The storage capacity of the Blume–Emery–Griffiths neural network

2005 J. Phys. A: Math. Gen. 38 3483

(http://iopscience.iop.org/0305-4470/38/16/002)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **38** (2005) 3483–3503 doi:10.1088/0305-4470/38/16/002

The storage capacity of the Blume–Emery–Griffiths neural network

Matthias Löwe[1] **and Franck Vermet**[2]

[1] Institut für Mathematische Statistik, Fachbereich Mathematik und Informatik, Universität Münster, Einsteinstr. 62, 48149 Münster, Germany

[2] Département de Mathématiques, Université de Bretagne Occidentale, 6, avenue Victor Le Gorgeu CS 93837 F-29238 Brest Cedex 3, France

E-mail: maloewe@math.uni-muenster.de and Franck.Vermet@univ-brest.fr

Received 15 November 2004, in final form 15 February 2005
Published 6 April 2005
Online at stacks.iop.org/JPhysA/38/3483

Abstract

We analyse the so-called Blume–Emery–Griffiths (BEG) neural network at zero temperature. An upper bound on its storage capacity is given if we want the stored patterns to be fixed points of the retrieval dynamics. Besides, we discuss a more liberal notion of storage capacity introduced by Newman (1988 *Neural Netw.* **1** 223–38) in the context of the Hopfield model (Hopfield 1982 *Proc. Natl Acad. Sci. USA* **79** 2554–8). We show that, similar to the findings in the neural networks literature, the BEG model with this notion of storage capacity can store a number of patterns proportional to the number of neurons in the model.

PACS numbers: 05.50.+q, 05.40.−a, 87.18.Sn
Mathematics Subject Classification: 82c32, 82B44, 60k35

1. Introduction

The storage capacity of binary neural networks, especially the Hopfield model, has been intensively studied in the recent probability, physics and neural networks literature (see, e.g., [N88, L94, L97, T95, T98, Lö98, Lö99, B99]). Lately one has also studied a particular non-binary case, where the patterns to be stored consist of an information whether the corresponding neuron has 'seen' a +1 or a −1 (for example whether it has seen a white or a black pixel) or whether the neuron has been inactive (0). As a result, the state space of the spins of the neural network changes to $\{0, -1, +1\}$, as opposed to $\{-1, +1\}$ in the case of the Hopfield model. The special role taken by 0 (namely to encode the inactivity of a neuron) makes the network different from the usual three-state Hopfield model (also called the Potts–Hopfield model, where 0 is just one of the possible 'colours' seen by a neuron) studied for instance in [FMP92] or [M96].

While the Hopfield model originally was introduced by Pastur and Figotin [FP77] as a simplified model of a spin glass and later on reinterpreted as a neural network by Hopfield [Ho82] (indeed it was Hopfield's work that attracted the interest of many researchers to that area), first models for the question raised above were studied by Baram and Sal'ee [BS92].

The basic idea behind all these models is to choose the information to be stored—which usually is referred to as patterns—as the local minima of an appropriate Hamiltonian (this will be further developed in the next section) on the set $\{0, \pm 1\}^N$, where the set $\{1, \ldots, N\}$ constitutes the set of neurons. Hence the retrieval dynamics (usually a Monte Carlo dynamics, possibly at zero temperature) will eventually converge to the stored patterns. Since, according to the other rules of neural networks, this is possible only up to a certain accuracy depending on the number of patterns to be stored, one can introduce the notion of a storage capacity of the model as the critical number of patterns (depending on N, of course) up to which the model can successfully reconstruct these patterns.

In this paper, we consider two distinct definitions of storage capacity at zero temperature. The first one (as shown by McEliece *et al* [MPRV87], also see [Bu94, V94, P96] for nice proofs for this result, [B99] for the corresponding upper bound and [Lö98, Lö99] for the case of biased or dependent patterns) leads to a storage capacity of at least $\frac{N}{\gamma \log N}$ for the Hopfield model with N neurons and randomly and independently chosen unbiased patterns. Secondly, the essential progress of Newman's work [N88] was that he was able to give a proof for the result of Amit, Gutfreund and Sompolinsky [AGS87] showing that the Hopfield model enables storage of even αN patterns (for an $\alpha > 0$ small enough), if small errors are tolerated. The value for α obtained by Newman was $\alpha = 0.056$. This has been improved by Loukianova [L94, L97] to $\alpha = 0.071$ by a refined large deviation analysis and by Talagrand [T95, T98] to $\alpha = 0.08$ by further improving on Loukianova's idea. Nevertheless, it remains an open question, whether the prediction of $\alpha = 0.14$ in [AGS87] on the basis of the replica method and computer simulations can be mathematically justified.

It has been argued that for character recognition, sparse patterns, that is patterns with an intensity of less than 100% for being ± 1, are more realistic than equiprobable binary ones. In [BS92], a first model for such ternary networks (networks with input patterns from $\{-1, 0, +1\}^N$) was introduced and studied. The authors came to the conclusion that their model of a neural network can store about $\frac{N}{\gamma p \log N}$ patterns, where p is the activity of the patterns, i.e. the probability that a fixed spin of a fixed pattern is different from zero. Here the first of the two notions of storage capacity described above is used. In particular, the authors conclude that the storage capacity of the networks *increases* when the activity is low.

However, other authors have proposed that an optimal Hamiltonian, guaranteeing the best retrieval properties for neural networks with multi-state neurons, can be achieved by maximizing the mutual information content of the networks (see, for instance, [DK00, BoVe02]). In this way, for two-state neurons the Hopfield model is retrieved while for the three-state problem described above, one obtains a spin glass version of the three-state Ising model introduced by Blume, Emery and Griffiths to model and study He^3–He^4 mixtures. We will refer to this model as the Blume–Emery–Griffiths (BEG) neural network (or spin glass). In [BCS03, BV03] the non-rigorous replica method is used to study the storage capacity of such networks. In this paper, the authors conclude that a number of αN patterns can be stored (where $\alpha > 0$ has to be chosen appropriately).

The goal of the present paper is to analyse rigorously the storage capacity of the BEG neural network (using both notions of storage capacity discussed above) and to compare it to the capacities found by Balam and Sal'ee in [BS92]. With this setup the results obtained resemble the results in the ordinary Hopfield model: the BEG neural network model is shown to have the capacity to store a number of patterns in the same order of magnitude as the number of patterns that can be stored in the Hopfield model. The bound on the storage capacity for the BEG network is found to be below the bound on the storage capacity of the network studied in [BS92], though similar techniques are employed in both cases. In the BEG model the capacity decreases when the activity becomes small.

This paper is organized in the following way: section 2 contains the basic setup, especially the type of patterns we consider, the definition of the models we have in mind, as well as the definition of storage capacity we use. Section 3 details our results concerning the storage capacity of the models introduced in section 2. Both the dynamic and the static notion of storage are analysed. The proofs are given lastly in section 4.

2. The basic setup

This section describes the two models we will be interested in as well as the two notions of storage capacity we use.

For the rest of this paper a network will always operate on N neurons, labelled by $1, \dots, N$. In this network one wants to store $M = M(N)$ patterns $\left(\xi_i^\mu\right)_{i=1,\dots,N}^{\mu=1,\dots,M}$, which for the sake of this paper will always be thought of as random elements on the space $\{-1, 0, +1\}^N$. The underlying probability distribution of these patterns will be such that makes all the random variables ξ_i^μ independent with

$$1 - p = \mathbb{P}\left(\xi_i^\mu = 0\right) \qquad \text{and} \qquad \mathbb{P}\left(\xi_i^\mu = 1\right) = \mathbb{P}\left(\xi_i^\mu = -1\right) = \frac{p}{2} \tag{2.1}$$

for some fixed $p \in (0, 1)$.

Correspondingly, for each $i = 1, \dots, N$ there is an associated spin σ_i that also takes its values in the space $\{-1, 0, +1\}$. A network will now be defined by a dynamics on the spin space $\{-1, 0, +1\}^N$, or equivalently (if possible), by a Hamiltonian, that is minimized by the dynamics.

Two different sorts of networks will be distinguished for our case of a ternary neuron.

2.1. The simple threshold network

The idea of the simple threshold network is to adapt the gradient dynamics from the Hopfield model of binary neurons, hence to switch a spin σ_i into the direction of his external field

$$h_i(\sigma) = \sum_{j \neq i} \sigma_j J_{ij} = \frac{1}{p^2 N} \sum_{j \neq i} \sum_{\mu=1}^{M} \sigma_j \xi_i^\mu \xi_j^\mu . \tag{2.2}$$

Nevertheless, to cope with the fact that now we have ternary neurons, we only switch σ_i into the direction of $h_i(\sigma)$ if this external field is strong enough. Otherwise we switch σ_i to 0. Formally the rule is the following: update the spins σ_i asynchronously, i.e. one after another, in a random fashion (each time the model is updated, an independent random variable from $\{1, \dots, N\}$ is drawn and we update the corresponding spin). Choose a (fixed) threshold $W > 0$. The updating rule for σ_i is then given by:

$$T_i(\sigma) = \begin{cases} 1 & \text{if} \quad Np^2 h_i(\sigma) > W \\ 0 & \text{if} \quad -W \leqslant Np^2 h_i(\sigma) \leqslant W \\ -1 & \text{if} \quad Np^2 h_i(\sigma) < -W \end{cases} \tag{2.3}$$

where for convenience we have multiplied the external field by the factor Np^2.

2.2. The Blume–Emery–Griffiths network

The BEG network was recently advertised to be the most effective network for storage of ternary data. In principle, it works like the simple threshold network described above with the difference that the threshold is computed from the patterns. It is given by the following updating rule: update the spins σ_i asynchronously at random. The updating rule for σ_i is then given by

$$T_i(\sigma) = \operatorname{sgn}(h_i(\sigma))\Theta(|h_i(\sigma)| + \theta_i(\sigma)). \tag{2.4}$$

Here Θ is the Heaviside function (which is 1 for positive x, and 0 otherwise),

$$\theta_i(\sigma) = \sum_{j\neq i} K_{ij}\sigma_j^2,$$

where

$$K_{ij} = \frac{1}{p^2(1-p)^2N}\sum_{\mu=1}^{M}\eta_i^\mu\eta_j^\mu, \qquad \text{and} \qquad \eta_i^\mu = \left(\xi_i^\mu\right)^2 - p,$$

and h_i is the external field defined in (2.2).

The dynamics described above minimizes the Hamiltonian function of the BEG network:

$$H_N(\sigma) = -\frac{1}{2p^2N}\sum_{i,j=1;i\neq j}^{N}\sum_{\mu=1}^{M(N)}\sigma_i\sigma_j\xi_i^\mu\xi_j^\mu - \frac{1}{2p^2(1-p)^2N}\sum_{i,j=1;,i\neq j}^{N}\sum_{\mu=1}^{M}\sigma_i^2\sigma_j^2\eta_i^\mu\eta_j^\mu. \tag{2.5}$$

Indeed, let $\sigma' = (\sigma_1,\ldots,\sigma_{i-1},T_i(\sigma),\sigma_{i+1},\ldots,\sigma_N)$. Since

$$H_N(\sigma) = -\frac{1}{2}\sum_{i,j=1;i\neq j}^{N} J_{ij}\sigma_i\sigma_j - \frac{1}{2}\sum_{i,j=1;,i\neq j}^{N} K_{ij}\sigma_i^2\sigma_j^2,$$

we have

$$\begin{aligned} H_N(\sigma') - H_N(\sigma) &= (\sigma_i - T_i(\sigma))\sum_{j;j\neq i} J_{ij}\sigma_j + \left(\sigma_i^2 - T_i(\sigma)^2\right)\sum_{j;j\neq i} K_{ij}\sigma_j^2 \\ &= (\sigma_i - T_i(\sigma))(h_i + (\sigma_i + T_i(\sigma))\theta_i(\sigma)). \end{aligned}$$

And using (2.4) it is easy to check that for all values of σ_i and $T_i(\sigma)$, the quantity $H_N(\sigma') - H_N(\sigma)$ is non-positive: the dynamics leads to a local minimum of H_N.

Remark 2.1. Dynamics (2.4) can be viewed as the limit as $\beta \to \infty$, i.e. the zero-temperature limit, of the following stochastic updating rule: for all $s \in \{-1, 0, +1\}$,

$$\begin{aligned} P[T_i(\sigma) = s] &= \frac{\exp(-\beta H_N(\sigma_1,\ldots,\sigma_{i-1},s,\sigma_{i+1},\ldots,\sigma_N))}{\sum_{t\in\{-1,0,1\}}\exp(-\beta H_N(\sigma_1,\ldots,\sigma_{i-1},t,\sigma_{i+1},\ldots,\sigma_N))} \\ &= \frac{\exp(\beta(h_i s + \theta_i s^2))}{\sum_{t\in\{-1,0,1\}}\exp(\beta(h_i t + \theta_i t^2))}. \end{aligned}$$

We introduce the following two definitions of storage capacity.

The dynamics $T := (T_i)_{i=1,\ldots,N}$ is considered to be the retrieval dynamics of the corresponding neural network, that is to say, given an input σ the network will 'recognize' this input as that pattern, that is found by (possibly many iterates of) T.

The least one could expect from such a dynamics is that the patterns themselves are stable under T, in such a way that, if we input one of the ξ^μ we also find ξ^μ, which implies that ξ^μ is a local minimum of the corresponding Hamiltonian—if there is one. The storage capacity in this concept is defined as the greatest number of patterns $M := M(N)$ such that all the patterns ξ^ν are stable in the above sense. Of course, this number depends on the randomly chosen patterns, so that in the following we will always speak of numbers $M(N)$ such that with probability converging to one (in some sense) a pattern is (or all the patterns are) stable. Let us remark that for the Hopfield model with one pattern ($M = 1$), the configurations ξ^1 and $-\xi^1$ are trivially the only global minima of the Hamiltonian for all values of N, but this is no longer the case for the BEG model: ξ^1 is not even necessarily a local minimum of H_N for a fixed value of N. We only expect probabilistic and asymptotic results.

The other approach to storage capacity is due to Amit *et al* [AGS87] and Newman [N88]. It takes into account small errors we are willing to accept in the restoration of the patterns (with the idea to increase the storage capacity). So we are satisfied if the retrieval dynamics converges to a configuration which is not too far away from the original patterns. Thus, for the BEG model introduced above, in this concept, a pattern ξ^ν is called stable if it is close to a local minimum of the Hamiltonian, or in other words if it is surrounded by a sufficiently high energy barrier. Technically speaking, we will call ξ^ν stable if there exists $\varepsilon > 0$ and $\delta > 0$ such that

$$\inf_{\sigma \in S_\delta(\xi^\nu)} H_N(\sigma) \geqslant H_N(\xi^\nu) + \varepsilon N. \tag{2.6}$$

Here the set $S_\delta(\xi^\nu)$, where the infimum is taken over, is the Hamming sphere of radius δN centred in ξ^ν. Again we use the notion of storage capacity for the maximal number $M(N)$ of patterns such that (2.6) holds true for all ξ^ν almost surely.

3. Results

In this section we give a lower bound on the number of patterns that can be stored in the simple threshold model and the BEG model introduced above. For the latter we analyse both notions of storage capacity.

First of all, let us consider the storage capacity of the simple threshold network introduced in section 2:

Theorem 3.1. *Assume the patterns ξ^μ are chosen at random as described above and that their number satisfies $M(N) = \frac{N}{\gamma \log N}$. Moreover let us choose the threshold as*

$$W = \frac{N(p-\varepsilon)}{2} \qquad \text{for some small} \quad \varepsilon > 0 \tag{3.1}$$

(and this choice is optimal).

Then for the simple threshold model the following assertions hold true:

(1) If $\gamma > 24p$

$$\mathbb{P}\left(\liminf_{N\to\infty}\left(\bigcap_{\mu=1}^{M(N)} T\xi^\mu = \xi^\mu\right)\right) = 1$$

i.e. the patterns are almost surely stable.

(2) If $\gamma > 16p$

$$\mathbb{P}\left(\bigcap_{\mu=1}^{M(N)} T\xi^\mu = \xi^\mu\right) = 1 - R_N$$

with $\lim_{N\to\infty} R_N = 0$.

(3) If $\gamma > 8p$ for every fixed $\mu = 1, \ldots, M$

$$\mathbb{P}(T\xi^\mu = \xi^\mu) = 1 - R_N$$

with $\lim_{N\to\infty} R_N = 0$.

Here T is the family of mappings (T_i) introduced in (2.3).

Remarks 3.2.

- Part (3) of theorem 3.1 has already been obtained in [BS92].
- Observe that the above bounds depend on p in such a way that the storage capacity increases, if p becomes small. This is after some considerations not too surprising, since in the limit $p \to 0$ (e.g. if $p \ll 1/N$) all the patterns are equal to the pattern with spins equal to 0, which is then easy to store.

We find that the situation in BEG model is quite similar. Here our result is the following.

Theorem 3.3. *Assume the patterns ξ^μ are chosen at random as described above and that their number satisfies $M(N) = \frac{N}{\gamma \log N}$.*

Define

$$\gamma^*(p) := \max\left\{\frac{2((1-p)^2+1)}{p}, \frac{p((1-p)^2+1)}{2(1-p)^2}\right\}.$$

Then for the BEG model the following assertions hold true:

(1) If $\gamma > 3\gamma^*(p)$

$$\mathbb{P}\left(\liminf_{N\to\infty}\left(\bigcap_{\mu=1}^{M(N)} T\xi^\mu = \xi^\mu\right)\right) = 1$$

i.e. the patterns are almost surely stable.

(2) If $\gamma > 2\gamma^*(p)$

$$\mathbb{P}\left(\bigcap_{\mu=1}^{M(N)} T\xi^\mu = \xi^\mu\right) = 1 - R_N$$

with $\lim_{N\to\infty} R_N = 0$.

(3) If $\gamma > \gamma^*(p)$ for every fixed $\mu = 1, \ldots, M$

$$\mathbb{P}(T\xi^\mu = \xi^\mu) = 1 - R_N$$

with $\lim_{N\to\infty} R_N = 0$.

Here T is the family of mappings (T_i) introduced in (2.4).

Remarks 3.4.

(1) Observe that other than in the simple threshold model discussed in theorem 3.1 the bound on $\gamma^*(p)$ in theorem 3.3 is increasing for $p \to 0$, implying that our bounds on the storage capacity of the BEG model become *smaller* for small p. Even if we do not have a proof of this statement (because we are not able to show the corresponding upper bounds for the storage capacity), it seems that the simple threshold model outperforms the BEG model.

(2) It might be expected that both the simple threshold model and the BEG model coincide with the Hopfield model in the limit $p \to 1$. This is however not the case. In the simple threshold model the optimal threshold (3.1) is *increasing* for $p \to 1$ in order to allow for restitution of those spins which are zero. This explains that for $p = 1$ the γ's in theorem 3.1 are larger than the corresponding γ's in the Hopfield model (implying a smaller storage capacity of the simple threshold model). The situation in the BEG model is even worse. Even though one could expect the θ_i's to converge to (the positive number) M/p^2, the variance coming from this term spoils the bounds on the storage capacity in such a way that γ^* converges to ∞ in the $p \to 1$ limit. One remarks that the divergence of the variance of $\theta_i(\xi^1)$ for $p \to 1$ is due to the prefactor $\frac{1}{(1-p)^2}$. This might raise the question of whether the performance of the BEG model can be improved by changing this factor. This in indeed the case: redoing the calculations for the proof of theorem 3.3 with $\frac{1}{(1-p)}$ instead of $\frac{1}{(1-p)^2}$, one obtains the bound

$$\gamma^*(p) = \max\left\{\frac{2}{p}, \frac{2p}{(2-p)^2}\right\}.$$

Note that this choice of $\gamma^*(p)$ is not only finite for $p = 1$ but even agrees with the optimal $\gamma^*(p) = 2$ in the Hopfield model (see [B99] for the optimality). This change in performance does not cause much surprise. Computing the variance of $\eta_i^\mu \eta_j^\mu$ we find $p^2(1-p)^2$. Therefore the term K_{ij} is not of order one, but rather of order $1/p(1-p)$ (ignoring the N dependence) in the original BEG network.

Recall that in [DK00] and [Bove02] a certain optimality of this model is shown, the value of the factors being obtained by optimizing the mutual information of the system. However, in [BV03], authors develop numerical estimations which suggest also that the prefactor $1/(p^2(1-p)^2)$ does not give the largest capacity.

It is also of interest to prove that not only the patterns ξ^μ are stable for T but also the dynamics can correct corrupted patterns: if x is not too different from ξ^1, the network may retrieve ξ^1. This is the principle of autoassociative memory. Using the same method as in the proof of the preceding theorems (to be given in section 4) we prove the following result:

Theorem 3.5. *Assume the patterns ξ^μ are chosen at random as described above and that $M = \frac{N}{\gamma \log N}$. Let $x \in \{-1, 0, 1\}^N$ a configuration and*

$$\delta_1 := \frac{1}{N}\operatorname{card}\{i : |\xi_i^1| = 1, x_i = 0\}, \qquad \delta_2 := \frac{1}{N}\operatorname{card}\{i : \xi_i^1 x_i = -1\}$$

and

$$\delta_3 := \frac{1}{N}\operatorname{card}\{i : \xi_i^1 = 0, |x_i| = 1\}.$$

Then

(1) For the threshold model with $W = \frac{N(p-\varepsilon)}{2}$ (and any $\varepsilon > 0$),

$$\mathbb{P}[T(x) \neq \xi^1] \to 0$$

as N goes to infinity under the conditions:

$$\delta_1 + 2\delta_2 < \frac{p}{2}, \quad \delta_3 < 1 - p, \qquad \gamma > \max\left\{\frac{2(p - \delta_1 + \delta_3)p^2}{(p/2 - \delta_1 - 2\delta_2)^2}, 8(p - \delta_1 + \delta_3)\right\}.$$

(2) For the BEG model,

$$\mathbb{P}[T(x) \neq \xi^1] \to 0$$

as N goes to infinity under the conditions:

$$\delta_1 + 2\delta_2 < p, \quad \delta_3 < 1 - p, \quad 2(p - \delta_1 - \delta_2) > \frac{p}{1-p}\delta_3, \quad p - \delta_1 - \frac{p}{1-p}\delta_3 > 0,$$

$$\gamma > \max\left\{\frac{2(p-\delta_1+\delta_3)(1+(1-p)^2)}{\left(p-\delta_1-\frac{p}{1-p}\delta_3\right)^2}, \frac{2p^2(p-\delta_1+\delta_3)(1+(1-p)^2)}{(1-p)^2\left(2(p-\delta_1-\delta_2)-\frac{p}{1-p}\delta_3\right)^2}, \frac{2p^2(p-\delta_1+\delta_3)}{(p-\delta_1-2\delta_2)^2}\right\}.$$

Remark 3.6. When all the δ_i go 0, we retrieve the conditions of the preceding theorems. But the dependence on the δ_i is complicated in the general case. However in the simple case $\delta_1 = \delta_3 = 0$ and $\delta_2 \neq 0$, we obtain for the BEG model a result that is reminiscent of a result known in the Hopfield model (e.g. [P95]): if $\delta_2 < p/2$, the correction is possible in one step and the capacity is inferior to a constant (depending on *p*) proportional to $(p - 2\delta_2)^2$.

Finally, we prove that the BEG model is able to store 'extensively many' patterns (i.e. *M(N)* grows like αN) provided that Newman's concept [N88] of storage is used.

Theorem 3.7. *Assume the patterns ξ^μ are chosen at random as described above. Then for the Hamiltonian of the BEG model defined in (2.5) the following holds true: if $M(N)$ scales like αN there exists an energy barrier around each of the patterns. I.e. there exists an $\alpha_c > 0$ (depending on p, but not on N) such that if $M(N) \leqslant \alpha_c N$, then there are $\varepsilon > 0$ and $0 < \delta < 1/2$ such that for the Hamiltonian of the BEG model (2.5) it holds*

$$\mathbb{P}\left(\liminf_{N\to\infty}\left(\bigcap_{\mu=1}^{M(N)}\bigcap_{\sigma\in S_\delta(\xi^\mu)}\{H_N(\sigma) \geqslant H_N(\xi^\mu\} + \varepsilon N)\right)\right) = 1$$

where $S_\delta(\xi^\mu)$ is the Hamming sphere of radius δN centred in ξ^μ.

Remark 3.8. Theorem 3.7 is a rigorous result in agreement with the findings in [BCS03] and [BV03] that the BEG model can store up to αN patterns. However, all these complementary results are difficult to compare. In [BCS03], the authors consider the BEG perceptron with general couplings, a model with the same dynamics as the model we study. Studying the basin of attraction of the patterns for the dynamics by a Gardner-type analysis [Ga88, ST03], they get that there exist interactions such that the model can store up to αN patterns. And in [BV03], following [AGS87], authors apply replica mean-field theory to calculate the free energy of the BEG model.

4. Proofs

The proofs of the theorems stated in section 3 are detailed and explained below. The techniques employed are adaptations of the usual large deviations analysis applied in [V94, P95, BS92, N88, L94], or [T98] in the case of the Hopfield model.

Proof of theorem 3.1. Without loss of generality it suffices to analyse stability of the first pattern ξ^1. Observe that—according to the definition of the dynamics *T* of the simple threshold model—the pattern ξ^1 is stable if and only if

$$T_i(\xi^1) = \xi_i^1 \qquad \text{for all} \quad i = 1, \ldots, N.$$

Multiplying all the ξ_i^μ by ξ_i^1 (a very simple 'gauge transformation'), we can assume that the ξ_i^1's are either 0 or 1. We distinguish these two cases.

Firstly, if $\xi_i^1 = 0, \xi_i^1$ is unstable if and only if $Np^2|h_i(\xi^1)| > W$. We estimate the probability of this event by the exponential Markov–Chebyshev inequality ([Gr92], theorem p 285, with $h(x) = \mathrm{e}^{tx}$). For all $t \geqslant 0$

$$\begin{aligned}\mathbb{P}\big(Np^2|h_i(\xi^1)| > W \,\big|\, \xi_i^1 = 0\big) &= 2\mathbb{P}\big(Np^2 h_i(\xi^1) > W \,\big|\, \xi_i^1 = 0\big) \\ &= 2\mathbb{P}\left(\sum_{j\neq i}\sum_{\mu\neq 1}\xi_j^1\xi_i^\mu\xi_j^\mu > W \,\Big|\, \xi_i^1 = 0\right) \\ &\leqslant 2\mathrm{e}^{-tW}\mathbb{E}\left(\exp t\sum_{j\neq i}\sum_{\mu\neq 1}\xi_j^1\xi_i^\mu\xi_j^\mu \,\Big|\, \xi_i^1 = 0\right). \qquad (4.1)\end{aligned}$$

Assume that exactly $0 \leqslant K \leqslant N$ of ξ_j^1 are equal to 1 (and the others are equal to 0) and without loss of generality assume that these are the first K, hence $\xi_1^1 = \cdots = \xi_K^1 = 1$ (which, of course, implies that $i \geqslant K+1$) and call this condition $\mathcal{K}$. Then

$$\begin{aligned}\mathbb{E}\left(\exp t\sum_{j\neq i}\sum_{\mu\neq 1}\xi_j^1\xi_i^\mu\xi_j^\mu \,\Big|\, \xi_i^1 = 0, \mathcal{K}\right) &= \mathbb{E}\left(\exp t\sum_{j=1}^{K}\sum_{\mu\neq 1}\xi_i^\mu\xi_j^\mu\right) \\ &= \prod_{\mu\neq 1}\mathbb{E}_{\xi_i^\mu}\prod_{j=1}^{K}\mathbb{E}_{\xi_j^\mu}\,\mathrm{e}^{t\xi_i^\mu\xi_j^\mu},\end{aligned}$$

where $\mathbb{E}_{\xi_j^\mu}$ denotes the expectation with respect to the random variable ξ_j^μ. Anticipating that $t > 0$ will be chosen small at the end, we expand the right-hand side as

$$\begin{aligned}\prod_{\mu\neq 1}\mathbb{E}_{\xi_i^\mu}\prod_{j=1}^{K}\mathbb{E}_{\xi_j^\mu}\,\mathrm{e}^{t\xi_i^\mu\xi_j^\mu} &\approx \prod_{\mu\neq 1}\mathbb{E}_{\xi_i^\mu}\left(1+\frac{p}{2}t^2\big(\xi_i^\mu\big)^2+\mathcal{O}(t^3)\right)^K \\ &= \prod_{\mu\neq 1}\left[p\left(1+\frac{p}{2}t^2+\mathcal{O}(t^3)\right)^K+(1-p)(1+\mathcal{O}(t^3))\right] \\ &\approx \left(1+\frac{Kp^2t^2}{2}+\mathcal{O}(Kt^3)\right)^{M-1} \\ &\leqslant \exp\left(\frac{K(M-1)p^2t^2}{2}+\mathcal{O}(KMt^3)\right).\end{aligned}$$

Thus

$$\mathbb{P}\big(Np^2|h_i(\xi^1)| > W \,\big|\, \xi_i^1 = 0, \mathcal{K}\big) \leqslant 2\exp(-tW)\exp\left(\frac{K(M-1)p^2t^2}{2}+\mathcal{O}(KMt^3)\right),$$

which for our choice of $t = \frac{W}{K(M-1)p^2}$ gives

$$\mathbb{P}\big(Np^2|h_i(\xi^1)| > W \,\big|\, \xi_i^1 = 0, \mathcal{K}\big) \leqslant 2\exp\left(-\frac{W^2}{2K(M-1)p^2}+\mathcal{O}\left(\frac{W^3}{K^2M^2p^6}\right)\right).$$

Secondly, if $\xi_i^1 = 1, \xi_i^1$ is unstable if and only if $Np^2h_i(\xi^1) \leqslant W$. Again we assume that $\xi_1^1 = \cdots = \xi_K^1 = 1$ and the other ξ_j^1's are 0 (and call this condition $\mathcal{K}$) and estimate the probability of this event by the exponential Markov–Chebyshev inequality: for all $t \geqslant 0$

$$\begin{aligned}
\mathbb{P}\big(Np^2 h_i(\xi^1) \leqslant W \,\big|\, \xi_i^1 = 1, \mathcal{K}\big) &\leqslant \mathrm{e}^{tW}\mathbb{E}\left(\exp -t \sum_{j\neq i}\sum_{\mu=1}^{M} \xi_j^1 \xi_i^\mu \xi_j^\mu \,\big|\, \xi_i^1 = 1, \mathcal{K}\right) \\
&\leqslant \mathrm{e}^{t(W-K)}\mathbb{E}\left(\exp -t \sum_{j=1}^{K}\sum_{\mu\neq 1}^{M} \xi_i^\mu \xi_j^\mu\right) \\
&\leqslant \mathrm{e}^{t(W-K)} \exp\left(\frac{K(M-1)p^2t^2}{2} + \mathcal{O}(KMt^3)\right) \qquad (4.2)
\end{aligned}$$

where we compute the moment generating function as above. Choosing $t = \frac{W-K}{K(M-1)p^2}$ yields

$$\mathbb{P}\big(Np^2 h_i(\xi^1) \leqslant W \,\big|\, \xi_i^1 = 1, \mathcal{K}\big) \leqslant \exp\left(-\frac{(W-K)^2}{2K(M-1)p^2} + \mathcal{O}\left(\frac{(W-K)^3}{K^2M^2p^6}\right)\right).$$

Now observe that with overwhelming probability, $\frac{K}{N} \in [p-\varepsilon, p+\varepsilon]$ for some small $\varepsilon > 0$ if N is large. Even more we have that with overwhelming probability

$$S^\mu := \sum_{j=1}^{N} \big|\xi_j^\mu\big| \in [(p-\varepsilon)N, (p+\varepsilon)N] \qquad (4.3)$$

for some small $\varepsilon > 0$ and all $\mu = 1, \ldots, M(N)$, if N gets large and for instance $M(N) \leqslant N$. Hence, for N large enough, with overwhelming probability in ξ

$$\begin{aligned}
\mathbb{P}(\xi^1 \text{ is not stable}) &= \mathbb{P}\big(\exists_{i=1,\ldots,N} T_i(\xi^1) \neq \xi_i^1\big) \\
&\leqslant N((1-p)+\varepsilon)\mathbb{P}\big(T_1(\xi^1) \neq \xi_1^1 \big| \xi_1^1 = 0\big) + N(p+\varepsilon)\mathbb{P}\big(T_1(\xi^1) \neq \xi_1^1 \big| \xi_1^1 = 1\big) \\
&\leqslant 2N((1-p)+\varepsilon)\exp\left(-\frac{W^2}{2N(p-\varepsilon)(M-1)p^2} + \mathcal{O}\left(\frac{W^3}{N^2(p-\varepsilon)^2M^2p^6}\right)\right) \\
&\quad + N(p+\varepsilon)\exp\left(-\frac{(W-N(p-\varepsilon))^2}{2N(p-\varepsilon)(M-1)p^2} + \mathcal{O}\left(\frac{(W-N(p-\varepsilon))^3}{N^2(p-\varepsilon)^2M^2p^6}\right)\right) \qquad (4.4)
\end{aligned}$$

Since we think of p being independent of N in order to have the summands in (4.4) of the same order, we choose the threshold as $W = N(p-\varepsilon)/2$ and we use the ansatz

$$M = M(N) = \frac{N}{\gamma \log N}$$

for some $\gamma > 0$. Then

$$\frac{W^3}{N^2(p-\varepsilon)^2M^2p^6} \to 0 \qquad \text{as well as} \qquad \frac{(W-N(p-\varepsilon))^3}{N^2(p-\varepsilon)^2M^2p^6} \to 0$$

and we come to

$$\mathbb{P}(\xi^1 \text{ is not stable}) \leqslant 2((1-p)+\varepsilon)N^{-\frac{\gamma(p-\varepsilon)}{8p^2}+1}(1+\varepsilon_N) + (p+\varepsilon)N^{-\frac{\gamma(p-\varepsilon)}{8p^2}+1}(1+\varepsilon_N)$$

for a sequence $\varepsilon_N \to 0$. If now $\gamma > 8p$ then also $\gamma > \frac{8p^2}{p-\varepsilon}$ for $\varepsilon > 0$ small enough, i.e. $-\frac{\gamma(p-\varepsilon)}{8p^2} + 1 < 0$ and therefore

$$\mathbb{P}(\xi^1 \text{ is not stable}) \to 0.$$

If we are now asking for all of the patterns to be stable simultaneously, using the above line of arguments together with (4.3), all this goes to show that

$$\begin{aligned}\mathbb{P}(\exists\mu : \xi^\mu \text{ is not stable}) &\leqslant \sum_{\mu=1}^{M} \mathbb{P}(\xi^\mu \text{ is not stable})\\ &= M\mathbb{P}(\xi^1 \text{ is not stable})\\ &\leqslant CN^{-\frac{\gamma(p-\varepsilon)}{8p^2}+2}(1+\varepsilon_N).\end{aligned}$$

Here C is constant (depending on p and ε). The latter expression goes to zero for $\gamma > \frac{16p^2}{p-\varepsilon}$ which can be achieved, if $\gamma > 16p$ by taking $\varepsilon > 0$ small enough.

Eventually, if we are heading for an almost sure result, considering that for $\gamma > 24p$ the sum $\sum_N N^{-\frac{\gamma(p-\varepsilon)}{8p^2}+2}$ is even finite and hence—using the Borel–Cantelli lemma—we can conclude that eventually all patterns ξ^μ, $\mu = 1, \ldots, M$ are stable with probability one. □

The proof in the case of the BEG network is rather similar in spirit to the above proof. However, the details need to be worked out carefully.

Proof of theorem 3.3. We treat the question of whether a spin is fixed under the dynamics T_i defined by (2.4) separately for the cases that this spin is 1 or 0 (again we can assume that the pattern we are interested in just takes values 0 or 1).

Let us first assume that $\xi_1^1 = 0$. According to the definition of the retrieval dynamics in the BEG-network $T_1(\xi^1) \neq 0$ if and only if

$$|h_1(\xi^1)| > -\frac{1}{p^2(1-p)^2N}\sum_{j\neq 1}\sum_{\mu=1}^{M}\eta_1^\mu\eta_j^\mu\left(\xi_j^1\right)^2.$$

Thus

$$\mathbb{P}\big(T_1(\xi^1)\neq 0\big|\xi_1^1=0\big) = \mathbb{P}\left(\left|\sum_{j\neq 1}\sum_{\mu=1}^{M}\xi_1^\mu\xi_j^\mu\xi_j^1\right| > -\frac{1}{(1-p)^2}\sum_{j\neq 1}\sum_{\mu=1}^{M}\eta_1^\mu\eta_j^\mu\left(\xi_j^1\right)^2\big|\xi_1^1=0\right).$$

Assuming that $\xi_j^1 = 1$ for $j = 2, \ldots, K \leqslant N$ and $\xi_j^1 = 0$ for $K < j \leqslant N$ (and calling this condition $\mathcal{K}$ again) yields, that for $\mu = 1$ one has $\sum_{j=2}^{K}\xi_1^1\xi_j^1\xi_j^1 = 0$ as well as $\frac{1}{(1-p)^2}\sum_{j=2}^{K}\eta_1^1\eta_j^1\left(\xi_j^1\right)^2 = -\frac{(K-1)p}{1-p}$. Therefore

$$\begin{aligned}&\mathbb{P}\big(T_1(\xi^1)\neq 0\big|\xi_1^1=0,\mathcal{K}\big)\\ &\quad= \mathbb{P}\left(\left|\sum_{j\neq 1}\sum_{\mu=1}^{M}\xi_1^\mu\xi_j^\mu\xi_j^1\right| > -\frac{1}{(1-p)^2}\sum_{j\neq 1}\sum_{\mu=1}^{M}\eta_1^\mu\eta_j^\mu\left(\xi_j^1\right)^2\big|\xi_1^1=0,\mathcal{K}\right)\\ &\quad\leqslant \mathbb{P}\left(\sum_{j=2}^{K}\sum_{\mu=2}^{M}\xi_1^\mu\xi_j^\mu + \frac{1}{(1-p)^2}\eta_1^\mu\eta_j^\mu > \frac{(K-1)p}{1-p}\right|\xi_1^1=0,\mathcal{K}\Bigg)\\ &\qquad+\mathbb{P}\left(\sum_{j=2}^{K}\sum_{\mu=2}^{M}\xi_1^\mu\xi_j^\mu - \frac{1}{(1-p)^2}\eta_1^\mu\eta_j^\mu < -\frac{(K-1)p}{1-p}\right|\xi_1^1=0,\mathcal{K}\Bigg).\end{aligned}$$

Along the lines of the proof of 3.1 one computes

$$\mathbb{E}\exp\left(t\left(\sum_{j=2}^{K}\sum_{\mu=2}^{M}\xi_1^\mu\xi_j^\mu+\frac{1}{(1-p)^2}\eta_1^\mu\eta_j^\mu\right)\right)$$
$$=\exp\left(\frac{t^2}{2}(M-1)(K-1)p^2\left(1+\frac{1}{(1-p)^2}\right)+\mathcal{O}(KMt^3)\right)$$

for $t\to 0$, which gives by the exponential Chebyshev inequality

$$\mathbb{P}\big(T_1(\xi^1)\neq 0\,\big|\,\xi_1^1=0,\mathcal{K}\big)\leqslant 2\exp\left(-\frac{K-1}{2(M-1)((1-p)^2+1)}+\mathcal{O}(KM^{-2})\right).$$

In the case that $\xi_1^1 = 1$ one has two possibilities that ξ_1^1 is not stable under the retrieval dynamics: either $\Theta\big(|h_1(\xi^1)|+\frac{1}{p^2(1-p)^2N}\sum_{j\neq 1}\sum_{\mu=1}^{M}\eta_1^\mu\eta_j^\mu\big(\xi_j^1\big)^2\big)=0$, or the corresponding Heaviside function $\Theta=1$, but $h_1(\xi^1)<0$. Hence the following bound can be obtained

$$\mathbb{P}\big(T_1(\xi^1)\neq 1\big|\xi_1^1=1,\mathcal{K}\big)\leqslant\mathbb{P}\left(\sum_{j=2}^{K}\sum_{\mu=2}^{M}\xi_1^\mu\xi_j^\mu+\frac{1}{(1-p)^2}\eta_1^\mu\eta_j^\mu<-2(K-1)\big|\xi_1^1=1,\mathcal{K}\right)$$
$$+\mathbb{P}\left(\sum_{j=2}^{K}\sum_{\mu=2}^{M}\xi_1^\mu\xi_j^\mu<-(K-1)\big|\xi_1^1=1,\mathcal{K}\right)\tag{4.5}$$

where we use the same condition $\mathcal{K}$ as above and that under the condition $\{\xi_1^1=1,\mathcal{K}\}$, one has that $\sum_{j\neq 1}\xi_1^1\xi_j^1\xi_j^1=K-1$ as well as $\frac{1}{(1-p)^2}\sum_{j\neq 1}\eta_1^1\eta_j^1\big(\xi_j^1\big)^2=K-1$. The two summands on the right-hand side of (4.5) are again estimated by the exponential Chebyshev inequality: using the computations from above one obtains

$$\mathbb{P}\big(T_1(\xi^1)\neq 1\,\big|\,\xi_1^1=1,\mathcal{K}\big)\leqslant\exp\left(-\frac{2(K-1)(1-p)^2}{(M-1)p^2((1-p)^2+1)}+\mathcal{O}(KM^{-2})\right)$$
$$+\exp\left(-\frac{K-1}{2p^2(M-1)}+\mathcal{O}(KM^{-2})\right).\tag{4.6}$$

Again with overwhelming probability $K\sim pN$ and (4.3) holds true. Thus putting $M=\frac{N}{\gamma\log N}$ we obtain that with overwhelming probability

$$\mathbb{P}\big(T_1(\xi^1)\neq 0\big|\xi_1^1=0,\mathcal{K}\big)\leqslant 2N^{-\frac{\gamma p}{2((1-p)^2+1)}}(1+\varepsilon_N)$$

as well as

$$\mathbb{P}\big(T_1(\xi^1)\neq 1\big|\xi_1^1=1,\mathcal{K}\big)\leqslant\Big(N^{-\frac{2\gamma(1-p)^2}{p((1-p)^2+1)}}+N^{-\frac{\gamma}{2p}}\Big)(1+\varepsilon_N)$$

for a sequence $\varepsilon_N\to 0$. In order to guarantee that ξ^1 is fixed we need to have all $\xi_i^1=1$ are mapped to 1 by T_i and that all $\xi_i^1=0$ are mapped to 0 by T_i. This can be guaranteed (with probability converging to one), if all of $N^{1-\frac{\gamma p}{2((1-p)^2+1)}}$, $N^{1-\frac{2\gamma(1-p)^2}{p((1-p)^2+1)}}$, and $N^{1-\frac{\gamma}{2p}}$ converge to zero. This is true if

$$\gamma>\max\left\{\frac{2((1-p)^2+1)}{p},\frac{p((1-p)^2+1)}{2(1-p)^2},2p\right\}=\gamma^*(p).$$

Hence for $\gamma>\gamma^*(p)$ we have that

$$\mathbb{P}\big(\exists_i:T_i(\xi^1)\neq\xi_i^1\big)\to 0$$

as $N \to \infty$. And the third term $2p$ may be omitted since $2((1-p)^2+1) > 2p^2$ for all $p \in]0,1[$.

If we ask for the probability that all patterns are stable with a probability converging to one, we even need to assure that $MN^{1-\frac{\gamma p}{2((1-p)^2+1)}}$, $MN^{1-\frac{2\gamma(1-p)^2}{p((1-p)^2+1)}}$ and $MN^{1-\frac{\gamma}{2p}}$ converge to zero. This is true if $\gamma > 2\gamma^*(p)$.

Eventually, if we are interested in the equation that all patterns become stable with probability one, we need to assure that $MN^{1-\frac{\gamma p}{2((1-p)^2+1)}}$, $MN^{1-\frac{2\gamma(1-p)^2}{p((1-p)^2+1)}}$ and $MN^{1-\frac{\gamma}{2p}}$ are even summable, to be able to apply a Borel–Cantelli argument as above. The latter is true if $\gamma > 3\gamma^*(p)$. □

Finally, we analyse the other kind of storage capacity.

Proof of theorem 3.7. For $J \subseteq \{1,\dots,N\}$, $|J| = \delta N$ (for simplicity we assume that δN is an integer), let us denote by $\xi_J^\mu = \left(\xi_{J,i}^\mu\right)_{i=1,\dots,N}$ the vector that disagrees with a given pattern ξ^μ on exactly the coordinates $j \in J$. For a given $\varepsilon > 0$ we are interested in the probability

$$\mathbb{P}\left(\bigcup_\mu \bigcup_J H_N\left(\xi_J^\mu\right) \leqslant H_N(\xi^\mu) + \varepsilon N\right) \leqslant M\mathbb{P}\left(\bigcup_J H_N\left(\xi_J^1\right) \leqslant H_N(\xi^1) + \varepsilon N\right).$$

By symmetry, we assume again that $\xi_i^1 \in \{0,1\}$ for all $i \in \{1,\dots,N\}$. Then

$$\mathbb{P}\left(\bigcup_J H\left(\xi_J^1\right) \leqslant H_N(\xi^1) + \varepsilon N\right) = \mathbb{P}\left(\bigcup_{J_1,J_2,J_3,J_4} H_N\left(\xi_J^1\right) \leqslant H_N(\xi^1) + \varepsilon N\right),$$

where

$$J = \bigcup_{k=1}^4 J_k, \qquad J_1 = \left\{i : \xi_i^1 = 1, \xi_{J,i}^1 = 0\right\}, \qquad J_2 = \{i : \xi_i^1 = 1, \xi_{J,i}^1 = -1\},$$

$$J_3 = \left\{i : \xi_i^1 = 0, \xi_{J,i}^1 = -1\right\}, \qquad \text{and} \qquad J_4 = \left\{i : \xi_i^1 = 0, \xi_{J,i}^1 = 1\right\}.$$

Let $S^1 = \mathrm{card}\left\{j : \xi_j^1 = 1\right\}$ and $k_i = \mathrm{card}(J_i)$, $i = 1,\dots,4$. Then

$$\mathbb{P}\left(\bigcup_J H_N\left(\xi_J^1\right) \leqslant H_N(\xi^1) + \varepsilon N\right) = \sum_{K=0}^N \mathbb{P}_K\left(\bigcup_J H_N\left(\xi_J^1\right) \leqslant H_N(\xi^1) + \varepsilon N\right)\mathbb{P}(S^1 = K) \tag{4.7}$$

where $P_K(\cdot)$ denotes the conditional probability under the condition that $S^1 = K$. Now

$$\sum_{K=0}^N \mathbb{P}_K\left(\bigcup_J H_N\left(\xi_J^1\right) \leqslant H_N(\xi^1) + \varepsilon N\right)\mathbb{P}(S^1 = K)$$
$$\leqslant \sum_{K=0}^N \sum_{k_1,k_2,k_3,k_4} d(k_1,k_2,k_3,k_4)\mathbb{P}_K\left(H_N\left(\xi_J^1\right) \leqslant H_N(\xi^1) + \varepsilon N\right)\mathbb{P}(S^1 = K),$$

where

$$d(k_1,k_2,k_3,k_4) = \binom{K}{k_1}\binom{K-k_1}{k_2}\binom{N-K}{k_3}\binom{N-K-k_3}{k_4}$$

and the sum is over k_1,k_2,k_3,k_4 such that $k_1+k_2 \leqslant K$, $k_3+k_4 \leqslant N-K$, $k_1+k_2+k_3+k_4 = \delta N$. For calculations we may choose ξ^1 and ξ_J^1 as particular configurations on the right-hand side: for instance $\xi_i^1 = 1$ for $i = 1,\dots,K$, and $J_1 = \{1,\dots,k_1\}$, $J_2 = \{k_1+1,\dots,k_1+k_2\}$,

$J_3 = \{K+1, \ldots, K+k_3\}$, $J_4 = \{K+k_3+1, \ldots, K+k_3+k_4\}$. For $N \in \mathbb{N}^*$, $\mu = 1, 2, \ldots, M$, and $\sigma \in \{-1, 0, +1\}^N$ define

$$\tilde{H}_N^\mu(\sigma) := -\frac{1}{2p^2N} \sum_{i,j=1;i\neq j}^N \sigma_i\sigma_j\xi_i^\mu\xi_j^\mu = -\frac{1}{2p^2N}\left(\left(\sum_{i=1}^N \sigma_i\xi_i^\mu\right)^2 - \sum_{i=1}^N \left(\sigma_i\xi_i^\mu\right)^2\right)$$

and

$$\overline{H}_N^\mu(\sigma) := -\frac{1}{2p^2(1-p)^2N} \sum_{i,j=1;i\neq j}^N \sigma_i^2\sigma_j^2\eta_i^\mu\eta_j^\mu$$
$$= -\frac{1}{2p^2(1-p)^2N}\left(\left(\sum_{i=1}^N \sigma_i^2\eta_i^\mu\right)^2 - \sum_{i=1}^N \left(\sigma_i^2\eta_i^\mu\right)^2\right).$$

Then

$$H_N(\sigma) = \sum_{\mu=1}^{M(N)} \tilde{H}_N^\mu(\sigma) + \overline{H}_N^\mu(\sigma).$$

A little computation shows that for $K = \lambda N$ (which we assume to be an integer) and $\delta_i = k_i/N$,

$$\tilde{H}_N^1(\xi^1) - \tilde{H}_N^1(\xi_J^1) = -N\left(\left(\frac{\delta_1+2\delta_2}{p}\right)\left(\frac{\lambda}{p} - \frac{\delta_1+2\delta_2}{2p}\right) + \mathcal{O}\left(\frac{1}{N}\right)\right),$$

and

$$\overline{H}_N^1(\xi^1) - \overline{H}_N^1(\xi_J^1) = -N\left(\left(\frac{\delta_1}{p} + \frac{\delta_3+\delta_4}{1-p}\right)\left(\frac{\lambda}{p} - \frac{\delta_1}{2p} - \frac{\delta_3+\delta_4}{2(1-p)}\right) + \mathcal{O}\left(\frac{1}{N}\right)\right).$$

Hence with $H_N^1(\sigma) := \tilde{H}_N^1(\sigma) + \overline{H}_N^1(\sigma)$ we obtain that

$$H_N^1(\xi^1) - H_N^1(\xi_J^1) = -N\left(c + \mathcal{O}\left(\frac{1}{N}\right)\right)$$

for a positive constant

$$c = c(p,\lambda,\delta_1,\delta_2,\delta_3,\delta_4) := \left(\frac{\delta_1+2\delta_2}{p}\right)\left(\frac{\lambda}{p} - \frac{\delta_1+2\delta_2}{2p}\right)$$
$$+ \left(\frac{\delta_1}{p} + \frac{\delta_3+\delta_4}{1-p}\right)\left(\frac{\lambda}{p} - \frac{\delta_1}{2p} - \frac{\delta_3+\delta_4}{2(1-p)}\right) \tag{4.8}$$

For $\mu = 2, \ldots, M$, we have

$$\tilde{H}_N^\mu(\xi^1) - \tilde{H}_N^\mu(\xi_J^1) = -\frac{1}{2p^2N}\left(\left(\sum_{i=1}^N \xi_i^\mu\xi_i^1\right)^2 - \left(\sum_{i=1}^N \xi_i^\mu\xi_{J,i}^1\right)^2\right) + \mathcal{O}(1)$$
$$= -\frac{1}{2p^2N}\left(\left(\sum_{i=1}^N \xi_i^\mu\xi_i^1\right) - \left(\sum_{i=1}^N \xi_i^\mu\xi_{J,i}^1\right)\left(\sum_{i=1}^N \xi_i^\mu\xi_i^1\right.\right.$$
$$\left.\left. + \left(\sum_{i=1}^N \xi_i^\mu\xi_{J,i}^1\right)\right) + \mathcal{O}(1)$$
$$= -\frac{1}{2p^2N}AB + \mathcal{O}(1),$$

where

$$A = \sum_{i\in J_1}\xi_i^\mu + 2\sum_{i\in J_2}\xi_i^\mu + \sum_{i\in J_3}\xi_i^\mu - \sum_{i\in J_4}\xi_i^\mu$$

and

$$B = \sum_{i\in J_1} \xi_i^\mu - \sum_{i\in J_3} \xi_i^\mu + \sum_{i\in J_4} \xi_i^\mu + 2\sum_{i\in J_5} \xi_i^\mu,$$

where $J_5 = \left\{i : \xi_i^1 = \xi_{J,i}^1 = 1\right\}$. Expanding the product AB, we get

$$\tilde{H}_N^\mu(\xi^1) - \tilde{H}_N^\mu\left(\xi_J^1\right) \geqslant -\frac{1}{2p^2N}\left(2P_{15}^\mu + 2P_{21}^\mu - 2P_{23}^\mu + 2P_{24}^\mu + 4P_{25}^\mu - P_{33}^\mu + 2P_{34}^\mu + 2P_{35}^\mu - P_{44}^\mu - 2P_{45}^\mu\right) + \mathcal{O}(1), \tag{4.9}$$

where

$$P_{kl}^\mu = \left(\sum_{i\in J_k} \xi_i^\mu\right)\left(\sum_{i\in J_l} \xi_i^\mu\right),$$

and we use $P_{11}^\mu \geqslant 0$.

Similarly, we obtain for the second part of the Hamiltonian

$$\overline{H}_N^\mu(\xi^1) - \overline{H}_N^\mu\left(\xi_J^1\right) = -\frac{1}{2p^2(1-p)^2N} CD + \mathcal{O}(1),$$

where

$$C = \sum_{i\in J_1} \eta_i^\mu - \sum_{i\in J_3\cup J_4} \eta_i^\mu$$

and

$$D = \sum_{i\in J_1} \eta_i^\mu + 2\sum_{i\in J_2\cup J_5} \eta_i^\mu + \sum_{i\in J_3\cup J_4} \eta_i^\mu.$$

Expanding the product CD, we get

$$\overline{H}_N^\mu(\xi^1) - \overline{H}_N^\mu\left(\xi_J^1\right) \geqslant -\frac{1}{2p^2(1-p)^2N}\left(2Q_{12}^\mu + Q_{14}^\mu + 2Q_{15}^\mu - 2Q_{32}^\mu - Q_{33}^\mu - 2Q_{34}^\mu - 2Q_{35}^\mu - Q_{41}^\mu - 2Q_{42}^\mu - Q_{44}^\mu - 2Q_{45}^\mu\right) + O(1), \tag{4.10}$$

where

$$Q_{kl}^\mu = \sum_{i\in J_k} \eta_i^\mu \sum_{i\in J_l} \eta_i^\mu,$$

and we use $Q_{11}^\mu \geqslant 0$. Then we have

$$\mathbb{P}_{\lambda N}\left(H_N\left(\xi_J^1\right) \leqslant H_N(\xi^1) + \varepsilon N\right) = \mathbb{P}_{\lambda N}\left(\sum_{\mu=2}^M H_N^\mu(\xi^1) - H_N^\mu\left(\xi_J^1\right) \geqslant (c-\varepsilon)N\left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right)\right),$$

and equations (4.9) and (4.10) give the bound

$$\mathbb{P}\left(H_N\left(\xi_J^1\right) \leqslant H_N(\xi^1) + \varepsilon N | S^1 = \lambda N\right) \leqslant \mathbb{P}\left(-\frac{1}{p^2N}\sum_{\mu=2}^M P_{15}^\mu \geqslant \frac{c-\varepsilon}{21}N\left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right)\right) + \cdots + \mathbb{P}\left(\frac{1}{p^2(1-p)^2N}\sum_{\mu=2}^M Q_{45}^\mu \geqslant \frac{c-\varepsilon}{21}N\left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right)\right).$$

Now again, for N large enough with overwhelming probability $S^1 \sim pN$. From (4.7) it immediately follows that it will be possible to restrict our considerations to the case $\lambda \in [p-\epsilon', p+\epsilon']$ for a fixed $\epsilon' > 0$ arbitrarily small.

Finally we need the bounds of lemma 4.1 stated below. It gives exponential bounds on the probabilities of interest.

Assume $M = \alpha N$, for some constant $\alpha > 0$ not depending on N. Then, if Z_N is a random variable such that $\mathbb{P}(Z_N \geqslant \gamma M) \leqslant \mathrm{e}^{-F(\gamma)M}$ for a$\gamma > 0$ and a positive function F, one immediately obtains that $\mathbb{P}(Z_N \geqslant \epsilon N) \leqslant \mathrm{e}^{-\alpha F(\frac{\epsilon}{\alpha})N}$. Hence to achieve the proof, we have to show that for α sufficiently small, there exist $\delta, \epsilon, \epsilon' > 0$ such that for all $\delta_1, \delta_2, \delta_3, \delta_4 > 0$, $\delta_1 + \delta_2 \leqslant \lambda$, $\delta_3 + \delta_4 \leqslant 1 - \lambda$, and $\delta_1 + \delta_2 + \delta_3 + \delta_4 =: \delta$, we have for F_i as in lemma 4.1

$$\alpha F_i\left(\frac{c(p,\lambda,\delta_1,\delta_2,\delta_3,\delta_4)-\epsilon}{21\alpha}\right) > Q(p,\lambda,\delta_1,\delta_2,\delta_3,\delta_4) \qquad \text{for} \quad i = 1,2,3,4, \tag{4.11}$$

where Q is defined by

$$\binom{K}{k_1}\binom{K-k_1}{k_2}\binom{N-K}{k_3}\binom{N-K-k_3}{k_4} =: \mathrm{e}^{N(Q+\mathcal{O}(\frac{1}{N}))}.$$

Using Stirling's formula, we get

$$\begin{aligned} Q &= -\sum_{i=1}^{4}\delta_i\log(\delta_i) + (\delta_1+\delta_2)\log(\lambda-\delta_1-\delta_2) - \lambda\log\left(1-\frac{\delta_1+\delta_2}{\lambda}\right) \\ &\quad + (\delta_3+\delta_4)\log(1-\lambda-\delta_3-\delta_4) - (1-\lambda)\log\left(1-\frac{\delta_3+\delta_4}{1-\lambda}\right) \\ &= -\sum_{i=1}^{4}\delta_i\log(\delta_i)\left(1-\frac{1+\log(p)}{\log(\delta_i)}\right) + \mathcal{O}((\delta_1+\delta_2)^2+(\delta_3+\delta_4)^2). \end{aligned}$$

Since the function $f(x) = -x\log(x)$ is increasing for $x \in [0, \mathrm{e}^{-1}]$, there exists a constant $c_6 > 0$ such that for $\delta > 0$ sufficiently small and all admissible δ_i,

$$Q \leqslant -c_6\delta\log(\delta). \tag{4.12}$$

Now there exist strictly positive constants $c_1, \ldots, c_5$ depending only on p and ϵ' such that

$$c - \epsilon > (1-c_5\delta)\sum_{i=1}^{4}c_i\delta_i - \epsilon > \frac{(1-c_5\delta)c^*}{2}\delta, \tag{4.13}$$

where $c^* = \min\{c_i, i = 1,\ldots,4\}$ and $\epsilon := \frac{(1-c_5\delta)c^*}{2}\delta$. Let us consider the different cases:

- The function F_1 verifies $\alpha F_1\left(\frac{\gamma}{\alpha}\right) \sim \frac{2\gamma p^2}{\sqrt{\gamma_1\gamma_2}}$ for small α. Hence to tackle the terms with P_{ij}, $i \neq j$, we get the condition

$$\frac{2p^2(c-\epsilon)}{21\sqrt{\delta_i\delta_j}Q} > 1, \tag{4.14}$$

which is equivalent to (4.11). Using (4.12) and (4.13), we have

$$\frac{2p^2(c-\epsilon)}{21\sqrt{\delta_i\delta_j}Q} > -\frac{p^2(1-c_5\delta)c^*}{21c_6}\frac{\delta}{\sqrt{\delta_i\delta_j}\delta\log(\delta)}.$$

For $i, j \in \{1,2,3,4\}$, we have

$$-\frac{\delta}{\sqrt{\delta_i\delta_j}\delta\log(\delta)} \geqslant -\frac{1}{\delta\log(\delta)} \longrightarrow +\infty, \qquad \text{as} \quad \delta \longrightarrow 0.$$

For $i \in \{1, 2, 3, 4\}$ and $j = 5$, since $\delta_5 := \lambda - \delta_1 - \delta_2 \leqslant p + \epsilon'$ we get

$$-\frac{\delta}{\sqrt{\delta_i \delta_j}\delta \log(\delta)} \geqslant -\frac{\sqrt{\delta}}{(p+\epsilon')\delta \log(\delta)} \longrightarrow +\infty, \qquad \text{as} \quad \delta \longrightarrow 0.$$

Hence (4.14) is verified for all P_{ij}, $i \neq j$ for δ sufficiently small.

- The function F_2 satisfies $\alpha F_2\left(\frac{\gamma}{\alpha}\right) = \frac{\gamma^2 p^4}{\alpha \gamma_1{}^2}$. Hence for the terms with P_{33} and P_{44}, using (4.12) and (4.13) we get
$$\frac{\alpha F_2\left(\frac{c-\epsilon}{21\alpha}\right)}{Q} \geqslant -\left(\frac{(1-c_5\delta)c^* p^2}{42}\right)^2 \frac{1}{c_6 \alpha \delta \log(\delta)},$$
which is > 1 for small δ and α.
- The function F_3 is the same as F_1 up to a constant depending only on p, which implies (4.11) for $i = 3$.
- The function F_4 satisfies $\alpha F_4\left(\frac{\gamma}{\alpha}\right) = \frac{\gamma}{v} - \frac{\alpha}{2} + \frac{\alpha}{2}\log\left(\frac{\alpha w}{2\gamma}\right) \sim \frac{\gamma}{w}$, for small α, where
$$w = \frac{(1+|2p-1|)^2}{4p^2(1-p)^2}\gamma_1.$$
Hence for the terms with Q_{ii}, we have to verify
$$\frac{4p^2(1-p)^2}{(1+|2p-1|)^2}\frac{c-\epsilon}{21\delta_i Q} > 1, \qquad \text{for} \quad i = 3, 4.$$
Using (4.12) and (4.13), we get
$$\frac{c-\epsilon}{21\delta_i Q} > -c_7 \frac{1}{\delta \log(\delta)},$$
for some $c_7 > 0$, which implies the result for δ sufficiently small. □

Finally let us state and prove the lemma we have just used.

Lemma 4.1. *Let* $p \in]0, 1[$ *and* $\left(\xi_i^\mu\right)_{i=1,\dots,N}^{\mu=1,\dots,M}$ *be independent random variables such that*
$$1 - p = \mathbb{P}\left(\xi_i^\mu = 0\right) \quad \textit{and} \quad \mathbb{P}\left(\xi_i^\mu = 1\right) = \mathbb{P}\left(\xi_i^\mu = -1\right) = \frac{p}{2}.$$
Let $I_1, I_2 \subset \{1, \dots, N\}$, $I_1 \cap I_2 = \emptyset$, *such that* $\mathrm{card}(I_1) = \gamma_1 N$ *and* $\mathrm{card}(I_2) = \gamma_2 N$. *Let* $P_k^\mu = \sum_{i \in I_k} \xi_i^\mu$ *and* $Q_k^\mu = \sum_{i \in I_k}\left(\left(\xi_i^\mu\right)^2 - p\right)$ *for* $k = 1, 2$. *Then for* $s \in \{-1, 1\}$, *for all* $\gamma > 0$,

(i)
$$\mathbb{P}\left(\frac{s}{2p^2 NM}\sum_{\mu=1}^{M} P_1^\mu P_2^\mu \geqslant \gamma\right) \leqslant \mathrm{e}^{-F_1(\gamma)M},$$
where
$$F_1(\gamma) = \frac{1}{2}\left(\sqrt{1+4u^2\gamma^2} - 1 + \log\left(\frac{\sqrt{1+4u^2\gamma^2}-1}{2u^2\gamma^2}\right)\right)$$
and
$$u = \frac{2p^2}{\sqrt{\gamma_1\gamma_2}}.$$

(ii)
$$\mathbb{P}\left(\frac{1}{p^2 NM}\sum_{\mu=1}^{M}\left(P_1^\mu\right)^2 \geqslant \gamma\right) \leqslant \mathrm{e}^{-F_2(\gamma)M}$$
where
$$F_2(\gamma) = \frac{\gamma^2 p^4}{\gamma_1{}^2},$$

(iii)

$$\mathbb{P}\left(\frac{s}{2p^2(1-p)^2NM}\sum_{\mu=1}^{M}Q_1^\mu Q_2^\mu \geqslant \gamma\right) \leqslant \mathrm{e}^{-F_3(\gamma)M},$$

where

$$F_3(\gamma) = \frac{1}{2}\left(\sqrt{1+4v^2\gamma^2} - 1 + \log\left(\frac{\sqrt{1+4v^2\gamma^2}-1}{2v^2\gamma^2}\right)\right)$$

and

$$v = \frac{8p^2(1-p)^2}{(1+|2p-1|)^2\sqrt{\gamma_1\gamma_2}}.$$

(iv)

$$\mathbb{P}\left(\frac{1}{2p^2(1-p)^2NM}\sum_{\mu=1}^{M}\left(\left(Q_1^\mu\right)^2 \geqslant \gamma\right)\right) \leqslant e^{-F_4(\gamma)M}$$

where

$$F_4(\gamma) = \frac{1}{2}\left(\left(\frac{2\gamma}{w} - 1 + \log\left(\frac{w}{2\gamma}\right),\right)\right.$$

and

$$w = \frac{(1+|2p-1|)^2}{4p^2(1-p)^2}\gamma_1.$$

Proof of lemma 4.1. To obtain the different bounds, we use the exponential Markov inequality: for a sum of independent variable $Z = \sum_{\mu=1}^{M} Z^\mu$, for all $\gamma > 0$,

$$\mathbb{P}(Z \geqslant \gamma M) \leqslant \inf_{t>0} \mathrm{e}^{-\gamma t M}\left(\mathbb{E}\left[\mathrm{e}^{tZ^1}\right]\right)^M.$$

Thus we have to evaluate the exponential moment in the four different cases and to mimimize in t. For (i) and (ii) we use the same method as in [N88]. More precisely we have

$$\mathbb{E}\left(\exp\left(\frac{t}{2p^2N}P_1^1P_2^1\right)\right) \leqslant \mathbb{E}\left(\exp\left(\frac{t\sqrt{\gamma_1\gamma_2}}{2p^2}\left(\frac{1}{\sqrt{\gamma_1 N}}\sum_{i\in I_1}X_i\right)\left(\frac{1}{\sqrt{\gamma_2 N}}\sum_{i\in I_2}X_i\right)\right)\right)$$
$$= \mathbb{E}\left(\exp\left(\frac{t\sqrt{\gamma_1\gamma_2}}{2p^2}X_1X_2\right)\right),$$

and

$$\mathbb{E}\left(\exp\left(\frac{t}{2p^2N}\left(P_1^1\right)^2\right)\right) \leqslant \mathbb{E}\left(\exp\left(\frac{t\gamma_1}{2p^2}\left(\frac{1}{\sqrt{\gamma_1 N}}\sum_{i\in I_1}X_i\right)^2\right)\right) = \mathbb{E}\left(\exp\left(\frac{t\gamma_1}{2p^2}(X_1)^2\right)\right),$$

where the X_i are independent standard normal random variables. Then exact calculations give the results. For (iii) we have

$$\mathbb{E}\left(\exp\left(\frac{t}{2p^2(1-p)^2N}Q_1^1Q_2^1\right)\right) = \mathbb{E}_{(\xi_i^1)_{i\in I_1}}\left(\mathbb{E}_{(\xi_i^1)_{i\in I_2}}\left(\exp\left(\frac{t}{2p^2(1-p)^2N}Q_1^1Q_2^1\right)\right)\right)$$
$$= \mathbb{E}_{(\xi_i^1)_{i\in I_1}}\left(\left(p\exp\left(\frac{t}{2p^2(1-p)N}Q_1^1\right)\right.\right.$$
$$\left.\left.+ (1-p)\exp\left(-\frac{t}{2p(1-p)^2N}Q_1^1\right)\right)^{\gamma_2 N}\right),$$

where $\mathbb{E}_{(\xi_i^1)_{i\in I_1}}$ denotes the expectation with respect to the random variables $\left(\xi_i^1\right)_{i\in I_1}$. Then using the inequality

$$p\,\mathrm{e}^{(1-p)t} + (1-p)\,\mathrm{e}^{-pt} \leqslant \cosh\left((1+|2p-1|)\frac{t}{2}\right) \leqslant \mathrm{e}^{(1+|2p-1|)^2\frac{t^2}{8}}, \quad \forall\, t\in\mathbb{R} \tag{4.15}$$

we get for a standard normal random variable X

$$\begin{aligned}
\mathbb{E}\left(\exp\left(\frac{t}{2p^2(1-p)^2N}Q_1^1Q_2^1\right)\right) &\leqslant \mathbb{E}_{(\xi_i^1)_{i\in I_1}}\left(\exp\left(\frac{(1+|2p-1|)^2\gamma_2t^2}{32p^4(1-p)^4N}\left(Q_1^1\right)^2\right)\right)\\
&= \mathbb{E}_X\left(\mathbb{E}_{(\xi_i^1)_{i\in I_1}}\left(\exp\left(\frac{(1+|2p-1|)\sqrt{\gamma_2}t}{4p^2(1-p)^2\sqrt{N}}Q_1^1X\right)\right)\right),\\
&= \mathbb{E}_X\Bigg(p\exp\left(\frac{(1+|2p-1|)\sqrt{\gamma_2}t}{4p^2(1-p)\sqrt{N}}X\right)\\
&\quad + (1-p)\exp\left(-\frac{(1+|2p-1|)\sqrt{\gamma_2}t}{4p(1-p)^2\sqrt{N}}X\right)^{\gamma_1N}\Bigg)\\
&\leqslant \mathbb{E}_X(\exp(c_8t^2X^2))\\
&= \frac{1}{\sqrt{1-2c_8t^2}},
\end{aligned}$$

where

$$c_8 = \frac{(1+|2p-1|)^4\gamma_1\gamma_2}{128p^4(1-p)^4N}.$$

To obtain (iv) we write for a standard normal random variable X

$$\begin{aligned}
\mathbb{E}\left(\exp\left(\frac{t}{2p^2(1-p)^2N}\left(Q_1^1\right)^2\right)\right) &= \mathbb{E}_X\mathbb{E}\left(\exp\left(\frac{\sqrt{t}}{p(1-p)\sqrt{N}}\left(\sum_{i\in I_1}\left(\left(\xi_i^\mu\right)^2-p\right)\right)X\right)\right)\\
&= \mathbb{E}_X\left(p\exp\left(X\frac{\sqrt{t}}{p\sqrt{N}}\right) + (1-p)\exp\left(-X\frac{\sqrt{t}}{(1-p)\sqrt{N}}\right)\right)^{\gamma_1N}.
\end{aligned}$$

Then using inequality (4.15) we get

$$\mathbb{E}\left(\exp\left(\frac{t}{2p^2(1-p)^2N}\left(Q_1^1\right)^2\right)\right) \leqslant \mathbb{E}_X\left(\mathrm{e}^{c_9tX^2}\right) = \frac{1}{\sqrt{1-2c_9t}},$$

with

$$c_9 = \frac{(1+|2p-1|)^2}{8p^2(1-p)^2}\gamma_1.$$

□

5. Concluding remarks

In this paper, we have compared the performance of the Blume–Emery–Griffiths neural network to the simplest three-state network, the simple threshold network. When the storage capacity is defined as the number of patterns that can be permitted such that they (one or all of them) are fixed points of the retrieval dynamics, both these networks show a storage capacity of $M = N/\gamma\log N$ for an appropriately chosen γ. Nevertheless for small values of p (the activity of the network), the simple threshold model outperforms the BEG model. As in the paper [BV03], this finding seems to contradict the optimality of the BEG model claimed in

[DK00, BoVe02]. However their optimality criterion differs from our simple storage capacity criterion.

Moreover we have proved that with a more liberal definition of storage capacity, the BEG is able to store $M = \alpha N$ patterns for some $\alpha > 0$. Although our approach is different and a direct comparison is not possible, this supports the results in [BCS03, BV03] that were found with the help of the replica method (even though with a probably different value for α).

Acknowledgments

This research was carried out while the first author was on a visit at the University of Brest, France. He is extremely grateful to the University of Brest for the invitation, the facilities and for the atmosphere provided, necessary for fruitful investigations. Moreover we thank the anonymous referees for valuable comments.

References

[Am89] Amari S 1989 Characteristics of sparsely encoded associative memory *Neural Netw.* **2** 451–7

[AGS87] Amit D J, Gutfreund G and Sompolinsky H 1987 Statistical mechanics of neural networks near saturation *Ann. Phys.* **173** 30–67

[BS92] Baram Y and Sal'ee D 1992 Lower bounds on the capacities of binary and ternary networks storing sparse random vectors *IEEE Inf. Theory* **38** 1633–47

[BoVe02] Bollé D and Verbeiren T 2002 An optimal Q-state neural network using mutual information *Phys. Lett.* A **297** 156–61

[BCS03] Bollé D, Prez Castillo I and Shim G M 2003 Optimal capacity in a Blume–Emery–Griffiths perpectron *Phys. Rev.* E **67** 036113

[BV03] Bollé D and Verbeiren T 2003 Thermodynamics of fully connected Blume–Emery–Griffiths neural networks *J. Phys. A: Math. Gen.* **36** 295–305

[B99] Bovier A 1999 Sharp upper bounds on perfect retrieval in the Hopfield model *J. Appl. Probab.* **36** 941–50

[Bu94] Burshtein D 1994 Nondirect convergence radius and number of iterations of the Hopfield associative memory *IEEE Trans. Inf. Theory* **40** 838–47

[DK00] Carreta D and Korutcheva E 2000 Three-state neural network: from mutual information to the Hamiltonian *Phys. Rev.* E **62** 2620–8

[FMP92] Ferrari P, Martinez S and Picco P 1992 A lower bound for the memory capacity in the Potts–Hopfield model *J. Stat. Phys.* **66** 1643–52

[FP77] Figotin A L and Pastur L A 1977 Exactly soluble model of a spin-glas *Sov. J. Low Temp. Phys.* **3** 378–83

[Ga88] Gardner E 1988 The space of interactions in neural network models *J. Phys. A: Math. Gen.* **21** 257–70

[Gr92] Grimmett G and Stirzaker D 1992 *Probability and Random Processes* 2nd edn (Oxford: Oxford University Press)

[HK91] van Hemmen L and Kühn R 1991 Collective phenomena in neural networks *Models of Neural Networks (Physics of Neural Networks)* ed E Domany, L v Hemmen and R Schulte (Berlin: Springer)

[Ho82] Hopfield J J 1982 Neural networks and physical systems with emergent collective computational abilities *Proc. Natl Acad. Sci. USA* **79** 2554–8

[L94] Loukianova D 1994 Capacité de mémoire dans le modèle de Hopfield *C. R. Acad. Sci. Paris* **318** 157–60

[L97] Loukianova D 1997 Lower bounds on the restitution error in the Hopfield model *Probab. Theory Rel. Fields* **107** 161–76

[Lö98] Löwe M 1998 On the storage capacity of Hopfield models with weakly correlated patterns *Ann. Appl. Probab.* **8** 1216–50

[Lö99] Löwe M 1999 On the storage capacity of the Hopfield model with biase patterns *IEEE Trans. Inf. Theory* **45** 314–8

[MPRV87] McEliece R, Posner E and Venkatesh E Rodemich S 1987 The capacity of the Hopfield associative memory *IEEE Inf. Theory* **33** 461–82

[M96] Martinez S 1996 Introduction to neural networks. Storage capacity and optimization *Disordered Systems Proce. the Summer School on Dynamical Systems and Frustrated Systems (Temuco, Chile, Dec. 30, 1991–Jan. 24, 1992) (Trav. Cours.* vol 53*)* ed Bamon Rodrigo *et al* (Paris: Hermann) pp 113–33

[N88] Newman C 1988 Memory capacity in neural networks *Neural Netw.* **1** 223–38

[P96] Petritis D 1996 Thermodynamic formalism of neural computing *Dynamics of Complex Interacting Systems. (Nonlinear Phenom. Comples Sys.* vol 2*)* ed E Goles *et al* (Dordrecht: Kluwer) pp 81–146

[ST03] Shcherbina S and Tirozzi B 2003 Rigorous solution of the Gardner problem *Commun. Math. Phys.* **234** 383–422

[T95] Talagrand M 1995 Résultats rigoureux pour le modèle de Hopfield *C. R. Acad. Sci. Paris* I **321** 309–12

[T98] Talagrand M 1998 Rigorous results of the Hopfield model with many patterns *Probab. Theory Rel. Fields* **110** 177–276

[V94] Vermet F 1994 Étude asymptotique d'un réseau neuronal: le modèle de mémoire associative de Hopfield *Thèse de l'Université de Rennes 1*