WWU
MÜNSTER

› Storage Capacities of Sparse Associative Memories

Judith Heusel

2017

wissen.leben
WWU Münster

Judith Heusel

Storage Capacities of Sparse Associative Memories

2017

Mathematik

Storage Capacities of Sparse Associative Memories

Inauguraldissertation
zur Erlangung des akademischen Grades eines Doktors
der Naturwissenschaften durch den Fachbereich
Mathematik und Informatik
der Westfälischen Wilhelms-Universität Münster

vorgelegt von
Judith Heusel
aus Mainz
- 2017 -

Dekan:	Prof. Dr. Xiaoyi Jiang
Erster Gutachter:	Prof. Dr. Matthias Löwe
Zweiter Gutachter:	Prof. Dr. Gerold Alsmeyer
Tag der mündlichen Prüfung:	
Tag der Promotion:	

Abstract

Models of autoassociative memories process information contained in a certain number of states of a neural network and are afterwards able to associate output states with given input states of this neural network. An important performance aspect of these models is the ability to reproduce the input configuration as ouput if it is one of the learned neural network states. This is measured by the capacity which indicates the maximal number of stored patterns until the described property is lost. A further performance aspect is the capability of reconstructing or correcting partially erased or corrupted versions of the stored configurations.

Various models of associative memories storing sparse patterns are studied. The storage capacities and error correcting behaviour are analysed with regard to the application of different storing mechanisms, retrieval dynamics and probability distributions of the stored patterns. We prove the existence of sharp bounds on the storage capacity and error correction, depending on the remaining relevant parameters of the models. Characteristical properties of the different models are examinend.

Acknowledgements

Ich danke in erster Linie meinem Doktorvater Matthias Löwe für seine geduldige, unterstützende, qualifizierte und humorvolle Betreuung, sein Vertrauen und die Möglichkeit, diese Arbeit bei ihm zu schreiben.

Merci aussi à Franck Vermet et Vincent Gripon pour une coopération créative et productive. Weiterer Dank gilt Gerold Alsmeyer für hilfreiche Verbesserungsvorschläge an dieser Arbeit. Danken möchte ich auch meiner gesamten Arbeitsgruppe, in der ich mich immer sehr wohl gefühlt habe.

Vielen Dank an Marten, Tim, Tim und Michael für wertvollen Rat bei so vielen Fragen und viel Spaß im Arbeitsalltag. Schließlich danke ich meinen Eltern, Schwestern und Großeltern, die in vielerlei Hinsicht zum Entstehen dieser Arbeit beigetragen haben und außerdem meinen Freunden, insbesondere Alex, Claudi, Eva, Hanna, Hathumar, Hendrik, Imke, Jana, Janna, Jessica, Lisa, Miriam, Nele, Nils, Sarah, Sina, Tobias und Theresa.

Contents

1 Introduction

With the objective to model the memory activities of a human brain various models of neural networks have been introduced and analysed during the last decades. A neural network usually consists of a simple graph $G = (V, E)$: the elements of the vertex set V are called neurons and are connected through a set of edges E. These take the role of the synapses. To model neural activities the neurons can take values in a state space S containing at least two elements. A state of the network is then described by a vector $\sigma = (\sigma_i)_{i \in V} \in S^V$, $\sigma_i \in S$ denoting the state of neuron $i \in V$. Additionally, for each edge $e \in E$, there is a variable J_e to be specified later, describing a certain relationship between the two neurons connected by e and called synaptic efficacy or synaptic weight (cf. [30]). It is possible to use additional functions indicated by the edges also containing information concerning the relationship between the connected neurons.

To make the neural network a model of an (associative) memory, it is supposed to store a certain amount of information. This information comes in the form of M states of the network $\xi^\mu \in S^V$, $1 \le \mu \le M$ (called patterns or messages). Hebb, a psychologist, postulated in [19] that learned information in a brain is reflected by the simultaneous firing of certain neurons and this in turn strengthens the connecting synapses. This so called Hebbian learning (see [41]) is a guiding principle in many models when constructing the J_e. The patterns are usually stored by determining the value J_e through a calculation rule, depending on the ξ^μ and this rule in turn describes the outcome of a Hebbian learning process. The J_e are called synaptic efficacies or synaptic weights. In particular, the value J_e is measurable with respect to ξ_i^μ, ξ_j^μ, $1 \le \mu \le M$, if $e = \{i, j\}$, $i, j \in V$. This property is called locality of the synaptic weights, see [30].

A memory is called associative if, confronted with some input pattern, it can produce an output pattern: in heteroassociative memories, the output patterns are of other forms than the input patterns, in autoassociative memories they have the same form (see [45]). An autoassociative or content-addressable memory should be able to recall a stored pattern, given only a part of it (see [25]) and to correct corrupted patterns (see [33]). To reach this, a dynamics T is defined on S^V. The dynamics, confronted with an input pattern $\sigma \in S^V$, can either update all the neurons in one step (synchronous updating/ parallel dynamics) or one after another (asynchronous updating/ sequential dynamics).

The model that attracted lots of interest to this research area and is the standard model for an associative memory, see [21], was introduced by Hopfield in 1982 and is nowadays called the Hopfield model. For some $N \in \mathbb{N}$, one considers the vertex set $V = \{1, \dots N\}$ together with the state space $S = \{-1, 1\}$. The state 1 signifies that a neuron is active (also called excited) and -1 signifies that it is inactive (also called quiescent). There are $M(N)$ patterns $\xi^1, \dots \xi^M \in S^V$ stored at random, independently and identically distributed according to the uniform distribution on S^V. The underlying graph is the complete graph on V and the synaptic weights are determined by the Hebbian learning rule

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

Given an input pattern σ, the local field at a neuron i is defined by

$$S_i(\sigma) := \sum_{j \neq i} J_{ij} \sigma_j.$$

The parallel dynamics $T = (T_1, \ldots, T_N)$ maps an input spin configuration $\sigma \in \{-1, 1\}^N$ to $(T_1(\sigma), \ldots, T_N(\sigma))$, with $T_i(\sigma) = \operatorname{sgn}(S_i(\sigma))$. Alternatively, a sequential dynamics $\bar{T} = \bar{T}_{\tau(N)} \circ \ldots \circ \bar{T}_{\tau(1)}$ can be used, with $\bar{T}_i(\sigma) = (\sigma_1, \ldots, \sigma_{i-1}, T_i(\sigma), \sigma_{i+1}, \ldots, \sigma_N)$ and an arbitrary permutation τ of the set V. The model can also be interpreted as a spin glass as introduced by Pastur and Figotin in [39]. The space S^V is thus also called configuration space and the states of the neurons spins. Concretely, the dynamics is related to a Hamiltonian (energy) function defined on $\{-1, 1\}^N$, namely

$$H(\sigma) := -\frac{1}{2} \sum_{i \neq j} J_{ij} \sigma_i \sigma_j = -\frac{1}{2} \sum_{i=1}^{N} \sigma_i S_i(\sigma)$$

such that, using the sequential dynamics, each step of the dynamics decreases the energy. Then the sequential dynamics converges to a local minimum of the Hamiltonian and a stored pattern is stable, this is, a fixed point of the dynamics, if it is a local minimum of the Hamiltonian. It can be interesting to analyse the energy landscape to determine the basins of attraction of the local minima: if the input pattern is within the basin of attraction of a local minimum, the dynamics started in this input pattern converges to the local minimum.

The maximal number of patterns that can be stored and correctly memorised is called capacity. There are different notions of capacity, depending on the exact requirements one has, e. g., if one claims that the stored patterns shall be perfectly memorised (they are fixed points of the dynamics or local minima of the Hamiltonian, respectively), or if one is willing to accept small errors (the patterns just have to be located close to local minima of the energy function, measured in Hamming distance). Another aspect of the notion of storage capacity is the requirement that all patterns are fixed points of the dynamics, in contrast to the less restrictive requirement that an arbitrary but fixed stored pattern is stable. A further, interesting question is if corrupted versions of the stored patterns are recovered by the dynamics or at least attracted to a local minimum near the stored pattern, respectively.

Hopfield suggested, supported by computer simulations, that the Hopfield model can store up to $\alpha^* \cdot N$ patterns, with $\alpha^* \approx 0.15$, and that all learned messages will be forgotten (the stored patterns are no longer close to fixed points of the dynamics) if $M(N) > \alpha^* \cdot N$. Using the non-rigorous replica trick, Amit et al. ([3], [4]) obtained similar results as Hopfield ($\alpha^* \approx 0.138$), namely that the Hopfield model can store up to αN patterns, with a similar value for α, if small errors are accepted. Newman in [35] was able to prove rigorously that the model is able to store αN patterns, if small errors are tolerated, giving a lower bound on $\alpha \geq 0.056$. This bound was later improved by Loukianova (see [28],

[29]) and Talagrand (see [42], [43]), up to 0.08. Concerning the more restrictive case where the patterns have to be fixed points of the dynamics and corrupted patterns have to be recovered exactly, McEliece et al. showed rigorously in [34] that a fixed but arbitrary pattern is exactly memorised and in addition a corrupted version of a stored pattern with a number of ϱN errors is corrected by the dynamics if $\varrho < \frac{1}{2}$ and the number of stored patterns is at most $\alpha N/\log(N)$ with α depending on ϱ, where $\alpha = \frac{1}{2}$ for $\varrho = 0$; a similar results holds for all the patterns if $\alpha < \frac{1}{4}$. Bovier, also considering the perfect retrieval, was able to show that this bound is sharp and the network is not able to recognise a stored pattern, with probability converging to 1 if $\alpha > \frac{1}{2}$ (see [11]).

Inspired by results of biological research concerning the human brain observing that only few neurons are involved when something is memorised, sparse models were proposed. On a vertex set of $N \in \mathbb{N}$ neurons with a state space S containing states that are either referred to as active or as inactive, with possibly more than one active state, these models store patterns with only a few active neurons per pattern, compared to the number N. Willshaw et al. [44], Palm [37]/ Palm and Sommer [38], Amari [2] and Okada [36], among others, proposed and/or analysed sparse models and came to the conclusion that their capacity is much larger than the one of the original Hopfield model. The grade of sparsity is measured by the activity of the stored patterns, this is the fraction of the expected number of activated neurons in a stored pattern divided by the total number of neurons N (see [2]). The models mentioned above either use $S = \{0, 1\}$ or $S = \{-1, 1\}$, with the 0 or respectively the -1 representing the inactive state. The two possible states -1 and 1 of the neurons in the Hopfield model described so far are interchangeable. If the inactive state is represented by a 0 instead of a -1, the synaptic efficacy belonging to the edge connecting the two neurons i and j is only influenced by such messages in which i and j are both activated. In the -1 case, the synaptic efficacy takes every stored message into account. Besides Okada, all the researchers named above considered the state space $\{0, 1\}$. On the edges of the complete graph on the vertex set $V = \{1, \dots, N\}$ they either used additive synaptic weights as Hopfield (Amari, Okada) or a binary learning rule (Willshaw et al., Palm, Palm and Sommer) with binary synaptic efficacies $J_{ij} \in \{0, 1\}$ indicating whether the two neurons belonging to an edge are at least once together activated in one of the stored patterns (then $J_{ij} = 1$) or not ($J_{ij} = 0$).

Initiated by the publication of several variations of a new sparse model of associative memory, see e.g. [17], we rigorously analyse two of these models with an activity depending on N and converging to 0. Chapters 2 and 4 are devoted to the models of Amari and Willshaw, both using the state space $S = \{0, 1\}$, the first one using the additive and the second one the binary learning rule. In this context the Willshaw model can be defined with two different dynamics. The first one uses a threshold algorithm that activates neurons whose local field, defined as in Hopfield's model, but with the binary synaptic weights, exceeds a given threshold. The second one is a Winner takes all (WTA) algorithm as proposed by Gripon, Berrou et al. for their model ([22]), activating a certain number of neurons that possess the highest local field. Both models, the one of Amari and the Willshaw model, can be considered in a version where stored patterns consist of a fixed number of active neurons per message and in a version where the states of the neurons of a stored message are independent and identically distributed random variables taking the active state with probability p_N, where p_N is the activity in the network with N neurons and tends to 0 as N goes to infinity.

The binary state space is expanded to $S = \{-1, 0, 1\}$, now with a 0 corresponding to the inactive state of a neuron and both, a 1 and a -1, corresponding to active states. We analyse two of these models in a sparse version, both using the vertex set $\{1, \dots N\}$ for some $N \in \mathbb{N}$ and the complete graph K_N on these neurons. The $M(N)$ stored patterns are chosen at random and for each $i \in \{1, \dots, N\}$ and $\mu \leq M(N)$ the i-th spin of the μ-th stored pattern is 0 with probability $1 - p_N$ and ± 1 with probability $1/2 \cdot p_N$, where p_N is the activity of the model using N neurons. As in the binary models, it is also possible to choose the stored patterns uniformly among the set of all spin configurations $\sigma \in \{-1, 0, 1\}^N$ with a fixed number $N \cdot p_N$ of active neurons.

The first model uses an additive learning rule like the Hopfield model. It is called Ternary simple model, was introduced and analysed for fixed activity p by Löwe and Vermet in [31] and is examined in Chapter 3 in the sparse case with activity $p_N = \log(N)/N$ depending on N. A more complicated model for the ternary state space is the model of Blume, Emery and Griffiths, proposed in [7] as a spin glass model to analyse the behaviour of liquid $\mathrm{He}^3 - \mathrm{He}^4$ mixtures. Using an additive learning rule with $J_{ij} = \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu$, the dynamics takes the variables K_{ij}, $i \neq j$, into account besides the synaptic weights. The variable K_{ij} counts common activities with a positive impact and common inactivities of the neurons i and j per message and with a negative impact messages in which either i or j is active and the other one is inactive. A neuron i is activated by the dynamics, if the absolute value of its local field $S_i(\sigma) = \sum_{j \neq i} \sigma_j J_{ij}$ at the input pattern σ, defined as in Amari's model and using the synaptic efficacies, exceeds the value of the function $\theta_i(\sigma) = \sum_{j \neq i} (\sigma_j)^2 K_{ij}$. The neuron is in this case assigned to the sign of the local field. While Bollé and Verbeiren in [8] and Bollé, Castillo and Shim in [9] stated that this model performs best compared to other three-state-networks and based their assertions on the non-rigorous replica theory, Löwe and Vermet in [31] rigorously showed the existence of lower bounds on the storage capacity of these two ternary models, for fixed p, observing that the capacity increases for small p in the Ternary simple model, whereas it decreases as p gets smaller in the BEG model. They supposed that for small p the BEG model is outperformed by the Ternary simple model. We will show that in the sparse case with activity $p_N = \log(N)/N$ depending on N the BEG model is indeed outperformed by the Ternary simple model, but can slightly be changed and then offers the second best capacity of the models examinend in this thesis. We prove that stored patterns are not stable in the original model in the considered sparse case and propose a modification that allows to store a high number of messages, compared to the models in chapters 2 - 4.

Over the past few years, a series of papers has been published by Gripon, Berrou et al. (see e. g., [1], [16], [17], [22], [23], [24], [45]): they deal with several variations of models whose special feature is a cluster structure. Inspired by the brain, a pattern to be memorised addresses neurons of different parts of the neural net. The number of clusters c is small compared to the number of neurons l contained in a cluster. They use $c = \log(l)$ and $N = cl$ neurons in total. The edge set includes all possible edges that connect two neurons which do not belong to the same cluster and all self-loops (an edge connecting a neuron to itself). The edges are usually allocated to binary synaptic weights, but other possibilities are imaginable, as proposed in [20]. The active neurons of an input pattern send signals through active edges which means that the synaptic weight of the edge is 1. The decision which edges are active is made by storing a number of patterns, and an edge connecting two neurons is active if and only if these two neurons are at least once part of (excited in) the same stored message. There are several ways to decide which neurons are activated after a step of the dynamics. Usually the decision is made per cluster by a WTA algorithm, in [17] still by counting all incoming signals and determining the neuron that collects the highest number of signals, later (e. g., [1], [23], [24], [45]) by counting, per neuron, the number of clusters from which it receives at least one signal. The second decision rule is more plausible because it prevents a high influence of errors and accounts for the fact that there is only one active neuron per cluster in a stored message. According to the authors and simulations they made, this decision rule shows a better error correcting behaviour (see [45]). It is also possible to use a threshold algorithm per cluster (see [23]), here also counting the number of clusters from which at least one signal is obtained instead of the total number of signals. In particular, the cluster structure in combination with the self-loops can be exploitet to guarantee stability of all stored patterns, independent of their number. Chapter 6 deals with different versions of this model, called Gripon-Berrou model or GB model, for short.

The authors state that their model performs better than the Hopfield model, which is indeed the case, but all sparse models analysed in this work possess much higher capacities than the Hopfield model. We show that the number of stored patterns can in all other models be at most $M(N) = \alpha N^2/\log(N)^2$ with a constant α to be determined and depending on the model, whereas the Hopfield model can only remember $\alpha N/\log(N)$ or αN patterns, depending on the notion of capacity. We have to remark that some versions of the models of chapters 2-5, namely the ones with independent and identically distributed spins, can only preserve this high capacity if the requirement is that an arbitrary chosen pattern is stable with probability converging to 1. As we will show in the corresponding chapters, this high capacity is lost if one claims that all patterns have to be stable with probability tending to 1. On the other hand, we will see that in the versions with fixed number of active neurons per stored pattern and in the GB model, the order of the capacity can also be maintained under the latter condition on the stability.

We give a detailed analysis of the capacity and the error correcting abilities of the different models. If $M = \alpha N^2/\log(N)^2$ patterns are stored, the variable α is called capacity variable. We show that there are sharp bounds on the capacity variables for each model besides the GB model with WTA algorithm (called SUM-of-MAX rule) and a version of the GB model with threshold algorithm, i.e., there is an $\alpha^* > 0$ depending on the model, such that for $\alpha < \alpha^*$ and appropriate choice of the threshold, if there is one, $\alpha N^2/\log(N)^2$ patterns can be stored such that an arbitrary one is a fixed point of the dynamics, with probability converging to 1. If $\alpha > \alpha^*$, an arbitrary stored pattern is instable with probability concerging to 1 independent of the choice of the threshold. We prove that the GB model with SUM-of-MAX rule and the GB model with threshold algorithm, using an appropriate threshold, can store all patterns such that they are stable. Of course it is not reasonable to store all possible patterns and if one expects the model to correct a certain number of errors in a stored pattern, the number of stored patterns is also bounded in these models.

Amari's model, the Ternary simple model, the BEG model in the adapted version used in this thesis, the Willshaw model with threshold dynamics and two versions of the GB model use thresholds in their dynamics. These thresholds are determined by a threshold variable γ. We determine the sets of admissible threshold variables and show that in each considered model in its version with i.i.d. spins, there are sharp bounds $\alpha^*(\gamma)$ on the capacity variables in dependence on the used threshold variable, such that a number $M = \alpha N^2/\log(N)^2$ of stored patterns leads to instability of an arbitrary stored pattern if $\alpha > \alpha^*(\gamma)$ and to stability, if $\alpha < \alpha^*(\gamma)$, both with high probability. Besides the stability of arbitrary stored patterns we also consider the stability of all stored patterns. We observe that in the models of chapters 2-5 it is advantageous to choose the stored patterns uniformly among the set of patterns with fixed number Np_N of active neurons instead of using independent and identically distributed random variables to determine the spins of a stored message. The capacity of the latter versions (i.i.d. spins) of the models decreases drastically if one wants all patterns to be stable, compared to the first versions of the models. The model in Chapter 6 has a fixed number of active neurons per stored message by construction.

Concerning the error correcting abilities of the models, we can give sharp bounds on α and γ (depending on the type of error) for the one step retrieval of corrupted patterns. It turns out that there are two principal types of errors that have to be distinguished: corruption of active and of inactive neurons, respectively. These influence the capacity in different ways, if one wants to correct a certain number of errors of this kind. The results obtained in the different chapters are summarised and compared in Chapter 7.

Notations and Conventions

In this thesis, the sizes of the networks depend on a natural number N as do the random variables in the various chapters. The results are asymptotic. An event is called to happen with high probability, if its probability tends to 1 as N tends to infinity. All random variables are defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$; notations of functions, synaptic efficacies, local fields, dynamics and random variables are usually made per chapter. In case of ambiguity, their affilitation is indicated more precisely. The following notations are used: for two functions f, $g : \mathbb{R} \to \mathbb{R}$, we write $g = \mathcal{O}(f)$, if

$$\exists C > 0, \exists x^* \in \mathbb{R} : \forall x > x^* : |g(x)| \leq C|f(x)|,$$

and $g = o(f)$, if

$$\lim_{x \to \infty} \frac{|g(x)|}{|f(x)|} = 0.$$

In addition, we write $f \approx g$ as $x \to \infty$, if

$$\lim_{x \to \infty} \frac{g(x)}{f(x)} = 1.$$

Finally, a function or a sequence of sets is called increasing or decreasing, respectively, if it is monotonically (not necessarily strictly) increasing/decreasing. Strict monotony is in each case indicated specifically.

2 Amari's Model

The first model relies on the complete graph built on the vertex set $V = \{1, \ldots, N\}$, $N \in \mathbb{N}$. The set of neuronal states is $S = \{0, 1\}$: a 1 corresponds to an active and a 0 to an inactive state, respectively. The Hopfield model (see [21]), the standard neural network, uses $\{-1, 1\}$ as its state space, with a -1 representing the inactive state. The latter model is symmetric under a spin flip (-1 to 1, and vice versa), which is not the case in the present model as we will see. Since the neurons are numbered from 1 to N, a certain spin configuration (also called state vector) $\sigma = (\sigma_1, \ldots, \sigma_N) \in \{0, 1\}^N$ correponds to the state of the neural net in which exactly the neurons $i_1, \ldots, i_r \in \{1, \ldots, N\}$ whose spins $\sigma_{i_1}, \ldots, \sigma_{i_r}$ are 1 are active and those whose spins are 0 are inactive. We aim at storing $M = M(N)$ messages in the network. The M patterns are denoted by $\xi^\mu = (\xi_1^\mu, \ldots, \xi_N^\mu)$, $1 \le \mu \le M$. Amari in [2] proposes a Hebbian learning rule, using the synaptic efficacies

$$J_{ij} := \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu, \quad i \neq j.$$

As Amari points out in his paper, this learning rule demonstrates the difference between the state spaces $\{0, 1\}$ and $\{-1, 1\}$: in the first case, the synaptic strength between the neurons i and j changes by the learning process of a given ξ^μ only if both neurons are activated. In the second case, each pattern contributes to their synaptic efficacy. Besides that, the essential characteristic of Amari's model is that the stored patterns will be **sparse**: a message consists of a very small number of activated neurons compared to the total number of neurons. Assuming that for fixed N, the $(\xi_j^\mu, 1 \le \mu \le M(N), 1 \le j \le N)$ are identically distributed Bernoulli random variables with parameter p_N, Amari defines sparsity by the fact that the probability

$$p_N = \mathbb{P}(\xi_j^\mu = 1)$$

converges to 0 as N tends to infinity. This is, the expected ratio of excited neurons in a stored pattern converges to 0, as N tends to infinity:

$$\mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N} \xi_j^\mu\right] \longrightarrow 0.$$

Amari mentions two possibilities to choose the probability distribution of $(\xi_j^\mu)_{j \le N, \mu \le M}$: in the first, the $\xi_j^\mu, j \le N, \mu \le M$ are all independent and identically distributed. A second possibility is to keep the activity

$$a_N := \frac{\sum_{j=1}^{N} \xi_j^\mu}{N},$$

that is, the ratio of activated neurons per message, fixed and to choose uniformly among the set of patterns with exactly $N \cdot a_N$ excited neurons. Amari asserts without giving a proof that the second version works much better than the first one. We will analyse both and compare them.

The parameter p_N is chosen as

$$p_N = \frac{\log(N)}{N},$$

as the extreme case of sparsity mentioned by Amari, additionally comparable to the model of Gripon and Berrou regarding this property because they choose almost the same activity in their model. In case of exactly c active neurons per message, c is chosen as $c = \lfloor p_N \cdot N \rfloor = \lfloor \log(N) \rfloor$. We will, to simplify expressions, without loss of generality assume that $\log(N) \in \mathbb{N}$.

The neural network can serve as an associative memory only if there is some given dynamics. Amari uses synchronous updating; here the dynamics is a map $T = (T_1, \ldots, T_N) : \{0,1\}^N \to \{0,1\}^N$. The involved T_i are maps $T_i : \{0,1\}^N \to \{0,1\}$ and T_i determines the state the i-th neuron will take. Alternatively, the spins can be updated sequentially, e. g., in random order or from 1 to N, this is, $\bar{T} = \bar{T}_N \circ \ldots \circ \bar{T}_1$, with $\bar{T}_i(\sigma) = (\sigma_1, \ldots, \sigma_{i-1}, T_i(\sigma), \sigma_{i+1}, \ldots, \sigma_N)$.

For each $\sigma \in \{0,1\}^N$ and each neuron $i \in \{1, \ldots, N\}$, the local field $S_i(\sigma)$ at the state vector σ is defined by

$$S_i(\sigma) := \sum_{j \neq i} J_{ij} \sigma_j.$$

A neuron i remains or is activated, if the local field $S_i(\sigma)$ is large enough, i.e. larger than a given threshold $h > 0$. Confronted with the input $\sigma = (\sigma_1, \ldots, \sigma_N)$, the i-th component of the dynamics $T = (T_1, \ldots, T_N)$ assigns to the i-th neuron the updated value

$$T_i(\sigma) = \Theta\left(\sum_{j \neq i} J_{ij} \sigma_j - h\right),$$

where the function Θ is the Heaviside function defined by

$$\Theta(x) = \begin{cases} 1 & x \geq 0, \\ 0 & x < 0. \end{cases}$$

The threshold h is chosen as $h = \gamma \log(N)$ with some γ, called a threshold variable, to be determined. This is reasonable because we expect $\log(N)$ activated neurons in a message and the synaptic efficacies among the activated neurons are non-zero because they are at least once part of the same message. The threshold should be chosen such that stored messages are fixed points of the dynamics, i.e. the local field of the excited neurons should exceed this bound, but it should be large enough to prohibit that the local fields of non-excited neurons of the message become larger than the threshold. This argument firstly suggests a choice of γ as an element of $(0, 1)$ which is consistent with the models in Chapter 3 and 4, where it is not possible to choose higher thresholds. It will however turn out that it is also possible to use a threshold variable $\gamma \geq 1$ as long as it is smaller than the critical threshold variable γ^* that will be determined in the following section.

The maximal number of stored patterns is supposed to be of size $N^2/\log(N)^2$, an assertion made by Amari and proven nonrigorously. We therefore presume a number of $M(N) = \alpha N^2/\log(N)^2$ stored patterns and will see that this is an appropriate choice because with the dynamics described above, there will be some $\alpha^*(\gamma)$ dependent on the threshold variable γ such that the stability of a stored pattern can be guaranteed with high probability if $\alpha < \alpha^*(\gamma)$ and a stored pattern is instable with high probability if $\alpha > \alpha^*(\gamma)$. From now on, the variable α is called capacity variable. The following section deals with the capacity - concerning two possible notions of it - of the above model and its error correcting behaviour. In the subsequent section, we will define a Hamiltonian function associated with the dynamics and show that there is, with high probability, an energy valley around a stored pattern.

2.1 Stability and Error Correction

A first question to approach is: how many patterns can be stored in the network until the system is overloaded with information? The most important task a network should perform is the recognition of the stored patterns. This means that the stored patterns should be fixed points of the dynamics T. We distinguish between the requirement that *all* stored messages should be fixed points and the less restrictive condition that a randomly chosen stored pattern should be a fixed point of the dynamics T. The capacity of the model then is the maximal number of messages until this property is lost (dependent on the notion of stability one requires). We begin with the analysis of the stored pattern's stability in Amari's model with independent spins ξ_j^μ, $1 \leq \mu \leq M$, $1 \leq j \leq N$, using the second definition of capacity.

We will show that it is possible to store up to $\alpha N^2/\log(N)^2$ patterns with a constant α to be chosen in dependence on γ such that an arbitrary pattern is stable with probability converging to 1 as N tends to infinity. As we will additionally see, there is, for each $\gamma \in (0,1)$, a critical value $\alpha^*(\gamma)$ strictly separating the two sets $(0, \alpha^*(\gamma))$ and $(\alpha^*(\gamma), \infty)$, such that each element of the first set is an admissible capacity variable (supposed that $M = \alpha N^2/\log(N)^2$ patterns are stored, the probability that an arbitrary stored pattern is stable tends to 1) and each element of the second set is an inadmissible capacity variable ($M = \alpha N^2/\log(N)^2$ leads to instability of an arbitrary stored pattern, with positive probability not tending to 0). Of course, M has to be a natural number and we mean by convention that $M = \lfloor \alpha N^2/\log(N)^2 \rfloor$ if we write $M = \alpha N^2/\log(N)^2$. We mostly write M instad of $M(N)$.

The distribution of the ξ_j^μ, $\mu \leq M(N)$, $j \leq N$ clearly depends on N and so do the patterns. We firstly analyse the situation and the behaviour of the dynamics for fixed N; the dependence on N of the ξ_j^μ is, as the one of p and M, not indicated in the notation. The derived results are asymptotic. An event with probability converging to 1 as the size of the network tends to infinity is said to happen with "high probability".

Considering the network with N neurons, the σ-algebra generated by ξ_j^μ, $1 \leq j \leq k$, $1 \leq \mu \leq M(N)$ is called $\bar{\mathcal{F}}_k^N$ and the one with respect to ξ_j^μ, $j \leq k$, $1 < \mu \leq M(N)$ is called $\mathcal{F}_k^N$.

We begin now by stating the following theorem:

Theorem 2.1 *(cf. [18]) In Amari's network using N neurons and the threshold $h = \gamma \log(N)$, $\gamma \in (0,1)$, assume that the number of stored messages is equal to $M = \alpha N^2/\log(N)^2$. Then, if $\alpha < \gamma$ satisfies*

$$-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha < -1, \tag{2.1}$$

an arbitrary stored message ξ^μ is stable with probability converging to 1:

$$\lim_{N\to\infty} \mathbb{P}(\forall i \leq N : T_i(\xi^\mu) = \xi_i^\mu) = 1. \tag{2.2}$$

On the other hand, the system gets unstable if either 1.) $\alpha < \gamma$ and

$$-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha > -1 \tag{2.3}$$

or if 2.) $\alpha \geq \gamma$. In these cases

$$\lim_{N\to\infty} \mathbb{P}(\exists i \leq N : T_i(\xi^\mu) \neq \xi_i^\mu) = 1 \tag{2.4}$$

for an arbitrary but fixed μ.

The critical value α_1^*, that is, the supremum of all α such that α is an admissible capacity variable for Amari's model using a threshold variable $\gamma \in (0,1)$, is equal to the root of the function

$$g(\alpha) = \log\left(\frac{1}{\alpha}\right) + \alpha - 2,$$

$$\alpha_1^* \approx 0.1585.$$

Concretely, for each $\alpha < \alpha_1^*$ there is some $\gamma \in (0,1)$, $\gamma > \alpha$, such that (2.1) is true and therefore (2.2) holds for the dynamics T using threshold γ if $M = \alpha N^2/\log(N)^2$ patterns have been stored in the system. For each $\alpha > \alpha_1^*$, (2.4) holds for the dynamics T using an arbitrary threshold $\gamma \in (0,1)$, if $M = \alpha N^2/\log(N)^2$.

Remark 2.2 *In particular, we could not only find upper and lower bounds on the capacity variable but the bounds obtained in Theorem 2.1 match. We have thus found sharp bounds on the capacity variable, depending on the used threshold variable γ.*

Until now, the threshold variable γ is restricted to the interval $(0,1)$. The result of Theorem 2.1 including the determined α_1^* can be helpful in the analysis of the Willshaw model in Chapter 4.

Proof of Theorem 2.1: We start by proving the first statement of the theorem. To this end, we fix a randomly chosen stored message: without loss of generality, let this pattern be ξ^1. To determine the probability that ξ^1 is stable, there are two cases to be

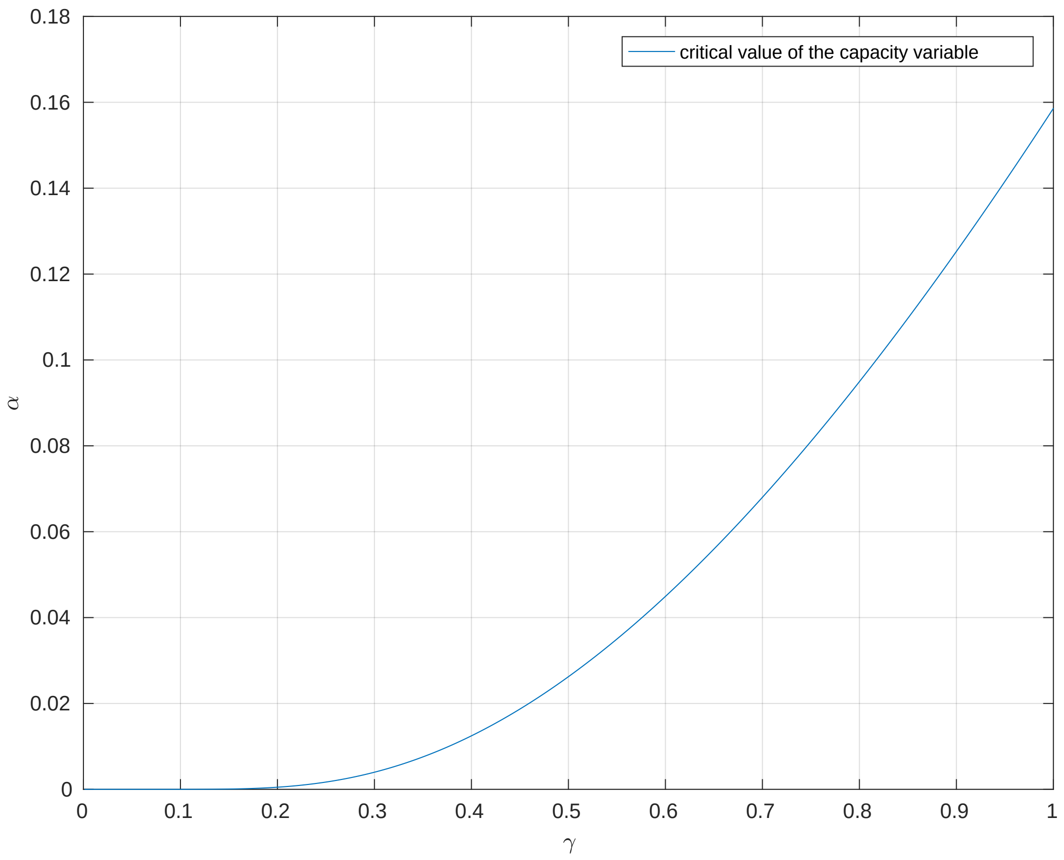

Figure 2.1: Critical capacity variable $\alpha^*(\gamma)$ in dependence on the threshold variable γ for Amari's model

distinguished: the stability of the neurons that are excited in ξ^1 and the stability of those that are not excited in ξ^1.

The local field depends considerably on the number of activated neurons in ξ^1. This is why we split the set

$$\left\{\exists i \in \{1,\dots,N\} : T_i(\xi^1) \neq \xi_i^1\right\}$$

into the intersections with the disjoint sets

$$\left\{\left|\log(N) - \sum_{j=1}^{N} \xi_j^1\right| \geq \delta \log(N)\right\} \quad \text{and} \quad \left\{\left|\log(N) - \sum_{j=1}^{N} \xi_j^1\right| < \delta \log(N)\right\},$$

for some fixed $\delta > 0$. Then we can bound the probability of an error:

$$\mathbb{P}\left[\exists i \in \{1,\dots,N\} : T_i(\xi^1) \neq \xi_i^1\right] \leq \mathbb{P}\left[\left|\log(N) - \sum_{j=1}^{N} \xi_j^1\right| \geq \delta \log(N)\right]$$
$$+ \mathbb{P}\left[\left\{\exists i \in \{1,\dots,N\} : T_i(\xi^1) \neq \xi_i^1\right\} \cap \left\{\left|\log(N) - \sum_{j=1}^{N} \xi_j^1\right| < \delta \log(N)\right\}\right].$$

Since the ξ_j^μ are independent Bernoulli random variables with success probability $p = \log N/N$, the first term disappears as N tends to infinity, due to the Chebyshev inequality.

Every activated neuron of ξ^1 will be stable, that is to say, $T_i(\xi^1) = \xi_i^1$ for each i with $\xi_i^1 = 1$, if the number of activated neurons in ξ^1 is big enough. Due to the condition $\gamma < 1$, the variable δ can be chosen such that $0 < \delta < 1 - \gamma$. The inequality

$$\left| \log(N) - \sum_{j=1}^{N} \xi_j^1 \right| < \delta \log(N) \tag{2.5}$$

implies $\sum_j^N \xi_j^1 > (1-\delta)\log(N)$, and for each i with $\xi_i^1 = 1$, we obtain

$$\begin{aligned} S_i(\xi^1) &= \sum_{j\neq i} J_{ij}\xi_j^1 = \sum_{j\neq i} \xi_j^1 \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu = \xi_i^1 \sum_{j\neq i} \xi_j^1 + \sum_{j\neq i} \xi_j^1 \sum_{\mu=2}^{M} \xi_i^\mu \xi_j^\mu \\ &\geq \sum_{j\neq i} \xi_j^1 > (1-\delta)\log(N) - 1 \geq \gamma \log(N) \end{aligned}$$

for N large enough, if (2.5) holds. The activated neurons in ξ^1 are therefore stable with probability converging to 1.

To facilitate the readibility while examining the behaviour of the inactive neurons of ξ^1, we will denote the event $\{\sum_{j=1}^N \xi_j^1/\log(N) \in (1-\delta, 1+\delta)\}$ by A_δ. In addition, for each $k \in \{1, \ldots, N\}$, the events

$$\Big\{ \sum_{j=1}^{N} \xi_j^1 = k \Big\}, \quad \Big\{ \sum_{j=1}^{N} \xi_j^1 = \sum_{j=1}^{k} \xi_j^1 = k \Big\}$$

are called $\bar{\mathcal{Z}}_k$ and $\mathcal{Z}_k$, respectively.

A neuron i is activated by the application of the dynamics to ξ^1 if $S_i(\xi^1) \geq \gamma \log(N)$. To obtain a stable ξ^1, all the inactive neurons of this pattern must remain inactive. The probability of the complement of this event can be bounded as follows:

$$\begin{aligned} &\mathbb{P}\left[\exists i \leq N : \xi_i^1 = 0, T_i(\xi^1) = 1\right] \leq \mathbb{P}\left[\{\exists i \leq N : \xi_i^1 = 0, T_i(\xi^1) = 1\} \cap A_\delta\right] + \mathbb{P}(A_\delta^c) \\ &= \sum_{k=\lceil(1-\delta)\log(N)\rceil}^{\lfloor(1+\delta)\log(N)\rfloor} \mathbb{P}\left[\{\exists i \leq N : \xi_i^1 = 0, T_i(\xi^1) = 1\} \cap \bar{\mathcal{Z}}_k\right] + \mathbb{P}(A_\delta^c) \\ &= \sum_{k=\lceil(1-\delta)\log(N)\rceil}^{\lfloor(1+\delta)\log(N)\rfloor} \mathbb{P}\left[\exists i \leq N : \xi_i^1 = 0, S_i(\xi^1) \geq \gamma\log(N) \Big| \bar{\mathcal{Z}}_k\right] \mathbb{P}(\bar{\mathcal{Z}}_k) + \mathbb{P}(A_\delta^c) \\ &\leq \sum_{k=\lceil(1-\delta)\log(N)\rceil}^{\lfloor(1+\delta)\log(N)\rfloor} (N-k)\,\mathbb{P}\left[S_N(\xi^1) \geq \gamma\log(N) \Big| \mathcal{Z}_k\right] \mathbb{P}(\bar{\mathcal{Z}}_k) + \mathbb{P}(A_\delta^c). \end{aligned}$$

The delimiters of the sum are chosen because the number of excited neurons in the mentioned message, $\sum_{j=1}^N \xi_j^1$, does only take values in $\mathbb{N}$. Without loss of generality, we assume for the rest of this part of the proof that $(1+\delta)\log(N) \in \mathbb{N}$. In the last line, the event

$\bar{\mathcal{Z}}_k$ is replaced by $\mathcal{Z}_k$ because it does not matter in probability which k of the N neurons are activated. We continue to estimate the probability:

$$\begin{aligned}
&\sum_{k=(1-\delta)\log(N)}^{(1+\delta)\log(N)} (N-k)\,\mathbb{P}\left[S_N(\xi^1) \geq \gamma\log(N)\middle|\mathcal{Z}_k\right]\mathbb{P}(\bar{\mathcal{Z}}_k) + \mathbb{P}(A_\delta^c)\\
=&\sum_{k=(1-\delta)\log(N)}^{(1+\delta)\log(N)} (N-k)\,\mathbb{P}\left[\sum_{j\leq N-1}\xi_j^1\sum_{\mu=1}^{M}\xi_N^\mu\xi_j^\mu \geq \gamma\log(N)\middle|\mathcal{Z}_k\right]\mathbb{P}(\bar{\mathcal{Z}}_k) + \mathbb{P}(A_\delta^c)\\
\leq&\max_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}}(N-k)\,\mathbb{P}\left[\sum_{j\leq k}\sum_{\mu=2}^{M}\xi_N^\mu\xi_j^\mu \geq \gamma\log(N)\right]\cdot\sum_{k=(1-\delta)\log(N)}^{(1+\delta)\log(N)}\mathbb{P}(\bar{\mathcal{Z}}_k) + \mathbb{P}(A_\delta^c)\\
\leq& N\,\mathbb{P}\left[\sum_{j\leq(1+\delta)\cdot\log(N)}\sum_{\mu=2}^{M}\xi_N^\mu\xi_j^\mu \geq \gamma\log(N)\right]\cdot\mathbb{P}(A_\delta) + \mathbb{P}(A_\delta^c).
\end{aligned}$$

The transition from the second to the third line is performed by taking into account that on $\mathcal{Z}_k$ we have $\xi_i^1 = 0$ for each $i > k$ and $\xi_i^1 = 1$ for $i \leq k$. Then the maximum of the conditional probabilities in the last line is attained for $k = (1+\delta)\log(N)$ because the sum $\sum_{j\leq k}\sum_{\mu=2}^{M}\xi_i^\mu\xi_j^\mu$ is increasing in k; the maximum is thus attained by choosing the maximal value for k.

Since $\mathbb{P}(A_\delta)$ tends to 1, we have to find a bound on the probability

$$\mathbb{P}\left[\sum_{j\leq(1+\delta)\cdot\log(N)}\sum_{\mu=2}^{M}\xi_N^\mu\xi_j^\mu \geq \gamma\log(N)\right].$$

Using the exponential Markov inequality (that will also be called exponential Chebychev inequality) for some $t > 0$, we obtain

$$\begin{aligned}
&\mathbb{P}\left[\sum_{j\leq(1+\delta)\cdot\log(N)}\sum_{\mu=2}^{M}\xi_N^\mu\xi_j^\mu \geq \gamma\log(N)\right] \leq \exp\left[-t\gamma\log(N)\right]\mathbb{E}\exp\left[t\sum_{j=1}^{(1+\delta)\log(N)}\sum_{\mu=2}^{M}\xi_N^\mu\xi_j^\mu\right]\\
&=\exp\left[-t\gamma\log(N)\right]\left(\mathbb{E}\exp\left[t\sum_{j=1}^{(1+\delta)\log(N)}\xi_N^M\xi_j^M\right]\right)^{M-1} \qquad (2.6)\\
&=\exp\left[-t\gamma\log(N)\right]\left[1-p+p\left(1-p+pe^t\right)^{(1+\delta)\log(N)}\right]^{M-1} \qquad (2.7)
\end{aligned}$$

by using independence of the patterns $\xi^\mu, 2 \leq \mu \leq M$ in (2.6) and independence of the spins $\xi_i^\mu, 1 \leq i \leq N$ of ξ^μ in (2.7). Conditionally on $\{\xi_N^\mu = 1\}$, the sum $\sum_{j=1}^{(1+\delta)\log(N)}\xi_N^\mu\xi_j^\mu$ is Binomially distributed with parameters p and $(1+\delta)\log(N)$ which results in the given exponential moment. To continue, (2.7) is at most

$$\exp\left[-t\gamma\log(N)\right]\left[1-p+p\left(1-p+pe^t\right)^{(1+\delta)\log(N)}\right]^{M-1}$$

$$\leq \exp\left[-t\gamma\log(N)\right]\left[1-p+pe^{p(e^t-1)(1+\delta)\log(N)}\right]^{M-1} \tag{2.8}$$

$$\leq \exp\left[-t\gamma\log(N)+(M-1)p\left(e^{p(e^t-1)(1+\delta)\log(N)}-1\right)\right] \tag{2.9}$$

$$=\exp\left[-t\gamma\log(N)+(M-1)p\left(p(e^t-1)(1+\delta)\log(N)+\mathcal{O}(\log(N)^2p^2)\right)\right], \tag{2.10}$$

due to the inequality $1+u\leq e^u$ for all $u\in\mathbb{R}$ in (2.8) as well as in (2.9) and the power series representation of the exponential in (2.10) where we assume t to not depend on N.

Using $M=\alpha N^2/\log(N)^2$, $p=\log(N)/N$ and in the last line once more the series representation of the exponential, we obtain

$$\begin{aligned}
&\exp\left[-t\gamma\log(N)+(M-1)p\left(p(e^t-1)(1+\delta)\log(N)+\mathcal{O}(\log(N)^2p^2)\right)\right]\\
&\leq \exp\left[-t\gamma\log(N)+Mp^2(e^t-1)(1+\delta)\log(N)+\mathcal{O}(\log(N)^2p)\right]\\
&=\exp\left[-t\gamma\log(N)+\alpha(e^t-1)(1+\delta)\log(N)+\mathcal{O}(\log(N)^3/N)\right]\\
&=\exp\left[\log(N)(-t\gamma+\alpha(e^t-1)(1+\delta))\right](1+\mathcal{O}(\log(N)^3/N)).
\end{aligned}$$

To find a good bound on the probability, the function

$$f_{\delta,\gamma,\alpha}(t)=-t\gamma+\alpha(e^t-1)(1+\delta)$$

has to be minimised in t. This yields as minimal argument

$$t^*_{\delta,\gamma,\alpha}=\log\left(\frac{\gamma}{\alpha(1+\delta)}\right).$$

By choosing $0<\delta<(\gamma-\alpha)/\alpha$, possible due to the condition $\alpha<\gamma$, $t^*_{\delta,\gamma,\alpha}$ is positive. Inserting $t^*_{\delta,\gamma,\alpha}$ into the above exponential then yields

$$\begin{aligned}
&\exp\left[\log(N)(-\gamma t^*_{\delta,\gamma,\alpha}+\alpha(e^{t^*_{\delta,\gamma,\alpha}}-1)(1+\delta))\right]\\
&=\exp\left[\log(N)\left(-\gamma\log\left(\frac{\gamma}{\alpha(1+\delta)}\right)+\gamma-\alpha(1+\delta)\right)\right].
\end{aligned}$$

The inequality

$$-\gamma\log\left(\frac{\gamma}{\alpha(1+\delta)}\right)+\gamma-\alpha(1+\delta)<-1 \tag{2.11}$$

is sufficient to let the probability converge to 0. Since δ can be chosen arbitrarily small, this condition can be obtained for each α satisfying inequality (2.1). This is all we need to prove the upper bound on α. By chosing $\delta>0$ small enough that the three conditions $\delta<1-\gamma$, $\delta<(\gamma-\alpha)/\alpha$ and (2.11) are fulfilled, we obtain $\mathbb{P}\left[\exists 1\leq i\leq N: T_i(\xi^1)\neq\xi^1_i\right]\longrightarrow 0$ for $\alpha<\gamma$ and $\gamma\in(0,1)$ such that (2.1) holds.

To show the reverse bound, we choose an arbitrary message, e. g., ξ^1, and analyse the probability of having an error at any neuron that is not activated in the message. With the above notation, the requested probability of having at least one error after the application of the dynamics is at least equal to

$$\mathbb{P}\left(\exists i\leq N: T_i(\xi^1)\neq\xi^1_i\right)\geq\mathbb{P}\left(\exists i\leq N:\xi^1_i=0, T_i(\xi^1)\neq 0\right)$$

$$
\begin{aligned}
&=\mathbb{P}\left(\{\exists i \le N : \xi_i^1 = 0, T_i(\xi^1) \neq 0\} \cap A_\delta\right) + \mathbb{P}\left(\{\exists i \le N : \xi_i^1 = 0, T_i(\xi^1) \neq 0\} \cap A_\delta^c\right) \\
&\ge\mathbb{P}\left(\{\exists i \le N : \xi_i^1 = 0, T_i(\xi^1) \neq 0\} \cap A_\delta\right) \\
&= \sum_{k=\lceil (1-\delta)\log(N)\rceil}^{\lfloor (1+\delta)\log(N)\rfloor} \mathbb{P}\left[\bar{\mathcal{Z}}_k\right] \mathbb{P}\left[\exists i \le N : \xi_i^1 = 0, T_i(\xi^1) \neq 0 \middle| \bar{\mathcal{Z}}_k\right] \\
&\ge\mathbb{P}(A_\delta) \cdot \min_{\substack{k\in\mathbb{N}:\lceil (1-\delta)\log(N)\rceil \le k \\ \le \lfloor (1+\delta)\log(N)\rfloor}} \mathbb{P}\left[\exists i \le N : \xi_i^1 = 0, T_i(\xi^1) \neq 0 \middle| \bar{\mathcal{Z}}_k\right]. \qquad (2.12)
\end{aligned}
$$

We have seen that the probability of A_δ tends to 1 and continue by determining the conditional probability of the last line for fixed k:

$$
\begin{aligned}
&\mathbb{P}\left[\exists i \le N : \xi_i^1 = 0, T_i(\xi^1) \neq 0 \middle| \bar{\mathcal{Z}}_k\right] = \mathbb{P}\left[\exists i \le N : \xi_i^1 = 0, T_i(\xi^1) \neq 0 \middle| \mathcal{Z}_k\right] \\
&=\mathbb{P}\left[\exists i \ge k+1 : T_i(\xi^1) \neq 0 | \mathcal{Z}_k\right] = 1 - \mathbb{P}\left[\forall i \ge k+1 : T_i(\xi^1) = 0 | \mathcal{Z}_k\right] \\
&=1 - \mathbb{P}\left[\forall i \ge k+1 : \sum_{\mu=2}^M \sum_{j \le k} \xi_i^\mu \xi_j^\mu < \gamma \log(N)\right]. \qquad (2.13)
\end{aligned}
$$

For a fixed realisation $(x_j^\mu)_{\mu\ge 2, j\le k} \in \{0,1\}^{k(M-1)}$ of $(\xi_j^\mu)_{\mu\ge 2, j\le k}$, the events

$$
\left\{\sum_{\mu=2}^M \sum_{j\le k} \xi_i^\mu \xi_j^\mu < \gamma \log(N)\right\}, \quad i > k
$$

are conditionally independent, given $\{(x_j^\mu)_{\mu\ge 2, j\le k} = (\xi_j^\mu)_{\mu\ge 2, j\le k}\}$. The last line in (2.13) is therefore equal to

$$
\begin{aligned}
&1 - \mathbb{P}\left[\forall i \ge k+1 : \sum_{\mu=2}^M \sum_{j\le k} \xi_i^\mu \xi_j^\mu < \gamma \log(N)\right] \\
&=1 - \mathbb{E}_{(\xi_j^\mu)_{\mu\ge 2, j\le k}}\left[\mathbb{P}\left(\forall i \ge k+1 : \sum_{\mu=2}^M \sum_{j\le k} \xi_i^\mu \xi_j^\mu < \gamma \log(N) \middle| \mathcal{F}_N^k\right)\right] \\
&=1 - \mathbb{E}_{(\xi_j^\mu)_{\mu\ge 2, j\le k}}\left[\mathbb{P}\left(\sum_{\mu=2}^M \sum_{j\le k} \xi_N^\mu \xi_j^\mu < \gamma \log(N) \middle| \mathcal{F}_N^k\right)^{N-k}\right] \\
&=1 - \mathbb{E}_{(\xi_j^\mu)_{\mu\ge 2, j\le k}}\left[\left[1 - \mathbb{P}\left(\sum_{\mu=2}^M \xi_N^\mu \sum_{j\le k} \xi_j^\mu \ge \gamma \log(N) \middle| \mathcal{F}_N^k\right)\right]^{N-k}\right]. \qquad (2.14)
\end{aligned}
$$

To simplify the computations, we note that

$$
\sum_{\mu=2}^M \xi_N^\mu \sum_{j\le k} \xi_j^\mu \ge \sum_{\mu>1:\sum_{j\le k}\xi_j^\mu = 1} \xi_N^\mu \sum_{j\le k} \xi_j^\mu
$$

and therefore

$$\mathbb{P}\left(\sum_{\mu=2}^{M}\xi_N^\mu\sum_{j\le k}\xi_j^\mu\ge\gamma\log(N)\Big|\mathcal{F}_N^k\right)\ge\mathbb{P}\left(\sum_{\mu>1:\sum_{j\le k}\xi_j^\mu=1}\xi_N^\mu\sum_{j\le k}\xi_j^\mu\ge\gamma\log(N)\Big|\mathcal{F}_N^k\right). \tag{2.15}$$

We continue with the analysis of the behaviour of the random variables $\sum_{j\le k}\xi_j^\mu$, $\mu\ge 2$, which are independent and identically Binomially distributed with parameters $p=\log(N)/N$ and k. Consequently, the parameters of the Binomially distributed random variable

$$X_1(k):=\sum_{\mu=2}^{M}\mathbb{1}_{\{\sum_{j\le k}\xi_j^\mu=1\}}$$

are

$$p_1(k)=kp(1-p)^{k-1}=kp+\mathcal{O}(k^2p^2)$$

and M, where we only (and also in the rest of the proof) consider the ks belonging to the set on which the minimum is taken in (2.12). Using the Chebyshev inequality, we obtain for each $\delta>0$

$$\lim_{N\to\infty}\mathbb{P}\left(\frac{X_1(k)}{\alpha k\frac{N}{\log(N)}}\notin(1-\delta,1+\delta)\right)=0.$$

The event

$$\left\{\frac{X_1(k)}{\alpha k\frac{N}{\log(N)}}\in(1-\delta,1+\delta)\right\}$$

is denoted by $B_\delta(k)$.

To return to (2.14), we first use (2.15) and the fact that the interior sum is 1 on the considered summands:

$$\begin{aligned}
&1-\mathbb{E}_{(\xi_j^\mu)_{\mu\ge2,j\le k}}\left[\left[1-\mathbb{P}\left(\sum_{\mu=2}^{M}\xi_N^\mu\sum_{j\le k}\xi_j^\mu\ge\gamma\log(N)\Big|\mathcal{F}_N^k\right)\right]^{N-k}\right]\\
\ge&1-\mathbb{E}_{(\xi_j^\mu)_{\mu\ge2,j\le k}}\left[\left[1-\mathbb{P}\left(\sum_{\mu>1:\sum_{j\le k}\xi_j^\mu=1}\xi_N^\mu\sum_{j\le k}\xi_j^\mu\ge\gamma\log(N)\Big|\mathcal{F}_N^k\right)\right]^{N-k}\right]\\
=&1-\mathbb{E}_{(\xi_j^\mu)_{\mu\ge2,j\le k}}\left[\left[1-\mathbb{P}\left(\sum_{\mu>1:\sum_{j\le k}\xi_j^\mu=1}\xi_N^\mu\ge\gamma\log(N)\Big|\mathcal{F}_N^k\right)\right]^{N-k}\right]
\end{aligned}$$

and, taking the expectation on the set $B_\delta(k)$, we obtain that the last line is at least

$$1-\mathbb{E}_{(\xi_j^\mu)_{\mu\ge2,j\le k}}\left[\mathbb{1}_{B_\delta(k)}\left[1-\mathbb{P}\left(\sum_{\mu>1:\sum_{j\le k}\xi_j^\mu=1}\xi_N^\mu\ge\gamma\log(N)\Big|\mathcal{F}_N^k\right)\right]^{N-k}\right]-\mathbb{P}(B_\delta(k)^c)$$

$$\geq 1-\mathbb{P}(B_\delta(k))\cdot\max_{B_\delta(k)}\left[1-\mathbb{P}\left(\sum_{\mu>1:\sum_{j\leq k}\xi_j^\mu=1}\xi_N^\mu\geq\gamma\log(N)\Big|\mathcal{F}_N^k\right)\right]^{N-k}-\mathbb{P}\left(B_\delta(k)^c\right)$$

$$\geq 1-\max_{B_\delta(k)}\left[1-\mathbb{P}\left(\sum_{\mu>1:\sum_{j\leq k}\xi_j^\mu=1}\xi_N^\mu\geq\gamma\log(N)\Big|\mathcal{F}_N^k\right)\right]^{N-k}-\mathbb{P}\left(B_\delta(k)^c\right). \qquad (2.16)$$

Hence we can limit our examinations to the set $B_\delta(k)$ because the probability of its complement vanishes as N tends to infinity.

Conditionally on $\mathcal{F}_N^k$, the remaining sum

$$X(k) := \sum_{\mu>1:\sum_{j\leq k}\xi_j^\mu=1}\xi_N^\mu$$

is Binomially distributed with parameters p and $X_1(k)$. But on the subset $B_\delta(k)$ on which we will consider $X_1(k)$, $X_1(k)$ is about $\alpha kN/\log(N)$; $X(k)$ is, given $X_1(k)$, therefore approximately Poisson distributed with parameter $X_1(k)p$.

To be more precise, we use Le Cam's Theorem:

Lemma 2.3 (Prohorov, Le Cam) *(see [40] and [27]) Let $\pi_\lambda(m)$ denote the probability weight of $m\in\mathbb{N}_0$ under a Poisson distribution with parameter λ and $\theta_{n,\tilde{p}}(m)$ its probability weight under a $Bin(n,\tilde{p})$ distribution. The total variation distance between these two variables is bounded by*

$$\sum_{m=0}^{\infty}\left|\theta_{n,\tilde{p}}(m)-\pi_\lambda(m)\right|\leq 2n\tilde{p}^2.$$

Keeping the notation of the Lemma, we obtain

$$\mathbb{P}\left(\sum_{\mu:\sum_{j\leq k}\xi_j^\mu=1}\xi_N^\mu\geq\gamma\log(N)\Big|\mathcal{F}_N^k\right)=\sum_{m=\lceil\gamma\log(N)\rceil}^{\infty}\theta_{X_1(k),\,p}\,(m)$$

$$\geq\sum_{m=\lceil\gamma\log(N)\rceil}^{\infty}\pi_{X_1(k)\,\cdot\,p}\,(m)-\sum_{m=0}^{\infty}\left|\theta_{X_1(k),\,p}\,(m)-\pi_{X_1(k)\,\cdot\,p}\,(m)\right|$$

$$\geq\sum_{m=\lceil\gamma\log(N)\rceil}^{\infty}\pi_{X_1(k)\,\cdot\,p}\,(m)-2X_1(k)\cdot p^2. \qquad (2.17)$$

We only consider the random variable $X_1(k)$ on the set $B_\delta(k)$ because its probability tends to 1. On $B_\delta(k)$, the last line of (2.17) is at least

$$\min_{B_\delta(k)}\sum_{m=\lceil\gamma\log(N)\rceil}^{\infty}\pi_{X_1(k)\,\cdot\,p}\,(m)-2X_1(k)\cdot p^2$$

$$\geq\sum_{m=\lceil\gamma\log(N)\rceil}^{\infty}\pi_{k(1-\delta)\alpha\frac{N}{\log(N)}\,\cdot\,p}\,(m)-2k(1+\delta)\alpha\frac{N}{\log(N)}\cdot p^2$$

$$= \sum_{m=\lceil \gamma \log(N) \rceil}^{\infty} \pi_{k(1-\delta)\alpha}(m) - 2k(1+\delta)\alpha \frac{\log(N)}{N}, \tag{2.18}$$

because the probability that a Poisson random variable exceeds the threshold $\gamma \log(N)$ increases with the parameter of the Poisson distribution. A Poisson random variable $Y_{n\lambda}$ with parameter $n \cdot \lambda$, $n \in \mathbb{N}$, is equal in distribution to a sum of n independent Poisson(λ) variables, which we call $Y_1, \dots, Y_n$. This justifies to apply Cramér's theorem, that is:

Lemma 2.4 (Cramér) *(see [13]) For independent and identically distributed random variables $Y_1, Y_2 \dots$ with finite moment generating function,*

$$\mathbb{E}\left[e^{tY_1}\right] < \infty, \quad t \in \mathbb{R}$$

and logarithmic moment generating function

$$\Lambda(t) = \log \mathbb{E}\left[e^{tY_1}\right],$$

define the Legendre transform

$$\Lambda^*(x) := \sup_{t \in \mathbb{R}} (tx - \Lambda(t)), \quad x \in \mathbb{R}.$$

Then, for $x > \mathbb{E}(Y_1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\sum_{i=1}^{n} Y_i \geq xn\right] = -\Lambda^*(x).$$

Taking for $n \in \mathbb{N}$ a Poisson distributed random variable $Y_{n\lambda} \sim \sum_{i=1}^{n} Y_i$, $Y_1, \dots, Y_n$ independent with $Y_i \sim \text{Poi}(\lambda)$, $i \leq n$, the moment generating function of Y_1 is equal to

$$\mathbb{E}\left[e^{tY_1}\right] = \sum_{m=0}^{\infty} \frac{e^{tm}\lambda^m}{m!} e^{-\lambda} = e^{\lambda(e^t-1)}$$

and the Legendre transform in dependence on the parameter λ, indicated as $\Lambda^*_\lambda(t)$, is calculated as

$$\Lambda^*_\lambda(t) = \sup_{t\in\mathbb{R}}(tx - \Lambda_\lambda(t)) = \sup_{t\in\mathbb{R}}(tx - \lambda(e^t - 1)) = x\log\left(\frac{x}{\lambda}\right) - \lambda\left(\frac{x}{\lambda} - 1\right) = x\log\left(\frac{x}{\lambda}\right) - x + \lambda.$$

This yields for each $x > \lambda$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[Y_{n\lambda} \geq xn\right] = -\,x\log\left(\frac{x}{\lambda}\right) + x - \lambda.$$

Taking the second term of the last line in (2.12), we remember that it suffices to show its convergence to 1 to show the second claim of the theorem. Taking into account (in)equalities (2.13), (2.14) and (2.16), we obtain

$$\lim_{N\to\infty} \min_{\substack{k\in\mathbb{N}:\lceil(1-\delta)\log(N)\rceil \\ \leq k \leq \lfloor(1+\delta)\log(N)\rfloor}} \mathbb{P}\left(\exists i \leq N : \xi^1_i = 0, T_i(\xi^1) \neq 0 \middle| \bar{\mathcal{Z}}_k\right)$$

$$\geq \lim_{N\to\infty} \min_{\substack{k\in\mathbb{N}:\lceil(1-\delta)\log(N)\rceil\\ \leq k\leq\lfloor(1+\delta)\log(N)\rfloor}} \left(1-\max_{B_\delta(k)}\left[1-\mathbb{P}\left(\sum_{\mu>1:\sum_{j\leq k}\xi_j^\mu=1} \xi_N^\mu \geq \gamma\log(N)\middle|\mathcal{F}_N^k\right)\right]^{N-k}\right), \tag{2.19}$$

because $\lim_{N\to\infty}\min_k \mathbb{P}\left[(B_\delta(k))^c\right]=0$. For these k, we deduce with the help of inequality (2.18), again denoting by Y_λ a Poisson random variable with parameter λ:

$$\begin{aligned}
&\max_{B_\delta(k)}\left[1-\mathbb{P}\left(\sum_{\mu>1:\sum_{j\leq k}\xi_j^\mu=1}^{M} \xi_N^\mu \geq \gamma\log(N)\middle|\mathcal{F}_N^k\right)\right]^{N-k}\\
&\leq\left[1-\sum_{m=\lceil\gamma\log(N)\rceil}^{\infty}\pi_{k(1-\delta)\alpha}(m)+2k(1+\delta)\alpha\frac{\log(N)}{N}\right]^{N-k}\\
&=\left[1-\mathbb{P}\big[Y_{k(1-\delta)\alpha}\geq\gamma\log(N)\big]+2k(1+\delta)\alpha\frac{\log(N)}{N}\right]^{N-k}\\
&=\exp\left[(N-k)\log\left(1-\mathbb{P}\big[Y_{k(1-\delta)\alpha}\geq\gamma\log(N)\big]+2k(1+\delta)\alpha\frac{\log(N)}{N}\right)\right].
\end{aligned} \tag{2.20}$$

By applying the series expansion of the logarithm, the logarithmic term in (2.20) is

$$\begin{aligned}
&\log\left(1-\mathbb{P}\big[Y_{k(1-\delta)\alpha}\geq\gamma\log(N)\big]+2k(1+\delta)\alpha\frac{\log(N)}{N}\right)\\
&=-\,\mathbb{P}\big[Y_{k(1-\delta)\alpha}\geq\gamma\log(N)\big]+2k(1+\delta)\alpha\frac{\log(N)}{N}\\
&\quad-\mathcal{O}\left[\left(\mathbb{P}\big(Y_{k(1-\delta)\alpha}\geq\gamma\log(N)\big)+2k(1+\delta)\alpha\frac{\log(N)}{N}\right)^2\right].
\end{aligned} \tag{2.21}$$

The limit in (2.19) is, by using (2.20) and (2.21), equal to 1 if

$$\lim_{N\to\infty}\max_{\substack{\lceil(1-\delta)\log(N)\rceil\\ \leq k\leq\\ \lfloor(1+\delta)\log(N)\rfloor}}\exp\left[(N-k)\left(-\mathbb{P}\big(Y_{k(1-\delta)\alpha}\geq\gamma\log(N)\big)+2k(1+\delta)\alpha\frac{\log(N)}{N}\right)\right]=0. \tag{2.22}$$

The minimum of the probabilities $\mathbb{P}\big(Y_{k(1-\delta)\alpha}\geq\gamma\log(N)\big)$ is attained by the minimal value of k, this is $k=\lceil(1-\delta)\log(N)\rceil$. Convergence to zero in (2.22) is consequently reached if

$$\lim_{N\to\infty}-\Big[N-(1+\delta)\log(N)\Big]\mathbb{P}\big(Y_{\lceil(1-\delta)\log(N)\rceil(1-\delta)\alpha}\geq\gamma\log(N)\big)+\log(N)^2=-\infty$$

which is fulfilled if

$$\liminf_{N\to\infty}\frac{\log\left(\mathbb{P}(Y_{\lceil(1-\delta)\log(N)\rceil(1-\delta)\alpha}\geq\gamma\log(N)\right)}{\log(N)}>-1. \tag{2.23}$$

Denoting by $\Lambda^*_{\alpha(1-\delta)}(x)$ the Legendre transform of a Poisson distributed random variable with parameter $\alpha(1-\delta)$ at argument x, $x > \alpha(1-\delta)$, Lemma 2.4 yields for each $\alpha < \gamma/(1-\delta)^2$

$$\lim_{N\to\infty} \frac{\log\left(\mathbb{P}(Y_{\lceil(1-\delta)\log(N)\rceil(1-\delta)\alpha} \geq \lceil(1-\delta)\log(N)\rceil\frac{\gamma}{1-\delta}\right)}{\lceil(1-\delta)\log(N)\rceil}$$
$$=\Lambda^*_{\alpha(1-\delta)}\left(\frac{\gamma}{1-\delta}\right) = -\frac{\gamma}{1-\delta}\log\left(\frac{\gamma}{\alpha(1-\delta)^2}\right) + \frac{\gamma}{1-\delta} - \alpha(1-\delta),$$

and this implies

$$\liminf_{N\to\infty} \frac{\log\left(\mathbb{P}(Y_{\lceil(1-\delta)\log(N)\rceil(1-\delta)\alpha} \geq \gamma\log(N)\right)}{\log(N)}$$
$$\geq (1-\delta)\liminf_{N\to\infty} \frac{\log\left(\mathbb{P}(Y_{\lceil(1-\delta)\log(N)\rceil(1-\delta)\alpha} \geq (1-\delta)\log(N)\frac{\gamma}{1-\delta}\right)}{\lceil(1-\delta)\log(N)\rceil}$$
$$\geq (1-\delta)\liminf_{N\to\infty} \frac{\log\left(\mathbb{P}(Y_{\lceil(1-\delta)\log(N)\rceil(1-\delta)\alpha} \geq \lceil(1-\delta)\log(N)\rceil\frac{\gamma}{1-\delta}\right)}{\lceil(1-\delta)\log(N)\rceil}$$
$$= (1-\delta)\Lambda^*_{\alpha(1-\delta)}\left(\frac{\gamma}{1-\delta}\right) = -\gamma\log\left(\frac{\gamma}{\alpha(1-\delta)^2}\right) + \gamma - \alpha(1-\delta)^2. \tag{2.24}$$

The value $\delta > 0$ is chosen a priori, but arbitrarily small, and it is thus for each α which fulfills (2.3), that is,

$$-\gamma\log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha > -1,$$

possible to find a suitable δ such that (2.24) is bigger than -1 which yields (2.23) and therefore (2.4).

Finally, the condition $\alpha < \gamma$ indicated in Theorem 2.1 is not only a sufficient, but also a necessary condition to achieve stability of the stored messages. To prove this, we assume that $\alpha \geq \gamma$. We will show in the next part of the proof that there is, for each $\gamma > 0$, an α' such that $0 < \alpha' < \gamma$ and such that (2.3) holds. The probability

$$\mathbb{P}(\exists i \geq k+1 : S_i(\xi^1) \geq \gamma\log(N)|\mathcal{Z}_k) = \mathbb{P}\left(\exists i \geq k+1 : \sum_{\mu=2}^{M}\sum_{j=1}^{k}\xi_i^\mu\xi_j^\mu\right)$$

is increasing in M. Since $\alpha > \alpha'$ and the above probability tends for each $k \in ((1-\delta)\log(N), (1+\delta)\log(N))$ to 1 if $M = \alpha' N^2/\log(N)^2$, it does if $M = \alpha N^2/\log(N)^2$. The condition

$$\alpha < \gamma \tag{2.25}$$

is therefore indeed necessary to ensure stability of the stored messages.

The critical value α_1^* is defined as the supremum of all $\alpha > 0$ such that α is an admissible capacity variable for Amari's model using a threshold variable $\gamma \in (0,1)$. The function $g : (0,\infty)^2 \to \mathbb{R}$,

$$g(\gamma, \alpha) := \gamma\log\left(\frac{\gamma}{\alpha}\right) - \gamma + \alpha - 1,$$

is continuous. The partial derivative of g with respect to α is equal to

$$\frac{\partial g}{\partial \alpha} = 1 - \frac{\gamma}{\alpha}$$

which is negative on $(0, \gamma)$. This implies that $g(\gamma, \cdot)$ is strictly decreasing on $(0, \gamma)$. Since $\lim_{\alpha \searrow 0} g(\gamma, \alpha) = \infty$ and $\lim_{\alpha \nearrow \gamma} g(\gamma, \alpha) = -1$, there is a unique root of $g(\gamma, \cdot)$ in $(0, \gamma)$ and the preimage of $\mathbb{R}_+$ belonging to the restricted function $g(\gamma, \cdot)$ to the interval $(0, \gamma)$ is of the form $(0, \alpha^*(\gamma))$. The variable $\alpha^*(\gamma)$ is the root of $g(\gamma, \cdot)$ in $(0, \gamma)$.

In particular, (2.2) is guaranteed for each $\alpha \in (0, \alpha^*(\gamma))$ used as a capacity variable. On the contrary, the interval $(\alpha^*(\gamma), \gamma)$ is nonempty. For each α in this interval, (2.4) holds. As we have seen in the previous part of the proof, the probability in (2.4) increases with the capacity variable α and therefore (2.4) also holds for $\alpha \geq \gamma$. This means that each $\alpha \in (\alpha^*(\gamma), \infty)$ is inadmissible.

To determine the critical value α_1^* of the capacity, we determine the partial derivative of g with respect to γ:

$$\frac{\partial g}{\partial \gamma} = \log\left(\frac{\gamma}{\alpha}\right)$$

which is positive for $\gamma > \alpha$. Let $\alpha^*(1)$ be the root of $g(1, \cdot)$ in $(0, 1)$, $\alpha^*(1) \approx 0.1585$. Then $\alpha_1^* \leq \alpha^*(1)$ because for each $\alpha > \alpha^*(1)$, $\alpha \in (0, 1)$, we have $g(1, \alpha) < g(1, \alpha^*(1)) = 0$ due to the fact that $g(1, \cdot)$ is strictly decreasing on $(0, 1)$. Since $g(\cdot, \alpha)$ is strictly increasing in γ on (α, ∞) and $\gamma \leq \alpha$ is inadmissible, there is no $\gamma \in (0, 1)$ such that α is an admissible capacity variable for γ.

To see that $\alpha_1^* = \alpha^*(1)$, we recall that $g(1, \cdot)$ is strictly decreasing on $(0, 1)$, that $g(1, \alpha^*(1)) = 0$ and that g is continuous. Hence for each $\alpha < \alpha^*(1)$ there is an $\gamma \in (\alpha, 1)$ such that $g(\gamma, \alpha) > 0$.

□

We have seen that for each $\gamma \in (0, 1)$, it is possible to store up to $M = \alpha N^2 / \log(N)^2$ patterns, if $\alpha < \alpha^*(\gamma)$, such that a randomly chosen one of the stored patterns is stable with probability converging to 1.

To reach a high storage capacity, it is advantageous to choose a big γ. So far we considered threshold variables $\gamma \in (0, 1)$; but there is a priori no reason for this choice. We can use threshold variables $\gamma \geq 1$ and reach a higher critical capacity variable $\alpha^* \approx 0.255$ for the model. This improves the maximal capacity obtained in Theorem 2.1, but will also give rise to a lower bound on α: we have to store at least a certain number of patterns to guarantee the stability of the stored messages, with high probability.

Theorem 2.5 *It is possible to use threshold variables $\gamma \geq 1$ to increase the storage capacity of the above model. For arbitrary $\gamma > 0$ there are sharp bounds on the capacity variables: with the root $\alpha^*(\gamma)$ of the function $g(\gamma, \cdot)$ in $(0, \gamma)$, each*

$$\alpha \in (\max(0, \gamma - 1), \alpha^*(\gamma))$$

is an admissible capacity variable for γ; each

$$\alpha \in (0, \max(0, \gamma - 1)) \cup (\alpha^*(\gamma), \infty)$$

is an inadmissible capacity variable (that is, (2.4) holds).

Let $\tilde{\gamma}$ *be the unique root of*

$$f(\gamma) = \frac{\gamma}{e^{2/\gamma}} - \gamma + 1,$$

$\tilde{\gamma} \approx 1.255$. *The critical threshold variable, defined by*

$$\gamma^* := \sup\{\gamma > 0 : \exists \alpha > 0 : \alpha \text{ is admissible for } \gamma\},$$

is equal to $\tilde{\gamma}$. *The critical capacity variable for the model is finally*

$$\alpha^* := \sup\{\alpha > 0 : \exists \gamma > 0 : \alpha \text{ is an admissible capacity variable for } \gamma\} = \gamma^* - 1 = \frac{\gamma^*}{e^{2/\gamma^*}}.$$

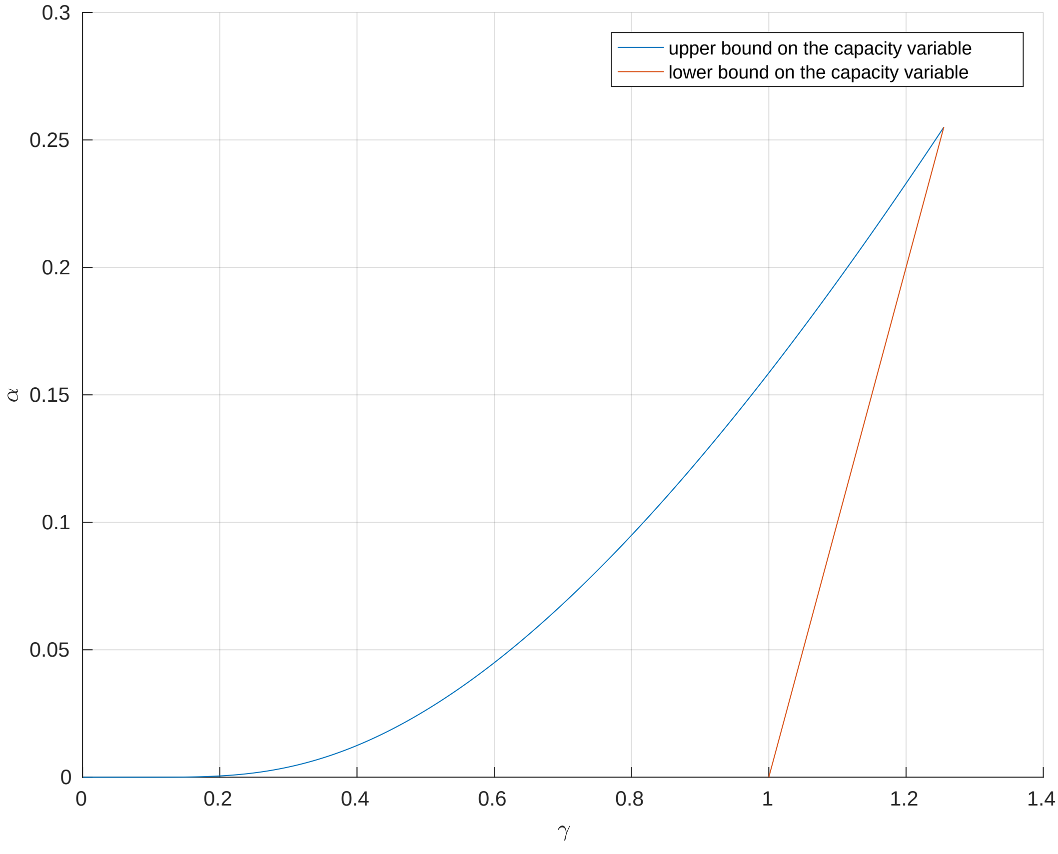

Figure 2.2: Upper and lower bounds on the capacity variable α in dependence on the threshold variable γ for Amari's model

Proof of Theorem 2.5: We aim to find upper and lower bounds on the capacity variables in dependence on γ. Given an arbitrary pattern ξ^μ, we distinguish between the stability of the active and of the inactive neurons of the message. The stability of the active neurons will result in a sharp lower bound and the stability of the inactive neurons in a sharp upper bound, respectively, on the capacity variable.

Let us begin with the stability of the inactive neurons of ξ^μ: to this end, we can use some of the results of Theorem 2.1. Firstly, we have shown for any $\alpha < \gamma$ satisfying

$$-\gamma \log(\gamma/\alpha) + \gamma - \alpha > -1,$$

that for arbitrary $\mu \in \{1, \ldots, M\}$

$$\lim_{N\to\infty} \mathbb{P}\left(\exists i \le N : T_i(\xi^\mu) \ne \xi_i^\mu\right) \ge \lim_{N\to\infty} \mathbb{P}\left(\exists i \le N : \xi_i^\mu = 0, T_i(\xi^\mu) \ne 0\right) = 1,$$

if $M = \alpha N^2/\log(N)^2$. In this part of the proof we did not use the condition $\gamma < 1$. Thus the condition (2.3) remains an upper bound on α and each $\alpha < \gamma$ fulfilling (2.3) is an inadmissible capacity variable. In addition, each $\alpha \ge \gamma$ is inadmissible because there is always an $\alpha < \gamma$ fulfilling (2.3) and the probability of the instability of the inactive neurons is increasing in M and therefore in α.

However, if threshold variables $\gamma \ge 1$ are used, the stability of the inactive neurons of pattern ξ^μ is still provided with high probability, if $\alpha < \gamma$ and

$$-\gamma \log(\gamma/\alpha) + \gamma - \alpha < -1;$$

the proof did not use the condition $\gamma < 1$, either. For this choice of α, we have for an arbitrarily chosen $\mu \in \{1, \ldots, M\}$:

$$\lim_{N\to\infty} \mathbb{P}\left(\exists i \le N : \xi_i^\mu = 0, T_i(\xi^\mu) \ne \xi_i^\mu\right) = 0.$$

Hence a sufficient constraint to guarantee stability of the inactive neurons of a message is

$$\alpha < \gamma, \quad -\gamma \log(\gamma/\alpha) + \gamma - \alpha < -1. \tag{2.26}$$

If we examine the behaviour of the active neurons of a stored pattern under the dynamics, the threshold variable $\gamma > 1$ causes a further condition on α: to keep the message's 1's, the constant α must exceed the value $\gamma - 1$. To show this, assume first $\gamma > 1$, $\alpha < \gamma - 1$. Without loss of generality, we consider ξ^1. With the notation of Theorem 2.1, we estimate

$$\begin{aligned}
&\mathbb{P}\left(\exists i \le N : \xi_i^1 = 1, T_i(\xi^1) \ne \xi_i^1\right)\\
\ge&\mathbb{P}(A_\delta) \cdot \min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left(\exists i \le N : \xi_i^1 = 1, T_i(\xi^1) \ne \xi_i^1 | \mathcal{Z}_k\right)\\
\ge&\mathbb{P}(A_\delta) \cdot \min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left[\sum_{\mu=1}^{M}\sum_{j=2}^{N} \xi_j^1 \xi_1^\mu \xi_j^\mu < \gamma \log(N) \middle| \mathcal{Z}_k\right]\\
\ge&\mathbb{P}(A_\delta) \cdot \min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left[k - 1 + \sum_{\mu=2}^{M}\sum_{j=2}^{k} \xi_1^\mu \xi_j^\mu < \gamma \log(N)\right]\\
\ge&\mathbb{P}(A_\delta) \cdot \left[1 - \max_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left[k + \sum_{\mu=2}^{M}\sum_{j=2}^{k} \xi_1^\mu \xi_j^\mu \ge \gamma \log(N)\right]\right].
\end{aligned}$$

The probability of the set A_δ tends to 1 and we are left with estimating the second probability in the last line for the considered k.

For some fixed k, the probability is bounded with the help of the exponential Chebyshev inequality:

$$\mathbb{P}\left[k+\sum_{\mu=2}^{M}\sum_{j=2}^{k}\xi_1^\mu\xi_j^\mu \geq \gamma\log(N)\right] \leq \exp\left[-t\gamma\log(N)+kt\right]\mathbb{E}\left[\exp\left(\sum_{j=2}^{k}t\xi_1^M\xi_j^M\right)\right]^{M-1},$$

due to the independence of the stored patterns; this is, for t not depending on N, as in the proof of Theorem 2.1 at most

$$\begin{aligned}&\exp\left[-t\gamma\log(N)+kt\right]\mathbb{E}\left[\exp\left(\sum_{j=2}^{k}t\xi_1^M\xi_j^M\right)\right]^{M-1}\\ \leq&\exp\left[-t\gamma\log(N)+kt\right]\exp\left[\alpha k(e^t-1)+\mathcal{O}\big(k^2\log(N)/N\big)\right].\end{aligned}$$

We just consider $k\in((1-\delta)\log(N),(1+\delta)\log(N))$ and the expression in the last line is increasing in k. An upper bound for the last line is therefore given by inserting $k=(1+\delta)\log(N)$. Minimising in t yields $t=\log[(\gamma-1-\delta)/(\alpha(1+\delta))]$ which is well defined if $\delta<\gamma-1$ and positive if $\gamma-1-\delta>\alpha(1+\delta)$. Since $\gamma>1$ and $\alpha<\gamma-1$, δ can be chosen such that $0<\delta<\gamma-1$, $0<\delta<(\gamma-1-\alpha)/(1+\alpha)$. Then the above expression converges for all k of the considered set to 0 as N tends to infinity, if

$$-(\gamma-1-\delta)\log\left(\frac{\gamma-1-\delta}{(1+\delta)\alpha}\right)+\alpha(1+\delta)\left(\frac{\gamma-1-\delta}{(1+\delta)\alpha}-1\right)<0.$$

The function $f(x)=-x\log(x)+x-1$ is negative on $\mathbb{R}_+$, so the condition is fulfilled for each choice of $\gamma,\delta,\alpha>0$ as long as $\delta<\gamma-1$. This yields (2.4) for

$$\alpha<\gamma-1. \tag{2.27}$$

If we want to choose $\gamma>1$, the usage of an α such that (2.27) holds leads to instability of the stored patterns.

On the other hand, we cannot use the proof of Theorem 2.1 to show the stability of the messages' 1's in the case $\gamma\geq 1$. If $\alpha>\gamma-1\geq 0$ and in the first step, $\gamma>1$, we have for fixed k and $t>0$:

$$\begin{aligned}&\mathbb{P}(\exists 1\leq i\leq k: T_i(\xi^1)\neq\xi_i^1|\mathcal{Z}_k)\leq k\mathbb{P}(T_1(\xi^1)\neq 1|\mathcal{Z}_k)\\ =&k\mathbb{P}\left[k-1+\sum_{\mu=2}^{M}\sum_{j=2}^{k}\xi_1^\mu\xi_j^\mu<\gamma\log(N)\right]=k\mathbb{P}\left[1-k-\sum_{\mu=2}^{M}\sum_{j=2}^{k}\xi_1^\mu\xi_j^\mu>-\gamma\log(N)\right]\\ \leq&k\exp\left[-(k-1)t+\log(N)\gamma t\right]\mathbb{E}\left[\exp\left(-t\sum_{j=2}^{k}\xi_1^M\xi_j^M\right)\right]^{M-1}\\ \leq&k\exp\left[-(k-1)t+\log(N)\gamma t+Mp^2(k-1)(e^{-t}-1)+\mathcal{O}(pk^2)\right].\end{aligned}$$

Again we consider A_δ whose probability converges to 1. The value of k can therefore be restricted to $((1-\delta)\log(N),(1+\delta)\log(N))$. This yields, for such k:

$$k\exp\Big[-(k-1)t+\log(N)\gamma t+Mp^2(k-1)(e^{-t}-1)+\mathcal{O}(pk^2)\Big]\le(1+\delta)\log(N)\cdot$$
$$\exp\Big[\log(N)\big(-(1-\delta-\gamma)t+\alpha(1-\delta)(e^{-t}-1)\big)+\mathcal{O}(p\log(N)^2)+t-\alpha(e^{-t}-1)\Big].$$

We choose $t=-\log((\gamma-1+\delta)/((1-\delta)\alpha))$. This is positive if $(\gamma-1+\delta)/(\alpha(1-\delta))<1$. This inequality can be satisfied for each $\alpha>\gamma-1$ by choosing a suitable $0<\delta<(1-\gamma+\alpha)/(1+\alpha)$. The convergence of the probability $\mathbb{P}(\exists i:\xi_i^1=1,T_i(\xi^1)\neq 1)$ to 0 follows now because the function $-x\log(x)+x-1$ is negative on $\mathbb{R}_+$.

If $\gamma=1$ and a threshold variable $\alpha>0$ is used, the pattern ξ^1 is also stable with high probability:

$$\begin{aligned}\mathbb{P}\left(\exists i:\xi_i^1=1,T_i(\xi^1)\neq 1\right)\le&\,\mathbb{P}(A_\delta)\cdot\max_{\substack{k/\log(N)\\ \in(1-\delta,1+\delta)}}k\mathbb{P}\left[k-1+\sum_{\mu=2}^{M}\sum_{j=2}^{k}\xi_1^\mu\xi_j^\mu<\log(N)\right]+\mathbb{P}(A_\delta^c)\\ \le&\,(1+\delta)\log(N)\mathbb{P}\left[\sum_{\mu=2}^{M}\sum_{j=2}^{\lceil(1-\delta)\log(N)\rceil}\xi_1^\mu\xi_j^\mu<\delta\log(N)\right]+\mathbb{P}(A_\delta^c)\end{aligned}$$

which is bounded from above by

$$(1+\delta)\log(N)\exp\Big[\log(N)\delta t+[(1-\delta)\log(N)-1]\alpha(e^{-t}-1)+\mathcal{O}(p\log(N)^2)\Big]+\mathbb{P}(A_\delta^c).$$

Choosing $\delta/(1-\delta)<\alpha$, we can use $t=-\log(\frac{\delta}{(1-\delta)\alpha})>0$, and the probability converges to 0.

It remains to compute the critical values of the parameters. First, we claim that the function f, defined by

$$f:\mathbb{R}_+\to\mathbb{R},\quad f(\gamma)=\frac{\gamma}{e^{2/\gamma}}-\gamma+1,$$

has a unique root and is strictly decreasing on $\mathbb{R}_+$. Indeed, its derivative is

$$f'(\gamma)=-1+\frac{2}{\gamma}e^{-2/\gamma}+e^{-2/\gamma}=-1+e^{-2/\gamma}\left(1+\frac{2}{\gamma}\right).$$

This is negative on $\mathbb{R}_+$ because $1+\frac{2}{\gamma}<e^{\frac{2}{\gamma}}$.

Additionally, the function f fulfills $\lim_{\gamma\to 0}f(\gamma)=1$ and $\lim_{\gamma\to\infty}f(\gamma)=-1$, so there is a unique root $\tilde\gamma$ of f in $\mathbb{R}_+$, and f is positive on $(0,\tilde\gamma)$ and negative on $(\tilde\gamma,\infty)$.

We recycle the continuous function $g(\gamma,\alpha)$ from the proof of Theorem 2.1, $g:(0,\infty)^2\to\mathbb{R}$,

$$g(\gamma,\alpha)=\gamma\log\left(\frac{\gamma}{\alpha}\right)-\gamma+\alpha-1.$$

We observed that $g(\gamma,\cdot)$ is strictly decreasing on $(0,\gamma)$ and strictly increasing on (γ,∞). The patterns are instable, if $\alpha\notin[\gamma-1,\gamma]$. But on the interval $[\gamma-1,\gamma]$, we know as result of the strict monotony of $g(\gamma,\cdot)$ on $(0,\gamma)$ and its continuity on $\mathbb{R}_+$

$$\max_{\alpha\in[\gamma-1,\gamma]}g(\gamma,\alpha)=g(\gamma,\gamma-1)=\gamma\log\left(\frac{\gamma}{\gamma-1}\right)-2.$$

Let $\gamma > \tilde{\gamma}$. We show that there is no $\alpha \in [\gamma-1, \gamma]$ such that $g(\gamma, \alpha) = \gamma \log\left(\frac{\gamma}{\alpha}\right) - \gamma + \alpha - 1 > 0$. Since the function f is strictly decreasing, we observe that $\gamma > \tilde{\gamma}$ implies $f(\gamma) < 0$ and therefore $\gamma/e^{2/\gamma} < \gamma - 1$. This yields in combination with the properties of the function g

$$\max_{\alpha \in [\gamma-1,\gamma]} \gamma \log\left(\frac{\gamma}{\alpha}\right) - \gamma + \alpha - 1 = \gamma \log\left(\frac{\gamma}{\gamma - 1}\right) - 2 < \gamma \log\left(\frac{\gamma}{\frac{\gamma}{e^{2/\gamma}}}\right) - 2 = 0.$$

So each $\alpha > 0$ is inadmissible for γ.

We have shown in the proof of Theorem 2.1 that $\gamma \in (0, 1)$ is admissible. For $\gamma \in [1, \tilde{\gamma})$, we choose $\alpha \in (\gamma-1, \gamma/e^{2/\gamma})$. This interval is non-empty: the strict monotony of f implies $f(\gamma) > 0$, i.e. $\gamma/e^{2/\gamma} > \gamma - 1$. Note that the inequality $\alpha < \gamma$ is trivially fulfilled because $\gamma/e^{2/\gamma} < \gamma$. With this choice of α we can conclude

$$g(\gamma, \alpha) = \gamma \log\left(\frac{\gamma}{\alpha}\right) - \gamma + \alpha - 1 > \gamma \log\left(\frac{\gamma}{\gamma/e^{2/\gamma}}\right) - \gamma + \gamma - 2 = \gamma \log\left(e^{2/\gamma}\right) - 2 = 0.$$

This shows that there is an admissible capacity variable for each threshold variable $\gamma \in [1, \tilde{\gamma})$. So $\gamma^* = \tilde{\gamma}$.

It remains to show that the critical value α^* is equal to $\gamma^* - 1$. This follows in analogy to the proof of Theorem 2.1 concerning α_1^*, if we keep in mind that $\gamma^* - 1 = \tilde{\gamma} - 1$ is the root of the function $g(\gamma^*, \cdot) = g(\tilde{\gamma}, \cdot)$, which follows immediately from the definition of $\tilde{\gamma}$ as root of f.

□

Remark 2.6 *The local field of a fixed neuron i, evaluated in a stored pattern ξ^μ, can be split up into a signal term, that is, the part of the local field coming from the stored message itself (for neuron i and the stored message ξ^μ, it is $\xi_i^\mu \sum_{j \neq i} \xi_j^\mu \xi_j^\mu$), and a noise term (the part of the local field coming from the remaining stored patterns).*

If threshold variables $\gamma \geq 1$ are used, the stability of one of a stored pattern's active neurons requires a certain amount of signals coming from the noise term. The noise terms of all the active neurons are high enough, with high probability, if $\alpha > \gamma - 1$. A reason for this phenomenon is that there are only about $\log(N)$ active neurons per stored pattern.

Pattern recognition is not the only task an associative memory is able to perform. If it is confronted with only partial information of a stored pattern, the retrieval dynamics should succeed in reconstructing the original pattern. If a spin is erased, the neuron begins in an inactive state. One step further, if a message is partially corrupted, i.e. there are some incorrect active and inactive spins in the input pattern, the network can under certain conditions reconstruct this stored message.

Theorem 2.7 *In Amari's model described above, assume that a threshold variable $0 < \gamma < \gamma^*$ is used and that the capacity variable α is fitted to the threshold variable γ according to the stability conditions derived in Theorem 2.1.*

1. a) If $\gamma < 1$, for each error rate $\varrho_1 \log(N)$, $\varrho_1 < 1 - \gamma$, and a fixed but arbitrary pattern ξ^μ, a fraction of $\varrho_1 \log(N)$ (which is assumed to be a natural number) of the active neurons can be deleted at random, i.e. set to the inactive state

0, and the dynamics will recover the original message in one step, with high probability. That is, if $\tilde{\xi}^\mu$ is obtained by deleting $\varrho_1 \log(N)$ excited neurons of ξ^μ at random,

$$\lim_{N\to\infty} \mathbb{P}(\forall i \leq N : T_i(\tilde{\xi}^\mu) = \xi_i^\mu) = 1.$$

b) If $0 < \gamma < \gamma^*$, the same holds true for a number of $\varrho_1 \log(N)$ deleted 1's if $\varrho_1 < 1$ and

$$\alpha > \frac{\gamma - 1 + \varrho_1}{1 - \varrho_1}.$$

This includes in particular the case $\gamma < 1$, $\varrho_1 > 1 - \gamma$. Note that the stability conditions $\alpha < \gamma$ and (2.1) must also be fulfilled.

c) On the contrary, using an arbitrary $0 < \gamma < \gamma^*$, the pattern cannot be recovered, with high probability, even not after multiple steps,

$$\lim_{N\to\infty} \mathbb{P}(\exists i \leq N : \forall n \geq 1 : T_i^{(n)}(\tilde{\xi}^\mu) \neq \xi_i^\mu) = 1$$

if

$$\alpha < \frac{\gamma - 1 + \varrho_1}{1 - \varrho_1}.$$

2. *For a threshold variable $0 < \gamma < \gamma^*$, a number of $\varrho_2 \log(N)$ spuriously activated neurons will be deactivated in one step with high probability if at most $\alpha N^2/\log(N)^2$ patterns are stored and the inequalities $\gamma/(1+\varrho_2) > \alpha$,*

$$-\gamma \log\left(\frac{\gamma}{(1+\varrho_2)\alpha}\right) + \gamma - \alpha(1+\varrho_2) < -1 \tag{2.28}$$

hold.

The bound is sharp concerning the one-step-retrieval: if $\gamma > \alpha(1+\varrho_2)$, but

$$-\gamma \log\left(\frac{\gamma}{(1+\varrho_2)\alpha}\right) + \gamma - \alpha(1+\varrho_2) > -1$$

or if $\gamma/(1+\varrho_2) \leq \alpha$, the pattern cannot be corrected in one step, with high probability:

$$\lim_{N\to\infty} \mathbb{P}(\forall i : T_i(\tilde{\xi}^\mu) = \xi_i^\mu) = 0.$$

3. *Finally, a corrupted version of a stored message, where $\varrho_1 \log(N)$ of the active neurons have been deactivated and $\varrho_2 \log(N)$ of the inactive neurons have been spuriously activated, can be returned in one step into the original message, with high probability, if $\gamma > \alpha(1 - \varrho_1 + \varrho_2)$,*

$$-\gamma \log\left(\frac{\gamma}{(1-\varrho_1+\varrho_2)\alpha}\right) + \gamma - \alpha(1 - \varrho_1 + \varrho_2) < -1$$

and additionally

$$\alpha > \frac{\gamma - 1 + \varrho_1}{1 - \varrho_1 + \varrho_2}$$

holds. The first two conditions are only supplementary to the stability conditions if $\varrho_1 < \varrho_2$, the third if $\varrho_2 < \varrho_1$. If the first two conditions are supplementary, they are sharp concerning the one step retrieval; if in contrary $\varrho_1 > \varrho_2$ and the third condition does not hold, the pattern is never corrected, with high probability.

Proof of Theorem 2.7: The proof is similar to the proof of Thm. 2.1. We begin with the first case, 1.a): let $\gamma < 1$ and $\varrho_1 < 1-\gamma$. Without loss of generality, we assume that the corrupted message is a partially deleted version of ξ^1 and that in case of k 1's in ξ^1 the patterns ξ^1 and $\tilde{\xi}^1$ are composed by

$$\xi_i^1 = 1, 1 \le i \le k; \qquad \xi_i^1 = 0, k+1 \le i \le N;$$
$$\tilde{\xi}_i^1 = 1, 1 \le i \le k - \varrho_1 \log(N); \qquad \tilde{\xi}_i^1 = 0, k - \varrho_1 \log(N) + 1 \le i \le N.$$

To recover the message ξ^1 from $\tilde{\xi}^1$, we need to reactivate the erased active neurons of ξ^1. The non-deleted 1's and the 0's must remain in their current state.

The local field of the deleted and non-deleted neurons (both activated in ξ^1) will show almost the same behaviour: for $i \in \{1, \dots, k\}$, we observe

$$S_i(\tilde{\xi}^1) = \sum_{j \neq i} J_{ij} \tilde{\xi}_j^1 = \sum_{j \neq i} \tilde{\xi}_j^1 \xi_i^1 \xi_j^1 + \sum_{\mu=2}^{M} \sum_{j \neq i} \tilde{\xi}_j^1 \xi_i^\mu \xi_j^\mu = k - \varrho_1 \log(N) - \tilde{\xi}_i^1 + \sum_{\substack{j=1 \\ j \neq i}}^{k-\varrho_1 \log(N)} \sum_{\mu=2}^{M} \xi_i^\mu \xi_j^\mu .$$

We call the part of the local field coming from the message ξ^1 signal term and the part coming from $\xi^\mu, \mu \ge 2$, noise term. The only difference between deleted and non-deleted neurons is that the signal term and the number of neurons from which signals can be received are both equal to $k - \varrho_1 \log(N)$ for the deleted neurons and to $k - \varrho_1 \log(N) - 1$ for the non-deleted neurons. For $\delta < 1 - \gamma - \varrho_1$, their local field is on A_δ at least

$$S_i(\tilde{\xi}^1) \ge k - \varrho_1 \log(N) - 1 \ge (1-\delta)\log(N) - \varrho_1 \log(N) - 1 \ge \gamma \log(N),$$

if N is large enough. Since $\mathbb{P}(A_\delta)$ tends to 1, the active neurons are recovered or remain active, with high probability. The probability that an inactive neuron of ξ^1 is activated by the dynamics, if $\tilde{\xi}^1$ is the input, is bounded by the corresponding probability when examining the stability of ξ^1, because there are less active neurons in $\tilde{\xi}^1$ than in ξ^1. Since the stability conditions are fulfilled, they remain inactive, with high probability. This concerns also 1.b) and 1.c).

We continue with 1.b). For $\varrho_1 > 1 - \gamma$ and a corrupted pattern obtained by deleting 1's, we can estimate, as in Theorem 2.5:

$$\mathbb{P}\left[\exists i \le N : \xi_i^1 = 1, T_i(\tilde{\xi}^1) = 0\right] \le \mathbb{P}(A_\delta) \max_{\substack{k \in \mathbb{N}: k/\log(N) \\ \in (1-\delta, 1+\delta)}} k \cdot \mathbb{P}\left[S_1(\tilde{\xi}^1) < \gamma \log(N) | \mathcal{Z}_k\right] + \mathbb{P}(A_\delta^c)$$
$$\le (1+\delta)\log(N) \exp\left[-\log(N) t \left(-\gamma + 1 - \delta - \varrho_1\right) + t\right] \cdot$$
$$\exp\left[(1-\delta-\varrho_1)\log(N)\alpha\left(e^{-t}-1\right) - \alpha\left(e^{-t}-1\right) + \mathcal{O}\left(\log(N)^2/N\right)\right] + \mathbb{P}(A_\delta^c).$$

We are in the same situation as in the proof of Theorem 2.5 concerning the stability of the active neurons using $\gamma > 1$. Similarly to this proof, we observe that the probability converges to 0 if

$$\alpha > \frac{\gamma - 1 + \varrho_1 + \delta}{1 - \varrho_1 - \delta}$$

by taking $t = -\log\left[(\gamma - 1 + \delta + \varrho_1)/(\alpha(1 - \delta - \varrho_1))\right]$ and using $-x\log(x) + x - 1 < 0$ for $x > 0$. Especially $t > 0$ by choosing a δ such that $\alpha > (\gamma - 1 + \delta + \varrho_1)/(1 - \delta - \varrho_1)$ which is possible due to the choice of α.

Concerning 1.c), if $\alpha < \frac{\gamma - 1 + \varrho_1}{1 - \varrho_1}$, there will be errors involving the activated neurons of ξ^1. We show that after one step of the dynamics, all activated neurons will be deactivated, with high probability. In addition, as shown in the proof concerning 1.a), none of the inactive neurons of ξ^1 will be activated by the dynamics, with high probability. If this both occurs, the message can never be recovered. The proof is very similar to the proof of Theorem 2.5 concerning the instability of the active neurons of ξ^1 if $\alpha < \gamma - 1$ and $\gamma > 1$. First,

$$\mathbb{P}(\exists i \le k : T_i(\tilde{\xi^1}) = 1 | \mathcal{Z}_k) \le k\mathbb{P}\left(k - \varrho_1 \log(N) + \sum_{\mu=2}^{M} \sum_{j=1}^{k - \varrho_1 \log(N)} \xi_k^\mu \xi_j^\mu \ge \gamma \log(N)\right).$$

The only difference of the probability on the right hand side to the probability in the corresponding part of the proof of Theorem 2.5 is the length of the interior sum and the reduced signal term coming from the message ξ^1. The proof can be repeated with this slight difference and we obtain that the term on the right hand side is bounded by

$$k \exp\left[-t(\gamma + \varrho_1)\log(N) + kt\right] \exp\left[\alpha(k - \varrho_1 \log(N))(e^t - 1) + \mathcal{O}\big(\log(N)^3/N\big)\right].$$

We only need to take into account $k \in ((1-\delta)\log(N), (1+\delta)\log(N))$; the maximum of the above expression is attained for $k = (1+\delta)\log(N)$. By choosing δ such that $\delta < \varrho_1 - 1 + \gamma$, $(\gamma + \varrho_1 - 1 - \delta)/(1 + \delta - \varrho_1) > \alpha$, the above probability tends to 0 by using

$$t = \log\left(\frac{\gamma + \varrho_1 - 1 - \delta}{(1 + \delta - \varrho_1)\alpha}\right) > 0.$$

This yields

$$\mathbb{P}(\exists i \le N : \forall n \ge 1 : T_i^{(n)}(\tilde{\xi^1}) \ne \xi_i^1) \longrightarrow 1$$

because the probability that $\tilde{\xi^1}$ is mapped to $(0, \ldots, 0)$ in the first step tends to 1. This completes the proofs of 1.a), 1.b) and 1.c).

Let us continue with 2.: we again assume that the active neurons in ξ^1 are the first k ones and the spuriously activated are the subsequent $\varrho_2 \log(N)$ ones. The local field of the neurons i, $1 \le i \le k$, is at least

$$S_i(\tilde{\xi^1}) = \sum_{j \ne i} \tilde{\xi}_j^1 \xi_i^1 \xi_j^1 + \sum_{\mu=2}^{M} \sum_{j \ne i, j=1}^{k + \varrho_2 \log(N)} \tilde{\xi}_j^1 \xi_i^\mu \xi_j^\mu \ge \sum_{j=1, j \ne i}^{k} 1 + \sum_{\mu=2}^{M} \sum_{j=1, j \ne i}^{k} \xi_i^\mu \xi_j^\mu = S_i(\xi^1).$$

Since α is chosen such that ξ^1 is stable with high probability, the active neurons of ξ^1 will also remain activated if the input pattern is $\tilde{\xi^1}$, with high probability.

The neurons $i > k$ get a potentially higher signal as result of the falsely activated neurons and the probability of an error consequently increases. The bound on the probability is derived in analogy to the one in the proof of Theorem 2.1 concerning the stability of the inactive neurons of a stored pattern. Let δ be small enough such that the inequality

$$\alpha < \frac{\gamma}{1+\delta+\varrho_2}$$

holds. The probability of A_δ tends to 1 and the probability of mapping a fixed inactive neuron of ξ^1 to 1 is increasing with the number of active neurons in the message, k. This yields

$$\mathbb{P}\left[\exists i : \xi_i^1 = 0, T_i(\tilde{\xi}^1) \neq \xi_i^1\right] \leq N\mathbb{P}\left[\sum_{\mu=2}^{M} \sum_{j=1}^{\lfloor(1+\delta)\log(N)\rfloor+\varrho_2\log(N)} \xi_N^\mu \xi_j^\mu \geq \gamma \log(N)\right] + \mathbb{P}(A_\delta^c)$$
$$\leq N \exp\left[\log(N)(-t\gamma + \alpha(e^t-1)(1+\delta+\varrho_2)) + \mathcal{O}(\log(N)^2/N)\right] + \mathbb{P}(A_\delta^c).$$

With $t = \log\left(\frac{\gamma}{\alpha(1+\delta+\varrho_2)}\right) > 0$, the probability tends to zero if

$$-\gamma \log\left(\frac{\gamma}{\alpha(1+\delta+\varrho_2)}\right) + \gamma - \alpha(1+\delta+\varrho_2) < -1.$$

This inequality can be fulfilled for each α such that the condition (2.28) holds.

The sharpness of the bound can be proven in exactly the same way as we have shown the sharpness of the stability bound: the only difference is the number of neurons from which signals can potentially come (we saw already in the previous part that the proof works exactly as the proof of Theorem 2.1). Instead of k, it is $k+\varrho_2\log(N)$. We condition on the random variables $\xi_j^\mu, \mu \geq 2, j \leq k+\varrho_2\log(N)$ and show that there exists, with probability tending to 1, a neuron i, $i > k+\varrho_2\log(N)$, that receives enough signals to be activated.

For the last part, 3., assume that message $\hat{\xi}^1$ is a corrupted version of ξ^1 with $\varrho_1\log(N)$ deleted 1's and $\varrho_2\log(N)$ spuriously activated neurons.

To correct the faulty positions of $\hat{\xi}^1$, it is necessary to recover the deleted 1's, to deactivate the spuriously activated neurons and to keep the values of the neurons whose activity has not been changed from ξ^1 to $\hat{\xi}^1$. Since the difference between the local fields of active neurons of ξ^1, deleted or not, is negligible, as well as the difference between the local field of the spuriously active neurons and the correctly inactive neurons in $\hat{\xi}^1$, we only distinguish between active and inactive neurons of ξ^1. The deleting causes a lower height of the signal term of the 1's of ξ^1; apart from that, we only have to observe that there are $k-\varrho_1\log(N)+\varrho_2\log(N)$ active neurons in the pattern. The proof of 1.c) can almost literally be repeated to show

$$\mathbb{P}\left(\exists i : \xi_i^1 = 1, T_i(\hat{\xi}^1) = 0\right) \longrightarrow 1$$

if $\alpha < \frac{\gamma-1+\varrho_1}{1-\varrho_1+\varrho_2}$ and the one of 1.b) to show

$$\mathbb{P}\left(\exists i \leq N : \xi_i^1 = 1, T_i(\hat{\xi}^1) = 0\right) \longrightarrow 0$$

if $\alpha > \frac{\gamma-1+\varrho_1}{1-\varrho_1+\varrho_2}$. If $\varrho_1 \geq \varrho_2$, the inactive neurons of ξ^1 remain inactive or are deactivated by the first step of the dynamics, with high probability, because α is chosen subject to the stability conditions of Theorem 2.1. If $\varrho_1 \geq \varrho_2$ and $\alpha < \frac{\gamma-1+\varrho_1}{1-\varrho_1+\varrho_2}$, the pattern is mapped to $(0,\dots,0)$ in one step, and it is stable, if $\alpha > \frac{\gamma-1+\varrho_1}{1-\varrho_1+\varrho_2}$, both with high probability.

If $\varrho_2 > \varrho_1$, the active neurons are with high probability stable, because $\alpha > \gamma - 1$. There are $k - \varrho_1 + \varrho_2 \log(N)$ active neurons in the pattern instead of k and the inactive neurons are or remain deactivated in the first step of the dynamics, with high probability, if $\gamma > \alpha(1-\varrho_1+\varrho_2)$ and

$$-\gamma \log\left(\frac{\gamma}{(1-\varrho_1+\varrho_2)\alpha}\right) + \gamma - \alpha(1-\varrho_1+\varrho_2) < -1.$$

This follows as in the proof of 2.). The bound is again sharp concerning the one step correction; this is proven analogously to the corresponding part of the proof of Theorem 2.1.

□

All the results obtained until now are results concerning the first notion of capacity. If we want instead all the messages to be stable with high probability, the model at hand looses the order of its capacity:

Proposition 2.8 *The stability of a stored pattern's 1's raises problems if one wants all patterns to be stable.*

1. For each $0 < \kappa < 1$, we have

$$\liminf_{N\to\infty} \mathbb{P}\left(\exists \mu \in \{1,\dots,M(N)\} : \sum_{j=1}^{N} \xi_j^\mu < \kappa \log(N)\right) > 0$$

if $N^\beta = \mathcal{O}(M)$, $\beta = \kappa\log(\kappa) - \kappa + 1$, and

$$\lim_{N\to\infty} \mathbb{P}\left(\exists \mu \in \{1,\dots,M(N)\} : \sum_{j=1}^{N} \xi_j^\mu < \kappa \log(N)\right) = 1$$

for $N^\beta = o(M)$, $\beta = \kappa\log(\kappa) - \kappa + 1$.

2. In addition, for arbitrary γ and α such that they fulfill the stability conditions of Theorem 2.5, $M = \alpha N^2/\log(N)^2$ stored patterns and used threshold $\gamma\log(N)$, the probability of having at least one instable message tends to 1. In particular we have for any choice of $\gamma, \alpha > 0$, threshold $\gamma\log(N)$ and $M(N) = \alpha N^2/\log(N)^2$ stored patterns

$$\lim_{N\to\infty} \mathbb{P}\left(\exists \mu \leq M(N) : T(\xi^\mu) \neq \xi^\mu\right) = 1.$$

3. Moreover, if a threshold $\gamma\log(N)$ with threshold variable $\gamma \geq 1$ is used, any choice of $M(N)$ such that $N^\beta = o(M(N))$ for arbitrary but fixed $\beta > 0$ leads to

$$\limsup_{N\to\infty} \mathbb{P}\left(\exists \mu \leq M(N) : T(\xi^\mu) \neq \xi^\mu\right) = 1.$$

A threshold variable $\gamma \geq 1$ is therefore not admissible for the second notion of capacity.

4. For a threshold variable $\gamma < 1$ and $M(N)$ stored patterns such that $N^{\beta'} = o(M(N))$ for some $\beta' > \gamma \log(\gamma) - \gamma + 1$, we have

$$\limsup_{N\to\infty} \mathbb{P}\left(\exists \mu \leq M(N) : T(\xi^\mu) \neq \xi^\mu\right) = 1.$$

However, if we use $\gamma < 1$ as threshold variable, we can reach that all patterns are stable, with high probability, if $M = o(N^\beta)$, $\beta = \gamma \log(\gamma) - \gamma + 1$.

Before proving this statement, we recall

Lemma 2.9 *(see [5], Chapter 1, (1.2).) Let X be a Binomially distributed random variable with parameters n and $\tilde{p}$ and Y a Poisson distributed random variable with parameter $n\tilde{p}$. Let A be a subset of $\mathbb{N}_0$. There is a constant C not depending on A, n and $\tilde{p}$, such that*

$$\begin{aligned}|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \leq & C \sum_{m\in A} \frac{(n\tilde{p})^m}{m!} e^{-n\tilde{p}} (n\tilde{p}^2 + m^2 n^{-1}) \\ \leq & C \mathbb{P}(Y \in A) \max_{m\in A} (n\tilde{p}^2 + m^2 n^{-1}).\end{aligned}$$

Proof of Proposition 2.8: *1.* Let $0 < \kappa < 1$. The probability of having at least one stored pattern that has less than $\kappa \log(N)$ active neurons is equal to

$$\begin{aligned}&\mathbb{P}\left(\exists \mu \leq M : \sum_{j=1}^N \xi_j^\mu < \kappa \log(N)\right) = 1 - \mathbb{P}\left(\forall \mu \leq M : \sum_{j=1}^N \xi_j^\mu \geq \kappa \log(N)\right) \\ =& 1 - \mathbb{P}\left(\sum_{j=1}^N \xi_j^1 \geq \kappa \log(N)\right)^M = 1 - \left[1 - \mathbb{P}\left(\sum_{j=1}^N \xi_j^1 < \kappa \log(N)\right)\right]^M.\end{aligned}$$

The number of active neurons in a message is Binomially distributed with parameters N and $\log(N)/N$. Denoting by $Y_{\log(N)}$ a Poisson distributed random variable with parameter $\log(N)$, Lemma 2.9 implies

$$\begin{aligned}&1 - \left[1 - \mathbb{P}\left(\sum_{j=1}^N \xi_j^1 < \kappa \log(N)\right)\right]^M \\ \geq& 1 - \left[1 - \mathbb{P}\left[Y_{\log(N)} < \kappa \log(N)\right]\left[1 - C\big(\log(N)^2/N + \kappa^2 \log(N)^2/N\big)\right]\right]^M.\end{aligned}$$

The application of Lemma 2.4 yields, as in the proof of Theorem 2.1, for $0 < \kappa < 1$,

$$\lim_{N\to\infty} \frac{1}{\log(N)} \log \mathbb{P}\left[Y_{\log(N)} < \kappa \log(N)\right] = -\kappa \log(\kappa) + \kappa - 1.$$

So we obtain

$$\liminf_{N\to\infty} \mathbb{P}\left(\exists \mu \in \{1, \ldots, M(N)\} : \sum_{j=1}^N \xi_j^\mu < \kappa \log(N)\right) > 0$$

if $N^\beta = \mathcal{O}(M)$, $\beta = \kappa \log(\kappa) - \kappa + 1$, and

$$\lim_{N\to\infty} \mathbb{P}\left(\exists \mu \in \{1,\dots,M(N)\} : \sum_{j=1}^{N} \xi_j^\mu < \kappa \log(N)\right) = 1$$

for $N^\beta = o(M)$, $\beta = \kappa \log(\kappa) - \kappa + 1$.

2. For $0 < \kappa < 1$, the expression $-\kappa \log(\kappa) + \kappa - 1$ is restricted to the interval $(-1, 0)$. Hence we know for $M = \alpha N^2 / \log(N)^2$

$$\lim_{N\to\infty} \mathbb{P}\left(\exists \mu \in \{1,\dots,M(N)\} : \sum_{j=1}^{N} \xi_j^\mu < \kappa \log(N)\right) = 1$$

for each $\kappa \in (0,1)$. For arbitrary α and γ such that the stability conditions of Theorem 2.1 hold, especially $\alpha < \gamma$, we choose $\kappa \in (0,1)$ such that $0 < \kappa < \gamma - \alpha$. Since the previous probability of having at least one pattern with less than $\kappa \log(N)$ active neurons tends to 1, we assume without loss of generality that ξ^1 has $k < \kappa \log(N)$ active neurons in exactly the first k places. As in the proof of Theorem 2.7, where we showed that a pattern is mapped to $(0,\dots,0)$ in one step if α is too small, we obtain for arbitrary $i \le k$, using that $0 < \alpha < \gamma - \kappa$:

$$\begin{aligned}\mathbb{P}(S_i(\xi^1) \ge \gamma \log(N)) =& \mathbb{P}\left(\sum_{\mu=2}^{M}\sum_{j\neq i}^{k} \xi_i^\mu \xi_j^\mu > \gamma \log(N) - k\right)\\ \le& \mathbb{P}\left(\sum_{\mu=2}^{M}\sum_{j\neq i}^{\log(N)} \xi_i^\mu \xi_j^\mu > (\gamma - \kappa) \log(N)\right) \longrightarrow 0.\end{aligned}$$

An arbitrary active neuron of the stored message is deactivated after the first step, with high probability and there is thus at least one instable stored pattern. If the stability conditions are not fulfilled by α and γ, one of the conditions *a)* $\gamma > \gamma^*$, *b)* $\gamma \le \gamma^*$ and $\alpha \in (0, \gamma - 1) \cup (\alpha^*(\gamma), \infty)$, *c)* $1 < \gamma \le \gamma^*$ and $\alpha = \gamma - 1$ or *d)* $\gamma \le \gamma^*$ and $\alpha = \alpha^*(\gamma)$ must be fulfilled. Conditions *a)* and *b)* imply that an arbitrary stored pattern is instable with high probability. In the two latter cases, there is again $0 < \kappa < \gamma - \alpha$ (recall that $\alpha^*(\gamma) < \gamma$) and we can continue the proof as in the case where the stability conditions hold.

3. Now suppose that a threshold variable $\gamma \ge 1$ is used. If $\gamma > \gamma^*$, an arbitrary stored pattern is instable with high probability. Let $1 \le \gamma \le \gamma^*$ and fix an arbitrary $\beta > 0$ and a sequence $M(N)_{N\in\mathbb{N}}$, $N^\beta = o(M(N))$. If there is an $\alpha > 0$ such that there is a subsequence $(N_l)_{l\in\mathbb{N}}$ with $M(N_l) \ge \alpha N_l^2 / \log(N_l)^2$, for all $l \in \mathbb{N}$, we know by the second part of the proof that the probability of an error is tending to 1 along this subsequence and $\limsup_{N\to\infty} \mathbb{P}(\exists \mu \le M(N) : T(\xi^\mu) \neq \xi^\mu) = 1$. If in contrary $N^\beta = o(M(N))$ and $M(N) = o(N^2/\log(N)^2)$, we first observe with the help of the first part of the proof that we have for $\kappa \in (0,1)$, $\beta = \kappa \log(\kappa) - \kappa + 1$

$$\lim_{N\to\infty} \mathbb{P}\left(\exists \mu \le M(N) : \sum_{j=1}^{N} \xi_j^\mu \le \kappa \log(N)\right) = 1.$$

Since $\gamma \geq 1$, we have $\gamma - \kappa > 0$ and due to the fact that $M(N) = o(N^2/\log(N)^2)$, another application of the exponential Chebyshev inequality shows that an arbitrary one of the active neurons is deactivated after the first step of the dynamics because it cannot allocate enough signals to make the local field exceed the threshold, with high probability.

4. Let now $0 < \gamma < 1$ and $\beta' > \gamma\log(\gamma) - \gamma + 1$. Let $M(N)$ be a sequence with $N^{\beta'} = o(M(N))$. We can as in the case $\gamma \geq 1$ assume that $M(N) = o(N^2/\log(N)^2)$; otherwise we see immediately that $\limsup_{N\to\infty} \mathbb{P}\left(\exists \mu \leq M(N) : T(\xi^\mu) \neq \xi^\mu\right) = 1$. Since $\beta' > \gamma\log(\gamma) - \gamma + 1$ there is an $\varepsilon > 0$ such that $\beta' \geq (\gamma - \varepsilon)\log(\gamma - \varepsilon) - \gamma + \varepsilon + 1$. Using the first result of the Proposition yields $\lim_{N\to\infty} \mathbb{P}(\exists \mu \leq M(N) : \sum_{j=1}^N \xi_j^\mu \leq (\gamma - \varepsilon)\log(N)) = 1$. The local field of an arbitrary active neuron of a stored pattern with at most $(\gamma - \varepsilon)\log(N)$ excited neurons does with high probability not exceed the threshold $\gamma\log(N)$ because $M(N) = o(N^2/\log(N)^2)$.

Finally if $M = o(N^\beta)$, $\beta = \gamma\log(\gamma) - \gamma + 1$, $\gamma \in (0,1)$, patterns are stored and the threshold is equal to $\gamma\log(N)$, the probability that there is at least one pattern that is not stable is at most

$$\begin{aligned}\mathbb{P}\left(\exists \mu \leq M, \exists i \leq N : T_i(\xi^\mu) \neq \xi_i^\mu\right) &\leq M\mathbb{P}\left[\sum_{j=1}^N \xi_j^1 \leq \gamma\log(N)\right] + M\mathbb{P}\left[\sum_{j=1}^N \xi_j^1 \geq 3\log(N)\right] \\ &+ M\mathbb{P}\left(\exists i \leq N : \xi_i^1 = 0, T_i(\xi^1) \neq 0 \middle| \sum_{j=1}^N \xi_j^1 < 3\log(N)\right),\end{aligned} \tag{2.29}$$

because the active neurons are trivially stable, if $\sum_{j=1}^N \xi_j^1 \geq \gamma\log(N)$. Again, with the help of the exponential Chebyshev inequality, we observe for $t > 0$

$$\begin{aligned}&M\mathbb{P}\left(\sum_{j=1}^N \xi_j^1 \leq \gamma\log(N)\right) \leq M\exp\left[\gamma\log(N)t\right]\left(1 - p + pe^{-t}\right)^N \\ \leq &M\exp\left[\gamma\log(N)t + Np(e^{-t} - 1)\right] \leq M\exp\left[\log(N)(-\gamma\log(\gamma) + \gamma - 1)\right]\end{aligned}$$

and

$$M\mathbb{P}\left(\sum_{j=1}^N \xi_j^1 \geq 3\log(N)\right) \leq M\exp\left[\log(N)(-3\log(3) + 3 - 1)\right].$$

For $\gamma \in (0,1)$, $-\beta = -\gamma\log(\gamma) + \gamma - 1$ only takes values in $(-1,0)$. Due to the choice of M and the fact that $-3\log(3) + 2 < -1$, the first two summands in (2.29) vanish. Finally,

$$\begin{aligned}&M\mathbb{P}\left(\exists i \leq N : \xi_i^1 = 0, T_i(\xi^1) \neq 0 \middle| \sum_{j=1}^N \xi_j^1 < 3\log(N)\right) \\ \leq &MN\mathbb{P}\left(\sum_{\mu=2}^M \sum_{j\leq 3\log(N)} \xi_N^\mu \xi_j^\mu \geq \gamma\log(N)\right) \\ \leq &MN\exp\left[-\gamma\log(N)t + Mp^2(e^t - 1)3\log(N) + \mathcal{O}(M\log(N)^5/N^3)\right]\end{aligned}$$

for arbitrary $t > 0$ not depending on N. This converges to 0 for the given choice of γ and M, using an arbitrary $t \geq (\beta + 1)/\gamma$. □

In his paper [2], Amari uses a version of the model in which the number of activated neurons is kept constant, i.e., each message has the same number of active neurons. This version of the model behaves like the model we analysed until now, but allows in addition to ensure the stability of every stored pattern without loosing the order of the capacity.

Proposition 2.10 *In the second version of Amari's model, where a stored message has exactly $c = \log(N)$ excited neurons instead of independent and identically distributed spins ξ_j^μ, $1 \leq j \leq N$, let the threshold variable be $\gamma \leq 1$ and the threshold γc. Suppose that $M = \alpha N^2/c^2$ patterns have been stored. Every pattern is stable with high probability, if $\alpha < \gamma$ and*

$$-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha < -3.$$

The results obtained for independent and identically distributed spins per message concerning the stability of one arbitrary stored pattern ξ^μ are also valid for this version of the model: using $\gamma < \gamma^*$, we have

$$\lim_{N\to\infty} \mathbb{P}(\forall i \leq N : T_i(\xi^\mu) = \xi_i^\mu) = 1$$

if $\max(0, \gamma - 1) < \alpha < \gamma$ *and*

$$-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha < -1,$$

whereas

$$\lim_{N\to\infty} \mathbb{P}(\exists i \leq N : T_i(\xi^\mu) \neq \xi_i^\mu) = 1$$

if either 1. $\alpha < \gamma$ and

$$-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha > -1,$$

2. $\alpha \geq \gamma$ or 3. $\gamma > 1$, $\alpha < \gamma - 1$.

Concerning this notion of capacity, the critical variables remain γ^* and α^* as obtained in Theorem 2.5.

The results obtained in Theorem 2.7 remain true for this version of the model.

Proof of Proposition 2.10: We do not give a detailed proof of the results already obtained for the other version of the model because most of it can be carried over from the corresponding proofs of the first version of the model. However, we focus on the differences between the models. Concerning the less restrictive definition of the storage capacity, the exponential Chebyshev inequality is used to show the lower bound concerning the stability of the 0's and the upper and lower bounds concerning the stability of the 1's. One can omit the conditioning on A_δ; this makes the proof slightly easier. Assume that the c activated neurons in ξ^1 are the first c neurons $1, \ldots, c$. For $i > c$, the following exponential moment is

$$\mathbb{E}\left[\exp\left(\sum_{\mu=2}^{M}\sum_{j=1}^{c} \xi_i^\mu \xi_j^\mu\right)\right] = \mathbb{E}\left[\exp\left(\sum_{j=1}^{c} \xi_i^M \xi_j^M\right)\right]^{M-1}$$

$$
\begin{aligned}
&= \left[1 - \frac{c}{N} + \frac{c}{N}\left(\sum_{n=0}^{c-1}\binom{c}{n}\prod_{m=1}^{n}\frac{c-m}{N-m}\prod_{k=0}^{c-n-1}\left(1-\frac{c-1-n}{N-1-n-k}\right)e^{tn}\right)\right]^{M-1} \\
&\leq \left[1 + \frac{c^2(c-1)}{N(N-1)}(e^t-1) + \mathcal{O}\left(\frac{c^5}{N^3}\right)\right]^M = \left[1 + \frac{c^2(c-1)}{N^2}(e^t-1) + \mathcal{O}\left(\frac{c^5}{N^3}\right)\right]^M \\
&\leq \exp\left[\alpha(e^t-1)(c-1) + \mathcal{O}(c^3/N)\right] = \exp\left[\alpha(e^t-1)(c-1)\right](1+\mathcal{O}(c^3/N)).
\end{aligned}
$$

We used in the second line that there are exactly c active neurons per message and in the third line that the summand for $n = 0$ is at most

$$
\prod_{k=0}^{c-1}\left(1 - \frac{c-1}{N-1-k}\right) = 1 - \sum_{k=0}^{c-1}\frac{c-1}{N-1-k} + \mathcal{O}\left(\frac{c^4}{N^2}\right) \leq 1 - c\frac{c-1}{N-1} + \mathcal{O}\left(\frac{c^4}{N^2}\right).
$$

Now all parts of the proof involving this exponential moment can repeated analogously to the proofs concerning the other version of the model. In particular, for $\gamma \leq 1$, the active neurons are automatically stable and

$$
\begin{aligned}
&\mathbb{P}\left(\exists \mu \leq M : \exists i \leq N : T_i(\xi^\mu) \neq \xi_i^\mu\right) \leq \mathbb{P}\left(\exists \mu \leq M : \exists i \leq N : \xi_i^\mu = 0, T_i(\xi^\mu) \neq 0\right) \\
\leq & MN \exp\left[-\gamma c t + \alpha(e^t-1)(c-1)\right](1+\mathcal{O}(c^3/N)) \\
\leq & \alpha N^3 \exp\left[-\gamma c t + \alpha(e^t-1)(c-1)\right](1+o(1)).
\end{aligned}
$$

Using $t = \log(\gamma/\alpha)$ for $\alpha < \gamma$, this shows the stability for each ξ^μ with high probability, if

$$
-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha < -3.
$$

For the upper bound, assume again that the c active neurons in ξ^1 are the first c ones. To show that (2.4) holds, if (2.3) is fulfilled, we condition on ξ_j^μ, $\mu \geq 2, j \leq c$. But in contrary to the version with independent and identically distributed spins, the sums $\sum_{j\leq c}\sum_{\mu=2}^M \xi_i^\mu \xi_j^\mu$, $i > c$, are no longer conditionally independent, given ξ_j^μ, $\mu \geq 2, j \leq c$.

The random variable X_1,

$$
X_1 = \sum_{\mu=2}^{M} \mathbb{1}_{\{\sum_{j\leq c}\xi_j^\mu = 1\}},
$$

used also in the corresponding part of the proof concerning the first version of the model, is again Binomially distributed, but now with parameters $M-1$ and

$$
\tilde{p}_1 = c\frac{c}{N}\prod_{k=0}^{c-2}\left(1 - \frac{c-1}{N-k-1}\right) = \frac{c^2}{N}(1+\mathcal{O}(c^2/N))
$$

The probability of the event

$$
B_\delta = \left\{\frac{X_1}{\alpha N} \in (1-\delta, 1+\delta)\right\}
$$

tends, again, to 1. We observe that

$$
\mathbb{P}\left(\exists i > c : \sum_{j\leq c}\sum_{\mu=2}^{M}\xi_i^\mu\xi_j^\mu \geq \gamma c\right) \geq \mathbb{P}\left(\exists i > c : \sum_{j\leq c}\sum_{\mu:\ \sum_{j\leq c}\xi_j^\mu = 1}\xi_i^\mu\xi_j^\mu \geq \gamma c\right)
$$

$$=\mathbb{P}\left(\exists i > c : \sum_{\mu:\, \sum_{j\le c}\xi_j^\mu = 1} \xi_i^\mu \ge \gamma c\right)$$

$$\ge \mathbb{P}(B_\delta)\cdot \min_{l\in\mathbb{N}:\, l/(\alpha N)\in(1-\delta,1+\delta)} \mathbb{P}\left(\exists i > c : \sum_{\mu:\, \sum_{j\le c}\xi_j^\mu = 1} \xi_i^\mu \ge \gamma c \,\middle|\, X_1 = l\right)$$

$$=\mathbb{P}(B_\delta)\cdot \min_{l\in\mathbb{N}:\, l/(\alpha N)\in(1-\delta,1+\delta)} \mathbb{P}\left(\exists i > c : \sum_{\mu=2}^{l+1} \xi_i^\mu \ge \gamma c \,\middle|\, \forall \mu\in\{2,\dots,l+1\} : \sum_{j\le c}\xi_j^\mu = 1\right)$$

$$=\mathbb{P}(B_\delta)\cdot \mathbb{P}\left(\exists i > c : \sum_{\mu=2}^{\lceil\alpha N(1-\delta)\rceil+1} \xi_i^\mu \ge \gamma c \,\middle|\, \forall \mu\in\{2,\dots,\lceil\alpha N(1-\delta)\rceil+1\} : \sum_{j\le c}\xi_j^\mu = 1\right).$$

We will use negative association of random variables to show that the upper bounds on α of this version of the model coincide with those of the first version. We therefore recall the definition and some properties of negatively associated random variables:

Definition 2.11 *(see [26]) A set of random variables $X_1, \dots, X_k$ is negatively associated if for every pair of disjoint subsets A_1, A_2 of $\{1, \dots, k\}$ and every pair of (coordinatewise) increasing functions f_1 and f_2*

$$\mathrm{Cov}(f_1(X_i : i\in A_1), f_2(X_j : j\in A_2)) \le 0.$$

Obviously, the increasing functions can equally be replaced by decreasing functions f_1, f_2.

Lemma 2.12 *(see [26], P_2, P_6 and P_7 and [10], Theorem 1.(8)) Negatively associated random variables possess the following properties:*

1. *For disjoint subsets $A_1, \dots, A_m$ of $\{1, \dots, k\}$ and increasing positive functions $f_1, \dots, f_m$, the inequality*

   $$\mathbb{E}\left(\prod_{i=1}^m f_i(X_j : j\in A_i)\right) \le \prod_{i=1}^m \mathbb{E}\left(f_i(X_j : j\in A_i)\right)$$

   holds. The inequality holds likewise for decreasing positive functions.

2. *Increasing functions defined on disjoint subsets of negatively associated random variables are negatively associated.*

3. *The union of independent sets of negatively associated random variables is negatively associated.*

4. *If $X_1, \dots, X_n$ are negatively associated nonnegative integer valued random variables, then*

   $$\mathbb{P}(X_i = 0, i = 1, \dots, n) \le \prod_{i=1}^n \mathbb{P}(X_i = 0).$$

For a fixed μ and conditionally on $\{\sum_{j\le c}\xi_j^\mu = 1\}$, the common distribution of the random variables ξ_i^μ, $i > c$, is Multivariate Hypergeometric (with parameters $N-c$ (number of characteristics), $c-1$ (number of drawings) and $(1,\dots,1)$ (vector of multiplicity of the characteristics)) and they are therefore negatively associated (see [26], 3.1(c)). So are ξ_i^μ, $2\le\mu\le\lceil\alpha N(1-\delta)\rceil+1$, $i>c$, given $\{\forall\mu\ge 2, \mu\le 1+\lceil\alpha N(1-\delta)\rceil : \sum_{j\le c}\xi_j^\mu = 1\}$ since these are unions of independent sets of negatively associated random variables. So the

$$\sum_{\mu=2}^{\lceil\alpha N(1-\delta)\rceil+1}\xi_i^\mu, \quad i>c$$

are conditionally negatively associated, given $\{\forall\mu\ge 2, \mu\le 1+\lceil\alpha N(1-\delta)\rceil : \sum_{j\le c}\xi_j^\mu = 1\}$, because they are increasing functions defined on disjoint subsets of negatively associated random variables. Finally, the same is true if we apply to $\sum_{\mu=2}^{\lceil\alpha N(1-\delta)\rceil+1}\xi_i^\mu$, $i>c$, the increasing functions $h_i(x)=h(x)=\mathbb{1}_{[\gamma c,\infty)}(x)$ and obtain the random variables

$$\mathbb{1}_{[\gamma c,\infty)}\left(\sum_{\mu=2}^{\lceil\alpha N(1-\delta)\rceil+1}\xi_i^\mu\right), \quad i>c.$$

These variables are in addition nonnegative and integer valued, and this implies due to Lemma 2.12, 4.:

$$\begin{aligned}
&\mathbb{P}\left(\exists i>c: \sum_{\mu=2}^{\lceil\alpha N(1-\delta)\rceil+1}\xi_i^\mu \ge \gamma c \middle| \forall\mu\in\{2,\dots,\lceil\alpha N(1-\delta)\rceil+1\} : \sum_{j\le c}\xi_j^\mu = 1\right)\\
=&1-\mathbb{P}\left(\forall i>c: \sum_{\mu=2}^{\lceil\alpha N(1-\delta)\rceil+1}\xi_i^\mu < \gamma c \middle| \forall\mu\in\{2,\dots,\lceil\alpha N(1-\delta)\rceil+1\} : \sum_{j\le c}\xi_j^\mu = 1\right)\\
\ge&1-\mathbb{P}\left(\sum_{\mu=2}^{\lceil\alpha N(1-\delta)\rceil+1}\xi_N^\mu < \gamma c \middle| \forall\mu\in\{2,\dots,\lceil\alpha N(1-\delta)\rceil+1\} : \sum_{j\le c}\xi_j^\mu = 1\right)^{N-c}.
\end{aligned}$$

The sum $\sum_{\mu=2}^{\lceil\alpha N(1-\delta)\rceil+1}\xi_N^\mu$ is, conditionally on $\{\forall\mu\ge 2, \mu\le\lceil\alpha N(1-\delta)\rceil+1 : \sum_{j\le c}\xi_j^\mu = 1\}$, Binomially distributed with parameters $\lceil\alpha N(1-\delta)\rceil$ and $\frac{c-1}{N-c}$. For an arbitrary $\varepsilon>0$, choose N big enough such that $\frac{c-1}{c}\frac{N}{N-c}>1-\varepsilon$. Let for $n\in\mathbb{N}$ and $\lambda\in(0,1)$ $R_{\lambda,n}$ denote a Binomial random variable with parameters n and λ. Then for $\lambda'\ge\lambda$ and $x\in\mathbb{R}$,

$$\mathbb{P}\left(R_{\lambda',n}\ge x\right)\ge\mathbb{P}\left(R_{\lambda,n}\ge x\right).$$

$R_{(1-\varepsilon)c/N,\lceil\alpha N(1-\delta)\rceil}$ is, as in the proof of Theorem 2.1, asymptotically Poisson distributed with parameter $(1-\varepsilon)c\lceil\alpha N(1-\delta)\rceil/N\ge(1-\varepsilon)(1-\delta)c\alpha$. The rest of the proof can now be adopted from the corresponding part of the proof of Theorem 2.1, using $(1-\varepsilon)(1-\delta)c$ instead of $(1-\delta)k$ as parameter of the Poisson distribution.

□

2.2 Hamiltonian Function and Energy Landscape

The dynamics of the neural network analysed in the previous section corresponds to an Hamiltonian function: if the updating rule of one step of the sequential dynamics has decided to change a state, this decreases the energy of the configuration. The sequential dynamics of the neural network therefore leads to local minima of the energy function. This corresponds to a spin glass notion of the associative memory: the dynamics corresponds to a Metropolis algorithm at zero temperature using the associated energy function. The stability of a pattern defined in the previous section signifies that this pattern is a local minimum of the Hamiltonian function. Newman, in [35], takes another approach to the definition of stability: a pattern is considered to be stable if it is not too far away from a local minimum, measured in Hamming distance. A small number of errors is tolerated. An energy barrier around the pattern is a sphere of a given Hamming distance centered in the pattern, such that the difference of the energy functions evaluated in a configuration on the sphere and the stored pattern exceeds a given amount, for each configuration on the sphere. If there is an energy barrier around the stored pattern, the gradient descent dynamics of the energy, started in the pattern, converges to a local minimum within the ball with the corresponding Hamming radius.

If we suppose a sequential updating, the following Hamiltonian is decreasing after an updating step of the dynamics:

Proposition 2.13 *The function* $H : \{0,1\}^N \to \mathbb{R}$,

$$H(\sigma) := -\frac{1}{2}\sum_{\substack{i\neq j\\ i,j=1}}^{N} \sigma_i\sigma_j J_{ij} + \gamma\log(N)\sum_{j=1}^{N}\sigma_j,$$

is decreasing along each step $\bar{T}_i$, $1 \le i \le N$, *of a sequential dynamics* $\bar{T}$.

Proof of Proposition 2.13: Indeed, using the sequential dynamics $\bar{T}_i(\sigma) = (\sigma_1, \dots, \sigma_{i-1}, T_i(\sigma), \sigma_{i+1}, \dots, \sigma_N)$, the difference of the Hamiltonians before and after the updating is

$$H(\sigma) - H(\bar{T}_i(\sigma)) = -(\sigma_i - T_i(\sigma))\sum_{j\le N, j\neq i}\sigma_j J_{ij} + (\sigma_i - T_i(\sigma))\gamma\log(N).$$

If neuron i is deactivated during this step, the difference is positive:

$$\begin{aligned} H(\sigma) - H(\bar{T}_i(\sigma)) &= -\,(\sigma_i - T_i(\sigma))\sum_{j\neq i}\sigma_j J_{ij} + (\sigma_i - T_i(\sigma))\gamma\log(N)\\ &= -\sum_{j\neq i}\sigma_j J_{ij} + \gamma\log(N) > 0 \end{aligned}$$

because the local field $S_i(\sigma) = \sum_{j\neq i}\sigma_j J_{ij}$ is smaller than the threshold $\gamma\log(N)$, if σ_i is flipped from 1 to 0. For the opposite situation, if the value of a neuron is upgraded from 0 to 1, we have

$$H(\sigma) - H(\bar{T}_i(\sigma)) = -\,(\sigma_i - T_i(\sigma))\sum_{j\neq i}\sigma_j J_{ij} + (\sigma_i - T_i(\sigma))\gamma\log(N)$$

$$= \sum_{j \neq i} \sigma_j J_{ij} - \gamma \log(N) \geq 0,$$

because then the local field $S_i(\sigma)$ is at least equal to the threshold $\gamma \log(N)$.

□

We want to find valleys of the energy function centered in the stored patterns including their depth and radii. We therefore consider a fixed pattern in a given Hamming distance of a stored pattern ξ^μ and analyse the difference of the Hamiltonians evaluated in this pattern and ξ^μ.

Proposition 2.14 *Consider the given neural network with Hamiltonian function H using the threshold $\gamma \log(N)$, $\gamma \in (0,1)$, and an arbitrary fixed $1 \leq \mu \leq M$. There are $\eta > 0$, $\varepsilon > 0$, $\alpha > 0$ such that the Hamiltonian function of some arbitrary pattern $\bar{\xi}^\mu$ in Hamming-distance $\eta \log(N)$ of ξ^μ exceeds the value of the message ξ^μ by at least $\varepsilon \log(N)^2$, with high probability, if $M \leq \alpha N^2 / \log(N)^2$ messages are stored in the network.*

This choice is maximal: for a Hamming distance $h(N)$, $\log(N) = o(h(N))$, and an arbitrary pattern $\bar{\xi}^\mu$ in Hamming distance $h(N)$ of ξ^μ, we have even

$$\lim_{N \to \infty} \mathbb{P}(H(\bar{\xi}^\mu) \leq H(\xi^\mu)) = 1.$$

Proof of Proposition 2.14: Take a stored message, e. g., ξ^1. To shorten expressions, we define for $\bar{\xi}^1$ the set J by $J := \{i : \bar{\xi}_i^1 \neq \xi_i^1\}$ and

$$J_1 = \{i : \bar{\xi}_i^1 = 1, \xi_i^1 = 0\}, \quad J_2 = \{i : \bar{\xi}_i^1 = 0, \xi_i^1 = 1\}, \quad J = J_1 \cup J_2.$$

We call $|J_1| = k_1$, $|J_2| = k_2$, $f = k_1 + k_2$. For fixed k, k_1 and k_2, we assume that the excited neurons of ξ^1 are exactly the first k ones and consider $\bar{\xi}^1$ such that $J_1 = \{k+1, \ldots, k+k_1\}$ and $J_2 = \{k - k_2, \ldots, k\}$. The difference between the Hamiltonians is, for fixed k, given by

$$H(\xi^1) - H(\bar{\xi}^1) = -\frac{1}{2} \sum_{\substack{i \neq j \\ i,j=1}}^{N} J_{ij} \left[\xi_i^1 \xi_j^1 - \bar{\xi}_i^1 \bar{\xi}_j^1\right] + \gamma \log(N) \sum_{j=1}^{N} \left[\xi_j^1 - \bar{\xi}_j^1\right]$$

$$= -\frac{1}{2} \left[\sum_{\substack{i \neq j \\ i,j=1}}^{N} \xi_i^1 \xi_j^1 \left[\xi_i^1 \xi_j^1 - \bar{\xi}_i^1 \bar{\xi}_j^1\right] + \sum_{\mu=2}^{M} \sum_{\substack{i \neq j \\ i,j=1}}^{N} \xi_i^\mu \xi_j^\mu \left[\xi_i^1 \xi_j^1 - \bar{\xi}_i^1 \bar{\xi}_j^1\right] \right] + \gamma \log(N)[k - [k - k_2 + k_1]]$$

$$= -\binom{k}{2} + \binom{k - k_2}{2} - \frac{1}{2} \sum_{\mu=2}^{M} \sum_{\substack{i \neq j \\ i,j=1}}^{N} \xi_i^\mu \xi_j^\mu \left[\xi_i^1 \xi_j^1 - \bar{\xi}_i^1 \bar{\xi}_j^1\right] + \gamma \log(N)(k_2 - k_1)$$

$$= -kk_2 + \frac{k_2^2 + k_2}{2} - \frac{1}{2} \sum_{\mu=2}^{M} \sum_{\substack{i \neq j \\ i,j=1}}^{N} \xi_i^\mu \xi_j^\mu \left[\xi_i^1 \xi_j^1 - \bar{\xi}_i^1 \bar{\xi}_j^1\right] + \gamma \log(N)(k_2 - k_1). \tag{2.30}$$

We define the random variables $A_\mu(k,k_2)$, $B_\mu(k,k_2)$ and $C_\mu(k,k_1)$ by

$$A_\mu(k,k_2) := \sum_{j=1}^{k-k_2} \xi_j^\mu, \quad B_\mu(k,k_2) := \sum_{j=k-k_2+1}^{k} \xi_j^\mu, \quad C_\mu(k,k_1) := \sum_{j=k+1}^{k+k_1} \xi_j^\mu.$$

Recalling that $A_\mu(k,k_2)$, $B_\mu(k,k_2)$ and $C_\mu(k,k_1)$ depend on k and k_2 or k and k_1, respectively, we fix k, k_1 and k_2 and write simply A_μ, B_μ and C_μ. We then can rewrite the random part of the difference of the energy functions:

$$\begin{aligned}
-\frac{1}{2}\sum_{\mu=2}^{M}\sum_{\substack{i\neq j\\ i,j=1}}^{N} \xi_i^\mu \xi_j^\mu \left[\xi_i^1\xi_j^1 - \bar{\xi}_i^1\bar{\xi}_j^1\right] &= -\frac{1}{2}\sum_{\mu=2}^{M}\sum_{\substack{i\neq j\\ i,j=1}}^{k} \xi_i^\mu \xi_j^\mu + \frac{1}{2}\sum_{\mu=2}^{M} \sum_{\substack{i\neq j\\ i,j\in\{1,\dots,k_2\}\\ \cup\{k+1,\dots,k+k_1\}}} \xi_i^\mu\xi_j^\mu \\
&= \sum_{\mu=2}^{M}\left[-\binom{A_\mu+B_\mu}{2} + \binom{A_\mu+C_\mu}{2}\right] \\
&= \frac{1}{2}\sum_{\mu=2}^{M}\left[C_\mu^2 + 2A_\mu C_\mu - C_\mu - B_\mu^2 - 2A_\mu B_\mu + B_\mu\right] \\
&= \sum_{\mu=2}^{M}\left[\binom{C_\mu}{2} + A_\mu C_\mu - \binom{B_\mu}{2} - A_\mu B_\mu\right]. \qquad (2.31)
\end{aligned}$$

The variables $(A_\mu, \mu \geq 1, B_\mu, \mu \geq 1, C_\mu, \mu \geq 1)$ are (for fixed k, k_1 and k_2) independent. For each μ, the distributions of A_μ, B_μ and C_μ are Binomial: for A_μ the parameters are $k - k_2$ and p, for B_μ the parameters are k_2 and p and for C_μ they are k_1 and p. The expectation of one summand of the preceding sum is thus

$$\mathbb{E}\left[\binom{C_\mu}{2} + A_\mu C_\mu - \binom{B_\mu}{2} - A_\mu B_\mu\right] = p^2 \cdot \left[\binom{k_1}{2} + (k-k_2)k_1 - \binom{k_2}{2} - (k-k_2)k_2\right].$$

For some fixed $k \leq (1+\delta)\log(N)$, assuming w.l.o.g. that $h(N)p \to 0$, a short computation yields

$$\begin{aligned}
&\mathbb{V}\left[\binom{C_\mu}{2} + A_\mu C_\mu - \binom{B_\mu}{2} - A_\mu B_\mu\right] \\
=&p^2\left(\binom{k_1}{2} + \binom{k_2}{2} + k_1(k-k_2) + k_2(k-k_2)\right) + \mathcal{O}(p^3(k_1^2(k-k_2) + (k-k_2)^2 k_1)).
\end{aligned}$$

Using the Chebyshev inequality, we can now show that the Hamiltonian on the Hamming sphere indeed exceeds the Hamiltonian in ξ^1 by at least $\varepsilon \log(N)^2$, if η, ε and α are appropriately chosen and the radius of the sphere is $\eta \log(N)$. This statement is also included in the next theorem and therefore not shown in detail. To show that this choice is maximal, assume that $\log(N) = o(h(N))$. Recalling that A_μ, B_μ and C_μ are defined in dependence on k, k_1 and k_2, the Chebyshev inequality yields

$$\lim_{N\to\infty} \max_{\substack{k,k_1,k_2\in\mathbb{N}: k/\\ \log(N)\in(1-\delta,1+\delta),\\ k_2\leq k, k_1=h(N)-k_2}} \mathbb{P}\left(\frac{\sum_{\mu=2}^{M}\binom{C_\mu}{2} + A_\mu C_\mu - \binom{B_\mu}{2} - A_\mu B_\mu}{\alpha\left(\binom{h(N)-k_2}{2} + \binom{k_2}{2} + h(N)(k-k_2)\right)} \notin (1-\delta, 1+\delta)\right) = 0.$$

Since $k_2 \leq k$, we observe that for $k \leq (1+\delta)\log(N)$, the expression

$$\alpha(1-\delta)\left[\binom{h(N)-k_2}{2}+\binom{k_2}{2}+h(N)(k-k_2)\right]-kk_2+\frac{k_2^2+k_2}{2}-\gamma\log(N)(h(N)-k_2)$$

tends to infinity, as N does. We finally obtain for large enough N and an arbitrary pattern $\bar{\xi^1}$ in Hamming distance $h(N)$ of ξ^1, chosen in dependence on k:

$$\mathbb{P}(H(\bar{\xi^1}) \leq H(\xi^1)) \geq \mathbb{P}(A_\delta) \min_{\substack{k/\log(N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}(H(\bar{\xi^1}) \leq H(\xi^1)|\mathcal{Z}_k) \geq \mathbb{P}(A_\delta)\cdot$$

$$\min_{\substack{k,k_1,k_2\in\mathbb{N}:k/\\ \log(N)\in(1-\delta,1+\delta),\\ k_2\leq k,k_1=h(N)-k_2}} \mathbb{P}\left(\frac{\sum_{\mu=2}^{M}\binom{C_\mu}{2}+A_\mu C_\mu-\binom{B_\mu}{2}-A_\mu B_\mu}{\alpha\left(\binom{h(N)-k_2}{2}+\binom{k_2}{2}+h(N)(k-k_2)\right)} \in (1-\delta,1+\delta)\right)$$

and both probabilities tend to 1, as N tends to infinity. The random variables A_μ, B_μ and C_μ are defined in dependence on k, k_1, k_2, as described in the proof.

□

The radii will thus be maximal of order $\log(N)$; the height of the valleys is of order $\log(N)^2$.

Theorem 2.15 *Consider the Hamiltonian in Amari's model using a threshold variable $0<\gamma<1$ and a number of stored patterns $M=\alpha N^2/\log(N)^2$ with a capacity variable α such that $\alpha<\gamma$,*

$$-\gamma\log\left(\frac{\gamma}{\alpha}\right)+\gamma-\alpha<-1.$$

Fix some arbitrary ξ^μ, $\mu\in\{1,\ldots,M\}$. Then there is with high probability an energy barrier around ξ^μ, that is, there are $\eta>0$ and $\varepsilon>0$ such that

$$\lim_{N\to\infty}\mathbb{P}(\forall\bar{\xi^\mu}\in S_{\lfloor\eta\log(N)\rfloor}(\xi^\mu): H(\xi^\mu)-H(\bar{\xi^\mu})\leq-\varepsilon\log(N)^2)=1.$$

$S_{\lfloor\eta\log(N)\rfloor}(\xi^\mu)$ describes the sphere with radius $\lfloor\eta\log(N)\rfloor$ with respect to the Hamming distance $d_{\mathcal{H}}(\cdot,\cdot)$ centered in ξ^μ.

Proof of Theorem 2.15: We first assume that ξ^1 consists of exactly k 1's and that these are located in the first k places. This event is, as in the previous proofs, denoted by $\mathcal{Z}_k$. Furthermore, we assume that $k=\rho\log(N)$ with $\rho\in(1-\delta,1+\delta)$ for some fixed $\delta>0$. For a pattern $\bar{\xi^\mu}$ in Hamming-distance $\eta\log(N)$ (w.l.o.g. we assume $\eta\log(N)\in\mathbb{N}$) we denote by $\eta_1\log(N)=k_1$ the number of neurons that are not excited in ξ^1, but in $\tilde{\xi^1}$, and by $\eta_2\log(N)=k_2$ the number of neurons whose spins have been changed from 1 to 0. The patterns in Hamming distance $\eta\log(N)$ correspond uniquely to the subsets $J\subseteq\{1,\ldots,N\}$, $|J|=\eta\log(N)$. The pattern that differs from ξ^1 exactly in the values of the neurons belonging to the subset J is denoted by ξ^1_J.

As shown in the previous proposition, the difference of the Hamiltonians (for fixed choices of k, k_1 and k_2) is given by

$$H(\xi^1) - H(\xi^1_J) = \sum_{\mu=2}^{M} \left[\binom{C^J_\mu}{2} - \binom{B^J_\mu}{2} + A^J_\mu C^J_\mu - A^J_\mu B^J_\mu \right] - kk_2 + \frac{k_2^2 + k_2}{2} + \gamma \log(N)(k_2 - k_1)$$

with

$$A^J_\mu := \sum_{i \in \{1,\dots,k\} \setminus J_2} \xi^\mu_i, \quad B^J_\mu := \sum_{i \in J_2} \xi^\mu_i, \quad C^J_\mu := \sum_{i \in J_1} \xi^\mu_i.$$

For fixed numbers k_1, k_2 and k, $k_1 + k_2 = \eta \log(N)$, there are $\binom{N-k}{k_1}\binom{k}{k_2}$ possibilities to choose patterns that differ from ξ^1 in exactly k_1 of the inactive neurons of ξ^1 and in exactly k_2 of the active neurons of ξ^1. The probability of having at least one such configuration whose Hamiltonian is below the Hamiltonian of ξ^1 plus the threshold $\varepsilon \log(N)^2$ is thus bounded by

$$\begin{aligned} &\mathbb{P}\left[\exists J_1 \subseteq \{k+1,\dots,N\}, |J_1| = k_1, J_2 \subseteq \{1,\dots,k\}, |J_2| = k_2 : \right. \\ &\qquad \left. H(\xi^1_J) \le H(\xi^1) + \varepsilon \log(N)^2 | \mathcal{Z}_k \right] \\ &\le \binom{N-k}{k_1}\binom{k}{k_2} \mathbb{P}\left(H(\xi^1_{\tilde{J}}) \le H(\xi^1) + \varepsilon \log(N)^2 | \mathcal{Z}_k\right) \end{aligned}$$

for an arbitrary subset $\tilde{J} = \tilde{J}_1 \cup \tilde{J}_2$, $\tilde{J}_1 \subseteq \{k+1,\dots N\}$, $\tilde{J}_2 \subseteq \{1,\dots,k\}$, with $|\tilde{J}_1| = k_1$, $|\tilde{J}_2| = k_2$. We take $\tilde{J}_2 = \{k - k_2 + 1, \dots, k\}$ and $\tilde{J}_1 = \{k+1,\dots,k+k_1\}$. To prove the assertion of the Theorem, we show that

$$\begin{aligned} \lim_{N\to\infty} \max_{\substack{k \in \mathbb{N}: k/\log(N) \\ \in (1-\delta, 1+\delta)}} \eta \log(N) \cdot \max_{0 \le k_1 \le \eta \log(N)} \mathbb{P}\left[\exists J_1 \subseteq \{k+1,\dots,N\}, |J_1| = k_1, \right. \\ \left. J_2 \subseteq \{1,\dots,k\}, |J_2| = k_2 = \eta\log(N) - k_1 : H(\xi^1_J) \le H(\xi^1) + \varepsilon \log(N)^2 | \mathcal{Z}_k \right] = 0. \end{aligned} \tag{2.32}$$

As the distributions of A^J_μ, B^J_μ and C^J_μ do not depend on the subset J for fixed k_1 and k_2, we renounce on the index J. We observe that we need to bound the probability $\mathbb{P}\left(H(\xi^1_{\tilde{J}}) \le H(\xi^1) + \varepsilon \log(N)^2 | \mathcal{Z}_k\right)$ for each choice of k_1 and k of the considered sets by $\exp[-\log(N)^2 \kappa]$, with $\kappa > \eta$. For this purpose, we determine the exponential moments of the random variables $\binom{C_\mu}{2} - \binom{B_\mu}{2} + A_\mu C_\mu - A_\mu B_\mu$, $\mu \ge 2$. First,

$$\begin{aligned} &\mathbb{E}\left[\exp\left(t\left(\binom{C_\mu}{2} - \binom{B_\mu}{2} + A_\mu C_\mu - A_\mu B_\mu\right)\right)\right] \\ &= \sum_{i_1=0}^{k-k_2}\sum_{i_2=0}^{k_2}\sum_{i_3=0}^{k_1} \binom{k-k_2}{i_1}\binom{k_2}{i_2}\binom{k_1}{i_3} p^{i_1+i_2+i_3}(1-p)^{k+k_1-i_1-i_2-i_3} e^{t\left(\binom{i_3}{2} - \binom{i_2}{2} - i_1 i_2 + i_1 i_3\right)} \end{aligned}$$

because A_μ, B_μ and C_μ are independent and Binomially distributed. Having in mind that $p = \log(N)/N$, we sort according to the power of p and bound the expectation by

$$\sum_{i_1=0}^{k-k_2}\sum_{i_2=0}^{k_2}\sum_{i_3=0}^{k_1} \binom{k-k_2}{i_1}\binom{k_2}{i_2}\binom{k_1}{i_3} p^{i_1+i_2+i_3}(1-p)^{k+k_1-i_1-i_2-i_3} e^{t\left(\binom{i_3}{2} - \binom{i_2}{2} - i_1 i_2 + i_1 i_3\right)}$$

$$
\begin{aligned}
\leq &(1-p)^{k+k_1} + p(1-p)^{k+k_1-1}(k+k_1) + p^2(1-p)^{k+k_1-2}\left[\binom{k_1}{2}e^t + k_1(k-k_2)e^t \right. \\
&\left. + \binom{k_2}{2}e^{-t} + k_2(k-k_2)e^{-t} + \binom{k-k_2}{2} + k_1k_2\right] \\
&+ \log(N)^3 \max_{\substack{i_1,i_2,i_3:i_1+i_2+i_3\geq 3\\ i_1\leq k-k_2, i_2\leq k_2, i_3\leq k_1}} \binom{k-k_2}{i_1}\binom{k_2}{i_2}\binom{k_1}{i_3} p^{i_1+i_2+i_3} e^{t\left(\binom{i_3}{2}-\binom{i_2}{2}+i_1i_3-i_1i_2\right)}.
\end{aligned}
$$

The last line is at most

$$
\begin{aligned}
&\log(N)^3 \max_{\substack{i_1,i_2,i_3:i_1+i_2+i_3\geq 3\\ i_1\leq k-k_2, i_2\leq k_2, i_3\leq k_1}} \binom{k-k_2}{i_1}\binom{k_2}{i_2}\binom{k_1}{i_3} p^{i_1+i_2+i_3} e^{t\left(\binom{i_3}{2}-\binom{i_2}{2}+i_1i_3-i_1i_2\right)} \\
\leq &\log(N)^3 \max_{\substack{i_1,i_2,i_3:i_1+i_2+i_3\geq 3\\ i_1\leq k-k_2, i_2\leq k_2, i_3\leq k_1}} k^{i_1}\, k_2^{i_2}\, k_1^{i_3} p^{i_1+i_2+i_3} e^{t\left(\binom{i_3}{2}-\binom{i_2}{2}+i_1i_3-i_1i_2\right)}. \qquad (2.33)
\end{aligned}
$$

To determine the maximal value of this expression, we first fix values t, i_1 and i_3. It will later turn out that t does not depend on N; this is also assumed in the proof.

Since $k_2p < 1$ and $i_1 \geq 0$, $t > 0$, one sees immediately that the expression decreases in i_2. We thus restrict the choice of i_2 to the set $\{0,1,2,3\}$ in order to determine the maximal value in (2.33).

For fixed $k \leq (1+\delta)\log(N)$, $t < \frac{1}{\eta}$, i_2 and i_3, the expression in (2.33) is decreasing in i_1: we consider

$$
p^{i_1}\, k^{i_1} e^{ti_1i_3 - i_1i_2} = \exp\left[i_1\left(\log\log(N) - \log(N) + \log(k) + ti_3 - ti_2\right)\right].
$$

Since $i_3 \leq \eta\log(N)$, $t < \frac{1}{\eta}$,

$$
\log\log(N) - \log(N) + \log(k) + ti_3 - ti_2 < 0
$$

for N large enough and each choice of i_2 and i_3 of the considered sets. So the whole expression in the last line of (2.33) is, for fixed i_2 and i_3 and large enough N, decreasing in i_1.

For fixed t, i_2 and i_1, we observe finally

$$
p^{i_3}\, k_1^{i_3} e^{t\left(\binom{i_3}{2}+i_1i_3\right)} = \exp\left[\frac{i_3^2}{2}t + i_3\left(\log\log(N) - \log(N) + \log(k_1) + ti_1 - \frac{t}{2}\right)\right].
$$

This is either maximal in $i_3 = k_1$ or in the smallest value of i_3, $i_3 \in \{0,1,2,3\}$, depending on i_1 and i_2 (because their sum must be at least 3).

For $t < \frac{1}{\eta}$, $k \leq (1+\delta)\log(N)$ and large enough N, the maximal argument in (2.33) is an element of the set

$$
\{(i_1,i_2,i_3) \in \{0,1,2,3\}^3 : i_1+i_2+i_3 = 3\} \cup \{(0,0,k_1)\}. \qquad (2.34)
$$

For each triple of the set on the left hand side, we have

$$
p^{i_1+i_2+i_3}\, k^{i_1}\, k_2^{i_2}\, k_1^{i_3} e^{t\left(\binom{i_3}{2}-\binom{i_2}{2}+i_1i_3-i_1i_2\right)} = \mathcal{O}\left(\frac{\log(N)^6}{N^3}\right).
$$

Inserting the last triple $(0, 0, k_1)$ in the right hand side of the last line in (2.33) yields

$$p^{i_1+i_2+i_3}\, k^{i_1}\, k_2^{i_2}\, k_1^{i_3} e^{t\left(\binom{i_3}{2}-\binom{i_2}{2}+i_1 i_3 - i_1 i_2\right)} \leq \exp\left[k_1\left(\log\log(N) - \log(N) + \log(k_1) + t\frac{k_1}{2}\right)\right]$$

and the argument of the exponential is again quadratic in k_1, which implies that it is either maximal in $k_1 = \eta\log(N)$ or in $k_1 = 3$. It remains to consider $k_1 = \eta\log(N)$; using that $t < \frac{1}{\eta}$, we have

$$\begin{aligned}
&\exp\left[\eta\log(N)\left(\log\log(N) - \log(N) + \log(\eta\log(N)) + t\frac{\eta\log(N)}{2}\right)\right]\\
\leq&\exp\left[\eta\log(N)\left(\log\log(N) - \frac{1}{2}\log(N) + \log(\eta\log(N))\right)\right] = o\left(\frac{\log(N)^6}{N^3}\right).
\end{aligned}$$

The maximum in (2.33) is therefore attained by one of the elements of $\{(i_1, i_2, i_3) \in \{0, 1, 2, 3\}^3 : i_1 + i_2 + i_3 = 3\}$. We can now estimate the exponential moment under the assumption $t < \frac{1}{\eta}$, using the previous calculations on page 46:

$$\begin{aligned}
&\mathbb{E}\left[\exp\left[t\left(\binom{C_\mu}{2} - \binom{B_\mu}{2} + A_\mu C_\mu - A_\mu B_\mu\right)\right]\right]\\
\leq&1 + p^2\left[\binom{k_1}{2}e^t + k_1(k-k_2)e^t + \binom{k_2}{2}e^{-t} + (k-k_2)k_2 e^{-t}\right.\\
&\left.+\binom{k-k_2}{2} + k_1 k_2 - \binom{k+k_1}{2}\right] + \mathcal{O}\left(\frac{\log(N)^9}{N^3}\right).
\end{aligned}$$

The $\binom{C_\mu}{2} - \binom{B_\mu}{2} + A_\mu C_\mu - A_\mu B_\mu$, $\mu \geq 2$, are independent, which yields

$$\begin{aligned}
&\mathbb{E}\left[\exp\left[t\sum_{\mu=2}^{M}\left(\binom{C_\mu}{2} - \binom{B_\mu}{2} + A_\mu C_\mu - A_\mu B_\mu\right)\right]\right]\\
=&\mathbb{E}\left[\exp\left[t\left(\binom{C_2}{2} - \binom{B_2}{2} + A_2 C_2 - A_2 B_2\right)\right]\right]^{M-1}\\
\leq&\exp\left[\alpha\left[\binom{k_1}{2}e^t + k_1(k-k_2)e^t + \binom{k_2}{2}e^{-t} + (k-k_2)k_2 e^{-t}\right.\right.\\
&\left.\left.+\binom{k-k_2}{2} + k_1 k_2 - \binom{k+k_1}{2}\right] + \mathcal{O}\left(\frac{\log(N)^7}{N}\right)\right].
\end{aligned}$$

We obtain for fixed k, k_1 and k_2 with the aid of the exponential Chebyshev inequality

$$\begin{aligned}
&\binom{N}{k_1}\binom{k}{k_2}\mathbb{P}\left(H(\xi^1_{\tilde{J}}) \leq H(\xi^1) + \varepsilon\log(N)^2 | \mathcal{Z}_k\right) \leq \exp\left[k_1\log(N) + k_2\log(k)\right]\cdot\\
&\exp\left[-t(kk_2 - k_2^2/2 - k_2/2 - \gamma\log(N)(k_2-k_1) - \varepsilon\log(N)^2) + \alpha\left[\binom{k_1}{2}e^t + k_1(k-k_2)e^t\right.\right.\\
&\left.\left.+\binom{k_2}{2}e^{-t} + (k-k_2)k_2e^{-t} + \binom{k-k_2}{2} + k_1k_2 - \binom{k+k_1}{2}\right] + \mathcal{O}\left(\log(N)^7/N\right)\right].
\end{aligned} \tag{2.35}$$

We know that $k_1 + k_2 = \eta \log(N)$ which means that the argument in the exponential is quadratic in k_1 if we replace k_2 by $\eta \log(N) - k_1$: the function in k_1 used as argument in the exponential is of the form $f(k_1) = ak_1^2 + bk_1 + c$, with

$$a = \alpha(2\sinh(t) + \cosh(t) - 1) + t/2 > 0.$$

The maximum is thus attained by $k_1 = 0$ or $k_1 = \eta \log(N)$.

If $k_1 = 0$, we use $k_2 = \eta \log(N)$ in (2.35). We only need to consider $k/\log(N) \in (1-\delta, 1+\delta)$; due to the equality $\binom{k_2}{2} + (k-k_2)k_2 = \binom{k}{2} - \binom{k-k_2}{2} = kk_2 - \frac{k_2^2+k_2}{2}$, (2.35) is

$$\begin{aligned}
&\exp\Bigg[k_2\log(k) - tk_2\Big(k - k_2/2 - 1/2 - \gamma\log(N)\Big) + t\varepsilon\log(N)^2 + \\
&\qquad \alpha\left(kk_2(e^{-t}-1) - \frac{k_2^2+k_2}{2}(e^{-t}-1)\right) + \mathcal{O}\left(\log(N)^7/N\right)\Bigg] \\
&= \exp\Bigg[t\varepsilon\log(N)^2 + \eta\log(N)\Bigg[\log(k) - t\left(k - \frac{\eta\log(N)}{2} - \frac{1}{2} - \gamma\log(N)\right) \\
&\qquad +\alpha\Big(k(e^{-t}-1) - \frac{\eta\log(N)+1}{2}(e^{-t}-1)\Big)\Bigg] + \mathcal{O}\left(\log(N)^7/N\right)\Bigg] \\
&\leq \exp\Bigg[t\varepsilon\log(N)^2 + \eta\log(N)\Bigg[\log[(1+\delta)\log(N)] - \log(N)t\left(1-\delta-\frac{\eta}{2}-\gamma\right) - \frac{1}{2}t \\
&\quad +\alpha\log(N)\Big((1-\delta)(e^{-t}-1) - \frac{\eta}{2}(e^{-t}-1)\Big) + \alpha\frac{1}{2}(e^{-t}-1)\Bigg] + \mathcal{O}\left(\log(N)^7/N\right)\Bigg].
\end{aligned}$$

Assuming that the maximum in (2.32) is attained in $k_1 = 0$, we observe that (2.32) is fulfilled if

$$\varepsilon < \eta(1 - \delta - \eta/2 - \gamma).$$

To see this, we consider all the terms of order $\log(N)^2$ in the argument of the exponential function and use an arbitrary $t > 0$; note that $e^{-t} - 1 < 0$ and $1 - \delta - \eta/2 > 0$ if δ is small enough, since $\eta < 1$. This condition can be fulfilled for suitable $\delta > 0$ if

$$\varepsilon < \eta(1 - \gamma - \eta/2)$$

and, to obtain a positive ε,

$$0 < \eta < 2(1-\gamma)$$

Since $\gamma < 1$, there are ε, η such that the conditions are fulfilled.

We continue with the second case, $k_1 = \eta \log(N)$, $k_2 = 0$. The exponential in (2.35) is at most

$$\begin{aligned}
&\exp\left[\log(N)\left[k_1 - t\gamma k_1 + t\varepsilon\log(N)\right]\right]\exp\left[\alpha\left[\binom{k_1}{2}e^t + k_1ke^t + \binom{k}{2} - \binom{k+k_1}{2}\right]\right] \\
&\leq \exp\left[\log(N)^2\left(\eta - t\gamma\eta + t\varepsilon + \alpha\frac{\eta^2}{2}(e^t-1) + \alpha(1+\delta)\eta(e^t-1)\right) + \mathcal{O}(\log(N))\right].
\end{aligned}$$

We used the maximal argument for k, $k = (1+\delta)\log(N)$; it is maximal because $e^t - 1 > 0$ for $t > 0$. Note also that $\binom{k}{2} - \binom{k+k_1}{2} = -kk_1 - \binom{k_1}{2}$.

Minimising the expression

$$-t\gamma\eta + t\varepsilon + \alpha\frac{\eta^2}{2}(e^t - 1) + \alpha(1+\delta)\eta(e^t - 1)$$

yields

$$t^*(\delta, \eta, \varepsilon, \gamma, \alpha) = \log\left(\frac{\gamma\eta - \varepsilon}{\eta\alpha(\eta/2 + 1 + \delta)}\right) = \log\left(\frac{\gamma - \frac{\varepsilon}{\eta}}{\alpha(\eta/2 + 1 + \delta)}\right)$$

as the extremal argument. This is positive if

$$\varepsilon < \eta\left(\gamma - \alpha(\eta/2 + 1 + \delta)\right).$$

Since $\alpha < \gamma$, there are positive ε, η and δ such that the inequality is fulfilled.

The probability in (2.32) converges to 0 as N tends to infinity, if

$$-t\gamma\eta + t\varepsilon + \alpha\frac{\eta^2}{2}(e^t - 1) + \alpha(1+\delta)\eta(e^t - 1) < -\eta$$

is satisfied for some $t > 0$. Inserting $t^*(\delta, \eta, \varepsilon, \gamma, \alpha)$ yields the condition

$$-\left(\gamma - \frac{\varepsilon}{\eta}\right)\log\left(\frac{\gamma - \frac{\varepsilon}{\eta}}{\alpha(\eta/2 + 1 + \delta)}\right) + \gamma - \frac{\varepsilon}{\eta} - \alpha\frac{\eta}{2} - \alpha(1+\delta) < -1.$$

If all derived conditions are fulfilled, the probability

$$\begin{aligned}
&\mathbb{P}(\exists\bar{\xi^\mu} : d_{\mathcal{H}}(\xi^\mu, \bar{\xi^\mu}) = \eta\log(N), H(\xi^\mu) - H(\bar{\xi^\mu}) \geq -\varepsilon\log(N)^2) \\
\leq &\mathbb{P}(A_\delta^c) + \max_{\substack{k/\log(N)\in \\ (1-\delta, 1+\delta)}} \mathbb{P}(\exists\bar{\xi^\mu} : d_{\mathcal{H}}(\xi^\mu, \bar{\xi^\mu}) = \eta\log(N), H(\xi^\mu) - H(\bar{\xi^\mu}) \geq -\varepsilon\log(N)^2|\mathcal{Z}_k)
\end{aligned}$$

tends to 0.

It remains to show that it is possible to choose the variables such that all the conditions are fulfilled. Setting $\bar{\gamma} := \gamma - \frac{\varepsilon}{\eta}$ and $\bar{\alpha} := \alpha\left(\frac{\eta}{2} + 1 + \delta\right)$, the inequality can be traced back to inequality (2.1) whose solutions have been analysed during the last section. For $\bar{\gamma} < \gamma^*$, the inequality is fulfilled if $\bar{\alpha} < \alpha^*(\bar{\gamma})$. In particular, for a pair $\gamma \in (0, 1)$ and $\alpha < \gamma$, such that (2.1) holds, there are $\varepsilon > 0$ and $\eta > 0$ such that all the conditions are fulfilled by choosing η and ε small enough, especially to reach $t^* < 1/\eta$.

□

Remark 2.16 *In the Hopfield model, the analysis of the Hamiltonian and the less restrictive notion of capacity allow to store αN instead of $\alpha N/\log(N)$ patterns such that they are stable (tolerating small errors). In contrary to the Hopfield model, we do not obtain better results concerning the bounds on α compared to the stability results of the previous section. The difference to the result of Theorem 2.7 is that Theorem 2.15 concerns all corrupted patterns in a given Hamming-distance and shows that their energy if higher than the one of the stored pattern, with high probability, whereas the results obtained in Theorem 2.7 concern an arbitrary fixed pattern with a certain number of errors.*

3 The Ternary Simple Model

This model is an adaption of Amari's network to ternary spin values as proposed in [31]. The network we consider consists again of N neurons, $V = \{1, \dots, N\}$, of which each one can take values in $S_2 = \{-1, 0, 1\}$. A spin $\sigma_i = 0$ signifies that neuron i is inactive; an active neuron can be in the states -1 or 1. The state 0 has a special role; this model is therefore different in spirit from the three state Hopfield model (Potts-Hopfield model, see e. g., [32]) using three equal states. We store randomly M patterns $\xi^\mu \in \{-1, 0, 1\}^N$, $1 \le \mu \le M$. The values of these patterns are indicated by random variables ξ_i^μ, $1 \le i \le N$, and the random variables $(\xi_i^\mu, 1 \le \mu \le M, 1 \le i \le N)$ are independent and identically distributed such that

$$\begin{aligned}
\mathbb{P}(\xi_i^\mu = 0) &= 1 - \frac{\log(N)}{N} = 1 - p_N, \\
\mathbb{P}(\xi_i^\mu = \pm 1) &= \frac{\log(N)}{2N} = \frac{1}{2} p_N.
\end{aligned}$$

We will have in mind that p_N depends on N but omit the index to achieve a better readability. Given an input spin configuration $\sigma \in \{-1, 0, 1\}^N$, the dynamics described in the next paragraph is applied. Just as in Amari's model, the edge set of the graph is $E = \{\{i, j\} : i, j \in \{1, \dots, N\}, i \neq j\}$ and on each edge, the synaptic efficacy J_{ij} is defined by

$$J_{ij} := \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu.$$

We define the local field $S_i(\sigma)$ in dependence on the synaptic efficacies by

$$S_i(\sigma) := \sum_{j=1, j\neq i}^{N} \sigma_j J_{ij} = \sum_{j \neq i} \sum_{\mu=1}^{M} \sigma_j \xi_i^\mu \xi_j^\mu.$$

The parallel dynamics T will now assign a non-zero value to neuron i if and only if the absolute value of the local field S_i is at least equal to a given threshold. This threshold will be chosen as $\gamma \log(N)$, with $\gamma \in (0, 1)$, as in the first suggestion of the previous model, because the expected number of neurons taking non-zero values per message is equal to $\log(N)$. Other than in Chapter 2, the choice of γ is indeed restricted to the interval $(0, 1)$, as we will see in Proposition 3.2. If the absolute value of $S_i(\sigma)$ exceeds the threshold, the neuron is mapped to the sign of the local field. That is,

$$T_i(\sigma) = \begin{cases} \pm 1 & S_i(\sigma) \gtreqless \pm\gamma \log(N) \\ 0 & |S_i(\sigma)| < \gamma \log(N). \end{cases}$$

Löwe and Vermet analysed this model in [31] for fixed p not depending on N and showed that stability (for one or for all patterns) can be reached as well as that a certain number of errors can be corrected, with high probability, if the number of stored patterns does not exceed a certain bound (depending on the number of errors). The threshold of their model cannot be utilised in the present, extremely sparse version of the model and has been replaced by $\gamma \log(N)$, as described above. We will see that this choice is optimal to obtain a maximal capacity.

The model described so far can be varied in the way that one chooses, as in Chapter 2, the patterns $\xi^1, \dots, \xi^M$ independently and uniformly from the set of patterns with exactly $c = \log(N)$ active neurons. We will show that the lower bounds on α concerning the stability and error correction for the model with independent spins are also valid for this second version and that the latter model outperforms the first version when stability of all patterns is required. The second version is thus better than the one with independent and identically distributed spins.

We first examine the sparse model with independent spins. We show the existence of sharp bounds on the capacity variables concerning the stability and error correction of a fixed stored pattern and prove afterwards that the lower bounds on the capacity variables are also valid for the second version of the model. The crucial difference between the two versions is that in the second one all the patterns can be stable with high probability without loosing the order of the capacity, whereas the first version cannot keep the same size of the number of stored patterns if the stability of all patterns (with high probability) is required.

Stability and Error Correction

A network should provide stability of the stored patterns: as in Chapter 2, it is either required for one arbitrary fixed pattern or for all patterns, depending on the notion of capacity. We are interested in perfect retrieval. As in Amari's model, we suggest a maximal number of $M = \alpha N^2 / \log(N)^2$ messages until the system looses any ability to recognise a stored pattern, where α has still to be determined. As we will see, this choice is justified and will lead to sharp bounds on α dependent on γ with the property that a randomly chosen message will be stable if α is smaller than this bound and that it will not be stable if α exceeds the bound, both with high probability. We state the following theorem:

Theorem 3.1 *In the Ternary simple model using $\gamma \in (0,1)$ to determine the threshold $\gamma \log(N)$, suppose that $M = \alpha N^2 / \log(N)^2$ patterns are stored, with α fulfilling*

$$\alpha < \frac{\gamma}{y_\gamma}, \tag{3.1}$$

where y_γ is the unique root of the function

$$g_\gamma : \mathbb{R}_+ \to \mathbb{R},\ g_\gamma(x) = -arsinh(x) + x^{-1}\left(cosh(arsinh(x)) - 1\right) + \frac{1}{\gamma}.$$

Then we have for each arbitrary but fixed of the stored messages ξ^μ:

$$\lim_{N\to\infty} \mathbb{P}\left(\forall i \leq N : T_i(\xi^\mu) = \xi_i^\mu\right) = 1. \tag{3.2}$$

This bound is sharp: for any

$$\alpha > \frac{\gamma}{y_\gamma},$$

the system looses the ability of recognising the stored patterns: for each arbitrary but fixed $\mu \in \{1, \ldots, M\}$, we have

$$\lim_{N\to\infty} \mathbb{P}\left(\exists i \leq N : T_i(\xi^\mu) \neq \xi_i^\mu\right) = 1. \tag{3.3}$$

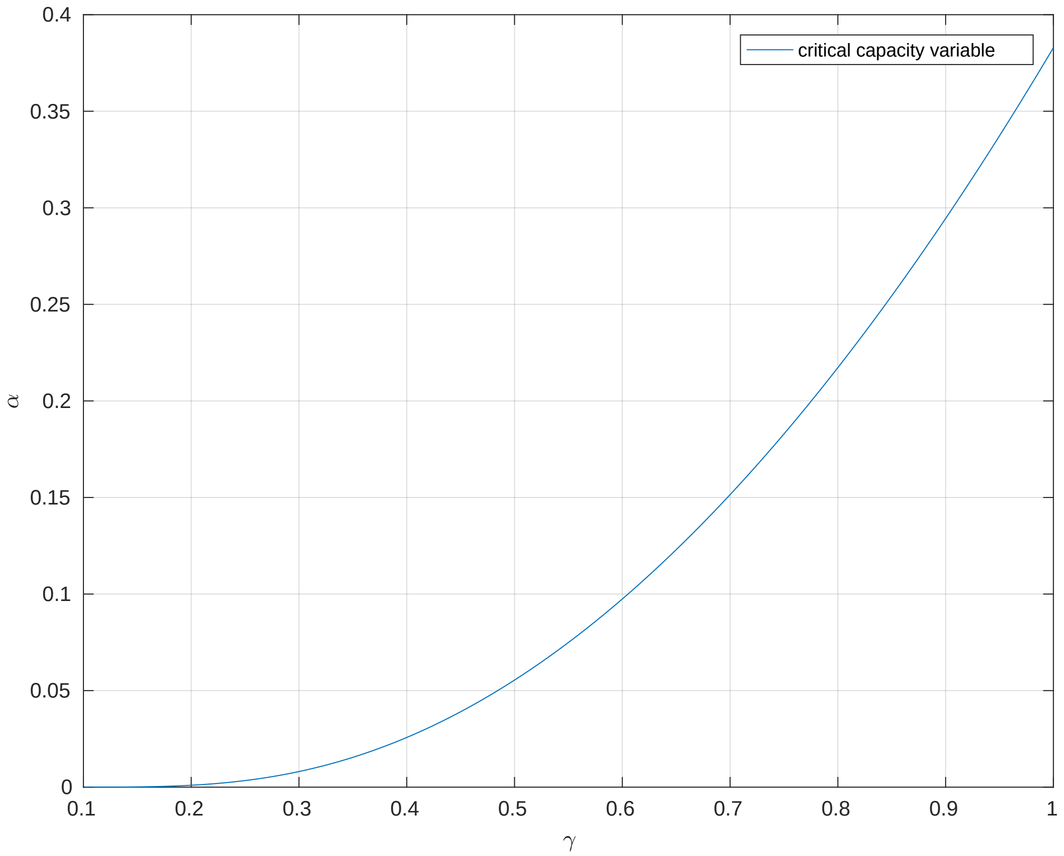

Figure 3.1: Critical capacity variable $\alpha^*(\gamma)$ in dependence on the threshold variable γ for the Ternary simple model

Proof of Theorem 3.1: Without loss of generality, we consider $\mu = 1$ and begin the proof by showing the first part of the theorem. We recall that for each $\delta > 0$ and as N tends to infinity,

$$\mathbb{P}\left(\left|\sum_{j=1}^{N} |\xi_j^1| - \log(N)\right| \geq \delta \log(N)\right) \longrightarrow 0.$$

For later use, the variable δ should be chosen such that $\delta < 1-\gamma$; there will be another condition that results during the proof. Let $\bar{\mathcal{Z}}_k := \{\sum_{j=1}^N |\xi_j^1| = k\}$ and $\mathcal{Z}_k := \{\sum_{j=1}^N |\xi_j^1| = \sum_{j=1}^k |\xi_j^1| = k\}$. Using the notation

$$A_\delta := \left\{ \Big| \sum_{j=1}^N |\xi_j^1| - \log(N) \Big| < \delta \log(N) \right\},$$

we estimate

$$\begin{aligned}
&\mathbb{P}\left(\exists i \le N : T_i(\xi^1) \neq \xi_i^1\right) \le \mathbb{P}(A_\delta^c) + \sum_{\substack{k\in\mathbb{N}: k/\log(N) \\ \in(1-\delta,1+\delta)}} \mathbb{P}\left(\bar{\mathcal{Z}}_k\right) \mathbb{P}\left(\exists i \le N : T_i(\xi^1) \neq \xi_i^1 | \mathcal{Z}_k\right) \\
\le& \mathbb{P}(A_\delta^c) + \mathbb{P}(A_\delta) \max_{\substack{k\in\mathbb{N}: k/\log(N) \\ \in(1-\delta,1+\delta)}} \mathbb{P}\left(\exists i \le N : T_i(\xi^1) \neq \xi_i^1 | \mathcal{Z}_k\right) \\
\le& \mathbb{P}(A_\delta^c) + \max_{\substack{k\in\mathbb{N}: k/\log(N) \\ \in(1-\delta,1+\delta)}} \mathbb{P}\left(\exists i \le N : T_i(\xi^1) \neq \xi_i^1 | \mathcal{Z}_k\right). \qquad (3.4)
\end{aligned}$$

We now examine $\mathbb{P}\left(T_i(\xi^1) \neq \xi_i^1 | \mathcal{Z}_k\right)$ for some arbitrary $i > k$ as well as for arbitrary $i \le k$, for fixed $k \in ((1-\delta)\log(N), (1+\delta)\log(N))$. First, let $i > k$. The probability can be written in the following way:

$$\begin{aligned}
&\mathbb{P}\left(T_i(\xi^1) \neq \xi_i^1 | \mathcal{Z}_k\right) = \mathbb{P}\left(T_i(\xi^1) \neq 0 | \mathcal{Z}_k\right) = \mathbb{P}\left(\left|S_i(\xi^1)\right| \ge \gamma \log(N) | \mathcal{Z}_k\right) \\
=& \mathbb{P}\left(\Big| \sum_{j\neq i} \sum_{\mu=1}^M \xi_j^1 \xi_i^\mu \xi_j^\mu \Big| \ge \gamma \log(N) \Big| \mathcal{Z}_k \right) = \mathbb{P}\left(\Big| \sum_{j=1}^k \sum_{\mu=2}^M \xi_j^1 \xi_i^\mu \xi_j^\mu \Big| \ge \gamma \log(N) \Big| \mathcal{Z}_k \right). \qquad (3.5)
\end{aligned}$$

Before we will continue, we state some facts that will play an important role during the proof.

First, for fixed i, the random variables $(\xi_j^1 \xi_i^\mu \xi_j^\mu,\ 2 \le \mu \le M,\ j \neq i)$ are conditionally independent, given $(\xi_i^\mu, \mu \ge 2,\ \xi_j^1, j \le N)$. To explain this, consider independent and identically distributed random variables Z_1, Z_2, Z_3 such that $\mathbb{P}(Z_1 = 1) = \mathbb{P}(Z_1 = -1) = \frac{1}{2}$. For each choice of $x_1, x_2, x_3 \in \{-1, 1\}$, the events $\{Z_1 Z_2 Z_3 = x_3\}$ and $\{Z_1 = x_1, Z_2 = x_2\}$ are independent. So, for fixed i, the conditional distribution of the random variables $\xi_j^1 \xi_i^\mu \xi_j^\mu$, $\mu \ge 2, j \neq i$, given $(\xi_i^\mu, \mu \ge 2,\ \xi_j^1, j \le N)$ is

$$\begin{aligned}
\mathbb{P}\left[\xi_{j'}^1 \xi_i^\nu \xi_{j'}^\nu = 0 \Big| \xi_i^\mu, \mu \ge 2, \xi_j^1, j \le N\right] &= \mathbb{P}\left[\xi_{j'}^1 \xi_i^\nu \xi_{j'}^\nu = 0 \Big| \xi_{j'}^1, \xi_i^\nu\right] = 1 - p\left|\xi_{j'}^1 \xi_i^\nu\right| \\
\mathbb{P}\left[\xi_{j'}^1 \xi_i^\nu \xi_{j'}^\nu = \pm 1 \Big| \xi_i^\mu, \mu \ge 2, \xi_j^1, j \le N\right] &= \mathbb{P}\left[\xi_{j'}^1 \xi_i^\nu \xi_{j'}^\nu = \pm 1 \Big| \xi_{j'}^1, \xi_i^\nu\right] = \left|\xi_{j'}^1 \xi_i^\nu\right| \cdot \frac{p}{2}.
\end{aligned}$$

This yields for arbitrary subsets $\{j_1, \ldots, j_{\tilde{r}}\} \subseteq \{1, \ldots, N\} \setminus \{i\}$ and $\{\mu_1, \ldots, \mu_{\tilde{s}}\} \subseteq \{2, \ldots, M\}$ and $x_{r,s} \in \{-1, 0, 1\}$, $r \le \tilde{r}$, $s \le \tilde{s}$:

$$\begin{aligned}
&\mathbb{P}\left(\bigcap_{r=1}^{\tilde{r}} \bigcap_{s=1}^{\tilde{s}} \{\xi_{j_r}^1 \xi_i^{\mu_s} \xi_{j_r}^{\mu_s} = x_{r,s}\} \Big| \xi_i^\mu, \mu \ge 2, \xi_j^1, j \le N \right) \\
=& \prod_{r,s=1}^{\tilde{r},\tilde{s}} \left[\mathbb{1}_{\{|x_{r,s}|=1\}} \frac{1}{2} \cdot p \cdot |\xi_{j_r}^1 \xi_i^{\mu_s}| + \mathbb{1}_{\{|x_{r,s}|=0\}} \left(1 - |\xi_{j_r}^1 \xi_i^{\mu_s}| \cdot p\right) \right]
\end{aligned}$$

$$= \prod_{r,s=1}^{\tilde r,\tilde s} \mathbb{P}\left(\xi^1_{j_r}\xi_i^{\mu_s}\xi_{j_r}^{\mu_s} = x_{r,s} \Big| \xi_i^\mu, \mu \geq 2, \xi_j^1, j \leq N \right).$$

Note that the $(\xi_j^1\xi_i^\mu\xi_j^\mu,\ 2 \leq \mu \leq M,\ j \neq i)$ are, for fixed i, also conditionally independent, given $(|\xi_i^\mu|, \mu \geq 2,\ |\xi_j^1|, j \leq N)$.

As next observation, the random variables

$$|\xi_i^\mu|, \quad \mu \geq 2,$$

are independent and identically Bernoulli distributed with parameter p. For each fixed i and $\delta > 0$, we conclude with the help of the exponential Chebyshev inequality and $t = \log(1+\delta) > 0$:

$$\begin{aligned}
&\mathbb{P}\left[\frac{\sum_{\mu=2}^M |\xi_i^\mu|}{\alpha N/\log(N)} \geq 1+\delta\right] \leq e^{-\alpha\frac{N}{\log(N)}t(1+\delta)}\mathbb{E}\left[\exp\left(t\sum_{\mu=2}^M |\xi_i^\mu|\right)\right] \\
=&e^{-\alpha\frac{N}{\log(N)}t(1+\delta)}\left[1-p+pe^t\right]^{M-1} \leq e^{-\alpha\frac{N}{\log(N)}t(1+\delta)}e^{Mp(e^t-1)} \\
=&e^{-\alpha\frac{N}{\log(N)}\left(t(1+\delta)-e^t+1\right)} = \exp\left[-\alpha\frac{N}{\log(N)}\left(\log(1+\delta)(1+\delta)-\delta\right)\right].
\end{aligned}$$

We used the independence of the $(\xi_i^\mu,\ \mu \geq 2)$ from the first to the second line, the inequality $1+x \leq e^x$, for all $x \in \mathbb{R}$ in the second line and the definition of M in the third. Analogously we obtain, with $t = -\log(1-\delta) > 0$

$$\begin{aligned}
&\mathbb{P}\left[\frac{\sum_{\mu=2}^M |\xi_i^\mu|}{\alpha N/\log(N)} \leq 1-\delta\right] \leq e^{\alpha\frac{N}{\log(N)}t(1-\delta)}\mathbb{E}\left[\exp\left(-t\sum_{\mu=2}^M |\xi_i^\mu|\right)\right] \\
\leq&e^{\alpha\frac{N}{\log(N)}t(1-\delta)}e^{(M-1)p(e^{-t}-1)} \leq \exp\left[-\alpha\frac{N}{\log(N)}\left((1-\delta)\log(1-\delta)+\delta\right)+p\delta\right].
\end{aligned}$$

Note that there is, in contrary to the first estimation, the (negligible) term $p\delta$, due to the fact that we consider $M-1$ and that $e^{-t}-1 < 0$ for $t > 0$. These two computations yield

$$\begin{aligned}
&\mathbb{P}\left[\frac{\sum_{\mu=2}^M |\xi_i^\mu|}{\alpha N/\log(N)} \notin (1-\delta, 1+\delta)\right] \\
\leq&2\exp\left[-\alpha\frac{N}{\log(N)}\cdot\min\Big((1+\delta)\log(1+\delta)-\delta; (1-\delta)\log(1-\delta)+\delta\Big)+p\delta\right]. \qquad (3.6)
\end{aligned}$$

Note that

$$\min\Big((1+\delta)\log(1+\delta)-\delta; (1-\delta)\log(1-\delta)+\delta\Big) > 0.$$

This bound will be needed later; the corresponding complementary event whose probability tends to 1 is called

$$B_\delta(i) := \left\{ \left| \sum_{\mu=2}^M |\xi_i^\mu| - \alpha\frac{N}{\log(N)} \right| < \delta\alpha\frac{N}{\log(N)} \right\}.$$

We want to examine the probability in (3.5) and intersect similarly to what we did in (3.4):

$$\begin{aligned}
&\mathbb{P}\left[\Big|\sum_{j=1}^{k}\sum_{\mu=2}^{M}\xi_j^1\xi_i^\mu\xi_j^\mu\Big| \geq \gamma\log(N)\Big|\mathcal{Z}_k\right] \leq \mathbb{P}(B_\delta(i)^c|\mathcal{Z}_k)+\\
&\sum_{\substack{l\in\mathbb{N}: l\log(N)/(\alpha N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left[\sum_{\mu=2}^{M}|\xi_i^\mu| = l\Big|\mathcal{Z}_k\right]\cdot\mathbb{P}\left[\Big|\sum_{j=1}^{k}\sum_{\mu=2}^{M}\xi_j^1\xi_i^\mu\xi_j^\mu\Big| \geq \gamma\log(N)\Big|\mathcal{Z}_k\cap\Big\{\sum_{\mu=2}^{M}|\xi_i^\mu| = l\Big\}\right]\\
\leq&\mathbb{P}(B_\delta(i)^c) + \max_{\substack{l\in\mathbb{N}: l\log(N)/(\alpha N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left[\Big|\sum_{j=1}^{k}\sum_{\mu=2}^{M}\xi_j^1\xi_i^\mu\xi_j^\mu\Big| \geq \gamma\log(N)\Big|\mathcal{Z}_k\cap\Big\{\sum_{\mu=2}^{M}|\xi_i^\mu| = l\Big\}\right].
\end{aligned}$$

In the last line, we used the fact that $\{\sum_{\mu=2}^{M}|\xi_i^\mu| = l\}$ and $\mathcal{Z}_k$ are independent.

Recall now that for fixed i, the $\xi_j^1\xi_i^\mu\xi_j^\mu$, $j\neq i$, $\mu\geq 2$, are conditionally independent, given $(\xi_i^\mu,\mu\geq 2,\xi_j^1,j\leq N)$. In addition, for fixed i, the conditional distribution of $\sum_{j\neq i}\sum_{\mu=2}^{M}\xi_j^1\xi_i^\mu\xi_j^\mu$, given $(\xi_i^\mu,\mu\geq 2,\xi_j^1,j\leq N)$, is completely determined by $\sum_{j\neq i}^{N}|\xi_j^1|$ and $\sum_{\mu=2}^{M}|\xi_i^\mu|$. In particular, if $\zeta_1,\zeta_2,\dots$ are independent and identically distributed with $\zeta_1\sim\xi_1^1$, we have, given $(\xi_i^\mu,\mu\geq 2,\xi_j^1,j\leq N)$,

$$\sum_{j\neq i}\sum_{\mu=2}^{M}\xi_j^1\xi_i^\mu\xi_j^\mu \sim \sum_{\substack{j\neq i:\\ \xi_j^1\neq 0}}\ \sum_{\mu>1:\xi_i^\mu\neq 0}\xi_j^\mu \sim \sum_{n=1}^{\sum_{j\neq i}^{N}|\xi_j^1|\cdot\sum_{\mu=2}^{M}|\xi_i^\mu|}\zeta_n,$$

where $\sim$ denotes identical distributions. This yields for $i>k$

$$\mathbb{P}\left[\Big|\sum_{j=1}^{k}\sum_{\mu=2}^{M}\xi_j^1\xi_i^\mu\xi_j^\mu\Big| \geq \gamma\log(N)\Big|\mathcal{Z}_k\cap\Big\{\sum_{\mu=2}^{M}|\xi_i^\mu| = l\Big\}\right] = \mathbb{P}\left[\Big|\sum_{j=1}^{k}\sum_{\mu=2}^{l+1}\xi_j^\mu\Big| \geq \gamma\log(N)\right]$$

and with the help of the exponential Chebyshev inequality, for $t>0$

$$\begin{aligned}
&\mathbb{P}\left[\Big|\sum_{j=1}^{k}\sum_{\mu=2}^{l+1}\xi_j^\mu\Big| \geq \gamma\log(N)\right] \leq 2e^{-\gamma\log(N)t}\quad \mathbb{E}\left[\exp\left(t\sum_{j=1}^{k}\sum_{\mu=2}^{l+1}\xi_j^\mu\right)\right]\\
=&2e^{-\gamma\log(N)t}\quad\mathbb{E}\left[\exp\left(t\xi_j^\mu\right)\right]^{lk} = 2e^{-\gamma\log(N)t}\left[1-p+\frac{1}{2}pe^t+\frac{1}{2}pe^{-t}\right]^{lk}\\
=&2e^{-\gamma\log(N)t}\left[1-p+p\cosh(t)\right]^{lk} = 2e^{-\gamma\log(N)t}\left[1+p(\cosh(t)-1)\right]^{lk}.
\end{aligned}\tag{3.7}$$

The function $\cosh(t)-1$ is positive on $\mathbb{R}_{\backslash 0}$. Using that $k\leq(1+\delta)\log(N)$, $l\leq(1+\delta)\alpha N/\log(N)$, the last line is therefore bounded by

$$2e^{-\gamma\log(N)t}\left[1+p(\cosh(t)-1)\right]^{\alpha N(1+\delta)^2}.$$

Combining the results (3.4) to (3.7) yields

$$\begin{aligned}
&\mathbb{P}(\exists i \le N : \xi_i^1 = 0, T_i(\xi^1) \neq \xi_i^1) \\
\le&\mathbb{P}(A_\delta^c) + N\left[\mathbb{P}(B_\delta(N)^c) + 2e^{-\gamma \log(N) t}\left[1 + p(\cosh(t) - 1)\right]^{\alpha N (1+\delta)^2}\right] \\
\le&\mathbb{P}(A_\delta^c) + N\left[\mathbb{P}(B_\delta(N)^c) + 2e^{-\gamma \log(N) t}\ \exp\left[\log(N)\alpha(1+\delta)^2(\cosh(t) - 1)\right]\right]. \qquad (3.8)
\end{aligned}$$

In the last line we again used the inequality $1 + x \le e^x$ for $x \in \mathbb{R}$. Due to the fact that $\mathbb{P}(A_\delta^c)$ vanishes as N tends to infinity and to the estimation in (3.6), the probability in (3.8) tends to 0 if

$$N 2e^{-\gamma \log(N) t}\ \exp\left[\log(N)\alpha(1+\delta)^2(\cosh(t) - 1)\right]$$

does. Let the function $f_{\delta,\alpha,\gamma}(t)$ be defined by

$$f_{\delta,\alpha,\gamma}(t) := -\gamma t + \alpha(1+\delta)^2(\cosh(t) - 1);$$

by inserting

$$t^*_{\delta,\alpha,\gamma} := \operatorname{arsinh}\left(\frac{\gamma}{(1+\delta)^2\alpha}\right),$$

which is positive for our choice of $\delta, \gamma, \alpha > 0$, the last line of (3.8) converges to 0 if

$$f_{\delta,\alpha,\gamma}(t^*_{\delta,\alpha,\gamma}) < -1.$$

To reformulate this condition in terms of the function g_γ defined in the theorem,

$$g_\gamma(x) = -\operatorname{arsinh}(x) + x^{-1}\left(\cosh(\operatorname{arsinh}(x)) - 1\right) + \frac{1}{\gamma},$$

we define

$$x_{\delta,\alpha,\gamma} := \frac{\gamma}{(1+\delta)^2\alpha}.$$

We can write

$$f_{\delta,\alpha,\gamma}(t^*_{\delta,\alpha,\gamma}) = f_{\delta,\alpha,\gamma}(\operatorname{arsinh}(x_{\delta,\alpha,\gamma})) = \gamma g_\gamma(x_{\delta,\alpha,\gamma}) - 1.$$

A sufficient constraint for the convergence to 0 of the probability in (3.8) is

$$g_\gamma(x_{\delta,\alpha,\gamma}) < 0. \qquad (3.9)$$

We will come back to this later and continue with the study of the behaviour of those neurons that are active in ξ^1. We start as in the last part, that is, with (3.4), and consider now $i \le k$. We thus analyse

$$\max_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left(\exists i \le k : T_i(\xi^1) \neq \xi_i^1 | \mathcal{Z}_k\right).$$

The local field for $i \le k$ is, conditional on $\mathcal{Z}_k$, equal to

$$S_i(\xi^1) = \sum_{\mu=1}^{M}\sum_{j\neq i}\xi_j^1\xi_i^\mu\xi_j^\mu = \xi_i^1\sum_{j\neq i}|\xi_j^1| + \sum_{\mu=2}^{M}\sum_{j\neq i}\xi_j^1\xi_i^\mu\xi_j^\mu = \xi_i^1(k-1) + \sum_{\mu=2}^{M}\sum_{j\neq i}^{k}\xi_j^1\xi_i^\mu\xi_j^\mu.$$

The variable δ was chosen such that $\delta < 1-\gamma$; the difference $1-\gamma-\delta$ is therefore positive. To obtain the stability of ξ_i^1, the sum $\sum_{\mu=2}^{M}\sum_{j\neq i}\xi_j^1\xi_i^\mu\xi_j^\mu$ must not become too large. In fact, a necessary condition for the instability of neuron $i \leq k$ is

$$\left|\sum_{\mu=2}^{M}\sum_{j\neq i}^{k}\xi_j^1\xi_i^\mu\xi_j^\mu\right| > k - \gamma\log(N) - 1,$$

and for $k > (1-\delta)\log(N)$, this event is contained in

$$\left|\sum_{\mu=2}^{M}\sum_{j\neq i}^{k}\xi_j^1\xi_i^\mu\xi_j^\mu\right| \geq (1-\gamma-\delta)\log(N) - 1.$$

In analogy to the computations for $i > k$, we estimate

$$\begin{aligned}&\max_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left(\exists i \leq k : T_i(\xi^1) \neq \xi_i^1 | \mathcal{Z}_k\right)\\ &\leq \max_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} k\left[\mathbb{P}(B_\delta(1)^c) + \max_{\substack{l\in\mathbb{N}:\log(N)/(\alpha N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left[\left|\sum_{j\neq i}^{k}\sum_{\mu=2}^{l+1}\xi_j^\mu\right| \geq (1-\gamma-\delta)\log(N)-1\right]\right].\end{aligned}$$

To achieve with high probability stability of the active neurons,

$$\max_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} k \max_{\substack{l\in\mathbb{N}:l\log(N)/(\alpha N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left[\left|\sum_{j\neq i}^{k}\sum_{\mu=2}^{l+1}\xi_j^\mu\right| \geq (1-\gamma-\delta)\log(N)-1\right]$$

must converge to 0. Again using an exponential Chebyshev inequality, this is bounded by

$$\begin{aligned}&\max_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} k \max_{\substack{l\in\mathbb{N}:l\log(N)/(\alpha N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left[\left|\sum_{j\neq i}^{k}\sum_{\mu=2}^{l+1}\xi_j^\mu\right| \geq (1-\gamma-\delta)\log(N)-1\right]\\ \leq &\max_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} k \max_{\substack{l\in\mathbb{N}:l\log(N)/(\alpha N)\\ \in(1-\delta,1+\delta)}} 2\exp\left[-\log(N)(1-\gamma-\delta)s + s\right]\exp\left[plk(\cosh(s)-1)\right]\end{aligned}$$

for $s > 0$. Since $\cosh(s)-1$ is positive for each $s > 0$, the expression in the last line is bounded by inserting $k = (1+\delta)\log(N)$ and $l = \frac{\alpha N}{\log(N)}(1+\delta)$. With $f_{\delta,\alpha,1-\gamma-\delta}(s) = -(1-\gamma-\delta)s + (1+\delta)^2\alpha(\cosh(s)-1)$ and

$$s^*_{\delta,\alpha,\gamma} := \operatorname{arsinh}\left(\frac{1-\gamma-\delta}{(1+\delta)^2\alpha}\right)$$

– note that $s^*_{\delta,\alpha,\gamma} > 0$ for each choice of δ, α with $1-\gamma > \delta > 0$, $\delta, \alpha > 0$ – the last line of the previous estimation is at most

$$\begin{aligned}&(2+2\delta)\log(N)\exp\left[s^*_{\delta,\alpha,\gamma} + \log(N)\left(-(1-\gamma-\delta)s^*_{\delta,\alpha,\gamma} + (1+\delta)^2\alpha(\cosh(s^*_{\delta,\alpha,\gamma})-1)\right)\right]\\ =&(2+2\delta)\log(N)\exp\left[s^*_{\delta,\alpha,\gamma} + \log(N)f_{\delta,\alpha,1-\gamma-\delta}(s^*_{\delta,\alpha,\gamma})\right]. \qquad (3.10)\end{aligned}$$

We define the continuous function $g : \mathbb{R} \to \mathbb{R}$,

$$g(x) := -x\text{arsinh}(x) + \cosh(\text{arsinh}(x)) - 1. \tag{3.11}$$

We observe $g(0) = 0$ and $g'(x) = -\text{arsinh}(x)$, so g is strictly decreasing on $\mathbb{R}_+$, strictly increasing on $\mathbb{R}_-$ and negative on $\mathbb{R}_{\setminus 0}$. This function and its properties will play an important role in this chapter, as first property, its negativity on $\mathbb{R}_+$. Recall that $x_{\delta,\alpha,1-\gamma-\delta} = \frac{1-\gamma-\delta}{(1+\delta)^2\alpha}$. The last line of (3.10) converges to 0 if

$$\begin{aligned} f_{\delta,\alpha,1-\gamma-\delta}((s^*_{\delta,\alpha,\gamma})) =& f_{\delta,\alpha,1-\gamma-\delta}(\text{arsinh}(x_{\delta,\alpha,1-\gamma-\delta})) \\ =& (1+\delta)^2\alpha \cdot g(x_{\delta,\alpha,1-\gamma-\delta}) < 0. \end{aligned}$$

Due to the negativity of g on $\mathbb{R}_+$, this condition is fulfilled for each choice of $\delta > 0$ such that $1 - \gamma > \delta$. The stability of the active neurons of ξ^1 will thus simply restrict the value of γ to the interval $(0, 1)$. This is because $\delta > 0$ can be chosen arbitrarily small in the proof of the theorem.

The remaining condition to examine is the one formulated in (3.9). Recall that $g_\gamma = -\text{arsinh}(x) + x^{-1}(\cosh(\text{arsinh}(x)) - 1) + \frac{1}{\gamma}$. We first observe that on $\mathbb{R}_+$, $g_\gamma(x) = \frac{1}{x}g(x) + \frac{1}{\gamma}$ and thus

$$g'_\gamma(x) = -\frac{1}{x^2}g(x) + \frac{1}{x}g'(x) = \frac{x\text{arsinh}(x) - \sqrt{1+x^2} + 1}{x^2} - \frac{\text{arsinh}(x)}{x} = x^{-2}(1 - \sqrt{x^2+1})$$

which is negative on $\mathbb{R}_+$. So g_γ is strictly decreasing on $\mathbb{R}_+$. In addition, $\lim_{x\searrow 0} g_\gamma(x) = \frac{1}{\gamma}$ and $\lim_{x\to\infty} g_\gamma(x) = -\infty$. The function has consequently a unique root $y_\gamma \in \mathbb{R}_+$ and is positive on $(0, y_\gamma)$ and negative on (y_γ, ∞).

If the capacity variable α is chosen such that

$$\alpha < \frac{\gamma}{y_\gamma},$$

there is a $\delta' > 0$ that satisfies

$$\frac{\gamma}{\alpha(1+\delta')^2} > y_\gamma.$$

For all $\delta > 0$, $\delta \le \delta'$, inequality (3.9) holds. This guarantees stability with high probability if we store at most $M = \alpha N^2/\log(N)^2$ patterns.

We just proved the first part of the theorem. To show the reverse bound, we consider again ξ^1 and recall that the probability of the event $A_\delta = \{\sum_{j=1}^N |\xi^1_j|/\log(N) \in (1-\delta, 1+\delta)\}$ tends to 1 for each $\delta > 0$ as N tends to infinity. So we will restrict our considerations to this event for some fixed $\delta > 0$ (the conditions on δ will result from the proof). By showing

$$\lim_{N\to\infty} \min_{\substack{k\in\mathbb{N}:k/\log(N) \\ \in(1-\delta,1+\delta)}} \mathbb{P}\left(\exists i > k : T_i(\xi^1) \neq 0 | \mathcal{Z}_k\right) = 1,$$

we obtain the assertion of the theorem.

As our next step, we assert that the random variables

$$X_1(k) = \sum_{\mu=2}^M \mathbb{1}_{\{\sum_{j\le k}|\xi^\mu_j|=1\}}, \quad X_2(k) = \sum_{\mu=2}^M \mathbb{1}_{\{\sum_{j\le k}|\xi^\mu_j|=2\}} \quad \text{and} \quad X_3(k) = \sum_{\mu=2}^M \mathbb{1}_{\{\sum_{j\le k}|\xi^\mu_j|>2\}}$$

are Binomially distributed with parameters $M-1$ and

$$p_1(k) = kp(1-p)^{k-1}, \quad p_2(k) = \binom{k}{2} p^2 (1-p)^{k-2}$$

and

$$\begin{aligned} p_3(k) &= 1-(1-p)^k - kp(1-p)^{k-1} - \binom{k}{2}p^2(1-p)^{k-2} \\ &= 1 - \left[1 - kp + \binom{k}{2}p^2 - \binom{k}{3}p^3 + \mathcal{O}(k^4p^4)\right] \\ &\quad - kp\left[1-(k-1)p + \binom{k-1}{2}p^2 + \mathcal{O}(k^3p^3)\right] - \binom{k}{2}p^2\left[1-(k-2)p + \mathcal{O}(k^2p^2)\right] \\ &= \binom{k}{3}p^3 + \mathcal{O}(k^4p^4), \end{aligned}$$

respectively. Using the Chebyshev inequality and $\mathbb{E}[X_1(k)] = (1-p)^{k-1}(Mpk - pk)$ as well as $\mathbb{V}[X_1(k)] = (M-1)kp(1-p)^{k-1}[1-kp(1-p)^{k-1}] \leq Mkp$, we know for any $\delta > 0$ and $k \leq (1+\delta)\log(N)$:

$$\begin{aligned} \mathbb{P}\left(\frac{X_1(k)}{\alpha k \frac{N}{\log(N)}} \notin (1-\delta, 1+\delta)\right) &\leq \frac{\mathbb{V}(X_1(k))}{[\delta\alpha k N/\log(N) - Mk(k-1)p^2 + \mathcal{O}(pk^3)]^2} \\ &\leq \frac{Mkp}{M^2k^2p^2[\delta - (k-1)p + \mathcal{O}(p^2k^2)]^2}. \end{aligned}$$

Note that the term $Mk(k-1)p^2 + \mathcal{O}(pk^3)$ in the denominator is due to the fact that the expectation is not exactly $Mkp = \alpha kN/\log(N)$. We obtain analogously, using $\mathbb{E}[X_2(k)] = \alpha\binom{k}{2}(1-p)^{k-2} + \mathcal{O}(p^2)$ and $\mathbb{V}[X_2(k)] \leq \alpha\binom{k}{2}$,

$$\mathbb{P}\left(\frac{X_2(k)}{\alpha\binom{k}{2}} \notin (1-\delta, 1+\delta)\right) \leq \frac{\alpha\binom{k}{2}}{[\delta\alpha\binom{k}{2} - \alpha\binom{k}{2}(k-2)p + \mathcal{O}(k^4p^2)]^2}.$$

Both probablities tend to 0 as N tends to infinity. Finally, $X_3(k)$ vanishes with probability converging to 1:

$$\mathbb{P}(X_3(k) \neq 0) \leq M\mathbb{P}\left(\sum_{j\leq k} |\xi_j^M| > 2\right) \leq M\binom{k}{3}p^3 + \mathcal{O}(Mk^4p^4) \longrightarrow 0$$

as N tends to infinity. We denote the sets

$$\left\{\frac{X_1(k)}{\alpha k\frac{N}{\log(N)}} \in (1-\delta, 1+\delta)\right\}, \quad \left\{\frac{X_2(k)}{\alpha\binom{k}{2}} \in (1-\delta,1+\delta)\right\}, \quad \{X_3(k) = 0\}$$

by $B_\delta(k)$, $C_\delta(k)$ and $D(k)$, respectively.

We observe by similar considerations as on page 54 that for $i \neq j$, $\mu \geq 2$

$$\mathbb{P}\left[\xi_j^1\xi_i^\mu\xi_j^\mu = 0 \middle| \xi_j^1\right] = 1 - p^2|\xi_j^1|, \quad \mathbb{P}\left[\xi_j^1\xi_i^\mu\xi_j^\mu = \pm 1 \middle| \xi_j^1\right] = \frac{1}{2}p^2|\xi_j^1|$$

and that conditionally on $\mathcal{Z}_k$,

$$(\xi_j^1\xi_i^\mu\xi_j^\mu, \mu \geq 2, j \leq k, i > k) \sim (\xi_i^\mu\xi_j^\mu, \mu \geq 2, j \leq k, i > k),$$

where $\sim$ denotes identical distributions. In particular, conditional on $\mathcal{Z}_k$,

$$\sum_{\mu=2}^{M}\sum_{j=1}^{N}\xi_j^1\xi_i^\mu\xi_j^\mu \sim \sum_{\mu=2}^{M}\sum_{j=1}^{k}\xi_i^\mu\xi_j^\mu.$$

This implies

$$\begin{aligned}
\mathbb{P}\left(\exists i > k : T_i(\xi^1) \neq \xi_i^1|\mathcal{Z}_k\right) =& 1 - \mathbb{P}\left(\forall i > k : T_i(\xi^1) = \xi_i^1|\mathcal{Z}_k\right) \\
=& 1 - \mathbb{P}\left(\forall i > k : |S_i(\xi^1)| < \gamma\log(N)|\mathcal{Z}_k\right) \\
=& 1 - \mathbb{P}\left[\forall i > k : \Big|\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_j^1\xi_i^\mu\xi_j^\mu\Big| < \gamma\log(N)\Big|\mathcal{Z}_k\right] \\
=& 1 - \mathbb{P}\left[\forall i > k : \Big|\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_i^\mu\xi_j^\mu\Big| < \gamma\log(N)\right].
\end{aligned}$$

The sums $\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_i^\mu\xi_j^\mu$, $i > k$, are conditionally independent, given $(\xi_j^\mu, j \leq k, \mu \geq 2)$. This explains the transformation from the second to the third line in the subsequent computation. Defining $\mathcal{F}_k^N := \sigma(\xi_j^\mu, j \leq k, \mu \geq 2)$, we estimate:

$$\begin{aligned}
&\mathbb{P}\left[\forall i > k : \Big|\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_i^\mu\xi_j^\mu\Big| < \gamma\log(N)\right] \\
=&\mathbb{E}_{(\xi_j^\mu, j\leq k, \mu\geq 2)}\left[\mathbb{P}\left[\forall i > k : \Big|\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_i^\mu\xi_j^\mu\Big| < \gamma\log(N)\Big|\mathcal{F}_k^N\right]\right] \\
=&\mathbb{E}_{(\xi_j^\mu, j\leq k, \mu\geq 2)}\left[\mathbb{P}\left[\Big|\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_{k+1}^\mu\xi_j^\mu\Big| < \gamma\log(N)\Big|\mathcal{F}_k^N\right]^{N-k}\right] \\
\leq& \max_{B_\delta(k)\cap C_\delta(k)\cap D(k)}\mathbb{P}\left[\Big|\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_{k+1}^\mu\xi_j^\mu\Big| < \gamma\log(N)\Big|\mathcal{F}_k^N\right]^{N-k} + \mathbb{P}((B_\delta(k)\cap C_\delta(k)\cap D(k))^c).
\end{aligned} \tag{3.12}$$

Since $\mathbb{P}((B_\delta(k)\cap C_\delta(k)\cap D(k))^c)$ vanishes for the k of the set of interest, it suffices to consider $B_\delta(k)\cap C_\delta(k)\cap D(k)$. It remains to show

$$\lim_{N\to\infty}\max_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}}\max_{B_\delta(k)\cap C_\delta(k)\cap D(k)}\mathbb{P}\left[\Big|\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_{k+1}^\mu\xi_j^\mu\Big| < \gamma\log(N)\Big|\mathcal{F}_k^N\right]^{N-k} = 0. \tag{3.13}$$

We split the sum $\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_{k+1}^\mu\xi_j^\mu$ into

$$\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_{k+1}^\mu\xi_j^\mu = \sum_{\mu>1:\sum_{j\leq k}|\xi_j^\mu|=0}\;\sum_{j\leq k}\xi_{k+1}^\mu\xi_j^\mu + \sum_{\mu>1:\sum_{j\leq k}|\xi_j^\mu|=1}\;\sum_{j\leq k}\xi_{k+1}^\mu\xi_j^\mu$$

$$+ \sum_{\mu>1:\sum_{j\le k}|\xi_j^\mu|=2} \sum_{j\le k} \xi_{k+1}^\mu \xi_j^\mu + \sum_{\mu>1:\sum_{j\le k}|\xi_j^\mu|>2} \sum_{j\le k} \xi_{k+1}^\mu \xi_j^\mu$$

$$= \sum_{\mu>1:\sum_{j\le k}|\xi_j^\mu|=1} \sum_{j\le k} \xi_{k+1}^\mu \xi_j^\mu + \sum_{\mu>1:\sum_{j\le k}|\xi_j^\mu|=2} \sum_{j\le k} \xi_{k+1}^\mu \xi_j^\mu + \sum_{\mu>1:\sum_{j\le k}|\xi_j^\mu|>2} \sum_{j\le k} \xi_{k+1}^\mu \xi_j^\mu \quad (3.14)$$

and analyse the three remaining parts.

The third sum is zero on the set $D(k)$. We continue with the second one. On $C_\delta(k)$, the following conditional probability of a contribution of the second sum to the total sum is bounded by

$$\begin{aligned} &\mathbb{P}\left(\sum_{\mu>1:\sum_{j\le k}|\xi_j^\mu|=2} \sum_{j\le k} \xi_{k+1}^\mu \xi_j^\mu \neq 0 \middle| \mathcal{F}_k^N \right) \\ \le &\mathbb{P}\left(\exists \mu > 1 : \sum_{j\le k} |\xi_j^\mu| = 2, \xi_{k+1}^\mu \neq 0 \middle| \mathcal{F}_k^N \right) \\ \le &X_2(k) \cdot \mathbb{P}(\xi_{k+1}^2 \neq 0) = X_2(k) \cdot p \le (1+\delta)\alpha \binom{k}{2} \cdot p. \end{aligned} \quad (3.15)$$

It remains to examine the first summand in (3.14),

$$\sum_{\mu>1:\sum_{j\le k}|\xi_j^\mu|=1} \sum_{j\le k} \xi_{k+1}^\mu \xi_j^\mu. \quad (3.16)$$

As explained in the next paragraph, this is conditionally on $\mathcal{F}_k^N$ distributed as a random walk with random length determined by a Binomial random variable with parameters $X_1(k)$ and p.

Assume that $X_1(k) = x_1$ and that $\sum_{j\le k} |\xi_j^\mu| = 1$ for $\mu \in \{\mu_1, \dots, \mu_{x_1}\}$. Then

$$\xi_{k+1}^{\mu_r} \sum_{j\le k} \xi_j^{\mu_r} = \xi_{k+1}^{\mu_r} \xi_{j_r}^{\mu_r},$$

$1 \le r \le x_1$. The index j_r is here chosen such that j_r is the one and only element of $\{1, \dots, k\}$ with $\xi_{j_r}^{\mu_r} \neq 0$.

A direct consequence of the observation at the beginning of the proof (p. 54) concerning the random variables Z_1 and Z_2 is that for each $r \le x_1$,

$$\mathbb{P}\left(\xi_{k+1}^{\mu_r} \xi_{j_r}^{\mu_r} = \pm 1 \middle| \xi_{j_r}^{\mu_r} \right) = \frac{1}{2} p |\xi_{j_r}^{\mu_r}|, \quad \mathbb{P}\left(\xi_{k+1}^{\mu_r} \xi_{j_r}^{\mu_r} = 0 \middle| \xi_{j_r}^{\mu_r} \right) = 1 - p |\xi_{j_r}^{\mu_r}|.$$

This yields

$$\sum_{\mu:\sum_{j\le k}|\xi_j^\mu|=1} \sum_{j\le k} \xi_{k+1}^\mu \xi_j^\mu \sim \sum_{\mu:\sum_{j\le k}|\xi_j^\mu|=1} \xi_{k+1}^\mu \sim \sum_{\mu=2}^{X_1(k)+1} \xi_{k+1}^\mu \quad (3.17)$$

and conditional on $(\xi_j^\mu, \mu \ge 2, j \le k)$, this is a sum of $X_1(k)$ independent and identically distributed random variables, each one distributed as ξ_{k+1}^2. Given $X_1(k)$, this is finally a random walk of random length, the length independent of the steps and the distribution of the length given by a Binomially distributed random variable with parameters p and $X_1(k)$.

As next step, we observe that the length-determining distribution can be approximated by a Poisson distribution with parameter $X_1(k)p$. For $x_1 \in \{1\ldots, M-1\}$, the total variation distance of a Poisson(x_1p) random variable whose probability weights are denoted by $\pi_{x_1p}(m)$, $m \in \mathbb{N}$, and the Binomially distributed random variable $\sum_{\mu=2}^{x_1+1} |\xi_{k+1}^{\mu}|$ is at most

$$\sum_{m=0}^{\infty} \left| \mathbb{P}\left(\sum_{\mu=2}^{x_1+1} |\xi_{k+1}^{\mu}| = m \right) - \pi_{x_1p}(m) \right| \leq 2p^2 x_1$$

(see Lemma 2.4). Let, for $x_1 \in \mathbb{N}$, R_{x_1} denote a Binomially distributed random variable with parameters x_1 and p, Y_{x_1p} denote a Poisson distributed random variable with parameter x_1p and let Z_n, $n \in \mathbb{N}$, be identically distributed random variables such that

$$\mathbb{P}(Z_n = 1) = \mathbb{P}(Z_n = -1) = \frac{1}{2}$$

and that

$$R_{x_1}, Y_{x_1p}, Z_1, Z_2, \ldots$$

are independent.

We continue with the analysis of (3.17): with the objective to examine a random walk whose length is given by a Poisson random variable (instead of a Binomial one), we observe that

$$\begin{aligned}
\mathbb{P}\left[\sum_{n=1}^{R_{x_1}} Z_n \geq \gamma \log(N)\right] &= \sum_{m=0}^{\infty} \mathbb{P}(R_{x_1} = m)\mathbb{P}\left[\sum_{n=1}^{m} Z_n \geq \gamma \log(N)\right] \\
&= \sum_{m=0}^{\infty} \left[\mathbb{P}\left[R_{x_1} = m\right] - \mathbb{P}\left[Y_{x_1p} = m\right]\right] \mathbb{P}\left[\sum_{n=1}^{m} Z_n \geq \gamma \log(N)\right] + \\
&\qquad \sum_{m=0}^{\infty} \mathbb{P}\left[Y_{x_1p} = m\right] \mathbb{P}\left[\sum_{n=1}^{m} Z_n \geq \gamma \log(N)\right] \\
&\geq -\sum_{m=0}^{\infty} \Big|\mathbb{P}\left[R_{x_1} = m\right] - \mathbb{P}\left[Y_{x_1p} = m\right]\Big| + \sum_{m=0}^{\infty} \mathbb{P}\left[Y_{x_1p} = m\right] \mathbb{P}\left[\sum_{n=1}^{m} Z_n \geq \gamma \log(N)\right] \\
&\geq -2p^2 x_1 + \mathbb{P}\left[\sum_{n=1}^{Y_{x_1p}} Z_n \geq \gamma \log(N)\right]. \qquad (3.18)
\end{aligned}$$

Our last analysis of this theorem is thus directed at a random walk of random length; this length is by now determined by a Poisson distributed random variable. This Poisson random variable, whose parameter will be about $\log(N)$, is distributed as the sum of independent Poisson random variables. Now the important observation is that it does not matter in distribution if we run one random walk of a length determined by a Poisson random variable with parameter $w > 0$, with $w = \sum_{r=1}^{s} w_r$, $w_1, \ldots, w_r > 0$, or if we run

s independent random walks where the length of the rth random walk is determined by a Poisson(w_r) random variable, and consider their sum. Formally, if $(Y_w, Z_n, n \in \mathbb{N})$ as well as $(Y_{w_1}, \ldots, Y_{w_s}, Z_n \in \mathbb{N})$ and $(Y_{w_1}, \ldots, Y_{w_s}, Z_{n,r}, 1 \leq r \leq s, n \in \mathbb{N})$ are in each case independent and $Z_{n,r} \sim Z_1$ for each $r \geq 1$, $n \geq 1$, this is

$$\sum_{n=1}^{Y_w} Z_n \sim \sum_{n=1}^{Y_{w_1}+\ldots+Y_{w_s}} Z_n \sim \sum_{r=1}^{s} \sum_{n=1}^{Y_{w_r}} Z_{n,r} =: \sum_{r=1}^{s} W_r.$$

$W_1, \ldots, W_s$ are then independent and W_r is a random walk with random length given by a Poisson(w_r) random variable (such that $Y_{w_r}, Z_{n,r}, n \geq 1$ are independent).

The next step is to see that on $B_\delta(k)$, the equality

$$\begin{aligned}
&\min_{B_\delta(k)} \quad \mathbb{P}\left(\left| \sum_{\mu: \sum_{j \leq k} |\xi_j^\mu| = 1} \sum_{j \leq k} \xi_{k+1}^\mu \xi_j^\mu \right| \geq \gamma \log(N) \middle| \mathcal{F}_k^N \right) \\
&= \min_{B_\delta(k)} \quad \mathbb{P}\left(\left| \sum_{\mu=2}^{X_1(k)+1} \xi_{k+1}^\mu \right| \geq \gamma \log(N) \middle| \mathcal{F}_k^N \right) \\
&= \min_{\substack{x_1 \in \mathbb{N}: x_1 \log(N)/(\alpha k N) \\ \in (1-\delta, 1+\delta)}} \quad \mathbb{P}\left(\left| \sum_{n=1}^{R_{x_1}} Z_n \right| \geq \gamma \log(N) \right)
\end{aligned} \tag{3.19}$$

holds, where $R_{x_1}, Z_1, \ldots$ are independent and distributed as described above. Combined with (3.18), we obtain

$$\begin{aligned}
&\min_{\substack{x_1 \in \mathbb{N}: x_1 \log(N)/(\alpha k N) \\ \in (1-\delta, 1+\delta)}} \quad \mathbb{P}\left(\left| \sum_{n=1}^{R_{x_1}} Z_n \right| \geq \gamma \log(N) \right) \\
&\geq \min_{\substack{x_1 \in \mathbb{N}: x_1 \log(N)/(\alpha k N) \\ \in (1-\delta, 1+\delta)}} \quad \mathbb{P}\left(\sum_{n=1}^{R_{x_1}} Z_n \geq \gamma \log(N) \right) \\
&\geq \min_{\substack{x_1 \in \mathbb{N}: x_1 \log(N)/(\alpha k N) \\ \in (1-\delta, 1+\delta)}} \mathbb{P}\left(\sum_{n=1}^{Y_{p x_1}} Z_n \geq \gamma \log(N) \right) - 2p^2(1+\delta)\alpha \frac{Nk}{\log(N)} \\
&\geq \min_{\rho \in (1-\delta, 1+\delta)} \mathbb{P}\left(\sum_{n=1}^{Y_{k\rho\alpha}} Z_n \geq \gamma \log(N) \right) - 2p(1+\delta)\alpha k.
\end{aligned} \tag{3.20}$$

Using the notation of the previous considerations, $Y_{k\rho\alpha}$ is distributed as the sum of k independent Poisson random variables with parameter $\rho\alpha$, each, i.e. $s = k$ and $w_r = \rho\alpha$, $r \in \{1, \ldots, k\}$.

These random walks with random length follow as independent and identically distributed random variables with, as we will see, finite moment generating function, a large deviation principle by the application of Cramér's theorem (see Lemma 2.4). The logarithmic moment generating function of a random walk with Poisson length of parameter

λ is equal to

$$\begin{aligned}\Lambda_\lambda(t) &= \log\left[\mathbb{E}\left(\exp t\sum_{n=1}^{Y_\lambda} Z_n\right)\right] = \log\left[\sum_{m=0}^{\infty}\mathbb{P}(Y_\lambda = m)\mathbb{E}\left(\exp\left(t\sum_{n=1}^{m} Z_n\right)\right)\right]\\ &= \log\left[\sum_{m=0}^{\infty}\mathbb{P}(Y_\lambda = m)\left(\mathbb{E}\left(\exp tZ_1\right)\right)^m\right] = \log\left[\sum_{m=0}^{\infty}\mathbb{P}(Y_\lambda = m)\left(\frac{1}{2}e^t + \frac{1}{2}e^{-t}\right)^m\right]\\ &= \log\left[e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}\left(\frac{1}{2}e^t + \frac{1}{2}e^{-t}\right)^m\right] = \log\left[e^{-\lambda+\lambda\cosh(t)}\right] = \lambda(\cosh(t)-1).\end{aligned}$$

The Legendre transform of Λ_λ is defined as

$$\Lambda_\lambda^*(x) = \sup_{t\in\mathbb{R}}(tx - \Lambda_\lambda(t)).$$

We determine

$$\frac{d}{dt}\quad(tx - \lambda(\cosh(t)-1)) = x - \lambda\cdot\sinh(t).$$

Since

$$\frac{d^2}{dt^2}\quad(tx - \lambda(\cosh(t)-1)) = -\lambda\cdot\cosh(t)$$

is negative on $\mathbb{R}$,

$$t = \operatorname{arsinh}\left(\frac{x}{\lambda}\right)$$

is the global maximum of $tx - \Lambda_\lambda(x)$, $t\in\mathbb{R}$. This yields

$$\Lambda_\lambda^*(x) = \operatorname{arsinh}\left(\frac{x}{\lambda}\right)x - \lambda\left(\cosh\left(\operatorname{arsinh}\left(\frac{x}{\lambda}\right)\right) - 1\right).$$

For independent random variables $Y_{\rho\alpha}^r$, $Z_{n,r}$, $n,r\geq 1$, and $Y_{k\rho\alpha}$, Z_n, $n,k\geq 1$, respectively, using the above notation, we obtain for each $\gamma > 0$ by the application of Cramér's theorem:

$$\lim_{k\to\infty}\frac{1}{k}\log\left[\mathbb{P}\left(\sum_{n=1}^{Y_{k\rho\alpha}} Z_n \geq \gamma k\right)\right] = \lim_{k\to\infty}\frac{1}{k}\log\left[\mathbb{P}\left(\sum_{r=1}^{k}\left(\sum_{n=1}^{Y_{\rho\alpha}^r} Z_{n,r}\right) \geq \gamma k\right)\right] = -\Lambda_{\rho\alpha}^*(\gamma).$$

We observe by simple transformations

$$\begin{aligned}&\frac{1}{\log(N)}\min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}}\log\left[\mathbb{P}\left(\sum_{r=1}^{k}\left(\sum_{n=1}^{Y_{\rho\alpha}^r} Z_{n,r}\right) \geq \gamma\log(N)\right)\right]\\ \geq&(1-\delta)\min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}}\frac{1}{\log(N)(1-\delta)}\log\left[\mathbb{P}\left(\sum_{r=1}^{k}\left(\sum_{n=1}^{Y_{\rho\alpha}^r} Z_{n,r}\right) \geq \gamma k\frac{1}{1-\delta}\right)\right]\\ \geq&(1-\delta)\min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}}\frac{1}{k}\log\left[\mathbb{P}\left(\sum_{r=1}^{k}\left(\sum_{n=1}^{Y_{\rho\alpha}^r} Z_{n,r}\right) \geq \gamma k\frac{1}{1-\delta}\right)\right]\end{aligned}\tag{3.21}$$

and therefore

$$\liminf_{N\to\infty} \min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \frac{1}{\log(N)} \log\left[\mathbb{P}\left(\sum_{r=1}^{k}\left(\sum_{n=1}^{Y^r_{\rho\alpha}} Z_{n,r}\right) \geq \gamma\log(N)\right)\right]$$
$$\geq -(1-\delta)\Lambda^*_{\rho\alpha}\left(\frac{\gamma}{1-\delta}\right). \tag{3.22}$$

For fixed $x > 0$, the function $F(\lambda, x) : \mathbb{R}^2_+ \to \mathbb{R}$, $F(\lambda, x) := \Lambda^*_\lambda(x)$ is decreasing in λ: $\frac{\partial F(\lambda,x)}{\partial\lambda} = 1 - \sqrt{1+\frac{x^2}{\lambda^2}}$ is negative for $x, \lambda > 0$. This yields

$$\min_{\rho\in(1-\delta,1+\delta)} \liminf_{N\to\infty} \min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \frac{1}{\log(N)} \log\left[\mathbb{P}\left(\sum_{n=1}^{Y_{k\rho\alpha}} Z_n \geq \gamma\log(N)\right)\right]$$
$$= \min_{\rho\in(1-\delta,1+\delta)} \liminf_{N\to\infty} \min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \frac{1}{\log(N)} \log\left[\mathbb{P}\left(\sum_{r=1}^{k}\left(\sum_{n=1}^{Y^r_{\rho\alpha}} Z_{n,r}\right) \geq \gamma\log(N)\right)\right]$$
$$\geq -(1-\delta)\Lambda^*_{(1-\delta)\alpha}\left(\frac{\gamma}{1-\delta}\right). \tag{3.23}$$

We finally resume and conclude, starting from (3.13) and using the considerations in (3.15), (3.19) and (3.20) on the three summands in (3.14):

$$\max_{\substack{B_\delta(k)\cap C_\delta(k)\\ \cap D(k)}} \left[\mathbb{P}\left(\Big|\sum_{\mu=2}^{M}\sum_{j\leq k} \xi^\mu_{k+1}\xi^\mu_j\Big| < \gamma\log(N)\ \Big|\mathcal{F}^N_k\right)\right]^{N-k}$$
$$\leq \left[1+(1+\delta)\alpha\binom{k}{2}p - \min_{\substack{x_1\in\mathbb{N}:x_1\log(N)/\\ (\alpha kN)\in(1-\delta,1+\delta)}} \mathbb{P}\left(\sum_{n=1}^{R_{x_1}} Z_n \geq \gamma\log(N)\right)\right]^{N-k}$$
$$\leq \left[1+(1+\delta)\alpha\binom{k}{2}p + 2p(1+\delta)\alpha k - \min_{\rho\in(1-\delta,1+\delta)} \mathbb{P}\left(\sum_{n=1}^{Y_{k\rho\alpha}} Z_n \geq \gamma\log(N)\right)\right]^{N-k}.$$

So there is an error with high probability, if

$$\min_{\rho\in(1-\delta,1+\delta)} \liminf_{N\to\infty} \min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \frac{1}{\log(N)} \log\left[\mathbb{P}\left(\sum_{n=1}^{Y_{k\rho\alpha}} Z_n \geq \gamma\log(N)\right)\right] > -1.$$

Due to (3.21) and (3.23), this is fulfilled if

$$-(1-\delta)\Lambda^*_{(1-\delta)\alpha}\left(\frac{\gamma}{1-\delta}\right) > -1.$$

Finally $F(\lambda, x)$ is continuous in λ and x, and

$$\lim_{\delta\searrow 0} -(1-\delta)\Lambda^*_{(1-\delta)\alpha}\left(\frac{\gamma}{1-\delta}\right) = -\Lambda^*_\alpha(\gamma).$$

Because δ can be chosen arbitrarily small, $\mathbb{P}\left(\exists i \le N : T_i(\xi^1) \neq \xi_i^1\right) \longrightarrow 1$ holds for any $\alpha > 0$ with

$$-\Lambda_\alpha^*(\gamma) > -1.$$

This condition can be written as

$$-\gamma \text{arsinh}\left(\frac{\gamma}{\alpha}\right) + \alpha\left(\cosh\left(\text{arsinh}\left(\frac{\gamma}{\alpha}\right)\right) - 1\right) > -1$$

and is fulfilled if $g_\gamma(\gamma/\alpha) > 0$, i.e. if $\alpha > \gamma/y_\gamma$.

□

Proposition 3.2 *Any threshold $\gamma \log(N)$ with $\gamma \ge 1$ leads to instability of an fixed but arbitrary stored pattern, with positive probability not converging to 0; each $\gamma \ge 1$ is thus an inadmissible threshold variable.*

Proof of Proposition 3.2 Using $\gamma \ge 1$ and $\mu = 1$,

$$\mathbb{P}\left(\sum_{j=1}^N |\xi_j^1| < \gamma \log(N)\right) \ge \mathbb{P}\left(\sum_{j=1}^N |\xi_j^1| < \log(N)\right) \longrightarrow \frac{1}{2}.$$

This is a result of the two observations

$$\left|\mathbb{P}\left(\sum_{j=1}^N |\xi_j^1| < \log(N)\right) - \mathbb{P}\left(Y_{\log(N)} < \log(N)\right)\right| \le 2p^2N = 2p\log(N),$$

$Y_{\log(N)} \sim \text{Poi}(\log(N))$, and

$$\mathbb{P}\left(Y_{\log(N)} < \log(N)\right) = \mathbb{P}\left(\frac{Y_{\log(N)} - \log(N)}{\sqrt{\log(N)}} < 0\right) \longrightarrow \frac{1}{2}$$

according to the Central Limit Theorem. We again assume that the non-zero entries are exactly in the first k places if $\sum_{j=1}^N |\xi_j^1| = k$. For an arbitrary $i \le k$, the local field of ξ^1 is

$$S_i(\xi^1) = \sum_{\mu=1}^M \sum_{\substack{j=1 \\ j\neq i}}^N \xi_j^1 \xi_i^\mu \xi_j^\mu = (k-1)\xi_i^1 + \sum_{\mu=2}^M \sum_{\substack{j=1 \\ j\neq i}}^k \xi_j^1 \xi_i^\mu \xi_j^\mu.$$

If $k < \gamma \log(N)$, the neuron is definitely instable if $\text{sgn}(\sum_{\mu=2}^M \sum_{j=1, j\neq i}^k \xi_j^1 \xi_i^\mu \xi_j^\mu) \neq \text{sgn}(\xi_i^1)$, and the probability of this event tends to $\frac{1}{2}$.

□

Proposition 3.3 *The set of permissible threshold variables γ is the interval $\Gamma = (0, 1)$. For each $\gamma \in (0, 1)$, the critical value $\alpha^*(\gamma)$, such that $(0, \alpha^*(\gamma))$ is the set of admissible capacity variables for γ and $(\alpha^*(\gamma), \infty)$ is the set of inadmissible capacity variables for γ, is $\alpha^*(\gamma) = \gamma/y_\gamma$ with the root y_γ of the function g_γ.*

The critical value $\alpha^* := \sup\{\alpha > 0 : \exists \gamma \in (0,1) : \alpha$ *is an admissible capacity variable for* $\gamma\}$ *is equal to*

$$\alpha^* = \frac{1}{y_1}, \quad -\text{arsinh}(y_1) + y_1^{-1}(\cosh(\text{arsinh}(y_1)) - 1) + 1 = 0, \quad \alpha^* \approx 0.3829. \tag{3.24}$$

In particular, if $\alpha > \alpha^*$, *we have for each* $\gamma \in (0,1)$ *used as threshold variable and arbitrary* μ

$$\lim_{N\to\infty} \mathbb{P}(\exists i \le N : T_i(\xi^\mu) \ne \xi_i^\mu) = 1$$

and if $\alpha < \alpha^*$, *there is some* $\gamma \in (0,1)$ *such that for the dynamics with threshold variable* γ, $M = \alpha N^2/\log(N)^2$ *stored patterns and arbitrary* $\mu \le M$,

$$\lim_{N\to\infty} \mathbb{P}(\exists i \le N : T_i(\xi^\mu) \ne \xi_i^\mu) = 0.$$

In this case, there is a nonempty interval $(\gamma^*(\alpha), 1)$ *of possible threshold variables* γ *such that* α *is an admissible capacity variable for* γ.

Proof of Proposition 3.3: First, we showed in the previous Proposition 3.2 that $\gamma \ge 1$ leads to instability of the stored messages. We derived in the proof of Theorem 3.1 that α fulfills the stability condition if $\alpha < \gamma/y_\gamma$, where y_γ is the unique root of g_γ in $\mathbb{R}_+$, and leads to instability of the stored patterns if $\alpha > \gamma/y_\gamma$. For each $\gamma \in (0,1)$, there is thus a nonempty interval $(0, \alpha^*(\gamma))$ with $\alpha^*(\gamma) = \gamma/y_\gamma$ such that (3.2) holds for each $\alpha \in (0, \alpha^*(\gamma))$ and (3.3) holds for each $\alpha \in (\alpha^*(\gamma), \infty)$. The set of admissible threshold variables is $(0,1)$.

To show that $\alpha^* = \frac{1}{y_1}$, define the function $G : \mathbb{R}_+^2 \to \mathbb{R}$,

$$G(\gamma, x) := g_\gamma(x), \quad g_\gamma(x) = -\text{arsinh}(x) + x^{-1}(\cosh(\text{arsinh}(y_1)) - 1) + \frac{1}{\gamma}.$$

G is continuous and for fixed x strictly decreasing in γ. For this reason, $\gamma < \gamma'$ implies $y_\gamma > y_{\gamma'}$ and therefore $\alpha^*(\gamma) < \alpha^*(\gamma')$.

Now let $\alpha > \frac{1}{y_1}$, y_1 as defined in (3.24). We show that there is no $\gamma \in (0,1)$ such that α is an admissible capacity variable for γ. Due to the definition of y_1 and the properties of y_γ we obtain

$$\alpha > \frac{1}{y_1} > \frac{\gamma}{y_1} > \frac{\gamma}{y_\gamma}$$

for each $\gamma \in (0,1)$ and consequently an arbitrary stored pattern is instable with high probability if α is used as capacity variable, for each $\gamma \in (0,1)$. This shows that $\alpha^* \le \frac{1}{y_1}$.

If in contrary $\alpha < \frac{1}{y_1}$, we show that there is some γ such that $\alpha \in (0, \alpha^*(\gamma))$. This implies $\alpha^* = \frac{1}{y_1}$. Due to the three facts that y_1 is the root of g_1, $\alpha < \frac{1}{y_1}$ and that g_1 is strictly decreasing, we have

$$g_1\left(\frac{1}{\alpha}\right) < g_1(y_1) = 0.$$

Since G is continuous, there is an $\gamma \in (0,1)$ such that $G(\gamma, \gamma/\alpha) < 0$ and α is an admissible capacity variable for γ. In addition, we saw that $\alpha^*(\gamma) < \alpha^*(\gamma')$ if $\gamma < \gamma'$. There is thus a nonempty interval $(\gamma^*(\alpha), 1)$, $\gamma^*(\alpha) = \inf\{\gamma \in (0,1) : \alpha \in (0, \alpha^*(\gamma))\}$, such that α is admissible for each $\gamma \in (\gamma^*(\alpha), 1)$ and for none $\gamma \in (0, \gamma^*(\alpha))$. □

Proposition 3.4 *The capacity in the model with independent and identically distributed spins per message drastically decreases if one wants all patterns to be stable. Precisely, in the model using threshold $\gamma \log(N)$, $\gamma \in (0,1)$, there is at least one stored pattern that is not stable, with positive probability not converging to 0, if $N^\beta = \mathcal{O}(M)$,*

$$\beta = \gamma \log(\gamma) - \gamma + 1.$$

In contrary, there are exclusively stable patterns, with high probabillity, if

$$M = \mathcal{O}(N^\beta)$$

for some β such that

$$-\gamma \log(\gamma) + \gamma - 1 < -\beta.$$

In particular, $\beta \in (0,1)$ because $\gamma \in (0,1)$.

Proof of Proposition 3.4: Concerning the proof of the last statement, assume that the threshold is $\gamma \log(N)$, $\gamma < 1$, and that $M \leq CN^\beta$, for some $C > 0$, and $-\beta > -\gamma \log(\gamma) + \gamma - 1$. Let $\delta > 0$ be chosen such that $-\beta > -(\gamma+\delta)\log(\gamma+\delta) + \gamma + \delta - 1$ and that $\gamma + \delta < 1$. The probability that a stored pattern has less than $(\gamma+\delta)\log(N)$ excited neurons is at most

$$\begin{aligned}\mathbb{P}\left(\sum_{j=1}^N |\xi_j^\mu| \leq (\gamma+\delta)\log(N)\right) &\leq \exp\left[t(\gamma+\delta)\log(N)\right](1-p+pe^{-t})^N \\ \leq \exp\left[\log(N)[t(\gamma+\delta)+(e^{-t}-1)]\right] &\leq N^{-(\gamma+\delta)\log(\gamma+\delta)+\gamma+\delta-1}.\end{aligned}$$

In the same way, we obtain

$$\mathbb{P}\left(\sum_{j=1}^N |\xi_j^\mu| \geq 3\log(N)\right) \leq N^{-3\log(3)+3-1}.$$

We observe that $-3\log(3)+2 \approx -1.3 < -(\gamma+\delta)\log(\gamma+\delta)+\gamma+\delta-1 \in (-1,0)$ for each $\gamma+\delta \in (0,1)$. The excited as well as the inactive neurons will now be stable in each pattern with high probability: for $s, t > 0$,

$$\begin{aligned}&\mathbb{P}\left(\exists \mu \leq M : \exists i \leq N : T_i(\xi^\mu) \neq \xi_i^\mu\right) \leq M\mathbb{P}\left(\exists i \leq N : T_i(\xi^1) \neq \xi_i^1\right) \\ \leq &2MN^{-(\gamma+\delta)\log(\gamma+\delta)+\gamma+\delta-1} + MN \max_{k\in\mathbb{N}:k/\log(N)\in(\gamma+\delta,3)} \mathbb{P}\left(\left|\sum_{\mu=2}^M \sum_{j\leq k} \xi_N^\mu \xi_j^\mu\right| \geq \gamma\log(N)\right) \\ &+ M \max_{k\in\mathbb{N}:k/\log(N)\in(\gamma+\delta,3)} k\,\mathbb{P}\left(\left|\sum_{\mu=2}^M \sum_{j=2}^k \xi_1^\mu \xi_j^\mu\right| \geq \delta\log(N)\right) \\ \leq &2MN^{-(\gamma+\delta)\log(\gamma+\delta)+\gamma+\delta-1} \\ &+MN\exp\left[-t\gamma\log(N) + M3\log(N)p^2(\cosh(t)-1) + \mathcal{O}\left(M\log(N)^2p^3\right)\right] \\ &+M3\log(N)\exp\left[-s\delta\log(N) + M3\log(N)p^2(\cosh(s)-1) + \mathcal{O}\left(M\log(N)^2p^3\right)\right].\end{aligned}$$

With arbitrary constants $t > \frac{\beta+1}{\gamma}$ and $s > \frac{\beta}{\delta}$ (that can be chosen independently of N), this tends to 0.

For the opposite bound, let us first observe that a pattern whose number of excited neurons is too low, that is, below $\gamma \log(N)$, is not stable with positive probability not tending to 0 because the signal term in the local field of the active neurons is not sufficiently high (it is $k-1$ if k neurons are active in the pattern). An arbitrary but fixed excited neuron of the pattern has a positive probability of not being stable because the noise term of the local field of ξ^μ in neuron i,

$$\sum_{\nu \neq \mu} \sum_{j \neq i} \xi_j^\mu \xi_i^\nu \xi_j^\nu$$

has, with positive probability ($\approx \frac{1}{2}$), the wrong sign. Hence it is enough to recall the results of Proposition 2.8. □

Proposition 3.5 *If the model is altered in the way that the random variables ξ_i^μ, $i \leq N$, are (for fixed μ) no longer independent, but there are exactly c excited neurons per message, the lower bounds on α concerning the stability and error correction obtained in Theorem 3.1 remain true. In addition, the stability of all patterns can be reached if*

$$\alpha < \frac{\gamma}{y_{\gamma/3}}, \quad \alpha < \frac{1-\gamma}{y_{(1-\gamma)/2}}$$

where y_γ is the unique root of the function

$$g_\gamma(x) = -arsinh(x) + x^{-1}\left(cosh(arsinh(x)) - 1\right) + \frac{1}{\gamma}.$$

Then

$$\mathbb{P}\left(\exists \mu \leq M : \exists i \leq N : T_i(\xi^\mu) \neq \xi_i^\mu\right) \longrightarrow 0.$$

Proof of Proposition 3.5 As in the proof of Proposition 2.10, we determine, for $i > c$, the following conditional exponential moment, given $\mathcal{Z}_c$:

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(t \sum_{\mu=2}^{M} \sum_{j=1}^{c} \xi_j^1 \xi_i^\mu \xi_j^\mu\right) \Big| \mathcal{Z}_c\right] = \mathbb{E}\left[\exp\left(t \sum_{j=1}^{c} \xi_i^M \xi_j^M\right)\right]^{M-1} \\
&= \left[1 - \frac{c}{N} + \frac{c}{N}\mathbb{E}\left(\exp\left(t \sum_{j=1}^{c} \xi_i^M \xi_j^M\right) \Big| \{\xi_i^M \neq 0\}\right)\right]^{M-1} \\
&= \left[1 - \frac{c}{N} + \frac{c}{N}\left(\sum_{n=0}^{c-1} \binom{c}{n} \prod_{m=1}^{n} \frac{c-m}{N-m} \prod_{k=0}^{c-n-1}\left(1 - \frac{c-1-n}{N-1-n-k}\right) \cosh(t)^n\right)\right]^{M-1} \\
&= \left[1 + \frac{c^2(c-1)}{N^2}(\cosh(t) - 1) + \mathcal{O}\left(\frac{c^5}{N^3}\right)\right]^{M-1} \\
&\leq \exp\left[\alpha(\cosh(t) - 1)(c-1) + \mathcal{O}(c^3/N)\right].
\end{aligned}$$

In the first line we used that $\sum_{j\leq c}\xi_j^1\xi_i^\mu\xi_j^\mu, \mu \geq 2$ are conditionally independent, given $\xi_j^1, j \leq c$, and that the conditional distribution of one of these sums, given $\mathcal{Z}_c$, is the same as the distribution of $\sum_{j\leq c}\xi_i^\mu\xi_j^\mu$. In the third line we used that the distribution of

$$\sum_{j\leq c}|\xi_j^\mu|$$

is conditionally Hypergeometric, given $\{|\xi_i^\mu| = 1\}$, with parameters $N - c - 1$ (failure), c (success) and $c - 1$ (number of drawings). Finally, $\xi_i^\mu\xi_j^\mu, j \leq c$ are conditionally independent, given $|\xi_i^\mu|, |\xi_j^\mu|, j \leq c$, and identically distributed with

$$\mathbb{P}\left(\xi_i^\mu\xi_j^\mu = 0\middle||\xi_i^\mu|, |\xi_j^\mu|\right) = 1 - |\xi_i^\mu||\xi_j^\mu|, \quad \mathbb{P}\left(\xi_i^\mu\xi_j^\mu = \pm 1\middle||\xi_i^\mu|, |\xi_j^\mu|\right) = \frac{1}{2}|\xi_i^\mu||\xi_j^\mu|.$$

Conditional on $\{|\xi_i^\mu| = 1, \sum_{j\leq k}|\xi_j^\mu| = n\}$, the exponential moment of $t\sum_{j\leq c}\xi_i^\mu\xi_j^\mu$ is thus $\cosh(t)^n$.

This can be used to show a fixed pattern's stability with high probability for small enough α, as in the proof of Theorem 3.1. The lower bounds on α remain true.

If we consider all messages at the same time, the probability of at least one error is bounded by

$$\begin{aligned}
&\mathbb{P}\left(\exists\mu \leq M, \exists i \leq N : T_i(\xi^\mu) \neq \xi_i^\mu\right) \\
\leq &M(N - c)\mathbb{P}\left(\exists i \leq N : \xi_i^1 = 0, T_i(\xi^1) \neq 0\right) + Mc\mathbb{P}\left(\exists i \leq N : \xi_i^1 \neq 0, T_i(\xi^1) = 0\right).
\end{aligned}$$

The first summand, concerning the inactive neurons, differs from the one concerning one fixed stored pattern by the factor $M = \alpha N^2/\log(N)^2$; using the exponential moment above, it tends to 0 if

$$-\gamma\operatorname{arsinh}\left(\frac{\gamma}{\alpha}\right) + \alpha\left(\cosh\left(\operatorname{arsinh}\left(\frac{\gamma}{\alpha}\right)\right) - 1\right) < -3.$$

Using the termini of the proof of Theorem 3.1, this condition can be written in terms of $g_{\gamma/3}$:

$$\gamma g_{\gamma/3}\left(\frac{\gamma}{\alpha}\right) < 0.$$

Since $g_{\gamma/3}$ is strictly decreasing on $\mathbb{R}_+$, this is for $\alpha, \gamma > 0$ equivalent to

$$\alpha < \frac{\gamma}{y_{\gamma/3}}.$$

Concerning the active neurons of a stored pattern, their noise term of the local field must not exceed the threshold $c - 1 - \gamma c = (1 - \gamma)c - 1$. We obtain the condition

$$-(1-\gamma)\operatorname{arsinh}\left(\frac{1-\gamma}{\alpha}\right) + \alpha\left(\cosh\left(\operatorname{arsinh}\left(\frac{1-\gamma}{\alpha}\right)\right) - 1\right) < -2.$$

In terms of $g_{(1-\gamma)/2}$, this is

$$(1-\gamma)g_{(1-\gamma)/2}\left(\frac{1-\gamma}{\alpha}\right) < 0$$

and this is finally fulfilled if

$$\alpha < \frac{1-\gamma}{y_{(1-\gamma)/2}}.$$

□

Choosing the stored patterns among all messages with exactly c active neurons thus improves the performance of the model.

In addition to the stability of the stored patterns, we are interested in the error-correcting abilities of the network. We distinguish three types of errors. Suppose that a stored message ξ^μ is destroyed. Then it can be corrupted by

1. deleting spins of neurons whose values are either 1 or -1, i.e. deactivate the corresponding neurons (erasure),
2. reversing spins of excited neurons, i.e multiply them by -1; and finally
3. spuriously activating neurons whose state in ξ^μ is 0: map them either to 1 or to -1.

We can prove the following result for these three types of errors and, finally, for a message concerned by all types of errors.

Corollary 3.6 (Corollary to Theorem 3.1:) *Let ξ^μ be an arbitrary stored pattern. Suppose that the number of stored messages is in any case at most $M = \alpha N^2/\log(N)^2$, $\alpha < \frac{\gamma}{y_\gamma}$, if $\gamma \in (0,1)$ is used as threshold, to guarantee stability of the stored messages (see Theorem 3.1).*

1. *Suppose that the pattern $\hat{\xi}_1^\mu$ is obtained by deleting at random $\varrho_1 \log(N)$ non-zero spins in ξ^μ. Then the errors can, with high probability, be corrected in one step by the retrieval dynamics, if $\varrho_1 < 1-\gamma$, where γ determines the threshold $\gamma \log(N)$. If $\varrho_1 > 1-\gamma$, the pattern is never corrected, with high probability.*

   Other than in the Hopfield model, it is in particular possible to correct a number of errors $\varrho_1 \log(N)$ (obtained by deleting entries), $\varrho_1 > 0.5$, more than half of the non-zero entries of the message, if the threshold is small enough.

2. *Concerning a pattern $\hat{\xi}_2^\mu$ obtained by multiplying $\varrho_2 \log(N)$ randomly chosen non-zero spins of ξ^μ by -1, we can recover the original message ξ^μ in one step, with high probability, if ϱ_2 satisfies*

   $$\varrho_2 < \frac{1-\gamma}{2}.$$

   The pattern is never corrected, with high probability, if $\varrho_2 > \frac{1-\gamma}{2}$. In this case of corruption it is not possible to correct more than $0.5\log(N)$ errors.

3. *If $\varrho_3 \log(N)$ of the inactive neurons of ξ^μ have been spuriously activated, we can correct these errors with high probability in the first step of the retrieval dynamics if α satisfies*

   $$\alpha < \frac{\gamma}{y_\gamma(1+\varrho_3)},$$

where y_γ is again the unique root of the function g_γ. This bound is also sharp concerning the one step retrieval: the message is not corrected in the first step, if $\alpha > \frac{\gamma}{y_\gamma(1+\varrho_3)}$.

4. *Finally, if $\hat{\xi}_4^\mu$ is a corrupted version of ξ^μ which combines these three types of errors, i.e.: $\varrho_1 \log(N)$ of the activated neurons have been deactivated, $\varrho_2 \log(N)$ of the spins have been inverted and $\varrho_3 \log(N)$ inactive neurons have been activated, either to 1 or to -1, the correct message will be recovered with high probability in one step, if ϱ_1 and ϱ_2 satisfy*

$$\varrho_1 + 2\varrho_2 < 1 - \gamma$$

and α satisfies

$$\alpha < \frac{\gamma}{(1 - \varrho_1 + \varrho_3)y_\gamma}.$$

The second condition is not supplementary to the stability condition (3.1), if $\varrho_1 \geq \varrho_3$. If $\varrho_1 < \varrho_3$ and $\alpha > \frac{\gamma}{(1-\varrho_1+\varrho_3)y_\gamma}$, the pattern is not corrected in one step, with high probability. Finally, the pattern is not corrected in the first step, with high probability, if $\varrho_1 + 2\varrho_2 > 1 - \gamma$.

Proof: Without loss of generality, we consider ξ^1, assuming that the active neurons of ξ^1 are exactly in the places $1, \ldots, k$ for some $k \in \mathbb{N}, k/\log(N) \in (1-\delta, 1+\delta)$. We keep the notation of Theorem 3.1.

We begin with the first case and assume that $\varrho_1 \log(N)$ of the active neurons of ξ^1 are deactivated in $\hat{\xi}_1^1$. We suppose that α fulfills the stability condition of Theorem 3.1. Then the inactive neurons will also be stable with high probability if some of the non-zero entries are deleted. To see this, consider for $i > k$

$$S_i(\hat{\xi}^1) := \sum_{j \neq i} \sum_{\mu=1}^{M} \hat{\xi}_j^1 \xi_i^\mu \xi_j^\mu = \sum_{j \leq k, \hat{\xi}_j^1 \neq 0} \sum_{\mu=2}^{M} \xi_j^1 \xi_i^\mu \xi_j^\mu.$$

The only difference to the proof of the previous theorem is that the exterior sum has $k - \varrho_1 \log(N)$ summands instead of k, if there are k activated neurons in the message. Using again A_δ and $B_\delta(i)$, we show by the same arguments as in the proof of Theorem 3.1 that the probability of turning a deactivated neuron into 1 or -1 is bounded by

$$\mathbb{P}(\exists i : \xi_i^1 = 0, T_i(\hat{\xi}^1) \neq 0) \leq \mathbb{P}(A_\delta^c) + N\mathbb{P}(B_\delta(N)^c) + \\ N \max_{\substack{k \in \mathbb{N}: k/\log(N) \\ \in (1-\delta, 1+\delta)}} \max_{\substack{l \in \mathbb{N}: l\log(N)/ \\ (\alpha N) \in (1-\delta, 1+\delta)}} e^{-t\gamma \log(N)} (1 + p(\cosh(t) - 1))^{l(k - \varrho_1 \log(N))}.$$

The term $\cosh(t) - 1$ is positive for $t > 0$, so

$$(1 + p(\cosh(t) - 1))^{l(k - \varrho_1 \log(N))} \leq (1 + p(\cosh(t) - 1))^{lk}$$

and the last line is upper bounded by inserting $l = \alpha N(1+\delta)/\log(N)$, $k = (1+\delta)\log(N)$. We are now in the same situation as in the proof of Theorem 3.1 concerning the stability of the inactive neurons. Since their stability is provided with high probability if (3.1) is fulfilled, they also are stable in $\hat{\xi}_1^1$, with probability converging to 1.

Concerning the active neurons, we have for any $i \leq k$

$$\begin{aligned} S_i(\hat{\xi}_1^1) =& \sum_{j\neq i}\sum_{\mu=1}^{M} \hat{\xi}_{1j}^1 \xi_i^\mu \xi_j^\mu = \sum_{j\leq k, j\neq i, \hat{\xi}_{1j}^1 \neq 0} \xi_j^1 \xi_i^1 \xi_j^1 + \sum_{j\leq k, j\neq i, \hat{\xi}_{1j}^1 \neq 0} \sum_{\mu=2}^{M} \xi_j^1 \xi_i^\mu \xi_j^\mu \\ =& \xi_i^1 (k - \varrho_1 \log(N) - |\hat{\xi}_{1i}^1|) + \sum_{j\leq k, j\neq i, \hat{\xi}_{1j}^1 \neq 0} \sum_{\mu=2}^{M} \xi_j^1 \xi_i^\mu \xi_j^\mu . \end{aligned}$$

The difference between active neurons of ξ^1 concerned by the deletion and active neurons of ξ^1 not concerned by the deletion again is negligible. To reach that an active neuron of ξ^1 is not corrected after the first step or does not keep its correct value, respectively, the sum on the right hand side of the last line must exceed the value $k - \varrho_1 \log(N) - \gamma \log(N) - 1$. There are two differences compared to the proof of Theorem 3.1: there are less summands in the random term of the local field and the threshold to exceed to reach instability is decreased by $\varrho_1 \log(N)$. We obtain analogously to this proof

$$\begin{aligned} &\max_{k\in\mathbb{N}: k/\log(N)\in(1-\delta,1+\delta)} \mathbb{P}(\exists i \leq N : \xi_i^1 \neq 0, T_i(\hat{\xi}^1) = 0 | \mathcal{Z}_k) \\ \leq& \max_{k\in\mathbb{N}: k/\log(N)\in(1-\delta,1+\delta)} k\,\mathbb{P}\left(\left| \sum_{j=1}^{k-\varrho_1 \log(N)} \sum_{\mu=2}^{M} \xi_1^\mu \xi_j^\mu \right| \geq \left(\frac{k}{\log(N)} - \varrho_1 - \gamma \right) \log(N) - 1 \right) \\ \leq& (1+\delta)\log(N) \left[\mathbb{P}(B_\delta(1)^c) + e^{-t(1-\delta-\varrho_1-\gamma)\log(N)+t} \left(1 + p(\cosh(t) - 1)\right)^{(1+\delta)(1+\delta-\varrho_1)\alpha N} \right]. \end{aligned}$$

The steps we made are the same as in the proof of Theorem 3.1 or the previous proof concerning the stability of the inactive neurons and are therefore not explained in detail.

With $t = \operatorname{arsinh}((1-\delta-\varrho_1-\gamma)/((1+\delta)(1+\delta-\varrho_1)\alpha))$, the wanted convergence is reached for each ϱ_1, $\varrho_1 < 1-\gamma$ if δ is chosen such that $0 < \delta < 1-\gamma-\varrho_1$ because the function g (see (3.11)) is negative on $\mathbb{R}_+$.

To show the reverse bound, consider $\varrho_1 > 1-\gamma$. For an active neuron of ξ^μ, the absolute value of the signal term coming from message ξ^μ is with high probability not big enough to exceed the threshold: on A_δ, it is at most $(1+\delta-\varrho_1)\log(N)$. So the probability of turning $\hat{\xi}_1^1$ into $(0,\ldots,0)$ is for $k \leq (1+\delta)\log(N)$ at least

$$\begin{aligned} &\mathbb{P}(\forall i \leq k : |S_i(\hat{\xi}_1^1)| < \gamma \log(N) | \mathcal{Z}_k) \geq 1 - k\mathbb{P}(|S_1(\hat{\xi}_1^1)| \geq \gamma \log(N) | \mathcal{Z}_k) \\ \geq& 1 - 2k\mathbb{P}\left[\sum_{\mu=2}^{M} \sum_{j=2}^{(1+\delta-\varrho_1)\log(N)+1} \xi_1^\mu \xi_j^\mu \geq (\gamma - 1 - \delta + \varrho_1)\log(N) \right]. \end{aligned}$$

Analogously to the computations concerning the case $\varrho_1 < 1-\gamma$, this vanishes as N tends to infinity, if $\delta < \varrho_1 + \gamma - 1$.

We do not prove the statements concerning the other three cases in detail, because the proofs are very similar to the one of Theorem 3.1 and to the first part of this proof. There are only two things to observe: first, for arbitrary $i \le N$, the probability that $\sum_{\mu=1}^{l} \xi_i^\mu$ exceeds a threshold, e. g.,

$$\mathbb{P}\left(\sum_{\mu=1}^{l} \xi_i^\mu \ge \gamma \log(N)\right),$$

is for each choice of $t > 0$ and $l' \le l$ bounded by

$$\mathbb{P}\left(\sum_{\mu=1}^{l} \xi_i^\mu \ge \gamma \log(N)\right) \le e^{-t\gamma \log(N)}(1 + p(\cosh(t) - 1))^l$$
$$\le e^{-t\gamma \log(N)}(1 + p(\cosh(t) - 1))^{l'}.$$

We can thus use upper bounds received for sums of bigger length (this is what we did in the first part of the proof when we explained the high probability of the stability of the inactive neurons in $\hat{\xi}^1$ by the one of the stability of the inactive neurons in ξ^1 where the length of the sum in the noise term of the local field is larger).

Second, we have to examine the effects of the corruption on the local fields in the different cases:

1. This case has already been examined.

2. If we multiply the values of $\varrho_2 \log(N)$ activated neurons by -1, which are, without loss of generality, neurons $i = 1, \dots, \varrho_2 \log(N)$, this does not cause any supplementary requirements on α to guarantee the stability of the inactive neurons of ξ^1. Their local field is for some $i > k$ equal to

$$S_i(\hat{\xi}_2^1) = \sum_{j \le k} \sum_{\mu=1}^{M} \hat{\xi}^1_{2j} \xi_i^\mu \xi_j^\mu = \sum_{j \le \varrho_2 \log(N)} \sum_{\mu=2}^{M} (-\xi_j^1) \xi_i^\mu \xi_j^\mu + \sum_{\varrho_2 + 1 \le j \le k} \sum_{\mu=2}^{M} \xi_j^1 \xi_i^\mu \xi_j^\mu$$
$$\sim \sum_{j \le k} \sum_{\mu=2}^{M} \xi_i^\mu \xi_j^\mu.$$

   The probability $\mathbb{P}(|S_i(\hat{\xi}_2^1)| > \gamma \log(N))$ remains the same as the corresponding probability in the proof concerning the stability of ξ^1 in Theorem 3.1 because the same number of neurons is active in the two patterns.

   For any active neuron $i \le k$ of ξ^1, the local field is

$$S_i(\hat{\xi}_2^1) = \left(k - 2\varrho_2 \log(N) - \mathbb{1}_{\hat{\xi}_i^1 = \xi_i^1}\right)\xi_i^1 - \sum_{j \ne i}^{\varrho_2 \cdot \log(N)} \sum_{\mu=2}^{M} \xi_j^1 \xi_i^\mu \xi_j^\mu + \sum_{\substack{j \ne i, j > \\ \varrho_2 \cdot \log(N)}}^{k} \sum_{\mu=2}^{M} \xi_j^1 \xi_i^\mu \xi_j^\mu.$$

   The remaining computations are thus made as in the proof of Theorem 3.1 concerning the stability of the active neurons of a message ξ^μ. The signal term is just decreased by $\varrho_2 \log(N)$; due to the results on page 54, the conditional distribution of the noise term, given $\mathcal{Z}_k$, does not change, compared to the distribution of the noise term of the input pattern ξ^1. To have $T_i(\hat{\xi}^1) \neq \xi_i^1$, it is necessary that

$$\Big| - \sum_{\substack{j\neq i}}^{\substack{\varrho_2\cdot\\ \log(N)}} \sum_{\mu=2}^{M} \xi_j^1 \xi_i^\mu \xi_j^\mu + \sum_{\substack{j\neq i, j>\\ \varrho_2\cdot\log(N)}}^{k} \sum_{\mu=2}^{M} \xi_j^1 \xi_i^\mu \xi_j^\mu \Big| > k - 2\varrho_2 \log(N) - \gamma \log(N) - 1.$$

The probability $\mathbb{P}(\exists i \leq k : T_i(\hat{\xi}^1) \neq \xi_i^1 | \mathcal{Z}_k)$ then vanishes for each $k \in \mathbb{N}, k/\log(N) \in (1-\delta, 1+\delta)$, if $1 - \delta - 2\varrho_2 - \gamma > 0$. Combined with the fact that A_δ tends to 1 for each $\delta > 0$ the assertion follows immediately, if $2\varrho_2 < 1 - \gamma$.

If $2\varrho_2 > 1 - \gamma$, for an active neuron in ξ^1, the absolute value of the signal term in $S_i(\hat{\xi}_2^1)$ coming from pattern ξ^1 is at most $k - 2\varrho_2 \log(N)$ and with high probability smaller than $(1 + \delta - 2\varrho_2)\log(N)$. Since δ can be chosen such that $\delta < \gamma - 1 + 2\varrho_2$, the probability of turning the pattern into $(0, \dots, 0)$ in the first step of the dynamics tends to 1.

3. If $\varrho_3 \log(N)$ of the inactive neurons have been activated, assuming that these are the neurons $k+1, \dots, k + \varrho_3 \log(N)$, the local field is for any $i > k$ equal to

$$S_i(\hat{\xi}_3^1) = \sum_{j \leq k} \sum_{\mu=2}^{M} \xi_j^1 \xi_i^\mu \xi_j^\mu + \sum_{\substack{k+1 \leq j \leq\\ k+\varrho_3 \log(N), j\neq i}} \sum_{\mu=2}^{M} \hat{\xi}_{3j}^1 \xi_i^\mu \xi_j^\mu .$$

We can thus refer to the proof of Theorem 3.1 concerning the stability of the inactive neurons, with the difference that the length of the exterior sum in the local field is $k + \varrho_3 \log(N)$ instead of k, because there are $k + \varrho_3 \log(N)$ active neurons. The exact value of $\hat{\xi}_{3i}^1$, $k + 1 \leq i \leq k + \varrho_3 \log(N)$, does not play any role but only the fact that the neuron is activated. We obtain

$$\begin{aligned}&\mathbb{P}(\exists i \leq N : \xi_i^1 = 0, T_i(\hat{\xi}_3^1) \neq 0)\\ \leq& N e^{-t\gamma \log(N)} e^{\log(N)(\cosh(t)-1)(1+\delta)(1+\delta+\varrho_3)\alpha} + \mathbb{P}(A_\delta^c) + N\mathbb{P}(B_\delta(N)^c);\end{aligned}$$

the only difference to (3.8) is that $\alpha(1 + \delta)$ is replaced by $\alpha(1 + \delta + \varrho_3)$. The convergence of this term to 0 is ensured if

$$\alpha(1 + \varrho_3) < y_\gamma, \text{ i.e. } \alpha < \frac{y_\gamma}{1 + \varrho_3}$$

as mentioned in the Lemma.

The excited neurons of ξ^1, i.e. neurons $1, \dots, k$, possess a local field that is equal to

$$S_i(\hat{\xi}_3^1) = (k-1) \cdot \xi_i^1 + \sum_{j \leq k, j \neq i} \sum_{\mu=2}^{M} \xi_j^1 \xi_i^\mu \xi_j^\mu + \sum_{\substack{k+1 \leq j \leq\\ k+\varrho_3 \log(N)}} \sum_{\mu=2}^{M} \hat{\xi}_{3j}^1 \xi_i^\mu \xi_j^\mu .$$

The signal term of the local field of these neurons remains unchanged compared to ξ^1 and the increased number of active neurons does not change the result: for $k/\log(N) \in (1-\delta, 1+\delta)$,

$$\begin{aligned}&\mathbb{P}\left(\exists i \leq k : T_i(\hat{\xi}_3^1) \neq \xi_i^1 | \mathcal{Z}_k\right)\\ \leq& 2(1+\delta)\log(N)\left(e^{-(1-\delta-\gamma)\log(N)t}e^{\log(N)(\cosh(t)-1)(1+\delta)(1+\delta+\varrho_3)\alpha} + \mathbb{P}(B_\delta(1)^c)\right)\end{aligned}$$

for each $t > 0$. Comparing this with (3.10), we observe that the probabiity tends to 0 if $1-\delta-\gamma > 0$ due to the negativity of the function g.

Concerning the reverse bound on α, consider the neurons that are not activated in $\hat{\xi}^1$. Then with high probability, there is at least one of them that is activated by the dynamics if the bound on α is exceeded. The proof works identically as in the proof of Theorem 3.1: the only difference is that potential signals can come from $k + \varrho_3 \log(N)$ instead of k neurons.

4. Finally we have to examine the case where all these errors are combined. Errors of the first and second kind (as considered in 1. and 2.) decrease the absolute value of the signal term of the local field of the active neurons by $\varrho_1 \log(N) + 2\varrho_2 \log(N)$. These errors are corrected and those active neurons of ξ^1 that are not affected by the corruption remain stable, if $\varrho_1 + 2\varrho_2 < 1-\gamma$. If $\varrho_1 + 2\varrho_2 > 1-\gamma$, the active neurons of ξ^1 are all deactivated in the first step of the dynamics, with high probability. This is proven analogously to 1. or 2.; the fact that the number of active neurons is increased by the spuriously activated neurons does not change the result.

   Concerning the inactive neurons, the relevant information is the number of active neurons in $\hat{\xi}^1$. If more neurons are deleted than spuriously activated, the inactive neurons of ξ^1 are inactive after one step of the dynamics, applied to $\hat{\xi}^1$, with high probability. If $\varrho_1 < \varrho_3$, we can continue as in 3.: the supplementary condition to the stability condition $\alpha < \gamma/y_\gamma$ then is

$$\alpha < \frac{\gamma}{y_\gamma(1-\varrho_1+\varrho_3)}$$

   to correct (if they have been corrupted), respectively keep (if they are the same in ξ^1 and $\hat{\xi}_4^1$) the inactive neurons of ξ^1. Again, this bound is sharp concerning the one step retrieval.

□

Remark 3.7 *The errors of the first and second kind (corruption of active neurons) affect the choice of* γ. *If* $\varrho_1 \log(N)$ *deleted neurons shall be corrected,* $\gamma < 1-\varrho_1$ *and if* $\varrho_2 \log(N)$ *reversed spins shall be corrected,* $\gamma < 1-2\varrho_2$. *For a fixed* γ *and* α *according to the stability conditions, it is possible to correct a certain number of errors, without changing* α. *On the contrary, an adaption of* α *cannot lead to stability in this case, if too many neurons are corrupted. In addition, the model is more vulnerable to a corruption in form of multiplication by* -1 *as to deletion.*

For fixed γ, α decreases antiproportionally with $1 + \varrho_3$, if $\varrho_3 \log(N)$ neurons are corrupted by the third type of error (spuriously activation) and shall be corrected. If $\alpha(\gamma)$ is the critical capacity variable for γ and $\varrho_3 \log(N)$ neurons are spuriously activated, the errors are corrected, with high probability, if $\alpha < \alpha^*(\gamma)/(1 + \varrho_3)$.

Remark 3.8 *The energy function*

$$H(\sigma) := -\frac{1}{2} \sum_{\substack{i \neq j \\ i,j=1}}^{N} \sigma_i \sigma_j J_{ij} + \gamma \log(N) \sum_{j=1}^{N} \sigma_j^2$$

is associated to the dynamics of the present model. It is possible to prove the existence of energy valleys of this Hamiltonian. However, the proof is very similar to the proof of Theorem 2.15 and does not lead to an increased capacity by a less restrictive notion of the capacity. The radii of the valleys are of order $\log(N)$, the depths of order $\log(N)^2$.

4 The Willshaw Model

The model to which this chapter is devoted was proposed in a paper by Willshaw (see [44]). As in chapters 2 and 3, the neuron set is $V = \{1, \dots, N\}$ for some $N \in \mathbb{N}$. The set of edges is enlarged by adding the self-loops $\{1\}, \dots \{N\}$. The model uses, as Amari's model, the configuration space $\{0,1\}^N$: the neurons can be active or inactive. The major difference to Amari's model is that also the synaptic weights are binary. A number of M patterns $\xi^1, \dots, \xi^M$ is stored; the information of their spins is used to built the synaptic weights. The weight J_{ij} does not depend on the number of messages in which neurons i and j are activated together; it rather indicates whether there is any μ with $\xi_i^\mu \xi_j^\mu = 1$ (that is to say, whether there is any message that contains both neurons) or not. Even though there is more information contained in Amari's model, Willshaw's model offers better storage capacities than Amari's model. This can be understood considering the fact that the additional information contained in Amari's model does not lead to any improvement of the performance but only favours errors: if one stabs the dynamics in a stored message, the neurons that are not excited in this message can accumulate more signals than in the Willshaw model. The synaptic efficacy J_{ij} counts the number of messages in which neurons i and j are both activated and the local fields can easier become too large. This is prevented in the Willshaw model, whereas the important information to recognise a stored pattern, i.e., the information that the participating neurons are contained in the same message, is still contained in the synaptic efficacy. All the neurons belonging to a stored message are interconnected which means that the synaptic efficacies among all these neurons are 1. Here it is important that the model is sparse and there are only a few synaptic weights that are 1.

As in the previous chapters, we can use two possible settings for the model. For the first one we assume that the (ξ_i^μ) are i.i.d $0-1$ random variables with success probability $p = \frac{\log N}{N}$. For the second version we choose the M messages uniformly at random from all sets of M messages with exactly $c = \log N$ active neurons. We begin with the analysis of the model with the first setting, because in this case the messages as well as all their spins are independent.

In the Willshaw model, the synaptic efficacies are defined by

$$J_{ij} = \Theta\left(\sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu - 1\right) = \begin{cases} 1 & \exists \mu : \xi_i^\mu \xi_j^\mu = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Θ is the Heaviside function, with $\Theta(x) = 1$ if $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$. Especially J_{ii} is defined in this model and it is 1 if there is at least one message that contains neuron i. The model benefits from admitting a self-signal, which means that we consider the local field $\bar{S}_i(\sigma) = \sum_{j=1}^{N} J_{ij}\sigma_j$ instead of $S_i(\sigma) = \sum_{j \neq i} J_{ij}\sigma_j$; we will see that it improves performance to modify S_i in this way. This modification is called "memory effect". Edges whose synaptic weight is 1 are denoted as active; we sometimes call an active edge a connection (even though, strictly speaking, each pair of neurons is connected in the graph, but possibly by an inactive edge that cannot transmit signals).

There are two different types of dynamics to be considered. The first one, the threshold dynamics comparable to the dynamics used e. g., in Amari's model, can be applied to both versions. Formally, we use a parallel dynamics whose i-th component sets for an input $\sigma \in \{0,1\}^N$

$$T_i(\sigma) = \Theta\left(\bar{S}_i(\sigma) - h\right)$$

with a fixed threshold $h = \gamma \log N$, for some $\gamma \in (0,1]$.

The second setting of the model also allows for the application of another retrieval dynamics. It keeps the neurons with the highest values of the local field $\bar{S}_i(\sigma)$ active. Formally, let the ordered values of $\bar{S}_i$, $1 \leq i \leq N$, be denoted by $\bar{S}_{(1)} \geq \bar{S}_{(2)} \geq \ldots \geq \bar{S}_{(c)} \geq \ldots \geq \bar{S}_{(N)}$. A neuron i is activated if and only if $\bar{S}_i(\sigma) \geq \bar{S}_{(c)}$. This procedure is called "Winner takes all"-algorithm (WTA algorithm, for short). The above dynamics is not appropriate for the first version of the model, because a stored pattern might consist of less than c active neurons. The WTA dynamics activates by construction at least c neurons and the stored pattern would definitely not be stable.

Finally, there is a variation of this second dynamics that is also applicable to the first setting where the ξ_j^μ, $\mu \leq M$, $j \leq N$, are independent. In this variation only the most active neurons are activated, that is, all neurons with a value $\bar{S}_i(\sigma) = \bar{S}_{(1)}$; all other neurons are deactivated. Compared to the WTA dynamics using $\bar{S}_{(c)}$, this dynamics using $\bar{S}_{(1)}$ does not change the first step (and therefore the behaviour of the one-step retrieval) if the input is a partially erased version $\tilde{\xi}^\mu$ of a stored message ξ^μ or a stored message. This is only true if $\bar{S}$ is used. The variation using $\bar{S}_{(1)}$ is also applicable if in the second setting the fixed number of neurons per message is unknown.

The first section deals with the convergence of the three different dynamics of this model; the second one contains an analysis of the stability and error correcting behaviour as well as a short view on the energy landscape of the Willshaw model using the threshold dynamics. In the third section, we analyse the Willshaw model with WTA algorithm and show that it offers the best behaviour of the binary models. In the whole chapter, we use the notation of A_δ, $\mathcal{Z}_k$ and $\bar{\mathcal{Z}}_k$ used in Chapter 2.

4.1 Dynamical Properties of the Willshaw Model

As described in the introduction of this chapter, we consider three different dynamics for the Willshaw model:

1. the threshold dynamics using the fixed threshold $h = \gamma \log(N)$,
2. the WTA dynamics using the variable threshold $h = \bar{S}_{(1)}$,
3. the WTA dynamics using the variable threshold $h = \bar{S}_{(c)}$.

The first version uses a fixed threshold, whereas the second and third work with variable thresholds that are determined in each step of the dynamics. Note that in all cases we consider the memory effect, i.e., use $\bar{S}$. We consider, in all three cases, a parallel dynamics.

In this section we show the following results:

1. choosing a fixed threshold h enforces convergence of the dynamics, if the input is a partially deleted stored pattern or a stored pattern,

2. choosing the threshold $h = \bar{S}_{(1)}$, the performance does not benefit from iterating more than once,

3. choosing a variable $h = \bar{S}_{(c)}$ can lead to oscillation effects in the dynamics.

Theorem 4.1 *The dynamics of the Willshaw model with threshold dynamics (with a fixed threshold h) definitely converges if the input pattern is a partially deleted version of a stored pattern ξ^μ or a stored pattern.*

Proof of Theorem 4.1: Let $\tilde{\xi}^\mu$ be a partially erased version of ξ^μ. We denote by k_μ the number of activated neurons of ξ^μ. If $k_\mu < h$, none of the neurons gets a signal that is higher than k_μ which means that $\tilde{\xi}^\mu$ is turned into $(0, \ldots, 0)$ in one step of the parallel dynamics. One step of the dynamics here means the updating of all neurons, so the pattern after one step is $(T_1(\tilde{\xi}^\mu), \ldots, T_N(\tilde{\xi}^\mu))$.

If $k_\mu \geq h$, we introduce the sequence

$$\begin{aligned} \tilde{\xi}^\mu(0) &:= \tilde{\xi}^\mu \\ \tilde{\xi}^\mu(t+1) &:= T\left(\tilde{\xi}^\mu(t)\right) = \left((T_1(\tilde{\xi}^\mu(t)), \ldots, T_N(\tilde{\xi}^\mu(t))\right) \qquad t \in \mathbb{N}, \end{aligned}$$

and the sequence $(a^\mu(t))_{t\geq 0}$,

$$a^\mu(t) = \{i \in \{1, \ldots, N\} : \tilde{\xi}^\mu_i(t) = 1\}, \qquad t \in \mathbb{N}_0.$$

We will show the following proposition:

Proposition 4.2 *If $h \leq k_\mu$, the sequence $(a^\mu(t))_{t\geq 0}$ is increasing with respect to inclusion.*

Proof: We prove the assertion by induction. Firstly, we observe that $a^\mu(0) \subseteq a^\mu(1)$: each activated neuron is part of the stored message ξ^μ and thus

$$\forall i, j \in a^\mu(0) : J_{ij} = 1$$

which implies $\bar{S}_i(\tilde{\xi}^\mu) = k_\mu \geq h$ for all $i \in a^\mu(0)$. This proves the base clause.

Then suppose that $a^\mu(t) \subseteq a^\mu(t+1)$ holds for some arbitrary $t \in \mathbb{N}$. By definition of the dynamics, each $i \in a^\mu(t+1)$ gets at least h signals. So we have, for fixed i, $|\{j \in a^\mu(t) : J_{ij} = 1\}| \geq h$. According to the induction hypothesis, $a^\mu(t) \subseteq a^\mu(t+1)$ and therefore $|\{j \in a^\mu(t+1) : J_{ij} = 1\}| \geq h$. This implies that $i \in a^\mu(t+2)$ and that $a^\mu(t+1) \subseteq a^\mu(t+2)$. □

Proposition 4.2 implies that $(a^\mu(t))_{t\geq 0}$ converges.

□

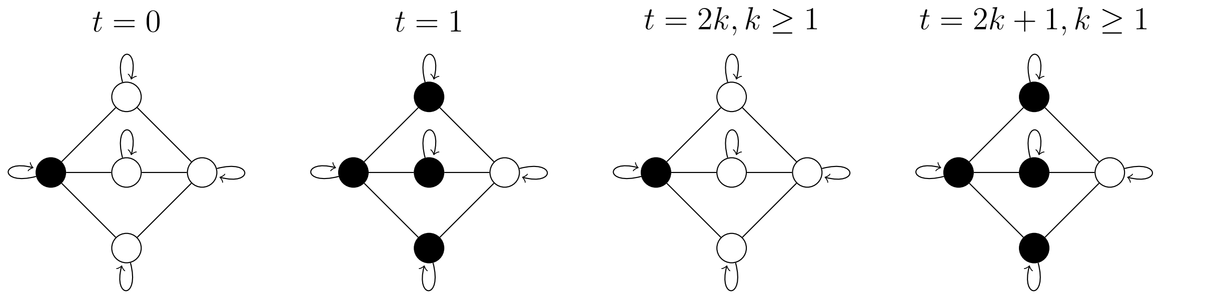

Figure 4.1: The oscillation of the WTA dynamics in the Willshaw model using $h = \bar{S}_{(1)}$. The model contains $N = 5$ neurons and the number of activated neurons in a stored message is $c = 2$. Active edges are visualised.

Theorem 4.3 *Choosing a variable h, that is, one of the WTA dynamics, can lead to oscillation effects, even if the input is a partially erased stored pattern.*

Proof: To see this, we propose an example where $N = 5$ and $c = 2$. We firstly choose the threshold $h = \bar{S}_{(1)}$. Let us consider the stored messages

$$(\xi^\mu)_{1\leq\mu\leq 6} = \left(\begin{pmatrix} 1\\1\\0\\0\\0 \end{pmatrix} \begin{pmatrix} 1\\0\\1\\0\\0 \end{pmatrix} \begin{pmatrix} 1\\0\\0\\1\\0 \end{pmatrix} \begin{pmatrix} 0\\1\\0\\0\\1 \end{pmatrix} \begin{pmatrix} 0\\0\\1\\0\\1 \end{pmatrix} \begin{pmatrix} 0\\0\\0\\1\\1 \end{pmatrix} \right).$$

Let the input pattern be the partially erased pattern

$$\tilde{\xi}^1(0) = \begin{pmatrix} 1\\0\\0\\0\\0 \end{pmatrix}.$$

Then, due to the memory effect, the subsequent two outputs of the dynamics are

$$\left(\tilde{\xi}^1(t)\right)_{0\leq t\leq 2} = \left(\begin{pmatrix} 1\\0\\0\\0\\0 \end{pmatrix} \begin{pmatrix} 1\\1\\1\\1\\0 \end{pmatrix} \begin{pmatrix} 1\\0\\0\\0\\0 \end{pmatrix} \right)$$

and thus $\tilde{\xi}^1(2t+1) = \tilde{\xi}^1(1)$, $\tilde{\xi}^1(2t) = \tilde{\xi}^1(2)$ for each $t \in \mathbb{N}$.

If we use the variable threshold $h = \bar{S}_{(c)}$ instead, in our example $\bar{S}_{(2)}$, this will also lead to an oscillation of the output patterns:

$$\left(\tilde{\xi}^1(t)\right)_{0\leq t\leq 4} = \left(\begin{pmatrix} 1\\0\\0\\0\\0 \end{pmatrix} \begin{pmatrix} 1\\1\\1\\1\\0 \end{pmatrix} \begin{pmatrix} 1\\0\\0\\0\\1 \end{pmatrix} \begin{pmatrix} 0\\1\\1\\1\\0 \end{pmatrix} \begin{pmatrix} 1\\0\\0\\0\\1 \end{pmatrix} \right),$$

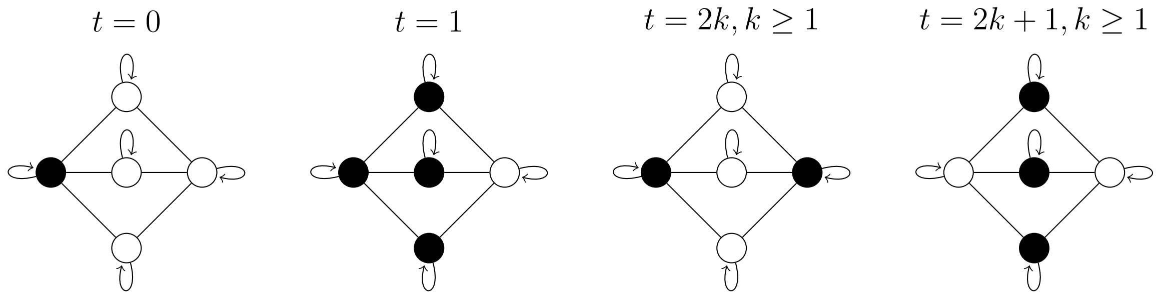

Figure 4.2: The oscillation of the WTA dynamics in the Willshaw model using $h = \bar{S}_{(2)}$. The model contains $N = 5$ neurons and the number of activated neurons in a stored message is $c = 2$.

here we have $\tilde{\xi}^1(2t) = \tilde{\xi}^1(2)$ and $\tilde{\xi}^1(2t+1) = \tilde{\xi}^1(3)$ for each $t \in \mathbb{N}$.

In the figures, only the active edges are drawn, even though actually all edges exist in the graph, but some of them are inactive and cannot transmit signals.

□

Interestingly, the WTA dynamics using $h = \bar{S}_{(1)}$ does not benefit from further iterations, if the input is a stored or a partially erased stored pattern.

Theorem 4.4 *Consider the Willshaw model where the threshold is chosen as $h = \bar{S}_{(1)}$. Choose as input a partially erased version $\tilde{\xi}^\mu \neq (0, \dots, 0)$ of a stored message ξ^μ or the stored pattern itself. Then the dynamics converges if and only if it converges in one step. In particular, it can only converge to ξ^μ if it does so in one iteration.*

Proof: We use the same notations as in the proof of Theorem 4.1. We denote by $h(t)$ the value of the threshold at step t. There are two cases to consider:

1. After the first iteration, the activated neurons in $\tilde{\xi}^\mu(1)$ are completely interconnected (in the sense that the synaptic weights corresponding to their connecting edges are all 1). Due to the memory effect, we have

   $$h(1) = \sum_{i=1}^{N} \tilde{\xi}_i^\mu(1).$$

   Since $h(0)$ is at most the number of activated neurons in the input pattern that is a partially erased version of a stored pattern and thus the edges among its activated neurons are all active, we observe that

   $$h(0) = \sum_{i=1}^{N} \tilde{\xi}_i^\mu.$$

   The neurons activated in $\tilde{\xi}^\mu(1)$ are thus exactly those neurons that are connected by active edges to each initially activated neuron. In particular, $a^\mu(0) \subseteq a^\mu(1)$.

The local field of the inactive neurons in $\tilde{\xi}^\mu(1)$ is thus smaller than $h(1)$ because they have not been activated by the first step of the dynamics and did not receive a signal from each of the initially activated neurons. Since additionally the threshold $h(1)$ is reached by each active neuron in $\tilde{\xi}^\mu(1)$, we observe $\tilde{\xi}^\mu(t) = \tilde{\xi}^\mu(1)$ for $t \geq 2$.

2. There are i' and j' such that $\tilde{\xi}^\mu_{i'}(1) = \tilde{\xi}^\mu_{j'}(1) = 1$ but $J_{i'j'} = 0$, i.e., after the first step, there are activated neurons whose connecting edge is inactive. In this case $\tilde{\xi}^\mu(1) \neq \xi^\mu$ because the edges between neurons of a stored pattern are active.

   We fix such a pair i' and j'. As we saw in the first part,

$$h(0) = \sum_{i=1}^{N} \tilde{\xi}^\mu_i$$

   and the neurons activated in $\tilde{\xi}^\mu(1)$ are those neurons that are connected by an active edge to each initially activated neuron. Consequently, $a^\mu(0) \subseteq a^\mu(1)$ and the activated neurons of ξ^μ are contained in $a^\mu(1)$.

   Since at least all neurons activated at step 0 are connected by active edges to all neurons activated at step 1, we obtain

$$h(1) = \sum_{i=1}^{N} \tilde{\xi}^\mu_i(1).$$

   Consequently, $\tilde{\xi}^\mu_i(2) = 0$ for each i that is not connected by an active edge to every other neuron of $a^\mu(1)$, at least for the two neurons i' and j'. This implies $\tilde{\xi}^\mu(1) \neq \tilde{\xi}^\mu(2)$. Since $h(1) = \sum_{i=1}^{N} \tilde{\xi}^\mu_i(1)$, the neurons activated in $\tilde{\xi}^\mu(2)$ are exactly those that receive signals from all neurons in $\tilde{\xi}^\mu(1)$. In particular the initially activated neurons are connected by active edges to all activated neurons in $\tilde{\xi}^\mu(1)$ and we obtain $a^\mu(0) \subseteq a^\mu(2)$. In addition, we saw that $a^\mu(0) \subseteq a^\mu(1)$. So each one of the activated neurons in $\tilde{\xi}^\mu(2)$ is also activated in $\tilde{\xi}^\mu(1)$ because it receives signals from each activated neuron in $\tilde{\xi}^\mu(1)$ and therefore also by each activated neuron in $\tilde{\xi}^\mu(0)$; this implies $a^\mu(2) \subseteq a^\mu(1)$.

   The threshold $h(2)$ is at most $\sum_{i=1}^{N} \tilde{\xi}^\mu_i(2)$. Since $a^\mu(2) \subseteq a^\mu(1)$ and each initially activated neuron is actively connected to each neuron of $\tilde{\xi}^\mu(1)$, hence also to each neuron of $\tilde{\xi}^\mu(2)$, it collects $\sum_{i=1}^{N} \tilde{\xi}^\mu_i(2)$ signals and we deduce

$$h(2) = \sum_{i=1}^{N} \tilde{\xi}^\mu_i(2).$$

   The neurons activated at step 3 are thus the neurons that get signals from each activated neuron of $\tilde{\xi}^\mu(2)$. But the activated neurons of $\tilde{\xi}^\mu(2)$ are connected by active edges to each activated neuron of $\tilde{\xi}^\mu(1)$, so $a^\mu(1) \subseteq a^\mu(3)$. Since $a^\mu(0) \subseteq a^\mu(2)$, each neuron connected to all neurons of $\tilde{\xi}^\mu(2)$ and therefore activated in $\tilde{\xi}^\mu(3)$ is connected to every $i \in a^\mu(0)$ and thus also activated in $\tilde{\xi}^\mu(1)$. We conclude $a^\mu(3) \subseteq a^\mu(1)$ and $\tilde{\xi}^\mu(1) = \tilde{\xi}^\mu(3)$. This yields

$$\tilde{\xi}^\mu(2t+1) = \tilde{\xi}^\mu(1) \qquad \text{and} \qquad \tilde{\xi}^\mu(2t) = \tilde{\xi}^\mu(2), \quad t \in \mathbb{N}.$$

We observed that $\tilde{\xi}^\mu(1) \neq \tilde{\xi}^\mu(2)$. The dynamics does thus not converge.

□

Remark 4.5 *In the Willshaw model with WTA dynamics using $\bar{S}_{(1)}$, the dynamics started in a partially erased stored pattern either converges in one step or does not converge. For the Hopfield model, Burshtein in [12] showed that a correction of a pattern in a certain Hamming distance ϱN of a stored message is corrected in at most q steps, q depending on ϱ and on the number of stored patterns. In particular, he showed that there is a distance such that the dynamics requires exactly 2 steps to correct the pattern. Multiple steps are indeed important in the Hopfield model, but do not to lead to an improved performance in the considered case.*

4.2 The Willshaw Model with Threshold Dynamics

The Willshaw model with threshold dynamics as described in the introduction offers higher bounds on α (in dependence on γ) for the capacity and also a better error correcting behaviour than Amari's model.

4.2.1 Capacity and Error Correction

A first approach to obtain a lower bound on the capacity of the Willshaw model can be made by using the results of the second chapter.

Proposition 4.6 *Suppose that in the Willshaw model with i.i.d. random variables $(\xi_j^\mu, \mu \leq M, j \leq N)$, the threshold dynamics with threshold $h = \gamma \log(N)$, $0 < \gamma < 1$, is used and $M = \alpha N^2/(\log N)^2$ patterns are stored. For an arbitrary message ξ^μ, we have*

$$\lim_{N\to\infty} \mathbb{P}(\forall i \leq N : T_i(\xi^\mu) = \xi_i^\mu) = 1$$

if α fulfills the two conditions

$$\alpha < \gamma$$

and

$$-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha < -1.$$

Proof: The statement of the proposition is a direct consequence of Theorem 2.1. The only fact one has to observe is that the synaptic efficacies among the activated neurons of ξ^1 are 1 and consequently the dynamics keeps these neurons activated if the message consists of enough activated neurons. Formally, if $\xi_i^1 = 1$, the local field is

$$\bar{S}_i(\xi^1) = \sum_{j=1}^N \xi_j^1 J_{ij} = \sum_{j=1}^N \xi_j^1,$$

because $J_{ij} = 1$ for all i, j with $\xi_j^1 = \xi_i^1 = 1$. Since $\gamma < 1$, the probability $\mathbb{P}[\sum_{j=1}^N \xi_j^1 \geq \gamma \log(N)]$ tends to 1.

On the other hand, for each i with $\xi_i^1 = 0$, the probability of turning ξ_i^1 into a 1 by the dynamics is given by

$$\mathbb{P}\left(T_i(\xi^1) = 0\right) = \mathbb{P}\left(\sum_{j=1}^{N} \xi_j^1 J_{ij} \geq \gamma \log(N)\right).$$

But the local field in Willshaw's model is bounded by the one of Amari's model: denoting by $\bar{S}^{\mathrm{Wi}}$ and S^{Am} the local fields and by J_{ij}^{Wi} and J_{ij}^{Am} the synaptic efficacies defined in the corresponding models, we observe

$$\bar{S}_i^{\mathrm{Wi}}(\xi^1) = \sum_{j=1}^{N} \xi_j^1 J_{ij}^{\mathrm{Wi}} = \sum_{j\neq i}^{N} \xi_j^1 J_{ij}^{\mathrm{Wi}} = \sum_{j\neq i}^{N} \xi_j^1 \,\Theta\left(\sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu - 1\right)$$
$$\leq \sum_{j\neq i}^{N} \xi_j^1 \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu = \sum_{j\neq i}^{N} \xi_j^1 J_{ij}^{\mathrm{Am}} = S_i^{\mathrm{Am}}(\xi^1).$$

So

$$\mathbb{P}\left(\exists i \leq N : \xi_i^1 = 0, T_i^{\mathrm{Wi}}(\xi^1) \neq 0\right) \leq \mathbb{P}\left(\exists i \leq N : \xi_i^1 = 0, T_i^{\mathrm{Am}}(\xi^1) \neq 0\right)$$

with the dynamics T^{Am} of Amari's model and T^{Wi} of the Willshaw model. Thus the lower bound on α obtained for Amari's model is also a lower bound on α for the Willshaw model with threshold dynamics.

□

Remark 4.7 *In the Willshaw model, it is not possible to use threshold variables $\gamma > 1$ because the synaptic efficacies are binary: the $\bar{S}_i$ are bounded by the number of activated neurons per message. If the messages are chosen uniformly from the set of patterns with exactly c non-zero spins, a threshold γc with $\gamma > 1$ implies convergence to $(0, \ldots, 0)$ of every stored pattern in the first step. If the spins are independent and identically Bernoulli distributed with parameter p, the probability of A_δ^c vanishes for each $\delta > 0$, especially for $0 < \delta < \gamma - 1$, and an arbitrary stored pattern is turned to $(0, \ldots, 0)$ in the first step, with probability converging to 1. If in this model $\gamma = 1$, the probability of turning ξ^μ into $(0, \ldots, 0)$ tends to $1/2$, whereas $\gamma = 1$ is an admissible threshold variable in the version with exactly c activated neurons per message. Depending on γ, the (preliminary) lower bounds on the storage capacity obtained in Proposition 4.6 are thus smaller as the ones obtained for Amari's model because there is no option to increase the value γ to reach a higher capacity. As we will see, Proposition 4.6 underestimates the real storage capacity of Willshaw's model that will be revealed to perform better than Amari's model, even by only using threshold variables $\gamma < 1$.*

For the Willshaw model with threshold dynamics and i.i.d. distributed spins ξ_j^μ, we try to improve the results obtained in Proposition 4.6 by dealing with the characteristical properties of the dynamics. The local field of a neuron counts the number of activated neurons to which it is connected (their common synaptic efficacy is positive). Fix a stored message, e. g., ξ^1, and some neuron i. To examine the behaviour of the dynamics, we will determine the distribution of the number of activated neurons in ξ^1 to which a neuron i is connected.

Lemma 4.8 *In the Willshaw model with i.i.d. spins ξ_j^μ, $\mu \leq M$, $j \leq N$, suppose that one of the messages, e. g., ξ^1, consists of k 1's and that $k \leq (1+\delta)\log(N)$ for some $\delta > 0$. Let $i \in \{1,\ldots,N\}$ be chosen such that $\xi_i^1 = 0$. The distribution of the number of active neurons in ξ^1 to which neuron i is connected is asymptotically Binomially distributed with parameters k and $\tilde{p} := 1 - e^{-\alpha}$. As $N \to \infty$, we have for $m \in \{0,\ldots,k\}$:*

$$\mathbb{P}\left(\sum_{j=1}^{N} \xi_j^1 J_{ij} = m \Big| \bar{\mathcal{Z}}_k\right) = \binom{k}{m}(1-e^{-\alpha})^m e^{-\alpha(k-m)}\left[1+o(1)\right].$$

For each i with $\xi_i^1 = 1$, we have $\sum_{j=1}^{N} \xi_j^1 J_{ij} = k$.

Proof of Lemma 4.8: Without loss of generality, we assume that the activated neurons are exactly the first k ones. Let $m \in \{0,\ldots,k\}$ and $i > k$. In addition, let $\bar{J}_{ij}$ denote the synaptic weight that only takes into account the messages $\xi^2,\ldots,\xi^M$:

$$\bar{J}_{ij} := \Theta\left(\sum_{\mu=2}^{M} \xi_i^\mu \xi_j^\mu - 1\right).$$

For $i > k$, define $\mathcal{X}(i) := \{\mu \geq 2 : \xi_i^\mu = 1\}$. The probability of interest is transformed into

$$\mathbb{P}\left[\sum_{j=1}^{N} \xi_j^1 J_{ij} = m \Big| \mathcal{Z}_k\right] = \mathbb{P}\left[\sum_{j=1}^{k} \bar{J}_{ij} = m\right]$$
$$= \binom{k}{m}\mathbb{P}\left[\forall j \leq m : \bar{J}_{ij} = 1, \forall j \in \{m+1,\ldots,k\} : \bar{J}_{ij} = 0\right]$$
$$= \binom{k}{m}\sum_{B\subseteq\{2,\ldots,M\}} \mathbb{P}\left[\mathcal{X}(i) = B\right]\mathbb{P}\left[\forall j \leq m : \bar{J}_{ij} = 1, \forall j > m, j \leq k : \bar{J}_{ij} = 0 \Big| \mathcal{X}(i) = B\right].$$

The distribution of the size of $\mathcal{X}(i)$ is Binomial with parameters $M-1$ and p. The probability

$$\mathbb{P}\left[\forall j \leq m : \bar{J}_{ij} = 1, \forall j > m, j \leq k : \bar{J}_{ij} = 0 \Big| \mathcal{X}(i) = B\right]$$

only depends on $|B|$. It is, for some arbitrary set $B \subseteq \{2,\ldots,M\}$ with $|B| = n$, equal to

$$\mathbb{P}\left[\forall j \leq m : \bar{J}_{ij} = 1, \forall j > m, j \leq k : \bar{J}_{ij} = 0 \Big| \mathcal{X}(i) = B\right] = \left[1-(1-p)^n\right]^m (1-p)^{n(k-m)}.$$

Consequently,

$$\sum_{B\subseteq\{2,\ldots,M\}} \mathbb{P}\left[\mathcal{X}(i) = B\right]\mathbb{P}\left[\forall j \leq m : \bar{J}_{ij} = 1, \forall j > m, j \leq k : \bar{J}_{ij} = 0 \Big| \mathcal{X}(i) = B\right]$$
$$= \sum_{n=0}^{M-1}\binom{M-1}{n} p^n (1-p)^{M-1-n}\left[1-(1-p)^n\right]^m (1-p)^{n(k-m)}$$
$$= \sum_{n=0}^{M-1}\binom{M-1}{n} p^n (1-p)^{M-1-n}(1-p)^{n(k-m)}\sum_{l=0}^{m}\binom{m}{l}(-1)^l(1-p)^{nl}$$

$$= \sum_{l=0}^{m} \binom{m}{l} (-1)^l \left[p(1-p)^{k-m+l} + 1 - p\right]^{M-1}.$$

We just applied the Binomial formula, the definition of the Binomial distribution and our observations until now. The next step takes into account that $k = \mathcal{O}(\log(N))$, $p = \log(N)/N$, the series representation of the logarithm (at argument $1+x$, for $|x| < 1$) and finally the size of M:

$$\begin{aligned}
&\left[p(1-p)^{k-m+l} + 1 - p\right]^{M-1} = \left[p\left(1 - p(k-m+l) + \mathcal{O}(p^2(k-m+l)^2\right) + 1 - p\right]^{M-1} \\
&= \left[1 - p^2(k-m+l) + \mathcal{O}\big(p^3(k-m+l)^2\big)\right]^{M-1} \\
&= \exp\left[(M-1)\log\left(1 - p^2(k-m+l) + \mathcal{O}\left(p^3(k-m+l)^2\right)\right)\right] \\
&= \exp\left[(M-1)\left(-p^2(k-m+l) + \mathcal{O}\left(p^3(k-m+l)^2\right)\right)\right] \\
&= \exp\left[-\alpha(k-m+l) + \mathcal{O}\left(p(k-m+l)^2\right)\right] \\
&= \exp\left[-\alpha(k-m+l)\right]\left(1 + \mathcal{O}\left(p(k-m+l)^2\right)\right).
\end{aligned}$$

This result is used to determine

$$\begin{aligned}
&\binom{k}{m} \sum_{l=0}^{m} \binom{m}{l} (-1)^l \left[p(1-p)^{k-m+l} + 1 - p\right]^{M-1} \\
&= \binom{k}{m} \sum_{l=0}^{m} \binom{m}{l} (-1)^l \exp\left[-\alpha(k-m+l)\right] \left(1 + \mathcal{O}\left(\frac{\log(N)}{N}(k-m+l)^2\right)\right) \\
&= \binom{k}{m} e^{-\alpha(k-m)} \left(1 - e^{-\alpha}\right)^m (1 + o(1)).
\end{aligned}$$

The second statement of the Lemma follows immediately because the synaptic efficacies among the active neurons of ξ^1 are 1.

□

With this information we can now improve the results of Proposition 4.6.

Theorem 4.9 *In the Willshaw model with i.i.d. random variables $\xi_j^\mu, \mu \geq 2, j \leq N$, threshold dynamics using the threshold $h = \gamma \log(N)$, $\gamma \in (0,1)$, and number of stored patterns $M = \alpha N^2/(\log N)^2$, an arbitrary but fixed pattern ξ^μ is stable with high probability, i.e.,*

$$\lim_{N\to\infty} \mathbb{P}\left(\forall i \leq N : T_i(\xi^\mu) = \xi_i^\mu\right) = 1 \tag{4.1}$$

if α satisfies

$$\alpha < -\log(1-\gamma)$$

and additionally

$$-\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right) < -1. \tag{4.2}$$

This bound is sharp: for each α with 1.) $\alpha < -\log(1-\gamma)$ and

$$-\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right) > -1 \tag{4.3}$$

or 2.) $\alpha \geq -\log(1-\gamma)$, an arbitrary stored message is not stable with probability converging to 1:

$$\lim_{N\to\infty} \mathbb{P}(\exists i \leq N : T_i(\xi^\mu) \neq \xi_i^\mu) = 1. \tag{4.4}$$

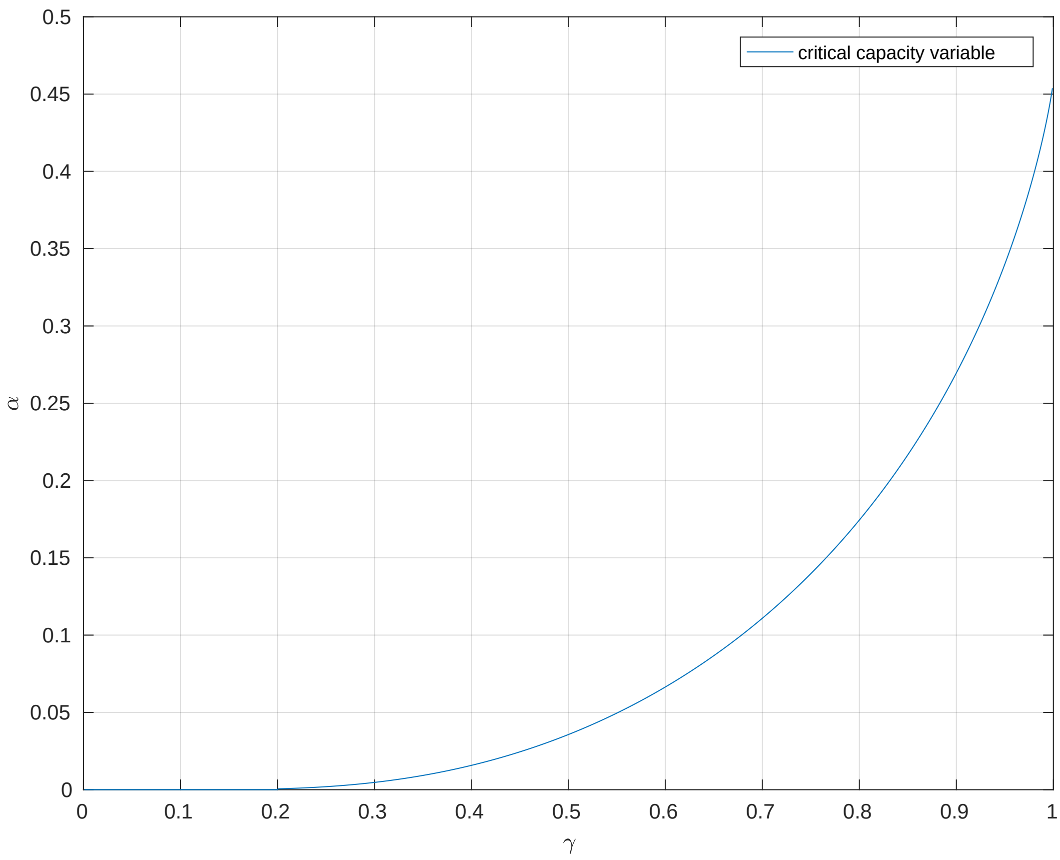

Figure 4.3: Critical capacity variable $\alpha^*(\gamma)$ in dependence on the threshold variable γ for the Willshaw model with threshold dynamics

Proof of Theorem 4.9: We consider an arbitrary μ, e. g., $\mu = 1$. The event A_δ denotes again $A_\delta = \left\{ \sum_{j=1}^N \xi_j^1 \in ((1-\delta)\log(N), (1+\delta)\log(N))\right\}$. The probability of A_δ tends to 1. In addition, the event $\{\sum_{j=1}^N \xi_j^1 = \sum_{j=1}^k \xi_j^1 = k\}$ is denoted by $\mathcal{Z}_k$.

We begin the proof by showing that the active neurons of ξ^1 are stable with high probability. To see that they receive with high probability a signal that is sufficiently strong, we consider some $\delta < 1-\gamma$ and obtain for i with $\xi_i^1 = 1$:

$$\mathbb{P}\left(T_i(\xi^1) = 1\right) = \mathbb{P}\left[\sum_{j=1}^N \xi_j^1 J_{ij} \geq \gamma \log(N)\right]$$

$$=\mathbb{P}\left[\sum_{j=1}^{N}\xi_j^1 \geq \gamma\log(N)\right] \geq \mathbb{P}(A_\delta),$$

which tends to 1. This holds in particular independently of the choice of α.

We continue with the behaviour of the inactive neurons of ξ^1. Firstly,

$$\begin{aligned}
&\mathbb{P}\left(\exists i \leq N : \xi_i^1 = 0, T_i(\xi^1) \neq \xi_i^1\right)\\
\leq&\mathbb{P}(A_\delta)\max_{k\in\mathbb{N}:k/\log(N)\in(1-\delta,1+\delta)}\mathbb{P}\left(\exists i \leq N : \xi_i^1 = 0, T_i(\xi^1) \neq \xi_i^1|\mathcal{Z}_k\right) + \mathbb{P}(A_\delta^c)\\
\leq&\max_{k\in\mathbb{N}:k/\log(N)\in(1-\delta,1+\delta)}\mathbb{P}\left(\exists i > k : T_i(\xi^1) \neq 0|\mathcal{Z}_k\right) + \mathbb{P}(A_\delta^c).
\end{aligned}$$

In Lemma 4.8 we showed that the asymptotic distribution of $\sum_{j=1}^{k}\bar{J}_{ij}$ for some fixed $i > k$ is Binomial with parameters k and $\tilde{p} = 1 - e^{-\alpha}$. Let a random variable with Binomial distribution with parameters k and $\tilde{p}$ be denoted by $R_{k,\tilde{p}}$. The conditional probability of the last line is then bounded by

$$\begin{aligned}
&\mathbb{P}\left(\exists i > k : T_i(\xi^1) \neq 0|\mathcal{Z}_k\right) \leq (N-k)\mathbb{P}\left(T_N(\xi^1) \neq 0|\mathcal{Z}_k\right)\\
=&(N-k)\mathbb{P}\left[\sum_{j=1}^{k}\bar{J}_{Nj} \geq \gamma\log(N)\right]\\
\leq&(N-k)\mathbb{P}\left[R_{k,\tilde{p}} \geq \gamma\log(N)\right]\cdot(1+o(1)).
\end{aligned}$$

The probability $\mathbb{P}\left[R_{k,\tilde{p}} \geq \gamma\log(N)\right]$ is maximal for the maximal value of k, $k = \lfloor(1+\delta)\log(N)\rfloor$, which is without loss of generality assumed to be a natural number. The probability is bounded with the help of the exponential Chebyshev inequality:

$$\begin{aligned}
&(N-k)\,\mathbb{P}\left[R_{k,\tilde{p}} \geq \gamma\log(N)\right]\cdot(1+o(1))\\
\leq&N\,\mathbb{P}\left[R_{(1+\delta)\log(N),\tilde{p}} \geq \gamma\log(N)\right]\cdot(1+o(1))\\
\leq&N\,\exp\left[-t\gamma\log(N)\right]\left[1-(1-e^{-\alpha})+(1-e^{-\alpha})e^t\right]^{(1+\delta)\log(N)}(1+o(1)).
\end{aligned}$$

Minimising in t yields

$$t^*_{\gamma,\delta,\alpha} = \log\left(\frac{\gamma e^{-\alpha}}{(1+\delta-\gamma)(1-e^{-\alpha})}\right)$$

as minimal argument; $t^*_{\gamma,\delta,\alpha}$ is positive, if $\alpha < -\log(1-\gamma/(1+\delta))$. We assumed that $\alpha < -\log(1-\gamma)$ and can thus choose an appropriate δ in the beginning of the proof.

Finally the condition

$$-\gamma\log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1+\delta-\gamma)\log\left(\frac{e^{-\alpha}}{1+\delta-\gamma}\right) + (1+\delta)\log(1+\delta) < -1$$

should be fulfilled in order to enforce the probability to converge to 0. Obviously,

$$\begin{aligned}
&\lim_{\delta\searrow 0} -\gamma\log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1+\delta-\gamma)\log\left(\frac{e^{-\alpha}}{1+\delta-\gamma}\right) + (1+\delta)\log(1+\delta)\\
&= -\gamma\log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right).
\end{aligned}$$

So, if the condition

$$-\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right) < -1$$

is fulfilled, as in the formulation of the theorem, δ can be chosen appropriately small to obtain

$$-\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1+\delta-\gamma)\log\left(\frac{e^{-\alpha}}{1+\delta-\gamma}\right) + (1+\delta)\log(1+\delta) < -1.$$

Thus we have

$$\lim_{N\to\infty} \mathbb{P}\left(\exists i \le N : T_i(\xi^1) \neq \xi_i^1\right) = 0$$

for this choice of α. This proves the first part of the theorem.

For the second part of the proof, we first assume that $\alpha < -\log(1-\gamma)$ and that (4.3) holds. The probability of an error is bounded by

$$\begin{aligned}\mathbb{P}\left(\exists i \le N : T_i(\xi^1) \neq \xi_i^1\right) \ge& \mathbb{P}\left(\exists i \le N : \xi_i^1 = 0, T_i(\xi^1) \neq 0\right)\\ \ge& \mathbb{P}(A_\delta) \min_{k\in\mathbb{N}: k/\log(N)\in(1-\delta,1+\delta)} \mathbb{P}\left(\exists i > k : T_i(\xi^1) \neq 0 | \mathcal{Z}_k\right)\\ \ge& \mathbb{P}(A_\delta) \min_{k\in\mathbb{N}: k/\log(N)\in(1-\delta,1+\delta)} \left[1 - \mathbb{P}\left(\forall i > k : T_i(\xi^1) = 0 | \mathcal{Z}_k\right)\right].\end{aligned} \tag{4.5}$$

For an arbitrary realisation $(x_j^\mu)_{\mu\ge 2, j\le k} \in \{0,1\}^{k(M-1)}$ of $(\xi_j^\mu)_{\mu\ge2,j\le k}$, the events $\left\{T_i(\xi^1) = 0\right\}$, $i > k$, are conditionally independent, given $\{(\xi_j^\mu)_{\mu\ge2,j\le k} = (x_j^\mu)_{\mu\ge2,j\le k}\}$. We therefore obtain

$$\begin{aligned}\mathbb{P}\left(\forall i > k : T_i(\xi^1) = 0 | \mathcal{Z}_k\right) =& \mathbb{P}\left(\forall i > k : \sum_{j=1}^k \bar{J}_{ij} < \gamma\log(N)\right)\\ =&\mathbb{E}_{(\xi_j^\mu)_{\mu\ge2,j\le k}}\left[\mathbb{P}\left(\forall i > k : \sum_{j=1}^k \bar{J}_{ij} < \gamma \log(N) \Big| \mathcal{F}_k^N\right)\right]\\ =&\mathbb{E}_{(\xi_j^\mu)_{\mu\ge2,j\le k}}\left[\left(\mathbb{P}\left(\sum_{j=1}^k \bar{J}_{Nj} < \gamma\log(N)\Big|\mathcal{F}_k^N\right)\right)^{N-k}\right]\\ =&\mathbb{E}_{(\xi_j^\mu)_{\mu\ge2,j\le k}}\left[\left(1 - \mathbb{P}\left(\sum_{j=1}^k \bar{J}_{Nj} \ge \gamma\log(N)\Big|\mathcal{F}_k^N\right)\right)^{N-k}\right].\end{aligned} \tag{4.6}$$

As we will see, it suffices to analyse the contribution of messages that contain exactly one of the k activated neurons of ξ^1, to $\sum_{j=1}^k \bar{J}_{Nj}$. We therefore examine the event

$$\bar{B}_\delta(k) := \left\{\forall j \le k : \sum_{\mu\ge2} \mathbb{1}_{\{\sum_{j\le k}\xi_j^\mu = 1\}}\, \xi_j^\mu \in \left[(1-\delta)\frac{\alpha N}{\log(N)}, \infty\right)\right\}.$$

The sum $\sum_{\mu\geq 2}\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\xi_j^\mu$ is for each fixed $j\leq k$ Binomially distributed with parameters $M-1$ and $\hat{p}:=p(1-p)^{k-1}$.

We bound the probability of $\bar{B}_\delta(k)^c$ to show that it tends to 0:

$$\begin{aligned}\mathbb{P}(\bar{B}_\delta(k)^c)=&\mathbb{P}\Big[\exists j\leq k:\sum_{\mu\geq 2}\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\xi_j^\mu\notin\Big[(1-\delta)\frac{\alpha N}{\log(N)},\infty\Big)\Big]\\ \leq&k\mathbb{P}\Big[\sum_{\mu\geq 2}\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\xi_1^\mu<(1-\delta)\frac{\alpha N}{\log(N)}\Big]\\ \leq&k\exp\left[Mp(1-\delta)t\right]\left(1+\hat{p}\left(e^{-t}-1\right)\right)^{M-1}\\ \leq&k\exp\left[Mp\left(-(1-\delta)\log\left(1-\delta\right)-\delta\right)+\mathcal{O}(Mp^2k)\right].\end{aligned}$$

This implies that $\mathbb{P}(\bar{B}_\delta(k)^c)$ tends to 0.

Now for some $i>k$, the number of active edges between neuron i and excited neurons of ξ^1 is at least

$$\sum_{j\leq k}\bar{J}_{ij}=\sum_{j\leq k}\Theta\left(\sum_{\mu=2}^M\xi_i^\mu\xi_j^\mu-1\right)\geq\sum_{j\leq k}\Theta\left(\sum_{\mu=2}^M\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\xi_i^\mu\xi_j^\mu-1\right).$$

In addition, for each $j\leq k$ and $i>k$, the conditional probabilities

$$\mathbb{P}\left[\Theta\left(\sum_{\mu=2}^M\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\xi_i^\mu\xi_j^\mu-1\right)=1\middle|\mathcal{F}_k^N\right]=1-(1-p)^{\sum_{\mu=2}^M\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\xi_j^\mu}$$

are increasing in $\sum_{\mu=2}^M\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\,\xi_j^\mu$. Furthermore, for fixed $i>k$, the events

$$\left\{\Theta\left(\sum_{\mu=2}^M\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\xi_i^\mu\xi_j^\mu-1\right)=1\right\},\quad j\leq k$$

are conditionally independent, given $\{(\xi_j^\mu)_{\mu\geq 2,j\leq k}=(x_j^\mu)_{\mu\geq 2,j\leq k}\}$ for an arbitrary realisation $(x_j^\mu)_{\mu\geq 2,j\leq k}$ of $(\xi_j^\mu)_{\mu\geq 2,j\leq k}$.

Without loss of generality, assume that $(1-\delta)\frac{\alpha N}{\log(N)}\in\mathbb{N}$. We define the event

$$\widehat{B}_\delta(k):=\left\{\forall j\leq k:\sum_{\mu\geq 2}\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\xi_j^\mu=(1-\delta)\frac{\alpha N}{\log(N)}\right\}.$$

Due to the conditional independence mentioned above and the fact that the conditional probability of an active edge between i and j, $j\leq k$, is increasing in $\sum_{\mu=2}^M\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\xi_j^\mu$, we obtain

$$\begin{aligned}&\min_{\bar{B}_\delta(k)}\mathbb{P}\left[\sum_{j\leq k}\Theta\left(\sum_{\mu=2}^M\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\xi_i^\mu\xi_j^\mu-1\right)\geq\gamma\log(N)\middle|\mathcal{F}_k^N\right]\\ \geq&\mathbb{P}\left[\sum_{j\leq k}\Theta\left(\sum_{\mu=2}^M\mathbb{1}_{\{\sum_{j\leq k}\xi_j^\mu=1\}}\xi_i^\mu\xi_j^\mu-1\right)\geq\gamma\log(N)\middle|\widehat{B}_\delta(k)\right].\end{aligned}$$

These considerations allow to bound the last line of (4.6):

$$
\begin{aligned}
&\mathbb{E}_{(\xi_j^\mu)_{\mu\ge 2, j\le k}}\left[\left(1-\mathbb{P}\left(\sum_{j\le k}\bar{J}_{Nj}\ge\gamma\log(N)\Big|\mathcal{F}_k^N\right)\right)^{N-k}\right]\\
\le&\mathbb{E}_{(\xi_j^\mu)_{\mu\ge 2, j\le k}}\left[\left(1-\mathbb{P}\left[\sum_{j\le k}\Theta\left(\sum_{\mu=2}^{M}\mathbb{1}_{\{\sum_{j\le k}\xi_j^\mu=1\}}\xi_N^\mu\xi_j^\mu-1\right)\ge\gamma\log(N)\Big|\mathcal{F}_k^N\right]\right)^{N-k}\right]\\
\le&\max_{\bar{B}_\delta(k)}\left(1-\mathbb{P}\left[\sum_{j\le k}\Theta\left(\sum_{\mu=2}^{M}\mathbb{1}_{\{\sum_{j\le k}\xi_j^\mu=1\}}\xi_N^\mu\xi_j^\mu-1\right)\ge\gamma\log(N)\Big|\mathcal{F}_k^N\right]\right)^{N-k}+\mathbb{P}\left(\bar{B}_\delta(k)^c\right)\\
\le&\left(1-\mathbb{P}\left[\sum_{j\le k}\Theta\left(\sum_{\mu=2}^{M}\mathbb{1}_{\{\sum_{j\le k}\xi_j^\mu=1\}}\xi_N^\mu\xi_j^\mu-1\right)\ge\gamma\log(N)\Big|\widehat{B}_\delta(k)\right]\right)^{N-k}+\mathbb{P}\left(\bar{B}_\delta(k)^c\right).
\end{aligned}
$$

The conditional distribution of $\sum_{j\le k}\Theta\left(\sum_{\mu=2}^{M}\mathbb{1}_{\{\sum_{j\le k}\xi_j^\mu=1\}}\xi_N^\mu\xi_j^\mu-1\right)$, given $\widehat{B}_\delta(k)$, is

$$
\begin{aligned}
&\mathbb{P}\left[\sum_{j\le k}\Theta\left(\sum_{\mu=2}^{M}\mathbb{1}_{\{\sum_{j\le k}\xi_j^\mu=1\}}\xi_N^\mu\xi_j^\mu-1\right)=m\Big|\widehat{B}_\delta(k)\right]\\
=&\binom{k}{m}\mathbb{P}\left[\forall j\le k:\Theta\left(\sum_{\mu=2}^{M}\mathbb{1}_{\{\sum_{j\le k}\xi_j^\mu=1\}}\xi_N^\mu\xi_j^\mu-1\right)=\mathbb{1}_{j\le m}\Big|\widehat{B}_\delta(k)\right]\\
=&\binom{k}{m}(1-p)^{(1-\delta)\alpha\frac{N}{\log(N)}(k-m)}\left[1-(1-p)^{(1-\delta)\alpha\frac{N}{\log(N)}}\right]^m.
\end{aligned}
\tag{4.7}
$$

With the help of the series expansion of the logarithm and the one of the exponential function, we obtain

$$
\begin{aligned}
&(1-p)^{(1-\delta)\alpha\frac{N}{\log(N)}(k-m)}=e^{\log(1-p)(1-\delta)\alpha\frac{N}{\log(N)}(k-m)}\\
=&e^{-(1-\delta)\alpha(k-m)+\mathcal{O}(p(k-m))}=e^{-(1-\delta)\alpha(k-m)}\left[1+\mathcal{O}(p(k-m))\right]=e^{-(1-\delta)\alpha(k-m)}[1+o(1)]
\end{aligned}
$$

and, using that $p=\log(N)/N$ and $m\le k=\mathcal{O}(\log(N))$,

$$
\begin{aligned}
&\left[1-(1-p)^{(1-\delta)\alpha\frac{N}{\log(N)}}\right]^m=\left[1-e^{\log(1-p)(1-\delta)\alpha\frac{N}{\log(N)}}\right]^m\\
=&\left[1-e^{-(1-\delta)\alpha-p\frac{1}{2}(1-\delta)\alpha+\mathcal{O}(p^2)}\right]^m=\left[1-e^{-(1-\delta)\alpha}\left(1-p\frac{1}{2}(1-\delta)\alpha+\mathcal{O}(p^2)\right)\right]^m\\
=&\left[1-e^{-(1-\delta)\alpha}\right]^m\cdot\left(1+\frac{e^{-(1-\delta)\alpha}(p\frac{1}{2}(1-\delta)\alpha+\mathcal{O}(p^2))}{1-e^{-(1-\delta)\alpha}}\right)^m\\
=&\left[1-e^{-(1-\delta)\alpha}\right]^m\cdot\left(1+\mathcal{O}(\log(N)^2/N)\right)=\left[1-e^{-\alpha(1-\delta)}\right]^m(1+o(1)).
\end{aligned}
\tag{4.8}
$$

Finally

$$
\binom{k}{m}(1-p)^{(1-\delta)\alpha\frac{N}{\log(N)}(k-m)}\left[1-(1-p)^{(1-\delta)\alpha\frac{N}{\log(N)}}\right]^m
$$

$$=\binom{k}{m} e^{-\alpha(1-\delta)(k-m)} \left[1-e^{-\alpha(1-\delta)}\right]^m (1+o(1)).$$

$R_{k,\,1-e^{-(1-\delta)\alpha}}$ again denotes a Binomially distributed random variable with parameters k and $1-e^{-\alpha(1-\delta)}$. Inequality (4.5) as well as the subsequent conclusions imply

$$\mathbb{P}\left(\exists i \leq N : T_i(\xi^1) \neq \xi_i^1\right) \geq \mathbb{P}(A_\delta) \min_{\substack{k\in\mathbb{N}: k/\log(N)\\ \in(1-\delta,1+\delta)}} \left[1-\mathbb{P}\left(\forall i>k : T_i(\xi^1)=\xi_i^1 | \mathcal{Z}_k\right)\right] \geq \mathbb{P}(A_\delta)\cdot$$
$$\min_{\substack{k\in\mathbb{N}: k/\log(N)\\ \in(1-\delta,1+\delta)}} \left[1-\left[1-\mathbb{P}\left(R_{k,\,1-e^{-(1-\delta)\alpha}} \geq \gamma\log(N)\right)(1+o(1))\right]^{N-k} - \mathbb{P}(\bar{B}_\delta(k)^c)\right].$$

The minimum of $\mathbb{P}\left(R_{k,\,1-e^{-(1-\delta)\alpha}} \geq \gamma\log(N)\right)$ is realised by the argument $k = \lceil(1-\delta)\log(N)\rceil$. The probability thus tends to 1 if the equation

$$\liminf_{N\to\infty} \frac{1}{\log(N)} \log\left[\mathbb{P}\left(R_{\lceil(1-\delta)\log(N)\rceil,\,1-e^{-(1-\delta)\alpha}} \geq \gamma\log(N)\right)\right] > -1 \tag{4.9}$$

holds. Using Lemma 2.4 and that the Legendre transform $\Lambda_p^*(x)$ of a Bernoulli random variable with parameter $p\in(0,1)$ is

$$\Lambda_p^*(x) = x\log\left(\frac{x}{p}\right) - (1-x)\log\left(\frac{1-p}{1-x}\right),$$

we see that

$$\liminf_{N\to\infty} \frac{1}{\log(N)} \log\left[\mathbb{P}\left(R_{\lceil(1-\delta)\log(N)\rceil,\,1-e^{-(1-\delta)\alpha}} \geq \gamma\log(N)\right)\right]$$
$$\geq (1-\delta)\liminf_{N\to\infty} \frac{1}{\lceil(1-\delta)\log(N)\rceil} \log\left[\mathbb{P}\left(R_{\lceil(1-\delta)\log(N)\rceil,\,1-e^{-(1-\delta)\alpha}} \geq \frac{\gamma}{1-\delta}\lceil(1-\delta)\log(N)\rceil\right)\right]$$
$$=-(1-\delta)\left[\frac{\gamma}{1-\delta}\log\left(\frac{\gamma}{(1-\delta)(1-e^{-\alpha(1-\delta)})}\right) - \left(1-\frac{\gamma}{1-\delta}\right)\log\left(\frac{e^{-\alpha(1-\delta)}}{1-\frac{\gamma}{1-\delta}}\right)\right]$$
$$=-\gamma\log\left(\frac{\gamma}{(1-\delta)(1-e^{-\alpha(1-\delta)})}\right) + (1-\delta-\gamma)\log\left(\frac{e^{-\alpha(1-\delta)}}{1-\frac{\gamma}{1-\delta}}\right)$$

for $\gamma > (1-\delta)(1-e^{-\alpha(1-\delta)})$, that is, $\alpha < -\frac{1}{1-\delta}\log(1-\frac{\gamma}{1-\delta})$ (which is fulfilled if δ is small enough and $\alpha < -\log(1-\gamma)$). Hence the condition (4.9) holds if

$$-\gamma\log\left(\frac{\gamma}{(1-\delta)(1-e^{-\alpha(1-\delta)})}\right) + (1-\delta-\gamma)\log\left(\frac{e^{-\alpha(1-\delta)}}{1-\frac{\gamma}{1-\delta}}\right) > -1. \tag{4.10}$$

If now

$$-\gamma\log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right) > -1,$$

as formulated in the Theorem, the variable δ can be chosen in dependence on γ and α such that (4.10) is fulfilled, since by continuity we have

$$\lim_{\delta\searrow 0} -\gamma\log\left(\frac{\gamma}{(1-\delta)(1-e^{-\alpha(1-\delta)})}\right) + (1-\delta-\gamma)\log\left(\frac{e^{-\alpha(1-\delta)}}{1-\frac{\gamma}{1-\delta}}\right)$$

$$= -\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right).$$

Finally, this yields

$$\lim_{N\to\infty} \mathbb{P}\left(\exists i \leq N : T_i(\xi^1) \neq \xi_i^1\right) \longrightarrow 1$$

for each $\alpha < -\log(1-\gamma)$ that satisfies (4.3).

If $\alpha \geq -\log(1-\gamma)$, we choose some $\alpha' < -\log(1-\gamma)$ such that (4.3) holds. We will see in the proof of the next proposition that this is possible for each $\gamma \in (0,1)$. The probability of having at least one inactive neuron that gets too many signals clearly increases in the number of stored patterns. Since this probability tends to 1 if $\alpha' N^2/\log(N)^2$ patterns are stored, it also does if $\alpha N^2/\log(N)^2$ patterns are stored. This finishes the proof.

□

Proposition 4.10 *In the situation of Theorem 4.9, there is for each* γ *a critical* $\alpha^*(\gamma)$*, such that (4.1) holds for* $0 < \alpha < \alpha^*(\gamma)$ *and (4.4) holds for* $\alpha > \alpha^*(\gamma)$*. In particular, the set of admissible capacity variables is nonempty for each* $\gamma \in (0,1)$*. Each* $\gamma \geq 1$ *is inadmissible.*

The critical value

$$\alpha^* := \sup\left\{\alpha > 0 : \exists\gamma \in (0,1) : \alpha \textit{ is an admissible capacity variable for } \gamma\right\}$$

is equal to

$$\alpha^* = -\log(1-e^{-1}) \approx 0.45.$$

For each $\alpha < \alpha^*$*, there is a nonempty interval* $(\gamma^*(\alpha), 1)$ *such that* $(\gamma^*(\alpha), 1)$ *contains admissible threshold variables for* α *and* $(0, \gamma^*(\alpha))$ *contains inadmissible threshold variables for* α*.*

Proof: The functions $\tilde{f} : \mathbb{R}_+ \to \mathbb{R}$, $f_\gamma : (0,\gamma] \to \mathbb{R}$, $\gamma \in (0,1)$,

$$\tilde{f}(\alpha) := 1 - e^{-\alpha}, \quad f_\gamma(x) := -\gamma \log\left(\frac{\gamma}{x}\right) + (1-\gamma)\log\left(\frac{1-x}{1-\gamma}\right)$$

are continuous and strictly increasing on $\mathbb{R}_+$ and strictly increasing on $(0,\gamma)$, respectively. In particular, $\lim_{x\searrow 0} f_\gamma(x) = -\infty$ and $f_\gamma(\gamma) = 0$. So, there is a unique root x_γ of $f_\gamma + 1$ in $(0,\gamma)$, and for $x \in (0,\gamma)$, we know

$$f_\gamma(x) < -1 \Leftrightarrow x < x_\gamma, \quad f_\gamma(x) > -1 \Leftrightarrow x > x_\gamma.$$

For fixed γ, the fact that $\tilde{f}$ is strictly increasing implies that (4.2) holds for $0 < \alpha < -\log(1-x_\gamma)$ and (4.3) holds for $\alpha > -\log(1-x_\gamma)$. In particular, $x_\gamma < \gamma$ and therefore $-\log(1-x_\gamma) < -\log(1-\gamma)$. Thus the whole set $(0, -\log(1-x_\gamma))$ contains exclusively admissible capacity variables and $(-\log(1-x_\gamma), \infty)$ only contains inadmissible capacity variables; $\alpha^*(\gamma) = -\log(1-x_\gamma)$. The choice of $\gamma \geq 1$ is inadmissible, because $\mathbb{P}(\sum_{j=1}^N \xi_j^1 < \log(N))$ tends to $1/2$ and the synaptic weights are binary.

Finally, to determine α^*, we state that for fixed α, the continuous function

$$f_\alpha : (0,1) \to \mathbb{R}, \quad f_\alpha(\gamma) := -\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right)$$

is strictly increasing on $(0, \frac{1}{1+e^2})$, strictly decreasing on $(\frac{1}{1+e^2}, 1)$ and

$$\lim_{\gamma \searrow 0} -\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right) = -\alpha > -1,$$

$$\lim_{\gamma \nearrow 1} -\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right) = \log\left(1-e^{-\alpha}\right).$$

This implies that the set of possible threshold variables for α is of the form $(\gamma^*(\alpha), 1)$ and that $\alpha^*(\gamma)$ is strictly increasing in γ. In particular, $\alpha^* = -\log(1-e^{-1})$, because for each $\alpha > -\log(1-e^{-1})$, we have $f_\alpha(\gamma) > -1$ for each $\gamma \in (0,1)$ and for $\alpha < -\log(1-e^{-1})$, there is some $\gamma \in (0,1)$ such that $f_\alpha(\gamma) < -1$.

□

Corollary 4.11 *In the Willshaw model with independent and identically distributed patterns and spins using the threshold dynamics with threshold $\gamma \log(N)$, $\gamma \in (0,1)$ let a number of $M = \alpha N^2/\log(N)^2$ patterns be stored, with α such that the stability conditions $\alpha < -\log(1-\gamma)$, (4.2) of Theorem 4.9 are fulfilled.*

1. *Let $\hat{\xi}^\mu$ be a partially erased version of an arbitrary message ξ^μ, where at most $\varrho_1 \log(N)$ of the 1s have been deleted at random, $0 < \varrho_1 < 1$. If $\varrho_1 < 1-\gamma$, the probability of a one step correction of $\hat{\xi}^\mu$ into ξ^μ tends to 1:*

$$\lim_{N\to\infty} \mathbb{P}\left(\forall i \le N : T_i(\hat{\xi}^\mu) = \xi_i^\mu\right) = 1.$$

   If on the other hand $\varrho_1 > 1-\gamma$,

$$\lim_{N\to\infty} \mathbb{P}\left(\exists i \le N : T_i(\hat{\xi}^\mu) \ne \xi_i^\mu\right) = 1.$$

   In this case we can even conclude

$$\lim_{N\to\infty} \mathbb{P}\left(\forall n \ge 1\ \exists i_n \le N : (T^n)_{i_n}(\hat{\xi}^\mu) \ne \xi_{i_n}^\mu\right) = 1.$$

   If $\varrho_1 = 1-\gamma$,

$$\liminf_{N\to\infty} \mathbb{P}\left(\forall n \ge 1\ \exists i_n \le N : (T^n)_{i_n}(\hat{\xi}^\mu) \ne \xi_{i_n}^\mu\right) > 0.$$

2. *If we consider a message $\tilde{\xi}^\mu$ obtained by randomly activating $\varrho_2 \log(N)$ additional neurons, $0 < \varrho_2 < 1$, the probability of correcting the message in one step to ξ^μ tends to 1 if α fulfills $\alpha < -\log(1-\gamma/(1+\varrho_2))$ and*

$$-\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1+\varrho_2-\gamma)\log\left(\frac{e^{-\alpha}}{1+\varrho_2-\gamma}\right) + (1+\varrho_2)\log(1+\varrho_2) < -1. \tag{4.11}$$

This bound is again sharp concerning the one step correction: if either $\alpha < -\log(1-\gamma/(1+\varrho_2))$ *and*

$$-\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1+\varrho_2-\gamma)\log\left(\frac{e^{-\alpha}}{1+\varrho_2-\gamma}\right) + (1+\varrho_2)\log(1+\varrho_2) > -1 \tag{4.12}$$

or $\alpha \geq -\log(1-\gamma/(1+\varrho_2))$*, the message is not corrected in one step, with high probability.*

3. *If finally* $\bar{\xi}^\mu$ *is a randomly corrupted message such that* $\varrho_1 \log(N)$ *activated neurons have been deactivated and* $\varrho_2 \log(N)$ *inactive neurons of* ξ^μ *have randomly been activated,* $0 < \varrho_1 < 1$*, the probability of a correction in one step tends to 1 if 1.)*

$$\varrho_1 < 1-\gamma$$

and, in case $\varrho_1 < \varrho_2$*, if the conditions 2.a)* $\alpha < -\log(1-\gamma/(1-\varrho_1+\varrho_2))$ *and 2.b)*

$$\begin{aligned} &-\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\varrho_1+\varrho_2-\gamma)\log\left(\frac{e^{-\alpha}}{1-\varrho_1+\varrho_2-\gamma}\right) \\ &\quad + (1-\varrho_1+\varrho_2)\log(1-\varrho_1+\varrho_2) < -1 \end{aligned}$$

are fulfilled. If $\varrho_2 > \varrho_1$*, the bound in 2. is sharp concerning the one-step retrieval.*

Proof: Concerning the first point, we cannot correct a number $\varrho_1 \log(N)$ of deleted entries if $\varrho_1 > 1-\gamma$: since the synaptic efficacies are binary, the local field of a neuron i can be at most $k - \varrho_1 \log(N)$, where k is the number of 1's in ξ^μ. For $0 < \delta < \varrho_1 - 1 + \gamma$ and each $k \leq (1+\delta)\log(N)$,

$$k - \varrho_1 \log(N) \leq (1+\delta-\varrho_1)\log(N) < \gamma \log(N).$$

Due to $\mathbb{P}(A_\delta) \to 1$ for $A_\delta = \{\sum_{j=1}^N \xi_j^\mu / \log(N) \in (1-\delta, 1+\delta)\}$, the pattern $\hat{\xi}^1$ is turned to $(0,\ldots,0)$ in the first step, with high probability:

$$\mathbb{P}\left(\forall i \leq N : T_i(\hat{\xi}^\mu) = 0\right) \geq \mathbb{P}(A_\delta) \longrightarrow 1.$$

In this case ξ^1 can never be recovered.

If $\varrho_1 = 1-\gamma$, $\hat{\xi}^\mu$ is turned to $(0,\ldots,0)$ with positive probability because

$$\lim_{N\to\infty} \mathbb{P}\left(\sum_{j=1}^N \xi_j^\mu < \log(N)\right) = \frac{1}{2}.$$

If in contrary $\varrho_1 < 1-\gamma$, we choose $\delta < 1-\gamma-\varrho_1$. The activated neurons of ξ^μ receive $k - \varrho_1 \log(N)$ signals if the input message is the corrupted pattern $\hat{\xi}^\mu$ and ξ^μ consists of k active neurons. On A_δ,

$$k - \varrho_1 \log(N) \geq (1-\delta-\varrho_1)\log(N) > \gamma \log(N);$$

so these neurons remain activated or are reactivated if they have been deactivated to obtain $\hat{\xi}^\mu$.

The inactive neurons of ξ^μ get a lower signal if $\hat{\xi}^\mu$ is the input vector of the dynamics compared to the situation where ξ^μ is the input and are thus stable if ξ^μ is stable, which is the case with high probability because one of the conditions of the theorem is that the stability conditions $\alpha < -\log(1-\gamma)$ and (4.2) are fulfilled.

Concerning 2., if in $\tilde{\xi}^\mu$ a number of $\varrho_2 \log(N)$ neurons is spuriously activated, the inactive neurons of ξ^μ are more likely to be activated by the dynamics. The only difference to the proof of the previous theorem is that the number of potential neurons from which signals can come is increased and equal to $k + \varrho_2 \log(N)$ instead of k if k is the number of activated neurons in ξ^μ. We can adopt Lemma 4.8 and see that the distribution of the number of activated neurons in $\tilde{\xi}^\mu$ to which a fixed neuron $i > k$ is connected by an active edge is asymptotically Binomial with parameters $k + \varrho_2 \log(N)$ and $1 - e^{-\alpha}$. The rest of the proof of Theorem 4.9 can almost literally be repeated and we obtain that the probability of a correction in one step tends to 0 if $\gamma > (1+\varrho_2)(1-e^{-\alpha})$ and (4.11) holds and to 1 if either $\gamma \le (1+\varrho_2)(1-e^{-\alpha})$ or if $\gamma > (1+\varrho_2)(1-e^{-\alpha})$ and (4.12) holds.

In the last case, where $\varrho_1 \log(N)$ neurons have been deactivated and $\varrho_2 \log(N)$ neurons have spuriously been activated, these two conditions are combined. The neurons belonging to ξ^μ get at least $k - \varrho_1 \log(N)$ signals, which exceeds the threshold with high probability if $\varrho_1 < 1-\gamma$, whereas the remaining neurons can potentially get signals from $k - \varrho_1 \log(N) + \varrho_2 \log(N)$ neurons, which leads in case $\varrho_2 > \varrho_1$ to the sharp bound formulated in the corollary.

□

Remark 4.12 *Again, if one wants to correct a certain number of erased entries, one must choose an appropriately small γ and the critical capacity variable $\alpha^*(\gamma)$ decreases with γ. A correction of a number of $\varrho \log(N)$ spuriously active neurons also decreases the maximal capacity variable: if x_γ is the root of $f_\gamma + 1$ in $(0,\gamma)$, as defined in the proof of Proposition 4.10, α must fulfill $\alpha < -\log(1 - x_\gamma/(1+\varrho))$ instead of $\alpha < -\log(1-x_\gamma)$.*

Theorem 4.13 *The bounds on α concerning the stability and error correction of the Willshaw model obtained so far are also valid if the $M = N^2/c^2$ messages are chosen uniformly among all patterns with exactly $c = \log(N)$ active neurons.*

Proof: We only give a short summary because the proof is very similar to the one of Theorem 4.9. We consider ξ^1, assuming that the first c neurons are active. Concerning the lower bound on α, we consider the random variables

$$X_2(c) = \sum_{\mu=2}^{M} \mathbb{1}_{\{\sum_{j\le c} \xi_j^\mu = 2\}} \quad \text{and} \quad X_3(c) = \sum_{\mu=2}^{M} \mathbb{1}_{\{\sum_{j\le c} \xi_j^\mu > 2\}},$$

each Binomially distributed with parameters $M-1$ and $\tilde{p}_2 = \binom{c}{2} \frac{c(c-1)}{N(N-1)} (1 + \mathcal{O}(c^2/N)) = \binom{c}{2} \frac{c^2}{N^2} (1 + \mathcal{O}(c^2/N))$ and $\tilde{p}_3 = \binom{c}{3} \frac{c^3}{N^3} (1 + \mathcal{O}(c^2/N))$, respectively. As in the previous proof of Theorem 3.1, die probabilities of the sets

$$C_\delta(c) = \left\{ \frac{X_2(c)}{\alpha \binom{c}{2}} \in (1-\delta, 1+\delta) \right\}, \quad D(c) = \{X_3(c) = 0\}$$

tend to 1. This is also true for the event

$$\tilde{B}_\delta(c) := \left\{\forall j \le c : \sum_{\mu\ge 2} \mathbb{1}_{\{\sum_{j\le c}\xi_j^\mu=1\}}\ \xi_j^\mu \in \Big(-\infty, (1+\delta)\frac{\alpha N}{\log(N)}\Big)\right\},$$

similar to the computation on page 92, using that for each $j \le c$, $\mathbb{P}(\mathbb{1}_{\{\sum_{j\le c}\xi_j^\mu=1\}}\ \xi_j^\mu = 1) = c/N(1+\mathcal{O}(c^2/N))$. To obtain stability of ξ^1, we need $\sum_{j=1}^c \bar{J}_{ij} < \gamma c$ for all $i > c$. We first consider the distribution of

$$\sum_{j\le c}\Theta\left(\sum_{\mu=2}^M \mathbb{1}_{\{\sum_{j\le c}\xi_j^\mu=1\}}\xi_i^\mu\xi_j^\mu - 1\right)$$

instead of $\sum_{j\le c}\bar{J}_{ij} = \sum_{j\le c}\Theta(\sum_{\mu=2}^M \xi_i^\mu\xi_j^\mu - 1)$.

Similarly to the considerations on pages 92 - 94 and the computations in (4.7) and (4.8), by replacing p (the probability of activating neuron i in a message, given that exactly one of the first c neurons is already active), by $(c-1)/(N-c)$, we can bound the probability

$$\mathbb{P}\left(\sum_{j\le c}\Theta\left(\sum_{\mu=2}^M \mathbb{1}_{\{\sum_{j\le c}\xi_j^\mu=1\}}\xi_i^\mu\xi_j^\mu - 1\right) \ge \gamma c \,\middle|\, \tilde{B}_\delta(c)\right) \le \mathbb{P}\left(R_{c,\,1-e^{-(1+\delta)\alpha}} \ge \gamma c\right)(1+o(1)).$$

Since additionally the following probability vanishes, as N tends to infinity:

$$\mathbb{P}\left(\exists i > c : \sum_{\mu:\sum_{j\le c}\xi_j^\mu=2} \xi_i^\mu > 1 \,\middle|\, C_\delta(c)\right) \le (N-c)\binom{\alpha(1+\delta)\binom{c}{2}}{2}p^2 \longrightarrow 0,$$

we obtain

$$\mathbb{P}\left(\exists i > c : \sum_{j=1}^c \bar{J}_{ij} \ge \gamma c\right) \le \mathbb{P}\left((B_\delta(c)\cap C_\delta(c)\cap D(c))^c\right) +$$
$$(N-c)\binom{\alpha(1+\delta)\binom{c}{2}}{2}p^2 + (N-c)\mathbb{P}\left(R_{c-2,\,1-e^{-(1+\delta)\alpha}} \ge \gamma c - 2\right)(1+o(1)).$$

This vanishes, if

$$-\gamma\log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right) < -1$$

by choosing an appropriate $\delta > 0$.

Concerning the opposite bound, the situation is more complicated as in the model with i.i.d. spins. First, we can proceed as in the proof of Theorem 4.9. Again the probability of the event $\bar{B}_\delta = \{\forall j \le k : \sum_{\mu\ge 2}\mathbb{1}_{\{\sum_{j\le k}\xi_j^\mu=1\}}\ \xi_j^\mu \in [(1-\delta)\alpha N/\log(N), \infty)\}$ tends to 1. Additionally, the sum $\sum_{j=1}^c \bar{J}_{ij}$ is estimated to be at least

$$\sum_{j=1}^c \bar{J}_{ij} \ge \sum_{j=1}^c \Theta\left(\sum_{\mu:\sum_{j=1}^c \xi_j^\mu=1} \xi_i^\mu\xi_j^\mu - 1\right).$$

Now, given a realisation of ξ_j^μ, $\mu \geq 2$, $j \leq c$, the events

$$\left\{\sum_{j=1}^{c} \Theta\Big(\sum_{\mu:\sum_{j=1}^{c} \xi_j^\mu=1} \xi_i^\mu \xi_j^\mu - 1\Big) \geq \gamma c\right\}, \quad i > c,$$

are *not* conditionally independent (in contrary to the other version of the model). For arbitrary $(x_j^\mu)_{j\leq c,\mu\geq 2} \in \{0,1\}^{c(M-1)}$, the (ξ_i^μ), $i > c$, $\mu \geq 2$, are conditionally negatively associated, given $\{(\xi_j^\mu)_{j\leq c,\mu\geq 2} = (x_j^\mu)_{j\leq c,\mu\geq 2}\}$: for fixed μ, $(\xi_i^\mu, i > c)$ is conditionally Multivariate Hypergeometrically distributed with parameters $N-c$ (number of characteristics), $c - \sum_{j\leq c} x_j^\mu$ (number of drawings) and $(1, \dots, 1)$ (vector of multiplicity of the characteristics), because there are exactly c active neurons per message. So the $(\xi_i^\mu$, $i > c$, $\mu \geq 2)$, are conditionally negatively associated as union of independent sets of the conditionally negatively associated random variables $(\xi_i^\mu, i > c), \mu \geq 2$. The variables

$$\sum_{j=1}^{c} \Theta\left(\sum_{\mu:\sum_{j=1}^{c} \xi_j^\mu=1} \xi_i^\mu \xi_j^\mu - 1\right), \quad i > c$$

are, for fixed $(\xi_j^\mu)_{j\leq N,\mu\geq 2}$, coordinatewise increasing functions using disjoint subsets of the negatively associated random variables $(\xi_i^\mu, i > c, \mu \geq 2)$ and therefore also conditionally negatively associated, given $\{(\xi_j^\mu)_{j\leq c,\mu\geq 2} = (x_j^\mu)_{j\leq c,\mu\geq 2}\}$. We obtain, with Lemma 2.12 4.,

$$\begin{aligned}
&\mathbb{P}\left[\forall i > c : \sum_{j=1}^{c} \Theta\left(\sum_{\mu:\sum_{j=1}^{c} \xi_j^\mu=1} \xi_i^\mu \xi_j^\mu - 1\right) < \gamma c \middle| \mathcal{F}_c^N\right]\\
=&\mathbb{P}\left[\forall i > c : \mathbb{1}_{\sum_{j=1}^{c} \Theta\left(\sum_{\mu:\sum_{j=1}^{c} \xi_j^\mu=1} \xi_i^\mu \xi_j^\mu - 1\right) \geq \gamma c} = 0 \middle| \mathcal{F}_c^N\right]\\
\leq&\mathbb{P}\left[\sum_{j=1}^{c} \Theta\left(\sum_{\mu:\sum_{j=1}^{c} \xi_j^\mu=1} \xi_N^\mu \xi_j^\mu - 1\right) < \gamma c \middle| \mathcal{F}_c^N\right]^{N-c}.
\end{aligned}$$

The rest of the proof of Theorem 4.9 can now be modified for this version of the model. We consider $\bar{B}_\delta$ and compute analogously the probabilities in (4.7) and (4.8), using $\frac{c-1}{N-c}$ instead of p in the computations, because we consider patterns with exactly one active neuron out of the neurons $1, \dots, c$. This is a negligible difference and does not change the result. We obtain the upper bound on α : the pattern is instable if

$$-\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right) > -1.$$

The proofs concerning the error correction can be adapted in the same way.

□

Theorem 4.14 *If at most* $M = o(N^\beta)$, $\beta = \gamma\log(\gamma) - \gamma + 1$, *messages are stored in the first setting of the model (i.i.d. spins), these are stable with high probability: then*

$$\lim_{N\to\infty} \mathbb{P}(\exists \mu \le M : T(\xi^\mu) \neq \xi^\mu) = 0.$$

If in contrary $N^\beta = o(M)$, $\beta = \gamma\log(\gamma) - \gamma + 1$, *there is, with high probability, at least one pattern that is not stable:*

$$\lim_{N\to\infty} \mathbb{P}(\exists \mu \le M : T(\xi^\mu) \neq \xi^\mu) = 1.$$

In particular, in this case

$$\lim_{N\to\infty} \mathbb{P}(\exists \mu \le M : T(\xi^\mu) = (0,\ldots,0)) = 1.$$

In the second version of the model, with exactly c active neurons per stored message, all patterns are stable with high probability, if $M \le \alpha N^2/\log(N)^2$, *such that*

$$-\gamma\log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right) < -3.$$

Proof: The first two statements follow immediately from Proposition 2.8. If on the one hand $M = o(N^\beta)$, $\beta = \gamma\log(\gamma) - \gamma + 1$, there are with high probability enough activated neurons ($\ge \gamma c$) in each message, and we showed in Proposition 2.8 that the inactive neurons in Amari's model are stable for this choice of M. Since their local field is in the Willshaw model smaller than in Amari's model, the messages are also stable in the Willshaw model, with high probability.

On the other hand, if M is too large, there are not enough activated neurons ($< \gamma c$) in at least one message and since the synaptic efficacies are binary, the activated neurons get in this case a signal that is too low; the whole message is turned into $(0,\ldots,0)$.

The last claim follows analogously to the computations of the precedent Theorem 4.13. To take the number of messages M into account, one to observes that $M\mathbb{P}(\tilde{B}_\delta(c)^c)$ and $M\mathbb{P}(C_\delta(c)^2)$ vanish, where $\tilde{B}_\delta(c)$ and $C_\delta(c)$ are defined as in the proof of Theorem 4.13); in addition, one has to consider messages with 3 and 4 active neurons in the set $\{1,\ldots,c\}$. However, these only lead to a negligible number of connections and do not change the result as N and c tend to infinity. In fact,

$$MN\mathbb{P}(R_{c-x,\,1-e^{-(1+\delta)\alpha}} \ge \gamma\log(N))$$

vanishes for fixed x (x is 11 in the worst case) if $-\gamma\log(\gamma/(1-e^{-\alpha})) + (1-\gamma)\log(e^{-\alpha}/(1-\gamma)) < -3$ and δ is appropriately chosen.

□

In the next section, we analyse the capacity of the Willshaw model with WTA dynamics. As we will see, the bound of the latter model is of another form as the one we found for the first version: the WTA dynamics outperforms the threshold dynamics. However, they both lead to the same maximal capacity for $\gamma \approx 1$, whereas they give different bounds for the error correcting case. This is reasonable because using a threshold variable near 1, an error concerning the inactive neurons of ξ^μ occurs if one of the inactive neurons is connected to almost all of the activated neurons. The WTA dynamics will produce an error if one of them is connected to every of the active neurons.

In computer simulations the threshold dynamics in the Willshaw model is outperformed by WTA (see [18]). Our theoretical results support this observation.

4.2.2 Energy function of the Willshaw Model with Threshold Dynamics

The Willshaw model with threshold dynamics is related to the following Hamiltonian function:

Proposition 4.15 *The function H, defined by*

$$H(\sigma) = -\frac{1}{2}\sum_{i\neq j}\sigma_i\sigma_j J_{ij} + \sum_{i=1}^{N}\sigma_i\left[-J_{ii} + \gamma\log(N)\right]$$

is decreasing along steps of the sequential dynamics $\bar{T} = \bar{T}_1 \circ \ldots \circ \bar{T}_N$.

The term J_{ii} is chosen for the reason that this model includes self-loops of the neurons.

Proof: Suppose that one spin of $\sigma \in \{0,1\}^N$ has been updated. Then, using $\bar{T}_i(\sigma) = (\sigma_1, \ldots, T_i(\sigma), \ldots, \sigma_N)$, the Hamiltonians evaluated in σ and $\bar{T}_i(\sigma)$ differ by

$$H(\bar{T}_i(\sigma)) - H(\sigma) = [T_i(\sigma) - \sigma_i]\left(-\sum_{j=1}^{N} J_{ij} + \gamma\log(N)\right)$$

which is, in each possible case of updating, non-positive due to the updating rules. Concretely, if an activated neuron is deactivated, we know that

$$T_i(\sigma) - \sigma_i = -1 \text{ and } \sum_{j=1}^{N} J_{ij} < \gamma\log(N);$$

in the opposite case, we have

$$T_i(\sigma) - \sigma_i = 1 \text{ and } \sum_{j=1}^{N} J_{ij} \geq \gamma\log(N)$$

which leads in both cases to nonpositivity of the difference of the values of the energy function.

□

The question then is: can we determine energy barriers, i.e. are there Hamming spheres around the message ξ^μ such that on each point on the Hamming sphere of a certain radius, the energy is at least higher than the energy of ξ^μ plus a given constant? We suppose, as in the previous models, a radius of $\eta\log(N)$, with $\eta < 1$. A pattern in Hamming-distance $\eta\log(N)$ of a given message, e. g., ξ^1, is uniquely characterised by the k_1 places where a neuron of ξ^1 has been activated and the k_2 places where a 1 has been deleted, with $k_1 + k_2 = \eta\log(N)$. These sets are denoted by J_1 and J_2, respectively. The pattern that differs from ξ^1 in exactly the spins of the neurons belonging to J is denoted by ξ^1_J. We assume without loss of generality that

$$\xi^1_i = 1 \text{ for } i \le k, \quad \xi^1_i = 0 \text{ for } i > k$$

for some fixed $k \in \mathbb{N}$, $k/\log(N) \in (1-\delta, 1+\delta)$, and consider

$$J_2 = \{k-k_2, \dots, k\}, \quad J_1 = \{k+1, \dots, k+k_1\}.$$

As in Amari's model, we can show that there exist energy valleys around a fixed but arbitrary stored pattern. As in Theorem 2.15, we do not obtain a higher capacity than concerning the perfect retrieval (Theorem 4.9), but the result shows that every corrupted pattern on the sphere has a higher energy than the stored pattern. In Corollary 4.11, we only consider a fixed corrupted pattern.

Theorem 4.16 *In the Willshaw model with threshold algorithm, for $\gamma \in (0,1)$ used as threshold variable and α as capacity variable, such that $\alpha < \gamma$ and*

$$-\gamma\log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha < -1,$$

there are $\eta > 0$ and $\varepsilon > 0$ such that

$$\lim_{N\to\infty} \mathbb{P}\left(\exists J \subseteq \{1,\dots,N\} : |J| = \eta\log(N), H(\xi^1) - H(\xi^1_J) \ge -\varepsilon\log(N)^2\right) = 0$$

for the Hamiltonian function defined in Proposition 4.15.

Proof: Suppose that there are k active neurons in ξ^1. Since $J_{ii} = 1$ for all neurons that are active in ξ^1 and $J_{ii} \le 1$, the difference of the Hamiltonian functions evaluated in ξ^1 and ξ^1_J is bounded from above by

$$\begin{aligned} H(\xi^1) - H(\xi^1_J) &= -\frac{1}{2}\sum_{i,j:i\ne j} J_{ij}\xi^1_i\xi^1_j + \sum_{i=1}^N \xi^1_i\left[-J_{ii} + \gamma\log(N)\right] \\ &\quad +\frac{1}{2}\sum_{i,j:i\ne j} J_{ij}\xi^1_{Ji}\xi^1_{Jj} - \sum_{i=1}^N \xi^1_{Ji}\left[-J_{ii} + \gamma\log(N)\right] \\ &\le -\binom{k}{2} + \binom{k-k_2}{2} + \left[\gamma\log(N) - 1\right]\left[k_2 - k_1\right] + \sum_{\substack{i,j:i\in\{1,\dots,k-k_2\},\\ j\in\{k+1,\dots,k+k_1\}}} J_{ij} + \frac{1}{2}\sum_{\substack{i,j:i\ne j,(i,j)\in\\ \{k+1,\dots,k+k_1\}^2}} J_{ij} \end{aligned}$$

$$= -kk_2 + \frac{k_2^2}{2} + \frac{k_2}{2} + [\gamma \log(N) - 1]\,[k_2 - k_1] + \sum_{\substack{i,j: i \in \{1,\ldots,k-k_2\},\\ j \in \{k+1,\ldots,k+k_1\}}} J_{ij} + \frac{1}{2} \sum_{\substack{i,j: i \neq j, (i,j) \in\\ \{k+1,\ldots,k+k_1\}^2}} J_{ij}.$$

The random variable

$$Y(k,k_1,k_2) := \sum_{\substack{i,j: i \in \{1,\ldots,k-k_2\},\\ j \in \{k+1,\ldots,k+k_1\}}} J_{ij} + \frac{1}{2} \sum_{\substack{i,j: i \neq j, (i,j) \in\\ \{k+1,\ldots,k+k_1\}^2}} J_{ij}$$

denotes the number of positive synaptic efficacies on edges within the set of neurons $\{k+1,\ldots,k+k_1\}$ (without loops) and between neurons of the latter set and neurons of the set $\{1,\ldots,k-k_2\}$. We are interested in estimating the following probability

$$\mathbb{P}\Bigg[Y(k,k_1,k_2) \geq kk_2 - k_2^2/2 - k_2/2 + [\gamma \log(N) - 1][k_1 - k_2] - \varepsilon \log(N)^2\Bigg]$$

such that the product of this estimation of the probability with the number of possible patterns on the Hamming sphere with radius $\eta \log(N)$ (assumed to be a natural number), centred in ξ^μ, converges to 0. Using

$$Y_\mu^{(1)}(k,k_1) := \sum_{j=k+1}^{k_1} \xi_j^\mu, \quad Y_\mu^{(2)}(k,k_2) := \sum_{j=1}^{k-k_2} \xi_j^\mu,$$

this number is bounded from above by

$$Y(k,k_1,k_2) \leq \sum_{\mu=2}^{M} \binom{Y_\mu^{(1)}(k,k_1)}{2} + Y_\mu^{(1)}(k,k_1) Y_\mu^{(2)}(k,k_2).$$

To be precise, the variables $Y_\mu^{(1)}$ and $Y_\mu^{(2)}$ depend, of course, on k_1, k_2 and k, but we omit this in the notation. To estimate the probability, we determine the moment generating function of the random variables: as in the proof of Theorem 2.15, we can for $t < \frac{1}{\eta}$ estimate:

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(t \sum_{\mu=2}^{M} \binom{Y_\mu^{(1)}}{2} + Y_\mu^{(1)} Y_\mu^{(2)}\right)\right] \\
&= \left[\sum_{i_1=0}^{k_1} \sum_{i_2=0}^{k-k_2} \binom{k_1}{i_1} \binom{k-k_2}{i_2} p^{i_1+i_2} (1-p)^{k_1+k-k_2-i_1-i_2} e^{t\left(\binom{i_1}{2} + i_1 i_2\right)}\right]^{M-1} \\
&= \Bigg[(1-p)^{k+k_1-k_2} + (k_1+k-k_2)p(1-p)^{k+k_1-k_2-1} + \binom{k_1}{2} p^2 e^t \\
&\qquad + \binom{k-k_2}{2} p^2 + (k-k_2)k_1 p^2 e^t + \mathcal{O}(p^3(k+k_1)^4)\Bigg]^{M-1} \\
&= \left[1 + p^2 \left(-\binom{k+k_1-k_2}{2} + \binom{k_1}{2} e^t + \binom{k-k_2}{2} + (k-k_2)k_1 e^t\right) + \mathcal{O}(p^3(k+k_1)^4)\right]^{M-1}
\end{aligned}$$

$$= \left[1 + p^2 \left(\binom{k_1}{2}(e^t - 1) + (k - k_2)k_1(e^t - 1)\right) + \mathcal{O}(p^3(k + k_1)^4)\right]^{M-1}$$
$$\leq \exp\left[\alpha \left(\binom{k_1}{2}(e^t - 1) + (k - k_2)k_1(e^t - 1)\right)\right](1 + o(1)).$$

By the application of the exponential Chebyshev inequality, we obtain

$$\begin{aligned} &\mathbb{P}\left[Y(k, k_1, k_2) \geq kk_2 - k_2^2/2 - k_2/2 + \gamma \log(N)(k_1 - k_2) - k_1 + k_2 - \varepsilon \log(N)^2\right] \\ &\leq \exp\left(-t\left[kk_2 - k_2^2/2 - k_2/2 + \gamma \log(N)(k_1 - k_2) - k_1 + k_2 - \varepsilon \log(N)^2\right]\right) \\ &\quad \exp\left[\alpha\left(\binom{k_1}{2}(e^t - 1) + (k - k_2)k_1(e^t - 1)\right)\right](1 + o(1)). \end{aligned} \tag{4.13}$$

By intersecting with A_δ, it suffices to consider

$$\begin{aligned} &\mathbb{P}\left(\exists J \subseteq \{1, \ldots, N\} : |J| = \eta \log(N), H(\xi^1) - H(\xi^1_J) \geq -\varepsilon \log(N)^2 | \mathcal{Z}_k\right) \\ &\leq \eta \log(N) \max_{k_1 \in \mathbb{N}: 0 \leq k_1 \leq \eta \log(N)} \binom{N}{k_1}\binom{k}{k_2} \mathbb{P}\left(H(\xi^1) - H(\xi^1_{\tilde{J}_{k_1}}) \geq -\varepsilon \log(N)^2 | \mathcal{Z}_k\right) \end{aligned} \tag{4.14}$$

for $k/\log(N) \in (1 - \delta, 1 + \delta)$, with $\tilde{J}_{k_1} := \{k - k_2 + 1, \ldots, k\} \cup \{k + 1, \ldots, k + k_1\}$, $k_2 = \eta \log(N) - k_1$. We take into account that $\eta \log(N) = k_1 + k_2$: similarly to the proof in Chapter 2, by inserting the last two lines of (4.13) and replacing k_2 by $\eta \log(N) - k_1$, we see that the maximal value is either attained in $k_1 = 0$ or in $k_1 = \eta \log(N)$.

If $k_1 = 0$, the second exponential in (4.13) vanishes. We observe

$$\eta \log(N) \binom{k}{\eta \log(N)} \mathbb{P}\left(H(\xi^1) - H(\xi^1_{\tilde{J}_0}) \geq -\varepsilon \log(N)^2 | \mathcal{Z}_k\right) \longrightarrow 0$$

if $kk_2 - k_2^2/2 - k_2^2 - \gamma \log(N)k_2 - \varepsilon \log(N)^2 > 0$, which is fulfilled if

$$\varepsilon < \varrho(1 - \delta - \varrho/2 - \gamma).$$

For $k_1 = \eta \log(N)$ and $k \leq (1 + \delta)\log(N)$, the probability in (4.13) is at most

$$\begin{aligned} &\exp\left[-t\gamma \log(N)k_1^2 + tk_1 + t\varepsilon \log(N)^2 + \alpha\left(\binom{k_1}{2}(e^t - 1) + kk_1(e^t - 1)\right)\right] \\ &\leq \exp\left[-t\log(N)^2(\gamma\eta - \varepsilon) + \alpha \log(N)^2(e^t - 1)\left(\frac{\eta^2}{2} + \eta(1 + \delta)\right) + \mathcal{O}(\log(N))\right]. \end{aligned}$$

Using $t = \log(\frac{\gamma\eta - \varepsilon}{\alpha(\eta^2/2 + \eta(1+\delta))})$, the last line in (4.14) vanishes, if

$$-\left(\gamma - \frac{\varepsilon}{\eta}\right)\log\left(\frac{\gamma - \varepsilon/\eta}{\alpha(\eta/2 + 1 + \delta)}\right) + \alpha\left(\frac{\gamma - \varepsilon/\eta}{\alpha(\eta/2 + 1 + \delta)} - 1\right)\left(\frac{\eta}{2} + 1 + \delta\right) < -1.$$

We are now in the same situation as in the proof of Theorem 2.15. It is possible to find $\varepsilon > 0, \eta > 0$ such that all the named conditions are fulfilled (including $t < 1/\eta$), if $\alpha < \gamma$ and $-\gamma \log(\gamma/\alpha) + \gamma - \alpha < -1$.

□

4.3 Capacity and Error Correction of the Willshaw Model with WTA Dynamics

Content of this section is the Willshaw model with WTA dynamics as described in the introduction of this chapter. We first consider the behaviour of the one step retrieval. Our first observation involves the difference between the dynamics using the variable thresholds $\bar{S}_{(c)}$ and $\bar{S}_{(1)}$:

Remark 4.17 *Consider the Willshaw model with WTA dynamics in the version with stored messages containing exactly c active neurons. There is no difference in the first step of the WTA dynamics between the usage of the variable thresholds $\bar{S}_{(c)}$ and $\bar{S}_{(1)}$, as long as we consider a partially erased version of a stored pattern or the stored pattern itself as input.*

This can easily be understood: assume that f of the activated neurons have been erased; $f = 0$ in the case of the correct pattern. The $\bar{S}_i$ are bounded by the number of activated neurons in the input, because the synaptic weights are binary. The highest possible value, that is, $c - f$, is at least attained by the c neurons of the stored pattern. Consequently we have $\bar{S}_{(c)} = \bar{S}_{(1)}$. This is only true if self-loops are accepted: otherwise the inactive neurons in a partially erased pattern get $c - f$, the non-erased $c - f - 1$ signals, and we cannot guarantee $\bar{S}_{(1)} = \bar{S}_{(c)}$. In particular, for all the results obtained for $\bar{S}_{(1)}$, it is important to respect self-loops.

The WTA dynamics using $\bar{S}_{(1)}$ can also be applied if the messages do not consist of exactly c neurons, but whose spins $(\xi_i^\mu, i \leq N)$, are independent and identically distributed. We begin with the analysis of this version, but can afterwards show that the obtained results are also valid for the model with stored patterns chosen uniformly from the set of all messages with exactly c active neurons.

Theorem 4.18 *Consider the Willshaw model with patterns such that ξ_j^μ, $\mu \geq 1$, $j \leq N$, are independent and identically distributed. Assume that in the model with WTA dynamics using $\bar{S}_{(1)}$ a number of $M = \alpha N^2/(\log N)^2$ patterns has been stored. Then for $\alpha < -\log(1 - e^{-1})$ and for any fixed μ*

$$\lim_{N\to\infty} \mathbb{P}\left(\forall i \leq N : T_i(\xi^\mu) = \xi_i^\mu\right) = 1.$$

This bound is sharp: for $\alpha > -\log(1 - e^{-1})$ and any fixed μ, we have

$$\lim_{N\to\infty} \mathbb{P}\left(\exists i \leq N : T_i(\xi^\mu) \neq \xi_i^\mu\right) = 1.$$

If $\varrho \log N$, $0 \leq \varrho < 1$ of the initially activated neurons of message ξ^μ have been erased at random to obtain $\tilde{\xi}^\mu$, we have for $\alpha < -\log(1 - e^{-1/(1-\varrho)})$ and any fixed μ:

$$\lim_{N\to\infty} \mathbb{P}\left(\forall i \leq N : T_i(\tilde{\xi}^\mu) = \xi_i^\mu\right) = 1.$$

Again, this bound is sharp: For $\alpha > -\log(1 - e^{-1/(1-\varrho)})$ we have for any fixed μ

$$\lim_{N\to\infty} \mathbb{P}\left(\exists i \leq N : T_i(\tilde{\xi}^\mu) \neq \xi_i^\mu\right) = 1.$$

Proof of Theorem 4.18: We just prove the second part of the theorem. By setting f to 0, we obtain the first part. We again assume that k neurons in the message ξ^1 are active, with $k/\log(N) \in (1-\delta, 1+\delta)$ for some small $\delta > 0$, and that $\xi_i^1 = 1$ for $i \le k$ and $\xi_i^1 = 0$ for $i \ge k+1$.

Assume that f of the k active neurons are erased. Let $\tilde{\xi}^1 \in \{0,1\}^N$ be the version of ξ^1 corrupted in the way that $\tilde{\xi}_i^1 = 1$ for $i \le k-f$ and $\tilde{\xi}_i^1 = 0$ for $i \ge k-f+1$. The $\bar{S}_i$ are bounded by

$$\bar{S}_i(\tilde{\xi}^1) = \sum_{j=1}^{k-f} J_{ij} \le k-f,$$

but the activated neurons of ξ^1 have positive synaptic efficacies among them coming from their common activation in ξ^1, so

$$\bar{S}_i(\tilde{\xi}^1) = k-f, \quad i \le k.$$

The value of $\bar{S}_{(1)}$ is thus equal to $k-f$ and therefore $T_i(\tilde{\xi}^1) = \xi_i^1$ for each $i \le k$.

Obviously, we have $T(\tilde{\xi}^1) \neq \xi^1$ if there exists some $i \ge k+1$, such that neuron i also attains $k-f$ signals, which is possible if and only if the synaptic efficacies between neuron i and each of the $k-f$ activated neurons is 1.

Lemma 4.8 can be modified to an arbitrary subset of the activated neurons of ξ^1, e. g., $\{1,\dots,k-f\}$. The probability that a fixed neuron i, $i > k$, has non-zero synaptic efficacies with every of the first $k-f$ neurons is then

$$\mathbb{P}\left(\sum_{j=1}^{k-f} \bar{J}_{ij} = k-f\right) = (1-e^{-\alpha})^{k-f}\left[1+o(1)\right].$$

This can be used to bound

$$\begin{aligned}
&\mathbb{P}\left(\exists i \ge k+1 : \sum_{j=1}^{k-f} \bar{J}_{ij} = k-f\right)\\
\le& N(1-e^{-\alpha})^{k-f}\left[1+o(1)\right]\\
=& \exp\left[\log(N) + (k-f)\log(1-e^{-\alpha})\right]\left[1+o(1)\right].
\end{aligned}$$

Taking the maximum over all $k \in \mathbb{N}$ such that $k/\log(N) \in (1-\delta, 1+\delta)$ yields

$$\begin{aligned}
&\mathbb{P}\left(\exists i \le N : \xi_i^1 = 0, T_i(\tilde{\xi}^1) = 1\right)\\
\le& \max_{k\in\mathbb{N}: k/\log(N)\in(1-\delta,1+\delta)} \mathbb{P}\left(\exists i > k : T_i(\tilde{\xi}^1) = 1 | \mathcal{Z}_k\right) + \mathbb{P}(A_\delta^c)\\
\le& \exp\left[\log(N) + (1-\delta-\varrho)\log(N)\log(1-e^{-\alpha})\right]\left[1+o(1)\right] + \mathbb{P}(A_\delta^c).
\end{aligned}$$

This tends to 0 if

$$(1-\delta-\varrho)\log(1-e^{-\alpha}) < -1, \text{ i.e. } \alpha < -\log\left(1-e^{-1/(1-\delta-\varrho)}\right).$$

Let $\alpha < -\log\left(1 - e^{-1/(1-\varrho)}\right)$. Then, due to the fact that $\mathbb{P}(A_\delta) \to 1$ and by choosing δ such that $(1-\delta-\varrho)\log(1-e^{-\alpha}) < -1$, we obtain

$$\mathbb{P}\left[\exists i \le N : T_i\left(\tilde{\xi^1}\right) \neq \xi_i^1\right] \to 0$$

as desired.

It remains to prove the reverse bound on the storage capacity. We assume that $M \geq \alpha N^2/\log(N)^2$ for some $\alpha > -\log(1-e^{-1/(1-\varrho)})$. To show that there will be an error with high probability, the proof of the corresponding part of Theorem 4.9 can almost literally be repeated. There are only two differences:

1. If $f \neq 0$, it suffices to condition on a realisation of the variables ξ_i^μ, $\mu \in \{1, \dots, M\}$, $i \le k - f$, instead of $i \le k$; the corresponding events change to $\bar{B}_\delta(k-f)$ and $\widehat{B}_\delta(k-f)$, respectively. The probability of $\bar{B}_\delta(k-f)$ tends, of course, also to 1, as it is included in $\bar{B}_\delta(k)$. The minimum of the conditional probabilities is attained on $\widehat{B}_\delta(k-f)$.

2. After having stated that

$$\sum_{j \le k-f} \Theta\left(\sum_{\mu=2}^{M} \xi_i^\mu \xi_j^\mu - 1\right) \geq \sum_{j \le k-f} \Theta\left(\sum_{\mu=2}^{M} \mathbb{1}_{\{\sum_{j\le k-f} \xi_j^\mu = 1\}} \xi_i^\mu \xi_j^\mu - 1\right),$$

   it suffices to compute the probability in (4.7) and (4.8) for $m = k - f$:

$$\begin{aligned}&\mathbb{P}\left[\sum_{j \le k-f} \Theta\left(\sum_{\mu=2}^{M} \mathbb{1}_{\{\sum_{j\le k-f} \xi_j^\mu = 1\}} \xi_i^\mu \xi_j^\mu - 1\right) = k - f \middle| \widehat{B}_\delta(k-f)\right] \\ &= \left[1 - e^{-\alpha(1-\delta)}\right]^{k-f} [1 + o(1)] .\end{aligned}$$

We obtain

$$\begin{aligned}&\mathbb{P}\left(\exists i \le N : T_i(\tilde{\xi}^1) \neq \xi_i^1\right) \\ \geq &\mathbb{P}(A_\delta) \min_{\substack{k/\log(N) \in \\ (1-\delta, 1+\delta)}} \left[1 - \mathbb{P}\left(\bar{B}_\delta(k-f)^c\right) - \left(1 - \left[1 - e^{-\alpha(1-\delta)}\right]^{k-f} [1 + o(1)]\right)^{N-k}\right]\end{aligned}$$

This converges to 1 if

$$(1-\delta-\varrho)\log\left[1 - e^{-\alpha(1-\delta)}\right] > -1. \tag{4.15}$$

We consider an α with

$$\alpha > -\log\left(1 - e^{-1/(1-\varrho)}\right);$$

obviously δ can be chosen small enough such that (4.15) holds and thus the probability converges to 1. This finishes the proof.

□

Proposition 4.19 *The results concerning the one-step retrieval of the WTA dynamics using $\bar{S}_{(1)}$ obtained in this section are also valid if the stored messages contain exactly c active neurons and either $\bar{S}_{(1)}$ or $\bar{S}_{(c)}$ is used: for a fixed stored pattern ξ^μ, stability is reached if $\alpha < -\log(1-e^{-1})$ and a partially erased pattern, with ϱc erased neurons, is corrected in one step, if $\alpha < -\log(1-e^{-1/(1-\varrho)})$. Both bounds are sharp.*

Proof: Take a stored pattern ξ^μ and assume that f (possibly $f = 0$) of the c active neurons have been deactivated. The pattern ξ^μ is not recovered in the first step if and only if there is a neuron not belonging to the pattern that is linked by active edges to each one of the active neurons of $\tilde{\xi}^\mu$. To prove that the bounds coincide, we refer to the proof of Theorem 4.13. The difference to this proof is that we are interested in the probability that none or at least one, respectively, of the neurons not belonging to ξ^μ is connected by an active edge to every one of the $c-f$ activated neurons of ξ^μ or $\tilde{\xi}^\mu$, instead of to at least γc of them.

□

Remark 4.20 *Both dynamics used in the Willshaw model allow to reach a better capacity than Amari's model and the Ternary simple model.*

If we are only interested in the stability of stored patterns, we set $\gamma \lessapprox 1$ and see that the threshold dynamics is as efficient as the WTA algorithm. But as soon as we analyse the error correcting abilities of the network, we see that the WTA algorithm outperforms the threshold dynamics and reaches explicitly better values. To correct $\varrho \log(N)$ deleted neurons, some $\gamma < 1-\varrho$ must be used, and the critical capacity variable in dependence on γ is smaller than $-\log(1-e^{-1/(1-\varrho)})$. However, this comparison is not fair because the WTA algorithm uses the information of all local fields to update a fixed neuron, whereas the threshold algorithm only needs the information of the local field of the concerned to update it.

The Willshaw model with WTA dynamics is also able to correct patterns with spuriously activated neurons:

Proposition 4.21 *In all versions of the Willshaw model with WTA dynamics (i.i.d. spins or patterns with exactly c activated neurons; dynamics using $h = \bar{S}_{(1)}$ or $\bar{S}_{(c)}$), a corrupted version of ξ^μ can be corrected in one step if*

1. $\varrho_2 c$ additional neurons are spuriously activated and $\alpha < -\log(1-1/(1+\varrho_2))$,

$$-\frac{1}{1+\varrho_2}\log\left(\frac{1}{(1+\varrho_2)(1-e^{-\alpha})}\right)+\left(1-\frac{1}{1+\varrho_2}\right)\log\left(\frac{e^{-\alpha}}{1-\frac{1}{1+\varrho_2}}\right)<-1$$

or if

2. $\varrho_1 c$ neurons of ξ^μ have been deleted, $\varrho_2 c$ ones are spuriously activated and $\alpha < -\log(1-\varrho_1/(1-\varrho_1+\varrho_2))$,

$$-\frac{1-\varrho_1}{1-\varrho_1+\varrho_2}\log\left(\frac{1-\varrho_1}{(1-\varrho_1+\varrho_2)(1-e^{-\alpha})}\right)+\left[1-\frac{1-\varrho_1}{1-\varrho_1+\varrho_2}\right]\log\left(\frac{e^{-\alpha}}{1-\frac{1-\varrho_1}{1-\varrho_1+\varrho_2}}\right)<-1.$$

Proof: The proofs concerning the i.i.d. spins and the patterns with exactly c activated neurons are very similar. We prove the proposition for the second version; in the first version it is necessary to take A_δ into account, but the result is not changed.

In both dynamics, the messages are corrected in one step, if none of the neurons not belonging to ξ^μ gets more signals than the neurons of ξ^μ. In the first case, the correct neurons get at least c signals, in the second case at least $(1-\varrho_1)c$ signals. Analogously to the proofs of Lemma 4.8 and of Theorem 4.9, we see that for an arbitrary neuron not belonging to ξ^μ, the number of incoming signals is asymptotically Binomially distributed with parameters $1-e^{-\alpha}$ and $(1+\varrho_2)c$ and $(1-\varrho_1+\varrho_2)c$, respectively. This yields the statement of the proposition.

□

Theorem 4.22 *In the Willshaw model with exactly c activated neurons per message and WTA dynamics using either $\bar{S}_{(1)}$ or $\bar{S}_{(c)}$, all patterns are stable with high probability, if*

$$\alpha < -\log\left(1-e^{-3}\right).$$

Proof: In each message, there are exactly c activated neurons. The pattern ξ^μ is not stable if there is at least one neuron not belonging to ξ^μ that is connected by active edges to each of its activated neurons. We consider ξ^1 and can proceed as in the proof of Theorem 4.13 or alternatively modify the proof of Lemma 4.8: the distribution of $\mathcal{X}(i)$ remains the same as in the model with i.i.d. spins (Binomial with parameters c/N and $M-1$), and for each $B \subseteq \{2,\dots,M-1\}$,

$$\mathbb{P}\left[\forall j \le c : \bar{J}_{ij} = 1 \middle| \mathcal{X}(i) = B\right]$$

is in the model with exactly c active neurons per message upper bounded by

$$\mathbb{P}\left[\bar{J}_{i1} = 1 \middle| \mathcal{X}(i) = B\right]^c,$$

because the $\bar{J}_{ij} = \Theta(\sum_{\mu=2}^M \xi_i^\mu \xi_j^\mu - 1)$, $j \le c$, are for each $(x^\mu)_{2\le\mu\le M} \in \{0,1\}^{M-1}$ conditionally negatively associated on $\{(\xi_i^\mu)_{2\le\mu\le M} = (x^\mu)_{2\le\mu\le M}\}$. We are then in the situation of the proof of Lemma 4.8 and only need to replace p by $\frac{c-1}{N-1}$ in the rest of the proof; this does not change the result. This yields, for $i > c$,

$$\mathbb{P}\left[\forall j \le c : \bar{J}_{ij} = 1\right] \le (1-e^{-\alpha})^c(1+o(1)).$$

So

$$\begin{aligned} &\mathbb{P}\left(\exists \mu \le M : \exists i \le N : T_i(\xi^\mu) \neq \xi_i^\mu\right) \\ \le& MN\left(1-e^{-\alpha}\right)^c(1+o(1)) \le \exp\left[3\log(N) + c\log(1-e^{-\alpha})\right](1+o(1)). \end{aligned}$$

This immediately yields the statement of the theorem.

□

Finally, with the help of the first section of this chapter, we can shortly analyse the behaviour of the dynamics concerning the mutiple step retrieval.

Proposition 4.23 *In the Willshaw model with either independent and identically distributed spins or exactly c activated neurons per stored message and the WTA dynamics with $h = \bar{S}_{(1)}$, suppose that the input is a partially erased version of a stored pattern with ϱc deleted spins. Then the pattern is corrected if and only if it is corrected in one step. In particular, the bounds on α derived in Theorem 4.18 are also sharp concerning the retrieval process involving multiple steps: if $\alpha < -\log(1 - e^{-1/(1-\varrho)})$, the message is corrected (in one step); if $\alpha > -\log(1 - e^{-1/(1-\varrho)})$, the dynamics does never converge to the correct pattern (both with high probability).*

In the second version, using $\bar{S}_{(c)}$, the pattern is corrected with high probability, if $\alpha < -\log(1 - e^{-1/(1-\varrho)})$. If it is not corrected in the first step, it can benefit from further steps of the dynamics.

Proof: We showed in the first section of this chapter that the dynamics using the variable threshold $h = \bar{S}_{(1)}$ either converges in the first step or does not converge. This shows that the sharp bounds derived concerning the one-step retrieval are also valid for the multiple-step retrieval.

Since the first step of $\bar{S}_{(c)}$ coincides with the first step of $\bar{S}_{(1)}$, the dynamics $\bar{S}_{(c)}$ converges in the first step if $\alpha < -\log(1 - e^{-1/(1-\varrho)})$.

□

5 The Sparse Blume-Emery-Griffiths Model

The model analysed in this chapter is called Blume-Emery-Griffiths model (Blume, Emery and Griffiths, [7]). It uses a ternary state space and has been introduced by Blume, Emery and Griffiths as a three-state spin glass model to study phase separation of liquid He^3-He^4 mixtures. The model has been analysed by Löwe and Vermet in [31] for fixed activity p not depending on the size of the network. They showed the existence of lower bounds on the capacity; these bounds are decreasing with the activity p.

Again, we consider a neural network of N neurons, $N \in \mathbb{N}$, $V = \{1, \ldots, N\}$. The state space of the neurons is the set $S = \{-1, 0, 1\}$ and the edge set is $E = \{\{i, j\} : i, j \in V, i \neq j\}$. The setup is firstly the same as in the Ternary simple threshold network: M patterns $\xi^\mu \in \{-1, 0, 1\}^N$, $1 \leq \mu \leq M$, whose values are indicated by the random variables ξ_j^μ, $1 \leq j \leq N$, $1 \leq \mu \leq M$, are stored. The $(\xi_j^\mu, 1 \leq \mu \leq M, 1 \leq j \leq N)$ are independent and identically distributed, such that

$$\mathbb{P}(\xi_j^\mu = 0) = 1 - p_N = 1 - \frac{\log(N)}{N},$$
$$\mathbb{P}(\xi_j^\mu = \pm 1) = \frac{1}{2} p_N = \frac{\log(N)}{2N}.$$

We will omit the index and write p instead of p_N but have in mind that p depends on N. The information of the stored patterns is processed in two expressions: on the one hand, in

$$J_{ij} := \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu, \quad i \neq j, \quad i, j \in \{1, \ldots, N\}$$

and on the other hand, in

$$K_{ij} := \frac{1}{(1-p)^2} \sum_{\mu=1}^{M} \eta_i^\mu \eta_j^\mu, \quad i \neq j, \quad i, j \in \{1, \ldots, N\}$$

with

$$\eta_i^\mu := (\xi_i^\mu)^2 - p.$$

This distinguishes this model from the previous models where only J_{ij} is built. The K_{ij} are an indicator of common activation in a pattern: stored messages in which both neurons i and j are active or inactive have a positive impact, messages where one neuron is active and the other one is inactive have a negative impact. Given a pattern $\sigma \in \{-1, 0, 1\}^N$, the local field is again given by

$$S_i(\sigma) := \sum_{j \neq i} J_{ij} \sigma_j = \sum_{j \neq i} \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu \sigma_j;$$

additionally, the function

$$\theta_i(\sigma) := \sum_{j \neq i} K_{ij} \sigma_j^2$$

is defined.

The dynamics originally used in the Blume-Emery-Griffiths model in [31] is the dynamics $\tilde{T}$: given a pattern $\sigma \in \{-1, 0, 1\}^N$, it assigns to neuron i the new spin

$$\tilde{T}_i(\sigma) := \operatorname{sgn}(S_i(\sigma))\, \Theta\left(\left|S_i(\sigma)\right| + \theta_i(\sigma)\right).$$

The Heaviside function Θ is again defined by $\Theta(x) = \mathbb{1}_{[0,\infty)}(x)$. A randomly chosen pattern has, in expectation, $\log(N)$ non-zero values (out of the total number of N neurons); this is why we will call it the sparse Blume-Emery-Griffiths model (BEG model, for short). As we will see, the model as defined so far is outperformed by the Ternary simple threshold model of Chapter 3, if we do not adapt it to the sparsity. The new dynamics is of a similar form, including additionally a threshold term: the $i-$th component of T assigns to neuron i the value

$$T_i(\sigma) := \operatorname{sgn}(S_i(\sigma))\, \Theta\left(\left|S_i(\sigma)\right| + \theta_i(\sigma) - \gamma \log(N)\right).$$

We can show that we obtain stability of an arbitrary stored pattern with high probability, if T and a suitable α are used, with the best maximal capacity of the four models analysed so far.

5.1 Stability of the Stored Patterns in the Original BEG Model

The BEG model in the extremely sparse version with $p_N = \log(N)/N$ depending on N first provides a remarkably worse capacity than the Ternary simple threshold model and Amari's model: if $M = \alpha N^2/\log(N)^2$ messages are stored, a randomly chosen message ξ^μ is not stable with positive probability not converging to zero. In fact, a randomly chosen inactive neuron of ξ^μ is mapped to a non-zero value with positive probability.

The BEG model does thus not work as sparse version with this grade of sparsity without being modified.

Proposition 5.1 *In the original BEG model with activity $p = \log(N)/N$ and $M = \alpha N^2/\log(N)^2$ stored patterns, the stored patterns are instable with positive probability: that is,*

$$\liminf_{N\to\infty} \mathbb{P}\left(\tilde{T}(\xi^\mu) \neq \xi^\mu\right) > 0$$

for a fixed but arbitrary $1 \leq \mu \leq M$.

Proof: We consider message ξ^1 and consider without loss of generality the case where ξ^1 consists of k active neurons which are the first neurons $1, \dots, k$. The corresponding event is denoted by $\mathcal{Z}_k$. It suffices to show that at least one of the inactive neurons is activated with non-vanishing probability. In fact, this is even true for an arbitrary neuron i, $i > k$.

To see this, we observe that an inactive neuron $i > k$ is mapped to a non-zero value if the Heaviside-function of $|S_i(\xi^1)| + \theta_i$ is 1: this happens if

$$|S_i(\xi^1)| + \theta_i(\xi^1) \geq 0$$

which is equivalent to

$$\Big|\sum_{\mu=1}^{M}\sum_{j\neq i}\xi_j^1\xi_i^\mu\xi_j^\mu\Big| \geq -\frac{1}{(1-p)^2}\sum_{\mu=1}^{M}\sum_{j\neq i}\left(\xi_j^1\right)^2\eta_i^\mu\eta_j^\mu.$$

We will now show that with positive, non-vanishing probability,

$$-\frac{1}{(1-p)^2}\sum_{\mu=1}^{M}\sum_{j\neq i}\left(\xi_j^1\right)^2\eta_i^\mu\eta_j^\mu < 0 \tag{5.1}$$

which implies $\Theta(|S_i(\xi^1)| + \theta_i(\xi^1)) = 1$.

To analyse the left hand side of (5.1), we first insert the values for $\mu = 1$, assuming that exactly the first k neurons in ξ^1 are activated:

$$-\frac{1}{(1-p)^2}\sum_{\mu=1}^{M}\sum_{j\neq i}\left(\xi_j^1\right)^2\eta_i^\mu\eta_j^\mu = \frac{kp}{1-p} - \frac{1}{(1-p)^2}\sum_{\mu=2}^{M}\sum_{j\leq k}\eta_i^\mu\eta_j^\mu$$

which is negative if

$$kp(1-p) < \sum_{\mu=2}^{M}\sum_{j\leq k}\eta_i^\mu\eta_j^\mu. \tag{5.2}$$

The left hand side of (5.2) is smaller than 1, if $k \leq (1+\delta)\log(N)$. Since

$$\mathbb{P}(A_\delta) = \mathbb{P}\left((1-\delta)\log(N) < \sum_{j=1}^{N}|\xi_j^1| < (1+\delta)\log(N)\right) \longrightarrow 1$$

as N tends to infinity, it suffices to show that

$$\begin{aligned}
&\liminf_{N\to\infty}\min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}}\mathbb{P}\left[\exists i > k : \tilde{T}_i(\xi^1) \neq 0 | \mathcal{Z}_k\right] \geq \liminf_{N\to\infty}\min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}}\mathbb{P}\left[\tilde{T}_N(\xi^1) \neq 0 | \mathcal{Z}_k\right]\\
&\geq \liminf_{N\to\infty}\min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}}\mathbb{P}\left(\sum_{\mu=2}^{M}\sum_{j\leq k}\eta_N^\mu\eta_j^\mu \geq 1\right)
\end{aligned} \tag{5.3}$$

is positive. To shorten the expressions, we denote by $U(N)$ the random variable

$$U(N) := \sum_{\mu=2}^{M}|\xi_N^\mu| \sim \mathrm{Bin}(M-1, p).$$

$U(N)$ is Binomially distributed with parameters $M-1=\alpha\frac{N^2}{\log(N)^2}-1$ and p. Without loss of generality, we replace $M-1$ by M to obtain a better readability. This does not change the result. Again, using the Chebyshev inequality, we obtain

$$\mathbb{P}\left(\frac{(1-\delta)N\alpha}{\log(N)}<U(N)<\frac{(1+\delta)N\alpha}{\log(N)}\right)\longrightarrow 1$$

for each $\delta>0$ as N tends to infinity. We again call the set

$$\left\{\frac{U(N)\log(N)}{N\alpha}\in(1-\delta,1+\delta)\right\}=:B_\delta(N).$$

It then suffices to consider the set $B_\delta(N)$: we observe

$$\mathbb{P}\left(\sum_{\mu=2}^{M}\sum_{j\le k}\eta_N^\mu\eta_j^\mu\ge 1\right)\ge\mathbb{P}(B_\delta(N))\min_{\substack{m\in\mathbb{N}:m\log(N)/\\(N\alpha)\in(1-\delta,1+\delta)}}\mathbb{P}\left[\sum_{\mu=2}^{M}\sum_{j\le k}\eta_N^\mu\eta_j^\mu\ge 1\middle|U(N)=m\right].$$

Now, for neuron N, we define the following two random variables, depending on k:

$$V(k,N):=\sum_{\mu:|\xi_N^\mu|=1}\sum_{j=1}^{k}|\xi_j^\mu|$$

and

$$W(k,N):=\sum_{\mu:|\xi_N^\mu|=0}\sum_{j=1}^{k}|\xi_j^\mu|.$$

We fix k and omit the reference to the dependence on N and k in the next computation in order to write V, W and U instead. The sum in (5.3) is transformed into

$$\begin{aligned}\sum_{\mu=2}^{M}\sum_{j=1}^{k}\eta_N^\mu\eta_j^\mu=&\sum_{\mu:|\xi_N^\mu|=1}\sum_{j=1}^{k}(1-p)\eta_j^\mu+\sum_{\mu:|\xi_N^\mu|=0}\sum_{j=1}^{k}(-p)\eta_j^\mu\\=&V\,(1-p)^2+(Uk-V)(-p)(1-p)+W(1-p)(-p)+((M-U)k-W)\,p^2\\=&V-Vp-Ukp-Wp+Mp^2k=V-Vp-Ukp-Wp+\alpha k.\end{aligned}\tag{5.4}$$

Assume that $U(N)=m$. Then $V(k,N)$ is Binomially distributed with parameters km and p. We can use the Berry-Esseen-bound:

Lemma 5.2 (Berry-Esseen) *[6],[15] Let $X_1,X_2\dots$ be independent and identically distributed random variables with*

$$\mathbb{E}(X_1)=0,\quad\mathbb{E}(X_1^2)=\sigma^2>0,\quad\mathbb{E}(|X_1|^3)<\infty.$$

Then there is a constant $C>0$ such that for all $n\in\mathbb{N}$, denoting by Φ the distribution function of the Standard Normal distribution,

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{\sqrt{\sigma^2 n}}\le x\right)-\Phi(x)\right|\le\frac{C\mathbb{E}(|X_1|^3)}{\sigma^3\sqrt{n}}.$$

This yields

$$\left|\mathbb{P}\left(\frac{V(k,N)-kmp}{\sqrt{p(1-p)km}} \leq b \middle| U(N)=m\right) - \Phi(b)\right| \leq \frac{Cp(1-p)\left((1-p)^2+p^2\right)}{p(1-p)\sqrt{p(1-p)km}}.$$

For k, m of the considered sets, i.e. fulfilling

$$|k-\log(N)| \leq \delta\log(N), \quad |m-\alpha N/\log(N)| \leq \delta\alpha N/\log(N),$$

the right hand side is at most equal to

$$\frac{Cp(1-p)\left((1-p)^2+p^2\right)}{p(1-p)\sqrt{p(1-p)km}} \leq \frac{C\left((1-p)^2+p^2\right)}{\sqrt{(1-p)\alpha(1-\delta)^2\log(N)}}$$

and therefore vanishing as N tends to infinity. In particular, this implies

$$\begin{aligned}&\mathbb{P}\left(V(k,N) \in \left(kmp+0.1\sqrt{kmp(1-p)}, kmp+3\sqrt{kmp(1-p)}\right) \middle| U(N)=m\right)\\ \geq&\Phi(3)-\Phi(0.1)-2C\frac{1-2p+2p^2}{\sqrt{(1-p)\alpha(1-\delta)^2\log(N)}}\end{aligned}$$

for the choice of k and m made above.

The same argumentation holds for $W(k,N)$ – conditionally on $\{U(N)=m\}$, $W(k,N)$ is Binomially distributed with parameters $(M-1-m)k$ and p. Without loss of generality, we again replace $(M-1-m)k$ by $(M-m)k$ for a better readability. The Berry Esseen bound is in this case

$$\left|\mathbb{P}\left(\frac{W(k,N)-(M-m)kp}{\sqrt{p(1-p)(M-m)k}} \leq b \middle| U(N)=m\right) - \Phi(b)\right| \leq \frac{Cp(1-p)\left((1-p)^2+p^2\right)}{\sqrt{p^3(1-p)^3(M-m)k}}$$

and this also tends to 0 for our choice of k and m. We conclude

$$\begin{aligned}&\mathbb{P}\Big(W(k,N) \in \Big((M-m)kp-3\sqrt{(M-m)kp(1-p)}, (M-m)kp-\\ &\qquad 0.1\sqrt{(M-m)kp(1-p)} \quad \Big| U(N)=m\Big)\\ &\geq\Phi(-0.1)-\Phi(-3)-2C\frac{1-2p+2p^2}{\sqrt{\log(N)(1-\delta)\alpha(1-p)[N/\log(N)-1-\delta]}}.\end{aligned}$$

We will see that for an arbitrary but fixed choice of k and m, such that k and m are chosen as described above, and with

$$F(k,m,N) := \left\{V(k,N) \in \left(kmp+0.1\sqrt{kmp(1-p)}, kmp+3\sqrt{kmp(1-p)}\right)\right\}$$

and

$$\begin{aligned}G(k,m,N) :=&\Big\{W(k,N) \in \Big((M-m)kp-3\sqrt{(M-m)kp(1-p)},\\ &(M-m)kp-0.1\sqrt{(M-m)kp(1-p)}\Big)\Big\},\end{aligned}$$

we have

$$\{U(N)=m\}\cap\Big\{\sum_{\mu=2}^{M}\sum_{j\le k}\eta_N^\mu\eta_j^\mu\ge 1\Big\}\quad\supseteq\quad\{U(N)=m\}\cap F(k,m,N)\cap G(k,m,N).$$

This yields

$$\begin{aligned}
&\liminf_{N\to\infty}\mathbb{P}\left(\exists i\le N:\xi_i^1=0,T_i(\xi^1)\neq 0\right)\\
\ge&\liminf_{N\to\infty}\mathbb{P}(A_\delta)\mathbb{P}(B_\delta(N))\min_{\substack{k,m\in\mathbb{N}:k/\log(N),\\ m\log(N)/(\alpha N)\in(1-\delta,1+\delta)}}\mathbb{P}\left[\sum_{\mu=2}^{M}\sum_{j\le k}\eta_N^\mu\eta_j^\mu\ge 1\Big|U(N)=m\right]\\
\ge&\liminf_{N\to\infty}\mathbb{P}(A_\delta)\mathbb{P}(B_\delta(N))\min_{\substack{k,m\in\mathbb{N}:k/\log(N),\\ m\log(N)/(\alpha N)\in(1-\delta,1+\delta)}}\mathbb{P}\left[F(k,m,N)\cap G(k,m,N)\Big|U(N)=m\right]\\
>&0.
\end{aligned}$$

To show the precedent claim, we write

$$U(N)=m=\rho_2\alpha N/\log(N),\quad k=\rho_1\log(N),$$

with $\rho_1,\rho_2\in(1-\delta,1+\delta)$, as well as

$$V(k,N)=\rho_1\rho_2\alpha\log(N)+\rho_3\sqrt{\rho_1\rho_2\alpha\log(N)(1-p)}$$

and

$$W(k,N)=\alpha\rho_1\left[N-\rho_2\log(N)\right]-\rho_4\sqrt{\alpha\rho_1\left[N-\rho_2\log(N)\right](1-p)}$$

$\rho_3,\rho_4\in(0.1,3)$. Recalling the last term of (5.4), we transform

$$\begin{aligned}
&V(k,N)-V(k,N)p-U(N)kp-W(k,N)p+\alpha k\\
=&\rho_3\sqrt{\rho_1\rho_2\alpha\log(N)(1-p)}-\frac{\log(N)^2}{N}\rho_1\rho_2\alpha-\rho_3\frac{\log(N)}{N}\sqrt{\rho_1\rho_2\alpha\log(N)(1-p)}\\
&-\alpha\rho_1\log(N)+\frac{\log(N)^2}{N}\rho_1\rho_2\alpha+\frac{\log(N)}{N}\rho_4\sqrt{\alpha\rho_1\left[N-\rho_2\log(N)\right](1-p)}+\alpha\rho_1\log(N)\\
=&\rho_3\sqrt{\rho_1\rho_2\alpha\log(N)(1-p)}+O\left(\frac{\log(N)}{\sqrt{N}}\right)\ge 1
\end{aligned}$$

if N is large enough.

Consequently, with positive probability not converging to zero, a randomly chosen inactive neuron of message ξ^1 is turned into a 1 or a -1 and ξ^1 is not a fixed point of the dynamics.

□

Remark 5.3 *The sparsity of the model poses problems in the original version. For a stored pattern used as input and an inactive neuron i of the pattern, the negative signal term in θ_i (the part of θ_i coming from the stored pattern itself) is too small: on the one hand, there are only a few active neurons, on the other hand, p is very small and the resulting threshold can easily be exceeded by the noise term of θ_i.*

5.2 Stability and Error Correction in the Sparsity Adopted BEG Model

To adopt the dynamics of the BEG model to the sparsity of the patterns, we add a threshold and change the dynamics slightly into

$$T_i(\sigma) := \operatorname{sgn}(S_i(\sigma))\,\Theta\left(\left|S_i(\sigma)\right| + \theta_i(\sigma) - \gamma\log(N)\right)$$

with the functions defined at the beginning of the chapter. This adoption now allows to store up to $M = \alpha N^2/\log(N)^2$ patterns such that an arbitrary one is stable with high probability if α is appropriately chosen and N tends to infinity:

Theorem 5.4 *In the BEG network with activity $p = \log(N)/N$, the dynamics $T = (T_1, \dots, T_N)$, defined by*

$$T_i(\sigma) := sgn(S_i(\sigma))\,\Theta\left(\left|S_i(\sigma)\right| + \theta_i(\sigma) - \gamma\log(N)\right),$$

provides

$$\lim_{N\to\infty} \mathbb{P}\left(\exists i \le N : T_i(\xi^\mu) \neq \xi_i^\mu\right) = 0$$

for any arbitrary but fixed μ, if the following conditions are fulfilled:

$$0 < \gamma < 2$$

and

$$\alpha < \frac{\gamma}{x_\gamma^* - 1}$$

with the unique root x_γ^* in $(1, \infty)$ of the function

$$g_\gamma(x) := x\left(1 + \frac{2}{\gamma} - \log(x)\right) - 1 - \frac{2}{\gamma}.$$

This bound is sharp: if

$$\alpha > \frac{\gamma}{x_\gamma^* - 1},$$

an arbitrary stored pattern is instable with high probability:

$$\lim_{N\to\infty} \mathbb{P}\left(\exists i \le N : T_i(\xi^\mu) \neq \xi_i^\mu\right) = 1.$$

Proof of Theorem 5.4: Without loss of generality, we consider message ξ^1. First, we observe that there are two principal types of errors which can occur, namely:

- a 0 is turned into a 1 or to a -1
- a non-zero spin is turned to a 0 or multiplied by (-1).

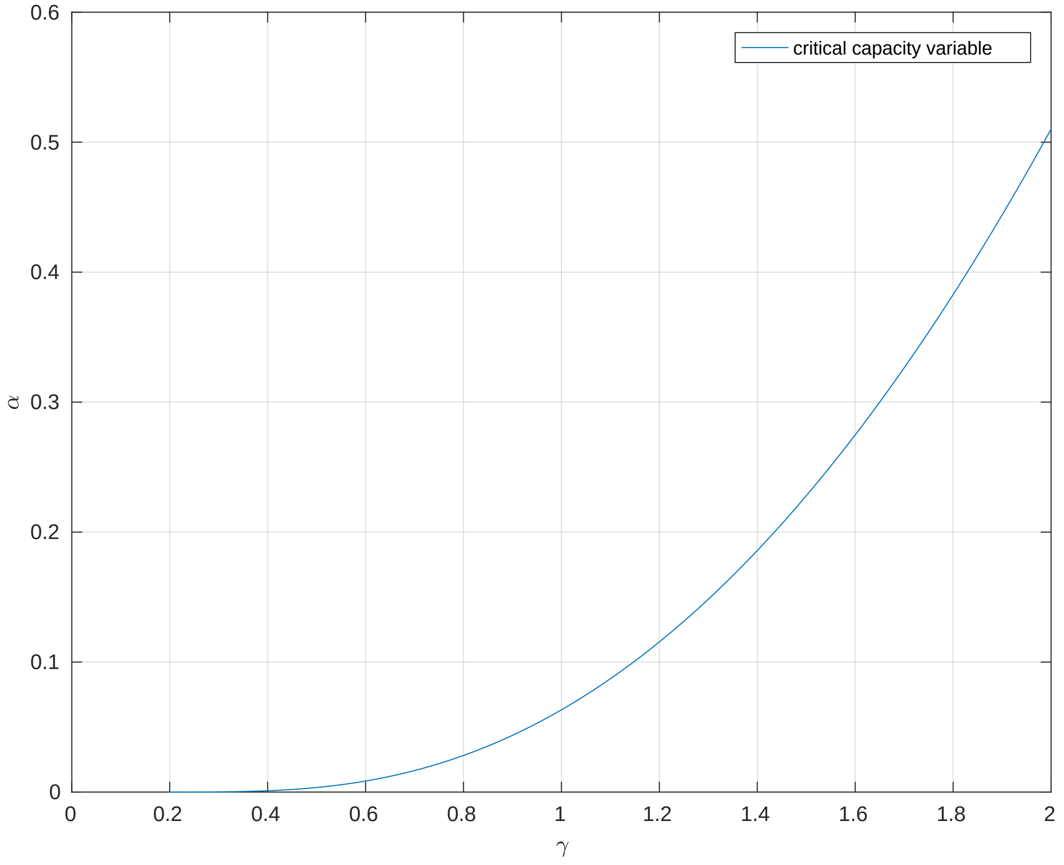

Figure 5.1: Critical capacity variable $\alpha^*(\gamma) = \gamma/(x^*_\gamma - 1)$ in dependence on the threshold variable γ for the BEG model

To begin with the proof, keeping the notation of the previous chapters, we recall that we can estimate

$$\begin{aligned}&\mathbb{P}\left(\exists i \leq N : T_i(\xi^1) \neq \xi^1_i\right)\\ \leq&\mathbb{P}(A_\delta) \max_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \mathbb{P}\left(\exists i \leq N : T_i(\xi^1) \neq \xi^1_i | \mathcal{Z}_k\right) + \mathbb{P}(A^c_\delta).\end{aligned}$$

The probability $\mathbb{P}(A^c_\delta)$ tends to 0 and it suffices to examine the conditional probabilities, given $\mathcal{Z}_k$, for k belonging to the set mentioned above.

We begin with the analysis of the first kind of error: to this purpose, we fix k and take some $i \geq k+1$, e. g., $i = k+1$. An error in this place occurs if $|S_{k+1}(\xi^1)| + \theta_{k+1}(\xi^1) - \gamma\log(N) \geq 0$, i.e., if

$$\left|\sum_{j\neq k+1}\sum_{\mu=1}^{M} \xi^1_j \xi^\mu_{k+1} \xi^\mu_j\right| \geq -\frac{1}{(1-p)^2} \sum_{j\neq k+1} \left(\xi^1_j\right)^2 \sum_{\mu=1}^{M} \eta^\mu_{k+1}\eta^\mu_j + \gamma\log(N). \tag{5.5}$$

Since we consider $i = k+1$ and $\mathcal{Z}_k$, i.e., $\xi^1_{k+1} = 0$ and additionally $\xi^1_j = 0$ for $j > k+1$, the left hand side of (5.5) is equal to

$$\left| \sum_{j \neq k+1} \sum_{\mu=1}^{M} \xi^1_j \xi^\mu_{k+1} \xi^\mu_j \right| = \left| \sum_{j \neq k+1} \xi^1_j \xi^1_j \xi^1_{k+1} + \sum_{j=1}^{k} \sum_{\mu=2}^{M} \xi^1_j \xi^\mu_{k+1} \xi^\mu_j \right| = \left| \sum_{j=1}^{k} \sum_{\mu=2}^{M} \xi^1_j \xi^\mu_{k+1} \xi^\mu_j \right|.$$

We multiply the random part of the term on the right-hand side of (5.5) by $(1-p)^2$ and observe that

$$- \sum_{j \neq k+1} \left(\xi^1_j\right)^2 \sum_{\mu=1}^{M} \eta^\mu_{k+1} \eta^\mu_j = - \sum_{j=1}^{k} \left[\eta^1_{k+1} \eta^1_j + \sum_{\mu=2}^{M} \eta^\mu_{k+1} \eta^\mu_j \right]$$
$$= - \sum_{j=1}^{k} \left[(-p)(1-p) + \sum_{\mu=2}^{M} \eta^\mu_{k+1} \eta^\mu_j \right] = kp(1-p) - \sum_{j=1}^{k} \sum_{\mu=2}^{M} \eta^\mu_{k+1} \eta^\mu_j .$$

After these transformations, the inequality (5.5) becomes

$$\left| \sum_{j=1}^{k} \sum_{\mu=2}^{M} \xi^1_j \xi^\mu_{k+1} \xi^\mu_j \right| > \frac{kp}{(1-p)} - \frac{1}{(1-p)^2} \sum_{j=1}^{k} \sum_{\mu=2}^{M} \eta^\mu_{k+1} \eta^\mu_j + \gamma \log(N).$$

This is either fulfilled if

$$\sum_{j=1}^{k} \sum_{\mu=2}^{M} \left[\xi^1_j \xi^\mu_{k+1} \xi^\mu_j + \frac{1}{(1-p)^2} \eta^\mu_{k+1} \eta^\mu_j \right] > \frac{kp}{(1-p)} + \gamma \log(N)$$

or if

$$\sum_{j=1}^{k} \sum_{\mu=2}^{M} \left[-\xi^1_j \xi^\mu_{k+1} \xi^\mu_j + \frac{1}{(1-p)^2} \eta^\mu_{k+1} \eta^\mu_j \right] > \frac{kp}{(1-p)} + \gamma \log(N).$$

We recall that for independent and identically distributed random variables Z_1, Z_2, Z_3 such that $\mathbb{P}(Z_1 = 1) = \mathbb{P}(Z_1 = -1) = \frac{1}{2}$, the events $\{Z_1 Z_2 Z_3 = x_3\}$ and $\{Z_1 = x_1, Z_2 = x_2\}$ are independent for each choice of $x_1, x_2, x_3 \in \{-1, 1\}$. The same is true for $\{Z_1 Z_2 Z_3 = x_3\}$ and $\{Z_1 = x_1\}$ for arbitrary $x_1, x_3 \in \{-1, 1\}$ as well as for $\{Z_1 Z_2 = x_2\}$ and $\{Z_1 = x_1\}$ for arbitrary $x_1, x_2 \in \{-1, 1\}$. In particular, the conditional distribution of $Z_1 Z_2 Z_3$ with respect to an arbitrary realisation of $Z_1 Z_2$ is the same as the distribution of Z_3.

So for fixed i, the conditional distribution of $\xi^1_{j'} \xi^\mu_i \xi^\mu_{j'}$, for $j' \neq i$, given ξ^1_j, $1 \leq j \leq N$, is

$$\mathbb{P}\left(\xi^1_{j'} \xi^\mu_i \xi^\mu_{j'} = \pm 1 \middle| \, \xi^1_j, j \leq N\right) = \frac{p^2}{2} |\xi^1_{j'}|, \quad \mathbb{P}\left(\xi^1_{j'} \xi^\mu_i \xi^\mu_{j'} = 0 \middle| \, \xi^1_j, j \leq N\right) = 1 - |\xi^1_{j'}| p^2$$

which is the distribution of $\xi^\mu_i \xi^\mu_{j'}$, if $|\xi^1_{j'}| = 1$. Given an arbitrary realisation of $\xi^1_j, j \leq N$, and for fixed i, the

$$(\xi^1_j \xi^\mu_i \xi^\mu_j, j \neq i, \mu \geq 2) \text{ and } (|\xi^1_j| \xi^\mu_i \xi^\mu_j, j \neq i, \mu \geq 2)$$

are identically distributed and we can conclude that conditional on $\mathcal{Z}_k$, the sum $\sum_{j\neq i}^{N}\sum_{\mu=2}^{M}\xi_j^1\xi_i^\mu\xi_j^\mu$ has the same distribution as

$$\sum_{j\neq i}^{k}\sum_{\mu=2}^{M}\xi_i^\mu\xi_j^\mu.$$

In addition, conditional on $\mathcal{Z}_k$, we have for some arbitrary i

$$\sum_{j=1,j\neq i}^{N}\sum_{\mu=2}^{M}\xi_j^1\xi_i^\mu\xi_j^\mu+\frac{1}{(1-p)^2}\left(\xi_j^1\right)^2\eta_i^\mu\eta_j^\mu\sim\sum_{j\leq k,j\neq i}\sum_{\mu=2}^{M}\xi_i^\mu\xi_j^\mu+\frac{1}{(1-p)^2}\eta_i^\mu\eta_j^\mu$$

and

$$\sum_{j\leq k,j\neq i}\sum_{\mu=2}^{M}\xi_i^\mu\xi_j^\mu+\frac{1}{(1-p)^2}\eta_i^\mu\eta_j^\mu\sim\sum_{j\leq k,j\neq i}\sum_{\mu=2}^{M}-\xi_i^\mu\xi_j^\mu+\frac{1}{(1-p)^2}\eta_i^\mu\eta_j^\mu,$$

because the sums

$$\sum_{j=1,j\neq i}^{N}\sum_{\mu=2}^{M}\left(\xi_j^1\right)^2\eta_i^\mu\eta_j^\mu,\qquad\sum_{j=1,j\neq i}^{k}\sum_{\mu=2}^{M}\eta_i^\mu\eta_j^\mu$$

are measurable with respect to $|\xi_j^\mu|$, $1\leq j\leq N$, $1\leq\mu\leq M$ and the sums

$$\sum_{j=1,j\neq i}^{N}\sum_{\mu=2}^{M}\xi_j^1\xi_i^\mu\xi_j^\mu,\qquad\sum_{j=1,j\neq i}^{k}\sum_{\mu=2}^{M}\xi_i^\mu\xi_j^\mu$$

are symmetrically distributed with respect to these random variables. We obtain for the inactive neurons of ξ^1:

$$\begin{aligned}&\mathbb{P}\left[\exists i\geq k+1:T_i(\xi^1)\neq 0|\mathcal{Z}_k\right]\\ \leq&N\mathbb{P}\left[T_{k+1}(\xi^1)\neq 0|\mathcal{Z}_k\right]\leq 2N\mathbb{P}\left[\sum_{j\leq k}\sum_{\mu=2}^{M}\xi_{k+1}^\mu\xi_j^\mu+\frac{1}{(1-p)^2}\eta_{k+1}^\mu\eta_j^\mu>\frac{kp}{1-p}+\gamma\log(N)\right].\end{aligned}$$

In the last line, we used the previous considerations. With the intention to give an upper bound on this probability, we use the exponential Chebyshev inequality: for $t>0$,

$$\begin{aligned}&\mathbb{P}\left[\sum_{j\leq k}\sum_{\mu=2}^{M}\xi_{k+1}^\mu\xi_j^\mu+\frac{1}{(1-p)^2}\eta_{k+1}^\mu\eta_j^\mu>\frac{kp}{1-p}+\gamma\log(N)\right]\\ \leq&\exp\left[-t\left(\frac{kp}{1-p}+\gamma\log(N)\right)\right]\mathbb{E}\left[\exp\left(t\sum_{j\leq k}\sum_{\mu=2}^{M}\xi_{k+1}^\mu\xi_j^\mu+\frac{1}{(1-p)^2}\eta_{k+1}^\mu\eta_j^\mu\right)\right].\end{aligned}$$

The M messages are independent: this yields

$$\mathbb{E}\left[\exp\left(t\sum_{j\leq k}\sum_{\mu=2}^{M}\xi_{k+1}^\mu\xi_j^\mu+\frac{1}{(1-p)^2}\eta_{k+1}^\mu\eta_j^\mu\right)\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\exp\left(t\sum_{j\leq k}\xi_{k+1}^M\xi_j^M + \frac{1}{(1-p)^2}\eta_{k+1}^M\eta_j^M\right)\right]^{M-1} \\
&= \left[(1-p)\cdot\left[(1-p)e^{t\frac{p^2}{(1-p)^2}} + pe^{-t\frac{p}{1-p}}\right]^k + p\cdot\left[(1-p)e^{-t\frac{p}{1-p}} + \frac{1}{2}pe^{2t} + \frac{1}{2}p\right]^k\right]^{M-1}.
\end{aligned}
$$

Remembering the argumentation at the beginning of the proof, it is enough to analyse the conditional probability for each $k \in \mathbb{N}$ satisfying $k/\log(N) \in (1-\delta, 1+\delta)$. We use the series representation of the exponential function in combination with the Binomial formula, for t not depending on N:

$$
\begin{aligned}
&\left[(1-p)\cdot\left((1-p)e^{t\frac{p^2}{(1-p)^2}} + pe^{-t\frac{p}{1-p}}\right)^k + p\cdot\left((1-p)e^{-t\frac{p}{1-p}} + \frac{1}{2}pe^{2t} + \frac{1}{2}p\right)^k\right]^{M-1} \\
&= \left[(1-p)\left[(1-p)\left(1+\frac{tp^2}{(1-p)^2} + \mathcal{O}(p^4)\right) + p\left(1 - \frac{tp}{1-p} + \mathcal{O}(p^2)\right)\right]^k\right. \\
&\qquad\left. + p\left[(1-p)\left(1 - \frac{tp}{1-p} + \mathcal{O}(p^2)\right) + \frac{1}{2}pe^{2t} + \frac{1}{2}p\right]^k\right]^{M-1} \\
&= \left[(1-p)\left[1+\mathcal{O}(p^3)\right]^k + p\left[1 - \frac{tp}{1-p} - \frac{1}{2}p + \frac{1}{2}pe^{2t} + \mathcal{O}(p^2)\right]^k\right]^{M-1} \\
&= \left[1 + kp^2\left(\frac{1}{2}e^{2t} - \frac{1}{2} - \frac{t}{1-p}\right) + \mathcal{O}(p^3k^2)\right]^{M-1} \\
&\leq \exp\left[k\alpha\left(\frac{1}{2}e^{2t} - \frac{1}{2} - \frac{t}{1-p}\right) + \mathcal{O}\left(Mp^3k^2\right)\right] = \exp\left[k\alpha\left(\frac{1}{2}e^{2t} - \frac{1}{2} - t\right) + \mathcal{O}\left(pk^2\right)\right].
\end{aligned}
\tag{5.6}
$$

In the last steps we used the inequality $1 + x \leq e^x$ for each $x \in \mathbb{R}$ as well as the number of stored patterns $M = \alpha N^2/\log(N)^2$.

In combination with the previous steps of the proof, the following conditional probability is at most

$$
\begin{aligned}
&\mathbb{P}\left(\exists i \geq k+1 : T_i(\xi^1) \neq 0 | \mathcal{Z}_k\right) \\
&\leq \exp\left(\log(N) + \log(2) - \frac{tkp}{1-p} - t\gamma\log(N) + k\alpha\left(\frac{1}{2}e^{2t} - \frac{1}{2} - t\right)\right)\left(1 + \mathcal{O}\left(pk^2\right)\right).
\end{aligned}
$$

Since we conditioned on A_δ, the second line must vanish for each k such that $k/\log(N) \in (1-\delta, 1+\delta)$. Anticipating that the variable t will not depend on N and assuming that $k = \rho\log(N)$ for some $\rho \in (1-\delta, 1+\delta)$, the exponent is equal to

$$
\log(N) - t\gamma\log(N) + \rho\log(N)\alpha\left(\frac{1}{2}e^{2t} - \frac{1}{2} - t\right) + o(\log(N)). \tag{5.7}
$$

Convergence to zero of $\mathbb{P}(\exists i \leq N : \xi_i^1 = 0, T_i(\xi^1) \neq 0)$ is reached if there is for each $\rho \in (1-\delta, 1+\delta)$ some $t_{\alpha,\gamma,\rho} > 0$ depending on ρ, γ and α such that

$$
h_{\alpha,\gamma,\rho}(t_{\alpha,\gamma,\rho}) := -t_{\alpha,\gamma,\rho}\gamma + \rho\alpha\left(\frac{1}{2}e^{2t_{\alpha,\gamma,\rho}} - \frac{1}{2} - t_{\alpha,\gamma,\rho}\right) < -1.
$$

Using the global minimum of $h_{\alpha,\gamma,\rho}$,

$$t^*_{\alpha,\gamma,\rho} := \frac{1}{2}\log\left(1+\frac{\gamma}{\rho\alpha}\right), \quad t^*_{\alpha,\gamma,\rho} > 0$$

the dominant term in (5.7) is equal to

$$\begin{aligned}
&\log(N) - t^*_{\alpha,\gamma,\rho}\gamma\log(N) + \rho\log(N)\alpha\left(\frac{1}{2}e^{2t^*_{\alpha,\gamma,\rho}} - \frac{1}{2} - t^*_{\alpha,\gamma,\rho}\right)\\
&= \log(N)\left[1-\gamma\frac{1}{2}\log\left(1+\frac{\gamma}{\rho\alpha}\right)+\frac{1}{2}\rho\alpha\left(1+\frac{\gamma}{\rho\alpha}\right)-\frac{1}{2}\rho\alpha-\frac{1}{2}\rho\alpha\log\left(1+\frac{\gamma}{\rho\alpha}\right)\right]\\
&= \log(N) f_{\alpha,\rho}(x_{\alpha,\gamma,\rho})
\end{aligned}$$

with

$$f_{\alpha,\rho}(x) := 1 + \frac{1}{2}\rho\alpha\left(-x\log(x) + x - 1\right)$$

and

$$x_{\alpha,\gamma,\rho} := 1 + \frac{\gamma}{\rho\alpha}.$$

To let the conditional probability tend to 0, the condition

$$f_{\alpha,\rho}(x_{\alpha,\gamma,\rho}) < 0$$

must be fulfilled for each $\rho \in (1-\delta, 1+\delta)$. For fixed γ and α, $f_{\alpha,\rho}(x_{\alpha,\gamma,\rho})$ is continuous in ρ, $\rho \in \mathbb{R}_+$. If for fixed γ, α, the inequality $f_{\alpha,1}(x_{\alpha,\gamma,1}) < 0$ holds true, then also

$$f_{\alpha,\rho}(x_{\alpha,\gamma,\rho}) < 0 \quad \forall \rho \in (1-\delta, 1+\delta)$$

if $\delta > 0$ is small enough. For a fixed pair of γ and α such that $f_{\alpha,1}(x_{\alpha,\gamma,1}) < 0$, δ can be chosen in dependence of α and γ. The satisfaction of the inequality $f_{\alpha,1}(x_{\alpha,\gamma,1}) < 0$ implies therefore the desired convergence to zero of the probability $\mathbb{P}\left(\exists i \le N : \xi_i^1 = 0, T_i(\xi^1) \neq 0\right)$.

It remains thus to determine the set $\{\alpha > 0 : f_{\alpha,1}(x_{\alpha,\gamma,1}) < 0\}$ for fixed γ. Multiplying by $2/\alpha$ and replacing $1/\alpha$ by $(x_{\alpha,\gamma,1}-1)/\gamma$ allows to reformulate the condition $f_{\alpha,1}(x_{\alpha,\gamma,1}) < 0$ in terms of the in Theorem 5.4 defined function: it holds if

$$g_\gamma(x_{\alpha,\gamma,1}) := x_{\alpha,\gamma,1}\left(1+\frac{2}{\gamma}-\log(x_{\alpha,\gamma,1})\right) - 1 - \frac{2}{\gamma} < 0.$$

The function $g_\gamma(0,\infty) \to \mathbb{R}$ has two roots, one is equal to 1, the other one is bigger than 1 (and bigger than the extremal point $\hat{x}_\gamma = e^{2/\gamma}$). The derivative $g'_\gamma(x) = 2/\gamma - \log(x)$ is positive on $(0, \hat{x}_\gamma)$ and negative on $(\hat{x}_\gamma, \infty)$. Let x^*_γ be the unique root of g_γ bigger than 1. The function g_γ is negative on the intervals $(0,1)$ and (x^*_γ, ∞) and positive on $(1, x^*_\gamma)$. We choose α depending on γ such that

$$\alpha < \frac{\gamma}{x^*_\gamma - 1}. \tag{5.8}$$

Then each pair of $\gamma > 0$ and $\alpha > 0$ chosen subject to these conditions provides

$$\mathbb{P}\left[\exists i \le N : \xi_i^1 = 0, T_i(\xi^1) \neq 0\right] \longrightarrow 0$$

as N tends to infinity. Due to conditions following from subsequent calculations, γ must be chosen such that

$$\gamma \in (0, 2). \tag{5.9}$$

We now analyse the second type of error. We choose an arbitrary neuron that is activated in ξ^1; assuming that $\mathcal{Z}_k$ holds, we take $i = 1$. The neuron is not mapped to its original value if either

$$|S_1(\xi^1)| + \theta_1(\xi^1) - \gamma \log(N) < 0 \tag{5.10}$$

or if the two conditions

$$|S_1(\xi^1)| + \theta_1(\xi^1) - \gamma \log(N) > 0 \tag{5.11}$$

and

$$\operatorname{sgn}(S_1(\xi^1)) \neq \xi_1^1 \tag{5.12}$$

hold.

Without loss of generality, we set $\xi_1^1 = 1$. Due to the symmetric distribution of $\sum_\mu \sum_j \xi_j^1 \xi_i^\mu \xi_j^\mu$, the case $\xi_1^1 = -1$ is run analogously. Inequality (5.10) is either satisfied if

$$|S_1(\xi^1)| = S_1(\xi^1), \;\; S_1(\xi^1) + \theta_1(\xi^1) - \gamma \log(N) < 0 \tag{5.13}$$

or if

$$|S_1(\xi^1)| = -S_1(\xi^1), \;\; -S_1(\xi^1) + \theta_1(\xi^1) - \gamma \log(N) < 0. \tag{5.14}$$

Condition (5.12) is necessary for (5.14). So the probability of having an error, given $\mathcal{Z}_k$ and $\xi_1^1 = 1$, is bounded by the sum of the two probabilities

$$\mathbb{P}\left(S_1(\xi^1) + \theta_1(\xi^1) - \gamma \log(N) < 0, |S_1(\xi^1)| = S_1(\xi^1) \Big| \mathcal{Z}_k \cap \{\xi_1^1 = 1\}\right) \tag{5.15}$$

and

$$\mathbb{P}\left(\operatorname{sgn}(S_1(\xi^1)) \neq 1 | \mathcal{Z}_k \cap \{\xi_1^1 = 1\}\right). \tag{5.16}$$

The probability in (5.16) is transformed into

$$\begin{aligned}
&\mathbb{P}\left[\operatorname{sgn}(S_1(\xi^1)) \neq 1 | \mathcal{Z}_k \cap \{\xi_1^1 = 1\}\right] \\
=&\mathbb{P}\left[k - 1 + \sum_{j=2}^{N} \sum_{\mu=2}^{M} \xi_j^1 \xi_1^\mu \xi_j^\mu < 0 \Big| \mathcal{Z}_k\right] = \mathbb{P}\left[- \sum_{j=2}^{N} \sum_{\mu=2}^{M} \xi_j^1 \xi_1^\mu \xi_j^\mu > k - 1 \Big| \mathcal{Z}_k\right].
\end{aligned}$$

Conditionally on $\mathcal{Z}_k$, the distribution of $\sum_{j=2}^{N} \sum_{\mu=2}^{M} \xi_j^1 \xi_1^\mu \xi_j^\mu$ is symmetric and identical with the distributions of

$$\sum_{j=2}^{k} \sum_{\mu=2}^{M} \xi_1^\mu \xi_j^\mu \overset{d}{\sim} - \sum_{j=2}^{k} \sum_{\mu=2}^{M} \xi_1^\mu \xi_j^\mu.$$

We saw in Chapter 3 that for s not depending on N

$$\begin{aligned}
&\mathbb{P}\left[-\sum_{j=2}^{N}\sum_{\mu=2}^{M}\xi_j^1\xi_1^\mu\xi_j^\mu > k-1\middle|\mathcal{Z}_k\right] = \mathbb{P}\left[\sum_{j=2}^{k}\sum_{\mu=2}^{M}\xi_1^\mu\xi_j^\mu > k-1\right]\\
&\leq \exp\left[s(-k+1)\right]\left[p\left(1-p+p\cdot\cosh(s)\right)^{k-1}+1-p\right]^{M-1}\\
&\leq \exp\left[-s(k-1)+(k-1)\alpha(\cosh(s)-1)+\mathcal{O}(pk^2)\right].
\end{aligned}$$

Minimizing the exponent in s yields $s^*_\alpha = \operatorname{arsinh}(1/\alpha)$ as global minimum. According to the definition of s^*_α, the equality

$$-s^*_\alpha + \alpha(\cosh(s^*_\alpha)-1) = -\operatorname{arsinh}(1/\alpha) + \alpha(\cosh(\operatorname{arsinh}(1/\alpha))-1)$$

holds. The right hand side is negative for every choice of $\alpha > 0$ because the function $g(x) = -x\operatorname{arsinh}(x) + \cosh(\operatorname{arsinh}(x)) - 1$ in Chapter 3 is negative on $\mathbb{R}_+$. The probability vanishes because k can be assumed to be of order $\log(N)$ and therefore tends to infinity.

It remains to examine the probability in (5.15). We show that

$$\begin{aligned}
&\mathbb{P}\left[|S_1(\xi^1)| = S_1(\xi^1),\ S_1(\xi^1)+\theta_1(\xi^1)-\gamma\log(N) < 0\middle|\mathcal{Z}_k\cap\{\xi_1^1=1\}\right]\\
\leq&\mathbb{P}\left[S_1(\xi^1)+\theta_1(\xi^1)-\gamma\log(N) < 0\middle|\mathcal{Z}_k\cap\{\xi_1^1=1\}\right]\\
=&\mathbb{P}\left[\sum_{j=2}^{N}\sum_{\mu=1}^{M}\xi_j^1\xi_j^\mu\xi_1^\mu + \frac{\left(\xi_j^1\right)^2}{(1-p)^2}\eta_1^\mu\eta_j^\mu - \gamma\log(N) < 0\middle|\mathcal{Z}_k\cap\{\xi_1^1=1\}\right]\\
=&\mathbb{P}\left[2(k-1)-\gamma\log(N)+\sum_{1<j\leq k}\sum_{\mu=2}^{M}\xi_j^1\xi_j^\mu\xi_1^\mu + \frac{\left(\xi_j^1\right)^2}{(1-p)^2}\eta_1^\mu\eta_j^\mu < 0\middle|\mathcal{Z}_k\right]\\
=&\mathbb{P}\left[-\left(\sum_{1<j\leq k}\sum_{\mu=2}^{M}\xi_j^\mu\xi_1^\mu + \frac{1}{(1-p)^2}\eta_1^\mu\eta_j^\mu\right) > 2k-2-\gamma\log(N)\right]. \qquad (5.17)
\end{aligned}$$

By the same arguments and the proof techniques we used to estimate the previous two probabilities, especially in (5.6), the probability in (5.17) is for $u > 0$ bounded by

$$\begin{aligned}
&\mathbb{P}\left[-\left(\sum_{1<j\leq k}\sum_{\mu=2}^{M}\xi_j^\mu\xi_1^\mu + \frac{1}{(1-p)^2}\eta_1^\mu\eta_j^\mu\right) > 2k-2-\gamma\log(N)\right]\\
\leq&\exp\left[u\left(-2k+2+\gamma\log(N)\right)+k\alpha\left(\frac{1}{2}e^{-2u}-\frac{1}{2}+u\right)\right]\left(1+\mathcal{O}(pk^2)\right).
\end{aligned}$$

Due to the condition of the theorem, $\gamma < 2$. If $\alpha > 2-\gamma$, let $0 < \delta < \gamma/(2-\alpha)-1$ and $\delta < (2-\gamma)/2$. Then, for some $k = \rho\log(N)$ of the considered set we choose

$$u^*_{\rho,\alpha,\gamma} = -\frac{1}{2}\log\left(1-\frac{2}{\alpha}+\frac{\gamma}{\rho\alpha}\right)$$

to minimise

$$u\left(-2k+\gamma\log(N)\right)+k\alpha\left(\frac{1}{2}e^{-2u}-\frac{1}{2}+u\right).$$

Inserting $u^*_{\rho,\alpha,\gamma}$ and $k = \rho\log(N)$ in the above expression leads to

$$\begin{aligned}
&u^*_{\rho,\alpha,\gamma}\left(-2\rho\log(N)+\gamma\log(N)\right)+\rho\log(N)\alpha\left(\frac{1}{2}e^{-2u^*_{\rho,\alpha,\gamma}}-\frac{1}{2}+u^*_{\rho,\alpha,\gamma}\right)\\
=&\frac{1}{2}\log(N)\rho\alpha\left[-\log\left(1-\frac{2}{\alpha}+\frac{\gamma}{\rho\alpha}\right)\left(1-\frac{2}{\alpha}+\frac{\gamma}{\rho\alpha}\right)-\frac{2}{\alpha}+\frac{\gamma}{\rho\alpha}\right]\\
=&\frac{1}{2}\log(N)\rho\alpha\left[-v_{\rho,\alpha,\gamma}\log(v_{\rho,\alpha,\gamma})+v_{\rho,\alpha,\gamma}-1\right],
\end{aligned}$$

$v_{\rho,\alpha,\gamma} = 1 - 2/\alpha + \gamma/(\rho\alpha)$. As we know, the function $-x\log(x) + x - 1$ is negative on $\mathbb{R}_+$. Since

$$2\rho - \gamma > 0, \quad \alpha > 2 - \frac{\gamma}{\rho}$$

due to the choice of δ, both conditions $u^*_{\rho,\alpha,\gamma} > 0$ and $v_{\rho,\alpha,\gamma} > 0$ hold.

If $0 < \alpha < 2 - \gamma$, the choice of $u > 0$ is irrelevant. For $0 < \delta < \frac{2-\gamma-\alpha}{2+\alpha}$, arbitrary $u > 0$ and on A_δ,

$$\begin{aligned}
&u\left(-2k+2+\gamma\log(N)\right)+k\alpha\left(\frac{1}{2}e^{-2u}-\frac{1}{2}+u\right)\\
\leq&u\log(N)\left(-2(1-\delta)+\gamma+(1+\delta)\alpha\right)+2u+(1-\delta)\log(N)\alpha\frac{1}{2}\left(e^{-2u}-1\right)\\
\leq&2u+(1-\delta)\log(N)\alpha\frac{1}{2}\left(e^{-2u}-1\right)
\end{aligned}$$

due to the choice of δ. The function $e^{-x} - 1$ is negative on $\mathbb{R}_+$, so the expression in the last line tends to $-\infty$ for fixed $u > 0$.

For $\alpha = 2 - \gamma$, we estimate

$$\begin{aligned}
&u\left(-2k+\gamma\log(N)\right)+k\alpha\left(\frac{1}{2}e^{-2u}-\frac{1}{2}+u\right)\\
=&u\left(-(\gamma+\alpha)k+\gamma\log(N)\right)+k\alpha\left(\frac{1}{2}e^{-2u}-\frac{1}{2}+u\right)\\
\leq&u\left(-\gamma(1-\delta)\log(N)+\gamma\log(N)\right)+(1-\delta)\log(N)\alpha\left(\frac{1}{2}e^{-2u}-\frac{1}{2}\right)
\end{aligned}$$

and choose $u = -1/2\log(\delta\gamma/((1-\delta)\alpha))$, if $\delta < \alpha/(\alpha+\gamma)$. As $\mathbb{P}(A_\delta)$ tends to 1 for each $\delta > 0$, this finally implies that

$$\mathbb{P}(\exists i \leq N : \xi_i^1 \neq 0, T_i(\xi^1) \neq \xi_i^1) \leq \mathbb{P}(A_\delta)(1+\delta)\log(N)\mathbb{P}(T_1(\xi^1) \neq \xi_1^1|\mathcal{Z}_k) \longrightarrow 0$$

for

$$0 < \gamma < 2$$

without a further condition on α.

In the second part of the proof we show the sharpness of the bound on α. It is based on the following observations:

1. The probability of the event A_δ converges to 1, for each $\delta > 0$.

2. The random variables

$$X_1(k) = \sum_{\mu=2}^{M} \mathbb{1}_{\{\sum_{j\le k} |\xi_j^\mu|=1\}}, \quad X_2(k) = \sum_{\mu=2}^{M} \mathbb{1}_{\{\sum_{j\le k} |\xi_j^\mu|=2\}}, \quad X_3(k) = \sum_{\mu=2}^{M} \mathbb{1}_{\{\sum_{j\le k} |\xi_j^\mu|>2\}}$$

are Binomially distributed with parameters $M-1$ and

$$p_1(k) = kp(1-p)^{k-1}, \quad p_2(k) = \binom{k}{2} p^2(1-p)^{k-2} \quad \text{and} \quad p_3(k) = \binom{k}{3} p^3 + \mathcal{O}(k^4p^4),$$

respectively, as in the proof of Theorem 3.1. We saw in Chapter 3 that the Chebyshev inequality implies for each $\delta > 0$

$$\mathbb{P}\left(\frac{X_1(k)}{\alpha k \frac{N}{\log(N)}} \notin (1-\delta, 1+\delta) \right) \le \frac{\log(N)}{(\delta - (k-1)p + \mathcal{O}(k^3p))^2 \alpha k N},$$

which tends to 0 as N, k tend to infinity for the given choice of $k \le (1+\delta)\log(N)$, as well as

$$\mathbb{P}\left(\frac{X_2(k)}{\alpha\binom{k}{2}} \notin (1-\delta, 1+\delta) \right) \le \frac{1}{(\delta - p(k-2) + \mathcal{O}(k^4p^2))^2 \alpha \binom{k}{2}} \longrightarrow 0$$

as N tends to infinity. Additionally one sees immediately

$$\mathbb{P}\left(X_3(k) \ne 0\right) \le M p_3(k) = M\left(\binom{k}{3} p^3 + \mathcal{O}(k^4p^4) \right) \longrightarrow 0$$

as N and $k \le (1+\delta)\log(N)$ tend to infinity.

The corresponding complementary sets

$$\left\{ \frac{X_1(k)}{\alpha k \frac{N}{\log(N)}} \in (1-\delta, 1+\delta) \right\}, \quad \left\{ \frac{X_2(k)}{\alpha\binom{k}{2}} \in (1-\delta, 1+\delta) \right\}, \quad \{X_3(k) = 0\}$$

are denoted by $B_\delta(k)$, $C_\delta(k)$ and $D(k)$.

3. For each $n \in \mathbb{N}$, $n \le M-1$ and arbitrary $i > k$, we estimate

$$\mathbb{P}\left[\sum_{\mu:\sum_{j\le k} |\xi_j^\mu|=2} \sum_{j\le k} |\xi_j^\mu \xi_i^\mu| > 0 \middle| X_2(k) = n \right] \le np.$$

This yields

$$\max_{\substack{k,n\in\mathbb{N}: k/\log(N),\\ n/(\alpha\binom{k}{2})\in(1-\delta,1+\delta)}} \mathbb{P}\left[\sum_{\mu:\sum_{j\le k} |\xi_j^\mu|=2} \sum_{j\le k} |\xi_j^\mu \xi_i^\mu| > 0 \middle| X_2(k) = n \right] \le (1+\delta)^3 \alpha \frac{\log(N)^3}{2N}. \tag{5.18}$$

This means that neuron $i > k$ is never activated in any message with more than one of the activated neurons of message 1, with high probability.

The event

$$\left\{ \sum_{\mu : \sum_{j\le k} |\xi_j^\mu| = 2} \sum_{j \le k} |\xi_j^\mu \xi_i^\mu| = 0 \right\}$$

is denoted by $C(k,i)$.

4. Let $X_4(k,i)$ be, for some $i > k$, defined as

$$X_4(k,i) := \sum_{\mu=2}^{M} \mathbb{1}_{\{\sum_{j\le k} |\xi_j^\mu| = 0\}} \mathbb{1}_{\{\xi_i^\mu \neq 0\}}.$$

The conditional distribution of $X_4(k,i)$, given $\mathcal{F}_k^N = \sigma(\xi_j^\mu, \mu \geq 2, j \leq k)$ is Binomial with parameters $M - 1 - X_1(k) - X_2(k) - X_3(k)$ and p. The conditional probability of

$$E_\delta(k,i) := \left\{ \frac{X_4(k,i)\log(N)}{\alpha N} \in (1-\delta, 1+\delta) \right\},$$

given the intersection of the sets $B_\delta(k)$, $C_\delta(k)$, $D(k)$ is at least

$$\begin{aligned} & \mathbb{P}\left(E_\delta(k,i) | B_\delta(k) \cap C_\delta(k) \cap D(k)\right) \\ \geq & \min_{B_\delta(k)\cap C_\delta(k) \cap D(k)} \mathbb{P}(E_\delta(k,i) \big| \mathcal{F}_k^N) \\ = & \min_{\substack{m,n\in\mathbb{N}: m\log(N)/(\alpha k N), \\ n/(\alpha\binom{k}{2})\in(1-\delta,1+\delta)}} \mathbb{P}\left(E_\delta(k,i) | X_1(k) = m, X_2(k) = n, X_3(k) = 0\right). \end{aligned} \tag{5.19}$$

Conditionally on $\{X_1(k) = m, X_2(k) = n, X_3(k) = 0\}$, $X_4(k,i)$ is a Binomially distributed random variable with parameters $M - 1 - m - n$ and p; applying the exponential Chebyshev inequality yields

$$\begin{aligned} & 1 - \max_{\substack{m,n\in\mathbb{N}: m\log(N)/(\alpha k N), \\ n/(\alpha\binom{k}{2})\in(1-\delta,1+\delta)}} \mathbb{P}\left[\frac{X_4(k,i)}{Mp} \geq 1+\delta \Big| X_1(k) = m, X_2(k) = n, X_3(k) = 0 \right] \\ \geq & 1 - \max_{\substack{m,n\in\mathbb{N}: m\log(N)/(\alpha k N), \\ n/(\alpha\binom{k}{2})\in(1-\delta,1+\delta)}} \exp[-t(1+\delta)Mp](1-p+pe^t)^{M-1-m-n} \\ \geq & 1 - \exp[-t(1+\delta)Mp](1-p+pe^t)^M \geq 1 - \exp[(-\log(1+\delta)(1+\delta)+\delta)Mp], \end{aligned} \tag{5.20}$$

as well as

$$\begin{aligned} & 1 - \max_{\substack{m,n\in\mathbb{N}: m\log(N)/(\alpha k N), \\ n/(\alpha\binom{k}{2})\in(1-\delta,1+\delta)}} \mathbb{P}\left[\frac{X_4(k,i)}{Mp} \leq 1-\delta \Big| X_1(k) = m, X_2(k) = n, X_3(k) = 0 \right] \\ \geq & 1 - \exp[t(1-\delta)Mp](1-p+pe^{-t})^{M-1-(1+\delta)\alpha\left(Nk/\log(N)+\binom{k}{2}\right)} \end{aligned}$$

$$\geq 1 - \exp[(-\log(1-\delta)(1-\delta)-\delta)Mp]\exp\left[\delta p + \delta(1+\delta)\alpha\left(k + \binom{k}{2}p\right)\right]. \tag{5.21}$$

In particular, $N \cdot \mathbb{P}(E_\delta(k,i)|B_\delta(k) \cap C_\delta(k) \cap D(k)) \longrightarrow 0$.

5. As penultimate point, we observe that

$$X_5(k,i) := \sum_{\mu:\sum_{j\leq k}|\xi_j^\mu|=1} |\xi_i^\mu|,$$

is, given $\mathcal{F}_k^N$, Binomially distributed with parameters p and $X_1(k)$. Assuming that $k \leq (1+\delta)\log(N)$ and that $X_1(k) \leq (1+\delta)\alpha(kN/\log(N))$, the random variable $X_5(k,i)$ is asymptotically Poisson distributed with Parameter $pX_1(k)$. Referring to Lemma 2.3 and again denoting by $\pi_\lambda(m)$, $\lambda > 0, m \in \mathbb{N}$, the probability weights of a Poisson random variable, the total variation distance of these two distributions is for some $i > k$ at most

$$\sum_{m=0}^{\infty}\left|\mathbb{P}\left(\sum_{\mu:\sum_{j\leq k}|\xi_j^\mu|=1}|\xi_i^\mu| = m\middle|\mathcal{F}_k^N\right) - \pi_{pX_1(k)}(m)\right| \leq 2p^2X_1(k).$$

6. Finally, the events

$$\left\{\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_i^\mu\xi_j^\mu + \frac{1}{(1-p)^2}\eta_i^\mu\eta_j^\mu < \gamma\log(N)\right\}, \quad i > k,$$

are conditionally independent, given $\mathcal{F}_k^N$.

First, we estimate

$$\begin{aligned}
&\mathbb{P}\left(\exists i \leq N : T_i(\xi^1) \neq 0\right) \geq \mathbb{P}\left(\exists i \leq N : \xi_i^1 = 0, T_i(\xi^1) \neq 0\right)\\
\geq&\mathbb{P}(A_\delta)\min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}}\mathbb{P}\left(\exists i \leq N : \xi_i^1 = 0, T_i(\xi^1) \neq 0|\mathcal{Z}_k\right)\\
=&\mathbb{P}(A_\delta)\min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}}\mathbb{P}\left(\exists i > k : \sum_{\mu=2}^{M}\sum_{j\leq N}\xi_j^1\xi_i^\mu\xi_j^\mu + \frac{1}{(1-p)^2}\left(\xi_j^1\right)^2\eta_i^\mu\eta_j^\mu \geq \gamma\log(N)\middle|\mathcal{Z}_k\right)\\
=&\mathbb{P}(A_\delta)\min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}}\mathbb{P}\left(\exists i > k : \sum_{\mu=2}^{M}\sum_{j\leq k}\xi_i^\mu\xi_j^\mu + \frac{1}{(1-p)^2}\eta_i^\mu\eta_j^\mu \geq \gamma\log(N)\right);
\end{aligned} \tag{5.22}$$

because conditionally on $\mathcal{Z}_k$,

$$\left(\sum_{\mu=2}^{M}\sum_{j\leq N}\xi_j^1\xi_i^\mu\xi_j^\mu + \frac{1}{(1-p)^2}\left(\xi_j^1\right)^2\eta_i^\mu\eta_j^\mu,\ i > k\right) \sim \left(\sum_{\mu=2}^{M}\sum_{j\leq k}\xi_i^\mu\xi_j^\mu + \frac{1}{(1-p)^2}\eta_i^\mu\eta_j^\mu,\ i > k\right),$$

as we saw in the first part of the proof.

So we can conclude, writing $B_\delta \cap C_\delta \cap D(k) := B_\delta(k) \cap C_\delta(k) \cap D(k)$ for short:

$$\begin{aligned}
&\mathbb{P}\left(\exists i > k : \sum_{\mu=2}^{M} \sum_{j \le k} \xi_i^\mu \xi_j^\mu + \frac{1}{(1-p)^2} \eta_i^\mu \eta_j^\mu \ge \gamma \log(N)\right) \\
=&1 - \mathbb{E}_{(\xi_j^\mu, j \le k, \mu \ge 2)}\left[\mathbb{P}\left(\sum_{\mu=2}^{M} \sum_{j \le k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2} \eta_N^\mu \eta_j^\mu < \gamma \log(N) \Big| \mathcal{F}_k^N\right)^{N-k}\right] \\
\ge&1 - \max_{\substack{B_\delta \cap C_\delta \\ \cap D(k)}} \mathbb{P}\left(\sum_{\mu=2}^{M} \sum_{j \le k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2} \eta_N^\mu \eta_j^\mu < \gamma \log(N) \Big| \mathcal{F}_k^N\right)^{N-k} - \mathbb{P}[(B_\delta \cap C_\delta \cap D(k))^c].
\end{aligned} \tag{5.23}$$

For every k in the considered set, the probability of $B_\delta(k) \cap C_\delta(k) \cap D(k)$ tends to 1 and we analyse the second part of the last line. The sum

$$\sum_{\mu=2}^{M} \sum_{j \le k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2} \eta_N^\mu \eta_j^\mu$$

can be split into the sums

$$\begin{aligned}
&\sum_{\mu: \sum_{j \le k} |\xi_j^\mu| = 0} \sum_{j \le k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2} \eta_N^\mu \eta_j^\mu + \sum_{\mu: \sum_{j \le k} |\xi_j^\mu| = 1} \sum_{j \le k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2} \eta_N^\mu \eta_j^\mu \\
+ &\sum_{\mu: \sum_{j \le k} |\xi_j^\mu| = 2} \sum_{j \le k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2} \eta_N^\mu \eta_j^\mu + \sum_{\mu > 2: \sum_{j \le k} |\xi_j^\mu| > 2} \sum_{j \le k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2} \eta_N^\mu \eta_j^\mu.
\end{aligned}$$

On $B_\delta(k) \cap C_\delta(k) \cap C(k,i) \cap D(k) \cap E_\delta(k,i)$, the last sum is zero; the penultimate sum is

$$\sum_{\mu: \sum_{j \le k} |\xi_j^\mu| = 2} \sum_{j \le k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2} \eta_N^\mu \eta_j^\mu = \left(-\frac{2p}{1-p} + (k-2)\frac{p^2}{(1-p)^2}\right) X_2(k) \tag{5.24}$$

and therefore tends to 0 on this set, because $X_2(k)/(\alpha\binom{k}{2}) \in (1-\delta, 1+\delta)$.

Concerning the first sum, we state that

$$\sum_{\mu=2}^{M} \mathbb{1}_{\sum_{j \le k} |\xi_j^\mu| = 0} = M - 1 - X_1(k) - X_2(k) - X_3(k)$$

and thus

$$\begin{aligned}
&\sum_{\mu: \sum_{j \le k} |\xi_j^\mu| = 0} \sum_{j \le k} \xi_N^\mu \xi_j^\mu + \frac{\eta_N^\mu \eta_j^\mu}{(1-p)^2} \\
&= [M - 1 - X_1(k) - X_2(k) - X_3(k) - X_4(k,N)]\, k \frac{p^2}{(1-p)^2} - X_4(k,N) k \frac{p}{1-p}
\end{aligned}$$

$$\geq \frac{Mkp^2}{(1-p)^2} - X_4(k,N)k\frac{p}{1-p} \geq \alpha k \frac{1}{(1-p)^2} - \alpha k \frac{1+\delta}{1-p} \tag{5.25}$$

on the considered set.

The remaining second sum is equal to

$$\sum_{\mu:\sum_{j\leq k}|\xi_j^\mu|=1} \sum_{j\leq k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2}\eta_N^\mu \eta_j^\mu$$
$$= \sum_{\substack{\mu:\sum_{j\leq k}|\xi_j^\mu| \\ =1,|\xi_N^\mu|=1}} \left(1 - p\frac{k-1}{1-p} + \sum_{j\leq k} \xi_N^\mu \xi_j^\mu\right) + \sum_{\substack{\mu:\sum_{j\leq k}|\xi_j^\mu| \\ =1,|\xi_N^\mu|=0}} \sum_{j\leq k} \frac{\eta_N^\mu \eta_j^\mu}{(1-p)^2}. \tag{5.26}$$

The right hand side of the last line in (5.26) is at least

$$\sum_{\substack{\mu:\sum_{j\leq k}|\xi_j^\mu| \\ =1,|\xi_N^\mu|=0}} \sum_{j\leq k} \frac{\eta_N^\mu \eta_j^\mu}{(1-p)^2} = (X_1(k) - X_5(k,N))\left((k-1)\frac{p^2}{(1-p)^2} - \frac{p}{1-p}\right) \geq X_1(k)\frac{-p}{1-p} \tag{5.27}$$

because $X_1(k) \geq X_5(k,N)$. It remains to examine the left hand side of the last line in (5.26); this is the last step in the proof.

Due to the results of (5.22) and (5.23), our goal is to show that

$$\lim_{N\to\infty} \max_{\substack{k/\log(N)\in \\ (1-\delta,1+\delta)}} \max_{B_\delta\cap C_\delta\cap D(k)} \left[1 - \mathbb{P}\left(\sum_{\mu=2}^{M}\sum_{j\leq k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2}\eta_N^\mu \eta_j^\mu \geq \gamma \log(N) \middle| \mathcal{F}_k^N\right)\right]^{N-k} = 0.$$

This is fulfilled if

$$\lim_{N\to\infty} \min_{\substack{k/\log(N)\in \\ (1-\delta,1+\delta)}} \min_{B_\delta\cap C_\delta\cap D(k)} \frac{\log\left[\mathbb{P}\left(\sum_{\mu=2}^{M}\sum_{j\leq k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2}\eta_N^\mu \eta_j^\mu \geq \gamma \log(N) \middle| \mathcal{F}_k^N\right)\right]}{\log(N)} > -1. \tag{5.28}$$

Due to the results in (5.24), (5.25), (5.26) and (5.27), the sum under consideration is, on $B_\delta(k) \cap C_\delta(k) \cap C(k,i) \cap D(k) \cap E_\delta(k,i)$, at least

$$\sum_{\mu=2}^{M}\sum_{j\leq k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2}\eta_N^\mu \eta_j^\mu \geq \left(-\frac{2p}{1-p} + (k-2)\frac{p^2}{(1-p)^2}\right)(1-\delta)\alpha\binom{k}{2} + \alpha k \frac{1}{(1-p)^2}$$
$$- \alpha k \frac{1+\delta}{1-p} - (1+\delta)\alpha k \frac{N}{\log(N)}\frac{p}{1-p} + \sum_{\substack{\mu:\sum_{j\leq k}|\xi_j^\mu| \\ =1,|\xi_N^\mu|=1}} \left(1 - p\frac{k-1}{1-p} + \sum_{j\leq k} \xi_N^\mu \xi_j^\mu\right)$$
$$\geq \alpha k \frac{-1-2\delta}{1-p} + \mathcal{O}(k^2 p) + \sum_{\substack{\mu:\sum_{j\leq k}|\xi_j^\mu| \\ =1,|\xi_N^\mu|=1}} \left(1 - p\frac{k-1}{1-p} + \sum_{j\leq k} \xi_N^\mu \xi_j^\mu\right). \tag{5.29}$$

We have seen in 4., (5.19), (5.20) and (5.21), using $\varepsilon_\delta = \min(\log(1+\delta)(1+\delta)+\delta; \log(1-\delta)(1-\delta)-\delta) > 0$, that

$$\max_{B_\delta(k)\cap C_\delta(k)\cap D(k)} \mathbb{P}(E_\delta(k,N)^c|\mathcal{F}_k^N) \leq \exp\left[-\varepsilon_\delta \alpha \frac{N}{\log(N)} + \mathcal{O}(k)\right]$$

and in (5.18), that

$$\max_{B_\delta(k)\cap C_\delta(k)\cap D(k)} \mathbb{P}(C(k,N)^c|\mathcal{F}_k^N) \leq (1+\delta)^3 \alpha \frac{\log(N)^3}{N}.$$

So we can conclude for the probability in (5.28), using additionally (5.29),

$$\begin{aligned}
&\min_{B_\delta\cap C_\delta\cap D(k)} \mathbb{P}\left(\sum_{\mu=2}^{M}\sum_{j\leq k} \xi_N^\mu \xi_j^\mu + \frac{1}{(1-p)^2}\eta_N^\mu \eta_j^\mu \geq \gamma \log(N) \big| \mathcal{F}_k^N\right) \\
&\geq \min_{B_\delta(k)} \mathbb{P}\Bigg(\sum_{\substack{\mu: \sum_{j\leq k}|\xi_j^\mu| \\ =1, |\xi_N^\mu|=1}} \left(1 - p\frac{k-1}{1-p} + \sum_{j\leq k} \xi_N^\mu \xi_j^\mu\right) \geq \gamma\log(N) + \frac{\alpha k(1+2\delta)}{1-p} + \mathcal{O}(k^2 p) \Big| \mathcal{F}_k^N \Bigg) \\
&\qquad - \exp\left[-\varepsilon_\delta \alpha \frac{N}{\log(N)} + \mathcal{O}(k)\right] - (1+\delta)^3 \alpha \frac{\log(N)^3}{N}.
\end{aligned} \tag{5.30}$$

Finally we consider the behaviour of the summand on the left hand side in the last line of (5.26), that is,

$$\sum_{\substack{\mu: \sum_{j\leq k}|\xi_j^\mu| \\ =1, |\xi_N^\mu|=1}} \left(1 - p\frac{k-1}{1-p} + \sum_{j\leq k} \xi_N^\mu \xi_j^\mu\right). \tag{5.31}$$

It is, conditionally on $\xi_j^\mu, j \leq k, \mu \geq 2$ and $|\xi_N^\mu|, \mu \geq 2$, distributed as a sum of $X_5(k,N)$ independent and identically distributed random variables $Z_n(p,k)$, $n \geq 1$, with distribution

$$\mathbb{P}\left(Z_n(p,k) = -p\frac{k-1}{1-p}\right) = \mathbb{P}\left(Z_n(p,k) = 2 - p\frac{k-1}{1-p}\right) = \frac{1}{2}.$$

$X_5(k,N)$ has still to be determined and is, given $X_1(k)$, Binomially distributed with parameters $X_1(k)$ and p. Consequently, the sum in (5.31) is, given $X_1(k)$, distributed as a random sum of random variables $Z_n(p,k)$, $n \geq 1$, and length determined by a $\mathrm{Bin}(X_1(k), p)$-distributed random variable $R_{X_1(k)}$, such that $R_{X_1(k)}, Z_1(p,k), \ldots$ are independent. This Binomial distribution can again be approximated by a Poisson distribution; the subsequent computations are therefore made for a Poisson distribution instead. Let additionally for each $\varepsilon \geq 0$ $Z_n(\varepsilon)$, $n \geq 1$, be independent and identically distributed such that

$$\mathbb{P}(Z_n(\varepsilon) = -\varepsilon) = \mathbb{P}(Z_n(\varepsilon) = 2 - \varepsilon) = \frac{1}{2}.$$

Let now Y_λ denote a Poisson random variable with parameter λ. For independent sets of random variables $(Y_{\lambda k}, Z_n(\varepsilon), n \geq 1)$, respectively $(Y_\lambda^r, Z_n^r(\varepsilon), r = 1, \ldots, k, n \geq 1)$, with $Y_\lambda^r \sim \mathrm{Poi}(\lambda)$, $Z_n^r(\varepsilon) \sim Z_1(\varepsilon)$, for each r and $n \geq 1$, the two following sums have the same distribution:

$$\sum_{n=1}^{Y_{\lambda k}} Z_n(\varepsilon) \sim \sum_{r=1}^{k} \sum_{n=1}^{Y_\lambda^r} Z_n^r(\varepsilon).$$

The $\sum_{n=1}^{Y_\lambda^r} Z_n^r(\varepsilon)$, $1 \leq r \leq k$, then are independent. This yields, with the help of Lemma 2.4:

$$\lim_{k\to\infty} \frac{1}{k} \log\left(\mathbb{P}\left(\sum_{n=1}^{Y_{\lambda k}} Z_n(\varepsilon) \geq xk\right)\right) = \lim_{k\to\infty} \frac{1}{k} \log\left(\mathbb{P}\left(\sum_{r=1}^{k}\sum_{n=1}^{Y_\lambda^r} Z_n^r(\varepsilon) \geq xk\right)\right) = -\Lambda_{\lambda,\varepsilon}^*(x)$$

with

$$\Lambda_{\lambda,\varepsilon}^*(x) = \sup_{t\in\mathbb{R}}\ tx - \log\left[\mathbb{E}\left(\exp t \sum_{n=1}^{Y_\lambda^1} Z_n^1(\varepsilon)\right)\right]$$

and if

$$x > \mathbb{E}\left(\sum_{n=1}^{Y_\lambda^1} Z_n^1(\varepsilon)\right), \quad \mathbb{E}\left(\exp t \sum_{n=1}^{Y_\lambda^1} Z_n^1(\varepsilon)\right) < \infty, \quad t \in \mathbb{R}.$$

By Wald's identity, the expectation of this sum is

$$\mathbb{E}\left(\sum_{n=1}^{Y_\lambda^1} Z_n^1(\varepsilon)\right) = \mathbb{E}\left(Y_\lambda^1\right)\mathbb{E}(Z_1^1(\varepsilon)) = \lambda(1-\varepsilon).$$

To determine the Legendre transform, consider the moment generating function of $\sum_{n=1}^{Y_\lambda^1} Z_n^1(\varepsilon)$:

$$\begin{aligned}\Lambda_{\lambda,\varepsilon}(t) :=& \mathbb{E}\left(\exp t \sum_{n=1}^{Y_\lambda^1} Z_n^1(\varepsilon)\right) = \sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!}\frac{1}{2^m} e^{-\varepsilon t m}\left(1+e^{2t}\right)^m \\ =& \exp\left[-\lambda + \frac{\lambda}{2}e^{-\varepsilon t}(e^{2t}+1)\right].\end{aligned}$$

Then

$$\Lambda_{\lambda,\varepsilon}^*(x) = \sup_{t\in\mathbb{R}}\left(tx - \log\left[\mathbb{E}\left(\exp t \sum_{n=1}^{Y_\lambda^1} Z_n^1(\varepsilon)\right)\right]\right) = \sup_{t\in\mathbb{R}}\left(tx + \lambda - \frac{\lambda}{2}e^{-\varepsilon t}(e^{2t}+1)\right).$$

We are for fixed $x > \lambda$ interested in $\lim_{\varepsilon\searrow 0}\Lambda_{\lambda,\varepsilon}^*(x)$. First, we observe that for $t \leq 0$,

$$tx + \lambda - \frac{\lambda}{2}e^{-\varepsilon t}(e^{2t}+1) \leq tx + \lambda - \frac{\lambda}{2}(e^{2t}+1) = tx - \log\left(\Lambda_{\lambda,0}(t)\right).$$

Define for fixed λ and $x > \lambda$ the continuous and differentiable functions $\psi_\varepsilon : \mathbb{R} \to \mathbb{R}$,

$$\psi_\varepsilon(t) := tx + \lambda - \frac{\lambda}{2} e^{-\varepsilon t}(e^{2t} + 1).$$

We easily compute for $0 < \varepsilon < 2$

$$\psi_\varepsilon'(t) = x + \varepsilon \frac{\lambda}{2} e^{-\varepsilon t} - (2-\varepsilon)\frac{\lambda}{2} e^{(2-\varepsilon)t}; \quad \lim_{t\to\infty} \psi_\varepsilon'(t) = -\infty; \quad \lim_{t\to-\infty} \psi_\varepsilon'(t) = \infty.$$

In addition, for this choice of ε, $\psi_\varepsilon''(t) < 0$ on $\mathbb{R}$. So

$$\sup_{t\in\mathbb{R}} \psi_\varepsilon(t) = \psi_\varepsilon(t_\varepsilon),$$

with the unique root t_ε of $\psi_\varepsilon'(t)$. For each $t \le 0$, $\varepsilon \ge 0$,

$$\psi_\varepsilon'(t) \ge x + \varepsilon \frac{\lambda}{2} - (2-\varepsilon)\frac{\lambda}{2} \ge x - \lambda > 0$$

and for each $t > \log(1 + 2x/\lambda)$ and $0 \le \varepsilon \le 1$,

$$\psi_\varepsilon'(t) \le x + \frac{\lambda}{2} - \frac{\lambda}{2} e^t < 0.$$

So for $x > \lambda$, $0 \le \varepsilon \le 1$

$$\sup_{t\in\mathbb{R}} \psi_\varepsilon(t) = \sup_{t\in[0,\log(1+2x/\lambda)]} \psi_\varepsilon(t).$$

On $[0, \log(1 + 2x/\lambda)]$, ψ_ε is uniformly convergent to ψ_0 as $\varepsilon \to 0$. So we can conclude

$$\lim_{\varepsilon\searrow 0} t_\varepsilon = t_0, \quad \lim_{\varepsilon\searrow 0} \psi_\varepsilon(t_\varepsilon) = \psi_0(t_0).$$

We finally determine the Legendre transform $\Lambda^*_{\lambda,0}$:

$$\begin{aligned}\Lambda^*_{\lambda,0}(x) =& \sup_{t\in\mathbb{R}} \; tx - \log\left[\mathbb{E}\left(\exp t \sum_{n=1}^{Y^1_\lambda} Z^1_n(0)\right)\right] \\ =& \frac{1}{2}\log\left(\frac{x}{\lambda}\right) x - \frac{\lambda}{2}\left(\frac{x}{\lambda} - 1\right),\end{aligned}$$

using $t_0 = \frac{1}{2}\log\left(\frac{x}{\lambda}\right)$.

We combine the recent conclusions and deduce for $\varepsilon > p(k-1)/(1-p)$, using in the first step 5. and that $\alpha k(1+2\delta)/(1-p) = \alpha k(1+2\delta) + \mathcal{O}(kp)$,

$$\begin{aligned}&\min_{B_\delta(k)} \mathbb{P}\left(\sum_{\substack{\mu:\sum_{j\le k}|\xi^\mu_j| \\ =1,|\xi^\mu_N|=1}} \left(1 - p\frac{k-1}{1-p} + \sum_{j\le k} \xi^\mu_N \xi^\mu_j\right) \ge \gamma\log(N) + \frac{\alpha k(1+2\delta)}{1-p} + \mathcal{O}(k^2 p) \middle| \mathcal{F}^N_k \right) \\ \ge& \min_{B_\delta(k)} \mathbb{P}\left(\sum_{n=1}^{Y_{p\cdot X_1(k)}} Z_n(k,p) \ge \gamma\log(N) + \alpha k(1+2\delta) + \mathcal{O}(k^2 p) \middle| \mathcal{F}^N_k \right) - 2p^2 X_1(k)\end{aligned}$$

$$\geq \min_{\rho\in(1-\delta,1+\delta)} \mathbb{P}\left(\sum_{n=1}^{Y_{\alpha\rho k}} Z_n(\varepsilon) \geq \gamma \log(N) + \alpha k(1+2\delta) + \mathcal{O}(k^2 p)\right) - 2p\alpha\rho k.$$

Since $2p\alpha\rho k \leq 2(1+\delta)^2 \alpha \log(N)/N$ and due to the estimation in (5.30), it suffices to show

$$\liminf_{N\to\infty} \min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \min_{\substack{\rho\in(1-\delta,\\ 1+\delta)}} \frac{\log\left[\mathbb{P}\left(\sum_{n=1}^{Y_{\alpha\rho k}} Z_n(\varepsilon) \geq \gamma\log(N) + \alpha k(1+2\delta) + \mathcal{O}(k^2p)\right)\right]}{\log(N)} > -1$$

to obtain convergence of the probability of instability of ξ^1 to 1. We analysed the behaviour of the random variable $\sum_{n=1}^{Y_{\lambda k}} Z_n(\varepsilon)$ and saw that

$$\begin{aligned}&\lim_{k\to\infty} \frac{\log\left(\mathbb{P}\left[\sum_{n=1}^{Y_{\rho\alpha k}} Z_n(\varepsilon) \geq (\frac{\gamma}{1-\delta} + \alpha(1+2\delta))k\right]\right)}{k}\\ &= -\Lambda^*_{\rho\alpha,\varepsilon}\left(\frac{\gamma}{1-\delta} + \alpha(1+2\delta)\right)\end{aligned}$$

if $\frac{\gamma}{1-\delta} + \alpha(1+2\delta) > \rho\alpha(1-\varepsilon)$. We saw that the maximal argument t_ε of ψ_ε was positive; one easily sees that $tx - \log(\Lambda_{\lambda,\varepsilon}(t))$ is decreasing in λ for fixed $\varepsilon < 1/2$ and $t > 0$, and therefore

$$\begin{aligned}&\min_{\substack{\rho\in(1-\delta,\\ 1+\delta)}} \liminf_{N\to\infty} \min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \frac{\log\left[\mathbb{P}\left(\sum_{n=1}^{Y_{\alpha\rho k}} Z_n(\varepsilon) \geq \gamma\log(N) + \alpha k(1+2\delta)\right)\right]}{\log(N)}\\ &\geq (1-\delta) \min_{\substack{\rho\in(1-\delta,\\ 1+\delta)}} \liminf_{N\to\infty} \min_{\substack{k\in\mathbb{N}:k/\log(N)\\ \in(1-\delta,1+\delta)}} \frac{\log\left[\mathbb{P}\left(\sum_{n=1}^{Y_{\alpha\rho k}} Z_n(\varepsilon) \geq \frac{\gamma}{1-\delta}k + \alpha k(1+2\delta)\right)\right]}{k}\\ &= (1-\delta) \min_{\rho\in(1-\delta,1+\delta)} \left(-\Lambda^*_{\rho\alpha,\varepsilon}\left(\frac{\gamma}{1-\delta} + \alpha(1+2\delta)\right)\right)\\ &\geq (1-\delta)\left(-\Lambda^*_{(1-\delta)\alpha,\varepsilon}\left(\frac{\gamma}{1-\delta} + \alpha(1+2\delta)\right)\right).\end{aligned}$$

Due to our considerations on $Z_n(\varepsilon)$, there is some $\varepsilon' > 0$ such that for all $0 \leq \varepsilon < \varepsilon'$,

$$-(1-\delta)\Lambda^*_{(1-\delta)\alpha,\varepsilon}\left(\frac{\gamma}{1-\delta} + \alpha(1+2\delta)\right) > -1,$$

if

$$-(1-\delta)\Lambda^*_{(1-\delta)\alpha,0}\left(\frac{\gamma}{1-\delta} + \alpha(1+2\delta)\right) > -1.$$

Since additionally $\Lambda^*_{(1-\delta)\alpha,0}(x)$ is continuous on $\mathbb{R}_+$, this condition is sufficient to guarantee the convergence of the probability of the instability of ξ^1 to 1 (the term $\mathcal{O}(k^2p)$ in the probability $\mathbb{P}\left(\sum_{n=1}^{Y_{\alpha\rho k}} Z_n(\varepsilon) \geq \gamma\log(N) + \alpha k(1+2\delta) + \mathcal{O}(k^2p)\right)$ is neglibile).

Using the function

$$f_{\alpha,(1-\delta)^2}(x) = 1 + \frac{1}{2}(1-\delta)^2\alpha(-x\log(x) + x - 1)$$

of the first part of the proof and

$$\bar{x}_{\alpha,\gamma,\delta} := \frac{\gamma}{\alpha(1-\delta)^2} + \frac{1+2\delta}{1-\delta},$$

a sufficient condition for the convergence of $\mathbb{P}(\exists i : T_i(\xi^1) \neq \xi_i^1)$ to 1 is

$$f_{\alpha,(1-\delta)^2}(\bar{x}_{\alpha,\gamma,\delta}) > 0.$$

Finally, if α and γ fulfill

$$f_{\alpha,1}(\bar{x}_{\alpha,\gamma,0}) = 1 + \frac{1}{2}\alpha\left[-\log\left(1+\frac{\gamma}{\alpha}\right)(1+\gamma/\alpha) + \frac{\gamma}{\alpha}\right] > 0, \tag{5.32}$$

$\delta > 0$ can be chosen small enough to obtain $f_{\alpha,(1-\delta)^2}(\bar{x}_{\alpha,\gamma,\delta}) > 0$. As on page 124, we observe that condition (5.32) is fulfilled if $g_\gamma(\bar{x}_{\alpha,\gamma,0}) > 0$, with $\bar{x}_{\alpha,\gamma,0} = 1 + \frac{\gamma}{\alpha}$ and $g_\gamma(x) = x\left(1 + \frac{2}{\gamma} - \log(x)\right) - 1 - \frac{2}{\gamma}$. Since g_γ is positive on $(1, x_\gamma^*)$, with the root x_γ^* of g_γ in $(1,\infty)$, the precedent condition holds if

$$\bar{x}_{\alpha,\gamma,0} \in (1, x_\gamma^*).$$

As $\bar{x}_{\alpha,\gamma,0} > 1$, this is true if

$$\alpha > \frac{\gamma}{x_\gamma^* - 1}.$$

We conclude that $\lim_{N\to\infty}\mathbb{P}(\exists i \leq N : T_i(\xi^1) \neq \xi_i^1) = 1$ holds if $\alpha > \frac{\gamma}{x_\gamma^*-1}$, as stated in the theorem.

□

Proposition 5.5 *The threshold variable γ cannot be chosen from $(2,\infty)$.*

Proof: Suppose that $\gamma > 2$. Choosing some $\delta < \frac{1}{2}(\gamma - 2)$, we will see that,

$$\lim_{N\to\infty} \max_{\substack{k\in\mathbb{N}:k/\log(N)\in\\(1-\delta,1+\delta)}} \mathbb{P}\left(T_1(\xi^1) \neq \xi_1^1 | \mathcal{Z}_k\right) = 1,$$

independently of the choice of α. Using (5.6), we easily conclude

$$\max_{\substack{k\in\mathbb{N}:k/\log(N)\in\\(1-\delta,1+\delta)}} \mathbb{P}\left(T_1(\xi^1) = \xi_1^1 | \mathcal{Z}_k\right) \leq \max_{\substack{k\in\mathbb{N}:k/\log(N)\in\\(1-\delta,1+\delta)}} \mathbb{P}\left(|S_1(\xi^1)| + \theta_1(\xi^1) \geq \gamma\log(N) | \mathcal{Z}_k\right)$$

$$\leq \max_{\substack{k\in\mathbb{N}:k/\log(N)\in\\(1-\delta,1+\delta)}} \mathbb{P}\left[\left(\sum_{1<j\leq k}\sum_{\mu=2}^{M} \xi_1^\mu\xi_j^\mu + \frac{1}{(1-p)^2}\eta_1^\mu\eta_j^\mu\right) \geq \gamma\log(N) + 2 - 2k\right]$$

$$\le \exp\left[-t\log(N)(\gamma-2-2\delta)\right]\exp\left[(1+\delta)\log(N)\alpha\left(\frac{1}{2}e^{2t}-\frac{1}{2}-t\right)+\mathcal{O}\left(p\log(N)^2\right)\right]$$
$$\le \exp\left[\log(N)\left((1+\delta)\alpha\frac{1}{2}\left(-w_{\gamma,\delta,\alpha}\log(w_{\gamma,\delta,\alpha})+w_{\gamma,\delta,\alpha}-1\right)\right)\right],$$

with $w_{\gamma,\delta,\alpha} := 1+\frac{\gamma-2(1+\delta)}{(1+\delta)\alpha}$ and after having used $t=\frac{1}{2}\log\left(w_{\gamma,\delta,\alpha}\right)>0$. The probability tends to 0.

□

Proposition 5.6 *The critical value* $\gamma^* = \sup\{\gamma>0 : \gamma$ *is an admissible threshold variable*$\}$ *is*

$$\gamma^* = 2.$$

The critical value $\alpha^* = \sup\{\alpha>0 : \alpha$ *is an admissible capacity variable for the model*$\}$ *is*

$$\alpha^* = \frac{2}{x_2^*-1} \approx 0.51.$$

Here x_2^* *is the root of*

$$g_2(x) = x\left(2-\log(x)\right)-2$$

in $(1,\infty)$.

For $\alpha\in(0,\alpha^*)$*, there is* $\gamma^*(\alpha)\in(0,2)$*, such that* $(0,\gamma^*(\alpha))$ *is a set of inadmissible and* $(\gamma^*(\alpha),2)$ *a set of admissible threshold variables for* α*.*

Proof: The first assertion follows immediately from the previous Theorem 5.4 and Proposition 5.5.

The second claim is proven by considering the function $G(\gamma,x)=g_\gamma(x)$ on $\mathbb{R}_+\times\mathbb{R}_{>1}$. This is strictly decreasing in γ for fixed x. Additionally, for fixed γ, we saw that $g_\gamma(x)$ is positive on $(1,x_\gamma^*)$ and negative on (x_γ^*,∞). Thus $\gamma<\gamma'$ implies $x_\gamma^*>x_{\gamma'}^*$. If in particular there is a $\gamma\in(0,2)$ such that α is admissible for γ, each $\gamma\in(\gamma^*(\alpha),2)$, with $\gamma^*(\alpha):=\inf\{\gamma\in(0,2):\alpha$ is admissible capacity variable for $\gamma\}$ is admissible for α.

If now $\alpha>2/(x_2^*-1)$, the inequality $1+\frac{2}{\alpha}<x_2^*$ holds. Choose $\gamma\in(0,2]$. Then, resulting from the two facts that G is decreasing in γ and that $1+\gamma/\alpha\in(1,x_2^*)$,

$$g_\gamma\left(1+\frac{\gamma}{\alpha}\right)\ge g_2\left(1+\frac{\gamma}{\alpha}\right)>0.$$

There is thus no $\gamma\le 2$ that allows to use α as a capacity variable. Since $\gamma>2$ leads in any case to instability, α is inadmissible.

For $\alpha<2/(x_2^*-1)$, the inequality $1+\frac{2}{\alpha}>x_2^*$ is true. So $g_2(1+\frac{2}{\alpha})<0$. We consider

$$g_\gamma\left(1+\frac{\gamma}{\alpha}\right)=f_{\alpha,1}(\bar{x}_{\alpha,\gamma,0}),\quad f_{\alpha,1}(\bar{x}_{\alpha,\gamma,0})=1+\frac{1}{2}\alpha\left[-\log\left(1+\frac{\gamma}{\alpha}\right)(1+\gamma/\alpha)+\frac{\gamma}{\alpha}\right],$$

$\bar{x}_{\alpha,\gamma,0}=1+\gamma/\alpha$, see (5.32); this function is for fixed α continuous as a function in γ on $\mathbb{R}_+$ and there is thus some $\gamma<2$ such that

$$f_{\alpha,1}(\bar{x}_{\alpha,\gamma,0})<0$$

holds, if $f_{\alpha,1}(\bar{x}_{\alpha,2,0})=g_2(1+2/\alpha)<0$. So α is an admissible capacity variable. □

Remark 5.7 *Besides the GB model in Chapter 6, the BEG model has the best maximal capacity. Besides the fact that it has a ternary state space, its good performance is owed to more information processed in the two functions* J_{ij} *and* K_{ij}*. In the sparse models, it is very rare that two neurons are activated together in one pattern. The BEG model uses this important information twice, in* J_{ij} *and* K_{ij} *and a stored pattern benefits from a high signal term.*

Corollary 5.8 (Corollary to Theorem 5.4:) *Suppose that in the BEG model with threshold* $\gamma \in (0,2)$, *there are* $M = \alpha N^2/\log(N)^2$ *patterns stored, with* α *such that* $\alpha < \gamma/(x^*_\gamma - 1)$. *Let* $\tilde{\xi}^\mu$ *be a faulty version of the stored pattern* ξ^μ.

1. *If in* $\tilde{\xi}^\mu$, *there are* $\varrho_1 \log(N)$ *of the active neurons of* ξ^μ *deactivated,* $\varrho_1 < 1 - \gamma/2$, *the pattern is corrected in one step, with high probability.*

2. *If* $\varrho_2 \log(N)$ *neurons are spuriously activated, the pattern is corrected in the first step, with high probability, if*

$$\alpha < \frac{\gamma}{(x^*_\gamma - 1)(1+\varrho_2)}.$$

   Here again x^*_γ *is the root of* g_γ *in* $(1,\infty)$ *as defined in Theorem 5.4.*

3. *If* $\varrho_3 \log(N)$ *neurons are mapped to the opposite value (a 1 is turned into a -1 and the other way around), the pattern is corrected in one step, if* $\varrho_3 < 1 - \frac{\gamma}{2}$ *and* $\varrho_3 < \frac{1}{2}$.

4. *If all these errors are combined, the pattern is mapped to* ξ^μ *in one step with high probability, if*

$$\varrho_1 + \varrho_3 < 1 - \frac{\gamma}{2}, \quad \varrho_1 + 2\varrho_3 < 1$$

   and, in case $\varrho_1 < \varrho_2$, *if* α *fulfills additionally*

$$\alpha < \frac{\gamma}{(x^*_\gamma - 1)(1+\varrho_2 - \varrho_1)}.$$

All bounds are sharp concerning the correction after one step of the parallel dynamics.

Proof: We just refer to the proof of Theorem 5.4 and point out the differences and similarities in the proofs. W.l.o.g. we take $\mu = 1$ and assume that $\mathcal{Z}_k$ holds.

1. The difference between active neurons of ξ^1 that are or are not deleted in $\tilde{\xi}^1$ is negligible. The signal terms of their local field and of the function θ_i is decreased by $\varrho_1 \log(N) - 1$ (respectively $\varrho_1 \log(N)$).

   The noise terms of $S_i(\tilde{\xi}^1)$ and $\theta_i(\tilde{\xi}^1)$ become

$$\sum_{\mu=2}^{M} \sum_{j \neq i, \tilde{\xi}^1_j \neq 0} \tilde{\xi}^1_j \xi^\mu_i \xi^\mu_j \sim \sum_{\mu=2}^{M} \sum_{j \neq i, \tilde{\xi}^1_j \neq 0} \xi^\mu_i \xi^\mu_j \quad \text{and} \quad \sum_{\mu=2}^{M} \sum_{j \neq i, \tilde{\xi}^1_j \neq 0} \frac{1}{(1-p)^2} \eta^\mu_i \eta^\mu_j .$$

The interior sums are of length $k - \varrho_1 \log(N) - 1$ (resp. $k - \varrho_1 \log(N)$) compared to $k-1$ in ξ^1. As in the proof of Theorem 5.4, we observe that these neurons are corrected in one step or remain correct if they have not been corrupted, with high probability, if $\varrho_1 < 1 - \gamma/2$. Note that $\varrho_1 < 1$ and the local fields of the active neurons of ξ^1 have the same sign as their spins, with high probability, if $\tilde{\xi}^1$ is the input. The sharpness is proven analogously to Proposition 5.5.

Concerning the inactive neurons of ξ^1, the signal term of $\theta_i(\tilde{\xi}^1)$ is $-(k - \varrho_1 \log(N))p/(1-p)$ instead of $-kp/(1-p)$ in ξ^1. This difference is not important for the proof; the exponential in (5.6) changes to

$$\exp\left[(k - \varrho_1 \log(N))\alpha\left(\frac{1}{2}e^{2t} - \frac{1}{2} - t\right) + \mathcal{O}(pk)\right],$$

because there are less active neurons in the pattern. This is upper bounded by the term in (5.6). Since α fulfills the stability condition of the model, these neurons remain inactive, with high probability.

2. The local field of the active neurons of ξ^1 is attached by this kind of error in terms of the number of involved neurons in the noise term; it increases from $k-1$ to $k - 1 + \varrho_2 \log(N)$, if there are k active neurons in ξ^1. The signal term of the function $\theta_i(\tilde{\xi}^1)$ does, compared to $\theta_i(\xi^1)$, only change by $\varrho_2 \log(N)\frac{-p}{1-p}$ and the random term is increased by

$$\sum_{\mu=2}^{M} \sum_{j:\tilde{\xi}_j^1 \neq 0, \xi_j^1 = 0} \eta_i^\mu \eta_j^\mu.$$

These changes do not raise problems for the stability of the active neurons of ξ^1; they remain correct with high probability.

For the inactive neurons, the proof is almost the same as the corresponding part of the proof of Theorem 5.4 concerning the stability of the inactive neurons. The signal term of $\theta_i(\tilde{\xi}^1)$ is only increased by $\varrho_2 \log(N)\frac{p^2}{(1-p)^2}$ compared to $\theta_i(\xi^1)$, which is negligible. The noise term of $S_i(\tilde{\xi}^1) + \theta_i(\tilde{\xi}^1)$ includes additionally the random variables

$$\sum_{\mu=2}^{M} \sum_{j \neq i:\tilde{\xi}_j^1 \neq 0, \xi_j^1 = 0} \tilde{\xi}_j^1 \xi_i^\mu \xi_j^\mu + \frac{1}{(1-p)^2}\eta_i^\mu \eta_j^\mu.$$

There are $\varrho_2 \log(N)$ additionally active neurons and the exponential in (5.6) changes to

$$\exp\left[(k + \varrho_2 \log(N))\alpha\left(\frac{1}{2}e^{2t} - \frac{1}{2} - t\right) + \mathcal{O}(pk)\right].$$

The rest of the proof is the same as in the proof of Theorem 5.4. Correction or respectively stability of the inactive neurons is reached if $\alpha < \gamma/[(x_\gamma^* - 1)(1 + \varrho_2)]$. The bound is sharp; this is also proven as the corresponding part of Theorem 5.4, using that there are $k + \varrho_2 \log(N)$ active neurons in $\tilde{\xi}^1$.

3. The local field of the activated neurons of ξ^1 changes, it is

$$\xi_i^1(k - 2\varrho_3 \log(N) + \mathbb{1}_{\tilde{\xi}^\mu \neq \xi_i^1}) + \sum_{\mu=2}^{M} \sum_{j \neq i}^{k} \tilde{\xi}_j^1 \xi_i^\mu \xi_j^\mu,$$

with

$$\sum_{\mu=2}^{M} \sum_{j \neq i}^{k} \tilde{\xi}_j^1 \xi_i^\mu \xi_j^\mu \sim \sum_{\mu=2}^{M} \sum_{j \neq i}^{k} \xi_i^\mu \xi_j^\mu.$$

The difference between active neurons of ξ^1 whose value is multiplied by -1 and those that are not affected by the corruption again is negligible. The function θ_i does not change. So the signal term of $S_i(\tilde{\xi}^1) + \theta_i(\tilde{\xi}^1)$ is decreased by $2\varrho_3 \log(N)$ compared to ξ^1. These neurons are active after one step of the dynamics, with high probability, if $2\varrho_3 < 2 - \gamma$. If $\varrho_3 > 1 - \gamma/2$, an arbitrary active neuron of ξ^1 is deactivated after the first step of the dynamics, with high probability: this follows as in the proof of Proposition 5.5. In addition, $\mathbb{P}(\forall i \leq k : \mathrm{sgn}(S_i(\tilde{\xi}^1)) = \mathrm{sgn}(\xi_i^1) | \mathcal{Z}_k)$ tends to 1, if $\varrho_3 < 1/2$. On the contrary, for an arbitrary active neuron i of ξ^1, we have $\mathbb{P}(\mathrm{sgn}(S_i(\tilde{\xi}^1)) \neq \mathrm{sgn}(\xi_i^1)) \nrightarrow 0$, if $\varrho_3 > 1/2$ (see Chapter 3).

The distribution of the local field of the inactive neurons of ξ^1 is the same in $\tilde{\xi}^1$ and ξ^1 and the function θ_i does not change for these neurons; since α is chosen according to the stability condition of the model, they remain inactive, with high probability.

4. The relevant changes for the excited neurons in ξ^1 are the decreased signal terms of the local field and of the function θ, which are equal to $k - \varrho_1 \log(N) - 2\varrho_3 \log(N)$ and $k - \varrho_1 \log(N) - \varrho_2 \log(N)p/(1-p)$, respectively, instead of $k - 1$. Since an error concerning one of these neurons can only occur if either the random part of $|S_i| + \theta_i$ falls below $\gamma \log(N) - 2k + 2(\varrho_1 + \varrho_3)\log(N) + \varrho_2 \log(N)p/(1-p)$ or if the random part of S_i exceeds $k - \varrho_1 \log(N) - 2\varrho_3 \log(N)$, we observe analogously to the proof of Theorem 5.4 that the corresponding probability vanishes, if $\varrho_1 + \varrho_3 < 1 - \gamma/2$ and if $\varrho_1 + 2\varrho_3 < 1$. If one of these conditions is not fulfilled, an arbitrary active neuron is not stable, with positive probability not concerging to 0 (in the case $\varrho_1 + \varrho_3 > 1 - \gamma/2$ it even tends to 1).

Concerning the inactive neurons, the signal term of θ_i changes, compared to ξ^1, from

$$\frac{-kp}{1-p} \quad \text{to} \quad \frac{-(k - \varrho_1 \log(N))p}{1-p} + \frac{\varrho_2 \log(N)p^2}{(1-p)^2}$$

which is negligible. There are $k - \varrho_1 \log(N) + \varrho_2 \log(N)$ active neurons in $\tilde{\xi}^1$ and the proof works as the corresponding part (stability of the inactive neurons) of Theorem 5.4, using $\alpha(1 + \varrho_2 - \varrho_1)$ instead of α. If $\varrho_2 > \varrho_1$, α must fulfill

$$\alpha < \frac{\gamma}{(x_\gamma^* - 1)(1 + \varrho_2 - \varrho_1)}$$

to guarantee the stability/correction of the inactive neurons of ξ^1 with high probability. If $\varrho_2 > \varrho_1$ and $\alpha > \gamma/[(x_\gamma^* - 1)(1 + \varrho_2 - \varrho_1)]$, the pattern is not corrected after one step of the dynamics, with high probability.

Remark 5.9 *The corruption of active neurons of a stored pattern affects the choice of* γ*: if one wants to correct a high number of errors of this kind, one must choose a small* γ*. Of course, this decreases the capacity. Interestingly, there is no difference respective the choice of* γ *between the deletion and multiplication by* -1 *(*$\varrho \log(N)$ *errors require* $\gamma < 2 - 2\varrho$*), in contrary to the model in Chapter 3, which is more vulnerable to the sign change than to the deletion: the deletion has a less negative impact on the local field and the correction of* $\varrho \log(N)$ *deleted neurons requires* $\gamma < 1 - \varrho$*, whereas the mulitplication by* -1 *requires* $\gamma < 1 - 2\varrho$*. However,* $\varrho \in (0, 1)$ *is in both models allowed for the deletion, in contrary to a restricted choice of* $\varrho \in (0, 1/2)$ *for the multiplication by* -1*.*

The spuriously activated neurons affect directly the bound on α*: the critical value decreases antiproportionally with* $1 + \varrho$ *if* $\varrho \log(N)$ *errors have to be corrected. There is no further condition on* ϱ*.*

Proposition 5.10 *There is at least one instable pattern with probability converging to 1 if there is some*

$$\beta > \frac{\gamma}{2} \log\left(\frac{\gamma}{2}\right) - \frac{\gamma}{2} + 1$$

such that $N^\beta = o(M(N))$*, where* $M(N)$ *denotes the number of stored patterns.*

In contrary, all patterns are stable with high probability, if $M = o(N^{\beta'})$*,* $\beta' < \frac{\gamma}{2} \log\left(\frac{\gamma}{2}\right) - \frac{\gamma}{2} + 1$*.*

Proof: To prove the first part, we can assume that $M(N) \leq N^2/\log(N)^2$ for almost all $N \in \mathbb{N}$, because otherwise an arbitrary stored pattern is instable (there is at least one spuriously activated neuron after the first step of the dynamics) with high probability.

Using Proposition 2.8, we have for $N^\beta = o(M(N))$ and $\varepsilon > 0$ such that $\beta \geq \left(\frac{\gamma}{2} - \varepsilon\right) \log\left(\frac{\gamma}{2} - \varepsilon\right) - \frac{\gamma}{2} + \varepsilon + 1$:

$$\lim_{N\to\infty} \mathbb{P}\left(\exists \mu \leq M(N) : \frac{1}{\log(N)} \sum_{j=1}^{N} |\xi_j^\mu| < \gamma/2 - \varepsilon\right) = 1.$$

Using now $M(N) \leq N^2/\log(N)^2$, we see analogously to Proposition 5.5 that an arbitrary active neuron of a pattern with $k < (\gamma/2 - \varepsilon)\log(N)$ excited neurons is inactive after the first step of the dynamics, with high probability.

If in contrary $M(N) = o(N^{\beta'})$, $\beta' < \frac{\gamma}{2}\log(\frac{\gamma}{2}) - \frac{\gamma}{2} + 1$, we have for $\varepsilon' > 0$ such that $\beta' \leq (\frac{\gamma}{2} + \varepsilon')\log(\frac{\gamma}{2} + \varepsilon') - \frac{\gamma}{2} - \varepsilon' + 1$

$$\lim_{N\to\infty} \mathbb{P}\left(\forall \mu \leq M(N) : \frac{1}{\log(N)} \sum_{j=1}^{N} |\xi_j^\mu| \geq \gamma/2 + \varepsilon'\right) = 1.$$

As in the proof of Proposition 3.4, using that $\mathbb{P}\left(\exists \mu \leq M(N) : \sum_{j=1}^N |\xi_j^\mu| \geq 3\log(N)\right) \to 0$, we observe that the stability of all stored patterns is provided if $M(N) = o(N^{\beta'})$, $\beta' < \frac{\gamma}{2}\log\left(\frac{\gamma}{2}\right) - \frac{\gamma}{2} + 1$.

To improve the performance, it is also possible to consider a version of the sparse BEG model with a fixed number $c \approx \log(N)$ of active neurons per stored pattern, using $p = c/N$ in the definition of the variables K_{ij} and η_i.

Proposition 5.11 *If the stored messages are chosen independently and uniformly from the set of patterns with exactly $c \approx \log(N)$ active neurons, if the number of stored patterns is $M = \alpha N^2/c^2$ and the threshold γc is used, each $\gamma < 2$ is admissible, each $\gamma > 2$ is inadmissible and for fixed γ, each α that fulfills the stability conditions of Theorem 5.4 is also admissible for γ for this second setting of the model.*

In addition, if the two conditions 1.)

$$\frac{1}{2}\alpha\left(-\left(1+\frac{\gamma}{\alpha}\right)\log\left(1+\frac{\gamma}{\alpha}\right)+\frac{\gamma}{\alpha}\right) < -3$$

and 2.)

$$-\text{arsinh}\left(\frac{1}{\alpha}\right)+\alpha\text{cosh}\left(\text{arsinh}\left(\frac{1}{\alpha}\right)\right)-\alpha < -2$$

are fulfilled and additionally either condition 3a.)

$$\frac{1}{2}\alpha\left(-\left(1-\frac{2-\gamma}{\alpha}\right)\log\left(1-\frac{2-\gamma}{\alpha}\right)-\frac{2-\gamma}{\alpha}\right) < -2 \quad \textit{and} \quad \alpha > 2-\gamma$$

or 3b.)

$$\alpha < 2-\gamma$$

hold, we have

$$\lim_{N\to\infty}\mathbb{P}(\exists\mu \leq M(N) : T(\xi^\mu) \neq \xi^\mu) = 0.$$

Proof: For a fixed neuron $i > c$, the exponential moments of

$$\sum_{\mu=2}^{M}\sum_{j\leq c}\xi_i^\mu\xi_j^\mu + \frac{1}{(1-c/N)^2}\eta_i^\mu\eta_j^\mu$$

are as in the previous models similar to those of the model with i.i.d. spins: we consider for fixed $i > c$

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(t\sum_{\mu=2}^{M}\sum_{j=1}^{c}\xi_i^\mu\xi_j^\mu+\frac{1}{(1-c/N)^2}\eta_i^\mu\eta_j^\mu\right)\Big|\mathcal{Z}_c\right]\\
&=\Bigg[\left(1-\frac{c}{N}\right)\sum_{i=0}^{c}\binom{c}{i}\prod_{n=0}^{i-1}\frac{c-n}{N-1-n}\prod_{m=0}^{c-i-1}\left(1-\frac{c-i}{N-1-i-m}\right)e^{-ti\frac{c/N}{1-c/N}+t(c-i)\frac{c^2/N^2}{(1-c/N)^2}}+\\
&\quad\frac{c}{N}\sum_{i=0}^{c-1}\binom{c}{i}\prod_{n=0}^{i-1}\frac{c-1-n}{N-1-n}\prod_{m=0}^{c-1-i}\left(1-\frac{c-1-i}{N-1-i-m}\right)e^{ti-t(c-i)\frac{c/N}{1-c/N}}\left(\frac{1}{2}e^t+\frac{1}{2}e^{-t}\right)^i\Bigg]^{M-1}\\
&=\left[\left(1-\frac{c}{N}\right)\left[1+\mathcal{O}\left(\frac{c^5}{N^3}\right)\right]+\frac{c}{N}\left[1-\frac{c^2}{N}-\frac{c^2}{N}t+\frac{c^2}{N}\frac{1}{2}e^t(e^t+e^{-t})+\mathcal{O}\left(\frac{c^5}{N^2}\right)\right]\right]^{M-1}\\
&=\left[1+\frac{c^3}{N^2}\left(\frac{1}{2}(e^{2t}+1)-1-t\right)+\mathcal{O}\left(\frac{c^6}{N^3}\right)\right]^{M-1}=e^{\alpha c\left[\frac{1}{2}(e^{2t}+1)-1-t\right]}(1+o(1)).
\end{aligned}$$

For fixed $i \leq c$ we obtain analogously

$$\mathbb{E}\left[\exp\left(t\sum_{\mu=2}^{M}\sum_{j\neq i}^{c}\xi_i^\mu\xi_j^\mu + \frac{1}{(1-c/N)^2}\eta_i^\mu\eta_j^\mu\right)\Big|\mathcal{Z}_c\right] \leq e^{\alpha(c-1)\left[\frac{1}{2}(e^{2t}+1)-1-t\right]}(1+o(1)).$$

Combined with the results of Proposition 3.5 and Theorem 5.4 we observe that stability is reached with high probability in this second setting of the model if the stability conditions of Theorem 5.4 are fulfilled and that particularly each $\gamma < 2$ is admissible. In addition, the same arguments as in the proof of Proposition 5.5 show that $\gamma > 2$ is inadmissible.

Concerning the second notion of capacity, the conditions to guarantee the stability of all patterns are obtained analogously to the conditions of Theorem 5.4: using $\mathbb{P}(\exists\mu : T(\xi^\mu) \neq \xi^\mu) \leq M\mathbb{P}(T(\xi^1) \neq \xi^1)$. Condition 1.) is sufficient to keep the stability of the inactive neurons (compare (5.7) and subsequent calculations) and conditions 2.) and 3.) for the stability of the active neurons ((5.16) and subsequent calculations for the second condition and finally (5.15), (5.17) and page 127 for the third condition).

□

6 The Gripon-Berrou Model

Gripon, Berrou et al. proposed in a series of papers, e. g., [16], [17], [22] or [23], several variations of a model relying on a cluster structure. The model uses $N = cl$ neurons organised in c clusters, each cluster containing l neurons. The patterns that are stored in the network consist of exactly one activated neuron per cluster. The authors propose a choice of $c = \log(l)$: c should not be a constant, as we will see in the next section, but also be as small as possible (see [17]). The patterns are therefore extremely sparse, as in the previous models. A further difference to the models considered so far is that the underlying graph is not the complete graph on $\{1, \ldots, N\}$: apart from self-loops, neurons are connected if and only if they belong to different clusters. Both, cluster structure and sparse patterns are properties of the network inspired by brain structure and activities.

Gripon and Berrou motivate the groups of neurons as an alphabet $\mathcal{A}$ of size l. Each cluster then consists of l neurons of which each one is associated to exactly one value of the alphabet. The patterns, also called messages, can be identified with words of length c of this alphabet. By identifying $\mathcal{A}$ with $\mathcal{A}' = \{1, \ldots, l\}$ by some bijective map, a message is associated with an element $\tilde{\xi} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_c)$ of $(\mathcal{A}')^c$. The value $\tilde{\xi}_a$, $1 \leq a \leq c$, indicates the one and only element in cluster a that is activated in (also referred as part of) the message. This representation is, to reach a more convenient notation, replaced by the following one: instead of $\tilde{\xi}$, we use $\xi \in \left(\{0,1\}^l\right)^c$, and $\tilde{\xi}$ is transformed into ξ through the map $\tilde{\xi} \to (e_{\tilde{\xi}_1}, \ldots, e_{\tilde{\xi}_c}) \in \{0,1\}^{cl}$, with the i-th l-dimensional unit vector e_i.

We assume the messages to be chosen independently and uniformly of the set of all valid (which means that exactly one neuron is activated per cluster) patterns in $\{0,1\}^{cl}$. The stored patterns are denoted by $\xi^1, \ldots, \xi^M$; $M = M(N)$ is the number of stored messages. In particular, these patterns are sparse because each one has c active neurons, only, with $c = \log(l) \approx \log(N)$, comparable to the models in chapters 2 - 5.

The model is originally close to the WTA algorithm in the Willshaw model, using binary synaptic efficacies and a sort of Winner takes all algorithm. Gripon and Berrou illustrate the storing process and the dynamics by a graph visualisation (see [17]): using the underlying graph structure of the model, a message ξ^μ is stored by establishing the complete graph $G(\xi^\mu)$ between its activated neurons and the corresponding self-loops. This means that each edge between two neurons being part of the message is activated, that is, J_e is set to 1. In this context, we sometimes speak of existing connections between the neurons or of a fully connected graph even though strictly speaking the edges already exist but are inactive. The set of neurons that could form a valid message together with the (activated) edge set of its complete graph is also referred as **clique**. Once activated, an edge cannot be erased. If it is already active, nothing is changed. This corresponds to the synaptic efficacies in the Willshaw model: they are also set to 1 if the corresponding neurons are at least once activated in the same message and remain at this value afterwards. The model can be defined with or without self-loops: taking them into account, each neuron that is once excited in a message is then connected to itself. In [23] Gripon and Berrou also propose a weighted self-loop.

For a given configuration of the network used as input, a dynamics can be applied. Given this input, each neuron receives a certain number of signals: signals are emitted by activated neurons and can only be transmitted through active edges. In particular, a neuron can only obtain signals from neurons of different clusters or from itself. However, the decisive information is not how many signals a neuron receives in total: the idea of the model is to count only the *number of different clusters* from which a neuron receives signals. Finally there are different options to proceed and to decide which neurons will be (or remain) activated, but the original one which turns out to provide the best capacity is to activate the neuron within a cluster that obtains the most signals from different clusters (this is called SUM-of-MAX rule). It is reasonable to count at most one incoming signal per cluster because the network aims to decide if messages belong to the stored ones or to repair deleted or corrupted messages. Let us e. g., consider a corrupted stored message as input pattern such that in at least one cluster there are at least $c-1$ spuriously activated neurons and that in one cluster the active neuron has been erased and no neuron is active. The fact that a message only consists of one activated neuron per cluster implies that a neuron in the erased cluster that is e. g., connected to c neurons of the cluster with spuriously activated neurons and to none of the other clusters should not win against a neuron that receives one signal from each of the $c-1$ clusters to which it does not belong because this neuron and the neurons that send signals to this neuron could form a correct stored message, whereas the first neuron cannot be part of a valid message formed by, besides the neuron itself, active neurons of the input. The second neuron should thus be more likely to be activated.

Formally, the GB model can be described as follows: on a vertex set $V = \{(a,i) : a \in \{1,\ldots,c\}, i \in \{1,\ldots,l\}\}$ the edge set is given by

$$E = \{\{(a,i),(b,j)\} : (a,i),(b,j) \in \{1,\ldots,c\} \times \{1,\ldots,l\} : a \neq b\} \cup \{\{(a,i)\} : (a,i) \in V\}.$$

Denoting the stored patterns by $\xi^1,\ldots,\xi^M$, $\xi^\mu = (\xi^\mu_{(1,1)},\ldots,\xi^\mu_{(1,l)},\xi^\mu_{(2,1)},\ldots,\xi^\mu_{(c,l)})$, the synaptic weights are for each $e \in E$ defined by

$$J_{(a,i),(b,j)} = \begin{cases} 1 & \exists \mu \in \{1,\ldots,M\} : \xi^\mu_{(a,i)}\xi^\mu_{(b,j)} = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The synaptic efficacy $J_{(a,i)(a,i)}$ is equal to 1 if and only if the neuron (a,i) (the i-th neuron of the a-th cluster) is part of at least one message ξ^μ. The synaptic efficacy $J_{(a,i),(b,j)}$ is 1 if and only if the i-th neuron of cluster a and the j-th neuron of cluster b have been part for at least one time in the same stored message.

The model Gripon and Berrou propose in e. g., [16] and [22] uses a WTA algorithm which is called the SUM-OF-MAX rule. The local field is, in dependence of the intensity of the self-loop influence, defined by

$$S(\beta)_{(a,i)}(\sigma) = \sum_{b=1,b\neq a}^{c} \Theta\left(\sum_{j=1}^{l} J_{(a,i),(b,j)}\sigma_{(b,j)} - 1\right) + \beta J_{(a,i),(a,i)}\sigma_{(a,i)}.$$

Θ again is the Heaviside function, $\Theta(x) = \mathbb{1}_{x\geq 0}$. We concentrate our analysis on the extremal cases $\beta = 1$ and $\beta = 0$, for which the local fields are called

$$\bar{S}_{(a,i)}(\sigma) := S(1)_{(a,i)}(\sigma) = \sum_{b=1,b\neq a}^{c} \Theta\left(\sum_{j=1}^{l} J_{(a,i),(b,j)}\sigma_{(b,j)} - 1\right) + J_{(a,i),(a,i)}\sigma_{(a,i)}$$

and

$$S_{(a,i)}(\sigma) := S(0)_{(a,i)}(\sigma) = \sum_{b=1,b\neq a}^{c} \Theta\left(\sum_{j=1}^{l} J_{(a,i),(b,j)}\sigma_{(b,j)} - 1\right).$$

By analysing both variations in comparison, we will see that the self-loops improve the stability results and additionally observe that the dynamics using $S(\beta)$, $\beta \in (0,1)$, behaves exactly like the one using $\bar{S}$. In dependence on $S(\beta)$, the parallel dynamics is defined by $T(\beta) = (T(\beta)_{(1,1)}, \ldots, T(\beta)_{(c,l)})$,

$$T(\beta)_{(a,i)}(\sigma) = \Theta(S(\beta)_{(a,i)}(\sigma) - h(a)), \text{ where } h(a) = \max\{S(\beta)_{(a,i)}, i = 1, \ldots, l\}.$$

If there is no ambiguity, $T(\beta)$ is also called T, for short.

There are several variations of the model, obtained by modifying the dynamics and sometimes also the synaptic efficacies. These models are also referred to as GB model because they use the cluster structure, but their characteristics are indicated by their specifying names. The model described so far is called the **GB model with binary synaptic efficacies and SUM-of-MAX rule**. Concerning this dynamics and local field, speaking in terms of the graph illustration, a pattern $\sigma' \in \{\sigma \in \{0,1\}^{cl} : \forall 1 \leq a \leq c : \sum_{i=1}^{l} \sigma_{(a,i)} = 1\}$ is considered to be stored if each edge of the complete graph spanned by its activated neurons (the corresponding clique) is contained in the set $\mathcal{M} := \{e \in E : J_e = 1\}$. We will see that in this case the dynamics cannot decide whether the pattern is one of the stored ones or not and will recognise it as stored.

The first variation was also proposed by Gripon and Berrou (see [23]). It uses a threshold dynamics while keeping the definition of the synaptic efficacies and the cluster structure. The dynamics is, for some $0 < \gamma < 1$, defined by

$$T(\beta)_{(a,i)}(\sigma) = \begin{cases} 1 & S(\beta)_{(a,i)} \geq \gamma c \\ 0 & \text{otherwise.} \end{cases}$$

We will call this model the **GB model with binary synaptic efficacies and threshold dynamics**.

A second variation of the model is obtained if the cluster structure is maintained, but the synaptic weights are replaced by

$$J_{(a,i),(b,j)} = \sum_{\mu=1}^{M} \xi^{\mu}_{(a,i)}\xi^{\mu}_{(b,j)}, \quad a \neq b, \quad J_{(a,i),(a,i)} = 0$$

as in the Hopfield model, Amari's model or the Ternary simple model. This is combined with a threshold dynamics:

$$T_{(a,i)}(\sigma) = \begin{cases} 1 & \sum_{b=1,b\neq a}^{c} \sum_{j=1}^{l} J_{(a,i),(b,j)}\sigma_{(b,j)} \geq \gamma(c-1) \\ 0 & \text{otherwise.} \end{cases}$$

This model was proposed and analysed by Löwe, Vermet and Heusel in [20]. This model is called the **GB model with weighted synaptic efficacies**. It turned out that it shows the same behaviour as Amari's model with fixed activity.

The different variations are analysed in the following sections. We will see that each of them corresponds to one of the models of the previous chapters and offers the same capacity variables but also benefits from the cluster structure and has some advantages compared to the corresponding model. In particular, the GB model with binary synaptic weights and SUM-of-MAX rule has some interesting properties. The first section deals with the GB model with weighted synaptic efficacies and the second with the the two binary versions of the model.

Finally it is also possible to adapt the model to the two ternary models while keeping the cluster structure. This will shortly be described in the last section of this chapter.

6.1 The GB Model with Weighted Synaptic Efficacies

A detailed analysis of the associative abilities of the GB-network with SUM-OF-MAX rule first seemed to be difficult. Therefore the following variant of this model has been proposed and analysed in [20].

As indicated in the name, the synaptic efficacy between two neurons used in the original model is replaced in this section by one counting every joint activation in a message of these two neurons. However, the cluster structure is maintained. The synaptic efficacies are given by

$$J_{(a,i),(b,j)} = \sum_{\mu=1}^{M} \xi^{\mu}_{(a,i)} \xi^{\mu}_{(b,j)}, \quad a \neq b; \quad J_{(a,i),(a,i)} = 0$$

and the local field of an input $\sigma \in \{0,1\}^{cl}$ at some point (a,i) is defined as

$$S_{(a,i)}(\sigma) = \sum_{a \neq b} \sum_{j=1}^{l} J_{(a,i),(b,j)} \sigma_{(b,j)}.$$

The dynamics activates a neuron if its local field exceeds the threshold $h = \gamma(c-1)$.

We will see that this model generally behaves like Amari's model. Assuming that the number of stored messages is $M = \alpha l^2$, the variable α must be chosen according to the same constraints as in Amari's model to achieve stability of an arbitrary stored pattern or to allow correction of a certain number of errors. However, the model has some advantages over Amari's model: it is possible to achieve stability *of every pattern* with high probability without loosing the order of the number of stored patterns but only by adapting the constant α. In the first version of Amari's model on the contrary, the size of M cannot be maintained: to ensure that all patterns are stable, M must be adjusted downwards to $M = N^{\beta}$, for some $\beta < 1$ depending on the threshold γ.

Theorem 6.1 *In the GB model with weighted synaptic efficacies and threshold dynamics, suppose that there are $M = \alpha l^2$ stored messages $\xi^1, \dots, \xi^M$. The threshold variable $\gamma > 0$ is chosen such that*

$$\gamma < \gamma^*$$

with the root γ^* *of* $f(\gamma) = \frac{\gamma}{e^{2/\gamma}} - \gamma + 1$, $\gamma^* \approx 1.255$. *If now* $\alpha < \gamma$,

$$-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha < -1 \tag{6.1}$$

and additionally

$$\alpha > \max(0, \gamma - 1),$$

an arbitrary stored message ξ^μ *is stable with probability converging to 1:*

$$\lim_{l\to\infty} \mathbb{P}\left(\forall (a,i) : T_{(a,i)}(\xi^\mu) = \xi^\mu_{(a,i)}\right) = 1.$$

On the other hand, if either 1.) $\alpha \geq \gamma$ *or 2.)* $\alpha < \gamma$ *and*

$$-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha > -1 \tag{6.2}$$

or 3.) in case $\gamma > 1$

$$0 < \alpha < \gamma - 1,$$

we have

$$\lim_{l\to\infty} \mathbb{P}\left(\exists (a,i) : T_{(a,i)}(\xi^\mu) \neq \xi^\mu_{(a,i)}\right) = 1.$$

Each message is stable with high probability,

$$\lim_{l\to\infty} \mathbb{P}\left(\forall \mu, (a,i) : T_{(a,i)}(\xi^\mu) = \xi^\mu_{(a,i)}\right) = 1$$

if we choose $\gamma \leq 1$ *and* $\alpha < \gamma$ *such that*

$$-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha < -3.$$

If we use the threshold γ *and a threshold variable* α *fulfilling the above stability conditions, a randomly chosen* $\tilde{\xi^\mu}$ *which has been built by deleting randomly* ϱc *entries of* ξ^μ *and by replacing them in* $\varrho_1 c$, $\varrho_1 \leq \varrho$ *clusters by another randomly chosen neuron of the corresponding cluster, is mapped with high probability directly to* ξ^μ*:*

$$\lim_{l\to\infty} \mathbb{P}\left(\forall (a,i) : T_{(a,i)}(\tilde{\xi^\mu}) = \xi^\mu_{(a,i)}\right) = 1,$$

if the inequality

$$\varrho < 1 - \gamma + \alpha(1 - \varrho_2)$$

holds, with $\varrho_2 = \varrho - \varrho_1$. *This is in particular guaranteed if* $\varrho < 1 - \gamma + \alpha$.

Proof of Theorem 6.1: Without loss of generality we consider $\mu = 1$ and assume that the first pattern is $\xi^1 = (e_1, \ldots, e_1)$. To begin with the proof, let $\gamma \leq 1$ and $\alpha < \gamma$.

There are two cases to consider. An active neuron should remain activated and an inactive neuron should remain deactivated. Let us start with the first case.

For $i = 1$ and $1 \leq a \leq c$,

$$S_{(a,1)}(\xi^1) = \sum_{b=1,b\neq a}^{c} \sum_{j=1}^{l} J_{(a,1),(b,j)} \xi^1_{(b,j)} = c - 1 + \sum_{b=1,b\neq a}^{c} \sum_{\mu=2}^{M} \xi^\mu_{(a,1)} \xi^\mu_{(b,1)},$$

so if we choose $\gamma \leq 1$, we immediately see that

$$\mathbb{P}(\forall 1 \leq a \leq c : T_{(a,1)}(\xi^\mu) = 1) = 1.$$

We consider $a = 1$ and $i = 2$ representatively for the case $\xi^1_{(a,i)} = 0$. Then

$$\begin{aligned}
\mathbb{P}\left(T_{(1,2)}(\xi^1) \neq \xi^1_{(1,2)}\right) &= \mathbb{P}\left(\sum_{b=2}^{c}\sum_{j=1}^{l} J_{(1,2),(b,j)}\xi^1_{(b,j)} \geq \gamma(c-1)\right) \\
=&\mathbb{P}\left(\sum_{b=2}^{c}\sum_{\mu=2}^{M} \xi^\mu_{(1,2)}\xi^\mu_{(b,1)} \geq \gamma(c-1)\right) \leq e^{-t\gamma(c-1)}\mathbb{E}\left[e^{t\sum_{b=2}^{c}\sum_{\mu=2}^{M}\xi^\mu_{(1,2)}\xi^\mu_{(b,1)}}\right] \\
=&e^{-t\gamma(c-1)}\mathbb{E}\left[e^{t\sum_{b=2}^{c}\xi^2_{(1,2)}\xi^2_{(b,1)}}\right]^{M-1} \qquad (6.3)
\end{aligned}$$

for arbitrary $t > 0$. Here we first applied the definition of the dynamics, the definition of J and the assumptions concerning ξ^1. Finally we used the exponential Chebyshev inequality and the independence of the messages. We compute $\mathbb{E}\left[e^{t\sum_{b=2}^{c}\xi^2_{(1,2)}\xi^2_{(b,1)}}\right]$ and obtain the following bound:

$$\begin{aligned}
\mathbb{E}\left[e^{t\sum_{b=2}^{c}\xi^2_{(1,2)}\xi^2_{(b,1)}}\right] &= \left(1-\frac{1}{l}\right)\cdot 1 + \frac{1}{l}\left(\mathbb{E}[e^{t\sum_{b=2}^{c}\xi^2_{(b,1)}}]\right) \\
=\left(1-\frac{1}{l}\right) + \frac{1}{l}\left(1-\frac{1}{l}+\frac{1}{l}e^t\right)^{c-1} &\leq \left(1-\frac{1}{l}\right) + \frac{1}{l}e^{(c-1)\cdot\frac{e^t-1}{l}},
\end{aligned}$$

due to the estimate $1+x \leq e^x$ for all $x \in \mathbb{R}$. Inserting this into (6.3) yields, by using once more $1+x \leq e^x$:

$$\mathbb{P}\left(T_{(1,2)}(\xi^1) \neq \xi^1_{(1,2)}\right) \leq e^{-t\gamma(c-1)}\left[\left(1-\frac{1}{l}\right)+\frac{1}{l}e^{(c-1)\cdot\frac{e^t-1}{l}}\right]^M \leq e^{-t\gamma(c-1)}e^{\frac{M}{l}\left(e^{(c-1)\cdot\frac{e^t-1}{l}}-1\right)}. \qquad (6.4)$$

For fixed t not depending on l and $l \to \infty$, expanding the exponential and taking into account $M = \alpha l^2$ yields

$$\begin{aligned}
\mathbb{P}\left(T_{(1,2)}(\xi^1) \neq \xi^1_{(1,2)}\right) &\leq \exp\left[-t\gamma(c-1)\right]\exp\left[\frac{M}{l}\left(e^{(c-1)\cdot\frac{e^t-1}{l}}-1\right)\right] \\
=&\exp\left[-t\gamma(c-1)+\frac{M}{l}\left((c-1)\frac{e^t-1}{l}+\mathcal{O}\left(\frac{c^2}{l^2}\right)\right)\right] \\
=&\exp\left[(c-1)\left(-t\gamma+\alpha(e^t-1)\right)+\mathcal{O}\left(\frac{c^2}{l}\right)\right] \\
=&\exp\left[(c-1)\left(-t\gamma+\alpha(e^t-1)\right)\right]\left[1+\mathcal{O}\left(\frac{c^2}{l}\right)\right].
\end{aligned}$$

The last line takes its minimum at $t = \log(\gamma/\alpha)$. The condition $t > 0$ is for this choice fulfilled if $\gamma > \alpha$. Since we have l neurons in each cluster, the probability of having an error is, as long as $\gamma \leq 1$ and $\alpha < \gamma$, at most

$$\begin{aligned}&\mathbb{P}\left(\exists(a,i): T_{(a,i)}(\xi^1)\neq\xi^1_{(a,i)}\right)\\ \leq& cl\cdot\mathbb{P}\left(T_{(1,2)}(\xi^1)\neq\xi^1_{(1,2)}\right)\leq\exp\left[\log(c)+c+(c-1)\left(-t\gamma+\alpha(e^t-1)\right)\right]\left[1+\mathcal{O}\left(\frac{c^2}{l}\right)\right].\end{aligned}$$

Using $t=\log(\gamma/\alpha)$, this vanishes if

$$-\gamma\log\left(\frac{\gamma}{\alpha}\right)+\gamma-\alpha<-1.$$

Let now $\gamma>1$, $\alpha>\max(0,\gamma-1)$ and $\alpha<\gamma$, $-\gamma\log(\gamma/\alpha)+\gamma-\alpha<-1$. We observe again that the local field of an activated neuron, i.e. $i=1$, is

$$S_{(a,1)}(\xi^1)=\sum_{b=1,b\neq a}^{c}\sum_{j=1}^{l}J_{(a,1),(b,j)}\xi^1_{(b,j)}=c-1+\sum_{b=1,b\neq a}^{c}\sum_{\mu=2}^{M}\xi^\mu_{(a,1)}\xi^\mu_{(b,1)}$$

and that the pattern is stable with high probability if

$$\mathbb{P}\left(\exists 1\leq a\leq c:\sum_{b=1,b\neq a}^{c}\sum_{\mu=2}^{M}\xi^\mu_{(a,1)}\xi^\mu_{(b,1)}<(\gamma-1)(c-1)\right)\longrightarrow 0.$$

In analogy to the exponential moment computed in the first part of the proof, the exponential moment of $-\sum_{b=1,b\neq 1}^{c}\sum_{\mu=2}^{M}\xi^\mu_{(1,1)}\xi^\mu_{(b,1)}$ is at most

$$\mathbb{E}\left[e^{-t\sum_{b=2}^{c}\sum_{\mu=2}^{M}\xi^\mu_{(1,1)}\xi^\mu_{(b,1)}}\right]\leq\exp\left[(c-1)\alpha\left(e^{-t}-1\right)\right]\left[1+\mathcal{O}\left(\frac{c^2}{l}\right)\right]$$

and the above probability is bounded by

$$\begin{aligned}&\mathbb{P}\left(\exists 1\leq a\leq c:\sum_{b=1,b\neq a}^{c}\sum_{\mu=2}^{M}\xi^\mu_{(a,1)}\xi^\mu_{(b,1)}<(\gamma-1)(c-1)\right)\\ &\leq c\cdot\exp\left[(\gamma-1)(c-1)t\right]\exp\left[(c-1)\alpha\left(e^{-t}-1\right)\right]\left[1+\mathcal{O}\left(\frac{c^2}{l}\right)\right].\end{aligned}$$

With $t=-\log\left(\frac{\gamma-1}{\alpha}\right)$, positive if $\gamma-1<\alpha$, this converges to 0. The part concerning the inactive neurons can be shown as in the first part of the proof. Thus an arbitrary pattern is stable if the conditions of Theorem 6.1 are fulfilled.

In addition, all the patterns are stable, that is,

$$\lim_{l\to\infty}\mathbb{P}\left(\exists 1\leq a\leq c,1\leq i\leq l,1\leq\mu\leq M:T_{(a,i)}(\xi^\mu)\neq\xi^\mu_{(a,i)}\right)=0,$$

if $\gamma\leq 1$ and

$$Mlce^{(c-1)(-\gamma\log(\gamma/\alpha)+\gamma-\alpha)}\longrightarrow 0.$$

This happens if

$$-\gamma\log\left(\frac{\gamma}{\alpha}\right)+\gamma-\alpha<-3.$$

Concerning the stability results, it remains to show that the derived bounds are sharp. First, assume that $\gamma > 1$ and $0 < \alpha < \gamma - 1$. Take (a,i) with $\xi^1_{(a,i)} = 1$, w.l.o.g. $(1,1)$. $\xi^1_{(1,1)}$ is not stable if

$$\sum_{\mu=2}^{M} \sum_{b=2}^{c} \xi^\mu_{(1,1)} \xi^\mu_{(b,1)} < (\gamma - 1)(c-1).$$

Now the probability is bounded, by applying the exponential Chebyshev inequality and the exponential moment obtained in the first part of the proof

$$\begin{aligned} &\mathbb{P}\left(\exists a \leq c : \sum_{\mu=2}^{M} \sum_{b \neq a}^{c} \xi^\mu_{(a,1)} \xi^\mu_{(b,1)} \geq (\gamma-1)(c-1) \right) \\ \leq & c \exp\left[-(\gamma-1)(c-1)t\right] \exp\left[(c-1)\alpha\left(e^t - 1\right)\right] \left[1 + \mathcal{O}\left(\frac{c^2}{l}\right)\right]. \end{aligned}$$

Taking $t = \log\left(\frac{\gamma-1}{\alpha}\right) > 0$ because $\gamma - 1 > \alpha$, this probability vanishes. Thus

$$\lim_{l\to\infty} \mathbb{P}\left(T(\xi^1) \neq \xi^1\right) = 1$$

for $\gamma - 1 > \alpha$. In particular, if $\alpha < \gamma - 1$, even

$$\lim_{l\to\infty} \mathbb{P}\left(T(\xi^1) = (0,\ldots,0)\right) = 1$$

holds.

Let now $\alpha < \gamma$ and $-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha > -1$. We saw in the first part of the proof that the lower bounds on α of this model coincide with those of Amari's model. The exponential moments of the local field have the same form in these two models. But also the upper bounds coincide. In comparison to the proof concerning Amari's model, the cluster structure complicates the situation. There are two main differences to Amari's model: on the one hand, we will just consider one fixed cluster and prove that there will already be an error with high probability. It is not necessary to consider all inactive neurons. On the other hand, the cluster structure implies that for some fixed cluster a, the $\xi^\mu_{(a,i)}, 1 \leq i \leq l$, are not independent. But these random variables are negatively associated which will be helpful in the proof.

So we fix some cluster, e. g., $a = 1$. We consider the random variables

$$\Theta\left(\sum_{b=2}^{c} \xi^\mu_{(b,1)} - 1\right), \quad \mu \geq 2$$

which are independent and identically Bernoulli distributed with parameter $p_1 = 1 - (1 - 1/l)^{c-1} = \frac{c-1}{l} + \mathcal{O}(c^2/l^2)$. As in Chapter 2, the probability

$$\mathbb{P}\left[\sum_{\mu=2}^{M} \Theta\left(\sum_{b=2}^{c} \xi^\mu_{(b,1)} - 1\right) < (1-\delta)\alpha(c-1)l\right] \longrightarrow 0.$$

The complement of the corresponding event is called

$$B_\delta := \Big\{ \sum_{\mu=2}^{M} \Theta\Big(\sum_{b=2}^{c} \xi^\mu_{(b,1)} - 1 \Big) \geq (1-\delta)\alpha(c-1)l \Big\}.$$

In addition, the variables $\xi^\mu_{(1,i)}, i \geq 2$, are for fixed μ negatively associated. This can easily be checked by computing the covariance for increasing functions f_1, f_2 due to the fact that the variables are binary and exactly one of them takes the value 1, each with equal probability or by observing that they are Multivariate Hypergeometrically distributed with parameters $l-1$ (number of different colours/characteristics), 1 (number of drawings) and $(1, \dots, 1)$ (multiplicity of the characteristics) (see [26]). Using Lemma 2.12 3.), we observe that ξ^μ_i, $\mu \geq 2$, $i \geq 2$, are negatively associated. Given $\xi^\mu_{(b,1)}, \mu \geq 2, b \geq 2$, the sums

$$\sum_{\mu=2}^{M} \Theta\left(\sum_{b=2}^{c} \xi^\mu_{(b,1)} - 1 \right) \cdot \xi^\mu_{(1,i)}, \quad i \geq 2$$

then are conditionally negatively associated because we applied increasing functions on disjoint subsets of the negatively associated variables. We write $\mathcal{F} = \sigma(\xi^\mu_{(b,1)}, \mu \geq 2, b \geq 2)$. This yields, using Lemma 2.12, 4.) in the penultimate step of the subsequent computation

$$\begin{aligned}
&\mathbb{P}\left[\forall i \geq 2 : \sum_{\mu=2}^{M} \sum_{b=2}^{c} \xi^\mu_{(1,i)} \xi^\mu_{(b,1)} < \gamma(c-1) \Big| \mathcal{F} \right] \\
\leq &\mathbb{P}\left[\forall i \geq 2 : \sum_{\mu=2}^{M} \Theta\left(\sum_{b=2}^{c} \xi^\mu_{(b,1)} - 1 \right) \cdot \xi^\mu_{(1,i)} < \gamma(c-1) \Big| \mathcal{F} \right] \\
\leq &\prod_{i=2}^{l} \mathbb{P}\left[\sum_{\mu=2}^{M} \Theta\left(\sum_{b=2}^{c} \xi^\mu_{(b,1)} - 1 \right) \cdot \xi^\mu_{(1,i)} < \gamma(c-1) \Big| \mathcal{F} \right] \\
= &\left(1 - \mathbb{P}\left[\sum_{\mu=2}^{M} \Theta\left(\sum_{b=2}^{c} \xi^\mu_{(b,1)} - 1 \right) \cdot \xi^\mu_{(1,2)} \geq \gamma(c-1) \Big| \mathcal{F} \right] \right)^l .
\end{aligned}$$

In addition, we have on B_δ

$$\begin{aligned}
\min_{B_\delta} \; &\mathbb{P}\left[\sum_{\mu=2}^{M} \Theta\left(\sum_{b=2}^{c} \xi^\mu_{(b,1)} - 1 \right) \cdot \xi^\mu_{(1,2)} \geq \gamma(c-1) \Big| \mathcal{F} \right] \\
\geq &\mathbb{P}\left[\sum_{\mu=2}^{\alpha(c-1)l(1-\delta)+1} \xi^\mu_{(1,2)} \geq \gamma(c-1) \right] .
\end{aligned}$$

This is true because the (conditional) probability of the event that the sum in the first line exceeds the threshold $\gamma(c-1)$ is increasing in the number of non-zero summands $\Theta(\sum_{b=2}^{c} \xi^\mu_{(b,1)} - 1)$, $\mu \geq 2$, so the minimum of the conditional probabilities on this set is attained for $\sum_{\mu=2}^{M} \Theta(\sum_{b=2}^{c} \xi^\mu_{(b,1)} - 1) = \alpha(c-1)l(1-\delta)$, assuming that $\alpha(c-1)l(1-\delta) \in \mathbb{N}$.

Now the proof is continued analogously to the proof of Theorem 2.1. The random variable $\sum_{\mu=2}^{\alpha(c-1)l(1-\delta)+1} \xi^{\mu}_{(1,2)}$ is approximately Poisson distributed with parameter $\alpha(c-1)(1-\delta)$. The total variation distance between the two distributions is at most $2\alpha c/l$. Let Y_λ denote a Poisson distributed random variable with parameter $\lambda > 0$. Combining the above results, the limit

$$\begin{aligned}&\lim_{l\to\infty} \mathbb{P}\left(\exists i \geq 2 : T_{(1,i)}(\xi^1) \neq \xi^1_{(1,i)}\right)\\ &\geq \lim_{l\to\infty} \mathbb{P}(B_\delta)\left(1 - \left[1 - \mathbb{P}\left(Y_{\alpha(1-\delta)(c-1)} \geq \gamma(c-1)\right) + 2\alpha c\frac{1}{l}\right]^l\right)\end{aligned}$$

is equal to 1 if

$$\lim_{l\to\infty} \frac{1}{\log(l)} \log\left(\mathbb{P}\left(Y_{(c-1)(1-\delta)\alpha} \geq \gamma(c-1)\right)\right) > -1.$$

Since $-\gamma \log\left(\frac{\gamma}{\alpha}\right) + \gamma - \alpha > -1$ and $\alpha < \gamma$, a suitable choice of δ yields

$$\lim_{l\to\infty} \frac{1}{\log(l)} \log\left(\mathbb{P}\left(Y_{(c-1)(1-\delta)\alpha} \geq \gamma(c-1)\right)\right) = -\gamma \log\left(\frac{\gamma}{\alpha(1-\delta)}\right) + \gamma - \alpha(1-\delta) > -1$$

and a stored pattern is instable, with high probability.

Let finally $\alpha \geq \gamma$ and $\gamma < \gamma^*$. As shown in Chapter 2, we observe that there is some $\alpha' < \gamma$ such that $-\gamma\log(\gamma/\alpha') + \gamma - \alpha' > -1$, which means that

$$\mathbb{P}\left(\exists i \leq l : \xi^1_{(1,i)} = 0, T_{(1,i)}(\xi^1) \neq 0\right) \longrightarrow 1$$

if $M = \alpha' l^2$ patterns are stored. Since $\alpha > \alpha'$, this also holds if $M = \alpha l^2$, because the probability that the correcponding neuron allocates enough signals to exceed $\gamma(c-1)$ increases in the number of stored patterns.

Concerning the error correction, let the pattern $\tilde{\xi}^1$ corrupted as described in the theorem be used as input for the dynamics. The inactive neurons of ξ^1 are or remain inactive with high probability, because at most c neurons are activated and the stability conditions are fulfilled. The activated neurons of ξ^1 are recovered or remain activated with high probability, because the local field of one of these neurons is

$$S_{(a,1)}(\tilde{\xi}^1) = (1-\varrho)(c-1) + \sum_{\mu=2}^{M} \sum_{\substack{b\neq a: \exists i_b \leq l:\\ \tilde{\xi}^1_{(b,i_b)}=1}} \xi^{\mu}_{(a,1)}\xi^{\mu}_{(b,i_b)},$$

which can exactly be treated as in the previous proof concerning the stability of the activated neurons if $\gamma > 1$. Note that there are $(1-\varrho_2)c$ clusters with exactly one active neuron and $\varrho_2 c$ clusters with no active neuron.

This finishes the proof of Theorem 6.1.

□

Corollary 6.2 *The critical values of the threshold and the capacity variables in this model are* $\gamma^* \approx 1.255$ *and* $\alpha^* \approx 0.255$*, respectively, obtained in Chapter 2.*

Proof: This results from the analysis of the relation of the capacity and threshold variables in the proof of Theorem 2.5.

Remark 6.3 *We saw that the model behaves exactly as Amari's model concerning the bounds on the capacity variables. The main difference to Amari's model is, on the one hand, that the number of active neurons per stored pattern is fixed (in contrast to the first version of Amari's model) and, on the other hand, that the location of the active neurons is restricted to the different clusters. The cluster structure cannot be exploitet to obtain better results. However, it improves the performance of the models in the next section.*

6.2 The Gripon-Berrou Model with Binary Synaptic Efficacies

6.2.1 The SUM-OF-MAX Rule

As already explained in the introduction of this chapter, the neurons of this network are organised in clusters and the updating dynamics activates the neuron(s) in each cluster that possess the highest local field within their cluster. The local field counts the number of clusters from which at least one signal is received. Each activated neuron that is linked to the neuron by an active edge sends a signal; the important feature of the model is that it is not decisive how many signals a neuron receives in total but the number of different clusters in which it is linked by an (active) edge to at least one activated neuron. The winner(s) within a cluster is/are activated and each other neuron is set to 0.

We will see that this model, as long as it counts self-loops, i.e., the local field $\bar{S}$ is used, guarantees the stability of all stored patterns, independently of their number. This is stated and proved in the following proposition:

Proposition 6.4 *In the GB model with WTA dynamics (SUM-OF-MAX rule) and local field $\bar{S}$ or $S(\beta)$, respectively, with an arbitrary $\beta \in (0, 1)$, each stored pattern is a fixed point of the dynamics, independent of the number of stored messages in the network.*

Proof of Proposition 6.4 We take one of the stored messages, e. g., ξ^1. There is exactly one activated neuron per cluster, and these c neurons are fully interconnected (by active edges) because they are all part of ξ^1. Each of them consequently gets one signal per cluster, in total c signals. A neuron that is not activated in ξ^1 does not get a signal from its own cluster because it is inactive and there are no connections within a cluster beside the self-loops. It can thus collect at most $c - 1$ signals. We therefore know that

$$\bar{S}_{(a,i)}(\xi^1) \begin{cases} = c & \xi^1_{(a,i)} = 1 \\ \leq c - 1 & \xi^1_{(a,i)} = 0. \end{cases}$$

This yields immediately $T(\xi^1) = \xi^1$.

Taking $\beta > 0$, we observe the same, because again for each neuron (a, i) with $\xi^1_{(a,i)} = 0$, the local field is at most

$$S(\beta)_{(a,i)}(\xi^1) \leq c - 1$$

and for a neuron (a, i) with $\xi^1_{(a,i)} = 1$, we have

$$S(\beta)_{(a,i)}(\xi^1) = c - 1 + \beta > c - 1.$$

□

Of course, it is not reasonable to store all possible valid messages; the network would then be needless. Besides the stability of the stored patterns, a task the network should perform is the error correction of partially corrupted messages. If too many patterns are stored, the network is intolerant to small errors.

To illustrate this problem, imagine that a stored pattern ξ^μ is defective in the following way: in one fixed cluster, the activated neuron has been replaced by another one. If this spurious neuron has connections to the rest of the neurons of ξ^μ, the message will not be recovered. If $\bar{S}$ is used, the correct neuron will even not be activated; if S is used instead, it will be activated, but the spurious neuron will never be deactivated. It is thus reasonable to restrict the number of stored patterns such that not too many of the edges are contained in $\mathcal{M}$.

In addition, patterns that are not stored should not be recognised as stored ones. The network cannot distinguish stored patterns from patterns whose edge set connecting the activated neurons is completely contained in $\mathcal{M}$. The probability of having activated all the edges corresponding to an arbitrary pattern should converge to 0.

We will analyse the mentioned probability and gather informations about an appropriate number of messages to be stored in the network. To this end, let ξ^0 be an arbitrary valid message. It is recognised as stored if each of the $\binom{c}{2}$ edges of the complete graph $G(\xi^0)$ connecting the c active neurons of the message is contained in $\mathcal{M}$: this event is denoted by $\mathcal{G}(\xi^0)$. Note that we do not consider the self-loops belonging to ξ^0 in $\mathcal{G}(\xi^0)$ because they are automatically contained in $\mathcal{M}$ if the other edges connecting all pairs of active neurons of ξ^0 are; then each neuron has been active in at least one stored message.

Theorem 6.5 *Consider the GB model with binary weights using $N = cl$ neurons organised in $c = \log(l)$ clusters of size l. Assume that the total number of stored patterns is equal to*

$$M = \alpha(\log c)l^2 = \alpha \log(\log l) \cdot l^2.$$

If $\alpha > 2$, a randomly chosen message (independent of the stored patterns) will be recognised as a stored message with probability converging to 1 as $l \to \infty$.

If $\alpha = 2$, with positive probability not tending to 0 a random message will be recognized as a stored message as $l \to \infty$.

If $\alpha < 2$, the probability that a random message will be recognized as stored goes to zero as $l \to \infty$.

In the proof of this theorem, we will use association of random variables (see e. g., [14]).

Definition 6.6 *(see [10] and [14]) A set of real valued random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is associated, if*

$$\mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X})) \geq 0$$

for every coordinatewise increasing functions f and g from $\mathbb{R}^n$ to $\mathbb{R}$ for which the corresponding expectations exist.

Lemma 6.7 *(see [14], P1, P4 and Theorem 1)*

1. *Any subset of associated random variables is again associated.*
2. *Independent random variables are associated.*
3. *Increasing functions of associated random variables are associated.*

For nonnegative integer valued respectively binary associated random variables, the following two Lemmata hold and can be applied in order to bound the probability $\mathbb{P}(\mathcal{G}(\xi^0))$.

Lemma 6.8 *(see [10], Theorem 1) Let $X_1, X_2, \ldots, X_n$ be associated nonnegative integer valued random variables. Then*

$$0 \leq \mathbb{P}[X_i = 0, i = 1, \ldots, n] - \prod_{i=1}^{n} \mathbb{P}[X_i = 0] \leq \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j).$$

Lemma 6.9 *(see [14], Theorem 4.1) Let $X_1, \ldots, X_n$ be associated binary random variables. Then*

$$\mathbb{P}[X_1 = 1, \ldots, X_n = 1] \geq \prod_{i=1}^{n} \mathbb{P}[X_i = 1].$$

Proof of Theorem 6.5 Let ξ^0 be a randomly chosen valid message. Without loss of generality we may assume that $\xi^0_{(a,1)} = 1$, for all $a = 1, \ldots, c$. The probability $\mathbb{P}(\mathcal{G}(\xi^0))$ is given by

$$\begin{aligned}
&\mathbb{P}(\mathcal{G}(\xi^0)) = \mathbb{P}(\forall a, b \in \{1, \ldots, c\}, a \neq b : \exists \mu \in \{1, \ldots, M\} : \xi^\mu_{(a,1)} \xi^\mu_{(b,1)} = 1) \\
=&\mathbb{P}(\forall a, b \in \{1, \ldots, c\}, a < b : \max_\mu \xi^\mu_{(a,1)} \xi^\mu_{(b,1)} = 1).
\end{aligned}$$

The $(\xi^\mu_{(a,1)},\ 1 \leq a \leq c,\ \mu \geq 1)$ are independent random variables. Building their product and taking the maximum of these products are increasing functions. Thus $(\max_\mu \xi^\mu_{(a,1)} \xi^\mu_{(b,1)},$ $a < b)$ are associated. The application of Lemma 6.9 yields

$$\begin{aligned}
\mathbb{P}\left(\forall a, b \in \{1, \ldots, c\}, a < b : \max_\mu \xi^\mu_{(a,1)} \xi^\mu_{(b,1)} = 1\right) &\geq \mathbb{P}\left(\max_\mu \xi^\mu_{(1,1)} \xi^\mu_{(2,1)} = 1\right)^{\binom{c}{2}} \\
&= \left(1 - \left(1 - \frac{1}{l^2}\right)^M\right)^{\binom{c}{2}}.
\end{aligned}$$

The right-hand side can be approximated by inserting $M = \alpha \log cl^2$ and using the series expansion of $\log(x)$ for $0 < x < 1$:

$$\left(1 - (1 - \tfrac{1}{l^2})^M\right)^{\binom{c}{2}} = \left(1 - e^{-\frac{M}{l^2} + \mathcal{O}(\frac{M}{l^4})}\right)^{\binom{c}{2}}$$

$$= \exp\Big(-\tbinom{c}{2}\left[e^{-\alpha\log(c)} + \mathcal{O}\left(e^{-2\alpha\log(c)}\right)\right]\Big) \approx \exp\left(-\tfrac{c^2}{2}e^{-\alpha\log c}\right).$$

This converges to 1 if $\alpha > 2$, and to $e^{-1/2}$ if $\alpha = 2$, which shows the first two statements of the theorem.

For the third statement we can use the upper bound given in Lemma 6.8. Define, for an edge e belonging to ξ^0, $e = \{(a,1),(b,1)\}$ for some $a \neq b$, the random variable $X_e = \max\{\xi^\mu_{(a,1)}\xi^\mu_{(b,1)} : \mu = 1,\dots,M\}$. X_e indicates if the edge e is contained in the set $\mathcal{M}$. Let

$$Z = \sum_{e\in V} X_e \quad \text{with } V = \{\{(a,1),(b,1)\}, a,b \in \{1,\dots,c\}, a \neq b\}.$$

Z denotes the number of edges belonging to $\mathcal{G}(\xi^0)$ that are contained in $\mathcal{M}$. ξ^0 is recognised by the system if and only if $Z = \binom{c}{2}$:

$$\mathbb{P}[\mathcal{G}(\xi^0)] = \mathbb{P}[Z = c(c-1)/2].$$

The random variables $Y_e = 1 - X_e$ are also associated (since for coordinatewise increasing functions f and g, $-f((1,\dots,1)-\mathbf{X})$ and $-g((1,\dots,1)-\mathbf{X})$ are coordinatewise increasing in $\mathbf{X}$) and nonnegative (even binary). Lemma 6.8 yields, denoting by $L = \binom{c}{2}$,

$$\mathbb{P}[\mathcal{G}(\xi^0)] = \mathbb{P}[Z = L] = \mathbb{P}\left[\sum_{e\in V} Y_e = 0\right] \leq \prod_{e\in V}\mathbb{P}[Y_e = 0] + \sum_{e\neq e',\, e,e'\in V}\mathrm{Cov}(Y_e, Y_{e'}).$$

Y_e is Bernoulli distributed with parameter $(1 - 1/l^2)^M$, X_e Bernoulli distributed with parameter $1 - (1 - 1/l^2)^M$. Denoting by d the parameter of X_e, the probability of $\mathcal{G}(\xi^0)$ is bounded by

$$\mathbb{P}[Z = L] \;\leq\; d^L + \sum_{e\neq e',\, e,e'\in V}\mathrm{Cov}(X_e, X_{e'}). \tag{6.5}$$

Concerning the covariances, notice that $\mathrm{Cov}(X_e, X_{e'}) = 0$, if e and e' are disjoint, which means that they do not share a common vertex. So we assume that $e = \{(a,1),(b,1)\}$ and $e' = \{(a,1),(b',1)\}$ and put $\mathcal{M}(a,1) := \{\mu : \xi^\mu_{(a,1)} = 1\}$. The number of elements $|\mathcal{M}(a,1)|$ in $\mathcal{M}(a,1)$ is Binomially distributed with parameters $1/l$ and M. On $\{\mathcal{M}(a,1) = B\}$, the events

$$\{\exists\mu \in \mathcal{M}(a,1) : \xi^\mu_{(b,1)} = 1\} \quad \text{and } \{\exists\nu \in \mathcal{M}(a,1) : \xi^\nu_{(b',1)} = 1\}$$

are independent and have equal probabilities. So

$$\begin{aligned}
\mathbb{E}(X_e X_{e'}) &= \mathbb{P}\Big(\exists\mu,\nu \in \mathcal{M}(a,1) : \xi^\mu_{(b,1)} = 1, \xi^\nu_{(b',1)} = 1\Big)\\
&= \sum_{r=0}^{M}\mathbb{P}\Big(\exists\mu,\nu \in \mathcal{M}(a,1) : \xi^\mu_{(b,1)}\xi^\nu_{(b',1)} = 1\Big|\, |\mathcal{M}(a,1)| = r\Big)\,\mathbb{P}\left(|\mathcal{M}(a,1)| = r\right)\\
&= \sum_{r=0}^{M}\mathbb{P}\Big(\exists\mu \in \mathcal{M}(a,1) : \xi^\mu_{(b,1)} = 1\Big|\, |\mathcal{M}(a,1)| = r\Big)^2\,\mathbb{P}\left(|\mathcal{M}(a,1)| = r\right)
\end{aligned}$$

$$= \sum_{r=0}^{M} \left(1 - \left(1 - \frac{1}{l}\right)^r\right)^2 \binom{M}{r} \left(\frac{1}{l}\right)^r \left(1 - \frac{1}{l}\right)^{M-r}.$$

The last line can be simplified to

$$\begin{aligned} \mathbb{E}(X_e X_{e'}) &= 1 - 2\sum_{r=0}^{M} \binom{M}{r} \left(\frac{1}{l}\right)^r \left(1 - \frac{1}{l}\right)^M + \sum_{r=0}^{M} \binom{M}{r} \left(\frac{1}{l}\right)^r \left(1 - \frac{1}{l}\right)^{M+r} \\ &= 1 - 2\left(1 - \frac{1}{l}\right)^M \left(1 + \frac{1}{l}\right)^M + \left(1 - \frac{1}{l}\right)^M \left(1 + \frac{1}{l}\left(1 - \frac{1}{l}\right)\right)^M \\ &= 1 - 2\left(1 - \frac{1}{l^2}\right)^M + \left(1 - \frac{2}{l^2} + \frac{1}{l^3}\right)^M. \end{aligned}$$

In addition,

$$(\mathbb{E}(X_e))^2 = (\mathbb{P}(X_e = 1))^2 = d^2 = \left(1 - \left(1 - \frac{1}{l^2}\right)^M\right)^2,$$

and together, this yields

$$\begin{aligned} \mathrm{Cov}(X_e, X_{e'}) &= 1 - 2\left(1 - \frac{1}{l^2}\right)^M + \left(1 - \frac{2}{l^2} + \frac{1}{l^3}\right)^M - \left(1 - \left(1 - \frac{1}{l^2}\right)^M\right)^2 \\ &= \left(1 - \frac{2}{l^2} + \frac{1}{l^3}\right)^M - \left(1 - \frac{2}{l^2} + \frac{1}{l^4}\right)^M \\ &= \exp\left(M \log\left(1 - \frac{2}{l^2} + \frac{1}{l^3}\right)\right) - \exp\left(M \log\left(1 - \frac{2}{l^2} + \frac{1}{l^4}\right)\right) \\ &= \exp\left(-\frac{2M}{l^2} + \frac{M}{l^3} + \mathcal{O}\left(\frac{M}{l^4}\right)\right) - \exp\left(-\frac{2M}{l^2} + \mathcal{O}\left(\frac{M}{l^4}\right)\right) \\ &= \exp\left(-\frac{2M}{l^2}\right)\left(\frac{M}{l^3} + \mathcal{O}\left(\frac{M}{l^4}\right)\right), \end{aligned}$$

after using the series expansion of the logarithm and the exponential. We now use

$$M = \alpha l^2 \log\log l$$

and obtain, by using $c = \log l \lessapprox \log N$ and $l = N/c$:

$$\begin{aligned} \sum_{\substack{e \neq e', \\ e,e' \in V}} \mathrm{Cov}(X_e, X_{e'}) &\leq 2\binom{c}{2}(c-2)\mathrm{Cov}(X_{(1,1)}, X_{(1,2)}) \\ &\leq c^3 \exp(-2\alpha \log\log l)\left[\frac{\alpha \log\log l}{l} + \mathcal{O}\left(\frac{\log\log l}{l^2}\right)\right] \\ &\leq c^3 \exp(-2\alpha \log\log l)\left[\alpha(\log\log N)\frac{c}{N} + \mathcal{O}\left(\frac{c^2 \log\log l}{N^2}\right)\right] \\ &\approx (\log N)^4 \frac{1}{N} \exp(-2\alpha \log\log N)\alpha(\log\log N)[1 + \mathcal{O}(c/N)]. \end{aligned}$$

Inserting this into (6.5), we obtain

$$\begin{aligned}\mathbb{P}[\mathcal{G}(\xi^0)] &= \mathbb{P}[Z = c(c-1)/2] \\ &\leq d^L + \sum_{e \neq e',\, e,e' \in V} \mathrm{Cov}(X_e, X_{e'}) \\ &\leq d^L + \frac{1}{N}\alpha(\log\log N)(\log N)^4 \exp(-2\alpha \log\log N)[1 + \mathcal{O}(c/N)] \\ &\leq d^L + \frac{1}{N}\alpha(\log\log N)(\log N)^{(4-2\alpha)}[1 + \mathcal{O}(c/N)].\end{aligned}$$

We remember that

$$d^L = \left(1 - \left(1 - \frac{1}{l^2}\right)^M\right)^{\binom{c}{2}} \approx \exp\left[-\frac{c^2}{2}e^{-\alpha \log(c)}\right] = \exp\left[-\frac{c^{2-\alpha}}{2}\right].$$

Obviously d^L converges to 0 if $\alpha < 2$. The covariance term clearly vanishes for this choice of M. Thus $\mathbb{P}[\mathcal{G}(\xi^0)]$ converges to 0 if $\alpha < 2$.

We are also interested in the size of $\mathbb{P}[\mathcal{G}(\xi^0)]$. For this choice of M, $\mathbb{P}[\mathcal{G}(\xi^0)]$ is exactly of order d^L if $\alpha > 1$ because in this case

$$\exp\left[c^{2-\alpha}/2\right] = o\left(N\right).$$

The probability $\mathbb{P}[\mathcal{G}(\xi^0)]$ converges to 0 and is exactly of order d^L if $\alpha \in (1, 2)$.

For $\alpha \leq 1$, we can state that $\mathbb{P}[\mathcal{G}(\xi^0)]$ converges to 0 but we cannot draw conclusions concerning the exact order of $\mathbb{P}[\mathcal{G}(\xi^0)]$.

□

Remark 6.10 The previous theorem shows that it is not reasonable to choose a constant c. In this case, the first part of the proof can be applied analogously: the edges remain positively associated and we obtain

$$\mathbb{P}[\mathcal{G}(\xi^0)] \geq d^L = \left(1 - \left(1 - \frac{1}{l^2}\right)^M\right)^{c(c-1)/2}.$$

The right hand side converges to a positive constant, if c is constant and if we either use $M = \alpha(\log\log l)l^2$ or $M = \alpha l^2$. A randomly chosen pattern ξ^0 not belonging to $\{\xi^\mu,\ \mu \in \{1, \ldots, M\}\}$, should not be recognised as a stored message with a positive probability. It is thus not reasonable to choose some c not depending on N.

Corollary 6.11 *In the GB model with cluster size l and number of clusters $c = \log(l)$ suppose that*

$$M = \alpha l^2$$

patterns have been stored. Then the probability that a randomly chosen message is recognised as a stored one converges to zero as l tends to infinity for each choice of α.

The estimation obtained by the application of Lemma 6.8 in the proof of Theorem 6.5 does not permit to draw conclusions concerning the exact order of $\mathbb{P}[\mathcal{G}(\xi^0)]$.

Proof of Corollary 6.11: The fact that the random variables Y_e, $e \in V$, are positively associated, does not depend on the number of stored messages M. Thus the inequality

$$\mathbb{P}(Z = L) \le d^L + \sum_{e \neq e',\, e,e' \in V} \mathrm{Cov}(Y_e, Y_{e'}) = d^L + \sum_{e \neq e',\, e,e' \in V} \mathrm{Cov}(X_e, X_{e'})$$

holds again, with $d = 1 - (1 - 1/l^2)^M$.

The computation of the covariances can also be used in this case: if e and e' are not disjoint, we obtain

$$\mathrm{Cov}(X_e, X_{e'}) = \exp\left[-2M/l^2\right](M/l^3 + \mathcal{O}(M/l^4)) = e^{-2\alpha}\left(\frac{\alpha}{l} + \mathcal{O}(l^{-2})\right);$$

if e and e' are disjoint, X_e and $X_{e'}$ are independent. There are on the whole $2\binom{c}{2}(c-2)$ non-zero summands in $\sum_{e,e' \in V} \mathrm{Cov}(X_e, X_{e'})$. The value of d^L is approximately

$$d^L = \left[1 - \left(1 - \frac{1}{l^2}\right)^M\right]^{\binom{c}{2}} \approx \left(1 - e^{-\alpha}\right)^{\binom{c}{2}}.$$

This yields immediately

$$\mathbb{P}(Z = L) \le d^L + c^3 \frac{e^{-2\alpha}\alpha}{l} + \mathcal{O}\left(l^{-2}\right) \approx \exp\left[\binom{c}{2}\log(1 - e^{-\alpha})\right] + c^3 \frac{e^{-2\alpha}\alpha}{l} + \mathcal{O}\left(l^{-2}\right).$$

This probability tends to zero for each $\alpha > 0$; nevertheless the second term dominates the first one and we cannot see if $\mathbb{P}(Z = L)$ is of the same size as d^L.

□

With the exception of Proposition 6.16, we will for the rest of this chapter assume that

$$M = \alpha l^2.$$

Gripon and Berrou suppose in their papers that the variables $X_e, e \in V$, behave for this choice of cluster size and the numbers of neurons per cluster and of stored patterns as independent ones and that the probability $\mathbb{P}[\mathcal{G}(\xi^0)]$ is well approximated by d^L (see [17]). If the edges X_e were independent, Z would be Binomially distributed with parameters $\binom{c}{2}$ and $d = 1 - (1 - 1/l^2)^M \approx 1 - e^{-\alpha}$. On the one hand, we can show a Weak Law of Large Numbers for Z, see Proposition 6.12. However, the last remark showed that we cannot confirm the assumption of Gripon and Berrou so far. As we will see in the Proposition 6.13, the conjecture of Gripon and Berrou is even not true: for $\alpha < -\log(1 - e^{-2}) \approx 0.14$, we can actually show the contrary.

Proposition 6.12 *Consider the GB model with $c = \log(l)$ clusters, l neurons per cluster and a random (valid) message $\xi^0 \in \{0,1\}^{cl}$. To point out that ξ^0 depends on c, we write during this proposition and its proof $\xi^0(c)$. Let $V(c)$ be the set of edges spanned by the neurons of $\xi^0(c)$ and $Z_c = \sum_{e \in V(c)} X_e$ as defined in the proof of Theorem 6.5. The number of stored messages is $M = \alpha l^2$. Then Z_c obeys a Weak Law of Large Numbers, i.e.*

$$\lim_{c \to \infty} \mathbb{P}\left(\left|\frac{Z_c - (1 - e^{-\alpha})\binom{c}{2}}{\binom{c}{2}}\right| > \varepsilon\right) = 0.$$

Proof of Proposition 6.12: As l tends to infinity, $(1-1/l^2)^M = (1-1/l^2)^{\alpha l^2}$ converges to $e^{-\alpha}$. For any fixed $\varepsilon > 0$, let l_1 be chosen such that

$$\left| \left(1 - \frac{1}{l^2}\right)^M - e^{-\alpha} \right| < \varepsilon/2$$

for $l > l_1$. Combined with the Chebyshev inequality, this yields for $l > l_1$

$$\begin{aligned}
&\mathbb{P}\left(\left| \frac{Z_c - (1-e^{-\alpha})\binom{c}{2}}{\binom{c}{2}} \right| > \varepsilon \right) \\
\leq &\mathbb{P}\left(\left| \frac{Z_c}{\binom{c}{2}} - \left(1 - \left(1 - \frac{1}{l^2}\right)^M\right) \right| + \left| e^{-\alpha} - \left(1 - \frac{1}{l^2}\right)^M \right| > \varepsilon \right) \\
\leq &\mathbb{P}\left(\left| \frac{Z_c}{\binom{c}{2}} - \left(1 - \left(1 - \frac{1}{l^2}\right)^M\right) \right| > \varepsilon/2 \right) \leq 4\frac{\mathbb{V}\left(Z_c/\binom{c}{2}\right)}{\varepsilon^2} \\
\leq &\frac{4}{\varepsilon^2 \binom{c}{2}^2} \left[\binom{c}{2} \left(1 - \left(1 - \frac{1}{l^2}\right)^M\right) \left(1 - \frac{1}{l^2}\right)^M + \sum_{e \neq e',\, e, e' \in V} \mathrm{Cov}(X_e, X_{e'}) \right] \\
\leq &\frac{4}{\varepsilon^2 \binom{c}{2}^2} \left[\binom{c}{2} \left(1 - \left(1 - \frac{1}{l^2}\right)^M\right) \left(1 - \frac{1}{l^2}\right)^M + e^{-2\alpha}\alpha\frac{c^3}{l} + \mathcal{O}(1/l^2) \right]
\end{aligned}$$

by the usage of the computations of the covariances in the proof of Corollary 6.11; this clearly vanishes as l tends to infinity.

□

Proposition 6.13 *Consider the GB model with $N = l\log(l)$ neurons organised in $c = \log(l)$ clusters with l neurons per cluster. Suppose that $M = \alpha l^2$. For each $\alpha < -\log(1-e^{-2})$, we have*

$$\lim_{l\to\infty} \frac{d^L}{\mathbb{P}(\mathcal{G}(\xi^0))} = 0,$$

which means

$$\mathbb{P}(\mathcal{G}(\xi^0)) \neq d^L(1 + o(1)).$$

Proof: First of all we recall that $d = \mathbb{P}(X_e X_{e'} = 1) = 1 - (1 - 1/l^2)^M$. The probability $\mathbb{P}(\mathcal{G}(\xi^0))$ is at least

$$\begin{aligned}
&\mathbb{P}(\mathcal{G}(\xi^0)) \geq \mathbb{P}\left(\exists \mu \in \{1, \ldots, M\} : \xi^\mu = \xi^0\right) \\
=&1 - \left(1 - \frac{1}{l^c}\right)^M = \frac{\alpha}{l^{c-2}} \left[1 + \mathcal{O}\left(\frac{1}{l^{c-2}}\right)\right] \\
=&e^{-c(c-2)+\log(\alpha)} \left[1 + \mathcal{O}\left(\frac{1}{l^{c-2}}\right)\right].
\end{aligned}$$

We took into account that $Ml^k \to 0$ for $k < -2$.

On the other hand, one can use the series expansion of the logarithm and $M = \alpha l^2$ to obtain

$$d^L = d^{\binom{c}{2}} = \left[1 - \left(1 - \frac{1}{l^2}\right)^M\right]^{\binom{c}{2}} = \left[1 - e^{-\alpha - (\alpha/2)l^{-2} + \mathcal{O}\left(l^{-4}\right)}\right]^{\binom{c}{2}} \approx e^{\binom{c}{2}\log\left(1-e^{-\alpha}\right)}.$$

This yields

$$\begin{aligned}\lim_{l\to\infty} \frac{d^L}{\mathbb{P}(\mathcal{G}(\xi^0))} &\leq \lim_{l\to\infty} \frac{e^{\binom{c}{2}\log\left(1-e^{-\alpha}\right)}}{e^{-c(c-2)+\log(\alpha)}\left[1+\mathcal{O}\left(l^{-c+2}\right)\right]} \\ &= \lim_{l\to\infty} \exp\left[c^2\left(\frac{1}{2}\log\left(1-e^{-\alpha}\right)+1\right) - c\left(\frac{\log\left(1-e^{-\alpha}\right)}{2}+2\right) - \log(\alpha)\right].\end{aligned}$$

This limit is equal to 0 if

$$\frac{1}{2}\log\left(1-e^{-\alpha}\right) < -1,$$

that is,

$$\alpha < -\log\left(1-e^{-2}\right).$$

The probability $\mathbb{P}(\mathcal{G}(\xi^0))$ is thus not well approximated by d^L if $\alpha < -\log(1-e^{-2}) \approx 0.1454$.

□

For $\alpha > -\log(1-e^{-2})$, we can find a bound that is clearly better than the one used in the proof of Theorem 6.5: The probability of $\mathcal{G}(\xi^0)$ is at most the square root of d^L.

Proposition 6.14 *Let t and κ be such that*

$$0 < t < 2, \quad 0 < \kappa < 1.$$

The probability $\mathbb{P}(\mathcal{G}(\xi^0))$ is, for $M = \alpha l^2$, bounded from above by

$$\begin{aligned}&\mathbb{P}\left(\mathcal{G}\left(\xi^0\right)\right) \\ &\leq \exp\left[-t\kappa\binom{c}{2}\right][1+o(1)] + \exp\left[(1-\kappa)\log(1-e^{-\alpha})\binom{c}{2}\right][1+o(1)].\end{aligned}$$

In particular, the optimal choice of κ and t is such that $-t\kappa = (1-\kappa)\log(1-e^{-\alpha})$. Such a choice is possible for each $\alpha > -\log(1-e^{-2})$; using $\kappa = \frac{1}{2}$, we obtain for each $\alpha > -\log(1-e^{-2})$

$$\mathbb{P}\left(\mathcal{G}\left(\xi^0\right)\right) \leq 2\exp\left[(1-\kappa)\log(1-e^{-\alpha})\binom{c}{2}\right][1+o(1)] \leq 2\sqrt{(1-e^{-\alpha})^{\binom{c}{2}}}(1+o(1)).$$

Proof of Proposition 6.14: Let $\xi^0 = (e_1, \ldots, e_1)$. We denote the number of active neurons of ξ^0 belonging to the μ-th pattern by

$$X_\mu := \Big|\sum_{a=1}^{c}\sum_{i=1}^{l} \xi^\mu_{(a,i)}\xi^0_{(a,i)}\Big| = \Big|\sum_{a=1}^{c}\xi^\mu_{(a,1)}\Big|.$$

The edge set of ξ^0 is again denoted by V and the edges in V are identified by the two clusters they connect: $e_{a,b}$ is the edge that connects $(a,1)$ and $(b,1)$. In addition, we call

$$Y = \left|\left\{e_{a,b} \in V : \exists \mu, X_\mu = 2 : \xi^\mu_{(a,1)}\xi^\mu_{(b,1)} = 1\right\}\right|.$$

For some $1 > \kappa > 0$, the probability $\mathbb{P}(\mathcal{G}(\xi^0))$ is bounded from above by

$$\mathbb{P}(\mathcal{G}(\xi^0)) \le \mathbb{P}\left(\sum_{\mu: X_\mu \ge 3} \binom{X_\mu}{2} \ge \kappa \binom{c}{2}\right) + \mathbb{P}\left(\mathcal{G}(\xi^0) \middle| \sum_{\mu: X_\mu \ge 3} \binom{X_\mu}{2} < \kappa \binom{c}{2}\right). \tag{6.6}$$

We first consider the first summand on the right hand side. The exponential moment of $\sum_{\mu: X_\mu \ge 3} \binom{X_\mu}{2}$ is, for $0 < t < 2$,

$$\begin{aligned}
&\mathbb{E}\left[\exp\, t \sum_{\mu: X_\mu \ge 3} \binom{X_\mu}{2}\right] = \mathbb{E}\left[\exp\, t\, \mathbb{1}_{X_1 \ge 3} \binom{X_1}{2}\right]^M \\
&= \left[\sum_{i=0}^{c} \binom{c}{i} \frac{1}{l^i} (1 - 1/l)^{c-i} \exp\left(t\, \mathbb{1}_{i \ge 3} \binom{i}{2}\right)\right]^M \\
&= \left[1 - \frac{1}{l^3}\binom{c}{3} + \mathcal{O}(l^{-4}c^4) + \sum_{i=3}^{c} \binom{c}{i} \frac{1}{l^i} (1 - 1/l)^{c-i} \exp\left(t \binom{i}{2}\right)\right]^M.
\end{aligned} \tag{6.7}$$

For fixed t and $i \in \{3, \ldots, c\}$, the expression

$$\begin{aligned}
\binom{c}{i} \frac{1}{l^i} (1 - 1/l)^{c-i} \exp\left(t \binom{i}{2}\right) &\le c^i \frac{1}{l^i} \exp\left(t \binom{i}{2}\right) \\
&= \exp\left[i \log(c) - ic + t(i^2/2 - i/2)\right]
\end{aligned}$$

is either maximal in $i = 3$ or $i = c$. Evaluating in these two arguments yields that for $t < 2$, the maximum is attained in $i = 3$. With this, (6.7) is bounded by

$$\begin{aligned}
&\left[1 - \frac{1}{l^3}\binom{c}{3} + \mathcal{O}(l^{-4}c^4) + \sum_{i=3}^{c} \binom{c}{i} \frac{1}{l^i} (1 - 1/l)^{c-i} \exp\left(t \binom{i}{2}\right)\right]^M \\
&\le \left[1 - \frac{1}{l^3}\binom{c}{3} + \mathcal{O}(l^{-4}c^4) + c \cdot c^3 \frac{1}{l^3} e^{3t}\right]^M \le \exp\left[M \frac{1}{l^3} c^3 (ce^{3t} - 1/6) + \mathcal{O}(c^4/l^2)\right].
\end{aligned} \tag{6.8}$$

With this,

$$\mathbb{P}\left(\sum_{\mu: X_\mu \ge 3} \binom{X_\mu}{2} \ge \kappa \binom{c}{2}\right) \le \exp\left[-t\kappa \binom{c}{2}\right] (1 + \mathcal{O}(c^4/l)).$$

For the second part, we observe that

$$\left|\left\{e_{a,b} \in V : \exists \mu, X_\mu > 2 : \xi^\mu_{(a,1)}\xi^\mu_{(b,1)} = 1\right\}\right| \le \sum_{\mu: X_\mu \ge 3} \binom{X_\mu}{2}.$$

Let V' be an arbitrary subset of V with $|V'| = (1-\kappa)\binom{c}{2}$ and

$$Y' := \left|\left\{e_{a,b} \in V' : \exists \mu, X_\mu = 2 : \xi^\mu_{(a,1)}\xi^\mu_{(b,1)} = 1\right\}\right|.$$

Then the second probability on the right hand side of (6.6) is bounded by

$$\mathbb{P}\left(\mathcal{G}(\xi^0)\middle| \sum_{\mu:X_\mu\geq 3}\binom{X_\mu}{2} < \kappa\binom{c}{2}\right) \leq \mathbb{P}\left(Y' = (1-\kappa)\binom{c}{2}\middle|\forall M : X_\mu \leq 2\right).$$

To estimate the probability on the right hand side, we define by $Z'_\mu := \left|\sum_{a<b:\, e_{ab}\in V'} \xi^\mu_{(a,1)}\xi^\mu_{(b,1)}\right|$ the number of edges of V' contained in ξ^μ and denote by

$$\begin{aligned} p' :=& \mathbb{P}(Z'_\mu = 1 \big| X_\mu \leq 2) = \frac{(1-\kappa)\binom{c}{2}1/l^2(1-1/l)^{c-2}}{(1-1/l)^c + c/l(1-1/l)^{c-1} + \binom{c}{2}1/l^2(1-1/l)^{c-2}} \\ =& \frac{(1-\kappa)\binom{c}{2}1/l^2}{1+(c-2)/l + 1/l^2(\binom{c}{2}+1-c)} = (1-\kappa)\binom{c}{2}\frac{1}{l^2}(1+\mathcal{O}(c/l)). \end{aligned}$$

In the next computation, using in the third line the number of sujective maps from a set with i elements to a set with $(1-\kappa)\binom{c}{2}$ elements and the total number of maps from the first set to the second one, we see that

$$\begin{aligned} &\mathbb{P}\left(Y' = (1-\kappa)\binom{c}{2}\middle|\forall\mu \leq M : X_\mu \leq 2\right) \\ =&\sum_{i=0}^{M}\binom{M}{i}(p')^i(1-p')^{M-i}\mathbb{P}\left(Y' = (1-\kappa)\binom{c}{2}\middle|\forall\mu\leq i : X_\mu = 2, \forall\mu > i : X_\mu < 2\right) \\ =&\sum_{i=0}^{M}\binom{M}{i}(p')^i(1-p')^{M-i}\frac{1}{(1-\kappa)^i\binom{c}{2}^i}\sum_{m=0}^{(1-\kappa)\binom{c}{2}}(-1)^m\binom{(1-\kappa)\binom{c}{2}}{m}\left((1-\kappa)\binom{c}{2}-m\right)^i \\ =&\sum_{m=0}^{(1-\kappa)\binom{c}{2}}(-1)^m\binom{(1-\kappa)\binom{c}{2}}{m}\left(p'\frac{(1-\kappa)\binom{c}{2}-m}{(1-\kappa)\binom{c}{2}}+1-p'\right)^M \\ =&\sum_{m=0}^{(1-\kappa)\binom{c}{2}}(-1)^m\binom{(1-\kappa)\binom{c}{2}}{m}\left(1-\frac{1}{l^2}m(1+\mathcal{O}(c/l))\right)^M \\ =&\sum_{m=0}^{(1-\kappa)\binom{c}{2}}(-1)^m\binom{(1-\kappa)\binom{c}{2}}{m}\exp\left[-\alpha m + \mathcal{O}(cm/l)\right] \\ =&\left(1-e^{-\alpha}\right)^{(1-\kappa)\binom{c}{2}}(1+o(1)). \end{aligned}$$

For $\alpha > -\log(1-e^{-2})$, there is some $0 < t < 2$, $0 < \kappa < 1$ such that $t\kappa = -(1-\kappa)\log(1-e^{-\alpha})$. Especially, for $\kappa = \frac{1}{2}$, we obtain with $t \in [-\log(1-e^{-\alpha}), 2)$

$$-(1-\kappa)\log(1-e^{-\alpha}) = -\frac{1}{2}\log(1-e^{-\alpha}) \leq t\frac{1}{2} = t\kappa$$

which implies that

$$\mathbb{P}\left(\mathcal{G}\left(\xi^0\right)\right) \le 2\exp\left[\frac{1}{2}\log(1-e^{-\alpha})\binom{c}{2}\right][1+o(1)] \le 2\sqrt{(1-e^{-\alpha})^{\binom{c}{2}}}\,[1+o(1)]\,.$$

This shows the claim of the proposition.

□

Remark 6.15 *The bound is improved if α increases. Taking $\alpha = -\log(1-e^{-1})$, which is the critical capacity variable for the GB model using binary synaptic weights and the WTA algorithm or the threshold algorithm, both with S instead of $\bar{S}$ (we will see this in Corollary 6.24 and in the next section), we obtain $\kappa = 0.187$ and*

$$\mathbb{P}\left(\mathcal{G}\left(\xi^0\right)\right) \le 2\left[\left(1-e^{-\alpha}\right)^{\binom{c}{2}}\right]^{0.813}(1+o(1)).$$

To make a first approach to the analysis of the error correcting abilities of the network, assume that ϱc bits of a message ξ^μ have been deleted. We will see that it is impossible to reconstruct this message, if there is a clique (not corresponding to ξ^μ) containing the remaining, non-deleted neurons of ξ^μ and whose edge set is contained in $\mathcal{M}$; this will happen with positive, not vanishing probability if M is too large.

Without loss of generality, we consider ξ^1 and assume that $\xi^1_{(a,1)} = 1$ for all clusters $1 \le a \le c$, whereas $\xi^1_{(a,i)} = 0$ if $i \ne 1$. Let us consider a partially erased pattern. Assume that the $\xi^1_{(a,1)}$ are still 1 for the clusters $1 \le a \le (1-\varrho)c$, $0 < \varrho < 1$ and that all other neurons are set to 0. We are interested in the probability that a randomly chosen completion of the partially deleted message, a pattern $\bar{\xi}^1$ as described below, is recognised by the network. We choose for each cluster $(1-\varrho)c+1 \le a \le c$ a neuron (a, i_a), $2 \le i_a \le l$ and set it to 1. Let $\mathcal{G}(\bar{\xi}^1)$ be the event that the message $\bar{\xi}^1$ having 1's in positions $(a,1)$, $1 \le a \le (1-\varrho)c$ and in (a, i_a), $(1-\varrho)c+1 \le a \le c$, is recognized by the system as a stored message.

The following proposition shows that the probability of $\mathcal{G}(\bar{\xi}^1)$ converges to a positive constant, if M is chosen too large. This confirms the bound on M derived in Theorem 6.5: if M exceeds the bound of this theorem, a partially erased message can never be corrected with positive probability not tending to 0 because the edges of an arbitrary ξ^0 containing the non-erased part of ξ^1 are stored in $\mathcal{M}$, with positive probability.

Proposition 6.16 *Let $\bar{\xi}^1$ be a message in the GB model that consists of $(1-\varrho)c$ neurons of ξ^1 and ϱc randomly chosen neurons, one in each erased cluster. Suppose that there are $M = \alpha l^2 \log c$ messages stored. Then $\mathbb{P}(\mathcal{G}(\bar{\xi}^1))$ tends to 0 if and only if $\alpha < 2$.*

If $M = \alpha l^2$ patterns are stored in the network, $\mathbb{P}(\mathcal{G}(\bar{\xi}^1))$ tends to 0 for any choice of the parameter α.

Proof of Proposition 6.16: Without self-loops, there are $\binom{c}{2}$ edges that belong to $\bar{\xi}^1$. Since only ϱc of the neurons are spurious, there are $\binom{(1-\varrho)c}{2}$ edges of $\bar{\xi}^1$ that also belong to the set of edges of ξ^1 and thus are part of $\mathcal{M}$. There remain $r(c,\varrho) := (1-\varrho)\varrho c^2 + \varrho c(\varrho c - 1)/2 = \frac{c^2}{2}(1-(1-\varrho)^2) - \frac{1}{2}c\varrho$ edges. To bound the probability of the event that these

$r(c, \varrho)$ edges exist in $\mathcal{M}$, we observe that positive association is still valid for a subset of positively associated random variables. Thus the probability $\mathbb{P}(\mathcal{G}(\bar{\xi}^1))$ is bounded from below by $(1-(1-\frac{1}{l^2})^M)^{r(c,\varrho)} = d^{\,r(c,\varrho)}$. More precisely, we have, for $M = \alpha l^2 \log(c)$,

$$\mathbb{P}(\mathcal{G}(\bar{\xi}^1)) \geq \left(1 - \left(1 - \frac{1}{l^2}\right)^M\right)^{r(c,\varrho)}$$
$$\approx \exp\left[-r(c,\varrho)e^{-\alpha \log(c)}\right] \approx \exp\left(-\frac{c^{2-\alpha}}{2}(1-(1-\varrho)^2)\right).$$

This implies that $\mathbb{P}(\mathcal{G}(\bar{\xi}^1))$ does not converge to 0 for $\alpha \geq 2$, as stated in the proposition.

The positive association can also be used to bound the probability from above: the same exponential inequality as in the proof of Theorem 6.5 also shows that $\mathbb{P}(\mathcal{G}(\bar{\xi}^1))$ is at most equal to $d^{\,r(c,\varrho)}$ plus a vanishing term. This term differs from that in the proof of Theorem 6.5 only by the number of non-zero covariance terms, which is larger in the proof of Theorem 6.5. Concretely, we have

$$\mathbb{P}(\mathcal{G}(\bar{\xi}^1)) \leq d^{\,r(c,\varrho)} + \frac{1}{N}\alpha(\log N)^{(4-2\alpha)} \log\log N(1 + \mathcal{O}(c/N)).$$

The second term vanishes and $\mathbb{P}(\mathcal{G}(\bar{\xi}^1))$ converges to 0 for $\alpha < 2$. Similarly we can conclude that $\mathbb{P}(\mathcal{G}(\bar{\xi}^1)) \to 0$ for $M = \alpha l^2$, for each choice of α.

□

Remark 6.17 *If $M = \alpha l^2 \log(c)$, we see as in Theorem 6.5 that $\mathbb{P}(\mathcal{G}(\bar{\xi}^1))$ is well approximated by $d^{\frac{c^2}{2}(1-(1-\varrho)^2)}$ and additionally the latter goes to 0, if $\alpha \in]1,2[$. For $M = \alpha l^2 \log(c)$, $\alpha \in]0,1[$ or if $M = \alpha l^2$, for arbitrary α, the additive error term in the upper bound vanishes slower than $d^{\frac{c^2}{2}(1-(1-\varrho)^2)}$.*

For the rest of this chapter, the number of stored patterns is assumed to be $M = \alpha l^2$. Concerning the subclique, the conjecture that the edges behave as independent (and thus $\mathbb{P}(\mathcal{G}(\bar{\xi}^1))$ is well approximated by $d^{\frac{c^2}{2}(1-(1-\varrho)^2)}$) is false for small α:

Proposition 6.18 *Consider the GB-model with $c = \log(l)$ clusters and l neurons per cluster, with $M = \alpha l^2$ messages stored in the network. Let $\bar{\xi}^1$ be a pattern consisting of $(1-\varrho)c$ of the original neurons of ξ^1 and of ϱc arbitrarily chosen neurons that differ from the neurons of ξ^1. $\bar{\xi}^1$ is valid, i.e. has exactly one activated neuron in each cluster.*

The probability $\mathbb{P}[\mathcal{G}(\bar{\xi}^1)]$ that the edges of $\mathcal{G}(\bar{\xi}^1)$ are contained in $\mathcal{M}$ is not well approximated by $d^{\frac{c^2}{2}(1-(1-\varrho)^2)}$ if

$$\alpha < -\log\left(1 - e^{-2/\varrho}\right).$$

Then

$$\frac{d^{\frac{c^2}{2}(1-(1-\varrho)^2)}}{\mathbb{P}(\mathcal{G}(\bar{\xi}^1))} \longrightarrow 0$$

as l tends to infinity.

Proof: Without loss of generality we consider ξ^1 and assume that this message consists exactly of the activated neurons $(a,1)$, $1 \le a \le c$ and that $\bar{\xi}^1_{(a,i)} = 1$ if $a \le (1-\varrho)c$, $i=1$ or $a > (1-\varrho)c$, $i=2$; otherwise $\bar{\xi}^1_{(a,i)} = 0$.

The probability $\mathbb{P}(\mathcal{G}(\bar{\xi}^1))$ is at least

$$\begin{aligned}&\mathbb{P}(\mathcal{G}\left(\bar{\xi}^1\right))\\ \ge&\mathbb{P}\left(\forall a > (1-\varrho)c : \xi^2_{(a,2)} = 1\right)\mathbb{P}\left(\forall a > (1-\varrho)c, b \le (1-\varrho)c : \exists \mu \ge 3 : \xi^\mu_{(a,2)}\xi^\mu_{(b,1)} = 1\right),\end{aligned}$$

this means, the whole set of spurious neurons is activated in the second message, and the rest of the edges, that is, the ones between the spurious neurons and the non-erased neurons of ξ^1, is activated during the storing process of the remaining $M-2$ messages. Let V' be the set of edges belonging to $\bar{\xi}^1$ and X'_e the indicator variable of having stored the edge e within the edges of the last $M-2$ messages:

$$X'_e = \mathbb{1}_{\{\exists \mu \ge 3 : \xi^\mu_{(a,i)}\xi^\mu_{(b,j)} = 1\}},$$

if $e = ((a,i),(b,j))$. The random variables X'_e, $e \in V'$, are positively associated. In addition, each subset of these variables is positively associated, e. g., the variables concerning the edges between the first $(1-\varrho)c$ clusters and the last ϱc clusters. This yields in particular

$$\begin{aligned}&\mathbb{P}\left[\forall a > (1-\varrho)c, b \le (1-\varrho)c : \exists \mu \ge 3 : \xi^\mu_{(a,2)}\xi^\mu_{(b,1)} = 1\right]\\ \ge&\mathbb{P}\left[\exists \mu \ge 3 : \xi^\mu_{(c,2)}\xi^\mu_{(1,1)} = 1\right]^{c^2\varrho(1-\varrho)}\end{aligned}$$

and the last probability is equal to

$$\mathbb{P}\left[\exists \mu \ge 3 : \xi^\mu_{(c,2)}\xi^\mu_{(1,1)} = 1\right]^{c^2\varrho(1-\varrho)} = \left[1 - \left(1 - \frac{1}{l^2}\right)^{M-2}\right]^{c^2\varrho(1-\varrho)}.$$

Putting these things together, we obtain

$$\begin{aligned}&\mathbb{P}(\mathcal{G}(\bar{\xi}^1))\\ \ge&\mathbb{P}(\forall a > (1-\varrho)c : \xi^2_{(a,2)} = 1)\mathbb{P}(\forall a > (1-\varrho)c, b \le (1-\varrho)c : \exists \mu \ge 3 : \xi^\mu_{(a,2)}\xi^\mu_{(b,1)} = 1)\\ \ge&\exp\left[-c^2\varrho\right]\left[1 - \left(1 - \frac{1}{l^2}\right)^{M-2}\right]^{c^2\varrho(1-\varrho)}\\ \approx&\exp\left[c^2\left(-\varrho + \varrho(1-\varrho)\log(1-e^{-\alpha})\right)\right].\end{aligned}$$

$\mathbb{P}(\mathcal{G}(\bar{\xi}^1))$ is definitely not well approximated by $d^{(1-(1-\varrho)^2)c^2/2}$ if the following limit

$$\begin{aligned}&\lim_{c\to\infty}\frac{d^{(1-(1-\varrho)^2)c^2/2}}{\mathbb{P}(\mathcal{G}(\bar{\xi}^1))} = \lim_{c\to\infty}\frac{d^{c^2(\varrho-\varrho^2/2)}}{\mathbb{P}(\mathcal{G}(\bar{\xi}^1))}\\ \le&\lim_{c\to\infty}\exp\left[c^2\log(1-e^{-\alpha})(\varrho-\varrho^2/2)\right]\exp\left[-c^2\left(-\varrho+\varrho(1-\varrho)\log(1-e^{-\alpha})\right)\right]\end{aligned}$$

$$= \lim_{c\to\infty} \exp\left[c^2 \log(1-e^{-\alpha})\varrho^2/2 + c^2\varrho\right]$$

is 0. This happens if

$$\log\left(1-e^{-\alpha}\right)\frac{\varrho}{2} < -1,$$

which is for $\rho, \alpha > 0$ equivalent to

$$\alpha < -\log\left(1-e^{-2/\varrho}\right).$$

□

Similarly to the case where we considered a complete pattern ξ^0, we can also give an upper bound on the probability of having stored the edges of a subclique.

Proposition 6.19 *Suppose that $\bar{\xi}^1$ consists, as in the previous proposition, of $(1-\varrho)c$ neurons of ξ^1 and ϱc neurons differing from those of ξ^1, exactly one in each remaining cluster and that $M = \alpha l^2$. The probability of the event that the remaining $r(c,\varrho) = c^2\varrho(1-\varrho/2) - c\varrho/2$ edges of the complete graph connecting the neurons of $\bar{\xi}^1$ are contained in $\mathcal{M}$ is bounded from above by*

$$\mathbb{P}\left(\mathcal{G}(\bar{\xi}^1)\right) \le \exp\left[-t\kappa r(c,\varrho)\right]\left[1+o(1)\right] + \exp\left[(1-\kappa)\log(1-e^{-\alpha})r(c,\varrho)\right]\left[1+o(1)\right].$$

The variables t and κ must fulfill

$$0 < t < \frac{1}{\varrho(1-\varrho/2)}, \quad 0 < \kappa < 1.$$

Proof: The proof of Proposition 6.14 can almost literally be repeated; the only two differences are that on the one hand, the number $\binom{c}{2}$ is replaced by $r(c,\varrho)$ and on the other hand, the bound concerning t changes. This is due to the bounds in (6.8) and in (6.7): the clique consists of the ϱc neurons in a deleted cluster and of the $(1-\varrho)c$ non-deleted neurons belonging to the message ξ^1. We are looking for a bound of

$$\sum_{i=3}^{c}\sum_{k=0}^{i}\binom{\varrho c}{k}\binom{(1-\varrho)c}{i-k}\frac{1}{l^i}(1-1/l)^{c-i}\exp\left(t\,k(i-k) + t\binom{k}{2}\right),$$

as in (6.7); here k denotes the number of active neurons belonging to the spurious neurons of $\bar{\xi}^1$ and i the total number of active neurons of $\bar{\xi}^1$ in a fixed stored pattern ξ^μ, $\mu \ge 2$. The difference to the term in (6.7) is that the number of edges that are activated if $i = k+i-k$ neurons of the clique are activated in one message, k in the ϱc deleted clusters and $i-k$ in the $(1-\varrho)c$ remaining clusters, is only equal to $k(i-k)+\binom{k}{2}$, because the edges among the neurons in the non-deleted clusters are already part of $\mathcal{M}$. For fixed i, we have

$$\frac{1}{l^i}\binom{\varrho c}{k}\binom{(1-\varrho)c}{i-k}\exp\left(t\,k(i-k)+t\binom{k}{2}\right) \le \frac{1}{l^i}\binom{c}{i}\exp\left(t\,k(i-k)+t\frac{k^2}{2}\right).$$

We fix $i \in \{3,\dots,c\}$. The expression $k(i-k)+k^2/2$ is, for $i \le \varrho c$, maximal in $k = i$. Note that $i \le c$, but the number of deleted clusters (and the maximal value of k) is just equal to ϱc. For $\varrho c \leq i \leq c$, the expression is maximal in $k = \varrho c$. So, for $i \leq \varrho c$, $0 \leq k \leq i$, we have

$$\frac{1}{l^i}\binom{c}{i}\exp\left(t\,k(i-k)+t\frac{k^2}{2}\right) \leq \exp\left[i\left(\log(c)-c\right)+t\frac{i^2}{2}\right]$$

and the right hand side is maximal in $i = \varrho c$ or $i = 3$. By using $t < \frac{2}{\varrho}$, the maximal argument (on $\{3, \ldots, \varrho c\}$) of the right hand side of the previous estimation is $i = 3$, if l is large enough.

For $i \geq \varrho c$, $0 \leq k \leq \varrho c$, we estimate

$$\frac{1}{l^i}\binom{c}{i}\exp\left(t\,k(i-k)+t\frac{k^2}{2}\right) \leq \exp\left[i\left(\log(c)-c\right)+t\varrho c(i-\varrho c)+t\frac{\varrho^2 c^2}{2}\right],$$

using $k = \varrho c$, as concluded above. The argument of the exponential function is linear in i, and the maximum is either attained in $i = \varrho c$ or $i = c$. Since we already examined $i = \varrho c$, we insert $i = c$ and see that for each $t < \frac{1}{\varrho(1-\varrho/2)}$ and l large enough

$$\max_{\substack{3\leq i\leq c,\\ 0\leq k\leq \varrho c}} \frac{1}{l^i}\binom{\varrho c}{k}\binom{(1-\varrho)c}{i-k}\exp\left(t\,k(i-k)+t\binom{k}{2}\right) \leq \frac{1}{l^3}c^3\exp(5t)\,.$$

The inequality

$$2 > 1/(1-\varrho/2)$$

for $\varrho < 1$ implies that the second condition on t is the stronger one.

□

The probability of having stored a (sub)clique in the network is interesting concerning the question if a corrupted pattern can be corrected after a certain number of steps. If the subclique (by which we mean the union of active edges belonging to the set of spurious neurons and the connections between the correct neurons in the non-deleted clusters and the spurious neurons) exists, there is no way to recover the correct message by the dynamics. This is considered in the next part of this section.

A message is retrieved in multiple steps if and only if there is no other clique than the one of the message containing the non-erased neurons. To see this, note that the correct neurons are activated in the first step, because the largest possible value of the local field is the number of neurons that can fire signals, and the correct neurons obtain a signal by each one of these. Due to the self-loops, they collect from the second step a number of c signals, which cannot be reached by a neuron not activated in the first updated version of the pattern. The neurons that are newly activated in the first step are exactly the ones that are connected to every one of the initially activated neurons. On the other hand, if there is a (valid) clique that contains the non-erased neurons and at least one neuron in a deleted cluster that does not belong to ξ^μ, all neurons of this clique are activated in the first step and remain activated, because they also obtain $(1-\varrho)c$ signals in the first step and c signals from the second step, each.

Proposition 6.20 *For an arbitrary stored pattern ξ^μ and a randomly deleted version $\tilde{\xi}^\mu$ of ξ^μ, where ϱc, $\varrho > 0$, neurons have been deactivated and if $\alpha < -\log(1-e^{-1/[(1-\varrho/2)(1-\varrho)]})$, then the pattern is with high probability recovered after finitely many steps.*

Proof: As already mentioned, a necessary condition of the event that the pattern is not recovered after finitely many steps is that a clique containing the non-deleted $(1-\varrho)c$ neurons exists. We assume that $\mu = 1$, that $\xi^1 = (e_1, \ldots, e_1)$ and that the deleted clusters are the clusters $(1-\varrho)c+1, \ldots, c$. To estimate the corresponding probability, we state

$$\begin{aligned}&\mathbb{P}\left(\exists \xi^0 \neq \xi^1 : \xi^0 \text{ valid}, \, \xi^0_{(1,1)} = \ldots = \xi^0_{((1-\varrho)c,1)} = 1, \, G(\xi^0) \in \mathcal{M}\right) \\ \leq &\sum_{i=1}^{\varrho c} \mathbb{P}\Big(\exists \xi^0 : \xi^0 \text{ valid}, \sum_{a=1}^{(1-\varrho)c} \xi^0_{(a,1)} = (1-\varrho)c, \sum_{a=(1-\varrho)c+1}^{c} \xi^0_{(a,1)} = \varrho c - i, \, G(\xi^0) \in \mathcal{M}\Big).\end{aligned}$$

Depending on i, there are $r'(c,i) := (c-i)i + \binom{i}{2}$ edges of $G(\xi^0)$ not contained in $G(\xi^1)$, if ξ^0 differs from ξ^1 in exactly i clusters. Let $\xi^0(i)$ be the pattern only differing from ξ^1 in the last i clusters, where the second neuron is active instead of the first one. Then

$$\begin{aligned}&\sum_{i=1}^{\varrho c} \mathbb{P}\Big(\exists \xi^0 : \xi^0 \text{ valid}, \sum_{a=1}^{(1-\varrho)c} \xi^0_{(a,1)} = (1-\varrho)c, \sum_{a=(1-\varrho)c+1}^{c} \xi^0_{(a,1)} = \varrho c - i, \, G(\xi^0) \in \mathcal{M}\Big) \\ \leq &\sum_{i=1}^{\varrho c} \binom{\varrho c}{i} l^i \, \mathbb{P}\left(G(\xi^0(i)) \in \mathcal{M}\right) \leq \varrho c \max_{1 \leq i \leq \varrho c} \binom{\varrho c}{i} l^i \, \mathbb{P}\left(G(\xi^0(i)) \in \mathcal{M}\right).\end{aligned}$$

For fixed i and $\xi^0(i)$, we can apply Proposition 6.19:

$$\mathbb{P}\left(G(\xi^0(i)) \in \mathcal{M}\right) \leq 2\left[\exp\left[\max\left\{-t\kappa r'(c,i), \, (1-\kappa)\log(1-e^{-\alpha})r'(c,i)\right\}\right]\right][1+o(1)],$$

with $\kappa \in (0,1)$ and $t < \frac{1}{\varrho(1-\varrho/2)}$. For $1 \leq i \leq \varrho c$, the two expressions

$$\binom{\varrho c}{i} l^i \exp\left[-t\kappa(ci - i^2/2 - i/2)\right], \quad \binom{\varrho c}{i} l^i \exp\left[(1-\kappa)\log(1-e^{-\alpha})(ci - i^2/2 - i/2)\right] \tag{6.9}$$

are both maximal either in $i = 1$ or $i = \varrho c$. For $\alpha < -\log(1 - e^{-1/[(1-\varrho/2)(1-\varrho)]})$, we know that

$$\log(1-e^{-\alpha}) < -\frac{1}{(1-\varrho)(1-\varrho/2)}$$

and we can find a $\kappa > \varrho$ such that still

$$(1-\kappa)\log(1-e^{-\alpha}) < -\frac{1}{(1-\varrho/2)}.$$

Furthermore, it is possible to choose $t < \frac{1}{\varrho(1-\varrho/2)}$ such that still

$$t\kappa > \frac{1}{1-\varrho/2}.$$

With this choice, we examine the expressions in (6.9), first for $i = 1$: the left hand side is

$$\varrho c \binom{\varrho c}{1} l^1 \exp\left[-t\kappa(c-1)\right] \leq \exp\left[2\log(\varrho c) - t\kappa(c-1) + c\right] \longrightarrow 0,$$

because $t\kappa > 1$, and the right hand side

$$\begin{aligned} &\varrho c\binom{\varrho c}{1} l^1 \exp\left[(1-\kappa)\log(1-e^{-\alpha})(c-1)\right] \\ \leq &\exp\left[2\log(\varrho c) + (1-\kappa)\log(1-e^{-\alpha})(c-1) + c\right] \longrightarrow 0, \end{aligned}$$

because $-(1-\kappa)\log(1-e^{-\alpha}) > 1$. For $i = \varrho c$, we obtain in (6.9) on the left hand side

$$\begin{aligned} &\varrho c\binom{\varrho c}{\varrho c} l^{\varrho c} \exp\left[-t\kappa(\varrho c^2 - \varrho^2 c^2/2 - \varrho c/2)\right] \\ = &\exp\left[\log(\varrho c) + \varrho c^2 - t\kappa(\varrho c^2 - \varrho^2 c^2/2 - \varrho c/2)\right] \longrightarrow 0, \end{aligned}$$

because

$$t\kappa > \frac{1}{1-\varrho/2}.$$

Finally, the right hand side of (6.9) is, for $i = \varrho c$, at most

$$\begin{aligned} &\varrho c\binom{\varrho c}{\varrho c} l^{\varrho c} \exp\left[(1-\kappa)\log(1-e^{-\alpha})(\varrho c^2 - \varrho^2 c^2/2 - \varrho c/2)\right] \\ = &\exp\left[\log(\varrho c) + \varrho c^2 + (1-\kappa)\log(1-e^{-\alpha})(\varrho c^2 - \varrho^2 c^2/2 - \varrho c/2)\right] \longrightarrow 0, \end{aligned}$$

because

$$(1-\kappa)\log(1-e^{-\alpha}) < -\frac{1}{(1-\varrho/2)}.$$

This proves the proposition.

□

The proof strategy is also used when proving that the GB model can correct a high number of spuriously activated neurons, see Proposition 6.27. Interestingly, the result of Proposition 6.20 is improved when considering the one step retrieval:

Theorem 6.21 *Consider the GB model using the SUM-of-MAX rule and synaptic efficacy $\bar{S}$. Fix $\mu \in \{1, \ldots, M\}$ and take a randomly deleted pattern $\tilde{\xi}^\mu$, such that f of the activated neurons of ξ^μ have been deactivated. Suppose that $f = \varrho c$. Then the probability of correcting $\tilde{\xi}^\mu$ in one step tends to one,*

$$\lim_{l\to\infty} \mathbb{P}\left(T(\tilde{\xi}^\mu) = \xi^\mu\right) = 1,$$

if

$$\alpha < -\log(1 - e^{-1/(1-\varrho)}).$$

This bound is sharp: $\tilde{\xi}^\mu$ is not corrected in one step,

$$\lim_{l\to\infty} \mathbb{P}\left(T(\tilde{\xi}^\mu) \neq \xi^\mu\right) = 1$$

for each fixed but arbitrary μ if

$$\alpha > -\log(1 - e^{-1/(1-\varrho)}).$$

The proof needs the following Lemma: we determine, for a fixed neuron, the asymptotic distribution of the number of active edges to the set of neurons belonging to a (fixed) clique:

Lemma 6.22 *Let ξ^μ be a stored message and (a,i) be a neuron that is inactive in ξ^μ. We define Y by*

$$Y = \sum_{b \neq a}^{c} \sum_{j=1}^{l} J_{(a,i),(b,j)} \xi^\mu_{(b,j)}.$$

Y is the number of neurons in clusters $b \neq a$ belonging to the message ξ^μ which are connected by an active edge to (a,i).

The distribution of Y is asymptotically Binomial with parameters $c-1$ and $1-e^{-\alpha}$:

$$\mathbb{P}(Y = m) = \begin{cases} \binom{c-1}{m} \left(1 - e^{-\alpha}\right)^m e^{-\alpha(c-1-m)}(1 + o(1)) & \text{for } m \in \{0, \dots, c-1\}, \\ 0 & \text{otherwise} \end{cases}$$

as $l \to \infty$.

In particular, the probability of having (a,i) completely connected with the neurons of an arbitrary but fixed message is, for $\xi^\mu = (e_{j_1}, \dots, e_{j_c})$,

$$\mathbb{P}(\forall b \in \{1, \dots, c\} \setminus \{a\} : J_{(a,i),(b,j_b)} = 1) = \left(1 - e^{-\alpha}\right)^{c-1} (1 + o(1))$$

as $l \to \infty$.

More generally, if we take a set of neurons $\{(b_1, j_1), \dots, (b_r, j_r)\}$ with the restriction that b_k, $1 \leq k \leq r$, are pairwise distinct, for a neuron (a,i) not situated in one of the clusters $b_1, \dots, b_r$, the variable $\tilde{Y}$, defined by

$$\tilde{Y} = \sum_{k=1}^{r} J_{(a,i),(b_k,j_k)},$$

is asymptotically Binomially distributed with parameters r and $1-e^{-\alpha}$:

$$\mathbb{P}\left(\tilde{Y} = m\right) = \binom{r}{m} \left(1 - e^{-\alpha}\right)^m e^{-\alpha(r-m)}(1 + o(1)), \quad m \in \{0, \dots, r\}.$$

Proof of Lemma 6.22: We choose w.l.o.g. $\mu = 1$, $\xi^1 = (e_1, \dots, e_1)$ and assume that $a = 1$ and $i = 2$. The random variable Y indicates the number of neurons belonging to the message ξ^1 which are connected by active edges to $(1,2)$. To determine the distribution of Y, we split the event $\{Y = m\}$ into the disjoint events

$$\left\{Y = m, \sum_{\mu=2}^{M} \xi^\mu_{(1,2)} = n\right\}, \quad n \in \{1, \dots, M-1\}$$

where n indicates the number of stored messages containing neuron $(1,2)$. The conditional probability of $\{Y = m\}$, given the event $\{\sum_{\mu=2}^{M} \xi^\mu_{(1,2)} = n\}$ is equal to

$$\mathbb{P}\left(Y = m \Big| \sum_{\mu=2}^{M} \xi^\mu_{(1,2)} = n\right) = \binom{c-1}{m} \left(1 - \left(1 - \frac{1}{l}\right)^n\right)^m \left(1 - \frac{1}{l}\right)^{n(c-1-m)}.$$

In addition, $\sum_{\mu=2}^{M} \xi^{\mu}_{(1,2)}$ is Binomially distributed with parameters $M-1$ and $\frac{1}{l}$. This yields

$$\begin{aligned}
\mathbb{P}(Y=m) &= \sum_{n=0}^{M-1} \mathbb{P}\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(1,2)} = n, Y=m\right) \\
&= \sum_{n=0}^{M-1} \mathbb{P}\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(1,2)} = n\right) \mathbb{P}\left(Y=m \middle| \sum_{\mu=2}^{M} \xi^{\mu}_{(1,2)} = n\right) \\
&= \sum_{n=0}^{M-1} \binom{M-1}{n} \frac{1}{l^n}\left(1-\frac{1}{l}\right)^{M-1-n} \binom{c-1}{m} \left[1-\left(1-\frac{1}{l}\right)^n\right]^m \left(1-\frac{1}{l}\right)^{n(c-1-m)}
\end{aligned}$$

and, using the Binomial formula, this is

$$\begin{aligned}
&= \binom{c-1}{m} \sum_{n=0}^{M-1} \binom{M-1}{n} \frac{1}{l^n}\left(1-\frac{1}{l}\right)^{M-1-n+n(c-1-m)} \sum_{k=0}^{m} \binom{m}{k} (-1)^k \left(1-\frac{1}{l}\right)^{nk} \\
&= \binom{c-1}{m} \sum_{k=0}^{m} \binom{m}{k} (-1)^k \sum_{n=0}^{M-1} \binom{M-1}{n} \frac{1}{l^n}\left(1-\frac{1}{l}\right)^{M-1-n+nk+n(c-1-m)} \\
&= \binom{c-1}{m} \sum_{k=0}^{m} \binom{m}{k} (-1)^k \left[\frac{1}{l}\left(1-\frac{1}{l}\right)^{k+c-1-m} + 1 - \frac{1}{l}\right]^{M-1} \\
&= \binom{c-1}{m} \sum_{k=0}^{m} \binom{m}{k} (-1)^k \left[\frac{1}{l} - \frac{k+c-1-m}{l^2} + \mathcal{O}((k+c-m)^2 l^{-3}) + 1 - \frac{1}{l}\right]^{M-1} \\
&= \binom{c-1}{m} \sum_{k=0}^{m} \binom{m}{k} (-1)^k \left[1 - \frac{k+c-1-m}{l^2} + \mathcal{O}((k+c-m)^2 l^{-3})\right]^{M-1}.
\end{aligned}$$

As in the proof of Lemma 4.8, we obtain, using $M = \alpha l^2$, with the help of the series expansions of the logarithm and of the exponential that

$$\begin{aligned}
&\binom{c-1}{m} \sum_{k=0}^{m} \binom{m}{k} (-1)^k \left[1 - \frac{k+c-1-m}{l^2} + \mathcal{O}((k+c-m)^2 l^{-3})\right]^{M-1} \\
&= \binom{c-1}{m} \sum_{k=0}^{m} \binom{m}{k} (-1)^k \exp\left[-\alpha(k+c-1-m)\right] \left(1 + \mathcal{O}\left(\frac{(k+c-m)^2}{l}\right)\right) \\
&= \binom{c-1}{m} \left(1-e^{-\alpha}\right)^m e^{-\alpha(c-1-m)} (1+o(1)),
\end{aligned}$$

in particular

$$\mathbb{P}(Y=c-1) = \left(1-e^{-\alpha}\right)^{c-1} (1+o(1)).$$

The same computation can be made for a subset of neurons belonging to a stored message (this will be needed if we consider a partially erased message) or, more generally, for a random set of neurons with the restriction that they belong to pairwise different clusters. The asymptotic distribution is a Binomial one with parameters $\tilde{p} = 1 - e^{-\alpha}$ and r, if r

is the number of neurons to which the fixed neuron should be connected. The proof can almost literally be repeated, in the latter case with the only difference that the number of messages to which neuron (a, i) can belong is equal to M instead of $M - 1$, but this does not change the result.

□

Proof of Theorem 6.21: We take a stored message, e. g., ξ^1, and assume that $f = \varrho c$ entries have been deleted at random. We assume w.l.o.g. that $\xi^1 = (e_1, \ldots, e_1)$. The message $\tilde{\xi^1}$ is not mapped to ξ^1 in one step of the dynamics if at least one neuron of one of the clusters which does not belong to the message ξ^1 is activated by the dynamics. Note that this can never happen in the clusters not concerned by the deletion of neurons (w.l.o.g. these are the clusters $1, \ldots, c - f$), because we respect self-loops. We note that the neurons $(1, 1), \ldots, (c - f, 1)$ are activated and each synaptic efficacy corresponding to the edges among them is positive, so they get $c - f$ signals whereas each other neuron of the cluster can only get up to $c - f - 1$ signals because it is not activated. This is, formally, for $1 \le a \le c$:

$$\bar{S}_{(a,1)}(\tilde{\xi^1}) = J_{(a,1),(a,1)}\tilde{\xi}^1_{(a,1)} + \sum_{\substack{b=1\\b\neq a}}^{c}\sum_{j=1}^{l} J_{(a,1),(b,j)}\tilde{\xi^1}_{(b,j)} = J_{(a,1),(a,1)}\tilde{\xi}^1_{(a,1)} + \sum_{\substack{b=1\\b\neq a}}^{c-f} J_{(a,1),(b,1)} = c - f$$

since $J_{(a,1),(b,1)} = 1$ for $1 \le a, b \le c$ as result of the fact that the neurons $(1, 1), \ldots, (c, 1)$ are part of the message ξ^1.

For $i > 1$, $a \le c - f$, we have

$$\bar{S}_{(a,i)}(\tilde{\xi^1}) = J_{(a,i),(a,i)}\tilde{\xi}^1_{(a,i)} + \sum_{\substack{b=1\\b\neq a}}^{c}\sum_{j=1}^{l} J_{(a,i),(b,j)}\tilde{\xi}^1_{(b,j)} = \sum_{\substack{b=1\\b\neq a}}^{c-f} J_{(a,i),(b,1)} \le c - f - 1.$$

It is indeed possible to produce errors in the remaining f clusters where the neurons have been deleted. The local field of the neurons $(a, 1)$, $a > c - f$, is again equal to $c - f$. This is the maximal attainable value because the total number of activated neurons is equal to $c - f$. But in the clusters $c - f + 1, \ldots, c$, this value can be reached by the other neurons of these clusters, $i \neq 1$. So an error occurs if there is a neuron in one of these clusters not belonging to ξ^1 which gets as well $c - f$ signals. This is the case if and only if this neuron is (actively) connected to each one of the $c - f$ activated neurons. By Lemma 6.22 the probability of this event is, for a fixed (a, i), asymptotically given by $(1 - e^{-\alpha})^{c-f}$.

The probability of an error is therefore bounded from above by

$$\begin{aligned} &\mathbb{P}\left(\exists (a, i) : T_{(a,i)}(\tilde{\xi^1}) \neq \xi^1_{(a,i)}\right) \le f(l - 1)\mathbb{P}\left(T_{(c,2)}(\tilde{\xi^1}) \neq \xi^1_{(c,2)}\right) \\ =&f(l - 1)(1 - e^{-\alpha})^{c-f}(1 + o(1)) \le \exp\left[\log(\varrho c) + c + (c - f)\log\left(1 - e^{-\alpha}\right)\right](1 + o(1)) \\ =&\exp\left[\log(\varrho c) + c + (1 - \varrho)c\log\left(1 - e^{-\alpha}\right)\right](1 + o(1)) \end{aligned}$$

and convergence to 0 is reached if $(1 - \varrho)\log(1 - e^{-\alpha}) < -1$; this is fulfilled if

$$\alpha < -\log\left(1 - e^{-1/(1-\varrho)}\right).$$

This shows the lower bound mentioned in Theorem 6.21.

For the upper bound, we fix a cluster in which the neuron has been deleted, e. g., the last cluster. The probability of an error is at least

$$\begin{aligned}
&\mathbb{P}\left(\exists(a,i) : T_{(a,i)}(\tilde{\xi^1}) \neq \xi^1_{(a,i)}\right) = 1 - \mathbb{P}\left(\forall(a,i) : T_{(a,i)}(\tilde{\xi^1}) = \xi^1_{(a,i)}\right)\\
\geq& 1 - \mathbb{P}\left(\forall i > 1 : T_{(c,i)}(\tilde{\xi^1}) = 0\right)\\
=& 1 - \mathbb{P}\left(\forall i > 1 : \exists b_i \in \{1,\dots,c-f\} : J_{(c,i),(b_i,1)} = 0\right).
\end{aligned}$$

The probability in the last line is bounded by

$$\begin{aligned}
&\mathbb{P}\left(\forall i > 1 : \exists b_i \in \{1,\dots,c-f\} : J_{(c,i),(b_i,1)} = 0\right)\\
\leq& \mathbb{P}\left(\exists b_2 \in \{1,\dots,c-f\} : J_{(c,2),(b_2,1)} = 0\right)^{l-1}.
\end{aligned}$$

This inequality is first explained by a plausibility argument: the fact that neuron $(c,2)$ is not connected to all of the neurons $(1,1),\dots,(c-f,1)$ gives us no information about the connection between neuron $(c,3)$ and $(1,1),\dots,(c-f,1)$ besides the fact that $(c,2)$ has possibly not been activated often enough. This is because different neurons of one cluster cannot be part of the same message and thus the connections of neuron $(c,2)$ give us no information on the connections of $(c,3),\dots,(c,l)$ to $(1,1),\dots,(c-1,1)$. If $(c,2)$ has not been activated so often, the probability of an activation of one of the neurons $(c,3),\dots,(c,l)$ increases and so does the probability of the connection of one of these neurons to the complete set $\{(1,1),\dots,(c-f,1)\}$. Certainly the fact that $(c,2)$ is not connected to each of the neurons of $\{(1,1),\dots,(c-f,1)\}$ does not decrease the probability that one of the neurons $(c,3),\dots,(c,l)$ is connected to all neurons $\{(1,1),\dots,(c-1,1)\}$.

To give a rigorous proof of this inequality, we recall that for fixed μ, the set of random variables $\xi^\mu_{(c,i)}$, $i \geq 2$, is negatively associated. So the variables $(\xi^\mu_{(c,i)}, i \geq 2, \mu \geq 2)$, are negatively associated (Lemma 2.12, 3.)). Taking increasing functions of disjoint subsets of negatively associated random variables preserves negative association, so

$$\sum_{\mu=2}^{M} \xi^\mu_{(c,i)}, \quad i \geq 2$$

are negatively associated.

The conditional probabilities

$$\mathbb{P}\left(\sum_{b=1}^{c-f} J_{(c,i),(b,1)} < c-f \,\middle|\, \sum_{\mu=2}^{M} \xi^\mu_{(c,i)} = k\right), \quad i \geq 2$$

are decreasing in k, and additionally we have for each tuple $(k_2,\dots,k_l) \in \{0,\dots,M-1\}^l$ with positive probability under the distribution of $(\sum_{\mu=2}^M \xi^\mu_{(c,2)},\dots,\sum_{\mu=2}^M \xi^\mu_{(c,l)})$, i.e. $\sum_{i=2}^l k_i \leq M-1$,

$$\mathbb{P}\left(\forall i \geq 2 : \sum_{b=1}^{c-f} J_{(c,i),(b,1)} < c-f \,\middle|\, \forall i \geq 2 : \sum_{\mu=2}^{M} \xi^\mu_{(c,i)} = k_i\right)$$

$$= \prod_{i=2}^{l} \mathbb{P}\left(\sum_{b=1}^{c-f} J_{(c,i),(b,1)} < c - f \Big| \sum_{\mu=2}^{M} \xi^{\mu}_{(c,i)} = k_i \right).$$

We define the decreasing functions $f_i : \{0, \ldots, M-1\} \longrightarrow \mathbb{R}$, $i > 1$:

$$f_i(k_i) = \mathbb{P}\left(\sum_{b=1}^{c-f} J_{(c,i),(b,1)} < c - f \Big| \sum_{\mu=2}^{M} \xi^{\mu}_{(c,i)} = k_i \right).$$

Using the first part of Lemma 2.12, we obtain the desired inequality:

$$\mathbb{P}\left(\forall i > 1 : \exists b_i \in \{1, \ldots, c-f\} : J_{(c,i),(b_i,1)} = 0\right) = \mathbb{P}\left(\forall i > 1 : \sum_{b=1}^{c-f} J_{(c,i),(b,1)} < c - f \right)$$

$$= \mathbb{E}_{\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(c,2)}, \ldots, \sum_{\mu=2}^{M} \xi^{\mu}_{(c,l)}\right)} \left[\prod_{i=2}^{l} f_i \left(\sum_{\mu=2}^{M} \xi^{\mu}_{(c,i)} \right) \right] \leq \prod_{i=2}^{l} \mathbb{E}_{\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(c,2)}, \ldots, \sum_{\mu=2}^{M} \xi^{\mu}_{(c,l)}\right)} \left[f_i \left(\sum_{\mu=2}^{M} \xi^{\mu}_{(c,i)} \right) \right]$$

$$= \prod_{i=2}^{l} \mathbb{P}\left(\sum_{b=1}^{c-f} J_{(c,i),(b,1)} < c - f \right) = \mathbb{P}\left(\exists b_2 \in \{1, \ldots, c-f\} : J_{(c,2),(b_2,1)} = 0\right)^{l-1}.$$

Back to the last line of the precedent calculation, we continue by applying Lemma 6.22:

$$1 - \mathbb{P}\left(\exists b \in \{1, \ldots, c-f\} : J_{(c,2),(b,1)} = 0\right)^{l-1} = 1 - \mathbb{P}\left(\sum_{b=1}^{c-f} J_{(c,2),(b,1)} \leq c - f - 1 \right)^{l-1}$$

$$= 1 - \left[1 - \mathbb{P}\left(\sum_{b=1}^{c-f} J_{(c,2),(b,1)} = c - f \right) \right]^{l-1} = 1 - \left[1 - \left(1 - e^{-\alpha}\right)^{c-f} (1 + o(1)) \right]^{l-1}$$

$$= 1 - \left[1 - \exp\left[(c-f) \log\left(1 - e^{-\alpha}\right)\right] (1 + o(1)) \right]^{l-1}.$$

A sufficient condition for the convergence to 1 of the expression in the last line is $(1 - \varrho) \log\left(1 - e^{-\alpha}\right) > -1$, and this is for $\varrho < 1$, $\alpha > 0$ equivalent to

$$\alpha > -\log\left(1 - e^{-1/(1-\varrho)}\right).$$

□

Remark 6.23 *The results of Theorem 6.21 are the same, if we use $S(\beta)$ instead of $\bar{S}$, $\beta \in (0,1)$. This follows immediately because the self-signal is, for the activated neurons, positive and can, as long as the activated neuron in the cluster is not erased, not be compensated by non-excited neurons.*

We can use most of the proof of Theorem 6.21 to make the following observations concerning the error correction of the GB model without self-loops:

Corollary 6.24 *1. The maximal capacity variable of the GB model without self-loops, using $M = \alpha l^2$ is*

$$\alpha^* = -\log(1 - e^{-1}).$$

For each $\alpha < \alpha^*$ and arbitrary but fixed μ, we have $\lim_{l \to \infty} \mathbb{P}\left(T(\xi^{\mu}) = \xi^{\mu}\right) = 1$.

2. *In contrast, if $M = \alpha l^2$ with $\alpha > \alpha^*$, a fixed stored pattern is not stable, with high probability: we have* $\lim_{l\to\infty} \mathbb{P}\left(T(\xi^\mu) \neq \xi^\mu\right) = 1$.

3. *If ϱc bits of ξ^μ are erased, for arbitrary μ, we can correct the partially deleted pattern $\tilde{\xi}^\mu$ in one step,*
$$\lim_{l\to\infty} \mathbb{P}\left(T(\tilde{\xi}^\mu) = \xi^\mu\right) = 1$$
if
$$\alpha < -\log(1 - e^{-1/(1-\varrho)}).$$

4. *On the contrary, the pattern is with high probability not corrected in one step,*
$$\lim_{l\to\infty} \mathbb{P}\left(T(\tilde{\xi}^\mu) \neq \xi^\mu\right) = 1,$$
if
$$\alpha > -\log(1 - e^{-1/(1-\varrho)}).$$

5. *The stability of all stored patterns is guaranteed with high probability if*
$$\alpha < -\log(1 - e^{-3}).$$

Remark 6.25 *This shows that the self-loops are advantageous for this model, because we can guarantee the stability of each stored pattern but do not loose error correcting abilities concerning the error correction in one step.*

Proof of Corollary 6.24: We can adopt the proof of Theorem 6.21 and will use the same notation and assumptions: we consider ξ^1 and assume that $\xi^1 = (e_1, \ldots, e_1)$ and that the entries of the last f cluster have been deleted and set $f = 0$ if we want to analyse the stability. We just have to reflect the differences between the two models. The maximal attainable value of $S_{(a,i)}(\xi^1)$ is, within each cluster, equal to $c - f$ (deleted clusters) or to $c-f-1$ (non-deleted clusters), respectively. Each of the neurons can attain this value, the activated neurons have no advantage by the self-loops. In particular, an error is possible in each cluster, not only in the ones concerned by the deletion of the activated neuron. In the case $f = 0$, this leads to the loss of the guaranteed stability.

In a fixed cluster, there occurs an error if and only if there is at least one neuron not belonging to the message ξ^μ that is actively connected to the $c - f$ activated neurons (or $c - f - 1$, respectively) in the other clusters. We get an upper bound of the probability of the failure of the correction in one step by

$$\begin{aligned}
&\mathbb{P}\left(\exists (a,i) : T_{(a,i)}(\tilde{\xi}^1) \neq \xi^1_{(a,i)}\right) \leq c(l-1)\mathbb{P}\left(T_{(1,2)}(\tilde{\xi}^1) \neq \xi^1_{(1,2)}\right) \\
=&c(l-1)\mathbb{P}\left(\sum_{b=2}^{c-f} J_{(1,2),(b,1)} = c - f - 1\right) \leq c(l-1)(1-e^{-\alpha})^{c-f-1}(1+o(1)),
\end{aligned}$$

the last line obtained by Lemma 6.22. This again vanishes if

$$(1-\varrho)\log\left(1 - e^{-\alpha}\right) < -1,$$

i.e.

$$\alpha < -\log\left(1 - e^{-1/(1-\varrho)}\right).$$

Especially, for $f = 0$, we obtain

$$\alpha < -\log\left(1 - e^{-1}\right).$$

For the upper bound, we again fix a cluster, e. g., the first one, and show that there will be an error with high probability if we use $\alpha > -\log\left(1 - e^{-1/(1-\varrho)}\right)$. Here the only difference to the model with self-loops is that in the latter model, an error can only occur in a cluster where the correct neuron has been deleted. Therefore there cannot be an error if a stored pattern ξ^μ is the input. In the model without self-loops, there can be an error in each cluster, despite the fact whether the correct neuron has been deleted or not. It is thus possible to get an incorrect output if the input pattern is a stored pattern ξ^μ and there is an upper bound of the number of stored patterns.

Since the proof only uses one fixed cluster and shows that there will be an error with high probability if too many patterns are stored, the proof is the same as the proof of Theorem 6.21, completed by the case $\varrho = 0$ in case where the input is a stored pattern. We obtain

$$\lim_{l\to\infty} \mathbb{P}\left(T(\tilde{\xi}^1) \neq \xi^1\right) \geq 1 - \lim_{l\to\infty}\left(1 - \left(1 - e^{-\alpha}\right)^{c-f-1}(1 + o(1))\right)^l = 1$$

if $\alpha > -\log(1 - e^{-1/(1-\varrho)})$, in particular if $\alpha > -\log(1 - e^{-1})$ in case $\varrho = 0$.

□

Remark 6.26 *For an arbitrary stored pattern ξ^μ and a randomly deleted version $\tilde{\xi}^\mu$ of ξ^μ, where ϱc, $\varrho > 0$, neurons have been deactivated, we observe concerning the retrieval process with multiple steps, using S or $\bar{S}$:*

1. *If $\alpha > -\log(1 - e^{-1})$, the message can never be recovered, with high probability.*
2. *If $\alpha < -\log(1 - e^{-1/(1-\varrho)})$, the pattern is corrected in one step, with high probability.*

Concerning *1.*, if $\alpha > -\log(1 - e^{-1})$, in an arbitrary cluster where the entry has been deleted, there is at least one neuron that is connected to every neuron of the other $c - 1$ clusters that belong to ξ^μ (see the proof of Corollary 6.24, *2.*). This neuron is activated in the first step because it is in particular connected to every one of the $c - \varrho c$ initially excited neurons. Since it is connected to every of the $c - 1$ neurons in different clusters being part of ξ^μ, it remains active; this error will never be corrected. The second statement is a result obtained in Theorem 6.21/Corollary 6.24.

The GB model with SUM-of-MAX rule and self-loops is particularly resistent to spuriously activated neurons. In the other models analysed in this thesis, a number of ϱc spuriously activated neurons decreases the capacity noticeably, mostly antiproportionally with $(1+\varrho)$. This is not the case in the GB model: it is possible to correct a comparatively high number of spuriously activated neurons.

Proposition 6.27 *In the GB model with self-loops and SUM-of-MAX rule and $M = \alpha l^2$ stored patterns, with some $\alpha > 0$, let the input pattern be a corrupted version $\tilde{\xi}^\mu$ of a stored message ξ^μ. In each cluster, there are up to ϱc neurons spuriously activated, for some fixed but arbitrary $\varrho > 0$. Then the pattern is corrected in one step, with high probability:*

$$\mathbb{P}(T(\tilde{\xi}^\mu) = \xi^\mu) \longrightarrow 1$$

as N tends to infinity.

Proof: We first state that the message is not recovered if and only if there is a clique within the activated neurons (spuriously or correctly active) that is not the message itself; this means, the clique forms a valid message and each of the connecting edges is active. If there is not such a clique, the correct neurons obtain each at least one signal per cluster, in total c signals. Neurons that do not belong to the active neurons of $\tilde{\xi}^\mu$ can only get up to $c - 1$ signals and remain inactive. The activated neurons will all be deactivated after some steps because there is no clique that includes a subset of the initially activated neurons. If there is in contrary such a clique, the contained neurons remain activated forever, as the neurons belonging to ξ^μ.

We proceed as in the proof of Proposition 6.20 and split the probability into a sum, the summands distinguished by the number of neurons that differ from the original message. The big difference to this proof is that the number of possible cliques that differ from ξ^μ in i clusters/neurons decreases drastically from $\binom{c}{i} l^i$ to at most $\binom{c}{i}(\varrho c)^i$ because we can only choose from the initially active neurons. Using the upper bound of the probability of $\mathcal{G}(\xi^0(i))$ of such a clique derived in Proposition 6.19, we observe that the probability of not being able to correct the pattern tends to 0.

□

6.2.2 The GB Model with Binary Synaptic Efficacies and Threshold Dynamics

This model was as well proposed by Gripon and Berrou, e. g., in [23]. It respects most of the structure of the model of the previous subsection. The only difference is that it does not use a WTA, but a threshold dynamics. This means that we keep the cluster structure and the definitions of the synaptic efficacies and the local field ($S_{(a,i)}$, $S(\beta)_{(a,i)}$ and $\bar{S}_{(a,i)}$, respectively). We recall

$$J_{(a,i),(b,j)} = \Theta\left(\sum_{\mu=1}^{M} \xi^\mu_{(a,i)} \xi^\mu_{(b,j)} - 1\right), \quad a \neq b \vee i = j,$$

$$S(\beta)_{(a,i)}(\sigma) = \sum_{b=1, b\neq a}^{c} \Theta\left(\sum_{j=1}^{l} J_{(a,i),(b,j)} \sigma_{(b,j)} - 1\right) + \beta J_{(a,i),(a,i)} \sigma_{(a,i)},$$

$S_{(a,i)}(\sigma) = S(0)_{(a,i)}(\sigma)$ and $\bar{S}_{(a,i)}(\sigma) = S(1)_{(a,i)}(\sigma)$.

The dynamics in this version of the GB model is defined by

$$T(\beta)_{(a,i)}(\sigma) = \Theta\left(S(\beta)_{(a,i)}(\sigma) - h\right).$$

The threshold h is chosen in dependence of the tasks the network should perform. As by the application of the SUM-OF-MAX rule and for the same reasons, choosing $h = c$ and $T(1)$ implies stability of all stored messages $\xi \in \{\xi^1, \dots, \xi^M\}$ without limiting the number M. We also shortly consider the dynamics $T(\beta)(\sigma)$, $\beta \in (0,1)$ using $S(\beta)$ and show that its performance is the same as the one of $T(1)$ and that $T(0)$ performs worse compared to $T(1)$. The dynamics $T(1)$ using $\bar{S}$ is called T in this section.

First, we examine the stability of the stored patterns. The model shows similar properties to the Willshaw model with threshold dynamics. Other than in the latter model, we can reach that every message is stable without limiting the number of stored patterns. This is reached by choosing $\bar{S}$ and $h = c$. If $h < c$, the network performs better than the Willshaw model with independent and identically distributed spins, because it is possible to find a bound on α such that all patterns are stable, with high probability. Concerning the stability and error correction of an arbitrary stored pattern, the bounds on α are the same as in the both versions of the Willshaw model with threshold dynamics.

We prove the following theorem:

Theorem 6.28 *In the GB model threshold dynamics, we observe for the different variations of the model, concerning the stability of the stored patterns and the error correction:*

1. *In the GB model with self-loops, threshold dynamics and binary synaptic efficacies, all the stored messages are stable, if we set the threshold to $h = c$.*

2. *The same is true for the model using $S(\beta)$, $\beta > 0$, by choosing $h = c - 1 + \beta$.*

3. *In the model without self-loops, i.e. $\beta = 0$, $M = \alpha l^2$ patterns can be stored such that a randomly chosen ξ^μ is stable with probability tending to 1. $\mathbb{P}(T(\xi^\mu) = \xi^\mu) \longrightarrow 1$, if*
$$\alpha < -\log(1 - e^{-1})$$
and the threshold $h = c - 1$ is used.

4. *For arbitrary $\beta \in [0,1]$, using $S(\beta)$, the threshold $h = \gamma c - 1 + \beta$, $\gamma \in (0,1)$, will provide that an arbitrary stored pattern is stable with high probability, if*
$$\alpha < -\log(1 - \gamma)$$
and
$$-\gamma \log\left(\frac{\gamma}{1 - e^{-\alpha}}\right) + (1 - \gamma)\log\left(\frac{e^{-\alpha}}{1 - \gamma}\right) < -1.$$

5. *For arbitrary $\beta \in [0,1]$, a randomly deleted message $\tilde{\xi^\mu}$ obtained by deleting ϱc entries, $\varrho > 0$, of ξ^μ can be corrected in one step with high probability, if $\varrho < 1 - \gamma$ and if α is chosen accordingly to the conditions in 4.). Then*
$$\lim_{l \to \infty} \mathbb{P}\left(T(\tilde{\xi^\mu}) = \xi^\mu\right) = 1.$$

6. *The bounds in 3.), 4.) and 5.) are sharp: we have, in the model without self-loops and* $h = c - 1$, $\mathbb{P}\left(T(0)(\xi^\mu) \neq \xi^\mu\right) \longrightarrow 1$, *if*

$$\alpha > -\log(1 - e^{-1})$$

and, for $\beta \in [0,1]$ *and* $h = \gamma c - 1 + \beta$, $\mathbb{P}\left(T(\beta)(\xi^\mu) \neq \xi^\mu\right) \longrightarrow 1$, *if a.)* $\alpha < -\log(1-\gamma)$ *and*

$$-\gamma \log\left(\frac{\gamma}{1 - e^{-\alpha}}\right) + (1 - \gamma)\log\left(\frac{e^{-\alpha}}{1 - \gamma}\right) > -1$$

or if b.) $\alpha \geq -\log(1-\gamma)$. *If a pattern is partially deleted and* $\varrho > 1 - \gamma$, *the pattern cannot be corrected: then* $\mathbb{P}(T(\beta)(\tilde{\xi^\mu}) = (0, \ldots, 0)) = 1$.

7. *Using the threshold* $h = \gamma c - 1 + \beta$, $\gamma \in (0,1)$ *and arbitrary* $\beta \in [0,1]$, *the whole set of stored messages is stable concerning the dynamics using* $S(\beta)$ *if* $\alpha < -\log(1-\gamma)$ *and*

$$-\gamma \log\left(\frac{\gamma}{1 - e^{-\alpha}}\right) + (1 - \gamma)\log\left(\frac{e^{-\alpha}}{1 - \gamma}\right) < -3 :$$

then

$$\lim_{l \to \infty} \mathbb{P}\left(\exists \mu \geq 1, \exists (a,i) : T(\beta)_{(a,i)}(\xi^\mu) \neq \xi^\mu_{(a,i)}\right) = 0.$$

Proof of Theorem 6.28: The first two statements follow immediately because they have already been proven in Proposition 6.4 concerning the SUM-OF-MAX rule. The neurons belonging to ξ^1 collect c signals, each, but the neurons not belonging to ξ^1 only can get $c - 1$ signals.

The model without self-loops, however, provides no guaranteed stability of the stored patterns. To achieve a preferably high number of stored messages (which will go to the expense of the error correcting abilities), we choose $h = c-1$ and are in the same situation as in the GB model with WTA dynamics without self-loops, see Corollary 6.24. This shows the third statement.

Concerning 4.) and 5.), if we choose a threshold $h = \gamma c - 1 + \beta$, with fixed $\gamma < 1$, and consider ξ^1 that is assumed to be the pattern $(e_1, \ldots, e_1)$, the activated neurons of ξ^1 are trivially stable. Their local field is

$$S(\beta)_{(a,1)}(\xi^1) = c - 1 + \beta > \gamma c - 1 + \beta.$$

If ϱc entries have been deleted, with $\varrho < 1 - \gamma$, it is still at least equal to

$$S(\beta)_{(a,1)}(\tilde{\xi}^1) \geq \min((1-\varrho)c - 1 + \beta, (1-\varrho)c) \geq (1-\varrho)c - 1 + \beta > \gamma c - 1 + \beta.$$

If $\varrho > 1 - \gamma$, none of the neurons can get enough signals and all neurons are deactivated in one step.

In 4.) and in 5.), the remaining neurons, these are the inactive neurons of ξ^1, are stable, if they are connected to less than $\gamma c - 1 + \beta$ of the activated neurons of ξ^1. Noticing that the probability of an error is decreasing if the number of deleted neurons increases, Lemma 6.22 bounds the probability of an error concerning the inactive neurons of ξ^1 (if either ξ^1 or $\tilde{\xi}^1$ is the input) by

$$\begin{aligned}\mathbb{P}\left(\exists(a,i), i\geq 2 : T(\beta)_{(a,i)}(\xi^1)=1\right) &\leq c(l-1)\mathbb{P}\left(\sum_{b=2}^{c} J_{(1,2),(b,1)} \geq \gamma c - 1 + \beta\right)\\ =&c(l-1)\mathbb{P}\left(R_{(1-e^{-\alpha}),c-1}\geq \gamma c - 1+\beta\right)(1+o(1))\end{aligned}$$

with a Binomially distributed random variable $R_{(1-e^{-\alpha}),c-1}$ having parameters $\tilde{p}=1-e^{-\alpha}$ and $c-1$. By the application of the exponential Chebyshev inequality, we obtain for $t>0$

$$\begin{aligned}\mathbb{P}\left(R_{(1-e^{-\alpha}),c-1}\geq \gamma c-1+\beta\right) &\leq \exp\left[-\gamma ct+(1-\beta)t\right]\left(1-\tilde{p}+\tilde{p}e^t\right)^{c-1}\\ &\leq \exp\left[-\gamma ct+(1-\beta)t+(c-1)\log(1+\tilde{p}(e^t-1))\right].\end{aligned}$$

We are now in the same situation as in the proof of Theorem 4.9. The probability converges to 0 if $\alpha<-\log(1-\gamma)$ and $-\gamma\log(\frac{\gamma}{1-e^{-\alpha}})+(1-\gamma)\log(\frac{e^{-\alpha}}{1-\gamma})<-1$.

Concerning 7.), the probability of the existence of an unstable message is bounded by

$$\mathbb{P}\left(\exists\mu,\exists(a,i): T(\beta)_{(a,i)}(\xi^\mu)\neq\xi^\mu_{(a,i)}\right)\leq Ml c\mathbb{P}\left(T(\beta)_{(a,i)}(\xi^\mu)\neq\xi^\mu_{(a,i)}\right)$$

for some fixed neuron (a,i) and message μ such that $\xi^\mu_{(a,i)}=0$. As in the previous part of the proof concerning 4.) and 5.), using $t=\log\frac{\gamma(1-\tilde{p})}{(1-\gamma)\tilde{p}}$ yields convergence to 0 if $\alpha<-\log(1-\gamma)$ and

$$-\gamma\log\frac{\gamma}{1-e^{-\alpha}}+(1-\gamma)\log\frac{e^{-\alpha}}{1-\gamma}<-3.$$

To show the sharpness of the bounds as claimed in 6.), we use again Lemma 6.22. We recall that for fixed a, e.g., $a=1$,

$$\sum_{\mu=2}^{M}\xi^\mu_{(1,i)},\quad i\geq 2$$

are negatively associated. The conditional probabilities

$$\mathbb{P}\left(\sum_{b=2}^{c}\Theta\left(\sum_{\mu=2}^{M}\xi^\mu_{(b,1)}\xi^\mu_{(1,i)}-1\right)<\gamma c\,\Big|\sum_{\mu=2}^{M}\xi^\mu_{(1,i)}=k\right)$$

are for each $i\geq 2$ decreasing in k. In addition, as in the proof of Theorem 6.21, for some vector $(k_2,\ldots,k_l)$ with $\mathbb{P}(\sum_{\mu=2}^{M}\xi^\mu_{(1,2)}=k_2,\ldots,\sum_{\mu=2}^{M}\xi^\mu_{(1,l)}=k_l)>0$, the following conditional probability is equal to

$$\begin{aligned}&\mathbb{P}\left[\forall i\geq 2:\sum_{b=2}^{c}\Theta\left(\sum_{\mu=2}^{M}\xi^\mu_{(b,1)}\xi^\mu_{(1,i)}-1\right)<\gamma c\,\Big|\sum_{\mu=2}^{M}\xi^\mu_{(1,2)}=k_2,\ldots,\sum_{\mu=2}^{M}\xi^\mu_{(1,l)}=k_l\right]\\ =&\prod_{i=2}^{l}\mathbb{P}\left[\sum_{b=2}^{c}\Theta\left(\sum_{\mu=2}^{M}\xi^\mu_{(b,1)}\xi^\mu_{(1,i)}-1\right)<\gamma c\,\Big|\sum_{\mu=2}^{M}\xi^\mu_{(1,i)}=k_i\right].\end{aligned}$$

This is because only one neuron of cluster 1 can be part of message μ. Taking the functions

$$f_i(k_i) = \mathbb{P}\left[\sum_{b=2}^{c} \Theta\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(b,1)}\xi^{\mu}_{(1,i)} - 1\right) < \gamma c \Big| \sum_{\mu=2}^{M} \xi^{\mu}_{(1,i)} = k_i\right],$$

and using Lemma 2.12, we conclude, as in the last part of the proof of Theorem 6.21:

$$\begin{aligned}&\mathbb{P}\left(\forall i \geq 2 : \sum_{b=2}^{c} \Theta\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(b,1)}\xi^{\mu}_{(1,i)} - 1\right) < \gamma c\right)\\ \leq &\prod_{i=2}^{l} \mathbb{P}\left(\sum_{b=2}^{c} \Theta\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(b,1)}\xi^{\mu}_{(1,i)} - 1\right) < \gamma c\right).\end{aligned}$$

With this result, we have

$$\begin{aligned}&\mathbb{P}\left(\exists i \geq 2 : \sum_{b=2}^{c} \Theta\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(b,1)}\xi^{\mu}_{(1,i)} - 1\right) \geq \gamma c)\right)\\ =&1 - \mathbb{P}\left(\forall i \geq 2 : \sum_{b=2}^{c} \Theta\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(b,1)}\xi^{\mu}_{(1,i)} - 1\right) < \gamma c\right)\\ \geq&1 - \mathbb{P}\left(\sum_{b=2}^{c} \Theta\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(b,1)}\xi^{\mu}_{(1,2)} - 1\right) < \gamma c\right)^{l-1}.\end{aligned}$$

By Lemma 6.22, the random variable $\sum_{b=2}^{c} \Theta(\sum_{\mu=2}^{M} \xi^{\mu}_{(b,1)}\xi^{\mu}_{(1,2)} - 1)$ is asymptotically distributed as a Binomial random variable with parameters $1-e^{-\alpha}$ and $c-1$. The probability of an error thus tends to 1 if

$$\left[1 - \mathbb{P}\left(\sum_{b=2}^{c} \Theta\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(b,1)}\xi^{\mu}_{(1,2)} - 1\right) \geq \gamma c\right)\right]^{l-1} \longrightarrow 0;$$

this is fulfilled if

$$\lim_{l\to\infty} \frac{1}{\log(l)} \quad \log\left(\mathbb{P}\left(\sum_{b=2}^{c} \Theta\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(b,1)}\xi^{\mu}_{(1,2)} - 1\right) \geq \gamma c\right)\right) > -1.$$

Using Lemma 2.4, the limit is, if $\alpha < -\log(1-\gamma)$,

$$\begin{aligned}&\lim_{l\to\infty} \frac{1}{\log(l)} \quad \log\left(\mathbb{P}\left(\sum_{b=2}^{c} \Theta\left(\sum_{\mu=2}^{M} \xi^{\mu}_{(b,1)}\xi^{\mu}_{(1,2)} - 1\right) \geq \gamma c\right)\right)\\ &= -\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right).\end{aligned}$$

So the probability tends to 1 if $-\gamma \log(\frac{\gamma}{1-e^{-\alpha}})+(1-\gamma)\log(\frac{e^{-\alpha}}{1-\gamma}) > -1$. The sharpness of the stability bound of the model using S and $\gamma = 1$ follows from the proof of Theorem 6.21.

□

Remark 6.29 *The results of the analysis concerning the sets of possible capacity variables in dependence on* γ *of Proposition 4.10 are also valid for the GB model using a threshold* $\gamma c - 1 + \beta$, $\gamma < 1$.

Remark 6.30 *The threshold dynamics does not allow for improvement after multiple steps, if the input was a stored pattern or a partially erased stored pattern. The pattern is either corrected in one step or never corrected.*

If the number of active neurons in the erased pattern does not exceed the threshold, the input is mapped to $(0, \ldots, 0)$ and never corrected. If on the other hand the number of active neurons exceeds the threshold, the initially activate neurons remain active, since the input is a stored or partially erased stored message and they receive signals from each other activated neuron. If the message is not corrected in one step, there is a spurious neuron that has been activated in the first step of the dynamics. It remains active after each following step, because it obtains enough signals in the first step and the initially active neurons will also remain active in each subsequent iteration.

Corollary 6.31 *If a threshold variable* $\gamma < 1$ *is used, it is as well possible to correct patterns with* $\varrho_1 c$ *incorrect entries and* $\varrho_2 c$ *deleted ones. We assume that exactly one neuron is activated in the* $\varrho_1 c$ *wrong clusters and it has replaced the correct neuron. Then the pattern can be corrected in one step if*

$$\varrho_1 + \varrho_2 < 1 - \gamma$$

and α *is chosen in order to guarantee stability of the stored patterns, i.e.*

$$-\gamma \log\left(\frac{\gamma}{1-e^{-\alpha}}\right) + (1-\gamma)\log\left(\frac{e^{-\alpha}}{1-\gamma}\right) < -1.$$

Proof: Trivially, the 1's (erased or not) in the stored message are corrected in one step (or remain) if the first condition holds. The non-excited neurons of the stored message are stable/corrected if they are connected to less than γc of the activated neurons; since there is at most one active neuron per cluster, this probability is bounded by the corresponding probability of the stability of the inactive neurons in a stored pattern and tends to 0 if the stability condition of Theorem 6.28 is fulfilled. This is, in any case, a condition that should be fulfilled by α, even if the correction would only need a less restrictive bound.

□

6.3 Other Variations of the GB Model

Comparable to the model treated in the first section of this chapter that corresponds to Amari's model, it is also possible to define analogons to the models of chapters 3 and 5. The neurons remain organised in clusters, and per message and cluster, there is exactly one activated neuron. An active neuron can take one of the two values -1 or 1, each one with equal probability. Each message has the form $\xi^\mu = (\xi^\mu_{(a,i)})_{(a,i)\in\{1,\dots,c\}\times\{1,\dots,l\}}$, where $\xi^\mu_{(a,i)}$ denotes the value of (a,i) in message μ. The messages are chosen independently and accordingly to the uniform distribution on the set of all messages with exactly one active neuron per cluster. The underlying graph is given by the edge set $E = \{\{(a,i),(b,j)\} : a \neq b,\, i,j \in \{1,\dots,l\}\}$. For the cluster versions of the Ternary simple and the BEG model, the synaptic weights $J_{(a,i)(b,j)}$ are defined by

$$J_{(a,i),(b,j)} = \sum_{\mu=1}^{M} \xi^\mu_{(a,i)}\xi^\mu_{(b,j)}.$$

In the BEG version, there are additionally the variables

$$K_{(a,i),(b,j)} = \frac{1}{(1-1/l)^2}\sum_{\mu=1}^{M} \eta^\mu_{(a,i)}\eta^\mu_{(b,j)}, \quad \eta^\mu_{(a,i)} = (\xi^\mu_{(a,i)})^2 - 1/l.$$

The dynamics T are defined as in the models analysed in the corresponding chapters. The results and bounds on α are expected to be the same as the ones obtained during the analysis of the corresponding models in chapters 3 and 5.

However, the cluster structure only particularly improves the performance and is advantageous to the models without cluster structure in the self-loop accepting version with SUM-of-MAX rule.

7 Comparison of the Different Models

There are different aspects to compare the different models, first of all, the critical capacity variable α^*. Though using threshold variables $\gamma \geq 1$, Amari's model offers the worst critical capacity variable $\alpha^*_{\mathrm{Am}} \approx 0.255$. Capacity variables $\gamma \geq 1$ used in this model use the random noise of the local field and have to be taken with care; if one only uses threshold variables $\gamma < 1$, the critical variable for this model is even less, $\alpha^*_{\mathrm{Am}_{<1}} \approx 0.1585$. The Ternary simple model performs better, with a critical capacity variable $\alpha^*_{\mathrm{Ts}} \approx 0.38$, but one has to keep in mind that the situation is not fair because the state space of the latter model is ternary. The Willshaw model with threshold dynamics or WTA algorithm in turn outperforms the Ternary simple model, which is remarkable, because it only has a binary state space. Its critical capacity variable α^*_{Wi} is approximately 0.45. The best capacity of chapters 2-5 is reached by the BEG model with $\alpha^*_{\mathrm{BEG}} \approx 0.51$. Finally, the GB model in its version with self-loops and SUM-of-MAX rule has a structure that allows to store an umlimited number of patterns such that each one is stable. Of course, too many stored patterns go at the expense of the error correcting abilities of the network. In the next paragraph, we compare this property for partially deleted patterns.

The sharp bounds on the capacity variable for the models using a threshold are visualised in Figure 7.1. In Amari's model using threshold variables $\gamma > 1$, α must additionally fulfill $\alpha > \gamma - 1$.

Concerning the error correcting abilities, let us suppose that a number of $\varrho \log(N)$ (ϱc in the different versions of the GB model) of the active neurons of a stored pattern has been deactivated. A number of $\varrho \log(N)$ deleted active neurons in a stored pattern can in all models using a threshold, except the BEG model, be corrected with high probability if $\gamma < 1 - \varrho$. In the BEG model $\varrho \log(N)$ deleted neurons can be corrected if $\gamma < 2 - 2\varrho$. These bounds for γ are in all models sharp, except in Amari's model (and its equivalent, the GB model with weighted synaptic efficacies). In this model the usage of $\gamma \geq 1 - \varrho$ is possible but requires a certain random noise and does not only rely on the signal coming from the stored message. We therefore compare the critical capacity variable for the threshold variable $\gamma = 1 - \varrho$ to the critical capacity variables of the other models. This is the critical threshold such that no minimal value for α is required and to keep the active neurons stable only by the signal term of the local field, without using the noise term. The bounds on γ determine in the remaining models with threshold algorithms the critical values for α, that is, the supremum of all α that can be used in order to guarantee a correction of this pattern, with probability converging to 1. In the Willshaw model with WTA algorithm and in the GB model with SUM-of-MAX rule, the critical variable α to correct $\varrho \log(N)$ errors is $\alpha(\varrho) = -\log(1 - e^{-1/(1-\varrho)})$, as determined in chapters 4 and 6. Figure 7.2 demonstrates the bounds on α: for all models except Amari's model, it shows

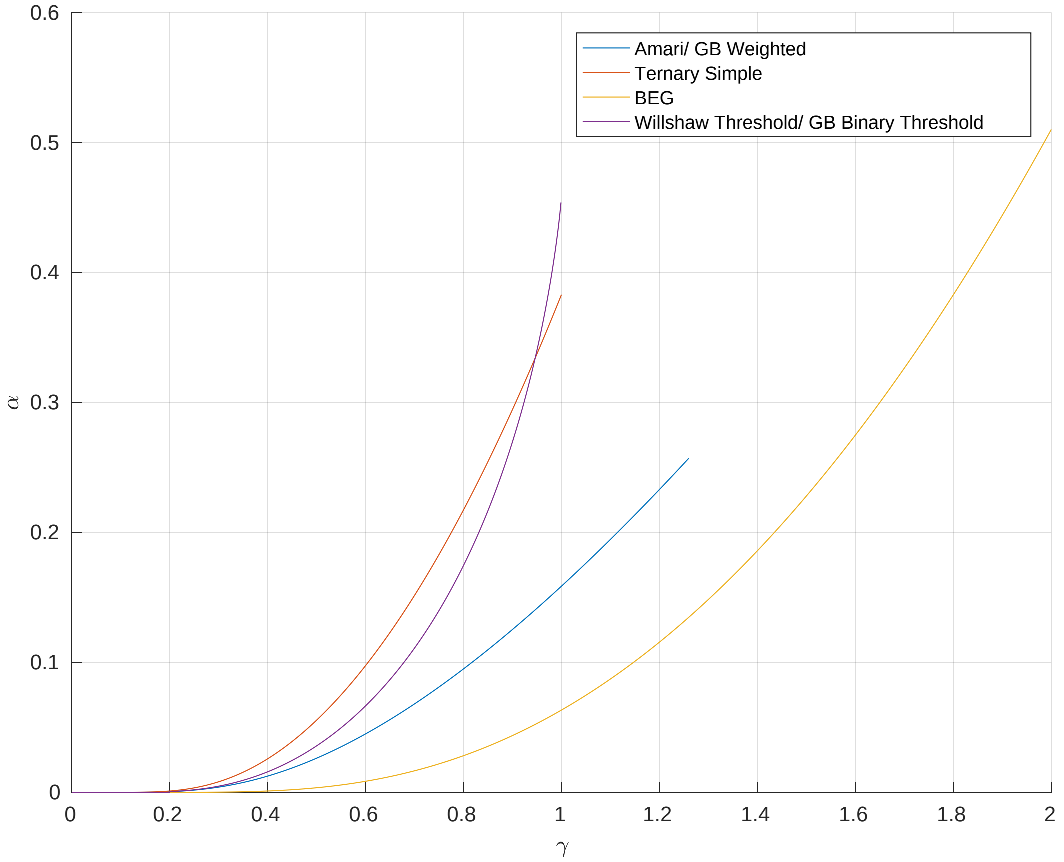

Figure 7.1: Critical capacity variables $\alpha^*(\gamma)$ of the different models in dependence on the threshold variables γ. In the Willshaw model with WTA dynamics, the critical capacity variable is $\alpha^*_{\text{Wi WTA}} \approx 0.45$. In the GB model with SUM-of-MAX rule and in the GB model with binary synaptic weights, self-loops, threshold dynamics and $\gamma = 1$, one can store all patterns such that they are stable.

the critical capacity variable, if $\varrho \log(N)$ active neurons are deleted and the message shall be corrected. We observe that Amari's model and the GB model with weighted synaptic efficacies perform again worst, followed for low error rates ($\varrho \lessapprox 0.05$) by the Ternary simple model and then by the Willshaw model with threshold dynamics (and its equivalent, the GB model with binary synaptic efficacies and threshold dynamics). The Ternary simple model and the Willshaw model with threshold dynamics change places for higher error rates. The Willshaw model with WTA algorithm and the GB model with SUM-of-MAX rule are second best for low error rates ($\varrho \lessapprox 0.07$) and best for higher error rates, offering the same critical value on α. The best values for α for low error rates and second best values for high error rates in dependence on ϱ are attained by the BEG model.

If a stored pattern is corrupted such that there are $\varrho \log(N)$ spuriously active neurons or even by a combination of erased and spuriously active neurons, all models are able to correct this pattern, if α and possibly γ are small enough. Here it is remarkable that the GB model with SUM-of-MAX rule and self-loops is very robust concerning errors of

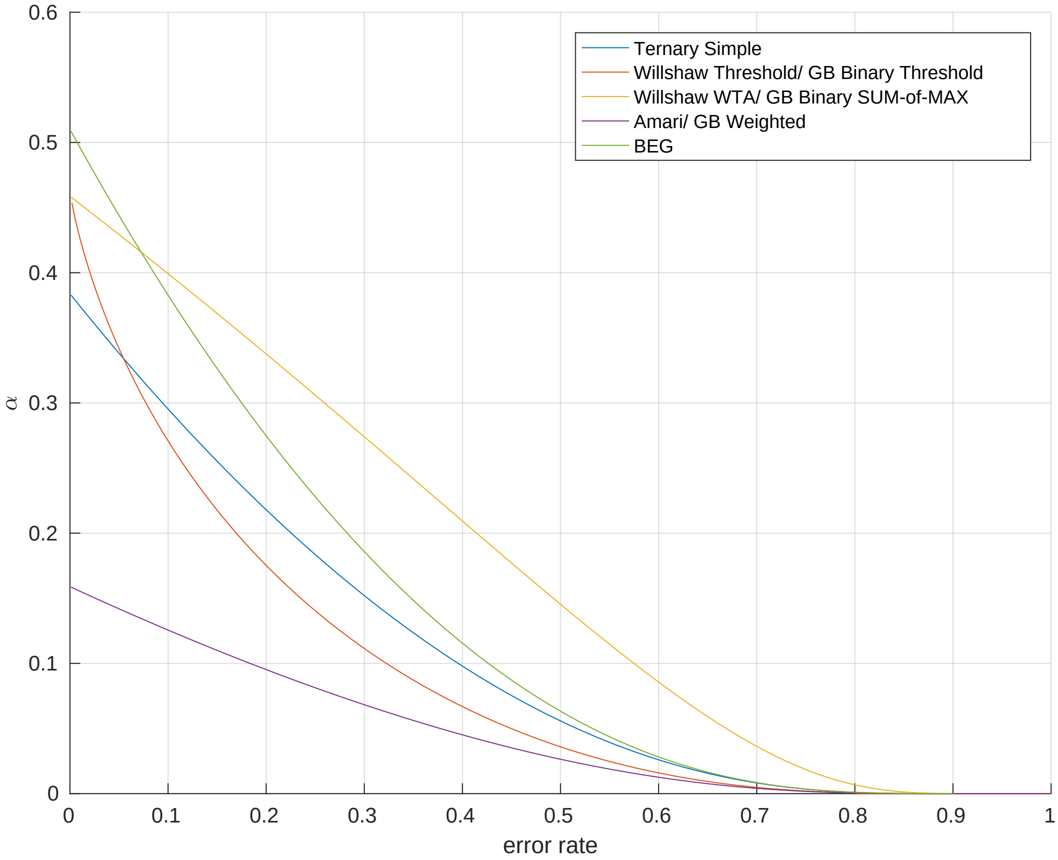

Figure 7.2: Capacity variables of the different models in dependence on ϱ, if a partially deleted stored pattern is to be corrected and $\varrho \log(N)$ of the active neurons of a stored pattern are deleted.

this kind. In all of the other models, a number of $\varrho \log(N)$ spuriously activated neurons considerably decreases the bound on α, mostly antiproportionally with $1 + \varrho$. This shows that the cluster structure improves the performance of the model and that the GB model with SUM-of-MAX rule has a further advantage to the Willshaw model besides the fact that stored patterns are automatically stable.

Comparing the results of the different models, we have to remark that the good performing WTA algorithm ignores the associative memory rule that the updating decision of a neuron should be made independently of the decisions concerning the other neurons. Therefore the comparison of the performances of these models to those using a threshold dynamics has to be taken with care. Anyway, - besides the GB model - the best performing network, the BEG model, uses a threshold algorithm. But this model has a ternary state space, what makes it difficult to compare it to the binary networks.

All variations of the models in chapters 2 - 5 where the stored patterns are uniformly chosen from the set of patterns with exactly $c \approx \log(N)$ active neurons perform clearly better than the ones with i.i.d. spins if it comes to stability of all patterns and exactly or at least as good concerning the stability of an arbitrary stored pattern.

Bibliography

[1] Behrooz Kamary Aliabadi, Claude Berrou, Vincent Gripon, and Xiaoran Jiang. Storing sparse messages in networks of neural cliques. *IEEE Transactions on neural networks and learning systems*, 25(5):980–989, 2014.

[2] Shun-Ichi Amari. Characteristics of sparsely encoded associative memory. *Neural Networks*, 2(6):451–457, 1989.

[3] Daniel J. Amit, Hanoch Gutfreund, and H. Sompolinsky. Spin-glass models of neural networks. *Phys. Rev. A*, 32:1007–1018, Aug 1985.

[4] Daniel J. Amit, Hanoch Gutfreund, and H. Sompolinsky. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Phys. Rev. Lett.*, 55:1530–1533, Sep 1985.

[5] A. D. Barbour, Lars Holst, and Svante Janson. *Poisson Approximation (Oxford Studies in Probability)*. Clarendon Press, 1992.

[6] Andrew C. Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941.

[7] M. Blume, V. J. Emery, and Robert B. Griffiths. Ising model for the λ transition and phase separation in he^3-he^4 mixtures. *Phys. Rev. A*, 4:1071–1077, Sep 1971.

[8] Désiré Bollé, Isaac P. Castillo, and G.M. Shim. Optimal capacity of the blume-emery-griffiths perceptron. *Physical Review E*, 67(3):036113, 2003.

[9] Désiré Bollé and Toni Verbeiren. Thermodynamics of fully connected blume–emery–griffiths neural networks. *Journal of Physics A: Mathematical and General*, 36(2):295, 2003.

[10] Michael V. Boutsikas and Markos V. Koutras. A bound for the distribution of the sum of discrete associated or negatively associated random variables. *Ann. Appl. Probab.*, 10(4):1137–1150, 2000.

[11] Anton Bovier. Sharp upper bounds on perfect retrieval in the hopfield model. *Journal of Applied Probability*, 36(3):941–950, 1999.

[12] David Burshtein. Nondirect convergence radius and number of iterations of the hopfield associative memory. *IEEE Trans. Inf. Theor.*, 40(3):838–847, September 2006.

[13] Harald Cramér. Sur un nouveau théoreme-limite de la théorie des probabilités. *Actualités scientifiques et industrielles*, 736(5-23):115, 1938.

[14] J. D. Esary, Frank Proschan, and D. W. Walkup. Association of random variables, with applications. *Ann. Math. Statist.*, 38:1466–1474, 1967.

[15] Carl-Gustaf Esseen. *On the Liapounoff limit of error in the theory of probability.* Almqvist & Wiksell, 1942.

[16] Vincent Gripon and Claude Berrou. A simple and efficient way to store many messages using neural cliques. In *CCMB 2011: IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain*, 2011.

[17] Vincent Gripon and Claude Berrou. Sparse neural networks with large learning diversity. *IEEE transactions on neural networks*, 22(7):1087–1096, Jul 2011.

[18] Vincent Gripon, Judith Heusel, Matthias Löwe, and Franck Vermet. A comparative study of sparse associative memories. *Journal of Statistical Physics*, 164(1):105–129, 2016.

[19] Donald O. Hebb. The organization of behavior, 1949.

[20] Judith Heusel, Matthias Löwe, and Franck Vermet. On the capacity of an associative memory model based on neural cliques. *Statist. Probab. Lett.*, 106:256–261, 2015.

[21] John J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[22] Hooman Jarollahi, Naoya Onizawa, Vincent Gripon, and Warren J Gross. Architecture and implementation of an associative memory using sparse clustered networks. In *2012 IEEE International Symposium on Circuits and Systems*, pages 2901–2904. IEEE, 2012.

[23] Hooman Jarollahi, Naoya Onizawa, Vincent Gripon, and Warren J Gross. Reduced-complexity binary-weight-coded associative memories. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2523–2527. IEEE, 2013.

[24] Xiaoran Jiang, Vincent Gripon, Claude Berrou, and Michael G. Rabbat. Storing sequences in binary tournament-based neural networks. *IEEE transactions on neural networks and learning systems*, 27(5):913–925, 2016.

[25] János Komlós and Ramamohan Paturi. Effect of connectivity in an associative memory model. *Journal of Computer and System Sciences*, 47(2):350 – 373, 1993.

[26] Frank Proschan Kumar Joag-Dev. Negative association of random variables with applications. *The Annals of Statistics*, 11(1):286–295, 1983.

[27] Lucien Le Cam. An approximation theorem for the poisson binomial distribution. *Pacific J. Math.*, 10(4):1181–1197, 1960.

[28] Daria Loukianova. Memory capacity in the hopfield model. *Comptes rendus de l' académie des sciences série I-Mathématique*, 318(2):157–160, 1994.

[29] Daria Loukianova. Lower bounds on the restitution error in the hopfield model. *Probability Theory and Related Fields*, 107(2):161–176, 1997.

[30] Matthias Löwe. On the storage capacity of hopfield models with correlated patterns. *The Annals of Applied Probability*, 8(4):1216–1250, 1998.

[31] Matthias Löwe and Franck Vermet. The storage capacity of the Blume-Emery-Griffiths neural network. *J. Phys. A*, 38(16):3483–3503, 2005.

[32] Matthias Löwe and Franck Vermet. The capacity of q-state potts neural networks with parallel retrieval dynamics. *Statistics & probability letters*, 77(14):1505–1514, 2007.

[33] Matthias Löwe and Franck Vermet. Capacity of an associative memory model on random graph architectures. *Bernoulli*, 21(3):1884–1910, 08 2015.

[34] Robert J. McEliece, Edward C. Posner, Eugene R. Rodemich, and Santosh S. Venkatesh. The capacity of the hopfield associative memory. *IEEE Trans. Inf. Theor.*, 33(4):461–482, July 1987.

[35] Charles M. Newman. Memory capacity in neural network models: Rigorous lower bounds. *Neural Networks*, 1(3):223–238, 1988.

[36] Masato Okada. Notions of associative memory and sparse coding. *Neural Networks*, 9(8):1429–1458, 1996.

[37] Günther Palm. On associative memory. *Biological Cybernetics*, 36(1):19–31, 1980.

[38] Günther Palm and Friedrich T. Sommer. Information capacity in recurrent mcculloch–pitts networks with sparsely coded memory states. *Network: Computation in Neural Systems*, 3(2):177–186, 1992.

[39] Leonid A. Pastur and Alexander L. Figotin. Exactly soluble model of a spin glass. *Sov. J. Low Temp. Phys*, 3(6):378–383, 1977.

[40] Yu V. Prohorov. Asymptotic behavior of the binomial distribution. *Selected Translations in Mathematical Statistics and Probability*, 1:87–95, 1953.

[41] Friedrich T. Sommer and Günther Palm. Improved bidirectional retrieval of sparse patterns stored by hebbian learning. *Neural Netw.*, 12(2):281–297, March 1999.

[42] Michel Talagrand. Résultats rigoureux pour le modele de hopfield. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique*, 321(1):109–112, 1995.

[43] Michel Talagrand. Rigorous results for the hopfield model with many patterns. *Probability theory and related fields*, 110(2):177–275, 1998.

[44] David J. Willshaw, Peter O. Buneman, and Hugh C. Longuet-Higgins. Non-Holographic Associative Memory. *Nature*, 222:960–962, June 1969.

[45] Zhe Yao, Vincent Gripon, and Michael G. Rabbat. A massively parallel associative memory based on sparse neural networks. *arXiv preprint arXiv:1303.7032*, 2013.

Tabellarischer Lebenslauf

Name:	Judith Heusel
Geboren:	4. Juli 1986 in Mainz
Eltern:	Wolfgang Heusel Lydia Heusel, geb. Hahn
Familienstand:	ledig
Schulbildung:	1993–1997: Grundschule St. Johann in Konz 1997–1999: Gymnasium Konz 1999–2006: Hindenburg-Gymnasium in Trier
Hochschulreife:	24. März 2006 in Trier
Studium:	2006 – 2012: Mathematik an der WWU Münster
Abschluss:	10. März 2012 als Diplom-Mathematikerin
Promotionsstudiengang:	Mathematik
Tätigkeiten:	April 2009 – September 2009, Oktober 2010 – März 2011: Studentische Hilfskraft, WWU Münster Oktober 2012 – April 2017: Wissenschaftliche Mitarbeiterin, WWU Münster
Beginn der Dissertation:	Oktober 2012 bei Prof. Dr. Matthias Löwe, WWU Münster