Probabilistic Approach to a Randomly Diluted Neural Network with Variable Activity

A Thesis Presented

by

Stefan Wolfgang Großkinsky

to

The Graduate School in Partial Fulfillment of the Requirements for the

Degree of

Master of Arts

in

Physics

State University of New York

at Stony Brook

August 1998

State University of New York

at Stony Brook

The Graduate School

Stefan Wolfgang Großkinsky

We, the thesis committee for the above candidate for the

Master's degree, hereby recommend

acceptance of this thesis.

Professor Robert Shrock (Advisor)
Institute for Theoretical Physics

Professor John Smith (Chair)
Institute for Theoretical Physics

Professor James Lukens
Department of Physics and Astronomy

This thesis is accepted by the Graduate School

Abstract of the Thesis

Probabilistic Approach to a Randomly Diluted Neural Network with Variable Activity

by

Stefan Wolfgang Großkinsky

Master of Arts
in
Physics

State University of New York
at Stony Brook

1998

The subject of study is a neural network model with neurons $S_i \in \{0, 1\}$, randomly diluted synapses and variable pattern activity. We look at the system with parallel updating using a probabilistic approach to solve the one step dynamics with one condensed pattern. We derive a dynamical version of the storage capacity and the mutual information content for arbitrary network states, and study the effect of noisy updating, giving an analytic result for the critical temperature. The analysis gives new insight in restrictions occuring during the retrieval process which is applied to the choice of proper threshold functions. The description is applicable to the whole retrieval process only in the limit of strong dilution. Using the dynamical results, we also consider the network in equilibrium and explain qualitatively the numerical solution of the fixed point equations.

Contents

List of Figures

Acknowledgments

I would like to thank Professor Shrock for advising me in many respects and helping me with useful discussions about statistical mechanics. He gave me the choice to study a subject of my interest and the work with him was very inspiring and motivated me to continue research in this area.

Special thanks to Miguel Maravall who supported my work by giving me access to his huge library on neural networks. His general suggestions and advice made it significantly easier for me to find interesting references and areas of study.

This work was partially supported by the german academic exchange program DAAD and the Swartz Foundation.

1 Introduction

Neural networks were originally introduced to model brain function but today artificial networks are also used for data processing. The step from the brain to a neural network model contains many simplifications and idealizations of which we try to summarize the most important ones in the following. After that we give a short overview of the studies on neural networks in physics and put the present work into context.

1.1 From the brain to a neural network model

The human brain consists of about 10^{11} nerve cells, the neurons, that communicate with each other by consecutive pulses of electrochemical potentials called action potentials with variable firing rate. An action potential of a presynaptic neuron is transmitted via the interneural connection, the synapse, causing the socalled postsynaptic potential (PSP) in the postsynaptic neuron. On average each neuron receives the signals of about 10^4 other neurons and integrates the sum of all PSPs over a period of time of the order of a few milliseconds. If this sum exceeds a certain level, the neural threshold, the neuron fires an action potential. The higher the stimulus of the neuron the higher is the firing rate of a series of action potentials. If a PSP is negative the contribution from the corresponding synapse inhibits firing of the postsynaptic neuron and is called inhibitory in contrast to an excitatory synapse. The conductance of a synapse is not fixed and there is evidence that learning is the result of a proper change in the synaptic strengths. According to the hypothesis of Hebb [1] a synapse

is strengthened if the pre- and postsynaptic neuron fire simultaneously.

To model the behavior of a system of N neurons one chooses a subset $M \subset \mathbf{R}$ that describes all possible firing states, including inactivity, of a single neuron and ascribes a number $S_i \in M$ to each neuron $i \in \{1, \ldots, N\}$. The subset M depends on the degree of simplification and in most cases it is an interval $(-q, q) \subset \mathbf{R}$ (graded response neurons) or even simpler a discrete subset like $\{0, 1\}$ (McCulloch-Pitts neurons). The strength of the synapse between neuron i and j is described by a number $J_{ij} \in \mathbf{R}$ where negative numbers usually apply for inhibitory and positive numbers for excitatory synapses. Therefore the PSP at neuron i caused by neuron j in state S_j is $J_{ij}S_j$. If the neurons are not connected with each other J_{ij} is just 0, so that the whole architecture of the network is carried by the connection matrix. There are two different kinds of architectures which were main subject of study in the literature:

- In a feedforward network the neurons are gathered in layers and signal processing occurs only forward from layer to layer. This is the preferred structure of most artificial networks and as there are often much less output than input neurons it is used for the classification of noisy data. But there are also parts in the sensory system that have a layered structure.

- In a recurrent or attractor neural network there is no underlying structure in the synaptic matrix and the neurons are recurrently connected. They form a dynamical system that eventually reaches a fixed configuration. This type models the associative recall of memorized network configurations called patterns from a similar stimulus. It is an idealization of parts of the brain like specific areas of neocortex.

The process of integration of all PSPs in the neuron and activating the action potential is called updating. It takes place at different times for different neurons and this complicated process can hardly be modelled in its full complexity. Hence one discretizes the time into time steps $t \in \mathbf{N_0}$ and idealizes that the whole update process takes place instantly in time. Therefore the new state of the neuron $S_i(t+1) = f(h_i(t) - Q_i)$ is simply a function of the sum of postsynaptic potentials $h_i(t) = \sum_{i=1}^{N} J_{ij} S_j(t)$ and the neural threshold Q_i at timestep t. The function f is called response or activation function and is usually step function. By introducing noise in the update process the function becomes sigmoidal. As far as modelling is concerned there are two possibilities to organize the update processes of all neurons:

- Parallel updating: All neurons are updated synchroneously in the same time step so that the new states of all neurons at time $t+1$ depend on the states at time t.

- Asynchronous updating: One neuron is randomly chosen and updated at each time step.

There are various possibilities to store patterns in a network differing on how and when to change the synaptic efficacies. For simplicity the retrieval process is often separated from the learning phase leading to a fixed synaptic matrix during retrieval. For further simplification the stored patterns are drawn at random constituting a quenched disorder during the retrieval process. More realistic systems where the synaptic modification takes place during retrieval are also studied but we will concentrate on the simple case above.

1.2 Studies of neural network models in physics

Particularly suitable for the study with methods of physics are recurrent networks where the patterns constitute a quenched disorder during the retrieval process. In the model proposed by Hopfield [5, 6] and Little [7] the neurons have only two possible states $S_i = \pm 1$ and the synaptic efficacies are determined according to Hebb's hypothesis $J_{ij} = \frac{1}{N}\sum_\nu \xi_i^\nu \xi_j^\nu (1 - \delta_{ij})$ yielding a symmetric connectivity matrix. Using asynchronous updating and a noisy response function like Glauber dynamics (see section 2.1 and [8]) with noise parameter β the network can be seen as an Ising spin glass with the Hamiltonian

$$H = -\frac{1}{2}\sum_{i,j} J_{ij} S_i S_j + \sum_i Q_i S_i.$$

Therefore the system was studied by means of equilibrium statistical mechanics and the attractors of the dynamics are described by the minima of the free energy function. This was first done by Amit et. al. [9, 10] and properties like maximal storage capacity, information content and temperature phase diagram were studied. But the Hopfield model contains some unrealistic simplifications:

- The probability that the elements of the stored patterns are active is $\frac{1}{2}$ but the mean firing rate in the brain is much smaller and in data processing the information is often carried only by a small part of the signal with a large background.

- In the Hopfield network the system is fully connected but connectivity in the brain is much smaller on average and the symmetry of the synaptic matrix is surely unrealistic.

- The Hopfield dynamics are symmetric with respect to $S_i \to -S_i$ and the different firing rates of a neuron, almost continuously valued, are described with only two states.

To overcome these weaknesses studies were extended to patterns with variable activity [11]. A more realistic representation of the active and inactive state of a neuron with $S_i \in \{0, 1\}$ was introduced. Such a model for variable activity was proposed by Tsodyks et al. [12] and further studied in [13] and it was found that it resembles the estimate for the maximal storage capacity for small pattern activities obtained by Gardner [14]. Accounting for realistic connectivities the symmetry of the synaptic matrix is lost, an essential ingredient for the existence of a Lyapunov energy function. In this case a dynamical approach was made and most studies were done using continuous time dynamics [15, 17, 18]. The effects of different kinds of dilution were compared in [19]. For parallel updating the one step dynamics can be solved exactly with a probabilistic dynamical approach. This was applied to diluted networks with graded response neurons in [20, 21]. More recently attention was shifted towards proper choices of neural thresholds in order to optimize the retrieval properties of the network [22, 23, 24].

In this work we study the model introduced by Tsodyks with random dilution and arbitrary pattern activity using a probabilistic approach. We model the retrieval process of a single pattern in the limit of strong dilution and derive dynamical properties of the network by analyzing the one step dynamical equations. We study the storage capacity and corresponding thresholds for noiseless dynamics in chapter 3. In chapter 4 we look at the effect of noisy

dynamics making an expansion for low temperatures and deriving the critical temperature. Former studies describe these properties in equilibrium which is of course of great interest. However, the analysis given here provides some insight into the constraints on network performance during retrieval which is especially interesting for the choice of proper threshold functions. The essential considerations to dynamical properties are confirmed by simulations with relatively high pattern activities to avoid finite size effects (see appendix A.1). Nevertheless the analysis is valid over the whole range of pattern activity in contrast to former studies which were often done in the limit of low activity. In chapter 5 we try to apply some of the dynamical results to the equilibrium situation and explain the solutions to the fixed point equations depending on the choice of the threshold. In the next section we give an exact definition and description of the model and address the validity of the description which is limited due to feedback loops.

2 Model and description

2.1 Dynamics and statics of the network

We study a neural network model consisting of N binary neurons $S_i \in \{0,1\}$, $i = 1, \ldots, N$. As usual we use discrete time steps and label them with integer numbers $t \in \mathbf{N_0}$. The state of the network at time t is denoted by $\mathbf{S}(t) = (S_1(t), \ldots, S_N(t))$. The neurons are connected with each other via the synaptic couplings or interaction constants $J_{ij} \in \mathbf{R}$. The value of neuron i in the next time step $S_i(t+1)$ is a function of the local field

$$h_{iN}(t) \;=\; \sum_{j=1}^{N} J_{ij} S_j(t).$$

For a deterministic update rule we use the step function $\Theta(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases}$ as activation function:

$$S_i(t+1) \;=\; \Theta(h_{iN}(t) - Q) \tag{1}$$

Here Q is the synaptic threshold or external field which is taken to be spatially uniform but can be time dependent. With this rule the new state of the network $\mathbf{S}(t+1)$ is completely determined by the old one $\mathbf{S}(t)$.

To model the randomness inherent in the time evolution of a real network we use a probabilistic update rule with the noisy response function from Glauber dynamics $g(x,\beta) = \frac{1}{1+\exp(-2\beta x)}$, where β is the noise parameter:

$$\begin{aligned} P(S_i(t+1) = 1) \;&=\; g(h_i(t) - Q, \beta) \\ P(S_i(t+1) = 0) \;&=\; 1 - g(h_i(t) - Q, \beta) \end{aligned} \tag{2}$$

Due to the equivalence with spin models one applies the term temperature for $T = \frac{1}{\beta}$ to quantify the amount of noise in the system. For $T \to 0$ we return to

the threshold step function, i.e. $g(x,\beta) \to \Theta(x)$ for $\beta \to \infty$. The neurons are updated in parallel so that $h_i(t)$ is computed for all i in every time step and the values $S_i(t+1)$ are assigned according to the update rule.

There are p patterns $\boldsymbol{\xi^\nu}$, $\nu = 1, \cdots, p$, with elements $\xi_i^\nu \in \{0,1\}$, $i = 1, \cdots, N$. The ξ_i^ν are independent, identically distributed random variables (IIDRV) with the distribution function $P(\xi_i^\nu) = a\delta(\xi_i^\nu - 1) + (1-a)\delta(\xi_i^\nu)$. Therefore the activity of every pattern

$$a_N^\nu \;=\; \frac{1}{N}\sum_{i=1}^{N} \xi_i^\nu$$

is a gaussian random variable with mean a and variance $\frac{1}{N}a(1-a)$. The elements are weakly correlated with $< \xi_i^\nu >= a$ and $< \xi_i^\nu \xi_j^\mu >= a^2 + a(1-a)\delta_{\nu\mu}\delta_{ij}$ for all $\nu, \mu \in \{1, \cdots, p\}$ and $i, j \in \{1, \ldots, N\}$.

The patterns are stored initially before retrieval so that the ξ_i^ν constitute a quenched disorder, i.e. fixed during the retrieval process. This can be achieved by choosing the couplings according to the Hebb rule (see chapter 1.2) but for patterns with variable activity former studies ([11, 12, 13]) showed that it is favorable for retrieval properties to use a modified version of this rule:

$$J_{ij} = \frac{c_{ij}(1-\delta_{ij})}{Nca(1-a)} \sum_{\nu=1}^{p} (\xi_i^\nu - a)(\xi_j^\nu - a) \tag{3}$$

The $c_{ij} \in \{0,1\}$ are also IIDRV with the distribution $P(c_{ij}) = c\delta(c_{ij} - 1) + (1-c)\delta(c_{ij})$ and account for the dilution of the network, i.e. the lack of a fraction of couplings. Therefore the number of connections per synapse over the number of neurons

$$C_N^i/N = \frac{1}{N}\sum_{j=1}^{N} c_{ij}$$

is a random variable with mean c and variance $\frac{1}{N}c(1-c)$ constituting a quenched, random dilution of the synapses.

The factor $ca(1-a)$ in the definition of the synapses turns out to be useful in order to keep the local field in the same order of magnitude over the whole range of pattern activity and network connectivity. Without this term we would have to adapt the threshold which means that it would depend very strongly on a and c.

2.2 Observables

To describe the network state with variable pattern activity accurately we have to use two observables. We define the following two overlaps between the network state $\mathbf{S}(t)$ and pattern $\boldsymbol{\xi^\nu}$:

$$m^\nu_{\uparrow N}(t) = \frac{1}{Na}\sum_{i=1}^{N}\xi_i^\nu S_i(t) \qquad m^\nu_{\downarrow N}(t) = \frac{1}{N(1-a)}\sum_{i=1}^{N}(\xi_i^\nu - 1)(S_i(t)-1)$$

If we take the network configuration $\mathbf{S}(t)$ at random from the same distribution function as the patterns but with activity

$$A_N(t) = \frac{1}{N}\sum_{i=1}^{N} S_i(t)$$

rather than a, $\mathbf{S}(t)$ and $\boldsymbol{\xi^\nu}$are uncorrelated for all $\nu \in \{1,\ldots,p\}$, i.e. $<\xi_i^\nu><S_i(t)>=<\xi_i^\nu S_i(t)>= aA_N(t)$. So the average overlaps between $\mathbf{S}(t)$ and any pattern ν are:

$$m^\nu_{\uparrow N}(t) = A \qquad m^\nu_{\downarrow N}(t) = 1-A$$

We assume that $\mathbf{S}(t)$ is uncorrelated with every pattern except one, the so-alled condensed pattern $\mu \in \{1,\ldots,p\}$. So the network has a macroscopic

overlap only with pattern μ, i.e. ${m_\uparrow}_N^\mu(t)$ and ${m_\downarrow}_N^\mu(t)$ are significantly bigger than their mean values for an uncorrelated network state. This situation occurs during the retrieval process of a single pattern.

With the assumption that the noncondensed patterns stay uncorrelated during the retrieval process the state of the network is now fully described by ${m_\uparrow}_N^\mu(t)$ and ${m_\downarrow}_N^\mu(t)$ and we take them as order parameters omitting the index μ from now on. They do not have the range $[0,1]$ because the activity a^μ of the condensed pattern μ is not exactly a. But to study the network theoretically we will use the limit $N, p \to \infty$ with a constant pattern loading $\alpha = \frac{p}{cN}$ and the observables become

$$m_\uparrow(t) = \lim_{N\to\infty} {m_\uparrow}_N(t) \qquad\qquad m_\downarrow(t) = \lim_{N\to\infty} {m_\downarrow}_N(t).$$

Due to self averaging we get for the pattern activity and the network connectivity:

$$a^\nu = \lim_{N\to\infty} a_N^\nu = a, \quad \nu = 1,\ldots,p \qquad\qquad \frac{C^i}{N} = \lim_{N\to\infty} \frac{C_N^i}{N} = c, \quad i = 1,\ldots,N$$

Therefore $m_\uparrow(t)$ and $m_\downarrow(t)$ have the proper range $[0,1]$ and we can easily get the current activity $A(t)$ of the network:

$$A(t) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} S_i(t) = a\, m_\uparrow(t) + (1-a)(1-m_\downarrow(t)) \tag{4}$$

This observable could also be used as an order parameter like it was done in former studies [12, 13, 15, 22], but the choice of $m_\uparrow$ and $m_\downarrow$ has some technical advantages.

The overlaps with a noncondensed pattern $\nu \neq \mu$ in the thermodynamic limit are ${m_\uparrow}^\nu(t) = A(t)$ and ${m_\downarrow}^\nu(t) = 1 - A(t)$. So for low activity the zero elements

of the patterns do not carry any appreciable information and the difference between the condensed and the noncondensed patterns is represented by $m_\uparrow$. We will often consider the network in a state where $A(t) = a$. Given $m_\uparrow$ this is achieved by setting:

$$m_\downarrow(t) = m_{\downarrow A}(t) = 1 - \frac{a}{(1-a)}(1 - m_\uparrow(t)) \tag{5}$$

2.3 Theoretical description

To model the time evolution of the network we consider only one update step and will therefore omit the time index t for simpler notation. The current network state is characterized by $m_\uparrow$ and $m_\downarrow$ and we need to calculate the local field h_i in the thermodynamic limit for the two cases $\xi_i^\mu = 1$ and $\xi_i^\mu = 0$. To do this we have to distinguish between the condensed and the uncondensed patterns and therefore we split the local field into signal part h_i^s and noise part h_i^n with $\delta = \frac{1}{a(1-a)}$:

$$h_i = \lim_{\substack{N,p\to\infty \\ \alpha=\frac{p}{cN}}} \frac{\delta}{Nc} \Bigg[\underbrace{\sum_{\substack{j=1 \\ j\neq i}}^{N} c_{ij}(\xi_i^\mu - a)(\xi_j^\mu - a)S_j}_{\to h_i^s} + \underbrace{\sum_{\substack{\nu=1 \\ \nu\neq\mu}}^{p} \sum_{\substack{j=1 \\ j\neq i}}^{N} c_{ij}(\xi_i^\nu - a)(\xi_j^\nu - a)S_j}_{\to h_i^n} \Bigg]$$

Due to self-averaging the signal part becomes:

$$h_i^s = \frac{\xi_i^\mu - a}{ca(1-a)} < (\xi_j^\mu - a)S_j >_{j\neq i}$$

where $< \ldots >_{j\neq i}$ denotes the average over the sites $j \neq i$. In contrast, the noise part has a positive variance even in the thermodynamic limit so we cannot apply self-averaging. But according to the central limit theorem h_i^n is a gaussian random variable with variance σ^2 and after doing the calculation we

see that the mean is zero $\ll h_i^n \gg_{\substack{j\neq i \\ \nu\neq\mu}} = 0$. Therefore the local field for $\xi_i^\mu = 1$ resp. $\xi_i^\mu = 0$ is normally distributed with mean $\mu_\uparrow$ resp. $\mu_\downarrow$ and variance σ^2:

$$\begin{aligned}
\mu_\uparrow &= h_i^s|_{\xi_i^\mu=1} = (1-a)(m_\uparrow + m_\downarrow - 1) \\
\mu_\downarrow &= h_i^s|_{\xi_i^\mu=0} = -a(m_\uparrow + m_\downarrow - 1) \\
\sigma^2 &= \ll {h_i}^2 \gg_{\substack{j\neq i \\ \nu\neq\mu}} - \ll h_i \gg^2_{\substack{j\neq i \\ \nu\neq\mu}} = \alpha A
\end{aligned} \tag{6}$$

The distribution functions for $h_i|_{\xi_i^\mu=1} - Q$ and $h_i|_{\xi_i^\mu=0} - Q$ are:

$$\begin{aligned}
\rho_\uparrow(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu_\uparrow+Q)^2}{2\sigma^2}\right) \\
\rho_\downarrow(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu_\downarrow+Q)^2}{2\sigma^2}\right)
\end{aligned} \tag{7}$$

With the dynamical equations (2) we can compute $m_\uparrow$ and $m_\downarrow$ for the next time step by averaging the activation function over the distributions of the $h_i - Q$:

$$\begin{aligned}
m_\uparrow(t+1) &= < g(h_i(t) - Q, \beta) >_{\rho_\uparrow} = \int_{-\infty}^{\infty} g(x,\beta)\rho_\uparrow(x)\, dx \\
m_\downarrow(t+1) &= < 1 - g(h_i(t) - Q, \beta) >_{\rho_\downarrow} = \int_{-\infty}^{\infty} (1 - g(x,\beta))\rho_\downarrow(x)\, dx
\end{aligned} \tag{8}$$

For the noiseless case $T = 0$ we can use the simple dynamical equation (1) and the integral reduces to:

$$\begin{aligned}
m_\uparrow(t+1) &= \int_{-\infty}^{\infty} \Theta(x)\rho_\uparrow(x)\, dx = \frac{1}{2}\left(1 + erf\left(\frac{\mu_\uparrow - Q}{\sqrt{2}\sigma}\right)\right) \\
m_\downarrow(t+1) &= \frac{1}{2}\left(1 + erf\left(\frac{Q - \mu_\downarrow}{\sqrt{2}\sigma}\right)\right)
\end{aligned} \tag{9}$$

If the system is in a state with $m_\uparrow + m_\downarrow = 1$ we have $A|_{m_\downarrow = 1 - m_\uparrow} = m_\uparrow$ and $m_\downarrow = 1 - A$ and therefore this is a characterization of an uncorrelated state. In this case we have $\mu_\uparrow = \mu_\downarrow = 0$ and thus $\rho_\uparrow(x) = \rho_\downarrow(x)$ leading to:

$$m_\uparrow(t+1) + m_\downarrow(t+1) = \int_{-\infty}^{\infty} \rho_\uparrow(x)\, dx = 1 \tag{10}$$

Hence $m_\uparrow + m_\downarrow = 1$ is a fixed line in phase space which is consistent with the assumption that noncondensed patterns stay noncondensed during retrieval.

2.4 Validity of the description

In the calculations above we considered the neurons to be independent of each other. This is certainly true for the first time step but after one iteration there may be correlations due to feedback loops. In the limit of strong dilution, so that the neurons do not have common ancestors in former time steps, the description is valid for the whole retrieval process. To achieve this the number of connections at each neuron C has to be of order $\ln N$ and therefore we have $c \sim \frac{\ln N}{N} \to 0$ for $N \to \infty$ [18, 20].

This can be seen by a short estimate of the probability that two neurons have common ancestors during retrieval. The total number of ancestors of all neurons after t time steps is

$$\sum_{i=1}^{t} C^i = \frac{C^{t+1} - 1}{C - 1} - 1 \approx C^t \qquad \text{for large } C.$$

The probability that any two ancestors of these are not equal is given by

$$\prod_{i=1}^{C^t} \frac{N - j}{N} \approx 1 - \frac{C^{2t}}{2N} + \mathcal{O}\left(\left(\frac{C^{2t}}{2N}\right)^2\right) \qquad \text{for large } N.$$

This is very close to one as long as $C^{2t} \ll 2N$. This is certainly true for all t in the limit $N \to \infty$ if we take $C \sim \ln N$. But as long as we have a finite connectivity c we have

$$t \ll \frac{\ln N}{2 \ln(cN)} \longrightarrow \frac{1}{2} \qquad \text{for} \quad N \to \infty$$

as a condition for uncorrelated neurons. So this assumption in infinite systems with connectivity $c > 0$ is only true for the first time step. However, in finite systems with $c \ll 1$ the situation is not as drastic.

Therefore the dynamical results in sections 3 and 4 are valid for the first time step for arbitrary network connectivity and can be applied to the whole retrieval process in the limit of strong dilution. The equilibrium properties discussed in chapter 5 are only valid in this limit.

3 Dynamics at zero temperature

In the following we look at the network in a particular state described by $m_\uparrow$ and $m_\downarrow$. We derive conditions on α and Q for $m_\uparrow$ and $m_\downarrow$ to improve with the next update using noiseless dynamics. The network is not in equilibrium, only if the two improvements happen to be zero (see chapter 5). Therefore the critical capacity and the corresponding threshold calculated in this section are dynamical variables, i.e. they depend on the state of the network and change during the retrieval process. We set $\Delta m_\uparrow = m_\uparrow(t+1) - m_\uparrow(t) > 0$ and $\Delta m_\downarrow = m_\downarrow(t+1) - m_\downarrow(t) > 0$ in equation (9) and get two conditions for improvement of $m_\uparrow$ and $m_\downarrow$ ($inverf(x)$ is the inverse error function, see appendix A.2):

$$\begin{aligned} \frac{\mu_\uparrow - Q}{\sigma} &> c_\uparrow := \sqrt{2}\, inverf(2m_\uparrow - 1) \\ \frac{Q - \mu_\downarrow}{\sigma} &> c_\downarrow := \sqrt{2}\, inverf(2m_\downarrow - 1) \end{aligned} \tag{11}$$

We will look at the threshold later. First we are interested in the highest possible storage level independent of Q, i.e. we assume that Q is chosen optimally.

3.1 Critical storage capacity

By adding the two equations (11) we get a condition for improvement that has to be satisfied independent of the choice of Q.

$$\mu_\uparrow - \mu_\downarrow > \sigma(c_\uparrow + c_\downarrow) \tag{12}$$

This inequality represents the idea that the distance between the means of the two distributions $\rho_\uparrow$ and $\rho_\downarrow$ of the local field must be big enough to ensure im-

provement of $m_\uparrow$ and $m_\downarrow$. After inserting the expressions for $\mu_\uparrow$, $\mu_\downarrow$ and σ from equation (6) we can solve for the critical value α_c where both improvements are just zero:

$$\begin{aligned} \alpha_c &= \frac{(m_\uparrow + m_\downarrow - 1)^2}{(c_\uparrow + c_\downarrow)^2 A} \qquad \text{for} \qquad m_\uparrow + m_\downarrow \neq 1 \\ &= \frac{1}{2\pi m_\uparrow}\, e^{-{c_\uparrow}^2/2} \qquad \text{for} \qquad m_\uparrow + m_\downarrow = 1 \end{aligned} \tag{13}$$

From the inequality (12) we get the following conditions for improvement of $m_\uparrow$ and $m_\downarrow$:

$$\alpha < \alpha_c \quad \text{if} \quad m_\uparrow + m_\downarrow > 1 \qquad\qquad \alpha > \alpha_c \quad \text{if} \quad m_\uparrow + m_\downarrow < 1 \tag{14}$$

This is the case because we have $c_\uparrow + c_\downarrow < 0$ for $m_\uparrow + m_\downarrow < 1$ and in this region we can have improvement for arbitrary high pattern loadings. But it is not very surprising because if the network is uncorrelated with the retrieval pattern we have $m_\uparrow + m_\downarrow = 1$. So if α is too high the network state always develops towards an uncorrelated state. Most of the time we will consider the case $m_\uparrow + m_\downarrow > 1$, which is the natural one as the interest is mainly in pattern retrieval with a similar stimulus.

The usual storage capacity in studies of the network in equilibrium is defined to be the maximal α for which a fixed point of the dynamics with macroscopic overlap with the retrieval pattern exists. Here the critical capacity is the pattern loading for which $\Delta m_\uparrow$ and $\Delta m_\downarrow$ are just 0, which gives some insight in the restrictions on the pattern loading occuring during the retrieval process. In the next chapter we look at possible choices of Q to realize α_c, now we study the dependence of α_c on the parameters $m_\downarrow$, $m_\uparrow$, a and c.

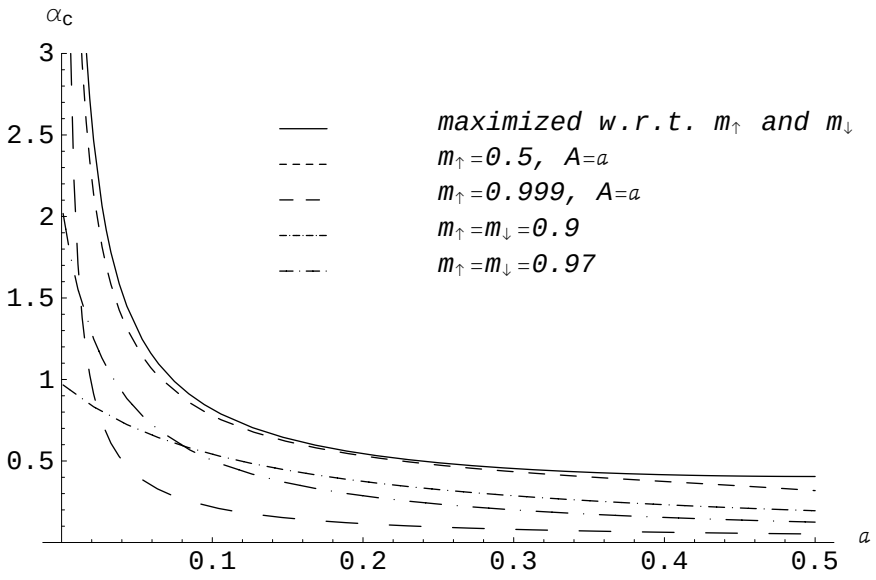

Figure 1: Critical storage capacity $\alpha_c(a)$ for $T = 0$ and several choices of $m_\uparrow$ and $m_\downarrow$.

- **Dependence on network connectivity**

  As we defined $\alpha = \frac{p}{cN}$ we see that the maximal number of storable patterns decreases proportional to the number of connections per neuron, because α_c is independent on c. This is in accordance with [19], where other types of dilution were also studied.

- **Dependence on pattern activity**

  For decreasing a the capacity increases monotonically if $m_\uparrow + m_\downarrow > 1$ and decreases if $m_\uparrow + m_\downarrow < 1$. If $m_\uparrow$ and $m_\downarrow$ are fixed we get the finite value $\alpha|_{a=0} = \frac{(m_\uparrow+m_\downarrow-1)^2}{(c_\uparrow+c_\downarrow)^2(1-m_\downarrow)}$ (see figure 1). When the activity of the network is equal to the one of the stored patterns we have $A = a$, $m_\downarrow = m_{\downarrow A}$ and therefore $c_\downarrow|_{m_{\downarrow A}} = \sqrt{2}\,inverf[1 - \frac{2a}{1-a}(1 - m_\uparrow)]$. By using the approximation $inverf\,(x) \approx \sqrt{-\ln(1-x)}$ for small x given in

appendix A.2 we get the following for $a \ll 0.5$:

$$(c_\uparrow + c_\downarrow|_{m_{\downarrow A}})^2 \approx (c_\downarrow|_{m_{\downarrow a}})^2 \approx -2\ln\left[\frac{2a}{(1-a)}(1-m_\uparrow)\right] \approx -2\ln a$$

This is only true if $m_\uparrow \ll 1-2a$ because otherwise we cannot neglect $c_\uparrow$. Now we get the leading behavior of α_c for the limit $a \to 0$:

$$\alpha_c \approx \frac{(m_\uparrow + m_{\downarrow A} - 1)^2}{(c_\downarrow|_{m_{\downarrow A}})^2 a} \approx \frac{{m_\uparrow}^2 c}{-2a\ln a} \qquad \text{with} \qquad m_\uparrow \ll 1-2a$$

This behavior is known from former studies [12, 13, 15, 11] as an approximation to the equilibrium storage capacity for small a and resembles the upper bound obtained by Gardner [14]. Here we see that in order to realize this high storage level we have to fix the network activity $A = a$ during the retrieval process. This can be done by adapting the threshold in every time step, studied in section 3.4. Other methods of activity control have also been studied [11, 23, 24].

- **Dependence on state of the network**

  As α_c is a dynamical variable it depends on the network state described by $m_\uparrow$ and $m_\downarrow$. If one of these two parameters is equal to 1, α_c is zero, because for $m_\uparrow, m_\downarrow \to 1$ we have $c_\uparrow, c_\downarrow \to \infty$. This results from the assumption of a gaussian distribution for the local field which is only true for $N \to \infty$. In that case we can have perfect retrieval only for a finite number of patterns. For finite N the distribution of the local field is discrete and bounded and therefore the description is only an approximation.

  The capacity α_c is differentiable with respect to $m_\uparrow$ and $m_\downarrow$ over the

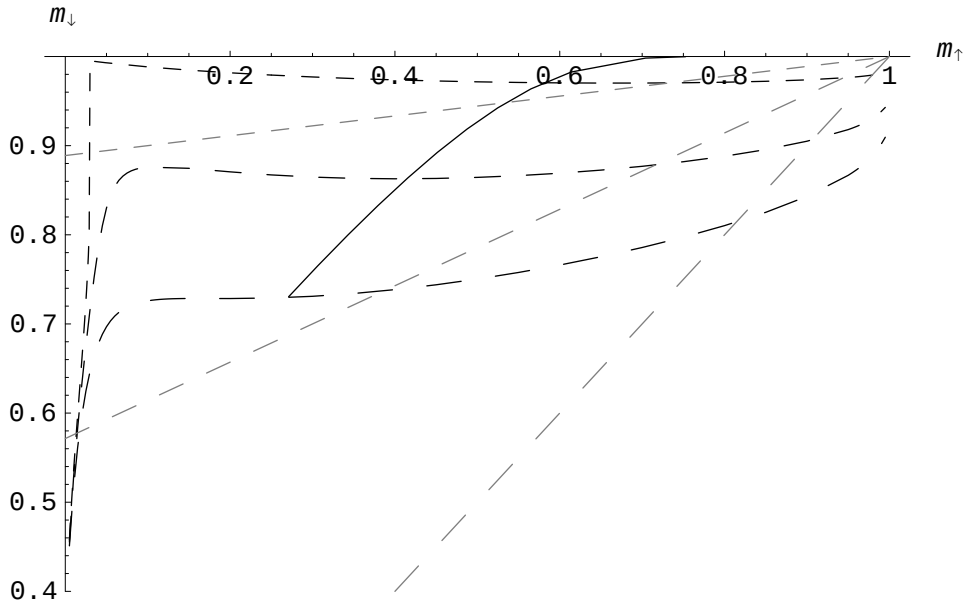

Figure 2: Comparison of $m_{\downarrow_{opt}}(m_\uparrow)$ (black broken lines) and $m_{\downarrow_A}(m_\uparrow)$ (gray broken lines) for $T = 0$, $a = 0.1$ - - -, $a = 0.3$ – – – and $a = 0.5$ — — —. Locations of maximized $\alpha_c(m_{\uparrow_{opt}}, m_{\downarrow_{opt}})$ for $a = 0, \ldots, 0.5$ ———.

whole parameter range $(0, 1)$. For fixed $m_\uparrow$, we have $\alpha_c = 0$ for $m_\downarrow = 0, 1$ so there has to be an optimal relation between $m_\uparrow$ and $m_\downarrow$ that maximizes α_c. By setting $\partial_{m_\downarrow}\alpha_c|_{m_{\downarrow_{opt}}} = 0$ we get the following condition that $m_{\downarrow_{opt}}$ has to satisfy:

$$2(c_\uparrow + c_\downarrow)A = (m_\uparrow + m_\downarrow - 1)[2(\partial_{m_\downarrow} c_\downarrow)A - (c_\uparrow + c_\downarrow)(1 - a)] \qquad (15)$$

We can solve the equation numerically for $m_{\downarrow_{opt}}$ and compare it to $m_{\downarrow_A}$ that ensures $A = a$ (Figure 2). We see that the two cases do not match, even for small a there is a considerable difference. So in contrast to what one could have expected the storage capacity is not maximized if the network activity is fixed to a.

In order to calculate the absolut maximum of α_c with respect to $m_\uparrow$ and

$m_\downarrow$ we solve (15) for $c_\uparrow + c_\downarrow$, insert in the expression for α_c (13) and maximize it with respect to $m_\uparrow$. We get the second condition between $m_\uparrow$ and $m_\downarrow$ that specifies the location of the maximum $(m_{\uparrow_{opt}}, m_{\downarrow_{opt}})$ which is drawn by a full line in figure 2.

$$m_\downarrow \;=\; 1 - \frac{a(1+a)}{(1-a)(2-a)} m_\uparrow \tag{16}$$

The maximized critical storage capacity $\alpha_c(m_{\uparrow_{opt}}, m_{\downarrow_{opt}})$ is also drawn in figure 1 (unbroken line). As α_c for $m_\downarrow = m_{\downarrow_A}$ already resembles the upper bound on α_c obtained by Gardner we have the same leading behavior of the maximized capacity for small activities.

3.2 Mutual information

For very small pattern activities the storage capacity may increase but the number of active neurons and the information represented by a single pattern decreases. Therefore the maximal amount of information storable in the network gives a more sensible characterization of performance. The information content of a single pattern I for specific binary probabilities is usually described as the logarithm of the degeneracy of a pattern:

$$\frac{I}{N} = \frac{1}{N} \ln \left(\begin{array}{c} N \\ aN \end{array} \right) \longrightarrow a \ln a + (1-a)\ln(1-a) \qquad \text{for } N \to \infty$$

So we get for the maximal information per synapse storable in the network in units of bits

$$i = \lim_{N\to\infty} I \frac{p_{max}}{cN^2 \ln 2} = -\frac{\alpha_c}{\ln 2}[a \ln a + (1-a)\ln(1-a)] \tag{17}$$

However, if the network is not in the perfect retrieval state with $m_\uparrow + m_\downarrow = 1$ information from the original patterns is lost. To account for this one uses the mutual information which can be defined as the negative logarithm of the conditional probability of choosing a network state with $m_\uparrow$ and $m_\downarrow$ given the activity A:

$$\begin{aligned}
\frac{I_m}{N} &= -\frac{1}{N}\ln\left[\begin{pmatrix} aN \\ m_\uparrow aN \end{pmatrix}\begin{pmatrix} (1-a)N \\ (1-m_\downarrow)(1-a)N \end{pmatrix}\Big/\begin{pmatrix} N \\ AN \end{pmatrix}\right] = \\
&\to -A\ln A - (1-A)\ln(1-A) + a(m_\uparrow \ln m_\uparrow + (1-m_\uparrow)\ln(1-m_\uparrow)) + \\
&\quad +(1-a)(m_\downarrow \ln m_\downarrow + (1-m_\downarrow)\ln(1-m_\downarrow)) \qquad \text{for } N \to \infty
\end{aligned}$$

This gives the average amount of information per pattern that can be obtained from the network in a given state $m_\uparrow$, $m_\downarrow$. A more theoretical definition of this quantity is given in [22]:

$$I_m = N(\mathcal{S}(\mathbf{S}) - < \mathcal{S}(\mathbf{S}|\xi) >_\xi)$$

Here $\mathcal{S}(\mathbf{S})$ is the entropy of the current state of the network which is only a function of the activity A and independent of site i and the stored patterns. $\mathcal{S}(\mathbf{S}|\xi)$ is the conditional entropy given a specific pattern ξ for retrieval.

$$\begin{aligned}
\mathcal{S}(\mathbf{S}) &= -\sum_{S_i=0,1} P(S_i)\ln P(S_i) \\
< \mathcal{S}(\mathbf{S}|\xi) >_\xi &= -\sum_{\xi_i=0,1} P(\xi_i)\left(\sum_{S_i=0,1} P(S_i|\xi_i)\ln P(S_i|\xi_i)\right)
\end{aligned}$$

This gives the same expression as before and with further simplification we get the following formula for the maximal mutual information content i_m per synapse of the network in units of bits:

$$\begin{aligned}
i_m = I_m\frac{p_{max}}{cN^2\ln 2} = \frac{\alpha_c}{c\ln 2}\{ \; & a\left[m_\uparrow \ln\tfrac{m_\uparrow}{A} + (1-m_\uparrow)\ln\tfrac{1-m_\uparrow}{1-A}\right] + \\
& +(1-a)\left[m_\downarrow \ln\tfrac{m_\downarrow}{1-A} + (1-m_\downarrow)\ln\tfrac{1-m_\downarrow}{A}\right]\} \qquad (18)
\end{aligned}$$

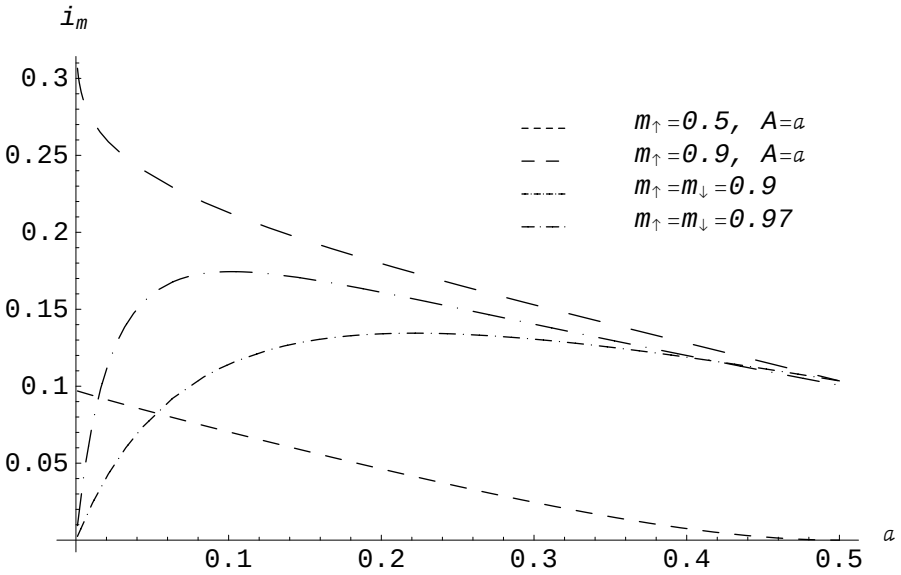

Figure 3: Mutual information content per synapse $i_m(a)$ for $T = 0$ and several values of $m_\uparrow$ and $m_\downarrow$.

This observable can be interpreted as the maximal possible information that can be obtained from the network, subject to the constraint of improvement of $m_\uparrow$ and $m_\downarrow$. As this is also a dynamical variable so we can compare it to the dynamical storage capacity and see that it has more reasonable properties for low pattern activity than α_c:

- If the network is in an uncorrelated state, i.e. $m_\uparrow + m_\downarrow = 1$ we have $i_m|_{m_\uparrow+m_\downarrow=1} = 0$ which makes sense as these states are only randomly correlated with the retrieval pattern.

- On the other hand i_m is bounded above by the information content of the undisturbed patterns i and if $m_\uparrow = m_\downarrow = 1$ we have $i_m = i$.

- We have $i_m|_{a=0,1} = 0$ for any fixed $m_\uparrow$ and $m_\downarrow$ as there is certainly no information stored when all pattern elements have the same value.

However for $m_\downarrow = m_{\downarrow_A}$ or $A = a$, using the approximation of α_c for small a we get:

$$i_m \approx \frac{{m_\uparrow}^2}{-2a \ln a \ln 2}(-a m_\uparrow \ln a) \longrightarrow \frac{{m_\uparrow}^3}{2 \ln 2} \qquad \text{for } a \to 0 \tag{19}$$

That means if we keep the network activity fixed $A = a$ during the retrieval process we can obtain a nonzero amount of information from the network for $a \to 0$ although the information content of a single pattern vanishes. This result is in accordance with former studies in equilibrium (see e.g. [13, 14]) and some examples are plotted in figure 3.

3.3 Involving the threshold

Now we look at the two conditions (11) separately. After inserting $\sigma = \sqrt{\alpha A}$ from (6) we can solve for $\alpha_\uparrow$ resp. $\alpha_\downarrow$ where $\Delta m_\uparrow = 0$ resp. $\Delta m_\downarrow = 0$:

$$\alpha_\uparrow = \frac{(\mu_\uparrow - Q)^2}{{c_\uparrow}^2 A} \qquad \alpha_\downarrow = \frac{(Q - \mu_\downarrow)^2}{{c_\downarrow}^2 A} \tag{20}$$

Because the conditions were squared only one branch of the parabolas $\alpha_\uparrow(Q)$ and $\alpha_\downarrow(Q)$ imposes a condition on α depending on the values of $m_\uparrow$ and $m_\downarrow$. We have to distinguish between $m_\uparrow > 0.5$ and $m_\uparrow < 0.5$. In the first case we have $c_\uparrow > 0$ and the condition for improvement of $m_\uparrow$ (11) gives an upper bound on α. In the second case the sign of the inequality is changed because $c_\uparrow < 0$ and therefore we have a lower bound on α for improvement of $m_\uparrow$:

$$m_\uparrow > 0.5 \Rightarrow \begin{array}{ll} \alpha < \alpha_\downarrow & \text{for } Q < \mu_\uparrow \\ \alpha = 0 & \text{for } Q \geq \mu_\uparrow \end{array} \qquad m_\uparrow < 0.5 \Rightarrow \begin{array}{ll} \alpha > \alpha_\downarrow & \text{for } Q > \mu_\uparrow \\ \alpha \geq 0 & \text{for } Q \leq \mu_\uparrow \end{array} \tag{21}$$

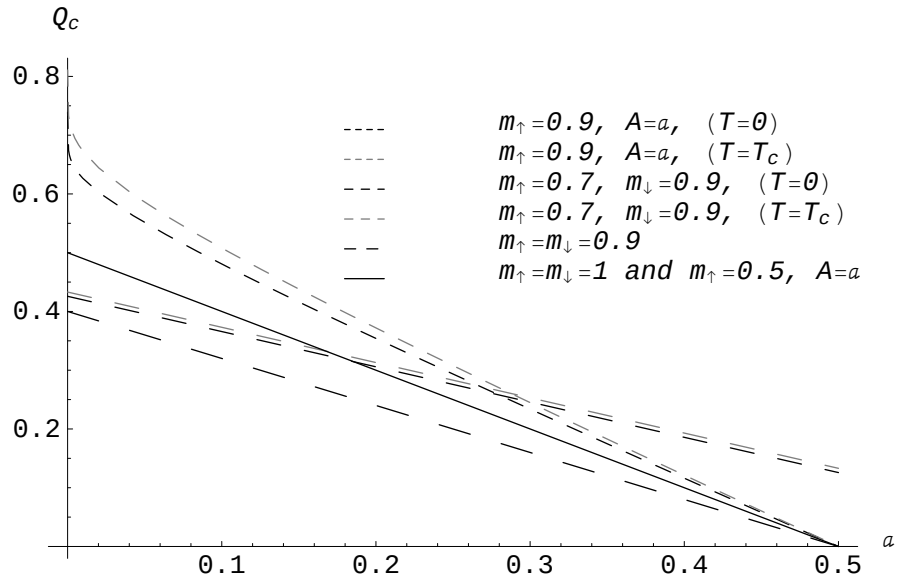

Figure 4: Comparison of critical thresholds $Q_c(a)$ between $T = 0$ and $T = T_c$ for several values of $m_\uparrow$ and $m_\downarrow$.

For $m_\uparrow = 0.5$ the parabola reduces to a vertical line at $\mu_\uparrow$ and we have improvement if $Q < \mu_\uparrow$ for unlimited pattern loading and no improvement for $Q > \mu_\uparrow$. The same applies for the conditions of improvement of $m_\downarrow$:

$$m_\downarrow > 0.5 \Rightarrow \begin{array}{ll} \alpha < \alpha_\uparrow & \text{for } \mu_\downarrow < Q \\ \alpha = 0 & \text{for } \mu_\downarrow \geq Q \end{array} \qquad m_\uparrow < 0.5 \Rightarrow \begin{array}{ll} \alpha > \alpha_\uparrow & \text{for } \mu_\downarrow > Q \\ \alpha \geq 0 & \text{for } \mu_\downarrow \leq Q \end{array} \tag{22}$$

For good retrieval qualities we would like to have improvements on both, $m_\uparrow$ and $m_\downarrow$, and this is ensured in a region of the (Q, α)-plane limited by the valid branches of $\alpha_\uparrow$ and $\alpha_\downarrow$ (see figure 5). Hence we have a lower and upper bound for the region of the threshold that will lead to improvements on $m_\uparrow$ and $m_\downarrow$ depending on α.

$$Q \in [\mu_\downarrow + c_\downarrow \sigma, \mu_\uparrow - c_\uparrow \sigma]$$

Unless $m_\uparrow = m_\downarrow = 0.5$ there is always an intersection of the valid branches of $\alpha_\uparrow$ and $\alpha_\downarrow$ where both improvements are equal to zero. The interval for Q

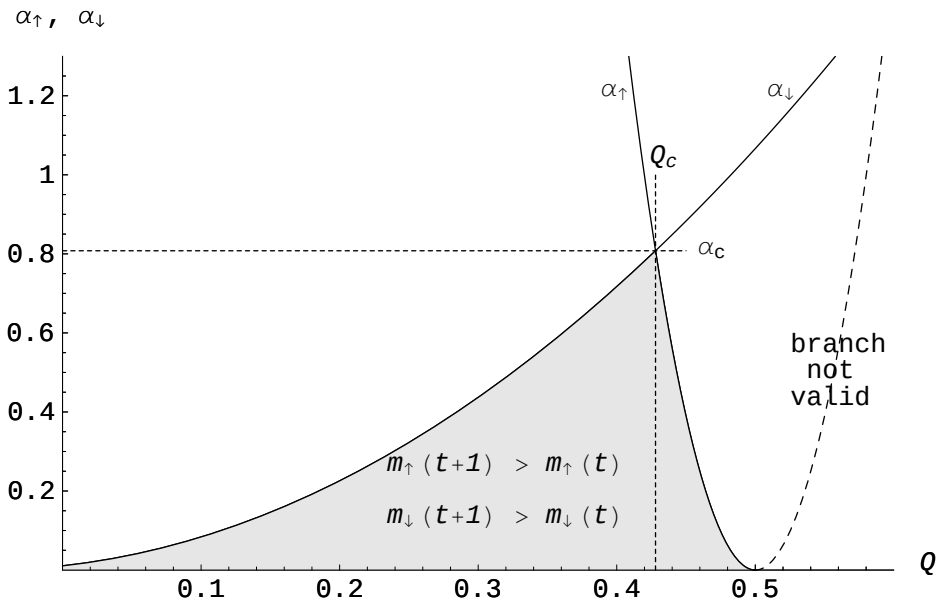

Figure 5: $\alpha_\uparrow(Q)$ and $\alpha_\downarrow(Q)$ for $m_\uparrow = 0.6$, $A = a = 0.1$ and $T = 0$. The region of simultaneous improvement is shaded.

reduces to a single point and $\alpha_\uparrow = \alpha_\downarrow$ are equal to the critical storage capacity α_c we discussed earlier. The value of Q at this point is:

$$\begin{aligned} Q_c &= (\frac{c_\downarrow}{c_\uparrow + c_\downarrow} - a)(m_\uparrow + m_\downarrow - 1) \qquad \text{for} \qquad m_\uparrow + m_\downarrow \neq 1 \\ &= \frac{c_\downarrow}{\sqrt{2\pi}}\, e^{-{c_\downarrow}^2/2} \qquad \text{for} \qquad m_\uparrow + m_\downarrow = 1 \end{aligned} \tag{23}$$

Like α_c, Q_c depends on $m_\uparrow$, $m_\downarrow$ and a. The dependence on a is linear and plotted in figure 4 in comparison to the value at $T = T_c$ (see section 4.1). For $m_\uparrow = m_\downarrow$ this reduces to $Q_c = (0.5 - a)(2m_\uparrow - 1)$ and at $m_\uparrow = 1$ finally to $Q_c = 0.5 - a$ which is known to be the optimal threshold near retrieval (see e.g. [3, 13]). For the case $m_\uparrow = 0.5$, $A = a$ we also get this dependence.

For $m_\uparrow + m_\downarrow > 1$ the region of simultaneous improvement is bounded below and above and the maximal possible storage is α_c. For $m_\uparrow + m_\downarrow < 1$ the region is only bounded below and α_c is the minimal possible pattern load-

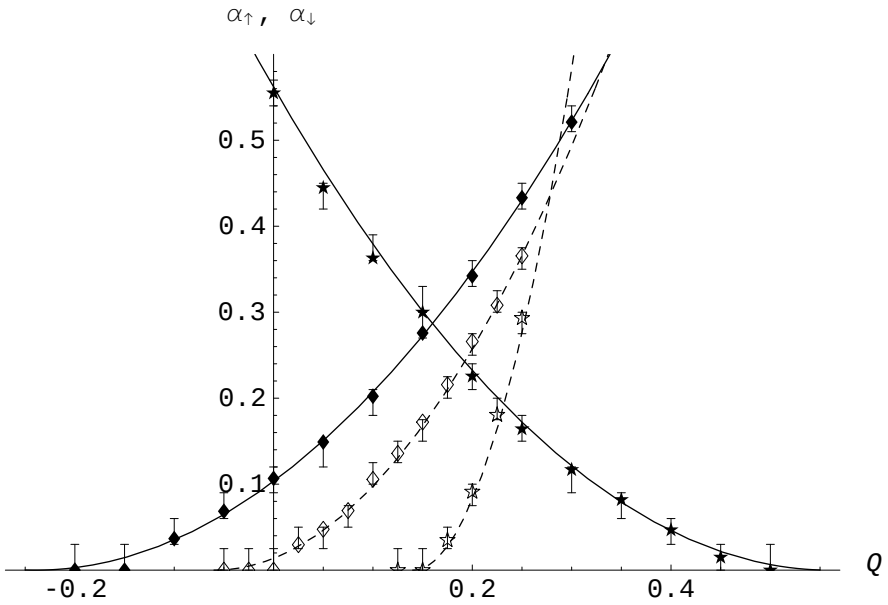

Figure 6: Simulations for $\alpha_\uparrow(Q)$ and $\alpha_\downarrow(Q)$ with $a = 0.3$, $T = 0$ and $m_\uparrow = m_\downarrow = 0.9$ —— (Data filled), $m_\uparrow = 0.3, m_\downarrow = 0.9$ - - - (Data unfilled). Data obtained with $N = 600$ and average over 10 networks each with at least 150 stimuli: The average improvement of $m_\uparrow$ and $m_\downarrow$ was recorded as a function of α for different values of Q. At $\alpha = \alpha_\uparrow$ resp. $\alpha_\downarrow$, $\Delta m_\uparrow$ resp. $\Delta m_\downarrow$ vanishes.

ing for improvement (see also (14)). The usual case for a condensed pattern $m_\uparrow, m_\downarrow > 0.5$ is plotted in figure 5 and the region where $m_\uparrow$ and $m_\downarrow$ improve simultaneously is shaded. In figure 6 our considerations are confirmed by simulations for $m_\uparrow = m_\downarrow = 0.9$ and $m_\uparrow = 0.3$, $m_\downarrow = 0.9$.

3.4 Proper choices of thresholds

Now we consider several choices of thresholds as functions of the state of the network which ensure $Q|_{\alpha=\alpha_c} = Q_c$, i.e. they allow for the critical storage capacity α_c (see Figure 7).

- **Critical threshold**

  The easiest thing is of course to choose $Q = Q_c$ which certainly meets the requirement. But if $m_\uparrow$ resp. $m_\downarrow < 0.5$ it does not improve if $\alpha < \alpha_c$ (see $m_\uparrow$ in figure 6). This choice of Q maximizes the term

  $$\Delta c_\uparrow + \Delta c_\downarrow = \frac{\partial c_\uparrow}{\partial m_\uparrow}\Delta m_\uparrow + \frac{\partial c_\downarrow}{\partial m_\downarrow}\Delta m_\downarrow.$$

- **Maximizing $\Delta m_\uparrow + \Delta m_\downarrow$**

  Just for comparison we will also look at $Q = Q_m := \frac{1}{2}(\mu_\uparrow + \mu_\downarrow)$ which comes from maximizing $\Delta m_\uparrow + \Delta m_\downarrow$. This seems to be a somehow 'natural' choice but it only fulfills $Q_m|_{\alpha=\alpha_c} = Q_c$ if $m_\downarrow = m_\uparrow$, where we have $c_\uparrow = c_\downarrow$ and therefore $Q_m = Q_c$.

- **Ensuring $A(t+1) = a$**

  Another possibility is to choose $Q = Q_a$ in order to get the network activity $A(t+1) = a$. As mentioned before keeping $A = a$ ensures a very high storage capacity for small a. In the retrieval state we also have $A = a$ so we can reach it with this threshold dependence. However the condition $Q_a|_{\alpha=\alpha_c} = Q_c$ is only obeyed for $A(t) = a$ because then $\Delta m_\uparrow$ and $\Delta m_\downarrow$ have the same sign for each α and we have $\Delta m_\uparrow, \Delta m_\downarrow \to 0$ for $\alpha \to \alpha_c$. But if $A(t) \neq a$ both improvements can't be equal to zero for the same α because then we would have $A(t+1) \neq a$.

- **Preserving $m_\uparrow/m_\downarrow$**

  We can also choose $Q = Q_r$ to preserve the ratio $\frac{m_\uparrow(t+1)}{m_\downarrow(t+1)} = \frac{m_\uparrow(t)}{m_\downarrow(t)} = r$. Therefore both improvements $\Delta m_\uparrow$ and $\Delta m_\downarrow$ must have the same sign and hence they both have to reach zero for $\alpha \to \alpha_c$. That means that

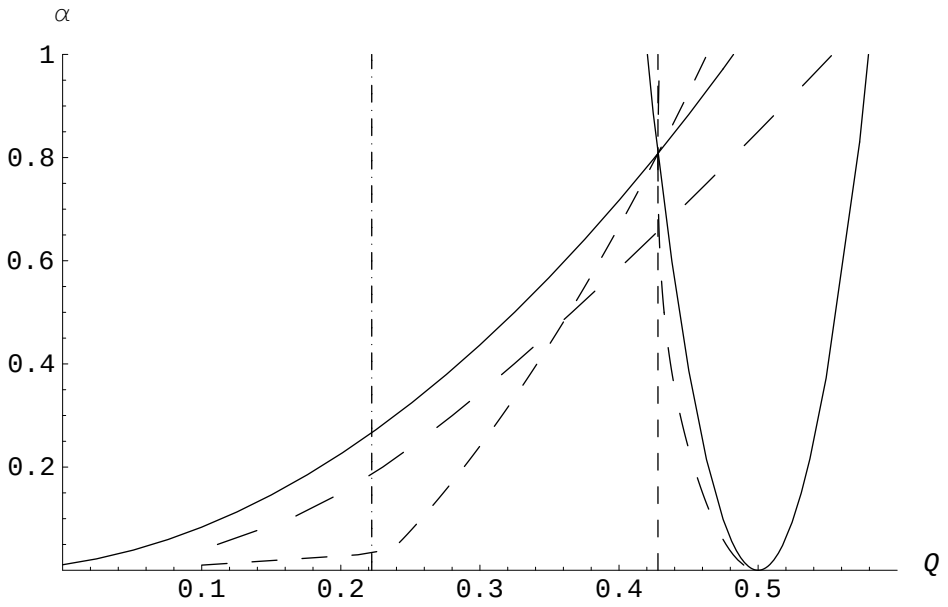

Figure 7: Comparison of thresholds for $m_\uparrow = 0.6$, $A = a = 0.1$ and $T = 0$. Threshold choices: Q_c - - -, Q_m - · - · -, Q_a – – –, Q_r — — —, Q_o —— —— ——.

$Q_r|_{\alpha=\alpha_c} = Q_c$ with the same argument as above. We look at this choice only because it is optimal, but if we start out with a ratio considerably different from 1 the attractor reached with this threshold won't be very close to the retrieval state. So the only case where $Q = Q_r$ is really interesting is for $r = 1$ where we have $Q_r = Q_c = Q_m$.

- **Ensuring $\mathbf{m_\downarrow(t+1) = m_{\downarrow opt}}$**

  The last possibility we look at is to choose $Q = Q_o$ so that $m_\downarrow(t+1) = m_{\downarrow opt}$ from equation (15) which maximizes α_c for the next time step. With the same arguments used above we will only have $Q_o|_{\alpha=\alpha_c} = Q_c$ if $m_\downarrow(t) = m_{\downarrow opt}$.

The last three possibilities do not give explicit values for the threshold so Q is determined numerically from each condition. They restrict the possible states

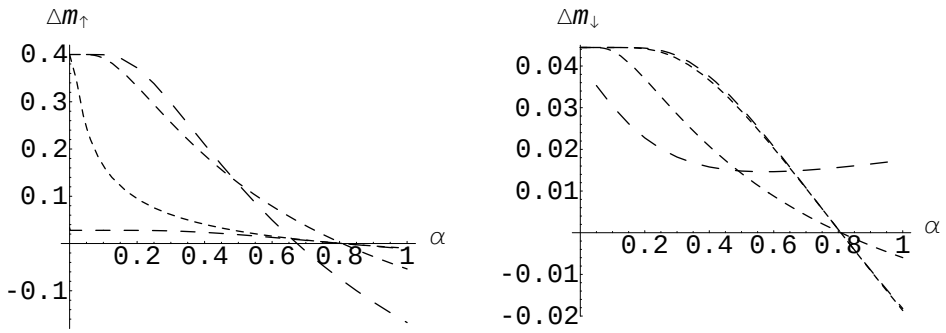

Figure 8: Comparison of improvements $\Delta m_\uparrow(\alpha)$ and $\Delta m_\downarrow(\alpha)$ for $m_\uparrow = 0.6$, $A = a = 0.1$. Threshold choices: Q_c - - -, Q_a – – –, Q_r — — —, Q_o —— —— ——.

of the network ensuring a fixed relation between $m_\uparrow$ and $m_\downarrow$. This will be advantageous when we study the fixed points of the dynamics in chapter 5.

In figure 7 we look at the case $m_\uparrow = 0.6$ and $m_\downarrow = m_{\downarrow_A}$. Therefore we have $Q_a|_{\alpha=\alpha_c} = Q_c$ and Q_o and Q_m do not fulfill the requirement. One can compare the different choices of thresholds by the improvement of $m_\uparrow$ and $m_\downarrow$ which was done in figure 8. As $m_\uparrow = 0.6$ the upper limit for $\Delta m_\uparrow$ is 0.4 and for $\Delta m_\downarrow$ it is $1 - m_{\downarrow_A} = 0.044$. For small α these limits are reached by Q_c, Q_a and Q_o. As expected Q_a gives the best improvements because it has the best position between the two parabolas in figure 7. The improvement of $m_\uparrow$ resp. $m_\downarrow$ is a monotonically increasing function of the distance between Q and $Q|_{\alpha=\alpha_\uparrow}$ resp. $Q|_{\alpha=\alpha_\downarrow}$. Q_c and Q_r give the highest possible storage capacity α_c in any case but the improvements $\Delta m_\uparrow$ look rather poor, especially of Q_r.

If we consider a case with $m_\uparrow = m_\downarrow$ the picture is symmetric with respect to the axis $Q = Q_c$ and therefore $Q_m = Q_r = Q_c$ give the best possible improvements $\Delta m_\uparrow = \Delta m_\downarrow$. Thus the optimal choice of the threshold during the retrieval process depends on the situation. If one has $m_\uparrow(0) \approx m_\downarrow(0)$ and

$a \approx 0.5$ the best possible choice for fast retrieval is just Q_c. But in the more interesting case of small pattern activity this choice limits the storage capacity compared to choosing Q_a. In this case it is best for the storage capacity to choose $Q = Q_a$, i.e. to keep $A = a$, independently of the initial condition. The retrieval time for this choice is fastest if one has an initial network activity close to the one of the stored patterns.

4 Dynamics for positive temperature

For positive temperature we can't solve the dynamical equations exactly like it was possible for $T = 0$ but we can get an analytical, exact expression for the critical temperature T_c. To calculate $\alpha_\uparrow$ and $\alpha_\downarrow$ we make an approximation for small T which is justified by the range of T_c.

4.1 Critical temperature

With increasing temperature the critical storage capacity decreases (see section 4.2) and we define T_c as the value of T where $\alpha_c = 0$. As α_c is a dynamical variable T_c also depends on the state of the network. The allowed region for the threshold reduces to a single point $Q_c|_{T=T_c}$ which we will also calculate. To evaluate T_c we have to look at the dynamical equation (8) for $T > 0$ in the limit $\alpha \to 0$. The distribution functions $\rho_\uparrow$ resp. $\rho_\downarrow$ reduce to delta functions because $\sigma^2 \sim \alpha$:

$$\rho_\uparrow(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-(x - \mu_\uparrow + Q)^2}{2\sigma^2} \quad \stackrel{\alpha\to 0}{\longrightarrow} \quad \delta(x - (\mu_\uparrow - Q))$$
$$\rho_\downarrow(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-(x - \mu_\downarrow + Q)^2}{2\sigma^2} \quad \stackrel{\alpha\to 0}{\longrightarrow} \quad \delta(x - (\mu_\downarrow - Q))$$

In this limit we can evaluate the integrals in the evolution equation (8) and get:

$$m_\downarrow(t+1) = \frac{1}{1 + e^{-2\beta(Q-\mu_\downarrow)}} \qquad m_\uparrow(t+1) = \frac{1}{1 + e^{-2\beta(\mu_\uparrow - Q)}}$$

By solving these equations for β and Q we get the critical temperature $T_c = \frac{1}{\beta_c}$ and the corresponding threshold $Q_c|_{T=T_c}$.

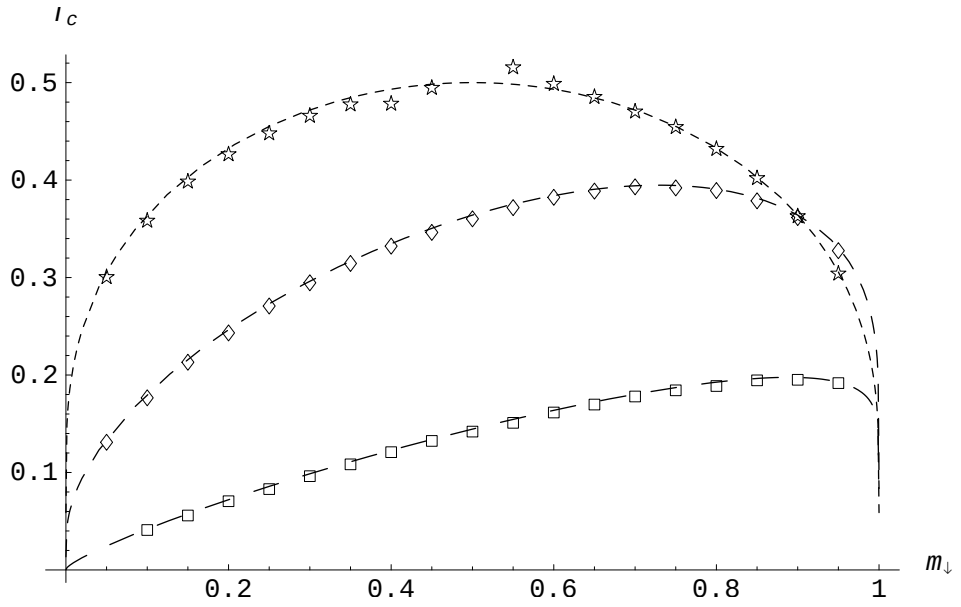

Figure 9: Critical temperature $T_c(m_\downarrow)$ for $m_\uparrow = 0.5$ - - -, $m_\uparrow = 0.9$ – – –, $m_\uparrow = 0.999$ — — —. Data with $N = 2000$, $\alpha = 0.0005$ and $m_\uparrow = 0.5$ ($\star$, $a = 0.3$ and $c = 0.5$), $m_\uparrow = 0.9$ ($\diamond$, $a = 0.3$ and $c = 1$) and $m_\uparrow = 0.999$ ($\Box$, $a = 0.5$ and $c = 1$).

For $m_\uparrow + m_\downarrow \neq 1$:

$$T_c = \frac{-2(m_\uparrow + m_\downarrow - 1)}{\ln(\frac{1}{m_\downarrow} - 1) + \ln(\frac{1}{m_\uparrow} - 1)} \tag{24}$$

$$Q_c|_{T=T_c} = \left(\frac{\ln(\frac{1}{m_\downarrow} - 1)}{\ln(\frac{1}{m_\downarrow} - 1) + \ln(\frac{1}{m_\uparrow} - 1)} - a \right) (m_\uparrow + m_\downarrow - 1) \tag{25}$$

For $m_\uparrow + m_\downarrow = 1$:

$$T_c = 2m_\uparrow(1 - m_\uparrow) \qquad Q_c|_{T=T_c} = m_\uparrow(1 - m_\uparrow) \ln\left(\frac{1}{m_\uparrow} - 1\right)$$

The critical temperature is only a function of the state of the network, i.e. $m_\uparrow$ and $m_\downarrow$, and independent of the pattern activity and the network connectivity. It is easy to see that the maximum value is $T_c = 0.5$ at $m_\uparrow = m_\downarrow = 0.5$ and if at least one of the two observables is equal to zero or one we have $T_c = 0$. We plotted $T_c(m_\downarrow)$ in figure 9 for several values of $m_\uparrow$ with confirmation by data

from simulations which were done at different values of a and c to demonstrate that it is independent on these parameters.

The critical threshold at $T = T_c$ has the same linear dependence on a as Q_c at $T = 0$ except for an additive constant depending on $m_\uparrow$ and $m_\downarrow$. In figure 4 in chapter 3.3 both are compared and we see that only for $m_\uparrow, m_\downarrow \to 1$ and $m_\uparrow \neq m_\downarrow$ there is a considerable difference. In the case $m_\uparrow = m_\downarrow$ both are equal for every a. This indicates that for increasing temperature the critical point in the (α, Q)-plane is mainly shifted to lower α with fixed Q, which is true to high accuracy as we will see in the next chapter.

4.2 Expansion for small temperatures

For $T > 0$ we have to solve the conditions $m_\uparrow(t+1) = m_\uparrow(t)$ and $m_\downarrow(t+1) = m_\downarrow(t)$ for $\alpha_\uparrow$ and $\alpha_\downarrow$ by using an approximation for low temperatures. As we saw before this is the interesting regime as $T_c \in [0, 0.5]$ for all possible network states. In the following we look at the first condition in order to get $\alpha_\uparrow$, the same can be done for $\alpha_\downarrow$. After some algebra given in the appendix we have the following approximation which is accurate for $T \ll \frac{2\sigma^2}{\mu_\uparrow - Q}$:

$$\begin{aligned} m_\uparrow(t+1) &= \int_{-\infty}^{\infty} \frac{1}{1+\exp\left(-2\beta x\right)}\, \rho_\uparrow(x)\, dx \approx \\ &\approx m_\uparrow(t+1)|_{T=0} - T^2 \frac{\mu_\uparrow - Q}{2\sqrt{2\pi}\sigma^3}\, e^{-(\frac{\mu_\uparrow - Q}{\sqrt{2}\sigma})^2} \frac{\pi^2}{12} \end{aligned} \tag{26}$$

By looking at the numerical solutions of $\alpha_\uparrow(Q)$ we see that the shape of the parabola is more or less unaltered, it is just shifted down. Therefore we make the ansatz $\alpha_\uparrow = \alpha_\uparrow|_{T=0} - f(T)$ where the function $f(T)$ accounts for the

downshift of $\alpha_\uparrow$. That means that we interpret the positive temperature as a source of noise in addition to that from the noncondensed patterns represented by α, as it was done in former studies (see e.g. [10]). Now we just take the dynamical equation for zero temperature and replace $\alpha_\uparrow|_{T=0}$ with $\alpha_\uparrow + f(T)$. We expand at $f(T) = 0$ to first order as f is small for small T:

$$\begin{aligned} m_\uparrow(t+1) &= \frac{1}{2}\left(1 + erf\left(\frac{\mu_\uparrow - Q}{\sqrt{2A(\alpha_\uparrow + f(T))}}\right)\right) \approx \\ &\approx m_\uparrow(t+1)|_{T=0} - \frac{\mu_\uparrow - Q}{2\sqrt{2\pi}\sigma^3}\, e^{-(\frac{\mu_\uparrow - Q}{\sqrt{2}\sigma})^2} A f(T) \end{aligned}$$

This term looks very similar to the expansion of the dynamical equation and by comparing the two we get the following approximation:

$$\alpha_{\uparrow_1} \approx \alpha_\uparrow|_{T=0} - \gamma_1 T^2 \qquad \alpha_{\downarrow_1} \approx \alpha_\downarrow|_{T=0} - \gamma_1 T^2 \qquad \text{where} \quad \gamma_1 = \frac{\pi^2}{12A} \tag{27}$$

The calculation for $\alpha_\downarrow$ gives the same correction to this order. We labeled the approximation with 1 because we can also think of another one.

By assuming a quadratic T-dependence of f we can make directly the ansatz $\alpha_\uparrow = \alpha_\uparrow|_{T=0} - \gamma_2 T^2$. Knowing the critical temperature we get the condition $\alpha_c|_{T=T_c} = \alpha_c|_{T=0} - \gamma_2 T_C^2 = 0$. After solving for γ_2 and inserting the expressions for α_c (13) and T_c (25) we have another approximation:

$$\begin{aligned} &\alpha_{\uparrow_2} \approx \alpha_\uparrow|_{T=0} - \gamma_2 T^2 \qquad\qquad \alpha_{\downarrow_2} \approx \alpha_\downarrow|_{T=0} - \gamma_2 T^2 \\ &\text{where} \qquad \gamma_2 = \frac{1}{4A} \frac{\left[\ln\left(\frac{1}{m_\uparrow} - 1\right) + \ln\left(\frac{1}{m_\downarrow} - 1\right)\right]^2}{(c_\uparrow + c_\downarrow)^2} \end{aligned} \tag{28}$$

In figure 10 both approximations are compared with the numerical solution and the case $T = 0$ for $m_\uparrow = 0.6$ and $m_\downarrow = m_{\downarrow_A} = 0.829$. Therefore we have $\gamma_1 = 0.822\frac{1}{A} > \gamma_2 = 0.679\frac{1}{A}$. The first approximation (27) gives the

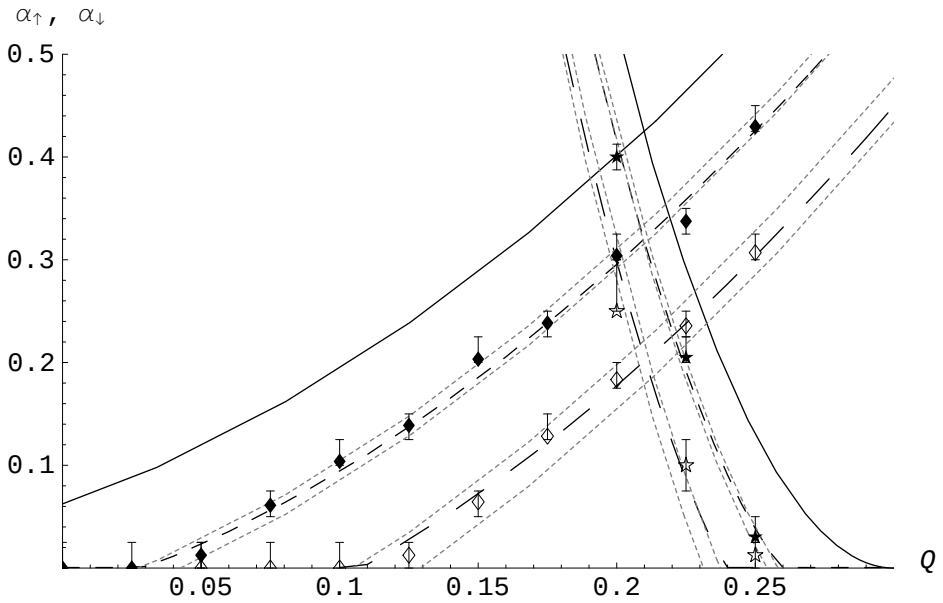

Figure 10: $\alpha_\uparrow(Q)$ and $\alpha_\downarrow(Q)$ for $m_\uparrow = 0.6$, $A = a = 0.3$ and $T = 0$ ——, $T = 0.2$ - - - (Data filled), $T = 0.3$ – – – (Data unfilled). Approximations —— (gray lines): $\alpha_{\uparrow_1}$ and $\alpha_{\downarrow_1}$ lower curves, $\alpha_{\uparrow_2}$ and $\alpha_{\downarrow_2}$ upper curves.

lower curves, the second (28) the upper ones. As expected (28) is a better approximation for small α as T_c was evaluated at $\alpha = 0$, and for $\alpha_{\uparrow_1}$ and $\alpha_{\downarrow_1}$ to be accurate we have the condition $\alpha \gg \max(\mu_\uparrow - Q, Q - \mu_\downarrow)\frac{1}{2A}T \approx 0.1$ from above. Therefore the first approximation is better for higher storage levels whereas the second one is less exact. We also see that with increasing temperature the ansatz of a simple downshift of the parabolas $\alpha_\uparrow(Q)$ and $\alpha_\downarrow(Q)$ becomes more and more inaccurate.

From the results for $\alpha_\uparrow$ and $\alpha_\downarrow$ we get the dependence of the critical storage capacity for $T > 0$, where both approximations are denoted by γ:

$$\alpha_c = \alpha_c|_{T=0} - \gamma T^2 \tag{29}$$

5 Equilibrium properties of the network

Now we are interested in the network in equilibrium, i.e. the fixed points of the dynamics. We can determine them by a numerical solution of the fixed point equations. Depending on the choice of the threshold we illustrate the fixed points and the basins of attraction in phase space. Systems with thresholds that constrain the dynamics show a simpler behavior than others. For simplicity we consider only zero temperature, but the qualitative analysis is easily extendable to cases with noisy updating.

In contrast to the dynamical analysis which is applicable for the first time step independent on the dilution level, the following considerations are only valid in the limit of strong dilution (see chapter 2.4).

5.1 Fixed point equations

To get the fixed points $\bar{m}_\uparrow$ and $\bar{m}_\downarrow$ of the dynamics we have to solve $m_\uparrow(t+1) = m_\uparrow(t)$ and $m_\downarrow(t+1) = m_\downarrow(t)$. After inserting the time evolution (8) we get the fixed point equations:

$$\bar{m}_\uparrow = \frac{1}{2}\left(1 + erf\left(\frac{\bar{\mu}_\uparrow - Q}{\sqrt{2}\bar{\sigma}}\right)\right) \qquad \bar{m}_\downarrow = \frac{1}{2}\left(1 + erf\left(\frac{Q - \bar{\mu}_\downarrow}{\sqrt{2}\bar{\sigma}}\right)\right) \tag{30}$$

Here $\bar{\mu}_\uparrow$, $\bar{\mu}_\downarrow$ and $\bar{\sigma}$ are taken at $m_\uparrow = \bar{m}_\uparrow$ and $m_\downarrow = \bar{m}_\downarrow$. These equations can be solved numerically for $\bar{m}_\uparrow$ and $\bar{m}_\downarrow$ and the solution depends on the choice of the threshold. The critical storage capacity in equilibrium $\bar{\alpha}_c$ is the value of α where the attractor fixed point with macroscopic overlap vanishes. This behavior is seen in figure 11.

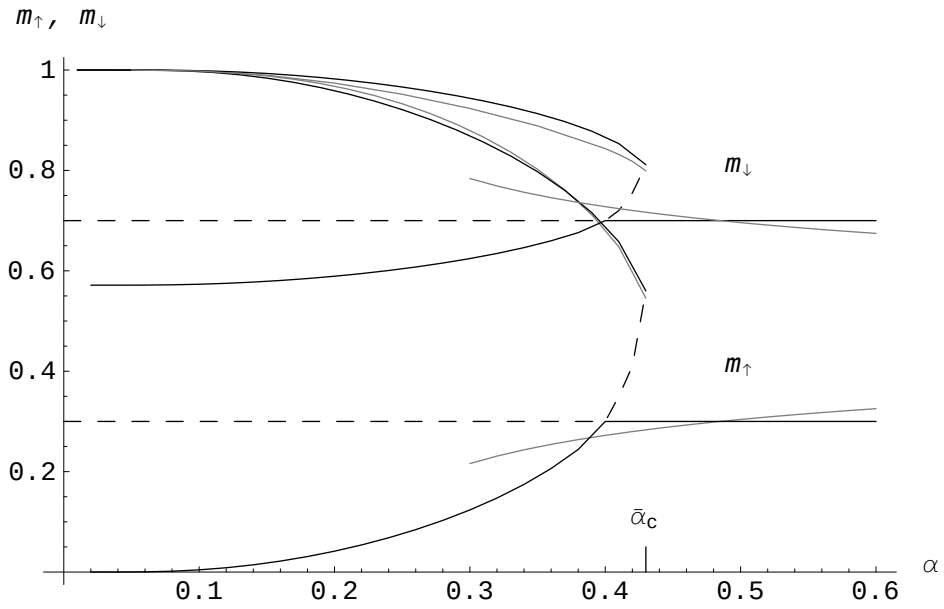

Figure 11: Fixed points $\bar{m}_\uparrow(\alpha)$ and $\bar{m}_\downarrow(\alpha)$ for $a = 0.3$, $T = 0$ and $Q = Q_a$ (- - - repellers, —— attractors), $Q = 0.2$ (——(gray) attractors).

For $Q = Q_a$ we plotted the attractor and repeller fixed points against the pattern loading. In addition to the attractor with macroscopic overlap there is another stable fixed point and the uncorrelated state is a repeller for $\alpha < \bar{\alpha}_c$. Near $\bar{\alpha}_c$ the last two become equal, the attractor destabilizes and the uncorrelated state becomes an attractor. At $\alpha = \bar{\alpha}_c$ the new repeller and the macroscopic attractor become equal, separated from the noncondensed stable fixed point. Both vanish, leaving the uncorrelated state as the only attractor. Because of this discontinuous behavior the underlying phase transition is of first order which is known from former studies [12, 15]. The behavior seen in this plot is the same for the three threshold choices Q_a, Q_o and Q_r that constrain the dynamics.

An exapmle for unconstrained dynamics is also plotted in figure 11 (gray lines).

The threshold is fixed during the retrieval process making the numerical solution more difficult. This is because the system can sample the whole phase space and the nature of the fixed points is not as clear as before, like shown in the next two chapters. Especially for small α the solutions become unstable and therefore we only plotted the attractors in figure 11 (gray lines) for $\alpha > 0.3$, choosing the optimal threshold for $m_\uparrow = m_\downarrow = 1$, $Q = 0.5 - a = 0.2$. Basically one has the same behavior, a first order phase transition at $\alpha = \bar{\alpha}_c$ from a macroscopic attractor to an uncorrelated state. In contrast to the case before where we had only one possible uncorrelated state due to constrained dynamics, the state is here only fixed by the relation $m_\uparrow + m_\downarrow = 1$. We see that it moves towards $(0.5, 0.5)$ for $\alpha \to \infty$.

From figure 11 one might get the impression that $\bar{\alpha}_c$ is the same for both cases, but in general this is not true. In the next chapters we try to explain the two cases applying some of the results of the dynamical considerations.

5.2 Constrained dynamics

If we choose one of the thresholds Q_a, Q_r or Q_o the network dynamics are constrained by a fixed relation between $m_\uparrow$ and $m_\downarrow$ and the accessible phase space is reduced to a line in the $(m_\uparrow, m_\downarrow)$-plane. Moreover we know that Q is always optimally chosen, i.e. $Q|_{\alpha_c} = Q_c$, and $\Delta m_\uparrow$ and $\Delta m_\downarrow$ have the same sign for all α. Hence it is useful to take the condition $\alpha = \alpha_c$ to get an implicit equation for a fixed point:

$$\alpha = \frac{(\bar{m}_\uparrow + \bar{m}_\downarrow - 1)^2}{(c_\uparrow|_{\bar{m}_\uparrow} + c_\downarrow|_{\bar{m}_\downarrow})^2 (a\bar{m}_\uparrow + (1-a)(1-\bar{m}_\downarrow))} \tag{31}$$

The numerical solution of this equation for different α gives the height lines of the function $\alpha_c(m_\uparrow, m_\downarrow)$. A fixed point for the network has to be somewhere on the height line corresponding to the loading α and in addition there is the fixed line $m_\uparrow + m_\downarrow = 1$ for uncorrelated states. In the region $m_\uparrow + m_\downarrow > 1$ we have the usual behavior where $m_\uparrow$ and $m_\downarrow$ only improve if $\alpha < \alpha_c$, i.e. the network state has to be inside the region enclosed by the corresponding height line. Whereas for $m_\uparrow + m_\downarrow < 1$ there is only improvement possible if $\alpha > \alpha_c$ as given in equation (14).

As the second condition to specify the fixed points we take the constraint relation between $m_\uparrow$ and $m_\downarrow$ and for the examples we studied before we have the following:

$$\begin{aligned} m_\downarrow = m_{\downarrow A} = 1 - \frac{a}{1-a}(1 - m_\uparrow) \qquad & \text{for } Q = Q_a \\ \frac{m_\uparrow}{m_\downarrow} = \frac{m_\uparrow(t=0)}{m_\downarrow(t=0)} = r \qquad & \text{for } Q = Q_r \\ \partial_{m_\downarrow} \alpha_c = 0 \quad (15) \qquad & \text{for } Q = Q_o \end{aligned} \tag{32}$$

In figure 12 we plot several height lines $\alpha = \alpha_c$ and $m_\uparrow + m_\downarrow = 1$ for $a = 0.3$. The choices of thresholds are Q_a, Q_o and Q_r with ratio $r = 1$ and the network dynamics are restricted to the corresponding line in the figure. The filled circles mark the attractor fixed points and the unfilled ones the repellers for $\alpha = 0.38$. They have no meaning for other pattern loadings and the other height lines are just there for illustration. The line for Q_r does not cross the height line $0.38 = \alpha_c$ and $m_\uparrow$ and $m_\downarrow$ always decrease for $m_\uparrow + m_\downarrow > 1$ and increase for $m_\uparrow + m_\downarrow < 1$. Therefore the point $(0.5, 0.5)$ is the only attractor if we choose $Q = Q_r$, because $\bar{\alpha}_c(Q_r) < 0.38$. The nature of the other fixed

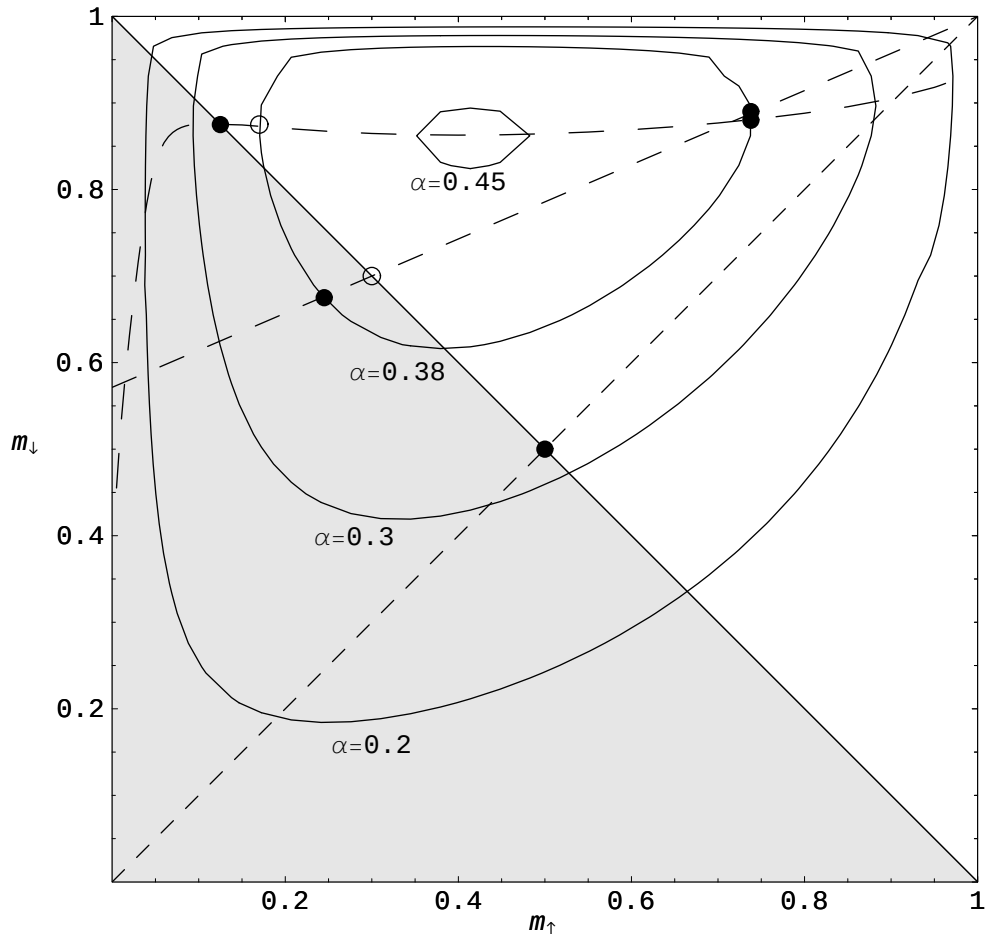

Figure 12: Fixed points in phase space with constrained dynamics for $a = 0.3$, and $T = 0$. —— height lines of $\alpha_c(m_\uparrow, m_\downarrow)$ and line for $m_\uparrow = m_\downarrow = 1$, shaded region for $m_\uparrow + m_\downarrow < 1$. Choices of thresholds: $Q_r(r = 1)$ - - -, Q_a – – –, Q_o — — —. Fixed points for $\alpha = 0.38$: attractor •, repeller ∘.

points is determined by the same arguments.

We see that for increasing α the attractor fixed point moves towards the line of noncondensed patterns and the basin of attraction becomes smaller. At the critical storage capacity for the particular choice of Q the basin of attraction vanishes and for $\alpha > \bar{\alpha}_c$ the attractor with macroscopic overlap is gone and the noncondensed fixed point is the new and only attractor. Hence at $\alpha = \bar{\alpha}_c$ the macroscopic equilibrium overlaps $m_\uparrow$ and $m_\downarrow$ jump to an uncorrelated state and therefore we have a first order phase transition which explains the solution of the fixed point equation in figure 11. Studies on the quality of retrieval and the basins of attraction were also done in [12, 15, 22].

In order to get a condition for $m_\uparrow$ and $m_\downarrow$ at the transition point we just insert the constraint relation for each threshold in the equation for α_c and maximize with respect to the remaining variable. For $Q = Q_o$ we already derived the location of the maximum storage capacity in (15) and (16). Using the abbreviations $c_\downarrow|_{m_{\downarrow A}} = c_{\downarrow A}$, $c_\uparrow|_{m_\uparrow = r m_\downarrow} = c_{\uparrow_r}$ and $A|_{m_\uparrow = r m_\downarrow} = A_r$ we get the following implicit conditions for the other two choices:

$$\begin{aligned} 1 &= (m_\uparrow - a)(\partial_{m_\uparrow} \ln(c_\uparrow + c_{\downarrow A})) \\ 2A_r &= (m_\downarrow(r+1) - 1)[2(\partial_{m_\downarrow} \ln(c_{\uparrow_r} + c_\downarrow))A_r + (a(r+1) - 1)] \end{aligned} \quad (33)$$

The second observable is easily determined by using the corresponding constraint equation (32). In Figure 13 we plotted $\bar{m}_\uparrow$ (black) and $\bar{m}_\downarrow$ (gray) at $\bar{\alpha}_c$ against the pattern activity a for the three threshold choices. For $Q = Q_r$ with $r = 1$ we have $\bar{m}_\uparrow = \bar{m}_\downarrow$ and therefore only one line. The other two choices give considerably smaller values for $\bar{m}_\uparrow$, but we have $\bar{m}_\uparrow \to 1$ for $a \to 0$, which means good retrieval qualities near $\bar{\alpha}_c$ for small activities in all three cases.

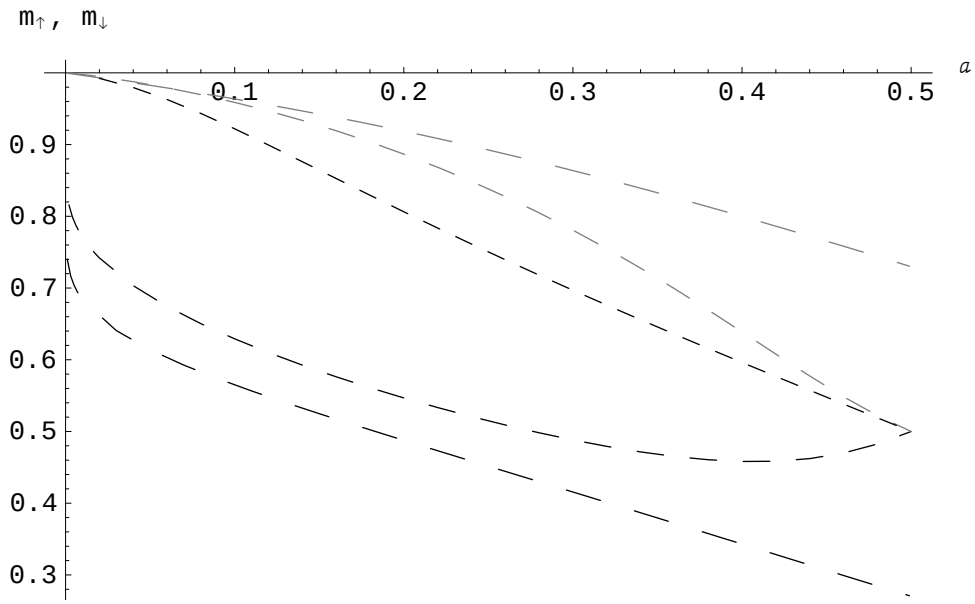

Figure 13: Comparison of location of phase transition $\bar{m}_\uparrow(a)|_{\bar{\alpha}_c}$ (black) and $\bar{m}_\downarrow(a)|_{\bar{\alpha}_c}$ (gray) for $Q = Q_r$ (r=1) - - -, $Q = Q_a$ – – –, $Q = Q_o$ — — —.

The values for $\bar{m}_\downarrow$ may be higher for Q_a and Q_o but this is not very important because the information is carried by the active neurons. Therefore the enhancement of the storage capacity for small a choosing Q_a and Q_o comes along with less accurate retrieval near $\bar{\alpha}_c$.

5.3 Unconstrained dynamics

For unconstrained dynamics the situation is more complicated because the system can sample the whole phase space. Given the threshold Q we can just take $Q = Q_c$ in addition to (31) as a second condition to specify the fixed points. This leads to:

$$Q = \left(\frac{c_\downarrow|_{\bar{m}_\downarrow}}{c_\uparrow|_{\bar{m}_\uparrow} + c_\downarrow|_{\bar{m}_\downarrow}} - a\right)(\bar{m}_\uparrow + \bar{m}_\downarrow - 1) \tag{34}$$

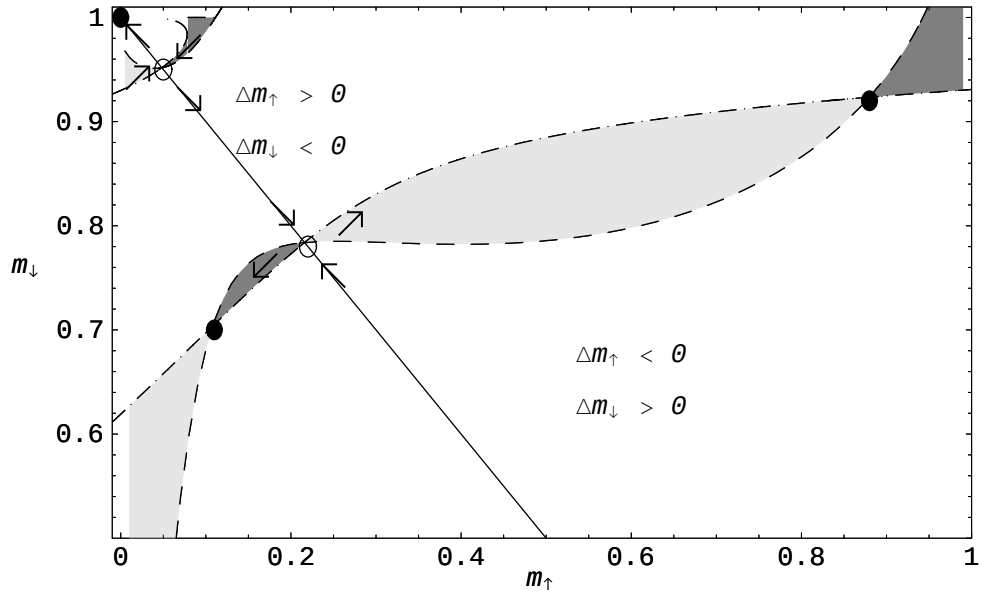

Figure 14: Fixed points in phase space with unconstrained dynamics for $a = 0.3$, $Q = 0.2$, $\alpha = 0.3$ and $T = 0$. $\Delta m_\uparrow = 0$ - - -, $\Delta m_\downarrow = 0$ - · - · -, Q_o. Fixed points: attractor •, saddle points ∘. Shaded regions: dark for $\Delta m_\uparrow$, $\Delta m_\downarrow < 0$, light for $\Delta m_\uparrow$, $\Delta m_\downarrow > 0$

But if we choose the dynamical threshold to be Q_c there is no second condition and the fixed points cannot be determined independent of the initial values $m_\uparrow(t = 0)$ and $m_\downarrow(t = 0)$. This is complicated to study and in addition the choice of Q_c is not very natural so we consider a case studied more often.

We take Q to be fixed during the retrieval process. Hence the threshold does not obey the condition $Q|_{\alpha_c} = Q_c$ and the improvements $\Delta m_\uparrow$ and $\Delta m_\downarrow$ do not have necessarily the same sign. Hence we could get the fixed points using (31) and (34) but the regions inside the height lines $\alpha = \alpha_c$ do not have the same meaning as before. The nature of the fixed points is more complex because the system can sample the whole phase space and to determine them, we look at the two improvements separately by using the conditions (20). We can solve

$\alpha_\uparrow = \alpha$ for $m_\downarrow$ and $\alpha_\downarrow = \alpha$ for $m_\uparrow$ and get:

$$\begin{aligned} m_\downarrow &= 1 - m_\uparrow + \frac{1}{1-a}\left[Q - c_\uparrow\frac{\alpha}{2} \overset{(+)}{\underset{-}{}} \frac{1}{2}\sqrt{c_\uparrow\alpha\left(c_\uparrow\alpha + 4m_\uparrow - 4Q\right)}\right] \\ m_\uparrow &= 1 - m_\downarrow + \frac{1}{a}\left[c_\downarrow\frac{\alpha}{2} - Q \overset{+}{\underset{(-)}{}} \frac{1}{2}\sqrt{c_\downarrow\alpha\left(c_\downarrow\alpha + 4(1-m_\downarrow) - 4Q\right)}\right] \end{aligned} \quad (35)$$

Like previously in (21) only one solution is valid respectively depending on the region in the $(m_\uparrow, m_\downarrow)$-plane. The condition without brackets () gives the macroscopic attractor. Considering this we end up with figure 14 for $a = 0.3$, $Q = 0.5 - a = 0.2$ and $\alpha = 0.3$.

The fixed points are at the intersections of the two curves and we have two attractors and two saddle points on the fixed line $m_\uparrow + m_\downarrow = 1$. The nature of the saddle points is indicated by the arrows in the plot. In addition to that the state $m_\uparrow = 0$ and $m_\downarrow = 1$ is always a fixed point for $Q > 0$ (in accordance with [13]) because here we have $\sigma = A = 0$ and $\mu_\uparrow = \mu_\downarrow = 0$. The regions where $m_\uparrow$ and $m_\downarrow$ improve simultaneously are shaded light and the regions where $\Delta m_\uparrow$ and $\Delta m_\downarrow$ are negative are shaded dark. Due to the discrete behavior of the system this is not a flow diagram and the basins of attraction of the fixed points cannot be seen in this plot.

For increasing α we get the same behavior like before with constrained dynamics. The attractor fixed point with $m_\uparrow + m_\downarrow < 1$ changes into a saddle point as it crosses the line of noncondensed patterns, and the macroscopic attractor moves towards the fixed line. Both meet at $\alpha = \bar{\alpha}_c$ and vanish, leaving the uncorrelated state as the only attractor besides the state with $m_\uparrow = 0$ and $m_\downarrow = 1$. This represents a first order phase transition.

For further increasing α the uncorrelated state moves towards $m_\uparrow = \frac{1}{2}$, $m_\downarrow = \frac{1}{2}$

which can be seen from equations (35). For $m_\uparrow + m_\downarrow = 1$ the conditions for the fixed point reduce to:

$$Q^2 = -m_\uparrow \alpha c_\uparrow \qquad \text{and} \qquad Q^2 = (1 - m_\downarrow)\alpha c_\downarrow$$

Because $m_\uparrow$ and $m_\downarrow$ are constrained to the interval $[0, 1]$ and Q is fixed we have to have $c_\uparrow, c_\downarrow \to 0$ in the limit $\alpha \to \infty$ which means $m_\uparrow, m_\downarrow \to \frac{1}{2}$. Expanding $c_\uparrow$ and $c_\downarrow$ at $\frac{1}{2}$ we get the leading behavior to first order:

$$m_\downarrow - \frac{1}{2} = \frac{1}{2} - m_\uparrow \approx \frac{1}{4} - \frac{1}{2}\sqrt{\frac{1}{4} - \frac{4Q^2}{\alpha\sqrt{2\pi}}} \qquad \text{for} \qquad \alpha \gg \frac{16Q^2}{\sqrt{2\pi}}$$

6 Conclusion

In this work we described a randomly diluted neural network model with variable pattern activity using a probabilistic approach to solve the one step dynamics with one condensed pattern. By carrying out the analysis as exactly as possible we were able to confirm with this relatively simple approach many of the previous results on this model, derived with different techniques and often restricted to special cases. Most of our results are valid for the whole range of the different parameters, except for the dilution level (see section 2.4), generalizing some of the former studies.

We got some new insight in the dynamical properties of the network, deriving the usual properties like critical storage capacity, information content and critical temperature for arbitrary network states not necessarily equilibrium. Special focus was on the resulting constraints on the threshold to optimize the network performance, a feature that was often overlooked in former studies. We used this to study the effects of choices of threshold functions during retrieval and in the equilibrium situation. The solution of the fixed point equations was explained in a qualitative way, focussing on the effect of different threshold functions in equilibrium.

We chose a neural network which is relatively simple, but has many realistic features for modelling the brain like explained before. Nevertheless the analysis is not restricted to this model, the probabilistic approach can be used for any kind of network with parallel updating. Further important is the site-independent description of the local field, in our case represented by $\mu_\uparrow$, $\mu_\downarrow$ and σ, excluding for example site-dependent thresholds. But the method is

easily extendable to more generalized models with graded response neurons (see [20, 21]), groups of patterns with different activities, any finite number of condensed patterns or sequential patterns.

It may be interesting to extend the presented analysis to these cases, making it possible to describe various network properties exactly with relatively easy computations. Especially the information on dynamical properties and restrictions on the parameters during retrieval may be very interesting. In addition to that the equilibrium study, presented here only in a qualitative way, may be extended to yield more quantitative results.

References

[1] D. O. Hebb. The organization of Behaviour. *Wiley, New York* (1949)

[2] M. F. Bear, B. W. Connors, M. A. Paradiso. Neuroscience: Exploring the Brain. *Williams & Wilkins, Baltimore* (1996)

[3] D. J. Amit. Modelling Brain Function. *Cambridge University Press, New York* (1989)

[4] J. Hertz, A. Krogh, R. G. Palmer. Introduction to the theory of neural computation. *Adison Wesley, Redwood City, CA* (1991)

[5] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **79**: 2554 - 2558 (1982)

[6] J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA* **81**: 3088 - 3092 (1984)

[7] W. A. Little. The existence of persistent states in the brain. *Math. Biosci.* **19**: 101 - 120 (1974)

[8] R. J. Glauber. Time-dependent statistics of the Ising model. *J. Math. Phys.* **4**: 294 - 307 (1963)

[9] D. J. Amit and H. Gutfreund, H. Sompolinsky. Spin-glass models of neural networks. *Phys. Rev. A*, **32(2)**: 1007 - 1018 (1985)

[10] D. J. Amit, H. Gutfreund and H. Sompolinsky. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Phys. Rev. Lett.* **55**: 1530 - 1533 (1985)

[11] D. J. Amit, H. Gutfreund and H. Sompolinsky. Information storage in neural networks with low level of activity. *Phys. Rev. A*, **35(5)**: 2293 - 2303 (1987)

[12] M. V. Tsodyks, M. V. Feigel'man. The Enhanced Storage Capacity in Neural Networks with Low Activity Level. *Europhys. Lett.*, **6**(2): 101 - 105 (1988)

[13] J. Buhmann, R. Divko and K. Schulten. Associative memory with high information content. *Phys. Rev. A*, **39**(5): 2689 - 2692 (1989)

[14] E. Gardner. The space of interactions in neural network models. *J. Phys. A: Math. Gen.* **21**: 257 - 270 (1988)

[15] M. V. Tsodyks. Associative Memory in Asymmetric Diluted Network with Low Level of Activity. *Europhys. Lett.*, **7**(3): 203 - 208 (1988)

[16] M. V. Tsodyks. Associative Memory in Neural Networks with the Hebbian Learning Rule. *Mod. Phys. Lett. B*, **3**(7): 555 - 560 (1989)

[17] R. Kree, A. Zippelius. Continuous-time dynamics of asymmetrically diluted neural networks. *Phys. Rev. A* **36(9)**: 4421 - 4427 (1987)

[18] B. Derrida, E. Gardner and A. Zippelius. An exactly solvable asymmetric neural network model. *Europhys. Lett.* **4(2)**: 167 - 173 (1987)

[19] M. Bouten, A. Engels, A. Komoda and R. Serneels. Quenched versus annealed dilution in neural networks. *J. Phys. A: Math. Gen.* **23**: 4643 - 4657 (1990)

[20] D. Bollé, B. Vinck and V. A. Zagrebnov. On the Parallel Dynamics of the Q-State Potts and Q-Ising Neural Networks. *J. Stat. Phys.*, **70**: 1099 - 1119 (1993)

[21] D. Bollé, G. M. Shim, B. Vinck and V. A. Zagrebnov. Retrieval and Chaos in Extremely Diluted Q-Ising Neural Networks. *J. Stat. Phys.*, **74**: 565 - 582 (1994)

[22] D. R. C. Dominguez and D. Bollé. Self-control in Sparsely Coded Networks. *Phys. Rev. Lett.*, **80**: 2961 (1998)

[23] K. Kitano and T. Aoyagi. Retrieval Dynamics of Neural Networks for Sparsely Coded Sequential Patterns. *cond-mat preprint 9805935*, (1998)

[24] M. Maravall. Sparsification from dilute connectivity in a neural network model of memory. *Submitted to Network*, (1998)

[25] J. R. Philip. The Function *inverfc* θ. *Australian J. of Phys.* **13**: 13 - 20 (1960)

Appendix

A.1 Corrections for finite N

To study finite size corrections we give the results for $\mu_\uparrow$, $\mu_\downarrow$ and σ for finite N in the case $\xi_i^\mu = 1$:

$$\begin{aligned}\mu_\uparrow &= (1-a)(m_\uparrow + m_\downarrow - 1)\frac{N-1}{N} \\ \sigma_\uparrow^2 &= (1-a)^2\frac{N-1}{N^2 c}\left[\frac{m_\uparrow}{a} + \frac{1-m_\downarrow}{1-a} - (m_\uparrow + m_\downarrow - 1)^2\right] + \frac{N-1}{N}\frac{p-1}{Nc}A\end{aligned}$$

The terms for $\xi_i^\mu = 0$ can simply be obtained by changing

$$m_\uparrow \longrightarrow 1 - m_\downarrow \qquad m_\downarrow \longrightarrow 1 - m_\uparrow \qquad a \longrightarrow 1 - a.$$

This is because the calculation only involves probabilities like $P(S_i = 1|\xi_i^\mu = 1) = m_\uparrow$ and $P(S_i = 1|\xi_i^\mu = 0) = 1 - m_\downarrow$ or combinations like $(1-a)P(\xi_i^\nu = 1) + aP(\xi_i^\nu = 0)$ which also obey this transformation. Therefore the only thing that changes is the first factor $(1-a)$ resp. $(1-a)^2$ of $\mu_\uparrow$ resp. $\sigma_\uparrow^2$.

In order to describe the system correctly with the ansatz of a gaussian distribution for the local field the numbers of active and inactive neurons must be large enough. With $a < 0.5$ we have to ensure $Na > 10$ as a minimum requirement.

In $\sigma_\uparrow$ there is a term $(1-a)^2\frac{m_\uparrow}{Nca}$ that becomes of order $\sigma \approx 10^{-1}$ for $a \approx \frac{10m_\uparrow}{Nc}$. For simulations with $N = 600$, $c = 1$ and $m_\uparrow = \mathcal{O}(1)$ the correction is of the same order as σ for $a \approx 0.02$ and can already be seen for $a = 0.1$. For high dilution the correction becomes even more important. In $\sigma_\downarrow$ there is no $\frac{1}{Na}$-correction because the factor in front is a^2 instead of $(1-a)^2$ so that the data for $\alpha_\downarrow$ can still be described by the results for infinite N.

Therefore we only present simulations with $a \geq 0.3$ so that the number of neurons can be relatively small. The analysis is nevertheless valid for the whole range of a and can be tested involving larger networks.

A.2 Expansions of *erf* and *inverf*

The error function and its inverse are defined as:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt \qquad \text{and} \qquad x = erf(inverf(x))$$

In reference [25] the inverse of the complementary error function $erfc = 1 - erf$ is studied and we give the results obtained there using $inverf(x) = inverfc(1 - x)$. Expansion around $x = 0$:

$$\begin{aligned} \frac{\sqrt{\pi}}{2} erf(x) &= x - \frac{x^3}{3} + \frac{x^5}{2!5} - \ldots = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)k!} \\ \frac{2}{\sqrt{\pi}} inverf(x) &= x + \frac{\pi x^3}{12} + \frac{7\pi^2 x^5}{480} + \frac{127\pi^3 x^7}{40320} \ldots \end{aligned}$$

Asymptotic behavior:

$$\begin{aligned} 1 - erf(x) &\approx \frac{1}{\sqrt{\pi} x} e^{-x^2} && \text{for } x \to \infty \\ inverf(x) &\approx \sqrt{-\ln(1-x) - \frac{1}{2} \ln[\pi(-\ln(1-x) - \ldots} && \text{for } x \to 1 \end{aligned}$$

A.3 Calculation for positive temperature

In section 4.2 we want to calculate $\alpha_\uparrow$ and $\alpha_\downarrow$ from the dynamical equations for $T > 0$. In the following we perform the calculation for $\alpha_\uparrow$ where $\alpha_\downarrow$ can be evaluated in exactly the same way. First we split the integral in the dynamical equation (8):

$$m_\uparrow(t+1) = \int_{-\infty}^{0} \frac{\exp(2\beta x)}{1 + \exp(2\beta x)} \rho_\uparrow(x)\, dx + \int_0^{\infty} \frac{1}{1 + \exp(-2\beta x)} \rho_\uparrow(x)\, dx$$

By using the geometric series $\frac{1}{1+y} = \sum_{n=0}^{\infty}(-1)^n y^n$ for $y < 1$ and replacing x by $-x$ in the first integral we get:

$$m_\uparrow(t+1) = \sum_{n=0}^{\infty}(-1)^n \int_0^\infty \left[e^{-2\beta x n}\rho_\uparrow(x) - e^{-2\beta x(n+1)}\rho_\uparrow(-x)\right] dx$$

Inserting $\rho_\uparrow(x)$ from equation (7) we have to calculate two gaussian integrals of the form

$$\int_0^\infty e^{-q_1 x - q_2 x^2}\, dx = \frac{\sqrt{\pi}}{2\sqrt{q_2}} \exp(\frac{q_1^2}{4q_2}) \left[1 - erf\left(\frac{q_1}{2\sqrt{q_2}}\right)\right].$$

After this we see that $m_\uparrow(t+1)|_{T=0}$ is the term in the sum for $n = 0$ and we get the expression

$$m_\uparrow(t+1) = \frac{1}{2}\left(1 + erf\left(\frac{\mu_\uparrow - Q}{\sqrt{2}\sigma}\right)\right) - \Delta = m_\uparrow(t+1)|_{T=0} - \Delta$$

with the correction term

$$\Delta = \frac{1}{2}\sum_{n=1}^{\infty}(-1)^n e^{2\beta^2\sigma^2 n^2} \left[e^{2\beta(\mu_\uparrow - Q)n}\left(1 - erf\left(\sqrt{2}\sigma\beta n + \tfrac{\mu_\uparrow - Q}{\sqrt{2}\sigma}\right)\right) - \right.$$
$$\left. - e^{-2\beta(\mu_\uparrow - Q)n}\left(1 - erf\left(\sqrt{2}\sigma\beta n - \tfrac{\mu_\uparrow - Q}{\sqrt{2}\sigma}\right)\right)\right].$$

This expression is still exact but in order to do the sum we have use the approximation $1 - erf(x) = \frac{\exp(-x^2)}{\sqrt{\pi}x}$ (see appendix A.2). This is accurate for $x \gg 1$ which means in our case $\beta \gg \frac{\mu_\uparrow - Q}{2\sigma^2}$. The expression for Δ simplifies to

$$\Delta = \frac{\mu_\uparrow - Q}{2\sqrt{2\pi}\beta^2\sigma^3}\, e^{-\left(\frac{\mu_\uparrow - Q}{\sqrt{2}\sigma}\right)^2} \sum_{n=1}^{\infty}(-1)^n \frac{1}{n^2 - \left(\frac{\mu_\uparrow - Q}{2\beta\sigma^2}\right)^2}.$$

The sum can be done exactly and we end up with the following, writing $T = \frac{1}{\beta}$:

$$\begin{aligned} m_\uparrow(t+1) &= m_\uparrow(t+1)|_{T=0} - T^2 \frac{\mu_\uparrow - Q}{2\sqrt{2\pi}\sigma^3}\, e^{-(\frac{\mu_\uparrow - Q}{\sqrt{2}\sigma})^2} \Gamma(T\pi\frac{\mu_\uparrow - Q}{2\sigma^2}) \\ \Gamma(x) &= \frac{x - \sin x}{2x^2 \sin x} = \frac{\pi^2}{12} + \frac{7\pi^4}{720}x^2 + \ldots \end{aligned} \tag{36}$$

For $T \ll \frac{2\sigma^2}{\mu_\uparrow - Q}$ the argument of Γ is small and we use the zero order approximation $\Gamma \approx \frac{\pi^2}{12}$.

A.4 Patterns with fixed activity

We also studied the network when the pattern activity is fixed to a. In this case the pattern elements ξ_i^ν are no longer IIDRVs, but we randomly choose exactly Na active elements of every pattern. The results for the mean and the variance of the local field yield an interesting difference:

$$\begin{aligned} \mu_\uparrow &= (1-a)(m_\uparrow + m_\downarrow - 1) - \alpha A \\ \mu_\downarrow &= -a(m_\uparrow + m_\downarrow - 1) - \alpha A \\ \sigma_\uparrow^2 = \sigma_\downarrow^2 = \sigma^2 &= \alpha A(1 - cA) \end{aligned} \tag{37}$$

We see that $\mu_\uparrow$ and $\mu_\downarrow$ depend on α and that σ^2 depends on the connectivity c. This is therefore also true for the critical storage capacity and as σ is slightly smaller than before α_c is enhanced. However, in the limit of strong dilution, $c \to 0$ (see section 2.4) the dependence on c vanishes. If we look at $\alpha_\uparrow$ and $\alpha_\downarrow$ we have to solve the conditions (20) for α.

$$\alpha_\uparrow = \frac{(\mu_\uparrow|_{\alpha=0} - \alpha_\uparrow A - Q)^2}{{c_\uparrow}^2 A(1 - Ac)} \qquad \alpha_\downarrow = \frac{(Q - \mu_\downarrow|_{\alpha=0} + \alpha_\downarrow A)^2}{{c_\downarrow}^2 A(1 - Ac)} \tag{38}$$

We still get parabolas but with twisted axes, $\alpha A + Q = const.$, which is shown in figure 15 and confirmed by simulations. There we have $m_\uparrow = 0.5$ and the parabola for $\alpha_\uparrow$ reduces to its axis. Due to the twist of the axis both branches of $\alpha_\downarrow$ become valid solutions and the upper branch is a lower bound for α, yielding improvement of $m_\downarrow$ for unlimited α. We also see that a change in connectivity affects the picture because $\alpha_\uparrow$ and $\alpha_\downarrow$ are no longer independent of c.

Therefore Q_c also depends on α and has the same slope like the axes of $\alpha_\uparrow$

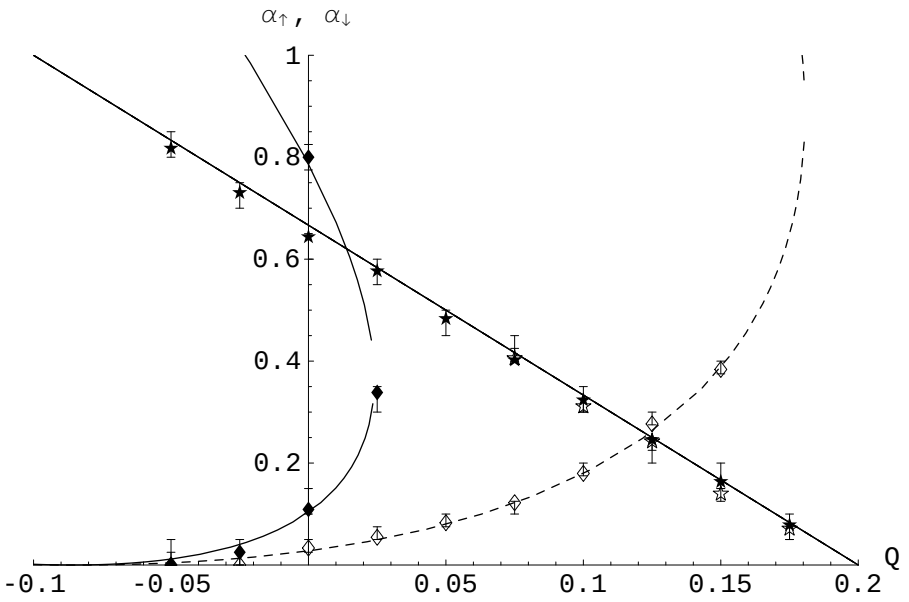

Figure 15: Effect of fixed pattern activity: $\alpha_\uparrow$ and $\alpha_\downarrow$ for $m_\uparrow = 0.5$, $A = a = 0.3$, $T = 0$ and $c = 1$ —— (data filled), $c = 0.5$ - - - (data unfilled).

and $\alpha_\downarrow$ in the (α, Q)-plane. The rest of the analysis of thresholds can be done exactly as before. For positive temperature the parabolas move down along the twisted axes. As the critical temperature was derived at $\alpha = 0$ the expression for T_c is exactly the same as before. The equilibrium analysis of this system can be done in the same way and will show qualitatively the same behavior.