Metastable states in the hierarchical Dyson model drive parallel processing in the hierarchical Hopfield network

Elena Agliari\textsuperscript{1}, Adriano Barra\textsuperscript{1}, Andrea Galluzzi\textsuperscript{2}, Francesco Guerra\textsuperscript{1}, Daniele Tantari\textsuperscript{2} and Flavia Tavani\textsuperscript{3}

\textsuperscript{1} Dipartimento di Fisica, Sapienza Università di Roma, Roma, Italy
\textsuperscript{2} Dipartimento di Matematica, Sapienza Università di Roma, Roma, Italy
\textsuperscript{3} Dipartimento di Scienze di base e applicate all’Ingegneria, Sapienza Università di Roma, Roma, Italy

E-mail: adriano.barra@roma1.infn.it

Received 23 July 2014, revised 9 October 2014
Accepted for publication 10 October 2014
Published 2 December 2014

Abstract

In this paper, we introduce and investigate the statistical mechanics of hierarchical neural networks. First, we approach these systems à la Mattis, by thinking of the Dyson model as a single-pattern hierarchical neural network. We also discuss the stability of different retrievable states as predicted by the related self-consistencies obtained both from a mean-field bound and from a bound that bypasses the mean-field limitation. The latter is worked out by properly reabsorbing the magnetization fluctuations related to higher levels of the hierarchy into effective fields for the lower levels. Remarkably, mixing Amit’s ansatz technique for selecting candidate-retrievable states with the interpolation procedure for solving for the free energy of these states, we prove that, due to gauge symmetry, the Dyson model accomplishes both serial and parallel processing. We extend this scenario to multiple stored patterns by implementing the Hebb prescription for learning within the couplings. This results in Hopfield-like networks constrained on a hierarchical topology, for which, by restricting to the low-storage regime where the number of patterns grows at its most logarithmic with the amount of neurons, we prove the existence of the thermodynamic limit for the free energy, and we give an explicit expression of its mean-field bound and of its related improved bound. We studied the resulting self-consistencies for the Mattis magnetizations, which act as order parameters, are studied and the stability of solutions is analyzed to get a picture of the overall retrieval capabilities of the system according to both mean-field and non-mean-field scenarios. Our main finding is that embedding the Hebbian rule on a hierarchical topology allows the
network to accomplish both serial and parallel processing. By tuning the level of fast noise affecting it or triggering the decay of the interactions with the distance among neurons, the system may switch from sequential retrieval to multitasking features, and vice versa. However, since these multitasking capabilities are basically due to the vanishing ‘dialogue’ between spins at long distance, this effective penury of links strongly penalizes the network’s capacity, with results bounded by low storage.

Keywords: neural networks, hierarchical models, interpolation techniques
PACS numbers: 05.20.-y, 07.05.Mh

(Some figures may appear in colour only in the online journal)

Introduction

Neural networks are a fascinating field of science that attract an incredibly large variety of scientists, including engineers who study electronics and robotics [1, 2], physicists who research statistical mechanics and stochastic processes [3, 4], mathematicians who work on learning algorithms and graph theory [5, 6], (neuro)biologists [7, 8], and (cognitive) psychologists [9, 10].

Tracing the genesis and evolution of neural networks is very difficult, probably due to the broad meaning they have acquired over the years. Scientists closer to the robotics branch often refer to either the McCulloch and Pitts model of perceptron [14] or the Rosenblatt version [15], while researchers closer to the neurobiology branch usually adopt the work of Hebb as a starting point [16].

On the other hand, scientists involved in statistical mechanics, which joined this field relatively recently, after a satisfactory picture of spin glasses was achieved [17, 18] in the 1980s, usually refer to either the seminal paper by Hopfield [19] or the celebrated work by Amit, Gutfreund, and Sompolinsky [20], where the statistical mechanical analysis of the Hopfield model is effectively carried out.

Confining ourselves to this last perspective and in a streamlined synthesis, the Hopfield model is a mean-field model where neurons are mimicked by binary (Ising) spins, whose possible states represent firing or quiescence, respectively [3, 5], and which interact pairwise via the Hebb prescription. This model acts as the harmonic oscillator for serial processing; once the system is allowed to relax, it spontaneously retrieves one of the stored patterns in suitable regions of the tunable parameters (e.g., low noise level and not-too-high storage load). The pattern retrieval depends on, for example, the initial state of the system. Recently, a generalization of this paradigm (i.e., the multitasking associative network [21]), appeared as a candidate mean-field network than can spontaneously perform parallel retrieval [22–27], meaning that it can retrieve more patterns at once without falling into spurious states [28, 29].

While these two networks perform in a fundamentally different way (serial versus parallel), they share the same mean-field statistical mechanics approximation. Each neuron interacts with all the others it is linked to with the same strength, unaware of any underlying topology, and independently of the actual pairwise distance among the neurons themselves.

4 Seminal ideas regarding automation can be found in the works of Lee from the 18th century, and all the way back to Descartes. More modern ideas regarding spontaneous cognition can be attributed to Turing [11], Von Neumann [12], and to the joint efforts of Minsky and Papert [13], just to name a few.
This limitation has always been considered as something that should be removed as soon as mathematical improvements to the available techniques would allow. Far from artificial intelligence, but related to the task of bypassing mean-field limitations, there is currently renewed interest in hierarchical models, and specifically in models where closer spins result in stronger links (see figure 1). Recently, Dyson’s pioneering work [30], where the hierarchical ferromagnet was introduced and its phase transition (splitting an ergodic region from a ferromagnetic one) was rigorously proven, has been extended to investigate extensions to spin-glasses [31, 33–38]. Although an analytical solution is still not available, researchers have made significant steps toward deep comprehension of hierarchical statistical mechanics [32, 39–45].

In this paper, we aim to analyze hierarchical neural networks in detail and we start this task by considering the statistical mechanics of the Dyson model from a novel perspective: we investigate its metastabilities.

In section 1, we deal with Dyson’s model. We introduce fundamental definitions, and then in the first subsection, we prove the existence of the thermodynamic limit of its related free energy within the spirit of the classical Guerra–Toninelli scheme [47]. In the following subsection, we explore the mean-field picture by mixing Amit’s ansatz technique [3] with our interpolation schemes [48]; the resulting technique allows us to think of the Dyson model as a single-pattern associative network, just as the Curie–Weiss (CW) plays its mean-field counterpart, thanks to the Mattis gauge [5]. A satisfactory picture of its related thermodynamic and retrieval capabilities is obtained and discussed. Remarkably, the intrinsic richness of (effectively) possible states in the Dyson model drives the system from serial processing to parallel processing, and yields the breakdown of the self-average for the order parameters. Note that parallel processing may appear strange for a one-pattern neural network. However, due to gauge-symmetry, there are actually two stored patterns, so if half of the network spins retrieve the original pattern and the other half retrieve its gauged version, it can be seen as a multitasking feature, as will become obvious when investigating the hierarchical Hopfield model.

Figure 1. Schematic representation of the hierarchical topology where the associative network insists. Green spots represent Ising neurons (N = 16 in this snapshot) while links are drawn with different thicknesses, mimicking various interaction strengths: the thicker the line, the stronger the link.
Subsection 1.3 follows the same pattern as subsection 1.2, but bypasses the mean-field limitation. In spite of the fact that we are unable to completely solve the statistical mechanics of this model, thanks to a new interpolation scheme developed in [39], we can partially account for the order parameter’s fluctuations level by level (of the hierarchy). The idea is to leverage the hierarchical structure of the model in order to account for these fluctuations. The latter are reabsorbed recursively, level by level, into an effective Hamiltonian for the underlying block spins whose thermodynamics remain solvable, thus improving the mean-field result. The only difference between these two scenarios is a different critical noise for ergodicity breaking, but the serial and parallel retrieval capabilities (namely, the existence of pure and metastable states) are preserved in both cases. Following the Dyson analysis, in section 2 we introduce the real hierarchical neural network: a hierarchical network with Hebbian couplings or, equivalently, the Hopfield model in a hierarchical setting. This system is studied in the low-storage regime, which is where the amount of stored patterns scales at its most logarithmical with contained in the amount of neurons contained in network. For this model, we first prove the existence of its free energy’s thermodynamic limit (subsection 2.1), then we move toward a mean-field (MF) scenario (subsection 2.2). We further we investigate the non-mean-field (NMF) scenario (subsection 2.3). In both cases, the model has an extremely rich phase diagram, where beyond standard serial retrieval (which we also accomplish), a number of parallel states suddenly appears by properly tuning the level of (fast) noise affecting the network. A discussion of these states and their stability analysis is included in section 2, while a discussion regarding the network’s capacity can be found in the conclusions section, which closes the paper.

1. Analysis of the Dyson hierarchical model

The Dyson hierarchical model (DHM) is a system composed at the microscopic level by $2^{k+1}$ Ising spins, $S_i = \pm 1$, with $i = 1, \ldots, 2^{k+1}$ embedded in a hierarchical topology. The Hamiltonian capturing the model is recursively introduced by the following definition.

**Definition 1.** The Hamiltonian of the DHM is defined by

$$H_{k+1}(\vec{S} | J, \sigma) = H_k(\vec{\bar{S}}_1) + H_k(\vec{\bar{S}}_2) - \frac{J}{2^{2k+1}} \sum_{i < j} S_i S_j,$$

where $J > 0$ and $\sigma \in (1/2, 1)$ are numbers tuning the interaction strength. Clearly $\bar{S}_1 \equiv \{S_i\}_{1 < i < 2^k}$, $\bar{S}_2 \equiv \{S_i\}_{2^k+1 < i < 2^{k+1}}$, and $H_0[\bar{S}] = 0$.

Thus, in this model, $\sigma$ triggers the decay of the interaction with the distance among spins, while $J$ uniformly rules the overall intensity of the couplings.

Note that this is explicitly a non-mean-field model, as the distance, $d_{ij}$, between two spins, $i, j$, ranges between 0 and $k$ (see figure 1). Indeed, it is possible to rewrite the Hamiltonian (1) in terms of the $d_{ij}$ as

$$H_k[\{S_1 \ldots S_{2^k}\}] = -\sum_{i < j} S_i S_j J_{ij}$$

(2)
\[ J_{ij} = \sum_{l=di}^{k} \left( \frac{J}{2^{l+l'}} \right) = J \left( d_{ij}, k, \sigma, J \right) = J \frac{4^{\sigma-d_{ij}} - 4^{-k_{a}}}{4^{k_{a}} - 1}. \] (3)

Once the Hamiltonian is given (in this paper we will refer mainly to the form (1)), it is possible to introduce the partition function, \( Z_{k+1}(\beta, J, \sigma) \), at finite volume \( k + 1 \) as

\[ Z_{k+1}(\beta, J, \sigma) = \sum_{S} \exp \left[ -\beta H_{k+1}(\vec{S}, J, \sigma) \right]. \] (4)

and the related free energy, \( f_{k+1}(\beta, J, \sigma, h) \), namely the intensive logarithm of the partition function, as

\[ f_{k+1}(\beta, J, \sigma) = \frac{1}{2^{k+1}} \log \sum_{S} \exp \left[ -\beta H_{k+1}(\vec{S}) + h \sum_{i=1}^{2^{k+1}} S_{i} \right]. \] (5)

We are interested in an explicit expression of the infinite volume limit of the intensive free energy, defined as

\[ f(\beta, J, \sigma) = \lim_{k \to \infty} f_{k+1}(\beta, J, \sigma), \]

in terms of suitably introduced magnetizations, \( m \), that act as order parameters for the theory; in order to satisfy thermodynamic prescriptions, we want to find the free energy minima\(^{5}\) w.r.t. these order parameters. To achieve this goal, we introduce the global magnetization, \( m \), defined as the limit \( m = \lim_{k \to \infty} m_{k+1} \), where

\[ m_{k+1} = \frac{1}{2^{k+1}} \sum_{i} S_{i}. \] (6)

Recursively, and with a little abuse of notation, the \( k \) magnetizations \( m_{1}, \ldots, m_{k} \) level by level (over \( k \) levels and starting to define them from the largest bulk), as the same \( k \to \infty \) limit of the following quantities (we explicitly write only the two upper magnetizations related to the two main clusters that the system reduces to whenever \( J \to 0 \), as seen in figure 1):

\[ m_{\text{left}} = \frac{1}{2^{k}} \sum_{i=1}^{2^{k}} S_{i}, \quad m_{\text{right}} = \frac{1}{2^{k}} \sum_{i=2^{k}+1}^{2^{k+1}} S_{i}. \] (7)

As a last point, thermodynamical averages will be denoted by the brackets \( \langle \cdot \rangle \), such that

\[ \langle m_{k+1}(\beta, J, \sigma) \rangle = \frac{\sum_{S} m_{k+1} e^{-\beta H_{k+1}(\vec{S}, J, \sigma)}}{Z_{k+1}(\beta, J, \sigma)}, \] (8)

and clearly \( \langle m(\beta, J, \sigma) \rangle = \lim_{k \to \infty} \langle m_{k+1}(\beta, J, \sigma) \rangle \).

1.1. The thermodynamic limit

This section aims to prove the existence of the thermodynamic limit for the free energy of the DHM. In spite of the fact that this result was achieved a long time ago by Gallavotti and Miracle-Sole [49], we exploit a different interpolating scheme with the pedagogical aim of highlighting the technique more than the result, as the technique will then be used to prove the

\(^{5}\) Note that as the free energy, strictly speaking, is \( f(\beta) = -\alpha(\beta) \), we actually look for maxima throughout this paper.
existence of the thermodynamic limit for the hierarchical Hopfield network. The main idea is that, since the interaction is ferromagnetic, the free energy is monotone in $k$, with the introduction of new levels of positive interactions.

**Theorem 1.** The thermodynamic limit of the DHM free energy does exist, and we call
\[
\lim_{k \to \infty} f_{k+1}(\beta, J, \sigma) = f(\beta, J, \sigma).
\]

To prove this statement, let us introduce a real scalar parameter, $t \in [0, 1]$, and the following interpolating function
\[
\varphi_{k+1,t}(\beta) = \frac{1}{2^{k+1}} \log \sum_{\vec{S}} \exp \left( \beta \left( -H_k(\vec{S}_1) - H_k(\vec{S}_2) \right) + \frac{J}{2} \gamma^{(k+1)(1-2\sigma)} m_{k+1}^2(\vec{S}) \right),
\]
with $m_{k+1} = \frac{1}{2^{k+1}} \sum_{l=1}^{2^{k+1}} S_l$, such that
\[
\varphi_{k+1,1} = f_{k+1},
\]
\[
\varphi_{k+1,0} = f_k,
\]
and
\[
0 \leq \frac{d\varphi_{k+1,t}}{dt} = \left( \beta - \frac{1}{2^{k+1}} \frac{\gamma^{(k+1)(1-2\sigma)} J}{2} m_{k+1}^2(\vec{S}) \right) \leq \frac{\beta J \gamma^{(k+1)(1-2\sigma)}}{2}.
\]

Since
\[
\varphi_{k+1,1}(h) = \varphi_{k+1,0}(h) + \int_0^1 \frac{d\varphi_{k+1,t}}{dt} dt,
\]
$f_{k+1} \geq f_k$ (the sequence is non decreasing), thus
\[
f_{k+1}(\beta, J, \sigma) \leq f_k(\beta, J, \sigma) + \frac{\beta J}{2} \gamma^{(k+1)(1-2\sigma)}.
\]

Iterating this argument over the levels, we obtain
\[
f_{k+1}(\beta, J, \sigma) \leq f_0(\beta, J, \sigma) + \frac{\beta J}{2} \sum_{l=1}^{k+1} 2^{l(1-2\sigma)}.
\]

In the limit of $k \to \infty$
\[
f \leq f_0 + \frac{\beta J}{2} \sum_{l=1}^{\infty} 2^{l(1-2\sigma)}.
\]

The series on the right of the preceding inequality converges, since $\sigma > \frac{1}{2}$. Hence,
\[
f(\beta, J, \sigma) \leq f_0(\beta, J, \sigma) + \frac{\beta J}{2} \frac{1}{1 - 2^{2(\sigma-1)}}.
\]

The sequence $f_k(\beta, J, \sigma)$ is bounded and non decreasing, so it admits a well-defined limit for $k \to \infty$. 
1.2. The mean-field scenario

This section aims to turn around classical results [30, 49, 50] and investigate metastabilities in the Dyson model at the mean-field level. To this end, two schemes must be merged. We start by following [39] to build an interpolating iterative scheme that returns the mean-field free energy in terms of a bound, and then we implement the Amit method of ansatz to evaluate, within the free energy landscape obtained by this interpolation, the stability and thermodynamic importance of two test-states: the (standard) ferromagnetic state (with all the spin aligned, \( m_{\text{left}} = m_{\text{right}} \)) and the simplest metastable state, which is a state where all the left spins (that is, the first 1, \( \ldots \), \( 2^k \) spins) are aligned with each other and opposite to the right spins (that is, the remaining \( 2^k + 1, \ldots , 2^{k+1} \) spins). The right spins are also aligned with each other, and hence, \( m_{\text{left}} = -m_{\text{right}} \). Operatively, we state definition 2.

**Definition 2.** Once considered a real scalar parameter, \( t \in [0, 1] \), we introduce the following interpolating Hamiltonian

\[
H_{k+1,t}(\vec{S}) = -\frac{J t}{2^{2k+1}} \sum_{i>j=1}^{2^{k+1}} S_i S_j - (1-t) \sum_{i=1}^{2^{k+1}} S_i + H_k(\vec{S}_1) + H_k(\vec{S}_2),
\]

such that for \( t = 1 \), the original system is recovered, while at \( t = 0 \), the two-body interaction is replaced by an effective but tractable one-body term. The possible presence of an external magnetic field can be accounted for simply by adding to the Hamiltonian a term \( \sigma \propto \sum_i h_i^2 \), with \( h \in \mathbb{R} \).

This prescription allows us to define an extended partition function as

\[
Z_{k+1,t}(h, \beta, J, \sigma) = \sum \exp \left\{ -\beta \left[ H_{k+1,t}(\vec{S}) + h \sum_{i=1}^{2^{k+1}} S_i \right] \right\},
\]

where the subscript \( t \) stresses its interpolative nature, and, analogously,

\[
\Phi_{k+1,t}(h, \beta, J, \sigma) = \frac{1}{2^{2k+1}} \log Z_{k+1,t}(h, \beta, J, \sigma).
\]

Since

\[
\Phi_{k+1,0}(h, \beta, J, \sigma) = \Phi_{k,1}(h + m J 2^{(k+1)(1-2\sigma)}, \beta, J, \sigma),
\]

as shown in [39], (discarding the dependence of \( \Phi \) by \( \beta, J, \sigma \) for simplicity) through a long but straightforward calculation, we arrive at

\[
\Phi_{k+1,1}(h) = \Phi_{k+1,0}(h) - \frac{\beta J}{2} \left( 2^{(k+1)(1-2\sigma)} m^2 + 2^{-2(k+1)\sigma} \right)
\]

\[
+ \frac{\beta J}{2} 2^{(k+1)(1-2\sigma)} \left( m_{k+1}(\vec{S}) - m \right)^2,
\]

\[
\geq \Phi_{k,1}(h + J m 2^{(k+1)(1-2\sigma)}) - \frac{\beta J}{2} \left( 2^{(k+1)(1-2\sigma)} m^2 + 2^{-2(k+1)\sigma} \right).
\]

Note that, in the last passage, we neglected, level by level, the source of the order parameter’s fluctuations \( \left( m_{k+1}(\vec{S}) - m \right)^2 \), which is positive definite, and thus we obtained a bound for the free energy.
For the sake of simplicity, we extended the meaning of the brackets to account for the interpolating structure coded in the Boltzmannfaktor of equation (18), by adding a subscript \( t \) to them: namely \( \langle \cdot \rangle \rightarrow \langle \cdot \rangle_t \).

In order to start investigating non standard stabilities, we must also note that \( \Phi_{k+1,0}(h) = \Phi_{k,1}^1(h + mJ 2^{(k+1)(1-2\sigma)}) \), but in principle we can also have two different contributions from the two groups of \( 2^k \) spins left and right). Thus, we should write more generally

\[
\Phi_{k+1,0}(h) = \frac{1}{2} \left[ \Phi_{k,1}^1 \left( h + mJ 2^{(k+1)(1-2\sigma)} \right) + \Phi_{k,1}^2 \left( h + mJ 2^{(k+1)(1-2\sigma)} \right) \right].
\]  

(22)

Now let us assume the Amit perspective [3] and suppose that these two subsystems have different magnetizations, \( m_{\text{left}} = m_1 \) and \( m_{\text{right}} = m_2 \), which are equal in modulus but opposite in sign (i.e., \( m_1 = -m_2 \)). This observation implies that, starting from the \( k \)th level, we can iterate the interpolating procedure in parallel on the two clusters using, respectively, \( m_1 \) and \( m_2 \) as trial parameters. Via this route, we obtain

\[
\begin{align*}
g_k'(h, \beta, J, \sigma) & \geq \log 2 + \frac{1}{2} \left\{ \log \cosh \left[ \beta h + \beta J \left( m_1 \sum_{l=1}^{k} 2^l(1-2\sigma) \right) \right] \right\} + \frac{1}{2} \left( \log \cosh |\beta h \right) \\
& \quad + \frac{\beta J}{2} \left( m_2 \sum_{l=1}^{k} 2^l(1-2\sigma) + 2^{(k+1)(1-2\sigma)} \right) \\
& \quad - \frac{\beta J}{2} \left( \sum_{l=1}^{k+1} 2^{-2\sigma} \right) \\
& \quad - \frac{\beta J}{2} \left( \frac{m_1^2 + m_2^2}{2} \right) \\
& = f(k, m_1, m_2; h, \beta, J, \sigma).
\end{align*}
\]  

(23)

Therefore, we have that \( g_k'(h, \beta, J, \sigma) \geq \sup_{m_1, m_2} f(k, m_1, m_2; h, \beta, J, \sigma) \), and we need to evaluate the optimal order parameters in order to have the best estimate of the free energy.
Taking the derivatives of the free energy with respect to $m, m_1,$ and $m_2,$ we obtain the self-consistent equations holding at the extremal points of $f(k, m, m_1, m_2|h, \beta, J, \sigma),$ which read as

\[
\begin{align*}
    m_1 &= \tanh \left[ \beta h + \beta J \left( m_1 \sum_{l=1}^{k} 2^{(1-2\sigma)l} + 2^{(k+1)(1-2\sigma)m} \right) \right], \\
    m_2 &= \tanh \left[ \beta h + \beta J \left( m_2 \sum_{l=1}^{k} 2^{(1-2\sigma)l} + 2^{(k+1)(1-2\sigma)m} \right) \right], \\
    m &= \frac{m_1 + m_2}{2},
\end{align*}
\]

where the third equation is only a linear combination of $m_1$ and $m_2$ that simply states that the global magnetization is the average of the magnetizations of the two main clusters.

It is easy to see that, at zero external field $h = 0$, the pure solution $m_1 = m_2 = m = m_P$, where the whole system has a non zero magnetization, and the antiparallel (metastable) solution $m_A = m_1 = -m_2$ and $m = 0$, where the system has two clusters with opposite magnetizations and no global magnetization, both exist.

Clearly, according to the value of the temperature, we can have a paramagnetic solution ($m_P = m_A = 0$), or two gauge-symmetric solutions for each of the two possible states ($\pm m_P, \pm m_A$); we therefore need to analyze the stability of these solutions, to determine if they are maxima or minima of $f(h = 0, \beta, J, \sigma)$.

Obtaining an explicit expression for the second derivatives to build the Hessian, $H(m_1, m_2)$, of $f(h = 0, \beta, J, \sigma)$ is a rather lengthy process, yet it is easy to see that the entries of the Hessian actually depend on $m_1^2$ and $m_2^2$ only; thus, they are independent of the sign of the two magnetizations, $m_1, m_2$. This means that, as the paramagnetic solution becomes unstable, both the pure (i.e., $m_{\text{left}} = m_{\text{right}}$) and antiparallel (mixture, i.e., $m_{\text{left}} = -m_{\text{right}}$) solutions become stable, which ensures the possibility of taking the thermodynamic limit in (24), and sheds light on the breaking of standard self-averaging [50].

In this case, we get the following theorem.

**Theorem 2.** The mean-field bound for the DHM free energy associated with the metastable state reads as

\[
f(h, \beta, J, \sigma) \geq \sup_{m_1, m_2, m} \lim_{k \to \infty} f(k, m, m_1, m_2) = \sup_{m_1, m_2} \log 2 + \frac{1}{2} \log \cosh \left( \beta h + \beta J C_{2\sigma-1} m_1 \right) + \frac{1}{2} \log \cosh \left( \beta h + \beta J C_{2\sigma-1} m_2 \right) - \frac{\beta J C_{2\sigma-1}}{2} \left( m_1^2 + m_2^2 \right),
\]

where $C_{\sigma} = \frac{2^{2\sigma}}{2^{2\sigma} - 1}$ and the trial parameters $m_1$, $m_2$ fulfill the self-consistencies (27) and $m$ is its symmetric linear combination.

The mean field bound for the DHM free energy associated with the ferromagnetic state can be obtained again simply by identifying $m_1 = m_2 = m$, and it reads as
In the thermodynamic limit, the last level of interaction (the largest in terms of the number of links, but the weakest in terms of their intensity) that would tend to keep \( m_1 \) and \( m_2 \) aligned vanishes. Thus, the system effectively behaves just as the sum of two non-interacting subsystems with independent magnetizations (see figure 2), satisfying the following proposition.

**Proposition 1.** The mixture state of the DHM has two independent order parameters, one for each larger cluster, whose self-consistencies read as

\[
m_{1,2} = \tanh(\beta h + \beta J C_{2\sigma-1} m_{1,2})
\]

(27)

If we want to determine the critical value, \( \beta_c \), that breaks ergodicity, we can expand them for \( k \to \infty \), and for \( h = 0 \), hence obtaining, in the limit \( m_{1,2} \to 0 \):

\[
\begin{align*}
  m_1 &\sim \beta J m_1 \frac{2^{1-2\sigma}}{1 - 2^{1-2\sigma}} + \mathcal{O}(m_1^3), \\
  m_2 &\sim \beta J m_2 \frac{2^{1-2\sigma}}{1 - 2^{1-2\sigma}} + \mathcal{O}(m_2^3),
\end{align*}
\]

such that we can write corollary 1.
Corollary 1. Mean-field criticality in the DHM has the classical critical exponent one half and critical temperature $\beta_c^{MF}$, given by

$$\beta_c^{MF} = 1 - \frac{2^{1-2\sigma}}{J 2^{1-2\sigma}}. \quad (28)$$

It is worth noticing, however, that generally, the mean-field picture does not hold in this hierarchical setting, despite the fact that it correctly reproduces the network behavior (even the critical exponent, one-half) as long as $\sigma < 3/4$ [32, 33].

One may still debate, however, that while the intensity of the upper links is negligible, it may still collapse the state of one cluster to the other, thus destroying metastability. This can happen, for instance, when we use a vanishing external field in a critical mean-field ferromagnet to select the phase by hand. In the appendix, we give a detailed explanation and a rigorous proof that this is not the case here. The DHM has links that are too evanescent to drive all the spins so that they always converge to the same sign and mixture states are preserved.

1.3. The not-mean-field scenario

The scope of this section bypasses mean-field limitations and shows that the outlined scenario is robust, even beyond the mean-field picture. We stress that we do not have a rigorous solution of the free energy, but rather a more stringent (with respect to the mean-field counterpart) analytical bound supported by extensive numerical simulations. In particular, we exploit the interpolative technology introduced in [39] to take into account at least part of the fluctuations of the order parameters, thus improving the previous description. In models beyond the mean field, the magnetization is no longer self-averaging and its fluctuations cannot be neglected. It is indeed the proliferation of these metastable states that both avoids the collapse of the order parameter probability distribution on a Dirac delta and breaks the self-averaging definition.

Let us start investigating the improved bound with the following definition.

**Definition 3.** Once we introduce two suitable real parameters, $t, x$, the interpolating Hamiltonian that we will consider to bypass the mean-field bound has the form

$$H_{k+1}(\vec{S}) = -tu(\vec{S}) - (1 - t)v(\vec{S}) + H_k(\vec{S}_1) + H_k(\vec{S}_2). \quad (29)$$

with

$$u(\vec{S}) = -\frac{J}{2^{2n(k+1)}} \sum_{i \neq j = 1}^{2^{k+1}} S_i S_j + \frac{xJ}{2 \cdot 2^{2n(k+1)}} \sum_{i,j=1}^{2^{k+1}} (S_i - m)(S_j - m).$$

$$v(\vec{S}) = \frac{J (1 + x)}{2 \cdot 2^{2n(k+1)}} \left[ \sum_{i,j=1}^{2^{k+1}} (S_i - m)(S_j - m) + \sum_{i,j=2^{k+1}+1}^{2^{k+1}} (S_i - m)(S_j - m) \right]$$

$$+ mJ 2^{(k+1)(1-2\sigma)} \sum_{j=1}^{2^{k+1}} S_j,$$

where $x \geq 0$ accounts for fluctuation resorption and $0 \leq t \leq 1$ plays as before.
The associated partition function and free energy are, respectively,

\[
Z_{k+1,\sigma}(x, h) = \sum_S \exp \left\{ -\beta \left[ h S + \frac{1}{2^k} \sum_{i=1}^{2^k+1} S_i \right] \right\},
\]

(30)

\[
\Phi_{k+1,\sigma}(x, h) = \frac{1}{2^{k+1}} \log Z_{k+1,\sigma}(x, h).
\]

(31)

The procedure that yields to the non-mean-field bound for the free energy allows us to obtain (see [39]) the following expression for the pure ferromagnetic case, where we again omitted the dependence by \(\beta, J, \sigma\) for simplicity,

\[
f_{k+1}(h, \beta, J, \sigma) \geq \Phi_{k,1} \left( \frac{1}{2^{2k}}, h + m 2^{(k+1)(1-2\sigma)} \right) - \frac{\beta J}{2} \left( 2^{(k+1)(1-2\sigma)} m^2 + 2^{-2\sigma(k+1)} \right).
\]

(32)

However, as shown for the preceding bound, let us now suppose that the system is split in two parts with two different magnetizations, \(m_{\text{left}} = m_1\) and \(m_{\text{right}} = m_2\). Resuming the same lines of reasoning found in the previous section of the paper, we obtain

\[
\Phi_{k,1} \left( \frac{1}{2^{2k}}, h + m 2^{(k+1)(1-2\sigma)} \right) = \frac{1}{2} \Phi_{k,1} \left( \frac{1}{2^{2k}}, h + m 2^{(k+1)(1-2\sigma)} \right)
\]

\[
+ \frac{1}{2} \Phi_{k,1} \left( \frac{1}{2^{2k}}, h + m 2^{(k+1)(1-2\sigma)} \right).
\]

(33)

From this point, we can iterate the previous scheme point by point up to the last level of the hierarchy, using as trial order parameter \(m_{1,2}\) for \(\Phi_{1,2}\), respectively. As a consequence, formula (32), which is derived within the ansatz of a pure ferromagnetic state, is generalized by the following expression

\[
f_{k+1}(h, \beta, J, \sigma) \geq \frac{1}{2} \Phi_{0,1} \left( \sum_{i=1}^{k+1} 2^{-2i\sigma}, h + J m_1 \sum_{i=1}^k 2^{(1-2\sigma)} + m J 2^{(k+1)(1-2\sigma)} \right)
\]

\[
+ \frac{1}{2} \Phi_{0,1} \left( \sum_{i=1}^{k+1} 2^{-2i\sigma}, h + J m_2 \sum_{i=1}^k 2^{(1-2\sigma)} + m J 2^{(k+1)(1-2\sigma)} \right)
\]

\[
- \frac{\beta J}{2} \sum_{i=1}^{k+1} 2^{(1-2\sigma)} \left( \frac{m_1^2 + m_2^2}{2} \right) - \frac{\beta J}{2} \sum_{i=1}^{k+1} 2^{-2i\sigma} m^2.
\]

(34)

An explicit representation for \(\Phi_{0,1}\) reads as

\[
\Phi_{0,1} \left( \sum_{i=1}^{k+1} 2^{-2i\sigma}, h + J m_1 \sum_{i=1}^k 2^{(1-2\sigma)} + m J 2^{(k+1)(1-2\sigma)} \right)
\]

\[
= \ln 2 + \frac{\beta J}{2} \left( 1 + m_1^2 \right) \sum_{i=1}^{k+1} 2^{-2i\sigma} + \log \cosh \left\{ \beta h + \beta m J 2^{(k+1)(1-2\sigma)} \right\}
\]

\[
+ \beta m_1 J \left( \sum_{i=1}^{k} 2^{(1-2\sigma)} - \sum_{i=1}^{k+1} 2^{-2i\sigma} \right).
\]

(35)
in such a way that

$$f_{k+1} \geq \log 2 + \frac{1}{2} \log \cosh \left\{ \beta h + \beta m J 2^{(k+1)(1-2\sigma)} \right\}$$

\[+ \beta m J \left[ \sum_{l=1}^{k} 2^{l(1-2\sigma)} - \sum_{l=1}^{k+1} 2^{-2\sigma} \right] \]

\[+ \frac{1}{2} \log \cosh \left\{ \beta h + \beta m J 2^{(k+1)(1-2\sigma)} \right\} \]

\[- \frac{\beta J}{2} \left[ \sum_{l=1}^{k} 2^{l(1-2\sigma)} - \sum_{l=1}^{k+1} 2^{-2\sigma} \right] \left( m_1^2 + m_2^2 \right) \]

\[- \frac{\beta J}{2} 2^{(k+1)(1-2\sigma)} m^2. \]  \quad (36)

In summary, in the thermodynamic limit, one has the following.

**Theorem 3.** The non-mean-field bound for the DHM’s free energy associated to the mixture state reads as

$$f(h, \beta, J, \sigma) \geq \sup_{m_1, m_2} \left\{ \log 2 + \frac{1}{2} \log \cosh \left[ \beta h + \beta m J (C_{2\sigma-1} - C_{2\sigma}) \right] \right\}$$

\[+ \frac{1}{2} \log \cosh \left[ \beta h + \beta m J (C_{2\sigma-1} - C_{2\sigma}) \right] \]

\[- \frac{\beta J}{2} (C_{2\sigma-1} - C_{2\sigma}) \left( m_1^2 + m_2^2 \right) \} , \] \quad (37)

where $C_\gamma = 2^{-\gamma}$, and the trial parameters $m_1, m_2$ respect the self-consistencies that we will outline in proposition 2. If we assume that the system lives within a pure state, by identifying $m_1 = m_2 = m$, we again find the non-mean-field bound shown in [39], which is

$$f(h, \beta, J, \sigma) \geq \sup_{m} \left\{ \log 2 + \log \cosh \left[ \beta h + \beta m J (C_{2\sigma-1} - C_{2\sigma}) \right] \right\}$$

\[- \frac{\beta J}{2} (C_{2\sigma-1} - C_{2\sigma}) m^2 \} . \] \quad (38)

Imposing thermodynamic stability. We obtain the following proposition.

**Proposition 2.** Even beyond the mean-field level of description, the mixture state of the DHM is described by two independent order parameters, one for each larger cluster, whose self-consistencies read as

$$m_{1,2} = \tanh \left( \beta h + \beta m J_{1,2} (C_{2\sigma-1} - C_{2\sigma}) \right) .$$ \quad (39)
As for the MF approximation, we are going to find the critical temperature, $\beta_c$. Considering the system at zero external field, $h = 0$, and thus writing

\[
\begin{align*}
  m_2 &\sim \beta J m_2 \left( \frac{1}{2^{2\sigma} - 1} - \frac{1}{2^{2\sigma} - 2^{2\sigma}} \right) + \mathcal{O}(m_2^2), \\
  m_3 &\sim \beta J m_3 \left( \frac{1}{2^{2\sigma} - 1} - \frac{1}{2^{2\sigma} - 2^{2\sigma}} \right) + \mathcal{O}(m_3^3).
\end{align*}
\]

we get the following.

**Corollary 2.** This non-mean-field criticality in the DHM has the classical exponent, but has a different critical temperature, $\beta_c^{\text{NMF}}$, given by the following formula:

\[
\beta_c^{\text{NMF}} = \frac{(2^{2\sigma} - 1)(1 - 2^{1 - 2\sigma})}{J}.
\]

Note that the non-mean-field interpolation we exploited returned classical (i.e., wrong) critical behavior. This is due to the assumption of self-averaging for the dimers lying at the lowest levels.

Comparing the values of $\beta_c^{\text{MF}}$ and $\beta_c^{\text{NMF}}$, we get the following bound

\[
\beta_c^{\text{NMF}} > \beta_c^{\text{MF}} \quad \Rightarrow \quad T_c^{\text{MF}} > T_c^{\text{NMF}}.
\]

We do not push the analysis any further here because we want to present a streamlined minimal theory, but the model admits a proliferation of metastable states, which are achievable by proceeding hierarchically with Amit’s ansatz, thus taking both the blocks built by $k^i$ spins and splitting them into sub-clusters of $2^{k-1}$ spins each, and so on. Correspondingly, the hierarchical neural network has a much richer phase diagram w.r.t. its mean-field counterpart. Further investigations can be found in [46].

2. Analysis of the Hopfield hierarchical model

As we saw in the previous section, the Dyson model has a rich variety of retrievable states, and by retrievable we mean that they are free-energy minima in the thermodynamic limit, and their basins of attraction are not negligible. Now we want to apply the previously presented analysis and the ideas that stemmed from the related findings to a hierarchical Hopfield model (HHM).

To complete this task, we need to introduce, in addition to $2^{i+1}$ dichotomic spins/neurons, $p$-quenched patterns, $\xi^\mu$, $\mu \in \{1, \ldots, p\}$, that do not participate in thermalization. These are vectors of length $2^{k+1}$, whose entries are extracted once for all from centered and symmetrical i.i.d. as

\[
P(\xi^\mu) = \frac{1}{2} \delta(\xi^\mu - 1) + \frac{1}{2} \delta(\xi^\mu + 1).
\]

Mirroring the previous section, the Hamiltonian of the HHM is also defined recursively by the following.
Definition 4. The Hamiltonian of the HHM is defined by

\[ H_{k+1}(\vec{S}) = H_k(\vec{S}) + H_k(\vec{S}_z) = \frac{1}{2^{2\mu(k+1)}} \sum_{\mu=1}^{2^{\mu+1}} \sum_{i=1}^{2^{\mu+1}} \xi_i^\mu \xi_j^\mu S_i S_j \]

with \( H_0(S) = 0; \sigma \in (1/2, 1) \) being the number tuning the interaction strength with the neuron’s distance, and \( p \) being the number of stored patterns. Accounting for the presence of external stimuli can be included simply within a one-body additional term in the Hamiltonian as \( \propto \sum_{\mu=1}^{2^{\mu+1}} \xi_i^\mu \xi_j^\mu \), and an overall survey of the stimuli is accomplished by summing over \( \mu \in (1, \ldots, p) \) all the \( h_\mu \).

Even in this context, we can again write the Hamiltonian of the HHM in terms of a distance, \( d_{ij} \), between the spin pair \((i, j)\) (see figure 1, panel B), obtaining

\[ H_k(\{S_1\ldots S_2^k\}) = \sum_{i<j} S_i S_j \left( \sum_{\mu=1}^{2^{\mu+1}} \frac{\xi_i^\mu \xi_j^\mu}{2^{\mu+1}} \right) = \sum_{i<j} S_i S_j \tilde{J}_{ij}. \]

where, keeping the previous expression (see equation (3)) to encode neuronal distance, it also holds that

\[ \tilde{J}_{ij} = \frac{4^{\sigma-d_{ij}/\sigma} - 4^{-k\sigma}}{4^\sigma - 1} \cdot \sum_{\mu=1}^{2^{\mu+1}} \xi_i^\mu \xi_j^\mu. \]

Hence, the Hebbian kernel on a hierarchical topology becomes modified by the distance-dependent weight, \( J(d_{ij}, k, \sigma) \). Before starting to implement our interpolative strategy, some definitions are in order.

Definition 5. We introduce the Mattis magnetizations (or Mattis overlaps) over the whole system as

\[ m_\mu(\vec{S}) = \frac{1}{2^{\mu+1}} \sum_{i=1}^{2^{\mu+1}} \xi_i^\mu S_i. \]

This definition can be extended trivially to the inner clusters, properly restricting the sum over the (pertinent) spins; for example, dealing with the two larger subclusters as before, we have

\[ m_\mu^{left} = \frac{1}{2^2} \sum_{i=1}^{2^2} \xi_i^\mu S_i, \quad m_\mu^{right} = \frac{1}{2^2} \sum_{j=2^2+1}^{2^{\mu+1}} \xi_j^\mu S_j. \]

2.1. The thermodynamic limit

As in the previous investigation, first we want to prove that the model is well defined, and specifically that the thermodynamic limit for the free energy exists. To complete this task, we have the following.
Theorem 4. The thermodynamic limit of the HHM’s free energy exists, and we call it 
\[ \lim_{k \to \infty} f_{k+1}(\beta, p, \sigma) = f(\beta, p, \sigma). \]

Let us write the Hamiltonian as 
\[ H_{k+1}(\vec{S}) = H_k(\vec{S}_1) + H_k(\vec{S}_2) = \frac{1}{2} 2^{(k+1)} 2^{(k+1)(1-2\nu)} \sum_{\mu=1}^{p} (m_{\mu}^{k+1}(\vec{S}))^2, \]
and let us consider the following interpolation, where again, for the sake of simplicity, we stress the dependence of the external fields \( h_{\mu} \) only and use the symbol \( E_\xi \) to denote averaging over the quenched patterns: 
\[ \Phi_{k+1}(\{ h_{\mu} \}) = \frac{1}{2^{k+1}} E_\xi \log \sum_{\vec{S}} \exp \left\{ \beta [H_k(\vec{S}_1) - H_k(\vec{S}_2)] \right\} + \frac{1}{2} 2^{(k+1)} 2^{(k+1)(1-2\nu)} \sum_{\mu=1}^{p} (m_{\mu}^{k+1}(\vec{S}))^2 + \sum_{\mu=1}^{p} h_{\mu} \xi S_{\mu} \right\}. \] (49)

We notice that 
\[ \Phi_{k+1,1}(h) = f_{k+1}, \] (50)
\[ \Phi_{k+1,0}(h) = f_k \] (51)
and that 
\[ \frac{d}{dt} \Phi_{k+1,t} = \left( \frac{1}{2} 2^{k+1} 2^{(k+1)(1-2\nu)} \sum_{\mu=1}^{p} \left( m_{\mu}^{k+1}(\vec{S}) \right)^2 \right) \geq 0 \] (52)
in such a way that 
\[ f_{k+1}(\beta, p, \sigma) \geq f_k(\beta, p, \sigma). \] Now we want to prove that \( f_{k+1}(\beta, p, \sigma) \) is bounded; it is enough to see that 
\[ f_{k+1}(\beta, p, \sigma) = f_k(\beta, p, \sigma) + \int_0^1 \frac{d}{dt} \Phi_{k+1,t}. \] (53)
Since we have 
\[ \frac{d}{dt} \Phi_{k+1,t} = \left( \beta \frac{2^{(k+1)} 2^{(k+1)(1-2\nu)}}{2} \sum_{\mu=1}^{p} \left( m_{\mu}^{k+1}(\vec{S}) \right)^2 \right) \leq \beta \mu \frac{2^{(k+1)} 2^{(k+1)(1-2\nu)}}{2}, \] (54)
we can write 
\[ f_{k+1}(\beta, p, \sigma) \leq f_k(\beta, p, \sigma) + \beta \mu \frac{2^{(k+1)(1-2\nu)}}{2}. \] (55)
Iterating this procedure over the levels, we get
\[ f_{k+1}(\beta, p, \sigma) \leq f_0(\beta, p, \sigma) + \frac{\beta p}{2} \sum_{i=1}^{k+1} 2^{i(1-2\sigma)}, \] (56)
such that, in the \( k \to \infty \) limit, we can write
\[ f \leq f_0 + \frac{\beta p}{2} \sum_{i=1}^{\infty} 2^{i(1-2\sigma)}. \]

Since \( \sigma > \frac{1}{2} \), the series on the right-hand side of the preceding inequality converges, and thus \( f(\beta, p, \sigma) \) is bounded by
\[ f(\beta, p, \sigma) \leq f_0 + \frac{\beta p}{2} \frac{1}{2^{(2\sigma-1)}} - 1 \]
and nonincreasing for (52). Thus, its thermodynamic limit exists.

**2.2. The mean-field scenario**

In this section, we will investigate the serial and parallel retrieval capabilities of the HHM at the mean-field level. As usual, we attain our goal by mixing the Amit ansatz technique in selecting suitable candidate states for retrieval with the interpolation technique.

**Definition 6.** Let us define the interpolating Hamiltonian, \( H_{k+1, t}(\vec{S}) \), as
\[
H_{k+1, t}(\vec{S}) = H_k(\vec{S}_1) + H_\mu(\vec{S}_2) - \frac{t}{2} \cdot 2^{(k+1)} \\
\times \sum_{\mu=1, i,j=1}^{2^{k+1}} \sum_{\xi_i}^{\mu} \xi_i S_i \xi_j - (1 - t) \cdot 2^{(k+1)(1-2\sigma)} \\
\times \sum_{\mu=1}^{2^{k+1}} m_\mu \sum_{i=1}^{2^{k+1}} \xi_i S_i. \] (57)

Clearly, we can associate such a Hamiltonian to an extended partition function, \( Z_{k+1, t}(h) \), and to an extended free energy, \( \Phi_{k+1, t}(h) \), as
\[
Z_{k+1, t}(\{h_\mu\}) = \sum_{\vec{S}} \exp \left\{ -\beta \left[ H_{k+1, t}(\vec{S}) + \sum_{\mu=1}^{2^{k+1}} h_\mu \sum_{i=1}^{2^{k+1}} \xi_i S_i \right] \right\}, \] (58)
\[
\Phi_{k+1, t}(\{h_\mu\}) = \frac{1}{2^{k+1}} \mathcal{F}_t \log Z_{k+1, t}(\{h_\mu\}), \] (59)
where, for the sake of simplicity, we stressed only the dependence by the fields. We can rewrite (57) as
\[ H_{k+1,t}(\vec{S}) = H_k(\vec{S}_t) + H_k(\vec{S}_s) - \frac{2^{2(k+1)t}}{2 \cdot 2^{2n(k+1)}} \times \sum_{\mu=1}^{\mathbf{p}} m_{k+1,\mu}(\vec{S}) - (1 - \gamma)2^{(k+1)2(1-2n)(k+1)} \times \sum_{\mu=1}^{\mathbf{p}} m_{k+1,\mu}(\vec{S}) m_{\mu}. \]  

(60)

It is easy to show that

\[ \Phi_{k+1,1}\left( \{ h_\mu \} \right) = f_{k+1}, \]  

(61)

\[ \Phi_{k+1,0}\left( \{ h_\mu \} \right) = \Phi_{k,1}\left( \{ h_\mu + 2^{(k+1)(1-2n)}m_\mu \} \right), \]  

(62)

and that

\[ \frac{d\Phi_{k+1,t}}{dt} = \frac{1}{2^{k+1} Z_{k+1,t}} \sum_{\vec{S}} \exp \left( -\beta \left( H_{k+1,t}(\vec{S}) \right) \right) \times \sum_{\mu=1}^{\mathbf{p}} h_\mu \sum_{i=1}^{2^{k+1}} \xi_i S_i \left( -\beta \frac{dH_{k+1,t}(\vec{S})}{dt} \right) \]

\[ = \frac{1}{2^{k+1} Z_{k+1,t}} \sum_{\vec{S}} \exp \left( -\beta \left( H_{k+1,t}(\vec{S}) \right) + \sum_{\mu=1}^{\mathbf{p}} h_\mu \sum_{i=1}^{2^{k+1}} \xi_i S_i \right) \]

\[ \times \left( \frac{\beta 2^{2(k+1)}}{2 \cdot 2^{2n(k+1)}} \sum_{\mu=1}^{\mathbf{p}} m_{k+1,\mu}(\vec{S}) - \beta 2^{(k+1)(1-2n)} \gamma^{k+1} \right) \]

\[ = \frac{\beta}{2} 2^{(k+1)(1-2n)} \left( \sum_{\mu=1}^{\mathbf{p}} m_{k+1,\mu}(\vec{S}) - 2m_\mu m_{k+1,\mu}(\vec{S}) + m_\mu^2 \right) \]

\[ = \frac{\beta}{2} \gamma^{(k+1)(1-2n)} \sum_{\mu=1}^{\mathbf{p}} m_\mu^2 \]

\[ = \frac{\beta}{2} \gamma^{(k+1)(1-2n)} \sum_{\mu=1}^{\mathbf{p}} \left( m_{\mu}^{k+1}(\vec{S}) - m_\mu \right)^2 \]

\[ = \frac{\beta}{2} \gamma^{(k+1)(1-2n)} \sum_{\mu=1}^{\mathbf{p}} m_\mu^2. \]
Since the term in the brackets above, \((\cdot)\), is nonnegative, we get
\[
\Phi_{k+1,1} = \Phi_{k+1,0} + \int_0^1 d\Phi_{k+1,1}(x, h) \frac{d}{dt} \mu(t)
\]
\[
\geq \Phi_{k,1} \left( \left\{ h_\mu + 2^{(k+1)(1-2\epsilon)} m_\mu \right\} \right) - \frac{\beta}{2} 2^{(k+1)(1-2\epsilon)} \sum_{\mu=1}^{p} m_\mu^2
\]
\[
\geq \Phi_{1,0} \left( \left\{ h_\mu + \sum_{l=2}^{k+1} 2^{(1-2\epsilon)} m_\mu \right\} \right) - \frac{\beta}{2} \sum_{l=1}^{k+1} 2^{(1-2\epsilon)} \sum_{\mu=1}^{p} m_\mu^2
\]
\[
= \Phi_{0,1} \left( \left\{ h_\mu + \sum_{l=1}^{k+1} 2^{(1-2\epsilon)} m_\mu \right\} \right) - \frac{\beta}{2} \sum_{l=1}^{k+1} 2^{(1-2\epsilon)} \sum_{\mu=1}^{p} m_\mu^2,
\]
where we used (62) recursively.

Now we can estimate the last term, \(\Phi_{0,1}\), in the following way
\[
\Phi_{0,1} \left( \left\{ h_\mu + \sum_{l=1}^{k+1} 2^{(1-2\epsilon)} m_\mu \right\} \right)
\]
\[
= E_\xi \log \sum_{S \in \{-1,1\}} \exp \left( \beta \sum_{\mu=1}^{p} \left( h_\mu + \sum_{l=1}^{k+1} 2^{(1-2\epsilon)} m_\mu \right) \xi^S \right) \quad (63)
\]
\[
= \log 2 + E_\xi \log \cosh \left( \beta \sum_{\mu=1}^{p} \left( h_\mu + \sum_{l=1}^{k+1} 2^{(1-2\epsilon)} m_\mu \right) \xi^2 \right), \quad (64)
\]
where \(E_\xi\) averages over the quenched patterns as usual.

In summary, we have
\[
f_{k+1} \geq \log 2 + E_\xi \log \cosh \left( \beta \sum_{\mu=1}^{p} \left( h_\mu + \sum_{l=1}^{k+1} 2^{(1-2\epsilon)} m_\mu \right) \xi^2 \right) - \frac{\beta}{2} \sum_{l=1}^{k+1} 2^{(1-2\epsilon)} \sum_{\mu=1}^{p} m_\mu^2, \quad (65)
\]
which is enough to state the next theorem.

**Theorem 5 (Mean-field bound for serial retrieval).** Given \(-1 \leq m_\mu \leq +1, \forall \mu = 1, \ldots, p\), the following relation holds
\[
f \left( \beta, \{ h_\mu \}, p \right) \geq \sup_{\{m^\mu\}} \log 2 + E_\xi \log \cosh \left( \beta \sum_{\mu=1}^{p} \left( h_\mu + C_{2\epsilon-1} m_\mu \right) \xi^2 \right) - \frac{\beta}{2} C_{2\epsilon-1} \sum_{\mu=1}^{p} m_\mu^2,
\]
where the optimal order parameters are the solutions of the system
\[
m^\mu = E_\xi \xi^\mu \tanh \left( \beta \sum_{\nu=1}^{p} \left( h_\nu + C_{2\epsilon-1} m_\nu \right) \xi^2 \right).
\]
which are the self-consistent equations of a standard Hopfield model with the rescaled temperature, \(\beta C_{2\epsilon-1}\).

Again, the critical temperature of the model with no external fields, separating the paramagnetic phase from the retrieval phase, can be obtained by expanding for small \(\{m^\mu\}\) to get
Hence, $\beta_{c}^{\text{MF}} = C_{2\sigma - 1}^{-1}$. As previously outlined for the DHM, it is possible to assume, for the $k$th level, two different classes of Mattis magnetizations, $m^\mu_{\text{left}} = m^\mu_1$ and $m^\mu_{\text{right}} = m^\mu_2$ such that $m^\mu = m^\mu_1 + m^\mu_2$, and then check the stability of this potential parallel retrieval of two patterns. Following this method, we write

$$\Phi_{k,1}(\{ h_\mu + 2^{k+1}(1-2\sigma)m^\mu \}) = \frac{1}{2}\Phi_{k,1}^1(\{ h_\mu + 2^{(k+1)(1-2\sigma)}m^\mu \}) + \frac{1}{2}\Phi_{k,1}^2(\{ h_\mu + 2^{(k+1)(1-2\sigma)}m^\mu \}).$$

Using the procedure developed in the previous analysis for both the elements of the sum and using, starting from the $k$th level, $m^\mu_{1,2}$ as the order parameters of $\Phi_{1,2}^k$, we obtain

$$f_{k+1} \geq \frac{1}{2}\Phi_{0,1}(\{ h_\mu + \sum_{i=1}^{k} 2^{(1-2\sigma)}m^\mu_{1} + 2^{(k+1)(1-2\sigma)}m^\mu_2 \})$$

$$+ \frac{1}{2}\Phi_{0,1}(\{ h_\mu + \sum_{i=1}^{k} 2^{(1-2\sigma)}m^\mu_{2} + 2^{(k+1)(1-2\sigma)}m^\mu_1 \} - \beta \sum_{\mu=1}^{p} m^\mu_1^{2} + m^\mu_2^{2} - \frac{1}{2} \frac{\beta}{2} 2\sum_{i=1}^{k} 2^{(1-2\sigma)} \sum_{\mu=1}^{p} (m^\mu_1)^2 + (m^\mu_2)^2. \quad (67)$$

Now, evaluating both the terms $\Phi_{0,1}$ and taking the infinite volume limit, we can finally state the next theorem.

**Theorem 6 (Mean-field bound for parallel retrieval).** Given $-1 \leq m_\mu \leq +1$, $\forall \mu = 1, \ldots, p$, the following relation holds

$$f(\beta, \{ h_\mu \}, p) \geq \sup_{\{m^\mu\}} \log 2 + E_\xi \log \cosh \left( \beta \sum_{\mu=1}^{p} (h_\mu + C_{2\sigma-1}m^\mu) \xi^\mu \right)$$

$$+ E_\xi \log \cosh \left( \beta \sum_{\mu=1}^{p} (h_\mu + C_{2\sigma-1}m^\mu) \xi^\mu \right)$$

$$- \frac{\beta}{2} C_{2\sigma-1} \sum_{\mu=1}^{p} m^\mu_1^{2} + m^\mu_2^{2} \quad (68)$$

representing the free energy of two effectively independent Hopfield models, one for each subcluster (left and right), whose optimal order parameters fulfill

$$m^\mu_{1,2} = E_\xi \xi^\mu \tanh \left( \beta \sum_{i=1}^{p} (h_i + C_{2\sigma-1}m^i_{1,2}) \xi^i \right)$$

and whose critical temperature is again $\beta_{c}^{\text{MF}} = C_{2\sigma - 1}^{-1}$.  

---

J. Phys. A: Math. Theor. 48 (2015) 015001 E. Agliari et al
2.3. The not-mean-field scenario

This section will bypass mean-field limitations and show that the outlined scenario is robust. To achieve this goal, mirroring the previous analysis on DHM, we will provide an improved (with respect to the mean-field counterpart) bound.

The idea underlying this non-mean-field bound is the same as the idea we used in the DHM, which is extensively explained in [39]. Let us start by introducing the following definition.

**Definition 7.** Let us take \( \chi \geq 0 \), which is a real scalar parameter related to order-parameter fluctuations, and \( t \in [0, 1] \), which allows the morphism between the tricky two-body coupling and the effective one-body interaction. Let us also introduce the following interpolating Hamiltonian

\[
H_{k+1,t} = -tu(\vec{S}) - (1 - t)v(\vec{S}) + H_k(\vec{S}_1) + H_k(\vec{S}_2)
\]  

(69)

with

\[
u(\vec{S}) = \frac{1}{2} \sum_{\mu=1}^{2^{k+1}} \sum_{i\leq j}^p (\xi^\mu S_i - m_\mu)(\xi^\mu S_j - m_\mu),
\]  

(70)

\[
u(\vec{S}) = \frac{(x + 1)}{2} \sum_{\mu=1}^{2^{k+1}} \sum_{i\leq j}^p (\xi^\mu S_i - m_\mu)(\xi^\mu S_j - m_\mu)
\]  

(71)

\[
u(\vec{S}) = \sum_{\mu=1}^{2^{k+1}} \sum_{i\leq j}^p (\xi^\mu S_i - m_\mu)(\xi^\mu S_j - m_\mu)
\]  

(72)

The partition function and free energy associated with the Hamiltonian (69) are, respectively,

\[
Z_{k+1,t}(x, \{ h_\mu \}) = \sum_{\vec{S}} \exp \left[ -\beta \left( H_{k+1,t}(\vec{S}) + \sum_{\mu=1}^{2^{k+1}} h_\mu \xi^\mu S_i \right) \right],
\]  

(73)

\[
\Phi_{k+1,t}(x, \{ h_\mu \}) = \frac{1}{2^{k+1}} \sum_{x} \log Z_{k+1,t}(x, \{ h_\mu \}).
\]  

(74)

As usual, we relate \( \Phi_{k+1,0} \) with \( \Phi_{k,1} \) as

\[
\Phi_{k+1,0}(x, \{ h_\mu \}) = \Phi_{k,1} \left( \frac{1 + x}{2^{k+1}} \right),
\]  

(75)
It is possible to show that the derivative of $\Phi_{k+1,t}$ with respect to $t$ is
\[
\frac{d \Phi_{k+1,t}}{dt}(x, t) = \frac{1}{2^{k+1} Z_{k+1,t}} \sum_{S} \exp \left( -\beta \left( H_{k+1,t}(S) + \sum_{\mu=1}^{p} h_{\mu} \sum_{i} \xi_{i}^{\mu} S_{i} \right) \right) + \sum_{\mu=1}^{p} h_{\mu} \sum_{i} \xi_{i}^{\mu} S_{i} \right) \left( \beta \mu(S) - \beta \nu(S) \right) \right) \right) \right)
= -\frac{\beta}{2} 2^{(k+1)(1-2\sigma)} \sum_{\mu=1}^{p} m_{\mu}^{2} + \frac{\beta(x + 1)}{2^{(k+1)(1+2\sigma)}} \times \sum_{\mu=1}^{p} \sum_{l} \sum_{k} \left( \left( \xi_{l}^{\mu} S_{l} - m_{\mu} \right) \left( \xi_{k}^{\mu} S_{k} - m_{\mu} \right) \right). \tag{76}
\]

Now we are going to neglect the fluctuation source, which contains $\left( \left( \xi_{l}^{\mu} S_{l} - m_{\mu} \right) \left( \xi_{k}^{\mu} S_{k} - m_{\mu} \right) \right)$, and is indicated by $C(k + 1, \beta, \sigma, \{m_{\mu}\})$. While in the pure ferromagnetic case, Griffiths inequalities hold [51, 52] and ensure that such a term is positively defined (thus allowing us to get the bound), in this context we are left with an approximation only, because for neural networks, Griffiths theories have not been developed yet. However, we stress that this is not a big deal; already at a mean-field level, while the true solution of the Hopfield model is expected to be full-replica symmetry breaking (RSB) (see [55–57, 60]), usually only its replica symmetric approximation is retained for practical purposes (where order parameter’s fluctuations are disregarded). It is indeed an approximation, and not a bound.

\[
f_{k+1} = \Phi_{k+1,1}(0, \{h_{\mu}\}) = \Phi_{k,1} \left( \frac{1}{2^{2\sigma}} \left\{ h_{\mu} + \beta m_{\mu} 2^{(k+1)(1-2\sigma)} \right\} \right) - \frac{\beta}{2} 2^{(k+1)(1-2\sigma)} \sum_{\mu=1}^{p} m_{\mu}^{2} + C \left( k + 1, \beta, \sigma, \{m_{\mu}\} \right) \tag{77}
\]

Iterating the procedure one arrives to:

\[
f_{k+1} = \Phi_{0,1} \left( \sum_{l=1}^{k+1} 2^{-2\sigma} \left\{ h_{\mu} + \beta m_{\mu} \sum_{l=1}^{k+1} 2^{l(1-2\sigma)} \right\} \right) - \frac{\beta}{2} \sum_{l=1}^{k+1} 2^{l(1-2\sigma)} \sum_{\mu=1}^{p} m_{\mu}^{2} + \sum_{l=1}^{k+1} C \left( 1, \beta, \sigma, \{m_{\mu}\} \right). \tag{78}
\]

Calculating the value of $\Phi_{0,1}$, using (73), (74), and (75), we get the following theorem.

Theorem 7 (Non-mean-field approximation for serial retrieval). Given $-1 \leq m_{\mu} \leq +1$, $\forall \mu = 1, \ldots, p$, the serial NMF-approximation for the HHM reads as

\[
f^{\text{NMF}}(\beta, \{h_{\mu}\}, p) = \sup_{\mu} \left[ \log 2 + E_{\xi} \log \cosh \left( \sum_{\mu=1}^{p} \left( h_{\mu} + \beta m_{\mu} (C_{2\sigma-1} - C_{2\sigma}) \right) \xi_{i}^{\mu} \right) \right. \right.

- \frac{\beta}{2} \sum_{\mu=1}^{p} m_{\mu}^{2} \left( C_{2\sigma-1} - C_{2\sigma} \right) \right].
\]
representing a Hopfield model at rescaled temperature, with optimal order parameters fulfilling

\[ m^\mu = E_\xi \xi^\mu \tanh \left( \beta \sum_{\nu=1}^{\nu} (\beta h_\nu + (C_{2\sigma-1} - C_{2\sigma}) m^\nu) \xi^\nu \right) \]

and critical temperature \( \beta^\text{NMF} = C_{2\sigma-1} - C_{2\sigma} \).

Again it is possible to generalize the serial retrieval, assuming two different families of Mattis magnetizations \( (m_\mu^{(1)}, m_\mu^{(2)})_{\mu=1}^{p} \) for the two blocks of spin under the \( k \)th level. Following this method and using the NMF interpolating procedure for the two blocks, we get

\[
f_{k+1} \left( \{ h_\mu \}, \beta, \sigma, p \right) = \log 2 + \frac{1}{2} E_\xi \log \cosh \left( \sum_{\mu=1}^{p} \left( \beta h_\mu + \beta m_\mu^p \right) \sum_{l=1}^{k} 2^{(1-2\sigma)} \right)
- \sum_{l=1}^{k+1} 2^{(-2\sigma)} + \beta m_\mu^p 2^{(k+1)(1-2\sigma)} \xi_\mu 
+ \frac{1}{2} E_\xi \log \left( \sum_{\mu=1}^{p} \left( \beta h_\mu + \beta m_\mu^p \right) \sum_{l=1}^{k} 2^{(1-2\sigma)} \right)
- \sum_{l=1}^{k+1} 2^{(-2\sigma)} + \beta m_\mu^p 2^{(k+1)(1-2\sigma)} \xi_\mu 
- \beta \left( \sum_{l=1}^{k} 2^{(1-2\sigma)} - \sum_{l=1}^{k+1} 2^{(-2\sigma)} \right) \sum_{\mu=1}^{p} m_\mu^2 + m_\mu^p^2 \frac{2}{2}
- \frac{1}{2} \sum_{\mu=1}^{p} m_\mu^2 + C \left( k + 1, \beta, \sigma, \{ m_\mu^p \} \right)
+ \frac{1}{2} \sum_{\mu=1}^{k} \left( C \left( l, \beta, \sigma, \{ m_\mu^l \} \right) + C \left( l, \beta, \sigma, \{ m_\mu^l \} \right) \right) \right) \] (79)

that, in the infinite volume limit where the interactions between the two blocks vanish, and again partially neglecting the correlations, brings us to the following definition.

**Definition 8 (Non-mean-field approximation for parallel retrieval).** Given \(-1 \leq m_\mu \leq +1, \forall \mu = 1, \ldots, p\), the parallel NMF approximation for the HHM reads as

\[
f \left( \{ h_\mu \}, \beta, \sigma, p \right) = \sup_{\{ m_\mu^l \}} \left( \log 2 + \frac{1}{2} E_\xi \log \cosh \left( \sum_{\mu=1}^{p} \left( \beta h_\mu + \beta m_\mu^p \right) (C_{2\sigma-1} - C_{2\sigma}) \right) \right)
+ \frac{1}{2} E_\xi \log \cosh \left( \sum_{\mu=1}^{p} \left( \beta h_\mu + \beta m_\mu^p \right) (C_{2\sigma-1} - C_{2\sigma}) \right)
- \frac{1}{2} \left( C_{2\sigma-1} - C_{2\sigma} \right) \sum_{\mu=1}^{p} m_\mu^2 + m_\mu^p^2 \frac{2}{2}, \right) \] (80)
that is, the free energy of two independent Hopfield models for each of the two subgroups of spins, with disentangled optimal order parameters satisfying

\[ m_{1,2}^\mu = \mathbb{E}_\xi m_{1,2}^\mu \tanh \left( \beta \sum_{\nu=1}^n \left( h_\nu + \left( C_{2\sigma-1} - C_{2\sigma} \right) m_{1,2}^\nu \right) \right), \]

and critical temperature \( \beta_{\text{NMF}}^c = C_{2\sigma-1} - C_{2\sigma} \).

### 3. Outlooks and conclusions

Originally, neural networks were developed on fully connected structures and embedded with mean-field constraints [19]. Later, as graph theory analyzed complex structures as small worlds [63] or scale-free networks [64], neural networks were readily implemented on these structures, too [62, 65, 66]. Hence, neurons were no longer fully connected, but the mean-field prescription was retained. Note that in those cases, parallel processing was extensive, up to \( P \sim N \), but pattern-vectors allowed (extensive) blank entries [21].

However, the quest to bypass mean-field limitations, driven already by clear physical arguments, has recently been strongly influenced directly by neurobiology, leading toward hierarchical prescriptions [53].

As a side note exact hierarchical models have recently experienced renewed interest in statistical mechanics, as structures where testing spin-glasses beyond the mean-field paradigm [31, 54] have implicitly offered the backbone for bypassing mean-field limitations in neural networks.

As we recently developed a new interpolation scheme for these structures [39] that, while not fully solving the model’s thermodynamics, allows us to overcome the mean-field picture while maintaining a formal description (i.e., the theorems and bounds available), we extended this technology to cover neural networks, and we combined it with the Amit technique of using an ansatz to investigate the candidate-retrievable states. This fusion resulted in a stronger method that allowed us to analyze both the ferromagnet on a hierarchical topology (DHM) and the neural network on a hierarchical topology (HHM).

Starting with the former (which we used as a test-guide for the latter), remarkably, beyond the previously discussed ferromagnetic scenario [30, 49, 50], we have shown that the model has a plethora of metastable states that become stable in the thermodynamic limit and forbid self-averaging for the magnetization in a way quite similar to the scenario derived for the overlap in mean-field spin-glasses [17, 48].

Filtering these results within the neural network perspective, we shown that these networks, where clusters of neurons located far apart essentially do not interact, perform both à la Hopfield [19], thus relaxing via a global rearrangement of all the spins in order to retrieve an extensive stored pattern, and in a multitasking fashion similar to the parallel processing performance shown by other (mean-field) associative networks, which we developed over the past two years in a series of papers [21–24, 27].

We want to add one final note of interest regarding the capacity of these networks. We have shown how it is possible to simultaneously recall two patterns by splitting the system into two subgroups, going down over the levels from the top. We have seen that, since the upper interaction is vanishing with enough velocity (see the appendix for more information), in the thermodynamic limit the two subgroups of neurons can be regarded as independent:

\[ \text{Note that hierarchical models in neural networks have already been discussed [58, 59, 61], but in those papers, the adjective referred to the correlated patterns and not to the neurons, which is completely different research.} \]
each one is governed by an Hopfield Hamiltonian and can choose to recall one of the memorized patterns. Clearly we could use the same argument iteratively and split the system in to more sub-sub-clusters, going down over the various levels. Crucially, what is fundamental is that at least the sum of the upper levels of the interactions remains vanishing in the infinite volume limit. If we split the system $M$ times, we have to use different order parameters for the magnetizations of the blocks, until the $k - M$ level where the system is divided into $2^M$ subgroups. The procedure keeps working as far as

$$\lim_{k \to \infty} \sum_{j=k-M}^{k} 2^{(1-2\alpha)} \sum_{\mu=1}^{p} m_{\mu} = 0. \quad (81)$$

Since the magnetizations are bounded, in the worst case we have

$$\sum_{j=k-M}^{k} 2^{(1-2\alpha)} \sum_{\mu=1}^{p} m_{\mu} \leq p \sum_{j=k-M}^{k} 2^{(1-2\alpha)} \leq p \sum_{j=k-M}^{\infty} 2^{(1-2\alpha)} 2^{(1-2\alpha)(k-M)p}. \quad (82)$$

If we want the system to handle up to $p$ patterns, we need $p$ different blocks of spins, and then $M = \log (p)$. So, for example, if $p = \mathcal{O}(k)$, $2^{(1-2\alpha)(k-\log(p))} p \to 0$ as $k \to \infty$.

Thus, the parallel processing ability works best with a logarithmic load of patterns, as far as $p = \mathcal{O}(k) = \mathcal{O}(\log N)$. However, such a bound, which is not as enormous as $P \sim N$ patterns that its serial counterpart can handle, is never reached in practice. Why do these networks have a restricted capacity?

The answer lies in the real fingerprint of not-mean-field spontaneous parallel processing. In order to explain the network’s ability to manage two patterns simultaneously, we used the argument that the upper links connecting the left and right communities are actually vanishing in the thermodynamic limit. While this is exciting for parallel processing capabilities because it allows us to divide the network into almost-disjointed communities, it is a disaster for the storage capacity because each time we use this argument, we effectively admit that a huge amount of synapses for storing memory are vanishing. Thus, we have found a novel balancing requirement in non-mean-field processing. Extreme parallel processing implies the smallest storage capacity, and vice versa. We aim to check this prediction in real networks in the future.

Acknowledgments

The authors are grateful to Sapienza Università di Roma, Gruppo Nazionale per la Fisica Matematica (GNFM-INdAM), and Istituto Nazionale di Fisica Nucleare (INFN) for supporting this work. EA and AB are indebted to Enzo Marinari and Federico Ricci-Tersenghi for a list of countless items, ranging from personal support to priceless discussions.

Appendix. Selection of a state

While we have shown that in the thermodynamic limit, the (intensive) energies associated with the ferromagnetic and the mixture states coincide, the metastable state is not forbidden thermodynamically (while its weight is negligible w.r.t. the ferromagnetic scenario, and the system must be trapped opportune with external fields in its basin to keep it in the large $k$
(limit). We must still face some questions. Let us consider the mixture state and approach the critical region from an ergodic scenario. The two clusters differ in magnetization, with one having \( m_1 > 0 \) and the other \( m_1 < 0 \), and only the upper (and hence weakest) link connects them. Does that one cluster act on the other, playing the role of an external field in mean-field schemes, thus selecting the phase and reversing the other cluster magnetization sign, which would result in the destruction of mixture states? In this appendix, we aim to show that this is not the case.

In an attempt to prove this thesis, we will address this question within the more familiar mean-field perspective (namely the CW model), and then we will enlarge the observation based on that example toward bipartite ferromagnetic systems; we will show that they continue holding. As a last step to obtain the result, we will compare the Dyson model (whose spins are locked in a mixture state with a bipartite ferromagnet to enlarge the present model to include the stability argument.

Under the critical temperature systems of spins whose dynamics is no longer ergodic, there is an equilibrium state (in the thermodynamic limit) that can be a mixture of several pure states. Each of these states has its own basin of attraction, in the sense that the system will reach one of the states, according to its initial configuration. As far as ferromagnetic systems are concerned, by adding a suitable external field, it is possible to select one of these pure states (i.e., the dynamics is forced into one of the attractors). Now we can ask when an external field is able to select a state. Consider a Glauber dynamics for a ferromagnetic system of \( N \) spins at zero temperature:

\[
S_i(t + 1) = \text{sgn} \left( h_i(S(t)) + h_N \right).
\]  

(A.1)

For example, we can keep in mind the case of the CW model where \( h_i(S(t)) \), the field acting on the \( i \)th spin, is the magnetization \( m(S(t)) = \frac{1}{N} \sum_{i=1}^{N} S_i(t) \). In that case, each initial configuration for which \( |h_N| > |h_i(S^0)| \) will follow the external field. If, for example, \( |h_N| < 1 \), there will always exist initial configuration \( (h_N < |h_i(S^0)|) \), which does not feel the influence of the external field. But, if we choose the initial configuration randomly and according to \( p(S^0) \), we can say that the field \( h_N \) selects the state if

\[
P_N \left( S^0 : |h_N| > |h_i(S^0)| \right) \rightarrow 1.
\]  

(A.2)

On the contrary, we will say that the field will not select the state if

\[
P_N \left( S^0 : |h_N| < |h_i(S^0)| \right) \rightarrow 1.
\]  

(A.3)

In what follows, we will consider \( P_N(S^0) = \prod_{i=1}^{N} p(S_i^0) \), with \( p(S) \) uniform in \([-1, 1]\). For the CW model, we can state the following.

**Theorem 8.** In the CW model, where \( h_i(S) = m(S) = \frac{1}{N} \sum_{i=1}^{N} S_i \), \( \forall \epsilon > 0 \),

- \( h_N : |h_N| > \frac{1}{N^{\epsilon/(1-\epsilon)}} \) selects the state;
- \( h_N : |h_N| < \frac{1}{N^{\epsilon/(1-\epsilon)}} \) does not select the state.
With regard to the first statement, we note that if $|h_N| > \frac{1}{N^{\frac{1}{2(1-\epsilon)}}}$,
\[
P_N\left( S^0; |h_N| > h_1(S^0) \right) = 1 - P_N\left( S^0; |h_1(S^0)| > |h_N| \right) \\
\geq 1 - P_N\left( S^0; m(S^0) > \frac{1}{N^{\frac{1}{2(1-\epsilon)}}} \right) \\
\geq 1 - N^{1-\epsilon}\mathbb{E}_N\left[ \sigma_2(S^0) \right] = 1 - N^{-\epsilon} \rightarrow 1, \tag{A.4}
\]
where we used Chebyshev inequality and the fact that $\mathbb{E}_N[\sigma_2(S^0)] = \frac{1}{N}$. With regard to the second statement, we note that, since $|h_N| < \frac{1}{N^{\frac{1}{2(1-\epsilon)}}}$,
\[
P_N\left( S^0; |h_N| > h_1(S^0) \right) \leq P_N\left( S^0; |m(S^0)| < \frac{1}{N^{\frac{1}{2(1+\epsilon)}}} \right) \\
= P_N\left( S^0; \sqrt{N}m(S^0) < \frac{1}{N^{\frac{1}{2}}} \right) \\
\rightarrow \mu_{N(0,1)}(\{ \epsilon < \frac{1}{N^{\frac{1}{2}}} \}) \rightarrow 0, \tag{A.5}
\]
where we just used the fact that the variable $\sqrt{N}m(S^0) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} S_i$ satisfies the central limit theorem (CLT) and tends in distribution to a $\mathcal{N}(0, 1)$ Gaussian variable.

We can repeat the same analysis in a mean-field bipartite ferromagnetic model where the interaction inside the parties (modulated by $J_{11}$ and $J_{22}$) and the ones among the parties (modulated by $J_{12}$) have different couplings. In that case, we can seek out the cases where $J_{12}$ can select the state where the two parties are aligned and not independent. If we consider, for example, a spin in the first party, we have for the Glauber dynamics

\[\text{Figure A.1.} \text{ Analysis of the susceptibility of the system, defined as } X = (m^2) - (m)^2, \text{ versus the noise level, } T \equiv \beta^{-1}, \text{ for various sizes (as reported in the legend) and } \sigma = 0.99. \text{ Left panel: } X(T) \text{ for the pure state. Right panel: } X(T) \text{ for the mixture state. Note that, while in the ferromagnetic (pure) case all the cuspids are on the same noise level no matter the value of } k, \text{ this is not the case for the mixture state because such a state is metastable, because the difference in the energy, } \Delta E, \text{ among the two states scales as } \propto 1/N^{3\alpha-1}. \text{ Hence, for } k \to \infty \text{ only, the mixture state becomes stable and its cusp happens at the same noise level of its pure counterpart.} \]
(i.e., we can repeat the same argument of the CW model, identifying the field sent by the second party, which is proportional to the magnetization \( m_2(S) \), as the external field). Thus, using the analogous version of the previous theorem, we see that \( J_{12} \) is able to select the state only if

\[
J_{12}(N) \left| m_2(S^0) \right| > \frac{1}{N^{\frac{1}{2(1-\epsilon)}}}.
\]

with probability 1. Since for the CLT, \( |m_2(S^0)| \) is \( O(\sqrt{N}) \) with probability 1, vanishing \( J_{12} \) will not be able to select the totally magnetized state. In that case, the system behaves exactly as two non interacting CW subsystems. Oversimplifying along this line and assuming a mean-field scenario for the inner blocks, the Dyson model could be considered from this point of view as a generalization of a bipartite model. In fact, if we divide the system into two subgroups of spins, we have that the external field (representing the last level of interaction) is proportional to \( J(N)m_N(S) \), while the internal field is a sum of contributions coming from all the submagnetizations. Since \( J(N) = N^1-2\sigma \) is vanishing in the thermodynamic limit, the two subgroups behave as if they were non interacting. This may cause confusion about the phase transition because the system, when not trapped within the pure state, crosses the critical line in the \( \beta, \sigma \) plane and moves from an ergodic region where the global magnetization is zero toward a mixture state where the magnetization is also zero. However, regarding the latter, the two sub-clusters have nonzero magnetizations, and even in this case, crossing the line returns to a canonical phase transition (see figure A.1). To give further proof of this delicate way of breaking ergodicity, we show further results from extensive Monte Carlo runs that confirm our scenario and are reported in figure A.1.

References

[1] Hagan M T, Demuth H B and Beale M H 1996 Neural Network Design (Boston, MA: PWS)

[2] Miller W T, Werbos P J and Sutton R S 1995 Neural Networks for Control (Cambridge, MA: MIT Press)

[3] Amit D J 1992 Modeling Brain Function: The World of Attractor Neural Network (Cambridge: Cambridge University Press)

[4] Hertz J and Palmer R 1991 Introduction to the theory of neural networks Lecture Notes

[5] Coolen A C C, Kuehn R and Sollich P 2005 Theory of Neural Information Processing Systems (Oxford: Oxford University Press)

[6] Ackley D H, Hinton G E and Sejnowski T J 1985 A learning algorithm for Boltzmann machines Cogn. Sci. 9 147–69

[7] Rolls E T and Treves A 1988 Neural Networks and Brain Function (Cambridge: Cambridge University Press)

[8] Tuckwell H C 2005 Introduction to Theoretical Neurobiology vols 1 and 2 (Cambridge: Cambridge University Press)

[9] Domhoff G W 2003 The Scientific Study of Dreams: Neural Networks, Cognitive Development, and Content Analysis (Washington, DC: American Psychological Association)

[10] Martindale C 1991 Cognitive Psychology: A Neural Network Approach (Pacific Grove, CA: Thomson Brooks/Cole Publishing)

[11] Turing A M 1950 Computing machinery and intelligence Mind 59 433–60

[12] von Neumann J 1951 The general and logical theory of automata Cerebral Mechanisms in Behavior (Oxford: Wiley) pp 1–41

[13] Minsky M L and Papert S A 1987 Perceptrons—Expanded Edition: An Introduction to Computational Geometry (Cambridge, MA: MIT Press)

[14] McCulloch W S and Pitts W 1943 A logical calculus of the ideas immanent in nervous activity Bull. Math. Biophys. 5 115–33
[15] Rosenblatt F 1958 The perceptron: a probabilistic model for information storage and organization in the brain Psychol. Rev. 65 386
[16] Hebb D O 1940 The Organization of Behavior: A Neuropsychological Theory (Psychology Press)
[17] Mezard M, Parisi G and Virasoro M A 1987 Spin Glass Theory and Beyond (Lecture Notes in Physics vol 9) (Singapore: World Scientific)
[18] Panchenko D 2013 The Sherrington–Kirkpatrick Model (New York: Springer)
[19] Hopfield J J 1982 Neural networks and physical systems with emergent collective computational abilities Proc. Natl Acad. Sci. USA 79 2554–8
[20] Amit D J, Gutfreund H and Sompolinsky H 1985 Spin-glass models of neural networks Phys. Rev. A 32 1007
[21] Agliari E, Barra A, Galluzzi A, Guerra F and Moauro F 2012 Multitasking associative networks Phys. Rev. Lett. 109 268101
[22] Sollich P, Tantari D, Annibale A and Barra A 2014 Extensive load in multitasking associative networks arXiv:1404.3654
[23] Agliari E, Annibale A, Barra A, Coolen A C C and Tantari D 2013 Immune networks: multitasking capabilities at medium load J. Phys. A: Math. Theor. 46 335103
[24] Agliari E, Annibale A, Barra A, Coolen A C C and Tantari D 2013 Immune networks: multitasking capabilities near saturation J. Phys. A: Math. Theor. 46 415003
[25] Agliari E, Barra A, De Antoni A and Galluzzi A 2013 Parallel retrieval of correlated patterns: from Hopfield networks to Boltzmann machines Neural Netw. 38 52–63
[26] Agliari E, Barra A, Del Ferraro G, Guerra F and Tantari D 2014 Anergy in self-directed B lymphocytes: a statistical mechanics perspective J. Theor. Biol. at press doi:10.1016/j.jtbi.2014.05.006
[27] Agliari E et al 2013 Parallel processing in immune networks Phys. Rev. E 87 042701
[28] Barra A, Genovese G, Guerra F and Tantari D 2012 How glassy are neural networks? J. Stat. Mech. P07009
[29] Barra A, Bernacchia A, Contucci P and Santucci E 2012 On the equivalence of Hopfield networks and Boltzmann machines Neural Netw. 34 1–9
[30] Dyson F 1969 Existence of a phase transition in a one dimensional Ising ferromagnet Commun. Math. Phys. 12 91–107
[31] Castellana M, Decelle A, Franz S, Mezard M and Parisi G 2012 Hierarchical random energy model of a spin glass Phys. Rev. Lett. 104 127206
[32] Castellana M 2011 Real-space renormalization group analysis of a non-mean-field spin-glass Europhys. Lett. 95 47014
[33] Kim D and Thompson C J 1977 Critical properties of Dyson’s hierarchical model J. Phys. A: Math. Gen. 10 1579
[34] Collet P and Eckmann J-P 1977 The $\epsilon$-expansion for the hierarchical model Commun. Math. Phys. 55 67–96
[35] Angelini M C, Parisi G and Ricci-Tersenghi F 2013 Ensemble renormalization group for disordered systems Phys. Rev. B 87 134201
[36] Parisi G and Rocchi J 2014 Critical exponents of the random field hierarchical model Phys. Rev. B 90 024203
[37] Decelle A, Parisi G and Rocchi J 2014 Ensemble renormalization group for the random-field hierarchical model Phys. Rev. E 89 032132
[38] Angelini M C, Ricci-Tersenghi F and Parisi G 2014 Relations between short-range and long-range Ising models Phys. Rev. E 89 062120
[39] Castellana M, Barra A and Guerra F 2014 Free-energy bounds for hierarchical spin models J. Stat. Phys. 155 211
[40] Franz S, Jorg T and Parisi G 2009 Overlap interfaces in hierarchical spin-glass models J. Stat. Mech. P02002
[41] Metz F L, Leuzzi L, Parisi G and Sacksteder V 2013 Transition between localized and extended states in the hierarchical Anderson model Phys. Rev. B 88 045103
[42] Metz F L, Leuzzi L and Parisi G 2014 The renormalization flow of the hierarchical Anderson model at weak disorder Phys. Rev. B 89 064201
[43] Monthus C and Garel T 2013 Dynamical barriers in the Dyson hierarchical model via real space renormalization J. Stat. Mech. P02023
[44] Monthus C and Garel T 2014 Scaling of the largest dynamical barrier in the one-dimensional long-range Ising spin-glass Phys. Rev. B 89 014408
[45] Mukamel D 2008 Notes on the statistical mechanics of systems with long-range interactions *Les Houches Lecture Notes* (arXiv:0905.1457)
[46] Agliari E, Barra A, Galluzzi A, Guerra F, Tantari D and Tavani F 2014 Hierarchical neural networks perform both serial and parallel processing *Neural Netw.* submitted (arXiv:1409.0227)
[47] Guerra F and Toninelli F L 2002 The thermodynamic limit in mean-field spin glasses *Commun. Math. Phys.* 230 71–79
[48] Guerra F 2003 Broken replica symmetry bounds in the mean-field spin glass model *Commun. Math. Phys.* 233 1–12
[49] Gallavotti G and Miracle-Sole S 1967 Statistical mechanics of lattice systems *Commun. Math. Phys.* 5 317–23
[50] Bleher P M and Sinai J G 1973 Investigation of the critical point in models of the type of Dyson’s hierarchical models *Commun. Math. Phys.* 33 23–42
[51] Griffiths R B 1967 Correlations in Ising ferromagnets: III. *Commun. Math. Phys.* 6 121–7
[52] Griffiths R B 1967 Correlations in Ising ferromagnets: II. external magnetic fields *J. Math. Phys.* 8 484
[53] Moretti P and Munoz M A 2013 Griffiths phases and the stretching of criticality in brain networks *Nat. Commun.* 4 2521
[54] Castellana M and Parisi G 2011 A renormalization group computation of the critical exponents of hierarchical spin glasses *Phys. Rev. E* 83 041134
[55] Barra A, Genovese G and Guerra F 2010 The replica symmetric approximation of the analogue neural network *J. Stat. Phys.* 140 784–96
[56] Barra A, Genovese G, Guerra F and Tantari D 2014 About a solvable mean-field model of a Gaussian spin glass *J. Phys. A: Math. Theor.* 47 155002
[57] Barra A, Genovese G and Guerra F 2012 Equilibrium statistical mechanics of bipartite spin systems *J. Phys. A: Math. Theor.* 44 245002
[58] Engel A 1990 Storage of hierarchically correlated patterns *J. Phys. A: Math. Gen.* 23 2587
[59] Gutfreund H 1988 Neural networks with hierarchically correlated patterns *Phys. Rev. A* 37 570
[60] Pastur L, Shcherbina M and Tirozzi B 1994 The replica-symmetric solution without replica trick for the Hopfield model *J. Stat. Phys.* 74 1161–83
[61] Perez-Vincente C J 1989 Hierarchical neural network with high storage capacity *Phys. Rev. A* 40 5356
[62] Agliari E, Asti L, Barra A, Burioni R and Uguzzoni G 2012 Analogue neural networks on correlated random graphs *J. Phys. A: Math. Theor.* 45 365001
[63] Barra A and Agliari E 2011 Equilibrium statistical mechanics on correlated random graphs *J. Stat. Mech.* 02 02027
[64] Barabasi A L, Albert R and Jeong H 1999 Mean-field theory for scale-free random networks *Physica A* 272 173–87
[65] Wemmenhove B and Coolen A C C 2003 Finite connectivity attractor neural networks *J. Phys. A: Math. Gen.* 36 9617–33
[66] Perez Castillo I, Wemmenhove B, Hatchett J P L, Coolen A C C, Skantzos N S and Nikoletopoulo T 2004 Analytic solution of attractor neural networks on scale-free graphs *J. Phys. A: Math. Gen.* 37 8789–99