevated, later on progressively lower PSPs. All this can be modelled by a time-dependent effective threshold $\vartheta(t)$ which is infinite during the absolute refractory period and which decreases asymptotically to its resting value ϑ according to $\vartheta(t) = \vartheta + f_\vartheta(t - t_0)$, with some function f_ϑ decreasing from ∞ to 0, as t goes from t_0 to ∞ . Equality of externally induced PSP U and $\vartheta(t)$ at $t = t^*$ defines the time t^* of the next spike emission. This gives $V = g(U) = 1/t^*(U)$ for the gain function, with $g(U) = 0$ for $U < \vartheta$, and, putting $V_{\max} = 1/t_0 = 1$, $g(U) = 1/(1 + f_\vartheta^{-1}(U - \vartheta))$ for $U \geq \vartheta$, rising asymptotically to $V_{\max} = 1$ as $U \rightarrow \infty$. The detailed shape of g will depend on the shape of f_ϑ , i.e. on details of the decay of the effective threshold to its resting value.

A more sophisticated and detailed variant of the preceding argument has been advanced by Amit and Tsodyks (1991a). For our purposes, however, the details are of no prime importance. Rather, it should be noted that—independently of fine details—gain functions typically emerging from such a description of neural refractoriness do under normal circumstances ($U < \infty$) give rise to firing rates *below* maximum.

Turning to a description of the collective behaviour of such systems, we introduce, as usual, a gain parameter γ to vary U -scales. Scaling the *deviation* of the PSP from threshold ϑ , we write our input-output relation in the form

$$V = g(\gamma(U - \vartheta)) \tag{33}$$

and we have $g(x) = 0$ for $x \leq 0$, while $g(x) > 0$ for $x > 0$, rising asymptotically to 1, as $x \rightarrow \infty$. This convention modifies $G(V)$ in (3), (4) to

$$G_i(V) = \int^V g_i^{-1}(V') dV' + \gamma \vartheta V. \tag{34}$$

The averaging of the free energy over the uncondensed patterns is not significantly affected by the more complicated statistics of the η_i^μ , since only the covariance $\langle (\eta_i^\mu - a)(\eta_j^\nu - a) \rangle = \delta_{ij} \delta_{\mu\nu} \Delta_0$ matters. The normalization of the J_{ij} in (32) is chosen precisely to compensate its effect. Here we only reproduce the fixed-point

equations that determine the phase structure of the theory (Kühn 1991, Amit and Tsodyks 1991b)

$$\begin{aligned}
 m_\nu &= \left\langle \left\langle (\eta_i^\nu - a) / \Delta_0 \hat{V}(\beta, z) \right\rangle \right\rangle & C &= \frac{1}{\sqrt{\alpha r}} \left\langle \left\langle z \hat{V}(\beta, z) \right\rangle \right\rangle \\
 q_1 &= \left\langle \left\langle \hat{V}^2(\beta, z) \right\rangle \right\rangle.
 \end{aligned}
 \tag{35}$$

Because of the explicit appearance of a threshold in (33) and (34), the $\hat{V} = \hat{V}(\beta, z)$ are now determined from the solutions of a slightly modified version of (19), namely

$$\hat{V} = \hat{V}(\beta, z) = g \left(\gamma \left[\sum_\nu m_\nu (\eta^\nu - a) + \sqrt{\alpha r} z + \tau \hat{V} - \vartheta \right] \right). \tag{36}$$

Asymptotic analysis of these equations in the limit $a \rightarrow 0$ reveals that the storage capacity of the low-activity low-rates system obeys

$$\alpha_c \sim \frac{\bar{V}^2}{2a \ln(1/a)} \quad \text{as } a \rightarrow 0 \tag{37}$$

where $\bar{V} = \mathcal{O}(1)$ is the (low) average firing rate of those neurons that should be firing in one of the low-activity retrieval states. The main line of reasoning leading to this result is as follows. Anticipating that $\alpha_c \sim (a \ln(1/a))^{-1}$ as $a \rightarrow 0$, we assume that $\alpha \simeq (xa \ln(1/a))^{-1}$ and try to determine the smallest x for which the fixed-point equations have retrieval solutions. The requirement that in retrieval solutions the neurons on $\eta^\nu = 0$ sites should be ‘off’ with sufficiently high probability to guarantee $q_1 = \mathcal{O}(a)$ relates x with ϑ . We find $x \geq 2/\vartheta^2$. This condition can be shown to be consistent with retrieval—which, of course, requires $(1 - a)m_\nu > \vartheta$ and $-am_\nu - \vartheta < 0$ for $m_\nu = \mathcal{O}(1)$ —provided ϑ is suitably chosen. Retrieval at low firing rates is, in addition, characterized by $m_\nu \simeq \bar{V}$, where the actual value of the mean firing rate \bar{V} at ‘on’-sites will depend on details of the gain function. For $a \ll 1$, this implies the bound $\vartheta \leq \bar{V}$ which, finally, gives $x \geq 2/\bar{V}^2$, and hence proves (37).

In summary, networks retrieving at low firing rates saturate the Gardner bound for the storage of low-activity patterns by order of magnitude. Low firing rates do, however, give rise to a systematic depression of α_c relative to standard low-activity systems due to the factor \bar{V}^2 appearing in (37).

4.2. Pseudo-inverse couplings

The second topic of the present section is devoted to the statistical mechanics of graded-response neurons coupled via pseudo-inverse synapses (Kohonen 1984, Personnaz *et al* 1985, Kanter and Sompolinsky 1987). That is, we assume that the couplings between neurons are given by

$$J_{ij} = \frac{1}{N} \sum_{\mu, \nu=1}^p \xi_i^\mu (C^{-1})_{\mu, \nu} \xi_j^\nu \quad i \neq j \quad J_{ii} = J_0 \tag{38}$$

where C is the correlation matrix of the stored patterns with elements

$$C_{\mu\nu} = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \xi_i^\nu .$$

As noted in section 2, self-couplings in analogue-neuron systems with continuous time dynamics enjoy a different status to those in systems of binary neurons with asynchronous dynamics. In the latter, self-couplings must be absent from the dynamical rule in order that a statistical-mechanical approach be applicable; in the case of Ising neurons, they may of course be included in the Lyapounov function, since there they only give a constant, i.e. $\{S_i\}$ -independent, contribution to the energy. In analogue-neuron systems, on the other hand, self-couplings may appear in the dynamical rule, and they must be included in the Hamiltonian as they appear in the dynamics.

Our analysis of the collective behaviour of the system again uses the general ideas outlined in section 2. In the details, we follow Kanter and Sompolinsky (1987), with modifications as usual to account for the continuous nature of the V_i , the most prominent difference being non-trivial diagonal entries in the matrix of Edwards-Anderson order parameters, as we have already encountered in the cases studied above.

Being interested in pattern retrieval capabilities, we assume that the system has macroscopic correlations with at most one of the stored patterns, say ξ^1 , and obtain the following expression for the free energy in the replica-symmetric approximation:

$$\begin{aligned} -\beta f(\beta) = & -\frac{\beta}{2} \frac{m^2}{1+x_0} - \frac{1}{2} \left[\ln(1+x_0) + \frac{x}{1+x_0} \right] + \frac{\alpha}{2} \frac{x}{x_0} \\ & - \frac{\alpha}{2} \ln \frac{\alpha}{x_0} - \frac{1}{2} (1-\alpha) \ln(1-\alpha) + \left\langle \left\langle \log \int d\tilde{\rho}(V) \right. \right. \\ & \left. \left. \times \exp \left\{ \beta \left[\left(\frac{m}{1+x_0} \xi^1 + \sqrt{\alpha r z} \right) V + \frac{1}{2} \frac{x_0}{1+x_0} V^2 \right] \right\} \right\rangle \right\rangle \end{aligned} \quad (39)$$

where m , q_0 , q_1 , x , and x_0 must solve the following set of fixed-point equations:

$$m = \left\langle \left\langle \xi^1 [V]_{\xi^1, z} \right\rangle \right\rangle \quad (40a)$$

$$q_0 = \left\langle \left\langle [V^2]_{\xi^1, z} \right\rangle \right\rangle \quad (40b)$$

$$q_1 = \left\langle \left\langle [V]_{\xi^1, z}^2 \right\rangle \right\rangle \quad (40c)$$

$$x = \frac{\beta(q_1 - m^2)x_0}{[1 - 2\alpha - C + 2x_0(1 - \alpha)]} \quad (40d)$$

$$1 + x_0 = \frac{1}{2(1 - \alpha)} \left\{ 1 + C + [(1 - C)^2 + 4C\alpha]^{1/2} \right\} . \quad (40e)$$

Here $C = \beta(q_0 - q_1)$, $r = (1/\alpha\beta)[x/(1+x_0)^2]$, and

$$[\dots]_{\xi^1, z} = \frac{\int d\rho(V)(\dots) \exp[-\beta H(V)]}{\int d\rho(V) \exp[-\beta H(V)]} \quad (41a)$$

with

$$H(V) = \gamma^{-1}G(V) - \left(\frac{m}{1+x_0}\xi^1 + \sqrt{\alpha r}z \right) V - \frac{1}{2} \left(\frac{x_0}{1+x_0} + J_0 - \alpha \right) V^2. \quad (41b)$$

Note that (40d, e) are as in Kanter and Sompolinsky (1987). As usual, we have to take the deterministic $\beta \rightarrow \infty$ limit, in order to obtain information about the collective properties of the net. In this limit, (40a-c) simplify to

$$m = \langle\langle \xi^1 \hat{V}(\xi^1, z) \rangle\rangle \quad C = \frac{1}{\sqrt{\alpha r}} \langle\langle z \hat{V}(\xi^1, z) \rangle\rangle \quad q_1 = \langle\langle \hat{V}^2(\xi^1, z) \rangle\rangle \quad (42)$$

where $\hat{V} = \hat{V}(\xi^1, z)$ is determined from

$$\hat{V} = \hat{V}(\xi^1, z) = g \left(\gamma \left[\frac{m}{1+x_0}\xi^1 + \sqrt{\alpha r}z + \left(\frac{x_0}{1+x_0} + J_0 - \alpha \right) \hat{V} \right] \right). \quad (43)$$

Alternatively, (42)–(43) may be reformulated in terms of distributions of local fields, as outlined in section 3.2.

As in the case of Ising neurons coupled by pseudo-inverse synapses, the fixed-point equations describing the collective properties of the network have stable retrieval solutions with $r \rightarrow 0$. In this limit, we find that the fixed-point equation for the retrieval overlap m decouples from the other equations:

$$m = g(\gamma(1 + J_0 - \alpha)m) \quad (44)$$

and we have $q_0 = q_1 = m^2$. Equation (44) gives a retrieval phase boundary in the γ - α -plane at $\gamma(1 + J_0 - \alpha) = 1$, $\alpha \leq 1$. That is, for $J_0 > 0$ we have a retrieval phase for sufficiently large γ up to the theoretically possible maximum $\alpha_c = 1$. If, on the other hand, $J_0 < 0$, then the retrieval phase exists only up to $\alpha_c = 1 - |J_0|$, the retrieval phase being completely suppressed for $J_0 < -1$, much as in the Hebb–Hopfield case.

5. Stochastic dynamics in continuous time

Up to this point, our investigation of the collective behaviour of analogue neuron systems has been entirely confined to networks with a continuous time dynamics given by the *deterministic* rule (1), (2), and our statistical-mechanical approach has implied a further restriction to cases where this dynamical rule is governed by a Lyapounov function.

The dynamics of biological neural networks as well as that of their electronic counterparts is, however, never completely free of stochasticity. Accordingly, the purpose of the present section is to see whether and how stochasticity can be incorporated into our investigation of collective behaviour of networks with dynamics formulated in continuous time. As before, we will not attempt to solve the full (stochastic) dynamical problem; rather our prime concern will be with *equilibrium* properties of networks with stochastic continuous time dynamics. Before embarking on the analysis, let us add a few more specific comments regarding its motivation.

In the context of neural modelling, it has been argued (Amit and Tsodyks 1991a) that the effects of noise on neural dynamics can be incorporated in the gain function, in the course of a reduction from a description of the dynamics in terms of spikes

to a description in terms of rates, so that (1), (2) would already represent the full story—including stochasticity, in which case adding fluctuating forces to (1), (2) might not make much sense. Recent results of Gerstner and van Hemmen (1992), on the other hand, indicate that such a reduction may not be valid in general. This being so, adding noise to (1), (2) appears to be one of the possibilities to model the effects of whatever fluctuations might have been swamped in the course of the spikes-to-rates reduction (or even not taken into account in the spikes description to begin with). Moreover, apart from this debate, it should be noted that most of what we are going to present below is *not* restricted to the large-system limit which has been instrumental in the arguments in favour of the possibility of a complete spikes-to-rates reduction.

In a wider context, we believe that what we are going to report here may be relevant to understanding the effects of noise in coupled nonlinear dynamical systems in general.

Guided by previous approaches, we investigate the possibility of stationary states, *assuming* that they satisfy a detailed balance condition (Peretto 1984). At least two ways of approaching this problem are possible, corresponding to formulations of the continuous-time dynamics *solely* in terms of neural potentials or *solely* in terms of neural firing rates. In both cases, a detailed balance condition unambiguously fixes conditions on both noise and synaptic organization: synapses must be ‘essentially’ symmetric, and the strength of the noise must be coupled to the dynamical processes in a manner depending on neural input–output relations. Under these conditions a remarkable reciprocity between potential dynamics and firing rate dynamics emerges: for the potential-dynamics, the invariant distribution is a Gibbs distribution over firing-rate space. For the noisy firing-rate dynamics, it turns out to be a Gibbs distribution over the space of neural potentials. Both distributions will be seen to exist only under unrealistic assumptions about the noise in the system.

Let us now turn to the details. The possibility to choose two different representations for the dynamics (1), (2) rests on the assumption that the neuronal input–output relation $V_i = g_i(\gamma_i U_i)$ is invertible, so that (1):

$$C_i \frac{dU_i}{dt} = \sum_{j=1}^N J_{ij} V_j - \frac{U_i}{R_i} + I_i = -\frac{\partial \mathcal{H}_N(\mathbf{V})}{\partial V_i} \quad (45)$$

can be formulated either entirely in terms of firing rates, namely

$$\frac{dV_i}{dt} = -\frac{\gamma_i}{C_i} g'_i \left(g_i^{-1}(V_i) \right) \frac{\partial \mathcal{H}_N(\mathbf{V})}{\partial V_i} \quad (46)$$

or entirely in terms of neural potentials,

$$\frac{dU_i}{dt} = -\frac{1}{C_i \gamma_i g'_i(\gamma_i U_i)} \frac{\partial \mathcal{H}_N(\mathbf{U})}{\partial U_i} \quad (47)$$

where $\mathcal{H}_N(\mathbf{U})$ is obtained from $\mathcal{H}_N(\mathbf{V})$ by replacing every V_i by $g_i(\gamma_i U_i)$. By adding stochastic forces to the right-hand sides of (46) and (47), we get a pair of stochastic differential equations,

$$\frac{dV_i}{dt} = -\frac{\gamma_i}{C_i} g'_i \left(g_i^{-1}(V_i) \right) \frac{\partial \mathcal{H}_N(\mathbf{V})}{\partial V_i} + \zeta_i^V(t) \quad (48)$$

and

$$\frac{dU_i}{dt} = -\frac{1}{C_i \gamma_i g'_i(\gamma_i U_i)} \frac{\partial \mathcal{H}_N(U)}{\partial U_i} + \zeta_i^U(t) \tag{49}$$

which *unlike* (46) and (47) describe physically non-equivalent situations. To see this, one notes that by using the transfer functions g_i to transform the V -representation (48) of the stochastic dynamics into a U -representation, one obtains stochastic differential equations for the U_i in which the noise is of a *multiplicative* rather than of an additive nature as in (49). The same is observed on transforming (49) into a V -representation. Neither (48) nor (49) are of the form of a stochastic gradient dynamics, giving rise to canonical Gibbs distributions generated by $\mathcal{H}_N(U)$ or $\mathcal{H}_N(V)$ if the noise were Gaussian white noise with V - or U -independent intensity given by $\langle \zeta_i^x(t) \zeta_j^x(t') \rangle = 2k_B T D_0 \delta(t - t')$, $x \in \{U, V\}$. We will, however, see below that under suitable conditions on the noise we shall nevertheless obtain Gibbs-type equilibrium distributions describing the stationary states of networks with stochastic continuous-time dynamics which satisfy detailed balance conditions, and we will encounter the reciprocity relation between potential dynamics and firing-rate dynamics announced earlier.

It turns out that to satisfy these conditions on the noise, we have to allow the stochastic forces in (48) and (49) to depend on the V_i and the U_i , respectively, so that, what *prima facie* appears to be purely additive noise, may in fact contain multiplicative contributions. This covers, in principle, a rather wide range of noise models. In what follows we will, however, restrict our attention to cases where the multiplicative contribution to the noise—if any—is local (in a sense to be specified below), and where the noise is uncorrelated (white) in space (i, j) and time (t). Our results and conclusions below will thus be fairly general, subject only to these three restrictive assumptions.

Let us consider the firing rate, i.e. the V -dynamics, first. We take the stochastic forces in (48) to be Gaussian white noise, with an intensity locally coupled to the dynamical process according to

$$\langle \zeta_i^V(t) \zeta_j^V(t') \rangle = \delta_{ij} D(V_i) \delta(t - t'). \tag{50}$$

The nature of the stationary distribution is studied in terms of the Fokker–Planck equation (Gardiner 1983) corresponding to (48), (50):

$$\frac{\partial \mathcal{P}(V, t)}{\partial t} = - \sum_i \frac{\partial}{\partial V_i} \left\{ K_i(V) \mathcal{P}(V, t) - \frac{1}{2} \frac{\partial}{\partial V_i} (D(V_i) \mathcal{P}(V, t)) \right\} \tag{51}$$

with a drift-term

$$K_i(V) = -\frac{\gamma_i}{C_i} g'_i \left(g_i^{-1}(V_i) \right) \frac{\partial \mathcal{H}_N(V)}{\partial V_i} \tag{52}$$

given by the non-fluctuating contribution to the force in (48), and a diffusive probability current determined by the statistics (50) of the stochastic forces. We are now searching for stationary distributions $\mathcal{P}(V)$ of (51) satisfying detailed balance, i.e. solutions for which the probability current itself—rather than just its divergence—is everywhere zero. For such solutions, the term in curly brackets in (51) must vanish separately for each i , implying

$$\frac{2}{D(V_i)} \left[K_i(V) - \frac{1}{2} \left(\frac{\partial}{\partial V_i} D(V_i) \right) \right] \mathcal{P}(V) = \frac{\partial}{\partial V_i} \mathcal{P}(V). \tag{53}$$

If we demand that $\mathcal{P}(\mathbf{V})$ be of Gibbs' canonical form,

$$\mathcal{P}(\mathbf{V}) = \mathcal{N} \exp(-\Phi(\mathbf{V})) \quad (54)$$

we obtain the following so-called potential condition (Graham and Haken 1971, Gardiner 1983) connecting $\Phi(\mathbf{V})$ with the drift and diffusion terms in (51):

$$\frac{2}{D(V_i)} K_i(\mathbf{V}) - \frac{\partial}{\partial V_i} \ln D(V_i) = -\frac{\partial}{\partial V_i} \Phi(\mathbf{V}) \quad 1 \leq i \leq N. \quad (55)$$

If, in addition, one demands that $\Phi(\mathbf{V})$ be twice continuously differentiable in the V_i , so that

$$\frac{\partial^2}{\partial V_i \partial V_j} \Phi(\mathbf{V}) = \frac{\partial^2}{\partial V_j \partial V_i} \Phi(\mathbf{V})$$

one obtains a criterion for solvability of (55), which *at the same time* fixes the strength of the noise in (50) to

$$D(V_i) = 2k_B T \frac{\gamma_i}{d_i C_i} g'_i \left(g_i^{-1}(V_i) \right). \quad (56)$$

Here the d_i are positive constants, $d_i > 0$, which must be chosen such that

$$\tilde{J}_{ij} := d_i J_{ij} = d_j J_{ji} = \tilde{J}_{ji} \quad (57)$$

for all pairs i, j . The condition that such a set of positive constants can indeed be found constitutes a *condition of solvability* on the J_{ij} , which we shall call 'essential synaptic symmetry' in view of (57). Obviously, if the J_{ij} are symmetric to begin with, then the d_i in (56) are i -independent and can be absorbed in the temperature parameter. Given that a set of d_i exists, (55) can be solved for $\Phi(\mathbf{V})$ to give

$$\Phi(\mathbf{V}) = \beta \tilde{\mathcal{H}}_N(\mathbf{V}) + \sum_i \ln D(V_i) \quad (58)$$

where

$$\tilde{\mathcal{H}}_N = -\frac{1}{2} \sum_{i,j=1}^N \tilde{J}_{ij} V_i V_j + \sum_{i=1}^N \frac{d_i}{\gamma_i R_i} G_i(V_i) - \sum_{i=1}^N d_i I_i V_i \quad (59)$$

is a slightly modified version of (3)—coinciding with (3) if the J_{ij} are symmetric to begin with. This gives

$$\mathcal{P}(\mathbf{V}) \prod_i dV_i = \mathcal{N} \exp(-\beta \tilde{\mathcal{H}}_N(\mathbf{V})) \prod_i \frac{dV_i}{g'_i \left(g_i^{-1}(V_i) \right)} \quad (60)$$

for the stationary distribution, or, using the input-output relation once more,

$$\mathcal{P}(\mathbf{V}) \prod_i dV_i = \mathcal{N} \exp(-\beta \tilde{\mathcal{H}}_N(\mathbf{U})) \prod_i dU_i \quad (61)$$

where we have absorbed further constants in the normalization factor. Thus, if the strength of the Gaussian white noise in (50) satisfies (56), then the firing-rate

dynamics has a stationary distribution satisfying detailed balance. This distribution is the Gibbs distribution over the space of neural potentials U_i generated by $\tilde{\mathcal{H}}_N(U)$.

Before turning to an interpretation of this result, let us briefly outline the analogous argument for the potential dynamics, i.e. the U -dynamics. As for the V -dynamics, we take the stochastic forces to be Gaussian white noise with an intensity coupled to the dynamical process via

$$\langle \zeta_i^U(t) \zeta_j^U(t') \rangle = \delta_{ij} D(U_i) \delta(t - t') \quad (62)$$

and search for a stationary solution of the corresponding Fokker-Planck equation

$$\frac{\partial \mathcal{P}(U, t)}{\partial t} = - \sum_i \frac{\partial}{\partial U_i} \left\{ K_i(U) \mathcal{P}(U, t) - \frac{1}{2} \frac{\partial}{\partial U_i} (D(U_i) \mathcal{P}(U, t)) \right\} \quad (63)$$

with

$$K_i(U) = - \frac{1}{C_i \gamma_i g'_i(\gamma_i U_i)} \frac{\partial \mathcal{H}_N(U)}{\partial U_i} \quad (64)$$

satisfying detailed balance. Writing

$$\mathcal{P}(U) = \mathcal{N} \exp(-\Phi(U)) \quad (65)$$

for this stationary distribution, we can formulate the detailed balance condition in terms of a potential condition as in (55),

$$\frac{2}{D(U_i)} K_i(U) - \frac{\partial}{\partial U_i} \ln D(U_i) = - \frac{\partial}{\partial U_i} \Phi(U) \quad 1 \leq i \leq N. \quad (66)$$

Demanding the same type of differentiability properties of $\Phi(U)$ as we had of $\Phi(V)$ before, we obtain a condition for solvability that fixes the strength of the noise in (62),

$$D(U_i) = \frac{2k_B T}{d_i C_i \gamma_i g'_i(\gamma_i U_i)}. \quad (67)$$

As for the firing-rate dynamics, the $d_i > 0$ must be chosen to satisfy (57), a condition that can be met if synapses are essentially symmetric. Under these conditions, we can solve (66) for $\Phi(U)$ to obtain

$$\Phi(U) = \beta \tilde{\mathcal{H}}_N(U) + \sum_i \ln D(U_i) \quad (68)$$

and from this, by making further use of the input-output relation, finally

$$\mathcal{P}(U) \prod_i dU_i = \mathcal{N} \exp(-\beta \tilde{\mathcal{H}}_N(V)) \prod_i dV_i \quad (69)$$

which is the formal counterpart of (61). That is, with a Gaussian white noise satisfying (67) in (62), the noisy potential dynamics has a stationary distribution satisfying

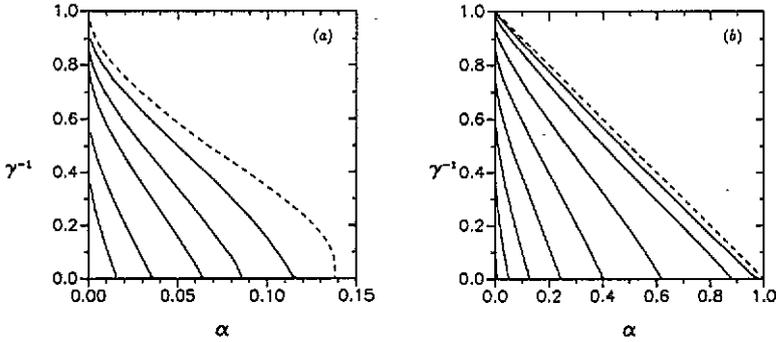


Figure 5. (a) Retrieval phase boundaries for a network with Hebb-Hopfield couplings and a stochastic U -dynamics in the α - γ -plane for various temperatures. From top to bottom, the curves correspond to $T = 0.00$ (broken), 0.05 , 0.10 , 0.15 , 0.20 , and 0.25 . The neural gain function is a hyperbolic tangent. (b) Retrieval phase boundaries for a network with Pseudo-inverse couplings and a stochastic U -dynamics in the α - γ -plane for various temperatures. Temperatures are $T = 0.000$ (broken), 0.001 , 0.01 , 0.05 , 0.10 , 0.15 , 0.20 , 0.25 , and 0.30 (from top to bottom). The neural response is taken to be a hyperbolic tangent.

detailed balance. This distribution is the Gibbs distribution over the space of neural outputs (firing rates) V_i generated by $\tilde{\mathcal{H}}_N(\mathbf{V})$.

In sections 3 and 4, we have evaluated free energies corresponding to (69) in cases where the synapses are symmetric, so that $\tilde{\mathcal{H}}_N(\mathbf{V}) = \mathcal{H}_N(\mathbf{V})$, with $\mathcal{H}_N(\mathbf{V})$ given by (3). There, we were only interested in the $\beta \rightarrow \infty$ limit of the theory. Free energies and fixed-point equations characterizing the phase structure of the models have, however, been derived, and can be evaluated for arbitrary β . At finite β , the collective behaviour thereby described is that of networks of analogue neurons with a noisy potential dynamics given by (49), (62).

With temperature we now have a third parameter, besides storage level α and gain parameter γ , which has its influence on network performance. In this enlarged parameter space we, too, have some degree of universality of network performance with respect to alterations of gain functions. For instance, returning to networks with Hebb-Hopfield couplings and to standard sigmoid gain functions—with $g(x) \rightarrow \pm 1$ for $x \rightarrow \pm\infty$ and $g'(x) \leq g'(0) = 1$ —as discussed in section 3, we find a third universal point of the retrieval phase boundary. Besides $\gamma_c^{-1}(T = 0, \alpha = 0) = 1$ and $\alpha_c(T = 0, \gamma^{-1} = 0) \simeq 0.138$, we also have $T_c(\gamma^{-1} = 0, \alpha = 0) = \frac{1}{3}$, irrespective of other details of the gain function. These three universal points on the axes of the three-dimensional parameter space give a rough first impression of the typical size of the retrieval region in parameter 3-space. Note that the critical temperature at infinite gain, where neural outputs are confined to ± 1 , is $T_c = \frac{1}{3}$ (which is incidentally also the critical temperature of the mean-field Heisenberg ferromagnet) rather than $T_c = 1$, as one might perhaps have expected on the basis of an analogy with the stochastic Ising network in this limit. The origin of this discrepancy lies in the fact that the dynamical variables in the present case are neural potentials rather than neural output levels as in the stochastic Ising network.

Details of the phase boundary, including the order of the phase transition at non-zero temperature, will of course depend on details of the transfer function chosen. In figure 5 we present boundaries of the retrieval phase in the α - γ plane for various

temperatures, both for networks with Hebb–Hopfield couplings and for networks with pseudo-inverse synapses.

All this may seem nice. At this point, however, we have to draw the readers' attention to the condition (67) on the variance of the Gaussian white noise under which the above results about the collective behaviour of networks with noisy potential dynamics apply. For gain functions which saturate as $U_i \rightarrow \pm\infty$, (67) demands that $D(U_i)$ diverges as $|U_i|$ becomes large. Note that for popular choices of input–output relations such as the hyperbolic tangent or the Fermi-function, the required divergence of $D(U_i)$ would even have to be exponentially fast. Neither for biological neural networks nor for their electronic counterparts, such properties of the noise seem in any way reasonable. From a pragmatic point of view, one might perhaps argue that neither biological nor electronic networks ever operate in the limits $U_i \rightarrow \pm\infty$, so that in the 'accessible regions of phase space', i.e. those to which the Gibbs distribution (69) does at all give appreciable weight, the condition (67) may not be as outlandish as it may seem at first sight. On the other hand, there are gain functions, in particular those modelled to encode neural refractoriness (see, for instance, section 4.1), for which the ' $g'(\gamma U_i) \rightarrow 0$ catastrophe' in (67) occurs at finite U_i , namely, immediately below threshold. Thus, for this case, the pragmatic point of view presents no way out of the dilemma.

For the stochastic V -dynamics (48), the condition (56) on the noise implies that the variance $D(V_i)$ of the noise, and with it the drift term (52), must vanish at saturation levels of the gain function, for which $g'(g^{-1}(V_i)) = g'(\gamma U_i) = 0$. As a consequence, a neuron with a firing rate V_i reaching such a saturation level, would henceforth never change its state, since both systematic and fluctuating forces exerted on it vanish. The network dynamics would therefore eventually completely freeze in neural saturation levels, whenever such saturation levels exist. In the stationary distribution (60), this freezing effect manifests itself in the fact that the *single-site measure* $dV_i/g'_i(g_i^{-1}(V_i))$ has non-integrable singularities at the saturation levels for every saturating input–output relation. This implies that the stationary distribution gives all the weight to these saturation levels, no matter what the other system parameters are. If we denote the possible configurations of saturation levels by $S = (S_i)$, then $\mathcal{P}(S) \propto \exp[-\beta\mathcal{H}_N(S)]$ would according to (60) give the probability that the system is frozen in S , independently of the gain parameter γ . For systems for which the gain functions lead to saturation levels $S_i \in \{\pm 1\}$ with $G(1) = G(-1)$, the statistics of asymptotic saturation configurations is thus given by the equilibrium distribution of the standard model with asynchronous stochastic dynamics!

As for the potential-dynamics, the detailed balance solution for the V -dynamics is not to be had without pathologies. In particular, the *vanishing* of the noise at saturation levels of the gain-functions appears to be an unrealistic feature of noise sources. On the other hand, this property at least—along with the vanishing of drift terms at saturation levels—must be regarded as a consequence of our transformation of a 'natural' U -dynamics in an unbounded domain into a V -dynamics, which is now defined in a bounded domain (if the $g(\gamma U_i)$ saturate), rendering detailed-balance conditions somewhat unnatural.

But with the 'natural' U -dynamics, too, the pathologies associated with detailed balance solutions remain. The obvious conclusion must then be that detailed balance solutions are not realistically to be observed in neural networks with continuous time dynamics. For biological nets, this is not surprising anyway, because for these the requirement of essential synaptic symmetry is unrealistic to begin with, so that the

pathologies of the noise appear as just a minor additional problematic issue. In the case of electronic networks, however, all building blocks can be designed to make the deterministic dynamics (1), (2) quantitatively correct, and to satisfy all conditions needed for detailed balance, *except* that on the noise. (There was one additional assumption we made on our way, namely the smoothness assumption on Φ in (54), resp. (65).) At least for systems with smooth input-output relations, this assumption seems to be innocuous, however, and our feeling is that an escape from the pathologies encountered should not be sought along the line of relaxing this particular condition.) Thus, perhaps surprisingly, stationary states satisfying detailed balance might not even be observed in networks built of electronic circuitry—unless, of course, relaxing our restrictive assumptions on the family of noise models taken into account could save the situation. Because of these restrictive assumptions, it must be admitted, our conclusions here must still be regarded as somewhat tentative. Nevertheless, the results of the present section and the foregoing discussion should shed some light on a number of difficulties that one is likely to encounter when dealing with noise in coupled nonlinear dynamic systems.

6. Summary and discussion

In the present paper, we have explored our previously developed statistical-mechanical approach to analysing the long-time behaviour of networks of analogue neurons governed by a set of RC -charging equations, by applying it to a networks varying with respect to learning rules and to the statistics of stored data.

In particular, we have provided further details about the analogue version of the Hopfield net, we have studied low-activity low-firing-rate systems, and we have analysed analogue neuron systems coupled via pseudo-inverse synapses.

Local field distributions were computed for the analogue version of the Hopfield model, and were found to deviate from the Gaussian form obtained for stochastic neurons in the context of the replica approach.

Replica symmetry of the retrieval solutions in networks with Hebb–Hopfield couplings was found unbroken down to fairly small values of inverse gain, whereas the spin-glass phase is unstable with respect to replica symmetry breaking *right where it appears*—much as in the standard model, if inverse-gain and temperature scales are identified.

Networks of analogue neurons storing low-activity patterns were observed to retrieve naturally at low firing rates, if gain functions are chosen such as to model neural refractoriness. For networks of this type the storage capacity scales as $\alpha_c \sim \bar{V}^2/2a|\ln a|$ as $a \rightarrow 0$, and is thus of the same order of magnitude as the Gardner bound for the storage of low-activity patterns, with a systematic depression proportional to the square of the average (low) firing rate (on a scale on which $V_{\max} = 1$).

Networks coupled by pseudo-inverse synapses can retrieve patterns up to the theoretically possible maximum $\alpha_c = 1$, if neural gain parameters are taken to be sufficiently high.

At this point, let us stress that our statistical-mechanical approach to analogue neuron systems can be applied to a very wide range of systems, namely those for which the set (1), (2) of RC charging equations is governed by a Lyapounov function. This requires synapses to be symmetric, and input-output relations to be monotone

non-decreasing, but otherwise *arbitrary*. While input–output relations should, in principle, increase for large $|U|$ not faster than linearly with $|U|$, to ensure that the Lyapounov function (3) is bounded from below, hence that the system is globally stable, this restriction can, in practice, be easily dispensed with. If one constrains the V_i integrations in (7) to some suitably large compact subset of \mathbb{R}^N , all formal manipulations carried out in the sequel go through unaltered, and the fixed-point equations describing the collective properties of the system remain valid *as they are*, except that *all* solutions (with bounded \dot{V} in (19), (21), etc) now describe truly metastable situations, there being no absolute minimum of $\mathcal{H}_N(\mathbf{V})$ for finite V_i , hence no true absolute stability of the system.

Lastly, the statistical-mechanical approach was generalized to include effects of fast stochastic noise. Equilibrium distributions satisfying detailed balance were shown to be unique and of Gibbs' canonical form, exhibiting a remarkable reciprocity between potential dynamics and firing-rate dynamics, as discussed in detail in the previous section. They do, however, exist only under unrealistic assumptions about the noise in the system. Hence, our conclusion is that equilibrium distributions satisfying detailed balance are not to be observed in systems with realistic noise sources, that is to say, noise sources within the class of noise models considered. Elucidating the nature of the probability currents that would be present in such systems—even if in equilibrium—would certainly be a project worth pursuing.

It is clear that the assumptions regarding properties of the noise sources we have made in section 5—though quite commonly adopted—should eventually be relaxed in the course of arriving at a solution that can be deemed fully satisfactory in general. As yet, we have not been able to do so.

In closing, it is perhaps worth pointing out that (3) is also a Lyapounov function of the *asynchronous* dynamics

$$V_i(t + \Delta t) = g_i \left(\gamma_i R_i \left[\sum_j J_{ij} V_j(t) + I_i \right] \right)$$

provided that the g_i are monotone increasing and that the J_{ij} are symmetric, with $J_{ii} = 0$. Under these conditions the Gibbs-distribution generated by $\mathcal{H}_N(\mathbf{V})$ is an equilibrium distribution for the asynchronous stochastic dynamics defined by the transition probability

$$\begin{aligned} \text{Prob}\{V_i(t + \Delta t) \in \{V', V' + dV'\} | \mathbf{V}\} \\ = \frac{\exp \left[\beta \left(U_i(\mathbf{V}) V' - \gamma_i^{-1} G_i(V') \right) \right] d\rho(V')}{\int d\rho(V') \exp \left[\beta \left(U_i(\mathbf{V}) V' - \gamma_i^{-1} G_i(V') \right) \right]} \end{aligned}$$

with $U_i(\mathbf{V}) = \sum_{j(\neq i)} J_{ij} V_j + I_i$, which satisfies detailed balance (see also Treves 1990a, b). This type of asynchronous dynamics does not suffer from the pathologies of the noise source we had to assume above, but it does, of course, leave the realm of networks with continuous time dynamics to which the present contribution has been devoted.

Acknowledgments

It is a pleasure to thank J L van Hemmen, H Horner, W Kinzel, R Kree, M Oppen, and A Zippelius for most valuable discussions and helpful advice.

This paper will be part of the PhD Thesis of SB, who also acknowledges financial support of the Volkswagenstiftung. The work of RK has been supported by the Deutsche Forschungsgemeinschaft through the Sonderforschungsbereich-123 'Stochastische Mathematische Modelle'.

References

- Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys., NY* **173** 30
 Amit D J and Tsodyks M V 1991a *Network* **2** 259
 Amit D J and Tsodyks M V 1991b *Network* **2** 275
 Bös S 1992 *PhD thesis (Giessen)* unpublished
 Bös S and Kühn R 1992 in preparation
 Bollé D, Dupont P and van Mourik J 1991 *Europhys. Lett.* **15** 893
 Bray A J, Sompolinsky H and Yu C 1986 *J. Phys. C: Solid State Phys.* **19** 6389
 Buhmann J, Divko R and Schulten K 1990 *Phys. Rev. A* **39** 2689
 Cohen M A and Grossberg S 1983 *IEEE Trans. on Systems, Man and Cybernetics SMC-13* 815
 de Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983
 Fukai T and Shiino M 1990 *Phys. Rev. A* **42** 7459
 Gardiner C W 1983 *Handbook of Stochastic Methods* (Berlin: Springer)
 Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
 Gerl F, Bauer K and Krez U 1992 *Z. Phys. B* **88** 339
 Gerstner W and van Hemmen J L 1992 *Network* **3** 139
 Graham R and Haken H 1971 *Z. Phys.* **243** 289
 Herz A 1991 *Phys. Rev. A* **44** 1415
 Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
 ——— 1984 *Proc. Natl Acad. Sci. USA* **81** 3088
 Hopfield J J and Tank D W 1985 *Biol. Cybern.* **52** 141
 Horner H 1989 *Z. Phys. B* **75** 133
 Kanter I and Sompolinsky H 1987 *Phys. Rev. A* **35** 380
 Koch C, Marroquin J and Yuille A 1986 *Proc. Natl Acad. Sci. USA* **83** 4263
 Kohonen T 1984 *Self-organization and Associative Memory* (Berlin: Springer)
 Kühn R 1990 *Statistical Mechanics of Neural Networks (Springer Lecture Notes in Physics 398)* ed L Garrido (Heidelberg: Springer) p 19
 Kühn R 1991 *Habilitationsschrift (Heidelberg)* unpublished
 Kühn R, Bös S and van Hemmen J L 1991 *Phys. Rev. A* **43** 2084
 Marcus C M, Waugh F M and Westervelt R M 1990 *Phys. Rev. A* **41** 3355
 Marcus C M and Westervelt R M 1989 *Phys. Rev. A* **40** 501
 Mead C 1989 *Analog VLSI and Neural Systems* (Reading, MA: Addison-Wesley)
 Mertens S, Köhler H M and Bös S 1991 *J. Phys. A: Math. Gen.* **24** 4941
 Peretto P 1984 *Biol. Cybern.* **50** 51
 Perez-Vicente C J and Amit D J 1989 *J. Phys. A: Math. Gen.* **22** 559
 Personnaz L, Guyon I and Dreyfus G 1985 *J. Physique Lett.* **46** 359
 Peterson C and Anderson J R 1987 *Complex Systems* **1** 995
 Rieger H 1990 *J. Phys. A: Math. Gen.* **23** L1273
 Shiino M and Fukai T 1990 *J. Phys. A: Math. Gen.* **23** L1009
 ——— 1992 *J. Phys. A: Math. Gen.* **25** L375
 Sompolinsky H, Crisanti A and Sommers H J 1988 *Phys. Rev. Lett.* **61** 259
 Soukoulis C M, Levin K and Grest G S 1983 *Phys. Rev. B* **28** 1495
 Tank D W and Hopfield J J 1985 *IEEE Trans. on Circuits and Systems CS-33* 533
 Treves A 1990a *J. Phys. A: Math. Gen.* **23** 2631
 ——— 1990b *Phys. Rev. A* **42** 2418
 Tsodyks M V and Feigel'man M V 1988 *Europhys. Lett.* **6** 101
 Waugh F M, Marcus C M and Westervelt R M 1990 *Phys. Rev. Lett.* **64** 1986