Complex Systems 4 (1990) 75–106

Comparison of Self-Organization and Optimization in Evolution and Neural Network Models

Elgar E. Pichler[*]
James D. Keeler[†]
John Ross[‡]
*Chemistry Department, Stanford University,
Stanford, CA 94305, USA*

Abstract. The Eigen model of macromolecular evolution is compared with the Little–Hopfield neural network model with discrete-state neurons. Similarities of these systems are shown by their description as Ising spin models. Both of the systems show self-organizing behavior in certain parameter regions. Energies for the single states can be defined in such a way that self-organization results in a localization around states with small energies. Therefore, both models can be used as optimization algorithms for complex combinatorial problems, and they can be interpreted as special cases of a more general optimization algorithm. The self-organization process depends on nonnegative transition frequency matrices, which describe transitions of the systems from one state to another. The transition frequencies are functions of a Gaussian noise source, which models an internal temperature. A standard deviation $\theta \geq 0$ for the distribution of the noise is necessary in both cases for an exhaustive search of the state space. However, the neural network can find local minima of the global energy function at $\theta = 0$, whereas the evolution model cannot do this because noise has to be present for migration to new states in the evolutionary run. Maximal values of θ, θ_{th}, above which no self-organization occurs, are given for both systems. In the evolution model there is a sharp decrease of θ_{th} with the size of the system. Self-organization is possible for the evolution model up to a certain system size even at noise levels which are higher than the critical noise level for the neural network model. External requirements determine effective noise

[*] Computer address (Internet): pichler@bogart.stanford.edu.

[†] Permanent address: Microelectronics and Computer Technology Corporation, 3500 West Balcones Center Drive, Austin, TX 78759, USA. Computer address (Internet): keeler@mcc.com.

[‡] To whom correspondence should be sent. Computer address (Internet): ross@cumulus. stanford.edu.

thresholds which are lower than the critical noise levels. Both systems relax to stationary states. Numerical simulations show that the dependence of the stationary state and the relaxation times on system parameters is different in the two models, although the same energy function is used for optimization, and that the probability of obtaining good solutions to optimization problems may be higher during the relaxation process than at the stationary state.

1. Introduction

Biological systems contain structures which can give rise to complex behavior. Interactions of a biological system with its environment can lead to macroscopic changes in the biological system. Such changes can be interpreted as adaptation of the system to external influences or constraints. We see this for example in evolutionary systems and in the brain and nervous system. These two systems process and store information, and adapt over time. One may ask the following questions: 1) are there similarities in the ways that models of such evolutionary and neuronal systems operate with information, and 2) can the wide varieties in adaptation and processing time be ascribed to differences in the models of the biological systems? We address both of these questions in the following by examining the Eigen model of an evolutionary system and a neural network model as it was defined by Little and Hopfield. We find that several similarities exist in these models: the dynamics in these two seemingly different systems occur on a surface that can be made to be topologically isomorphic; however, the motion on this surface is different, which results in different dynamics and different adaptation in the two models.

Prebiotic evolution in the Darwinian sense has been described by Eigen [1] for macromolecules which may be polynucleotide sequences. In Eigen's evolution model the macromolecules can be described as one-dimensional Ising spin chains [2,3]. The macromolecules are able to reproduce on themselves and thereby generate other macromolecules as offspring. The length of the spin chains is assumed to be constant during the evolutionary process, and only point mutations in the form of flipping of spins on the spin chains are allowed; insertions and deletions do not occur in this model. Every one of the possible spin chains is assigned a certain fitness value, as measured by the rate of reproduction, which reflects the extent of adaptation of each macromolecule to the environment. The fitness of each of the possible different species can be evaluated according to different functions of the spin chain characteristics [3,4] or can simply be set to a certain value. The evolutionary process takes place in an evolution reactor (ER) which is a continuously stirred tank reactor with inflow and outflow. Through the inflow, energy-rich molecules are introduced into the system. With these molecules a replication machinery, operating at a given error rate, produces new macromolecules from old ones, which are used as templates. The environmental constraint is given by the requirement of constant organization, i.e. the outflow is regulated in such a way that the overall concentration of macromolecules in the ER stays constant. This selective pressure drives the ER system toward mixtures of macromolecules with high replication rates. For this ER model, the survival of a single species is therefore a function of the replication rate of each of the species and of the error rate. Models of this kind are self-organizing, i.e. the competition among the species leads to the selection of a so called quasi-species [1,5–7], which corresponds to the calculated distribution of states at the stationary state of the system. Various methods have been employed to describe the behavior of this ER model: phenomenological kinetic equations, stochastic descriptions of the temporal development of the system [4,8], and stationary state statistics [3] of the ER model all give qualitatively the same results; in particular, they all show a limitation of the size of the species for a given error rate and selection of a quasi-species only below a certain error rate [3,6–11].

Neural networks (NN) have been used to describe collective phenomena of "neuron"-like elements [12–15]. In the discrete state case neurons in an NN are two-state model devices where the two states of a neuron loosely correspond to the firing and nonfiring states of biological neurons. The model neurons are fully interconnected, and each neuron is able to update its state according to a given algorithm. In most of these models the neurons have a threshold to the input from other neurons, so that they become active or inactive if the input exceeds the threshold value or is below it, respectively. The size of an NN is given by a finite number of neurons. These kinds of NN can be described as Ising spin systems as well [16–19].

The transitions of the NN from one state to another state depend on the connections between the single neurons. In most NN, learning algorithms are employed which change the connections in such a way as to ensure stable fixed points at desirable locations in state space. If the NN is to be used as a content-addressable memory, the memory states can be defined for systems without noise as stable fixed points with the so-called Hebbian learning rule [12]. If we start the NN close enough [20] to one of the memory states, the system will undergo a self-organization process, i.e. the NN will relax to the memory state. During that process the NN minimizes an energy or cost function [12]. In the case of zero noise, stable fixed points of the system correspond to minima of this energy function [12,13]. In the presence of noise, the system can localize around states with low associated energy but also has a finite probability of being in states with higher energies in the stationary state [21].

In this article we compare the ER model of Eigen with the NN model of Little, which is equivalent to the Hopfield model with synchronous updating. These models are special examples of systems which can be used for finding solutions in complex optimization tasks [22–27]. Therefore we compare the self-organization process and the capacities as optimization algorithms of both systems.

We review Ising spin system descriptions of the models (see for example [3,19]) in section 2; to make a comparison between the models possible we define noise in the ER model in a new way in terms of noise with a Gaussian distribution. First we discuss a general model of optimization in systems with discrete states. Then we derive the transition matrices for both types of systems in the case of a special energy function. The temporal behavior of the models is governed by these transition matrices, with which it is also possible to calculate the probability that a system will be in a certain state at a certain time. The dependence of the transition matrices on noise and energy functions is pointed out. For our description we assume a Gaussian noise source with mean of zero and standard deviation θ. Energy or cost functions are defined in such a way that they assign a value to every possible state.

In section 3 we discuss various properties of the transition frequency matrices and consider the importance of noise or the existence of temperature, noise thresholds, and the relaxation process in the ER and the NN model. Noise is an indispensable factor if a big portion of the state space needs to be searched for low energy values, i.e. if good solutions are sought to an optimization problem. In fact, noise *has to* be present in the ER case in order for optimization to lead to states which were not represented at time $t = 0$; otherwise, at $\theta = 0$, optimization leads only to selection of the best of the states represented at time $t = 0$ in the ER model. For both algorithms, critical noise thresholds can be defined and above these no self-organization occurs. Finally, we address the question of relaxation times and usefulness of the models as optimization algorithms, and we discuss results from numerical solutions of the equations of the ER and NN model.

2. Description of the systems

2.1 General description

In the following we look at the temporal development of systems which can only be in a finite number of discrete macrostates, henceforth called states for simplicity. A state consists of a specification of n microstates. We assume that the number N of possible states is equal to the number of all combinations of types of microstates,

$$N = r^n, \tag{2.1}$$

where r is the number of different types of microstates. A state which can be obtained by such a combination is represented by the vector

$$\mathrm{s}_k = \begin{pmatrix} s_1^k \\ \vdots \\ s_n^k \end{pmatrix}, \qquad k \in \{1, 2, \ldots, N\}. \tag{2.2}$$

To describe the temporal behavior of the system we introduce the vector

$$\sigma(t) = \begin{pmatrix} \sigma_1(t) \\ \vdots \\ \sigma_n(t) \end{pmatrix} \tag{2.3}$$

to denote the state of the system at time t. $\sigma_i(t)$ is therefore the type of the ith microstate at time t.

In the example of an ER model the microstates are the single symbolic nucleotides, 1, or -1, either purine or pyrimidine residues, on different positions of a polynucleotide, e.g. an RNA-strand. In the case of the NN model a microstate is one of the neurons in the NN; each single neuron can be in a firing (1) or in a nonfiring (-1) state. Since there are only two different kinds of microstates, both models have been treated as special spin systems [3,14,19]; with $r = 2$, the number of possible states of the systems is $N = 2^n$.

An energy function can be associated with every one of the states:

$$E_k = E(\mathbf{s}_k) \qquad \forall k. \tag{2.4}$$

In the optimal case the values of the energy function are known for all states and adaptation should lead to states close to or equal to states for which the energy function has a deep minimum. For most cases only a few values of the energy function may be known: the dependence of the energy function on numerous system parameters might make it difficult to estimate the behavior of the energy function in some or all regions of the state space, or the state size may be so big that explicit knowledge of the energy function at every possible state is impossible. Even though an energy function may be easily formulated, the behavior of the system or its movement in state-energy space may still not be obvious. This situation occurs frequently in complex combinatorial optimization problems, for which the analytical form of an energy function over a usually huge state space is known, but where it is not obvious for which states the energy function is a minimum. In these cases optimization algorithms are used to find states with low energy values. A good optimization algorithm has to find good solutions, i.e. deep or absolute minima in the energy function E_k. It has to achieve this without a complete search of the state space, since such a search quickly becomes impossible because of the exponential growth of the number of possible states with n (see (2.1)). The algorithm is only useful if it can do this in a reasonably short time and with acceptable computational effort.

ER and NN systems can be modeled in such a way that adaptation from an arbitrary initial state can lead to minimization of an energy function, or in other words, can lead to good solutions of an optimization problem. The updating process is quite different in these two models, but the general process can be characterized by a transition matrix $\mathbf{W} = \|w_{lk}\|_1^N$: the elements of that matrix are the transition frequencies:

$$w_{lk} = (\text{quality term})_k (\text{transition term})_{lk} \tag{2.5}$$

for the transition from state $\mathbf{s}_k$ to state $\mathbf{s}_l$ per unit time. The (quality term)$_k$ describes how good a solution $\mathbf{s}_k$ is, and the (transition term)$_{lk}$ gives the probability for the transition from state $\mathbf{s}_k$ to state $\mathbf{s}_l$. Both terms in (2.5) can be different in different optimization algorithms, although the optimization problem itself as well as the energy function are the same in these models.

This will be shown in later sections when we discuss the transition frequencies in the ER and the NN model.

The (transition term)$_{lk}$ in (2.5) depends on the updating rule for the system, which is

$$\sigma_i(t+1) = g\left[h_i(t)\right], \tag{2.6}$$

a function of the input $h_i(t)$. We will define one of many possible functions g in section 2.2.

We introduce probabilities of representation,

$$\psi(t) = P\begin{pmatrix} \boldsymbol{\sigma}(t) = \mathbf{s}_1 \\ \vdots \\ \boldsymbol{\sigma}(t) = \mathbf{s}_N \end{pmatrix}, \tag{2.7}$$

to describe the updating process if stochastic elements are included, as they will be in the ER and NN model in the form of noise. $\psi_k(t) = P\left[\boldsymbol{\sigma}(t) = \mathbf{s}_k\right]$ of (2.7) is the probability that the system is in state $\mathbf{s}_k$ at time t.

In order to generate a test function for the energy, we use the Hopfield Hamiltonian

$$E_k = -\frac{1}{2}\sum_{\substack{i,j=1 \\ i\neq j}}^{n} T_{ij} s_i^k s_j^k \tag{2.8}$$

where

$$T_{ij} = \frac{J}{n}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu,\ \forall i \neq j \qquad \text{and} \qquad T_{ii} = 0,\ \forall i. \tag{2.9}$$

J/n is a parameter for modulating the interaction strength between different spins. J is assumed to be nonnegative. The $\xi_i^\mu \in \{-1, 1\}, \forall \mu, i$ are chosen randomly with equal probability and form states $\boldsymbol{\xi}_\mu$ of the systems. Therefore, the $\boldsymbol{\xi}_\mu$ form a randomly chosen subset of the set of possible states $\mathbf{s}_k$. Equation (2.9), the Hebbian learning rule, constitutes an algorithm which defines energy minima for the $2p$ states $\pm\boldsymbol{\xi}_\mu$.

We now review and develop descriptions of an NN and an ER model which fit this general formalism. Transition frequencies are derived by formulating (quality term)s and (transition term)s of (2.5) for both models. We use the same variable names whenever possible and point at differences in the variables by using the superscripts "$\hat{\ }$" for the ER model and "$\tilde{\ }$" for the NN model.

2.2 The neural network model

In this section we describe the Little NN model [14,15]. This NN consists of two-state model devices, the neurons, which are updated synchronously at each timestep. The accuracy of the updating process depends on the the amount of noise added to the system. Noise is taken to be a stochastic variable with a Gaussian probability distribution with mean equal to zero and standard deviation θ; natural fluctuations like the ones caused by "imperfect" connections, stochastic firing, and others are modeled by the noise term [21]. The updating function is

$$\sigma_i(t+1) = g\left[h_i(t)\right] = \begin{cases} 1, & \text{if } \ h_i(t) + (\text{noise term})_i(t) > 0; \\ \sigma_i(t), & \text{if } \ h_i(t) + (\text{noise term})_i(t) = 0; \\ -1, & \text{if } \ h_i(t) + (\text{noise term})_i(t) < 0; \end{cases} \quad \forall i. \tag{2.10}$$

The input function for the NN model, $\widetilde{h}_i(t)$, is defined as

$$\widetilde{h}_i(t) = \sum_{j=1}^{n} T_{ij}\sigma_j(t), \qquad \forall i \tag{2.11}$$

where all the (noise term)$_i(t)$ have the same probability distribution (see figure 1). Transition frequencies between two arbitrary states $\mathbf{s}_k$ and $\mathbf{s}_l$ can therefore be written as

$$\begin{aligned} \widetilde{w}_{lk} &= \prod_{i=1}^{n} Q\left(-s_i^l \widetilde{h}_i^k\right) \\ &= \prod_{i=1}^{n} \left\{ \tfrac{1}{2}\left[1 - \operatorname{erf}\left(-\tfrac{s_i^l \widetilde{h}_i^k}{\theta\sqrt{2}}\right)\right]\right\} \end{aligned} \tag{2.12}$$

where

$$\begin{aligned} Q(x) &= \tfrac{1}{\theta\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{1}{2}\left(\frac{x'}{\theta}\right)^2} dx' \\ &= \tfrac{1}{2}\left[1 - \operatorname{erf}\left(\tfrac{x}{\theta\sqrt{2}}\right)\right] \end{aligned} \tag{2.13}$$

and

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{2.14}$$

Here $Q(-s_i^l \widetilde{h}_i^k)$ is the probability that $\sigma_i(t+1) = s_i^l$ if $\boldsymbol{\sigma}(t) = \mathbf{s}^k$ (compare also with figure 1).

For presentation purposes and for ease of comparison with the ER model description we now derive a standard description of transition frequencies in the NN case. With the approximation

$$\frac{1}{2}\left[1 - \operatorname{erf}(x)\right] \simeq \frac{1}{1+e^x}, \tag{2.15}$$

where the difference between the two sides of (2.15) is 0.01 at the maximum [28], we have from (2.12)

$$\begin{aligned} \widetilde{w}_{lk} &= \prod_{i=1}^{n} \left[\frac{1}{1+\exp\left(-\frac{1}{\theta\sqrt{2}} s_i^l \widetilde{h}_i^k\right)}\right] \\ &= \frac{\exp\left(\frac{1}{2}\frac{1}{\theta\sqrt{2}} \sum_{i=1}^{n} s_i^l \widetilde{h}_i^k\right)}{\prod_{i=1}^{n}\left[\exp\left(\frac{1}{2}\frac{1}{\theta\sqrt{2}} s_i^l \widetilde{h}_i^k\right) + \exp\left(-\frac{1}{2}\frac{1}{\theta\sqrt{2}} s_i^l \widetilde{h}_i^k\right)\right]}. \end{aligned} \tag{2.16}$$

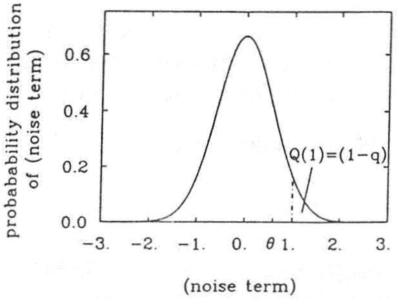

Figure 1: Normal distribution of noise for $\theta = 0.6$ and probability of incorrect updating, $(1-q) = Q(1)$, for the ER model (see equations 2.25 and 2.31).

With the use of

$$\prod_{i=1}^{n}\left(e^{x_i}+e^{-x_i}\right)=\sum_{\{s_i\}}\exp\left(\sum_{i=1}^{n}s_i x_i\right) \tag{2.17}$$

Little and Shaw [15,28] found

$$\begin{aligned}\widetilde{w}_{lk} &= \frac{\exp\left(\frac{1}{2}\frac{1}{\theta\sqrt{2}}\sum_{i=1}^{n}s_i^l\widetilde{h}_i^k\right)}{\sum_{j'=1}^{n}\exp\left(\frac{1}{2}\frac{1}{\theta\sqrt{2}}\sum_{i=1}^{n}s_i^{j'}\widetilde{h}_i^k\right)}\\ &= \frac{\exp\left[-\frac{1}{\theta\sqrt{2}}\widetilde{H}(l|k)\right]}{\sum_{j'=1}^{n}\exp\left[-\frac{1}{\theta\sqrt{2}}\widetilde{H}(j'|k)\right]}.\end{aligned} \tag{2.18}$$

Now we substitute for the T_{ij} from (2.9) in (2.18) and get for the transition frequencies

$$\widetilde{w}_{lk}=\frac{\exp\left(\frac{1}{\theta\sqrt{2}}\frac{1}{2}\frac{J}{n}\sum_{\mu=1}^{p}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\xi_i^\mu\xi_j^\mu s_i^l s_j^k\right)}{\sum_{j'=1}^{n}\exp\left(\frac{1}{\theta\sqrt{2}}\frac{1}{2}\frac{J}{n}\sum_{\mu=1}^{p}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\xi_i^\mu\xi_j^\mu s_i^{j'} s_j^k\right)}. \tag{2.19}$$

Hence, by comparison of (2.5) with (2.12) and (2.19) we see that the transition frequeny $\widetilde{w}_{lk}$ in the NN case is the probability for a transitions from state $\mathbf{s}_k$ to state $\mathbf{s}_l$ per unit time. They all contain the same $(\text{quality term})_k = 1, \quad \forall k$. Therefore, transitions from states of higher energy to states of lower energy must be more probable than vice versa in order for optimization to occur. Peretto [28] showed that this is the case for NN as described here and derived Hamiltonians for the low and the high noise level.

2.3 The evolution model

In this section we review the ER model description as it was introduced by Eigen [1] and recently interpreted by Leuthäusser [2,3] as an Ising spin system. In Eigen's model of macromolecular evolution [1] exemplary macromolecules of length n are made up of only two kinds of monomers, which we signify by 1 and -1, in analogy to the two different orientations of a spin. The behavior of the system is governed by the N coupled nonlinear differential equations,

$$\dot{c}_k = (b_k - v_k)c_k + \sum_{l \neq k}^{N} \widehat{w}_{kl} c_l - \sum_{l \neq k}^{N} \widehat{w}_{lk} c_k - c_k \sum_{l=1}^{N} (b_l - v_l) c_l, \qquad (2.20)$$

where the c_k are the normalized concentration variables for the species $\mathbf{s}_k$, i.e., $\sum_{k=1}^{N} c_k = 1$; the b_k are their replication; and the v_k their decay rate coefficients. With $\widehat{w}_{lk}$ we denote the rate coefficient for a mutation that produces a molecule of species $\mathbf{s}_l$ through a replication process with a molecule $\mathbf{s}_k$ as a template. The last term in (2.20) is the outflux term which ensures a constant overall concentration of species in the ER.

In the case $v_k = v, \forall k$ (2.20) can be simplified to

$$\dot{c}_k = b_k c_k + \sum_{l \neq k}^{N} \widehat{w}_{kl} c_l - \sum_{l \neq k}^{N} \widehat{w}_{lk} c_k - c_k \sum_{l=1}^{N} b_l c_l. \qquad (2.21)$$

The constraint of constant overall organization imposes a selection pressure which drives the system towards species mixtures with high average replication rates. With the transformation $y_k(t) = c_k(t) \exp \left[\int_0^t \sum_{k=1}^{N} b_k c_k(\tau)\, d\tau \right]$ the above rate equation becomes linear [29,30],

$$\dot{\mathbf{y}}(t) = \widehat{\mathbf{W}} \mathbf{y}(t). \qquad (2.22)$$

If the evolution process is assumed to be discrete in time (2.22) becomes

$$\mathbf{y}(t) = \widehat{\mathbf{W}}^t \mathbf{y}(0). \qquad (2.23)$$

The $c_k(t)$ can easily be recovered from (2.23) with

$$c_k(t) = \frac{y_k(t)}{\sum\limits_{i=1}^{N} y_i(t)}. \qquad (2.24)$$

If it is assumed that the replication machinery in the ER copies every spin of the template strand with an average replication accuracy q, $q \in [0, 1]$ then we can write

$$\widehat{w}_{lk} = b_k q^{n - d_{lk}} (1 - q)^{d_{lk}}. \qquad (2.25)$$

Here d_{lk} is the Hamming distance between $\mathbf{s}_k$ and $\mathbf{s}_l$,

$$\begin{aligned} d_{lk} &= d(\mathbf{s}_l, \mathbf{s}_k) \\ &= \tfrac{1}{2}\left(n - \sum_{i=1}^{n} s_i^l s_i^k\right), \end{aligned} \tag{2.26}$$

i.e., the number of spins in which these two states differ from each other. So, $q^{n-d_{lk}}$ in (2.25) is the probability that the spins which have the same orientation for $\mathbf{s}_l$ and $\mathbf{s}_k$ are copied correctly; $(1-q)^{d_{lk}}$ is the probability that the spins in which they differ are copied incorrectly.

To compare the ER model with the NN model we rewrite the description developed so far. In analogy to the description of the NN model we use the updating function from (2.10). We define the input function,

$$\widehat{h}_i(t) = \sum_{j=1}^{n} \widehat{T}_{ij}\sigma_j(t), \qquad \forall i \tag{2.27}$$

for the ER model, and assume that

$$\widehat{T}_{ij} = 0,\ \forall i \neq j \qquad \text{and} \qquad \widehat{T}_{ii} = 1,\ \forall i. \tag{2.28}$$

This is tantamount to saying that the ER model is a one-dimensional Ising spin system with nearest-neighbor interactions in the direction of development of time [3]. A (transition term)$_{lk}$ for the ER model can now be formulated as in the NN case,

$$(\text{transition term})_{lk} = \prod_{i=1}^{n} Q\left(-s_i^l \widehat{h}_i^k\right). \tag{2.29}$$

With (2.27) and (2.28) and considering the (quality term)$_k$ we get for (2.25)

$$\begin{aligned} \widehat{w}_{lk} &= b_k \prod_{i=1}^{n} Q\left(-s_i^l \widehat{h}_i^k\right) \\ &= b_k \prod_{i=1}^{n} Q\left(-s_i^l s_i^k\right). \end{aligned} \tag{2.30}$$

The accuracy of replication, q, is now defined in a new way in terms of a Gaussian error integral (compare 2.13; see also figure 1):

$$q = Q(-1). \tag{2.31}$$

This means that q can only be in the interval $[0.5, 1]$. However, the original ER model also allows for replication for which incorrect updating is more likely than correct updating; e.g., exact complementary replication at $q = 0$ causes every spin to be reversed in an updating process, $\boldsymbol{\sigma}(t+1) = -\boldsymbol{\sigma}(t)$. If replication with a higher probability for incorrect than for correct updating is to be described, the $\widehat{T}_{ii}$ would have to be set to -1.

Using the same substitutions as in section 2.2 we can express $\widehat{w}_{lk}$ as

$$\begin{aligned} \widehat{w}_{lk} &= b_k \frac{\exp\left(\frac{1}{2}\frac{1}{\theta\sqrt{2}}\sum_{i=1}^{n} s_i^l s_i^k\right)}{\sum_{j'=1}^{n} \exp\left(\frac{1}{2}\frac{1}{\theta\sqrt{2}}\sum_{i=1}^{n} s_i^{j'} s_i^k\right)} \\ &= b_k \frac{\exp\left[-\frac{1}{\theta\sqrt{2}}\widehat{H}(l|k)\right]}{\sum_{j'=1}^{n} \exp\left[-\frac{1}{\theta\sqrt{2}}\widehat{H}(j'|k)\right]}. \end{aligned} \tag{2.32}$$

We use (2.8) to model the b_k by setting

$$\begin{aligned} b_k &= e^{-\frac{1}{n}E_k} \\ &= \exp\left(\frac{1}{n}\frac{1}{2}\frac{J}{n}\sum_{\substack{i,j=1 \\ i\neq j}}^{n}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu s_i^k s_j^k\right). \end{aligned} \tag{2.33}$$

This relation between b_k and E_k was introduced by Leuthäusser [3]. The scaling factor $1/n$ prevents exponential growth of b_k with n and reflects experimental findings which show that the b_k stay fairly constant over a wide range of n. We point out that any inverse relation of b_k and E_k would result in higher fitnesses for species with low energies. The agreement of model and real system certainly depends on the specific relation of b_k to E_k but other relations than the one used in (2.33) might well result in better performance as an optimization algorithm. We finally get from (2.32) and (2.33)

$$\widehat{w}_{lk} = \exp\left(\frac{1}{n}\frac{1}{2}\frac{J}{n}\sum_{\mu=1}^{p}\sum_{\substack{i,j=1 \\ i\neq j}}^{n} \xi_i^\mu \xi_j^\mu s_i^k s_j^k\right) \frac{\exp\left(\frac{1}{2}\frac{1}{\theta\sqrt{2}}\sum_{i=1}^{n} s_i^l s_i^k\right)}{\sum_{j'=1}^{n} \exp\left(\frac{1}{2}\frac{1}{\theta\sqrt{2}}\sum_{i=1}^{n} s_i^{j'} s_i^k\right)}. \tag{2.34}$$

We see now that, unlike the transition frequencies in the NN case, the transition frequencies in the ER model are determined by two terms (compare equations 2.5, 2.19, and 2.34). The first factor in (2.34), the (quality term)$_k$, is a maximum for states which are equal to the stored states $\pm\boldsymbol{\xi}_\mu$; the second factor in (2.34), the (transition term)$_{lk}$, is a maximum for $l = k$, i.e. when transitions occur from one state to the same state. That means that the diagonal terms of $\widehat{\mathbf{W}}$ are always greater than the other matrix elements in the same row, $\widehat{w}_{kk} > \widehat{w}_{lk}$, $\forall l \neq k$. We can interpret the updating process in the ER model as a conservative process: any given state tends to update preferably to itself; a (transition term)$_{lk}$, $\forall l \neq k$ is only nonzero if noise is present and decreases with increasing Hamming distance between state $\mathbf{s}_k$ and state $\mathbf{s}_l$. The (transition term)$_{lk}$ is independent of the energies E_k and E_l. Only the competition between different states, reflected in their different (quality term)s or in their different replication rate constants, brings about a development of the system toward states with low associated energies.

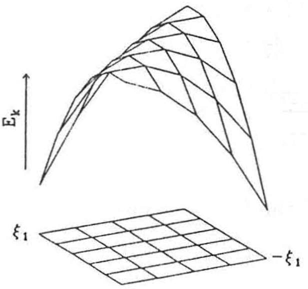

Figure 2: Energy surface, E_k, over a two-dimensional cut through the n-dimensional hypercube for $n = 9$ and one stored pattern, $p = 1$. The two-dimensional cut through the hypercube is represented in the bottom grid: crosspoints on this grid correspond to vertices of the hypercube, which are possible states of the system; Hamming distances between two states are given by the minimum number of edges connecting them; an edge is a line connecting two adjacent crosspoints. The two corners under the energy minima correspond to $\pm\boldsymbol{\xi}_1$.

3. Comparison and simulations

In this section of the article we discuss various characteristics of the ER and the NN models and emphasize similarities as well as differences by using results from exemplary numerical calculations. For our numerical simulations we use, if not stated otherwise, the simplest Hopfield Hamiltonian with one trained state, $p = 1$, as an energy surface. Thereby an energy surface with two minima of equal maximal depth is created; no spurious states or local minima [12,16,19] exist (see figure 2). In the simulations we use the exact transition frequency matrices for both models, (2.12) and (2.30), respectively. The scaling factor for the strength of spin interactions is set to $J = 1$ for all the simulations. We monitor the following variables during a calculation:

> The probability that the system is in state $\boldsymbol{\xi}_1$ at time t, $P[\boldsymbol{\sigma}(t) = \boldsymbol{\xi}_1]$. This is the representation probability of the system at the energy minimum at a certain time.
>
> The probability that the system is closer to $\boldsymbol{\xi}_1$ than to $-\boldsymbol{\xi}_1$,

$$P\left[\boldsymbol{\sigma}(t) \in \{\mathbf{s}_k^+\}\right] = \sum_{k=1}^{N} \psi_k^+(t), \tag{3.1}$$

where the probabilities $\psi_k^\pm(t)$ are given by

$$\psi_k^\pm(t) = \begin{cases} \psi_k(t), & \text{if } \; d[\sigma(t), \pm\boldsymbol{\xi}_1]; > d[\sigma(t), \mp\boldsymbol{\xi}_1] \\ \psi_k(t)/2, & \text{if } \; d[\sigma(t), \pm\boldsymbol{\xi}_1]; = d[\sigma(t), \mp\boldsymbol{\xi}_1] \\ 0, & \text{if } \; d[\sigma(t), \pm\boldsymbol{\xi}_1]. < d[\sigma(t), \mp\boldsymbol{\xi}_1] \end{cases} \tag{3.2}$$

The statistical average of the overlap of the system with $\boldsymbol{\xi}_1$ at time t,

$$m(t) = \frac{1}{n} \sum_{k=1}^{N} \sum_{i=1}^{n} \xi_i^1 s_i^k \psi_k(t). \tag{3.3}$$

The statistical average of the overlap of the system at time t with the pattern $\pm\boldsymbol{\xi}_1$, if the system is closer to $\pm\boldsymbol{\xi}_1$ than to the complementary pattern $\mp\boldsymbol{\xi}_1$,

$$m^\pm(t) = \frac{1}{n} \sum_{k=1}^{N} \sum_{i=1}^{n} \pm\xi_i^1 s_i^k \psi_k^\pm(t). \tag{3.4}$$

At the stationary state ($t = \infty$) the statistical average of the overlap with $\boldsymbol{\xi}_1$ is $m(t) = 0$, because the system is equally localized at $\boldsymbol{\xi}_1$ and $-\boldsymbol{\xi}_1$; therefore $m(t)$ is no longer a good measure of the degree of localization around minimum energy states. $m^\pm(t)$ gives us information about the width of localization around a stored pattern as t goes to ∞, when the system approaches the stationary state (see section 3.1 below).

The statistical average of the normalized overall energy of the system at time t,

$$E(t) = \frac{2}{J(n-1)} \sum_{k=1}^{N} E_k \psi_k(t), \tag{3.5}$$

is an important measure of the performance of the system, if it is used as an optimization algorithm for minimizing the energy function E.

3.1 Transition matrices

We begin the comparison of NN and ER with a discussion of transition matrices. Matrices of the form of (2.19) for the NN and (2.33) for the ER are nonnegative, and for most parameter sets they can be classified as positive matrices. Since for the NN model $\sum_{l=1}^{N} \widetilde{w}_{lk} = 1 \; \forall k$, $\widetilde{\mathbf{W}}$ is a stochastic matrix and updating is a Markov process [31], which can be described by

$$\widetilde{\psi}(t) = \widetilde{\mathbf{W}}^t \widetilde{\psi}(0). \tag{3.6}$$

In the case of the ER model

$$\widehat{\mathbf{W}} = \widehat{\mathbf{W}}'\mathbf{B} \tag{3.7}$$

where $\widehat{\mathbf{W}}'$ is the stochastic matrix of (transition term)s and $\mathbf{B}$ is a diagonal matrix with elements b_k. Therefore $\widehat{\mathbf{W}}$ is only a stochastic matrix if all the $b_k = 1$. In general this is not the case; then the probabilities of representation must be calculated with (2.24) and

$$\widehat{\psi}(t) = \mathbf{c}(t) \tag{3.8}$$

because the c_k are normalized concentration variables per definition.

Processes like the ones described by (3.6) and (2.23),

$$\mathbf{x}(t) = \mathbf{W}^t\mathbf{x}(0), \tag{3.9}$$

can be rewritten in terms of eigenvalues λ_k and eigenvectors φ_k of the matrix $\mathbf{W}$,

$$\mathbf{x}(t) = a_1\varphi_1\lambda_1^t + a_2\varphi_2\lambda_2^t + \cdots. \tag{3.10}$$

These processes undergo a relaxation to a stationary state. A system has reached its stationary state if the probabilities of representation do not change with time any more. Theoretically this will only be the case at $t = \infty$; practically, the $\psi(t)$ change only marginally after a finite number of updating steps. Therefore we say that the system has relaxed to its stationary state if the magnitude of changes in some system variables is lower than a criterial value (see section 3.5). We will see that ER and NN systems relax to stationary states but, because of the different dynamics of these systems, relaxation times and stationary states are different in the two models.

3.2 Noise dependence

In the case of no noise, $\theta = 0$, the transition frequency matrices become sparse matrices. In the ER case we get a diagonal matrix with $\widehat{w}_{kk} = b_k$, $\forall k$. Selection occurs only between species which are present at the beginning of an evolutionary run. The space of possible sequences, including states with higher b_k, cannot be explored because mutations to these states are not possible: the replication machinery works with an accuracy $q = 1$. If the system is in one of the N states with probability 1 at any time, it will stay in this state for all future times. For the NN model updating to other states is still possible, even in the case $\theta = 0$. The transition frequency matrix has entries $\widetilde{w}_{lk} = 1$ only if $s_i^l = g(\widetilde{h}_i^k)$, $\forall i$, and $\widetilde{w}_{lk} = 0$ otherwise. Since the updating function (2.10) leads to states of lower or equal energy [12], updating usually occurs until a state with a local energy minimum is reached. In our simulations we use systems with an odd number of microstates and one stored pattern, $p = 1$. For such NN updating at $\theta = 0$ leads with probability 1 to the energy minimum state which is closer to the initial state. Oscillations between two states with equal energies can occur at $\theta = 0$, and an example of such an oscillation is described in [28]. Oscillations of this type cannot be observed for $\theta = 0$ for the ER model. Energy minima which are reached by NN at $\theta = 0$ are not necessarily global minima. Therefore, especially in cases with complicated energy functions and many local minima, the updating process with $\theta = 0$ may not result in satisfactory solutions. This is a problem (e.g., in NN which have been trained with too many patterns) so that the requirement of pseudo-orthogonality of the patterns is no longer fulfilled. In that case, spurious states [12], combinations of patterns, may be generated which correspond to local minima in the energy function [19].

Noise plays an important role in both systems. In the ER model it makes updating from one state to another possible, and in both models noise allows escape, with finite probability, from states corresponding to local minima in the energy function. An appropriate search of the state space can only occur if noise is present, i.e. $\theta > 0$, in the system, see (2.19) and (2.34). As soon as noise is introduced in the system updating with errors generally allows transitions from any state to any other state. The negative side of this advantage is that, even as t goes to ∞, not only states at energy minima are populated but also nearby states. The width of the distribution around an energy minimum state depends on the level of noise and is different for the two systems (see also sections 3.3 and 3.6).

In the NN model the probability of flipping of one spin in an updating process depends on the "signal-to-noise" ratio for this spin, $\tilde{h}_i/\theta$ (see equations 2.12 and 2.19, [21]). For the ER model, flipping of any spin is equally likely in the updating process for all states (see equations 2.30 and 2.34).

3.3 Noise threshold

The amount of noise which can be added to either system in order to have the possibility of improving optimization in the course of the self-organization process is limited by the acceptable width of the distribution around good solutions at the stationary state (see section 3.6) and ultimately by noise thresholds which can be defined for both systems. Self-organization occurs only below this threshold, characterized by θ_{th}; above the noise threshold the updating process leads to a random walk on the n-dimensional hypercube: no gained information is preserved, and the distribution around good solutions is no higher than around bad solutions at the stationary state. Therefore the noise threshold at θ_{th} represents an upper limit for the amount of acceptable noise in the discussed systems. If the models are used as optimization algorithms a lower effective threshold value, θ_{eff}, is defined by the acceptable width of the distribution around states with minimal associated energy: the probabilities of representation at minimum energy states are higher if the noise is kept at low levels. If an energy minimum state has to be represented with a minimum probability, noise levels have to be used which allow sufficient localization. At higher noise levels, on the other hand, transitions to other states are more likely and a bigger part of the state space is searched in less time. Therefore, both algorithms are used most efficiently at noise levels just below the effective noise level.

We have reported a noise threshold for the Hopfield NN [21], which is similar to a Little NN except the neurons are updated randomly and asyn-

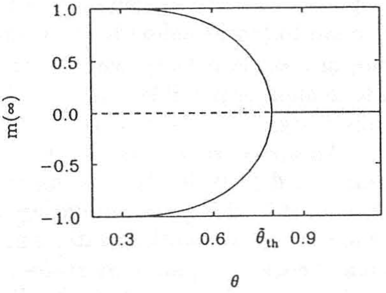

Figure 3: Bifurcation diagram of the statistical average of the overlap of the system with ξ_1 at time $t = \infty$, $m(\infty)$, and standard deviation of the noise threshold, $\widetilde{\theta}_{th}$, of the NN model as the size of the system goes to infinity, $n \to \infty$ (see equation 3.11, compare [21]); $J = 1$. Solid lines correspond to stable solutions and the dashed line to unstable fixed points.

chronously, one neuron at a time. The long-term behavior of the Hopfield NN is equivalent to that of the Little NN [17,28]. In [21] we calculated a noise threshold at

$$\widetilde{\theta}_{th} = J\sqrt{\frac{2}{\pi}}, \tag{3.11}$$

for noise with a Gaussian distribution. Below this noise threshold self-organization takes place and the final statistical average of the overlap of the system with a stored pattern, or the pattern complementary to the stored pattern, is nonzero. Figure 3 shows stable fixed points for the statistical average of the overlap (see 3.3) with the stored pattern as a function of θ for an infinitely large NN.

In an ER the replication rate for a state at an energy minimum has to be greater than the average replication rate of the system without this state, otherwise self-organization and localization around this state do not occur [1,11]. For the Ising model description of the ER model, with the same replication rates as in (2.33), a threshold copying accuracy,

$$q_{th} = \frac{n}{J+n}, \tag{3.12}$$

has been derived in [3] for the limit $J/n \ll 1$. For a given n we can calculate a critical error threshold, q_{th}, or a critical standard deviation for the noise distribution, $\widehat{\theta}_{th}$, as we introduced it. Alternatively, we can calculate a maximum system size, $n_{\max}$, for a given q with (3.12); $n_{\max} = \lfloor n \rfloor$ is the largest

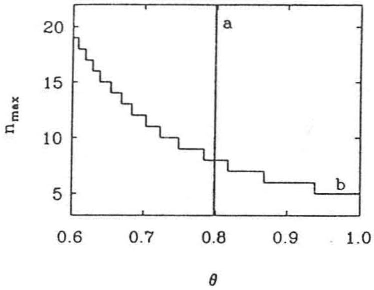

Figure 4: Maximum size of the system, $n_{\max}$, as a function of the standard deviation of the noise threshold, θ_{th}, for the NN (a) and ER models (b) (see equations 3.11 and 3.12).

integer which is not greater than n. Below q_{th} or above $n_{\max}$ a localized quasi-species can no longer be selected. Figure 4 shows $n_{\max}$ as a function of the noise distribution standard deviation, θ, for the ER and the NN model at $J = 1$. We see that $\widehat{\theta}_{th}$ for the ER is a function of n, whereas $\widetilde{\theta}_{th}$ for the NN is a constant.

We find for the specific example of $J = 1$ (compare figure 4) that for small systems, $n < 10$, lower noise levels are necessary to achieve organization for the ER model than for the NN model: e.g. for $n = 10$, $\widetilde{\theta}_{th} \simeq 0.8$ and $\widehat{\theta}_{th} \simeq 0.74$. Most optimization problems require system sizes with larger n. Therefore, ways to circumvent the problem of low noise thresholds should be very useful for some applications, and we mention ways to increase the noise threshold for both models in section 3.4. In section 3.6 we discuss why the ER model performs worse at large system sizes than at small system sizes for a given noise level.

We also note that small size ER models, $n_{\max} \leq 8$, are able to localize at values $\theta > \widetilde{\theta}_{th}$. To see when and for which n the ER model still shows self-organizing behavior even when the noise level is too high for the NN, we calculate $n_{\max}[\widetilde{\theta}_{th}(J)]$, the maximum system size for the ER at the noise threshold for the NN. From (2.25) and (2.31) we have for $q(\widetilde{\theta}_{th})$, the accuracy of replication for the ER model at the noise threshold of the NN model,

$$q\left(\widetilde{\theta}_{th}\right) = Q\left(-1; \widetilde{\theta}_{th}\right). \tag{3.13}$$

With (3.12) and (3.13) we get

$$n_{\max}\left[\widetilde{\theta}_{th}(J)\right] = \left\lfloor \frac{Jq\left[\widetilde{\theta}_{th}(J)\right]}{1 - q\left[\widetilde{\theta}_{th}(J)\right]} \right\rfloor. \tag{3.14}$$

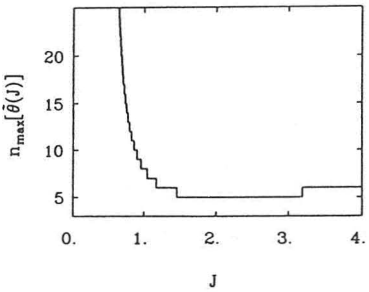

Figure 5: Maximum size of the system for the ER model at $\tilde{\theta}_{th}(J)$, $n_{\max}[\tilde{\theta}_{th}(J)]$ (see equation 3.14): for a given J self-organization is possible for the ER model up to $n_{\max}[\tilde{\theta}_{th}(J)]$, but it is not possible for the NN model at $\theta > \tilde{\theta}_{th}(J)$.

If the system size of the ER is less than $n_{\max}[\tilde{\theta}_{th}(J)]$ self-organization is possible in the ER model for a standard deviation of the noise distribution which is greater or equal than $\tilde{\theta}_{th}(J)$. Figure 5 shows a plot of $n_{\max}[\tilde{\theta}_{th}(J)]$ as a function of J. For all J there exist ER model sizes for which self-organization is possible at noise levels which are higher than the threshold noise levels for the NN model. The smallest "advantage" of this kind for the ER model occurs if J is approximately in the interval $[1.5, 3.2]$ which corresponds to a $\tilde{\theta}_{th}$-range of $[1.2, 2.6]$ or a q-range of $[0.80, 0.65]$. For this J-interval $n_{\max}[\tilde{\theta}_{th}(J)]$ is 5. A region in which self-organization may occur for significantly longer sequences in the ER model than for the NN model lies in the regime of low J- and high q-values. For the regime of high J- and low q-values (3.14) is no longer a good approximation for $n_{\max}$ since $n_{\max} \to J$: $q[\tilde{\theta}_{th}(J)]$ approaches 0.5 as J goes to ∞. Therefore the ER model may have a significant advantage over the NN model if optimization is sought at low spin interaction parameters and low noise levels.

3.4 Higher-order interaction

We mention higher-order interactions as an interesting but certainly not fully explored way of dealing with low noise thresholds, in particular how to operate with self-organization above these thresholds. We have shown in [21] how neuron–neuron interactions via secondary connections, T_{ijk}, can be used to move the error threshold to higher values. With these connections the abilities of an NN to function as a content-addressable memory can be retained above the threshold calculated for systems with first-order input function terms.

Higher-order interactions, $\hat{w}_{klm}$, symbolizing species–species interactions, have been proposed for the ER model [5,9,10]. Recently it has been shown [32] that these interactions may also cause the noise threshold to rise.

Although the type of these interactions are completely different in the two models, the resulting qualitative result is the same. In both models noise thresholds are shifted to higher values and hysteresis effects in the statistical average of the overlap with a stored pattern, $m(t)$, can be observed.

3.5 Relaxation process

The temporal development of the ER and NN models can be described with eigenvalues λ_k and eigenvectors $\boldsymbol{\varphi}_k$ of the matrix $\mathbf{W}$ (see section 3.1 and equation 3.10). The probabilities of representation $\boldsymbol{\psi}(t)$ can be calculated by using (3.10):

$$\begin{aligned}\boldsymbol{\psi}(t) = \frac{\mathbf{x}(t)}{\sum_{i=1}^{n} x_i(t)} &= \frac{a_1\boldsymbol{\varphi}_1\lambda_1^t + a_2\boldsymbol{\varphi}_2\lambda_2^t + \cdots}{\sum_{i=1}^{n} a_1\varphi_i^1\lambda_1^t + a_2\varphi_i^2\lambda_2^t + \cdots} \\ &= \frac{\boldsymbol{\varphi}_1 + \left(\frac{a_2}{a_1}\right)\left(\frac{\lambda_2}{\lambda_1}\right)^t\boldsymbol{\varphi}_2 + \cdots}{\sum_{i=1}^{n} \varphi_i^1 + \left(\frac{a_2}{a_1}\right)\left(\frac{\lambda_2}{\lambda_1}\right)^t\varphi_i^2 + \cdots}.\end{aligned} \tag{3.15}$$

(3.15) shows that the rate of relaxation of each system to its stationary state of probabilities of representation depends on the ratio of the second largest to the largest eigenvalue of $\mathbf{W}$, $\left(\frac{\lambda_2}{\lambda_1}\right)$, and on $\left(\frac{a_2}{a_1}\right)$, which in turn depends on the overlap of the system with $\boldsymbol{\varphi}_1$ at $t = 0$. For small systems determination of the temporal development and the stationary state is possible if all the terms in (3.15) can be calculated. For systems with large n it becomes impossible to calculate the complete transition matrix because N grows exponentially with n. We therefore investigate the behavior of both systems for small n ($n \leq 10$) and try to extrapolate our results to larger systems. We use a standard deviation of the noise distribution, $\theta = 0.6$, which allows self-organization for both models (compare figure 4). The presence of noise makes it possible for both systems to update to other states and explore the whole state space. For practical purposes we say that the system has reached its stationary state, in our simulations, when the iterated processes in (2.23) or (3.6) yield a distribution of states, $\boldsymbol{\psi}(t)$, for which $|m^+(t) - m^-(t)| \leq 0.01$ and $|\psi_k(t+1) - \psi_k(t)| < 0.01 \cdot \psi_k(t)$, $\forall k$. We record the system variables listed at the beginning of section 3 (figure 6), and the relaxation times, t_R, the number of timesteps from beginning to end of a simulation, when the system has reached its stationary state (figure 7).

Figure 6 shows initial stages of optimization runs for the ER and NN models. In figure 6A (ER model) and figure 6C (NN model), for $n = 5$, the system is started in a state with a maximum Hamming distance from $\boldsymbol{\xi}_1$ such that $d[\boldsymbol{\sigma}(t), \boldsymbol{\xi}_1] < d[\boldsymbol{\sigma}(t), -\boldsymbol{\xi}_1]$. In figure 6B (ER model) and figure 6D (NN model), for $n = 7$, the system is started in one of the next nearest-neighbor states of $\boldsymbol{\xi}_1$ so that $d[\boldsymbol{\sigma}(t), \boldsymbol{\xi}_1] = 1$. In all cases we observe an early

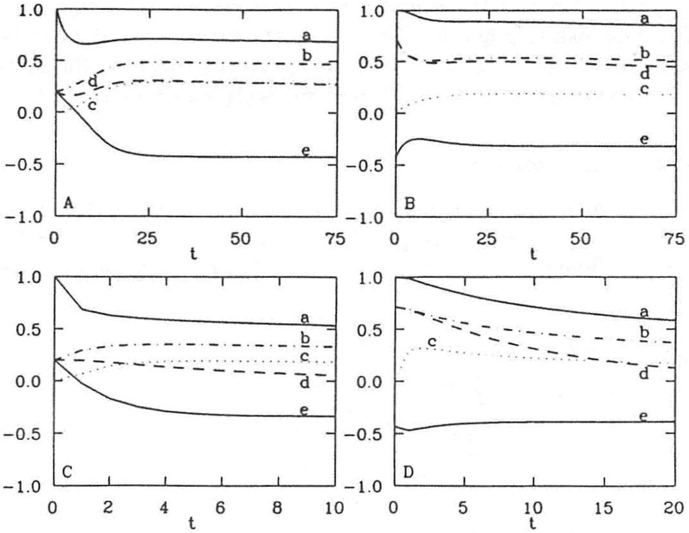

Figure 6: Initial stages of the relaxation process in the ER and NN model for different state size, n, and different initial conditions $\sigma(0)$; $J = 1$, $\theta = 0.6$, and $p = 1$ for all simulations. The size of the systems, n, and the Hamming distance between the initial state and an energy minimum state, $d[\sigma(0), \xi_1]$ (see equation 2.26) vary in the simulations. The curves are recordings of the following system variables: (A) $P\left[\sigma(t) \in \{s_k^+\}\right]$ (3.1), (B) $m^+(t)$ (3.4), (C) the probability of finding the system in the state ξ_1 at time t, $P[\sigma(t) = \xi_1]$, (D) $m(t)$ (3.3), (E) $E(t)$ (3.5). Diagrams: **A:** ER model, $n = 5$, $d[\sigma(0), \xi_1] = 2$; **B:** ER model, $n = 7$, $d[\sigma(0), \xi_1] = 1$; **C:** NN model, $n = 5$, $d[\sigma(0), \xi_1] = 2$; **D:** NN model, $n = 7$, $d[\sigma(0), \xi_1] = 1$.

localization at the stored pattern, ξ_1, which causes attainment of maxima in curves c for $P[\sigma(t) = \xi_1]$, e.g. in figures 6A and 6D at early times. Further indicators of this early localization are maxima in curves b and curves d for $m^+(t)$ and $m(t)$, respectively, in figures 6A, B, D, and a minimum in curve e for $E(t)$ in figure 6d. Updating of a state in the ER model leads to states around this state: newly generated states closer to an energy minimum have a higher b_k than states which are further away from an energy minimum state and will have a selective advantage. Because the initial state is closer to ξ_1 than to $-\xi_1$, localization occurs first around the stored pattern. In the NN model updating is most likely to occur to the closest energy minimum

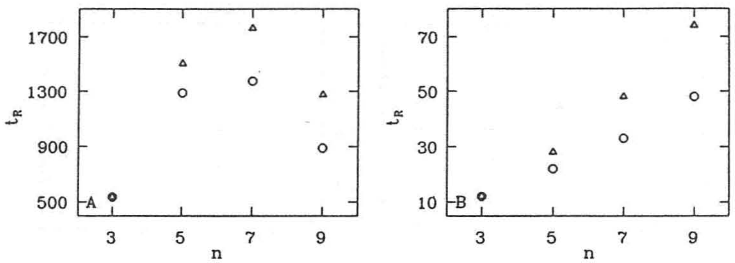

Figure 7: Relaxation times, t_R, for ER (diagram A) and NN model (diagram B) for relaxation from the initial state, $\sigma(0)$, to the stationary state vs. size of the system, n. $J = 1$, $\theta = 0.6$, and $p = 1$ for all simulations. The Hamming distance from the initial state to an energy minimum state is $d[\sigma(0), \xi_1] = (n-1)/2$ (symbol $\triangle$), and $d[\sigma(0), \xi_1] = 1$ (symbol $\bigcirc$). The relaxation criteria were: $|m^+(t) - m^-(t)| \leq 0.01$ and $|\psi_k(t+1) - \psi_k(t)| < 0.01 \cdot \psi_k(t)\ \forall k$.

state, and only noise allows a spreadout from this state. Therefore, early localization occurs in general much sooner for the NN than for the ER model. A further consequence of the updating algorithm in the ER model is a short-term spreadout effect before early localization. In the earliest phases of a simulation the influence of the (transition term)s may be much larger than that of the (quality term)s. As soon as states with high enough replication rate coefficients, b_k, are sufficiently represented, the influence of the noise controlled (transition term)s becomes secondary, which leads to the early localization around the stored pattern. Minima in curves a, b, and d in figures 6A and 6B for $P\left[\sigma(t) \in \{\mathrm{s}_k^+\}\right]$, $m^+(t)$ and $m(t)$, respectively, as well as the maximum in curve e for $E(t)$ in figure 6B are indicative of this effect which does not occur in the NN model.

After the early localization at the stored pattern, both systems relax to their stationary states. The relaxation to the stationary state occurs on a larger timescale than the early localization. At the stationary state the systems are equally localized at the two energy minima. We can see this in the long-term behavior of both systems: as $t \to \infty$ we have $P\left[\sigma(t) \in \{\mathrm{s}_k^+\}\right] = 0.5$, $m(t) = 0$ (see figure 8) and $m^+(t) = m^-(t)$. However, the amount of localization around an energy minimum is different for the two systems (see section 3.6).

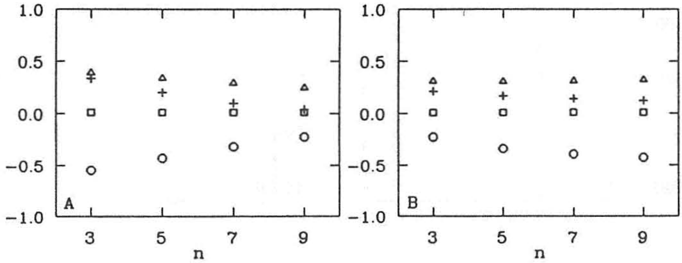

Figure 8: Stationary state values of $m^+(t)$ (symbol $\triangle$, s.a. (3.4)), the probability of finding the system in the state ξ_1 at time t, $P[\sigma(t) = \xi_1]$ (symbol $+$), $m(t)$ (symbol $\square$, s.a. (3.3)), and $E(t)$ (symbol $\bigcirc$, s.a. (3.5)) for simulations with the ER (diagram A) and NN model (diagram B); $J = 1$, $\theta = 0.6$, and $p = 1$ for all simulations.

Relaxation times, t_R, are in general different for the ER and NN models (compare figure 7). At least for small systems the NN relaxes significantly faster to its stationary state than the ER model at high θ-values, $\theta > 0.5$. For the NN model relaxation times seem to grow exponentially with n. In the ER model, relaxation times for a certain θ reach a maximum and decline after that as the system size increases. For small system sizes, relaxation times increase with n, as they do in the NN case. The decline of relaxation times with increasing system size in the ER model is a result of the approach to the maximum state size, $n_{\max}$, for a given θ. Increases in relaxation times in the NN case are caused by a growth in the input signals to a single neuron, $\widetilde{h}_i$, which grows with n for states with the same Hamming distance to the stored pattern. The result is a faster relaxation to the next energy minimum and a slower transition process to other states with the same energy (compare sections 3.2 and 3.3).

3.6 Use of ER and NN models as optimization algorithms

In section 2 we have defined one of many possible energy functions, which assigns energy values E_k to every possible state (see e.g. figure 2 and figure 10). The ER as well as the NN model can be interpreted as algorithms used to minimize this energy function, i.e. find states with low energies. Optimization algorithms have to do exactly this, therefore these two models can be understood as tools used to find good or optimal solutions to a combinatorial cost problem. If a different energy function has to be minimized, different sets of connections T_{ij} have to be used in the NN case to make a minimization of the new energy function possible. That this can be done has been shown in a variety of applications (see e.g. [23,25–27]). For the ER model, any inverse relation between E_k and b_k, as in (2.33), is acceptable. Which set of possible T_{ij} and b_k are to be preferred for an energy function is not always obvious; sometimes it is a matter of taste, and often it is a matter of luck to find a useful one. In this and the following section we discuss the performance of both models as optimization algorithms, if we use the energy function defined in section 2.1. We describe performance not only in terms of statistical average of the overall energy, $E(t)$, but also in terms of probability to be in an energy minimum state, $P\left[\boldsymbol{\sigma}(t)=\boldsymbol{\xi}_1\right]$, and in terms of overlap with this state, $m^+(t)$ and $m(t)$, since good performance has different meanings for different optimization problems; e.g., one could imagine a problem having the same energy function as the one we discussed in (2.8) with the exception that some of the possible states, which may have low energies, are physically meaningless: in this case we are interested rather in $P\left[\boldsymbol{\sigma}(t)=\boldsymbol{\xi}_1\right]$ than in $E(t)$ or $m^+(t)$.

We observe, that minimal values of $E(t)$ do not necessarily coincide with maximal values of $P\left[\boldsymbol{\sigma}(t)=\boldsymbol{\xi}_1\right]$, $m^+(t)$ or of $m(t)$ (see e.g. figure 6D) in the course of a simulation. Also stationary state values (figure 8) for $P\left[\boldsymbol{\sigma}(t)=\boldsymbol{\xi}_1\right]$, $m^+(t)$ or of $m(t)$ do not have to be equal to extremal values during a simulation (figure 9). Stationary state values are a valuable indicator for how well a system would perform if it were to be used as a black box for an optimization algorithm, with the answer to a problem being taken after system variables have approached asymptotic values. Furthermore, stationary state values provide us with good clues on the extent of localization around energy minima. Maximal values are important because efficient use of the algorithm also means intermediate recording of good solutions. By comparison of figures 8 and 9 we find that better solutions to the optimization problem are supplied by the algorithms with higher probability at an early time in the simulations, well before the stationary state is reached, than at the stationary state itself. For problems for which we know the general shape of the energy function, it may therefore be possible to stop an optimization algorithm before the stationary state is reached. But a general rule cannot be given as to when to stop the relaxation process in order to get good solutions with a high probability.

Both systems yield more extremal intermediate values the closer the process is started to one of the energy minima (compare figures 9A with figure 9B and figure 9C with 9D). This is to be expected (compare section 3.5) since early localization around the stored pattern is more likely the smaller the Hamming distance of the initial state to an energy minimum, $d[\boldsymbol{\sigma}(0),\boldsymbol{\xi}_1]$.

The ER model gives better performance as an optimization algorithm for small n, but the NN model outperforms the ER model at large n, for the chosen set of parameters and our range of n (compare figure 9A with 9C and figure 9B with 9D): at $n=3$ maximal intermediate values of $P\left[\boldsymbol{\sigma}(t)=\boldsymbol{\xi}_1\right]$, $m(t)$, and $m^+(t)$ are higher for the ER model than for the NN model, and minimal intermediate values of $E(t)$ are lower in the ER model than in the

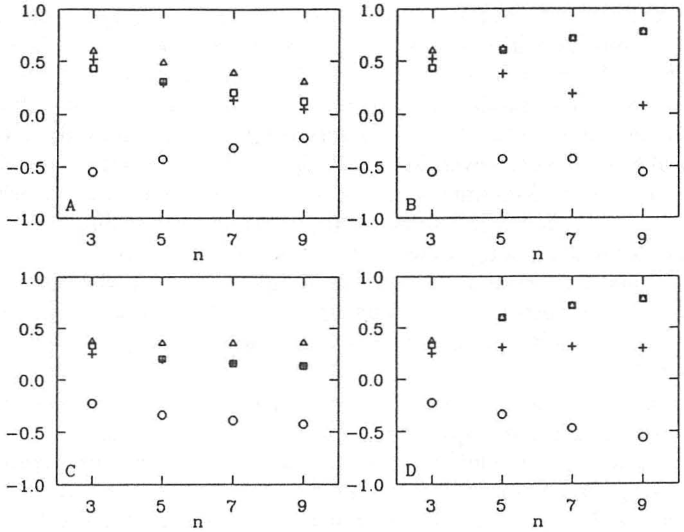

Figure 9: Maximal values of $P[\sigma(t)=\xi_1]$, the probability of finding the system in the state ξ_1 at time t, (symbol $+$), $m(t)$ (symbol $\square$, s.a. (3.3)), and $m^+(t)$ (symbol Δ, s.a. (3.4)), and minimal values of $E(t)$ (symbol $\bigcirc$) for simulations with the ER and NN model for different state size, n, and different initial conditions, $\sigma(0)$; $J=1$, $\theta=0.6$, and $p=1$ for all simulations. Diagrams: (A) ER model, $d[\sigma(0),\xi_1]=(n-1)/2$; (B) ER model, $d[\sigma(0),\xi_1]=1$; (C) NN model, $d[\sigma(0),\xi_1]=(n-1)/2$; (D) NN model, $d[\sigma(0),\xi_1]=1$.

NN model; at $n=5$ extremal intermediate values of these observables are approximately the same for both models; for $n \geq 7$ the situation is reversed with higher maximal intermediate values of $P[\sigma(t)=\xi_1]$, $m(t)$, and $m^+(t)$ for the NN model than for the ER model, and with lower minimal intermediate values of $E(t)$ for the NN model than for the ER model. As the system size gets bigger, the system size approaches the maximum system size in the case of the ER model, $n \to n_{\max}$, and the width of localization around an energy minimum state grows. This effect is clearly visible well below the theoretical value of $n_{\max}$ in our simulations, e.g. see figure 8A and figures 9A, B; for the parameter sets used in these figures 3.12 gives a maximum system size, $n_{\max}$, of 19. This means that the noise level has to be adjusted in the ER model, not only for different energy functions but also for different system sizes. The NN model on the other hand gives the same statistical average of the overlap, $m^+(t)$, even as the size of a problem increases (see figures 8B

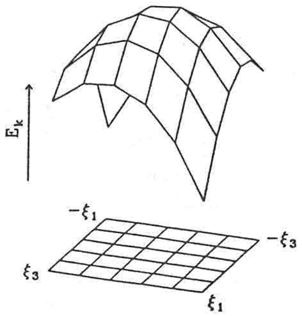

Figure 10: Energy surface, E_k, over a two-dimensional cut through the n-dimensional hypercube (s.a. explanations for figure 2) for $n = 10$ and three stored patterns, $p = 3$ with $\xi_1 = \xi_2$ and $d(\xi_1, \xi_3) = n/2$. The two corners under the deeper energy minima correspond to $\pm\xi_1$, and the other two corners correspond to $\pm\xi_3$.

and 4); but $P[\sigma(t) = \xi_1]$ decreases with n, making adjustment of the noise level necessary if $P[\sigma(t) = \xi_1]$ has to be maximized.

A dynamic internal noise regulation, as it is employed in simulated annealing [33], may therefore be useful for many optimization problems: as the probability of representation of good solutions increases the noise level is decreased. This approach leads to a better localization around energy minima states. Problems with this method can occur if the noise reduction occurs too rapidly; the system might then get trapped in local minima for all practical purposes, because transitions to other states become very unlikely; optimal solutions may not be found.

3.7 Energy surfaces with local minima

In order to compare the two models in their performance on an energy landscape with local minima, we choose a Hopfield Hamiltonian with three stored patterns, $p = 3$, for the energy function. We use systems with an even number of microstates and choose the patterns so that $\xi_1 = \xi_2$ and $d(\xi_1, \xi_3) = n/2$; spurious states are not created because ξ_1 and ξ_3 are exactly orthogonal. The energy minima at $\pm\xi_1$ are exactly twice as deep as the local energy minima at $\pm\xi_3$ (see figure 10). In this case critical noise thresholds, θ_{th}, are no longer the same as the ones pointed out in section 3.3; also relaxation times and stationary states must be different in cases of energy functions with multiple minima compared to cases of energy functions with only two energy minima of equal depth, which are on exactly opposite corners of the n-dimensional hypercube, as discussed above.

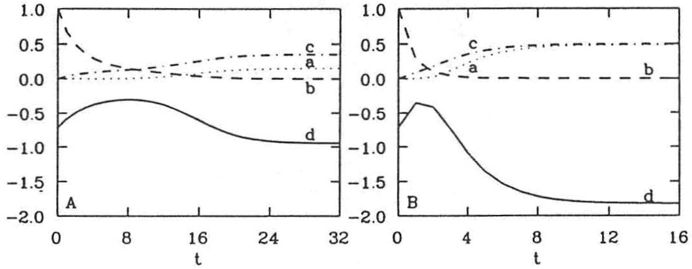

Figure 11: Initial stages of the relaxation process in the ER (diagram A) and NN model (diagram B) for an energy landscape with four minima; two minima are twice as deep as the other two local minima ($p = 3$, $\xi_1 = \xi_2$ and $d(\xi_1, \xi_3) = n/2$. $J = 1$, $\theta = 0.6$, $n = 8$. Initially the systems are in one of the local energy minimum states, $\sigma(0) = \xi_3$. The curves are recordings of the following system variables: (A) the probability of finding the system in the state ξ_1 at time t, $P[\sigma(t) = \xi_1]$, (B) the probability of finding the system in the state ξ_3 at time t, $P[\sigma(t) = \xi_3]$, (D) $m^+(t)$ (s.a. (3.4)), (D) $E(t)$ (s.a. (3.5)).

In our simulations we always started the systems in one of the local minima. Initial stages of these optimization runs are recorded in figure 11. We find again that it takes the ER model longer to migrate from the local to the deepest minima states. This can be seen in curves d of figure 11A (ER model) and figure 11B (NN model). The maxima in curves d reflect the high energy values of states between the local and the deepest minima (see figure 10) over which the systems migrate during relaxation. Changes in $P[\sigma(t) = \xi_1]$ (curves a), $P[\sigma(t) = \xi_3]$ (curves b), and $m^+(t)$ (curves c) are slower in the ER than in the NN case. Accordingly, this results in longer relaxation times, t_R, for the ER model than for the NN model (compare figure 12A and 12B), although the relaxation times are now of the same order of magnitude for $n \leq 10$ in the two models, compared to differences of two orders of magnitude in the t_R in the cases discussed in section 3.6 (compare also figure 7). Also, relaxation times increase only linearly with n for the ER model before they start decreasing, whereas they again increase exponentially for the NN model.

The performance in terms of finding best solutions or minimizing $E(t)$ is still higher for the NN than for the ER model. Stationary state values of $m^+(t)$, $P[\sigma(t) = \xi_1]$, $P[\sigma(t) = \xi_3]$, and $E(t)$ are shown in figure 13. For both models the low energy minimum states $\pm\xi_3$ are practically not represented

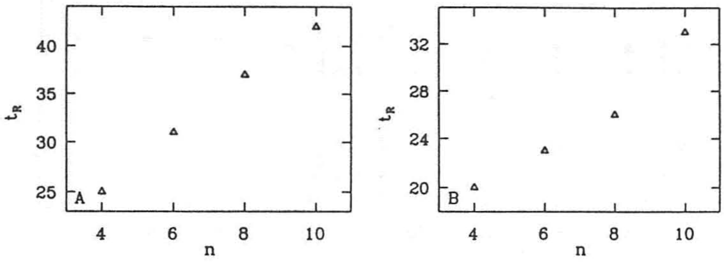

Figure 12: Relaxation times, t_R, for ER (diagram A) and NN models (diagram B) for relaxation from the initial state, $\sigma(0) = \xi_3$, to the stationary state. $J = 1$, $\theta = 0.6$, and $p = 3$ ($\xi_1 = \xi_2$ and $d(\xi_1, \xi_3) = n/2$ for all simulations. The relaxation criterion is $|\psi_k(t+1) - \psi_k(t)| < 0.01 \cdot \psi_k(t)\ \forall k$.

at the stationary state, $P[\sigma(t) = \xi_3] \simeq 0 \ll P[\sigma(t) = \xi_1]$, but the NN model gives significantly better values for $P[\sigma(t) = \xi_1]$ and $E(t)$ than the ER model. For the NN model at system sizes $n \geq 6$ and with the used set of parameters, the system localizes almost completely at the two deeper energy minimum states, $\pm\xi_1$, at the stationary state. This can be seen in figure 13 for system sizes $n \geq 6$ for which $P[\sigma(t) = \xi_1]$ and $m^+(t)$ are close to the maximum values of 0.5 for the stationary state.

In general, it is easier for both systems to migrate from one energy minimum state to another if low energy states exist between them. Therefore the relaxation process to the stationary states of these systems is shorter in the case of optimization on an energy surface with local minima than in the case of optimization on energy surfaces with few energy minima which are isolated and far away from each other.

4. Summary

We have reviewed the Eigen ER model and the Little–Hopfield NN, and we have outlined a general description of optimization algorithms for complex combinatorial problems which contains both models as special cases; a formulation in terms of spin systems is used for both models to point out similarities and differences. These models, as well as their natural counterparts, show self-organizing behavior, which is defined as localization of the system around certain states under given constraints. An energy function can be associated with the possible states of the systems in such a way that the systems tend toward states with low associated energies in the course of

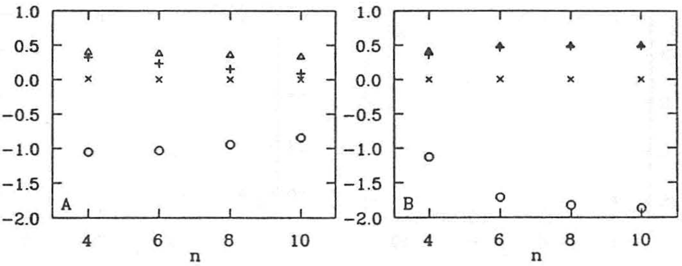

Figure 13: Stationary state values of $m^+(t)$ (symbol $\triangle$, s.a. (3.4)), the probability of finding the system in the state ξ_1 at time t, $P\left[\sigma(t) = \xi_1\right]$ (symbol $+$), the probability of finding the system in the state ξ_3 at time t, $P\left[\sigma(t) = \xi_3\right]$ (symbol $\times$), and $E(t)$ (symbol $\bigcirc$, s.a. (3.5)) for simulations with the ER (diagram A) and NN model (diagram B); $J = 1$, $\theta = 0.6$, and $p = 3$ ($\xi_1 = \xi_2$ and $d(\xi_1, \xi_3) = n/2$ for all simulations.

the localization process. This process can be thought of as an optimization process and the models can be used as optimization algorithms for complex combinatorial problems.

The self-organization process is governed by transition frequency matrices, which are derived for the two models: they are nonnegative and in the case of the NN that matrix is stochastic. Simple Hopfield Hamiltonians, generated by the Hebbian learning rule, are used for our comparison of the models.

By defining input functions and updating rules for the single microstates of a system, specific interactions between the microstates are imposed on the models. The NN is a highly interactive system in which the types of all microstates at one time have an influence on the updating of one single microstate. Interactions inside a macrostate of the ER model are only defined from every microstate to itself. In the ER model, (transition term)s between two states are independent of the energy of the two states and depend only on the Hamming distance between the two states. Selection pressure drives the system toward mixtures of states with high average (quality term)s or with lower associated energy values. In the NN, no selective advantage is given for any state. Updating occurs with higher probability to states with equal or lower energy than to states with higher energy.

Noise is an essential factor in both systems if freezing at suboptimal state-solutions is to be prevented. Although the NN finds local minima even in the absence of noise, such states may, for example, correspond to bad solutions of an optimization problem with many local minima. For the ER model, on the other hand, movement through state space is only possible in the presence of noise. Therefore noise is generally a positive feature of these systems, enabling them to reach energy minimum states, and plays the same role as temperature in a statistical system.

For both systems, noise thresholds can be defined. Localization is only possible below these thresholds. In the ER case the system size is limited by the amount of noise present; bigger systems have lower thresholds. In the NN the noise threshold is constant for all sizes of the system at a certain spin interaction parameter J. For example, with the Hopfield Hamiltonian as energy function and the noise level at $\theta = 0.6$, the ER system can operate at higher noise levels than the NN only for small system sizes, i.e. $n < 8$. In order for the ER to localize at bigger sizes of the system, a very small θ has to be used that may not be available for certain applications. Depending on J, a maximum system size can be given for the ER model up to which self-organization occurs in the ER model even if self-organization is not possible in the NN case. This effect may be most pronounced at regimes of low J-values and low noise levels. Whether the overlap with good sequences is higher in the NN than in the ER model depends strongly on the system size.

Higher noise levels make it possible to relax to the stationary state faster, but for noise levels above the threshold a random updating process follows. The environment or the requirement of usefulness of the optimization algorithm determines an effective noise threshold which is lower than the critical threshold. Both models operate most efficiently at noise levels, just below this effective noise threshold. At these levels good solutions are found quickest and with a sufficiently high probability.

In our simulations we find relaxation times up to two orders of magnitude larger for the ER than for the NN model at smaller system sizes, i.e. $n \leq 10$, but they are expected to grow exponentially with n for constant θ in the NN case. Relaxation very often occurs in two phases: a quick relaxation to nearby energy minima and a slow relaxation to the stationary state (see figure 6, and compare figures 8 and 9). Thus, expectation values for an overlap with good solutions may be higher at a very early phase in the relaxation process than in the stationary state. This suggests a dynamically self-regulating noise level as in simulated annealing to achieve faster and better localization around energy minima states.

Optimization algorithms are always limited by the way a problem is formulated for the algorithms: e.g., one numerical problem might be solved by using one of several computer programs which again might be run on one of several different computers. Although the numerical problem stays the same, the implementation of the problem, the choice of a specific program and a specific computer on which it is executed, might be quite different. This can well result in a different computational effort to solve the problem and might possibly lead to answers of different precision. We have used the same energy functions for two different algorithms. Our formulation of the (quality term)s and the (transition term)s resulted in better performance of the NN in big parts of the parameter space. Similarly, we imagine different adaptations of the same problem for the same model, resulting in different performances of the model. For examples we refer the reader to various implementations of the traveling salesperson problem on an ER model [24,34,35] and to implementations of the same problem on NN [23,25,26]. The ER versions do not operate with the mutation operator for single microstates which we discussed (compare equation 2.25) but with operators which may change a whole range of microstates in the course of one updating step. The NN implementations work with continuous state neuron NN [23,25,26] in which microstates can take on real values between zero and one. Some of the ER algorithms may not find solutions to the traveling salesperson problem very quickly [24,34], but all of them always give valid tours. The NN algorithm as described in [25,26] may give illegal tours and is more likely to do so as the size of the problem increases [36]. The implementation on a NN as described in [23], on the other hand, does not produce impossible tours.

In this and other contexts, capacities of hybrid ER–NN algorithms should be investigated. We believe that algorithms which use ER and NN models at different times of an optimization process, as well as models which combine features of both algorithms, may perform better in many optimization tasks than a single algorithm alone.

Acknowledgments

This work was supported in part by the National Science Foundation and the Air Force Office of Scientific Research. E.P. acknowledges support by the Exportakademie der Österreichischen Bundeswirtschaftskammer and the Österreichischen Bundesministerium für Wissenschaft und Forschung.

References

[1] M. Eigen, "Self-organization of matter and the evolution of biological macromolecules," *Naturwissenschaften*, **58** (1971) 465–523.

[2] I. Leuthäusser, "An exact correspondence between Eigen's evolution model and a two-dimensional Ising system," *Journal of Chemical Physics*, **84**(3) (1986) 1884–1885.

[3] I. Leuthäusser, "Statistical mechanics of Eigen's evolution model," *Journal of Statistical Physics*, **48**(1/2) (1987) 343–360.

[4] W. Fontana and P. Schuster, "A computer model of evolutionary optimization," *Biophysical Chemistry*, **26** (1987) 123–147.

[5] M. Eigen and P. Schuster, "The hypercycle, a principle of natural self-organization, Part A: Emergence of the hypercycle," *Naturwissenschaften*, **64** (1977) 541–565.

[6] M. Eigen and P. Schuster, "Stages of emerging life – Five principles of early evolution," *Journal of Molecular Evolution*, **19** (1982) 47–61.

[7] M. Eigen, J. McCaskill, and P. Schuster, "Molecular quasi-species," *Journal of Physical Chemistry*, **92** (1988) 6881–6891.

[8] L. Demetrius, P. Schuster, and K. Sigmund, "Polynucleotide evolution and branching processes," *Bulletin of Mathematical Biology*, **47**(2) (1985) 239–262.

[9] M. Eigen and P. Schuster, "The hypercycle, a principle of natural self-organization, Part B: The abstract hypercycle," *Naturwissenschaften*, **65** (1978) 7–41.

[10] M. Eigen and P. Schuster, "The hypercycle, a principle of natural self-organization, Part C: The realistic hypercycle," *Naturwissenschaften*, **65** (1978) 341–369.

[11] J. Swetina and P. Schuster, "Self-replication with errors — A model for polynucleotide replication," *Biophysical Chemistry*, **16** (1982) 329–345.

[12] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences USA*, **79** (1982) 2554–2558.

[13] J.J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proceedings of the National Academy of Sciences USA*, **81** (1984) 3088–3092.

[14] W.A. Little, "The existence of persistent states in the brain," *Mathematical Biosciences*, **19** (1974) 101–120.

[15] W.A. Little and G.L. Shaw, "Analytic study of the memory storage capacity of a neural network," *Mathematical Biosciences*, **39** (1978) 281–290.

[16] D.J. Amit, H. Gutfreund, and H. Sompolinsky, "Spin-glass models of neural networks," *Physical Review A*, **32**(2) (1985) 1007–1018.

[17] D.J. Amit, H. Gutfreund, and H. Sompolinsky, "Storing infinite numbers of patterns in a spin-glass model of neural networks," *Physical Review Letters*, **55**(14) (1985) 1530–1533.

[18] D.J. Amit, H. Gutfreund, and H. Sompolinsky, "Information storage in neural networks with low levels of activity," *Physical Review A*, **35**(5) (1987) 2293–2303.

[19] D.J. Amit, H. Gutfreund, and H. Sompolinsky, "Statistical mechanics of neural networks near saturation," *Annals of Physics*, **173** (1987) 30–67.

[20] M. Cottrell, "Stability and attractivity in associative memory networks," *Biological Cybernetics*, **58** (1988) 129–139.

[21] J.D. Keeler, E.E. Pichler, and J. Ross, "Noise in neural networks: Thresholds, hysteresis, and neuromodulation of signal-to-noise," *Proceedings of the National Academy of Sciences USA*, **86** (1989) 1712–1716.

[22] D.G. Bounds, "New optimization methods from physics and biology," *Nature*, **329** (1987) 215–219.

[23] R. Durbin and D. Willshaw, "An analogue approach to the travelling salesman problem using an elastic net method," *Nature*, **326** (1987) 689–691.

[24] W. Fontana, *Ein Computermodell der Evolutionären Optimierung* (Dissertation, University of Vienna, Austria, 1987).

[25] J.J. Hopfield and D.W. Tank, "'Neural' computation of decisions in optimization problems," *Biological Cybernetics*, **52** (1985) 141–152.

[26] J.J. Hopfield and D.W. Tank, "Computing with neural circuits: A model," *Science*, **233** (1986) 625–633.

[27] W. Jeffrey and R. Rosner, "Optimization algorithms: Simulated annealing and neural network processing," *The Astrophysical Journal*, **310** (1986) 473–481.

[28] P. Peretto, "Collective properties of neural networks: A statistical physics approach," *Biological Cybernetics*, **50** (1984) 51–62.

[29] B.L. Jones, R.H. Enns, and S.S. Ragnekar, "On the theory of selection of coupled macromolecular systems," *Bulletin of Mathematical Biology*, **38** (1976) 15–28.

[30] C.J. Thompson and J.L. McBride, "On Eigens's theory of the self-organization of matter and the evolution of biological macromolecules," *Mathematical Biosciences*, **21** (1974) 127–142.

[31] F.R. Gantmacher, *The Theory of Matrices*, Vols. 1 and 2 (Chelsea Publishing Company, New York, 1959).

[32] A. García-Tejedor, F. Morán, and F. Montero, "Influence of the hypercyclic organization on the error threshold," *Journal of Theoretical Biology*, **127** (1987) 393–402.

[33] S. Kirkpatrick, C.D. Gelatt Jr., and M.P. Vecchi, "Optimization by simulated annealing," *Science*, **220** (1983) 671–680.

[34] W. Fontana, W. Schnabl, and P. Schuster, "Physical aspects of evolutionary optimization and adaptation," *Physical Review A*, **40**(6) (1989) 3301–3321.

[35] D.B. Fogel, "An evolutionary approach to the traveling salesman problem," *Biological Cybernetics*, **60** (1988) 139–144.

[36] G.V. Wilson and G.S. Pawley, "On the stability of the travelling salesman problem algorithm of Hopfield and Tank," *Biological Cybernetics*, **58** (1988) 63–70.