

## Multitasking attractor networks with neuronal threshold noise

Elena Agliari<sup>a,b</sup>, Adriano Barra<sup>c,d,\*</sup>, Andrea Galluzzi<sup>e</sup>, Marco Isopi<sup>e</sup>

<sup>a</sup> Università di Parma, Dipartimento di Fisica, Italy

<sup>b</sup> Università di Parma, INFN Gruppo di Parma, Italy

<sup>c</sup> Sapienza Università di Roma, Dipartimento di Fisica, Italy

<sup>d</sup> Sapienza Università di Roma, GNFM Gruppo di Roma, Italy

<sup>e</sup> Sapienza Università di Roma, Dipartimento di Matematica, Italy

### ARTICLE INFO

#### Article history:

Received 16 April 2013

Revised and accepted 22 September 2013

#### Keywords:

Hopfield model

Statistical mechanics

Multitasking networks

### ABSTRACT

We consider the multitasking associative network in the low-storage limit and we study its phase diagram with respect to the noise level  $T$  and the degree  $d$  of dilution in pattern entries. We find that the system is characterized by a rich variety of stable states, including pure states, parallel retrieval states, hierarchically organized states and symmetric mixtures (remarkably, both even and odd), whose complexity increases as the number of patterns  $P$  grows. The analysis is performed both analytically and numerically: Exploiting techniques based on partial differential equations, we are able to get the self-consistencies for the order parameters. Such self-consistency equations are then solved and the solutions are further checked through stability theory to catalog their organizations into the phase diagram, which is outlined at the end. This is a further step towards the understanding of spontaneous parallel processing in associative networks.

© 2013 Elsevier Ltd. All rights reserved.

## 1. Introduction

The paradigm, introduced almost three decades ago by Amit, Gutfreund and Sompolinsky (Amit, 1989; Amit, Gutfreund, & Sompolinsky, 1985), of analyzing neural networks through techniques stemming from statistical mechanics of disordered systems (mainly the celebrated Replica Trick Mézard, Parisi, & Virasoro, 1987 for the Hopfield model Hopfield, 1982) has been so prolific that its applications went far beyond Artificial Intelligence and Robotics, overlapping Statistical Inference (Cheng & Titterington, 1994), System Biology (Nolfi & Floreano, 2000), Financial Market Planning (Trippi & Turban, 1992), Theoretical Immunology (Agliari, Barra, Guerra, & Moauro, 2011) and much more.

As a result, research in this field is under continuous development, ranging from the diverse applications outlined above, to a deeper and deeper understanding of the core-theory behind. For the sake of reaching results closer to experimental neuroscience outcomes, scientists involved in the field tried to bypass the rather crude mean field description of a fully connected network of interacting neurons, embedding them in diluted topologies as Erdős-Rényi graphs (Sompolinsky, 1986), small-worlds (Nikoleopoulos et al., 2004) or even finitely connected graphs

(Wemmenhove & Coolen, 2003). The main point was showing robustness of the mean-field paradigm even in these diluted, and in some sense “closer to biology”, versions and this was indeed successfully achieved (with the exception of too extreme degrees of dilution, where the associative capacities of the network trivially break down).

Recently, a mapping between Hopfield networks and Boltzmann machines (Barra, Bernacchia, Santucci, & Contucci, 2012) allowed the introduction of dilution into associative networks from a different perspective with respect to standard link removal à la Sompolinsky (1986) or à la Coolen (Nikoleopoulos et al., 2004; Wemmenhove & Coolen, 2003). In fact, while in their papers these authors perform dilution directly on the Hopfield network, through the equivalence with the Boltzmann machine, one may perform link dilution on the Boltzmann machine and then map the latter back into the associative Hopfield-like network (Agliari, Barra, Galluzzi, Guerra, & Moauro, 2012). Remarkably, the resulting model still works as an associative performer, as the Hebbian structure is preserved, but its capabilities are quite different from the standard scenario. In particular, the resulting associative network may still be fully connected but the stored patterns of information display entries which, beyond coding information through digital values  $\pm 1$ , can also be blank (Agliari, Barra, De Antoni, & Galluzzi, 2012; Agliari, Barra, Galluzzi et al., 2012). In fact, any missing link in the bipartite Boltzmann machine corresponds to a blank entry in the related pattern of the associative network.

Now, while standard (i.e., performed directly on the Hopfield network) dilution does not change qualitatively the system

\* Corresponding author at: Sapienza Università di Roma, Dipartimento di Fisica, Italy. Tel.: +39 0649914366.

E-mail address: [adriano.barra@roma1.infn.it](mailto:adriano.barra@roma1.infn.it) (A. Barra).

performances, the behavior of the system resulting from hidden (i.e., performed on the underlying Boltzmann machine) dilution becomes “multitasking” because retrieval of a single pattern, say  $\xi^1$ , does not exhaust all the neurons, and the ones coupled with the blank entries of  $\xi^1$  are free to align with  $\xi^2$ , whose entries will partially be blank as well, hence eliciting, in turn, the retrieval of  $\xi^3$  and so on up to a parallel logarithmic (with respect to the volume of the network  $N$ ) load of all the stored patterns.

As a consequence, by tuning the degree of dilution in the hidden Boltzmann network and the level of noise in the directed network, the system exhibits a very rich phase diagram, whose investigation is the subject of the present work.

The paper is organized as follows. In Section 2, we review the multitasking networks introduced in Agliari, Barra, Galluzzi et al. (2012) highlighting their main features and providing a rigorous solution for their thermodynamics through a novel technique based on mapping the statistical mechanical problem into a diffusion problem and then solving the latter through standard partial differential equation methods. In Section 3, the solutions obtained in the previous section are investigated. In particular, we discuss the emergence of spurious states for these multitasking networks. Then, in Section 4, we describe the analytical technique used to study the stability of the retrieval states, which are found to be solutions of the system. In Section 5, agreement with Monte Carlo simulations is presented, while, finally, Section 6 is devoted to a summary and a discussion of the results.

## 2. The multitasking associative network

In the conventional Hopfield model (see, e.g., Amit, 1989, Coolen, Kühn, & Sollich, 2005), one considers a network of  $N$  neurons, where each neuron  $\sigma_i$  can take two states, namely,  $\sigma_i = +1$  (firing) and  $\sigma_i = -1$  (quiescent). Neuronal states are given by the set of variables  $\sigma = (\sigma_1, \dots, \sigma_N)$ . Each neuron is located on a node of a complete graph and the synaptic connection between two arbitrary neurons, say,  $\sigma_i$  and  $\sigma_j$ , is defined by the following Hebb rule (Amit, 1989):

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu, \quad (1)$$

where  $\xi^\mu = (\xi_1, \dots, \xi_N)$  denotes the set of memorized patterns, each specified by a label  $\mu = 1, \dots, P$ . The entries are dichotomic, i.e.,  $\xi_i^\mu \in \{+1, -1\}$ , chosen randomly and independently with equal probability, namely, for any  $i$  and  $\mu$ ,

$$P(\xi_i^\mu) = \frac{1}{2}(\delta_{\xi_i^\mu - 1} + \delta_{\xi_i^\mu + 1}), \quad (2)$$

where the Kronecker delta  $\delta_x$  equals 1 iff  $x = 0$ , otherwise it is zero. Patterns are assumed as quenched, that is, the performance of the network is analyzed keeping the synaptic values fixed.

The Hamiltonian describing this system is

$$\begin{aligned} H_N(\sigma, \xi) &= - \sum_{i=1}^N \sum_{i>j=1}^N J_{ij} \sigma_i \sigma_j \\ &= - \frac{1}{2N} \sum_{i,j=1}^{N,N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j, \end{aligned} \quad (3)$$

so that the signal (i.e. the field) acting on neuron  $i$  is

$$h_i(\sigma, \xi) = \sum_{\substack{j=1 \\ j \neq i}}^N J_{ij} \sigma_j. \quad (4)$$

The evolution of the system is ruled by a stochastic dynamics, according to which the probability that the activity of a neuron  $i$

assumes the value  $\sigma_i$  is

$$P(\sigma_i; \sigma, \xi, \beta) = \frac{1}{2}[1 + \tanh(\beta h_i \sigma_i)], \quad (5)$$

where  $\beta \equiv T^{-1}$  tunes the level of noise (hence  $T$  plays as temperature as in classical physics) such that for  $\beta \rightarrow 0$  the system behaves completely randomly, while for  $\beta \rightarrow \infty$  it becomes noiseless and deterministic; notice that the noiseless limit of Eq. (5) is  $\sigma_i(t+1) = \text{sign}[h_i(t)]$ .

The main feature of the model described by Eqs. (3) and (5) is its ability to work as an associative memory. More precisely, the patterns are said to be memorized if each of the network configurations  $\sigma_i = \xi_i^\mu$  for  $i = 1, \dots, N$ , for every one of the  $P$  patterns labeled by  $\mu$ , is a fixed point of the dynamics. Introducing the overlap  $m^\mu$  between the state of neurons  $\sigma$  and one of the patterns  $\xi^\mu$ , as

$$m^\mu = \frac{1}{N}(\sigma \cdot \xi^\mu) = \frac{1}{N} \sum_{i=1}^N \sigma_i \xi_i^\mu, \quad (6)$$

such a pattern is said to be retrieved if, in the thermodynamic limit,  $m^\mu = \mathcal{O}(1)$ . Given the definition (6), the Hamiltonian (3) can also be written as

$$H_N(\sigma, \xi) = -N \sum_{\mu=1}^P (m^\mu)^2 + P = -N m^2 + P. \quad (7)$$

The analytical investigation of the system is usually carried out in the thermodynamic limit  $N \rightarrow \infty$ , consistently with the fact that real networks are comprised of a very large number of neurons. Dealing with this limit, it is convenient to specify the relative number of stored patterns, namely  $P/N$  and to define the ratio  $\alpha = \lim_{N \rightarrow \infty} P/N$ . The case  $\alpha = 0$ , corresponding to a number  $P$  of stored patterns scaling sub-linearly with respect to the amount of performing neurons  $N$ , is often referred to as “low storage”. Conversely, the case of finite  $\alpha$  is often referred to as “high storage”. In particular, in the former case ( $\alpha = 0$ ), the overall behavior of the standard Hopfield model is ruled only by the noise  $T \equiv 1/\beta$  and the so-called pure-state ansatz

$$\mathbf{m} = (m, 0, \dots, 0), \quad (8)$$

always corresponds to a stable solution for  $T < 1$ ; the order in the entries is purely conventional and here we assume that the first pattern is the one stimulated.

Let us now move on and generalize the system described above in order to account for the existence of blank entries in the patterns  $\xi$ . More precisely, we replace Eq. (2) by

$$P(\xi_i^\mu) = \frac{1-d}{2} \delta_{\xi_i^\mu - 1} + \frac{1-d}{2} \delta_{\xi_i^\mu + 1} + d \delta_{\xi_i^\mu}, \quad (9)$$

where  $d$  encodes the degree of “dilution” in pattern entries. Patterns are still assumed as quenched and, of course, the definitions of the Hamiltonian (3) and of the overlaps (6), with the dynamics provided by (5) still hold.

As discussed in Agliari, Barra, De Antoni et al. (2012), Agliari, Barra, Galluzzi et al. (2012), this kind of extension has strong biological motivations and also yields highly non-trivial thermodynamic outcomes. In fact, the distribution in Eq. (2) necessarily implies that the retrieval of a unique pattern does employ all the available neurons, so that no resources are left for further tasks. Conversely, with Eq. (9) the retrieval of one pattern still allows available neurons (i.e., those corresponding to the blank entries of the retrieved pattern), which can be used to recall other patterns up to the exhaustion of all neurons. The resulting network is therefore able to process several patterns simultaneously.

In particular, in the low-storage regime, it was shown both analytically (via density of states analysis) and numerically (via Monte Carlo simulations) (Agliari, Barra, Galluzzi et al., 2012), that

the system evolves towards an equilibrium state where several patterns are simultaneously retrieved. In the noiseless limit  $T = 0$  and for  $d$  not too large, the equilibrium state is characterized by a hierarchical overlap

$$\mathbf{m} = (1 - d)(1, d, d^2, \dots, 0), \quad (10)$$

hereafter referred to as the “parallel ansatz”. On the other hand, in the presence of noise or for large degrees of dilution in pattern entries, this state ceases to be a stable solution for the system and different states, possibly spurious, emerge. The aim of this work is to highlight the equilibrium states of this system as a function of the parameters  $d$  and  $T$ , and finally build a phase diagram; for this task we develop, first, a rigorous mathematical treatment for calculating the free energy of the model and then obtain the self-consistencies constraining the phase diagram; then, we solve these equations both numerically and with a stability analysis. In this way we are able to draw the phase diagram, whose peculiarities lie in the stability of both even and odd mixtures of spurious states (in proper regions of the parameters) and the formation of parallel spurious states. Both these results generalize the standard counterpart of classical Hopfield networks.

Findings are double-checked through Monte Carlo runs that are in excellent agreement with the picture we obtained.

## 2.1. Statistical mechanics analysis through Fourier technique

We solve the general model described by the Hamiltonian (3), with patterns diluted according to (9), in the low storage regime  $P \sim \log N$ , such that the limit  $\alpha = \lim_{N \rightarrow \infty} P/N = 0$  holds.<sup>1</sup>

Due to the formal analogy with statistical-mechanics models for magnetic systems (Amit, 1989), in the following neurons will be also referred to as spins.

As standard in disordered statistical mechanics, we introduce three types of average for an observable  $o(\sigma, \xi)$ :

(i) the Boltzmann average  $\omega(o) = \sum_{\sigma} o(\sigma, \xi) \exp[-\beta \mathcal{H}(\sigma; \xi)] / Z_{N,P}(\beta, d)$ , where

$$Z_{N,P}(\beta, d) = \sum_{\{\sigma\}} \exp[-\beta H_N(\sigma, \xi)]$$

is called the “partition function”, (ii) the average  $\mathbb{E}$  performed over the quenched disordered couplings  $\xi$ , (iii) the global expectation  $\mathbb{E}\omega(o)$  defined by the brackets  $\langle o \rangle_{\xi}$ .

Given these definitions, for the average energy of the system  $E$  we can write  $E \equiv \lim_{N \rightarrow \infty} (\langle H_N(\sigma, \xi) \rangle / N)$ .

Also, we are interested in finding an explicit expression for the order parameters of the model, namely the averaged  $P$  Mattis magnetizations

$$\langle m^{\mu} \rangle = \lim_{N \rightarrow \infty} \mathbb{E}\omega \left( \frac{1}{N} \sum_{j=1}^N \xi_j^{\mu} \sigma_j \right). \quad (11)$$

For this task we need to introduce the statistical pressure

$$\alpha(\beta, d) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln(Z_{N,P}(\beta, d)),$$

which is immediately related to the free energy per site  $f(\beta, d)$  by the relation  $f(\beta, d) = -\alpha(\beta, d)/\beta$  because, by maximizing  $\alpha(\beta, d)$  with respect to the  $P$  magnetizations  $\langle m^{\mu} \rangle$ , we get exactly

<sup>1</sup> Results outlined within this scaling can be extended with little effort to the whole region  $P \sim N^{\gamma}$ , with  $\gamma < 1$ , such that the constraint  $\alpha = 0$  is preserved, as realized in the Willshaw model (Willshaw & von der Malsburg, 1976) concerning neural sparse coding.

Note further that there is a deep similarity with the Potts model with pairwise interaction (Wu, 1982).

the self-consistency equations for these order parameters, whose solutions will give us a picture of the phase diagram.

In the past decades, scientists involved in disordered statistical mechanics investigations, even beyond Artificial Intelligence, laid down several strands for solving this kind of problems, and nowadays a plethora of techniques is available. We extend early ideas of Guerra, on the lines developed in Genovese and Barra (2009), consisting of modeling disordered statistical mechanics through dynamical system theory and in particular, here, we are going to proceed as follows:

Our statistical-mechanics problem is mapped into a diffusive problem embedded in a  $P$ -dimensional space and with given, known, boundaries. We solve the diffusive problem via a standard Green-propagator technique, and then we will map back the obtained solutions in terms of their original statistical mechanics meaning.

For this task, let us introduce and consider a generalized Boltzmann factor  $B_N(\mathbf{x}, t)$  depending on  $P + 1$  parameters  $\mathbf{x}, t$  (which we think of as *generalized  $P$ -dimensional Euclidean space and time*)

$$B_N(\mathbf{x}, t; \xi, \sigma)$$

$$= \exp \left( \frac{t}{2N} \sum_{i \neq j}^N \sigma_i \sigma_j \sum_{\mu=1}^P \xi_i^{\mu} \xi_j^{\mu} + \sum_{\mu=1}^P x_{\mu} \sum_{j=1}^N \xi_j^{\mu} \sigma_j \right), \quad (12)$$

and the generalized statistical pressure

$$\alpha_N(\mathbf{x}, t) = \frac{1}{N} \ln \left[ \sum_{\{\sigma\}} B_N(\mathbf{x}, t; \xi, \sigma) \right]. \quad (13)$$

Notice that, for proper values of  $\mathbf{x}, t$ , namely  $\mathbf{x} = 0$  and  $t = \beta$ , classical statistical mechanics is recovered as

$$\alpha(\beta, d) = \lim_{N \rightarrow \infty} \alpha_N(\mathbf{x} = 0, t = \beta)$$

$$= \lim_{N \rightarrow \infty} \frac{1}{N} \ln \left[ \sum_{\{\sigma\}} B_N(\mathbf{x} = 0, t = \beta; \xi, \sigma) \right].$$

In the same way, the average  $\langle \cdot \rangle_{(\mathbf{x}, t)}$  will be denoted by  $\langle \cdot \rangle$ , whenever evaluated in the sense of statistical mechanics, namely

$$\langle o \rangle_{(\mathbf{x}, t)} = \frac{\sum_{\{\sigma\}} o(\sigma, \xi) B_N(\mathbf{x}, t; \xi, \sigma)}{\sum_{\{\sigma\}} B_N(\mathbf{x}, t; \xi, \sigma)}, \quad (14)$$

$$\langle o \rangle = \frac{\sum_{\{\sigma\}} o(\sigma, \xi) \exp[-\beta H(\sigma, \xi)]}{\sum_{\{\sigma\}} \exp[-\beta H(\sigma, \xi)]} = \langle o \rangle_{(\mathbf{x}=0, t=\beta)}. \quad (15)$$

It is immediate to see that the following equations hold:

$$\partial_t \alpha_N(\mathbf{x}, t) = \frac{1}{2} \sum_{\mu} \langle m_{\mu}^2 \rangle_{(\mathbf{x}, t)}, \quad (16)$$

$$\partial_{x_{\mu}} \alpha_N(\mathbf{x}, t) = \langle m_{\mu} \rangle_{(\mathbf{x}, t)},$$

and, defining a vector  $\Gamma_N(\mathbf{x}, t)$  of elements  $\Gamma_N^{\mu}(\mathbf{x}, t) \equiv -\partial_{x_{\mu}} \alpha_N(\mathbf{x}, t)$ , by construction  $\Gamma_N^{\mu}(\mathbf{x}, t)$  obeys the following equation:

$$\begin{aligned} \partial_t \Gamma_N^{\mu}(\mathbf{x}, t) + \sum_{v=1}^P \Gamma_N^v(\mathbf{x}, t) [\partial_{x_v} \Gamma_N^{\mu}(\mathbf{x}, t)] \\ = \frac{1}{2N} \sum_{v=1}^P \partial_{x_v}^2 \Gamma_N^{\mu}(\mathbf{x}, t), \end{aligned} \quad (17)$$

which happens to be in the form of a Burgers equation for the vector  $\Gamma_N(\mathbf{x}, t)$  with a kinematic viscosity  $(2N)^{-1}$ . As is well known,

the Burgers equation can be mapped into a  $P$ -dimensional diffusive problem using the Cole–Hopf transformation (Genovese & Barra, 2009) as follows:

$$\psi_N(\mathbf{x}, t) = \exp \left[ -N \int dx_\mu \Gamma_N^\mu(\mathbf{x}, t) \right] = \exp[N\alpha_N(\mathbf{x}, t)], \quad (18)$$

and its  $t$  and  $x$  streaming read off as

$$\begin{aligned} \partial_t \psi_N(\mathbf{x}, t) &= N(\partial_t \alpha_N(\mathbf{x}, t)) \psi(\mathbf{x}, t), \\ \partial_{x_\mu} \psi_N(\mathbf{x}, t) &= N(\partial_{x_\mu} \alpha_N(\mathbf{x}, t)) \psi(\mathbf{x}, t), \end{aligned} \quad (19)$$

in such a way that

$$\begin{aligned} \partial_{x_\mu x_\nu}^2 \psi_N(\mathbf{x}, t) &= N\psi_N(\mathbf{x}, t) \left\{ \partial_{x_\mu x_\nu}^2 \alpha_N(\mathbf{x}, t) + N[\partial_{x_\mu} \alpha_N(\mathbf{x}, t)] \right. \\ &\quad \times \left. [\partial_{x_\nu} \alpha_N(\mathbf{x}, t)] \right\}. \end{aligned} \quad (20)$$

Now, from Eqs. (19), (20) we get

$$\partial_t \psi_N(\mathbf{x}, t) - \frac{1}{2N} \sum_\mu \left[ \partial_{x_\mu}^2 \psi_N(\mathbf{x}, t) \right] = 0. \quad (21)$$

Therefore, we established a reformulation of the problem of calculating the thermodynamic potential  $\alpha(\beta, d)$  over the equilibrium configuration of the order parameters for an attractor network model in terms of a diffusion equation for the function  $\psi_N(\mathbf{x}, t)$ , namely the Cole–Hopf transform of the Mattis magnetizations, with a diffusion coefficient  $D = (2N)^{-1}$ , that is

$$\partial_t \psi_N(\mathbf{x}, t) - D\nabla^2 \psi_N(\mathbf{x}, t) = 0,$$

$$\psi_N(\mathbf{x}, 0) = \sum_{\{\sigma\}} \exp \left( \sum_\mu x_\mu \sum_j \xi_j^\mu \sigma_j \right). \quad (22)$$

We solve this Cauchy problem (22) through standard techniques: first, we map the diffusive equation in the Fourier space, then we calculate the Green propagator for the homogeneous configuration, and finally we will inverse-transform the solution.

Let us consider the Fourier transform:

$$\begin{aligned} \tilde{\psi}_N(\mathbf{k}, t) &= \int_{\mathbb{R}^P} d^P x \exp \left( -i \sum_\mu x_\mu k_\mu \right) \psi_N(\mathbf{x}, t), \\ \psi_N(\mathbf{x}, t) &= \frac{1}{(2\pi)^P} \int_{\mathbb{R}^P} d^P k \exp \left( i \sum_\mu x_\mu k_\mu \right) \tilde{\psi}_N(\mathbf{k}, t), \end{aligned} \quad (23)$$

and the related Green problem:

$$\partial_t \tilde{G}(\mathbf{k}, t) + Dk^2 \tilde{G}(\mathbf{k}, t) = \delta(t), \quad (24)$$

where  $\tilde{G}(\mathbf{k}, t)$  is the Green propagator in the  $k$ -space, which can be decomposed as

$$\tilde{G}(\mathbf{k}, t) = \tilde{G}_R(\mathbf{k}, t) + \tilde{G}_S(\mathbf{k}, t), \quad (25)$$

with  $\tilde{G}_R(\mathbf{k}, t)$  the general solution of the homogeneous problem and  $\tilde{G}_S(\mathbf{k}, t)$  a particular solution of the non-homogeneous problem. Hence, the full solution will be

$$\psi_N(\mathbf{x}, t) = \int_{\mathbb{R}^P} d^P x' G_R(\mathbf{x} - \mathbf{x}', t) \psi_N(\mathbf{x}', 0), \quad (26)$$

where the function  $\tilde{G}_R(\mathbf{k}, t)$  fulfills

$$\begin{aligned} \partial_t \tilde{G}_R(\mathbf{k}, t) - Dk^2 \tilde{G}_R(\mathbf{k}, t) &= 0, \\ \tilde{G}_R(\mathbf{k}, 0) &= 1, \end{aligned} \quad (27)$$

hence

$$\tilde{G}(\mathbf{k}, t) = \exp(-Dk^2 t),$$

$$G(\mathbf{x}, t) = \frac{1}{(2\sqrt{\pi D t})^P} \exp \left( \frac{-\mathbf{x}^2}{4Dt} \right). \quad (28)$$

Therefore, we get

$$\psi_N(\mathbf{x}, t) = \left( \frac{N}{2\pi t} \right)^{\frac{P}{2}} \int \left( \prod_{\mu=1}^P dx'_\mu \right) \exp \left[ -N\Phi(\mathbf{x}', \mathbf{x}, t) \right], \quad (29)$$

$$\begin{aligned} \Phi(\mathbf{x}', \mathbf{x}, t) &= \frac{\sum_\mu (x_\mu - x'_\mu)^2}{2t} - \ln 2 \\ &\quad - \frac{1}{N} \sum_{j=1}^N \ln \left[ \cosh \left( \sum_{\mu=1}^P x'_\mu \xi_j^\mu \right) \right] \end{aligned} \quad (30)$$

and

$$\alpha_N(\mathbf{x}, t) = \frac{1}{N} \ln [\psi_N(\mathbf{x}, t)]. \quad (31)$$

We can solve now the saddle-point equation

$$\alpha(\mathbf{x}, t) = \lim_{N \rightarrow \infty} \alpha_N(\mathbf{x}, t) = \text{Extr}\{\Phi\}, \quad (32)$$

where we neglected  $\mathcal{O}(N^{-1})$  terms, as we performed the thermodynamic limit. Finally, by replacing  $t = \beta$  and  $\mathbf{x} = 0$  and  $x'_v = \beta \langle m_v \rangle$  (hence the original statistical mechanics framework), we obtain the following expressions for the statistical pressure

$$\begin{aligned} \alpha(\beta, d) &= \frac{\beta}{2} \sum_\mu \langle m_\mu \rangle^2 - \ln(2) \\ &\quad - \left\langle \ln \left[ \cosh \left( \beta \sum_{\mu=1}^P \langle m_\mu \rangle \xi^\mu \right) \right] \right\rangle_\xi, \end{aligned} \quad (33)$$

whose extremization offers immediately the  $P$  desired self-consistency equations for all the  $\langle m_v \rangle$ ,

$$\langle m_v \rangle = \left\langle \xi^v \tanh \left( \beta \sum_{\mu=1}^P \xi^\mu \langle m_\mu \rangle \right) \right\rangle_\xi \quad \forall \mu \in [1, P], \quad (34)$$

where with the index  $\xi$  we emphasized once more that the disorder average over the quenched patterns is performed as well.

Of course, the self-consistency equations (34) recover those obtained in Agliari, Barra, Galluzzi et al. (2012) via different analytical techniques, where they were also shown to yield the parallel ansatz (10), which, in turn, can be formally written as

$$\sigma_i = \xi_i^1 + \sum_{v=2}^P \xi_i^v \prod_{\mu=1}^{v-1} \delta(\xi_i^\mu), \quad (35)$$

and it will be referred to as  $\sigma^{(P)}$ .

The parallel ansatz (10) can be understood rather intuitively. To fix ideas let us assume zero noise level and that one pattern, say  $\mu = 1$ , is perfectly retrieved. This means that the related average magnetization is  $m_1 = (1 - d)$ , while a fraction  $d$  of spins is still available and they can arrange to retrieve a further pattern, say  $\mu = 2$ . Again, not all of them can match non-null entries in pattern  $\xi^2$  and the related average magnetization is  $m_2 = d(1 - d)$ . Proceeding in the same way, for all spins, we get the parallel state. Notice that, the number  $K$  of patterns which are, at least partially, retrieved does not necessarily equal  $P$ . In fact, due to discreteness, it must be  $d^{K-1}(1 - d) \leq 1/N$ , namely at least one spin must be aligned with  $\xi^K$ , and this implies  $K \lesssim \log N$ .

Such a hierarchical, parallel, fashion for alignment, providing an overall energy (see Eq. (7))

$$\begin{aligned} E^{(P)} &= -N \sum_{k=1}^P [(1 - d)d^{k-1}]^2 + P \\ &= -N \frac{(1 - d^{2P})(1 - d)}{1 + d} + P, \end{aligned} \quad (36)$$

is more optimal than a *uniform* alignment of spins amongst the available patterns, as this case would yield  $m_k = (1 - d)/P$  for any  $k$  and an overall energy

$$E^{(U)} = -N \sum_{k=1}^P \left( \frac{1-d}{P} \right)^2 + P = -\frac{(1-d)^2 N}{P} + P, \quad (37)$$

being  $(1 - d^{2+2P}) > (1 - d^2)/P$ .

On the other hand, as we will see in Section 3.1, when  $d > d_c \approx 1/2$ , the state (10) is no longer stable and spurious states do emerge.

Before proceeding, it is worth stressing that, although the parallel state (10) displays non-zero overlap with several patterns, it is deeply different from, and must not be confused with, a spurious state in standard Hopfield networks. In fact, in the former case, at least one pattern is completely retrieved, while in spurious states, the overlap with each memory pattern involved is only partial.

Moreover, in standard Hopfield networks, spurious states are somewhat undesirable because they provide corrupted information with respect to the best retrieval achievable where one, and only one, pattern is exactly retrieved. Conversely, in our model, the retrieval of more than one pattern is unavoidable (for finite  $d$  and  $\beta \rightarrow \infty$ ) and the quality of retrieval may be excellent (perfect) in the case of patterns poorly (not) overlapping.

Finally, and most importantly, for  $\beta \rightarrow \infty$  and in a wide region of dilution, the parallel state  $\sigma^{(P)}$  corresponds to a global minimum for the energy. This is not the case for an arbitrary mixture of states.

### 3. The emergence of spurious states

In Section 2.1, we explained why we expect the parallel state (35) to occur, exploiting the fact that each pattern tends to align as many spins among those still available. Actually, this intuitive approach yields the correct picture for  $T = 0$  (no fast noise) and not-too-large  $d$ , while when either  $T$  or the degree of dilution are large enough, the system can relax to a state where only one pattern is retrieved or fall into a spurious state where several patterns are partially retrieved, but none exactly. For instance, when patterns are sparse, none of them can generate an attraction basin strong enough to align all available spins, in such a way that stationary, mixture states can emerge.

These states are discussed in the following subsections and in Section 4 the analysis will be made quantitative.

#### 3.1. The failure of parallel retrieval

Let us start from the noiseless case and consider the state (35) corresponding to the parallel ansatz (10): we notice that, on average, there exists a fraction  $2[(1 - d)/2]^P$  of spins  $\sigma_i$  corresponding to the entries  $\xi_i^1 = 1, \xi_i^k = -1, \forall k \in [1, P]$  (and analogously for the “gauged” case  $\xi_i^1 = -1, \xi_i^k = +1$ ) and expected to be aligned with the first entry  $\xi_i^1$ , in such a way that the overall field insisting on each of them is  $h_i = m_1 - m_2 - m_3 - \dots - m_P$ . Of course, such spins are the most unstable, and, at zero noise level, they flip whenever  $h_i$  happens to be negative, that is, when  $m_1 < \sum_{k=2}^P m_k$ . Exploiting the ansatz  $m_k = d^{k-1}(1 - d)$ , this can be written as

$$h_i = (1 - d) \left[ 1 - \frac{d - d^P}{1 - d} \right] = 1 - 2d + d^P, \quad (38)$$

which becomes negative for a value of dilution  $d_c(P)$ , which converges exponentially from above to  $1/2$  as  $P$  gets large. From this point onwards, the first pattern is no longer completely retrieved and the system fails to parallel retrieve (according to the definition

in Eq. (35)). Therefore, when  $d \geq d_c(P)$ , genuine spurious states emerge and the system relaxes to states which correspond to mixtures of  $p \leq P$  patterns, but none of them is completely retrieved (at least up to extreme values of dilution). As we will see in Section 4.4, the transition at  $d_c(P)$  is first order.

Moreover, from Eq. (38) we find that the case  $P = 2$  has no solution in the range  $d \in [0, 1]$ , meaning that the parallel-retrieval state is always a stable solution in the zero noise limit; on the other hand,  $d_c(3) \approx 0.62$ ,  $d_c(4) \approx 0.54$  and so on.

Such phenomenology concerns relatively large degrees of dilution, yet, the presence of noise can also destabilize the true parallel-retrieval state (10) in the regime of small degrees of dilution. In fact, we expect that the spins aligned according to the  $k$ th pattern associated with a magnetization  $m_k = d^{k-1}(1 - d)$  will lose stability at noise levels  $T > d^{k-1}(1 - d)$ . In particular, at  $T > d(1 - d)$ , only one pattern will be retrieved and the pure state is somehow recovered. As we will see in Section 4.4, such estimates are correct for small  $d$ .

### 3.2. Symmetric mixtures

Typical spurious states emerging in standard associative networks are the so-called symmetric mixtures of  $p \leq P$  states, which can be described as

$$\sigma_i = \text{sign} \left( \sum_{\mu=1}^p \xi_i^\mu \right), \quad (39)$$

and it will be referred to as  $\sigma^{(S)}$ . We anticipate that the symmetric mixture turns out to emerge also in the diluted model under investigation.

Now, in the standard Hopfield model, odd mixtures of  $p$  patterns, are metastable, i.e. their energies are higher than those of the pure patterns, and, moreover, the smaller  $p$ , the more energetically favorable the mixture. On the other hand, even mixtures of  $p$  patterns are unstable (they are saddle-points of the energy). The instability of even mixtures is often associated with the fact that, for a macroscopic fraction of spins,  $\sigma^{(S)}$  is not defined due to the ambiguity of the sign. For instance, when  $p = 2$ ,  $\sum_{\mu=1}^p \xi_i^\mu$  turns out to be null for half of the spins and the related values are defined stochastically according to the distribution

$$P(\sigma_i) = \frac{1}{2} (\delta_{\sigma_i+1} + \delta_{\sigma_i-1}). \quad (40)$$

However, as we will show in Section 4.3, this is not the case for this diluted model as it displays wide regions in the parameter space ( $d, T$ ) where even and/or odd symmetric mixtures are stable.

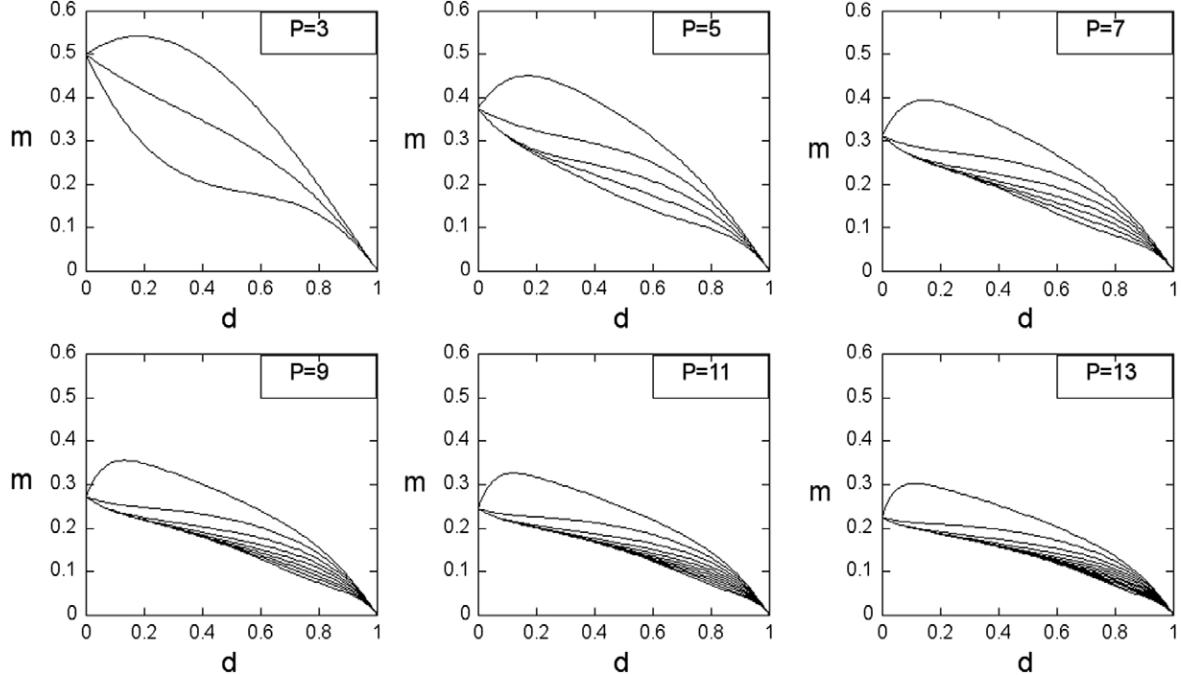
#### 3.3. A “hybrid” spurious state

As we will see in Section 4.3, the symmetric mixture  $\sigma^{(S)}$  can become unstable and relax to a different spurious state which is a “hybrid” state between the symmetric mixture  $\sigma^{(S)}$  and the parallel state  $\sigma^{(P)}$ .

To begin and fix ideas, let us set  $P = 3$  and start from the state  $\sigma_i = \text{sign}(\xi_i^1 + \xi_i^2 + \xi_i^3)$ . In the presence of dilution the argument  $\xi_i^1 + \xi_i^2 + \xi_i^3$  can be zero and in that situation one can adopt the following hierarchical rule: take  $\sigma_i = \xi_i^1$  provided that  $\xi_i^1 \neq 0$ ; otherwise, if  $\xi_i^1 = 0$ , then take  $\sigma_i = \xi_i^2$  provided that  $\xi_i^2 \neq 0$ ; otherwise, if also  $\xi_i^2 = 0$ , then take  $\sigma_i = \xi_i^3$  provided that  $\xi_i^3 \neq 0$ ; otherwise, if also  $\xi_i^3 \neq 0$ , then put  $\sigma_i = \pm 1$  with probability  $1/2$ . In this way we can build a state, generally defined for any  $P$ , and, being  $\Xi = \sum_i \xi_i^\mu$ , it can be written as

$$\begin{aligned} \sigma_i = & (1 - \delta_{\Xi,0}) \text{sign}(\Xi) + \delta_{\Xi,0} [\xi_i^1 + \delta_{\xi_i^1,0} \xi_i^2 \\ & + \delta_{\xi_i^1,0} \delta_{\xi_i^2,0} \xi_i^3 + \dots], \end{aligned} \quad (41)$$

which will be referred to as  $\sigma^{(H)}$ .



**Fig. 1.** Mattis magnetizations  $m$  versus dilution  $d$ , according to the analytical expression derived in Section 3.3. Each panel refers to a different value of  $P$ , as specified.

The related average Mattis magnetizations can be calculated as the sum of one contribution  $m_0$  (the same for any  $\mu$ ) deriving from the spins corresponding to non-ambiguous sign function (i.e.,  $\Xi \neq 0$ ), and another contribution accounting for hierarchical corrections (i.e.,  $\Xi = 0$ ). Let us focus on the first term:

$$\begin{aligned} m_0 &= \langle \xi^\mu \text{sign}(\Xi) \rangle_\xi \\ &= \frac{1-d}{2} \left\langle \text{sign} \left( 1 + \sum_{v \neq \mu}^P \xi^\mu \right) - \text{sign} \left( -1 + \sum_{v \neq \mu}^P \xi^\mu \right) \right\rangle_\xi \quad (42) \\ &= (1-d) \left[ \mathcal{P} \left( \sum_{v \neq \mu}^P \xi^\nu < 1 \right) - \mathcal{P} \left( \sum_{v \neq \mu}^P \xi^\nu > 1 \right) \right], \quad (43) \end{aligned}$$

where, in the last step, we exploited the implicit symmetry in pattern entries and  $\mathcal{P}(\sum_{v \neq \mu}^P \xi^\nu \geqslant 1)$  represents the probability that the specified inequality is verified over the distribution (9). The latter quantity can also be looked at as the probability for a symmetric random walk with holding probability  $d$  to be at distance  $\geqslant 1$  from its origin after a time span  $P-1$ . Hence, we get

$$m_0 = (1-d)[\mathcal{P}(0 \rightarrow 0, P-1) + \mathcal{P}(0 \rightarrow 1, P-1)], \quad (44)$$

where  $\mathcal{P}(x_0 \rightarrow x, t)$  is the probability for a symmetric random walk with stopping probability  $d$  to move from site  $x_0$  to site  $x$  in  $t$  steps, namely

$$\begin{aligned} \mathcal{P}(x_0 \rightarrow x, t) &= \sum_{s=0}^{t-(x-x_0)} \frac{t!}{s! \left( \frac{t-s-(x-x_0)}{2} \right)! \left( \frac{t-s+(x-x_0)}{2} \right)!} \\ &\times d^s \left( \frac{1-d}{2} \right)^{t-s}. \quad (45) \end{aligned}$$

The second contribution to the magnetization is  $(1-d) \sum_{k=1}^{P-1} \mathcal{P}(0 \rightarrow 1, P-k) d^{k-1}$ .

Finally, by summing the two contributions we find the following expressions for  $P = 3$

$$m_1 = \frac{1}{2}(1+d-3d^2+d^3), \quad (46)$$

$$m_2 = \frac{1}{2}(1-d)(1+d^2), \quad (47)$$

$$m_3 = \frac{1}{2}(1-3d+5d^2-3d^3), \quad (48)$$

and for  $P = 5$

$$m_1 = \frac{1}{8}(3+9d-42d^2+74d^3-65d^4+21d^5), \quad (49)$$

$$m_2 = \frac{1}{8}(1-d)(3+6d^2-d^4), \quad (50)$$

$$m_3 = \frac{1}{8}(1-d)(3-4d+18d^2-20d^3+11d^4), \quad (51)$$

$$m_4 = \frac{1}{8}(1-d)(3-4d+18d^2-28d^3+19d^4), \quad (52)$$

$$m_5 = \frac{1}{8}(1-d)(3-4d+18d^2-36d^3+27d^4). \quad (53)$$

The expressions for arbitrary  $P$  can be analogously calculated exactly and some examples are shown in Fig. 1.

We expect  $\sigma^H$  to become globally stable in the region of very large dilutions ( $d > d_H(P)$ ); intuitively, the dilution must be large enough to make magnetizations rather close to each other in such a way that the least signaled spins corresponding to  $(-, -, \dots, -, +, +, \dots, +)$  (overall  $(P-1)/2$  negative entries and  $(P+1)/2$  positive entries) are stable. This means  $\sum_i (1 - \delta_{\Xi,0}) \text{sign}(\Xi) \xi_i^\mu / N > \sum_{k=1}^{(P-1)/2} \varphi_k (P+1)/(P-k)$ , where  $\varphi_k = 2 \sum_l [(1-d)/2]^{2l} d^{P-2l} (P-k)! / [!(l-1)!(P-k-2l+1)!]$  and  $P$  is odd. This condition is fulfilled for values of dilution larger than  $d_H(P)$ , which converges to 1 as  $P$  gets larger, hence, in order to

tackle this limit, dilution must become a function of the system size  $d \rightarrow d(N)$ . In this case the network itself becomes diluted as well and different techniques are required; this will not be discussed in this paper.

#### 4. Stability analysis on the organization of the states

The set of solutions for self-consistent equations (34) describes states whose stability may vary strongly. In fact, provided the network has reached them, in the noiseless limit (of whatever kind) it would persist in those states. However, the equations do not contain any information about whether the solutions will be stable against small perturbations, that is to say if the system will indeed really thermalize on these states or will fall apart more or less quickly. In order to evaluate their stability we need to check the second derivative of the free energy (Amit, 1989). More precisely, we further need to build up the so called “stability matrix”  $\mathbf{A}$  with elements

$$A^{\mu\nu} = \frac{\partial^2 f_\beta(\vec{m})}{\partial m^\mu \partial m^\nu}. \quad (54)$$

Then, we evaluate and diagonalize  $\mathbf{A}$  at a point  $\vec{m}$ , representing a particular solution of the self-consistency equation (34), in order to determine whether  $\vec{m}$  is stable or not. With  $\{E_\mu\}_{\mu=1,\dots,P}$ , the set of related eigenvalues,  $\vec{m}$  is stable whenever all of them are positive.

Now, from Eqs. (33) and (54), remembering that  $\alpha(\beta, d) = -\beta f(\beta, d)$ , we find straightforwardly

$$A^{\mu\nu} = [1 - \beta(1 - d)]\delta^{\mu\nu} + \beta Q^{\mu\nu}, \quad (55)$$

where

$$Q^{\mu\nu} = \langle \xi^\mu \xi^\nu \tanh^2(\beta \vec{\xi} \cdot \vec{m}) \rangle_\xi. \quad (56)$$

Of course when  $d = 0$  we recover  $A^{\mu\nu} = (1 - \beta)\delta^{\mu\nu} + \langle \xi^\mu \xi^\nu \tanh^2(\beta \vec{\xi} \cdot \vec{m}) \rangle_\xi$ , namely the result known for the standard Hopfield model.

We now consider several states, known to be solutions of self-consistency equation (34) and check their stability. In this way we will find constraints in the region  $(T, d)$  where those states are stable and then we will build up the phase diagram.

##### 4.1. Paramagnetic state

Let us start with the paramagnetic state, which is described by

$$\vec{m} = \vec{0}; \quad (57)$$

this state trivially fulfills Eq. (34).

By replacing this expression in Eqs. (55) and (56) we find

$$A^{\mu\nu} = \delta_{\mu\nu}[1 - \beta(1 - d)]. \quad (58)$$

Therefore, in this case,  $\mathbf{A}$  is diagonal and its eigenvalues are directly  $E_\mu = A^{\mu\mu} = 1 - \beta(1 - d)$ ,  $\forall \nu \in [1, P]$ . We can conclude the paramagnetic state exists and is stable in the region  $1 - \beta(1 - d) > 0$ , that is (remembering that  $T = \beta^{-1}$ )

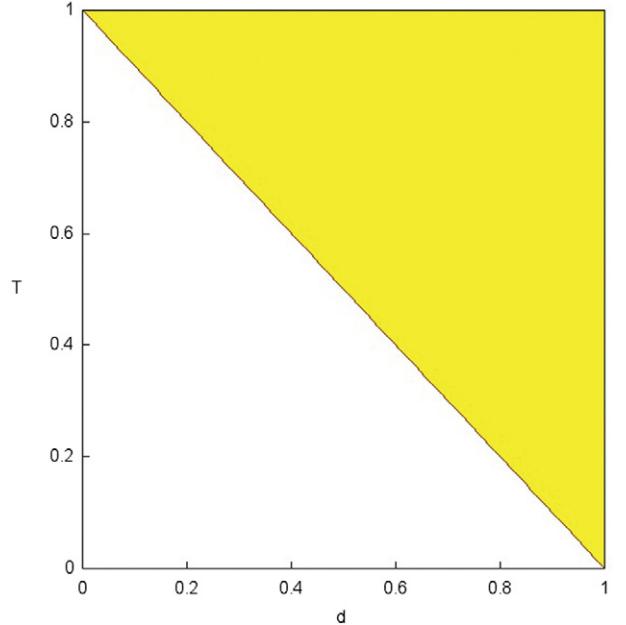
$$\text{PM stability} \Rightarrow T > 1 - d. \quad (59)$$

This region is highlighted in Fig. 2.

##### 4.2. Pure state

Let us now consider the pure state, that is any of the  $P$  configurations

$$\vec{m} = m(1, \vec{0}), \quad (60)$$



**Fig. 2.** (Color online) In the parameter space  $(T, d)$  we highlighted the region where the paramagnetic state exists and is stable. As proved in Section 4.1, this region includes points fulfilling  $T < 1 - d$ ; notice that this result is independent of  $P$ .

$m$  being the extent of the overlap, which, in general, depends on  $d$  and on  $T$ . The related self-consistency equations are

$$m^\mu = (1 - d) \tanh(\beta m^\mu), \quad (61)$$

$$m^{\nu \neq \mu} = 0. \quad (62)$$

The first equation has a solution in the whole half-plane  $T > 1 - d$ , and this ensures that, in the same region, the pure state exists. In order to check its stability, we calculate the stability matrix finding

$$A^{\mu\nu} = 0 \vee \mu \neq \nu \quad (63)$$

$$A^{\mu\mu} = 1 - \beta(1 - d)[1 - \tanh^2(\beta m^\mu)] \quad (64)$$

$$A^{\nu\nu} = 1 - \beta(1 - d)[1 - (1 - d) \tanh^2(\beta m^\nu)]. \quad (65)$$

Therefore  $\mathbf{A}$  is diagonal and the eigenvalues are  $E_\mu = A^{\mu\mu}$  and  $E_\nu = A^{\nu\nu}$ . Notice that these eigenvalues do not depend on  $P$  and that  $E_\mu \geq E_\nu$ , so that the analysis can be restricted on  $E_\nu$ . Requiring the positivity for  $E_\nu$ , we get the region in the plane  $(T, d)$ , where the pure state is stable; such a region is shown in Fig. 3. We stress that this result is universal with respect to  $P$  (in the low-storage regime).

##### 4.3. Symmetric state

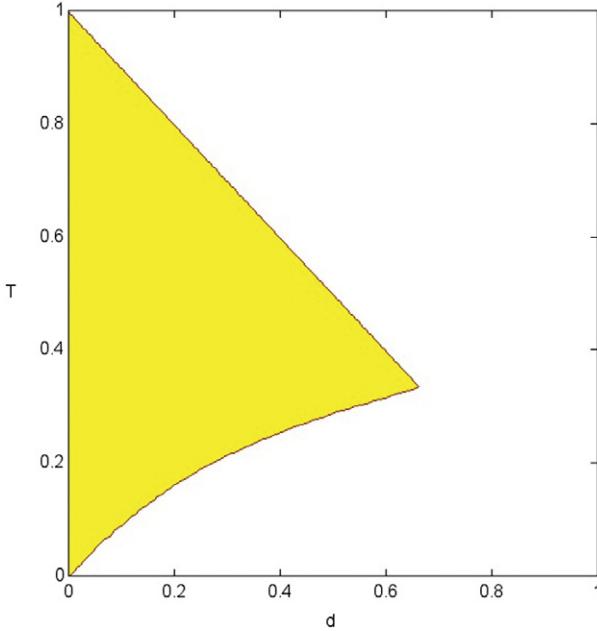
A symmetric mixture of states corresponds to configurations leading to

$$\vec{m} = m(d, T)(1, 1, 1, \dots, 1, 0, \dots, 0), \quad (66)$$

where  $p \leq P$  order parameters are equivalent and non-null, while the remaining  $P - p$  are vanishing.

Let us start with the case  $p = P = 3$ , yielding  $\vec{m} = m(d, T)(1, 1, 1)$ . In this special case the three self-consistency equations collapse on

$$\begin{aligned} m(d, T) = 2 \left( \frac{1-d}{2} \right)^3 & [\tanh^2(3\beta m) + \tanh^2(\beta m)] \\ & + d \left( \frac{1-d}{2} \right)^2 \tanh^2(2\beta m) \\ & + 2 \left( \frac{1-d}{2} \right) d^2 \tanh^2(\beta m) \end{aligned} \quad (67)$$



**Fig. 3.** (Color online) In the parameter space  $(T, d)$  we highlighted the region where the pure state exists and is stable. This result was found by numerically solving the self-consistency equation (34) and the inequality  $E_v > 0$ , where  $E_v$  is the smallest eigenvalues of the stability matrix  $\mathbf{A}$  (see Eq. (65)); notice that this result is independent of  $P$ .

and the matrix  $\mathbf{A}$  reads as

$$\begin{pmatrix} a & b & b \\ b & a & b \\ b & b & a \end{pmatrix} \quad (68)$$

$a$  and  $b$  being parameters related to  $m$ ,  $d$  and  $\beta$ . More precisely, the eigenvalues of  $\mathbf{A}$  are  $(a+2b, a-b, a-b)$ , which can be written as

$$\begin{aligned} a-b &= 1-\beta(1-d)+2\beta\left\{\tanh^2(2\beta m)d\left(\frac{1-d}{2}\right)^2 + \tanh^2(\beta m)\left[\frac{d^2(1-d)}{2}+4\left(\frac{1-d}{2}\right)^3\right]\right\}, \end{aligned} \quad (68)$$

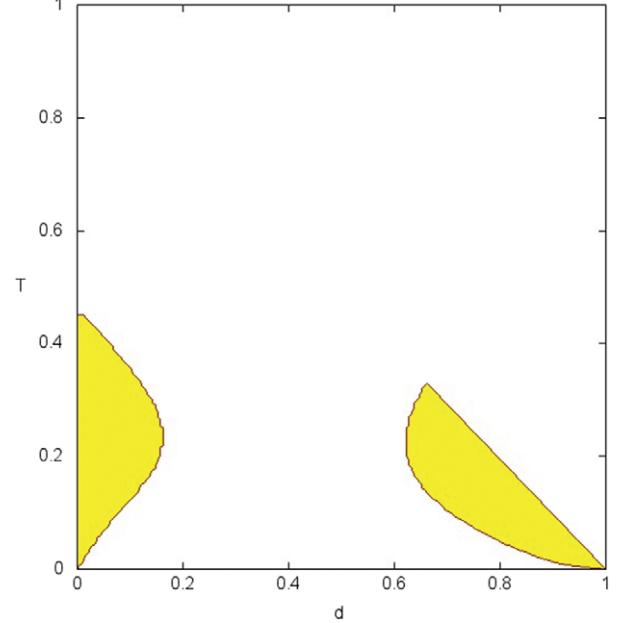
$$\begin{aligned} a+2b &= 1-\beta(1-d)+2\beta\left\{\tanh^2(3\beta m)3\left(\frac{1-d}{2}\right)^3 + \tanh^2(\beta m)\left[\frac{d^2(1-d)}{2}+\left(\frac{1-d}{2}\right)^3\right] + 8d\beta\tanh^2(2\beta m)\left(\frac{1-d}{2}\right)^2\right\}. \end{aligned} \quad (69)$$

The conditions for the existence and the stability of the symmetric, odd mixture with  $p = P = 3$ , yield a system of equations which was solved numerically and the region where such conditions are all fulfilled is shown in Fig. 4. Notice that the region is actually made up of two disconnected parts, each displaying peculiar features, as explained later.

This result is robust with respect to  $P$ , with  $P$  odd and  $p = P$ .

We can further generalize the analysis by considering  $P > p$ ,  $p$  still being odd. In this case we get the following stability matrix

$$\begin{pmatrix} a & b & b & 0 \\ b & a & b & 0 \\ b & b & a & 0 \\ 0 & 0 & 0 & c \end{pmatrix} \quad (70)$$



**Fig. 4.** (Color online) In the parameter space  $(T, d)$  we highlighted the region where the symmetric state  $\sigma^{(S)}$ , for the special case  $p = P = 3$ , exists and is stable. Notice that two disconnected regions emerge: the one corresponding to lower values of dilution derives from the fact that  $p$  is odd, while the one corresponding to larger values of dilution from the fact that  $p = P$ .

with eigenvalues  $(a-b, a-b, a+2b, c)$ , where

$$\begin{aligned} c &= 1-\beta(1-d)\left\{1-2\left[\left(\frac{1-d}{2}\right)^3[\tanh^2(3m)+3\tanh^2(m)] + d\left(\frac{1-d}{2}\right)^23\tanh^2(2m)+3\frac{1-d}{2}d^2\tanh^2(m)\right] \times \left[1-2\left(\frac{1-d}{2}\right)^3[\tanh^2(3\beta m)+3\tanh^2(\beta m)] + 3d\left(\frac{1-d}{2}\right)^2\tanh^2(2\beta m)+3\frac{1-d}{2}d^2\tanh^2(\beta m)\right]\right\} \end{aligned} \quad (71)$$

has degeneracy  $P - p$ .

Such states ( $p < P$ ,  $p$  odd) are stable only at small  $d$ . This is due to the fact that the eigenvalue  $c$  occurs only when  $p < P$  and it reads as ( $\mu > p$ ):

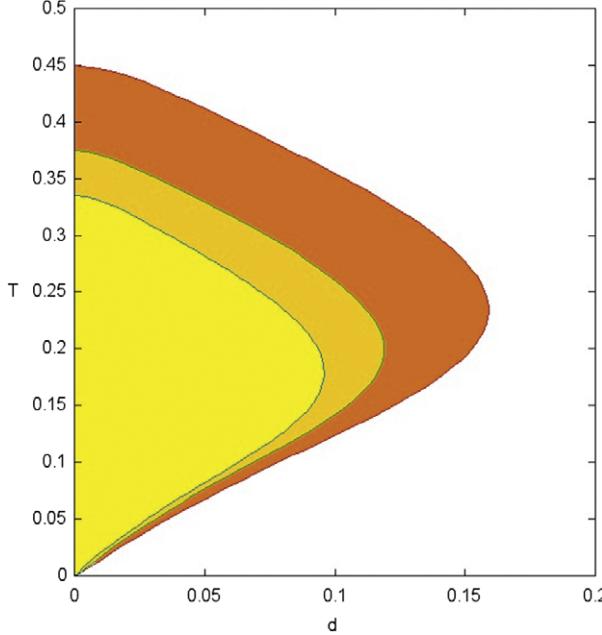
$$\begin{aligned} A^{\mu\mu} &= [1-\beta(1-d)]+\beta\langle(\xi^\mu)^2\rangle_\xi\left\langle\tanh^2\left[\beta m\sum_v^p\xi^v\right]\right\rangle_\xi \\ &= [1-\beta(1-d)]+\beta(1-d)\left\langle\tanh^2\left[\beta m\sum_v^p\xi^v\right]\right\rangle_\xi. \end{aligned} \quad (72)$$

Thus, one can see that the r.h.s. term contains factors  $(1-d)$  at least of second order in such a way that when  $d$  is close to 1, i.e. for high dilution, and  $T < 1-d$ , such a term becomes negative. On the other hand, in the case  $\mu \leq p$ , we get

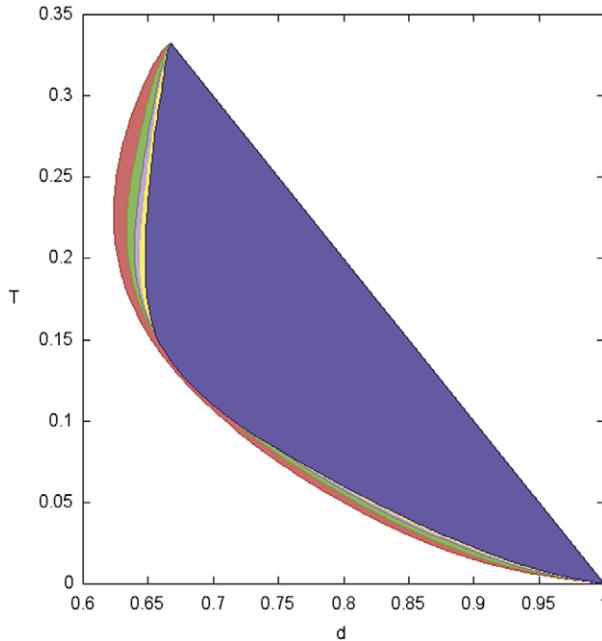
$$A^{\mu\mu} = [1-\beta(1-d)]+\beta\left\langle(\xi^\mu)^2\tanh^2\left[\beta m\sum_{v=1}^p\xi^v\right]\right\rangle_\xi$$

and therefore the r.h.s. term contains even first order terms  $(1-d)$ , which are comparable with  $\beta(1-d)$ .

Moreover, we find that the  $p$ -component, odd symmetric state exists and is stable in a region of the space  $(T, d)$ , which gets smaller and smaller as  $p$  grows (see Fig. 5). The emergence of such



**Fig. 5.** (Color online) In this plot we focused on the region of the parameter space  $(T, d)$ , where odd symmetric spurious states exist and are stable. In particular, we chose  $P = 7$  and we considered any possible odd mixture, i.e.  $p = 3, p = 5$  and  $p = 7$ ; each value of  $p$  is represented by a different curve. Notice that the smaller  $p$  the wider the region, analogously to the standard Hopfield model.



**Fig. 6.** (Color online) In this plot we focused on the region of the parameter space  $(T, d)$ , where symmetric spurious state with  $p = P$  exist and are stable. In particular, we chose  $P = 7$  and we considered any possible mixture, i.e.  $p = 3, p = 4, p = 5, p = 6$  and  $p = 7$ ; each value of  $p$  is represented by a different curve. Notice that the smaller  $p$  the wider the region, yet the region tends to an “asymptotic shape”.

states can be seen as a feature of robustness of the standard Hopfield model with respect to dilution.

Finally, the case  $P = p$  always admits a region of existence and stability in the regime of high dilution. The latter region is independent of the parity and depends slightly on  $P$  (see Fig. 5). The emergence of such states is due to the failure of hierarchical retrieval, namely uniformity prevails (see Fig. 6).

#### 4.4. Parallel state

The parallel-retrieval state can be looked at as the extension to arbitrary values of  $d$  of the pure state holding for the special case  $d = 0$ . We recall that in the noiseless limit the parallel-retrieval state can be described as

$$\vec{m} = (1 - d, (1 - d)d, (1 - d)d^2, \dots, (1 - d)d^P). \quad (73)$$

In this case the stability matrix is diagonal with terms:

$$A^{\mu\mu} = 1 - \beta(1 - d) + \beta \langle (\xi^\mu)^2 \tanh^2[\beta(1 - d) \times (\xi^1 + d\xi^2 + \dots + d^P \xi^P)] \rangle, \quad (74)$$

and, consistently, taking the limit  $\beta \rightarrow \infty$ , we get the simplified form

$$A^{\mu\mu} = \lim_{\beta \rightarrow \infty} = 1 - \beta(1 - d) + \beta \langle (\xi^\mu)^2 \times (1 - \delta[(\xi^1 + d\xi^2 + \dots + d^P \xi^P)]) \rangle. \quad (75)$$

Now, the third term in the r.h.s. is either  $\beta \langle (\xi^\mu)^2 \rangle = \beta(1 - d)$  (when the polynomial of order  $P$  is zero) or 0; the latter case would trivially yield  $A^{\mu\mu} < 0$ . Therefore, in the limit  $\beta \rightarrow \infty$  the stability of the parallel-retrieval state is constrained by the smallest real root  $\in [0, 1]$  of the polynomial  $\xi^1 + d\xi^2 + \dots + d^P \xi^P$  with  $\xi^i = 1, 0, -1$ . This corresponds to  $\xi^1 = 1$  and  $\xi^i = -1, \forall i > 1$ , under gauge symmetry and returns the same result found, from a more empirical point of view, in Section 3.1. More precisely, the critical dilution converges exponentially to  $1/2$  as  $P$  grows.

In particular, for  $P = 3$  we find that the parallel-retrieval state exists and is stable in the interval  $d \in (0, \frac{\sqrt{5}-1}{2}) \simeq (0, 0.618)$ . The point  $d_c(3) = \frac{\sqrt{5}-1}{2}$  corresponds to the unique real root in  $(0, 1)$ .

When noise is introduced, the critical dilution  $d_c$ , separating the parallel-retrieval state from spurious states, is shifted towards larger values, as suggested by Eq. (74). On the opposite side, namely in the regime of small dilution, the parallel state is progressively depleted and, as the temperature is increased, magnetizations vanish, starting from  $m_P$ , and proceeding up to  $m_2$ . One can distinguish a set of temperatures  $T_p(d) < T_{p-1}(d) < \dots < T_2(d) < T_1(d)$ , such that when  $T > T_k(d)$ , all magnetizations  $m_i, \forall i \leq k$  are null on average. Hence, above  $T_2(d)$  the pure state retrieval is recovered, while above  $T_1(d) = 1 - d$  the paramagnetic state emerges.

In Fig. 7 we highlight the region of the parameter space  $(T, d)$  where such parallel states exist and are stable. This was obtained numerically for the case  $P = 5$ ; for larger values of  $P$  the region is slightly restricted to account for the shift in  $d_c$ .

#### 5. Monte Carlo simulations

In this section we discuss details of Monte Carlo simulations.

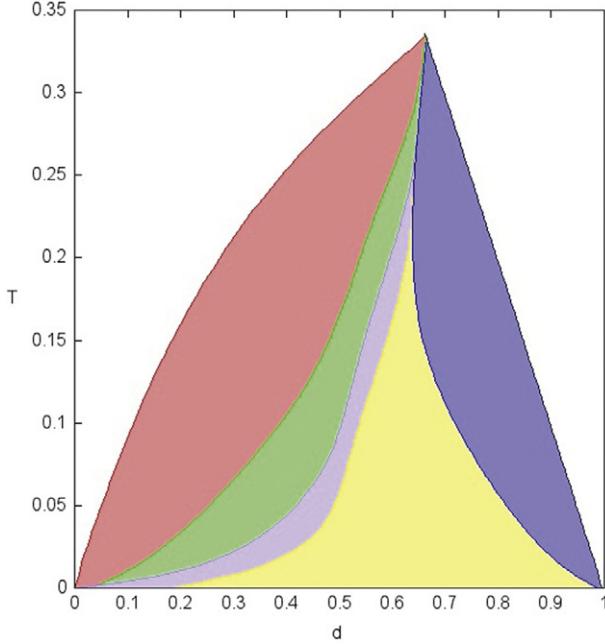
All the simulations were performed on an Ubuntu Linux system with Intel Core I7, 3.2 GHz, 12 CPU, Nvidia-Fermi technology, 12 Gb RAM and OpenMP libraries. The simulations were carried out sequentially according to the following algorithm:

##### 1. Building and storing of the coupling matrix.

First, we generate  $P$  patterns according to the distribution ( $d = 0$ ):

$$P(\xi_i^\mu) = \frac{1}{2} \delta_{(\xi_i^\mu - 1)} + \frac{1}{2} \delta_{(\xi_i^\mu + 1)}, \quad (76)$$

then, we build a char-matrix  $J_{ij} = \sum_\mu \xi_i^\mu \xi_j^\mu$  with entries ranging  $\in [0, 2P + 1]$  and acting as key pointing to another hash-matrix  $\tilde{J}_{ij}$  where the  $N(N - 1)/2$  real numbers accounting for the Hebb interactions (see Eq. (16)) are stored. If the amount of



**Fig. 7.** (Color online) In this plot we focused on the region of the parameter space  $(T, d)$ , where parallel retrieval states exist and are stable. In particular, we chose  $P = 5$  and we considered any possible state with  $k = 2, k = 3, k = 4$  and  $k = 5$  non-null magnetization.

patterns does not exceed  $P = 256$ , i.e. one byte, it is then possible to account for  $10^5$  neurons with no need of swapping on the hard disk (which would noticeably affect the performance of the simulation). This condition is fulfilled for the low storage regime we are interested in.

#### 2. Initialize the network status.

We checked the two standard approaches: The first is to initialize the network in a (assumed) fixed point of the dynamics, namely

$$\sigma_i = \xi_i^1 \quad \forall i \in [1, \dots, N], \quad (77)$$

and check its evolution: This gives information on the structure of the basins of attraction of the minima as we vary the dilution (see Point 5).

The second approach is to initialize the network randomly: We set  $\sigma_i = 1$  with probability 0.5 and  $\sigma_i = -1$  otherwise. This is a standard procedure to follow the relaxation to a fixed point with no initial assumption and gives information on the structure of the basins of attraction of the minima at fixed dilution.

#### 3. Evolution dynamics.

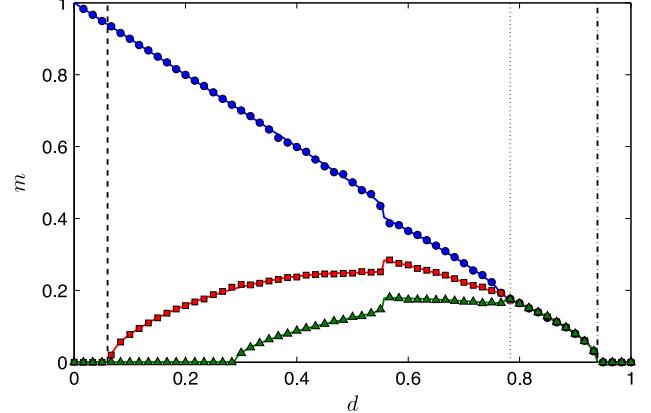
The spin status evolves according to a standard (random and sequential) Glauber dynamics for Ising-like systems (Amit, 1989): At each time interval, the neuronal state is updated according to its input signals, where the probability of the unit's activity is equal to a rectified value of the input (logit transfer function), i.e.

$$\Pr[\sigma_i(t) = \pm 1] = \frac{1}{1 + \exp\left[\mp 2\beta \sum_j J_{ij} \sigma_j\right]}. \quad (78)$$

The field-updating process is managed by a linked list whose parsing is parallelized through OpenMP.

#### 4. Convergence of the simulation.

Due to the peculiar structure of the fields induced by pattern dilution, the field insisting on a given neuron may be zero and the related spin would flip indefinitely. To avoid this pathological situation we skip the updating of these “paramagnetic” spins



**Fig. 8.** (Color online) Data from Monte Carlo simulations (symbols) and analytical predictions (solid lines) obtained for a system with  $P = 3$  patterns and set at a temperature  $T = 0.06$  are compared. Simulations are performed on a set of  $10^5$  spins. The dashed line at  $d \approx 0.06$  marks the boundary of the pure state regime; the dotted line at  $d \approx 0.78$  marks the onset of the symmetric phase; the semi-dashed line at  $d \approx 0.94$  marks the onset of the paramagnetic phase.

and focus on the remaining ones: In the zero noise limit convergence is almost immediate, such that when the whole ensemble of neurons remains unchanged for the whole  $N$ -length of the update cycle, dynamics is stopped and the resulting  $P$  pattern overlaps are printed on a file.

Relaxation at non-zero noise is checked through the linked list (see next step): The pointer of each neuron that is aligned with its own field is stored, the ones of neurons with no net fields are removed from the linked list, while all the other spins, mismatched to their own fields, are added to the linked list.

#### 5. Making the $P$ patterns sparser.

There can be two deeply different ways of increasing dilution. The first is a Bernoullian approach and essentially if one starts from a dilution  $d = 0.45$  towards a dilution  $d = 0.5$  (just as a concrete example) one may forget the starting information and generate a random pattern with on average one half of zero entries; the second is a Markovian dilution by which one needs to start from the previous coupling matrix (and patterns) diluted at  $d = 0.45$  and increases dilution on that structure.

Dilution is tuned at steps of 0.01, ranging from  $d = 0$  to  $d = 1$ . We take as the state of the network the last equilibrium state, then go to point (3).

Through Markovian dilution, we can follow the evolution of the pure Hopfield attractors while tuning  $d$ . In general, the results obtained via numerical simulations are in perfect agreement with the theory: This point is not surprising, as, due to the load storage regime,  $\lim_{N \rightarrow \infty} P/N = 0$ , hence replica symmetry is never broken and our solution is the real solution of the model (no approximations have been made, see Fig. 8).

## 6. Discussion

In this work we explored the retrieval capabilities of the multitasking associative network introduced in Agliari, Barra, Galluzzi et al. (2012). Such a system is characterized by (quenched) patterns which display a fraction  $d$  of null entries: interestingly, by paying the price of reducing the amount of information stored within each pattern (by a fraction  $d$ ), we get a system able to retrieve several patterns at the same time.

At zero noise level ( $T = 0$ ), and for a relatively low degrees of dilution, the system converges to an equilibrium state characterized by overlap  $\mathbf{m} = ((1-d), (1-d)d, \dots, (1-d)d^k, (1-d)d^{P-1})$ , where  $P$  is the number of stored patterns. Although this state displays non-null overlap with several patterns, it does not represent

a spurious state, as can be seen by noticing, for instance, that this state allows the complete retrieval of at least one pattern. However, through a careful inspection, we proved in this paper that there are regions in the  $(T, d)$  plane where genuine spurious states occur, hence a clear picture of the phase diagram becomes a fundamental issue in order to make the model ready for practical implementations.

A remarkable difference with respect to standard (serial processing) neural networks lies in the stability of mixture states: both even and odd mixtures are stable, which – within the world of spurious states – was a somewhat desired, and expected, result as there is neither a biological reason, nor a prescription from robotics, to weight differently odd and even mixtures (whose difference lies in the gauge invariance of the standard Hopfield model, which is broken within our framework due to the partial blankness of the pattern entries). Another expected feature, which we confirmed in this paper, is the emergence of parallel spurious states beyond standard ones from classical neural network theory: This is the natural generalization of the latter when moving from serial to parallel processing.

Beyond these somewhat expected results, the phase diagram of the model is still very rich and composed by several not-overlapping regions where the retrieval states are deeply differently structured: Beyond the paramagnetic state and the pure state, the system is able to achieve both a hierarchical organization of pattern retrievals (for intermediate values of dilution) and a completely symmetric parallel state (for high values of dilution), which act as the basis for the outlined mixtures when raising the noise level above thresholds whose value depends on the load of the network  $P$ .

These findings have been obtained developing a new strategy for computing the free energy of the model by which, imposing thermodynamic principles (hence extremizing the latter over the order parameters of the model), self-consistency has been obtained. The whole procedure has been strongly based on techniques stemming from partial differential equation theory. In particular, the key idea is showing that the noise-derivatives of the statistical pressure obey Burgers' equations, which can be solved through the Cole–Hopf transformation. The latter maps the evolution of the free energy over the noise into a diffusion problem which can be addressed through standard Green integration in momenta space and then pushed back in the original framework.

In the future, effort must still be made in order to achieve a clear scenario in the hyper-diluted regime, namely where the dilution

scales as a function of the volume (the amount of neurons), which cannot be accomplished through the techniques we presented here as saddle point integration is no longer useful. We plan to report on this research soon.

## Acknowledgments

This work is supported by FIRB grant RBFR08EKEV. Sapienza Universita' di Roma and INFN are acknowledged too for partial financial support.

## References

- Agliari, E., Barra, A., De Antoni, A., & Galluzzi, A. (2012). Parallel retrieval of correlated patterns: from Hopfield networks to Boltzmann machines. *Neural Networks*, 38, 52–63.
- Agliari, E., Barra, A., Galluzzi, A., Guerra, F., & Moauro, F. (2012). Multitasking associative networks. *Physical Review Letters*, 109, 268101.
- Agliari, E., Barra, A., Guerra, F., & Moauro, F. (2011). A thermodynamic perspective of immune capabilities. *Journal of Theoretical Biology*, 287, 48–63.
- Amit, D. (1989). *Modeling brain function*. Cambridge University Press.
- Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1985). Storing infinite numbers of patterns in a spin-glass model of neural networks. *Physical Review Letters*, 55, 1530–1533.
- Barra, A., Bernacchia, A., Santucci, E., & Contucci, P. (2012). On the equivalence of hopfield networks and Boltzmann machines. *Neural Networks*.
- Cheng, B., & Titterington, D. M. (1994). Neural networks: a review from a statistical perspective. *Statistical Science*, 9(1), 2–30.
- Coolen, A. C. C., Kühn, R., & Sollich, P. (2005). *Theory of neural information processing systems*. Oxford University Press.
- Genovese, G., & Barra, A. (2009). A mechanical approach to mean field models. *Journal of Mathematical Physics*, 50, 053303.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79, 2554–2558.
- Mézard, M., Parisi, G., & Virasoro, M. A. (1987). *Spin glass theory and beyond*. Singapore: World Scientific.
- Nikolaopoulos, T., Coolen, A. C. C., Pérez-Castillo, I., Skantzos, N. S., Hatchett, J. P. L., & Wemmenhove, B. (2004). Replicated transfer matrix analysis of ising spin models on 'small world' lattices. *Journal of Physics A (Mathematical and General)*, 37, 6455.
- Nolfi, S., & Floreano, D. (2000). *Evolutionary robotics: the biology, intelligence, and technology of self-organizing machines*.
- Sompolinsky, H. (1986). Neural networks with non-linear synapses and a static noise. *Physical Review A*, 34, 2571.
- Trippi, R. R., & Turban, E. (1992). *Neural networks in finance and investing: using artificial intelligence to improve real world performance*. New York, NY, USA: McGraw-Hill, Inc.
- Wemmenhove, B., & Coolen, A. C. C. (2003). Finite connectivity attractor neural networks. *Journal of Physics. A. Mathematical and Theoretical*, 36(9617).
- Willshaw, D. J., & von der Malsburg, C. (1976). How patterned neural connections can be set up by self-organization. *Proceedings of the Royal Society of London. Series B*, 194, 431–445.
- Wu, Fa-Yueh (1982). The Potts model. *Reviews of Modern Physics*, 54, 235.