Sapienza
Università di Roma

Facoltà di Scienze Matematiche Fisiche e Naturali

Corso di Dottorato in Matematica

Statistical mechanics of multitasking networks

Advisors

Prof. Francesco Guerra

Dott. Adriano Barra

Dott. Elena Agliari

Candidate

Daniele Tantari

ID number: 1168739

Abstract

The mean field Hopfield model is the paradigm for serial processing networks: a system able to retrieve, one at a time, previously stored patterns of information. On the other side, multitasking associative networks (retrieving several patterns of informations at the same time) are crucial for the understunding of real biological systems and for the development of parallel processing artificial machines. In this thesis, I will introduce two different ways to build parallel processing associative networks: by diluting the patterns entries (Diluted Hopfield Model) or by introducing a topological structure (Hierarchical Hopfield Model) towards a non mean-field interaction among agents. From a statistical mechanics perspective, passing through the analogy between the Hopfield model and a bipartite spin glass system, we analyze, in the former case, the capability of the network by varying the level of load and dilution (from low to high storage, from fully connected to finite connectivity regime), while, in the last case, we investigate the possibility of switching from a serial to a parallel processing regime by varying the level of the temperature.

Contents

Part I

Introduction

The study of the principles behind the information processing in complex networks of simple interacting, decision-making agents, be these cells in biological systems (for example neurons in brain and other nervous tissues or lymphocytes of the immune system) or components in artificial networks (electronnic processors or even softwares), is considered as one of the most interdisciplinary scientific challenge. The reason is that in this multifaced area of research, that involves biologists (particularly neuro-biologists [92, 99] and cognitive psychologists [52, 74]), computer scientists and engineers (mainly interested in electronics and robotics [67, 81]), physicists (mainly involved in statistical mechanics and stochastic processes[15, 69]) and mathematicians (mainly working in learning algorithms and graph theory [47, 2]), each discipline gets and offers at the same time something interesting and worthwhile for the collaboration with the others: biologists benefit of the new tools offered by the more quantitative sciences, computer scientists or engineers find inspiration from biology and techniques from physics and math, physicists and mathematicians meet new interesting challenges, developing new theories and discovering other application domains. However, this kind of interdisciplinarity has also its drawbacks: it is very difficult to keep the disciplines involved connected each others, beacause of the existence of language barriers or motivation differences, and, as a result, several important discoveries have to be made more then once, before they find themselves recognized as such.

Historically, the term complex network have always had a biological flavour being first associated to the neural one: the study of *neural information processing systems* probably started with the birth of programmable computing machines around the second world war. It came to be realized that programmable machines might be made to *think*, and, conversely, that human thinking could perhaps be understood in the language of programmable machines. According to this framework, the brain (or alternatively an other biological network) is considered as a piece of hardware, performing quite sophisticated information processing tasks, using microscopic elements interacting each others. On the other hand biological networks and conventional computer systems have a lot of differences. For example, in conventional computers each individual operation is performed, as a rule, sequentially, so that failure of any part of the chain is mostly fatal, in contrast with biological networks where microscopic agents operate in *parallel*. An other important difference is that conventional computers can only execute a detailed specification of orders, the program, requiring the programmer to know exactly which data can be expected and how to respond, while biological network can adapt quite well to changing circumstances. Finally there is biological robustness against physical hardware failure.

Roughly speaking, one can distinguish three types of motivation for studying neural networks. Biologists aim at understanding information processing in real biological networks; Engineers and computer scientists would like to use the principles behind neural information processing for designing adaptive software and artificialinformation processing systems which can learn and which are structured in such a way that the processors can operate efficiently in parallel; theoretical physicists and mathematicians are challenged by the fondamental new problems posed by neural network models, which exhibit a rich and highly non-trivial behaviour.

Tracing the genesis and evolution of neural networks back in time is very difficult, probably due to the broad meaning they have acquired along the years: scientists closer to the robotics branch often refer to the W. McCulloch and W. Pitts model of perceptron [75] (or the F. Rosenblatt version [93]) as the first systematic investigation on the information processing capabilities of neural networks, while researchers closer to the neurobiology branch adopt usually the D. Hebb work as a starting point [68]. On the other hand, scientists involved in statistical mechanics, that joined the community in relatively recent times (after a satisfactory picture of spin glasses was achieved [79, 88]), usually refer to the seminal paper by Hopfield [70] or to the celebrated work by Amit, Gutfreund and Sompolinsky [16], where the statistical mechanical analysis of the Hopfield model is effectively carried out.

The Hopfield model rapidly became the "harmonic oscillators" of parallel processing: neurons, thought of as "binary nodes" (spins) of a network, behave collectively to retrieve information, the latter being spread over the synapses, thought of as the interconnections among nodes. However, common intuition of parallel processing is not only the underlying parallel work performed by neurons to retrieve, say, an image on a book, but rather, for instance, to retrieve the image and, while keeping the book securely in hand, noticing beyond its edges the room where we are reading, still maintaining available resources for further retrieves as a safety mechanism. Standard Hopfield networks are not able to accomplish this kind of parallel processing [4, 6, 13]. Goal of this work is to relax this kind of limitation so to extend standard Hopfield model, the paradigm of neural network, toward multitasking capabilities, whose interest goes far beyond the artificial intelligence framework [95, 3, 5, 8].

Before proceeding, we recall some interesting results about the standard Hopfield model. To this task, following [29], we start from the Curie-Wiess model, showing that toy models for paramagnetic-ferromagnetic transition [9] are natural prototypes for the autonomous storage/

retrieval of information patterns in neural networks model, studied from a statistical mechanic point of view.

Storing the first pattern: Curie-Weiss paradigm.

The statistical mechanical analysis of the Curie-Weiss model (CW) can be summarized as follows: starting from a microscopic formulation of the system, i.e. N dichotomic spins $\sigma_i \in \{-1, 1\}$, $i = 1 \cdots N$, their pairwise couplings $J_{ij} \equiv J$, and possibly an external field h, we define an Hamiltonian

$$H_N(\sigma|J,h) = -\frac{J}{N}\sum_{i<j}^{N}\sigma_i\sigma_j \tag{1}$$

and derive an explicit expression for the (macroscopic) free energy

$$A(\beta) = \lim_{N\to\infty}\frac{1}{N}\ln\sum_{\{\sigma\}}^{2^N}\exp\left[-\beta H_N(\sigma|J,h)\right], \tag{2}$$

where the sum is performed over the set $\{\sigma\}$ of all 2^N possible spin configurations, each weighted by the Boltzmann factor $\exp[-\beta H_N(\sigma|J,h)]$ that tests the likelihood of the related configuration. From expression (2), we can derive the whole thermodynamic quantities and in particular the phase-diagrams, that is, we are able to discern regions in the space of tunable parameters (e.g. temperature/noise level) where the system behaves as a paramagnet or as a ferromagnet.

Thermodynamical averages, denoted with the symbol $\langle . \rangle$, provide for a given observable the expected value, namely the value to be compared with measures in an experiment. For instance, for the magnetization $m(\sigma) \equiv \sum_{i=1}^{N}\sigma_i/N$ we have

$$\langle m(\beta)\rangle = \frac{\sum_\sigma m(\sigma)e^{-\beta H_N(\sigma|J)}}{\sum_\sigma e^{-\beta H_N(\sigma|J)}}. \tag{3}$$

When $\beta \to \infty$ the system is noiseless (zero temperature) hence spins feel reciprocally without errors and the system behaves ferromagnetically ($|\langle m\rangle| \to 1$), while when $\beta \to 0$ the system behaves completely random (infinite temperature), thus interactions can not be felt and the system is a paramagnet ($\langle m\rangle \to 0$). In between a phase transition appears.

An explicit expression for the free energy in terms of $\langle m\rangle$ can be obtained carrying out summations in (2) and taking the *thermodynamic limit* $N \to \infty$ as

$$A(\beta) = \sup_m\{\ln 2 + \ln\cosh[\beta(m+h)] - \frac{\beta J}{2}m^2\}. \tag{4}$$

The optimal order parameter, that is the mean magnetization of the system $\langle m \rangle$, satysfies

$$\langle m \rangle = \tanh[\beta(J\langle m \rangle + h)]. \tag{5}$$

This expression returns the average behavior of a spin in a magnetic field. In order to see that a phase transition between paramagnetic and ferromagnetic states actually exists, we can fix $h = 0$ (and pose $J = 1$ for simplicity) and expand the r.h.s. of eq. 5 to get

$$\langle m \rangle \propto \pm\sqrt{\beta J - 1}. \tag{6}$$

Thus, while the noise level is higher than one ($\beta < \beta_c \equiv 1$ or $T > T_c \equiv 1$) the only solution is $\langle m \rangle = 0$, while, as far as the noise is lowered below its critical threshold β_c, two different-from-zero branches of solutions appear for the magnetization and the system becomes a ferromagnet. The branch effectively chosen by the system usually depends on the sign of the external field or boundary fluctuations.

Clearly, the lowest energy minima correspond to the two configurations with all spins aligned, either upwards ($\sigma_i = +1, \forall i$) or downwards ($\sigma_i = -1, \forall i$), these configurations being symmetric under spin-flip $\sigma_i \to -\sigma_i$. Therefore, the thermodynamics of the Curie-Weiss model is solved: energy minimization tends to align the spins (as the lowest energy states are the two ordered ones), however entropy maximization tends to randomize the spins (as the higher the entropy, the most disordered the states, with half spins up and half spins down): the interplay between the two principles is driven by the level of noise introduced in the system and this is in turn ruled by the tunable parameter $\beta \equiv 1/T$ as coded in the definition of free energy.

A crucial bridge between condensed matter and neural network could now be sighted: one could think at each spin as a basic neuron, retaining only its ability to spike such that $\sigma_i = +1$ and $\sigma_i = -1$ represent firing and quiescence, respectively, and associate to each equilibrium configuration of this spin system a *stored pattern* of information. The reward is that, in this way, the spontaneous (i.e. thermodynamical) tendency of the network to relax on free-energy minima can be related to the spontaneous retrieval of the stored pattern, such that the cognitive capability emerges as a natural consequence of physical principles.

The route from Curie-Weiss to Hopfield

Actually, the Hamiltonian (1) would encode for a rather poor model of neural network as it would account for only two stored patterns, corresponding to the two possible minima (that in turn would represent pathological network's behavior with all the neurons contemporarily completely firing of completely silenced).

This criticism can be easily overcome thanks to the Mattis-gauge, namely a re-definition of the spins via $\sigma_i \to \xi_i^1 \sigma_i$, where $\xi_i^1 = \pm 1$ are random entries extracted with equal probability and kept fixed in the network (in statistical mechanics these are called *quenched* variables to stress that they do not contribute to thermalization). Fixing $J \equiv 1$ for simplicity, the Mattis Hamiltonian reads as

$$H_N^{Mattis}(\sigma|\xi) = -\frac{1}{N} \sum_{i<j}^{N} \xi_i^1 \xi_j^1 \sigma_i \sigma_j - h \sum_{i}^{N} \xi_i^1 \sigma_i. \tag{7}$$

The Mattis magnetization is defined as $m_1 = \sum_i \xi_i^1 \sigma_i$. To inspect its lowest energy minima, we perform a comparison with the CW model: in terms of the (standard) magnetization, the Curie-Weiss model reads as $H_N^{CW} \sim -(N/2)m^2 - hm$ and, analogously we can write $H_N^{Mattis}(\sigma|\xi)$ in terms of Mattis magnetization as $H_N^{Mattis} \sim -(N/2)m_1^2 - hm_1$. It is then evident that, in the low noise limit (namely where collective properties may emerge), as the minimum of free energy is achieved in the Curie-Weiss model for $\langle m \rangle \to \pm 1$, the same holds in the Mattis model for $\langle m_1 \rangle \to \pm 1$. However, this implies that now spins tend to align parallel (or antiparallel) to the vector ξ^1, hence if the latter is, say, $\xi^1 = (+1, -1, -1, -1, +1, +1)$ in a model with $N = 6$, the equilibrium configurations of the network will be $\sigma = (+1, -1, -1, -1, +1, +1)$ and $\sigma = (-1, +1, +1, +1, -1, -1)$, the latter due to the gauge symmetry $\sigma_i \to -\sigma_i$ enjoyed by the Hamiltonian. Thus, the network relaxes autonomously to a state where some of its neurons are firing while others are quiescent, according to the *stored pattern* ξ^1.

If we want a network able to cope with P patterns, the starting Hamiltonian should have simply the sum over these P previously stored[1] patterns, namely

$$H_N(\sigma|\xi) = -\frac{1}{N} \sum_{i<j}^{N} \left(\sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu \right) \sigma_i \sigma_j, \tag{8}$$

where we neglect the external field ($h = 0$) for simplicity. As we will see in the next section, this Hamiltonian constitutes indeed the Hopfield model, namely the harmonic oscillator of neural networks, whose coupling matrix is called *Hebb matrix* as encodes the Hebb prescription for neural organization.

[1]The part of neural network's theory we are analyzing is meant for spontaneous retrieval of already stored information -grouped into patterns (pragmatically vectors)-. Clearly it is assumed that the network has already overpass the learning stage.

Despite the extension to the case $P > 1$ is formally straightforward, the investigation of the system as P grows becomes by far more tricky. Indeed, neural networks belong to the so-called "complex systems" realm. We propose that complex behaviors can be distinguished by simple behaviors as for the latter the number of free-energy minima of the system *does not scale* with the volume N, while for complex systems the number of free-energy minima *does scale* with the volume according to a proper function of N. For instance, the Curie-Weiss/Mattis model has two minima only and it constitutes the paradigmatic example for a simple system. As a counterpart, the prototype of complex system is the Sherrington-Kirkpatrick model (SK) that has an amount of minima that scales $\propto \exp(cN)$ with $c \neq f(N)$, and its Hamiltonian reads as

$$H_N^{SK}(\sigma|J) = \frac{1}{\sqrt{N}} \sum_{i<j}^{N} J_{ij}\sigma_i\sigma_j, \tag{9}$$

where, crucially, coupling are Gaussian distributed as $P(J_{ij}) \equiv \mathcal{N}[0,1]$. This implies that links can be either positive (hence favoring parallel spin configuration) as well as negative (hence favoring anti-parallel spin configuration), thus, in the large N limit, with large probability, spins will receive conflicting signals and we speak about "frustrated networks".

The mean-field statistical mechanics for the low-noise behavior of spin-glasses has been first described by Giorgio Parisi and it predicts a hierarchical organization of states and a relaxational dynamics spread over many timescales. Here we just need to know that their natural order parameter is no longer the magnetization (as these systems do not magnetize), but the *overlap* q_{ab}, as we are explaining. Spin glasses are balanced ensembles of ferromagnets and antiferromagnets (this can also be seen mathematically as $P(J)$ is symmetric around zero) and, as a result, $\langle m \rangle$ is always equal to zero, on the other hand, a comparison between two realizations of the system (pertaining to the same coupling set) is meaningful because at large temperatures it is expected to be zero, as everything is uncorrelated, but at low temperature their overlap is strictly non-zero as spins freeze in disordered but correlated states. More precisely, given two "replicas" of the system, labeled as a and b, their overlap q_{ab} can be defined as the scalar product between the related spin configurations, namely as $q_{ab} = (1/N)\sum_i^N \sigma_i^a \sigma_i^b$, thus the mean-field spin glass has a completely random paramagnetic phase, with $\langle q \rangle \equiv 0$ and a "glassy phase" with $\langle q \rangle > 0$ split by a phase transition at $\beta_c = T_c = 1$.

The Sherrington-Kirkpatrick model displays a large number of minima as expected for a cognitive system, yet it is not suitable to act as a cognitive system because its states are too "disordered". We look for an Hamiltonian whose minima are not purely random like those in

SK, as they must represent ordered stored patterns (hence like the CW ones), but the amount of these minima must be possibly extensive in the number of spins/neurons N (as in the SK and at contrary with CW), hence we need to retain a "ferromagnetic flavor" within a "glassy panorama": we need *something in between.*

Remarkably, the Hopfield model defined by the Hamiltonian (8) lies exactly in between a Curie-Weiss model and a Sherrington-Kirkpatrick model. Let us see why: when $P = 1$ the Hopfield model recovers the Mattis model, which is nothing but a gauge-transformed Curie-Weiss model. Conversely, when $P \to \infty$, $(1/\sqrt{N}) \sum_{\mu}^{P} \xi_i^{\mu} \xi_j^{\mu} \to \mathcal{N}[0, 1]$, by the standard central limit theorem, and the Hopfield model recovers the Sherrington-Kirkpatrick one. In between these two limits the system behaves as an associative network.

Such a crossover between CW (or Mattis) and SK models, requires for its investigation both the P Mattis magnetization $\langle m_{\mu} \rangle$, $\mu = (1, ..., P)$ (for quantifying retrieval of the whole stored patterns, that is the *vocabulary*), and the two-replica overlaps $\langle q_{ab} \rangle$ (to control the glassyness growth if the vocabulary gets enlarged), as well as a tunable parameter measuring the ratio between the stored patterns and the amount of available neurons, namely $\alpha = \lim_{N \to \infty} P/N$, also referred to as *network capacity.*

As far as P scales sub-linearly with N, i.e. in the low storage regime defined by $\alpha = 0$, the phase diagram is ruled by the noise level β only: for $\beta < \beta_c$ the system is a paramagnet, with $\langle m_{\mu} \rangle = 0$ and $\langle q_{ab} \rangle = 0$, while for $\beta > \beta_c$ the system performs as an attractor network, with $\langle m_{\mu} \rangle \neq 0$ for a given μ (selected by the external field) and $\langle q_{ab} \rangle = 0$. In this regime no dangerous glassy phase is lurking, yet the model is able to store only a tiny amount of patterns as the capacity is sub-linear with the network volume N.
Conversely, when P scales linearly with N, i.e. in the high-storage regime defined by $\alpha > 0$, the phase diagram lives in the α, β plane.When α is small enough the system is expected to behave similarly to $\alpha = 0$ hence as an associative network (with a particular Mattis magnetization positive but with also the two-replica overlap slightly positive as the glassy nature is intrinsic for $\alpha > 0$). For α large enough ($\alpha > \alpha_c(\beta), \alpha_c(\beta \to \infty) \sim 0.14$) however, the Hopfield model collapses on the Sherrington-Kirkpatrick model as expected, hence with the Mattis magnetizations brutally reduced to zero and the two-replica overlap close to one. The transition to the spin-glass phase is often called "blackout scenario" in neural network community. Making these predictions quantitative is a non-trivial task in statistical mechanics and, nowadays several techniques are available, among which we quote the replica-trick (originally used by the pioneers Amit-Gutfreund-Sompolinsky [16]), the martingale method (originally devel-

oped by Pastur, Sherbina and Tirozzi [89]) and the cavity field technique (recently developed by Guerra in [25, 26, 30]).

Although the Hopfield model has been extensively studied since it was introduced in [70], both from a physical [17, 33, 47, 51] and a more mathematical [14, 89, 90, 34, 35, 36, 37, 96, 97] point of view, from the rigorous perspective many points about its properties remain unsolved, which also prompts further efforts in developing new mathematical techniques and different physical perspectives. Obtaining rigourus results is simpler if we introduce an analogical version of the standard Hopfield model, namely a mean-field structure with N dichotomic neurons (spins) interconnected through Hebbian couplings whose P patterns are stored according to a standard Gaussian $N[0,1]$. Some of these results can be easily translated for the original dichotomic model.

Turning to the applied side, despite the fact that in neural networks (in their original artificial intelligence framework) the interest in continuous patterns is reduced or moved to rotators (e.g. Kuramoto oscillators [73]), as digital processing by Ising spins works as a better approximation for the standard integrate and fire models of neurons [39], in several other fields of science (as, for instance, in chemical kinetics [49, 50] or theoretical immunology [7]) continuous values of patterns can instead be preferred ([33, 46]) and a rigorous mathematical control of completely continuous models (namely with both continuous patterns and neurons) has to be certainly considered, for further investigations, as necessary and worthwhile. For the moment, we limit ourselves in presenting a clear scenario for the hybrid model made of by continuous patterns and dichotomic variables.

Chapter 1

Hopfield model and bipartite networks

Within our approach, the equilibrium statistical mechanics of the neural network is shown to be equivalent to the one of a bipartite spin glass [27, 28] whose parts consist of the original N neurons (belonging to the first party, hence made of by dichotomic variables) and the other hand P Gaussians that give rise to the second part (hence consisting of continuous variables). This is a very useful perspective in order to model real biological networks: in the second part of this thesis, we will use the fact that a biological system in which two or more parties of units interact can be be mapped and studied into a marginalized neural Hopfield-like network concerning just one party. In this way it is possible to read the features of a bipartite systems in terms of attractors of a neural networks able to retrieve patterns of informations previously stored.

Before proceeding in that direction, mainly following [25], a walk in the opposite direction can be very interesting, understanding how an Hopfield model is equivalent to a bipartite system and investigating the deep relation among this two statistical mechanics model.

1.1 The analogical Hopfield model

We introduce a large network of N two-state neurons $(1,..,N) \ni i \to \sigma_i = \pm 1$, which are thought of as quiescent when their value is -1 or spiking when their value is $+1$. They interact throughout a synaptic matrix J_{ij} defined according to the Hebb rule for learning [68, 70]

$$J_{ij} = \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu. \tag{1.1}$$

Each random variable $\xi^\mu = \{\xi_1^\mu, .., \xi_N^\mu\}$ represents a learned pattern: While in the standard literature these patterns are usually chosen at random independently with values ± 1 taken with equal probability $1/2$, we chose them as taking real values with a unit Gaussian probability distribution, *i.e.*

$$d\mu(\xi_i^\mu) = \frac{1}{\sqrt{2\pi}} e^{-(\xi_i^\mu)^2/2}. \quad (1.2)$$

The analysis of the network assumes that the system has already stored p patterns (no learning is investigated here), and we will be interested in the case in which this number asymptotically increases linearly with respect to the system size (high storage level), so that $p/N \to \alpha$ as $N \to \infty$, where $\alpha \geq 0$ is a parameter of the theory denoting the storage level.
The Hamiltonian of the model has a mean-field structure and involves interactions between any pair of sites according to the definition

$$H_N(\sigma; \xi) = -\frac{1}{N} \sum_{\mu=1}^{p} \sum_{i<j}^{N} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j. \quad (1.3)$$

1.1.1 Morphism in the bipartite model

By splitting the summations $\sum_{i<j}^N = \frac{1}{2}\sum_{ij}^N - \frac{1}{2}\sum_i^N \delta_{ij}$ in the Hamiltonian (1.3), we can introduce and write the partition function $Z_{N,p}(\beta;\xi)$ in the following form

$$\begin{aligned} Z_{N,p}(\beta;\xi) &= \sum_\sigma \exp\Big(\frac{\beta}{2N} \sum_{\mu=1}^{p} \sum_{ij}^{N} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j - \frac{\beta}{2N} \sum_{\mu=1}^{p} \sum_{i}^{N} (\xi_i^\mu)^2 \Big) \quad (1.4) \\ &= \tilde{Z}_{N,p}(\beta;\xi) e^{\frac{-\beta}{2N} \sum_{\mu=1}^{p} \sum_{i=1}^{N} (\xi_i^\mu)^2} \end{aligned}$$

where $\beta \geq 0$ is the inverse temperature, and denotes here the level of noise in the network. We have defined

$$\tilde{Z}_{N,p}(\beta;\xi) = \sum_\sigma \exp(\frac{\beta}{2N} \sum_{\mu=1}^{p} \sum_{ij}^{N} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j). \quad (1.5)$$

Notice that the last term at the r.h.s. of eq. (1.4) does not depend on the particular state of the network, hence the control of the last term can be easily obtained [30] and simply adds a factor $\alpha\beta/2$ to the free energy.
Consequently we focus just on $\tilde{Z}(\beta;\xi)$. Let us apply the Hubbard-Stratonovich lemma [54] to linearize with respect to the bilinear quenched memories carried by the $\xi_i^\mu \xi_j^\mu$.
We can write

$$\tilde{Z}_{N,p}(\beta;\xi) = \sum_\sigma \int (\prod_{\mu=1}^{p} \frac{dz_\mu \exp(-z_\mu^2/2)}{\sqrt{2\pi}}) \exp(\sqrt{\beta/N} \sum_{i,\mu} \xi_i^\mu \sigma_i z_\mu). \quad (1.6)$$

For a generic function F of the neurons, we define the Boltzmann state $\omega_\beta(F)$ at a given level of noise β as the average

$$\omega_\beta(F) = \omega(F) = (Z_{N,p}(\beta;\xi))^{-1} \sum_\sigma F(\sigma) e^{-\beta H_N(\sigma;\xi)} \tag{1.7}$$

and often we will drop the subscript β for the sake of simplicity. The s-replicated Boltzmann state is defined as the product state $\Omega = \omega^1 \times \omega^2 \times ... \times \omega^s$, in which all the single Boltzmann states are at the same noise level β^{-1} and share an identical sample of quenched memories ξ. For the sake of clearness, given a function F of the neurons of the s replicas and using the symbol $a \in [1,..,s]$ to label replicas, such an average can be written as

$$\Omega(F(\sigma^1,...,\sigma^s)) = \frac{1}{Z_{N,p}^s} \sum_{\sigma^1}\sum_{\sigma^2}...\sum_{\sigma^s} F(\sigma^1,...,\sigma^s) \exp(-\beta \sum_{a=1}^s H_N(\sigma^a,\xi)). \tag{1.8}$$

The average over the quenched memories will be denoted by $\mathbb{E}$ and for a generic function of these memories $F(\xi)$ can be written as

$$\mathbb{E}[F(\xi)] = \int (\prod_{\mu=1}^p \prod_{i=1}^N \frac{d\xi_i^\mu e^{-\frac{(\xi_i^\mu)^2}{2}}}{\sqrt{2\pi}}) F(\xi) = \int F(\xi) d\mu(\xi), \tag{1.9}$$

with $\mathbb{E}[\xi_i^\mu] = 0$ and $\mathbb{E}[(\xi_i^\mu)^2] = 1$.
Hereafter we will often denote the average over the gaussian spins as $d\mu(z)$. We use the symbol $\langle . \rangle$ to mean $\langle . \rangle = \mathbb{E}\Omega(.)$.
We recall that in the thermodynamic limit it is assumed

$$\lim_{N\to\infty} \frac{p}{N} = \alpha,$$

α being a given real number, which acts as free parameter of the theory.

1.1.2 The thermodynamical observables

The main quantities of interest are the intensive pressure, defined as

$$\lim_{N\to\infty} A_{N,p}(\beta,\xi) = -\beta \lim_{N\to\infty} f_{N,p}(\beta,\xi) = \lim_{N\to\infty} \frac{1}{N} \ln Z_{N,p}(\beta;\xi), \tag{1.10}$$

the quenched intensive pressure, defined as

$$\lim_{N\to\infty} A^*_{N,p}(\beta) = -\beta \lim_{N\to\infty} f^*_{N,p}(\beta) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E} \ln Z_{N,p}(\beta;\xi), \tag{1.11}$$

and the annealed intensive pressure, defined as

$$\lim_{N\to\infty} \bar{A}_{N,p}(\beta) = -\beta \lim_{N\to\infty} \bar{f}_{N,p}(\beta) = \lim_{N\to\infty} \frac{1}{N} \ln \mathbb{E} Z_{N,p}(\beta;\xi). \tag{1.12}$$

According to thermodynamics, here $f_{N,p}(\beta,\xi) = u_{N,p}(\beta,\xi) - \beta^{-1} s_{N,p}(\beta,\xi)$ is the free energy density, $u_{N,p}(\beta,\xi)$ is the internal energy density and $s_{N,p}(\beta,\xi)$ is the intensive entropy (the star and the bar denote the quenched and the annealed evaluations as well).

According to the exploited bipartite nature of the Hopfield model, we introduce two other order parameters: the first is the overlap between the replicated neurons, defined as

$$q_{ab} = \frac{1}{N}\sum_{i=1}^{N} \sigma_i^a \sigma_i^b \in [-1,+1], \tag{1.13}$$

and the second the overlap between the replicated Gaussian variables z, defined as

$$p_{ab} = \frac{1}{p}\sum_{\mu=1}^{p} z_\mu^a z_\mu^b \in (-\infty,+\infty). \tag{1.14}$$

These overlaps play a considerable role in the theory as they can express thermodynamical quantities.

1.2 A detailed description of the annealed region

1.2.1 The interpolation scheme for the annealing

In this section we present the main idea of the work, used here to get a complete control of the high-temperature region: We interpolate between the neural network (described in terms of a bipartite spin glass) and a system consisting of two separate spin glasses, one dichotomic and one Gaussian. Note that, by the Jensen inequality, namely

$$\mathbb{E} \ln Z_{N,p}(\beta) \le \ln \mathbb{E} Z_{N,p}(\beta),$$

we can write

$$A^*_{N,p} \le \frac{1}{N} \ln \mathbb{E} \sum_\sigma \int \prod_{\mu=1}^{p} d\mu(z_\mu) e^{\sqrt{\frac{\beta}{N}}\sum_{i\mu}\xi_i^\mu \sigma_i z_\mu} = \ln 2 - \frac{p}{2N}\log(1-\beta), \tag{1.15}$$

where we emphasize that the integral inside eq. (1.15) exists only for $\beta < 1$.

The $N \to \infty$ limit then offers immediately $\lim_{N\to\infty} A^*_{N,p}(\beta) \le \ln 2 - \alpha \ln(1-\beta)/2$. The next step is to use interpolation to prove the validity of the Jensen bound in the whole region defined by the line $\beta_c = 1/(1+\sqrt{\alpha})$, which defines the boundary of the validity of the annealed approximation, in complete agreement with the well known picture of Amit, Gutfreund and Sompolinsky [15][17].

To understand which is the proper interpolating structure, let us note that the exponent of the Boltzmann factor yields a family of random variables indexed by the configurations (σ, z). For a given realization of the noise, $H(\sigma, z|\xi) = \sqrt{\frac{\beta}{N}}\sum_{i\mu}\xi_{i,\mu}\sigma_i z_\mu$ is a randomly centered variable with variance

$$\mathbb{E}(H(\sigma, z|\xi)H(\sigma', z'|\xi)) = \frac{\beta}{N}\sum_{i\mu}\sigma_i\sigma'_i z_\mu z'_\mu = \beta p q_{\sigma\sigma'}p_{zz'}.$$

The presence of the product $q_{\sigma\sigma'}p_{zz'}$ in the variance suggests the correct interpolating structure among this bipartite network and two other independent spin glasses, namely a Sherrington-Kirkpatrick model with variance $q^2_{\sigma\sigma'}$ and another spin glass model with Gaussian spin and variance $p^2_{zz'}$. It is in fact clear that a proper interpolating structure can be held by

$$\begin{aligned}\varphi_N(t) &= \frac{1}{N}\mathbb{E}\ln\sum_\sigma\int\prod_{\mu=1}^{p}d\mu(z_\mu)\exp\big(\sqrt{t}\sqrt{\frac{\beta}{N}}\sum_{i\mu}\xi_i^\mu\sigma_i z_\mu\big) \qquad (1.16)\\ &\cdot\quad \exp\big(\sqrt{1-t}(\beta_1\sqrt{\frac{N}{2}}K(\sigma)+\beta_2\sqrt{\frac{p}{2}}\bar{K}(z))\big)\\ &\cdot\quad \exp\big((1-t)(\frac{p\beta}{2}p_{zz}-\frac{p\beta_2^2}{4}p^2_{zz})\big),\end{aligned}$$

where we have set

$$K(\sigma) = \frac{1}{N}\sum_{ij}J_{ij}\sigma_i\sigma_j$$

and

$$\bar{K}(z) = \frac{1}{p}\sum_{ij}\bar{J}_{ij}z_i z_j$$

and the average $\mathbb{E}$ is taken with respect to all the i.i.d. normal random variables $\xi_{ij}, J_{ij}, \bar{J}_{ij}$. The interpolation is performed such that for $t = 1$ the interpolating structure $\varphi(t = 1)$ returns the free energy of the bipartite model, namely of the neural network, while for $t = 0$ it coincides with a factorization in an SK spin glass and a (suitably regularized) Gaussian one (see Appendix A); β_1, β_2, which will be then fixed as opportune noise levels, for the moment are simply free parameters.

As in [26][64], the plan is now to evaluate the flow under a changing t of the interpolating structure in order to get a positive defined sum rule by tuning opportunely β_1, β_2; hence, if we generalize the states as $\langle . \rangle_t = \mathbb{E}\Omega_t$, where the subscript t accounts for the extended interpolating structure defined in (1.16) we can write

$$\begin{aligned}\frac{d\varphi_N(t)}{dt} &= \frac{1}{N}\frac{1}{2}\beta p\Big(\langle p_{zz}\rangle_t - \langle q_{\sigma\sigma'}p_{zz'}\rangle_t\Big) - \frac{1}{4}\beta_1^2\Big(1-\langle q^2_{\sigma\sigma'}\rangle_t\Big) + \qquad (1.17)\\ &- \frac{p}{N}\frac{1}{4}\beta_2^2\Big(\langle p^2_{zz}\rangle_t - \langle p^2_{zz'}\rangle_t\Big) + \frac{p}{N}\frac{1}{4}\beta_2^2\langle p^2_{zz}\rangle_t - \frac{\beta}{2}\frac{p}{N}\langle p_{zz}\rangle_t, \qquad (1.18)\end{aligned}$$

then, calling $\alpha = p/N$ even at finite size N (with a little language abuse), we can write

$$\frac{d\varphi_N(t)}{dt} = -\frac{\beta_1^2}{4} + \frac{1}{4}\langle \beta_1^2 q_{\sigma\sigma'}^2 + \alpha\beta_2^2 p_{zz'}^2 - 2\alpha\beta q_{\sigma\sigma'} p_{zz'}\rangle_t. \tag{1.19}$$

If we now impose on β_1, β_2 the constraint $\beta_1\beta_2 = \sqrt{\alpha}\beta$ we get a perfect square in the brackets of the flow under a changing t, and calling $S_t(\alpha,\beta) = \langle(\beta_1 q_{\sigma\sigma'} - \sqrt{\alpha}\beta_2 p_{zz'})^2\rangle_t$ the source term, we can write

$$\frac{d\varphi_N}{dt} \geq -\frac{1}{4}\beta_1^2 + S_t(\alpha,\beta). \tag{1.20}$$

We can then integrate back between $[0,1]$ to get the following inequality

$$\begin{aligned}
\varphi_N(1) &= \frac{1}{N}\mathbb{E}\ln\sum_\sigma \int \prod_\mu^p d\mu(z_\mu) e^{\sqrt{\frac{\beta}{N}}\sum_{i\mu}\xi_i^\mu \sigma_i z_\mu} \\
&\geq \frac{1}{N}\mathbb{E}\ln\sum_\sigma e^{\beta_1\sqrt{\frac{N}{2}}K(\sigma)} \\
&- \frac{\beta_1^2}{4} + \frac{p}{N}\frac{1}{p}\mathbb{E}\ln\int\prod_\mu d\mu(z_\mu) e^{\beta_2\sqrt{\frac{p}{2}}\bar{K}(z)} e^{-\frac{\beta_2^2 p}{4}p_{zz'}} e^{\frac{p}{2}\beta p_{zz}},
\end{aligned}$$

under the constraint $\beta_1\beta_2 = \sqrt{\alpha}\beta$.

Note that $K(\sigma)$ in the above expression defines the SK-model, while the last term defines the regularized Gaussian spin glass deeply investigated in Appendix A (see also [60]).

Now the advantages of this interpolation scheme become evident: As we have extremely satisfactory descriptions of the two independent models, namely the SK and the Gaussian spin glass, by these properties we can infer the behavior of the neural network (again thought of as the bipartite spin glass).

In particular, we know that the free energies of each single part spin glass approach their annealed expression in the region where $\beta_1 \leq 1$ [98] and $\beta + \beta_2 \leq 1$. Within this region, at the r.h.s. of eq. (1.21) we get, in the thermodynamic limit, exactly $\ln 2 - (\alpha/2)\ln(1-\beta)$. Furthermore, if α and β respect the constraint $\beta(1+\sqrt{\alpha}) \leq 1$, then finding β_1, β_2 such that the conditions $(A), (B), (C)$ hold, being

$$\beta_1\beta_2 = \sqrt{\alpha}\beta \;\; (A), \quad \beta_1 \leq 1 \;\; (B), \quad \beta + \beta_1 \leq 1 \;\; (C),$$

is certainly possible. In particular, using the SK critical behavior for the sake of simplicity, hence posing $\beta_1 = 1$, and setting $\beta_2 = \sqrt{\alpha}\beta$, conditions (A) and (B) are automatically satisfied and, for the latter, being $\beta_2 = \sqrt{\alpha}\beta$, we get

$$\beta + \beta_2 \equiv \beta + \sqrt{\alpha}\beta = \beta(1+\sqrt{\alpha}) \leq 1,$$

such that also condition (C) is verified. We can then state the following

Theorem 1.2.1. *In the α, β plane there exist a critical line, defined by*

$$\beta_c(\alpha) = \frac{1}{1+\sqrt{\alpha}}, \tag{1.21}$$

such that for $\beta \leq \beta_c(\alpha)$ the annealed approximation of the free energy holds

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\ln\sum_\sigma \int \prod_\mu d\mu(z_\mu) e^{\left(\sqrt{\frac{\beta}{N}}\sum_{i\mu}\xi_i^\mu \sigma_i z_\mu\right)} = \ln 2 - \frac{\alpha}{2}\ln(1-\beta). \tag{1.22}$$

Remark 1. *We stress that the Borel-Cantelli lemma allows straightforwardly to determine the correct annealed regions for the SK model [98] and, through a careful check of convergence of the integral defining the partition function, the same holds for the Gaussian case too (see Appendix A); however, the direct application of the Borel-Cantelli argument on the neural network gives a weaker result as shown for instance in [30]. The interpolation scheme allows to exploit and transfer the results for the SK and Gaussian models to the neural network, and enlarges the area of validity of the annealed expression for the free energy to the whole expected region, obtained e.g. via the replica method [15].*

1.2.2 The control of the annealed region

As a consequence, we can now extend the previous results exposed in [30] to the whole annealed region: Summarizing, we get the following

Theorem 1.2.2. *There exists $\beta_c(\alpha)$, defined by eq. (1.21), such that for $\beta < \beta_c(\alpha)$ we have the following limits for the intensive free energy, internal energy and entropy, as $N \to \infty$ and $p/N \to \alpha > 0$:*

$$\begin{aligned} -\beta \lim_{N\to\infty} f_{N,p}(\beta;\xi) &= \lim_{N\to\infty} N^{-1}\ln Z_{N,p}(\beta;\xi) \qquad (1.23) \\ &= \ln 2 - (\alpha/2)\ln(1-\beta) - (\alpha\beta/2), \\ \lim_{N\to\infty} u_{N,p}(\beta;\xi) &= -\lim_{N\to\infty} N^{-1}\partial_\beta \ln Z_{N,p}(\beta;\xi) \qquad (1.24) \\ &= -\alpha\beta/(2(1-\beta)), \\ \lim_{N\to\infty} s_{N,p}(\beta;\xi) &= \lim_{N\to\infty} N^{-1}(\ln Z_{N,p}(\beta;\xi) - \beta\partial_\beta \ln Z_{N,p}(\beta;\xi)) \qquad (1.25) \\ &= \ln 2 - (\alpha/2)\ln(1-\beta) - (\alpha\beta^2)/(2(1-\beta)) - (\alpha\beta/2), \end{aligned}$$

ξ-almost surely. The same limits hold for the quenched averages, so that in particular

$$\lim_{N\to\infty} N^{-1}\mathbb{E}\ln Z_{N,p}(\beta;\xi) = \ln 2 - \frac{\alpha}{2}\ln(1-\beta) - \frac{\alpha\beta}{2},$$

where, in all these formulas, the last term, namely $-\alpha\beta/2$, arises due to the diagonal contribution of the complete partition function (1.4).

Theorem 1.2.3. *There exists $\beta_c(\alpha)$, defined by eq. (1.21), such that for $\beta < \beta_c(\alpha)$ we have the following convergence in distribution*

$$\ln \tilde{Z}_{N,p}(\beta;\xi) - \ln \mathbb{E}\tilde{Z}_{N,p}(\beta;\xi) \to C(\beta) + \chi S(\beta) \tag{1.26}$$

where χ is a unit Gaussian in $\mathcal{N}[0,1]$ and

$$C(\beta) = -\frac{1}{2}\ln\sqrt{1/(1-\sigma^2\beta^2\alpha)} \tag{1.27}$$

$$S(\beta) = \Big(\ln\sqrt{1/(1-\sigma^2\beta^2\alpha)}\Big)^{\frac{1}{2}}, \tag{1.28}$$

with $\sigma = (1-\beta)^{-1}$.

1.3 Extension to the replica symmetric solution

Once the correct interpolating structure is understood, and spurred by the observation that the replica symmetric expression for the quenched free energy of the three models, namely the analogical neural network, the SK spin glass and the Gaussian one, are well known and investigated (for instance in [64][21][59][22][30][24]) we want to push further the equivalence among neural network and spin glasses, giving a complete picture also of the replica symmetric approximation.

To this task, let us recall that the replica symmetric approximation of the quenched free energy of the analogical neural network $A^{RS}_{NN}(\alpha,\beta)$ is given by the following expression [30]

$$\begin{aligned} A^{RS}_{NN}(\alpha,\beta) &= \ln 2 + \int d\mu(z)\ln\cosh(z\sqrt{\alpha\beta\bar{p}}) + \frac{\alpha}{2}\ln(\frac{1}{1-\beta(1-\bar{q})}) + \\ &+ \frac{\alpha\beta}{2}\frac{\bar{q}}{1-\beta(1-\bar{q})} - \frac{\alpha\beta}{2}\bar{p}(1-\bar{q}), \end{aligned} \tag{1.29}$$

where the order parameters denoted with a bar (to mean their RS approximation) are given by

$$\bar{q} = \int d\mu(z)\tanh^2\Big(z\sqrt{\alpha\beta\bar{p}}\Big), \tag{1.30}$$

$$\bar{p} = \beta\bar{q}/\Big(1-\beta(1-\bar{q})\Big)^2. \tag{1.31}$$

Let us introduce further β_1 and β_2 as

$$\beta_1 = \frac{\sqrt{\alpha}\beta}{1-\beta(1-\bar{q})}, \tag{1.32}$$

$$\beta_2 = 1-\beta(1-\bar{q}), \tag{1.33}$$

such that $\beta_1\beta_2 = \sqrt{\alpha}\beta$. We need also the RS approximation $A_{SK}^{RS}(\beta_1)$ of the quenched free energy of the SK model, at the noise level β_1, namely

$$A_{SK}^{RS}(\beta_1) = \ln 2 + \int d\mu(z) \ln\cosh(\beta_1\sqrt{\bar{q}_{SK}}z) + \frac{1}{4}\beta_1^2(1-\bar{q}_{SK})^2, \tag{1.34}$$

where

$$\bar{q}_{SK} = \int d\mu(z) \tanh^2\left(\beta_1 z\sqrt{\bar{q}_{SK}}\right). \tag{1.35}$$

By a direct comparison among the overlap expressions (1.30, 1.35) we immediately conclude that we must have

$$\beta_1^2\bar{q}_{SK} = \alpha\beta\bar{p},$$

which indeed holds as it can be verified easily, bearing in mind the expression (1.31) and (1.32) for $\bar{p}$ and β_1.

As a last ingredient we need to introduce also the replica symmetric expression $A_{Gauss}^{RS}(\beta_2,\beta)$ of the Gaussian spin glass at a noise level β_2 as (see Appendix A)

$$A_{Gauss}^{RS}(\beta_2,\beta) = \frac{1}{2}\ln\sigma + \frac{1}{2}\beta_2^2\bar{p}_G\sigma^2 + \frac{1}{4}\beta_2^2\bar{p}_G^2, \tag{1.36}$$

where

$$\bar{p}_G = (\beta_2-(1-\beta))/\beta_2^2, \tag{1.37}$$

$$\sigma^2 = 1/(1-\beta+\beta^2\bar{p}_G). \tag{1.38}$$

Note that the definition of the overlap between continuous variables encoded by eq. (1.31) is in perfect agreement with the same overlap defined within the framework of eq.(1.37), because, being $\beta_2 = 1-\beta(1-\bar{q})$, we can write

$$\bar{p}_{Gauss} = \frac{\beta_2-(1-\beta)}{\beta_2^2} = \frac{1-\beta(1-\bar{q})-(1-\beta)}{(1-\beta(1-\bar{q}))^2} = \frac{\beta\bar{q}}{(1-\beta(1-\bar{q}))^2}. \tag{1.39}$$

As a consequence, through a direct verification by comparison (that we omit as it is long and straightforward), we can state the following

Theorem 1.3.1. *Fixed, at noise level β, β_1 and β_2 as in (1.32) and (1.33), the replica symmetric approximation of the quenched free energy of the analogical neural network can be linearly decomposed in terms of the replica symmetric approximation of the Sherrington-Kirkpatrick quenched free energy, at noise level β_1, and the replica symmetric approximation of the quenched free energy of the Gaussian spin glass, at noise level β_2, such that*

$$A^{RS}_{NN}(\beta) = A^{RS}_{SK}(\beta_1) - \frac{1}{4}\beta_1^2 + \alpha A_{Gauss}(\beta_2, \beta), \tag{1.40}$$

and the inequality (1.21) becomes an identity for the RS behavior.

Remark 2. *We stress that the above Theorem is in agreement with the sum rule (1.20) of Section 2 as, in the replica symmetric approximation, $q_{\sigma\sigma'} = \bar{q}$ and $p_{zz'} = \bar{p}$, hence*

$$\beta_1\bar{q} - \sqrt{\alpha}\beta_2\bar{p} = \frac{\sqrt{\alpha}\beta\bar{q}}{(1-\beta(1-\bar{q}))^2} - \sqrt{\alpha}\Big(1-\beta(1-\bar{q})\Big)\frac{\beta\bar{q}}{(1-\beta(1-\bar{q}))^2} = 0. \tag{1.41}$$

Remark 3. *Approaching the high-temperature region we have $\bar{q} \to 0$ and $\bar{p} \to 0$, and clearly $\beta \to 1/(1+\sqrt{\alpha})$. As a consequence we have*

$$\begin{aligned} \beta_2 &= 1-\beta(1-\bar{q}) \to 1 - 1/(1+\sqrt{\alpha}), && (1.42)\\ \beta_1 &= \frac{\sqrt{\alpha}\beta}{1-\beta(1-\bar{q})} \to 1, && (1.43) \end{aligned}$$

then $\beta + \beta_2 \to 1$, such that also the single-party counterparts approaches their critical points.

Coherently, inside the annealed region we get $\bar{q} = 0$, then with the expressions for β_1, β_2 we can write $\beta_2 + \beta = 1$ that is the boundary of the annealed region for the Gaussian spin glass, while $\beta_1 = \sqrt{\alpha}\beta/(1-\beta)$ because $\beta \le 1/(1+\sqrt{\alpha})$ we get $\beta_1 \le 1$, namely the annealed region of the SK model.

These results open very interesting perspectives. The structure of the neural network as a linear combination of spin glasses is very rich: in fact we know that, as the SK model presents a very glassy full RSB structure [65], in the Gaussian one this is absent, since the true solution is infact RS even without external field (see Appendix A). Thus one could aspect in our analogical neural network a competition of these two effects: rather a new feature in the complex systems scenario, that has to be deeply investigated. Furthermore, the analogical model shares many features with the original Hopfield model (which is even harder from a mathematical point of view) for which one could study in what measure this structure is preserved. Future outlooks should cover also the completely analogical model in order to develop mathematical techniques beyond the standard ones required in artificial intelligence and closer to system biology.

Part II

Multitasking networks

Chapter 2

Diluted Hopfield models

In the previous chapter we illustrated the connection between the Hopfield model and a bipartite spin glass system. We started from a system of two sets of spins, $\sigma_i,\ i = 1,\dots,N$ and $\tau_\mu,\ \mu = 1,\dots,P$, connected by links ξ_i^μ and described by the Spin Glass Hamiltonian $H_{SG}(\boldsymbol{\sigma},\boldsymbol{\tau}|\boldsymbol{\xi}) \propto -\sum_{i,\mu}\xi_i^\mu\sigma_i\tau_\mu$. Marginalizing over $\boldsymbol{\tau}$ in the partition function $Z = \sum_{\boldsymbol{\sigma},\boldsymbol{\tau}} e^{-\beta H_{SG}(\boldsymbol{\sigma},\boldsymbol{\tau}|\boldsymbol{\xi})} = \sum_{\boldsymbol{\sigma}} e^{-\beta H_{NN}(\boldsymbol{\sigma}|\boldsymbol{\xi})}$, the $\boldsymbol{\sigma}$ represent a neural network with Hamiltonian $H_{NN}(\boldsymbol{\sigma}|\boldsymbol{\xi}) = -\beta^{-1}\sum_\mu \ln[2\cosh(\beta\sum_i \xi_i^\mu\sigma_i)]$ or, up to an additive constant, $H_{NN}(\boldsymbol{\sigma}|\boldsymbol{\xi})= -\frac{\beta}{2}\sum_{\mu,i,j}(\xi_i^\mu\xi_j^\mu)\sigma_i\sigma_j+\dots$. Higher order interactions are not written explicitly here; these are fully absent if the τ_i are continuous rather than discrete and have a Gaussian prior. Bipartite network are very useful in modelling biological systems, where agents of a different nature interact each other. In these kind of models the ξ's describe the topology of the interaction among different species. In a fully connected model, as the Hopfield one, each units can interact with all the others. This hypothesis is far from being realistic because the biology is not embedded in a mean field structure: there is always a kind of distance and the interaction surely decades between units that are very *far* each others. Consequently the concept of neighborhood is very important. The simpler modification we can carry to the model is reproducing at least the correct number of neighbors of a single agent of the bipartite network. To this task we consider diluted patterns entries, i.e. $\xi_i^\mu \in \{0,\pm1\}$, where the probability of a null entry is different from zero. As we are going to see, the most important result is that, while standard neural networks retrieved patterns sequentially (one at time), associative networks with diluted patterns, as resulting from the marginalization of a diluted bipartite spin glass, are able to accomplish parallel retrieval in appropriate dilution and load regimes [4, 3, 5].

2.1 Medium storage regime in extremely diluted connectivity: retrieval region

This section, mainly following [3], is dedicated to the statistical mechanics analysis of a network composed of N binary spins (σ_i, $i = 1, \dots, N$) and P patterns, such that the number ratio scales as

$$\lim_{N\to\infty} P/N^{\delta} = \alpha, \qquad \delta \in (0,1), \quad \alpha > 0. \tag{2.1}$$

We refer to it as the medium storage regime, meaning that the number of memorized patterns is growing with the size of the network but less than extensively. The effective interaction is described by the Hamiltonian

$$\mathcal{H}(\boldsymbol{\sigma}|\xi) = -\frac{1}{2N^{\tau}} \sum_{i,j=1}^{N} \sum_{\mu=1}^{P} \xi_i^{\mu} \xi_j^{\mu} \sigma_i \sigma_j, \tag{2.2}$$

where the patterns entries $\xi_i^{\mu} \in \{0, \pm 1\}$ are quenched random variables, independently and identically distributed according to

$$\mathbb{P}(\xi_i^{\mu} = 1) = \mathbb{P}(\xi_i^{\mu} = -1) = c/2N^{\gamma}, \qquad \mathbb{P}(\xi_i^{\mu} = 0) = 1 - c/N^{\gamma} \tag{2.3}$$

with $\gamma \in [0, 1)$. The parameter τ must be chosen such that $\mathcal{H}(\boldsymbol{\sigma}|\xi)$ scales linearly with N, and must therefore depend on γ and δ. Heuristically, since the number of non-zero entries $\mathcal{N}_{nz}$ in a generic pattern $(\xi_1^{\mu}, \dots, \xi_N^{\mu})$ is $\mathcal{O}(N^{1-\gamma})$, we expect that the network can retrieve a number of patterns of order $\mathcal{O}(N/\mathcal{N}_{nz}) = \mathcal{O}(N^{\gamma})$. We therefore expect to see changes in τ only when crossing the region in the (γ, δ) plane where pattern sparseness prevails over storage load (i.e. $\delta < \gamma$, where the system can recall all patterns), to the opposite situation, where the load is too high and frustration by multiple inputs on the same entry drives the network to saturation (i.e. $\delta > \gamma$). To validate this scenario, we carry out a statistical mechanical analysis, based on computing the free energy

$$f(\beta) = -\lim_{N\to\infty} \frac{1}{\beta N} \langle \log Z_N(\beta, \xi) \rangle_{\xi}. \tag{2.4}$$

2.1.1 Free energy computation and physical meaning of the parameters

If the number of patterns is sufficiently small compared to N, i.e. $\delta < 1$, we do not need the replica method; we can simply apply the steepest descent technique using the $P \ll N/\log N$ Mattis magnetizations as order parameters:

$$f(\beta) = -\frac{1}{\beta}\log 2 - \lim_{N\to\infty} \frac{1}{\beta N} \log \int d\mathbf{m}\; e^{-\frac{1}{2}\mathbf{m}^2 + N\left\langle \log\cosh\left(\sqrt{\beta/N^{\tau}}\,\boldsymbol{\xi}\cdot\mathbf{m}\right)\right\rangle_{\xi}}. \tag{2.5}$$

with $\mathbf{m} = (m_1, \ldots, m_{N_B})$, $\boldsymbol{\xi} = (\xi^1, \ldots, \xi^{N_B})$ and $\boldsymbol{\xi} \cdot \mathbf{m} = \sum_\mu \xi^\mu m_\mu$. Rescaling of the order parameters via $m_\mu \to m_\mu \sqrt{\beta} c N^{\tau/2+\theta}$ then gives

$$f(\beta) \quad = \quad -\frac{1}{\beta}\log 2 - \lim_{N\to\infty}\frac{1}{\beta N}\log\int \mathrm{d}\mathbf{m}\; e^{N\left(-\frac{\beta c^2}{2}N^{\tau+2\theta-1}\mathbf{m}^2 + \langle \log\cosh(\beta c N^\theta \boldsymbol{\xi}\cdot\mathbf{m})\rangle_\xi\right)}, \tag{2.6}$$

Hence, provided the limit exists, we may write via steepest descent integration:

$$f(\beta) = -\frac{1}{\beta}\log 2 - \frac{1}{\beta}\lim_{N\to\infty}\mathrm{extr}_{\mathbf{m}}\left[\left\langle \log\cosh\left(\beta c N^\theta \boldsymbol{\xi}\cdot\mathbf{m}\right)\right\rangle_\xi - \frac{\beta c^2}{2}N^{\tau+2\theta-1}\mathbf{m}^2\right]. \tag{2.7}$$

Differentiation with respect to the m_μ gives the self consistent equations for the extremum:

$$m_\mu = \frac{N^{1-\tau-\theta}}{c}\langle \xi^\mu \tanh(\beta c N^\theta \boldsymbol{\xi}\cdot\mathbf{m})\rangle_\xi. \tag{2.8}$$

With the additional new parameter θ, we now have two parameters with which to control separately two types of normalization: the normalization of the Hamiltonian, via τ, and the normalization of the order parameters, controlled by θ. To carry out this task properly, we need to understand the physical meaning of the order parameters. This is done in the usual way, by adding suitable external fields to the Hamiltonian:

$$\mathcal{H} \to \mathcal{H} - \sum_{\mu=1}^P \lambda_\mu \sum_{i=1}^N \xi_i^\mu \sigma_i \tag{2.9}$$

Now, with $\langle g(\boldsymbol{\sigma})\rangle_\sigma = Z_N^{-1}(\beta,\xi)\sum_{\boldsymbol{\sigma}} \mathrm{e}^{-\beta\mathcal{H}(\boldsymbol{\sigma}|\xi)}g(\boldsymbol{\sigma})$ and the corresponding new free energy $f(\beta,\boldsymbol{\lambda})$,

$$\lim_{N\to\infty}\frac{1}{N}\langle\sum_{i=1}^N \xi_i^\mu\sigma_i\rangle_\sigma = -\frac{\partial f(\beta,\boldsymbol{\lambda})}{\partial\lambda_\mu}\Big|_{\boldsymbol{\lambda}=0}, \tag{2.10}$$

with the short-hand $\boldsymbol{\lambda} = (\lambda_1 \ldots, \lambda_P)$. The new free energy is then found to be

$$f(\beta,\boldsymbol{\lambda}) = -\frac{1}{\beta}\log 2 - \frac{1}{\beta}\lim_{N\to\infty}\mathrm{extr}_{\mathbf{m}}\left[\left\langle \log\cosh\left(\beta\boldsymbol{\xi}\cdot[cN^\theta\mathbf{m}+\boldsymbol{\lambda}]\right)\right\rangle_\xi - \frac{\beta c^2}{2}N^{\tau+2\theta-1}\mathbf{m}^2\right] \tag{2.11}$$

Upon differentiation with respect to λ_μ we find (2.10) taking the form

$$\lim_{N\to\infty}\frac{1}{N}\langle\sum_{i=1}^N \xi_i^\mu\sigma_i\rangle_\sigma = \lim_{N\to\infty}\langle\xi^\mu\tanh(\beta c N^\theta\boldsymbol{\xi}\cdot\mathbf{m})\rangle_\xi. \tag{2.12}$$

We can then use expression (2.8) for m_μ to obtain the physical meaning of our order parameters:

$$\begin{aligned} m^\mu &= \lim_{N\to\infty}\frac{N^{1-(\tau+\theta)}}{c}\langle\xi^\mu\tanh(\beta c N^\theta\boldsymbol{\xi}\cdot\mathbf{m})\rangle_\xi \\ &= \lim_{N\to\infty}\langle\frac{1}{cN^{\tau+\theta}}\sum_{i=1}^N\xi_i^\mu\sigma_i\rangle_\sigma. \end{aligned} \tag{2.13}$$

Let us summarize the status of the various remaining control parameters in the theory, in the interest of transparency. Our model has three given external parameters:

- $\gamma \in [0,1)$: this quantifies the dilution of stored patterns, via $\mathbb{P}(\xi_i^\mu \neq 0) = cN^{-\gamma}$,
- $\delta \in (0,1)$ and $\alpha > 0$: these determine the number of stored patterns, via $\lim_{N\to\infty} P/N^\delta = \alpha$.

It also has two 'internal' parameters, which must be set in such a way for the statistical mechanical calculation to be self-consistent, i.e. such that various quantities scale in the physically correct way for $N \to \infty$:

- $\tau \geq 0$: this must ensure that the energy $\mathcal{H} = -\langle \frac{1}{2}N^{-\tau}\sum_{\mu=1}^P(\sum_{i=1}^N \xi_i^\mu \sigma_i)^2\rangle_\sigma$ scales as $\mathcal{O}(N)$,
- $\theta \geq 0$: this must ensure that the order parameter $m_\mu = \big\langle (1/cN^{\tau+\theta})\sum_i \xi_i^\mu \sigma_i \big\rangle_\sigma$ are of order $\mathcal{O}(1)$.

2.1.2 Setting of internal scaling parameters

To find the appropriate values for the internal scaling parameters θ and τ we return to the order parameter equation (2.8) and carry out the average over ξ^μ. This gives

$$m_\mu = \frac{N^{1-\tau-\theta}}{c}\langle \tanh\Big(\beta c N^\theta \Big((\xi^\mu)^2 m_\mu + \xi^\mu \sum_{\nu\neq\mu}^P \xi^\nu m_\nu\Big)\Big)\rangle_\xi. \tag{2.14}$$

$$= N^{1-\tau-\theta-\gamma}\langle \tanh\big(\beta c N^\theta \big(m_\mu + \sum_{\nu\neq\mu}^P \xi^\nu m_\nu\big)\big)\rangle_\xi. \tag{2.15}$$

Having non-vanishing m_μ in the limit $N \to \infty$ clearly demands $\theta + \tau \leq 1-\gamma$. If $\theta > 0$ the m_μ will become independent of β, which means that any phase transitions occur ar zero or infinite noise levels, i.e. we would not have defined the scaling of our Hamiltonian correctly. Similarly, if $\theta + \tau < 1-\gamma$ the effective local fields acting upon the σ_i (viz. the arguments of the hyperbolic tangent) and therefore also the expectation values $\langle \sigma_i \rangle_\sigma$, would be vanishingly weak. We therefore conclude that a natural ansatz for the free exponents is:

$$(\tau, \theta) = (1-\gamma, 0) \tag{2.16}$$

This simplifies the order parameter equation to

$$m_\mu = \langle \tanh\big(\beta c\big(m_\mu + \sum_{\nu\neq\mu}^P \xi^\nu m_\nu\big)\big)\rangle_\xi. \tag{2.17}$$

Let us analyze this equation further. Since $\mathbb{P}(\xi_i^\mu \neq 0) \sim N^{-\gamma}$ with $\gamma > 0$, we can for $N \to \infty$ replace in (2.14) the sum over $\nu \neq \mu$ with the sum over all μ; the difference is negligible in the thermodynamic limit. In this way it becomes clear that for each solution of (2.14) we have $m_\mu \in \{-m, 0, m\}$. Using the invariance of the free energy under $m_\mu \to -m_\mu$, we can from now on focus on solutions with non-negative magnetizations. If we denote with $K \leq P$ the number of μ with $m_\mu \neq 0$, then the value of $m > 0$ is to be solved from

$$m \;\; = \;\; \langle \tanh\big(\beta c m\big(1 + \sum_{\nu=1}^{K} \xi^\nu\big)\big)\rangle_\xi. \tag{2.18}$$

It is not a priori obvious how the number K of nonzero magnetizations can or will scale with N. We therefore set $K = \phi N^{\delta'}$, in which the condition $K \leq P$ then places the following conditions on ϕ and δ': $\delta' \in [0, \delta]$, and $\phi \in [0, \infty)$ if $\delta' < \delta$ or $\phi \in [0, \alpha]$ if $\delta' = \delta$. We expect that if K is too large, equation (2.18) will only have the trivial solution for $N \to \infty$, so there will be further conditions on ϕ and δ' for the system to operate properly. If $\delta' > \gamma$, the noise due to other condensed patterns (i.e. the sum over ν) becomes too high, and m can only be zero:

$$\mathbb{E}\left[\big(\sum_{\mu=1}^{K} \xi^\mu\big)^2\right] = \sum_{\mu=1}^{K} \mathbb{E}[\xi^{\mu 2}] = \phi c \frac{N^{\delta'}}{N^\gamma} \to \infty. \tag{2.19}$$

On the other hand, if $\delta' < \gamma$ this noise becomes negligible, and (2.18) reduces to the Curie-Weiss equation, whose solution is just the Mattis magnetization [45, 15, 20]. It follows that the critical case is the one where when $\delta' = \gamma$. Here we have for $N \to \infty$ the following equation for m:

$$m \;\; = \;\; \sum_{k \in \mathcal{Z}} \pi(k|\phi) \tanh(\beta c m(1 + k)) \tag{2.20}$$

with the following discrete noise distribution, which obeys $\pi(-k|\phi) = \pi(k|\phi)$:

$$\pi(k|\phi) \;\; = \;\; \big\langle \delta_{k, \sum_{\mu=1}^{\infty} \xi^\mu} \big\rangle_\xi \tag{2.21}$$

2.1.3 Computation of the noise distribution $\pi(k)$

Given its symmetry, we only need to calculate $\pi(k|\phi)$ for $k \geq 0$:

$$
\begin{aligned}
\pi(k|\phi) &= \lim_{K\to\infty}\int_{-\pi}^{\pi}\frac{\mathrm{d}\psi}{2\pi}e^{-i\psi k}\left\langle e^{i\psi\xi}\right\rangle_{\xi}^{K} = \lim_{K\to\infty}\int_{-\pi}^{\pi}\frac{\mathrm{d}\psi}{2\pi}e^{-i\psi k}\big(1+\frac{c\phi}{K}(\cos\psi-1)\big)^{K} \\
&= \int_{-\pi}^{\pi}\frac{\mathrm{d}\psi}{2\pi}e^{-i\psi k+\phi c(\cos\psi-1)} \\
&= e^{-\phi c}\int_{-\pi}^{\pi}\frac{\mathrm{d}\psi}{2\pi}e^{-i\psi k}\sum_{n\geq 0}\frac{(\phi c)^n}{2^n n!}(e^{i\psi}+e^{-i\psi})^n \\
&= e^{-\phi c}\int_{-\pi}^{\pi}\frac{\mathrm{d}\psi}{2\pi}e^{-i\psi k}\sum_{n\geq 0}\frac{(\phi c)^n}{2^n n!}\sum_{l\leq n}\frac{n!}{l!(n-l)!}e^{-i\psi(k-n+2l)} \\
&= e^{-\phi c}\sum_{n\geq 0}\sum_{l\leq n}\left(\frac{\phi c}{2}\right)^{n}\frac{1}{l!(n-l)!}\delta_{n,k+2l} \\
&= e^{-\phi c}\sum_{l\geq 0}\left(\frac{\phi c}{2}\right)^{k+2l}\frac{1}{l!(k+l)!} \;=\; e^{-\phi c}\,\mathcal{I}_k(\phi c)
\end{aligned}
\tag{2.22}
$$

where $\mathcal{I}_k(x)$ is the k-th modified Bessel function of the first kind [1]. These modified Bessel functions obey

$$
\begin{aligned}
2\frac{k}{x}\mathcal{I}_k(x) &= \mathcal{I}_{k-1}(x)-\mathcal{I}_{k+1}(x), \\
2\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{I}_k(x) &= \mathcal{I}_{k-1}(x)+\mathcal{I}_{k+1}(x).
\end{aligned}
\tag{2.23}
$$

The first identity leads to a useful recursive equation for $\pi(k|\phi)$, and the second identity simplifies our calculation of derivatives of $\pi(k|\phi)$ with respect to ϕ, respectively:

$$
\pi(k-1|\phi)-\pi(k+1|\phi)-2\pi(k|\phi)\frac{k}{\phi c}=0, \tag{2.24}
$$

$$
\frac{\mathrm{d}}{\mathrm{d}\phi}\pi(k|\phi)=c\Big(\frac{1}{2}\pi(k-1|\phi)+\frac{1}{2}\pi(k+1|\phi)-\pi(k|\phi)\Big) \tag{2.25}
$$

2.1.4 Retrieval in the zero noise limit

To emphasize the dependence of the recall overlap on ϕ, viz. the relative storage load, we will from now on write $m \to m_\phi$. With the abbreviation $\langle g(k)\rangle_k = \sum_k \pi(k|\phi)g(k)$, and using (2.24) and the symmetry of $\pi(k|\phi)$, we can transfer our equation (2.20) into a more convenient form:

$$
\begin{aligned}
m_\phi &= \frac{1}{2}\langle\Big[\tanh(\beta c m_\phi(1+k))+\tanh(\beta c m_\phi(1-k))\Big]\rangle_k \\
&= \frac{1}{2}\sum_{k\in\mathcal{Z}}\Big[\pi(k-1|\phi)-\pi(k+1|\phi)\Big]\tanh(\beta c m_\phi k) = \frac{1}{\phi c}\langle k\tanh(\beta c m_\phi k)\rangle_k
\end{aligned}
\tag{2.26}
$$

In the zero noise limit $\beta \to \infty$, where $\tanh(\beta y) \to \mathrm{sgn}(y)$, this reduces to $m_\phi = \frac{1}{\phi c}\langle|k|\rangle_k$, or, equivalently,

$$\begin{aligned} m_\phi &= \lim_{\beta\to\infty} \langle \tanh(\beta c m(1{+}k))\rangle_k = \langle \mathrm{sign}(1{+}k)\rangle_k \\ &= \sum_{k>-1}\pi(k) - \sum_{k<-1}\pi(k) \;=\; \pi(0|\phi) + \pi(1|\phi), \end{aligned} \tag{2.27}$$

Hence we always have a nonzero rescaled magnetization, for any relative storage load ϕ. To determine for which value of ϕ this state is most stable, we have to insert this solution into the zero temperature formula for the free energy and find the minimum with respect to ϕ. Here, with $m_\mu = m_\phi$ for all $\mu \le K = \phi N^\gamma$ and $m_\mu = 0$ for $\mu > K$, the free energy (2.7) takes asymptotically the form

$$f(\beta) = \frac{1}{2}c^2\phi m_\phi^2 - \frac{1}{\beta}\langle \log\cosh\left(\beta c m_\phi k\right)\rangle_k - \frac{1}{\beta}\log 2 \tag{2.28}$$

So for $\beta \to \infty$, and using our above identity $\langle|k|\rangle_k = \phi c m$, we find that the energy density is

$$u(\phi) \;=\; \lim_{\beta\to\infty} f(\beta) = \frac{1}{2}c^2\phi m_\phi^2 - cm\langle|k|\rangle_k = -\frac{1}{2}c^2\phi m_\phi^2 \tag{2.29}$$

$$\;=\; -\frac{1}{2}c^2\phi\big(\pi(0|\phi){+}\pi(1|\phi)\big)^2 \tag{2.30}$$

To see how this depends on ϕ we may use (2.25), and find

$$\begin{aligned} \frac{1}{c^2}\frac{\mathrm{d}}{\mathrm{d}\phi}u(\phi) &= -\frac{1}{2}m_\phi^2 - \phi m_\phi \frac{\mathrm{d}}{\mathrm{d}\phi}\big(\pi(0|\phi){+}\pi(1|\phi)\big) \\ &= -\frac{1}{2}m_\phi^2 - \phi c m_\phi\Big(-\frac{1}{2}\pi(0|\phi) + \frac{1}{2}\pi(2|\phi)\Big) = -\frac{1}{2}m_\phi^2 + m_\phi\pi(1|\phi) \\ &= -\frac{1}{2}m_\phi\Big(m_\phi - 2\pi(1|\phi)\Big) = -\frac{1}{2}m_\phi\Big(\pi(0|\phi) - \pi(1|\phi)\Big) < 0 \end{aligned} \tag{2.31}$$

The energy density $u(\phi)$ is apparently a decreasing function of ϕ, which reaches its minimum when the number of condensed patterns is maximal, at $\phi = \alpha$. However, the amplitude of each recalled pattern will also decrease for larger values of ϕ:

$$\frac{\mathrm{d}}{\mathrm{d}\phi}m_\phi \;=\; \frac{\mathrm{d}}{\mathrm{d}\phi}\pi(0|\phi) + \frac{\mathrm{d}}{\mathrm{d}\phi}\pi(1|\phi) = -\pi(1|\phi)/\phi < 0 \tag{2.32}$$

Hence m_ϕ starts at $m_0 = 1$, due to $\pi(k|0) = \delta_{k,0}$, and then decays monotonically with ϕ. Moreover, it follows from $\langle|k|\rangle_k^2 \le \langle k^2\rangle_k = \langle\sum_{\mu\le K}(\xi^\mu)^2\rangle_\xi = \phi c$ that

$$m_\phi = \langle|k|\rangle_k/\phi c \;\le\; 1/\sqrt{\phi c}, \qquad u(\phi) = -\frac{1}{2}c^2\phi m_\phi^2 \;\ge\; -\frac{1}{2}c \tag{2.33}$$

If we increase the number of condensed patterns, the corresponding magnetizations decrease in such a way that the energy density remains finite.

2.1.5 Retrieval at nonzero noise levels

To find the critical noise level (if any) where pattern recall sets in, we return to equation (2.8), which for $(\tau, \theta) = (1-\gamma, 0)$ and written in vector notation becomes

$$\mathbf{m} = \frac{N^\gamma}{c} \langle \boldsymbol{\xi} \tanh(\beta c \boldsymbol{\xi} \cdot \mathbf{m}) \rangle_\xi. \tag{2.34}$$

We take the inner product on both sides with $\mathbf{m}$ and obtain a simple inequality:

$$\begin{aligned} \mathbf{m}^2 &= \frac{N^\gamma}{c} \langle (\boldsymbol{\xi} \cdot \mathbf{m}) \tanh(\beta c \boldsymbol{\xi} \cdot \mathbf{m}) \rangle_\xi \\ &= \beta N^\gamma \langle (\boldsymbol{\xi} \cdot \mathbf{m})^2 \int_0^1 \mathrm{d}x \; [1 - \tanh^2(\beta c x \boldsymbol{\xi} \cdot \mathbf{m})] \rangle_\xi \\ &\leq \beta N^\gamma \langle (\boldsymbol{\xi} \cdot \mathbf{m})^2 \rangle_\xi = \beta c \mathbf{m}^2 \end{aligned} \tag{2.35}$$

Since $\mathbf{m}^2(1 - \beta c) \leq 0$, we are sure that $\mathbf{m} = 0$ for $\beta c \leq 1$. At $\beta c = 1$ nontrivial solutions of the previously studied symmetric type are found to bifurcate continuously from the trivial solution. This can be seen by expanding the amplitude equation (2.26) for small m:

$$\begin{aligned} m_\phi &= \frac{1}{\phi c} \langle k \tanh(\beta c m_\phi k) \rangle_k \\ &= \beta c m_\phi - \frac{1}{3} \beta^3 c^2 m_\phi^3 \langle k^4 \rangle_k / \phi + \mathcal{O}(m_\phi^4) \end{aligned} \tag{2.36}$$

This shows that the symmetric solutions indeed bifurcate via a second-order transition, at the ϕ-independent critical temperature $T_c = c$, with amplitude $m_\phi \propto (\beta c - 1)^{\frac{1}{2}}$ as $\beta c \to 1$. All the above predictions are confirmed by the results of numerical simulations, and by solving the order parameter equations and calculating the free energy numerically, see Figure 2.1.

We can now summarize the phase diagram in terms of the scaling exponents (γ, δ). The number of stored patterns is $P = \alpha N^\delta$, of which $K = \phi N^{\delta'}$ can be recalled simultaneously, with $\delta' = \min(\gamma, \delta)$:

$\delta < \gamma$: $\phi_{\max} = \alpha$, all stored patterns recalled simultaneously, with Curie-Weiss overlap m

$\delta = \gamma$: $\phi_{\max} = \alpha$, all stored patterns recalled simultaneously, with reduced but finite m

$\delta > \gamma$: $\phi_{\max} = \infty$, at most ϕN^γ patterns recalled simultaneously, with $\phi \to \infty$ and $m_\phi \to 0$

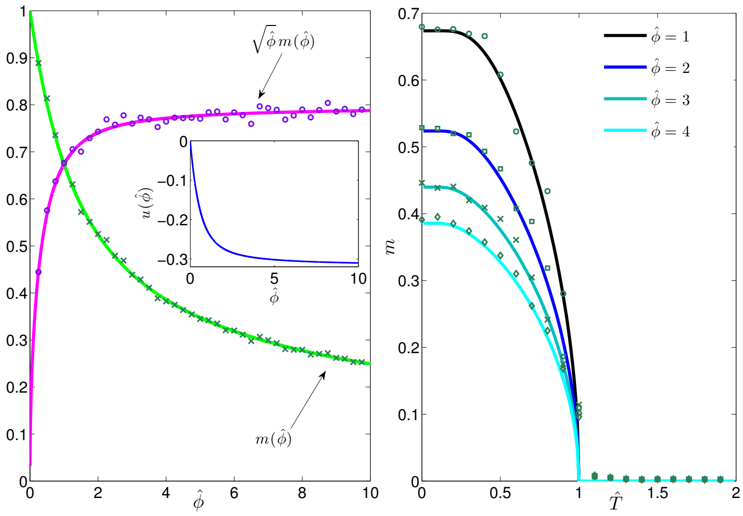

Figure 2.1: Left: energy density u versus the relative fraction of retrieved patterns, in terms of $\hat{\phi} = c\phi$ and $\hat{T} = T/c = 1/\beta c$. The minimum energy density is reached when $\hat{\phi}$ is maximal, i.e. when *all* stored patterns are simultaneously retrieved, but with decreasing amplitude for each. Right: critical noise levels for different values of $\hat{\phi}$, confirming that $\hat{T}_c^{-1} = \hat{\beta}_c = 1$, independently of $\hat{\phi}$. In both the panels, solid lines represent our theoretical predictions, while symbols represent data from numerical simulations on systems with $N = 5\times 10^4$, $\gamma = \delta = 0.45$, $c = 2$ and with with standard sequential Glauber dynamics.

2.2 High storage regime in extremely diluted connectivity: absence of retrieval

In this section we consider the same network, composed of N binary spins (σ_i, $i = 1, \dots, N$) and P patterns, but now at high storage load:

$$\lim_{N\to\infty} P/N = \alpha, \qquad \alpha > 0 \tag{2.37}$$

The effective interaction is described by the Hamiltonian (2.2), and the entries $\xi_i^\mu \in \{0, \pm 1\}$ are generated again from (2.3), i.e. in the extremely diluted regime $\gamma < 1$. Again we must choose τ such that the Hamiltonian will be of order N. Heuristically, since the number of non-zero entries $\mathcal{N}_{\rm nz}$ in a typical pattern $(\xi_1^\mu, \dots, \xi_N^\mu)$ scales as $\mathcal{O}(N^{1-\gamma})$, the number of patterns with non overlapping entries (i.e. those we expect to recall) will scale as $\mathcal{O}(N/\mathcal{N}_{nz}) = \mathcal{O}(N^\gamma)$. The contribution from $K = \mathcal{O}(N^\gamma)$ such condensed patterns to the Hamiltonian would then scale as

$$\mathcal{H}_C \sim N^{-\tau}\sum_{\mu=1}^{K}(\sum_{i=1}^{N}\xi_i^\mu\sigma_i)^2 \sim N^{-\tau}K\mathcal{N}_{\rm nz}^2 \sim N^{-\tau}N^{\gamma}N^{2(1-\gamma)} \sim N^{2-\gamma-\tau}$$

The non-condensed patterns, of which there are $N_{\rm nc} = P - N_c \sim P = \mathcal{O}(N)$, are expected to contribute

$$\mathcal{H}_{NC} \sim N^{-\tau}\sum_{\mu=1}^{N_{\rm nc}}(\sum_{i=1}^{N}\xi_i^\mu\sigma_i)^2 \sim N^{-\tau}N_{\rm nc}\sqrt{\mathcal{N}_{nz}}^2 \sim N^{-\tau}NN^{1-\gamma} \sim N^{2-\gamma-\tau}.$$

Thus, we expect to have an extensive Hamiltonian for $\tau = 1-\gamma$.

2.2.1 Replica-symmetric theory

In the scaling regime $P = \alpha N$ we can no longer use saddle-point arguments directly in the calculation of the free energy. Instead we calculate the free energy for typical patterns realizations, i.e. the average

$$\overline{f} = -\lim_{N\to\infty}\frac{1}{\beta N}\overline{\log Z_N(\beta,\xi)}, \tag{2.38}$$

Here $\overline{\cdots}$ indicates averaging over all $\{\xi_i^\mu\}$, according to the measure (2.3). The average over the disorder is done with the replica method, for $K = \mathcal{O}(N^\gamma)$; full details are given in Appendix B. We solve the model at the replica symmetric (RS) level, which implies the assumption that the system has at most a finite number of ergodic sectors for $N \to \infty$, giving

$$\beta\overline{f}_{\rm RS} = \lim_{N\to\infty}\text{extr}_{\mathbf{m},q,r}\ \beta\hat{f}_{\rm RS}(\mathbf{m},q,r), \tag{2.39}$$

$$\begin{aligned}\beta\hat{f}_{\rm RS}(\mathbf{m},q,r) = & -\log 2 + \frac{1}{2}\alpha r(\beta c)^2(1-q) + \frac{\beta c^2}{2N^\gamma}\mathbf{m}^2 - \frac{\alpha}{2}\Big(\frac{\beta cq}{1-\beta c(1-q)} - \log[1-\beta c(1-q)]\Big) \\ & -\Big\langle\int \mathrm{D}z\ \log\cosh[\beta c(\mathbf{m}\cdot\boldsymbol{\xi} + z\sqrt{\alpha r})]\Big\rangle_\xi\end{aligned} \tag{2.40}$$

in which $\mathbf{m} = (m_1,\ldots,m_K)$ denotes the vector of $K = \phi N^\gamma$ condensed (i.e. potentially recalled) patterns, $\boldsymbol{\xi} = (\xi^1,\ldots,\xi^K)$, and $\mathrm{D}z = (2\pi)^{-1/2}\mathrm{e}^{-z^2/2}\mathrm{d}z$. As in the analysis of standard Hopfield networks, this involves the Edward-Anderson spin-glass order parameter q [45, 15] and the Amit-Gutfreund-Sompolinsky uncondensed-noise order parameter r [45, 15]. We obtain self-consistent equations for the remaining RS order parameters (m,q,r) simply by extremizing $\hat{f}_{\mathrm{RS}}(\mathbf{m}, q, r)$, which leads to

$$\begin{aligned}
m^\mu &= \frac{N^\gamma}{c}\Big\langle \xi^\mu \int \mathrm{D}z\ \tanh[\beta c(\mathbf{m}\cdot\boldsymbol{\xi}+z\sqrt{\alpha r})]\Big\rangle_\xi, \\
q &= \Big\langle \int \mathrm{D}z\ \tanh^2[\beta c(\mathbf{m}\cdot\boldsymbol{\xi}+z\sqrt{\alpha r})]\Big\rangle_\xi, \\
r &= \frac{q}{[1-\beta c(1-q)]^2}. \qquad (2.41)
\end{aligned}$$

As before we deal with the equation for m^μ by using the identity $\xi^\mu \tanh(A) = \tanh(\xi^\mu A)$ (since $\xi^\mu \in \{-1, 0, 1\}$) and by separating the term $m^\mu \xi^\mu$ from the sum $\mathbf{m}\cdot\boldsymbol{\xi}$:

$$\begin{aligned}
m^\mu &= \frac{N^\gamma}{c}\Big\langle \int \mathrm{D}z\ \tanh[\beta c(m^\mu(\xi^\mu)^2 + \sum_{\nu\neq\mu\leq K} m^\nu\xi^\nu\xi^\mu + z\xi^\mu\sqrt{\alpha r})]\Big\rangle_\xi \\
&= \Big\langle \int \mathrm{D}z\ \tanh[\beta c(m^\mu + \sum_{\nu\neq\mu\leq K} m^\nu\xi^\nu + z\sqrt{\alpha r})]\Big\rangle_\xi \\
&= \Big\langle \int \mathrm{D}z\ \tanh\Big[\beta c\Big(m^\mu + \sum_{\nu=1}^{K} m^\nu\xi^\nu + z\sqrt{\alpha r}\Big)\Big]\Big\rangle_\xi + \mathcal{O}(N^{-\gamma})
\end{aligned}$$

Again we see that for $N \to \infty$ we will only retain solutions with $m^\mu \in \{-m, 0, m\}$ for all $\mu \leq K$. Given the trivial sign and pattern label permutation invariances, we can without loss of generality consider only non-negative magnetizations, and look for solutions where $m^\mu = m$ for $\mu = 1 \leq K$ and zero otherwise. We then find

$$m = \sum_{k=-\infty}^{\infty} \pi(k) \int \mathrm{D}z\ \tanh[\beta c(m + mk + z\sqrt{\alpha r})] \qquad (2.42)$$

with $\pi(k)$ given in (2.22). We can now use the manipulations employed in the previous section, to find

$$\begin{aligned}
m &= \Big\langle \frac{k}{\phi} \int \mathrm{D}z\ \tanh[\beta c(mk + z\sqrt{\alpha r})]\Big\rangle_k \qquad (2.43) \\
q &= \Big\langle \int \mathrm{D}z\ \tanh^2[\beta c(mk + z\sqrt{\alpha r})]\Big\rangle_k, \qquad (2.44) \\
r &= \frac{q}{[1-\beta c(1-q)]^2}.
\end{aligned}$$

The corresponding free energy assumes the form

$$\begin{aligned}
\beta \hat{f}_{\mathrm{RS}}(m, q, r) &= -\log 2 + \frac{1}{2}\alpha r(\beta c)^2(1-q) + \frac{1}{2}\beta c^2\phi m^2 - \frac{\alpha}{2}\Big(\frac{\beta c q}{1-\beta c(1-q)} - \log[1-\beta c(1-q)]\Big) \\
&\quad - \Big\langle \int \mathrm{D}z\ \log\cosh[\beta c(mk + z\sqrt{\alpha r})]\Big\rangle_k \qquad (2.45)
\end{aligned}$$

Note that we recover the equations of the medium storage regime simply by putting $\alpha = 0$.

2.2.2 The zero noise limit

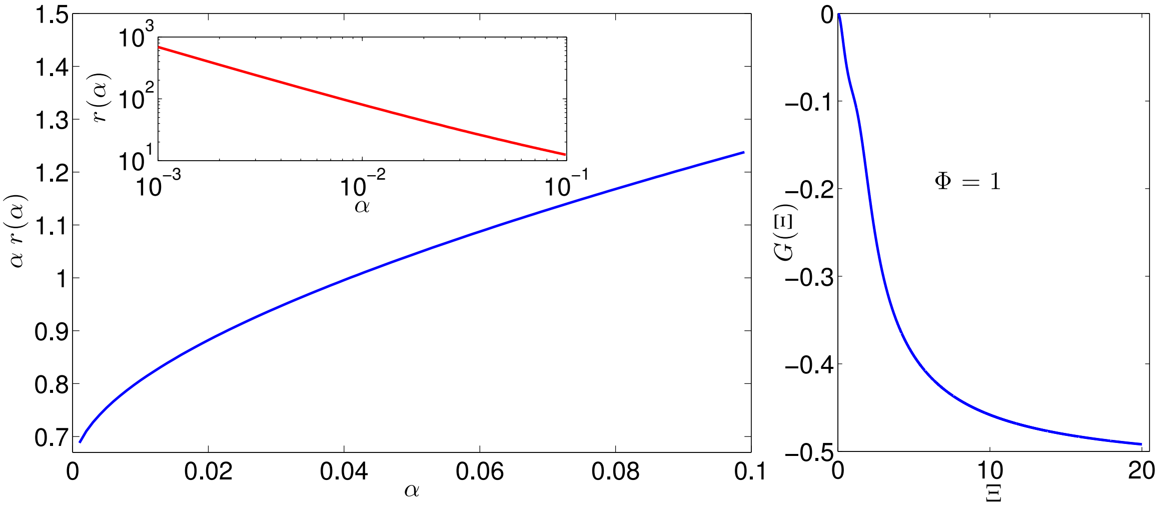

Figure 2.2: Left panel: Behavior of $\alpha r(\alpha)$ versus α in the spin-glass state (the inset shows only $r(\alpha)$ versus α), as calculated from the RS order parameter equations. This shows that $r(\alpha)$ goes to infinity as α approaches zero, such that $\alpha r(\alpha)$ remains positive; this means that the noise due to non-condensed patterns can never be neglected. Right panel: behavior of the function $G(\Xi)$ versus Ξ. Since $G(\Xi) < 0$ for $\alpha > 0$, equation (2.50) cannot have a solution for $\alpha > 0$, and hence no pattern recall is possible even at zero noise.

We now show that in the high storage case the system behaves as a spin glass, even in the zero temperature limit $\beta \to \infty$ where the retrieval capability should be largest. From (2.44) we deduce that $q \to 1$ in the zero noise limit, while the quantity $C = \beta c(1-q)$ remains finite. Let us first send $\beta \to \infty$ in equation (2.43):

$$m = \Big\langle \frac{k}{\phi} \int \mathrm{D}z \; \mathrm{sgn}\Big[mk + \frac{z\sqrt{\alpha}}{1-C}\Big]\Big\rangle_k = \Big\langle \frac{k}{\phi} \mathrm{Erf}\Big(\frac{mk(1-C)}{\sqrt{2\alpha}}\Big)\Big\rangle_k, \tag{2.46}$$

with the error integral $\mathrm{Erf}(x) = (2/\sqrt{\pi}) \int_0^x \mathrm{d}t \; \mathrm{e}^{-t^2}$. A second equation for the pair (m, C) follows from (2.44):

$$\begin{aligned} C &= \lim_{\beta\to\infty} \beta c \Big\langle 1 - \int \mathrm{D}z \; \tanh^2[\beta c(mk + z\sqrt{\alpha r})]\Big\rangle_k \\ &= \lim_{\beta\to\infty} \frac{\partial}{\partial m} \Big\langle \frac{1}{k} \int \mathrm{D}z \; \tanh\Big[\beta c\Big(mk + \frac{z\sqrt{\alpha q}}{1-C}\Big)\Big]\Big\rangle_k, \\ &= \frac{\partial}{\partial m} \Big\langle \frac{1}{k} \mathrm{Erf}\Big(\frac{mk(1-C)}{\sqrt{2\alpha}}\Big)\Big\rangle_k \\ &= \sqrt{\frac{2}{\alpha\pi}}(1-C) \Big\langle \exp\Big(-\frac{m^2k^2(1-C)^2}{2\alpha}\Big)\Big\rangle_k \end{aligned} \tag{2.47}$$

We thus have two coupled nonlinear equations (2.46,2.47), for the two zero temperature order parameters m and C. They can be further reduced by introducing the variable $\Xi = m(1-C)/\sqrt{2\alpha}$, with which we obtain

$$m = \Big\langle \frac{k}{\phi} \mathrm{Erf}(k\Xi)\Big\rangle_k \tag{2.48}$$

and rewriting $\Xi = m(1-C)/\sqrt{2\alpha}$ gives

$$C = 1 - \frac{\sqrt{2\alpha}\Xi}{m} = 1 - \sqrt{2\alpha}\Xi \Big\langle \frac{k}{\phi} \mathrm{Erf}(k\Xi)\Big\rangle_k^{-1}. \tag{2.49}$$

Using (2.47) and excluding the trivial solution $\Xi = 0$ (which always exists, but represents the spin glass state without pattern recall) we obtain after some simple algebra just a single equation, to be solved for Ξ:

$$\sqrt{2\alpha} = G(\Xi) = \frac{1}{\Xi}\Big\langle \frac{k}{\phi} \mathrm{Erf}(k\Xi)\Big\rangle_k - \frac{2}{\sqrt{\pi}}\Big\langle \mathrm{e}^{-k^2\Xi^2}\Big\rangle_k \tag{2.50}$$

One easily shows that

$$\lim_{\Xi\to 0} G(\Xi) = 0, \qquad \lim_{\Xi\to\infty} G(\Xi) = -\frac{2}{\sqrt{\pi}}\pi(0|\phi). \tag{2.51}$$

In fact further analytical and numerical investigation reveals that for $\Xi > 0$ the function $G(\Xi)$ is strictly negative; see Figure 2.2. Hence there can be no $m \neq 0$ solution for $\alpha > 0$, so the system cannot recall the patterns in the present scaling regime $P = \alpha N$.

2.3 High storage regime in a finite connectivity: replica approach

In this section, mainly following [5], we turn to a statistical mechanics analysis of an Hopfield like network, again near the saturation regime ($P = \alpha N$), but in a finite connectivity regime, i.e. the pattern entries $\xi_i^\mu \in \{-1, 0, 1\}$ are quenched random variables, identically and independently distributed according to

$$P(\xi_i^\mu = 1) = P(\xi_i^\mu = -1) = \frac{c}{2N}, \qquad P(\xi_i^\mu = 0) = 1 - \frac{c}{N}, \tag{2.52}$$

with c finite. The effective interactions among spins is described by the usual Hamiltonian

$$\mathcal{H}(\boldsymbol{\sigma}|\xi) = -\frac{1}{2c}\sum_{i,j}^{N}\sum_{\mu}^{P}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j. \tag{2.53}$$

It is not a priori obvious that solving this model analytically will be possible. Most methods for spins systems on finitely connected heterogeneous graphs rely (explicitly or implicitly) on these being locally tree-like; due to the pattern dilution, the underlying topology of the system (2.53) is a heterogeneous graph with many short loops.

The Hamiltonian is normalized correctly: since the term $\sum_{i=1}^N \xi_i^\mu \sigma_i$ is $\mathcal{O}(1)$ both for condensed and non condensed patterns [3], (2.53) is indeed extensive in N. The aim of this section is to compute the disorder-averaged free energy $\overline{f}$, at inverse temperature $\beta = T^{-1}$, where $\overline{\cdots}$ denotes averaging over the αN^2 variables $\{\xi_i^\mu\}$ and

$$f = -\lim_{N\to\infty}\frac{1}{\beta N}\log Z_N(\beta,\xi), \tag{2.54}$$

where $Z_N(\beta, \xi)$ is the partition function

$$Z_N(\beta,\xi) = \sum_{\boldsymbol{\sigma}\in\{-1,1\}^N} \mathrm{e}^{\frac{\beta}{2c}\sum_{\mu=1}^{\alpha N}\left(\sum_{i=1}^N \xi_i^\mu\sigma_i\right)^2}. \tag{2.55}$$

The state of the system can be characterized in terms of the αN (non-normalised) Mattis magnetizations, i.e. the overlaps between the system configuration and each pattern

$$M_\mu(\boldsymbol{\sigma}) = \sum_{i=1}^{N}\xi_i^\mu\sigma_i. \tag{2.56}$$

However, since in the high load regime the number of overlaps is extensive, it is more convenient to work with the overlap distribution

$$P(M|\boldsymbol{\sigma}) = \frac{1}{\alpha N}\sum_{\mu=1}^{\alpha N}\delta_{M_\mu(\boldsymbol{\sigma}),M}. \tag{2.57}$$

Although $M_\mu(\boldsymbol{\sigma})$ can take (discrete) values in the whole range $\{-N, -N+1, \cdots, N\}$, we expect that, due to dilution, the number of values that the $M_\mu(\boldsymbol{\sigma})$ assume remains effectively finite for large N, so that (2.57) represents an effective finite number of order parameters. In order to probe responses of the system to selected perturbations we introduce external fields $\{\psi_\mu\}$ coupled to the overlaps $\{M_\mu(\boldsymbol{\sigma})\}$, so we consider the extended Hamiltonian

$$\mathcal{H}(\sigma,\xi) = -\frac{1}{2c}\sum_{i,j}^{N}\sum_{\mu}^{\alpha N}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j - \sum_{\mu=1}^{\alpha N}\psi_\mu M_\mu(\boldsymbol{\sigma}). \tag{2.58}$$

We also define the field distribution $P(\psi)$ and the joint distribution $P(M,\psi|\boldsymbol{\sigma})$ of magnetizations and fields (and of which $P(\psi)$ is a marginal):

$$P(\psi) = \frac{1}{\alpha N}\sum_{\mu=1}^{\alpha N}\delta(\psi-\psi_\mu), \qquad P(M,\psi|\boldsymbol{\sigma}) = \frac{1}{\alpha N}\sum_{\mu=1}^{\alpha N}\delta_{M,M_\mu(\boldsymbol{\sigma})}\delta(\psi-\psi_\mu). \tag{2.59}$$

2.3.1 The free energy

The free energy per spin (2.54) for the Hamiltonian (2.58) can be writen as

$$f \;=\; -\lim_{N\to\infty}\frac{1}{\beta N}\log\sum_{\boldsymbol{\sigma}} \mathrm{e}^{\frac{\beta}{2c}\sum_{\mu=1}^{\alpha N}M_\mu^2(\boldsymbol{\sigma})+\beta\sum_{\mu=1}^{\alpha N}\psi_\mu M_\mu(\boldsymbol{\sigma})}. \tag{2.60}$$

We insert the following integrals of delta-functions written in Fourier representation

$$\begin{aligned}1 \;&=\; \prod_M\prod_\psi\int \mathrm{d}P(M,\psi)\;\delta\Big[P(M,\boldsymbol{\psi}) - \frac{1}{\alpha N}\sum_{\mu=1}^{\alpha N}\delta_{M,M_\mu(\boldsymbol{\sigma})}\delta(\psi-\psi_\mu)\Big] \\ &=\; \prod_M\prod_\psi\int\frac{\mathrm{d}P(M,\psi)\mathrm{d}\hat{P}(M,\psi)}{2\pi/\Delta N}\;\mathrm{e}^{\mathrm{i}N\Delta\hat{P}(M,\psi)[P(M,\psi)-\frac{1}{\alpha N}\sum_{\mu=1}^{\alpha N}\delta_{M,M_\mu(\boldsymbol{\sigma})}\delta(\psi-\psi_\mu)]}.\end{aligned} \tag{2.61}$$

In the limit $\Delta\to 0$ we use $\Delta\sum_\psi\ldots\to\int\mathrm{d}\psi\ldots$, and we define the path integral measure $\{\mathrm{d}P\mathrm{d}\hat{P}\} = \lim_{\Delta\to 0}\mathrm{d}P(M,\psi)\mathrm{d}\hat{P}(M,\psi)\Delta N/2\pi$. This gives us

$$1 \;=\; \int\{\mathrm{d}P\mathrm{d}\hat{P}\}\;\mathrm{e}^{\mathrm{i}N\int\mathrm{d}\psi\sum_M\hat{P}(M,\psi)P(M,\psi)-\frac{i}{\alpha}\sum_{\mu=1}^{\alpha N}\hat{P}(M_\mu(\boldsymbol{\sigma}),\psi_\mu)}. \tag{2.62}$$

Insertion into (2.60) leads us to an expression for f involving the density of states $\Omega[\hat{P}]$:

$$f \;=\; -\lim_{N\to\infty}\frac{1}{\beta N}\log\int\{\mathrm{d}P\mathrm{d}\hat{P}\}\,\mathrm{e}^{N\left\{\mathrm{i}\int\mathrm{d}\psi\sum_M P(M,\psi)\hat{P}(M,\psi)+\beta\alpha\int\mathrm{d}\psi\sum_M P(M,\psi)\left(\frac{M^2}{2c}+M\psi\right)+\Omega[\hat{P}]\right\}} \tag{2.63}$$

$$\Omega[\hat{P}] \;=\; \lim_{N\to\infty}\frac{1}{N}\log\sum_{\boldsymbol{\sigma}}\mathrm{e}^{-\frac{i}{\alpha}\sum_\mu\hat{P}(M_\mu(\boldsymbol{\sigma}),\psi_\mu)}. \tag{2.64}$$

Hence via steepest descent integration for $N \to \infty$, and after avering the result over the disorder, we obtain:

$$\overline{f} = -\frac{1}{\beta}\mathrm{extr}_{\{P,\hat{P}\}}\Big\{\mathrm{i}\int\mathrm{d}\psi\,\sum_{M} P(M,\psi)\hat{P}(M,\psi) + \beta\alpha\int\mathrm{d}\psi\,\sum_{M} P(M,\psi)(\frac{M^2}{2c} + M\psi) + \overline{\Omega}[\hat{P}]\Big\}, \tag{2.65}$$

with

$$\overline{\Omega}[\hat{P}] \;=\; \lim_{N\to\infty}\frac{1}{N}\,\overline{\log\sum_{\boldsymbol{\sigma}}\mathrm{e}^{-\frac{\mathrm{i}}{\alpha}\sum_{\mu}\hat{P}(M_\mu(\boldsymbol{\sigma}),\psi_\mu)}}. \tag{2.66}$$

Working out the functional saddle-point equations that define the extremum in (2.65) gives

$$\hat{P}(M,\psi) = \mathrm{i}\alpha\beta\Big(\frac{M^2}{2c} + M\psi\Big), \qquad P(M,\psi) = \mathrm{i}\frac{\delta\overline{\Omega}[\hat{P}]}{\delta\hat{P}(M,\psi)}, \tag{2.67}$$

and inserting the first of these equations into (2.65) leads us to

$$\overline{f} = -\frac{1}{\beta}\overline{\Omega}[\hat{P}]\Big|_{\hat{P}(M,\psi)=\mathrm{i}\alpha\beta(\frac{M^2}{2c}+M\psi)}. \tag{2.68}$$

Hence calculating the disorder-averaged free-energy boils down to calculating (2.66). This can be done using the replica method, which is based on the identity $\overline{\log Z} = \lim_{n\to 0} n^{-1}\log\overline{Z^n}$, yielding

$$\overline{\Omega}[\hat{P}] \;=\; \lim_{N\to\infty}\lim_{n\to 0}\frac{1}{Nn}\log\sum_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^n}\overline{\mathrm{e}^{-\frac{\mathrm{i}}{\alpha}\sum_{\alpha=1}^{n}\sum_{\mu=1}^{\alpha N}\hat{P}(M_\mu(\boldsymbol{\sigma}^\alpha),\psi_\mu)}}. \tag{2.69}$$

The free energy (2.68) could also have been calculated directly from (2.60), by taking the average over disorder and using the replica identity. The advantage of working with the log-density of states is that, working out $\overline{\Omega}[\hat{P}]$ first for arbitrary functions $\hat{P}$ gives us via (2.67) a formula for the distribution $P(M,\psi)$, from which we can obtain useful information on the system retrieval phases and response to external perturbations. Finally we set $\hat{P}(M,\psi) = \mathrm{i}\alpha\beta\chi(M,\psi)$ with a real-valued function χ, to compactify our equations, with which we can write our problem as follows

$$\overline{f} = f[\chi]\Big|_{\chi(M,\psi)=\frac{M^2}{2c}+M\psi} \qquad f[\chi] = -\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{\beta Nn}\log\sum_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^n}\overline{\mathrm{e}^{\beta\sum_{\alpha=1}^{n}\sum_{\mu=1}^{\alpha N}\chi(M_\mu(\boldsymbol{\sigma}^\alpha),\psi_\mu)}} \tag{2.70}$$

$$P(M,\psi) = -\frac{1}{\alpha}\frac{\delta f[\chi]}{\delta\chi}\Big|_{\chi(M,\psi)=\frac{M^2}{2c}+M\psi}. \tag{2.71}$$

For simple tests of (2.70) and (2.71) in special limits see C.1.

2.3.2 Derivation of saddle-point equations

From now on, unless indicated otherwise, all summations and products over α, i, and μ will be understood to imply $\alpha = 1\ldots n$, $i = 1\ldots N$, and $\mu = 1\ldots \alpha N$, respectively. We next need to introduce order parameters that allow us to carry out the disorder average in (2.70). The simplest choice is to isolate the overlaps themselves by inserting

$$1 \;=\; \prod_{\alpha\mu}\Big[\sum_{M_{\alpha\mu}=-N}^{N}\delta_{M_{\alpha\mu},\sum_i \xi_i^\mu \sigma_i^\alpha}\Big] = \prod_{\alpha\mu}\Big[\sum_{M_{\alpha\mu}=-N}^{N}\int_{-\pi}^{\pi}\frac{\mathrm{d}\omega_{\alpha\mu}}{2\pi}\mathrm{e}^{\mathrm{i}\omega_{\alpha\mu}(M_{\alpha\mu}-\sum_i \xi_i^\mu\sigma_i^\alpha)}\Big]. \tag{2.72}$$

This gives

$$f[\chi] \;=\; -\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{\beta N n}\log\Big\{\prod_{\alpha\mu}\Big[\sum_{M_{\alpha\mu}=-\infty}^{\infty}\int_{-\pi}^{\pi}\frac{\mathrm{d}\omega_{\alpha\mu}}{2\pi}\Big]\mathrm{e}^{\mathrm{i}\sum_{\alpha\mu}\omega_{\alpha\mu}M_{\alpha\mu}+\sum_{\alpha\mu}\beta\chi(M_\mu^\alpha,\psi_\mu)}$$
$$\times\sum_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^n}\overline{\mathrm{e}^{-\mathrm{i}\sum_i\sum_{\alpha\mu}\omega_{\alpha\mu}\xi_i^\mu\sigma_i^\alpha}}\Big\}. \tag{2.73}$$

We can carry out the disorder average

$$\overline{\mathrm{e}^{-\mathrm{i}\sum_i\sum_{\alpha\mu}\omega_{\alpha\mu}\xi_i^\mu\sigma_i^\alpha}} \;=\; \prod_{i\mu}\Big\{1-\frac{c}{N}+\frac{c}{2N}\Big(\mathrm{e}^{\mathrm{i}\sum_\alpha\omega_{\alpha\mu}\sigma_i^\alpha}+\mathrm{e}^{-\mathrm{i}\sum_\alpha\omega_{\alpha\mu}\sigma_i^\alpha}\Big)\Big\}$$
$$=\; \mathrm{e}^{\frac{c}{N}\sum_{i\mu}\left[\cos(\sum_\alpha\omega_{\alpha\mu}\sigma_i^\alpha)-1\right]+\mathcal{O}(N^0)}, \tag{2.74}$$

which leads us to

$$f[\chi] \;=\; -\lim_{n\to 0}\lim_{N\to\infty}\frac{1}{\beta N n}\log\Big\{\prod_{\alpha\mu}\Big[\sum_{M_{\alpha\mu}}\int_{-\pi}^{\pi}\frac{\mathrm{d}\omega_{\alpha\mu}}{2\pi}\Big].\mathrm{e}^{\mathrm{i}\sum_{\alpha\mu}\omega_{\alpha\mu}M_{\alpha\mu}+\sum_{\alpha\mu}\beta\chi(M_\mu^\alpha,\psi_\mu)}$$
$$\times\Big[\sum_{\sigma_1\ldots\sigma_n}\mathrm{e}^{\frac{c}{N}\sum_\mu\left[\cos(\sum_\alpha\omega_{\alpha\mu}\sigma_\alpha)-1\right]}\Big]^N\Big\} \tag{2.75}$$

where we have also interchanged the limits $n\to 0$ and $N\to\infty$, as usually done to progress in the calculation by using the saddle-point method. We next introduce n-dimensional vectors: $\boldsymbol{\sigma} = (\sigma_1,\ldots,\sigma_n) \in \{-1,1\}^n$, $\mathbf{M}^\mu = (M_{1\mu},\ldots,M_{n\mu}) \in \mathbb{Z}^n$ and $\boldsymbol{\omega}^\mu = (\omega_{1\mu},\ldots,\omega_{n\mu}) \in [-\pi,\pi]^n$. This allows us to write (2.75) as

$$f[\chi] \;=\; -\lim_{n\to 0}\lim_{N\to\infty}\frac{1}{\beta N n}\log\Big\{\prod_\mu\Big[\sum_{\mathbf{M}^\mu}\int_{-\pi}^{\pi}\frac{\mathrm{d}\boldsymbol{\omega}^\mu}{(2\pi)^n}\Big]\cdot\mathrm{e}^{\mathrm{i}\sum_\mu\boldsymbol{\omega}^\mu\cdot\mathbf{M}^\mu+\sum_\mu\beta\chi(\mathbf{M}^\mu,\psi_\mu)}$$
$$\times\Big[\sum_{\boldsymbol{\sigma}}\mathrm{e}^{\frac{c}{N}\sum_\mu[\cos(\boldsymbol{\omega}^\mu\cdot\boldsymbol{\sigma})-1]}\Big]^N\Big\}. \tag{2.76}$$

This last expression invites us to introduce the distribution $P(\boldsymbol{\omega})=(\alpha N)^{-1}\sum_\mu \delta(\boldsymbol{\omega}-\boldsymbol{\omega}^\mu)$, for $\boldsymbol{\omega}\in[-\pi,\pi]^n$, via path integrals. We therefore insert

$$\begin{aligned}1 &= \prod_{\boldsymbol{\omega}}\int \mathrm{d}P(\boldsymbol{\omega})\;\delta\Big[P(\boldsymbol{\omega})-\frac{1}{\alpha N}\sum_\mu \delta(\boldsymbol{\omega}-\boldsymbol{\omega}^\mu)\Big] \\ &= \prod_{\boldsymbol{\omega}}\int \frac{\mathrm{d}P(\boldsymbol{\omega})\mathrm{d}\hat{P}(\boldsymbol{\omega})}{2\pi/\Delta N}\mathrm{e}^{\mathrm{i}N\Delta\hat{P}(\boldsymbol{\omega})\Big[P(\boldsymbol{\omega})-\frac{1}{\alpha N}\sum_\mu\delta(\boldsymbol{\omega}-\boldsymbol{\omega}^\mu)\Big]}. \qquad (2.77)\end{aligned}$$

In the limit $\Delta\to 0$ we use $\Delta\sum_{\boldsymbol{\omega}}\ldots\to\int\mathrm{d}\boldsymbol{\omega}\ldots$, and we define the usual path integral measure $\{\mathrm{d}P\mathrm{d}\hat{P}\}=\lim_{\Delta\to 0}\mathrm{d}P(\boldsymbol{\omega})\mathrm{d}\hat{P}(\boldsymbol{\omega})\Delta N/2\pi$. This converts the above to

$$1 = \int\{\mathrm{d}P\mathrm{d}\hat{P}\}\;\mathrm{e}^{\mathrm{i}N\int\mathrm{d}\boldsymbol{\omega}\;\hat{P}(\boldsymbol{\omega})P(\boldsymbol{\omega})-(i/\alpha)\sum_\mu\hat{P}(\boldsymbol{\omega}^\mu)}. \qquad (2.78)$$

and upon insertion into (2.76) we get

$$\begin{aligned}f[\chi] &= -\lim_{n\to 0}\lim_{N\to\infty}\frac{1}{\beta N n}\log\int\{\mathrm{d}P\mathrm{d}\hat{P}\}\;\mathrm{e}^{\mathrm{i}N\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\;\hat{P}(\boldsymbol{\omega})P(\boldsymbol{\omega})}\Big[\sum_{\boldsymbol{\sigma}}\mathrm{e}^{\alpha c\int\mathrm{d}\boldsymbol{\omega}\;P(\boldsymbol{\omega})[\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})-1]}\Big]^N \\ &\quad\times\prod_{\mu=1}^{\alpha N}\Big(\sum_{\mathbf{M}}\int_{-\pi}^{\pi}\frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}+\sum_\alpha\beta\chi(M^\alpha,\psi_\mu)-\frac{\mathrm{i}}{\alpha}\hat{P}(\boldsymbol{\omega})}\Big). \qquad (2.79)\end{aligned}$$

In the limit $N\to\infty$, evaluation of the integrals by steepest descent leads to

$$f[\chi] = -\lim_{n\to 0}\frac{1}{\beta n}\mathrm{extr}_{\{P,\hat{P}\}}\;\Psi_n[\{P,\hat{P}\}], \qquad (2.80)$$

$$\begin{aligned}\Psi_n[\{P,\hat{P}\}] &= \mathrm{i}\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\;\hat{P}(\boldsymbol{\omega})P(\boldsymbol{\omega})+\alpha\Big\langle\log\Big(\sum_{\mathbf{M}}\int_{-\pi}^{\pi}\frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}+\sum_\alpha\beta\chi(M^\alpha,\psi)-\frac{i}{\alpha}\hat{P}(\boldsymbol{\omega})}\Big)\Big\rangle_\psi \\ &\quad+\log\Big(\sum_{\boldsymbol{\sigma}}\mathrm{e}^{\alpha c\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\;P(\boldsymbol{\omega})[\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})-1]}\Big), \qquad (2.81)\end{aligned}$$

in which $\langle\ldots\rangle_\psi=\int\mathrm{d}\psi\,P(\psi)\ldots$. We mostly write $\langle\ldots\rangle$ in what follows, when there is no risk of ambiguities. The saddle-point equations are found by functional variation of Ψ_n with respect to P and $\hat{P}$, leading to

$$\hat{P}(\boldsymbol{\omega}) = \mathrm{i}c\alpha\;\frac{\sum_{\boldsymbol{\sigma}}\big[\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})-1\big]\mathrm{e}^{\alpha c\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}'\;P(\boldsymbol{\omega}')[\cos(\boldsymbol{\omega}'\cdot\boldsymbol{\sigma})-1]}}{\sum_{\boldsymbol{\sigma}}\mathrm{e}^{\alpha c\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}'\;P(\boldsymbol{\omega}')[\cos(\boldsymbol{\omega}'\cdot\boldsymbol{\sigma})-1]}}, \qquad (2.82)$$

$$P(\boldsymbol{\omega}) = \Bigg\langle\frac{\sum_{\mathbf{M}}\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}+\sum_\alpha\beta\chi(M^\alpha,\psi)-\frac{\mathrm{i}}{\alpha}\hat{P}(\boldsymbol{\omega})}}{\sum_{\mathbf{M}}\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}'\;\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}'\cdot\mathbf{M}+\sum_\alpha\beta\chi(M^\alpha,\psi)-\frac{\mathrm{i}}{\alpha}\hat{P}(\boldsymbol{\omega}')}}\Bigg\rangle. \qquad (2.83)$$

The joint distribution of fields and magnetizations now follows directly from (2.71) and (2.80, 2.81), and is seen to require only knowledge of the conjugate order parameters $\hat{P}(\boldsymbol{\omega})$:

$$\frac{P(M,\psi)}{P(\psi)} = \lim_{n\to 0}\left.\frac{\sum_{\mathbf{M}}\Big(\frac{1}{n}\sum_\gamma\delta_{M,M_\gamma}\Big)\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\;\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}+\beta\sum_\alpha\chi(M^\alpha,\psi)-\frac{\mathrm{i}}{\alpha}\hat{P}(\boldsymbol{\omega})}}{\sum_{\mathbf{M}}\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\;\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}+\beta\sum_\alpha\chi(M^\alpha,\psi)-\frac{\mathrm{i}}{\alpha}\hat{P}(\boldsymbol{\omega})}}\right|_{\chi=\frac{M^2}{2c}+\psi M} \qquad (2.84)$$

Thus the right-hand side is an expression for $P(M|\psi)$. A last simple transformation $F(\boldsymbol{\omega}) = -\frac{\mathrm{i}}{c\alpha}\hat{P}(\boldsymbol{\omega}) + 1$ converts the saddle point equations into

$$F(\boldsymbol{\omega}) = \frac{\sum_{\boldsymbol{\sigma}} \cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma}) \mathrm{e}^{\alpha c \int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega}' \; P(\boldsymbol{\omega}') \cos(\boldsymbol{\omega}'\cdot\boldsymbol{\sigma})}}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^{\alpha c \int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega}' \; P(\boldsymbol{\omega}') \cos(\boldsymbol{\omega}'\cdot\boldsymbol{\sigma})}}, \tag{2.85}$$

$$P(\boldsymbol{\omega}) = \left\langle \frac{\mathrm{e}^{cF(\boldsymbol{\omega})} \prod_\alpha D_\psi(\omega_\alpha|\beta)}{\int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega}' \; \mathrm{e}^{cF(\boldsymbol{\omega}')} \prod_\alpha D_\psi(\omega_\alpha|\beta)} \right\rangle, \tag{2.86}$$

where we have introduced

$$D_\psi(\omega|\beta) = \frac{1}{2\pi} \sum_{M\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}\omega M + \beta\chi(M,\psi)}. \tag{2.87}$$

Similarly, (2.84) and (2.80) can now be expressed as, respectively,

$$P(M|\psi) = \lim_{n\to 0} \left. \frac{\sum_{\mathbf{M}} \left(\frac{1}{n}\sum_\gamma \delta_{M,M_\gamma}\right) \int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega} \; \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M} + \beta\sum_\alpha \chi(M^\alpha,\psi) + cF(\boldsymbol{\omega})}}{\sum_{\mathbf{M}} \int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega} \; \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M} + \beta\sum_\alpha \chi(M^\alpha,\psi) + cF(\boldsymbol{\omega})}} \right|_{\chi = M^2/2c + M\psi}, \tag{2.88}$$

and

$$\begin{aligned} f[\chi] &= -\lim_{n\to 0} \frac{1}{\beta n} \Big\{ -c\alpha \int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega} \; F(\boldsymbol{\omega}) P(\boldsymbol{\omega}) + \log\Big(\sum_{\boldsymbol{\sigma}} \mathrm{e}^{\alpha c \int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega} \; P(\boldsymbol{\omega})[\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma}) - 1]} \Big) \\ &\qquad + \alpha \Big\langle \log\Big(\sum_{\mathbf{M}} \int_{-\pi}^{\pi} \frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n} \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M} + \sum_\alpha \beta\chi(M^\alpha,\psi) + cF(\boldsymbol{\omega})} \Big) \Big\rangle \Big\}. \end{aligned} \tag{2.89}$$

We note that the saddle-point equations guarantee that $P(\boldsymbol{\omega})$ is normalised correctly on $[-\pi,\pi]^n$, while for $F(\boldsymbol{\omega})$ we have (see C.2)

$$\int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega} \, F(\boldsymbol{\omega}) = 0. \tag{2.90}$$

We observe that in the absence of external fields, i.e. for $\psi = 0$, the function (2.87) is real and symmetric:

$$D_0(\omega|\beta) = \frac{1}{2\pi} \sum_{M\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}\omega M + \frac{\beta}{2c}M^2} \in \mathbb{R}, \qquad \forall\omega\in[-\pi,\pi]: \quad D_0(-\omega|\beta) = D_0(\omega|\beta). \tag{2.91}$$

The introduction of external fields breaks the symmetry of $D_\psi(\omega|\beta)$ under the transformation $\omega \to -\omega$.

2.3.3 The RS ansatz – route I

To solve the saddle point equations for $n \to 0$ we need to make an ansatz on the form of the order parameter functions $P(\boldsymbol{\omega})$ and $F(\boldsymbol{\omega})$. Since the conditioned overlap distribution (2.88) depends on $F(\boldsymbol{\omega})$ only, a first route to proceed is eliminating the order function $P(\boldsymbol{\omega})$ from our equations and making a replica-symmetric (RS) ansatz for $F(\boldsymbol{\omega})$. Since $\boldsymbol{\omega} \in [-\pi,\pi]^n$ is continuous, the RS ansatz for $F(\boldsymbol{\omega})$ reads:

$$F(\boldsymbol{\omega}) \quad = \quad \int \{\mathrm{d}\pi\}\ W[\{\pi\}] \prod_{\alpha=1}^{n} \pi(\omega_\alpha), \tag{2.92}$$

where $W[\ldots]$ is a measure over functions, normalised according to $\int \{\mathrm{d}\pi\}\ W[\{\pi\}] = 1$ and nonzero (in view of (2.90)) only for functions $\pi(\ldots)$ that are real and obey $\int_{-\pi}^{\pi} \mathrm{d}\omega\ \pi(\omega) = 0$. The RS ansatz (2.92) is to be inserted into the saddle point equations. Insertion into (2.86) gives, with a normalization factor $C_n(\psi)$,

$$\begin{aligned} P(\boldsymbol{\omega}) \quad &= \quad \Big\langle C_n^{-1}(\psi) \prod_\alpha D_\psi(\omega_\alpha|\beta)\ \mathrm{e}^{c\int\{d\pi\}\,W[\{\pi\}]\prod_\alpha \pi(\omega_\alpha)} \Big\rangle \\ &= \quad \Big\langle C_n^{-1}(\psi) \prod_\alpha D_\psi(\omega_\alpha|\beta) \sum_{k\geq 0} \frac{c^k}{k!} \Big[\int \{\mathrm{d}\pi\}\, W[\{\pi\}] \prod_\alpha \pi(\omega_\alpha) \Big]^k \Big\rangle \\ &= \quad \Big\langle C_n^{-1}(\psi) \sum_{k\geq 0} \frac{c^k}{k!} \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}] \right] \prod_\alpha R_k(\omega_\alpha) \Big\rangle, \end{aligned} \tag{2.93}$$

with

$$R_k(\omega) = D_\psi(\omega|\beta) \prod_{\ell=1}^{k} \pi_\ell(\omega). \tag{2.94}$$

Next we turn to (2.85). We first work out for $\boldsymbol{\sigma} \in \{-1,1\}^n$ the quantity

$$\begin{aligned} L(\boldsymbol{\sigma}) \quad &= \quad \alpha c \int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega}\ P(\boldsymbol{\omega}) \cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma}) \\ &= \quad \alpha c \langle C_n^{-1}(\psi) \sum_{k\geq 0} \frac{c^k}{k!} \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}] \right] \\ &\quad \times \Big[\frac{1}{2} \prod_\alpha \int_{-\pi}^{\pi} \mathrm{d}\omega_\alpha\, R_k(\omega_\alpha) \mathrm{e}^{\mathrm{i}\omega_\alpha \sigma^\alpha} + \frac{1}{2} \prod_\alpha \int_{-\pi}^{\pi} \mathrm{d}\omega_\alpha\, R_k(\omega_\alpha) \mathrm{e}^{-\mathrm{i}\omega_\alpha \sigma^\alpha} \Big] \rangle, \end{aligned} \tag{2.95}$$

with $\int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega}\, P(\boldsymbol{\omega}) = 1$ requiring $L(\mathbf{0}) = \alpha c$. For Ising spins one can use the general identity

$$\tilde{R}_k(\sigma) = \int_{-\pi}^{\pi} \mathrm{d}\omega\, R_k(\omega) \mathrm{e}^{\mathrm{i}\omega\sigma} = B(\{R_k\}) \mathrm{e}^{\mathrm{i}A(\{R_k\})\sigma}, \tag{2.96}$$

where B and A are, respectively, the absolute value and the argument of the complex function $\tilde{R}_k$ evaluated at the point 1, $\tilde{R}_k(1) = |\tilde{R}_k(1)|\, \mathrm{e}^{\mathrm{i}\phi_{\tilde{R}(1)}}$, i.e.

$$B(\{R_k\}) = |\tilde{R}_k(1)|, \qquad A(\{R_k\}) = \phi_{\tilde{R}(1)} = \arctan\Big(\frac{\mathrm{Im}[\tilde{R}_k(1)]}{\mathrm{Re}[\tilde{R}_k(1)]} \Big). \tag{2.97}$$

This simplifies (2.95) to

$$L(\boldsymbol{\sigma}) = \alpha c \langle C_n^{-1}(\psi) \sum_{k\geq 0} \frac{c^k}{k!} \int \prod_{\ell=1}^{k} \Big[\{\mathrm{d}\pi_\ell\} W(\{\pi_\ell\})\Big] B^n(\{R_k\}) \cos\Big[A(\{R_k\}) \sum_\alpha \sigma^\alpha\Big]\rangle. \quad (2.98)$$

In order to have $L(\mathbf{0}) = \alpha c$ in the limit $n \to 0$, one must have $C_0(\psi) = e^c\ \forall\psi$. Inserting $L(\boldsymbol{\sigma})$ into (2.85) gives

$$K_n F(\boldsymbol{\omega}) = \sum_{\boldsymbol{\sigma}} \cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma}) \mathrm{e}^{c\alpha\langle C_n^{-1}(\psi) \sum_{k\geq 0} \frac{c^k}{k!} \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}]\right] B^n(\{R_k\}) \cos\left[A(\{R_k\}) \sum_\alpha \sigma^\alpha\right]\rangle}, \quad (2.99)$$

with

$$K_n = \sum_{\boldsymbol{\sigma}} \mathrm{e}^{c\alpha\langle C_n^{-1}(\psi) \sum_{k\geq 0} \frac{c^k}{k!} \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}]\right] B^n(\{R_k\}) \cos\left[A(\{R_k\}) \sum_\alpha \sigma^\alpha\right]\rangle}. \quad (2.100)$$

Upon isolating the term $\sum_\alpha \sigma^\alpha$ via $\sum_m \int_{-\pi}^{\pi} \frac{\mathrm{d}\theta}{2\pi} \mathrm{e}^{\mathrm{i}m\theta - \mathrm{i}\theta \sum_\alpha \sigma^\alpha} = 1$ we obtain

$$\begin{aligned} K_n F(\boldsymbol{\omega}) &= \sum_m \int_{-\pi}^{\pi} \frac{\mathrm{d}\theta}{2\pi}\ \mathrm{e}^{\mathrm{i}m\theta + c\alpha\langle C_n^{-1}(\psi) \sum_{k\geq 0} \frac{c^k}{k!} \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}]\right] B^n(\{R_k\}) \cos[A(\{R_k\}) m]\rangle} \\ &\qquad \times \sum_{\boldsymbol{\sigma}} \mathrm{e}^{-\mathrm{i}\theta \sum_\alpha \sigma^\alpha} \Big(\frac{1}{2} \mathrm{e}^{\mathrm{i} \sum_\alpha \sigma^\alpha \omega^\alpha} + \frac{1}{2} \mathrm{e}^{-\mathrm{i} \sum_\alpha \sigma^\alpha \omega^\alpha}\Big) \\ &= 2^{n-1} \sum_m \int_{-\pi}^{\pi} \frac{\mathrm{d}\theta}{2\pi} \mathrm{e}^{\mathrm{i}m\theta + c\alpha\langle C_n^{-1}(\psi) \sum_{k\geq 0} \frac{c^k}{k!} \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}]\right] B^n(\{R_k\}) \cos[A(\{R_k\}) m]\rangle} \\ &\qquad \times \Big[\prod_\alpha \cos(\omega^\alpha - \theta) + \prod_\alpha \cos(\omega^\alpha + \theta)\Big]. \end{aligned} \quad (2.101)$$

The two terms inside the square brackets in the last line yield identical contributions to the θ-integral, so

$$K_n F(\boldsymbol{\omega}) = 2^n \sum_m \int_{-\pi}^{\pi} \frac{\mathrm{d}\theta}{2\pi}\ \mathrm{e}^{\mathrm{i}m\theta + c\alpha\langle C_n^{-1}(\psi) \sum_{k\geq 0} \frac{c^k}{k!} \int \prod_{\ell=1}^{k} \left[\{d\pi_\ell\} W[\{\pi_\ell\}]\right] B^n(\{R_k\}) \cos[A(\{R_k\}) m]\rangle} \prod_\alpha \cos(\omega^\alpha - \theta), \quad (2.102)$$

with K_0 simply following from the demand $F(\boldsymbol{\omega} = \mathbf{0}) = 1$, as required by (2.85). Next we insert

$$1 = \int \{\mathrm{d}\pi\} \prod_\omega \delta\Big[\pi(\theta) - \cos(\omega - \theta)\Big], \quad (2.103)$$

where we have used the symbolic notation $\prod_\omega \delta[\pi(\omega) - f(\omega)]$ for the functional version of the δ-distribution, as defined by the identity $\int \{\mathrm{d}\pi\}\, G[\{\pi\}] \prod_\omega \delta[\pi(\omega) - f(\omega)] = G[\{f\}]$. This leads us to

$$\begin{aligned} K_n F(\boldsymbol{\omega}) &= 2^n \sum_m \int_{-\pi}^{\pi} \frac{\mathrm{d}\theta}{2\pi} \mathrm{e}^{\mathrm{i}m\theta + c\alpha\langle C_n^{-1}(\psi) \sum_{k\geq 0} \frac{c^k}{k!} \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}]\right] B^n(\{R_k\}) \cos[A(\{R_k\}) m]\rangle} \\ &\qquad \times \int \{\mathrm{d}\pi\} \prod_\omega \delta\Big[\pi(\theta) - \cos(\omega - \theta)\Big] \prod_\alpha \pi(\omega^\alpha). \end{aligned} \quad (2.104)$$

Substituting (2.92) for $F(\boldsymbol{\omega})$ in the left-hand side of this last equation shows that in the replica limit $n \to 0$, our RS ansatz indeed generates a saddle point if

$$W[\{\pi\}] \;=\; \int_{-\pi}^{\pi} \frac{\mathrm{d}\theta}{2\pi}\; \lambda(\theta|W) \prod_{\omega} \delta\Big[\pi(\omega) - \cos(\omega-\theta)\Big], \tag{2.105}$$

with the short-hand

$$\lambda(\theta|W) = K_0^{-1} \sum_{m\in\mathbb{Z}} \mathrm{e}^{im\theta + c\alpha \sum_{k\geq 0} \frac{c^k e^{-c}}{k!} \langle \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}]\right] \cos[A(\{R_k\})m]\rangle}. \tag{2.106}$$

The constant K_0 follows simply from normalisation, which now takes the form $\int_{-\pi}^{\pi} \frac{\mathrm{d}\theta}{2\pi}\, \lambda(\theta|W) = 1$, giving

$$\begin{aligned} K_0 &= \int \frac{\mathrm{d}\theta}{2\pi} \sum_{m\in\mathbb{Z}} \mathrm{e}^{im\theta + c\alpha \sum_{k\geq 0} \frac{c^k e^{-c}}{k!} \langle \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}]\right] \cos[A(\{R_k\})m]\rangle} \\ &= \sum_{m\in\mathbb{Z}} \delta_{m,0}\; \mathrm{e}^{c\alpha \sum_{k\geq 0} \frac{c^k e^{-c}}{k!} \langle \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}]\right] \cos[A(\{R_k\})m]\rangle} = \mathrm{e}^{c\alpha}. \end{aligned} \tag{2.107}$$

We then arrive at

$$\lambda(\theta|W) = \sum_{m\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}m\theta + c\alpha \sum_{k\geq 0} \frac{c^k e^{-c}}{k!} \langle \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}]\right] \left[\cos[A(\{R_k\})m] - 1\right]\rangle}. \tag{2.108}$$

It is convenient to write $D(\omega|\beta) = D'(\omega|\beta) + \mathrm{i}D''(\omega|\beta)$, with $D'(\omega|\beta) = \mathrm{Re}[D(\omega|\beta)]$ and $D''(\omega|\beta) = \mathrm{Im}[D(\omega|\beta)]$. Similarly, we write $R_k(\omega) = R'_k(\omega) + \mathrm{i}R''_k(\omega)$. We note that for $\chi(M,\psi) = M^2/2c + M\psi$ the function $D_\psi(\omega|\beta)$ defined in (2.87) has several useful properties, e.g.

$$\forall \omega \in [-\pi,\pi]: \qquad D'_\psi(-\omega|\beta) = D'_\psi(\omega|\beta), \qquad D''_\psi(-\omega|x) = -D''_\psi(\omega|x), \tag{2.109}$$

$$\int_{-\pi}^{\pi} \mathrm{d}\omega\; D_\psi(\omega|\beta) = \sum_{M\in\mathbb{Z}} \mathrm{e}^{\beta\chi(M,\psi)} \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi}\; \mathrm{e}^{\mathrm{i}\omega M} = \sum_{M\in\mathbb{Z}} \mathrm{e}^{\beta\chi(M,\psi)} \delta_{M,0} = 1, \tag{2.110}$$

$$D_\psi(\omega|0) = \frac{1}{2\pi} \sum_{M\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}\omega M} = \delta(\omega) \quad \text{for } \omega \in [-\pi,\pi]. \tag{2.111}$$

From (2.97) we have

$$A(\{R_k\}) = \arctan\Big[\frac{\mathrm{Im}[\tilde{R}_k(1)]}{\mathrm{Re}[\tilde{R}_k(1)]}\Big] = \arctan\Big[\frac{\int_{-\pi}^{\pi}\mathrm{d}\omega\,[R'_k(\omega)\sin\omega + R''_k(\omega)\cos\omega]}{\int_{-\pi}^{\pi}\mathrm{d}\omega\,[R'_k(\omega)\cos\omega - R''_k(\omega)\sin\omega]}\Big], \tag{2.112}$$

and insertion in (2.108) gives

$$\lambda(\theta|W) = \sum_{m\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}m\theta + c\alpha \sum_{k\geq 0} \frac{c^k e^{-c}}{k!} \int \prod_{\ell=1}^{k} \left[\{\mathrm{d}\pi_\ell\} W[\{\pi_\ell\}]\right] \left\{\cos[m \arctan f_k(\{\pi_1,\ldots,\pi_k\})] - 1\right\}}, \tag{2.113}$$

with

$$f_k(\{\pi_1,\ldots,\pi_k\}) = \frac{\int_{-\pi}^{\pi}\mathrm{d}\omega\,[D'(\omega|\beta)\sin\omega + D''(\omega|\beta)\cos\omega]\prod_{\ell=1}^{k}\pi_\ell(\omega)}{\int_{-\pi}^{\pi}\mathrm{d}\omega\,[D'(\omega|\beta)\cos\omega - D''(\omega|\beta)\sin\omega]\prod_{\ell=1}^{k}\pi_\ell(\omega)}. \tag{2.114}$$

For high temperatures $D'(\omega|0) = \delta(\omega)$ and $D''(\omega|0) = 0$, so $f_k(\{\pi_1,\ldots,\pi_k\}) = 0$ and $\lambda(\theta|W) = \delta(\theta)$. Hence

$$\beta = 0: \qquad W[\{\pi\}] = \prod_\omega \delta\Big[\pi(\omega) - \cos(\omega)\Big]. \tag{2.115}$$

We note that for any symmetric set of functions $\{\pi_1,\ldots,\pi_k\}$ one has, from (2.114), $f_k(\{\pi_1,\ldots,\pi_k\}) = 0$ due to the symmetry properties (2.109) of D_ψ, and thus $\lambda(\theta|W) = \delta(\theta)$. Hence, (2.115) is a solution of (2.105) for *all* temperatures, and the only solution at infinite temperature.

2.3.4 Conditioned distribution of overlaps

In order to give a physical interpretation to the RS solution (2.92,2.115), we consider the conditioned overlap distribution (2.88). Insertion of (2.115) into (2.92) gives

$$F(\boldsymbol{\omega}) = \int\{\mathrm{d}\pi\}\ W[\{\pi\}]\prod_\alpha \pi(\omega_\alpha) = \prod_\alpha \cos(\omega_\alpha),$$

and subsequent insertion into (2.88) leads to, with C_n and $\tilde{C}_n$ representing normalization constants,

$$\begin{aligned}
P(M|\psi) &= \lim_{n\to 0} C_n^{-1}\sum_{\mathbf{M}}\Big(\frac{1}{n}\sum_{\gamma=1}^{n}\delta_{M,M_\gamma}\Big)\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\ \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}+\beta\sum_\alpha\chi(M_\alpha,\psi)}\sum_{k\geq 0}\frac{c^k}{k!}\prod_\alpha\cos^k(\omega_\alpha) \\
&= \lim_{n\to 0}\frac{\tilde{C}_n^{-1}}{n}\sum_{k\geq 0}\frac{c^k}{k!}\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\prod_\alpha\cos^k(\omega_\alpha)\int_{-\pi}^{\pi}\mathrm{d}\lambda\ \mathrm{e}^{\mathrm{i}\lambda M}\sum_{\gamma=1}^{n}\sum_{M_\gamma\in\mathbb{Z}}\mathrm{e}^{\mathrm{i}(\omega_\gamma-\lambda)M_\gamma+\chi(M_\gamma,\psi)} \\
&\qquad\qquad\times\prod_{\alpha\neq\gamma}\sum_{M_\alpha}\mathrm{e}^{\mathrm{i}\omega_\alpha M_\alpha+\chi(M^\alpha,\psi)} \\
&= \lim_{n\to 0}\frac{C_n^{-1}}{n}\sum_{k\geq 0}\frac{c^k}{k!}\int_{-\pi}^{\pi}\mathrm{d}\lambda\ \mathrm{e}^{\mathrm{i}\lambda M}\sum_{\gamma=1}^{n}\int_{-\pi}^{\pi}\mathrm{d}\omega_\gamma\ \cos^k(\omega_\gamma)D_\psi(\omega_\gamma-\lambda|\beta) \\
&\qquad\qquad\times\prod_{\alpha\neq\gamma}\int_{-\pi}^{\pi}\mathrm{d}\omega_\alpha\ \cos^k(\omega_\alpha)D_\psi(\omega_\alpha|\beta) \\
&= \lim_{n\to 0} C_n^{-1}\sum_{k\geq 0}\frac{c^k}{k!}\int_{-\pi}^{\pi}\mathrm{d}\lambda\ \mathrm{e}^{\mathrm{i}\lambda M}I_k(\lambda,\beta)I_k^{n-1}(0,\beta),
\end{aligned} \tag{2.116}$$

with

$$
\begin{aligned}
I_k(\lambda,\beta) &= \int_{-\pi}^{\pi}\mathrm{d}\omega\;\cos^k(\omega)D_\psi(\omega-\lambda|\beta)=\frac{1}{2^k}\sum_{n=0}^{k}\binom{k}{n}\int_{-\pi}^{\pi}\mathrm{d}\omega\;\mathrm{e}^{-\mathrm{i}\omega(k-2n)}\sum_{m\in\mathbb{Z}}\mathrm{e}^{\mathrm{i}(\omega-\lambda)m+\beta\chi(m,\psi)} \\
&= \frac{1}{2^k}\sum_{n=0}^{k}\binom{k}{n}\mathrm{e}^{-\mathrm{i}\lambda(k-2n)+\beta\chi(k-2n,\psi)}=\frac{1}{2^k}\sum_{m=-k}^{k}\binom{k}{\frac{k-m}{2}}\mathrm{e}^{-\mathrm{i}\lambda m+\beta\chi(m,\psi)}. \qquad (2.117)
\end{aligned}
$$

We can now work out

$$
\int_{-\pi}^{\pi}\mathrm{d}\lambda\;\mathrm{e}^{\mathrm{i}\lambda M}I_k(\lambda,\beta)=\begin{cases} 2^{-k}\binom{k}{(k-M)/2}\mathrm{e}^{\beta\chi(M,\psi)} & \text{if } |M|\le k \\ 0 & \text{if } |M|>k \end{cases}, \qquad (2.118)
$$

and obtain our desired formula for $P(M|\psi)$ corresponding to the saddle-point (2.115), in which the normalisation constant comes out as $C_0=\mathrm{e}^c$. The result then is

$$
P(M|\psi)=\sum_{k\ge|M|}\mathrm{e}^{-c}\frac{c^k}{k!}\;\frac{\binom{k}{(k-M)/2}\mathrm{e}^{\beta\chi(M,\psi)}}{\sum_{m=-k}^{k}\binom{k}{(k-m)/2}\mathrm{e}^{\beta\chi(m,\psi)}}. \qquad (2.119)
$$

We can rewrite this result, with the short-hand $p_c(k)=\mathrm{e}^{-c}c^k/k!$, in the more intuitive form

$$
P(M|\psi) = \sum_{k\ge0}p_c(k)P(M|k,\psi), \qquad (2.120)
$$

$$
P(M|k,\psi) = \theta(k-|M|+\frac{1}{2})\;\frac{\binom{k}{(k-M)/2}\mathrm{e}^{\beta\chi(M,\psi)}}{\sum_{m=-k}^{k}\binom{k}{(k-m)/2}\mathrm{e}^{\beta\chi(m,\psi)}}. \qquad (2.121)
$$

We recognise that $p_c(k)$ is the asymptotic probability that any pattern $(\xi_1^\mu,\ldots,\xi_N^\mu)$ has k non-zero entries; since each pattern has N independent entries with probability c/N to be nonzero, k will for $N\to\infty$ indeed be a Poissonian random variable with average c. Hence, $P(M|k,\psi)$ is the conditional probability to have an overlap of value M, given the pattern concerned has k non-zero entries and is triggered by an external field ψ. We have apparently mapped the neural network with N neurons and $P=\alpha N$ diluted stored patterns to a system of k neurons with a single undiluted binary pattern. We will see that this is due to the fact that in the regime where replica-symmetric theory holds one is always able, as a consequence of the dilution, to decompose the original system into an extensive number of independent finite-sized subsystems, each recalling one particular pattern.

The solution (2.115), leading to (2.121), is a saddle-point for any temperature. At infinite temperatures it is the only solution, and simplifies further. For $\beta = 0$ expression (2.121) gives

$$P(M|k,\psi) = 2^{-k}\binom{k}{(k-M)/2}\theta(k-|M|+\frac{1}{2}), \tag{2.122}$$

which is the probability that a system of k spins has an overlap M with an undiluted stored pattern, if each spin behaves completely randomly. This describes, as expected, an immune network behaving as a paramagnet, i.e. unable to retrieve stored strategies. For the distribution of overlaps we find

$$P(M) = \mathrm{e}^{-c}\sum_{k\geq|M|}\frac{(\frac{1}{2}c)^k}{k!}\binom{k}{(k-M)/2}. \tag{2.123}$$

In the limit $\beta \to \infty$, the sum in the denominator of (2.121) is dominated by the value of m which maximises $\chi(m,\psi) = m^2/2c+\psi m$, being $m{=}k\ \mathrm{sgn}(\psi)$ if $\psi \neq 0$ and $m{=}\pm k$ for $\psi = 0$. In either case we obtain

$$\sum_{m=-k}^{k}\binom{k}{(k-m)/2}\mathrm{e}^{\beta\chi(m,\psi)} \sim \begin{cases} \mathrm{e}^{\beta(k^2/2c+k|\psi|)} & \psi \neq 0 \\ 2\mathrm{e}^{\beta k^2/2c} & \psi = 0, k \neq 0. \end{cases} \tag{2.124}$$

Substitution into (2.121) and (2.120) subsequently gives

$$\begin{aligned}\lim_{\beta\to\infty} P(M|\psi) &= \lim_{\beta\to\infty}\begin{cases} \mathrm{e}^{-c}\sum_{k\geq|M|}\frac{c^k}{k!}\binom{k}{(k-M)/2}\mathrm{e}^{-\beta(k^2-M^2)/2c-\beta|\psi|(k-\mathrm{sgn}(\psi)M)} & \text{if } \psi \neq 0 \\ \frac{1}{2}\mathrm{e}^{-c}\sum_{k\geq|M|}\frac{c^k}{k!}\binom{k}{(k-M)/2}\mathrm{e}^{-\beta(k^2-M^2)/2c} & \text{if } \psi = 0, M \neq 0 \end{cases} \\ &= \begin{cases} \mathrm{e}^{-c} & \text{if } M = 0 \\ \theta(M\psi)\ \mathrm{e}^{-c}c^{|M|}/|M|! & \text{if } \psi \neq 0,\ M \neq 0 \\ \frac{1}{2}\mathrm{e}^{-c}c^{|M|}/|M|! & \text{if } \psi = 0,\ M \neq 0 \end{cases}.\end{aligned} \tag{2.125}$$

Similarly we have

$$\psi \neq 0: \qquad P(M|k,\psi) = \delta_{|M|,k}\Big(\delta_{M,0} + \theta(\psi M)(1-\delta_{M,0})\Big), \tag{2.126}$$

$$\psi = 0: \qquad P(M|k,\psi) = \delta_{|M|,k}\Big(\delta_{M,0} + \frac{1}{2}(1-\delta_{M,0})\Big). \tag{2.127}$$

For $k > 0$ this describes error-free activation or inhibition of a stored strategy with k nonzero entries.

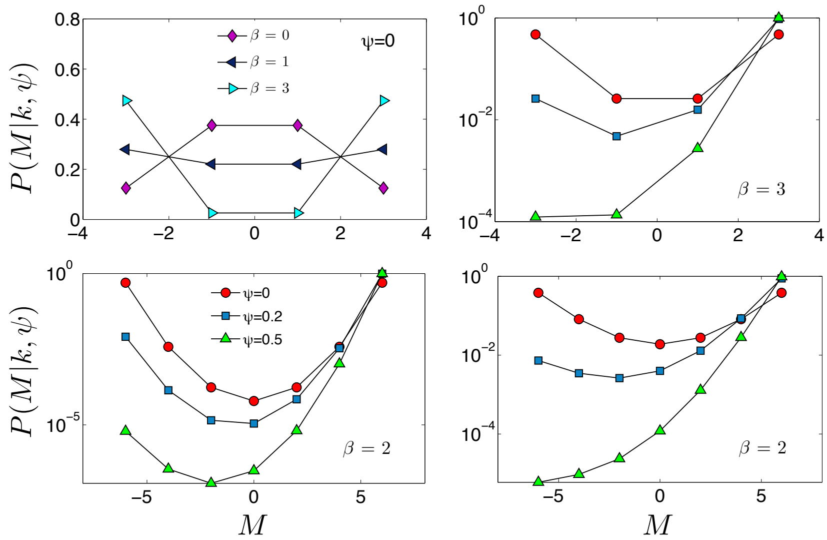

Figure 2.3: Conditioned overlap distribution $P(M|k,\psi)$ corresponding to the state (2.92,2.115), as given by formula (2.121). Top panels refer to $k = c = 3$. Left: $\beta = 0, 1, 3$ and $\psi = 0$; Right: $\beta = 3$ and $\psi = 0, 0.2, 0.5$. Bottom panels refer to $\psi = 0, 0.2, 0.5$ and $\beta = 2$. Left: $c = 3, k = 6$. Right: $c = k = 6$. Note that $M \in \{-k, -k+1, \ldots, k-1, k\}$, so that the lines connecting markers are only guides to the eye.

For intermediate temperatures a plot of (2.121) shows that without external fields, $P(M|0)$ acquires two symmetric peaks at large overlaps (in absolute value), as β is increased from $\beta = 0$; see Fig. 2.3, top left panel. Unlike typical magnetic systems in the thermodynamic limit, there is no spontaneous ergodicity breaking at $\psi = 0$; the system acts effectively as an extensive number of independent finite subsystems, each devoted to a single pattern. Each size-k subsystem oscillates randomly between the the two peaks in $P(M|0)$, with a characteristic switching timescale $t_k \sim e^{\beta k^2/2c}$, which grows with the size k of the subsystem and remains finite at finite temperature.

Introducing a field ψ reduces the overlap peak at M values opposite in sign to the field; this peak will eventually disappear for sufficiently strong fields (Fig. 2.3, top right panel). The field-induced asymmetry in the height of the two peaks increases at smaller temperatures and larger sizes (Fig. 2.3, bottom panels). Thus, external fields trigger the system towards either activation or inhibition of a strategy, whereas in their absence the system oscillates stochastically between the two.

2.3.5 Alternative formulation of the theory before the RS ansatz

The approach developed in the previous section led to transparent formulae for the distribution of overlaps in the RS state (2.115), and even allows us to derive analytically the condition defining the (continuous) phase transition where (2.115) ceases to hold (see C.3). However, the states beyond the transition point are better described within an alternative (but mathematically equivalent) formulation of the theory. This alternative approach is based on formulating our equations first in terms of the following quantities:

$$L(\boldsymbol{\sigma}) = \alpha c \int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega}\ P(\boldsymbol{\omega}) \cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma}), \qquad Q(\boldsymbol{\omega}) = \mathrm{e}^{cF(\boldsymbol{\omega})}. \tag{2.128}$$

Both $P(\boldsymbol{\omega})$ and $Q(\boldsymbol{\omega})$ are only defined for $\boldsymbol{\omega} \in [-\pi,\pi]^n$. In terms of (2.128) we can write our earlier saddle point equations (2.86, 2.85) as

$$P(\boldsymbol{\omega}) = \Big\langle \frac{Q(\boldsymbol{\omega}) \sum_{\mathbf{M}\in\mathbb{Z}^n} \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}+\sum_\alpha \chi(M_\alpha,\psi)}}{\int_{\pi}^{\pi} \mathrm{d}\boldsymbol{\omega}'\, Q(\boldsymbol{\omega}') \sum_{\mathbf{M}\in\mathbb{Z}^n} \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}'\cdot\mathbf{M}+\sum_\alpha \chi(M_\alpha,\psi)}} \Big\rangle_\psi, \tag{2.129}$$

$$\log Q(\boldsymbol{\omega}) = c\, \frac{\sum_{\boldsymbol{\sigma}\in\{-1,1\}^n} \cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma}) \mathrm{e}^{L(\boldsymbol{\sigma})}}{\sum_{\boldsymbol{\sigma}\in\{-1,1\}^n} \mathrm{e}^{L(\boldsymbol{\sigma})}}, \tag{2.130}$$

and the free energy (2.89) as

$$\begin{aligned} f[\chi] = & -\lim_{n\to 0} \frac{1}{\beta n} \Big\{ \log\Big(\sum_{\boldsymbol{\sigma}} \mathrm{e}^{L(\boldsymbol{\sigma})-c\alpha}\Big) - \frac{\sum_{\boldsymbol{\sigma}} L(\boldsymbol{\sigma}) \mathrm{e}^{L(\boldsymbol{\sigma})}}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^{L(\boldsymbol{\sigma})}} \\ & + \alpha \Big\langle \log\Big(\sum_{\mathbf{M}} \int_{-\pi}^{\pi} \frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\ \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}+\sum_\alpha \beta\chi(M_\alpha,\psi)} Q(\boldsymbol{\omega})\Big)\Big\rangle_\psi \Big\}, \end{aligned} \tag{2.131}$$

where we used $\alpha\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\, P(\boldsymbol{\omega}) \log Q(\boldsymbol{\omega}) = \sum_{\boldsymbol{\sigma}} L(\boldsymbol{\sigma})\mathrm{e}^{L(\boldsymbol{\sigma})}/\sum_{\boldsymbol{\sigma}} \mathrm{e}^{L(\boldsymbol{\sigma})}$. Clearly $\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\ P(\boldsymbol{\omega}) = 1$, $Q(\boldsymbol{\omega}) \in \mathbb{R}$, $Q(-\boldsymbol{\omega}) = Q(\boldsymbol{\omega})$, and $Q(\mathbf{0}) = \mathrm{e}^c$. We can now switch from the order parameter $Q(\boldsymbol{\omega})$ to a new order parameter $\tilde{Q}(\mathbf{M})$, defined on $\mathbf{M} \in \mathbb{Z}^n$, via the following one-to-one transformations:

$$\tilde{Q}(\mathbf{M}) = \int_{-\pi}^{\pi} \frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\ Q(\boldsymbol{\omega}) \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}}, \qquad Q(\boldsymbol{\omega}) = \sum_{\mathbf{M}\in\mathbb{Z}^n} \tilde{Q}(\mathbf{M}) \mathrm{e}^{-\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}}. \tag{2.132}$$

The validity of these equations follows from the two identities $(2\pi)^{-1}\int_{-\pi}^{\pi}\mathrm{d}\omega\ \mathrm{e}^{\mathrm{i}\omega m} = \delta_{m0}$ for $m \in \mathbb{Z}$, and $(2\pi)^{-1}\sum_{M\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}\omega M} = \delta(\omega)$ for $\omega \in [-\pi,\pi]$. By construction we now have $\sum_{\mathbf{M}} \tilde{Q}(\mathbf{M}) = \mathrm{e}^c$. Moreover, since $Q(-\boldsymbol{\omega}) = Q(\boldsymbol{\omega})$ we also know that $\tilde{Q}(\mathbf{M}) = (2\pi)^{-n}\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\, Q(\boldsymbol{\omega}) \cos(\boldsymbol{\omega}\cdot\mathbf{M}) \in \mathbb{R}$. One can write the saddle point equations in terms of these order functions (see C.4

for details):

$$\tilde{Q}(\mathbf{M}) = \int_{-\pi}^{\pi} \mathrm{d}\boldsymbol{\omega}\, \cos(\boldsymbol{\omega}\cdot\mathbf{M}) \exp\Big[c\, \frac{\sum_{\boldsymbol{\sigma}} \cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})\mathrm{e}^{L(\boldsymbol{\sigma})}}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^{L(\boldsymbol{\sigma})}}\Big], \tag{2.133}$$

$$L(\boldsymbol{\sigma}) = \alpha c\, \mathrm{e}^{\frac{\beta n}{2c}} \Big\langle \frac{\sum_{\mathbf{M}} \tilde{Q}(\mathbf{M})\mathrm{e}^{\beta\sum_\alpha \chi(M_\alpha,\psi)} \cosh[\beta(\frac{1}{c}\mathbf{M}\cdot\boldsymbol{\sigma} + \psi\sum_\alpha \sigma^\alpha)]}{\sum_{\mathbf{M}} \tilde{Q}(\mathbf{M})\mathrm{e}^{\beta\sum_\alpha \chi(M_\alpha,\psi)}} \Big\rangle_\psi. \tag{2.134}$$

and the free energy reads

$$f[\chi] = -\lim_{n\to 0} \frac{1}{\beta n}\Big\{ \log\sum_{\boldsymbol{\sigma}} \mathrm{e}^{L(\boldsymbol{\sigma})-c\alpha} - \frac{\sum_{\boldsymbol{\sigma}} L(\boldsymbol{\sigma})\mathrm{e}^{L(\boldsymbol{\sigma})}}{\sum_{\boldsymbol{\sigma}} \mathrm{e}^{L(\boldsymbol{\sigma})}} + \alpha\Big\langle \log\Big[\sum_{\mathbf{M}} \mathrm{e}^{\sum_\alpha \beta\chi(M_\alpha,\psi)}\tilde{Q}(\mathbf{M})\Big]\Big\rangle_\psi \Big\}. \tag{2.135}$$

From (2.88) we find that the distribution of overlaps can be written as

$$P(M|\psi) = \lim_{n\to 0} \frac{\sum_{\mathbf{M}} \Big(\frac{1}{n}\sum_{\gamma=1}^n \delta_{M,M_\gamma}\Big)\mathrm{e}^{\beta\sum_\alpha \chi(M_\alpha,\psi)}\tilde{Q}(\mathbf{M})}{\sum_{\mathbf{M}} \mathrm{e}^{\beta\sum_\alpha \chi(M_\alpha,\psi)}\tilde{Q}(\mathbf{M})}\Bigg|_{\chi(M,\psi)=M^2/2c+M\psi}. \tag{2.136}$$

In C.5 we confirm the correctness of (2.136) in several special limits.

2.3.6 The RS ansatz – route II

We now try to construct the RS solution of our new equations (2.134, 2.133), by applying the RS ansatz to the functions $L(\boldsymbol{\sigma})$ and $\tilde{Q}(\mathbf{M})$:

$$L(\boldsymbol{\sigma}) = \alpha c \int \mathrm{d}h\, W(h) \prod_{\alpha=1}^{n} \mathrm{e}^{\beta h \sigma^\alpha}, \qquad \tilde{Q}(\mathbf{M}) = \mathrm{e}^c \int \{\mathrm{d}\pi\} W[\{\pi\}] \prod_\alpha \pi(M^\alpha), \tag{2.137}$$

with $\int \mathrm{d}h\, W(h) = 1$, $W(h) = W(-h)$, and with a (normalised) functional measure $W[\pi]$ that is only non-zero for functions $\pi(M)$ that are themselves normalised according to $\sum_{M\in\mathbb{Z}} \pi(M) = 1$. This ansatz meets the requirements $L(-\boldsymbol{\sigma}) = L(\boldsymbol{\sigma})$, $L(\mathbf{0}) = \alpha c$ and $\sum_{\mathbf{M}} \tilde{Q}(\mathbf{M}) = \mathrm{e}^c$, and is the most general form of the functions $L(\boldsymbol{\sigma})$ and $\tilde{Q}(\mathbf{M})$ that is invariant under all replica permutations. The advantage of this second formulation of the theory is that it allows us to work with a distribution $W(h)$ of effective fields, instead of functional measures over distributions, which have easier physical interpretations, and are more easy to solve numerically from self-consistent equations.

We relegate to C.6 all the details of the derivation of the RS equations, based on the form (2.137), the results of which can be summarised as follows. The RS functional measure $W[\pi]$

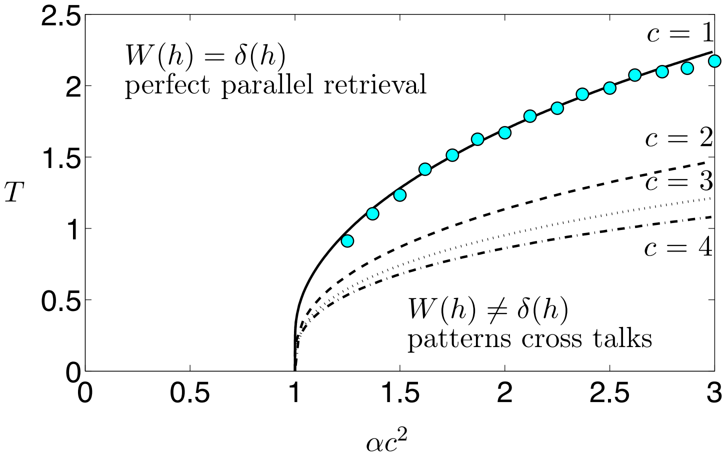

Figure 2.4: Transition lines (2.141) for $c = 1, 2, 3, 4$, in the $(\alpha c^2, T)$ plane, with $T = \beta^{-1}$. The distribution $W(h)$ represents the statistics of the interfering fields among different patterns, which are caused by increased connectivity in the graph $\mathcal{G}$. If $W(h) = \delta(h)$, spins are controlled via signaling patterns that can act independently; we see that this is possible even above the percolation threshold if the temperature (i.e. the signalling noise) is nonzero. Circles: transition calculated via numerical solution of (2.140) for $c = 1$ (see section 2.3.7).

and the field distribution $W(h)$ obey the following closed equations:

$$W(h) = \int\{\mathrm{d}\pi\}W[\pi]\Big\langle\Big\langle\delta\Big[h-\tau\psi-\frac{1}{2\beta}\log\Big(\frac{\sum_M \pi(M)\mathrm{e}^{\beta(M^2/2c+M(\psi+\tau/c))}}{\sum_M \pi(M)\mathrm{e}^{\beta(M^2/2c+M(\psi-\tau/c))}}\Big)\Big]\Big\rangle_\psi\Big\rangle_{\tau=\pm1}, \quad (2.138)$$

$$W[\pi] = e^{-c}\sum_{k\geq0}\frac{c^k}{k!}e^{-\alpha ck}\sum_{r\geq0}\frac{(\alpha c)^r}{r!}\int_{-\infty}^{\infty}\mathrm{d}h_1\ldots\mathrm{d}h_r\Big[\prod_{s\leq r}W(h_s)\Big]\sum_{\ell_1\ldots\ell_r\leq k}$$
$$\times\prod_M\delta\left[\pi(M)-\frac{\langle\mathrm{e}^{\beta\sum_{s\leq r}h_s\sigma_{\ell_s}}\delta_{M,\sum_{\ell\leq k}\sigma_\ell}\rangle_{\sigma_1\ldots\sigma_k}}{\langle\mathrm{e}^{\beta\sum_{s\leq r}h_s\sigma_{\ell_s}}\rangle_{\sigma_1\ldots\sigma_k}}\right]. \quad (2.139)$$

Both $W(h)$ and $W[\pi]$ are correctly normalised, $W(h) = W(-h)$, and $W[\pi]$ allows only for functions π such that $\pi(M) = \pi(-M)$ and $\sum_M \pi(M) = 1$. We can substitute the second equation into the first and eliminate the functional measure $W[\pi]$, leaving us with a compact RS equation for the field distribution $W(h)$ only:

$$W(h) = \mathrm{e}^{-c}\sum_{k\geq0}\frac{c^k}{k!}\mathrm{e}^{-\alpha ck}\sum_{r\geq0}\frac{(\alpha c)^r}{r!}\int_{-\infty}^{\infty}\mathrm{d}h_1\ldots\mathrm{d}h_r\Big[\prod_{s\leq r}W(h_s)\Big]\sum_{\ell_1\ldots\ell_r\leq k} \quad (2.140)$$
$$\times\Big\langle\Big\langle\delta\left[h-\tau\psi-\frac{1}{2\beta}\log\left(\frac{\langle\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta(\sum_{\ell\leq k}\tau_\ell)(\psi+\tau/c)+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\rangle_{\tau_1\ldots\tau_k=\pm1}}{\langle\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta(\sum_{\ell\leq k}\tau_\ell)(\psi-\tau/c)+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\rangle_{\tau_1\ldots\tau_k=\pm1}}\right)\right]\Big\rangle_\psi\Big\rangle_{\tau=\pm1}.$$

We see that $W(h) = \delta(h)$ is a solution of (2.140) for any temperature; one easily confirms that this is in fact the earlier state (2.115), recovered within the alternative formulation of the theory. If we inspect continuous bifurcations of new solutions with moments $m_r = \int \mathrm{d}h\ h^r W(h)$ different from zero, we find (see C.7) a second order transition along the critical surface in the (α, β, c)-space defined by

$$1 \ = \ \alpha c^2 \sum_{k\geq 0} \mathrm{e}^{-c} \frac{c^k}{k!} \left\{ \frac{\int \mathrm{D}z\ \tanh(z\sqrt{\beta/c}+\beta/c) \cosh^{k+1}(z\sqrt{\beta/c}+\beta/c)}{\int \mathrm{D}z\ \cosh^{k+1}(z\sqrt{\beta/c}+\beta/c)} \right\}^2 . \quad (2.141)$$

We note that the right-hand side obeys $0 \leq \mathrm{RHS} \leq \alpha c^2$, with $\lim_{\beta\to 0} \mathrm{RHS} = 0$ and $\lim_{\beta\to\infty} \mathrm{RHS} = \alpha c^2$. Hence a transition at finite temperature $T_c(\alpha, c) = \beta_c^{-1}(\alpha, c) > 0$ exists to a new state with $W(h) \neq \delta(h)$ as soon as $\alpha c^2 > 1$. The critical temperature becomes zero when $\alpha c^2 = 1$, consistent with the percolation threshold [86] derived from the network analysis. We show in C.8 that the critical surface (2.141) is indeed identical to the one found in (C.26), within the approach involving functional distributions.

Finally, within the new formulation of the theory, the replica-symmetric field-conditioned overlap distribution is found to be

$$\begin{aligned} P(M|\psi) \ &= \ \lim_{n\to 0} \frac{\int \{\mathrm{d}\pi\}\ W[\pi]\ \left(\sum_{M'} \pi(M')\mathrm{e}^{\beta(M'^2/2c+\psi M')}\right)^{n-1} \pi(M)\mathrm{e}^{\beta(M^2/2c+\psi M)}}{\int \{\mathrm{d}\pi\}\ W[\pi]\ \left(\sum_{M'} \pi(M')\mathrm{e}^{\beta(M'^2/2c+\psi M')}\right)^{n}} \\ &= \ \int \{\mathrm{d}\pi\} W[\pi] \Big\{ \frac{\pi(M)\mathrm{e}^{\beta(M^2/2c+\psi M)}}{\sum_{M'} \pi(M')\mathrm{e}^{\beta(M'^2/2c+\psi M')}} \Big\}. \end{aligned} \quad (2.142)$$

Insertion of (2.139) allows us to eliminate the functional measure in favour of effective field distributions:

$$\begin{aligned} P(M|\psi) \ &= \ \mathrm{e}^{-c} \sum_{k\geq 0} \frac{c^k}{k!} \mathrm{e}^{-\alpha c k} \sum_{r\geq 0} \frac{(\alpha c)^r}{r!} \int_{-\infty}^{\infty} \mathrm{d}h_1 \ldots \mathrm{d}h_r \Big[\prod_{s\leq r} W(h_s)\Big] \sum_{\ell_1\ldots\ell_r\leq k} \\ &\quad \times \left\{ \frac{\langle \mathrm{e}^{\beta\sum_{s\leq r} h_s \sigma_{\ell_s}} \delta_{M,\sum_{\ell\leq k}\sigma_\ell} \rangle_{\sigma_1\ldots\sigma_k} \mathrm{e}^{\beta(M^2/2c+\psi M)}}{\sum_{M'} \langle \mathrm{e}^{\beta\sum_{s\leq r} h_s \sigma_{\ell_s}} \delta_{M',\sum_{\ell\leq k}\sigma_\ell} \rangle_{\sigma_1\ldots\sigma_k} \mathrm{e}^{\beta(M'^2/2c+\psi M')}} \right\} \\ &= \ \mathrm{e}^{-c} \sum_{k\geq 0} \frac{c^k}{k!} \mathrm{e}^{-\alpha c k} \sum_{r\geq 0} \frac{(\alpha c)^r}{r!} \int_{-\infty}^{\infty} \mathrm{d}h_1 \ldots \mathrm{d}h_r \Big[\prod_{s\leq r} W(h_s)\Big] \sum_{\ell_1\ldots\ell_r\leq k} \\ &\quad \times \left\{ \frac{\langle \delta_{M,\sum_{\ell\leq k}\tau_\ell}\ \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta\psi\sum_{\ell\leq k}\tau_\ell+\beta\sum_{s\leq r} h_s\tau_{\ell_s}} \rangle_{\tau_1\ldots\tau_k=\pm 1}}{\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta\psi\sum_{\ell\leq k}\tau_\ell+\beta\sum_{s\leq r} h_s\tau_{\ell_s}} \rangle_{\tau_1\ldots\tau_k=\pm 1}} \right\} \end{aligned} \quad (2.143)$$

Again, we can rewrite this result (2.143) in the form (2.120), which is more useful to investigate the system's performance since it quantifies the statistics of overlaps relative to their maximum value k, with

$$P(M|k,\psi) = \mathrm{e}^{-\alpha ck}\sum_{r\geq 0}\frac{(\alpha c)^r}{r!}\int_{-\infty}^{\infty}\mathrm{d}h_1\ldots\mathrm{d}h_r\Big[\prod_{s\leq r}W(h_s)\Big]\sum_{\ell_1\ldots\ell_r\leq k}$$
$$\times\left\{\frac{\left\langle\delta_{M,\sum_{\ell\leq k}\tau_\ell}\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta\psi\sum_{\ell\leq k}\tau_\ell+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\right\rangle_{\tau_1\ldots\tau_k=\pm 1}}{\left\langle\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta\psi\sum_{\ell\leq k}\tau_\ell+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\right\rangle_{\tau_1\ldots\tau_k=\pm 1}}\right\} \quad (2.144)$$

The latter formula shows very clearly that h is to be interpreted as an interference field among different patterns, which is caused by overlapping signalling in the bi-partite graph $\mathcal{B}$ and leads to clique interactions in the effective H-H graph $\mathcal{G}$. Fortunately, we see in Figure 2.4 that even above the percolation threshold $\alpha c^2 = 1$, and even in the interfering phase the signalling performance of the system degrades only smoothly (see the section below).

2.3.7 Population dynamics calculation of the cross-talk field distribution

We solve numerically equation (2.140) for the interference field distribution $W(h)$ with a population dynamics algorithm [80], which is based on interpreting (2.140) as the fixed-point equation of a stochastic process and simulating this process numerically. One observes that (2.140) has the structural form

$$W(h) = \left\langle\left\langle\delta\left[h - h(k,r,\mathbf{h},\boldsymbol{\ell},\tau,\psi)\right]\right\rangle\right\rangle_{k,r,\mathbf{h},\boldsymbol{\ell},\tau,\psi}, \quad (2.145)$$

with the following set of random variables:

- $k \sim \text{Poisson}(c)$
- $r \sim \text{Poisson}(\alpha ck)$
- $\mathbf{h} = (h_1,\ldots,h_r):$ r i.i.d. random fields with probability density $W(h)$
- $\boldsymbol{\ell} = (\ell_1,\ldots,\ell_r):$ r i.i.d. discrete random variables, distributed uniformly over $\{1,\ldots,k\}$
- $\tau:$ dichotomic random variable, , distributed uniformly over $\{-1,1\}$
- $\psi:$ distributed according to $P(\psi)$

and with

$$h(k,r,\mathbf{h},\boldsymbol{\ell},\tau,\psi) = \tau\psi + \frac{1}{2\beta}\log\left(\frac{\left\langle\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta(\sum_{\ell\leq k}\tau_\ell)(\psi+\tau/c)+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\right\rangle_{\tau_1\ldots\tau_k=\pm 1}}{\left\langle\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta(\sum_{\ell\leq k}\tau_\ell)(\psi-\tau/c)+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\right\rangle_{\tau_1\ldots\tau_k=\pm 1}}\right).$$

We approximate $W(h)$ by the empirical field frequencies computed from a large number (i.e. a population) of fields, which are made to evolve by repeated numerical iteration of a stochastic map. We start by initialising S fields $h_s \in \mathbb{R}$, with $s = 1, \ldots, S$, randomly with uniform probabilities over the interval $[-h_{\max}, h_{\max}]$. Their empirical distribution then represents the zero-step approximation $W_0(h)$ of $W(h)$. We then evolve the fields stochastically via the following Markovian process, giving at each step n an empirical distribution $W_n(h)$ which as n increases given an increasingly precise approximation of the invariant measure $W(h)$:

- choose randomly the variables $k, r, \boldsymbol{\ell}, \tau, \psi$ according to their (known) probability distributions
- choose randomly r fields $\mathbf{h} = h_1, \ldots, h_r$ from the S fields available, i.e. draw r fields from the probability distribution $W_{n-1}(h)$ of the previous step
- compute $h(k, r, \mathbf{h}, \boldsymbol{\ell}, \tau, \psi)$
- choose randomly one field from the set of the S available, and set its value to $h(k, r, \mathbf{h}, \boldsymbol{\ell}, \tau, \psi)$

We iterate the procedure until convergence, checking every $\mathcal{O}(S^2)$ steps the distance between different $W_n(h)$, and speed up the computation of $h(k, r, \mathbf{h}, \boldsymbol{\ell}, \tau, \psi)$ by rewriting it as

$$\begin{aligned} h(k,r,\mathbf{h},\boldsymbol{\ell},\tau,\psi) &= \tau\psi + \frac{1}{2\beta}\log\left(\frac{\int \mathrm{D}z\ \left\langle \mathrm{e}^{z\sqrt{\beta/c}\sum_{\ell\leq k}\tau_\ell+\beta(\sum_{\ell\leq k}\tau_\ell)(\psi+\tau/c)+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\right\rangle_{\tau_1\ldots\tau_k=\pm1}}{\int \mathrm{D}z\ \left\langle \mathrm{e}^{z\sqrt{\beta/c}\sum_{\ell\leq k}\tau_\ell+\beta(\sum_{\ell\leq k}\tau_\ell)(\psi-\tau/c)+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\right\rangle_{\tau_1\ldots\tau_k=\pm1}}\right) \\ &= \tau\psi + \frac{1}{2\beta}\log\left(\frac{\int \mathrm{D}z\ \prod_{\ell\leq k}\cosh[z\sqrt{\beta/c}+\beta(\psi+\tau/c)+\beta\sum_{s\leq r}h_s\delta_{\ell\ell_s}]}{\int \mathrm{D}z\ \prod_{\ell\leq k}\cosh[z\sqrt{\beta/c}+\beta(\psi-\tau/c)+\beta\sum_{s\leq r}h_s\delta_{\ell\ell_s}]}\right) \end{aligned} \tag{2.146}$$

which requires Gaussian integration instead of the average over $\{\tau_1, \ldots, \tau_k\}$. Having computed $W(h)$, we can build $P(M|\psi)$ using equation (2.143). The latter can be rewritten as

$$\begin{aligned} P(M|\psi) &= \left\langle\!\!\left\langle \frac{\left\langle \delta_{M,\sum_{\ell\leq k}\tau_\ell}\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta\psi\sum_{\ell\leq k}\tau_\ell+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\right\rangle_{\tau_1\ldots\tau_k=\pm1}}{\left\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta\psi\sum_{\ell\leq k}\tau_\ell+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\right\rangle_{\tau_1\ldots\tau_k=\pm1}}\right\rangle\!\!\right\rangle_{k,r,\mathbf{h},\boldsymbol{\ell},\psi} \\ &= \left\langle\!\!\left\langle \frac{\left\langle \delta_{M,\sum_{\ell\leq k}\tau_\ell}\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta\psi\sum_{\ell\leq k}\tau_\ell+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\right\rangle_{\tau_1\ldots\tau_k=\pm1}}{Z(k,r,\mathbf{h},\boldsymbol{\ell},\psi)}\right\rangle\!\!\right\rangle_{k,r,\mathbf{h},\boldsymbol{\ell},\psi} \end{aligned} \tag{2.147}$$

with $Z(\ldots) = \int \mathrm{D}z\ \prod_{\ell\leq k}\cosh[z\sqrt{\frac{\beta}{c}}+\beta(\psi-\tau/c)+\beta\sum_{s\leq r}h_s\delta_{\ell\ell_s}]$ as determined as in (2.146). Hence we can carry out the ensemble average over the parameters $\{\boldsymbol{\tau}, k, r, \mathbf{h}, \boldsymbol{\ell}, \psi\}$ in this last expression as an arithmetic average over a large number L of samples drawn from their joint distribution (we choose $L = \mathcal{O}(10^7)$). The distribution (2.144) is handled in the same way, and can be rewritten as

$$P(M|k,\psi) = \left\langle\!\!\left\langle \frac{\left\langle \delta_{M,\sum_{\ell\leq k}\tau_\ell} \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta\psi\sum_{\ell\leq k}\tau_\ell+\beta\sum_{s\leq r}h_s\tau_{\ell_s}} \right\rangle_{\tau_1\ldots\tau_k=\pm1}}{Z(k,r,\mathbf{h},\boldsymbol{\ell},\psi)} \right\rangle\!\!\right\rangle_{r,\mathbf{h},\boldsymbol{\ell},\psi}, \quad (2.148)$$

i.e. upon simply omitting the averaging over k.

In the interest of transparancy and an intuitive understanding, it helps to identify the physical meaning of the random variables involved in the above stochastic process. Given a subsystem of k spins linked to a particular cytokine pattern (say pattern $\mu = 1$, without loss of generality), we may ask how many other patterns $\mu \neq 1$ interfere with it. This number is the cardinality of the set

$$R = \left\{\xi_i^\mu,\ i{=}1,\ldots,N;\ \mu{=}2,\ldots,\alpha N:\ \xi_i^\mu\xi_i^1 \neq 0\right\}. \quad (2.149)$$

With each of the k spins (labelled by i, with $\xi_i^1 \neq 0$) correspond $\alpha N-1$ cytokine variables ξ_i^μ with $\mu > 1$. Hence we have, for a set of k spins, $k(\alpha N-1)$ independent possibilities to generate interfering cytokine signals, each nonzero with probability c/N. Thus, for $N \to \infty$ the number of possible interferences is a Poissonian random variable with mean αck, which is recognised to be the variable r. For each value of r we next ask on *which* of the k spins each interference acts, i.e. which are the r indices i such that $\xi_i^\mu\xi_i^1 \neq 0$ for some $\mu > 1$. Each i refers to one of the k spins selected, so we can describe this situation by r random variables ℓ_s, with $s = 1,\ldots,r$, each distributed uniformly in $\{1,\ldots,k\}$, with are recognised as the vector $\boldsymbol{\ell}$. The parameters k, r and $\boldsymbol{\ell}$ considered so far depend only on the (quenched) structure of the B-H network. By conditioning on these random variables we can write

$$\begin{aligned} P(M|\psi) &= \sum_{k=0}^{\infty}\mathrm{e}^{-c}\frac{c^k}{k!}\sum_{r=0}^{\infty}\mathrm{e}^{-\alpha ck}\frac{(\alpha ck)^r}{r!}\sum_{\ell_1,\ldots,\ell_r=1}^{k}k^{-r}P(M|k,r,\boldsymbol{\ell},\psi) \\ &= \left\langle\!\!\left\langle \sum_{\boldsymbol{\sigma}}\delta_{M,\sum_{\ell=1}^k\xi_\ell^1\sigma_\ell}Z^{-1}(k,r,\boldsymbol{\ell},\psi)\mathrm{e}^{-\beta H(\boldsymbol{\sigma}|k,r,\boldsymbol{\ell},\psi)} \right\rangle\!\!\right\rangle_{k,r,\boldsymbol{\ell}}. \end{aligned} \quad (2.150)$$

Inside the brackets we have the overlap M of a single pattern ($\mu = 1$) with k non-null entries, whose correlation with the other patterns is specified uniquely by the parameters $(k, r, \boldsymbol{\ell})$. We can write the effective Hamiltonian governing this k-spin subsystem by isolating in the Hamiltonian (2.58) $\mu = 1$ contribution:

$$H_{\text{eff}}(\boldsymbol{\sigma}) = -M_1^2(\boldsymbol{\sigma})/2c - \psi M_1(\boldsymbol{\sigma}) - \sum_{i=1}^{k}\sigma_i\sum_{\mu>2}\xi_i^\mu(M_\mu(\boldsymbol{\sigma})/c + \psi_\mu). \quad (2.151)$$

Upon transforming $\tau_\ell = \xi^1_\ell \sigma_\ell$, and defining $h^\mu_\ell(\boldsymbol{\tau}) = \xi^1_\ell \xi^\mu_\ell (M_\mu/c + \psi_\mu)$, and using the meaning of the parameters r and ℓ_s, we arrive at a description involving r non zero fields $h_s(\boldsymbol{\tau})$, each acting on a spin ℓ_s:

$$H_{\text{eff}}(\tau_1, \ldots, \tau_k) = -(\sum_{\ell \leq k} \tau_\ell)^2/2c - \psi \sum_{\ell \leq k} \tau_\ell - \sum_{s \leq r} h_s(\boldsymbol{\tau}) \tau_{\ell_s}. \tag{2.152}$$

If we then regard each field $h_s(\boldsymbol{\tau})$ as a independent random field (conditional on $(k, r, \boldsymbol{\ell})$), with probability distribution $W(h_s)$, we arrive at

$$P(M) = \left\langle\!\!\left\langle \int \mathrm{d}\mathbf{h}\; W(\mathbf{h}) \left\langle \delta_{M, \sum_{\ell=1}^k \tau_\ell} \frac{\mathrm{e}^{\beta(\sum_{\ell \leq k} \tau_\ell)^2/2c + \beta\psi \sum_{\ell \leq k} \tau_\ell + \beta \sum_{s \leq r} h_s \tau_{\ell_s}}}{Z(k, r, \boldsymbol{\ell}, \psi)} \right\rangle_{\boldsymbol{\tau}} \right\rangle\!\!\right\rangle_{k, r, \boldsymbol{\ell}}. \tag{2.153}$$

This is exactly equation (2.143) obtained within the RS ansatz. Hence the parameters $\mathbf{h}$ in (2.145) represent the effective fields induced by the interferences among the patterns. The only difference between the rigorous RS derivation and the above heuristic one is that in the former we effectively find $W(\mathbf{h}) = \prod_{s \leq r} W(h_s)$, i.e. the random fields are independent. This may not always be the case: if we recall the definition of the r effective fields, viz. $h^\mu_\ell(\boldsymbol{\tau}) = \xi^1_\ell \xi^\mu_\ell (M_\mu/c + \psi_\mu)$, we see that as soon as different patterns have more then one spin in common, their interference fields will not be independent. One therefore expects that the RS equation is no longer exact if the bi-partite network is not-tree like but contains loops. Finally we note that in the absence of external fields, the effective fields take values in $\mathbb{Z}/c$, so that $W(h)$ becomes a superposition of delta functions, consistently with numerical results in Fig. 2.6. This allows in principle a rewriting of the self-consistency equation (2.140) in terms of the amplitudes of such superposition, which are scalar parameters rather than distribution and may be easier to find numerically.

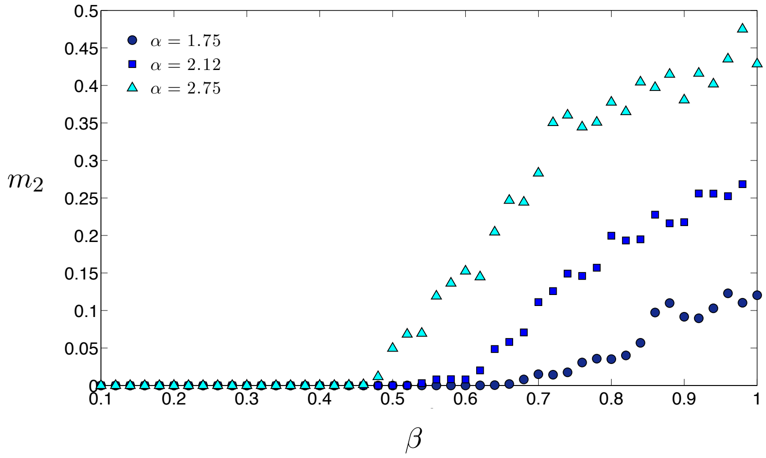

Figure 2.5: Widths (variances) $m_2 = \int \mathrm{d}h\ W(h)h^2$ of the distribution of interference field, shown as markers versus the inverse temperature β for different values of α. In all cases $c = 1$. The values of m_2 are calculated from the population dynamics solution of (2.145), and are (modulo finite size fluctuations in population dynamics algorithm) in excellent agreement with (2.141). The latter predicts that for the α-values considered and for $c = 1$ the widths m_2 should become nonzero at: $\beta_c = 0.6634$ (for $\alpha = 1.75$), $\beta_c = 0.5639$ (for $\alpha = 2.12$), and $\beta_c = 0.4707$ (for $\alpha = 2.75$).

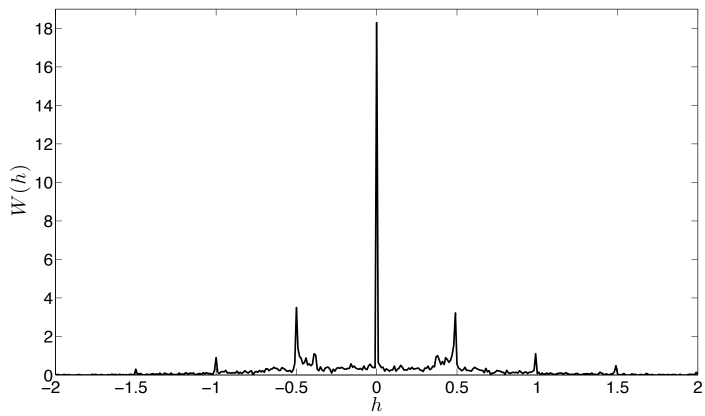

Figure 2.6: The interference field distribution $W(h)$ below the critical temperature and in the absence of external fields, as calculated (approximately) via the population dynamics algorithm, for $c = 2$, $\alpha = 2$ and $\beta = 6.2$. Note that the support of $W(h)$ is $\mathbb{Z}/c$. One indeed observes the weight of $W(h)$ being concentrated on these points; due to the finite population size in the algorithm (here $S = 5000$) one finds small nonzero values for $h \notin \mathbb{Z}/c$ due to finite size fluctuations.

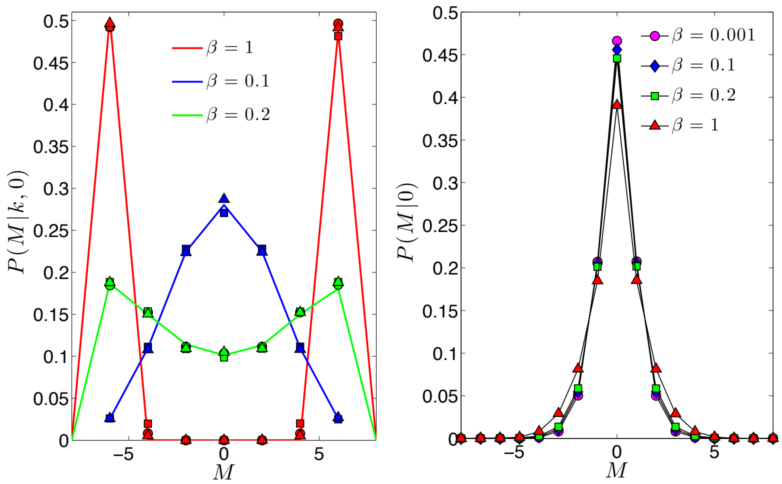

Figure 2.7: Left: degree-conditioned conditioned overlap distribution $P(M|k,0)$ in the under-percolated regime, for $k = 6$, $c = 1$, and different β values (see legend), without external fields. Solid lines: theoretical predictions. Markers: results of measuring the overlap statistics in Monte-Carlo simulations of the spin system with Hamiltonian (2.53), with $N = 3.10^4$ spins. Different symbols represent different values of α, namely $\alpha = 0.005$ (bullets), $\alpha = 0.008$ (squares) and $\alpha = 0.011$ (triangles). The theory predicts that here $P(M|k,0)$ is independent of α, which we find confirmed. Right panel: overlap distribution $P(M|0)$ at zero field in the under-percolated regime, for $k = 6$, $c = 1$ and $\alpha = 0.5$, and different temperatures (see legend). Note that $M \in \mathbb{Z}$, so line segments are only guides to the eye.

2.3.8 Critical line, overlap distributions, and interference field distribution

First we use the population dynamics algorithm to validate the location of the critical line (2.141). To do so we keep α fixed and compute $W(h)$ for different values of the inverse temperature β. From the solution we compute $m_2 = \int dh\ h^2 W(h)$, and determine for which β-value it becomes nonzero (starting from the high temperature phase), i.e. where patterns cross-talk sets in. The result is shown in Figure 2.5, which reveals excellent agreement between the predicted bifurcation temperatures (2.141) and those obtained from population dynamics. We also see that there is no evidence for discontinuous transitions. In Figure 2.4 we plotted the bifurcation temperatures obtained via population dynamics versus αc^2 (markers), together with the full transition lines predicted by (2.141) and again see excellent agreement between the two.

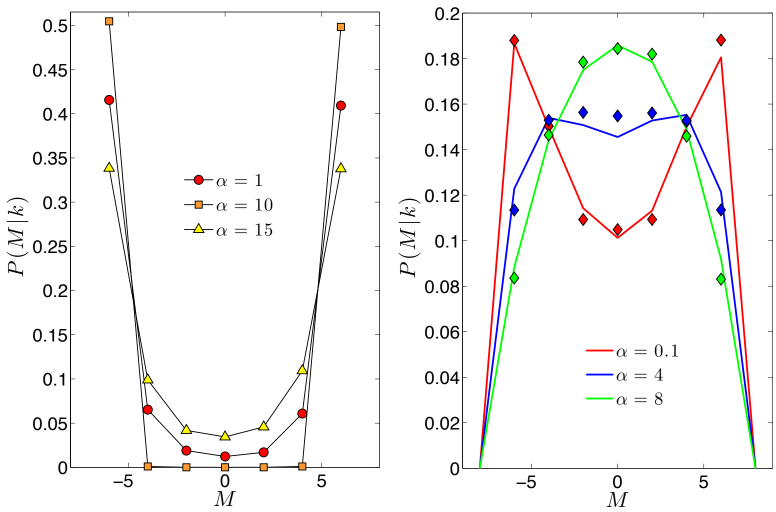

Figure 2.8: Left panel: overlap distribution $P(M|k)$ at zero field in the over-percolated regime, for $k = 6$, $c = 1$ and $\beta = 0.8$, and different α values (see legend). Right: the same distribution at $\beta = 0.8$, but now for $k = 6$, $c = 3$, and different α values (see legend). Note the different vertical axis scales of the two panels. Solid lines: theoretical predictions, calculated via the population dynamics method. Markers: results of measuring the overlap statistics in Monte-Carlo simulations of the spin system with Hamiltonian (2.53), with $N = 3.10^4$. The theory predicts that here $P(M|k, 0)$ is no longer independent of α, which we find confirmed. Note that $M \in \mathbb{Z}$, so line segments are only guides to the eye.

In the under-percolated regime $\alpha c^2 < 1$, there is no possbility of a phase transition and the only solution of (2.140) is $W(h) = \delta(h)$. Both equations (2.143,2.144) then lose their dependence on α, and after some simple manipulations we recover our earlier results (2.120,2.121). In Figure 2.7 we test our predictions for the overlap statistics against the results of numerical (Monte-Carlo) simulations of the spin process defined by Hamiltonian (2.53), in the absence of external fields. There is excellent agreement between theory and numerical experiment. Comparison of $P(M|k, 0)$ to $P(M|0)$ shows that the former changes shape as the inverse temperature β is increased from zero, from a single peak at $M = 0$ to two symmetric peaks, showing that the system behaviour at high versus low noise levels is very different. In contrast, $P(M|0)$ has always a maximum in $M = 0$, due to the Poissonian distribution of k, and does not capture the two different behaviours. Hence $P(M|k, 0)$ is the most useful quantifier of retrieval behavior, which from now on we will simply denote in the absence of external fields as $P(M|k)$.

When $\alpha c^2 > 1$, and below the critical line defined by equation (2.141), the solution of equation (2.140) in the absence of external fields will exhibit $W(h) \neq \delta(h)$, see Figure 2.6. As a consequence, the effective Boltzmann factor governing the behavior of a set of k spins, linked to a single pattern, acquires a term $\beta \sum_{s \leq r} h_s \tau_{\ell_s}$ (see equation (2.148)). This term means that each subsystem is no longer isolated as in the underpercolated regime, but feels the interference due to the other patterns in the form of effective random fields. Numerical results for $P(M|k)$ in the overpercolated regime, including comparisons between population dynamics calculations and measurements taken in numerical similations (involving spin systems with $N = 3.10^4$) are shown in Figure 2.8. Again we observe excellent agreement. Moreover, we see that in the regime of patterns cross talks the system's signalling preformance degrades only gracefully; provided α is not yet too large, the overlap distribution maintains its bimodal form.

2.4 High storage regime in a finite connectivity: cavity approach.

In the previous section, diluted associative networks have been studied via replica analysis, for pattern-independent dilution. This setting only accounts for special structures of the underlying bipartite graph, with all degrees in each set drawn from the same Poisson distribution. Here we adapt cavity (i.e. belief-propagation) methods to analyze the more general scenario where degrees in the two sets of spins have arbitrary distributions, thus allowing for a significantly broader range of bipartite structures, and corresponding marginalized associative networks.

We consider an equilibrated system of N binary neurons $\sigma_i = \pm 1$ at temperature (fast noise) $T = 1/\beta$, with Hamiltonian

$$H(\boldsymbol{\sigma}|\boldsymbol{\xi}) = -\frac{1}{2} \sum_{i,j} \sum_{\mu}^{P} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j,$$

where pattern entries $\{\xi_i^\mu\}$ are sparse (i.e. the number of non-zero entries of a pattern is finite). We can then use a factor graph representation of the Boltzmann weight as $\prod_\mu F_\mu$, with factors

$$F_\mu = e^{(\beta/2) \sum_{i,j \in O(\mu)} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j} = \langle e^{z \sum_{i \in O(\mu)} \xi_i^\mu \sigma_i} \rangle_z, \tag{2.154}$$

where $O(\mu) = \{i : \xi_i^\mu \neq 0\}$ and z is a zero mean Gaussian variable with variance β. We denote by $e_\mu = |O(\mu)|$ the degree of a pattern μ and by $d_i = |N(i)|$ the degree of a neuron i, with $N(i) = \{\mu : \xi_i^\mu \neq 0\}$. We consider random graph ensembles with given degree distributions $P(d)$ and $P(e)$, and nonzero ξ's independently and identically distributed (i.i.d.). Conservation of links demands $N\langle d\rangle = P\langle e\rangle$ where averages are taken over $P(d)$ and $P(e)$. The *message* from factor μ to node j is the cavity distribution $P_\mu(\sigma_j)$ of σ_j when this is coupled to factor μ only, which we can parametrize by an effective field $\psi_{\mu\to j}$. The message from node j to factor μ is the cavity distribution $P_{\backslash\mu}(\sigma_j)$ of σ_j when coupled to all factors except μ, which we can parametrize by the field $\phi_{j\to\mu}$. The cavity equations are then [80]

$$P_\mu(\sigma_j) = \mathrm{Tr}_{\{\sigma_k\}} F_\mu(\sigma_j, \{\sigma_k\}) \prod_{k\in O(\mu)\backslash j} P_{\backslash\mu}(\sigma_k), \tag{2.155}$$

$$P_{\backslash\nu}(\sigma_j) = \prod_{\mu\in N(j)\backslash\nu} P_\mu(\sigma_j), \tag{2.156}$$

Given the site factorization, conditional on z, of the factors (2.154), translating these to equations for the effective fields is straightforward:

$$\psi_{\mu\to j} = \tanh^{-1}\langle\sigma_j\rangle_\mu = \tag{2.157}$$

$$\tanh^{-1}\frac{\langle \sinh(z\xi_j^\mu)\prod_{k\in M(\mu)\backslash j}\cosh(\phi_{k\to\mu}+z\xi_k^\mu)\rangle_z}{\langle \cosh(z\xi_j^\mu)\prod_{k\in M(\mu)\backslash j}\cosh(\phi_{k\to\mu}+z\xi_k^\mu)\rangle_z},$$

$$\phi_{j\to\nu} = \sum_{\mu\in N(j)\backslash\nu}\psi_{\mu\to j}. \tag{2.158}$$

These equations, once iterated to convergence, are exact on tree graphs. They will also become exact on graphs sampled from our ensemble in the thermodynamic limit, because the sparsity of the ξ_i^μ makes the graphs locally tree-like, with typical loop lengths that diverge (logarithmically) with N [55, 78].

For large N, we can describe the solution of the cavity equations on any fixed graph – and hence also the quenched average over the graph ensemble and the nonzero pattern entries ξ_i^μ – in terms of the distribution of messages or fields, $W_\psi(\psi)$ and $W_\phi(\phi)$. Denoting by $\Psi(\{\phi_{k\to\mu}\}, \{\xi_k^\mu\}, \xi_j^\mu)$ the r.h.s. of (2.157), convergence of the cavity iterations then implies the self-consistency equation

$$W_\psi(\psi) = \sum_e \frac{eP(e)}{\langle e\rangle}\langle\delta\left(\psi - \Psi(\phi_1, ..., \phi_{e-1}, \xi^1, ..., \xi^e)\right)\rangle,$$

where the average is over i.i.d. values of the (nonzero) $\xi^1, ..., \xi^d$ and over i.i.d. $\phi_1, ..., \phi_{e-1}$

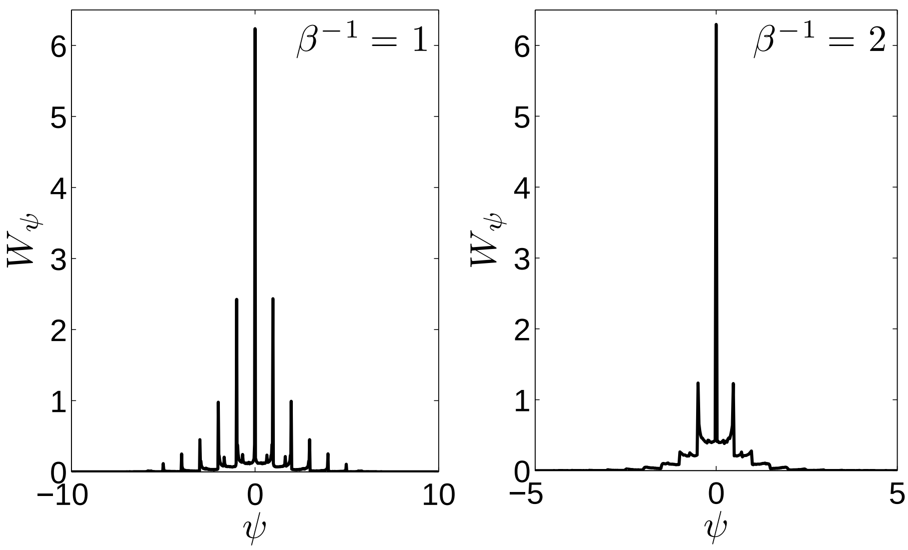

Figure 2.9: Histograms $W_\psi(\psi)$ of the field ψ for $\alpha = 8$, $c = 2$ and $\beta^{-1} = 1, 2$, as shown in figure.

drawn from $W_\phi(\phi)$, and similarly

$$W_\phi(\phi) = \sum_d \frac{dP(d)}{\langle d \rangle} \langle \delta \left(\phi - \sum_{\mu=1}^{d-1} \psi_\mu \right) \rangle,$$

where the average is over i.i.d. $\psi_1, ..., \psi_{d-1}$ drawn from $W_\psi(\psi)$. Field distributions can then be obtained numerically by population dynamics (PD) [80]. For symmetric ξ-distributions, a delta function at the origin for both W_ψ, W_ϕ is always a solution, and we find this to be stable at high temperatures. At low T, the ψ can become large (see Fig. 2.9), hence also the ϕ, and the spins σ_i, will typically be strongly polarized. The fields $\beta \xi_i^\mu \sum_{j \in O(\mu) \setminus i} \xi_j^\mu \sigma_j$ then fluctuate little, and the ψ as suitable averages of these fields cluster near multiples of β (for $\xi = \pm 1$).

Our main interest is in the retrieval properties, encoded in the fluctuating pattern overlaps $m_\mu = \sum_{i \in M(\mu)} \xi_i^\mu \sigma_i$. Since the joint distribution of the σ_i in $M(\mu)$ is $F_\mu(\{\sigma_i\}) \prod_{i \in M(\mu)} P_{\setminus\mu}(\sigma_i)$, the distribution of the pattern overlap m_μ is

$$\frac{\mathrm{Tr}_{\{\sigma_i\}} \left\langle \delta(m_\mu - m) \exp(\sum_{i \in M(\mu)} (\xi_i^\mu z + \phi_{i \to \mu}) \sigma_i) \right\rangle_z}{\mathrm{Tr}_{\{\sigma_i\}} \left\langle \exp(\sum_{i \in M(\mu)} (\xi_i^\mu z + \phi_{i \to \mu}) \sigma_i) \right\rangle_z}. \tag{2.159}$$

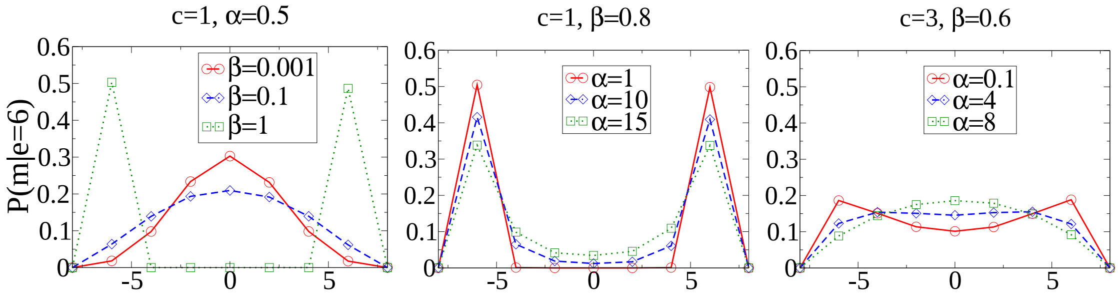

Figure 2.10: $P(m|e = 6)$ above (left) and crossing (middle and right) the critical line for different values of β and α, respectively. Full red (dashed blue and dotted green) curves in the middle and right panels refer to temperatures above (below) the critical line.

Defining this as $\mathcal{P}(m, \{\phi_{i\to\mu}\}, \{\xi_i^\mu\})$, in the graph ensemble we have

$$P(m) = \sum_e P(e) \left\langle \mathcal{P}(m, \phi_1, \ldots, \phi_e, \xi_1, \ldots, \xi_e) \right\rangle \ . \tag{2.160}$$

The average here can be read as $P(m|e)$, the overlap distribution for patterns with fixed degree e. Whenever $W_\phi(\phi) = \delta(\phi)$, $P(m|e)$ is the overlap distribution for an "effectively isolated" subsystem of size e: the neurons storing each pattern $\boldsymbol{\xi}^\mu$ can retrieve this independently of other patterns, even though the number of patterns is extensive. Retrieval within each group of neurons is strongest at low temperatures (see Fig. 2.10 left) as expected on general grounds. Once nonzero ϕ appear neuron groups are no longer independent: intuitively, cross-talk interference between patterns emerges.

2.4.1 Bifurcation

When the "parallel processor" solution with zero cavity fields ϕ becomes unstable, a bifurcation to a different stable solution occurs. Depending on the external parameters, this can be seen in the first or second moment of the field distribution. Expanding for small fields we get

$$\Psi(\{\phi_{k\to\mu}\}, \{\xi_k^\mu\}, \xi_j^\mu) \approx \sum_{k\in O(\mu)\setminus j} \phi_{k\to\mu}\, \Xi(\xi_k^\mu, \xi_j^\mu, \{\xi_l^\mu\}),$$

with coefficients $\Xi(\xi_k^\mu, \xi_j^\mu, \{\xi_l^\mu\})$ given by

$$\frac{\langle \sinh(z\xi_j^\mu) \sinh(z\xi_k^\mu) \prod_{l\in O(\mu)\setminus\{j,k\}} \cosh(z\xi_l^\mu)\rangle_z}{\langle \prod_{l\in O(\mu)} \cosh(z\xi_l^\mu)\rangle_z} \ .$$

The self-consistency relations for the field distributions W_ψ and W_ϕ then show that as long as the mean fields are small, they are related to leading order by

$$\begin{aligned} \langle\psi\rangle &= \langle\phi\rangle \sum_e P(e)\frac{e(e-1)}{\langle e\rangle}\langle\Xi(\xi_1,\ldots,\xi_e)\rangle, &(2.161)\\ \langle\phi\rangle &= B_d\langle\psi\rangle, &(2.162)\end{aligned}$$

where $B_d = \sum_d P(d)d(d-1)/\langle d\rangle$ is one of the two branching ratios of our locally tree-like graphs, the other being $B_e = \sum_e P(e)e(e-1)/\langle e\rangle$. If the means are zero then the onset of nonzero fields is detected by the variances, which are related to leading order by

$$\begin{aligned} \langle\psi^2\rangle &= \langle\phi^2\rangle \sum_e P(e)\frac{e(e-1)}{\langle e\rangle}\langle\Xi^2(\xi_1,\ldots,\xi_e)\rangle, &(2.163)\\ \langle\phi^2\rangle &= B_d\langle\psi^2\rangle. &(2.164)\end{aligned}$$

2.4.2 Symmetric pattern distributions

When the ξ are symmetrically distributed, then also the field distributions are always symmetric and there can be no instability from growing means; cf. (2.161). The bifurcation has to result from the growth of the variances, which from (2.164) occurs at $A = 1$ with

$$A = B_d \sum_e P(e)\frac{e(e-1)}{\langle e\rangle}\langle\Xi^2(\xi_1,\ldots,\xi_e)\rangle. \qquad (2.165)$$

This factorizes as $A = B_d A_e(\beta)$ with the dependence on the noise and the distribution of the e's contained in the second factor $A_e(\beta)$. For $\beta \to 0$ the variance of z goes to zero and $A_e(0) = 0$. For $\beta \to \infty$, the z-averages are dominated by large values of z where $\sinh^2(z) \approx \cosh^2(z)$, so $A_e(\infty) = B_e$. Hence there is no bifurcation when $B_d B_e < 1$, in agreement with the general bipartite tree percolation condition [86]. For the case $P(\xi_i^\mu = \pm 1) = c/(2N)$ considered in the previous section, the distributions of pattern degrees e and neuron degrees d are Poisson(c) and Poisson(αc), respectively, so $B_d = \alpha c$, $B_e = c$ and there is no bifurcation for $\alpha c^2 < 1$. The network acts as a parallel processor here for *any* T because the bipartite network consists of finite clusters of interacting spins in which there is no interference between different patterns. At higher connectivity, the critical line defined by $A = 1$ indicates the temperature above which this lack of interference persists even though the network now has a giant connected component. Fig. 2.11 (left) compares theory to PD results, where we locate the transition as the onset of nonzero second moments of the field distributions. The impact of the transition on the overlap probability distribution of a pattern with fixed e can be seen from the PD results in Fig. 2.10 (middle and right panels). Crossing the transition line, parallel retrieval is accomplished at low temperatures, but it degrades when α is increased (see shrinking peaks in the middle panel), or c is increased, eventually fading away for sufficiently large α and c (right panel).

One advantage of our present method is that we can easily investigate the parallel processing capabilities of a bipartite graph with arbitrary degrees $\{e_\mu\}$. Here we have a pattern-dependent dilution of the links $P(\boldsymbol{\xi}) \propto \prod_{i,\mu} P(\xi_i^\mu) \prod_\mu \delta_{e_\mu, \sum_i |\xi_i^\mu|}$ with

$$P(\xi_i^\mu) = \frac{e_\mu}{2N}(\delta_{\xi_i^\mu,1} + \delta_{\xi_i^\mu,-1}) + (1 - \frac{e_\mu}{N})\delta_{\xi_i^\mu,0} \tag{2.166}$$

leading to $P(d) = \text{Poisson}(\alpha\langle e\rangle)$ while $P(e) = P^{-1}\sum_\mu \delta_{e,e_\mu}$. If we keep the mean degree fixed at $\langle e\rangle = c$, the critical point for $\beta \to \infty$ is found at

$$B_d B_e = \alpha c(\langle e^2\rangle / c - 1) = \alpha\left[c(c-1) + \text{Var}(e)\right] = 1$$

while for large α one obtains for the critical line $\beta_c^{-1}(\alpha) \approx \sqrt{\alpha}\sqrt{c(c-1) + \text{Var}(e)}$. Similar results are obtained with soft constraints e_μ on the degrees, i.e. by dropping the delta function constraint in $P(\boldsymbol{\xi})$ before (2.166): one now finds $B_d B_e = \alpha(c^2 + \text{Var}(e))$ and $\beta_c^{-1}(\alpha) \approx \sqrt{\alpha}\sqrt{c^2 + \text{Var}(e)}$. In both cases, the region where parallel retrieval is obtained is larger for degree distributions with smaller variance; the optimal situation occurs when all patterns have exactly the same number c of non zero entries (Fig. 2.11, right): notably, scale free networks (best performing for information spreading [86]) are not optimal for information processing.

2.4.3 Non-symmetric pattern distributions

To introduce a degree of asymmetry $a \in [-1, +1]$ in the pattern distribution, we next take for the nonzero pattern entries $P(\xi_i^\mu = \pm 1) = (1 \pm a)/2$. Evaluating the ξ-average $\langle \Xi(\ldots)\rangle$ in (2.161), the condition for a transition to nonzero field means then becomes

$$1 = a^2 B_d \sum_e P(e) \frac{e(e-1)}{\langle e\rangle} \frac{\langle \sinh^2(z)\cosh^{e-2}(z)\rangle_z}{\langle \cosh^e(z)\rangle_z} \tag{2.167}$$

At zero temperature the bifurcation occurs when $B_d B_e = a^{-2}$; when a tends to zero the transition point goes to infinity and we retrieve the symmetric case. Beyond the bifurcation, non-centered field probability distributions (see Fig. 2.12) produce a non-zero global magnetization typical of ferromagnetic systems. One has to bear in mind, however, that even with a bias in the pattern entry distribution a bifurcation to growing field variances at zero

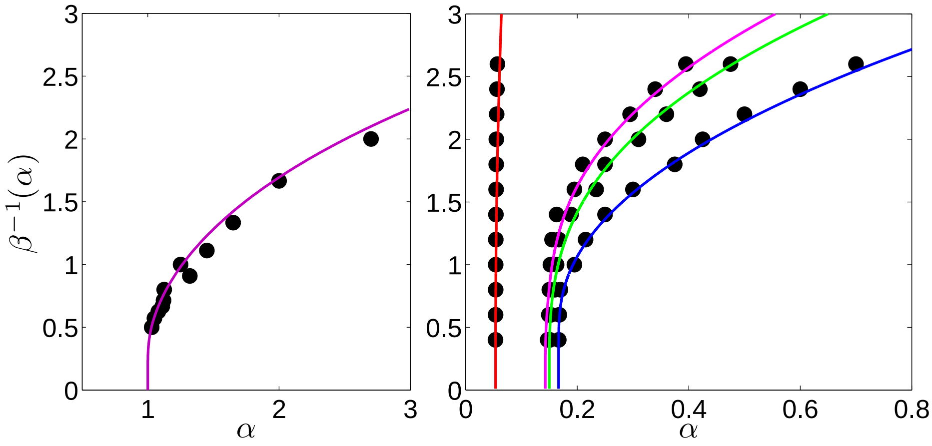

Figure 2.11: Transition lines (theory, with symbols from PD numerics) for different pattern degree distributions. Left: $e \sim \text{Poisson}(c{=}1)$. Right: changing $P(e)$ at constant $\langle e \rangle = 3$; $P(e) = \delta_{e,3}$ (blue); $P(e) = (\delta_{e,2} + \delta_{e,3} + \delta_{e,4})/3$ (green); $P(e) = (\delta_{e,2} + \delta_{e,4})/2$ (pink); $P(e)$ power law as in preferential attachment graphs, with $\langle e^2 \rangle = 21.66$ (orange).

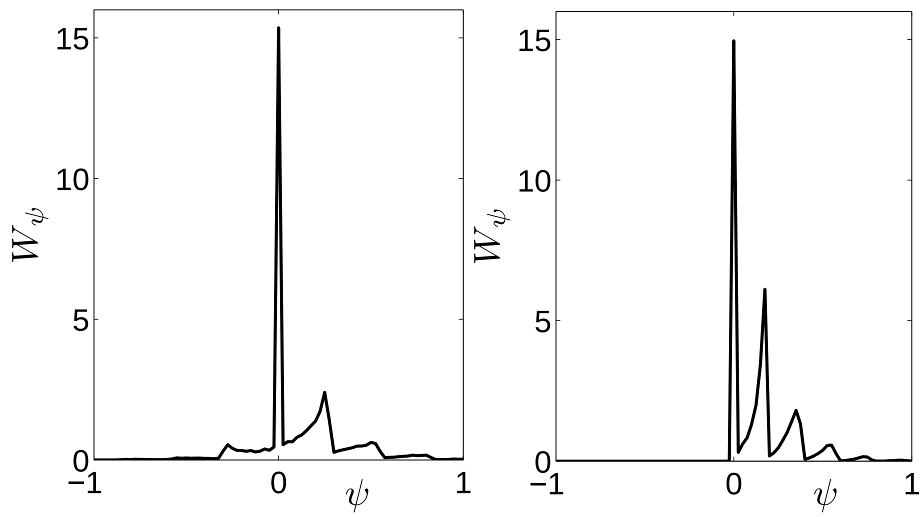

Figure 2.12: Histogram of the fields ψ in the ferromagnetic region, for $c = 1$, $\beta = 1$ and different levels of bias: $a = 0.9$ with $\alpha = 9$ (left) and $a = 1$ with $\alpha = 8$ (right). Field distributions are obtained by PD starting from positive fields, to break the gauge symmetry. For $a = 1$ (right) there are only positive fields as expected: when all patterns have positive entries there are no conflicting signals, even above the percolation threshold.

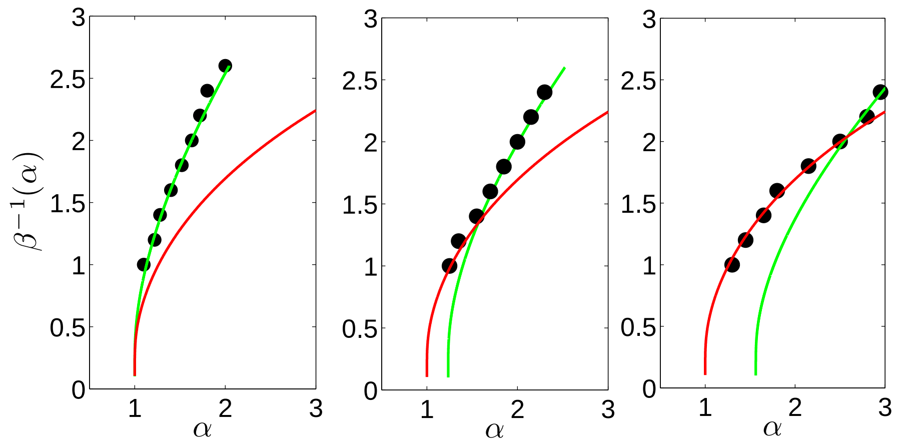

Figure 2.13: Transition lines to growing field means (theory, green) and variances (theory, red), showing a good match to numerical PD data (dots); here $c = 1$ and pattern bias $a = 1, 0.95, 0.9$ from left to right. The first line to be crossed from high $T = \beta^{-1}$ gives the physical transition.

means can occur; the physical bifurcation is the one occurring first on lowering T. Numerical evaluation shows that both bifurcation temperatures increase with α. For large α one can then resort to a low-β expansion: $\langle \sinh^2(z)\cosh^{e-2}(z)\rangle \approx \langle z^2\rangle = \beta$, $\langle\cosh^e(z)\rangle \approx 1$. This gives for the growing mean bifurcation condition $1 \approx B_d B_e \beta a^2$ while for the growing variance bifurcation one gets $1 \approx B_d B_e \beta^2$. For Poisson graphs $B_d B_e = \alpha c^2$, giving the transition lines $\beta_{c,1}^{-1}(\alpha) \approx c^2 a^2 \alpha$ and $\beta_{c,2}^{-1}(\alpha) \approx c\sqrt{\alpha}$ for large α. In the presence of a nonzero pattern bias a these cross at $\alpha = 1/(ca^2)$, with the bifurcation to growing means occurring first for larger α. The existence of this crossing is confirmed by numerical evaluation of (2.165) and (2.167) for finite α in Fig. 2.13.

Summarizing, we have developed a cavity/belief-propagation framework to analyze finitely connected bipartite spin glasses, with arbitrary structure (arbitrary pattern degree distributions) and an arbitrary degree of asymmetry in the link distribution, finding that:

-Extensive multitasking features appear quite naturally in these systems.

-A transition surface separates the region in (α, β, c)-space where the network is capable of parallel extensive retrieval, from the region where pattern interference affects the network performance as a parallel processor.

-Homogeneous degree distributions in the bipartite network favors parallel retrieval.

-A biased distribution of the sparse pattern entries can yield a macroscopic net magnetization and shrinks the region of parameter space where no pattern cross-talk occurs. However, in this ferromagnetic region, pattern cross-talk may result in a constructive interference between patterns, which does not disrupt the parallel retrieval performed by the network.

Chapter 3

Hierarchical Hopfield model

While the Hopfield 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 suitably regions of the tunable parameters, e.g. low noise level and not-too-high storage load), the diluted Hopfield models, as a generalization of this paradigm, appear as candidates mean-field multitasking networks able to perform spontaneously parallel retrieval, i.e. to retrieve more patterns at once (without falling into spurious states). While these two networks perform in a crucial different way (serial versus parallel), they share the same mean-field statistical mechanics approximation: each unit interacts with all the others it is linked to with the same strength, unaware of any underlying topology, namely independently of the actual pair-wise distance among the neurons themselves. This limitation has always been considered as something to remove as soon as mathematical improvements of available techniques would allow.

Infact, in the last decade, extensive research on complexity in networks has evidenced (among many results [87, 100]) the widespread of modular structures and the importance of quasi-independent communities in many research areas such as neuroscience [38, 72, 101], biochemistry [58] and genetics [44], just to cite a few. In particular, the modular, hierarchical architecture of cortical neural networks has nowadays been analyzed in depths [82], yet the beauty revealed by this investigation is not captured by the statistical mechanics of neural networks, nor standard ones (i.e. performing single pattern retrieval) [70, 15] neither diluted ones (performing multiple patterns retrieval) [4, 47], for the lacking of a proper definition of metric distance among neurons.

Far from Artificial Intelligence, but exactly to this task (i.e. bypassing mean field limitations), a renewal interest is nowadays raised for hierarchical models, namely models in which the closer the spins the stronger their links (Fig. 3.1). Starting from the pioneering Dyson work [53], where the hierarchical ferromagnet was introduced and its phase transition (splitting an ergodic region from a ferromagnetic one) rigorously proven, recently its extensions to spin-glasses have also been investigated [41, 71, 43, 18, 91, 48, 19]. Although an analytical solution is still not available, giant step forward toward a deep comprehension of the hierarchical statistical mechanics have been obtained [42, 40, 56, 76, 77, 83, 84, 85]. Here we investigate the retrieval capabilities of an Hopfield model embedded in a hiararchical structure. Following [10, 11, 12], we start studying the Dyson Hierarchical model (DHM) and we show that, as soon as ergodicity is broken, beyond the pure ferromagnetic state (largely discussed in the past, see e.g., [57, 32]), a number of metastable states suddenly appear and become stable in the thermodynamic limit. The emergence of such states stems from the weak ties connecting distant neurons, which, in the thermodynamic limit, effectively get split into detached modules. As a result, if the latter are initialized with opposite magnetizations, they remain stable.

This is a crucial point because, once implemented the Hebbian prescription [70, 15] to account for multiple pattern storage, it allows proving that the system not only is able to retrieve any single pattern at a time as a standard Hopfield model, but its communities can perform autonomously, hence making the simultaneous retrieval of multiple patterns feasible too. We stress that this feature is essentially due to the notion of metric the system is endowed with, differently from the multiple retrieval performed by the mean-field multitasking networks which require blank pattern entries [4, 3, 5, 47]. Therefore, the hierarchical neural network is able to perform both as a serial processor and as a multitasking processor.

3.1 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 $\sigma_i = \pm 1$, with $i = 1, ..., 2^{k+1}$ embedded in a hierarchical topology. The Hamiltonian capturing the model is recursively introduced by the following

Definition 3.1.1. *The Hamiltonian of Dyson's Hierarchical Model (DHM) is defined by*

$$H_{k+1}(\vec{\sigma}|J,\rho) = H_k(\vec{\sigma_1}) + H_k(\vec{\sigma_2}) - \frac{J}{2^{2\rho(k+1)}} \sum_{i<j=1}^{2^{k+1}} \sigma_i \sigma_j, \tag{3.1}$$

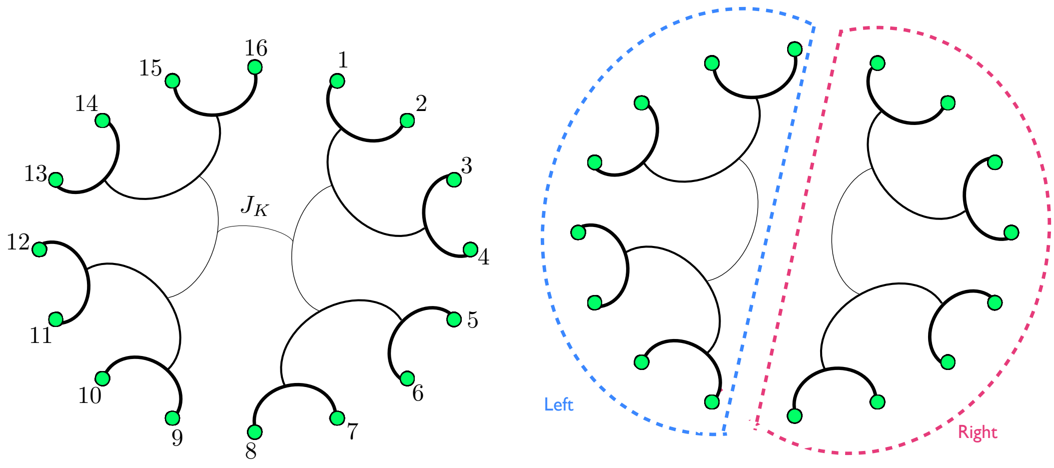

Figure 3.1: Schematic representation of the hierarchical topology where the associative network insists. Green spots represent Ising neurons ($N = 16$ in this shapshot) while links are drawn with different thickness mimicking various interaction strengths: The thicker the line, the stronger the link.

where $J > 0$ *and* $\rho \in (1/2, 1)$ *are numbers tuning the interaction strength. Clearly* $\vec{\sigma_1} \equiv \{\sigma_i\}_{1 \leq i \leq 2^k}$, $\vec{\sigma_2} \equiv \{\sigma_j\}_{2^k+1 \leq j \leq 2^{k+1}}$ *and* $H_0[\sigma] = 0$.

Thus, in this model, ρ triggers the decay of the interaction with the distance among spins, while J uniformly rules the overall intensity of the couplings. Note that this model is explicitly a non-mean-field model as the distance $d_{i,j}$ between two spins i, j ranges in 0 and k (see Fig.3.1). Indeed, it is possible to re-write the Hamiltonian (3.1) in terms of the $d_{i,j}$ as

$$H_k[\{\sigma_1 ... \sigma_{2^k}\}] = -\sum_{i<j} \sigma_i \sigma_j J_{ij} \tag{3.2}$$

$$J_{ij} = \sum_{l=d_{i,j}}^{k} \left(\frac{J}{2^{2\rho l}} \right) = J(d_{ij}, k, \rho, J) = J \frac{4^{\rho - d_{i,j}\rho} - 4^{-k\rho}}{4^{\rho} - 1}. \tag{3.3}$$

Once the Hamiltonian is given (and we refer mainly to the form (3.1)), it is possible to introduce the partition function $Z_{k+1}(\beta, J, \rho)$ at finite volume $k + 1$ as

$$Z_{k+1}(\beta, J, \rho) = \sum_{\vec{\sigma}}^{2^{k+1}} \exp\left[-\beta H_{k+1}(\vec{\sigma}|J, \rho)\right], \tag{3.4}$$

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

$$f_{k+1}(\beta, J, \rho) = \frac{1}{2^{k+1}} \log \sum_{\vec{\sigma}} \exp \left[-\beta H_{k+1}(\vec{\sigma}) + h \sum_{i=1}^{2^{k+1}} \sigma_i \right]. \tag{3.5}$$

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

$$f(\beta, J, \rho) = \lim_{k \to \infty} f_{k+1}(\beta, J, \rho),$$

in terms of suitably introduced magnetizations. To this task we introduce the global magnetization m, defined as the limit $m = \lim_{k\to\infty} m_{k+1}$ where

$$m_{k+1} = \frac{1}{2^{k+1}} \sum_{i}^{2^{k+1}} \sigma_i, \tag{3.6}$$

and, recursively and with a little abuse of notation, the k magnetizations $m_a, ..., m_k$ level by level (over k levels and starting to defined them from the largest bulk), as the same $k \to \infty$ limit of the following quantities (we write explicitly only the two upper magnetizations related to the two main clusters the system reduces to whenever $J_K \to 0$):

$$m_{left} = \frac{1}{2^k} \sum_{i=1}^{2^k} \sigma_i, \quad m_{right} = \frac{1}{2^k} \sum_{i=2^k+1}^{2^{k+1}} \sigma_i. \tag{3.7}$$

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

$$\langle m_{k+1}(\beta, J, \rho) \rangle = \frac{\sum_{\vec{\sigma}} m_{k+1} e^{-\beta H_{k+1}(\vec{\sigma}|J,\rho)}}{Z_{k+1}(\beta, J, \rho)}, \tag{3.8}$$

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

3.1.1 The thermodynamic limit

The existence of the thermodynamic limit for the free energy of the DHM has been achieved a long time ago by Gallavotti and Miracle-Sole [57]. We exploit here a different interpolating scheme with the pedagogical aim of highlighting the technique more than the result itself as it will then be used to prove the 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 3.1.2. *The thermodynamic limit of the DHM free energy does exist and we call*

$$\lim_{k \to \infty} f_{k+1}(\beta, J, \rho) = f(\beta, J, \rho).$$

Proof. To prove this statement let us introduce a real scalar parameter $t \in [0,1]$ and the following interpolating function

$$\Phi_{k+1,t}(\beta) = \frac{1}{2^{k+1}} \log \sum_{\vec{\sigma}} \exp(\beta(-H_k(\vec{\sigma_1}) - H_k(\vec{\sigma_2}) + \frac{tJ}{2} 2^{(k+1)} 2^{(k+1)(1-2\rho)} m_{k+1}^2(\vec{\sigma})), \quad (3.9)$$

with $m_{k+1} = \frac{1}{2^{k+1}} \sum_{l=1}^{2^{k+1}} \sigma_l$, such that

$$\Phi_{k+1,1} = f_{k+1}, \quad (3.10)$$

$$\Phi_{k+1,0} = f_k, \quad (3.11)$$

and

$$0 \leq \frac{d\Phi_{k+1,t}}{dt} = \left\langle \beta \frac{1}{2^{k+1}} \frac{2^{(k+1)} 2^{(k+1)(1-2\rho)} J}{2} m_{k+1}^2(\vec{\sigma}) \right\rangle_t \leq \frac{\beta J 2^{(k+1)(1-2\rho)}}{2}. \quad (3.12)$$

Since

$$\Phi_{k+1,1}(h) = \Phi_{k+1,0}(h) + \int_0^1 \frac{d\Phi_{k+1,t}}{dt} dt,$$

$f_{k+1} \geq f_k$ (the sequence is non-decreasing), thus

$$f_{k+1}(\beta, J, \rho) \leq f_k(\beta, J, \rho) + \frac{\beta J}{2} 2^{(k+1)(1-2\rho)}. \quad (3.13)$$

Iterating this argument over the levels we obtain

$$f_{k+1}(\beta, J, \rho) \leq f_0(\beta, J, \rho) + \frac{\beta J}{2} \sum_{l=1}^{k+1} 2^{l(1-2\rho)}. \quad (3.14)$$

In the limit of $k \to \infty$

$$f \leq f_0 + \frac{\beta J}{2} \sum_{l=1}^{\infty} 2^{l(1-2\rho)}. \quad (3.15)$$

The series on the right of the above inequality converges, since $\rho > \frac{1}{2}$, hence

$$f(\beta, J, \rho) \leq f_0(\beta, J, \rho) + \frac{\beta J}{2} \frac{1}{1 - 2^{(2\rho-1)}}. \quad (3.16)$$

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

□

3.1.2 The mean-field scenario

In this Section we investigate the equilibrium states of the Dyson model at the mean-field level, in particular we check the existence and the stability of two test-states: the (standard) ferromagnetic state (with all the spin aligned, hence $m_{left} = m_{right}$) and the simplest mixed state, namely a state where all the left spins (that is the first $1, ..., 2^k$ spins) are aligned each other and opposite to the right spins (that is the remaining $2^k + 1, ..., 2^{k+1}$ spins), which -in turn- are aligned each other too (hence $m_{left} = -m_{right}$).

Definition 3.1.3. *Once considered a real scalar parameter $t \in [0,1]$, we introduce the following interpolating Hamiltonian*

$$H_{k+1,t}(\vec{\sigma}) = -\frac{Jt}{2^{2\rho(k+1)}} \sum_{i>j=1}^{2^{k+1}} \sigma_i \sigma_j - (1-t) m J 2^{(k+1)(1-2\rho)} \sum_{i=1}^{2^{k+1}} \sigma_i + H_k(\vec{\sigma_1}) + H_k(\vec{\sigma_2}), \quad (3.17)$$

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 simply by adding to the Hamiltonian a term $\propto h \sum_i^{2^{k+1}} \sigma_i$, with $h \in \mathbb{R}$.

This prescription allows defining an extended partition function as

$$Z_{k+1,t}(h, \beta, J, \rho) = \sum_{\vec{\sigma}} \exp\{-\beta [H_{k+1,t}(\vec{\sigma}) + h \sum_{i=1}^{2^{k+1}} \sigma_i]\}, \quad (3.18)$$

where the subscript t stresses its interpolative nature, and, analogously,

$$\Phi_{k+1,t}(h, \beta, J, \rho) = \frac{1}{2^{k+1}} \log Z_{k+1,t}(h, \beta, J, \rho). \quad (3.19)$$

Since

$$\Phi_{k+1,0}(h, \beta, J, \rho) = \Phi_{k,1}(h + mJ2^{(k+1)(1-2\rho)}, \beta, J, \rho), \quad (3.20)$$

as shown in [40], (discarding the dependence of Φ by β, J, ρ for simplicity) through a long but straightforward calculation, we arrive to

$$\begin{aligned} \Phi_{k+1,1}(h) &= \Phi_{k+1,0}(h) - \frac{\beta J}{2}(2^{(k+1)(1-2\rho)} m^2 + 2^{-2(k+1)\rho}) + \frac{\beta J}{2} 2^{(k+1)(1-2\rho)} \left\langle (m_{k+1}(\vec{\sigma}) - m)^2 \right\rangle_t \\ &\geq \Phi_{k,1}(h + Jm2^{(k+1)(1-2\rho)}) - \frac{\beta J}{2}(2^{(k+1)(1-2\rho)} m^2 + 2^{-2(k+1)\rho}). \end{aligned} \quad (3.21)$$

Note that, in the last passage, we neglected level by level the source of order parameter's fluctuations $\left\langle (m_{k+1}(\vec{\sigma}) - m)^2 \right\rangle_t$ -which is positive definite- thus we obtained a bound for the free energy.

For the seek of simplicity we extended the meaning of the brackets to account also for the interpolating structure coded in the Boltzmannfaktor of eq.(3.18), by adding to them a subscript t, namely $\langle\cdot\rangle \rightarrow \langle\cdot\rangle_t$.

In order to start investigating non-standard stabilities, note further that $\Phi_{k+1,0}(h) = \Phi_{k,1}(h + mJ2^{(k+1)(1-2\rho)})$ but in principle we can have also two different contributions from the two groups of 2^k spins (*left* and *right*) thus we should write more generally

$$\Phi_{k+1,0}(h) = \frac{1}{2}\left[\Phi^1_{k,1}(h + mJ2^{(k+1)(1-2\rho)}) + \Phi^2_{k,1}(h + mJ2^{(k+1)(1-2\rho)})\right]. \tag{3.22}$$

Now let us assume the Amit perspective [15] and suppose that these two subsystems have different magnetizazions $m_{left} = m_1$ and $m_{right} = m_2$ (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

$$\begin{aligned}
\Phi_{k+1,1}(h) &\geq \frac{1}{2}\Phi_{0,1}\left\{h + J\left[\sum_{l=1}^{k} 2^{l(1-2\rho)}m_1 + 2^{(k+1)(1-2\rho)}m\right]\right\} \\
&+ \frac{1}{2}\Phi_{0,1}\left\{h + J\left[\sum_{l=1}^{k} 2^{l(1-2\rho)}m_2 + 2^{(k+1)(1-2\rho)}m\right]\right\} \\
&- \frac{\beta J}{2}\left[2^{(k+1)(1-2\rho)}m^2 + \sum_{l=1}^{k+1} 2^{-2l\rho}\right] - \frac{\beta J}{2}\sum_{l=1}^{k} 2^{l(1-2\rho)}\left(\frac{m_1^2 + m_2^2}{2}\right),
\end{aligned} \tag{3.23}$$

that is

$$\begin{aligned}
f_{k+1}(h,\beta,J,\rho) &\geq \log 2 + \frac{1}{2}\left\{\log\cosh\left[\beta h + \beta J\left(m_1\sum_{l=1}^{k} 2^{l(1-2\rho)} + 2^{(k+1)(1-2\rho)}m\right)\right]\right\} + \\
&+ \frac{1}{2}\left\{\log\cosh\left[\beta h + \beta J\left(m_2\sum_{l=1}^{k} 2^{l(1-2\rho)} + 2^{(k+1)(1-2\rho)}m\right)\right]\right\} + \\
&- \frac{\beta J}{2}\left[2^{(k+1)(1-2\rho)}m^2 + \sum_{l=1}^{k+1} 2^{-2l\rho}\right] - \frac{\beta J}{2}\sum_{l=1}^{k} 2^{l(1-2\rho)}\left(\frac{m_1^2 + m_2^2}{2}\right) \\
&= f(k, m, m_1, m_2|h,\beta,J,\rho).
\end{aligned} \tag{3.24}$$

Therefore, we have that $f_{k+1}(h,\beta,J,\rho) \geq \sup_{m,m_1,m_2} f(k,m,m_1,m_2|h,\beta,J,\rho)$ and we need to evaluate the optimal order parameters in order to have the best free energy estimate.

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 $f(k, m, m_1, m_2|h, \beta, J, \rho)$, which read as

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

where the third equation is only a linear combination of m_1 and m_2 and it simply states that the global magnetization is the average of the ones 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 (meta-stable) one $m_A = m_1 = -m_2$ 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 $(\pm m_P, \pm m_A)$. Taking the thermodynamic limit we get the following

Theorem 3.1.4. *The mean-field bound for the DHM free energy associated to the meta-stable state reads as*

$$\begin{aligned} f(h, \beta, J, \rho) \geq \sup_{m_1, m_2, m} & \quad \lim_{k\to\infty} f(k, m, m_1, m_2) \\ = \sup_{m_1, m_2} & \quad \log 2 + \frac{1}{2}\log\cosh(\beta h + \beta J C_{2\rho-1} m_1) \\ & + \quad \frac{1}{2}\log\cosh((\beta h + \beta J C_{2\rho-1} m_2) - \frac{\beta J C_{2\rho}}{2} - \frac{\beta J C_{2\rho-1}}{2}(\frac{m_1^2 + m_2^2}{2}), \end{aligned} \tag{3.25}$$

where $C_y = \frac{2^{-y}}{1-2^{-y}}$. The mean field bound for the DHM free energy associated to the ferromagnetic state can be obtained again simply by identifying $m_1 = m_2 = m$.

In the thermodynamic limit, the last level of interaction (the largest in number of links but the weakest as for 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 satisfying the following

Proposition 3.1.5. *The mixture state of the DHM has two independent order parameters, one for each larger cluster, whose self-consistencies read as*

$$m_{1,2} = \tanh\left(\beta h + \beta J C_{2\rho-1} m_{1,2}\right). \tag{3.26}$$

One step forward, if we want to find out the critical value β_c that breaks ergodicity, we can expand them for $k \to \infty$, and for $h = 0$, hence obtaining, in the limit $m_{1,2} \to 0$:

$$\left\{ \begin{array}{l} m_1 \sim \beta J m_1 \frac{2^{1-2\rho}}{1-2^{1-2\rho}} + \mathcal{O}(m_1^3), \\ \\ m_2 \sim \beta J m_2 \frac{2^{1-2\rho}}{1-2^{1-2\rho}} + \mathcal{O}(m_2^3), \end{array} \right\}$$

such that we can write the next

Corollary 3.1.6. *Mean-field criticality in the DHM has the classical critical exponent one half and critical temperature β_c^{MF} given by*

$$\beta_c^{MF} = \frac{1 - 2^{1-2\rho}}{J 2^{1-2\rho}}. \tag{3.27}$$

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), as for instance happens when we use a vanishing external field in a critical mean-field ferromagnet to select the phase by hand. In the appendix D we give a detailed explanation, and a rigorous proof, that this is not the case here: The DHM has links too *evanescent* to drive all the spins to converge always to the same sign and mixture states are preserved.

3.1.3 The not-mean-field scenario

In this Section we bypass the mean-field limitations and show that the outlined scenario is robust even beyond the mean-field picture. We stress that it is not 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 [40] to take into account (at least a) part of the fluctuations of the order parameters (thus improving the previous description) as, in models beyond mean-field, the magnetization is no longer self-averaging and its fluctuations can not be neglected. Let us start investigating the improved bound with the following

Definition 3.1.7. *Once introduced two suitable real parameters t, x, the interpolating Hamiltonian that we are going to consider to bypass the mean-field bound has the form*

$$H_{k+1,t}(\vec{\sigma}) = -tu(\vec{\sigma}) - (1-t)v(\vec{\sigma}) + H_k(\vec{\sigma_1}) + H_k(\vec{\sigma_2}), \tag{3.28}$$

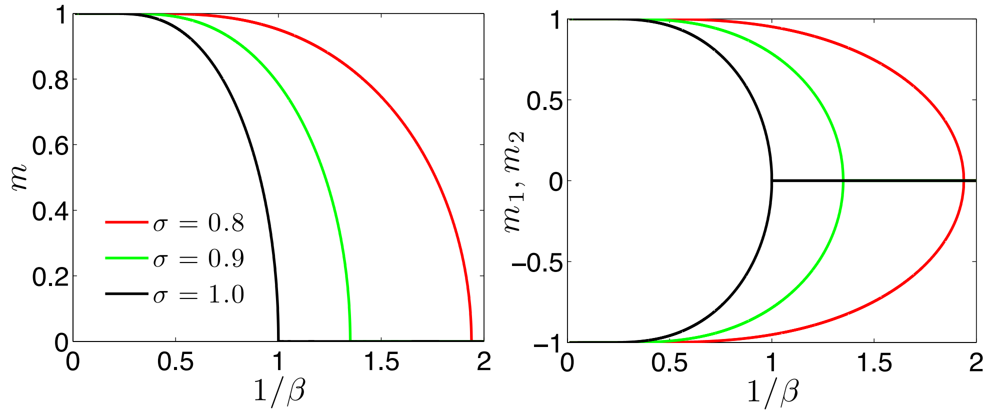

Figure 3.2: Behavior of the magnetizations for the Dyson model within the non-mean-field scenario. Left panel: Pure state (serial processing). Right panel: Mixture state (parallel processing). Note that the difference in energy among the pure state and the mixture state scales as $\Delta E \propto 2^{(k+1)\cdot(1-2\rho)}$, thus -in the thermodynamic limit- the parallel state becomes effectively stable.

with

$$\begin{aligned} u(\vec{\sigma}) &= \frac{J}{2^{2\rho(k+1)}} \sum_{i>j=1}^{2^{(k+1)}} \sigma_i\sigma_j + \frac{xJ}{2 \cdot 2^{2\rho(k+1)}} \sum_{i,j=1}^{2^{k+1}} (\sigma_i - m)(\sigma_j - m), \\ v(\vec{\sigma}) &= \frac{J(1+x)}{2 \cdot 2^{2\rho(k+1)}} \left[\sum_{i,j=1}^{2^k} (\sigma_i - m)(\sigma_j - m) + \sum_{i,j=2^k+1}^{2^{k+1}} (\sigma_i - m)(\sigma_j - m) \right] + mJ2^{(k+1)(1-2\rho)} \sum_{i=1}^{2^{k+1}} \sigma_i, \end{aligned}$$

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,

$$Z_{k+1,t}(x,h) = \sum_{\vec{\sigma}} \exp\left\{ -\beta \left[H_{k+1,t}(\vec{\sigma}) + h \sum_{i=1}^{2^{k+1}} \sigma_i \right] \right\}, \tag{3.29}$$

$$\Phi_{k+1,t}(x,h) = \frac{1}{2^{k+1}} \log Z_{k+1,t}(x,h). \tag{3.30}$$

The procedure that yields to the non-mean-field bound for the free energy permits to obtain (see [40]) the following expression for the pure ferromagnetic case (where again we omitted the dependence by $\beta,\ J,\ \rho$ for simplicity)

$$f_{k+1}(h,\beta,J,\rho) \geq \Phi_{k,1}(\frac{1}{2^{2\rho}}, h + m2^{(k+1)(1-2\rho)}) - \frac{\beta J}{2}(2^{(k+1)(1-2\rho)}m^2 + 2^{-2\rho(k+1)}). \tag{3.31}$$

However, as shown for the previous bound, let us now suppose that the system is split in two parts, with two different magnetizations $m_{left} = m_1$ and $m_{right} = m_2$: resuming the same lines of reasoning of the previous Section, we obtain

$$\Phi_{k,1}(\frac{1}{2^{2\rho}}, h + m2^{(k+1)(1-2\rho)}) = \frac{1}{2}\Phi^1_{k,1}(\frac{1}{2^{2\rho}}, h + m2^{(k+1)(1-2\rho)}) + \frac{1}{2}\Phi^2_{k,1}(\frac{1}{2^{2\rho}}, h + m2^{(k+1)(1-2\rho)}). \tag{3.32}$$

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 (3.31), derived within the ansatz of a pure ferromagnetic state, is generalized by the following expression

$$\begin{aligned} f_{k+1}(h, \beta, J, \rho) \geq & \frac{1}{2}\Phi_{0,1}(\sum_{l=1}^{k+1} 2^{-2l\rho}, h + Jm_1 \sum_{l=1}^{k} 2^{l(1-2\rho)} + mJ2^{(k+1)(1-2\rho)}) + \\ & + \frac{1}{2}\Phi_{0,1}(\sum_{l=1}^{k+1} 2^{-2l\rho}, h + Jm_2 \sum_{l=1}^{k} 2^{l(1-2\rho)} + mJ2^{(k+1)(1-2\rho)}) + \\ & - \frac{\beta J}{2}\sum_{l=1}^{k} 2^{l(1-2\rho)}(\frac{m_1^2 + m_2^2}{2}) - \frac{\beta J}{2}\sum_{l=1}^{k+1} 2^{-2l\rho} - \frac{\beta J}{2}2^{(k+1)(1-2\rho)}m^2. \end{aligned}$$

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

$$\begin{aligned} & \Phi_{0,1}(\sum_{l=1}^{k+1} 2^{-2l\rho}, h + Jm_1 \sum_{l=1}^{k} 2^{l(1-2\rho)} + mJ2^{(k+1)(1-2\rho)}) = \ln 2 + \\ & + \frac{\beta J}{2}(1 + m_1^2)\sum_{l=1}^{k+1} 2^{-2l\rho} + \log\cosh\left\{\beta h + \beta mJ2^{(k+1)(1-2\rho)} + \beta m_1 J\left[\sum_{l=1}^{k} 2^{l(1-2\rho)} - \sum_{l=1}^{k+1} 2^{-2l\rho}\right]\right\} \end{aligned}$$

in such a way that

$$\begin{aligned} f_{k+1} \geq \log 2 \; & + \; \frac{1}{2}\log\cosh\left\{\beta h + \beta mJ2^{(k+1)(1-2\rho)} + \beta m_1 J\left[\sum_{l=1}^{k} 2^{l(1-2\rho)} - \sum_{l=1}^{k+1} 2^{-2l\rho}\right]\right\} + \\ & + \; \frac{1}{2}\log\cosh\left\{\beta h + \beta mJ2^{(k+1)(1-2\rho)} + \beta m_2 J\left[\sum_{l=1}^{k} 2^{l(1-2\rho)} - \sum_{l=1}^{k+1} 2^{-2l\rho}\right]\right\} + \\ & - \; \frac{\beta J}{2}\left[\sum_{l=1}^{k} 2^{l(1-2\rho)} - \sum_{l=1}^{k+1} 2^{-2l\rho}\right]\left(\frac{m_1^2 + m_2^2}{2}\right) + \\ & - \; \frac{\beta J}{2}2^{(k+1)(1-2\rho)}m^2. \end{aligned} \tag{3.35}$$

Summarizing, in the thermodynamic limit one has the following

Theorem 3.1.8. *The non-mean-field bound for the DHM's free energy associated to the mixture state reads as*

$$f(h,\beta,J,\rho) \geq \sup_{m_1,m_2} \Big\{ \log 2 + \frac{1}{2}\log\cosh\left[\beta h + \beta m_1 J(C_{2\rho-1}-C_{2\rho})\right] + \tag{3.36}$$
$$+ \frac{1}{2}\log\cosh\left[\beta h + \beta m_2 J(C_{2\rho-1}-C_{2\rho})\right] - \frac{\beta J}{2}(C_{2\rho-1}-C_{2\rho})\left(\frac{m_1^2+m_2^2}{2}\right)\Big\},$$

where $C_y = \frac{2^{-y}}{1-2^{-y}}$.

Proposition 3.1.9. *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*

$$m_{1,2} = \tanh(\beta h + \beta J m_{1,2}(C_{2\rho-1}-C_{2\rho})). \tag{3.37}$$

As for the MF approximation, we are going to find the critical temperature β_c; considering the system at zero external field $h=0$, thus writing

$$\begin{cases} m_1 \sim \beta J m_1(\frac{1}{2^{2\rho-1}-1} - \frac{1}{2^{2\rho}-2^{-2\rho}}) + \mathcal{O}(m_1^3), \\ \\ m_2 \sim \beta J m_2(\frac{1}{2^{2\rho-1}-1} - \frac{1}{2^{2\rho}-2^{-2\rho}}) + \mathcal{O}(m_2^3), \end{cases}$$

so to get the following

Corollary 3.1.10. *This non-mean-field criticality, in the DHM, has the classical exponent too but a different critical temperature* β_c^{NMF} *given by the following formula:*

$$\beta_c^{NMF} = \frac{(2^{2\rho}-1)(1-2^{1-2\rho})}{J}. \tag{3.38}$$

3.2 Analysis of the Hopfield hierarchical model

As we saw in the previous Section, the Dyson model has a rich variety of states, whit the same free energy in the thermodynamic limit. Now we want to apply the analysis previously outlined and the ideas that stemmed from the related findings to a *Hierarchical Hopfield Model.*

To this task we need to introduce, beyond 2^{k+1} dichotomic spins/neurons, also p quenched patterns ξ^μ, $\mu \in (1,...,p)$, 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

$$P(\xi_i^\mu) = \frac{1}{2}\delta(\xi_i^\mu - 1) + \frac{1}{2}\delta(\xi_i^\mu + 1). \tag{3.39}$$

Mirroring the previous Section, the Hamiltonian of the hierarchical Hopfield model is as well defined recursively by the following

Definition 3.2.1. *The Hamiltionian of Hierarchical Hopfield model (HHM) is defined by*

$$H_{k+1}(\vec{\sigma}) = H_k(\vec{\sigma_1}) + H_k(\vec{\sigma_2}) - \frac{1}{2}\frac{1}{2^{2\rho(k+1)}}\sum_{\mu=1}^{p}\sum_{i,j=1}^{2^{k+1}}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j \tag{3.40}$$

with $H_0(\sigma) = 0$; $\rho \in (1/2, 1)$ is a number tuning the interaction strength with the neuron's distance, and p is 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 $\propto h_\mu \sum_i^{2^{k+1}} \xi_i^\mu \sigma_i$, and a survey overall the stimuli is accomplished summing over $\mu \in (1, ..., p)$ all the h_μ.

Even in this context, we can again write the Hamiltonian of the HHM in terms of a distance $d_{i,j}$ between the spin pair (i, j) obtaining

$$H_k[\{\sigma_1...\sigma_{2^k}\}] = -\sum_{i<j}\sigma_i\sigma_j\left[\sum_{l=d_{i,j}}^{k}\left(\frac{\sum_{\mu=1}^{P}\xi_i^\mu\xi_j^\mu}{2^{2\rho l}}\right)\right] = -\sum_{i<j}\sigma_i\sigma_j\widetilde{J_{ij}}, \tag{3.41}$$

$$\widetilde{J_{ij}} = \sum_{l=d_{i,j}}^{k}\left(\frac{\sum_{\mu=1}^{P}\xi_i^\mu\xi_j^\mu}{2^{2\rho l}}\right) = J(d_{i,j}, k, \rho)\sum_{\mu=1}^{P}\xi_i^\mu\xi_j^\mu \tag{3.42}$$

where, the Hebbian kernel on a hierarchical topology becomes modified by the distance-dependent weight

$$J(d_{i,j}, k, \rho) = \frac{4^{\rho - d_{i,j}\rho} - 4^{-k\rho}}{4^\rho - 1} \cdot . \tag{3.43}$$

Definition 3.2.2. *We introduce the Mattis magnetizations (or Mattis overlaps), over the whole system, as*

$$m_\mu(\vec{\sigma}) = \frac{1}{2^{k+1}}\sum_{i=1}^{2^{k+1}}\xi_i^\mu\sigma_i. \tag{3.44}$$

The definition can be extended trivially to the inner clusters restricting properly the sum over the spins, e.g. dealing with the two larger sub-clusters as before we have

$$m_{left}^\mu = \frac{1}{2^k}\sum_{i=1}^{2^k}\xi_i^\mu\sigma_i, \qquad m_{right}^\mu = \frac{1}{2^k}\sum_{j=2^k+1}^{2^{k+1}}\xi_j^\mu\sigma_j. \tag{3.45}$$

3.2.1 The thermodynamic limit

As for the previous investigation, at first we want to prove that the model is well defined, namely that the thermodynamic limit for the free energy exists. To this task we have the following

Theorem 3.2.3. *The thermodynamic limit of the HHM's free energy exists and we call*

$$\lim_{k\to\infty} f_{k+1}(\beta,p,\rho) = f(\beta,p,\rho).$$

Let us write the Hamiltonian as

$$H_{k+1}(\vec{\sigma}) = H_k(\vec{\sigma_1}) + H_k(\vec{\sigma_2}) - \frac{1}{2}2^{(k+1)}2^{(k+1)(1-2\rho)}\sum_{\mu=1}^{p}(m_\mu^{k+1}(\vec{\sigma}))^2,$$

and let us consider the following interpolation, where again -for the sake of simplicity- hereafter we stress the dependence by the external fields $\{h_\mu\}$ only and use the symbol $\mathbb{E}_\xi$ to denote averaging over the quenched patterns:

$$\begin{aligned}&\Phi_{k+1,t}(\{h_\mu\}) = \qquad (3.46)\\ = \;&\frac{1}{2^{k+1}}\mathbb{E}_\xi \log\sum_{\vec{\sigma}}\exp\left\{\beta\left[-H_k(\vec{\sigma_1}) - H_k(\vec{\sigma_2}) + t\frac{1}{2}2^{(k+1)}2^{(k+1)(1-2\rho)}\sum_{\mu=1}^{p}(m_\mu^{k+1}(\vec{\sigma}))^2 + \sum_{\mu=1}^{p}h_\mu\xi_i^\mu\sigma_i\right]\right\}.\end{aligned}$$

We notice that

$$\Phi_{k+1,1}(h) = f_{k+1}, \qquad (3.47)$$

$$\Phi_{k+1,0}(h) = f_k \qquad (3.48)$$

and that

$$\frac{d}{dt}\Phi_{k+1,t} = \left\langle \frac{1}{2^{k+1}}\frac{2^{(k+1)}2^{(k+1)(1-2\rho)}}{2}\sum_{\mu=1}^{p}[m_\mu^{k+1}(\vec{\sigma})]^2\right\rangle_t \geq 0. \qquad (3.49)$$

in such a way that $f_{k+1}(\beta,p,\rho) \geq f_k(\beta,p,\rho)$. Now we want to prove that $f_{k+1}(\beta,p,\rho)$ is bounded: it is enough to see that

$$f_{k+1}(\beta,p,\rho) = f_k(\beta,p,\rho) + \int_0^1 \frac{d}{dt}\Phi_{k+1,t}: \qquad (3.50)$$

Since we have

$$\frac{d}{dt}\Phi_{k+1,t} = \left\langle \beta\frac{2^{(k+1)}2^{(k+1)(1-2\rho)}}{2}\sum_{\mu=1}^{p}(m_\mu^{k+1}(\vec{\sigma}))^2\right\rangle_t \leq \beta p\frac{2^{(k+1)}2^{(k+1)(1-2\rho)}}{2}, \qquad (3.51)$$

we can write

$$f_{k+1}(\beta,p,\rho) \leq f_k(\beta,p,\rho) + \beta p \frac{2^{(k+1)(1-2\rho)}}{2}. \tag{3.52}$$

Iterating this procedure over the levels we get

$$f_{k+1}(\beta,p,\rho) \leq f_0(\beta,p,\rho) + \frac{\beta p}{2}\sum_{l=1}^{k+1} 2^{l(1-2\rho)}, \tag{3.53}$$

such that, in the $k \to \infty$ limit, we can write

$$f \leq f_0 + \frac{\beta p}{2}\sum_{l=1}^{\infty} 2^{l(1-2\rho)}.$$

Since $\rho > \frac{1}{2}$ the series on the r.h.s. of the above inequality converges, thus $f(\beta,p,\rho)$ is bounded by

$$f(\beta,p,\rho) \leq f_0 + \frac{\beta p}{2}\frac{1}{2^{(2\rho-1)}-1}$$

and non increasing for (3.49), then its thermodynamic limit exists.

3.2.2 The mean-field scenario

Plan of this Section is to investigate the serial and parallel retrieval capabilities in the HHM at the mean-field level. As usual, we obtain our goal by mixing the Amit ansatz technique (in selecting suitably candidate states for retrieval) with the interpolation technique.

Definition 3.2.4. *Let us define the interpolating Hamiltonian $H_{k+1,t}(\vec{\sigma})$ as*

$$H_{k+1,t}(\vec{\sigma}) = H_k(\vec{\sigma_1}) + H_k(\vec{\sigma_2}) - \frac{t}{2\cdot 2^{2\rho(k+1)}}\sum_{\mu=1}^{p}\sum_{i,j=1}^{2^{k+1}}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j - (1-t)\cdot 2^{(k+1)(1-2\rho)}\sum_{\mu=1}^{p} m_\mu \sum_{i=1}^{2^{k+1}}\xi_i^\mu\sigma_i, \tag{3.54}$$

Clearly, we can associate such an Hamiltonian to an extended partition function $Z_{k+1,t}(h)$ and to an extended free energy $\Phi_{k+1,t}(h)$ as

$$Z_{k+1,t}(\{h_\mu\}) = \sum_{\vec{\sigma}} \exp\left\{-\beta\left[H_{k+1,t}(\vec{\sigma}) + \sum_{\mu=1}^{p} h_\mu \sum_{i=1}^{2^{k+1}}\xi_i^\mu\sigma_i\right]\right\}, \tag{3.55}$$

$$\Phi_{k+1,t}(\{h_\mu\}) = \frac{1}{2^{k+1}}\mathbb{E}_\xi \log Z_{k+1,t}(\{h_\mu\}), \tag{3.56}$$

where, for the sake of simplicity, we stressed only the dependence by the fields. We can rewrite (3.54) as

$$H_{k+1,t}(\vec{\sigma}) = H_k(\vec{\sigma_1})+H_k(\vec{\sigma_2})-\frac{2^{2(k+1)}t}{2\cdot 2^{2\rho(k+1)}}\sum_{\mu=1}^{p} m^2_{k+1,\mu}(\vec{\sigma})-(1-t)2^{(k+1)}2^{(1-2\rho)(k+1)}\sum_{\mu=1}^{p} m_{k+1,\mu}(\vec{\sigma})m_\mu, \tag{3.57}$$

It is easy to show that

$$\Phi_{k+1,1}(\{h_\mu\}) = f_{k+1}, \tag{3.58}$$

$$\Phi_{k+1,0}(\{h_\mu\}) = \Phi_{k,1}(\{h_\mu + 2^{(k+1)(1-2\rho)}m_\mu\}), \tag{3.59}$$

and that

$$\begin{aligned}
\frac{d\Phi_{k+1,t}}{dt} &= \frac{1}{2^{k+1}}\frac{1}{Z_{k+1,t}}\sum_{\vec{\sigma}}\exp(-\beta(H_{k+1,t}(\vec{\sigma})+\sum_{\mu=1}^{p} h_\mu \sum_{i=1}^{2^{k+1}}\xi_i^\mu\sigma_i))(-\beta\frac{dH_{k+1,t}(\vec{\sigma})}{dt}) \\
&= \frac{1}{2^{k+1}}\frac{1}{Z_{k+1,t}}\sum_{\vec{\sigma}}\exp(-\beta(H_{k+1,t}(\vec{\sigma})+\sum_{\mu=1}^{p} h_\mu \sum_{i=1}^{2^{k+1}}\xi_i^\mu\sigma_i))\times \\
&\quad \times(\frac{\beta 2^{2(k+1)}}{2\cdot 2^{2\rho(k+1)}}\sum_{\mu=1}^{p} m^2_{k+1,\mu}(\vec{\sigma})-\beta 2^{(k+1)(1-2\rho)}2^{(k+1)}\sum_{\mu=1}^{p} m_\mu m_{k+1,\mu}(\vec{\sigma})) \\
&= \frac{\beta}{2}2^{(k+1)(1-2\rho)}\left\langle \sum_{\mu=1}^{p} m^2_{k+1,\mu}(\vec{\sigma}) - 2m_\mu m_{k+1,\mu}(\vec{\sigma}) + m_\mu^2 \right\rangle_t - \frac{\beta}{2}2^{(k+1)(1-2\rho)}\sum_{\mu=1}^{p} m_\mu^2 \\
&= \frac{\beta}{2}2^{(k+1)(1-2\rho)}\sum_{\mu=1}^{p}\left\langle (m_\mu^{k+1}(\vec{\sigma})-m_\mu)^2\right\rangle_t - \frac{\beta}{2}2^{(k+1)(1-2\rho)}\sum_{\mu=1}^{p} m_\mu^2.
\end{aligned}$$

Since the term in the brackets above $\langle\cdot\rangle_t$ is nonnegative, we get

$$\begin{aligned}
\Phi_{k+1,1} &= \Phi_{k+1,0} + \int_0^1 \frac{d\Phi_{k+1,t}(x,h)}{dt}dt \\
&\geq \Phi_{k,1}(\{h_\mu + 2^{(k+1)(1-2\rho)}m_\mu\}) - \frac{\beta}{2}2^{(k+1)(1-2\rho)}\sum_{\mu=1}^{p} m_\mu^2 \\
&\geq \Phi_{1,0}(\{h_\mu + \sum_{l=2}^{k+1} 2^{l(1-2\rho)}m_\mu\}) - \frac{\beta}{2}\sum_{l=2}^{k+1} 2^{l(1-2\rho)}\sum_{\mu=1}^{p} m_\mu^2 \\
&= \Phi_{0,1}(\{h_\mu + \sum_{l=1}^{k+1} 2^{l(1-2\rho)}m_\mu\}) - \frac{\beta}{2}\sum_{l=1}^{k+1} 2^{l(1-2\rho)}\sum_{\mu=1}^{p} m_\mu^2,
\end{aligned}$$

where we used (3.59) recursively.

Now we can estimate the last term, $\Phi_{0,1}(\{h_\mu + \sum_{l=1}^{k+1} 2^{l(1-2\rho)} m_\mu\})$, in the following way

$$\Phi_{0,1}(\{h_\mu + \sum_{l=1}^{k+1} 2^{l(1-2\rho)} m_\mu\}) = \mathbb{E}_\xi \log \sum_{S\in\{-1,1\}} \exp(\beta \sum_{\mu=1}^{p} (h_\mu + \sum_{l=1}^{k+1} 2^{l(1-2\rho)} m_\mu)\xi^\mu S) \quad (3.60)$$

$$= \log 2 + \mathbb{E}_\xi \log\cosh(\beta \sum_{\mu=1}^{p} (h_\mu + \sum_{l=1}^{k+1} 2^{l(1-2\rho)} m_\mu)\xi^\mu), \quad (3.61)$$

where $\mathbb{E}_\xi$ averages over the quenched patterns as usual.
Summarizing we have

$$f_{k+1} \geq \log 2 + \mathbb{E}_\xi \log\cosh(\beta \sum_{\mu=1}^{p} (h_\mu + \sum_{l=1}^{k+1} 2^{l(1-2\rho)} m_\mu)\xi^\mu) - \frac{\beta}{2} \sum_{l=1}^{k+1} 2^{l(1-2\rho)} \sum_{\mu=1}^{p} m_\mu^2. \quad (3.62)$$

which is enough to state the next

Theorem 3.2.5. *(Mean Field Bound for Serial Retrieval) Given $-1 \leq m_\mu \leq +1$, $\forall \mu = 1, ..., p$ the following relation holds*

$$f(\beta, \{h_\mu\}, p) \geq \sup_{\{m^\mu\}} \left[\log 2 + \mathbb{E}_\xi \log\cosh(\beta \sum_{\mu=1}^{p} (h_\mu + C_{2\rho-1} m_\mu)\xi^\mu) - \frac{\beta}{2} C_{2\rho-1} \sum_{\mu=1}^{p} m_\mu^2 \right],$$

where the optimal order parameters are the solutions of the system

$$m^\mu = \mathbb{E}_\xi \xi^\mu \tanh(\beta \sum_{\nu=1}^{p} (h_\nu + C_{2\rho-1} m^\nu)\xi^\nu),$$

that are the self-consistent equations of a standard Hopfield model with rescaled temperature $\beta C_{2\rho-1}$.

Again the critical temperature of the model with no external fields, separating the paramagnetic phase from the retrieval one, can be obtained expanding for small $\{m^\mu\}$, so to get

$$m^\mu = \mathbb{E}_\xi[\beta C_{2\rho-1} \xi^\mu \sum_{\nu=1}^{p} (\xi^\nu m^\nu)] + \mathcal{O}(m^{\mu 2}) = \beta C_{2\rho-1} + \mathcal{O}(m^{\mu 2}) \quad (3.63)$$

hence $\beta_c^{MF} = C_{2\rho-1}^{-1}$. As previously outlined for the DHM, it is possible to assume -for the k^{th} level- two different classes of Mattis magnetizations $m^\mu_{left} = m^\mu_1$ and $m^\mu_{right} = m^\mu_2$ 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 way we write

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

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 order parameters of $\Phi^{1,2}$ we obtain

$$\begin{aligned} f_{k+1} &\geq \frac{1}{2}\Phi_{0,1}(\{h_\mu + \sum_{l=1}^{k} 2^{l(1-2\rho)} m_1^\mu + 2^{(k+1)(1-2\rho)} m^\mu\}) + \frac{1}{2}\Phi_{0,1}(\{h_\mu + \sum_{l=1}^{k} 2^{l(1-2\rho)} m_2^\mu \\ &+ 2^{(k+1)(1-2\rho)} m^\mu\}) - \frac{\beta}{2} 2^{(k+1)(1-2\rho)} \sum_{\mu=1}^{p} m_\mu^2 - \frac{\beta}{2} \sum_{l=1}^{k} 2^{l(1-2\rho)} \sum_{\mu=1}^{p} \frac{(m_1^\mu)^2 + (m_1^\mu)^2}{2} \end{aligned} \quad (3.64)$$

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

Theorem 3.2.6. *(Mean Field Bound for Parallel Retrieval) Given $-1 \leq m_\mu \leq +1$, $\forall \mu = 1, ..., p$ the following relation holds*

$$\begin{aligned} f(\beta, \{h_\mu\}, p) &\geq \sup_{\{m^\mu\}} [\log 2 + \mathbb{E}_\xi \log\cosh(\beta \sum_{\mu=1}^{p} (h_\mu + C_{2\rho-1} m_1^\mu) \xi^\mu) \\ &+ \mathbb{E}_\xi \log\cosh(\beta \sum_{\mu=1}^{p} (h_\mu + C_{2\rho-1} m_2^\mu) \xi^\mu) - \frac{\beta}{2} C_{2\rho-1} \sum_{\mu=1}^{p} \frac{m_1^{\mu 2} + m_2^{\mu 2}}{2}] \end{aligned} \quad (3.65)$$

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

$$m^\mu_{1,2} = \mathbb{E}_\xi \xi^\mu \tanh(\beta \sum_{\nu=1}^{p} (h_\nu + C_{2\rho-1} m^\nu_{1,2}) \xi^\nu)$$

and whose critical temperature is again $\beta_c^{MF} = C_{2\rho-1}^{-1}$.

3.2.3 The not-mean-field scenario

Scope of the present Section is to bypass mean-field limitations and show that the outlined scenario is robust. To this task, mirroring the previous analysis on DHM, here we provide an improved (with respect to the mean-field counterpart) bound.
The idea underlying this non-mean-field bound is the same that we used in the DHM, extensively explained in [40]. Let us start introducing the following

Definition 3.2.7. *Let us take $x \geq 0$ -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-, and let us introduce also the following interpolating Hamiltonian*

$$H_{k+1,t} = -tu(\vec{\sigma}) - (1-t)v(\vec{\sigma}) + H_k(\vec{\sigma_1}) + H_k(\vec{\sigma_2}) \quad (3.66)$$

with

$$
\begin{aligned}
u(\vec{\sigma}) &= \frac{1}{2\cdot 2^{2\rho(k+1)}}\sum_{\mu=1}^{p}\sum_{i,j}^{2^{k+1}}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j+\frac{x}{2\cdot 2^{2\rho(k+1)}}\sum_{\mu=1}^{p}\sum_{i,j=1}^{2^{k+1}}(\xi_i^\mu\sigma_i-m_\mu)(\xi_j^\mu\sigma_j-m_\mu), && (3.67)\\
v(\vec{\sigma}) &= \frac{(x+1)}{2\cdot 2^{2\rho(k+1)}}\Big(\sum_{\mu=1}^{p}\sum_{i,j=1}^{2^{k}}(\xi_i^\mu\sigma_i-m_\mu)(\xi_j^\mu\sigma_j-m_\mu)+\sum_{i,j=2^k+1}^{2^{k+1}}(\xi_i^\mu\sigma_i-m_\mu)(\xi_j^\mu\sigma_j-m_\mu)\Big)\\
&+ \sum_{\mu=1}^{p}m_\mu 2^{(k+1)(1-2\rho)}\sum_{i=1}^{2^{k+1}}\xi_i^\mu\sigma_i. && (3.68)
\end{aligned}
$$

The partition function and free energy associated to the Hamiltonian (3.66) are, respectively,

$$
\begin{aligned}
Z_{k+1,t}(x,\{h_\mu\}) &= \sum_{\vec{\sigma}}\exp\Big(-\beta\big(H_{k+1,t}(\vec{\sigma})+\sum_{\mu=1}^{p}\sum_{i}^{2^{k+1}}h_i^\mu\xi_i^\mu\sigma_i\big)\Big), && (3.69)\\
\Phi_{k+1,t}(x,\{h_\mu\}) &= \frac{1}{2^{k+1}}\mathbb{E}_\xi\log Z_{k+1,t}(x,\{h_\mu\}). && (3.70)
\end{aligned}
$$

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

$$
\Phi_{k+1,0}(x,\{h_\mu\})=\Phi_{k,1}\Big(\frac{1+x}{2^{2\rho}},\{h_\mu+m_\mu 2^{(k+1)(1-2\rho)}\}\Big). \tag{3.71}
$$

It is possible to show that the derivative of $\Phi_{k+1,t}$ with respect to t is

$$
\begin{aligned}
\frac{d\Phi_{k+1,t}}{dt}(x,t) &= \frac{1}{2^{k+1}}\frac{1}{Z_{k+1,t}}\sum_{\vec{\sigma}}\exp\Big(-\beta\big(H_{k+1,t}(\vec{\sigma})+\sum_{\mu=1}^{p}h_\mu\sum_{i}^{2^{k+1}}\xi_i^\mu\sigma_i\big)\Big)(\beta u(\vec{\sigma})-\beta v(\vec{\sigma}))\\
&= -\frac{\beta}{2}2^{(k+1)(1-2\rho)}\sum_{\mu=1}^{p}m_\mu^2\\
&+ \frac{\beta(x+1)}{2^{(k+1)(1+2\rho)}}\sum_{\mu=1}^{p}\sum_{1\le i\le 2^k}\sum_{2^k+1\le j\le 2^{k+1}}\Big\langle(\xi_i^\mu\sigma_i-m_\mu)(\xi_j^\mu\sigma_j-m_\mu)\Big\rangle_t && (3.72)
\end{aligned}
$$

and consequently

$$
\begin{aligned}
f_{k+1}=\Phi_{k+1,1}(0,\{h_\mu\}) &= \Phi_{k,1}\Big(\frac{1}{2^{2\rho}},\{h_\mu+\beta m_\mu 2^{(k+1)(1-2\rho)}\}\Big)-\frac{\beta}{2}2^{(k+1)(1-2\rho)}\sum_{\mu=1}^{p}m_\mu^2\\
&+ \mathcal{C}(k+1,\beta,\rho,\{m_\mu\}) && (3.73)
\end{aligned}
$$

Iterating the procedure one arrives to:

$$
\begin{aligned}
f_{k+1} &= \Phi_{0,1}\Big(\sum_{l=1}^{k+1}2^{-2l\rho},\{h_\mu+\beta m_\mu\sum_{l=1}^{k+1}2^{l(1-2\rho)}\}\Big)-\frac{\beta}{2}\sum_{l=1}^{k+1}2^{l(1-2\rho)}\sum_{\mu=1}^{p}m_\mu^2\\
&+ \sum_{l=1}^{k+1}\mathcal{C}(l,\beta,\rho,\{m_\mu\}). && (3.74)
\end{aligned}
$$

Now we are going to neglect the fluctuation source, containing $\left\langle (\xi_i^\mu \sigma_i - m_\mu)(\xi_j^\mu \sigma_j - m_\mu) \right\rangle_t$, and that we indicate with $\mathcal{C}(k+1, \beta, \rho, \{m_\mu\})$: at difference with before, while in the pure ferromagnetic case Griffiths inequalities hold [61, 62] and ensure that such a term is positive defined (thus allowing us to get the bound), in this context -as for neural networks Griffiths theory have not yet been developed- we are left with an approximation only. Calculating the value of $\Phi_{0,1}$, using the (3.69), (3.70) and (3.71) we get the following

Theorem 3.2.8. *(Non-mean field approximation for Serial retrieval) Given $-1 \le m_\mu \le +1$, $\forall \mu = 1, ..., p$ the Serial NMF-approximation for the Hierarchical Hopfield model reads as*

$$f^{NMF}(\beta, \{h_\mu\}, p) = \sup_m \left[\log 2 + \mathbb{E}_\xi \log\cosh(\sum_{\mu=1}^p (h_\mu + \beta m_\mu (C_{2\rho-1} - C_{2\rho}))\xi^\mu) - \frac{\beta}{2} \sum_{\mu=1}^p m_\mu^2 (C_{2\rho-1} - C_{2\rho}) \right],$$

representing an Hopfield model at rescaled temperature, with optimal order parameters fulfilling

$$m^\mu = \mathbb{E}_\xi \xi^\mu \tanh(\beta \sum_{\nu=1}^p (\beta h_\nu + (C_{2\rho-1} - C_{2\rho}) m^\nu) \xi^\nu)$$

and critical temperature $\beta_c^{NMF}(C_{2\rho-1} - C_{2\rho}) = 1$.

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

$$\begin{aligned}
f_{k+1}(\{h_\mu\}, \beta, \rho, p) &= \log 2 + \frac{1}{2}\mathbb{E}_\xi \log\cosh(\sum_{\mu=1}^p (\beta h_\mu + \beta m_1^\mu (\sum_{l=1}^k 2^{l(1-2\rho)} - \sum_{l=1}^{k+1} 2^{l(-2\rho)}) + \beta m^\mu 2^{(k+1)(1-2\rho)}) \xi^\mu) \\
&+ \frac{1}{2}\mathbb{E}_\xi \log\cosh(\sum_{\mu=1}^p (\beta h_\mu + \beta m_2^\mu (\sum_{l=1}^k 2^{l(1-2\rho)} - \sum_{l=1}^{k+1} 2^{l(-2\rho)}) + \beta m^\mu 2^{(k+1)(1-2\rho)}) \xi^\mu) \\
&- \frac{\beta}{2} (\sum_{l=1}^k 2^{l(1-2\rho)} - \sum_{l=1}^{k+1} 2^{l(-2\rho)}) \sum_{\mu=1}^p \frac{m_1^{\mu 2} + m_2^{\mu 2}}{2} - \frac{\beta}{2} 2^{(k+1)(1-2\rho)} \sum_{\mu=1}^p m_\mu^2 \\
&+ \mathcal{C}(k+1, \beta, \rho, \{m_\mu\}) + \frac{1}{2} \sum_{l=1}^k \left(\mathcal{C}(l, \beta, \rho, \{m_\mu^1\}) + \mathcal{C}(l, \beta, \rho, \{m_\mu^2\}) \right)
\end{aligned}$$

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

Definition 3.2.9. *(Non mean field approximation for Parallel retrieval) Given $-1 \le m_\mu \le +1$, $\forall \mu = 1, ..., p$ the Parrallel NMF-approximation for the Hierarchical Hopfield model reads* *as*

$$\begin{aligned} f(\{h_\mu\},\beta,\rho,p) &= \sup_{\{m^\mu_{1,2}\}} \Big\{ \log 2 + \frac{1}{2}\mathbb{E}_\xi \log\cosh\Big[\sum_{\mu=1}^{p}(\beta h_\mu + \beta m_1^\mu (C_{2\rho-1}-C_{2\rho})\Big] \\ &+ \frac{1}{2}\mathbb{E}_\xi \log\cosh\Big[\sum_{\mu=1}^{p}(\beta h_\mu + \beta m_2^\mu (C_{2\rho-1}-C_{2\rho})\Big] \\ &- \frac{\beta}{2}(C_{2\rho-1}-C_{2\rho})\sum_{\mu=1}^{p}\frac{m_1^{\mu 2}+m_2^{\mu 2}}{2}\Big\}, \qquad (3.75) \end{aligned}$$

i.e., the free energy of two independent Hopfield models for each of the two subgroups of spins, with disentangled optimal order parameters satisfying

$$m^\mu_{1,2} = \mathbb{E}_\xi \xi^\mu \tanh(\beta\sum_{\nu=1}^{p}(h_\nu + (C_{2\rho-1}-C_{2\rho})m^\nu_{1,2}\xi^\nu),$$

and critical temperature $\beta_c^{NMF}(C_{2\rho-1}-C_{2\rho}) = 1$.

A last note of interest, regards the capacity of these networks: we have shown how it is possible to recall simultaneously two patterns by spitting the system into two subgroups, going down over the levels from the top and we have seen that, since the upper interaction is vanishing with enough velocity, in the thermodynamic limit the two subgroups of neurons can be thought of as *independent*: 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 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 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

$$\lim_{k\to\infty}\sum_{l=k-M}^{k} 2^{l(1-2\rho)}\sum_{\mu=1}^{p} m_l^\mu = 0. \qquad (3.76)$$

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

$$\begin{aligned} \sum_{l=k-M}^{k} 2^{l(1-2\rho)}\sum_{\mu=1}^{p} m_l^\mu &\le p\sum_{l=k-M}^{k} 2^{l(1-2\rho)} \\ &\le p\sum_{l=k-M}^{\infty} 2^{l(1-2\rho)} \propto 2^{(1-2\rho)(k-M)}p : \qquad (3.77) \end{aligned}$$

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 $p=\mathcal{O}(k)$, $2^{(1-2\rho)(k-\log(p))}p \to 0$ as $k\to\infty$.

Part III

Conclusions

In this thesis, starting from the mean field Hopfield model as the harmonic oscillator of complex network able to retrieve, one at a time, patterns of information, I have discussed about the needed for multitasking associative networks able to manage several patterns of information at the same time and I have investigated the possibility to build them, approaching the problem from a statistical mechanics perspective.

In Chapter 1 I recalled some interesting results about the standard Hopfield model, in particular its equivalence with a bipartite mean field spin glass model, the first (dichotomic) party representing spins and the second (gaussian) party involving patterns. I've shown how such a kind of mapping can be useful for a rigourus investigation of the model, cause it gives us the possibility to use well known results about the monopartite spin glass systems like the Sherrington-Kirkpatrick and the Gaussian spin glass model. Moreover, from the bipartite effective network it is possible to understand the key of the multiple patterns retrieval mechanism, i.e. the existence of weakly interacting subcommunities of spins/neurons. I've shown how these kind of communities can emerge in two different ways: introducing a suitable level of dilution in the bipartite interactions (and consequently in the patterns entries), or by going away from the mean field framework and introducing a different topology in such a way that a non-uniform distance among spins emerges.

There are several future direction starting from this mapping because from a rigorous perspective the Hopfield model is still considered a mathematical challenge. First of all the results obtained in Chapter 1 have to be generalized beyond the replica symmetric approximation, i.e. if the free energy of the analogical Hopfield model can be written as a convex linear combination of the SK and the Gaussian model also in the RSB framework. In this case, since the Gaussian model was proven to be replica symmetric, the complexity of the Hopfield model would be related only to that of the SK one, and this could be very interesting for proving the existence of the Hopfield thermodynamic limit. Moreover all those results have to be carried beyond the analogical (gaussian couplings) limitations.

In Chapter 2 I introduced the Diluted Hopfield model, investigating different regimes of load (number of stored patterns) and dilution: from the low to the high storage regime, from the fully connected to the finite connectivity regime. In particular, in the case when the bipartite network is sparse,i.e. each spin has a finite degree, I've shown that in a large part of the parameter space of noise, dilution and storage load, delimited by a critical surface, the network behaves as an extensive parallel processor, retrieving all P patterns in parallel without falling into spurious states due to pattern cross-talk and typical of the structural glassiness built into the network. In the final section I used a cavity/ belief propagation method including also pattern asymmetry and heterogeneous dilution, leading to a significantly broader range of network structures to which the theory can be applied.

The most interesting future directions regarding the Diluted Hopfield models come from the applications of such a models to real biological system like the immune system. In particular one may think to generalize the dilution procedure allowing correlations between different patterns or introducing a level of interaction among them: this would correspond to consider bipartite systems with intra-party interactions too [27].

In Chapter 3 I investigated the statistical mechanics of hierarchical neural networks. First, I approached these systems a la Mattis, by thinking at the Dyson model as a single-pattern hierarchical neural network and I discussed the stability of equilibrium states different from the ferromagnetic one. One step forward, I extended this scenario toward multiple stored patterns by implementing the Hebb prescription for learning within the couplings, resulting in an Hopfield-like networks constrained on a hierarchical topology. The main finding is that embedding the Hebbian rule on a hierarchical topology allows the network to accomplish both serial and parallel processing, depending on the level of fast noise affecting the system and the decay of the interactions with the distance among neurons.

Hierarchical neural networks are a very recent argument and there are a lot of future directions to follow. First of all these kind of results have to be linked and compared with the standard Renormalization-Group approach to the problem of hierarchical topology, indeed, the existence of stable states different from the ferromagnetic one in the Dyson model can be related perhaps to the lack of self-average in the magnetization. Finally there are several issues still not properly investigated, like for example the network's capacity, i.e. the number of patterns the system is able to retrieve at the same time, bacause up to now there are only numerical results or partial hints coming from the interpolating procedure.

Appendix A

The Gaussian SK model

Following [24], we introduce a system on N sites, whose generic configuration is defined by spin variables $z_i \in \mathbb{R}$, $i = 1, 2, ..., N$ attached on each site. We call the external quenched disorder a set of N^2 independent and identical distributed random variables J_{ij}, defined for each couple of sites (i, j). We assume each J_{ij} to be a centered unit Gaussian $\mathcal{N}(0, 1)$ i.e.

$$\mathbb{E}(J_{ij}) = 0, \quad \mathbb{E}(J_{ij}^2) = 1.$$

We give to the z variables an *a priori* unit Gaussian distribution, $d\mu(z) = d\mu(z_1)...d\mu(z_N)$, $d\mu(z_i) = (2\pi)^{-\frac{1}{2}} \exp(-z_i^2/2)$.

Then, according to the interpolation needs explained in [25], we define the random partition function

$$Z_N(\beta, J, \lambda) = \int d\mu(z) \exp\left[\beta \frac{1}{\sqrt{2N}} \sum_{i,j}^N J_{ij} z_i z_j - \frac{\beta^2}{4N} (\sum_{i=1}^N z_i^2)^2 + \frac{\lambda}{2} \sum_{i=1}^N z_i^2\right]. \tag{A.1}$$

Here, in the first piece of the *Boltzmannfaktor* we have the usual long range spin-spin interaction of the mean field spin glass model at inverse temperature β, extended to soft spins. The second piece, arises in a very natural way during the interpolation procedure between the analogical neural network and a couple of spin glasses, of dichotomic and soft nature, respectively [25]. This terms acts as an efficient smooth cut-off, preventing any divergence of the integral on soft spin at ∞. Notice that we have normalized the cut-off term so that in the annealing procedure, characterized by $\mathbb{E}Z_N(\beta, J, \lambda)$, where $\mathbb{E}$ are averages on the J variables, the contribution from the first term is exactly cancelled by the second term. But we can consider more general cases through a simple rescaling of the Gaussian variables. Finally, the parameter λ is a Lagrangian multiplier, inserted for the sake of convenience, in order to modify the scale of the soft spin, as it is sometime useful.

All the thermodynamic properties of the model are codified in the partition function, so that we can introduce the (quenched average of the) free energy per site $f_N(\beta)$, the Boltzmann state ω_J and the auxiliary function $A_N(\beta)$ (conventionally called the "pressure"), according to the definition

$$-\beta f_N(\beta) = A_N(\beta) = N^{-1}\mathbb{E}\log Z_N(\beta, J), \tag{A.2}$$

$$\omega_J(O) = Z_N^{-1}\int d\mu(z)O(z)\exp\big[-\beta H_N(z,J) - \frac{\beta^2}{4N}(\sum_{i=1}^N z_i^2)^2 + \frac{\lambda}{2}\sum_{i=1}^N z_i^2\big], \tag{A.3}$$

where O is a generic observable function of the z's. In the notation ω_J, we have stressed the dependence of the Boltzmann state on the external noise J, but, of course, there is also a dependence on β and N.
Let us now introduce the important concept of replicas. Consider a generic number s of independent copies of the system, characterized by the spin variables $z_i^{(1)}, \dots z_i^{(s)}$ distributed according to the product state

$$\Omega_J = \omega_J^{(1)} \dots \omega_J^{(s)}, \tag{A.4}$$

where all $\omega_J^{(\alpha)}$ act on each one $z_i^{(\alpha)}$'s, and are subject to the same sample J of the external noise. Finally, for a generic smooth function $F(z_i^{(1)}, \dots z_i^{(s)})$ of the replicated spin variables, we define the $\langle . \rangle$ average as

$$\left\langle F(z_i^{(1)}, \dots z_i^{(s)})\right\rangle = \mathbb{E}\Omega_J(F(z_i^{(1)}, \dots z_i^{(s)})). \tag{A.5}$$

We also define the overlap q between replicas:

$$q_{z^a z^b} = \frac{1}{N}\sum_{i=1}^N z_i^a z_i^b,$$

so that we can write

$$Z_N(\beta,\lambda,J) = \int d\mu(z)\exp\left(\beta\sqrt{\frac{N}{2}}\mathcal{K}(z) - \frac{\beta^2}{4N}(\sum_{i=1}^N z_i^2)^2 + \frac{\lambda}{2}\sum_{i=1}^N z_i^2\right), \tag{A.6}$$

where $\mathcal{K}(z)$ is a family of centered gaussian random variables with covariances $S_{zz'} = \mathbb{E}[\mathcal{K}(z)\mathcal{K}(z')] = q_{zz'}^2$ and the regularization term is just $\frac{1}{2}\frac{\beta^2 N}{2}q_{zz}^2 = \frac{1}{2}\frac{\beta^2 N}{2}S_{zz}$. Notice that the diagonal part of the variance is not trivial, since $q_{zz}^2 = \frac{1}{N^2}\left(\sum_{i=1}^N z^2\right)^2$.

A.1 Thermodynamic Limit

The aim of this section is to show how to get a rigorous control of the infinite volume limit of the free energy f_N (or similarly A_N). The main idea, inspired by [66], is to compare A_N, A_{N_1} and A_{N_2}, with $N = N_1 + N_2$. For this purpose we consider both the original N site system and two independent subsystems made of by N_1 and N_2 soft spins respectively, so to define

$$
\begin{aligned}
Z_N(t) = \quad \int d\mu(z) \quad & \exp\left(\beta\sqrt{\frac{t}{2N}}\sum_{i,j=1}^{N} J_{ij}z_iz_j - t\frac{\beta^2}{4N}(\sum_{i=1}^{N} z_i^2)^2\right) \\
& \exp\left(\beta\sqrt{\frac{1-t}{2N_1}}\sum_{i,j=1}^{N_1} J'_{ij}z_iz_j - (1-t)\frac{\beta^2}{4N_1}(\sum_{i=1}^{N_1} z_i^2)^2\right) \\
& \exp\left(\beta\sqrt{\frac{1-t}{2N_2}}\sum_{i,j=N_1+1}^{N} J''_{ij}z_iz_j - (1-t)\frac{\beta^2}{4N_2}(\sum_{i=N_1+1}^{N} z_i^2)^2\right) \\
& \exp\left(\frac{\lambda}{2}\sum_{i=1}^{N} z_i^2\right),
\end{aligned} \tag{A.7}
$$

with $0 \leq t \leq 1$. The partition function $Z_N(t)$ interpolates between the original N-spin model (obtained for $t = 1$) and the two subsystems (of sizes N_1 and N_2, obtained for $t = 0$) equipped with independent noises J' and J'', both independent of J. Notice that the quartic term does participate in the interpolation.

Now we follow the standard strategy, based on differentiation with respect to the interpolating parameter t, and witness an almost miraculous cancellation between terms coming from the differentiation of the quartic interaction and the diagonal part of the spin glass contribution. We omit the details, and state the following main result

Theorem A.1.1. *The following super-additivity property holds*

$$
NA_N(\beta, h) \geq N_1A_{N_1}(\beta, h) + N_2A_{N_2}(\beta, h). \tag{A.8}
$$

As it is very well known, the super-additivity property gives an immediate control of the thermodynamic limit [94], and we can state the next

Theorem A.1.2. *The thermodynamic limit for $A_N(\beta, h)$ exists and equals its* sup, *i.e.*

$$
\lim_{N\to\infty} A_N(\beta, h) = A(\beta, h) = \sup_N A_N(\beta, h). \tag{A.9}
$$

A.2 High Temperature behavior

We start to analyze our model characterizing the high temperature regime, at small β. First we define the annealed free energy of the model

$$-\beta f_N^A(\beta,\lambda) = A_N^A(\beta,\lambda) = \frac{1}{N}\log \mathbb{E} Z_N(\beta,\lambda,J), \tag{A.10}$$

that can be easily computed as in the following

Proposition A.2.1. *For $\lambda < 1$ the annealed free energy of the model in the thermodynamic limit is well defined and coincides with*

$$-\beta f^A(\beta,\lambda) = \lim_{N\to\infty} A_N^A(\beta,\lambda) = -\frac{1}{2}\log(1-\lambda). \tag{A.11}$$

Proof. It is enough to notice that

$$\mathbb{E}_J Z_N = \int d\mu(z) \exp\left(\frac{\lambda}{2}\sum_{i=1}^N z_i^2\right) = (1-\lambda)^{-\frac{N}{2}}. \tag{A.12}$$

This follows from the cancellation under annealing of the first term in the *Boltzmannfaktor* with the second. In fact, we have

$$\mathbb{E}\exp\left(\frac{\beta}{\sqrt{2N}}\sum_{i,j=1}^N J_{ij} z_i z_j\right) = \exp\left(\frac{\beta^2}{4N}\sum_{i,j=1}^N z_i^2 z_j^2\right) = \exp\left(\frac{\beta^2}{4N}(\sum_{i=1}^N z_i^2)^2\right).$$

Thus $(A.11)$ follows from $(A.10)$ and the proposition is proven. □

We define the high temperature regime as the region in the (β,λ) plane where the quenched free energy is equal to the annealed one. We already know that the annealed approximation is an upper bound for the pressure, in fact a simple application of the Jensen inequality shows that

$$\frac{1}{N}\mathbb{E}\log Z_N(\beta,\lambda;J) \le \frac{1}{N}\log\mathbb{E} Z_N(\beta,\lambda,J) = A_N^A(\beta,\lambda). \tag{A.13}$$

On the other side we have that

$$\frac{1}{N}\mathbb{E}\log Z_N(\beta,\lambda;J) \ge \frac{1}{N}\mathbb{E}\log Z'_N(\beta,\lambda;J), \tag{A.14}$$

where $Z'_N(\beta,\lambda;J)$ is an auxiliary partition function in which diagonal terms of the spin-spin interaction are neglected, *i.e.*

$$\begin{aligned} Z'_N(\beta,\lambda;J) &= \int d\mu(z)\exp\left(-\frac{1}{\sqrt{2N}}\sum_{i\neq j}^N J_{ij}z_i z_j - \frac{\beta^2}{4N}(\sum_{i=1}^N z_i^2)^2 + \frac{\lambda}{2}\sum_{i=1}^N z_i^2\right) \\ &= \int d\mu(z)\exp\left(-\frac{1}{\sqrt{N}}\sum_{i<j}^N J_{ij}z_i z_j - \frac{\beta^2}{4N}(\sum_{i=1}^N z_i^2)^2 + \frac{\lambda}{2}\sum_{i=1}^N z_i^2\right), \end{aligned}$$

where we have noted that $\frac{1}{\sqrt{2}}(J_{ij}+J_{ji})$ is a centered gaussian random variable $\mathcal{N}(0,1)$ that we have simply denoted by J_{ij}. Inequality (A.14) follows still from Jensen inequality on the J_{ii} noises:

$$\begin{aligned}\mathbb{E}\log Z_N(\beta,\lambda;J) &= \mathbb{E}_{J_{ij}}\mathbb{E}_{J_{ii}}\log Z_N(\beta,\lambda;J_{ij},J_{ii})\\ &\geq \mathbb{E}_{J_{ij}}\log Z_N(\beta,\lambda;J_{ij},\mathbb{E}_{J_{ii}}[J_{ii}])\\ &= \mathbb{E}_{J_{ij}}\log Z_N(\beta,\lambda;J_{ij},0)=\mathbb{E}\log Z'_N(\beta,\lambda;J).\end{aligned}$$

Note that the auxiliary partition function Z'_N gives the same annealed approximation of Z_N; in fact we have the following

Proposition A.2.2. *For $\lambda<1$,*

$$\lim_{N\to\infty}\frac{1}{N}\log\mathbb{E}_J Z'_N(\beta,\lambda;J)=-\frac{1}{2}\log(1-\lambda)=A^A(\beta,\lambda). \tag{A.15}$$

Proof.

$$\begin{aligned}\mathbb{E}Z'_N &= \int d\mu(z)\exp\left(\frac{\lambda}{2}\sum_{i=1}^N z_i^2-\frac{\beta^2}{4N}\sum_{i=1}^N z_i^4\right)\\ &= (1-\lambda)^{-\frac{N}{2}}\left(\int\frac{dz}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2-\frac{\beta^2}{4N(1-\lambda)^2}z^4}\right)^N,\end{aligned} \tag{A.16}$$

where the diagonal quartic term appears because it is not compensated in the annealing.

Now, putting $\beta_\lambda=\frac{\beta}{1-\lambda}$, we notice that the function in the integral

$$\int\frac{dz}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2-\frac{\beta_\lambda^2}{4N}z^4} \tag{A.17}$$

tends to 1 uniformly for $0\leq\lambda<1$, when N grows to infinity, and so the integral. That completes the proof. $\square$

Now, we can control the high temperature region of Z'_N studying the fluctuations of the random variable $Z'_N/\mathbb{E}Z'_N$, according to the Borel-Cantelli lemma approach [30][98]. The following lemma holds:

Lemma A.2.3. *For $\beta_\lambda=\frac{\beta}{1-\lambda}\leq 1$ we have*

$$\limsup_{N\to\infty}\frac{\mathbb{E}(Z'^2_N)}{\mathbb{E}^2(Z'_N)}\leq\frac{1}{\sqrt{1-\beta_\lambda^2}}. \tag{A.18}$$

The proof follows from a direct standard calculation along the line exploited in the case of the Sherrington-Kirkpatrick model, and will not be reported here in detail for the sake of conciseness. Lemma $A.2.3$ is a sufficient condition to state the following

Lemma A.2.4. *In the region of the (β, λ) plane defined by $\beta_\lambda < 1$, i.e. $\beta < 1 - \lambda$,*

$$\lim_{N\to\infty} \frac{1}{N} \log Z'_N(\beta, \lambda; J) = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{E} Z'_N(\beta, \lambda; J) = A^A(\beta, \lambda), \tag{A.19}$$

J-almost surely.

In fact, by following a standard procedure, it is enough to consider that for a sequence on non-negative random variables u_N, normalized to $\mathbb{E}u_N$, for which the second momenta are uniformly bounded $\mathbb{E}u_N^2 \leq c$, we have, by Borel-Cantelli lemma, that almost surely $\lim_{N\to\infty} N^{-1} \log u_N = 0$. In our case, we have to define $u_N = Z'_N/\mathbb{E}(Z'_N)$, and the rest follows smoothly.

Thanks to inequalities (A.13) and (A.14), we have proven the following main

Theorem A.2.5. *In the thermodynamic limit, J-almost surely, the free energy of the Gaussian spin glass model does coincide with the annealed one*

$$\lim_{N\to\infty} \frac{1}{N} \log Z_N(\beta, \lambda; J) = -\beta f(\beta, \lambda) = -\frac{1}{2} \log(1 - \lambda), \tag{A.20}$$

in the region of the (β, λ) plane defined by $\beta < 1 - \lambda$.

Therefore, by avoiding the diagonal terms in the interaction, which clearly do not suffer to be annealed, we can have a control of the ergodic region, by exploiting the usual method based on Borel-Cantelli lemma. However, a complete control of the ergodic region can be also achieved by using the strategy developed in Section 5.

A.3 The replica symmetric form for the free energy

In this section we introduce the replica symmetric expression for the free energy density, and give a sum rule connecting it with the true free energy together with an error term of definite sign. For this purpose, we apply the well known interpolation scheme [64][59][23][26] to compare the original two-body interaction with a one-body interaction system. Concretely, we define, for $t \in [0, 1]$ and a generic parameter $\bar{q} \geq 0$, which will be recognized after the optimization as the self-averaged overlap, the interpolating function

$$\begin{aligned}\varphi_N(t) &= \frac{1}{N}\mathbb{E}\log\int d\mu(z)\exp\left(\beta\sqrt{t}\sqrt{\frac{N}{2}}\mathcal{K}(z)-t\frac{\beta^2}{4N}(\sum_{i=1}^N z_i^2)^2\right)\\ &\quad \exp\left(\beta\sqrt{1-t}\sqrt{\bar{q}}\sum_{i=1}^N J'_i z_i-(1-t)\frac{\beta^2\bar{q}}{2}\sum_{i=1}^N z_i^2\right)\\ &\quad \exp\left(\frac{\lambda}{2}\sum_{i=1}^N z_i^2\right).\end{aligned} \tag{A.21}$$

Here the external cavity fields on each site J'_i are i.i.d. unit Gaussian random variables, also independent from all J_{ij}. We encode in $\mathbb{E}$ the averages with respect to both J and J'. At $t=1$ the interpolating function (A.21) recovers the original system, while at $t=0$ it accounts for a simpler factorized one-body model and we can easily get through a simple calculation

$$\varphi_N(0)=\log(\sigma)+\frac{1}{2}\beta^2\bar{q}\sigma^2, \tag{A.22}$$

with $\sigma=(1-\lambda+\beta^2\bar{q})^{-\frac{1}{2}}$. Therefore $\varphi_N(t)$ fulfills the following boundary conditions:

$$\begin{aligned}\varphi_N(1) &= A_N(\beta,\lambda),\\ \varphi_N(0) &= \log(\sigma)+\frac{1}{2}\beta^2\bar{q}\sigma^2.\end{aligned} \tag{A.23}$$

Now we proceed according to the usual strategy, by evaluating the derivative with respect to t, and then integrating in the interval $[0,1]$. We use the notation $\langle . \rangle_t = \mathbb{E}\Omega_t(.)$, where $\Omega_t(.)$ is the replicated Boltzmann state according to the interpolating system appearing in $(A.21)$. By taking the t derivative, we obtain

$$\frac{d}{dt}\varphi_N(t)=\frac{\beta^2}{4}\bar{q}^2-\frac{\beta^2}{4}\left\langle (q_{12}-\bar{q})^2\right\rangle_t, \tag{A.24}$$

and by integration:

Theorem A.3.1. *For every $\bar{q}\in\mathcal{D}_{\beta,\lambda}\equiv\left\{\bar{q}\in\mathbb{R}^+ : 1-\lambda+\beta^2\bar{q}>0\right\}$ let us define the trial function*

$$\tilde{A}(\beta,\lambda,\bar{q})=\log(\sigma)+\frac{1}{2}\beta^2\bar{q}\sigma^2+\frac{\beta^2}{4}\bar{q}^2, \tag{A.25}$$

with $\sigma=(1-\lambda+\beta^2\bar{q})^{-\frac{1}{2}}$. Then, $\forall N$ and $\forall\bar{q}\in\mathcal{D}_{\beta,\lambda}$, the quenched free energy of the mean field gaussian spin glass model defined in $(A.1)$ fulfills the sum rule

$$A_N(\beta,\lambda)=-\beta f_N(\beta,\lambda)=\tilde{A}(\beta,\lambda,\bar{q})-\frac{\beta^2}{4}\int_0^1 dt\left\langle (q_{12}-\bar{q})^2\right\rangle_t. \tag{A.26}$$

Moreover, $\forall\bar{q}\in\mathcal{D}_{\beta,\lambda}$, $\tilde{A}(\beta,\lambda,\bar{q})$ is an upper bound for $A_N(\beta,\lambda)$ uniformly in N, i.e.

$$A_N(\beta,\lambda)=-\beta f_N(\beta,\lambda)\le\tilde{A}(\beta,\lambda,\bar{q}). \tag{A.27}$$

Since the bound $(A.27)$ is uniform in N, then it is true also in the thermodynamic limit. The error term in $(A.26)$ reduces to the overlap's fluctuations around $\bar{q}$. We can minimize this error, or equivalently optimize the estimate in $(A.27)$, by taking the value of $\bar{q}$ that minimize $\tilde{A}(\beta,\lambda,\bar{q})$. For this purpose we state the following

Proposition A.3.2. *The minimum for $\tilde{A}(\beta,\lambda,\bar{q})$, as a function of $\bar{q}$, is reached at $\bar{q}=0$ for $\beta\leq 1-\lambda$, while for $\beta>1-\lambda$ the minimum is $\bar{q}=\frac{\beta-(1-\lambda)}{\beta^2}$.*

Proof. We study $\tilde{A}(\beta,\lambda,\bar{q})$ as a function of $\bar{q}^2$. A simple calculation gives

$$\frac{\partial}{\partial\bar{q}^2}\tilde{A}(\beta,\lambda,\bar{q})=\frac{1}{2\bar{q}}\frac{\partial}{\partial\bar{q}}\tilde{A}(\beta,\lambda,\bar{q})=\frac{\beta^2}{4}\left(1-\frac{\beta^2}{(1-\lambda+\beta^2\bar{q})^2}\right).$$

Therefore $\frac{\partial}{\partial\bar{q}^2}\tilde{A}(\beta,\lambda,\bar{q})$ is increasing, and $\tilde{A}$ is a convex function of $\bar{q}^2$. At $\bar{q}=0$ we have that

$$\frac{\partial}{\partial\bar{q}^2}\tilde{A}(\beta,\lambda,\bar{q}^2)|_{\bar{q}=0}=\frac{\beta^2}{4}\left(1-\frac{\beta^2}{(1-\lambda)^2}\right)=\frac{\beta^2}{4}\left(1-\beta_\lambda^2\right).$$

Due to the convexity of $\tilde{A}$ in $\bar{q}^2$, the minimum is achieved at $\bar{q}=0$ for $\beta_\lambda<1$ and at $\bar{q}=\frac{\beta-(1-\lambda)}{\beta^2}$ for $\beta_\lambda>1$, where the derivative has a zero. □

By combining the information of Theorem $A.3.1$ and Proposition $A.3.2$ we have the proof of the following important result.

Theorem A.3.3. *The replica symmetric expression for the free energy is well defined by the following variational principle:*

$$A^{RS}(\beta,\lambda)=\inf_{\bar{q}\in\mathcal{D}_{\beta,\lambda}}\tilde{A}(\beta,\lambda,\bar{q}),\tag{A.28}$$

where

$$\tilde{A}(\beta,\lambda,\bar{q})=\log(\sigma)+\frac{1}{2}\beta^2\bar{q}\sigma^2+\frac{\beta^2}{4}\bar{q}^2,\tag{A.29}$$

with $\sigma(\beta,\lambda,\bar{q})$ defined in $(A.25)$. The minimum is achieved at $\bar{q}=0$ for $\beta\leq 1-\lambda$ and at $\bar{q}=\frac{\beta-(1-\lambda)}{\beta^2}$ otherwise. Moreover the replica symmetric approximation is an upper bound for $A(\beta,\lambda)$, in fact, uniformly in N,

$$A_N(\beta,\lambda)=-\beta f_N(\beta,\lambda)\leq A^{RS}(\beta,\lambda).\tag{A.30}$$

Notice that at the optimal point $\bar{q}=\frac{\beta-(1-\lambda)}{\beta^2}$ we have $\beta\sigma^2=1$.

For $\beta_\lambda<1$ the replica symmetric free energy reduces to the annealed one, that, accordingly with Theorem $A.2.4$, coincides with the thermodynamic limit of the true free energy in such a region. Note that $\bar{q}=\frac{\beta-(1-\lambda)}{\beta^2}$ is also the optimal value for $\lambda>1$, in fact $1-\lambda+\beta^2\bar{q}=\beta>0$ such that $\bar{q}\in\mathcal{D}_{\beta,\lambda}$ and the RS expression is well defined. In this case we see that $\bar{q}\to\infty$ when $\beta\to 0$.

A.4 Fully Broken Replica Symmetry

The Gaussian model is unique, in that it allows to explicitly calculate the fully broken replica symmetry trial functional, which should give an improvement on the replica symmetric bound for the free energy density. As a matter of fact, as a consequence of an elementary sum rule, it will be shown that the fully broken replica symmetry variational principle gives the same result as the replica symmetric functional. Replica symmetry breaking is not realized in the Gaussian case. We give some details about the procedure, since it can be generalized beyond the simple model considered here, as for example in the case of the coupling of the Sherrington-Kirkpatrick interaction with a neural network and a Gaussian spin glass, as shown in [25], and work in preparation.

First of all we introduce the convex space $\mathcal{X}$ of functional order parameters x, as nondecreasing functions of the auxiliary variable q in the $[0,1]$ interval, i.e.

$$\mathcal{X} \ni x : [0,Q] \ni q \to x(q) \in [0,1], \tag{A.31}$$

We have to think x as connected at the end to the distribution function for the overlap. We will consider the case of piecewise constant functional order parameters, characterized by an integer K and two sequence of numbers, $q_0, q_1, \ldots, q_K$ and $m_1, \ldots, m_K$, satisfying

$$0 = q_0 \leq q_1 \ldots \leq q_K = Q, \quad 0 \leq m_1 \ldots \leq m_K \leq 1, \tag{A.32}$$

such that $x(q) = m_i$ for $q \in [q_{i-1}, q_i]$. It is useful to define also $m_0 = 0$ and $m_{K+1} = 1$. The replica symmetric case correspond to $K = 2$, $q_1 = \bar{q}$, $m_1 = 0$ and $m_2 = 1$, where overlap self-averages around $\bar{q}$; the case $K = 3$, with two possible value (q_1 and q_2) for the overlap, is the first level of replica symmetry breaking, and so on. Now, following the interpolation scheme in [65], we consider a generic piecewise constant x and we introduce the interpolating partition function

$$\begin{aligned}
\tilde{Z}_N(t;x) &= \int d\mu(z) \exp\left(\beta\sqrt{\frac{t}{2N}}\sum_{i,j=1}^{N} J_{ij} z_i z_j - t\frac{\beta^2}{4N}(\sum_{i=1}^{N} z_i^2)^2\right) \\
&\quad \exp\left(\beta\sqrt{1-t}\sum_{a=1}^{K}\sqrt{q_a - q_{a-1}}\sum_{i=1}^{N} J_i^a z_i - (1-t)\frac{\beta^2 Q}{2}\sum_{i=1}^{N} z_i^2\right) \\
&\quad \exp\left(\beta h\sum_{i=1}^{N} z_i + \frac{\lambda}{2}\sum_{i=1}^{N} z_i^2\right),
\end{aligned} \tag{A.33}$$

where $t \in [0,1]$. Here we have introduced additional independent unit gaussian random variables J_i^a, $a = 1, \ldots, K$, $i = 1, \ldots, N$. Let us call $\mathbb{E}_a$ the average with respect to all the random variables J_i^a, $i = 1, \dots, N$ and $\mathbb{E}_0$ the average with respect to all the J_{ij}. We denote with $\mathbb{E}$ the average with respect to all the J. Now we define recursively the random variables

$$Z_K = \tilde{Z}_N; \;\; Z_{K-1} = (\mathbb{E}_K Z_K^{m_k})^{\frac{1}{m_k}}; \;\; \dots; \;\; Z_0 = (\mathbb{E}_1 Z_1^{m_1})^{\frac{1}{m_1}}, \tag{A.34}$$

where each Z_a depends only on the external noise J_{ij} and on the J_i^b for $b \leq a$. Finally we define the auxiliary interpolating function

$$\varphi_N(t; x) = \frac{1}{N} \mathbb{E}_0 \log Z_0(t; x(q)), \tag{A.35}$$

that is completely averaged out with respect of all the external noises. Notice that, at $t = 1$, we recover the original $A_N(\beta, \lambda)$, while, at $t = 0$, we have a factorized expression in terms of a solvable one body interaction problem. Thus, we have the possibility to find a sum rule for the free energy in the fully broken replica case through

$$A_N(\beta, \lambda) = \varphi_N(0, x) + \int_0^1 dt \frac{d}{dt} \varphi_N(t), \tag{A.36}$$

after calculating the t-derivative of $\varphi_N(t, x)$. For this purpose we need some additional definitions. Let us introduce the random variables

$$f_a = \frac{Z_a^{m_a}}{\mathbb{E}_a Z_a^{m_a}}, \qquad a = 1, \dots, K, \tag{A.37}$$

and notice that they depend only on the J_i^b for $b \leq a$ and they are normalized, $\mathbb{E} f_a = 1$. Moreover we consider the t-dependent state ω associated to the *Boltzmannfaktor* defined in $(A.33)$ and its replicated Ω. A very important rule is played by the following states $\tilde{\omega}_a$, with $a = 1, \dots, K$, and their replicated $\tilde{\Omega}_a$, defined as

$$\tilde{\omega}_K(.) = \omega(.); \quad \tilde{\omega}_a = \mathbb{E}_{a+1} \dots \mathbb{E}_K (f_{a+1} \dots f_K \omega(.)). \tag{A.38}$$

Finally we define the generalized $\langle . \rangle_a$ average as

$$\langle . \rangle_a = \mathbb{E}(f_1 \dots f_a \tilde{\Omega}_a(.)). \tag{A.39}$$

We now proceed exactly as in the Sherrington-Kirkpatrick case [65], and reach the following

Theorem A.4.1. *The t-derivative of $\varphi_N(t)$, defined in $(A.35)$, is given by*

$$\begin{aligned} \frac{d}{dt} \varphi_N(t) &= \frac{\beta^2}{4} \sum_{a=1}^{K} (m_{a+1} - m_a) q_a^2 \\ &- \frac{\beta^2}{4} \sum_{a=1}^{K} (m_{a+1} - m_a) \left\langle (q_{12} - q_a)^2 \right\rangle_a . \end{aligned} \tag{A.40}$$

Theorem A.4.2. *In the thermodynamic limit, for every functional order parameter x of the type $(A.32)$, the following sum rule holds*

$$\begin{aligned} A(\beta,\lambda) &= \varphi(0;x)+\frac{\beta^2}{4}\sum_{a=1}^{K}(m_{a+1}-m_a)q_a^2 \\ &- \frac{\beta^2}{4}\sum_{a=1}^{K}(m_{a+1}-m_a)\int_0^1 \left\langle (q_{12}-q_a)^2\right\rangle_a dt, \end{aligned} \tag{A.41}$$

and, consequently, we have the following bound for the free energy density:

$$-\beta f(\beta,\lambda)=A(\beta,\lambda)\leq \varphi(0;x)+\frac{\beta^2}{4}\sum_{a=1}^{K}(m_{a+1}-m_a)q_a^2. \tag{A.42}$$

Clearly, Theorem A.4.2 follows from Theorem A.4.1 by taking into account $(A.36)$ and noting that the error term, containing overlap fluctuations around every q_a, has a definite sign. Now we give the expression for $\varphi_N(0;x)$, as in the following

Theorem A.4.3. *For any choice of the piecewise functional order parameter x, the initial condition $\varphi_N(0;x)$ is given by*

$$\varphi_N(0;x)=\log\sigma(Q)+f(0,0;x), \tag{A.43}$$

where $f(q,y;x)$ is the solution of the Parisi equation, i.e. the nonlinear anti-parabolic partial differential equation

$$\partial_q f(q,y)+\frac{1}{2}\big(f''(q,y)+x(q){f'}^2(q,y)\big)=0, \tag{A.44}$$

with final condition at $q=Q$

$$f(Q,y)=\frac{\beta^2}{2}\sigma^2(Q)y^2, \tag{A.45}$$

and $\sigma(Q)=(1-\lambda+\beta^2Q)^{-\frac{1}{2}}$, with the obvious restriction on Q to have a positive $\sigma(Q)$.

Proof. Since the *Boltzmannfaktor* factorizes at $t=0$, we have that

$$\begin{aligned} \tilde{Z}_N(0;x) &= \int d\mu(z)\exp\left(\frac{(\lambda-\beta^2Q)}{2}\sum_{i=1}^{N}z_i^2\right)\exp\left(\beta\sum_{a=1}^{K}\sqrt{q_a-q_{a-1}}\sum_{i=1}^{N}J_i^a z_i\right) \\ &= \prod_{i=1}^{N}\sigma(Q)\exp\left(\frac{\beta^2}{2}\sigma^2(Q)(\sum_{a=1}^{K}\sqrt{q_a-q_{a-1}}J_i^a)^2\right) \\ &\equiv \prod_{i=1}^{N}\sigma(Q)\exp\left(f(Q,\sum_{a=1}^{K}\sqrt{q_a-q_{a-1}}J_i^a)\right). \end{aligned} \tag{A.46}$$

From the definition (A.35) of the interpolating function $\varphi_N(t;x)$, we note that, due to the $1/N$ factor, we can evaluate the (A.46) on a single site only. The $\sigma(Q)$ goes to form the $\log\sigma(Q)$

term and what remains is just the solution of the Parisi equation, evaluated at $y = 0$, and propagated from $q = Q$ to $q = 0$ through a series of gaussian integration as in [65]. $\square$

Due to the Gaussian character of all integrations involved in this procedure, we can exactly solve equation (A.44) with final condition (A.45) to find $f(0, 0; x)$ and so $\varphi_N(0; x)$. In fact we give the following

Lemma A.4.4. *For any functional order parameter $x \in \mathcal{X}$, the solution of equation (A.44) with final condition (A.45), evaluated at $y = 0$ and $q = 0$ is given by*

$$f(0,0;x) = \frac{1}{2}\beta^2\sigma^2(Q)\int_0^Q \frac{dq}{1 - \beta^2\sigma^2(Q)\int_q^Q x(q')dq'}. \tag{A.47}$$

Proof. We look for a solution of (A.44) of the form $f(q, y) = a(q) + \frac{1}{2}b(q)y^2$. Since f must fulfill final condition (A.45), it has to be $a(Q) = 0$ and $b(Q) = \beta^2\sigma^2(Q)$. From

$$\begin{aligned} & \partial_q f(q,y) + \frac{1}{2}\big(f''(q,y) + x(q)f'^2(q,y)\big) \\ = \ & a'(q) + \frac{1}{2}b(q) + \frac{1}{2}y^2\left(b'(q) + x(q)b^2(q)\right), \end{aligned} \tag{A.48}$$

we see that $f(q, y)$ is a solution of (A.44) if $a(q)$ and $b(q)$ are solutions of the ordinary differential equation system

$$a'(q) + \frac{1}{2}b(q) = 0 \tag{A.49}$$

$$b'(q) + x(q)b^2(q) = 0, \tag{A.50}$$

with final conditions $a(Q) = 0$ and $b(Q) = \beta^2\sigma^2(Q)$. Integrating equation (A.50) we obtain

$$\frac{1}{b(q)} = \frac{1}{\beta^2\sigma^2(Q)} - \int_q^Q x(q')dq'. \tag{A.51}$$

Putting (A.51) into equation (A.49) and integrating, we have the proof. $\square$

Finally, from the continuity of $f(q, y; x)$ with respect to the choice of the functional order parameter x (see [65], [63]) and noticing that

$$\frac{\beta^2}{4}\sum_{a=1}^{K}(m_{a+1} - m_a)q_a^2 = \frac{\beta^2}{4}Q^2 - \frac{\beta^2}{2}\int_0^Q qx(q)dq, \tag{A.52}$$

we can use Theorem A.4.2 for stating our first result

Theorem A.4.5. *The pressure of the mean field gaussian spin glass model is bounded by:*

$$A(\beta,\lambda) \leq \inf_{x\in\mathcal{X}} \hat{A}(\beta,\lambda;x), \tag{A.53}$$

with

$$\begin{aligned}\hat{A}(\beta,\lambda;x) &= \log\sigma(Q) + \frac{1}{2}\beta^2\sigma^2(Q)\int_0^Q \frac{dq}{1-\beta^2\sigma^2(Q)\int_q^Q x(q')dq'} \\ &+ \frac{\beta^2}{4}Q^2 - \frac{\beta^2}{2}\int_0^Q qx(q)dq. \end{aligned}\tag{A.54}$$

Moreover the infimum is attained exactly at the RS functional order parameter $x=0$, $0\leq q<q^{RS}=\bar{q}$, $x=1$ elsewhere.

Proof. The bound is a direct consequence of all the results in this section. So we will focus our attention only on its last part, that is

$$A^{RS}(\beta,\lambda) \leq \inf_{x\in\mathcal{X}} \hat{A}(\beta,\lambda;x).$$

Let us notice that, without any loss of generality, we can assume Q as large as we like, in particular $Q\geq\bar{q}$, where $\bar{q}$ is the replica symmetric expression, with value $\bar{q}=\frac{\beta-(1-\lambda)}{\beta^2}$ outside of the ergodic region.. In fact, let us write $\hat{A}(\beta,\lambda;x)$ in the following equivalent form

$$\begin{aligned}\hat{A}(\beta,\lambda;x) &= \log\sigma(Q) + \frac{1}{2}\beta^2\int_0^Q \frac{dq}{1-\lambda+\beta^2 Q-\beta^2\int_q^Q x(q')dq'} \\ &+ \frac{\beta^2}{4}Q^2 - \frac{\beta^2}{2}\int_0^Q qx(q)dq. \end{aligned}\tag{A.55}$$

If we take $Q'>Q$, and define $x'(q)=x(q)$ for $0\leq q\leq Q$, and $x'(q)=1$ for $Q<q\leq Q'$, we can immediately check the equality

$$\hat{A}(\beta,\lambda;x') = \hat{A}(\beta,\lambda;x).$$

It is enough to split the integral $\int_0^{Q'} dq$ in the definition of $\hat{A}(\beta,\lambda;x')$ as a sum $\int_0^Q dq+\int_Q^{Q'} dq$, and take into account the definition of x'. Then we have the following sum rule, connecting a generic broken replica trial $\hat{A}(\beta,\lambda;x)$, with the optimal replica symmetric expression, corresponding to $\bar{x}(q)=0$ for $0\leq q\leq\bar{q}$, and $\bar{x}(q)=1$ for $\bar{q}<q\leq Q$,

$$\hat{A}(\beta,\lambda;x) = \hat{A}(\beta,\lambda;\bar{x}) + R,$$

with a nonnegative R. The proof is elementary, and can be easily obtained by splitting the integral $\int_0^Q dq$ in the definition of $\hat{A}(\beta,\lambda;x)$ as a sum $\int_0^{\bar{q}} dq+\int_{\bar{q}}^Q dq$, and taking into account the definition of $\bar{x}$. This shows that the replica symmetry breaking does not improve the optimization. Of course the error term R becomes zero when $x=\bar{x}$. □

The fact that symmetry breaking does not improve the result of the replica symmetric expression has very far reaching consequences. In fact, we have the following main result

Theorem A.4.6. *In the infinite volume limit the pressure of the mean field gaussian spin glass model is given by its replica symmetric expression*

$$A(\beta,\lambda) = A^{RS}(\beta,\lambda) \le \inf_{x\in\mathcal{X}} \hat{A}(\beta,\lambda;x), \tag{A.56}$$

where $A^{RS}(\beta,\lambda)$ *is defined in Theorem A.3.3.*

Proof. We follow a very brilliant strategy developed by Talagrand (see for example Chap. 2.11 in [98]), in order to give the precise boundary of the replica symmetric region in the Sherrington-Kirkpatrick mean field spin glass model. This strategy is based on the consideration of two coupled replicas with a generic fixed overlap constraint. For the pressure of the system with the two coupled replicas it is possible to develop broken replica symmetry bounds, as a generalization of those introduced in [65]. The final result is that the boundary region where replica symmetry holds is given by the occurence of a lowering of the pressure, with respect to its replica symmetric value, for a trial order parameter of the type $x(q) = 0$ for $0 \le q \le \bar{q}$, $x(q) = m$, for $\bar{q} < q \le \tilde{q}$, $x(q) = 1$ for $\tilde{q} < q \le 1$, where $\bar{q}$ is the replica symmetric overlap, and $m, \tilde{q}$ are suitable parameters, allowing to lower the pressure. In other words, the absence of a one level replica symmetry breaking assures that we are in the replica symmetric region. In our case, we can follow Talagrand procedure and conclude the validity of the replica symmetric expression for the pressure, for any value of the involved parameters, since we have already shown that there can be no lowering of the pressure by any symmetry breaking *Ansatz*. Notice that this proof develops completely inside the general frame given by the interpolation procedure, and the resultant broken replica symmetry bounds introduced in [65]. □

These results are in full agreement with those found by Ben Arous, Dembo and Guionnet in [31], where they exploit the rotational symmetry of the model and employ a clever method of large deviations, in a purely probabilistic setting.

Appendix B

Free energy evaluation in the high storage regime

In this appendix we calculate the free energy per spin of the system characterised by the Hamiltonian (2.2), within the replica-symmetric (RS) ansatz, for the scaling regime $P = \alpha N$. Let us start by introducing the partition function $Z_N(\beta, \xi)$ and the disorder-averaged free energy $\overline{f}$:

$$
\begin{aligned}
Z_N(\beta,\xi) &= \sum_{\boldsymbol{\sigma}} \mathrm{e}^{\frac{1}{2}\beta N^{-\tau}\sum_{i,j=1}^{N}\sum_{\mu=1}^{P}\xi_i^\mu \xi_j^\mu \sigma_i \sigma_j} && \text{(B.1)}\\
\overline{f} &= -\lim_{N\to\infty}\frac{1}{\beta N}\overline{\log Z_N(\beta,\xi)}, && \text{(B.2)}
\end{aligned}
$$

where $\overline{\cdots}$ denotes averaging over the randomly generated $\{\xi_i^\mu\}$. If we use the replica identity $\overline{\log Z} = \lim_{n\to 0} n^{-1}\log \overline{Z^n}$, and separate the contributions from the K condensed patterns from those of the $\alpha N - K$ non-condensed ones we get

$$
\begin{aligned}
\overline{f} &= -\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{\beta n N}\log \sum_{\boldsymbol{\sigma}^1,\cdots,\boldsymbol{\sigma}^n} \overline{\mathrm{e}^{\frac{1}{2}\beta N^{-\tau}\sum_{i,j=1}^{N}\sum_{\mu=1}^{P}\sum_{\alpha=1}^{n}\xi_i^\mu\xi_j^\mu\sigma_i^\alpha\sigma_j^\alpha}}\\
&= -\frac{1}{\beta}\log 2 - \lim_{N\to\infty}\lim_{n\to 0}\frac{1}{\beta n N}\log\Big\langle \mathrm{e}^{\frac{1}{2}\beta N^{-\tau}\sum_{\mu=1}^{K}\sum_{\alpha=1}^{n}(\sum_{i=1}^{N}\xi_i^\mu\sigma_i^\alpha)^2}\\
&\qquad\times\overline{\mathrm{e}^{\frac{1}{2}\beta N^{-\tau}\sum_{\mu>K}^{P}\sum_{\alpha=1}^{n}(\sum_{i=1}^{N}\xi_i^\mu\sigma_i^\alpha)^2}}\Big\rangle_{\boldsymbol{\sigma}^1,\ldots,\boldsymbol{\sigma}^n}. && \text{(B.3)}
\end{aligned}
$$

We compute the non-condensed contributions first, using the standard tool of Gaussian linearisation, and the usual short-hands $\mathrm{D}z=(2\pi)^{-1/2}\mathrm{e}^{-z^2/2}\mathrm{d}z$ and $\mathrm{D}\mathbf{z}=\prod_{\alpha=1}^n \mathrm{D}z_\alpha$:

$$
\begin{aligned}
\Xi &= \overline{\mathrm{e}^{\frac{1}{2}\beta N^{-\tau}\sum_{\mu>K}\sum_{\alpha=1}^n(\sum_{i=1}^N \xi_i^\mu\sigma_i^\alpha)^2}} = \Big[\overline{\mathrm{e}^{\frac{1}{2}\beta N^{-\tau}\sum_{\alpha=1}^n(\sum_{i=1}^N \xi_i\sigma_i^\alpha)^2}}\Big]^{P-K} \\
&= \Big[\int \mathrm{D}\mathbf{z}\ \overline{\mathrm{e}^{\sqrt{\beta}N^{-\tau/2}\sum_{\alpha=1}^n z_\alpha\sum_{i=1}^N\xi_i\sigma_i^\alpha}}\Big]^{P-K} \\
&= \Big[\int \mathrm{D}\mathbf{z}\ \prod_{i=1}^N\Big(1-cN^{-\gamma}+cN^{-\gamma}\cosh(\sqrt{\beta}N^{-\tau/2}\sum_{\alpha=1}^n\sigma_i^\alpha z_\alpha)\Big)\Big]^{P-K} \\
&= \Big[\int \mathrm{D}\mathbf{z}\ \prod_{i=1}^N\Big[1+\frac{1}{2}\beta cN^{-\gamma-\tau}\big(\sum_{\alpha=1}^n\sigma_i^\alpha z_\alpha\big)^2+\mathcal{O}(N^{-2\tau-\gamma})\Big]\Big]^{P-K} \\
&= \Big[\int \mathrm{D}\mathbf{z}\ \mathrm{e}^{\frac{1}{2}\beta cN^{-\gamma-\tau}\sum_{\alpha,\beta=1}^n z_\alpha z_\beta\sum_{i=1}^N\sigma_i^\alpha\sigma_i^\beta+\mathcal{O}(N^{1-2\tau-\gamma})}\Big]^{P-K}.
\end{aligned}
\tag{B.4}
$$

Now it is evident, as in our earlier calculations, that the correct scaling for large N requires choosing $\tau=1-\gamma$. For the correction term in the exponent this gives $\mathcal{O}(N^{1-2\tau-\gamma})=\mathcal{O}(N^{\gamma-1})$, which is indeed vanishing since $\gamma<1$. We now arrive at

$$
\Xi = \exp\Big\{(P-K)\log\int \mathrm{D}\mathbf{z}\ \mathrm{e}^{\frac{1}{2}\beta cN^{-1}\sum_{\alpha,\beta=1}^n z_\alpha z_\beta\sum_{i=1}^N\sigma_i^\alpha\sigma_i^\beta+\mathcal{O}(N^{\gamma-1})}\Big\}
\tag{B.5}
$$

We next introduce n^2 parameters $\{q_{\alpha\beta}\}$ and their conjugates $\{\hat{q}_{\alpha\beta}\}$, by inserting partitions of unity:

$$
1=\prod_{\alpha\beta}\int \mathrm{d}q_{\alpha\beta}\ \delta\big(q_{\alpha\beta}-\frac{1}{N}\sum_{i=1}^N\sigma_i^\alpha\sigma_i^\beta\big)=\int\Big[\prod_{\alpha\beta}\frac{\mathrm{d}q_{\alpha\beta}\mathrm{d}\hat{q}_{\alpha\beta}}{2\pi/N}\Big]\mathrm{e}^{\mathrm{i}N\sum_{\alpha,\beta}\hat{q}_{\alpha\beta}(q_{\alpha\beta}-\frac{1}{N}\sum_i\sigma_i^\alpha\sigma_i^\beta)}.
\tag{B.6}
$$

Substituting (B.6) into (B.5) gives the contribution to the partition function of non-condensed patterns:

$$
\begin{aligned}
\Xi = \int\Big[\prod_{\alpha\beta}\mathrm{d}q_{\alpha\beta}\mathrm{d}\hat{q}_{\alpha\beta}\Big]&\mathrm{e}^{\mathrm{i}N\sum_{\alpha,\beta}\hat{q}_{\alpha\beta}q_{\alpha\beta}+(P-K)\log\int\mathrm{D}\mathbf{z}\ \mathrm{e}^{\frac{1}{2}\beta c\sum_{\alpha,\beta=1}^n z_\alpha q_{\alpha\beta}z_\beta}+\mathcal{O}(N^\gamma)} \\
&\times\mathrm{e}^{-\mathrm{i}\sum_i\sum_{\alpha,\beta}\sigma_i^\alpha\hat{q}_{\alpha\beta}\sigma_i^\beta}.
\end{aligned}
\tag{B.7}
$$

The contribution from condensed pattern, see (B.3), is

$$
\mathrm{e}^{\frac{1}{2}\beta N^{\gamma-1}\sum_{\mu\le K}\sum_{\alpha=1}^n(\sum_{i=1}^N\xi_i^\mu\sigma_i^\alpha)^2}=\int\mathbf{Dm}\ \mathrm{e}^{\sqrt{\beta}N^{(\gamma-1)/2}\sum_{\mu\le K}\sum_{\alpha=1}^n\sum_{i=1}^N\xi_i^\mu\sigma_i^\alpha m_\alpha^\mu},
\tag{B.8}
$$

with $\mathbf{m}=\{m_\alpha^\mu\}\in\mathbb{R}^{nK}$. If we rescale $m_\alpha^\mu\to c\sqrt{\beta}N^{(1-\gamma)/2}m_\alpha^\mu$ this becomes

$$
\Big(c^2\beta N^{1-\gamma}\Big)^{nK/2}\int \mathrm{d}\mathbf{m}\ \mathrm{e}^{-\frac{1}{2}\beta c^2N^{1-\gamma}\mathbf{m}^2+\beta c\sum_{\mu\le K}\sum_{\alpha=1}^n\sum_{i=1}^N\xi_i^\mu\sigma_i^\alpha m_\alpha^\mu}.
\tag{B.9}
$$

Inserting (B.7,B.9) into (B.3) gives the following expression for the free energy per spin:

$$
\begin{aligned}
\overline{f} \;=\; & -\frac{1}{\beta}\log 2 - \lim_{N\to\infty}\Big\{\frac{K/2}{\beta N}\log(c^2\beta N^{1-\gamma}) \\
& + \lim_{n\to 0}\frac{1}{\beta n N}\log\int \mathrm{d}\mathbf{m}\Big[\prod_{\alpha\beta}\mathrm{d}q_{\alpha\beta}\mathrm{d}\hat{q}_{\alpha\beta}\Big] \\
& \times \mathrm{e}^{N\Big[\mathrm{i}\sum_{\alpha,\beta}\hat{q}_{\alpha\beta}q_{\alpha\beta}+\frac{P-K}{N}\log\int \mathrm{D}\mathbf{z}\ \mathrm{e}^{\frac{1}{2}\beta c\sum_{\alpha,\beta=1}^{n} z_\alpha q_{\alpha\beta} z_\beta}-\frac{1}{2}\beta c^2 N^{-\gamma}\mathbf{m}^2\Big]} \\
& \times\prod_{i=1}^{N}\Big\langle \mathrm{e}^{\beta c\sum_{\mu\le K}\sum_{\alpha=1}^{n}\xi_i^\mu\sigma_i^\alpha m_\alpha^\mu-\mathrm{i}\sum_{\alpha,\beta}\sigma_i^\alpha\hat{q}_{\alpha\beta}\sigma_i^\beta}.\Big\rangle_{\sigma_i^1,\ldots,\sigma_i^n}\Big\}
\end{aligned}
\tag{B.10}
$$

The number of order parameters being integrated over is of order K, so corrections to the saddle-point contribution will be of order $\mathcal{O}(K\log N/N)$. To proceed via steepest descent we must therefore impose $K \ll N/\log N$. Since also the energy term $N^{-\gamma}\sum_{\mu\le K}\mathbf{m}^2$ should be of order one, as well as the individual components of $\mathbf{m}$, the only natural choice is $K=\mathcal{O}(N^\gamma)$. Under this scaling condition we then find

$$
\overline{f} \;=\; -\frac{1}{\beta}\log 2 - \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{\beta n}\mathrm{extr}_{\mathbf{m},q,\hat{q}}\hat{f}(\mathbf{m},\{q,\hat{q}\}) \tag{B.11}
$$

with

$$
\hat{f}(\mathbf{m},\{q,\hat{q}\}) \;=\; \mathrm{i}\sum_{\alpha,\beta}\hat{q}_{\alpha\beta}q_{\alpha\beta}+\alpha\log\int \mathrm{D}\mathbf{z}\ \mathrm{e}^{\frac{1}{2}\beta c\sum_{\alpha,\beta=1}^{n}z_\alpha q_{\alpha\beta}z_\beta}-\frac{\beta c^2}{2N^\gamma}\sum_{\alpha=1}^{n}\sum_{\mu\le K}(m_\alpha^\mu)^2 \tag{B.12}
$$

$$
+\Big\langle\log\Big\langle \mathrm{e}^{\beta c\sum_{\mu\le K}\sum_{\alpha=1}^{n}\xi^\mu\sigma^\alpha m_\alpha^\mu-\mathrm{i}\sum_{\alpha,\beta}\sigma^\alpha\hat{q}_{\alpha\beta}\sigma^\beta}.\Big\rangle_{\sigma^1,\ldots,\sigma^n}\Big\rangle_{\xi} \tag{B.13}
$$

Now we can use the replica symmetry ansatz, and demand that the relevant saddle-point is of the form

$$
m_\alpha^\mu = m^\mu, \qquad q_{\alpha\beta}=\delta_{\alpha\beta}+q(1-\delta_{\alpha\beta}), \qquad \hat{q}_{\alpha\beta}=\frac{\mathrm{i}\alpha(\beta c)^2}{2}[R\delta_{\alpha\beta}+r(1-\delta_{\alpha\beta})], \tag{B.14}
$$

From now on we will denote $\mathbf{m}=(m^1,\ldots,m^K)$ and $\boldsymbol{\xi}=(\xi^1,\ldots,\xi^K)$. After some simple algebra we can take the limit $n\to 0$, and find that our free energy simplifies to

$$
\beta\overline{f}_{\mathrm{RS}} \;=\; \lim_{N\to\infty}\mathrm{extr}_{m,q,r}\ \hat{f}_{\mathrm{RS}}(\mathbf{m},q,r) \tag{B.15}
$$

with

$$
\begin{aligned}
\hat{f}_{\mathrm{RS}}(\mathbf{m},q,r) \;=\; & -\log 2+\frac{1}{2}\alpha r(\beta c)^2(1-q)+\frac{\beta c^2}{2N^\gamma}\mathbf{m}^2-\frac{\alpha}{2}\Big(\frac{\beta c q}{1-\beta c(1-q)}-\log[1-\beta c(1-q)]\Big) \\
& -\Big\langle\int \mathrm{D}z\ \log\cosh[\beta c(\mathbf{m}\cdot\boldsymbol{\xi}+z\sqrt{\alpha r})]\Big\rangle_{\xi}
\end{aligned}
\tag{B.16}
$$

Appendix C

Finite connectivity: Replica approach

C.1 Simple limits

Here we work out the theory in some simple limits, which can be worked out independently, to test more complicated stages of our general calculation:

- The paramagnetic state at $\beta = 0$:

$$\lim_{\beta\to 0} \beta\overline{f} \quad = \quad -\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{Nn}\log\sum_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^n} 1 = -\log 2. \tag{C.1}$$

The conditioned overlap distribution at $\beta = 0$ would be

$$\begin{aligned}
P(M|\psi) &= \frac{1}{P(\psi)}\lim_{N\to\infty}\frac{1}{\alpha N}\sum_{\mu=1}^{\alpha N}\delta(\psi-\psi_\mu)\int_{-\pi}^{\pi}\frac{\mathrm{d}\phi}{2\pi}\mathrm{e}^{\mathrm{i}M\phi}\ 2^{-N}\sum_{\boldsymbol{\sigma}}\overline{\mathrm{e}^{-\mathrm{i}\phi\sum_i \xi_i^\mu\sigma_i}} \\
&= \lim_{N\to\infty}\int_{-\pi}^{\pi}\frac{\mathrm{d}\phi}{2\pi}\mathrm{e}^{\mathrm{i}M\phi}\ \Big(1+\frac{c}{N}[\cos(\phi)-1]\Big)^N \\
&= \int_{-\pi}^{\pi}\frac{\mathrm{d}\phi}{2\pi}\mathrm{e}^{\mathrm{i}M\phi+c[\cos(\phi)-1]} = \mathrm{e}^{-c}\sum_{k\geq 0}\frac{c^k}{k!}\int_{-\pi}^{\pi}\frac{\mathrm{d}\phi}{2\pi}\mathrm{e}^{\mathrm{i}M\phi}\langle\mathrm{e}^{\mathrm{i}\phi\sigma}\rangle_{\sigma=\pm 1}^k \\
&= \mathrm{e}^{-c}\sum_{k\geq 0}\frac{c^k}{k!}\Big\langle\delta_{M,\sum_{\ell\leq k}\sigma_k}\Big\rangle_{\sigma_1\ldots\sigma_k=\pm 1}.
\end{aligned} \tag{C.2}$$

- The case of external fields only:

This simply corresponds to removing the M_μ^2 terms, and gives

$$\begin{aligned}
\overline{f} &= -\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{\beta N n}\log\overline{\prod_{i\alpha}\Big(\sum_\sigma \mathrm{e}^{\beta\sigma\sum_{\mu=1}^{\alpha N}\psi_\mu\xi_i^\mu}\Big)} \\
&= -\frac{1}{\beta}\log 2-\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{\beta n}\log\ \overline{\cosh^n\Big(\beta\sum_{\mu\le\alpha N}\psi_\mu\xi^\mu\Big)} \\
&= -\frac{1}{\beta}\log 2-\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{\beta n}\log\int\frac{\mathrm{d}h\mathrm{d}\hat{h}}{2\pi}\ \mathrm{e}^{\mathrm{i}\hat{h}h}\cosh^n(\beta h)\ \overline{\mathrm{e}^{-\mathrm{i}\hat{h}\sum_{\mu\le\alpha N}\psi_\mu\xi^\mu}} \\
&= -\frac{1}{\beta}\log 2-\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{\beta n}\log\int\frac{\mathrm{d}h\mathrm{d}\hat{h}}{2\pi}\ \mathrm{e}^{\mathrm{i}\hat{h}h}\cosh^n(\beta h)\prod_{\mu=1}^{\alpha N}\Big(1+\frac{c}{N}[\cos(\hat{h}\psi_\mu)-1]\Big) \\
&= -\frac{1}{\beta}\log 2-\lim_{n\to 0}\frac{1}{\beta n}\log\int\frac{\mathrm{d}h\mathrm{d}\hat{h}}{2\pi}\mathrm{e}^{\mathrm{i}\hat{h}h}\cosh^n(\beta h)\mathrm{e}^{\alpha c\int\mathrm{d}\psi\ P(\psi)[\cos(\hat{h}\psi)-1]} \\
&= -\frac{1}{\beta}\log 2-\lim_{n\to 0}\frac{1}{\beta n}\log\int\frac{\mathrm{d}h\mathrm{d}\hat{h}}{2\pi}\mathrm{e}^{\mathrm{i}\hat{h}h+\alpha c\int\mathrm{d}\psi\ P(\psi)[\cos(\hat{h}\psi)-1]}\Big\{1+n\log\cosh(\beta h)+\mathcal{O}(n^2)\Big\} \\
&= -\frac{1}{\beta}\log 2-\frac{1}{\beta}\int dh\ W(h)\log\cosh(\beta h), \qquad\qquad \text{(C.3)}
\end{aligned}$$

with the effective field distribution

$$\begin{aligned}
W(h) &= \int\frac{\mathrm{d}\hat{h}}{2\pi}\ \mathrm{e}^{\mathrm{i}\hat{h}h+\alpha c\int\mathrm{d}\psi\ P(\psi)[\cos(\hat{h}\psi)-1]} \\
&= \mathrm{e}^{-\alpha c}\sum_{k\ge 0}\frac{(\alpha c)^k}{k!}\int\Big[\prod_{\ell\le k}P(\psi_\ell)\mathrm{d}\psi_\ell\Big]\int\frac{\mathrm{d}\hat{h}}{2\pi}\mathrm{e}^{\mathrm{i}\hat{h}h}\prod_{\ell\le k}\cos(\hat{h}\psi_\ell) \\
&= \mathrm{e}^{-\alpha c}\sum_{k\ge 0}\frac{(\alpha c)^k}{k!}\int\Big[\prod_{\ell\le k}P(\psi_\ell)\mathrm{d}\psi_\ell\Big]\Big\langle\int\frac{\mathrm{d}\hat{h}}{2\pi}\mathrm{e}^{\mathrm{i}\hat{h}(h-\sum_{\ell\le k}\psi_\ell\sigma_\ell)}\Big\rangle_{\sigma_1\ldots\sigma_k=\pm 1} \\
&= \sum_{k\ge 0}\mathrm{e}^{-\alpha c}\frac{(\alpha c)^k}{k!}\Big\langle\Big\langle\delta\Big[h-\sum_{\ell\le k}\psi_\ell\sigma_\ell)\Big]\Big\rangle_{\psi_1\ldots\psi_k}\Big\rangle_{\sigma_1\ldots\sigma_k=\pm 1}. \qquad\qquad \text{(C.4)}
\end{aligned}$$

C.2 Normalization of $F(\boldsymbol{\omega})$

In this appendix we derive equation (2.90). It follows from

$$\begin{aligned}
\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\ \cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma}) &= \int\frac{\mathrm{d}m\mathrm{d}\hat{m}}{2\pi}\ \mathrm{e}^{\mathrm{i}m\hat{m}}\cos(m)\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\,\mathrm{e}^{-\mathrm{i}\hat{m}\boldsymbol{\omega}\cdot\boldsymbol{\sigma}} = \int\frac{\mathrm{d}m\mathrm{d}\hat{m}}{2\pi}\ \mathrm{e}^{\mathrm{i}m\hat{m}}\cos(m)\Big[\frac{2c}{\hat{m}}\sin(\hat{m}\pi)\Big]^n \\
&= \int\mathrm{d}\hat{m}\ \frac{\delta(\hat{m}+1)+\delta(\hat{m}-1)}{2}\Big[\frac{2c}{\hat{m}}\sin(\hat{m}\pi)\Big]^n = 0, \qquad\qquad \text{(C.5)}
\end{aligned}$$

where we isolated $\boldsymbol{\sigma}\cdot\boldsymbol{\omega}$ via $1=(2\pi)^{-1}\int\mathrm{d}m\mathrm{d}\hat{m}\,\mathrm{e}^{im\hat{m}-i\hat{m}\boldsymbol{\omega}\cdot\boldsymbol{\sigma}}$ and used

$$\begin{aligned}
\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\,\mathrm{e}^{-\mathrm{i}\hat{m}\boldsymbol{\omega}\cdot\boldsymbol{\sigma}} &= \prod_{\alpha=1}^{n}\int_{-\pi}^{\pi}\mathrm{d}\omega_\alpha\,\mathrm{e}^{-\mathrm{i}\hat{m}\omega_\alpha\sigma_\alpha} = \prod_{\alpha=1}^{n}\Big(2\int_0^{\pi}\mathrm{d}\omega_\alpha\ \cos(\hat{m}\omega_\alpha\sigma_\alpha)\Big) \\
&= \prod_{\alpha=1}^{n}\Big(\frac{2c\sigma^\alpha}{\hat{m}}\sin(\hat{m}\pi\sigma^\alpha)\Big) = \Big[\frac{2c}{\hat{m}}\sin(\hat{m}\pi)\Big]^n. \qquad\qquad \text{(C.6)}
\end{aligned}$$

C.3 Continuous RS phase transitions via route I

Here we derive the equation for the continuous phase transitions in the absence of external fields, i.e. for $P(\psi) = \delta(\psi)$, away from the solution (2.115). At the transition, the the function $D_0(\omega|\beta)$, which we will denote simply as $D(\omega|\beta)$, still satisfies (2.91). Continuous bifurcations away from (2.115) can be identified via a functional moment expansion. We transform

$$\pi(\omega) \to \cos(\omega) + \Delta(\omega), \tag{C.7}$$

with $f_k(\{\pi_1, \ldots, \pi_\ell\}) \to \tilde{f}_k(\{\Delta_1, \ldots, \Delta_k\})$, $W[\{\pi\}] \to \tilde{W}[\{\Delta\}]$, and $\tilde{W}[\{\Delta\}] = 0$ as soon as $\int \mathrm{d}\omega\ \Delta(\omega) \neq 0$ (because $\int \mathrm{d}\omega\, \pi(\omega) = 1$), and $\lambda(\theta|W) \to \tilde{\lambda}(\theta|\tilde{W})$. We expand our equations in powers of the functional moments $\varrho(\omega_1, \ldots, \omega_r) = \int \{\mathrm{d}\Delta\}\ \tilde{W}[\{\Delta\}]\Delta(\omega_1)\ldots\Delta(\omega_r)$. One assumes that close to the transition there exists some small ϵ such that $\varrho(\omega_1, \ldots, \omega_r) = \mathcal{O}(\epsilon^r)$. If the lowest bifurcating is of order ϵ^1, we obtain, upon multiplying (2.105) by Δ and subsequently integrating over Δ:

$$\begin{aligned}
\varrho(\omega) &= \int \{\mathrm{d}\Delta\}\, \Delta(\omega) \int \frac{\mathrm{d}\theta}{2\pi}\ \tilde{\lambda}(\theta|\tilde{W}) \prod_\omega \delta\Big[\Delta(\omega) + \cos(\omega) - \cos(\omega - \theta)\Big] \\
&= \int \frac{\mathrm{d}\theta}{2\pi}\, \tilde{\lambda}(\theta|\tilde{W})[\cos(\omega - \theta) - \cos(\omega)] = \cos(\omega) \int \frac{\mathrm{d}\theta}{2\pi}\, \tilde{\lambda}(\theta|\tilde{W})[\cos\theta - 1],
\end{aligned} \tag{C.8}$$

where we used the invariance under $\theta \to -\theta$ of

$$\tilde{\lambda}(\theta|\tilde{W}) = \sum_{m \in \mathbb{Z}} \mathrm{e}^{im\theta + c\alpha \sum_{k\geq 0} \frac{c^k \mathrm{e}^{-c}}{k!} \int \prod_{\ell=1}^k \left[\{\mathrm{d}\Delta_\ell\}\tilde{W}[\{\Delta_\ell\}]\right] \{\cos[m \arctan \tilde{f}_k(\{\Delta_1, \ldots, \Delta_k\})] - 1\}}. \tag{C.9}$$

The solution of (C.8) is clearly $\varrho(\omega) = \phi \cos(\omega)$, with

$$\phi = \int \frac{\mathrm{d}\theta}{2\pi}\, \tilde{\lambda}(\theta|\tilde{W})[\cos(\theta) - 1], \tag{C.10}$$

which we need to evaluate further by expanding $\tilde{\lambda}(\theta|\tilde{W})$ for small ϵ. Conversely, if the lowest bifurcating order is ϵ^2 one must focus on

$$\begin{aligned}
\varrho(\omega_1, \omega_2) &= \int \{\mathrm{d}\Delta\}\, \Delta(\omega_1)\Delta(\omega_2) \int \frac{\mathrm{d}\theta}{2\pi}\ \tilde{\lambda}(\theta|\tilde{W}) \prod_\omega \delta\Big[\Delta(\omega) - \cos(\omega) - \cos(\omega - \theta)\Big] \\
&= \cos(\omega_1)\cos(\omega_2) \int \frac{\mathrm{d}\theta}{2\pi}\, \tilde{\lambda}(\theta|\tilde{W})[\cos(\theta) - 1]^2 + \sin(\omega_1)\sin(\omega_2) \int \frac{\mathrm{d}\theta}{2\pi}\, \tilde{\lambda}(\theta|\tilde{W}) \sin^2(\theta).
\end{aligned} \tag{C.11}$$

We first inspect (C.10). Transforming each π_ℓ in (2.114) according to (C.7), we have

$$\prod_{\ell=1}^k \pi_\ell(\omega) = \prod_{\ell=1}^k [\cos(\omega) + \Delta_\ell(\omega)] = \cos^k(\omega)\Big[1 + \sum_{\ell=1}^k \frac{\Delta_\ell(\omega)}{\cos(\omega)}\Big] + \mathcal{O}(\Delta^2). \tag{C.12}$$

Inserting this result into (2.114), and using the properties (2.91), allows us to expand $\tilde{f}_k(\{\Delta_1,\ldots,\Delta_k\})$:

$$\tilde{f}_k(\{\Delta_1,\ldots,\Delta_k\}) = \frac{\sum_{\ell=1}^k \int_{-\pi}^{\pi} \mathrm{d}\omega\ \sin(\omega)\cos^{k-1}(\omega)\Delta_\ell(\omega)D(\omega|\beta)}{\int_{-\pi}^{\pi} \mathrm{d}\omega\ \cos^{k+1}(\omega)D(\omega|\beta)} + \mathcal{O}(\Delta^2). \tag{C.13}$$

We substitute the above into (C.9) and expand $\cos(m\arctan(x)) = 1-\frac{1}{2}m^2x^2+\mathcal{O}(x^4)$. Upon introducing

$$\mathcal{I}_k = \int \prod_{\ell=1}^k \left[\{\mathrm{d}\Delta_\ell\}\tilde{W}[\{\Delta_\ell\}]\right]\Big[\sum_{s=1}^k \int_{-\pi}^{\pi} \mathrm{d}\omega\ \sin(\omega)\cos^{k-1}(\omega)\Delta_s(\omega)D(\omega|\beta)\Big]^2, \tag{C.14}$$

$$A_k = \int_{-\pi}^{\pi} \mathrm{d}\omega\ \cos^{k+1}(\omega)D(\omega|\beta), \tag{C.15}$$

we see that $\mathcal{I}_k = \mathcal{O}(\epsilon^2)$, so we can now expand $\tilde{\lambda}(\theta|\tilde{W})$ as

$$\begin{aligned}
\tilde{\lambda}(\theta|\tilde{W}) &= \sum_{m\in\mathbb{Z}} \exp\Big[im\theta - \frac{c\alpha}{2}m^2\sum_{k\geq 0}\frac{\mathrm{e}^{-c}c^k}{k!}\frac{\mathcal{I}_k}{A_k^2} + \mathcal{O}(\epsilon^4)\Big] \\
&= \sum_{m\in\mathbb{Z}} \mathrm{e}^{im\theta}\Big[1 - \frac{c\alpha}{2}m^2\sum_{k\geq 0}\frac{\mathrm{e}^{-c}c^k}{k!}\frac{\mathcal{I}_k}{A_k^2} + \mathcal{O}(\epsilon^4)\Big] \\
&= 2\pi\delta(\theta) + \pi\alpha c\ \delta''(\theta)\sum_{k\geq 0}\frac{\mathrm{e}^{-c}c^k}{k!}\frac{\mathcal{I}_k}{A_k^2} + \mathcal{O}(\epsilon^4).
\end{aligned} \tag{C.16}$$

Next we need to work out the factors $\mathcal{I}_k$. Using the functional moment definition $\varrho(\omega_1,\ldots,\omega_r) = \int\{\mathrm{d}\Delta\}\ \tilde{W}[\{\Delta\}]\Delta(\omega_1)\ldots\Delta(\omega_r)$, one may write

$$\begin{aligned}
&\int \prod_{\ell=1}^k \left[\{\mathrm{d}\Delta_\ell\}\tilde{W}[\{\Delta_\ell\}]\right] \sum_{r,s=1}^k \Delta_r(\omega')\Delta_s(\omega'') \\
&= \sum_r \int \{\mathrm{d}\Delta_r\}\ \tilde{W}[\{\Delta_r\}]\Delta_r(\omega')\Delta_r(\omega'') + \sum_{r\neq s}\int \{\mathrm{d}\Delta_r\}\{\mathrm{d}\Delta_s\}\ \tilde{W}[\{\Delta_r\}]\tilde{W}[\{\Delta_s\}]\Delta_r(\omega')\Delta_s(\omega'') \\
&= k\varrho(\omega',\omega'') + k(k-1)\varrho(\omega')\varrho(\omega'').
\end{aligned} \tag{C.17}$$

This allows us to work out (C.14) further:

$$\begin{aligned}
\mathcal{I}_k &= k\int_{-\pi}^{\pi} \mathrm{d}\omega'\mathrm{d}\omega''\ \sin(\omega')\cos^{k-1}(\omega')D(\omega'|\beta)\sin(\omega'')\cos^{k-1}(\omega'')D(\omega''|\beta)\psi(\omega',\omega'') \\
&\quad + k(k-1)\Big[\int_{-\pi}^{\pi} \mathrm{d}\omega'\ \sin(\omega')\cos^{k-1}(\omega')D(\omega'|\beta)\psi(\omega')\Big]^2 \\
&= k\int_{-\pi}^{\pi} \mathrm{d}\omega'\mathrm{d}\omega''\ D(\omega'|\beta)D(\omega''|\beta)\sin(\omega')\cos^{k-1}(\omega')\psi(\omega',\omega'')\sin(\omega'')\cos^{k-1}(\omega''),
\end{aligned} \tag{C.18}$$

where in the last equality we have used the symmetry of $D(\omega|\beta)$ and $\varrho(\omega) = \phi\cos(\omega)$. Inserting this last expression in (C.16) and shifting the summation index $k \to k+1$ then leads to

$$\tilde{\lambda}(\theta|\tilde{W}) = 2\pi\delta(\theta)+\pi\alpha c^2\delta''(\theta)S(\{\varrho\})+\mathcal{O}(\epsilon^4), \tag{C.19}$$

$$S(\{\varrho\}) = \sum_{k\geq 0}\frac{\mathrm{e}^{-c}c^k}{k!}\frac{\int_{-\pi}^{\pi}\mathrm{d}\omega'\mathrm{d}\omega''\, D(\omega'|\beta)D(\omega''|\beta)\sin(\omega')\cos^k(\omega')\varrho(\omega',\omega'')\sin(\omega'')\cos^k(\omega'')}{\left[\int_{-\pi}^{\pi}\mathrm{d}\omega\, D(\omega|\beta)\cos^{k+2}(\omega)\right]^2} \tag{C.20}$$

To make further progress we need to calculate $\varrho(\omega',\omega'')$. We can first simplify (C.11) using (C.10), giving

$$\varrho(\omega_1,\omega_2) = \phi'\sin(\omega_1)\sin(\omega_2)-(2\phi+\phi')\cos(\omega_1)\cos(\omega_2), \tag{C.21}$$

where we defined

$$\phi' = \int_{-\pi}^{\pi}\frac{\mathrm{d}\theta}{2\pi}\,\tilde{\lambda}(\theta|\tilde{W})\,\sin^2(\theta). \tag{C.22}$$

With this we can simplify (C.20) to

$$S(\{\varrho\}) = \phi'\sum_{k\geq 0}\frac{\mathrm{e}^{-c}c^k}{k!}\frac{\left[\int_{-\pi}^{\pi}\mathrm{d}\omega\, D(\omega|\beta)\sin^2(\omega)\cos^k(\omega)\right]^2}{\left[\int_{-\pi}^{\pi}\mathrm{d}\omega\, D(\omega|\beta)\cos^{k+2}(\omega)\right]^2}. \tag{C.23}$$

Together with (C.19), this allows us to established equations from which to solve the two amplitudes ϕ and ϕ', by substitution into (C.10) and (C.22). This results in, after intergation by parts over θ:

$$\begin{aligned}
\phi &= \frac{1}{2}\alpha c^2 S(\{\varrho\})\int_{-\pi}^{\pi}\mathrm{d}\theta\,[\cos(\theta)-1]\delta''(\theta)+\mathcal{O}(\epsilon^4) = -\frac{1}{2}\alpha c^2 S(\{\varrho\})+\mathcal{O}(\epsilon^4)\\
&= -\frac{1}{2}\alpha c^2\phi'\sum_{k\geq 0}\frac{\mathrm{e}^{-c}c^k}{k!}\frac{\left[\int_{-\pi}^{\pi}\mathrm{d}\omega\, D(\omega|\beta)\sin^2(\omega)\cos^k(\omega)\right]^2}{\left[\int_{-\pi}^{\pi}\mathrm{d}\omega\, D(\omega|\beta)\cos^{k+2}(\omega)\right]^2}+\mathcal{O}(\epsilon^4) \qquad \text{(C.24)}\\
\phi' &= \frac{1}{2}\alpha c^2 S(\{\varrho\})\int_{-\pi}^{\pi}\mathrm{d}\theta\,\sin^2(\theta)\delta''(\theta)+\mathcal{O}(\epsilon^4)\\
&= \alpha c^2\phi'\sum_{k\geq 0}\frac{\mathrm{e}^{-c}c^k}{k!}\frac{\left[\int_{-\pi}^{\pi}\mathrm{d}\omega\, D(\omega|\beta)\sin^2(\omega)\cos^k(\omega)\right]^2}{\left[\int_{-\pi}^{\pi}\mathrm{d}\omega\, D(\omega|\beta)\cos^{k+2}(\omega)\right]^2}+\mathcal{O}(\epsilon^4). \qquad \text{(C.25)}
\end{aligned}$$

Since $\phi'=0$ immediately implies that $\phi=0$, the only possible continuous bifurcation must be the first instance where $\phi'\neq 0$. According to the above equation this $\mathcal{O}(\epsilon^2)$ bifurcation happens when

$$1 = \alpha c^2\sum_{k\geq 0}\frac{\mathrm{e}^{-c}c^k}{k!}\left[\frac{\int_{-\pi}^{\pi}\mathrm{d}\omega\sin^2(\omega)\cos^k(\omega)D(\omega|\beta)}{\int_{-\pi}^{\pi}\mathrm{d}\omega\cos^{k+2}(\omega)D(\omega|\beta)}\right]^2, \tag{C.26}$$

with $D(\omega|\beta)=(2\pi)^{-1}\sum_{m\in\mathbb{Z}}\cos(m\omega)\mathrm{e}^{\beta m^2/2c}$. Equation (C.26) defines the transition point, where the system will leave the state (2.115). The right-hand side of (C.26) obeys $\lim_{\beta\to 0}\mathrm{RHS}=0$. In C.7 we show that $\lim_{\beta\to\infty}\mathrm{RHS}=\alpha c^2$, so a transition at finite temperature $T_c=\beta_c^{-1}>0$ exists to a new state with $W[\{\pi\}]\neq\prod_\omega\delta[\pi(\omega)-\cos(\omega)]$ as soon as $\alpha c^2>1$. The critical temperature becomes zero when $\alpha c^2=1$.

C.4 Saddle point equations in terms of $L(\boldsymbol{\sigma})$

Here we derive equation (2.134), starting from the definition (2.128) and relation (2.129):

$$
\begin{aligned}
L(\boldsymbol{\sigma}) &= \alpha c\left\langle\frac{\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})Q(\boldsymbol{\omega})\sum_{\mathbf{M}}\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}+\sum_\alpha\chi(M_\alpha,\psi)}}{\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\,Q(\boldsymbol{\omega})\sum_{\mathbf{M}}\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}+\sum_\alpha\chi(M_\alpha,\psi)}}\right\rangle_\psi\\
&= \alpha c\left\langle\frac{\int\mathrm{d}\boldsymbol{\omega}\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})\sum_{\mathbf{M}'}\tilde{Q}(\mathbf{M}')\sum_{\mathbf{M}}\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot(\mathbf{M}-\mathbf{M}')+\sum_\alpha\chi(M_\alpha,\psi)}}{\int\mathrm{d}\boldsymbol{\omega}\sum_{\mathbf{M}'}\tilde{Q}(\mathbf{M}')\sum_{\mathbf{M}}\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot(\mathbf{M}-\mathbf{M}')+\sum_\alpha\chi(M_\alpha,\psi)}}\right\rangle_\psi. \qquad \text{(C.27)}
\end{aligned}
$$

We can then work out the integrals

$$
\begin{aligned}
\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\;\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot(\mathbf{M}-\mathbf{M}')} &= \frac{1}{2}\int_{-\pi}^{\pi}\mathrm{d}\boldsymbol{\omega}\;(\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{\sigma}}+\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{\sigma}})\mathrm{e}^{i\boldsymbol{\omega}\cdot(\mathbf{M}-\mathbf{M}')}\\
&= \pi(\delta_{\mathbf{M}',\mathbf{M}+\boldsymbol{\sigma}}+\delta_{\mathbf{M}',\mathbf{M}-\boldsymbol{\sigma}}), \qquad \text{(C.28)}
\end{aligned}
$$

and substituting into (C.27) gives

$$
\begin{aligned}
L(\boldsymbol{\sigma}) &= \frac{1}{2}\alpha c\left\langle\frac{\sum_{\mathbf{M}}\left[\tilde{Q}(\mathbf{M}+\boldsymbol{\sigma})\mathrm{e}^{\beta\sum_\alpha\chi(M_\alpha,\psi)}+\tilde{Q}(\mathbf{M}-\boldsymbol{\sigma})\mathrm{e}^{\beta\sum_\alpha\chi(M_\alpha,\psi)}\right]}{\sum_{\mathbf{M}}\tilde{Q}(\mathbf{M})\mathrm{e}^{\sum_\alpha\chi(M_\alpha,\psi)}}\right\rangle_\psi\\
&= \frac{1}{2}\alpha c\left\langle\frac{\sum_{\mathbf{M}}\tilde{Q}(\mathbf{M})\left[\mathrm{e}^{\beta\sum_\alpha\chi(M_\alpha-\sigma^\alpha,\psi)}+\mathrm{e}^{\beta\sum_\alpha\chi(M_\alpha+\sigma^\alpha,\psi)}\right]}{\sum_{\mathbf{M}}\tilde{Q}(\mathbf{M})\mathrm{e}^{\sum_\alpha\chi(M_\alpha,\psi)}}\right\rangle_\psi\\
&= c\alpha\left\langle\mathrm{e}^{\frac{\beta n}{2c}}\;\frac{\sum_{\mathbf{M}}\tilde{Q}(\mathbf{M})\mathrm{e}^{\beta\sum_\alpha\chi(M_\alpha,\psi)}\cosh[\beta(\frac{1}{c}\mathbf{M}\cdot\boldsymbol{\sigma}+\psi\sum_\alpha\sigma^\alpha)]}{\sum_{\mathbf{M}}\tilde{Q}(\mathbf{M})\mathrm{e}^{\beta\sum_\alpha\chi(M_\alpha,\psi)}}\right\rangle_\psi. \qquad \text{(C.29)}
\end{aligned}
$$

C.5 Simple limits to test the replica theory

Here we inspect several simple limits to test our results for the overlap distribution and the free energy.

- Infinite temperature:

  Using $\lim_{\beta\to 0}L(\boldsymbol{\sigma})=\alpha c$ and $\sum_{\mathbf{M}}\tilde{Q}(\mathbf{M})=\mathrm{e}^c$ in (2.135) we immediately find the correct free energy

$$\lim_{\beta\to 0}\beta\overline{f}_{\rm RSB} = -\lim_{n\to 0}\frac{1}{n}\Big\{\log\sum_{\boldsymbol{\sigma}}1\Big\} = -\log 2. \tag{C.30}$$

Moreover, from (2.133) we can extract

$$\begin{aligned}
\lim_{\beta\to 0}\tilde{Q}(\mathbf{M}) &= \int_{-\pi}^{\pi}\frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\ \cos(\boldsymbol{\omega}\cdot\mathbf{M})\ \mathrm{e}^{c\ 2^{-n}\sum_{\boldsymbol{\sigma}}\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})} \\
&= \sum_{k\geq 0}\frac{c^k}{k!}2^{-nk}\sum_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^k}\int_{-\pi}^{\pi}\frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\ \cos(\boldsymbol{\omega}\cdot\mathbf{M})\ \prod_{\ell\leq k}\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma}^\ell) \\
&= \sum_{k\geq 0}\frac{c^k}{k!}\ 2^{-nk}\sum_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^k}\delta_{\mathbf{M},\sum_{\ell\leq k}\boldsymbol{\sigma}^\ell} = \sum_{k\geq 0}\frac{c^k}{k!}\prod_{\alpha=1}^{n}\langle\delta_{M_\alpha,\sum_{\ell\leq k}\sigma_\ell}\rangle_{\sigma_1\ldots\sigma_k=\pm 1}.
\end{aligned}\tag{C.31}$$

Hence, it now follows from (2.136) that

$$\begin{aligned}
\lim_{\beta\to 0}P(M|\psi) &= \lim_{n\to 0}\frac{1}{n}\sum_{\gamma=1}^{n}\frac{\sum_{\mathbf{M}\in\mathbb{Z}^n}\sum_{k\geq 0}\frac{c^k}{k!}\prod_{\alpha=1}^{n}\langle\delta_{M_\alpha,\sum_{\ell\leq k}\sigma_\ell}\rangle_{\sigma_1\ldots\sigma_k=\pm 1}\ \delta_{M,M_\gamma}}{\sum_{\mathbf{M}\in\mathbb{Z}^n}\sum_{k\geq 0}\frac{c^k}{k!}\prod_{\alpha=1}^{n}\langle\delta_{M_\alpha,\sum_{\ell\leq k}\sigma_\ell}\rangle_{\sigma_1\ldots\sigma_k=\pm 1}} \\
&= \mathrm{e}^{-c}\sum_{k\geq 0}\frac{c^k}{k!}\langle\delta_{M,\sum_{\ell\leq k}\sigma_\ell}\rangle_{\sigma_1\ldots\sigma_k=\pm 1}.
\end{aligned}\tag{C.32}$$

This coincide with our RS expression, as it should since at high temperature the RS ansatz is exact.

- External fields only:

In the case of having only external fields we simply remove all terms that come from the interaction energy in (C.29), obtaining

$$L(\boldsymbol{\sigma}) = \alpha c\ \langle\cosh\big[\beta\psi\sum_\alpha\sigma_\alpha\big]\rangle_\psi. \tag{C.33}$$

Inserting this into (2.133), and introducing the normalised measure

$$\lambda(\boldsymbol{\sigma}) = \frac{\mathrm{e}^{\alpha c\ \langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi}}{\sum_{\boldsymbol{\sigma}'}\mathrm{e}^{\alpha c\ \langle\cosh[\beta\psi\sum_\alpha\sigma'_\alpha]\rangle_\psi}}, \tag{C.34}$$

we get

$$
\begin{aligned}
\tilde{Q}(\mathbf{M}) &= \int_{-\pi}^{\pi} \frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\ \cos(\boldsymbol{\omega}\cdot\mathbf{M})\ \mathrm{e}^{c\sum_{\boldsymbol{\sigma}}\lambda(\boldsymbol{\sigma})\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})} \\
&= \sum_{k\geq 0}\frac{c^k}{k!}\int_{-\pi}^{\pi}\frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\ \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}}\Big(\sum_{\boldsymbol{\sigma}}\lambda(\boldsymbol{\sigma})\mathrm{e}^{-\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{\sigma}}\Big)^k \\
&= \sum_{k\geq 0}\frac{c^k}{k!}\int_{-\pi}^{\pi}\frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\ \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{M}}\sum_{\boldsymbol{\sigma}_1\ldots\boldsymbol{\sigma}_k}\Big[\prod_{\ell=1}^{k}\lambda(\boldsymbol{\sigma}_\ell)\Big]\mathrm{e}^{-\mathrm{i}\boldsymbol{\omega}\cdot\sum_{\ell\leq k}\boldsymbol{\sigma}_\ell} \\
&= \sum_{k\geq 0}\frac{c^k}{k!}\sum_{\boldsymbol{\sigma}_1\ldots\boldsymbol{\sigma}_k}\Big[\prod_{\ell=1}^{k}\lambda(\boldsymbol{\sigma}_\ell)\Big]\delta_{\mathbf{M},\sum_{\ell\leq k}\boldsymbol{\sigma}_\ell}. \qquad \text{(C.35)}
\end{aligned}
$$

This then gives for the free energy, upon removing the interaction energy:

$$
\begin{aligned}
\overline{f}_{\mathrm{RSB}} &= -\lim_{n\to 0}\frac{1}{\beta n}\Big\{\alpha\Big\langle\log\sum_{\mathbf{M}\in\mathbb{Z}^n}\tilde{Q}(\mathbf{M})\ \mathrm{e}^{\beta\psi\sum_\alpha M_\alpha}\Big\rangle_\psi \\
&\quad +\log\sum_{\boldsymbol{\sigma}}\mathrm{e}^{\alpha c[\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi-1]}-\alpha c\sum_{\boldsymbol{\sigma}}\lambda(\boldsymbol{\sigma})\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi\Big\} \\
&= -\lim_{n\to 0}\frac{1}{\beta n}\Big\{\alpha\Big\langle\log\Big[\sum_{k\geq 0}\frac{c^k}{k!}\sum_{\boldsymbol{\sigma}_1\ldots\boldsymbol{\sigma}_k}\Big[\prod_{\ell=1}^{k}\lambda(\boldsymbol{\sigma}_\ell)\Big]\sum_{\mathbf{M}\in\mathbb{Z}^n}\delta_{\mathbf{M},\sum_{\ell\leq k}\boldsymbol{\sigma}_\ell}\mathrm{e}^{\beta\psi\sum_\alpha M_\alpha}\Big]\Big\rangle_\psi \\
&\quad +\log\sum_{\boldsymbol{\sigma}}\mathrm{e}^{\alpha c[\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi-1]}-\alpha c\sum_{\boldsymbol{\sigma}}\lambda(\boldsymbol{\sigma})\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi\Big\} \\
&= -\lim_{n\to 0}\frac{1}{\beta n}\Big\{\alpha\Big\langle\log\Big[\sum_{k\geq 0}\frac{c^k}{k!}\Big(\sum_{\boldsymbol{\sigma}}\lambda(\boldsymbol{\sigma})\mathrm{e}^{\beta\psi\sum_\alpha\sigma_\alpha}\Big)^k\Big]\Big\rangle_\psi \\
&\quad +\log\sum_{\boldsymbol{\sigma}}\mathrm{e}^{\alpha c[\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi-1]}-\alpha c\sum_{\boldsymbol{\sigma}}\lambda(\boldsymbol{\sigma})\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi\Big\} \\
&= -\lim_{n\to 0}\frac{1}{\beta n}\Big\{\alpha c\Big\langle\sum_{\boldsymbol{\sigma}}\lambda(\boldsymbol{\sigma})\mathrm{e}^{\beta\psi\sum_\alpha\sigma_\alpha}\Big\rangle_\psi-\alpha c \\
&\quad +\log\sum_{\boldsymbol{\sigma}}\mathrm{e}^{\alpha c[\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi-1]}-\alpha c\sum_{\boldsymbol{\sigma}}\lambda(\boldsymbol{\sigma})\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi\Big\} \\
&= -\lim_{n\to 0}\frac{1}{\beta n}\Big\{\alpha c\sum_{\boldsymbol{\sigma}}\lambda(\boldsymbol{\sigma})\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi-\alpha c \\
&\quad +\log\sum_{\boldsymbol{\sigma}}\mathrm{e}^{\alpha c[\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi-1]}-\alpha c\sum_{\boldsymbol{\sigma}}\lambda(\boldsymbol{\sigma})\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi\Big\} \\
&= -\lim_{n\to 0}\frac{1}{\beta n}\Big\{\log\sum_{\boldsymbol{\sigma}}\mathrm{e}^{\alpha c\ [\langle\cosh[\beta\psi\sum_\alpha\sigma_\alpha]\rangle_\psi-1]}\Big\}, \qquad \text{(C.36)}
\end{aligned}
$$

where in the penultimate step we used $\lambda(\boldsymbol{\sigma})=\lambda(-\boldsymbol{\sigma})$. We next use the following replica identity, which is proved via Taylor expansion of even non-negative analytical functions

$F(x)$ that have $F(0)=1$:

$$\lim_{n\to 0} n^{-1}\log\big\langle F(\sum_{\alpha=1}^{n}\sigma_\alpha)\big\rangle_{\sigma_1\ldots\sigma_n=\pm 1} = \sum_{k>0}\frac{F^{(k)}(0)}{k!}\Big(\frac{\mathrm{d}^k}{\mathrm{d}x^k}\log\cosh(x)\Big)\Big|_{x=0}. \tag{C.37}$$

Application to the function $F(z)=\exp[\alpha c\langle\cosh[\beta\psi z]\rangle_\psi-\alpha c]$ gives

$$\begin{aligned}
\overline{f}_{\mathrm{RSB}} &= -\frac{1}{\beta}\log 2-\frac{\mathrm{e}^{-\alpha c}}{\beta}\lim_{x,z\to 0}\sum_{k>0}\frac{1}{k!}\Big(\frac{\mathrm{d}^k}{\mathrm{d}x^k}\log\cosh(x)\Big)\frac{\mathrm{d}^k}{\mathrm{d}z^k}\mathrm{e}^{\alpha c\langle\cosh(\beta\psi z)\rangle_\psi} \\
&= -\frac{1}{\beta}\log 2-\frac{\mathrm{e}^{-\alpha c}}{\beta}\sum_{\ell\geq 0}\frac{(\alpha c)^\ell}{\ell!}\lim_{x,z\to 0}\sum_{k>0}\frac{1}{k!}\Big(\frac{\mathrm{d}^k}{\mathrm{d}x^k}\log\cosh(x)\Big)\frac{\mathrm{d}^k}{\mathrm{d}z^k}\langle\cosh(\beta\psi z)\rangle_\psi^\ell \\
&= -\frac{1}{\beta}\log 2-\frac{\mathrm{e}^{-\alpha c}}{\beta}\sum_{\ell\geq 0}\frac{(\alpha c)^\ell}{\ell!}\lim_{x,z\to 0}\sum_{k>0}\frac{1}{k!}\Big(\frac{\mathrm{d}^k}{\mathrm{d}x^k}\log\cosh(x)\Big)\frac{\mathrm{d}^k}{\mathrm{d}z^k}\langle\langle e^{\beta\psi\sum_{r\leq\ell}\sigma_r z_r}\rangle_{\psi_1\ldots\psi_\ell}\rangle_{\sigma_1\ldots\sigma_\ell=\pm 1} \\
&= -\frac{1}{\beta}\log 2-\frac{\mathrm{e}^{-\alpha c}}{\beta}\sum_{\ell\geq 0}\frac{(\alpha c)^\ell}{\ell!}\Big\langle\Big\langle\sum_{k>0}\frac{1}{k!}\Big(\lim_{x\to 0}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\log\cosh(x)\Big)\Big(\beta\sum_{r\leq\ell}\sigma_r\psi_r\Big)^k\Big\rangle_{\psi_1\ldots\psi_\ell}\Big\rangle_{\sigma_1\ldots\sigma_\ell=\pm 1} \\
&= -\frac{1}{\beta}\log 2-\frac{\mathrm{e}^{-\alpha c}}{\beta}\sum_{\ell\geq 0}\frac{(\alpha c)^\ell}{\ell!}\Big\langle\Big\langle\log\cosh\Big(\beta\sum_{r\leq\ell}\sigma_r\psi_r\Big)\Big\rangle_{\psi_1\ldots\psi_\ell}\Big\rangle_{\sigma_1\ldots\sigma_\ell=\pm 1} \\
&= -\frac{1}{\beta}\log 2-\frac{1}{\beta}\int\mathrm{d}h\ W(h)\log\cosh(\beta h),
\end{aligned} \tag{C.38}$$

with

$$W(h) = \sum_{k\geq 0}e^{-\alpha c}\frac{(\alpha c)^k}{k!}\Big\langle\Big\langle\delta\Big[h-\sum_{\ell\leq k}\psi_\ell\sigma_\ell\Big]\Big\rangle_{\psi_1\ldots\psi_k}\Big\rangle_{\sigma_1\ldots\sigma_k=\pm 1}. \tag{C.39}$$

This recovers correctly the solution of external fields only.

C.6 Derivation of RS equations via route II

The RS ansatz converts the saddle point equation (2.134) into

$$\begin{aligned}
\int\mathrm{d}h\ W(h)\mathrm{e}^{\beta h\sum_\alpha\sigma_\alpha} &= \mathrm{e}^{\beta n/2c}\Big\langle\Big\langle\int\{\mathrm{d}\pi\}W[\pi]\prod_\alpha\Big(\sum_M\pi(M)\mathrm{e}^{\beta(M^2/2c+\psi M+\tau(\psi+M/c)\sigma_\alpha)}\Big)\Big\rangle_\psi\Big\rangle_{\tau=\pm 1} \\
&= \mathrm{e}^{\beta n/2c}\Big\langle\Big\langle\int\{\mathrm{d}\pi\}W[\pi]\Big(\sum_M\pi(M)\mathrm{e}^{\beta(M^2/2c+\psi M+\tau(\psi+M/c))}\Big)^{\frac{1}{2}n+\frac{1}{2}\sum_\alpha\sigma_\alpha} \\
&\qquad\times\Big(\sum_M\pi(M)\mathrm{e}^{\beta(M^2/2c+\psi M-\tau(\psi+M/c))}\Big)^{\frac{1}{2}n-\frac{1}{2}\sum_\alpha\sigma_\alpha}\Big\rangle_\psi\Big\rangle_{\tau=\pm 1} \\
&= \mathrm{e}^{\beta n/2c}\Big\langle\Big\langle\int\{d\pi\}W[\pi]\Big(\frac{\sum_M\pi(M)\mathrm{e}^{\beta(M^2/2c+\psi M+\tau(\psi+M/c))}}{\sum_M\pi(M)\mathrm{e}^{\beta(M^2/2c+\psi M-\tau(\psi+M/c))}}\Big)^{\frac{1}{2}\sum_\alpha\sigma_\alpha}\Big\rangle_\psi\Big\rangle_{\tau=\pm 1} \\
&= \mathrm{e}^{\beta n/2c}\int\mathrm{d}h\ \mathrm{e}^{\beta h\sum_\alpha\sigma_\alpha}\Big\langle\Big\langle\int\{\mathrm{d}\pi\}W[\pi]\delta\Big[h-\frac{1}{2\beta}\log\Big(\frac{\sum_M\pi(M)\mathrm{e}^{\beta(M^2/2c+\psi M+\tau(\psi+M/c))}}{\sum_M\pi(M)\mathrm{e}^{\beta(M^2/2c+\psi M-\tau(\psi+M/c))}}\Big)\Big]\Big\rangle_\psi\Big\rangle_{\tau=\pm 1}.
\end{aligned} \tag{C.40}$$

We conclude after sending $n\to 0$ that

$$W(h) \;=\; \Big\langle\!\Big\langle \int\!\{\mathrm{d}\pi\}W[\pi]\;\delta\Big[h-\frac{1}{2\beta}\log\Big(\frac{\sum_M \pi(M)\mathrm{e}^{\beta(M^2/2c+\psi M+\tau(\psi+M/c))}}{\sum_M \pi(M)\mathrm{e}^{\beta(M^2/2c+\psi M-\tau(\psi+M/c))}}\Big)\Big]\Big\rangle_{\psi}\Big\rangle_{\tau=\pm 1} \tag{C.41}$$

$W(h)$ is indeed symmetric. Next we turn to equation (2.133), where we require quantities of the form

$$\varrho(\boldsymbol{\omega}) \;=\; \sum_{\boldsymbol{\sigma}}\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})\mathrm{e}^{L(\boldsymbol{\sigma})}=\sum_{\boldsymbol{\sigma}}\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})\mathrm{e}^{\alpha c\int \mathrm{d}h\;W(h)\mathrm{e}^{\beta h\sum_\alpha \sigma_\alpha}}. \tag{C.42}$$

In fact we will need only the ratio $\varrho(\boldsymbol{\omega})/\varrho(\mathbf{0})$. We note that

$$\varrho(\mathbf{0}) \;=\; 2^n\sum_{k\geq 0}\frac{(\alpha c)^k}{k!}\int\!\mathrm{d}h_1\ldots\mathrm{d}h_k\Big[\prod_{\ell\leq k}W(h_k)\Big]\cosh^n\Big(\beta\sum_{\ell\leq k}h_\ell\Big)=\mathrm{e}^{\alpha c+\mathcal{O}(n)}. \tag{C.43}$$

We can hence write the RS version of our first saddle-point equation as follows, using $W(h)=W(-h)$:

$$\begin{aligned}
\int\!\{\mathrm{d}\pi\}W[\pi]\prod_{\alpha=1}^n\pi(M_\alpha) \;&=\; \mathrm{e}^{-c}\int_{-\pi}^{\pi}\frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\cos(\boldsymbol{\omega}\cdot\mathbf{M})\mathrm{e}^{c\mathrm{e}^{-\alpha c+\mathcal{O}(n)}\varrho(\boldsymbol{\omega})}\\
&=\; \mathrm{e}^{-c+\mathcal{O}(n)}\sum_{k\geq 0}\frac{c^k}{k!}\int_{-\pi}^{\pi}\frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\cos(\boldsymbol{\omega}\cdot\mathbf{M})\Big\langle\cos(\boldsymbol{\omega}\cdot\boldsymbol{\sigma})\mathrm{e}^{\alpha c\int\mathrm{d}h\;W(h)[\mathrm{e}^{\beta h\sum_\alpha\sigma_\alpha}-1]}\Big\rangle_{\boldsymbol{\sigma}}^k\\
&= \mathrm{e}^{-c+\mathcal{O}(n)}\sum_{k\geq 0}\frac{c^k}{k!}\Big\langle\!\Big\langle\int_{-\pi}^{\pi}\frac{\mathrm{d}\boldsymbol{\omega}}{(2\pi)^n}\mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot(\tau\mathbf{M}-\sum_{\ell\leq k}\tau_\ell\boldsymbol{\sigma}^\ell)}\mathrm{e}^{\alpha c\sum_{\ell\leq k}\int\mathrm{d}h\;W(h)[\mathrm{e}^{\beta h\sum_\alpha\sigma_\alpha^\ell}-1]}\Big\rangle_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^k}\Big\rangle_{\tau,\tau_1\ldots\tau_k=\pm 1}\\
&=\; \mathrm{e}^{-c+\mathcal{O}(n)}\sum_{k\geq 0}\frac{c^k}{k!}\mathrm{e}^{-\alpha ck}\Big\langle\mathrm{e}^{\alpha c\sum_{\ell\leq k}\int\mathrm{d}h\;W(h)\mathrm{e}^{\beta h\sum_\alpha\sigma_\alpha^\ell}}\delta_{\mathbf{M},\sum_{\ell\leq k}\boldsymbol{\sigma}^\ell}\Big\rangle_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^k}\\
&=\; \mathrm{e}^{-c+\mathcal{O}(n)}\sum_{k\geq 0}\frac{c^k}{k!}\mathrm{e}^{-\alpha ck}\sum_{r\geq 0}\frac{(\alpha c)^r}{r!}\Big\langle\Big(\int\mathrm{d}h\;W(h)\sum_{\ell\leq k}\mathrm{e}^{\beta h\sum_\alpha\sigma_\alpha^\ell}\Big)^r\delta_{\mathbf{M},\sum_{\ell\leq k}\boldsymbol{\sigma}^\ell}\Big\rangle_{\boldsymbol{\sigma}^1\ldots\boldsymbol{\sigma}^k}\\
&= \mathrm{e}^{-c+\mathcal{O}(n)}\sum_{k\geq 0}\frac{c^k}{k!}\mathrm{e}^{-\alpha ck}\sum_{r\geq 0}\frac{(\alpha c)^r}{r!}\int\mathrm{d}h_1\ldots\mathrm{d}h_r\Big[\prod_{s\leq r}W(h_s)\Big]\sum_{\ell_1\ldots\ell_r\leq k}\prod_\alpha\Big\langle\mathrm{e}^{\beta\sum_{s\leq r}h_s\sigma_{\ell_s}}\delta_{M_\alpha,\sum_{\ell\leq k}\sigma_\ell}\Big\rangle_{\sigma_1\ldots\sigma_k}\\
&= \mathrm{e}^{-c+\mathcal{O}(n)}\sum_{k\geq 0}\frac{c^k}{k!}\mathrm{e}^{-\alpha ck}\sum_{r\geq 0}\frac{(\alpha c)^r}{r!}\int\mathrm{d}h_1\ldots\mathrm{d}h_r\Big[\prod_{s\leq r}W(h_s)\Big]\sum_{\ell_1\ldots\ell_r\leq k}\prod_\alpha\left\{\frac{\langle\mathrm{e}^{\beta\sum_{s\leq r}h_s\sigma_{\ell_s}}\delta_{M_\alpha,\sum_{\ell\leq k}\sigma_\ell}\rangle_{\sigma_1\ldots\sigma_k}}{\langle\mathrm{e}^{\beta\sum_{s\leq r}h_s\sigma_{\ell_s}}\rangle_{\sigma_1\ldots\sigma_k}}\right\}\\
&= \int\!\{\mathrm{d}\pi\}\Big(\prod_\alpha\pi(M_\alpha)\Big)\mathrm{e}^{-c+\mathcal{O}(n)}\sum_{k\geq 0}\frac{c^k}{k!}\mathrm{e}^{-\alpha ck}\sum_{r\geq 0}\frac{(\alpha c)^r}{r!}\int\mathrm{d}h_1\ldots\mathrm{d}h_r\Big[\prod_{s\leq r}W(h_s)\Big]\sum_{\ell_1\ldots\ell_r\leq k}\\
&\qquad\times\prod_M\delta\left[\pi(M)-\frac{\langle\mathrm{e}^{\beta\sum_{s\leq r}h_s\sigma_{\ell_s}}\delta_{M,\sum_{\ell\leq k}\sigma_\ell}\rangle_{\sigma_1\ldots\sigma_k}}{\langle\mathrm{e}^{\beta\sum_{s\leq r}h_s\sigma_{\ell_s}}\rangle_{\sigma_1\ldots\sigma_k}}\right].
\end{aligned} \tag{C.44}$$

We thus conclude that for $n \to 0$ the following equation for $W[\pi]$ solves our saddle-point problem:

$$
\begin{aligned}
W[\pi] \;=\; & \mathrm{e}^{-c}\sum_{k\geq 0}\frac{c^k}{k!}\mathrm{e}^{-\alpha c k}\sum_{r\geq 0}\frac{(\alpha c)^r}{r!}\int_{-\infty}^{\infty}\mathrm{d}h_1\ldots\mathrm{d}h_r\Big[\prod_{s\leq r}W(h_s)\Big]\sum_{\ell_1\ldots\ell_r\leq k} \\
& \times\prod_M \delta\left[\pi(M)-\frac{\langle \mathrm{e}^{\beta\sum_{s\leq r}h_s\sigma_{\ell_s}}\delta_{M,\sum_{\ell\leq k}\sigma_\ell}\rangle_{\sigma_1\ldots\sigma_k}}{\langle \mathrm{e}^{\beta\sum_{s\leq r}h_s\sigma_{\ell_s}}\rangle_{\sigma_1\ldots\sigma_k}}\right]. \qquad (C.45)
\end{aligned}
$$

Everything is properly normalised, and if $W(h) = W(-h)$ the measure $W[\pi]$ is seen to permit only real-valued distributions $\pi(M)$ such that $\pi(M) \in [0,\infty)$ and $\pi(-M) = \pi(M)$ for all $M \in \mathbb{Z}$.

C.7 Continuous RS phase transitions via route II

Here we work with the order parameter equation that is written in terms of $W(h)$ only, i.e. (2.140), and look for phase transitions in the absence of external fields. For $P(\psi) = \delta(\psi)$ we must solve $W(h)$ from

$$
\begin{aligned}
W(h) \;=\; & \mathrm{e}^{-c}\sum_{k\geq 0}\frac{c^k}{k!}\mathrm{e}^{-\alpha c k}\sum_{r\geq 0}\frac{(\alpha c)^r}{r!}\int_{-\infty}^{\infty}\mathrm{d}h_1\ldots\mathrm{d}h_r\Big[\prod_{s\leq r}W(h_s)\Big]\sum_{\ell_1\ldots\ell_r\leq k} \\
& \times\Big\langle\delta\left[h-\frac{1}{2\beta}\log\left(\frac{\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\rangle_{\tau_1\ldots\tau_k=\pm 1}}{\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c-\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c+\beta\sum_{s\leq r}h_s\tau_{\ell_s}}\rangle_{\tau_1\ldots\tau_k=\pm 1}}\right)\right]\Big\rangle_{\tau=\pm 1} \qquad (C.46)
\end{aligned}
$$

Clearly $W(h) = \delta(h)$ solves this equation for any temperature. Due to $W(h) = W(-h)$, we will always have $\int \mathrm{d}h\; W(h)h = 0$, so the first bifurcation away from $W(h) = \delta(h)$ is expected to be in the second moment. We write $h = \epsilon y$, with $0 < \epsilon \ll 1$, and expand in powers of ϵ. Upon setting $W(h) = \epsilon^{-1}\tilde{W}(h/\epsilon)$ we have

$$
\begin{aligned}
\tilde{W}(y) \;=\; & \mathrm{e}^{-c}\sum_{k\geq 0}\frac{c^k}{k!}\mathrm{e}^{-\alpha c k}\sum_{r\geq 0}\frac{(\alpha c)^r}{r!}\int_{-\infty}^{\infty}\mathrm{d}y_1\ldots\mathrm{d}y_r\Big[\prod_{s\leq r}\tilde{W}(y_s)\Big]\sum_{\ell_1\ldots\ell_r\leq k} \\
& \times\Big\langle\delta\left[y-\frac{1}{2\beta\epsilon}\log\left(\frac{\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c+\beta\epsilon\sum_{s\leq r}y_s\tau_{\ell_s}}\rangle_{\tau_1\ldots\tau_k=\pm 1}}{\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c-\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c+\beta\epsilon\sum_{s\leq r}y_s\tau_{\ell_s}}\rangle_{\tau_1\ldots\tau_k=\pm 1}}\right)\right]\Big\rangle_{\tau=\pm 1} \qquad (C.47)
\end{aligned}
$$

Next we expand the logarithm in the last line. To leading orders in ϵ we obtain

$$
\begin{aligned}
\frac{1}{2\beta\epsilon}\log\Big(\ldots\Big) &= \frac{1}{2\beta\epsilon}\log\left(\frac{\big\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c}\big[1+\beta\epsilon\sum_{s\leq r}y_s\tau_{\ell_s}\big]\big\rangle_{\tau_1\ldots\tau_k=\pm1}}{\big\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c-\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c}\big[1+\beta\epsilon\sum_{s\leq r}y_s\tau_{\ell_s}\big]\big\rangle_{\tau_1\ldots\tau_k=\pm1}}\right)\\
&= \frac{1}{2\beta\epsilon}\log\left(\frac{1+\beta\epsilon\sum_{s\leq r}y_s\frac{\langle \tau_{\ell_s}\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c}\rangle_{\tau_1\ldots\tau_k=\pm1}}{\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c}\rangle_{\tau_1\ldots\tau_k=\pm1}}}{1+\beta\epsilon\sum_{s\leq r}y_s\frac{\langle \tau_{\ell_s}\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c-\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c}\rangle_{\tau_1\ldots\tau_k=\pm1}}{\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c-\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c}\rangle_{\tau_1\ldots\tau_k=\pm1}}}\right)\\
= \frac{1}{2}\sum_{s\leq r}y_s&\left\{\frac{\langle \tau_{\ell_s}\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c}\rangle_{\tau_1\ldots\tau_k}}{\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c+\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c}\rangle_{\tau_1\ldots\tau_k}}-\frac{\langle \tau_{\ell_s}\mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c-\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c}\rangle_{\tau_1\ldots\tau_k}}{\langle \mathrm{e}^{\beta(\sum_{\ell\leq k}\tau_\ell)^2/2c-\beta(\sum_{\ell\leq k}\tau_\ell)\tau/c}\rangle_{\tau_1\ldots\tau_k}}\right\}\\
&= \tau\sum_{s\leq r}y_s\left\{\frac{\int\mathrm{Dz}\ \tanh(z\sqrt{\beta/c}+\beta/c)\cosh^k(z\sqrt{\beta/c}+\beta/c)}{\int\mathrm{Dz}\ \cosh^k(z\sqrt{\beta/c}+\beta/c)}\right\}. \qquad \text{(C.48)}
\end{aligned}
$$

Hence our order parameter equation (C.47) becomes

$$
\begin{aligned}
\tilde{W}(y) &= \mathrm{e}^{-c}\sum_{k\geq0}\frac{c^k}{k!}\mathrm{e}^{-\alpha ck}\sum_{r\geq0}\frac{(\alpha c)^r}{r!}\int_{-\infty}^{\infty}\mathrm{d}y_1\ldots\mathrm{d}y_r\Big[\prod_{s\leq r}\tilde{W}(y_s)\Big]\sum_{\ell_1\ldots\ell_r\leq k}\\
&\times\Big\langle\delta\left[y-\tau\sum_{s\leq r}y_s\left\{\frac{\int\mathrm{Dz}\ \tanh(z\sqrt{\beta/c}+\beta/c)\cosh^k(z\sqrt{\beta/c}+\beta/c)}{\int\mathrm{Dz}\ \cosh^k(z\sqrt{\beta/c}+\beta/c)}\right\}\right]\Big\rangle_{\tau=\pm1} \qquad \text{(C.49)}
\end{aligned}
$$

The first potential type of bifurcation away from $W(h)=\delta(h)$ would have $\int\mathrm{d}h\ W(h)h=\epsilon\int\mathrm{d}y\ \tilde{W}(y)y\equiv\epsilon m_1\neq0$. However, we see that mutiplying both sides of (C.49) by y, followed by integration, immediately gives $m_1=0$. Thus, as expected, a bifurcation leading to a function $W(h)$ with $\int\mathrm{d}h\ W(h)h\neq0$ is impossible.

Any continous bifurcation will consequently have $\int\mathrm{d}h\ W(h)h=0$ and $\int\mathrm{d}h\ W(h)h^2=\epsilon^2\int\mathrm{d}y\ \tilde{W}(y)y^2\equiv\epsilon^2m_2\neq0$. Multiplication of equation (C.49) by y^2, followed by integration over y gives

$$
\begin{aligned}
m_2 &= \mathrm{e}^{-c}\sum_{k\geq0}\frac{c^k}{k!}\mathrm{e}^{-\alpha ck}\sum_{r\geq0}\frac{(\alpha c)^r}{r!}\int_{-\infty}^{\infty}\mathrm{d}y_1\ldots\mathrm{d}y_r\Big[\prod_{s\leq r}\tilde{W}(y_s)\Big]\sum_{\ell_1\ldots\ell_r\leq k}\\
&\times\sum_{s\leq r}y_s^2\Big\langle\left\{\frac{\int\mathrm{Dz}\ \tanh(z\sqrt{\beta/c}+\beta/c)\cosh^k(z\sqrt{\beta/c}+\beta/c)}{\int\mathrm{Dz}\ \cosh^k(z\sqrt{\beta/c}+\beta/c)}\right\}^2\Big\rangle_{\tau=\pm1}. \qquad \text{(C.50)}
\end{aligned}
$$

So now we get a bifurcation when

$$
1 = \alpha c^2\sum_{k\geq0}e^{-c}\frac{c^k}{k!}\left\{\frac{\int Dz\ \tanh(z\sqrt{\beta/c}+\beta/c)\cosh^{k+1}(z\sqrt{\beta/c}+\beta/c)}{\int Dz\ \cosh^{k+1}(z\sqrt{\beta/c}+\beta/c)}\right\}^2. \qquad \text{(C.51)}
$$

We note that the right-hand side of (C.51) obeys $\lim_{\beta\to 0}\mathrm{RHS}=0$ and $\lim_{\beta\to\infty}\mathrm{RHS}=\alpha c^2$. Hence a transition at finite temperature $T_c(\alpha,c)>0$ exists to a new state with $W(h)\neq\delta(h)$ as soon as $\alpha c^2>1$. The critical temperature becomes zero when $\alpha c^2=1$, so $T_c(\alpha,1/\sqrt{\alpha})=0$ for all $\alpha\geq 0$. For large c, using $\tanh x=x+\mathcal{O}(x^3)$ and $\cosh x=1+x^2/2$, valid for small x, we have $T_c=\sqrt{\alpha}$.

C.8 Coincidence of the two formulae for the transition line

In order to prove that the two expressions (C.51) and (C.26) for the RS transition line are identical, as they should be, we show that

$$\left\{\frac{\int_{-\pi}^{\pi}\mathrm{d}\omega\ \sin^2(\omega)\cos^k(\omega)D(\omega|\beta)}{\int_{-\pi}^{\pi}\mathrm{d}\omega\ \cos^{k+2}(\omega)D(\omega|\beta)}\right\}^2=\left\{\frac{\int\mathrm{D}z\ \tanh(z\sqrt{\beta/c}+\beta/c)\cosh^{k+1}(z\sqrt{\beta/c}+\beta/c)}{\int\mathrm{D}z\ \cosh^{k+1}(z\sqrt{\beta/c}+\beta/c)}\right\}^2 \quad (C.52)$$

where $\mathrm{D}z=(2\pi)^{-1/2}\mathrm{e}^{-z^2/2}\,dz$. We can rewrite the argument of the curly brackets on the right-hand side, which we will denote as A, as

$$\begin{aligned}A &= \frac{\int\mathrm{D}z\ \sinh(z\sqrt{\beta/c}+\beta/c)\cosh^k(z\sqrt{\beta/c}+\beta/c)}{\int\mathrm{D}z\ \cosh^{k+1}(z\sqrt{\beta/c}+\beta/c)}\\ &= \frac{\int\mathrm{D}z\ \langle\tau_{k+1}e^{(z\sqrt{\beta/c}+\beta/c)\sum_{\ell\leq k+1}\tau_\ell}\rangle_{\tau_1\ldots\tau_{k+1}=\pm 1}}{\int\mathrm{D}z\ \langle\mathrm{e}^{(z\sqrt{\beta/c}+\beta/c)\sum_{\ell\leq k+1}\tau_\ell}\rangle_{\tau_1\ldots\tau_{k+1}=\pm 1}}\\ &= \frac{\langle\tau_{k+1}\mathrm{e}^{(\beta/2c)(\sum_{\ell\leq k+1}\tau_\ell)^2+(\beta/c)\sum_{\ell\leq k+1}\tau_\ell}\rangle_{\tau_1\ldots\tau_{k+1}=\pm 1}}{\langle\mathrm{e}^{(\beta/2c)(\sum_{\ell\leq k+1}\tau_\ell)^2+(\beta/c)\sum_{\ell\leq k+1}\tau_\ell}\rangle_{\tau_1\ldots\tau_{k+1}=\pm 1}},\end{aligned} \quad (C.53)$$

where we have carried out the Gaussian integrations. Next we insert $1=\sum_{M\in\mathbb{Z}}\delta_{M,\sum_{\ell\leq k+1}\tau_\ell}$, and write the Kronecker delta in integral form. This gives

$$\begin{aligned}A &= \frac{\sum_{M\in\mathbb{Z}}\mathrm{e}^{(\beta/2c)M^2+(\beta/c)M}\int_{-\pi}^{\pi}\mathrm{d}\omega\,\mathrm{e}^{\mathrm{i}\omega M}\langle\tau_{k+1}\mathrm{e}^{-\mathrm{i}\omega\sum_{\ell\leq k+1}\tau_\ell}\rangle_{\tau_1\ldots\tau_{k+1}=\pm 1}}{\sum_{M\in\mathbb{Z}}\mathrm{e}^{(\beta/2c)M^2+(\beta/c)M}\int_{-\pi}^{\pi}\mathrm{d}\omega\ \mathrm{e}^{\mathrm{i}\omega M}\langle\mathrm{e}^{-\mathrm{i}\omega\sum_{\ell\leq k+1}\tau_\ell}\rangle_{\tau_1\ldots\tau_{k+1}=\pm 1}}\\ &= -\mathrm{i}\,\frac{\sum_{M\in\mathbb{Z}}\mathrm{e}^{(\beta/2c)M^2+(\beta/c)M}\int_{-\pi}^{\pi}\mathrm{d}\omega\ \mathrm{e}^{\mathrm{i}\omega M}\cos^k(\omega)\sin(\omega)}{\sum_{M\in\mathbb{Z}}\mathrm{e}^{(\beta/2c)M^2+(\beta/c)M}\int_{-\pi}^{\pi}\mathrm{d}\omega\ \mathrm{e}^{\mathrm{i}\omega M}\cos^{k+1}(\omega)}.\end{aligned} \quad (C.54)$$

By completing the square, $\sum_M\mathrm{e}^{(\beta/2c)M^2+(\beta/c)M}=\mathrm{e}^{-\beta/(2c)}\sum_M e^{(\beta/2c)(M+1)^2}$, shifting the summation index $M\to M-1$, and using the symmetry properties (2.91) of $D(\omega|\beta)$ at zero fields, we finally get

$$\begin{aligned} A &= -\mathrm{i}\frac{\sum_{M\in\mathbb{Z}} \mathrm{e}^{(\beta/2c)M^2} \int_{-\pi}^{\pi}\mathrm{d}\omega\ \mathrm{e}^{\mathrm{i}\omega(M-1)}\cos^k(\omega)\sin(\omega)}{\sum_{M\in\mathbb{Z}} \mathrm{e}^{(\beta/2c)M^2} \int_{-\pi}^{\pi}\mathrm{d}\omega\ \mathrm{e}^{\mathrm{i}\omega(M-1)}\cos^{k+1}(\omega)} \\ &= -\mathrm{i}\frac{\int_{-\pi}^{\pi}\mathrm{d}\omega\ D(\omega|\beta)\cos^k(\omega)\sin(\omega)[\cos(\omega)-\mathrm{i}\sin(\omega)]}{\int_{-\pi}^{\pi}\mathrm{d}\omega\ D(\omega|\beta)\cos^{k+1}(\omega)[\cos(\omega)-\mathrm{i}\sin(\omega)]} \\ &= -\frac{\int_{-\pi}^{\pi}\mathrm{d}\omega\ D(\omega|\beta)\cos^k(\omega)\sin^2(\omega)}{\int_{-\pi}^{\pi}\mathrm{d}\omega\ D(\omega|\beta)\cos^{k+2}(\omega)}, \end{aligned} \tag{C.55}$$

which proves (C.52).

Appendix D

Selection of a state

While we have shown that, in the thermodynamic limit, the (intensive) energies associated to the two states that we used as example (the ferromagnetic and the mixture states) do coincide, thus thermodynamically the metastable state is not forbidden (while its weight is negligible w.r.t. the ferromagnetic scenario, and the system must be trapped opportunely with external fields in its basin to keep it in the large k limit), we still have to face the following addressable question: Let us consider the mixture state and approach the critical region from the ergodic scenario: the two clusters differ in magnetization, one has $m_1 > 0$ and the other $m_1 < 0$ and there is the just the upper (hence weakest) link connecting them. Maybe that one cluster acts on the other playing as an external field in mean-field schemes, thus selecting the phase /reversing the other cluster magnetization sign), which would result in destruction of mixture states? Aim of this note is to show that this is not the case.
The way we pave to prove this statement is the following: at first we will address this question within the more familiar mean-field perspective (namely considering the Curie-Weiss model), then we will enlarge the observation stemmed in that example toward bipartite ferromagnetic systems and 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 so to enlarge to the present model the stability argument.

Under the critical temperature systems of spins whose dynamics is no longer ergodic have 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 them, according with its initial configuration. As far as ferromagnetic systems are concerned, adding a suitable external field, it is possible to select one of this pure states,

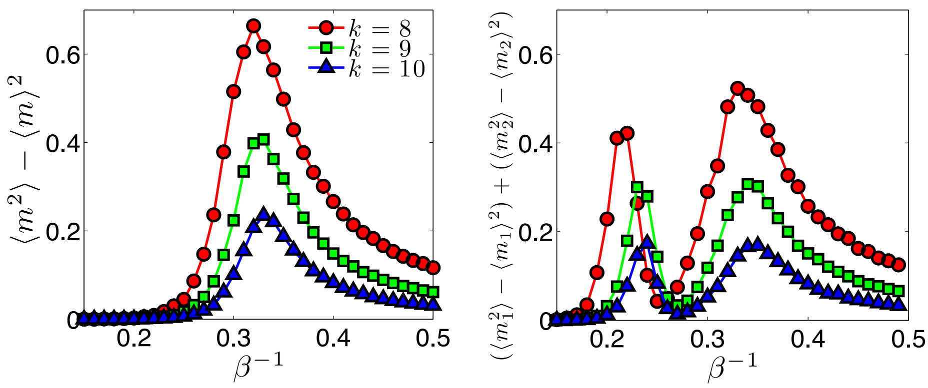

Figure D.1: Analysis of the susceptibility of the system, defined as $X = \langle m^2 \rangle - \langle m \rangle^2$, versus the noise level $T \equiv \beta^{-1}$, for various sizes (as reported in the legend) and $\rho = 0.99$. Left panel: $X(T)$ for the pure state. Right panel: $X(T)$ for the mixture state. Note that, while in the ferromagnetic (pure) case all the cuspids are on the same noise level whatever k, this is not the case for the mixture state because such a state is metastable as the difference in the energy ΔE among the two states scales as $\propto 1/N^{2\rho-1}$, hence only for $k \to \infty$ the mixture state becomes stable and its cuspid happens at the same noise level of the pure counterpart.

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 or not. Consider a Glauber dynamics for a ferromagnetic system of N spins at zero temperature:

$$\sigma_i(t+1) = \text{sgn}(h_i(S(t)) + h_N). \tag{D.1}$$

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 $m(S(t)) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i(t)$. In that case each initial configuration for which $|h_N| > |h_i(S^0)|$ will follow the external field. In general (if for example $|h_N| < 1$) there will always exist initial configuration ($|h_N| < |h_i(S^0)|$) that do not feel the influence of the external field, but, if we choose the initial configuration randomly and accordingly to $P_N(S^0)$, we can say that the field h_N selects the state if

$$P_N(S^0 : |h_N| > |h_i(S^0)|) \stackrel{N\to\infty}{\to} 1. \tag{D.2}$$

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

$$P_N(S^0 : |h_N| < |h_i(S^0)|) \stackrel{N\to\infty}{\to} 1. \tag{D.3}$$

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

Theorem D.0.1. *In the CW model, where* $h_i(S) = m(S) = \frac{1}{N}\sum_{i=1}^N \sigma_i$, $\forall \epsilon > 0$,

$h_N : |h_N| > \frac{1}{N^{\frac{1}{2}(1-\epsilon)}}$ *selects the state;*

$h_N : |h_N| < \frac{1}{N^{\frac{1}{2}(1+\epsilon)}}$ *does not select the state.*

For what concerns the first statement we note that, if $|h_N| > \frac{1}{N^{\frac{1}{2}(1-\epsilon)}}$

$$\begin{aligned} P_N(S^0 : |h_N| > |h_i(S^0)|) &= 1 - P_N\left(S^0 : |h_i(S^0)| > |h_N|\right) \geq 1 - P_N\left(S^0 : |m(S^0)| > \frac{1}{N^{\frac{1}{2}(1-\epsilon)}}\right) \\ &\geq 1 - N^{1-\epsilon}\mathbb{E}_N[m^2(S^0)] = 1 - N^{-\epsilon} \overset{N\to\infty}{\to} 1, \end{aligned} \tag{D.4}$$

where we used Chebyshev inequality and the fact that $\mathbb{E}_N[m^2(S^0)] = \frac{1}{N}$. For the second statement we note that, since $|h_N| < \frac{1}{N^{\frac{1}{2}(1+\epsilon)}}$,

$$\begin{aligned} P_N(S^0 : |h_N| > |h_i(S^0)|) &\leq P_N\left(S^0 : |m(S^0)| < \frac{1}{N^{\frac{1}{2}(1+\epsilon)}}\right) \\ &= P_N\left(S^0 : |\sqrt{N}m(S^0)| < \frac{1}{N^{\frac{\epsilon}{2}}}\right) \\ &\to \mu_{\mathcal{N}(0,1)}\left(|z| < \frac{1}{N^{\frac{\epsilon}{2}}}\right) \overset{N\to\infty}{\to} 0, \end{aligned} \tag{D.5}$$

where we just used the fact that the variable $\sqrt{N}m(S^0) = \frac{1}{\sqrt{N}}\sum_{i=1}^N S_i^0$ satysfies the 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 ask in which case J_{12} is able to 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

$$\sigma_i(t+1) = \mathrm{sgn}(h_i(S(t)) + h_N) = \mathrm{sgn}(J_{11}m_1(S(t)) + J_{12}m_2(S(t))), \tag{D.6}$$

i.e. we can repeat the same argument of the CW model identifying the field sent by the second party (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

$$J_{12}(N)|m_2(S^0)| > \frac{1}{N^{\frac{1}{2}(1-\epsilon)}}, \tag{D.7}$$

with probability 1. Since for the CLT $|m_2(S^0)|$ is $\mathcal{O}(\sqrt{N})$ with probability one, 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, namely 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 $J(N) = N^{1-2\rho}$ is vanishing in the thermodynamic limit, the two subgroups behave as they were non interacting: this may puzzle about the phase transition as the system -when not trapped within the pure state- crossing the critical line (in the β, ρ plane) moves from an ergodic region -where the global magnetization is zero- toward a mixture state where again is zero. However, regarding the latter, the two sub-clusters have not-zero magnetizations and even in this case, crossing the line returns in a canonical phase transition. 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 Fig. D.

Bibliography

[1] Abramowitz, M. and Stegun, IA *Handbook of Mathematical Functions* (Dover, New York) (1972)

[2] D.H. Ackley, G.E. Hinton, T.J. Sejnowski, *A learning algorithm for boltzmann machines*, Cognitive Science **9**:147-169 (1985).

[3] E. Agliari, A. Annibale, A. Barra, A.C.C Coolen, D. Tantari, *Immune networks: multitasking capabilities at medium load*, J. Phys. A: Math. Theor. 46 335101 (2013)

[4] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, F. Moauro, *Multitasking associative networks*, Phys. Rev. Lett. **109**(26), 268101 (2012).

[5] E. Agliari, A. Annibale, A. Barra, A.C.C. Coolen, D. Tantari, *Immune networks: multitasking capabilities near saturation*, J. Phys. A **46**(41), 415003 (2013).

[6] E. Agliari, A. Barra, A. De Antoni, A. Galluzzi, *Parallel retrieval of correlated patterns: From Hopfield networks to Boltzmann machines*, Neural Networks **38**, 52-63 (2013).

[7] E. Agliari, A. Barra, F. Guerra, F. Moauro, *A thermodynamical perspective of immune capabilities*, J. Theor. Biol. **287**, 48, 2011.

[8] E. Agliari, A. Barra, G. Del Ferraro, F. Guerra, D. Tantari, *Anergy in self-directed B lymphocytes: A statistical mechanics perspective*, to appear in J. Theor. Biol. $doi.org$/10.1016/$j.jtbi$.2014.05.006, (2014).

[9] E.Agliari, A. Barra, A.Galluzzi, A.Pizzoferrato, D.Tantari, *Ferromagnetic models for cooperative behavior: Revisiting Universality in complex phenomena*, proceedings of the Conference "Mathematical models and methods for Planet Heart", INdAM (2013)

[10] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani, *From Dyson to Hopfield: Processing on hierarchical networks*, to appear in to Physical Review Letters (2014).

[11] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani, *Meta-stable states in the hierarchical Dyson model drive parallel processing in the hierarchical Hopfield network*, to appear in J. Phys. A: Math. Theor (2014).

[12] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani, *Hierarchical neural networks perform both serial and parallel processing*, Submitted to Neural Networks (2014).

[13] E. Agliari et al., *Parallel processing in immune networks*, Phys. Rev. E **87**(4), 042701 (2013).

[14] S. Albeverio, B. Tirozzi, B. Zegarlinski *Rigorous results for the free energy in the Hopfield model*, Comm. Math. Phys. **150**, 337 (1992).

[15] D.J. Amit, *Modeling brain function: The world of attractor neural network*, Cambridge Univerisity Press, (1992).

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

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

[18] M.C. Angelini, G. Parisi, F. Ricci-Tersenghi, *Ensemble renormalization group for disordered systems*, Phys. Rev. B **87**:13, 134201 (2013).

[19] M.C. Angelini, F. Ricci-Tersenghi, G. Parisi, *Relations between short-range and long-range Ising models*, Phys. Rev. E **89**, 062120, (2014).

[20] Barra A 2008 *J. Stat. Phys.* **132** 787-809

[21] A. Barra, *Irreducible free energy expansion and overlap locking in mean field spin glasses*, J. Stat. Phys. **123**, 601-614 (2006).

[22] A.Barra, G.Del Ferraro, D.Tantari, *Mean field spin glasses treated with PDE techniques*, The European Physical Journal B 86:332 (2013).

[23] A. Barra, G. Genovese, F. Guerra, *Equilibrium statistical mechanics of bipartite spin systems*, J. Phys. A: Math. Theor. **44**, 245002 (2011).

[24] A. Barra, G. Genovese, F. Guerra, D. Tantari, *About a Solvable Mean Field Model of a Gaussian Spin Glass*, J. Phys. A **47**, 155002 (2014).

[25] A. Barra, G. Genovese, F. Guerra, D. Tantari, *How glassy are neural networks?*, J. Stat. Mech. P07009 (2012).

[26] A. Barra, G. Genovese, F. Guerra, *The replica symmetric approximation of the analogical neural network*, J. Stat. Phys. **140**:784-796, (2010).

[27] A. Barra, P.Contucci, E.Mingione, D.Tantari, *Multi-species mean-field spin-glasses. Rigorous results*, to appear in Annales Henry PoincarÃ¨ (2014).

[28] A.Barra, A. Galluzzi, F. Guerra, A. Pizzoferrato and D. Tantari, *Mean field bipartite spin models treated with mechanical techniques*, Eur. Phys. J. B 87: 74 (2014)

[29] A.Barra, A. Galluzzi, A. Pizzoferrato and D. Tantari, *A walk in the statistical mechanical formulation of neural networks*, Contribute to the proceeding of the conference: NCTA (2014).

[30] A. Barra, F. Guerra, *About the ergodic regime in the analogical Hopfield neural networks. Moments of the partition function*, J. Math. Phys. **49**, 125217 (2008).

[31] G. Ben Arous, A. Dembo, A. Guionnet, *Aging of spherical spin glasses*, Prob. Theor. Related Fields 120, 1, (2001).

[32] P.M. Bleher, J.G. Sinai, *Comm. Math. Phys.* **33**, 23 (1973).

[33] D. Bollé, T. M. Nieuwenhuizen, I. Perez-Castillo, T . Verbeiren, *A spherical Hopfield model*, J. Phys. A **36**, 10269, (2003).

[34] A. Bovier, B. Niederhauser, *The spin-glass phase-transition in the Hopfield model with p-spin interactions*, Adv. Theor. Math. Phys. **5**, $1001 - 1046$ (2001).

[35] A. Bovier, A.C.D. van Enter and B. Niederhauser, *Stochastic symmetry-breaking in a Gaussian Hopfield-model*, J. Stat. Phys. **95**, 181-213 (1999).

[36] A. Bovier, V. Gayrard *An almost sure central limit theorem for the Hopfield model*, Markov Proc. Rel. Fields **3**, 151-173 (1997).

[37] A. Bovier *Self-averaging in a class of generalized Hopfield models*, J. Phys. A **27**, 7069-7077 (1994).

[38] E. Bullmore, O. Sporns, *Nat. Rev. Neurosci.* **13**, 366 (2012).

[39] A. N. Burkitt, *A review of the integrate-and-fire neuron model*, Biol Cybern **95**, 1, (2006).

[40] M. Castellana, A. Barra, F. Guerra, *Free-energy bounds for Hierarchical Spin Models*, J. Stat. Phys. **155**, 211 (2014).

[41] M. Castellana, A. Decelle, S. Franz, M Mezard, G. Parisi, *Hierarchical random energy model of a spin glass*, Phys. Rev. Lett. **104**(12), 127206 (2012).

[42] M. Castellana, *Real-space renormalization group analysis of a non-mean-field spin-glass*, Europhys. Lett. **95**(4), 47014 (2011).

[43] P. Collet, J-P. Eckmann, *The ϵ-expansion for the Hierarchical Model*, Comm. Math. Phys. **55**:1, 67-96, (1977).

[44] C. Conaco, et al., *Proc. Natl. A. Soc.* **109**, 10612 (2012).

[45] Coolen ACC 2001 in *Handbook of Biological Physics 4* (F Moss and S Gielen Eds, Elsevier, Amsterdam) 531-59

[46] A.C.C. Coolen, A.J. Noest, G.B. de Vries, *Modelling Chemical Modulation of Neural Processes*, Network **4**, 101, (1993).

[47] A.C.C. Coolen, R. Kuehn, P. Sollich, *Theory of neural information processing systems*, Oxford Press (2005).

[48] A. Decelle, G. Parisi, J. Rocchi, *Ensemble renormalization group for the random-field hierarchical model*, Phys. Rev. E **89**:3, 032132, (2014).

[49] D. De Martino, M. Figliuzzi, A. De Martino, E. Marinari, *Computing fluxes and chemical potential distributions in biochemical networks: energy balance analysis of the human red blood cell*, submitted. Available at arxiv:1107.2330 (2012).

[50] A. Di Biasio, E. Agliari, A. Barra, R. Burioni, *Cooperativity in chemical kinetics*, Theor. Chem. Acc. **131**, 1104, (2012).

[51] R. Dietrich, M. Opper, and H. Sompolinsky. *Statistical mechanics of support vector networks*, Phys. rev. lett. **82**, 2975, (1999).

[52] G.W. Domhoff, *Neural networks, cognitive development, and content analysis*, American Psychological Association (2003).

[53] F. Dyson, *Existence of a phase transition in a one dimensional Ising ferromagnet*, Comm. Math. Phys. **12**, 91-107 (1969).

[54] R.S. Ellis, *Large deviations and statistical mechanics*, Springer, New York, 1985.

[55] U. Ferrari, et al. *Phys. Rev. B* **88**, 184201, (2013).

[56] S. Franz, T. Jorg, G. Parisi, *Overlap interfaces in hierarchical spin-glass models*, J. Stat. Mech. P02002 (2009).

[57] G. Gallavotti, S. Miracle-Sole', *Statistical mechanics of lattice systems*, Comm. Math. Phys. **5**(5):317-323 (1967).

[58] L. K. Gallos, et al., *Proc. Natl. A. Soc.* **104**, 7746 (2007).

[59] G. Genovese, A. Barra, *A mechanical approach to mean field spin models*, J. Math. Phys. **50**, 365234 (2009).

[60] G.Genovese, D. Tantari, *Spherical and Gaussian Spin Glasses*, Submitted to Mathematical Physics, Analysis and Geometry (2014).

[61] R. B. Griffiths, *Correlations in Ising ferromagnets. III*, Comm. Math. Phys. **6**(2):121-127 (1967).

[62] R. B. Griffiths, *Correlations in Ising ferromagnets. II: External magnetic fields*, J. Math. Phys **8**:484 (1967).

[63] F. Guerra, *An Introduction to Mean Field Spin Glass Theory: Methods and Results*, A. Bovier et al. eds, Les Houches, Session LXXXIII, (2005).

[64] F. Guerra, *Sum rules for the free energy in the mean field spin glass model*, in *Mathematical Physics in Mathematics and Physics: Quantum and Operator Algebraic Aspects*, Fields Institute Communications **30**, American Mathematical Society (2001).

[65] F. Guerra, *Broken Replica Symmetry Bounds in the Mean Field Spin Glass Model*, Commun, Math. Phys. **233:1**, 1-12 (2003).

[66] F. Guerra, F. L. Toninelli, *The Thermodynamic Limit in Mean Field Spin Glass Models, Comm. Math. Phys.* **230**, 71 (2002).

[67] M.T. Hagan, H.B. Demuth, M.H. and Beale, *Neural network design*, Pws Pub., Boston. Mass. (1996).

[68] D.O. Hebb, *The organization of behavior: A neuropsychological theory*, Psychology Press (1940).

[69] J. Hertz, R. Palmer, *Introduction to the theory of neural networks*, Lecture Notes (1991).

[70] J.J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, P.N.A.S. **79**, 2554-2558, (1982).

[71] D. Kim, C.J. Thompson, *Critical properties of Dyson's hierarchical model*, J. Phys. A **10**:9, 1579, (1977).

[72] A. Kumar, et al., *Trends Neurosci.* **36**, 579 (2013).

[73] J.A. Acebron, L. Bonilla, C. Perez-Vicente, Prez; F. Ritort, R. Spigler, *The Kuramoto model: a simple paradigm for synchronization phenomena*, Rev. Mod. Phys. **77**, 137, (2005).

[74] C. Martindale, *Cognitive psychology: A neuralnetwork approach*, Thomson Brooks/Cole Publishing, (1991).

[75] W.S. McCulloch, W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bull. Math. Biophys. **5**(4):115-133 (1943).

[76] F.L. Metz, L. Leuzzi, G. Parisi, V. Sacksteder, *Transition between localized and extended states in the hierarchical Anderson model*, Phys. Rev. B **88**, 045103, (2013).

[77] F.L. Metz, L. Leuzzi, G. Parisi, *The renormalization flow of the hierarchical Anderson model at weak disorder*, Phys. Rev. B **89**, 064201, (2014).

[78] M.Mezard, A.Montanari *Information, Physics and Computation*, Oxford University Press, New York, (2009).

[79] M. Mezard, G. Parisi, M.A. Virasoro, *Spin glass theory and beyond*, World Scientific, Singapore, Lect. Notes Phys. 9 (1987).

[80] Mézard M, Parisi, G 2001 *Eur. Phys. J. B* **20** 217

[81] W.T. Miller, P.J. Werbos, R.S. Sutton, *Neural networks for control*, MIT press (1995).

[82] P. Moretti, M.A. Munoz, *Nature Comm.* **4**, 2521 (2013).

[83] C. Monthus, T. Garel, *Dynamical barriers in the Dyson hierarchical model via real space renormalization*, J. Stat. Mech. P02023, (2013).

[84] C. Monthus, T. Garel, *Scaling of the largest dynamical barrier in the one-dimensional long-range Ising spin-glass*, Phys. Rev. B **89**, 014408, (2014).

[85] D. Mukamel, *Notes on the Statistical Mechanics of Systems with Long-Range Interactions*, Les Houches Lecture Notes, arXiv:0905.1457, (2008).

[86] M. E. J. Newman, et al., *Phys. Rev. E*, **64(2)**:026118, (2001).

[87] M. Newman, *Networks*, Oxford University Press (2010).

[88] D. Panchenko, *The Sherrington-Kirkpatrick model*, Springer-Verlag, New York, (2013).

[89] Pastur, L., Shcherbina, M., and Tirozzi, B. *The replica-symmetric solution without replica trick for the hopfield model.* Journal of Statistical Physics, **74**:1161-1183 (1994).

[90] L.Pastur, M. Scherbina, B. Tirozzi, *On the replica symmetric equations for the Hopfield model* J. Math. Phys. **40**, 3930-3947 (1999).

[91] G. Parisi, J. Rocchi J, *Critical exponents of the random field hierarchical model*, Phys. Rev. B **90**, 024203, (2014).

[92] E.T. Rolls, A. Treves, *Neural networks and brain function*, Cambridge Press (1988).

[93] F. Rosenblatt, *The perceptron: a probabilistic model for information storage and organization in the brain*, Psychol. Rev. **65**(6):386 (1958).

[94] D. Ruelle. *Statistical Mechanics. Rigorous results*, W.A. Benjamin Inc., New York, 1969.

[95] P. Sollich, D. Tantari, A. Annibale, A. Barra, *Extensive load in multitasking associative networks*, to appear in PRL (2014).

[96] M. Talagrand, *Rigourous results for the Hopfield model with many patterns*, Probab. Th. Relat. Fields **110**, 177-276 (1998).

[97] M. Talagrand, *Exponential inequalities and convergence of moments in the replica-symmetric regime of the Hopfield model*, Ann. Probab. **38**, 1393-1469 (2000).

[98] M. Talagrand, *Spin glasses: a challenge for mathematicians. Cavity and mean field models*, Springer-Verlag, (2003).

[99] H.C.Tuckwell, *Introduction to theoretical neurobiology, Vol.*1*, Vol.*2, Cambridge University Press (2005).

[100] A. Vespignani, G. Caldarelli, *Large Scale Structure and Dynamics of Complex Networks*, World Sci. Publ. (2007).

[101] M. Zhao, et al., *Phys. Rev. E* **84**, 016109 (2011).