Editors' Suggestion

Phase diagram of restricted Boltzmann machines and generalized Hopfield networks with arbitrary priors

Adriano Barra,[1,*] Giuseppe Genovese,[2,†] Peter Sollich,[3,‡] and Daniele Tantari[4,§]

[1]*Dipartimento di Matematica e Fisica Ennio De Giorgi, Università del Salento, 73100 Lecce, Italy*
[2]*Institut für Mathematik, Universität Zürich, CH-8057 Zürich, Switzerland*
[3]*Department of Mathematics, King's College London, WC2R 2LS London, United Kingdom*
[4]*Scuola Normale Superiore, Centro Ennio de Giorgi, Piazza dei Cavalieri 3, I-56100 Pisa, Italy*

(Received 1 August 2017; published 20 February 2018)

Restricted Boltzmann machines are described by the Gibbs measure of a bipartite spin glass, which in turn can be seen as a generalized Hopfield network. This equivalence allows us to characterize the state of these systems in terms of their retrieval capabilities, both at low and high load, of pure states. We study the paramagnetic-spin glass and the spin glass-retrieval phase transitions, as the pattern (i.e., weight) distribution and spin (i.e., unit) priors vary smoothly from Gaussian real variables to Boolean discrete variables. Our analysis shows that the presence of a retrieval phase is robust and not peculiar to the standard Hopfield model with Boolean patterns. The retrieval region becomes larger when the pattern entries and retrieval units get more peaked and, conversely, when the hidden units acquire a broader prior and therefore have a stronger response to high fields. Moreover, at low load retrieval always exists below some critical temperature, for every pattern distribution ranging from the Boolean to the Gaussian case.

DOI: 10.1103/PhysRevE.97.022310

I. INTRODUCTION

The genesis of modern Artificial Intelligence (AI) can be traced quite far back in time. Beyond the pioneering and historical contributions around the beginning of the last century, the most celebrated milestones are the neuron model of McCulloch and Pitts [1], the Rosenblatt perceptron [2], and the Hebb learning rule [3]. The latter was, in turn, exploited by Hopfield many years later to write his famous paper on neural networks from the connectionist perspective [4]. There has been a growing stream of studies of neural networks ever since, with the subject attracting the interest of various communities, from biology to signal processing and information theory [5–8]. The physics angle on the topic is mainly represented by the statistical mechanics of spin glasses [9]. In particular, problems of great biological and technological relevance, such as the capability to learn or retrieve memories, find a simple formulation in a genuine statistical mechanics language [4–7,11–13].

However, the models used to implement these two crucial features of neural networks—learning and retrieval—often start from quite different assumptions. For instance, in modern machine-learning approaches such as *deep learning* [8,10], network weights are normally taken as real, enabling the use of gradient descent for learning and inference. On the other side, the standard theory of pattern retrieval, as exemplified by the Amit-Gutfreund-Sompolinsky analysis of associative neural networks [11,12], assumes Boolean patterns. Nevertheless, the two most utilized models for machine learning and retrieval, i.e., restricted Boltzmann machines (RBMs) and associative Hopfield networks are known to be equivalent [14–18]. Their relation is easily understood from the point of view of bipartite spin glasses: on the one hand, the Gibbs measure of such systems is the same as the one of Restricted Boltzmann Machines, while on the other hand, bipartite spin glasses constitute a class of disordered systems in which the Hopfield model for neural networks can be embedded.

For this reason we analyze in this paper spin glasses defined on a bipartite network. We study the retrieval in these networks while varying both spin or unit priors and pattern or weight distributions continuously between the Boolean and the real Gaussian limits. We show that the presence of a ferromagnetic region of retrieval is not peculiar to the standard Hopfield model, but occurs also in the case of continuous units and weights when these take the form of a Gaussian "softening" of Boolean variables. Moreover, while retrieval disappears for Gaussian weights at high load, in the low-load limit our generalized Hopfield networks always have a retrieval phase throughout the entire range of pattern distributions, ranging from the Boolean to the Gaussian cases. This implies a degree of robustness in the machine-learning setup, where weights evolve on real axes and one usually works at low load, i.e., with a small number of features, to avoid overfitting [10,19].

A. Generalized Hopfield models and restricted Boltzmann machines

The Hopfield model introduced in Ref. [4] is a celebrated paradigm for neural networks in which the neurons are represented by N spins, taking values ±1. The energy function of the system is defined in terms of p so-called patterns, denoted by $\boldsymbol{\xi}^{\mu}$, $\mu = 1, \ldots, p$. It is natural to take the patterns

[*] adriano.barra@unisalento.it

[†] giuseppe.genovese@math.uzh.ch

[‡] peter.sollich@kcl.ac.uk

[§] Present address: Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy; daniele.tantari@sns.it

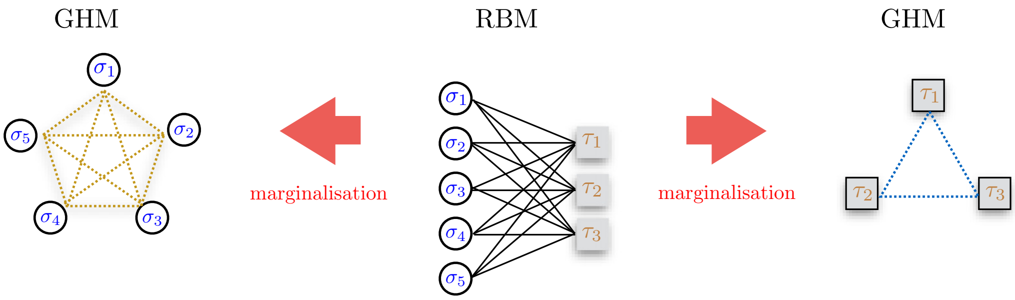

FIG. 1. Three equivalent architectures of neural networks: in a restricted Boltzmann machine (RBM) (consisting of $N_1 = 5$ σ variables and $N_2 = 3$ τ variables in the figure) the role of hidden and visible units can be exchanged and marginalizing over the hidden units one obtains two dual generalized Holpfield models (GHMs), where the visible layer of the RBM constitutes the network and the hidden layer determines the interaction.

to be N-dimensional random vectors with independent and identically distributed components, which makes the Hopfield model a spin glass. Given an instance of the patterns, the Hamiltonian and the Gibbs measure of this system are

$$H_{N,p}(\sigma|\xi) := -\sum_{\mu=1}^{p} N m_\mu^2, \qquad G_{N,p}(\sigma|\xi) := \frac{e^{-\beta H_{N,p}(\sigma|\xi)}}{\mathbb{E}_\sigma e^{-\beta H_{N,p}(\sigma|\xi)}}, \tag{1}$$

where $\beta > 0$ is the inverse temperature, $\beta = 1/T$, $\mathbb{E}_\sigma$ denotes the statistical expectation with respect to the spin configurations in $\{-1,1\}^N$, and

$$m_\mu := \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu \sigma_i$$

are the pattern overlaps, or Mattis magnetizations [20]. Intuitively, the spin configurations selected by this Hamiltonian have the best possible overlap with the quenched patterns. In particular when the Gibbs average of m_μ is nonzero for some μ we say that this pattern is being retrieved. For a short but comprehensive summary of the main known results on this model we refer to Section II B of Ref. [16].

A generalization of the Hopfield model is obtained by replacing m^2 in Eq. (1) with a generic even function $u(m)$:

$$H_{N,p}(\sigma|\xi) = -\sum_{\mu=1}^{p} u(\sqrt{N} m_\mu). \tag{2}$$

It is physically interesting, but not necessary, to consider convex u [2,5,7,21–23]. Any convex, even and smooth u can be expressed as the cumulant generating function of a sub-Gaussian symmetric probability distribution with unit variance [24]. Interpreting the random variables with this distribution as ancillary spins, we obtain a correspondence between generalized Hopfield models and bipartite spin-glass models. The latter are defined as follows: consider a bipartite system, with one part containing N_1 spins denoted σ and the other N_2 spins written as τ. Also, let $N = N_1 + N_2$, $\alpha = N_2/N$ and define the partition function

$$Z_{N_1,N_2}(\beta;\boldsymbol{\xi}) = \mathbb{E}_{\sigma,\tau} \exp\left(\sqrt{\frac{\beta}{N}} \sum_{i=1}^{N_1}\sum_{\mu=1}^{N_2} \xi_i^\mu \sigma_i \tau_\mu\right). \tag{3}$$

Setting $u(x) = \ln \mathbb{E}_{\tau_1} e^{x\tau_1}$, the cumulant-generating function of the random variable τ_1, and marginalizing over all τ, we clearly obtain the partition function of a generalized Hopfield model with interaction u, as claimed. Therefore, we can think of the ξ_i^μ as patterns, each entry being independently drawn from $P_\xi(\xi_i^\mu)$. On the other hand, Eq. (3) can be viewed as a restricted Boltzmann machine, where a layer of visible units $\boldsymbol{\sigma}$ interacts with a layer of hidden units $\boldsymbol{\tau}$ through the weights $\boldsymbol{\xi}$ (see Fig. 1).

The standard Hopfield model is recovered when the ξ and the σ are binary and the τ_μ are Gaussian variables, but we study in this paper a much larger class of priors $P_\sigma(\sigma_i)$, $P_\tau(\tau_\mu)$, and $P_\xi(\xi_i^\mu)$. This corresponds in the generalized Hopfield model to varying the pattern distribution, the spin prior and the form of the interaction u.

Here we investigate the general phase diagram, especially with regards to the existence of a retrieval phase (focusing on single pattern retrieval) and its interplay with the spin-glass phase. Similar models of RBMs with generic priors have recently been studied using belief propagation and related methods in Refs. [16–18,25].

B. Model and RS equations

We shall use random variables that interpolate between Gaussian and binary distributions. Let $\Omega \in [0,1]$, $g \sim N(0,1)$, and ε be a symmetric random variable taking values ± 1. We define ζ as

$$\zeta(\Omega) = \sqrt{\Omega} g + \sqrt{1-\Omega}\varepsilon,$$

and we denote by $\mathcal{D}(\Omega)$ its probability distribution. Of course, $\mathbb{E}[\zeta] = 0$ and $\mathbb{E}[\zeta^2] = 1$ for all Ω.

Throughout we will draw both the patterns and the spins from $\mathcal{D}(\Omega)$, i.e., $\xi_i^\mu \sim \mathcal{D}(\Omega_\xi)$, $\sigma_i \sim \mathcal{D}(\Omega_\sigma)$, and $\tau_\mu \sim \mathcal{D}(\Omega_\tau)$ for $\Omega_\xi, \Omega_\sigma, \Omega_\tau \in [0,1]$. It will be useful to define the shorthand $\delta = \sqrt{1-\Omega_\xi}$.

To allow for retrieval phases in our analysis, we assume there are some numbers ℓ_1 and ℓ_2 of condensed patterns with pattern overlaps or Mattis magnetizations,

$$m^\mu(\boldsymbol{\sigma}) = \frac{1}{N_1}\sum_{i=1}^{N_1} \xi_i^\mu \sigma_i, \quad \mu = 1,\dots,\ell_2, \tag{4}$$

$$n^i(\boldsymbol{\tau}) = \frac{1}{N_2}\sum_{\mu=1}^{N_2} \xi_i^\mu \tau_\mu, \quad i = 1,\dots,\ell_1, \tag{5}$$

of order unity. We consider, for the sake of simplicity, the possible retrieval of a single pattern, i.e., $(\ell_1,\ell_2) = (1,0)$ or $(0,1)$; this is known as a pure-state ansatz. The general case of mixed states is a straightforward generalization [6] and can be considered a finer characterization of the retrieval region we are going to describe. On the other hand, the possible presence of frozen but disordered states (spin glass region) can be described by introducing the overlaps,

$$q(\boldsymbol{\sigma}^a,\boldsymbol{\sigma}^b) = \frac{1}{N_1}\sum_{i=1}^{N_1}\sigma_i^a\sigma_i^b, \quad r(\boldsymbol{\tau}^a,\boldsymbol{\tau}^b) = \frac{1}{N_2}\sum_{\mu=1}^{N_2}\tau_\mu^a\tau_\mu^b, \tag{6}$$

between two configurations $(\boldsymbol{\sigma}^a,\boldsymbol{\tau}^a)$ and $(\boldsymbol{\sigma}^b,\boldsymbol{\tau}^b)$ sampled from the Gibbs measure with the same pattern realisation, and the self-overlaps $Q(\boldsymbol{\sigma})$ and $R(\boldsymbol{\tau})$ in the case $a = b$. From a fairly standard replica calculation and the replica symmetry assumption (see Appendix A for more details), one gets that in the thermodynamic limit the Gibbs averages of the order parameters converge to the solutions of the following system:

$$m = \langle \xi \langle \sigma \rangle_{\sigma|z,\xi}\rangle_{z,\xi}, \tag{7}$$

$$n = \langle \xi \langle \tau \rangle_{\tau|\eta,\xi}\rangle_{\eta,\xi}, \tag{8}$$

$$q = \big\langle \langle \sigma \rangle^2_{\sigma|z,\xi}\big\rangle_{z,\xi}, \tag{9}$$

$$r = \big\langle \langle \tau \rangle^2_{\tau|\eta,\xi}\big\rangle_{\eta,\xi}, \tag{10}$$

$$Q = \langle \langle \sigma^2 \rangle_{\sigma|z,\xi}\rangle_{z,\xi}, \tag{11}$$

$$R = \langle \langle \tau^2 \rangle_{\tau|\eta,\xi}\rangle_{\eta,\xi}. \tag{12}$$

Here z and η are standard Gaussian random variables, while ξ is sampled from P_ξ. The distributions of σ and τ being averaged over are proportional to, respectively,

$$P_\sigma(\sigma)e^{\beta(1-\alpha)\Omega_\tau m\xi\sigma+\sqrt{\beta\alpha r}\, z\sigma+\beta\alpha(R-r)\sigma^2/2}, \tag{13}$$

$$P_\tau(\tau)e^{\beta\alpha\Omega_\sigma n\xi\tau+\sqrt{\beta(1-\alpha)q}\,\eta\tau+\beta(1-\alpha)(Q-q)\tau^2/2}. \tag{14}$$

These equations are valid also for more general spin priors $P_\sigma(\sigma)$ and $P_\tau(\tau)$, provided one then defines Ω_σ (and similarly Ω_τ) as the high-field response of the spins, in the sense that the average of σ over $P_\sigma(\sigma)e^{h\sigma}$ approaches $\Omega_\sigma h$ for large h.

We will repeatedly need averages over the distribution Eqs. (13) and (14). Taking the first as an example, the prior as defined is $P_\sigma(\sigma) \propto \sum_\varepsilon \exp[-(\sigma - \varepsilon\sqrt{1-\Omega_\sigma})^2/(2\Omega_\sigma)]$. Thus, the distribution Eq. (13) of σ has the generic form

$$Z_\sigma^{-1}\sum_\varepsilon e^{-\sigma^2/(2\gamma_\sigma)+(\phi_\sigma\varepsilon+h_\sigma)\sigma}, \tag{15}$$

where we have set $\phi_\sigma = \sqrt{1-\Omega_\sigma}/\Omega_\sigma$ and

$$\begin{aligned}\gamma_\sigma^{-1} &= \Omega_\sigma^{-1} - \beta\alpha(R-r),\\ h_\sigma &= \beta(1-\alpha)\Omega_\tau m\xi + \sqrt{\beta\alpha r}\, z.\end{aligned} \tag{16}$$

Averages over the distribution Eq. (15) then follow from the effective single-spin partition function

$$Z_\sigma = \int d\sigma \sum_\varepsilon e^{-\sigma^2/(2\gamma_\sigma)+(\phi_\sigma\varepsilon+h_\sigma)\sigma} \propto \sum_\varepsilon e^{\gamma_\sigma(\phi_\sigma\varepsilon+h_\sigma)^2/2}, \tag{17}$$

giving

$$\langle \sigma \rangle_{\sigma|z,\xi} = \partial_{h_\sigma} \ln Z_\sigma = \frac{\sum_\varepsilon \gamma_\sigma(\phi_\sigma\varepsilon + h_\sigma)e^{\gamma_\sigma(\phi_\sigma\varepsilon+h_\sigma)^2/2}}{\sum_\varepsilon e^{\gamma_\sigma(\phi_\sigma\varepsilon+h_\sigma)^2/2}} \tag{18}$$

$$= \gamma_\sigma h_\sigma + \gamma_\sigma\phi_\sigma \tanh(\gamma_\sigma\phi_\sigma h_\sigma). \tag{19}$$

The average of σ^2 can similarly be found from

$$\begin{aligned}\langle \sigma^2 \rangle_{\sigma|z,\xi} - \langle \sigma \rangle^2_{\sigma|z,\xi} &= \partial^2_{h_\sigma} \ln Z_\sigma = \partial_{h_\sigma}\langle \sigma \rangle_{\sigma|z,\xi}\\ &= \gamma_\sigma + \gamma_\sigma^2\phi_\sigma^2[1-\tanh^2(\gamma_\sigma\phi_\sigma h_\sigma)],\end{aligned} \tag{20}$$

hence

$$\langle \sigma^2 \rangle_{\sigma|z,\xi} = \gamma_\sigma + \gamma_\sigma^2\big(h_\sigma^2 + \phi_\sigma^2\big) + 2\gamma_\sigma^2\phi_\sigma h_\sigma \tanh(\gamma_\sigma\phi_\sigma h_\sigma). \tag{21}$$

Analogous results hold for the averages of τ over the distribution Eq. (14).

The RBM and equivalent Hopfield model defined above generalizes a number of existing models that are included as special cases. For $\Omega_\sigma = 0$, $\Omega_\tau = 1$, and $\Omega_\xi = 0$, we recover the standard Hopfield model, while if $\Omega_\xi = 1$, then we have the analog Hopfield model studied in Refs. [15,26,27] (see also Ref. [28] for the associated Mattis model). For $\Omega_\sigma = \Omega_\tau = 0$ we recover the bipartite Sherrington-Kirkpatrick (SK) model studied in Refs. [29,30]. In this case it is known that the thermodynamics is not affected by the pattern distribution [31]. Throughout this paper we consider only fully connected networks: results on the sparse case restricted to the Hopfield model can be found in Refs. [32–34].

C. Summary and further comments

The aim of this paper is to study the phase diagram of restricted Boltzmann machines with generic priors and pattern or weight distributions as defined above. In general, one expects three phases: a high-temperature (or paramagnetic) phase, in which the free energy equals its annealed bound and all the order parameters are zero; a glassy phase, where all pattern overlaps are still zero but replica symmetry breaking (RSB) is expected; and finally a retrieval phase, in which the overlap still has a glassy structure, but now one or more pattern overlaps have nonzero mean values. The precise organization of the thermodynamic states is unknown in the glassy and retrieval regions. In particular, while in the glassy phase it is supposed to be similar to the one of the SK model [9,23], the understanding of the retrieval phase remains severely limited [6,23,35] and represents an open challenge for theoretical and mathematical physics.

Throughout the paper the starting point for our analysis will be Eqs. (7)–(12). We will study them analytically and numerically in the various regimes.

The high-temperature transition is well understood by exact methods for the standard Hopfield model [6,35], for the analog Hopfield model 14,26,27], and for the bipartite SK model [29]. Moving beyond these special cases, in Sec. II we give a theoretical prediction for the transition of the order parameter q as the distributions of the priors and patterns vary. We will see that the transition is independent of the particular pattern distribution. We find explicit expressions for the transition line for $\Omega_\sigma = 0$ (one layer made of ±1 spins) and (with totally different methods in Appendix B) for $\Omega_\sigma = \Omega_\tau = 1$. The remaining intermediate cases are studied by numerically solving the self-consistency Eqs. (7)–(12) for the order parameters.

Next we analyze the retrieval region considering the retrieval of one single pattern. A simple argument shows that no retrieval is possible for $\Omega_\sigma = \Omega_\tau = 0$: retrieval requires giving up an $O(N)$ amount of entropy in the σ-system. This is worthwhile only if we can gain an extensive amount of energy. The pattern being retrieved gives a field of $O(\sqrt{N})$ acting on a τ spin, so the response of it needs to be also $O(\sqrt{N})$ to get an overall $O(N)$ energy gain. This is impossible for binary τ, but *is* possible for τ with a Gaussian tail, for which the cumulant generating function grows quadratically at infinity. Hence, we always consider $\Omega_\tau > 0$.

In Sec. III we look at the low-load regime in which $\alpha = 0,1$. It turns out that the transitions in m and the replica overlap q occur at the same temperature $T = \Omega_\tau$, and both transitions are continuous (as is known to be correct for the standard Hopfield model [6,35]). First we concern ourselves with binary spins $\Omega_\sigma = 0$, then we analyze $\Omega_\sigma > 0$, where it turns out that we need to add an appropriate spherical cutoff.

In Sec. IV we study numerically the retrieval transition at high load, i.e., $\alpha \in (0,1)$, so that the number of patterns is proportional to the system size. First we vary separately Ω_τ and Ω_ξ while keeping $\Omega_\sigma = 0$ fixed. At $\Omega_\tau = 1$ we see absence of retrieval in this analog Hopfield model ($\Omega_\xi = 1$), as expected from Ref. [27]. An analysis at $T = 0$ shows, furthermore, that the most efficient retrieval is given by the standard Hopfield model.

Moving on to $\Omega_\sigma > 0$ we find that the model is well-defined only for high temperature. However, it is interesting that while for $\Omega_\sigma = \Omega_\tau = 1$ (Gaussian bipartite model) the divergence of the partition function coincides with the glassy transition, in the intermediate cases there is still a region of retrieval in the phase diagram. Finally, when we regularize the model, again with a spherical cutoff, we observe a standard retrieval phase, with a reentrant behavior of the transition line. The latter would suggest a RSB scenario, as in the standard Hopfield model [36].

In Appendix A we derive Eqs. (7)–(12) and in Appendix B we briefly analyze the Gaussian bipartite spin glass via Legendre duality, a method introduced for the spherical spin glass in Ref. [37].

The model we analyze is, for Ω_σ, a neural network with soft spins. Soft-spin networks were introduced at an early stage of the development of the field by Hopfield in Ref. [38], but were not much studied in the sequel. From the (bipartite) spin-glass perspective, soft spins (spherical or Gaussian) permit analytic methods to be more easily applied, compared to the more commonly studied binary ±1 spins. Indeed, there is a substantial number of results in the literature. In Refs. [39] and [40] (see also Ref. [23]) two similar models of spherical neural networks are introduced, with spherical spins and quadratic (-like) interactions. The authors find the free energy to be RS and no retrieval region. However, in Ref. [39] it is noted that retrieval appears when a quartic term is added to the Hamiltonian. More recently in Ref. [41] a spherical spin-glass model was considered with random interaction given by a Wishart random matrix, which is closely related with the work in Refs. [23,39,40]. The authors find the free energy (which one can argue to be RS by comparison with the Wigner matrix case [37,42,43]) and its fluctuations for all temperatures. No retrieval is observed. Finally, a spherical bipartite spin glass is analyzed in Ref. [44] for high temperatures, far from the critical point, and the authors find the free energy in a variational form. Interestingly enough, for this model our analysis yields the same paramagnetic or spin-glass transition line as for the bipartite SK model, and no retrieval (see the Secs. IV C, IV D, and Appendix B).

As for the RSB scenario there are only a few results for bipartite models. To the best of our knowledge these are limited to 1RSB for the standard Hopfield model (see Refs. [6,45]) and to a partial mathematical investigation of the bipartite (in fact, multipartite) SK model [30,46]. Therefore, we will restrict ourselves only to RS approximations, where needed.

II. TRANSITION TO THE SPIN-GLASS PHASE

At very high temperature ($\beta = 0$) the distribution Eqs. (13) and (14) have no external effective fields and the thermodynamic state is completely random, with order parameters $m = n = q = r = 0$. Lowering the temperature, a spin-glass transition to a frozen but disordered state takes place, creating nonzero overlaps q and r while m and n remain zero. Assuming this transition is continuous, we can linearize Eqs. (9) and (10) for small q and r:

$$q \sim \beta\alpha r\langle\sigma^2\rangle_0^2 + o(r), \tag{22}$$

$$r \sim \beta(1-\alpha)q\langle\tau^2\rangle_0^2 + o(q). \tag{23}$$

Here $\langle\,\rangle_0$ denotes the expectation value with respect to Eqs. (13) and (14) with $q = r = 0$ (in particular without the random field). The resulting transition criterion is

$$1 = \beta^2\alpha(1-\alpha)\langle\sigma^2\rangle_0^2\langle\tau^2\rangle_0^2 = \beta^2\alpha(1-\alpha)Q^2R^2, \tag{24}$$

where, using Eqs. (11) and (12), Q and R are the solutions of

$$Q = \langle\sigma^2\rangle_0 = Z_\sigma^{-1}\mathbb{E}_\sigma\sigma^2 e^{\beta\alpha R\sigma^2/2}, \tag{25}$$

$$R = \langle\tau^2\rangle_0 = Z_\tau^{-1}\mathbb{E}_\tau\tau^2 e^{\beta(1-\alpha)Q\tau^2/2}. \tag{26}$$

This result does not depend on the particular pattern distribution $P_\xi(\xi)$ (see also Ref. [47]), but it does clearly involve the spin priors. With these priors fixed, the transition takes place at an inverse temperature $\beta_c(\alpha) > 0$ that is a function of α. For $\beta < \beta_c(\alpha)$ one finds that the self overlaps are the solutions of

$$Q = \left(1 - \beta\alpha\Omega_\sigma^2 R\right)/(1-\beta\alpha\Omega_\sigma R)^2, \tag{27}$$

$$R = \left(1 - \beta(1-\alpha)\Omega_\tau^2 Q\right)/(1-\beta(1-\alpha)\Omega_\tau Q)^2. \tag{28}$$

The relation Eq. (27) can be derived directly from Eq. (21) with $\gamma_\sigma^{-1} = \Omega_\sigma^{-1} - \beta\alpha R$ and $h_\sigma = 0$, and similarly for Eq. (28). Solving Eqs. (27) and (28) together with Eq. (24), $T_c(\alpha) = 1/\beta_c(\alpha)$ satisfies

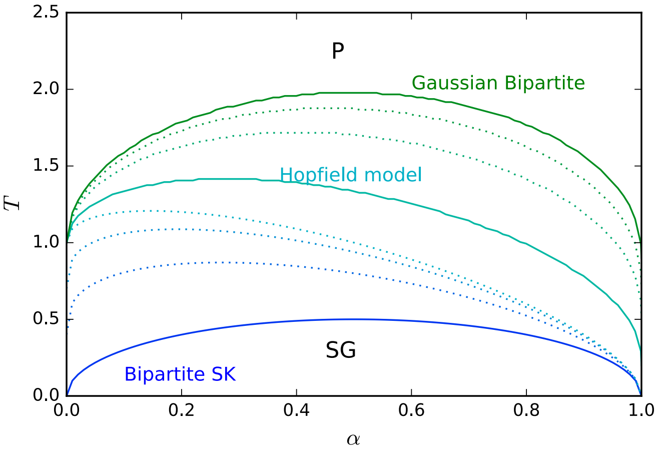

FIG. 2. Paramagnetic (P)–spin-glass (SG) transition line $T_c(\alpha)$ for different spin priors. The values of the relevant parameters, Ω_τ and Ω_σ, can be read off from each curve as $\Omega_\tau = T_c(0)$ and $\Omega_\sigma = T_c(1)$. The two continuous lines (lower, $\Omega_\sigma = \Omega_\tau = 0$, bipartite SK; upper, $\Omega_\sigma = \Omega_\tau = 1$, bipartite Gaussian) are completely symmetric with respect to exchange of the two network layers, i.e., the transformation $\alpha \to 1-\alpha$. In the middle the Hopfield critical line, $\Omega_\sigma = 0$, $\Omega_\tau = 1$.

(i) $\lim_{\alpha\to 0} T_c(\alpha) = \Omega_\tau, \quad \lim_{\alpha\to 1} T_c(\alpha) = \Omega_\sigma,$

(ii) $\lim_{\Omega_\sigma\to 0} T_c(\alpha) = \frac{1}{2}(2(1-\alpha)\Omega_\tau + \sqrt{\alpha(1-\alpha)}$
$+ [\alpha(1-\alpha) + 4\Omega_\tau(1-\Omega_\tau)(1-\alpha)\sqrt{\alpha(1-\alpha)}]^{1/2}),$

and of course the symmetric expression for $\Omega_\tau \to 0$, which is obtained by replacing in (ii) Ω_τ by Ω_σ and α by $1-\alpha$.

Relation (ii) recovers a number of known special cases. For $\Omega_\sigma = \Omega_\tau = 0$, one gets the critical line of the bipartite SK model $T_c = \sqrt{\alpha(1-\alpha)}$ as found in Ref. [29] (see also Ref. [30]). When $\Omega_\sigma = 0$ and $\Omega_\tau = 1$, one has the standard Hopfield model and finds $T_c = 1 - \alpha + \sqrt{\alpha(1-\alpha)}$ [12,15]. The case of both Gaussian priors ($\Omega_\sigma = \Omega_\tau = 1$) can be found independently using the Legendre duality between the Gaussian bipartite spin-glass model and the spherical Hopfield model studied in Refs. [39–41]; see Appendix B. The general bimodal case can be analyzed numerically; the results are shown in Fig. 2.

III. TRANSITION TO RETRIEVAL I: LOW LOAD

In the low-load regime the size of one layer is negligible with respect to the total size of the system, $N_1 \ll N$ or $N_2 \ll N$, corresponding, respectively, to $\alpha = 1$ or $\alpha = 0$. In this case it is possible to obtain Eq. (7) without any RS approximation since the model becomes a generalized ferromagnet. This can be studied in terms of the pattern overlaps only without the need to consider q and r [6,35]. Focusing on $\alpha = 0$ and linearizing Eq. (7) in m we get

$$m = \beta\Omega_\tau m + O(m^3),$$

which shows a bifurcation at $T = \Omega_\tau$. As is known in special cases, it therefore remains true for generic Ω_σ, Ω_τ, and Ω_ξ that the spin-glass and low-load retrieval transitions occur at the same temperature.

We next consider the strength of retrieval at temperatures below the transition: the inner average of Eq. (7) is, using Eq. (19) with $\gamma_\sigma = \Omega_\sigma$,

$$\langle\sigma\rangle_{\sigma|\xi} = \Omega_\sigma\beta\Omega_\tau m\xi + \sqrt{1-\Omega_\sigma}\tanh(\sqrt{1-\Omega_\sigma}\beta\Omega_\tau m\xi).$$

To carry out the remaining average over ξ, which by assumption is drawn from the bimodal distribution $\mathcal{D}(\Omega_\xi)$ with peaks at $\pm\delta = \pm\sqrt{1-\Omega_\xi}$, we set (see Sec. I B) $\xi = \delta\varepsilon + \sqrt{\Omega_\xi}g$. As $\langle\sigma\rangle_{\sigma|\xi}$ is odd in ξ, the two possible values of $\varepsilon = \pm 1$ give the same contribution to $\langle\xi\langle\sigma\rangle_{\sigma|\xi}\rangle$ and we have to average only over g. After an integration by parts, this gives

$$m = f_{\beta,\mathbf{\Omega}}(m), \tag{29}$$

with

$$\begin{aligned} f_{\beta,\mathbf{\Omega}}(m) &= \beta\Omega_\sigma\Omega_\tau m + \sqrt{1-\Omega_\sigma}\{\delta\,\bar{t}(\beta\sqrt{1-\Omega_\sigma}\Omega_\tau\delta\, m, \sqrt{v}) \\ &\quad + \beta\sqrt{1-\Omega_\sigma}\Omega_\tau\Omega_\xi m[1 - \overline{t^2}(\beta\sqrt{1-\Omega_\sigma}\Omega_\tau\delta\, m, \sqrt{v})]\}. \end{aligned} \tag{30}$$

Here have introduced the abbreviations

$$\bar{t}(a,b) = \langle\tanh(a+b\,g)\rangle_g, \qquad \overline{t^2}(a,b) = \langle\tanh^2(a+b\,g)\rangle_g, \tag{31}$$

where the averages are over a zero mean, unit variance Gaussian random variable g. We have also defined

$$v = \beta^2(1-\Omega_\sigma)\Omega_\tau^2\Omega_\xi m^2. \tag{32}$$

For binary spins ($\Omega_\sigma = 0$), $|\sigma| = 1$ and so $f_{\beta,\mathbf{\Omega}}(m) = \langle\xi\langle\sigma\rangle_{\sigma|\xi}\rangle$ is bounded (between $-\langle|\xi|\rangle$ and $+\langle|\xi|\rangle$). This ensures that a nontrivial solution m of Eq. (29) always exists below the retrieval transition. The zero temperature limit of m can be found explicitly: for $\beta \to \infty$, $\langle\sigma\rangle_{\sigma|\xi} \to \text{sgn}(m\xi)$ so $f_{\beta,\mathbf{\Omega}}(m) \to \text{sgn}(m)\langle\xi\,\text{sgn}(\xi)\rangle$, and therefore $m \to \pm\langle|\xi|\rangle$ with

$$\langle|\xi|\rangle = \sqrt{\frac{2\Omega_\xi}{\pi}}e^{-\delta^2/(2\Omega_\xi)} + \delta\,\text{erf}\left(\frac{\delta}{\sqrt{2\Omega_\xi}}\right). \tag{33}$$

For generic soft spins ($\Omega_\sigma > 0$), on the other hand, $f_{\beta,\mathbf{\Omega}}(m)$ is no longer bounded but grows as $\beta\Omega_\sigma\Omega_\tau m$ for large $|m|$. The spontaneous magnetization, which is the solution of $m = f_{\beta,\mathbf{\Omega}}(m)$, therefore diverges at $T_c = \Omega_\sigma\Omega_\tau$ as temperature is lowered; see Fig. 3. For lower T the model is ill-defined as we are going to see in more detail in the next section, thus we need to regularize the spin distribution in at least one network layer.

To fix the choice of regularization we note that for a large system, every rotationally invariant weight on the vector of σ-spins is equivalent to a rigid constraint at some fixed radius. Without loss of generality we therefore regularize by multiplying the σ-prior by the spherical constraint $\delta(N - \sum_{i=1}^N \sigma_i^2)$. The resulting prior still depends on Ω_σ; at $\Omega_\sigma = 1$ it is a uniform distribution on the sphere and we obtain the spherical Hopfield model studied in Refs. [39–41]. At $\Omega_\sigma = 0$, on the other hand, the regularization constraint is redundant and we recover the standard Hopfield model. One can now analyze the regularized model using similar replica computations to those above. The only difference is an extra Gaussian factor $e^{-\omega\sigma^2/2}$ in the effective σ-spin distribution. Here ω is a Lagrange multiplier that is determined from the spherical constraint $Q = 1$. It changes the variance of the two Gaussian peaks from

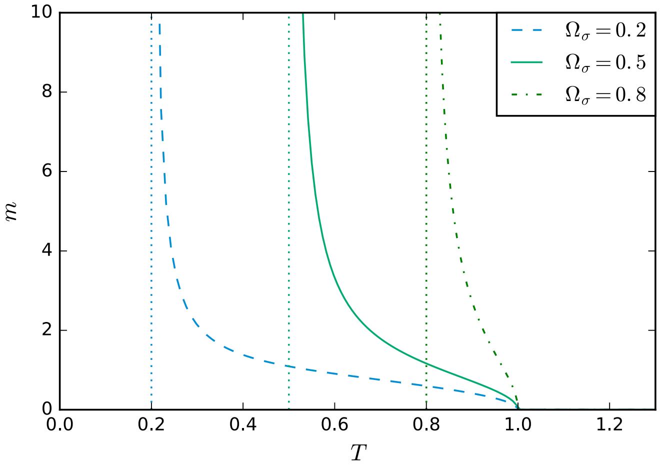

FIG. 3. Soft bipartite model at low load ($\alpha = 0$): for a generic pattern distribution (here $\delta = 0.5$) a spontaneous magnetization appears at $T = \Omega_\tau$, diverging at $T = \Omega_\sigma\Omega_\tau$.

Ω_σ to $\gamma_\sigma = (\Omega_\sigma^{-1} + \omega)^{-1}$. Accordingly, instead of $f_{\beta,\mathbf{\Omega}}$ in Eq. (30) one obtains a modified function

$$f_{\beta,\mathbf{\Omega},\gamma_\sigma}(m) = \beta\gamma_\sigma\Omega_\tau m + \gamma_\sigma\phi_\sigma\{\delta\,\bar{t}(\beta\gamma_\sigma\phi_\sigma\Omega_\tau\delta\, m, \sqrt{v}) + \beta\gamma_\sigma\phi_\sigma\Omega_\tau\Omega_\xi m[1 - \overline{t^2}(\dots)]\}, \tag{34}$$

where the arguments of $\overline{t^2}$ are the same as for $\bar{t}$. Note that the first term of Eq. (30) has become $\beta\gamma_\sigma\Omega_\tau m$ and all occurrences of $\sqrt{1 - \Omega_\sigma} = \Omega_\sigma\phi_\sigma$ have been replaced by $\gamma_\sigma\phi_\sigma$. Accordingly, also v now has the more general form

$$v = \beta^2\gamma_\sigma^2\phi_\sigma^2\Omega_\tau^2\Omega_\xi m^2. \tag{35}$$

The value of ω or equivalently γ_σ is defined from the condition $Q = \langle\sigma^2\rangle_{\sigma,\xi} = 1$, where Q can be worked out using Eq. (21) as

$$Q = \gamma_\sigma + \gamma_\sigma^2(\beta^2\Omega_\tau^2 m^2 + \phi_\sigma^2) + 2\beta\gamma_\sigma^2\phi_\sigma\Omega_\tau\delta\, m\,\bar{t}(\dots) + 2\beta^2\gamma_\sigma^3\phi_\sigma^2\Omega_\tau^2\Omega_\xi m^2[1 - \overline{t^2}(\dots)]. \tag{36}$$

The last two terms are proportional to the last two terms in Eq. (34), and hence to $(1 - \beta\gamma_\sigma\Omega_\tau)m$; if one traces back through the derivation this comes from the fact that both results are proportional to $\langle h_\sigma \tanh(\gamma_\sigma\phi_\sigma h_\sigma)\rangle$. With this simplification one obtains the equivalent expression

$$Q = \gamma_\sigma + \gamma_\sigma^2\phi_\sigma^2 + \beta\gamma_\sigma\Omega_\tau m^2(2 - \beta\gamma_\sigma\Omega_\tau) = 1. \tag{37}$$

For $\Omega_\sigma \to 0$, one has $\gamma_\sigma \approx \Omega_\sigma$, which vanishes as $\Omega_\sigma \to 0$, while $\gamma_\sigma\phi_\sigma = \sqrt{1 - \Omega_\sigma}\gamma_\sigma/\Omega_\sigma \to 1$. For this limiting case of Boolean σ-spins the constraint Eq. (37) is therefore automatically satisfied as expected. More generally, while Eq. (34) suggests the asymptotic behavior $f_{\beta,\mathbf{\Omega},\gamma_\sigma}(m) \approx \beta\gamma_\sigma\Omega_\tau m$ for $m \to \infty$, this first term is not the leading contribution because $\gamma_\sigma \sim 1/m^2$ for large m. Instead, the last two terms in Eq. (34) dominate, giving a nonzero constant asymptote. Near $m = 0$, on the other hand, $f_{\beta,\mathbf{\Omega},\gamma_\sigma}(m)$ goes as $\beta\Omega_\tau(\gamma_\sigma + \gamma_\sigma^2\phi_\sigma^2)m$. From Eq. (37), $\gamma_\sigma + \gamma_\sigma^2\phi_\sigma^2 = 1 + O(m^2)$, thus the ferromagnetic transition remains at $T_c = \Omega_\tau$ in the model with the spherical constraint. (One easily checks that $\gamma_\sigma + \gamma_\sigma^2\phi_\sigma^2 = 1$ implies as the physical solution $\gamma_\sigma = \Omega_\sigma$, so that the regularizer ω increases smoothly from zero at the transition.) For temperatures below T_c, one generally has to find m numerically. Results are shown in Fig. 4. As expected for a regularized model, m remains finite at all T. In the low-temperature limit it always reaches its maximum value $m \to 1$. One can easily check this from Eqs. (34) and (37): the latter implies for $m = 1$ that $\beta\gamma_\sigma\Omega_\tau \to 1$ (see the lower plots in Fig. 4). Hence, the first term on the right-hand side of Eq. (34) also approaches unity as it should from $m = f_{\beta,\mathbf{\Omega},\gamma_\sigma}(m)$ while the other terms in Eq. (34) vanish in the limit.

IV. TRANSITION TO RETRIEVAL II: HIGH LOAD

Now we study the entire phase diagram of the model, in particular with regards to the presence and stability of a retrieval region. We now use the full definition of γ_σ and h_σ from Eq. (16), along with the analogous definition for γ_τ:

$$\gamma_\sigma^{-1} = \Omega_\sigma^{-1} - \beta\alpha(R - r), \quad \gamma_\tau^{-1} = \Omega_\tau^{-1} - \beta(1 - \alpha)(Q - q),$$
$$h_\sigma = \beta(1 - \alpha)\Omega_\tau m\xi + \sqrt{\beta\alpha r}\, z. \tag{38}$$

Furthermore, we abbreviate the variance of the Gaussian part of $\gamma_\sigma\phi_\sigma h_\sigma$ as

$$v = \beta^2(1 - \alpha)^2\gamma_\sigma^2\phi_\sigma^2\Omega_\tau^2\Omega_\xi m^2 + \beta\alpha\gamma_\sigma^2\phi_\sigma^2 r, \tag{39}$$

where compared to Eq. (32) we again have the replacement of $\sqrt{1 - \Omega_\sigma}$ by $\gamma_\sigma\phi_\sigma$, and otherwise the incorporation of the α-dependence and the new term proportional to r. Then, taking the averages with respect to ξ and z we have, using Eqs. (19) and (21) and integrating by parts where appropriate,

$$\begin{aligned} m &= \langle\xi\langle\sigma\rangle_{\sigma|z,\xi}\rangle_{z,\xi} \\ &= \beta(1 - \alpha)\gamma_\sigma\Omega_\tau m + \gamma_\sigma\phi_\sigma\{\delta\,\bar{t}[\beta(1 - \alpha)\gamma_\sigma\phi_\sigma\Omega_\tau\delta\, m, \sqrt{v}] + \beta(1 - \alpha)\gamma_\sigma\phi_\sigma\Omega_\tau\Omega_\xi m[1 - \overline{t^2}(\dots)]\}, \\ q &= \left\langle\langle\sigma\rangle_{\sigma|z,\xi}^2\right\rangle_{z,\xi} = \langle[\gamma_\sigma h_\sigma + \gamma_\sigma\phi_\sigma \tanh(\gamma_\sigma\phi_\sigma h_\sigma)]^2\rangle_{z,\xi} \\ &= \beta^2(1 - \alpha)^2\gamma_\sigma^2\Omega_\tau^2 m^2 + \beta\alpha\gamma_\sigma^2 r + \gamma_\sigma^2\phi_\sigma^2\overline{t^2}(\dots) + 2\beta(1 - \alpha)\gamma_\sigma^2\phi_\sigma\Omega_\tau\delta\, m\,\bar{t}(\dots) + 2\gamma_\sigma v[1 - \overline{t^2}(\dots)] \end{aligned}$$

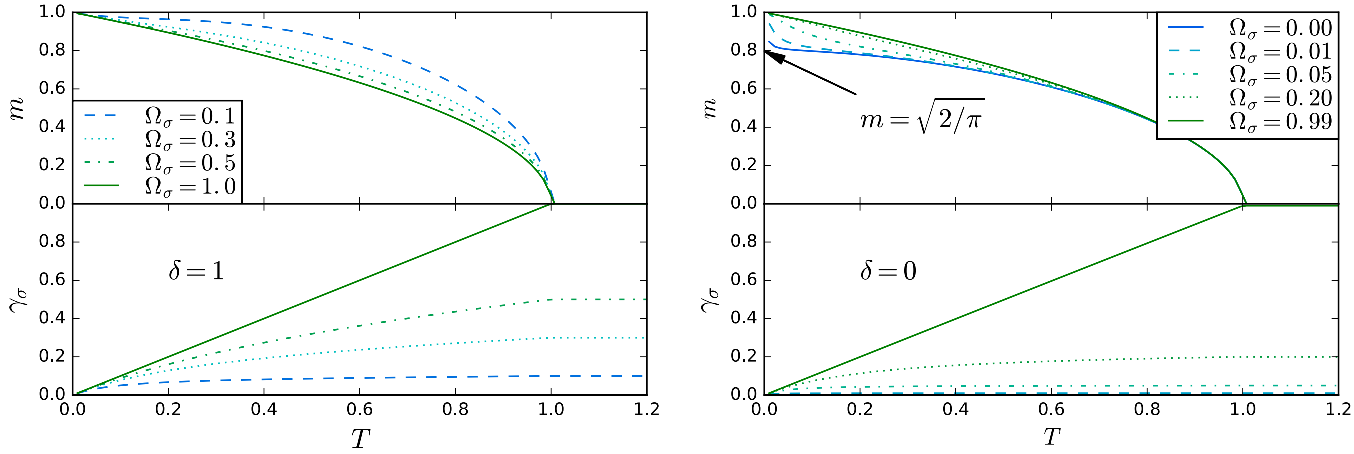

FIG. 4. Soft model with spherical constraint at low load ($\alpha = 0$). Spontaneous magnetization still occurs at $T = \Omega_\tau$ increasing until $T = 0$. Left panels $\delta = 1$, right panels $\delta = 0$. As $\Omega_\sigma \to 0$, m approaches the value Eq. (33) at low T. But at any $\Omega_\sigma > 0$, m eventually peels off from this asymptote to reach $m = 1$ for $T \to 0$. Lower panels show the behavior of γ_σ: it tends to zero linearly at low temperature, $\gamma_\sigma \approx T/\Omega_\tau$, while for $T \geq \Omega_\tau$, $\gamma_\sigma = \Omega_\sigma$.

$$= \beta(1-\alpha)\gamma_\sigma\Omega_\tau(2-\beta(1-\alpha)\Omega_\tau\gamma_\sigma)m^2 + \beta\alpha\gamma_\sigma^2\left(1+2\gamma_\sigma\phi_\sigma^2\right)r + \gamma_\sigma^2\phi_\sigma^2(1-2\beta\alpha\gamma_\sigma r)\overline{t^2}(\ldots),$$

$$Q = \langle\langle\sigma^2\rangle_{\sigma|z,\xi}\rangle_{z,\xi} = q + \gamma_\sigma + \gamma_\sigma^2\phi_\sigma^2[1-\overline{t^2}(\ldots)],$$

where all tanh averages $\bar{t}$ and $\overline{t^2}$ are evaluated for the same parameters, as given in the equation for m. In the final expression for q we have eliminated the $\bar{t}$ term using the expression for m. Repeating the same argument for the effective distribution of the τ spins, we get the equations for the other order parameters simply by exchanging labels appropriately and replacing α with $1-\alpha$, bearing in mind also that the corresponding magnetization parameter is $n = 0$. This gives the following additional equations:

$$r = \beta(1-\alpha)\gamma_\tau^2\left(1+2\gamma_\tau\phi_\tau^2\right)q + \gamma_\tau^2\phi_\tau^2[1-2\beta(1-\alpha)\gamma_\tau q]\bar{t}^2 \times [0,\gamma_\tau\phi_\tau\sqrt{\beta(1-\alpha)q}], \tag{40}$$

$$R = r + \gamma_\tau + \gamma_\tau^2\phi_\tau^2\{1-\overline{t^2}[0,\gamma_\tau\phi_\tau\sqrt{\beta(1-\alpha)q}]\}. \tag{41}$$

A. One Boolean layer

In the case where the σ-spins are Boolean, $\Omega_\sigma = 0$, the saddle-point Eqs. (40) simplify considerably. From Eq. (38), one has as before $\gamma_\sigma \approx \Omega_\sigma \to 0$ and $\gamma_\sigma\phi_\sigma \to 1$. This leads to

$$m = \delta\,\bar{t}[\beta(1-\alpha)\Omega_\tau\delta\, m,\sqrt{v}] + \beta(1-\alpha)\Omega_\tau\Omega_\xi m[1-\overline{t^2}(\ldots)], \tag{42}$$

$$q = \overline{t^2}(\ldots), \tag{43}$$

where after inserting Eq. (40) for r the Gaussian field variance can be written as $v = \beta^2(1-\alpha)^2 V$ with

$$V = \Omega_\tau^2\Omega_\xi m^2 + \frac{\alpha}{1-\alpha}\gamma_\tau^2\left(1+2\gamma_\tau\phi_\tau^2\right)q + \frac{\alpha}{1-\alpha}\gamma_\tau^2\phi_\tau^2\{[\beta(1-\alpha)]^{-1} - 2\gamma_\tau q\}\overline{t^2}[0,\gamma_\tau\phi_\tau\sqrt{\beta(1-\alpha)q}]. \tag{44}$$

Solutions of Eq. (42) are shown in Fig. 5. Starting from the standard Hopfield phase diagram ($\Omega_\xi = 0$ and $\Omega_\tau = 1$) the retrieval region gradually disappears with decreasing Ω_ξ or increasing Ω_τ. In the first case it shifts toward the T axis, as the critical temperature for $\alpha = 0$ is independent of Ω_ξ. In the second case, both the retrieval and spin-glass transition lines shift toward the α axis, as the critical α at $T = 0$ is independent of Ω_τ as we will see shortly.

B. Zero-temperature limit

Useful insight into the $\Omega_\sigma = 0$ case can be obtained by further specializing to the limit $T \to 0$ (i.e., $\beta \to \infty$). In this scenario,

$$\bar{t}(\beta a,\beta b) \to \langle\text{sgn}(a+b\eta)\rangle = \text{erf}(a/\sqrt{2}b), \tag{45}$$

and, putting $w = \beta(a+bg)$,

$$\beta[1-\bar{t}^2(\beta a,\beta b)] = \beta\int\frac{dw}{\sqrt{2\pi}\beta b}\exp[-(w-\beta a)^2/(2\beta^2b^2)][1-\tanh^2(w)] \to \int\frac{dw}{\sqrt{2\pi}b}\exp[-a^2/(2b^2)][1-\tanh^2(w)] = \frac{\sqrt{2}}{\sqrt{\pi}b}\exp[-a^2/(2b^2)]. \tag{46}$$

If we set $v = \beta^2(1-\alpha)^2 V$ as before and then apply the above large-β identities in the Eq. (42) for m we get

$$m = \delta\ \text{erf}(\Omega_\tau\delta\, m/\sqrt{2V}) + \Omega_\tau\Omega_\xi m\frac{\sqrt{2}}{\sqrt{\pi V}}s \times \exp(-\Omega_\tau^2\delta^2 m^2/2V). \tag{47}$$

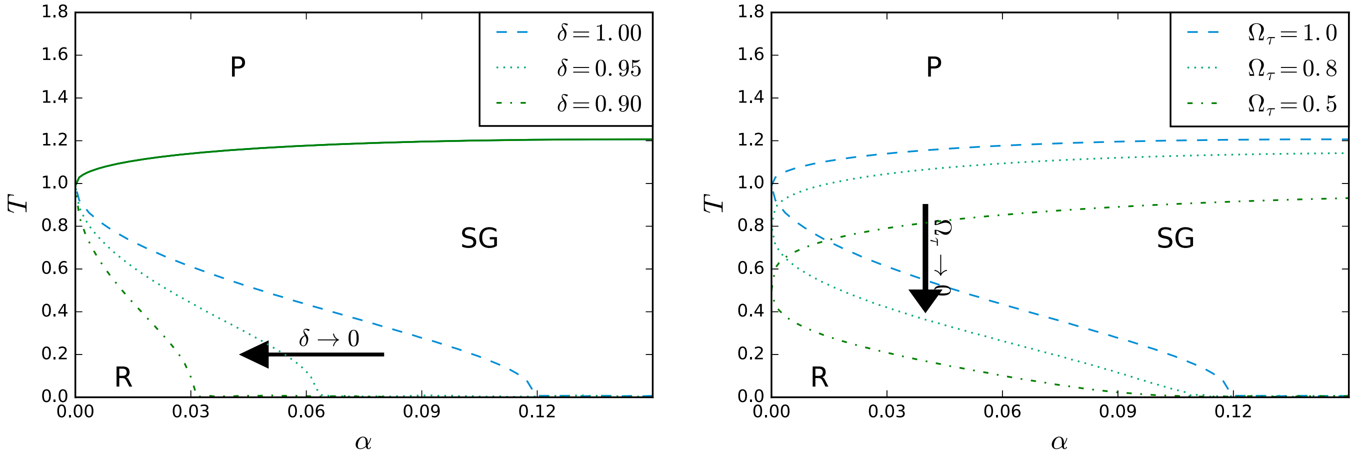

FIG. 5. Phase diagrams with one Boolean layer ($\Omega_\sigma = 0$), showing the paramagnetic (P), spin-glass (SG), and retrieval (R) regions. Left panel: (T,α) phase diagram for $\Omega_\tau = 1$ and different values of δ. The retrieval transition line moves toward the T axis as δ decreases while the critical temperature at $\alpha = 0$ remains fixed. Right panel: phase diagram for $\delta = 1$ and different values of Ω_τ. Both transition lines move toward the α axis as Ω_τ decreases while now the critical load at $T = 0$ is fixed.

Equation (43) for q has a limit in terms of $C = \beta(1-\alpha)(1-q)$:

$$C = \frac{\sqrt{2}}{\sqrt{\pi}\, v} \exp\left(-\Omega_\tau^2 \delta^2 m^2/2V\right). \tag{48}$$

Finally, for V in Eq. (44) the zero-temperature limit is simple as $\overline{t^2}(0,\beta b) \to 1$ and $q \to 1$, giving

$$V = \Omega_\tau^2 \Omega_\xi m^2 + \frac{\alpha}{1-\alpha}\gamma_\tau^2 = \Omega_\tau^2 \Omega_\xi m^2 + \frac{\alpha}{1-\alpha}(\Omega_\tau^{-1} - C)^{-2}. \tag{49}$$

One can reduce these three equations to a single one for $x = \Omega_\tau m/\sqrt{2V}$, which reads

$$x = F_{\delta,\alpha}(x),$$

$$F_{\delta,\alpha}(x) = \frac{\delta\, \mathrm{erf}(\delta x) - \frac{2}{\sqrt{\pi}} x \delta^2 e^{-\delta^2 x^2}}{[2\alpha + 2(1-\delta^2)(\delta\, \mathrm{erf}(\delta x) - \frac{2}{\sqrt{\pi}} x\delta^2 e^{-\delta^2 x^2})^2]^{1/2}}, \tag{50}$$

We leave the derivation of this result to the end of this section. One sees that $F_{\delta,\alpha}(x)$ is strictly increasing, starting from zero and approaching $\delta/\sqrt{2\alpha + 2(1-\delta^2)\delta^2}$ for large x (Fig. 6). Note also that Ω_τ has no effect on the value of x and only affects the coefficient in the linear relation between x and m.

For fixed δ, a first-order phase transition occurs in the self-consistency condition Eq. (50) as α increases. The transition value $\alpha_c(\delta)$ is largest for $\delta = 1$ and decreases to zero quite rapidly as $\delta \to 0$; see Fig. 7. For $\alpha < \alpha_c(\delta)$ a nonzero solution of Eq. (50) exists, with x (thus m) growing as α decreases. In particular, as $\alpha \to 0$, $x = F_{\delta,\alpha}(x) \to 1/\sqrt{2(1-\delta^2)} = 1/\sqrt{2\Omega_\xi}$. In this low-load limit one then recovers for m Eq. (33) as we show below.

We remark that since for any $0 < \alpha < 1$, $F_{\delta,a}(x) \to 0$ as $\delta \to 0$, one also has $m \to 0$ (with a first-order phase transition; see Fig. 7). For $\alpha = 0$, on the other hand, we see from Eq. (47) that $m \to \sqrt{2/\pi}$ as $\delta \to 0$, which is consistent with the data shown in Fig. 7. Thus, the Hopfield model retrieves Gaussian patterns only for $\alpha = 0$, but not at high load.

We close this section by outlining the derivation of Eq. (50). Bearing in mind $\delta = \sqrt{1 - \Omega_\xi}$, Eq. (47) for m becomes

$$m = \delta\, \mathrm{erf}(\delta\, x) + (1-\delta^2)\frac{2}{\sqrt{\pi}} x e^{-\delta^2 x^2}, \tag{51}$$

while for C one gets

$$C = \frac{2}{\sqrt{\pi}} \frac{x}{\Omega_\tau m} \exp(-\delta^2 x^2). \tag{52}$$

Thus,

$$\begin{aligned} V &= \Omega_\tau^2(1-\delta^2)m^2 + \frac{\alpha}{1-\alpha}\big[\Omega_\tau^{-1} - 2x/(\sqrt{\pi}\,\Omega_\tau m) \\ &\quad \times \exp(-\delta^2 x^2)\big]^{-2} \\ &= \Omega_\tau^2 m^2 \left\{1 - \delta^2 + \frac{\alpha}{1-\alpha}[m - (2/\sqrt{\pi})x \exp(-\delta^2 x^2)]^{-2}\right\} \\ &= \Omega_\tau^2 m^2 \bigg\{1 - \delta^2 + \frac{\alpha}{1-\alpha}[\delta\, \mathrm{erf}(\delta x) - (2/\sqrt{\pi})\delta^2 x \\ &\quad \times \exp(-\delta^2 x^2)]^{-2}\bigg\}. \end{aligned} \tag{53}$$

Now we set

$$\begin{aligned} F_{\delta,\alpha}(x) = \bigg\{ & 2(1-\delta^2) + 2(1-\alpha) \\ & \times \left[\delta\, \mathrm{erf}(\delta x) - \frac{2}{\sqrt{\pi}} x \delta^2 e^{-\delta^2 x^2}\right]^{-2}\bigg\}^{-1/2}, \end{aligned} \tag{54}$$

and we readily get Eq. (50).

C. Soft models

Models with both Gaussian spins are typically ill-defined at low temperature, due to the occurrence of negative eigenvalues in the interaction matrix. In the fully Gaussian model ($\Omega_\sigma = \Omega_\tau = 1$) the line where the partition function diverges coincides exactly with the paramagnetic to spin-glass transition. In this case, the distributions $P(\sigma|z,\xi)$ and $P(\tau|\eta,\xi)$ of Eqs. (13) and (14) are, respectively,

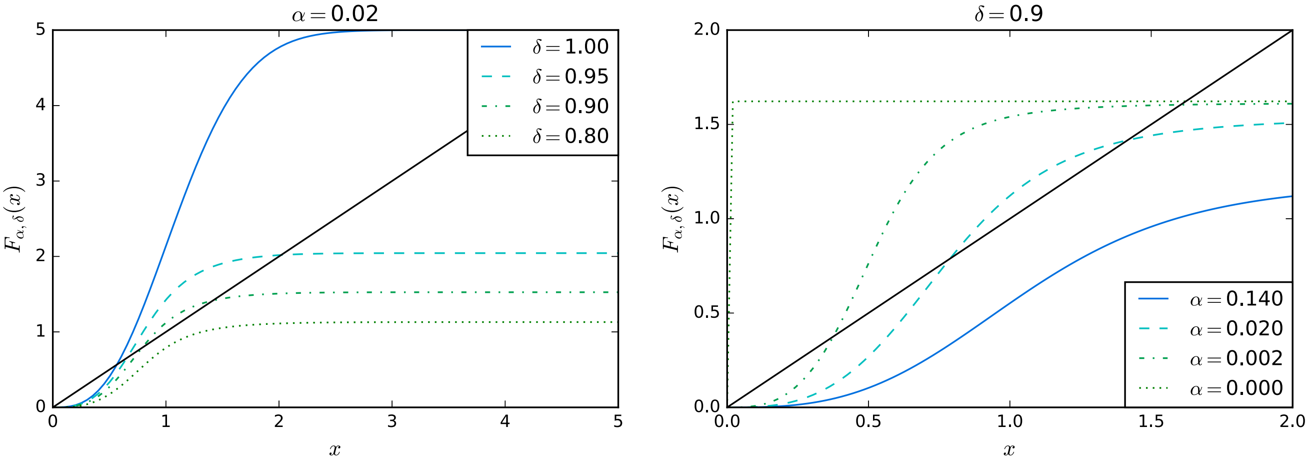

FIG. 6. Plot of $F_{\delta,\alpha}(x)$. It tends uniformly to zero as $\delta \to 0$ at fixed α (left panel), while it approaches $1/\sqrt{2(1-\delta^2)}$ as $\alpha \to 0$ at fixed δ (right panel).

proportional to

$$e^{\beta(1-\alpha)\Omega_\tau m\xi\sigma+\sqrt{\beta\alpha r}z\sigma-\frac{1}{2}[1-\beta\alpha(R-r)]\sigma^2}, \tag{55}$$

$$e^{\beta\alpha\Omega_\sigma n\xi\tau+\sqrt{\beta(1-\alpha)q}\tau\eta-\frac{1}{2}[1-\beta(1-\alpha)(Q-q)]\tau^2}. \tag{56}$$

Both these distributions are therefore Gaussian with variances Σ_σ, Σ_τ, defined by $\Sigma_\sigma^{-1} = 1 - \beta\alpha(R-r)$ and $\Sigma_\tau^{-1} = 1 - \beta(1-\alpha)(Q-q)$. The equations for R and Q read

$$Q = \langle\langle\sigma^2\rangle_{\sigma|z,\xi}\rangle_{z,\xi} = q + \Sigma_\sigma, \tag{57}$$

$$R = \langle\langle\tau^2\rangle_{\tau|\eta}\rangle_{\eta,\xi} = r + \Sigma_\tau. \tag{58}$$

Thus,

$$\begin{aligned}\Sigma_\sigma &= \frac{1}{1-\beta\alpha\Sigma_\tau},\\ \Sigma_\tau &= \frac{1}{1-\beta(1-\alpha)\Sigma_\sigma},\end{aligned} \tag{59}$$

and one has to study the equation $\mathcal{I}(\Sigma_\sigma) = \Sigma_\sigma$, where

$$\mathcal{I}(\Sigma_\sigma) = \frac{1-\beta(1-\alpha)\Sigma_\sigma}{1-\beta\alpha-\beta(1-\alpha)\Sigma_\sigma}. \tag{60}$$

The function $\mathcal{I}(x)$ is a hyperbola diverging at $x = (1-\beta\alpha)/\beta(1-\alpha)$; see Fig. 8. It is positive only for x below this value, so this is the range we need to consider as $\Sigma_\sigma > 0$. For small β one has a solution near $\Sigma_\sigma = 1$, which increases with β. At some $\hat{\beta}_c$, $\mathcal{I}(x)$ becomes tangent to x and for still larger β there are no intersections. After some calculations using Eq. (60) one finds for the threshold $\hat{\beta}_c$

$$\begin{aligned}1 &= \hat{\beta}_c^2\alpha(1-\alpha)\Sigma_\sigma^2\left(\frac{1}{1-\beta(1-\alpha)\Sigma_\sigma}\right)^2\\ &= \hat{\beta}_c^2\alpha(1-\alpha)\Sigma_\sigma^2\Sigma_\tau^2,\end{aligned} \tag{61}$$

which exactly coincides with the paramagnetic / spin glass transition temperature Eq. (24) as anticipated.

We note that we can compute the divergence of the partition function of the model also directly, by diagonalizing the

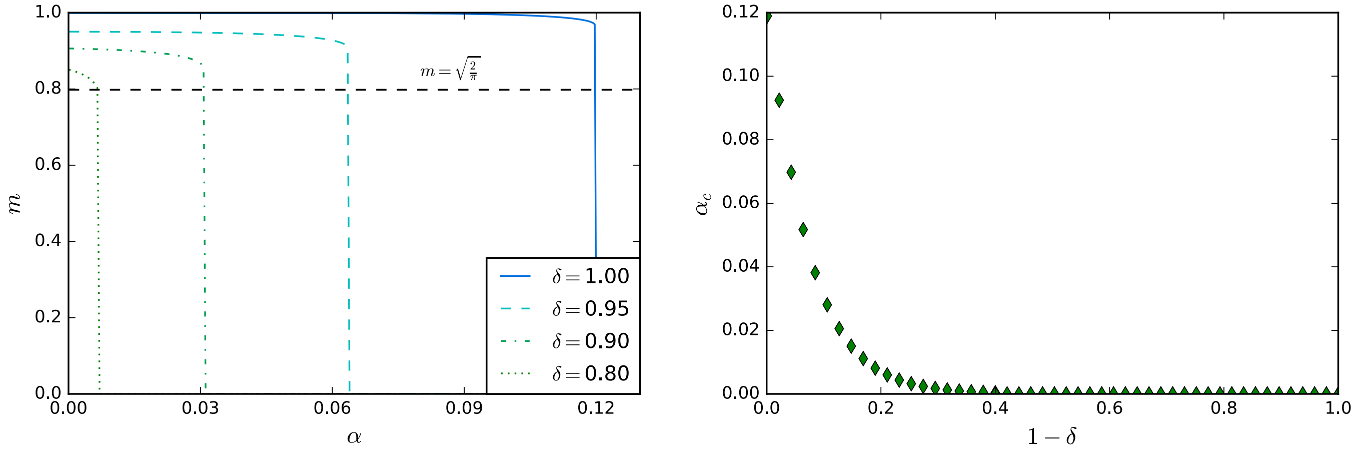

FIG. 7. Left panel: magnetization vs. α for different values of δ; at $\alpha_c(\delta)$ a first-order phase transition occurs. The low-load pattern overlap $m(\alpha = 0;\delta)$ tends to $\sqrt{\frac{2}{\pi}}$ as $\delta \to 0$. Right panel: $\alpha_c(\delta)$ plotted versus $1-\delta$: $\alpha_c(1) = 0.12\ldots$, while rapidly $\alpha_c(\delta) \to 0$ as $\delta \to 0$.

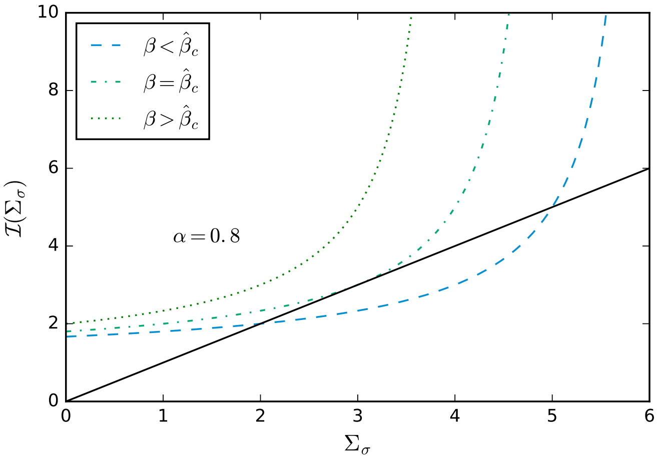

FIG. 8. $\mathcal{I}(x)$ vs. x for different values of β. At $\hat{\beta}_c$, $\mathcal{I}(x)$ is tangent to x.

interaction matrix (i.e., the weight matrix):

$$Z_N(\beta,\alpha;\xi) = \mathbb{E}_{\sigma,\tau} \exp\left(\sqrt{\frac{\beta}{N}} \sum_{i=1}^{N_1}\sum_{\mu=1}^{N_2} \xi_i^\mu \sigma_i \tau_\mu\right)$$
$$= \mathbb{E}_\sigma \exp\left(\frac{\beta}{2N} \sum_{i,j=1}^{N_1}\sum_{\mu=1}^{N_2} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j\right)$$
$$= \mathbb{E}_\sigma \exp\left(\frac{\beta\alpha}{2} \sum_{i,j=1}^{N_1} M_{ij}\sigma_i\sigma_j\right). \tag{62}$$

Here $\boldsymbol{M} = \frac{1}{N_2}\boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}}$ is a Wishart matrix so its empirical eigenvalue spectrum converges to the Marchenko-Pastur distribution for large N_1, which is nonzero only between $(1 \pm \sqrt{(1-\alpha)/\alpha})^2$. Using a suitable orthogonal transformation on the spin variables we can diagonalise $\boldsymbol{M}$, so that

$$Z_N(\beta,\alpha;\xi) = \mathbb{E}_\sigma \exp\left(\frac{\beta\alpha}{2}\sum_i^{N_1} \lambda_i \sigma_i^2\right).$$

This is well-defined as long as $\max_i [\beta(1-\alpha)\lambda_i] < 1$. Using the largest eigenvalue from Marchenko-Pastur, $\max_i \lambda_i = (1 + \sqrt{(1-\alpha)/\alpha})^2$ for large N, we get for the critical temperature

$$T_c(\alpha) = (\sqrt{\alpha} + \sqrt{1-\alpha})^2.$$

It can be checked that the spin-glass transition line numerically computed in Sec. II coincides with $T = T_c(\alpha)$. In the general case $0 < \Omega_\sigma, \Omega_\tau < 1$ we simply remark that (recall that the g are $\mathcal{N}(0,1)$ and $\varepsilon = \pm 1$)

$$\sum_{i=1}^{N_1}\sum_{\mu=1}^{N_2} \xi_i^\mu \sigma_i \tau_\mu$$
$$= \sqrt{(1-\Omega_\sigma)(1-\Omega_\tau)} \sum_{i=1}^{N_1}\sum_{\mu=1}^{N_2} \xi_i^\mu \varepsilon_i \varepsilon_\mu \tag{63}$$
$$+ [\sqrt{\Omega_\tau(1-\Omega_\sigma)} + \sqrt{\Omega_\sigma(1-\Omega_\tau)}] \sum_{i=1}^{N_1}\sum_{\mu=1}^{N_2} \xi_i^\mu g_i \varepsilon_\mu \tag{64}$$
$$+ \sqrt{\Omega_\sigma\Omega_\tau} \sum_{i=1}^{N_1}\sum_{\mu=1}^{N_2} \xi_i^\mu g_i g_\mu. \tag{65}$$

Of course, the first two terms have well-defined thermodynamical properties for all T, so we just need to rescale T_c as

$$T_c(\alpha) = \Omega_\sigma\Omega_\tau(\sqrt{\alpha} + \sqrt{1-\alpha})^2. \tag{66}$$

When $\alpha \in \{0,1\}$ we recover the result obtained at low load, where a divergence in m appears at $T_c = \Omega_\sigma\Omega_\tau$ (see Fig. 3); Eq. (66) generalizes this result to high load. Note that this critical temperature is in general lower than the one for the paramagnetic to spin-glass transition: they coincide only in the fully Gaussian case, while in all other cases the system first enters the ordered phase before encountering the singularity as T is lowered.

D. Spherical constraints

As before we can remove the singularity in the partition function by adding the spherical constraint $\delta(N - \sum_{i=1}^N \sigma_i^2)$ to the σ-prior. Equations (40) remain valid with the replacement

$$\gamma_\sigma^{-1} = \Omega_\sigma^{-1} - \beta\alpha(R - r) + \omega,$$

with $\omega \geq 0$ (or directly γ_σ, see also Sec. III) satisfying

$$Q = q + \gamma_\sigma + \gamma_\sigma^2\phi_\sigma^2[1 - \overline{t^2}(\beta(1-\alpha)\gamma_\sigma\phi_\sigma\Omega_\tau\delta m, \sqrt{v})] = 1. \tag{67}$$

For binary σ, i.e., $\Omega_\sigma \to 0$, one has $\gamma_\sigma\phi_\sigma \to 1$ and the constraint Eq. (67) is automatically satisfied. For Gaussian σ ($\Omega_\sigma = 1$), on the other hand, $\phi_\sigma = 0$, and hence $\gamma_\sigma = 1 - q$.

Starting from the low-load solution $\alpha = 0$ and increasing α, it is possible to find numerically the solution of the Eqs. (40) and the constraint Eq. (67). The results, presented in Fig. 9, indicate that the retrieval region is robust also in the high-load regime, disappearing as $\Omega_\sigma \to 1$. The retrieval transition line exhibits re-entrant behavior as in the standard Hopfield model, which might point to underlying RSB effects [36].

In principle one can ask further what happens in a model where *both* layers have a spherical constraint. In this case we simply need to put an additional Gaussian factor $e^{-\omega_\tau\tau^2/2}$ into the effective τ-spin distribution, where the additional Lagrange multiplier ω_τ can be found by fixing the radius $R = 1$. As a consequence, the paramagnetic to spin-glass transition line Eq. (24) becomes

$$\beta^2\alpha(1-\alpha)Q^2R^2 = \beta^2\alpha(1-\alpha) = 1. \tag{68}$$

This is valid for the bipartite SK model ($\Omega_\sigma = \Omega_\tau = 0$) but also for generic Ω_σ and Ω_τ. As $T_c = \sqrt{\alpha(1-\alpha)} \to 0$ for $\alpha \to 0$ and retrieval is expected only *below* the paramagnetic to spin-glass transition, this indicates that the double spherical constraint removes the possibility of a retrieval phase, even for low load. What is happening is that the high-field response Ω_τ is weakened and becomes $\gamma_\tau^0 = \Omega_\tau/(1 + \Omega_\tau\omega_\tau)$. Moreover, Eq. (40) still apply if we replace Ω_τ by γ_τ^0 and set

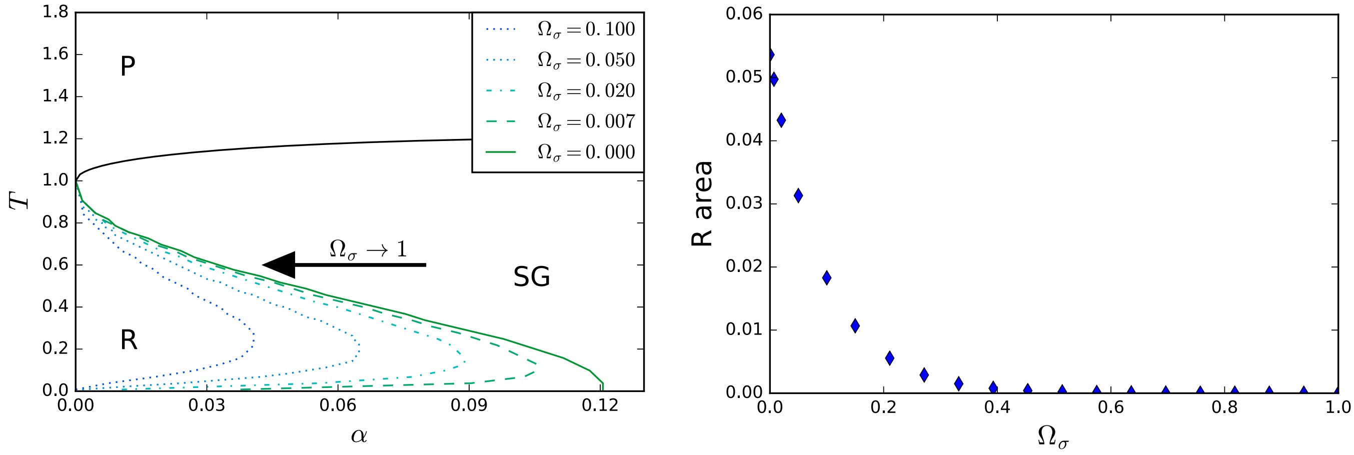

FIG. 9. Phase diagram with the paramagnetic (P), spin-glass (SG), and retrieval (R) regions of the soft model with a spherical constraint on the σ layer for different Ω_σ and fixed $\Omega_\tau = \delta = 1$. The area of the retrieval region shrinks exponentially as Ω_σ is increased from 0.

$\gamma_\tau^{-1} = \Omega_\tau^{-1} - \beta(1-\alpha)(1-q) + \omega_\tau$. In the paramagnetic regime γ_σ and γ_τ satisfy

$$Q = \gamma_\sigma + \gamma_\sigma^2 \phi_\sigma^2 = 1 \quad \to \quad \gamma_\sigma = \Omega_\sigma,$$
$$R = \gamma_\tau + \gamma_\tau^2 \phi_\tau^2 = 1 \quad \to \quad \gamma_\tau = \Omega_\tau, \tag{69}$$

while $q = 0$, giving for the response $\gamma_\tau^0 = 1/(\gamma_\tau^{-1} + \beta(1-\alpha)) = [\Omega_\tau^{-1} + \beta(1-\alpha)]^{-1}$. This is not sufficient for retrieval, not even at low load ($\alpha = 0$) where $\beta\gamma_\tau^0 = \beta\Omega_\tau/(1+\beta\Omega_\tau) < 1$ and the critical temperature is $T = 0$ ($\beta \to \infty$). Intuitively, because of the spherical cutoff the tail of the hidden units is simply not sufficient to give, after marginalizing out the visible units, an appropriate function u (see Sec. I A) to get spontaneous magnetization in the low-load ferromagnetic model.

V. CONCLUSIONS AND OUTLOOK

In this paper we have investigated the phase diagram of restricted Boltzmann machines with different unit and weight distributions, ranging from centered (real) Gaussian to Boolean variables. We highlighted the retrieval capabilities of these networks, using their duality with generalized Hopfield models.

Our analysis is mainly based on the study of the self-consistency relations for the order parameters and offers a nearly complete description of the properties of these systems. For this rather large class of models we have drawn the phase diagram, which is made up of three phases, namely paramagnetic, spin-glass, and retrieval, and studied the phase transitions between them.

We stress that, while in associative neural networks patterns are often restricted to the binary case, there is at present much research activity in the area of Boltzmann machines with real weights. Our analysis shows that retrieval is possible at high load for any pattern distribution interpolating between Boolean and Gaussian statistics. In the Gaussian case high-load retrieval fails but is recovered even here at low load.

A complete analysis of the paramagnetic–spin-glass transition and the spin-glass–retrieval transition is very useful for the study of modern deep neural networks, where the crucial learning phase is often initiated with a step of unsupervised learning using restricted Boltzmann machines [8,10]. A first attempt to link the properties of the phase diagram to the challenges of training a restricted Boltzmann machine from data and extracting statistically relevant features can be found in Ref. [48].

ACKNOWLEDGMENTS

A.B. acknowledges partial financial support by National Group of Mathematical Physics GNFM-INDAM. G.G. is supported by the NCCR SwissMAP. D.T. is supported by Scuola Normale Superiore and National Group of Mathematical Physics GNFM-INdAM.

APPENDIX A: DERIVATION OF EQUATIONS (7)–(12)

Consider a bipartite system with N_1 σ-spins and N_2 τ-spins, $N = N_1 + N_2$, $\alpha = N_2/N$ and partition function

$$Z_N(\beta,\alpha;\xi) = \mathbb{E}_{\sigma,\tau} \exp\left(\sqrt{\frac{\beta}{N}} \sum_{i=1}^{N_1} \sum_{\mu=1}^{N_2} \xi_i^\mu \sigma_i \tau_\mu\right), \tag{A1}$$

with the expectation being over generic spin distributions $P_\sigma(\sigma)$ and $P_\tau(\tau)$. We assume there are $\ell_1 = O(1)$ condensed patterns associated with the first ℓ_1 σ-variables and similarly ℓ_2 condensed patterns associated with the first ℓ_2 τ-variables, and two families of overlaps

$$m^\mu(\sigma) = \frac{1}{N_1} \sum_{i=\ell_1}^{N_1} \xi_i^\mu \sigma_i, \quad n^i(\tau) = \frac{1}{N_2} \sum_{\mu=\ell_2}^{N_2} \xi_i^\mu \tau_\mu, \tag{A2}$$

and

$$q_{\alpha\beta} = \frac{1}{N_1} \sum_{i=\ell_1}^{N_1} \sigma_i^\alpha \sigma_i^\beta \quad r_{\alpha\beta} = \frac{1}{N_2} \sum_{\mu=\ell_2}^{N_2} \tau_\mu^\alpha \tau_\mu^\beta. \tag{A3}$$

Then (all sums over non-condensed patterns start from $i = l_1 + 1$ and $\mu = l_2 + 1$ but we drop the $+1$ to save space)

$$Z_N(\beta,\alpha;\xi) = \mathbb{E}_{\sigma,\tau} \exp\left(\sqrt{\frac{\beta}{N}}\sum_{i=1}^{\ell_1}\sum_{\mu=\ell_2}^{N_2}\xi_i^\mu\sigma_i\tau_\mu + \sqrt{\frac{\beta}{N}}\sum_{\mu=1}^{\ell_2}\sum_{i=\ell_1}^{N_1}\xi_i^\mu\sigma_i\tau_\mu\right)\exp\left(\sqrt{\frac{\beta}{N}}\sum_{i=\ell_1}^{N_1}\sum_{\mu=\ell_2}^{N_2}\xi_i^\mu\sigma_i\tau_\mu + \sqrt{\frac{\beta}{N}}\sum_{i=1}^{\ell_1}\sum_{\mu=1}^{\ell_2}\xi_i^\mu\sigma_i\tau_\mu\right)$$

$$\sim \mathbb{E}_{\sigma,\tau}\exp\left(N_2\sqrt{\frac{\beta}{N}}\sum_{i=1}^{\ell_1}n^i(\tau)\sigma_i + N_1\sqrt{\frac{\beta}{N}}\sum_{\mu=1}^{\ell_2}m^\mu(\sigma)\tau_\mu + \sqrt{\frac{\beta}{N}}\sum_{i=\ell_1}^{N_1}\sum_{\mu=\ell_2}^{N_2}\xi_i^\mu\sigma_i\tau_\mu\right), \tag{A4}$$

where we have neglected the last, nonextensive, term of Eq. (A4). Constraining the values of the overlaps we get

$$Z_N = \int\{dm^\mu,d\hat{m}^\mu,dn^i,d\hat{n}^i\}\exp\left[-iN\left(\sum_{i=1}^{\ell_1}n^i\hat{n}^i + \sum_{\mu=1}^{\ell_2}m^\mu\hat{m}^\mu\right)\right]\mathbb{E}_{\sigma,\tau}\exp\left(N_2\sqrt{\frac{\beta}{N}}\sum_{i=1}^{\ell_1}n^i\sigma_i + N_1\sqrt{\frac{\beta}{N}}\sum_{\mu<\ell_2}m^\mu\tau_\mu\right)$$

$$\times\,\mathbb{E}_{\sigma,\tau}\exp\left(\frac{i}{\alpha}\sum_{i=1}^{\ell_1}\hat{n}^i\sum_{\mu=\ell_2}^{N_2}\xi_i^\mu\tau_\mu + \frac{i}{1-\alpha}\sum_{\mu=1}^{\ell_2}\hat{m}^\mu\sum_{i=\ell_1}^{N_1}\xi_i^\mu\sigma_i + \sqrt{\frac{\beta}{N}}\sum_{i=\ell_1}^{N_1}\sum_{\mu=\ell_2}^{N_2}\xi_i^\mu\sigma_i\tau_\mu\right). \tag{A5}$$

We recall the definition of $\Omega_{\sigma,\tau}$ and $u_{\sigma,\tau}$ from the Introduction: $u_{\sigma,\tau}$ is the cumulant generating function of $P_{\sigma,\tau}$, to wit $u_{\sigma,\tau}(h) = \ln\mathbb{E}_{P_{\sigma,\tau}}[e^{hx}]$ and

$$\lim_{N\to\infty}\frac{1}{N}u_{\sigma,\tau}(\sqrt{N}x) = \frac{\Omega_{\sigma,\tau}x^2}{2}. \tag{A6}$$

Then the terms in the second line of Eq. (A5) become

$$\mathbb{E}_{\sigma,\tau}\exp\left(N_2\sqrt{\frac{\beta}{N}}\sum_{i=1}^{\ell_1}n^i\sigma_i + N_1\sqrt{\frac{\beta}{N}}\sum_{\mu=1^{l_2}}m^\mu\tau_\mu\right) = \exp\left[\frac{\beta N}{2}\left(\alpha^2\Omega_\sigma\sum_{i=1}^{\ell_1}n_i^2 + (1-\alpha)^2\Omega_\tau\sum_{\mu=1}^{\ell_2}m_\mu^2\right)\right], \tag{A7}$$

while, after introducing replicas and averaging over the disorder, the last term in Eq. (A5) gives (with u_ξ the cumulant generating function associated with the patterns)

$$\mathbb{E}_\xi\exp\left(\sqrt{\frac{\beta}{N}}\sum_{\alpha=1}^{n}\sum_{i=\ell_1}^{N_1}\sum_{\mu=\ell_2}^{N_2}\xi_i^\mu\sigma_i^\alpha\tau_\mu^\alpha\right) = \exp\left[\sum_{i=\ell_1}^{N_1}\sum_{\mu=\ell_2}^{N_2}u_\xi\left(\sqrt{\frac{\beta}{N}}\sum_{\alpha=1}^{n}\sigma_i^\alpha\tau_\mu^\alpha\right)\right]$$

$$\sim \exp\left(\sum_{i=\ell_1}^{N_1}\sum_{\mu=\ell_2}^{N_2}\frac{\beta}{2N}\sum_{\alpha,\beta=1}^{n}\sigma_i^\alpha\sigma_i^\beta\tau_\mu^\alpha\tau_\mu^\beta\right).$$

(Here we have used that the patterns have unit variance, hence $u_\xi(x) = x^2 + \ldots$, and neglected corrections in $1/N$.) This term becomes $\exp[\frac{\beta N}{2}\alpha(1-\alpha)\sum_{\alpha\beta}q_{\alpha\beta}r_{\alpha\beta}]$ once it is expressed in terms of the order parameters q and r, bearing in mind that the *missing* spins $\sigma_1,\ldots,\sigma_{\ell_1}$ and $\tau_1,\ldots,\tau_{\ell_2}$ constitute a vanishing fraction of the total number. Now averaging over spin variables we get the other two terms in the last line of Eq. (A5), where we also include the contributions from constraining the q and r order parameters:

$$\mathbb{E}_\sigma\exp\left(\frac{i}{1-\alpha}\sum_{\alpha=1}^{n}\sum_{\mu=1}^{\ell_2}\hat{m}_\alpha^\mu\sum_{i=\ell_1}^{N_1}\xi_i^\mu\sigma_i^\alpha + \frac{i}{1-\alpha}\sum_{\alpha,\beta=1}^{n}\hat{q}_{\alpha\beta}\sum_{i=\ell_1}^{N_1}\sigma_i^\alpha\sigma_i^\beta\right)$$

$$= \exp\left(N(1-\alpha)\left\langle\ln\mathbb{E}_\sigma e^{\frac{i}{1-\alpha}(\sum_{\alpha=1}^{n}\sum_{\mu=1}^{\ell_2}\hat{m}_\alpha^\mu\xi^\mu\sigma^\alpha + \sum_{\alpha,\beta=1}^{n}\hat{q}_{\alpha\beta}\sigma^\alpha\sigma^\beta)}\right\rangle_\xi\right)$$

and

$$\mathbb{E}_\tau\exp\left(\frac{i}{(1-\alpha)}\sum_{\alpha=1}^{n}\sum_{i=1}^{\ell_1}\hat{n}_\alpha^i\sum_{\mu=\ell_2}^{N_2}\xi_i^\mu\tau_\mu^\alpha + \frac{i}{(1-\alpha)}\sum_{\alpha,\beta=1}^{n}\hat{r}_{\alpha\beta}\sum_{\mu=\ell_2}^{N_2}\tau_\mu^\alpha\tau_\mu^\beta\right)$$

$$= \exp\left(\alpha N\left\langle\ln\mathbb{E}_\tau e^{\frac{i}{\alpha}(\sum_{\alpha=1}^{n}\sum_{i=1}^{\ell_1}\hat{n}_\alpha^i\xi^i\tau^\alpha + \sum_{\alpha,\beta=1}^{n}\hat{r}_{\alpha\beta}\tau^\alpha\tau^\beta)}\right\rangle_\xi\right).$$

Collecting all the terms we get an expression for $\mathbb{E}[Z_N^n]$ which depends on the parameters m_α^μ, n_α^i, $q_{\alpha\beta}$, and $r_{\alpha\beta}$:

$$\mathbb{E}[Z_N^n] = \int\{dm_\mu^\alpha,d\hat{m}_\mu^\alpha\}\{dq_{\alpha\beta},d\hat{q}_{\alpha\beta}\}e^{Nf(\{m_\alpha^\mu\},\{n_\alpha^i\},\{q_{\alpha\beta}\},\{r_{\alpha\beta}\})}, \tag{A8}$$

with

$$\begin{aligned} &f\left(\{m_\alpha^\mu\},\{n_\alpha^i\},\{q_{\alpha,\beta}\},\{r_{\alpha\beta}\}\right)\\ &= -\frac{\beta}{2}\Omega_\tau(1-\alpha)^2\sum_{\mu=1}^{\ell_2} m_\alpha^{\mu\,2} - \frac{\beta}{2}\Omega_\sigma\alpha^2\sum_{i=1}^{\ell_1} {n_\alpha^i}^2 - \frac{\beta}{2}\alpha(1-\alpha)\sum_{\alpha,\beta=1} q_{\alpha\beta}r_{\alpha\beta}\\ &\quad + (1-\alpha)\left\langle \ln \mathbb{E}_\sigma e^{\beta(1-\alpha)\Omega_\tau \sum_{\alpha=1}^n (m\cdot\xi)\sigma^\alpha + \frac{\beta\alpha}{2}\sum_{\alpha,\beta=1}^n r_{\alpha\beta}\sigma^\alpha\sigma^\beta}\right\rangle_\xi + \alpha\left\langle \ln \mathbb{E}_\tau e^{\beta\alpha\Omega_\sigma \sum_{\alpha=1}^n (n\cdot\xi)\tau^\alpha + \frac{\beta(1-\alpha)}{2}\sum_{\alpha,\beta=1}^n q_{\alpha\beta}\tau^\alpha\tau^\beta}\right\rangle_\xi. \end{aligned} \tag{A9}$$

By a saddle-point calculation we obtain immediately

$$i\hat{m}_\alpha^\mu = \beta(1-\alpha)^2\Omega_\tau m_\alpha^\mu \quad i\hat{n}_\alpha^i = \beta\alpha^2\Omega_\sigma n_\alpha^i \quad i\hat{q}_{\alpha\beta} = \frac{\beta}{2}\alpha(1-\alpha)r_{\alpha\beta} \quad i\hat{r}_{\alpha\beta} = \frac{\beta}{2}\alpha(1-\alpha)q_{\alpha\beta}, \tag{A10}$$

and in the RS ansatz, assuming that

$$m_\alpha^\mu = m^\mu \quad n_\alpha^i = n^i \quad q_{ab} = Q\delta_{\alpha\beta} + q(1-\delta_{\alpha,\beta}) \quad r_{ab} = R\delta_{\alpha\beta} + r(1-\delta_{\alpha,\beta}), \tag{A11}$$

taking the limit $n \to 0$ and extremizing Eq. (A9) we get the saddle-point Eqs. (7)– (12)

APPENDIX B: GAUSSIAN BIPARTITE AND SPHERICAL HOPFIELD MODEL

The bipartite system with Gaussian priors on both layers ($\Omega_\sigma = \Omega_\tau = 1$) can be related to a spherical Hopfield model [39–41] via Legendre duality as in Ref. [37]. In fact, integrating over the radius $r\sqrt{N}$ we have

$$\begin{aligned} Z_g(\beta) &= \int dr \frac{e^{-Nr^2/2}}{\sqrt{2\pi}^N}\int d\Sigma_{r\sqrt{N}}(\sigma)e^{-\beta\mathcal{H}(\sigma)} = \int dr \frac{e^{-Nr^2/2}}{\sqrt{2\pi}^N} Z_s^{r\sqrt{N}}(\beta)\\ &= \int dr \frac{e^{-Nr^2/2}}{\sqrt{2\pi}^N} r^{N-1}\int d\Sigma_{\sqrt{N}}(\sigma)e^{-\beta r^2\mathcal{H}(\sigma)} = \int dr \frac{e^{-Nr^2/2}}{\sqrt{2\pi}^N} r^{N-1} Z_s^{\sqrt{N}}(\beta r^2), \end{aligned} \tag{B1}$$

where $d\Sigma_r(\sigma)$ is the uniform measure over the sphere of radius r and (Z_g, Z_s) are, respectively, the partition functions of the Gaussian and spherical models. Thus, the two free energies, f_g and f_s, are related by

$$-\beta f_g = \sup_r \left[-\tfrac{1}{2}r^2 - \tfrac{1}{2}\ln(2\pi) + \ln(r) - \beta f_s(\beta r^2)\right], \tag{B2}$$

and so the Gaussian free energy comes from the spherical free energy calculated at the optimal radius given by

$$r^2 = \frac{1}{1 - 2\beta\partial_\beta(-\beta f_s(\beta))|_{\beta r^2}}. \tag{B3}$$

Since $r^2 = Q$, the self overlap of the σ-spins (first layer), and using the expression for the spherical free energy from [39,41] we have, in the high-temperature region,

$$Q = \frac{1}{1 - \beta\alpha\frac{1}{1-\beta(1-\alpha)Q}} = \frac{1}{1-\beta\alpha R(Q)} \quad \text{and} \quad R(Q) = \frac{1}{1-\beta(1-\alpha)Q}. \tag{B4}$$

These are exactly Eqs. (27) and (28) with $\Omega_\sigma = \Omega_\tau = 1$. Moreover, again from Refs. [39,41] the critical line for the spherical model is given by $[1-\beta(1-\alpha)]^2 = \beta^2\alpha(1-\alpha)$. Thus, we obtain the critical line for the Gaussian model Eq. (24) by replacing $\beta \to \beta Q$:

$$\frac{\beta^2\alpha(1-\alpha)Q^2}{[1-\beta(1-\alpha)Q]^2} = 1 = \beta^2\alpha(1-\alpha)Q^2R^2. \tag{B5}$$

[1] W. S. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. **5**, 115 (1943).

[2] F. Rosenblatt, The Perceptron: A probabilistic model for information storage and organization in the brain, Psych. Rev. **65**, 386 (1958).

[3] D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).

[4] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).

[5] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*, Santa Fe Institute Studies in the Sciences of Complexity; Lecture Notes, Redwood City, CA (Addison-Wesley, Boston, 1991).

[6] A. C. C. Coolen, R. Kühn, and P. Sollich, *Theory of Neural Information Processing* (Oxford University Press, Oxford, 2005).

[7] A. Engel and C. Van den Broeck, *Statistical Mechanics of Learning* (Cambridge Press, Cambridge, 2001).

[8] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Google book (2016).

[9] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).

[10] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature **521**, 436 (2015).

[11] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Spin-glass model of neural networks, Phys. Rev. A **32**, 1007 (1985).

[12] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Storing Infinite Numbers of Patterns in a Spin Glass Model of Neural Networks, Phys. Rev. Lett. **55**, 1530 (1985).

[13] H. S. Seung, H. Sompolinsky, and N. Tishby, Statistical mechanics of learning from examples, Phys. Rev. A **45**, 6056 (1992).

[14] A. Barra, A. Bernacchia, E. Santucci, and P. Contucci, On the equivalence among Hopfield neural networks and restricted Boltzman machines, Neural Networks **34**, 1 (2012).

[15] A. Barra, G. Genovese, F. Guerra, and D. Tantari, How glassy are neural networks? J. Stat. Mech. (2012) P07009.

[16] M. Mezard, Mean-field message-passing equations in the Hopfield model and its generalizations, Phys. Rev. E **95**, 022117 (2017).

[17] H. Huang, Statistical mechanics of unsupervised feature learning in a restricted Boltzmann machine with binary synapses, J. Stat. Mech. (2017) 053302.

[18] M. Gabrié, E. W. Tramel, and F. Krzakala, Training restricted Boltzmann machine via the Thouless-Anderson-Palmer free energy, in *Advances in Neural Information Processing Systems* (Neural Information Processing Systems, 2015), pp. 640–648.

[19] M. Welling and C. A. Sutton, Learning in Markov random fields with contrastive free energies, *Proceedings of the 10th International Workshop on AI and Statistics* (AISTATS'05, 2005), pp. 397-404.

[20] D. C. Mattis, Solvable spin system with random interactions, Phys. Lett. A **56**, 421 (1976).

[21] E. Gardner, The space of interactions in neural network models, J. Phys. A **21**, 257 (1988).

[22] E. Gardner and B. Derrida, Optimal storage properties of neural network models, J. Phys. A **21**, 271 (1988).

[23] M. Talagrand, *Mean Field Models for Spin Glasses*, Vol. 1, 2 (Springer-Verlag, Berlin/Heidelberg, 2011).

[24] G. Genovese and D. Tantari, Non-Convex Multipartite Ferromagnets, J. Stat. Phys. **163**, 492 (2016).

[25] E. W. Tramel, A. Manoel, F. Caltagirone, M. Gabrié, and F. Krzakala, Inferring sparsity: Compressed sensing using generalized restricted Boltzmann machines, *Information Theory Workshop (ITW)* (IEEE, Cambridge, UK, 2016), pp. 265–269.

[26] A. Barra and F. Guerra, About the ergodic regime in the analogical Hopfield neural networks: Moments of the partition function, J. Math. Phys. **49**, 125217 (2008).

[27] A. Barra, G. Genovese, and F. Guerra, The replica symmetric approximation of the analogical neural network, J. Stat. Phys. **140**, 784 (2010).

[28] A. Bovier, A. C. D. van Enter, and B. Niederhauser, Stochastic symmetry-breaking in a Gaussian Hopfield model, J. Stat. Phys. **95**, 181 (1999).

[29] A. Barra, G. Genovese, and F. Guerra, Equilibrium statistical mechanics of bipartite spin systems, J. Phys. A: Math. Theor. **44**, 245002 (2011).

[30] A. Barra, P. Contucci, E. Mingione, and D. Tantari, Multi-Species Mean Field Spin Glasses. Rigorous Results, Annales Henri Poincaré **16**, 691 (2015).

[31] G. Genovese, Universality in Bipartite Mean Field Spin Glasses, J. Math. Phys. **53**, 123304 (2012).

[32] P. Sollich, D. Tantari, A. Annibale, and A. Barra, Extensive Parallel Processing on Scale free Networks, Phys. Rev. Lett. **113**, 238106 (2014).

[33] E. Agliari, A. Annibale, A. Barra, A. C. C. Coolen, and D. Tantari, Immune networks: Multitasking capabilities near saturation, J. Phys. A: Math. Theor **46**, 415003 (2013).

[34] E. Agliari, A. Annibale, A. Barra, A. C. C. Coolen, and D. Tantari, Immune networks: multi-tasking capabilities at medium load, J. Phys. A: Math. Theor. **46**, 335101 (2013).

[35] A. Bovier, *Statistical Mechanics of Disordered System. A Mathematical Perspective* (Cambridge University Press, Cambridge, 2006).

[36] J.-P. Naef and A. Canning, Reentrant spin glass behaviour in the replica symmetric solution of the Hopfield neural network model, J. Phys. I **2**, 247 (1992).

[37] G. Genovese and D. Tantari, Legendre duality of spherical and gaussian spin glasses, Math. Phys. Anal. Geom. **18**, 1 (2015).

[38] J. J. Hopfield, Neurons with graded response have collective computational properties like those of two-state neurons, Proc. Natl. Acad. Sci. USA **81**, 3088 (1984).

[39] D. Bollé, T. M. Nieuwenhuizen, and I. P. Castillo, A spherical Hopfield model, J. Phys. A **36**, 10269 (2003).

[40] M. Shcherbina and B. Tirozzi, Rigorous Solution of the Gardner Problem, Commun. Math. Phys. **234**, 383 (2003).

[41] J. Baik and J. O. Lee, Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model, J. Stat. Phys. **165**, 185 (2016).

[42] G. B. Arous, A. Dembo, and A. Guionnet, Aging of spherical spin glasses, Probab. Theory Relat. Fields **120**, 1 (2001).

[43] A. Barra, G. Genovese, F. Guerra, and D. Tantari, About a solvable mean field model of a Gaussian spin glass, J. Phys. A: Math. Theor. **47**, 155002 (2014).

[44] A. Auffinger and W.-K. Chen, Free energy and complexity of spherical bipartite models, J. Stat. Phys. **157**, 40 (2014).

[45] A. Crisanti, D. J. Amit, and H. Gutfreund, Saturation level of the hopfield model for neural network, Europhys. Lett. **2**, 337 (1986).

[46] D. Panchenko, The free energy in a multi-species sherrington-kirkpatrick model, Ann. Prob. **43**, 3494 (2015).

[47] E. Agliari, A. Barra, C. Longo, and D. Tantari, Neural Networks retrieving binary patterns in a sea of real ones, J. Stat. Phys. **168**, 1085 (2017).

[48] A. Barra, G. Genovese, P. Sollich, and D. Tantari, Phase transitions in restricted Boltzmann machines with generic priors, Phys. Rev. E **96**, 042156 (2017).